Relativistic corrections to energy spectrum of hydrogen due to the full one-photon-exchange interaction

Zi-Wen Zhang and Hai-Qing Zhou*

School of Physics, Southeast University, NanJing 211189, China

(Dated: March 5, 2024)

In this work, we estimate the relativistic corrections to the energy spectrum of hydrogen resulting from the full one-photon-exchange interaction using a highly precise numerical method. In the frame of the effective Schrodinger-like equation, which is derived exactly from the Bethe-Salpeter equation in quantum electrodynamics, we express the effective potential corresponding to the full one-photon-exchange interaction in terms of eight scalar functions. Unlike the usual calculations performed in the references, where the effective potential is expanded in terms of momenta order by order, we retain the exact momentum dependence in the effective potential to estimate its corrections to the energy spectrum of hydrogen using a highly precise numerical method. We also use the same numerical method to double-check the results in the non-relativistic case to ensure accuracy. We discuss the comparison of the numerical results with those obtained using the usual bound-state perturbative theory. Our calculations suggest that it is possible to accurately account for all the relativistic contributions using this method. It would be interesting to extent these calculations to positronium, muonic hydrogen, and cases involving nuclear structure and radiative corrections.

arXiv:2403.02003v1 [nucl-th] 4 Mar 2024

I. INTRODUCTION

The study of the energy spectrum of hydrogen-like atoms has been pivotal in the development of quantum mechanics and has remained important for over a century. However, dealing with bound states in a pure quantum field theory, especially when the non-relativistic expansion is not valid, can still be challenging. In the past fifteen years, there have been significant improvements in the precise experimental measurements of

* E-mail: zhouhq@seu.edu.cn

the Lamb shifts of hydrogen and muonic hydrogen, presenting many challenges [1–8]. Theoretically, the bound-state perturbative theory is commonly used to estimate energy corrections beyond the Coulomb potential (see recent reviews and books [9–11] and the references therein). To reliably estimate these corrections, the effective Schrodinger-like equation [12] or the effective Dirac-like equations [13], which are derived exactly from the Bethe-Salpeter (BS) equation [14, 15] in quantum electrodynamics (QED) or non-relativistic quantum electrodynamics (NRQED) [16], should be employed, as there is no analytical solution for the physical BS equation.

Throughout history, various other methods have also been employed to investigate corrections beyond the Coulomb potential. These include the use of the Dirac equation with an effective potential [17], the external field approximation [9], the quasi-potential approach [18–20], and the Foldy-Wouthuysen transformation method [21–23], among others. However, in this work, we focus our discussion within the framework of the effective Schrodinger-like equation, which is derived exactly from the Bethe-Salpeter equation in QED [12].

In the bound-state perturbative theory, the interaction kernel is expanded order by order in terms of the coupling constant α_e, velocities $\vec{p}_i/m_{e,p}$, and m_e/m_p, where $\vec{p}_i$ represents the three momenta of the system, and m_e and m_p denote the masses of the electron and proton, respectively. The expansion of the interaction kernel in terms of momenta implies that only contributions from low momenta are accurately considered, while contributions from high momenta need to be treated separately.

On the other hand, the wave functions in the Schrodinger equation with the Coulomb potential are valid for the entire momentum range. The integration of the wave functions and interaction kernels encompasses contributions from both low and high momenta. This can introduce additional ultraviolet (UV) divergences, which differ somewhat from the usual UV divergences in the scattering amplitudes arising from loop integrals in radiative corrections. To handle these UV divergences, regularization methods are employed [24–26]. These UV divergences eventually cancel each other out, and the final results are independent of the regularization procedure.

In this study, we use the one-photon-exchange (OPE) interaction as an example to perform a numerical calculation of the energy spectrum. Instead of expanding the OPE interaction in terms of momenta and m_e/m_p, we retain it in its full form. This approach

allows us to include all relativistic corrections at the order of α_e, and it avoids any additional UV divergences since the full OPE interaction kernel remains valid across the entire momentum range. This calculation can be regarded as a full result incorporating the OPE interaction, and it provides an interesting basis for comparison with the results obtained using the bound-state perturbative theory.

The paper is organized as follows: In Section II, we introduce the framework used in our calculations. In Section III, we present the numerical results for specific states, showcasing their properties. We also provide a brief discussion on the comparison of these numerical results with those obtained using the bound-state perturbative approach.

II. BASIC FORMULA

A. The effective Schrodinger-like equation

In QED, the bound states of the ep system in the center frame can be accurately described by the following effective Schrodinger-like equation [12]:

$$\left[P_0 - E_1^{(e)} - E_1^{(p)}\right] \phi_{\bar{\lambda},\bar{\mu}}(\vec{p}_1) = \int \frac{d^3\vec{p}_3}{(2\pi)^3} i\widetilde{K}_{\bar{\lambda}\lambda,\bar{\mu}\mu}(\vec{p}_1,\vec{p}_3,P)\phi_{\lambda,\mu}(\vec{p}_3), \tag{1}$$

where $P = (P_0, \vec{0}) \equiv p_1 + p_2 = p_3 + p_4$, $p_{1,2,3,4}$ are the momenta of final and intermediate fermions as shown in Fig.1, $E_i^{(e)} \equiv \sqrt{\vec{p}_i^2 + m_e^2}$, $E_i^{(p)} \equiv \sqrt{\vec{p}_i^2 + m_p^2}$. The interaction kernel $\widetilde{K}_{\bar{\lambda}\lambda,\bar{\mu}\mu}(\vec{p}_1,\vec{p}_3,P)$ is defined as

$$\widetilde{K}_{\bar{\lambda}\lambda,\bar{\mu}\mu}(\vec{p}_1,\vec{p}_3,P) \equiv \frac{\bar{u}_{\bar{\alpha}}(\vec{p}_1,m_e,\bar{\lambda})\bar{u}_{\bar{\beta}}(-\vec{p}_1,m_p,\bar{\mu})}{\sqrt{4E_1^{(e)}E_1^{(p)}}} \bar{K}_{\bar{\alpha}\alpha,\bar{\beta}\beta}(\vec{p}_1,\vec{p}_3,P) \frac{u_\alpha(\vec{p}_3,m_e,\lambda)u_\beta(-\vec{p}_3,m_p,\mu)}{\sqrt{4E_3^{(e)}E_3^{(p)}}}, \tag{2}$$

where

$$\begin{aligned} \bar{K}(\vec{p}_1,\vec{p}_3,P) &\equiv \bar{K}(p_1,p_3,P)\Big|_{p_1^0=p_3^0=\tau_e P_0}, \\ \bar{K} &\equiv [I - K_{\rm BS}(G_0 - \bar{G}_0)]^{-1} K_{\rm BS}, \end{aligned} \tag{3}$$

with $\tau_e \equiv m_e/(m_e+m_p)$ and $K_{\rm BS}$ the usual two-body BS irreducible kernel. Here, we directly make the assumption that the inverse $[I - K_{\rm BS}(G_0 - \bar{G}_0)]^{-1}$ exists, and that $\bar{K}$ can be expressed in a perturbative form as follows:

$$\bar{K}(p_1,p_3,P) = K_{\rm BS}(p_1,p_3,P) + \int \frac{{\rm d}^4k}{(2\pi)^4} K_{\rm BS}(p_1,k,P)[G_0(k,P) - \bar{G}_0(k,P)]K_{\rm BS}(k,p_3,P) + \cdots \tag{4}$$

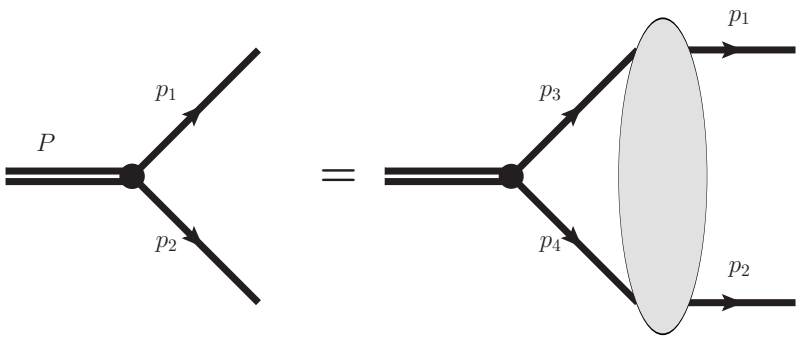


FIG. 1: Diagram for the BS equation.

where

$$
\begin{aligned}
G_0(k,P) &\equiv S^{(e)}_{full}(k,m_e)S^{(p)}_{full}(P-k,m_p),\\
\bar{G}_0(k,P) &\equiv 2\pi i\delta(k^0)\Big[\frac{\Lambda^{(e)}_+(\vec{k})\Lambda^{(p)}_+(-\vec{k})}{P_0-E_e(\vec{k})-E_p(\vec{k})}-\frac{\Lambda^{(e)}_-(\vec{k})\Lambda^{(p)}_-(-\vec{k})}{P_0+E_e(\vec{k})+E_p(\vec{k})}\Big]\\
&\approx 2\pi i\delta(k^0)\frac{\Lambda^{(e)}_+(\vec{k})\Lambda^{(p)}_+(-\vec{k})}{P_0-E_e(\vec{k})-E_p(\vec{k})}.
\end{aligned}
\tag{5}
$$

In the above expressions, $S^{(e,p)}_{full}$ are the full propagators of the electron and proton [10], respectively, and $\Lambda^{(a)}_\pm(\vec{k})$ are

$$
\Lambda^{(a)}_\pm(\vec{k}) \equiv \Big[E^{(a)}(\vec{k})\gamma^0 \mp (\vec{k}\cdot\vec{\gamma}-m_a)\Big]^{(a)}/2E^{(a)}(\vec{k}),
\tag{6}
$$

with $E^{(a)}(\vec{k}) \equiv \sqrt{\vec{k}^2+m_a^2}$.

In the leading order of α_e, $\vec{p}_i/m_{e,p}$, and m_e/m_p, the interaction kernel $\widetilde{K}$ can be chosen as

$$
i\widetilde{K}^{(0)}_{\bar{\lambda}\lambda,\bar{\mu}\mu}(\vec{p}_1,\vec{p}_3,P) \equiv \delta_{\bar{\lambda}\lambda}\delta_{\bar{\mu}\mu}\frac{-e^2}{|\vec{p}_1-\vec{p}_3|^2}\frac{1}{[N(\vec{p}_1)N(\vec{p}_3)]^{1/2}},
\tag{7}
$$

where

$$
N(\vec{p}_i) \equiv \frac{[P_0+E^{(e)}_i+E^{(p)}_i][P_0^2-(E^{(e)}_i+E^{(p)}_i)^2]}{2P_0\,[P_0^2-(m_e-m_p)^2]}.
\tag{8}
$$

By utilizing this effective interaction kernel and expanding the equation in terms of momenta and the bound energy, the effective Schrodinger-like equation Eq. (1) simplifies to the Schrodinger equation with the Coulomb potential. The contributions from interaction kernels involving higher orders of α_e, $\vec{p}_i/m_{e,p}$, and m_e/m_p beyond $\widetilde{K}^{(0)}(\vec{p}_1,\vec{p}_3,P)$ can be estimated systematically using the bound-state perturbative theory, order by order. This approach can also be applied in NRQED, where the power counting and analytical calculations are more straightforward.

B. The full interaction kernel in the leading order of α_e

In the bound-state perturbative theory, the interaction kernels are expanded order by order in terms of $\vec{p}_i/m_{e,p}$, and the corresponding matrix elements often exhibit divergences. To handle these divergences, regular methods are employed during the intermediate calculations. It is important to note that these divergences arise naturally due to the slow decrease of the wave functions of the bound states with respect to momenta $\vec{p}_i$, whereas the expansion of the interaction kernels in terms of momenta is not valid in the high momentum region.

In this study, we adopt a different approach by not expanding the interaction kernel $\widetilde{K}(\vec{p}_1, \vec{p}_3, P)$ in terms of $\vec{p}_i/m_{e,p}$ and m_e/m_p. Instead, we employ a numerical method to directly solve Eq. (1) using the leading-order α_e expression of $\widetilde{K}(\vec{p}_1, \vec{p}_3, P)$.

In the leading order of α_e, only the OPE interaction contributes. As an approximation, we consider the proton to be a point-like particle in this study. Therefore, we have:

$$K^{(\mathrm{OPE})}_{\mathrm{BS},\bar{\alpha}\alpha,\bar{\beta}\beta}(p_1, p_3, P) = (-ie\gamma^{\mu}_{\bar{\alpha}\alpha}) D_{\mu\nu}(p_1 - p_3)(+ie\gamma^{\nu}_{\bar{\beta}\beta}), \tag{9}$$

and

$$\bar{K}^{(\mathrm{LO})}(\vec{p}_1, \vec{p}_3, P) = K^{(\mathrm{OPE})}_{\mathrm{BS}}(p_1, p_3, P)\Big|_{p_1^0 = p_3^0 = \tau_e P_0}. \tag{10}$$

Taking the Coulomb gauge, we have

$$\bar{K}^{(\mathrm{LO})}_{\bar{\alpha}\alpha,\bar{\beta}\beta}(\vec{p}_1, \vec{p}_3, P) = (-ie\gamma^{0}_{\bar{\alpha}\alpha}) \frac{i}{|\vec{q}|^2} (+ie\gamma^{0}_{\bar{\beta}\beta}) + (-ie\gamma^{i}_{\bar{\alpha}\alpha}) [\frac{iq^i q^j}{|\vec{q}|^4} - \frac{i\delta^{ij}}{|\vec{q}|^2}] (+ie\gamma^{j}_{\bar{\beta}\beta}), \tag{11}$$

with $q = p_1 - p_3$. Here, $q^0 = 0$ due to the special choice of the momenta with $p_1^0 = p_3^0 = \tau_e P_0$.

After performing the necessary calculations, we obtain the following result:

$$\begin{aligned} i\widetilde{K}_{\bar{\lambda}\lambda,\bar{\mu}\mu}(\vec{p}_1, \vec{p}_3, P) &\equiv \chi^{(e)\dagger}(\bar{\lambda}) \chi^{(p)\dagger}(\bar{\mu}) V(\vec{p}_1, \vec{p}_3) \chi^{(p)}(\mu) \chi^{(e)}(\lambda) \\ &= \sum_{i=1}^{8} \frac{e^2 C_i [\chi^{(e)\dagger}(\bar{\lambda}) \chi^{(p)\dagger}(\bar{\mu}) T_i \chi^{(p)}(\mu) \chi^{(e)}(\lambda)]}{4 z_1^2 \sqrt{E_1^{(e)} E_3^{(e)} E_1^{(p)} E_3^{(p)} X_1^{(e)} X_3^{(e)} X_1^{(p)} X_3^{(p)}}}, \end{aligned} \tag{12}$$

where χ is the Pauli spinor and

$$\begin{aligned}
T_1 &= 1, \\
T_2 &= i(\vec{p}_1 \times \vec{p}_3) \cdot \sigma^{(e)}, \\
T_3 &= i(\vec{p}_1 \times \vec{p}_3) \cdot \sigma^{(p)}, \\
T_4 &= \vec{p}_1 \cdot \sigma^{(e)} \vec{p}_1 \cdot \sigma^{(p)}, \\
T_5 &= \vec{p}_1 \cdot \sigma^{(e)} \vec{p}_3 \cdot \sigma^{(p)}, \\
T_6 &= \vec{p}_3 \cdot \sigma^{(e)} \vec{p}_1 \cdot \sigma^{(p)}, \\
T_7 &= \vec{p}_3 \cdot \sigma^{(e)} \vec{p}_3 \cdot \sigma^{(p)}, \\
T_8 &= \sigma^{(e)} \cdot \sigma^{(p)},
\end{aligned} \tag{13}$$

and

$$\begin{aligned}
C_1 &= (z_2 - z_1)(\vec{p}_1 \cdot \vec{p}_3)^2 - z_1 z_3 \vec{p}_1 \cdot \vec{p}_3 - z_2 |\vec{p}_1|^2 |\vec{p}_3|^2 - z_1 X_1^{(e)} X_3^{(e)} X_1^{(p)} X_3^{(p)}, \\
C_2 &= -(E_1^{(e)} + E_3^{(e)})(X_1^{(p)} + X_3^{(p)}) \vec{p}_1 \cdot \vec{p}_3 + z_1 \vec{p}_1 \cdot \vec{p}_3 + (m_e z_1 + t_e)(X_1^{(p)} + X_3^{(p)}) + z_1 X_1^{(p)} X_3^{(p)}, \\
C_3 &= -(E_1^{(p)} + E_3^{(p)})(X_1^{(e)} + X_3^{(e)}) \vec{p}_1 \cdot \vec{p}_3 + z_1 \vec{p}_1 \cdot \vec{p}_3 + (m_p z_1 + t_p)(X_1^{(e)} + X_3^{(e)}) + z_1 X_1^{(e)} X_3^{(e)}, \\
C_4 &= 2E_3^{(e)} E_3^{(p)} \vec{p}_1 \cdot \vec{p}_3 + E_1^{(e)}(E_1^{(p)} - E_3^{(p)}) |\vec{p}_3|^2 - E_3^{(e)}(m_p z_1 + t_p) - m_e X_3^{(p)} z_1, \\
C_5 &= -\vec{p}_1 \cdot \vec{p}_3 [E_1^{(e)}(E_1^{(p)} + E_3^{(p)}) + E_3^{(e)}(E_3^{(p)} - E_1^{(p)})] + E_1^{(e)} E_3^{(p)} z_4 + E_1^{(e)} m_p z_1 + m_e X_3^{(p)} z_1, \\
C_6 &= -\vec{p}_1 \cdot \vec{p}_3 [E_1^{(e)}(E_1^{(p)} - E_3^{(p)}) + E_3^{(e)}(E_3^{(p)} + E_1^{(p)})] + E_3^{(e)} E_1^{(p)} z_4 + E_3^{(e)} m_p z_1 + m_e X_1^{(p)} z_1, \\
C_7 &= 2E_1^{(e)} E_1^{(p)} \vec{p}_1 \cdot \vec{p}_3 + E_3^{(e)}(E_3^{(p)} - E_1^{(p)}) |\vec{p}_1|^2 - E_1^{(e)}(m_p z_1 + t_p) - m_e X_1^{(p)} z_1, \\
C_8 &= (-\vec{p}_1 \cdot \vec{p}_3 (E_1^{(e)} + E_3^{(e)}) + m_e z_1 + t_e)(-\vec{p}_1 \cdot \vec{p}_3 (E_1^{(p)} + E_3^{(p)}) + m_p z_1 + t_p),
\end{aligned} \tag{14}$$

with

$$\begin{aligned}
X_i^{(a)} &\equiv E_i^{(a)} + m_a, \\
z_1 &\equiv |\vec{p}_1|^2 + |\vec{p}_3|^2 - 2\vec{p}_1 \cdot \vec{p}_3, \\
z_2 &\equiv (X_1^{(e)} + X_3^{(e)})(X_1^{(p)} + X_3^{(p)}), \\
z_3 &\equiv X_1^{(e)} X_3^{(e)} + X_1^{(p)} X_3^{(p)}, \\
z_4 &\equiv |\vec{p}_1|^2 + |\vec{p}_3|^2, \\
t_e &\equiv E_1^{(e)} |\vec{p}_3|^2 + E_3^{(e)} |\vec{p}_1|^2, \\
t_p &\equiv E_1^{(p)} |\vec{p}_3|^2 + E_3^{(p)} |\vec{p}_1|^2.
\end{aligned} \tag{15}$$

By expanding the potential V in terms of $\vec{p}_i/m_{e,p}$, it can be verified that the leading-order term corresponds to the Coulomb potential, while the next-leading-order terms

correspond to the Breit potential. The higher-order terms correspond to the effective potential beyond the Breit potential. Similarly, the expansion results also correspond to the amplitude in NRQED at the leading order of α_e, with higher orders of $\vec{p}_i/m_{e,p}$.

C. Energy correction in the bound-state perturbative theory

Expanding $H_p \equiv V + (E_1^{(e)} + E_1^{(p)} - m_e - m_p)(2\pi)^3\delta^3(\vec{p}_1 - \vec{p}_3)$ in terms of momenta yields the following results:

$$H_p = H_p^{(2)} + H_p^{(4)} + ..., \tag{16}$$

where

$$\begin{aligned} H_p^{(2)} &= [\frac{\vec{p}_1^2}{2m_e} + \frac{\vec{p}_1^2}{2m_p}](2\pi)^3\delta^3(\vec{p}_1 - \vec{p}_3) - \frac{e^2}{\vec{q}^2}, \\ H_p^{(4)} &= H_{p,0}^{(4)} + H_{p,1}^{(4)} + H_{p,2}^{(4)}, \end{aligned} \tag{17}$$

with

$$\begin{aligned} H_{p,0}^{(4)}(\vec{p}_1, \vec{q}) =& -\Big[\frac{\vec{p}_1^4}{8m_e^3} + \frac{\vec{p}_1^4}{8m_p^3}\Big](2\pi)^3\delta^3(\vec{p}_1 - \vec{p}_3) + \frac{e^2}{8}\Big[\frac{1}{m_e^2} + \frac{1}{m_p^2}\Big] + \frac{e^2}{m_e m_p}\Big[\frac{(\vec{q}\cdot\vec{p}_1)^2}{\vec{q}^4} - \frac{\vec{p}_1^2}{\vec{q}^2}\Big], \\ H_{p,1}^{(4)}(\vec{p}_1, \vec{q}) =& \Big[\frac{e^2}{4m_e^2} + \frac{e^2}{2m_e m_p}\Big]\frac{1}{\vec{q}^2} i\sigma^{(e)}\cdot(\vec{q}\times\vec{p}_1), \\ H_{p,2}^{(4)}(\vec{p}_1, \vec{q}) =& -\frac{e^2}{4m_e m_p}\Big[\frac{\sigma^{(e)}\cdot\vec{q}\sigma^{(p)}\cdot\vec{q}}{\vec{q}^2} - \sigma^{(e)}\cdot\sigma^{(p)}\Big] + \Big[\frac{e^2}{4m_p^2} + \frac{e^2}{2m_e m_p}\Big]\frac{i\sigma^{(p)}\cdot(\vec{q}\times\vec{p}_1)}{\vec{q}^2}. \end{aligned} \tag{18}$$

Here, $H_p^{(4)}$ corresponds to the Breit potential in momentum space.

In the perturbative bound-state theory, the energy contribution of the hydrogen due to the OPE interaction can be written as:

$$E_n = E_n^{(2)} + E_n^{(4)} + E_n^{(6)} + ... \tag{19}$$

where

$$\begin{aligned} E_n^{(2)} &= \langle n,l,j,F|H^{(2)}|n,l,j,F\rangle = -\frac{\alpha_e^2\mu}{2n^2}, \\ E_n^{(4)} &= \langle n,l,j,F|H^{(4)}|n,l,j,F\rangle, \\ E_n^{(6)} &= \langle n,l,j,F|H^{(4)}Q(E_n^{(2)} - H^{(2)})^{-1}QH^{(4)}|n,l,j,F\rangle + \langle n,l,j,F|H^{(6)}|n,l,j,F\rangle. \end{aligned} \tag{20}$$

Here, $\alpha_e \equiv \frac{e^2}{4\pi}, \mu = \frac{m_e m_p}{m_e + m_p}$ and Q is a projection operator on a subspace orthogonal to $|n,l,j,F\rangle$.

The calculation of $E^{(4)}_{n,i}$ can be performed directly, and the analytic results are expressed as:

$$E^{(4)}_n = E^{(4)}_{n,0} + E^{(4)}_{n,1} + E^{(4)}_{n,2}, \tag{21}$$

where

$$\begin{aligned}
E^{(4)}_{n,0} &= \frac{\alpha_e^4 \mu}{2n^3}\Big[1 - \frac{(2-6n)\mu}{n(m_e+m_p)}\Big], \\
E^{(4)}_{n,1} &= \frac{2\alpha_e^4 \mu m_p(2m_e+m_p)}{n^3(m_e+m_p)^2}\frac{j-l}{(2l+1)(2j+1)}(1-\delta_{0l}), \\
E^{(4)}_{n,2} &= \frac{\alpha_e^4 \mu^2}{n^3(m_e+m_p)}\frac{1}{(2l+1)(2F+1)}\begin{cases} \frac{4}{3}[F^2+F-\frac{3}{2}]\delta_{0l}, & l=0 \\ -\frac{1}{l}, & l\neq 0, F=l+1 \\ -\frac{1}{(l+1)}, & l\neq 0, F=l-1. \end{cases}
\end{aligned} \tag{22}$$

The calculation of $E^{(6)}_n$ is somewhat complex [27–30], and there is a UV divergence in the middle matrix elements for the S wave, as explained above. In this work, we will not compare these results.

D. The form of the wave function

The wave function in Eq. (1) can be written as:

$$\phi_{\bar{\lambda},\bar{\mu}}(\vec{p}_1) = \sum_{\bar{\alpha},\bar{\beta}} \chi^{(e)\dagger}_{\bar{\alpha}}(\bar{\lambda})\chi^{(p)\dagger}_{\bar{\beta}}(\bar{\mu})\Phi_{\bar{\alpha},\bar{\beta}}(\vec{p}_1), \tag{23}$$

which results in the following equation:

$$\Big[P_0 - E^{(e)}_1 - E^{(p)}_1\Big]\,\Phi_{\bar{\alpha},\bar{\beta}}(\vec{p}_1) = \int \frac{d^3\vec{p}_3}{(2\pi)^3} V_{\bar{\alpha}\alpha,\bar{\beta}\beta}(\vec{p}_1,\vec{p}_3)\Phi_{\alpha,\beta}(\vec{p}_3). \tag{24}$$

To compare with the results in the references, we label the state of the system as nL^F_j. For $F = j \pm \frac{1}{2} = l \pm 1$, we have the following form for $\Phi_{\alpha,\beta}(\vec{p}_3)$:

$$\Phi_{\alpha,\beta}(\vec{p}_3) = \sum_{s^e_z,s^p_z} \chi_\alpha(s^p_z)\chi_\beta(s^e_z)\langle jj_z|lm, ss^e_z\rangle\langle FF_z|jj_z, ss^p_z\rangle\Phi_{nlm}(\vec{p}_3). \tag{25}$$

When the effective potential V does not include spin-dependent terms, the above form of the wave function gives the same result as the following form:

$$\Phi_{\alpha,\beta}(\vec{p}_3) \to \Phi_{nlm}(\vec{p}_3). \tag{26}$$

We would also like to mention that the above form of the wave functions corresponds to the quantum numbers in non-relativistic quantum mechanics. When discussing states in QED, one can consider J^P as good quantum numbers to determine the most general form of the wave function $\Phi_{\alpha,\beta}$ for solving the equation. In this case, effects such as the mixing of S wave and D wave naturally appear. However, in this work, we limit our discussion to the relativistic effects when considering the above form of the wave functions.

E. Numerical method

To calculate the energy contributions beyond the bound-state perturbative theory, we expand $\Phi_{nlm}(\vec{p})$ as:

$$\Phi_{nlm}(\vec{p}) \approx \sum_{i=l+1}^{n_{max}} c_{ni}\phi^{(0)}_{ilm}(\vec{p}) \equiv \sum_{i=l+1}^{n_{max}} c_{ni}\phi^{(0)}_{il}(|\vec{p}|)Y_{lm}(\Omega_{\vec{p}}), \tag{27}$$

where

$$\phi^{(0)}_{il}(|\vec{p}|) \equiv \bar{p}^l a^{3/2} N_{il} \frac{l!}{\sqrt{2\pi}} \times \frac{4^{l+1}}{(\bar{p}^2+\delta^2)^{l+2}} C^{l+1}_{n-l-1}(\frac{\bar{p}^2-\delta^2}{\bar{p}^2+\delta^2}), \tag{28}$$

and

$$N_{il} = \frac{2}{i^2}\sqrt{\frac{(i-l-1)!}{(i+l)!}}, \quad \bar{p} = ap, \tag{29}$$

with $a = \frac{1}{\mu e^2}$ the Bohr radius.

By choosing a specific value of n_{max}, one can calculate the following matrix:

$$\int d^3\vec{p}_1 \Phi^{(0)\dagger}_{\bar{\alpha},\bar{\beta};\bar{i}}(\vec{p}_1)[E^{(e)}_1 + E^{(p)}_1]\delta_{\bar{\alpha}\alpha}\delta_{\bar{\beta}\beta}\Phi^{(0)}_{\alpha,\beta;i}(\vec{p}_1) + \int \frac{d^3\vec{p}_1 d^3\vec{p}_3}{(2\pi)^3}\Phi^{(0)\dagger}_{\bar{\alpha},\bar{\beta};\bar{i}}(\vec{p}_1)V_{\bar{\alpha}\alpha,\bar{\beta}\beta}(\vec{p}_1,\vec{p}_3)\Phi^{(0)}_{\alpha,\beta;i}(\vec{p}_3), \tag{30}$$

where $\Phi^{(0)}_{\alpha,\beta;i}(\vec{p}_1)$ is simply $\Phi_{\alpha,\beta}(\vec{p}_1)$ in Eq. (25) after replacing $\Phi_{nlm}(\vec{p}_1)$ with $\phi^{(0)}_{il}(\vec{p}_1)Y_{lm}$. The integration over the angles $\Omega_{\vec{p}1}$ and $\Omega_{\vec{p}_3}$ can be performed analytically, while the integration over $|\vec{p_1}|$ and $|\vec{p}_3|$ can be done numerically with high precision. After diagonalizing this matrix, the energy spectrum $E_n \equiv P0, n - m_e - m_p$ is obtained approximately.

III. NUMERICAL RESULTS AND DISCUSSION

In our numerical calculation, we set $n_{max} = 100$, and the physical constants are taken as $m_e = 0.510998950$ MeV, $m_p = 938.27208816$ MeV, and $\alpha_e = \frac{1}{137.035999084}$. The relative precision of the numerical calculation for each matrix element reaches 10^{-20}, ensuring that

the absolute precision of the matrix elements is better than 10^{-18} eV. This guarantees the reliability of the numerical results. The numerical precision is also tested for the matrix elements involving the Coulomb and Breit potentials.

To compare the results with those obtained using bound-state perturbative theory, we decompose the effective potential into three terms as follows:

$$V \equiv V_0 + V_1 + V_2, \tag{31}$$

where V_0 is spin-independent, V_1 depends only on the electron spin, and V_2 depends on the proton spin. We label the corresponding energy contributions due to $V_0, V_0+V_1, V_0+V_1+V_2$ as $E_{n,0}, E_{n,1}, E_{n,2}$, respectively. The contributions beyond the Breit potential are expressed by the following quantities:

$$\begin{aligned}
\Delta E_{n,0} &\equiv E_{n,0} - E_n^{(2)} - E_{n,0}^{(4)}, \\
\Delta E_{n,1} &\equiv E_{n,1} - E_n^{(2)} - E_{n,0}^{(4)} - E_{n,1}^{(4)}, \\
\Delta E_{n,2} &\equiv E_{n,2} - E_n^{(2)} - E_{n,0}^{(4)} - E_{n,1}^{(4)} - E_{n,2}^{(4)}.
\end{aligned} \tag{32}$$

n \ peV	$\Delta E_{n,0}(S)$	$\Delta E_{n,0}(P)$	$\Delta E_{n,0}(D)$
$n=1$	-2349361	-	-
$n=2$	-289552	813	-
$n=3$	-85154	424	13
$n=4$	-35738	242	16
$n=5$	-18225	151	14

TABLE I: The numeric results for the energy corrections $\Delta E_{n,0}(l)$, where the unit is peV (10^{-12}eV).

We present the numeric results for $\Delta E_{n,i}$ in Tabs. I, II and III, respectively, where $l = 0, 1, 2$ or S, P, D waves, $j = l + 1/2$ and $F = l + 1$ are considered. For the purpose of a more direct comparison, we also provide the contributions with specific orders as follows:

$$\begin{aligned}
\alpha_e^5 \mu &\sim 11\ \mu\text{eV}, \\
\alpha_e^6 \mu &\sim 77\ \text{neV}, \\
\alpha_e^6 \mu \frac{m_e}{m_p} &\sim 42\ \text{peV}.
\end{aligned} \tag{33}$$

n \ peV	$\Delta E_{n,1}(S)$	$\Delta E_{n,1}(P)$	$\Delta E_{n,1}(D)$
$n=1$	-2349361	-	-
$n=2$	-289552	-44	-
$n=3$	-85154	50	3
$n=4$	-35738	48	6
$n=5$	-18225	37	6

TABLE II: The numeric results for the energy correction $\Delta E_{n,1}(l)$ with $j = l + 1/2$, where the unit is peV (10^{-12}eV).

n \ peV	$\Delta E_{n,2}(S)$	$\Delta E_{n,2}(P)$	$\Delta E_{n,2}(D)$
$n=1$	-2357360	-	-
$n=2$	-290555	-45	-
$n=3$	-85452	50	3
$n=4$	-35864	48	6
$n=5$	-18289	37	6

TABLE III: The numeric results for the energy correction $\Delta E_{n,2}(l)$ with $F = j+1/2 = l+1$,where the unit is peV (10^{-12}eV).

The results in Tab. I, II and III clearly demonstrate a notable property whereby the contributions in the S wave are significantly larger than those in the P and D waves.

In terms of magnitude, the contributions $\Delta E_{n,0}(S)$ are approximately $-\frac{\alpha_e^5\mu}{4.5n^3} \sim -\frac{30\alpha_e^6\mu}{n^3}$, which are larger than the contributions at the order of $\frac{\alpha_e^6\mu}{n^3}$. The contributions $\Delta E_{n,0}(P)$ are on the order of $\frac{\alpha_e^6\mu}{20n^2}$. On the other hand, the contributions $\Delta E_{n,0}(D)$ weakly depend on n and are much smaller than $\alpha_e^6\mu$.

The results for $\Delta E_{n,1}(S)$ and $\Delta E_{n,2}(S)$ are similar to $\Delta E_{n,0}(S)$, indicating that the contributions from spin-independent terms are most significant for the S wave. This is expected, as the contributions from electron spin are zero for the S wave, and the contributions from proton spin are greatly suppressed.

The results for $\Delta E_{n,1}(P)$ and $\Delta E_{n,2}(P)$ are similar to each other and significantly differ from $\Delta E_{n,0}(P)$. This indicates that in the P wave, the contributions from the electron

spin are of the same order as the spin-independent contributions. The spin-dependent contributions in the D wave are similar to those in the P wave.

To illustrate the contributions more explicitly, we define the following terms:

$$\begin{aligned} \Delta E_n^{fin} &\equiv \Delta E_{n,1} - \Delta E_{n,0} = E_{n,1} - E_{n,0} - E_{n,1}^{(4)}, \\ \Delta E_n^{hyf} &\equiv \Delta E_{n,2} - \Delta E_{n,1} = E_{n,2} - E_{n,1} - E_{n,2}^{(4)}. \end{aligned} \tag{34}$$

These terms reflect the corrections to the fine structure and hyperfine structure beyond the Breit potential, respectively.

n \ peV	$\Delta E_n^{fin}(S)$	$\Delta E_n^{fin}(P)$	$\Delta E_n^{fin}(D)$
$n=1$	0	-	-
$n=2$	0	-857	-
$n=3$	0	-374	-10
$n=4$	0	-194	-10
$n=5$	0	-114	-8

TABLE IV: Numeric results for the energy correction $\Delta E_n^{fin}(l)$ with $j = l + 1/2$, where the unit is peV (10^{-12}eV).

n \ peV	$\Delta E_n^{hyf}(S)$	$\Delta E_n^{hyf}(P)$	$\Delta E_n^{hyf}(D)$
$n=1$	-7999	-	-
$n=2$	-1003	-0.66	-
$n=3$	-298	-0.27	-0.007
$n=4$	-126	-0.14	-0.007
$n=5$	-65	-0.08	-0.005

TABLE V: Numeric results for the energy correction $\Delta E_n^{hyf}(l)$ with $F = l + 1$, where the unit is peV (10^{-12}eV).

The numeric results for ΔE_n^{fin} and ΔE_n^{hyf} are presented in Tab. IV and V, respectively. The numerical results in Tab. IV indicate that the contributions to the fine structure beyond the Breit potential are approximately $-\frac{\alpha_e^6 \mu}{22 n^2}$ in the P wave and around -10 peV

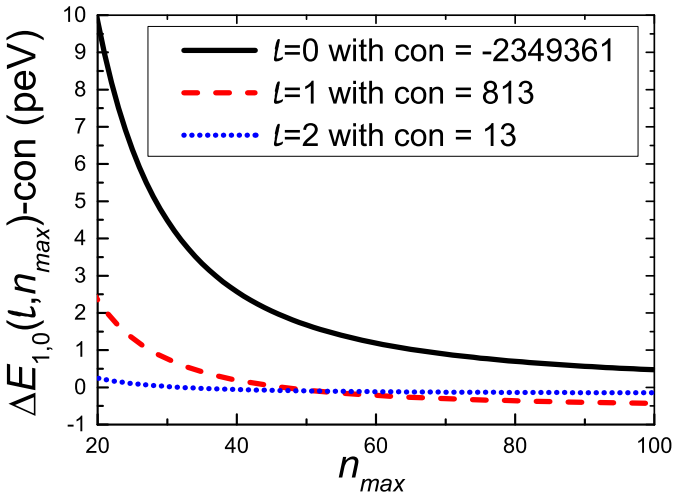


FIG. 2: Numeric results for $\Delta E_{1,0}(l, n_{\max})$ vs. n_{max} with $l = 0, 1$ and 2, where the unit is peV (10^{-12}eV).

in the D wave. These values are much smaller compared to the contributions of $-\frac{\alpha_e^6 \mu}{n^2}$. The numeric results in Tab. V show that the contributions to the hyperfine structure beyond the Breit potential are approximately $-\frac{200\alpha_e^6\mu}{n^3}\frac{m_e}{m_p}$ for the S wave, while the contributions are negligible for the P wave and D wave cases.

To demonstrate the uncertainty arising from the approximation of finite $n_{\max}$, we plot $\Delta E_{1,0}(l, n_{max})$ as a function of n_{max} in Fig. 3. The results clearly indicate that the uncertainty is smaller than 0.5peV when $n_{max} > 60$, suggesting that the approximation is reliable.

In our calculation, we do not expand the OPE interaction kernel order by order in momenta, but instead solve the effective Schrodinger-like equation using numerical method. This approach enables us to include all contributions from ladder diagrams with the approximated propagator $\bar{G}_0$ and the full photon propagator, while excluding the crossed diagrams. Consequently, the comparison of our results with those obtained from bound-state perturbation theory is not straightforward.

For a consistent comparison between the two approaches, one would need to separate the contributions in bound-state perturbation theory based on the diagram types and momentum regions, and then compare them with our results to assess the relativistic contributions obtained through the expansion in bound-state perturbation theory. How-

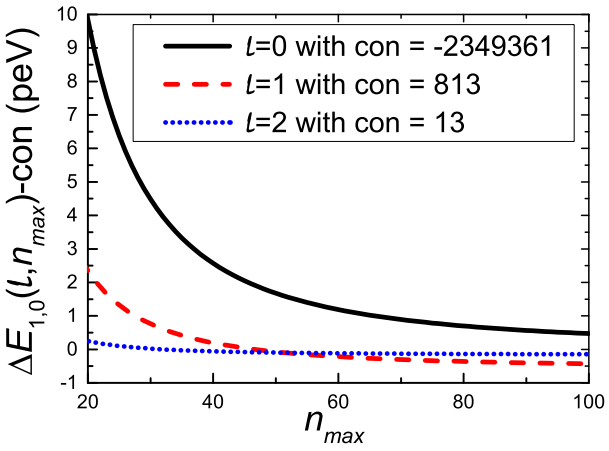


FIG. 3: Numeric results for $\Delta E_{1,0}(l, n_{\max})$ vs. n_{max} with $l = 0, 1$ and 2, where the unit is peV (10^{-12}eV).

ever, such a direct separation and comparison fall beyond the scope of this work.

In summary, the effective potential associated with the full OPE interaction in momentum space is expressed through eight scalar functions. The expansions of these scalar functions directly correspond to the $ep \to ep$ amplitude in NRQED or the quasi-potential at order α_e and any desired order of momenta. Our precise numerical calculations suggest that it is possible to capture all relativistic contributions using this method. Extending these calculations to positronium, muonic hydrogen, and cases involving nuclear structure and radiative corrections would be interesting directions for future research.

IV. ACKNOWLEDGMENTS

H.Q.Z. would like to thank Zhi-Hui Guo for his helpful suggestions and discussions. This work is funded in part by the National Natural Science Foundations of China under Grants No. 12075058, No. 12150013.

[1] Pohl R, *et al.*, Nature **466**, 213 (2010).

[2] Aldo Antognini, *et al.*, Science **339**, 417 (2013).

[3] A. Beyer, L. Maisenbacher, A. Matveev, R. Pohl, K. Khabarova, A. Grinin, T. Lamour, D.C. Yost, T.W. Hansch, N. Kolachevsky et al., Science 358, 79 (2017).

[4] H. Fleurbaey, S. Galtier, S. Thomas, M. Bonnaud, L. Julien, F. m. c. Biraben, F . m. c. Nez, M. Abgrall, J. Guena, Phys. Rev. Lett. 120, 183001 (2018).

[5] N. Bezginov, T. Valdez, M. Horbatsch, A. Marsman, A.C. Vutha, E.A. Hessels, Science 365, 1007 (2019).

[6] A. Grinin, A. Matveev, D.C. Yost, L. Maisenbacher, V. Wirthl, R. Pohl, T.W. Hansch, T. Udem, Science 370, 1061 (2020).

[7] A. D. Brandt, S. F. Cooper, C. Rasor, Z. Burkley, D. C. Yost and A. Matveev, Phys. Rev. Lett. **128**, 023001 (2022).

[8] R. G. Bullis, C. Rasor, W. L. Tavis, S. A. Johnson, M. R. Weiss and D. C. Yost, Phys. Rev. Lett. **130**, 203001 (2023).

[9] M. I. Eides, H. Grotch and V. A. Shelyuto, "Theory of Light Hydrogenic Bound States," Springer Tracts Mod. Phys. **222**,1 (2007), ISBN 978-3-540-45269-0, 978-3-540-45270-6.

[10] U. D. Jentschura and G. S. Adkins, "Quantum Electrodynamics: Atoms, Lasers and Gravity," World Scientific, 2022, ISBN 978-981-12-5225-9, 978-981-12-5227-3.

[11] K. Pachucki, V. Lensky, F. Hagelstein, S. S. Li Muli, S. Bacca and R. Pohl, Rev. Mod. Phys. **96**, 015001 (2024).

[12] H. W. E. Caswell and G. P. Lepage, Phys. Rev. A **18**, 810 (1978).

[13] G. P. Lepage, Phys. Rev. A **16**, 863 (1977).

[14] E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).

[15] M. Gell-Mann and F. Low, Phys. Rev. **84**, 350 (1951).

[16] W. Caswell and G. Lepage, Phys. Lett. B 167, 437 (1986).

[17] H. Grotch and D. r. Yennie, Rev. Mod. Phys. **41**, 350 (1969).

[18] R. N. Faustov, A. P. Martynenko, Teor.Mat.Fiz. 64,179 (1985).

[19] R. N. Faustov, A. P. Martynenko, Zh.Eksp.Teor.Fiz. 115, 1221 (1999).

[20] A. P. Martynenko, Jour. Exp. Theor. Phys. 101, 1021 (2005).

[21] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. 78, 29 (1950).

[22] K. Pachucki, Phys. Rev. A **71**, 012503 (2005).

[23] W. Zhou, X. Mei and H. Qiao, J. Phys. B **56**, 045001 (2023).

[24] K. Pachucki, Phys. Rev. A **56**, 297 (1997).

[25] A. Czarnecki, K. Melnikov and A. Yelkhovsky, Phys. Rev. A **59**, 4316 (1999).

[26] G. S. Adkins, M. F. Alam, C. Larison and R. Sun, Phys. Rev. A **101**, 042511 (2020).

[27] Z. X. Zhong, W. P. Zhou and X. S. Mei, Phys. Rev. A **98**, no.3, 032502 (2018) [erratum: Phys. Rev. A **101**, no.6, 069901 (2020)].

[28] W. Zhou, X. Mei and H. Qiao, Phys. Rev. A **100**, 012513 (2019).

[29] M. Haidar, Z. X. Zhong, V. I. Korobov and J. P. Karr, Phys. Rev. A **101**, 022501 (2020).

[30] G. S. Adkins, J. Gomprecht, Y. Li and E. Shinn, Phys. Rev. Lett. **130**, 023004 (2023).

CALT-TH 2024-009

Generalized Symmetry in Dynamical Gravity

Clifford Cheung,[a] **Maria Derda,**[a] **Joon-Hwi Kim,**[a]
Vinicius Nevoa,[a] **Ira Rothstein,**[b] **Nabha Shah**[a]

[a] *Walter Burke Institute for Theoretical Physics,*
California Institute of Technology, Pasadena, CA 91125
[b] *Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213*

Abstract: We explore generalized symmetry in the context of nonlinear dynamical gravity. Our basic strategy is to transcribe known results from Yang-Mills theory directly to gravity via the tetrad formalism, which recasts general relativity as a gauge theory of the local Lorentz group. By analogy, we deduce that gravity exhibits a one-form symmetry implemented by an operator U_α labeled by a center element α of the Lorentz group and associated with a certain area measured in Planck units. The corresponding charged line operator W_ρ is the holonomy in a spin representation ρ, which is the gravitational analog of a Wilson loop. The topological linking of U_α and W_ρ has an elegant physical interpretation from classical gravitation: the former materializes an exotic chiral cosmic string defect whose quantized conical deficit angle is measured by the latter. We verify this claim explicitly in an AdS-Schwarzschild black hole background. Notably, our conclusions imply that the standard model exhibits a new symmetry of nature at scales below the lightest neutrino mass. More generally, the absence of global symmetries in quantum gravity suggests that the gravitational one-form symmetry is either gauged or explicitly broken. The latter mandates the existence of fermions. Finally, we comment on generalizations to magnetic higher-form or higher-group gravitational symmetries.

arXiv:2403.01837v1 [hep-th] 4 Mar 2024

Contents

1 Introduction **2**

2 Gauge Theory **4**

2.1 Covariant Formalism 5

2.1.1 Line and Symmetry Operators 5

2.1.2 Ward Identity 9

2.1.3 Explicit Examples 13

2.2 Canonical Formalism 18

2.2.1 Phase Space 18

2.2.2 Ward Identity 18

3 Gravity **21**

3.1 Covariant Formalism 21

3.1.1 Line and Symmetry Operators 24

3.1.2 Ward Identity 26

3.1.3 Chiral Cosmic String 28

3.1.4 Linking and Conical Deficit Angle 30

3.1.5 Explicit Examples 32

3.2 Canonical Formalism 38

3.2.1 Phase Space 38

3.2.2 Ward Identity 39

3.3 Symmetry Breaking 40

4 Future Directions **41**

A Notations and Conventions **44**

B Geometry of Linking **47**

1 Introduction

Symmetry has long been a vital tool for investigating complex physical systems, particularly at strong coupling. Historically, most efforts in this expansive subject have focused on conventional symmetries, which act on *local* operators. The standard model of physics exhibits numerous exact and approximate symmetries of this type, for example relating to charge in electromagnetism and chiral symmetry in the strong interactions.

In the past decade, however, the fundamental concept of symmetry has broadened considerably [1–10]. As described in the seminal work of [11], it is now understood that the traditional formulation of symmetry is actually the tip of a colossal iceberg. Rather, there exists a rich patchwork of so-called higher-form symmetries whose distinguishing feature is that they act intrinsically on extended objects described by *nonlocal* operators supported on lines, surfaces, and membranes. Since higher-form symmetries act trivially on local operators, their physical implications are sometimes quite subtle to diagnose. From this point of view, the standard symmetries found in most quantum field theory textbooks are brusquely relegated to the special case of zero-form symmetry.

The growing body of work on generalized symmetries has revealed new perspectives on a broad spectrum of assorted phenomena in quantum field theory, including phase transitions [12–17], anomalies [18–29], and symmetry breaking [30–34]. Recent work has even explored new opportunities for physics beyond the standard model, for example in the context of flavor physics [35], neutrinos [36], and axions [37–46]. Such efforts are a welcome development, as they attempt to draw an explicit connection between highly formal developments in mathematical physics and high-energy physics of actual experimental relevance. That said, the constraints imposed by generalized symmetry on particle physics models tend to be explicable via more conventional means. This is perhaps not so surprising—these models are easily embedded within renormalizable theories in which all is calculable and there are no surprises to be had or which require explanation.

Gravity, on the other hand, is another story. Far less is understood about its putative ultraviolet completion. Consequently, the only truly theory-agnostic approach is to retreat to safely low energies, where gravitational dynamics are described universally by an effective field theory of gravitons on a fixed background, augmented by possible higher-derivative corrections. For example, see [47, 48] for a review of this perspective. The effective field theory of gravity is clearly a natural target for understanding generalized symmetry in a refreshingly different context. There has, however, been relatively little effort in this vein.[1] Some notable exceptions include interesting recent work studying the higher-form symmetries associated with parity [49] and topology change [50], as well as generalizations of continuous higher-form symmetries to *linearized gravity* [51–53].

In this paper, we extend the now well-established insights of higher-form symmetry

[1] The bulk of work that makes reference to both gravity and generalized symmetries has focused on the implications of swampland conjectures. In this picture, one posits a quantum field theory that exhibits certain generalized global symmetries. The conjectured absence of global symmetries in a theory of quantum gravity then imposes constraints on the theory in order to explicitly break or gauge these symmetries. Though interesting, this subject is not the topic of the present work.

in gauge theory to the effective theory of *nonlinear gravity* in four-dimensional spacetime. Our key ingredient is the well-known fact that gravity can itself be recast in gauge theoretic language. As history would have it, this perspective carries dual meanings. On the one hand, gravity is a theory of diffeomorphisms, nonlinearly realized by a self-interacting, massless spin two field. Since diffeomorphisms are a redundancy, they are on occasion referred to as a gauge symmetry, though colloquially and not in the strict technical sense. On the other hand, it is well-known that gravity can also be described by a bona fide gauge theory of local Lorentz transformations, which is the so-called Palatini formalism for the tetrad and spin connection. Formally, these descriptions are equivalent[2] since gauge symmetry is, after all, pure redundancy and no redundancy is more valid than any other.[3] As we will see, tetradic Palatini gravity is perfectly suited to our purposes because we can work in lockstep analogy with the familiar approach taken in gauge theory. For concreteness, the bulk of our analysis will be in Euclidean signature, though we will toggle to Lorentzian signature on and off when needed. Our conclusions for gravity are as follows.

First and foremost, our central claim is that tetradic Palatini gravity exhibits an electric one-form symmetry described by the center subgroup $Z(G)$ of the Lorentz group G.[4] This one-form symmetry depends crucially on the signature and global structure of G. For example, in Euclidean signature the center is nontrivial when we consider $Z(\mathrm{SO}(4)) = \mathbb{Z}_2$ or $Z(\mathrm{Spin}(4)) = \mathbb{Z}_2 \times \mathbb{Z}_2$, while in Lorentzian signature the center is nontrivial for $Z(\mathrm{SL}(2,\mathbb{C})) = \mathbb{Z}_2$. In all cases, these center subgroups have a zero-form symmetry action as various parities on Lorentz vector and spinor indices.

Second, we show how the one-form symmetry of gravity is implemented by a topological symmetry operator U_α. This object is constructed explicitly in terms of the local degrees of freedom as the exponential of a certain area operator for a closed surface measured in Planck units and labeled by an element of the center α. The symmetry operator U_α acts on a line operator W_ρ known as the spin holonomy, which is simply a Wilson loop for the spin connection computed in a spin representation ρ along a chosen contour. While U_α generates a *global* one-form symmetry, it can be implemented as a field transformation that is precisely the form of a local Lorentz transformation, but with nontrivial winding that precludes it from being a genuine local Lorentz transformation. Using this "twisted local Lorentz transformation", we show that W_ρ transforms by a center-valued phase that depends only on the topological linking of the surface and curve which define U_α and W_ρ, respectively. We prove the Ward identity for U_α and W_ρ using both covariant and canonical

[2] At low energies, general relativity and the tetradic Palatini formalism are equivalent classically and quantum mechanically since they both reproduce a local effective field theory of a massless spin two particle. As is well-known, the dynamics of such a theory are uniquely fixed, up to unknown Wilson coefficients.

[3] Of course, one can always start from the tetradic Palatini formalism and simply *integrate out* the spin connection and *gauge fix* the tetrad algebraically, thus reverting to the usual metric description of gravity. Doing so should yield the same physics, since these redundancies are unphysical. However, for our analysis it will be far more illuminating to keep the tetrad and spin connection since we will be especially interested in the gravitational interactions of fermions and their worldlines, which play an absolutely essential role in our construction. Said another way, the pure metric formulation is poorly equipped to describe fermions.

[4] In an abuse of notation, we will hereafter refer to the gauge group G of the tetradic Palatini formalism as the "Lorentz group" even though we will consider both Euclidean and Lorentzian signatures.

approaches. Notably, this proof is valid to all orders in perturbation theory, at least within the context of the effective field theory description of gravity[5] where the topology and dimension of spacetime are preserved.

Thirdly, we show that the interplay of U_α and W_ρ has a remarkably simple interpretation in terms of *classical gravitation*. The symmetry operator U_α creates a defect in spacetime that is a chiral version of a cosmic string defect, and serves as a certain gravitational analog of the Dirac string. The tension of U_α is quantized so as to induce a π deficit angle which is directly measured by the spin holonomy W_ρ as the center-valued linking number. We then compute the linking number by evaluating W_ρ on various spacetimes, including an AdS-Schwarzschild background. The topological nature of U_α implies that its linking with W_ρ arises purely from contributions at leading order in the so-called self-force expansion, where U_α is treated as a nondynamical background. Furthermore, this implies that higher order self-force corrections are vanishing, so evidently the classical deficit angle is not quantum corrected at any perturbative order.

Last but not least, we discuss the breaking of the gravitational one-form symmetry. As expected, explicit breaking requires a local operator in the representation ρ that renders the spin holonomy W_ρ "endable," thus unspooling its linking with U_α. Physically, this corresponds to the screening of the spin holonomy by spinning particles. Interestingly, the spin holonomy in the vector representation is automatically screened in pure gravity by orbital angular momentum. This mirrors the phenomenon in gauge theory where adjoint Wilson lines are screened by the gluon field itself. On the other hand, holonomies in the spinor representation are endable only by local fermionic operators. If no such operators exist, then the one-form gravitational symmetry is exact. Remarkably, this implies the emergence of a hidden symmetry of the real world: below the lightest neutrino mass, there is a gravitational one-form symmetry under which spinor holonomies are charged. More generally, in a theory in which the gravitational one-form symmetry is not gauged, the conjectured absence of exact global symmetries in quantum gravity directly implies the existence of fermions.

2 Gauge Theory

In this section, we present a self-contained review of one-form symmetries in gauge theory. Other treatments can be found in the literature [54–60]. We start with a covariant analysis expressed in the language of path integrals, followed by a treatment in terms of canonical quantization. Because this section is mostly—though not entirely—a recap of known results from gauge theory, it may be skipped by readers interested only in our new findings, which pertain to gravity. However, we note that this gauge theory warm up forms a concrete road map for our parallel analysis of gravity later on.

[5] As is well-known, quantum corrections are perfectly well-defined even within a low-energy effective field theory, provided one enforces systematic power counting. In the effective field theory of gravity, most quantum corrections are ultraviolet sensitive and thus absorbed into incalculable counterterms. However, there also exist calculable long-distance quantum corrections [47].

2.1 Covariant Formalism

To begin, let us consider Yang-Mills theory for a gauge group G which is a connected matrix Lie group. We take spacetime to be a Riemannian four-manifold $\mathcal{M}$ of Euclidean signature. Our discussion will apply irrespective of whether or not the background spacetime is curved, provided it is nondynamical. As is well-known, this theory admits a first-order formulation in terms of a one-form gauge connection and an auxiliary two-form field,

$$A^a = A^a{}_\mu \, dx^\mu \qquad \text{and} \qquad B_a = \frac{1}{2} B_{a\mu\nu} \, dx^\mu \wedge dx^\nu \,, \tag{2.1}$$

valued in the adjoint and the coadjoint of G, respectively, so $a, b, \ldots \in \{1, \ldots, \dim(G)\}$. The action is the integral over $\mathcal{M}$ of the Lagrangian four-form,

$$L = \frac{1}{g^2} B_a \wedge F^a - \frac{1}{2g^2} B_a \wedge *B^a \qquad \text{where} \qquad F^a = dA^a + \frac{1}{2} f^a{}_{bc} \, A^b \wedge A^c \,, \tag{2.2}$$

where g is the gauge coupling. Throughout, color indices are raised and lowered by the Killing form, while the Hodge dual $*$ with respect to the background metric acts on spacetime indices. Integrating out the auxiliary two-form field enforces $B_a = *F_a$,[6] and plugging this back in to Eq. (2.2), we obtain $(1/2g^2) \, F_a \wedge *F^a$, which is the textbook Lagrangian for Yang-Mills theory in a fixed background spacetime.

2.1.1 Line and Symmetry Operators

In Yang-Mills theory, the one-form symmetry group is identified with the center of the gauge group, $Z(G)$. By definition, the one-form symmetry acts on extended objects rather than local operators. The relevant charged object is the one-dimensional line operator,

$$W_\rho(\mathcal{C}) = \mathrm{tr}_\rho \, \mathrm{P} \exp\left(\oint_{\mathcal{C}} A \right) , \tag{2.3}$$

which is a path-ordered Wilson loop along a closed contour $\mathcal{C}$. Here ρ denotes the irreducible representation in which the trace and exponentiation are defined.[7]

Meanwhile, the one-form symmetry transformation is implemented by a corresponding symmetry operator,

$$U_\alpha(\mathcal{S}) \,, \tag{2.4}$$

which is an instance of the Gukov-Witten operator [61]. This operator is supported on a two-dimensional surface $\mathcal{S}$ and labeled by a center element $\alpha \in Z(G)$. We will present a concrete formula for $U_\alpha(\mathcal{S})$ in terms of explicit fields later on. But for the moment, let

[6] Here we emphasize to the reader that despite appearances B *does not* denote the magnetic field. Rather, it is the two-form of the "BF" formulation of Yang-Mills theory, so $\int B = \int *F$ and $\int *B = \int F$ denote electric and magnetic fluxes when integrated over a spatial surface, respectively.

[7] Note that there is no factor of i in the exponential map because we are using an anti-Hermitian convention for the Lie algebra generators.

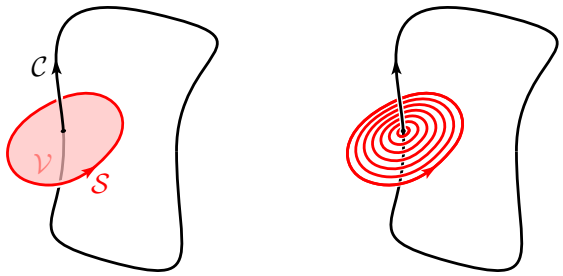


Figure 1. *Left:* The linking number between a one-dimensional contour $\mathcal{C}$ and an exact codimension two surface $\mathcal{S} = \partial\mathcal{V}$ is equal to the number of intersections between the coboundary $\mathcal{V}$ and $\mathcal{C}$. *Right:* The coboundary $\mathcal{V}$ defines a homotopy for shrinking the surface $\mathcal{S}$ to a point on $\mathcal{C}$.

us abstractly describe the symmetry operator in terms of the defining property that it generates the following transformation on the line operator,

$$W_\rho(\mathcal{C}) \;\mapsto\; \rho(\alpha)^{\text{link}(\mathcal{C},\mathcal{S})}\, W_\rho(\mathcal{C})\,, \tag{2.5}$$

where $\rho(\alpha)$ is the representation of the center element α as a complex phase, and we have defined $\text{link}(\mathcal{C},\mathcal{S})$ to be the linking number between the contour $\mathcal{C}$ and the surface $\mathcal{S}$. Crucially, since the linking number is a topological invariant, so too is the operator $U_\alpha(\mathcal{S})$, in the sense that the surface of its support $\mathcal{S}$ can be deformed arbitrarily to yield the same action on the Wilson loop $W_\rho(\mathcal{C})$ provided it does not degenerate with $\mathcal{C}$.

In this paper, we will always assume that the two-dimensional support of the symmetry operator is not only closed but also exact, so the surface $\mathcal{S} = \partial\mathcal{V}$ is the boundary of a three-dimensional volume $\mathcal{V}$. This is required so that the symmetry operator can be contracted continuously into an infinitesimal two-sphere enclosing the line operator, as depicted in Fig. 1. Intuitively, this deformation corresponds to the physical measurement of the electric charge of a body by computing the electric flux flowing through an infinitesimal two-sphere enclosing it. As a consequence, we see that

$$\text{link}(\mathcal{C},\mathcal{S}) = \text{int}(\mathcal{C},\mathcal{V}) \qquad \text{where} \qquad \mathcal{S} = \partial\mathcal{V}\,. \tag{2.6}$$

so the linking number between $\mathcal{S}$ and $\mathcal{C}$ is equal to the intersection number between $\mathcal{C}$ and the coboundary $\mathcal{V}$.

In Maxwell theory, it is well-known that the one-form symmetry is implemented by a shift of the gauge field by a "flat connection", which is *closed* wherever it is well-defined and nonsingular, but crucially *not exact*. A key fact that we now emphasize is that this can be realized as a transformation of the fields that takes the form of a gauge transformation for a multivalued—that is, winding—gauge parameter, which is hence is not globally defined. For instance, consider the map $A \mapsto A + d\chi$. If the zero-form parameter χ exhibits nontrivial winding, then $d\chi$ is not, despite its appearance, an exact form. For example, we might choose $\chi = \varepsilon\,\phi$, where ε is a constant and ϕ is the azimuthal angle in cylindrical coordinates. Crucially, $d\phi = (x\,dy - y\,dx)/(x^2+y^2)$ is not exact because its integral around a closed circular loop, $\int_{S^1} d\phi = 2\pi$, is nonzero.

There are two distinct ways to interpret this winding connection. From the mathematician's perspective, $d\phi$ would be described as a closed-but-not-exact one-form defined on the x-y plane with the origin excised, which is $\mathbb{R}^2 \setminus 0$. However, throughout this paper we will adopt the equivalent physicist's picture, which is to instead specify the exterior derivative of $d\phi$ as a *distributional* two-form in the entire x-y plane *without* excising the origin. That is, by demanding that Stokes theorem apply, we deduce $dd\phi = 2\pi\,\delta(x)\delta(y)\,dx\wedge dy$.[8] In this picture, under the transformation $A \mapsto A + \varepsilon\, d\phi$ the field strength shifts by $dd\chi = 2\pi\varepsilon\,\delta(x)\delta(y)\,dx\wedge dy$, which describes a magnetic flux tube, or a Dirac string. The total flux $2\pi\varepsilon$ of this Dirac string can be arbitrary and is measured by the induced phase on the Wilson loops. We emphasize that in this more physical picture, the Maxwell action is always defined over all of space without the excision of any particular support. Furthermore, the shift of the field strength $F \mapsto F + 2\pi\varepsilon\,\delta(x)\delta(y)\,dx\wedge dy$ correctly describes the fact that the one-form symmetry transformation of the gauge field, $A \mapsto A + \varepsilon\, d\phi$, does not leave the Lagrangian invariant on the locus of the Dirac string.[9]

An exactly analogous construction applies to Yang-Mills theory, which we now describe. In particular, in this case the one-form symmetry is realized by

$$A^a \;\mapsto\; (\Omega^{-1}A\Omega)^a + (\Omega^{-1}d\Omega)^a \qquad \text{and} \qquad B_a \;\mapsto\; (\Omega^{-1}B\,\Omega)_a\,, \tag{2.7}$$

where Ω is a zero-form parameter which is valued in the gauge group G and approaches the identity at infinity.[10] Here we also stipulate the crucial additional condition that Ω is multivalued and exhibits nontrivial winding. In the presence of winding, a global definition of Ω requires a collection of multiple charts which define it on *subregions* of spacetime, but together yield an atlas for all points. For any particular subregion chart, the corresponding function will necessarily have a branch cut residing on some volume, which we define to be $\mathcal{V}$. The boundary of $\mathcal{V}$ then coincides with $\mathcal{S}$ in this subregion, which is to say $\mathcal{S} = \partial\mathcal{V}$. So practically, when we define an explicit function for Ω in a given subregion we can deduce $\mathcal{S}$ directly from the branch cut hypersurface $\mathcal{V}$. For subregions outside of this particular chart for Ω—which in many cases includes asymptotic infinity—we can say nothing until we define another chart for Ω in that other patch.

Concretely, we will consider Ω which exhibits a discontinuity across $\mathcal{V}$ such that

$$\lim_{\mathcal{P}_\pm \to \mathcal{P}} \Omega(\mathcal{P}_+)\,\Omega^{-1}(\mathcal{P}_-) \;=\; \alpha \qquad \text{where} \qquad \alpha \in Z(G)\,, \tag{2.8}$$

where $\mathcal{P}_+$ and $\mathcal{P}_-$ are points infinitesimally displaced away from the same point on $\mathcal{V}$, but in opposite directions. Since Ω is multivalued, $d\Omega$ is not exact. The seemingly innocuous caveat implies that Eq. (2.7) is *not* a gauge transformation in the traditional sense, despite

[8] For the more mathematically inclined, this delta function expression can be thought of as a shorthand for stating a cocycle condition. A deformation retract of the triple overlap turns into the support of the delta function. See also the discussion in [61].

[9] For quantized values of ε, the shift of the gauge field by a flat connection is an invariance of the path integral and thus corresponds to a bona fide gauge transformation. If ε is nonquantized, then the shift of the gauge field implements the global one-form symmetry.

[10] The notations $\Omega^{-1}A\Omega$ and $\Omega^{-1}B\Omega$ here signify adjoint and coadjoint actions, which is validated by the fact that we specialize in matrix Lie groups and algebras. See App. A for a comment.

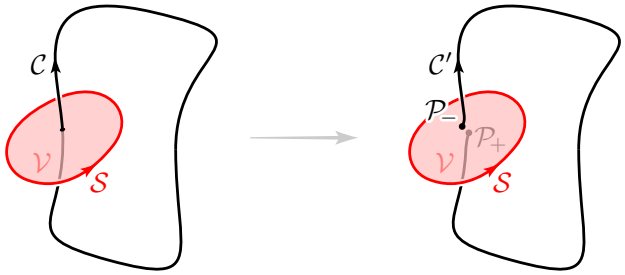


Figure 2. Infinitesimal opening of the closed contour $\mathcal{C}$ at the intersection point $\mathcal{C} \cap \mathcal{V}$.

its appearance. In particular, it does not leave Wilson loops invariant, which is why it corresponds to a global one-form symmetry. In some of the literature, the transformation defined in Eq. (2.7) is sometimes referred to as a "large gauge transformation" [62] in analogy with instanton configurations which support topological winding in a similar fashion. For the present work we refer to Eq. (2.7) as a "twisted gauge transformation," on account of the structural form of Eqs. (2.7) and (2.8).

To understand how Eq. (2.7) is equivalent to Eq. (2.5), consider a Wilson loop for a contour $\mathcal{C}$ that intersects with the coboundary $\mathcal{V}$ exactly once, so $\text{int}(\mathcal{C}, \mathcal{V}) = 1$. From Eq. (2.6), we see that this also means $\mathcal{C}$ *links* exactly once with $\mathcal{S}$, so $\text{link}(\mathcal{C}, \mathcal{S}) = 1$. We then find that the Wilson loop transforms as

$$\begin{aligned} W_\rho(\mathcal{C}) \;&\mapsto\; \lim_{\mathcal{C}' \to \mathcal{C}} \text{tr}_\rho \left[\text{Pexp} \left(\int_{\mathcal{C}'} \Omega^{-1} A \, \Omega + \Omega^{-1} d\Omega \right) \right] , \\ &=\; \lim_{\mathcal{C}' \to \mathcal{C}} \text{tr}_\rho \left[\Omega^{-1}(\mathcal{P}_-) \, \text{Pexp} \left(\int_{\mathcal{C}'} A \right) \Omega(\mathcal{P}_+) \right] , \end{aligned} \tag{2.9}$$

where $\mathcal{C}'$ is nearly identical to $\mathcal{C}$ except that it has been infinitesimally "cut open" in the vicinity of $\mathcal{V}$ such that $\partial \mathcal{C}' = \mathcal{P}_+ - \mathcal{P}_-$, as depicted in Fig. 2.[11] Cyclically permuting the terms inside the trace, we find that the Wilson loop transforms as

$$W_\rho(\mathcal{C}) \;\mapsto\; \lim_{\mathcal{C}' \to \mathcal{C}} \text{tr}_\rho \left[\alpha \, \text{Pexp} \left(\int_{\mathcal{C}'} A \right) \right] \;=\; \rho(\alpha) \, W_\rho(\mathcal{C}) \, , \tag{2.10}$$

where $\rho(\alpha)$ enters with a single power because $\text{int}(\mathcal{C}, \mathcal{V}) = 1$. As a result, we find that Eq. (2.10) is precisely the desired transformation of the Wilson loop for $\text{link}(\mathcal{C}, \mathcal{S}) = \text{int}(\mathcal{C}, \mathcal{V}) = 1$. When generalized to arbitrary linking number, the above calculation establishes the one-form symmetry transformation law for Wilson loops defined in Eq. (2.5).

The astute reader will notice that it was essential that the mismatch in the twisted gauge transformation is valued in a center element $\alpha \in Z(G)$, to make the symmetry operator topological. Otherwise, the branch cut $\mathcal{V}$ cannot be arbitrarily chosen, since $W_\rho(\mathcal{C})$ will transform differently under Ω depending on precisely where $\mathcal{C}$ has been cut open to yield $\mathcal{C}'$. In other words, if α were an arbitrary group element, its placement in the Wilson loop would matter and thus the corresponding transformation would not be topological.

[11] App. B contains a detailed accounting of the various signs and orientations associated with the curves and surfaces shown here.

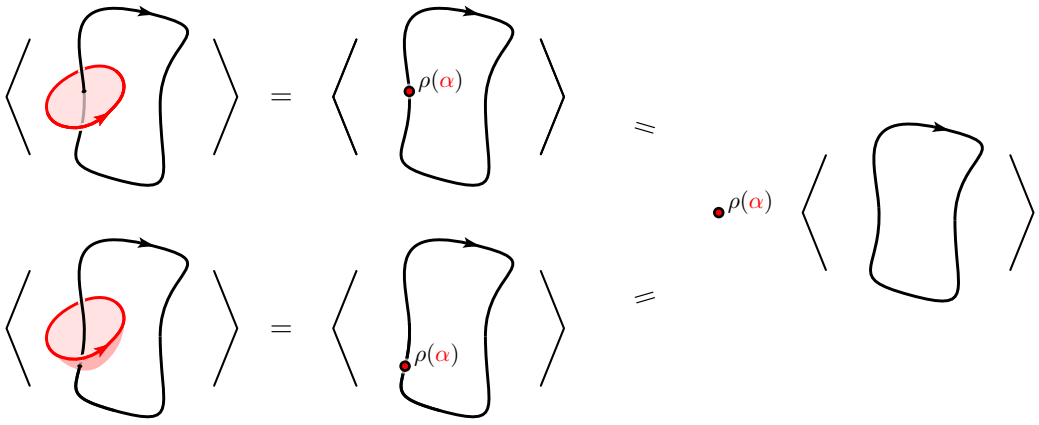


Figure 3. The one-form symmetry operator essentially inserts $\rho(\alpha)$ into the trace of the Wilson loop, where $\alpha \in Z(G)$ is a center element. The location of this insertion can be arbitrary since α commutes with all elements of G.

Before moving on, let us comment on a likely point of confusion. We have implemented the one-form symmetry transformation using a twisted gauge transformation Ω that is multivalued with a discontinuity center-valued in α. However, we saw earlier that the symmetry operator $U_\alpha(\mathcal{S})$ that implements this transformation should be labeled solely by the center element α rather than a whole zero-form parameter Ω. Why does the twisted gauge transformation depend on Ω rather than just its twist α? The resolution to this puzzle is that the naive Ω dependence in the twisted gauge transformation is spurious. Since $U_\alpha(\mathcal{S})$ is a topological surface operator, it only links with one-dimensional objects. The only such gauge invariant objects are Wilson loops, and we have already shown that the action of the twisted gauge transformation on Wilson loops *only* depends on the center element α, and not the details of Ω. Hence, different choices of Ω which have the same twist valued in α are physically indistinguishable. In other words, symmetry operator is gauge invariant despite the appearance of a reference structure Ω.

2.1.2 Ward Identity

Next, let us now derive the Ward identity which encodes the interplay between the symmetry operator $U_\alpha(\mathcal{S})$ and the line operator $W_\rho(\mathcal{C})$. Our goal is to prove that

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\big\rangle = \rho(\alpha)^{\mathrm{link}(\mathcal{C},\mathcal{S})}\big\langle W_\rho(\mathcal{C})\big\rangle\,, \tag{2.11}$$

where the brackets denote the path integral over all fields, so for example

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\big\rangle = \int \mathcal{D}A\,\mathcal{D}B\; e^{-S}\, U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\,. \tag{2.12}$$

Obviously, Eq. (2.11) is simply the transformation law for Wilson loops in Eq. (2.5), expressed in the language of covariant path integrals.

Earlier, we asserted that the one-form symmetry transformation in Eq. (2.5) is equivalent to the twisted gauge transformation defined in Eq. (2.7). The latter is implemented

by the symmetry operator, which can be written in the explicit form,

$$U_\alpha(\mathcal{S}) = \exp\left(\frac{2\pi}{g^2}\int_{\mathcal{S}} \lambda^a B_a\right) \qquad \text{where} \qquad e^{2\pi\lambda} = \alpha \in Z(G)\,, \tag{2.13}$$

where λ^a is an adjoint-valued zero-form. Our claim is that the Ward identity in Eq. (2.11) follows mechanically from the definition of the line and symmetry operators in Eqs. (2.3) and (2.13), and furthermore we can see directly how the one-form symmetry transformation arises as a twisted gauge transformation. As before, the physical interpretation of Eq. (2.13) is that it inserts a Dirac string or vortex [63–65] into spacetime. Note that Eq. (2.13) is a generalization of the center symmetry operator described in [54], which utilized temporal winding, and is also a special case of the family of symmetry operators constructed in [66].

The definition of $U_\alpha(\mathcal{S})$ in Eq. (2.13) may appear strange since the right-hand side is not manifestly a function of just the center element α. Rather, it depends on a color reference λ which has been chosen to exponentiate to α. Even worse, λ seems to specify an arbitrary vector in color space that naively violates gauge invariance. However, exactly like we saw for the twisted gauge transformation, the dependence on λ is actually spurious. The properties of $U_\alpha(\mathcal{S})$ are dictated entirely by its action on Wilson loops, which we will see depends only on α. Hence, any distinct choices of λ with the same twist α are physically equivalent.[12] We will see this borne out explicitly in the subsequent calculation.

The proof of the Ward identity is as follows. To begin, we apply the twisted gauge transformation defined in Eq. (2.7), under which the field strength becomes

$$F^a \;\mapsto\; (\Omega^{-1}F\,\Omega)^a + (\Omega^{-1}d d\Omega)^a\,. \tag{2.14}$$

As noted earlier, $dd\Omega = 0$ everywhere that $d\Omega$ is well-defined. However, there are regions of spacetime where it is ill-defined. Again, this is analogous to $d\phi$ in polar coordinates, which is closed, not exact, and ill-defined at the origin. Furthermore, it can be illuminating to consider a distributional interpretation of $dd\phi$ which has nontrivial delta function support at precisely at the origin, so the integral of it over a disc yields $\int_{D^2} dd\phi = \int_{S^1} d\phi = 2\pi$.

Similarly, for the case of our twisted gauge transformation, $dd\Omega$ has localized support on the surface $\mathcal{S}$ so in particular

$$(\Omega^{-1}d d\Omega)^a = 2\pi\lambda^a\,\delta(\mathcal{S}) \qquad \text{for any } \lambda^a \text{ such that} \qquad e^{2\pi\lambda} = \alpha \in Z(G)\,. \tag{2.15}$$

Here $\delta(\mathcal{S})$ is the two-form generalization of the Dirac delta distribution that peaks on $\mathcal{S}$ [59, 61], whose defining feature is that

$$\int_{\mathcal{M}} \varphi \wedge \delta(\mathcal{S}) = \int_{\mathcal{S}} \varphi\,, \tag{2.16}$$

for any two-form φ in $\mathcal{M}$.

[12] This is similar to what occurs in the case of instantons in gauge theory. These pure gauge configurations necessarily specify some explicit path in color space, and hence appear naively color breaking. However, only the topological winding number is the invariant label on these configurations.

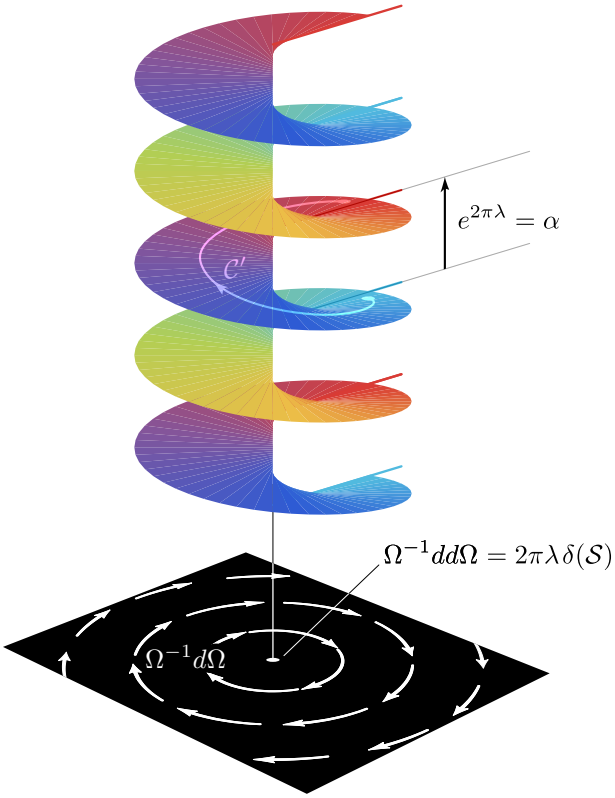


Figure 4. The twist of the multivalued gauge transformation can be characterized by the "curl" $\Omega^{-1}dd\Omega$ of the flat connection $\Omega^{-1}d\Omega$, which localizes along $\mathcal{S}$.

The equivalence between Eq. (2.8) and Eq. (2.15) can be understood intuitively from the visualization given in Fig. 4. The multivaluedness condition in Eq. (2.8) implies that the "gradient" $\Omega^{-1}d\Omega$ swirls about $\mathcal{S}$, and in turn, its "curl" $\Omega^{-1}dd\Omega$ localizes along $\mathcal{S}$ as a delta function, describing a Dirac string. A proof of this equivalence is left to App. B, which is essentially a colored generalization of our simpler example $dd\phi = 2\pi\,\delta(x)\delta(y)\,dx\wedge dy$. We will also provide explicit examples of Ω later on.

With this understanding, we observe that that Eq. (2.13) can be written as

$$U_\alpha(\mathcal{S}) = \exp\left(\frac{1}{g^2}\int_{\mathcal{M}} B_a \wedge 2\pi\lambda^a\delta(\mathcal{S})\right) = \exp\left(\frac{1}{g^2}\int_{\mathcal{M}} B_a \wedge (\Omega^{-1}dd\Omega)^a\right), \tag{2.17}$$

from which we can now revisit the left-hand side of the Ward identity in Eq. (2.12). The factor involving the action and the symmetry operator combine to give

$$e^{-S}\,U_\alpha(\mathcal{S}) = \exp\left(-\frac{1}{g^2}\int_{\mathcal{M}} B_a \wedge (F - \Omega^{-1}dd\Omega)^a - \frac{1}{2}\,B_a \wedge *B^a\right). \tag{2.18}$$

The twisted gauge transformation of the field strength in Eq. (2.14) eats up the $\Omega^{-1}dd\Omega$ term and *absorbs* the symmetry operator into the action, so

$$e^{-S}\,U_\alpha(\mathcal{S}) \;\mapsto\; e^{-S}\,. \tag{2.19}$$

Combing Eq. (2.5) and Eq. (2.19), we see that the left-hand side of the Ward identity in Eq. (2.12) transforms to

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\big\rangle \;\mapsto\; \int \mathcal{D}A\,\mathcal{D}B\; e^{-S}\rho(\alpha)^{\text{link}(\mathcal{C},\mathcal{S})}\, W_\rho(\mathcal{C})\,, \tag{2.20}$$

which proves the Ward identity in Eq. (2.11).

In summary, Eqs. (2.7) and (2.15) specify a change of field variables that *absorbs* the symmetry operator into the action as Eq. (2.19). Note that Eq. (2.19) clearly shows the action is *not invariant* under Eq. (2.7), as the Lagrangian four-form in Eq. (2.2) changes on the support of the surface $\mathcal{S}$. Hence we learn again explicitly that Eq. (2.7) is not a typical gauge transformation in Yang-Mills theory, which would leave the Lagrangian four-form at all points in spacetime invariant. Of course, if we excise the region $\mathcal{S}$ from spacetime, then the twisted gauge transformation in Eq. (2.7) becomes a bona fide gauge transformation in the resulting punctured manifold, as mentioned earlier in our discussion of Maxwell theory.

A few more remarks are in order. Firstly, it should be clear from the above logic that the twisted gauge transformation in Eq. (2.7) describes the action of the one-form symmetry operator $U_\alpha(\mathcal{S})$ on *arbitrary* operators. That is, if the twisted gauge transformation sends an operator $\mathcal{O}$ to $\mathcal{O}_\Omega$, then there is a corresponding generalized Ward identity,

$$\big\langle U_\alpha(\mathcal{S})\,\mathcal{O}\big\rangle = \big\langle \mathcal{O}_\Omega\big\rangle\,, \tag{2.21}$$

reiterating the fact that the one-form symmetry operator bridges between different center-twisted topological sectors of the gauge bundle. In fact, the Ward identity of Wilson loops in Eq. (2.11) can be regarded as a corollary of Eq. (2.21). Also, although Eq. (2.21) applies to local operators, it should be understood that any such point-supported operator is actually invariant under the one-form symmetry because the twisted gauge transformation can always be locally untwisted by a conventional gauge transformation. Geometrically, this reflects the fact that a point does not link with a codimension two object.

Secondly, the derivation of the Ward identity turns out to be very simple in the case of Maxwell theory with an exact contour $\mathcal{C}$. In this case the Wilson loop can be rewritten as a surface integral of the field strength F, as is familiar from the computation of the Aharonov-Bohm phase of a particle induced by a magnetic flux. The twisted gauge transformation of the field strength in Eq. (2.14) then creates a localized flux tube on the support of $\mathcal{S}$ whose integral over the surface yields the desired linking number. This is consistent with the above proof through a duality in the linking number computation described in App. B.

Thirdly, let us elaborate on the group multiplication rule for the symmetry operator, $U_{\alpha_1}(\mathcal{S})U_{\alpha_2}(\mathcal{S}) = U_{\alpha_1\alpha_2}(\mathcal{S})$, which is required axiomatically. Since our symmetry operator implements a twisted gauge transformation, the composition of two such transformations automatically yields a third, so we know a priori that the group composition law is valid. However, establishing this more directly in terms of the expression in Eq. (2.13) is more subtle. In particular, suppose $U_{\alpha_1}(\mathcal{S})$ and $U_{\alpha_2}(\mathcal{S})$ are realized with the color reference vectors λ_1 and λ_2. The product $U_{\alpha_1}(\mathcal{S})U_{\alpha_2}(\mathcal{S})$ naively exponentiates to symmetry operator with the color reference vector $\lambda_1 + \lambda_2$, which confusingly is not guaranteed to exponentiate

to a center element in general. However, this puzzle is resolved by realizing that $\lambda_1 + \lambda_2$ would correspond to a twisted gauge transformation with an irrational period. As described at length in Sec. 2.1.1 and in Fig. 3, the periods of the twisted gauge transformations must be center-valued in order for the symmetry operator to be *topological*. Otherwise line operators will not transform correctly.

In order to properly implement sequential twisted gauge transformations, the representative Lie algebra elements for $U_{\alpha_1}(\mathcal{S})$ and $U_{\alpha_2}(\mathcal{S})$ can be chosen as

$$\alpha_1 = e^{2\pi {\Omega_2}^{-1}\lambda_1\Omega_2} \qquad \text{and} \qquad \alpha_2 = e^{2\pi\lambda_2}\,. \tag{2.22}$$

These color references merge into a new one, ${\Omega_2}^{-1}\lambda_1\Omega_2 + \lambda_2$, which corresponds to the desired composite twisted gauge transformation,

$$\Omega = \Omega_1\Omega_2 \quad \Longrightarrow \quad \Omega^{-1}d\Omega = 2\pi({\Omega_2}^{-1}\lambda_1\Omega_2 + \lambda_2)^a\,\delta(\mathcal{S})\,. \tag{2.23}$$

This establishes the group composition law. In summary, the topological nature and gauge invariance of the symmetry operator together implies that the representative Lie algebra elements for the center elements in the group composition equation should be chosen in a specific form such that the composability and closure of center-twisted gauge transformations are correctly realized.

Last but not least, we realize that the above derivation of the Ward identity implies a remarkably simple and universal recipe for deducing the symmetry operator directly from the action itself. In particular, starting from any "BF"-type Lagrangian of the form $B_a \wedge F^a + f(B)$, we can define the symmetry operator as the object which is generated by a twisted gauge transformation. While the one-form symmetry operator is exactly eliminated by a twisted gauge transformation of the action, the $f(B)$ term will only serve as a spectator. This observation indeed is the key insight that will allow us to identify an explicit a one-form symmetry operator for gravity in Sec. 3.

2.1.3 Explicit Examples

As summarized in Eq. (2.21), the one-form symmetry of Yang-Mills theory is implemented by a twisted gauge transformation Ω that winds nontrivially with a mismatch valued in the center element $\alpha \in Z(G)$. In this section, we will explicitly construct some examples of Ω and apply them to various classical backgrounds. The resulting twisted backgrounds will reveal some illuminating physical interpretations for the one-form symmetry operator itself. For concreteness, we specialize in gauge group $G = \mathrm{SU}(N)$, whose center is $Z(G) = \mathbb{Z}_N$.

A. Symmetry Operator as Thin Solenoid

Suppose the spacetime is flat Euclidean space $\mathcal{M} = \mathbb{R}^4$, equipped with Cartesian coordinates $x^\mu = (x, y, z, t)$. Consider a twisted gauge transformation,

$$\Omega = \exp\left(\frac{k}{N}\,c\phi\right) \qquad \text{where} \qquad e^{2\pi c} = \mathbb{1}\,, \tag{2.24}$$

where ϕ is the azimuthal angle such that $\tan\phi = y/x$, and k is an integer. Here we have defined c to be an element of the Lie algebra of SU(N) that exponentiates to the identity via $e^{2\pi c} = \mathbb{1}$, so in the fundamental representation we might have [54]

$$c^i{}_j \,=\, i\,\mathrm{diag}(1,1,\cdots,-N+1)\,, \tag{2.25}$$

where $i, j, \ldots$ denote fundamental indices.

To demonstrate the transformation, let us consider a *trivial background* corresponding to $A^a = 0$, $B_a = 0$, where all the Wilson loops are trivial as $W_\rho(\mathcal{C}) = \dim\rho$, in particular $W_{\text{fund}}(\mathcal{C}) = N$ for the fundamental representation. However, applying Eqs. (2.7) and (2.14) on this trivial background with Ω in Eq. (2.24), we obtain a *nontrivial background*,

$$\begin{aligned} A^a &= (\Omega^{-1}d\Omega)^a = \frac{k}{N}\,c^a\,d\phi = \frac{k}{N}\,c^a\,\frac{x\,dy - y\,dx}{x^2+y^2} \\ F^a &= (\Omega^{-1}dd\Omega)^a = \frac{k}{N}\,c^a\,dd\phi = \frac{k}{N}\,2\pi c^a\,\delta(x)\delta(y)\,dx\wedge dy\,, \end{aligned} \tag{2.26}$$

with $B_a = 0$ still vanishing. For instance, consider a contour $\mathcal{C}$ that loops the z-axis once. Then the Wilson loop for $\mathcal{C}$, in the fundamental representation, is given by

$$\mathrm{tr}_{\text{fund}}\,\mathrm{P}\exp\left(\frac{k}{N}\,c\int_0^{2\pi} d\phi\right) = e^{2\pi ik/N}N\,, \tag{2.27}$$

in the twisted background described in Eq. (2.26). This demonstrates how the one-form symmetry transformation on Wilson loops arise from a twisted gauge transformation.

Interestingly, from the field strength in Eq. (2.26) we see that the resulting twisted field configuration describes a line of color flux flowing through the z-axis for all times t. Hence, we conclude that the physical interpretation of Ω is that it spontaneously excites an infinitely thin, straight and static solenoid from the vacuum. In turn, this implies that the one-form symmetry operator inserts a colored Dirac string into spacetime. With this interpretation, the Ward identity in Eq. (2.11) can be understood as the measurement of the Aharonov-Bohm phase by the Wilson loop in the background of the Dirac string. Each time the Wilson loop winds about this thin solenoid, we accrue an additional phase factor of $e^{2\pi ik/N}$, describing the center twist represented as a complex phase in the fundamental representation:

$$(\Omega(\phi=2\pi)\,\Omega^{-1}(\phi=0))^i{}_j \,=\, e^{2\pi ik/N}\,\delta^i{}_j\,. \tag{2.28}$$

A few remarks are in order. Firstly, it is worth noting that in Eq. (2.24) we can shift k by pN for $p\in\mathbb{Z}$, and this realizes various solenoids of different "strengths" but all yielding the same monodromy. This is an instantiation of a comment made earlier, which is that different Ω can realize the same α. Note also that a shift by pN can be implemented by a gauge transformation which is not continuously connected to the identity.

Secondly, since the linking between line and symmetry operators is topological, all of our results must be insensitive to homeomorphic deformations of their corresponding

integration surfaces. For this reason it is an amusing check to consider the twisted gauge transformation for a static but "wiggly" Dirac string,

$$\Omega = \exp\left(\frac{k}{N}\, c\phi\right) \qquad \text{where} \qquad \tan\phi = \frac{y - Y(z)}{x - X(z)}\,, \tag{2.29}$$

where $X(z)$ and $Y(z)$ describe a static line in space that is not necessarily straight. Here a simple calculation shows that the discontinuity in the twisted gauge transformation $\Omega^{-1} d d\Omega$ is proportional to

$$\delta(x - X(z))\, \delta(y - Y(z))\, d(x - X(z)) \wedge d(y - Y(z))\,, \tag{2.30}$$

which is a Dirac string that is not straight. It is obvious that the Aharonov-Bohm phase computed by the Ward identity is not modified by these wiggles. Going a step further, one can also promote the parameterization of the discontinuity to $X(z,t)$ and $Y(z,t)$ corresponding to time-dependent wiggles of the Dirac string. This case also accords with the general formula in Eq. (B.2).

In principle, the most general possible twisted gauge transformation can have a generator c that varies across spacetime. This variation in color space is perfectly possible and should also not alter the monodromies as long as it is properly derived form a multivalued transformation in accordance with Eq. (2.15).

Finally, let us take stock of the physical interpretation of the above calculation. The linking number between $U_\alpha(\mathcal{S})$ and $W_\rho(\mathcal{C})$ is typically interpreted as the center electric flux $U_\alpha(\mathcal{S})$ measured in the presence of the worldline of a colored particle given by $W_\rho(\mathcal{C})$. Interestingly, here we arrive at a dual, but completely equivalent picture: instead, $W_\rho(\mathcal{C})$ is the Aharonov-Bohm phase computed for a color Dirac string created by $U_\alpha(\mathcal{S})$.

Yet, clearly we have been cavalier about the global topology of $\mathcal{S}$ while demonstrating this example. In particular, because the thin solenoid extends off to infinity, we have not actually stipulated whether or how $\mathcal{S}$ "winds back" to form a closed surface. However, importantly, $\mathcal{S}$ must be closed in order for $U_\alpha(\mathcal{S})$ to be a topological operator. So why did the example of the solenoid yield the correct picture, despite the fact that the global structure of $\mathcal{S}$ was not specified?

Physically speaking, this setup yielded a sensible result because we effectively zoomed into a local region of $\mathcal{S}$ which links with $\mathcal{C}$ and measured the associated Aharanov-Bohm phase. That is, in the neighborhood of any point on $\mathcal{S}$, the surface appears as an infinite plane, and Ω is simply described by Eq. (2.24). As long as the Wilson loop does not deviate substantially from this region, the Aharanov-Bohm phase will be completely ignorant of how the ends of the flux tube reattach—or possibly even terminate—in some distant region. This is why we could obtain the correct transformation of the Wilson loop despite ignoring the global topology of $\mathcal{S}$.

Mathematically speaking, Eq. (2.24) should be understood as an expression for Ω in a certain patch on spacetime, which notably does not include the point at infinity. The details of "winding back" for the closure of $\mathcal{S}$ are contained in the charts for Ω which cover those other patches, which we have not defined explicitly. As a result, with the knowledge

of Ω in a single patch, Wilson loops are explicitly computable only when restricted to regions in this patch.

B. Symmetry Operator as Time Monodromy

Another interesting example is Yang-Mills theory at finite temperature, described by a compact product manifold with compactified Euclidean time,

$$\mathcal{M} = \mathcal{M}_3 \times \mathrm{S}^1\,, \quad t \sim t + \beta\,. \tag{2.31}$$

In addition to the trivial vacuum, there is an infinite set of gauge equivalence classes for the background gauge field. For example, consider

$$A^a = n\,\frac{2\pi c^a}{\beta}\,dt \qquad \text{where} \qquad n \in \mathbb{Z}\,. \tag{2.32}$$

In this background, Wilson loops winding about the thermal circle are trivial, as $e^{2\pi n c} = \mathbb{1}$. Meanwhile, consider the following twisted gauge transformation:

$$\Omega = \exp\left(\frac{2\pi k c}{N}\frac{t}{\beta}\right). \tag{2.33}$$

This maps the background considered in Eq. (2.32) to

$$A^a = \left(n + \frac{k}{N}\right)\frac{2\pi c^a}{\beta}\,dt\,, \tag{2.34}$$

so the Wilson loops gain a nontrivial phase factor of $e^{2\pi i k/N}$ per each thermal circle. Hence, the symmetry operator has induced a monodromy in the time direction.

Note that Eqs. (2.32) and (2.34) can be obtained by identifying the ends of a flat gauge field configuration in $\mathcal{M}_3$ times an interval with a twisted boundary condition. An implementation of this construction in Lorentzian signature can be found in [54].

An interesting feature of this example which is absent from the previous one is that the Wilson loops do not generally admit a coboundary. That is, they can be closed but not exact. However, the construction of the one-form symmetry in terms of a multivalued gauge transformation still applies.

Again, in the more rigorous sense Eq. (2.33) should be taken as the specification of Ω in a certain patch, say a ball in $\mathcal{M}_3$ times S^1. Then the surface support of the symmetry operator can reside at a time slice along the boundary of a large volume that goes beyond the ball.

C. Symmetry Operator as Circular Loop

In the examples considered thus far, the symmetry operator exhibited support on a surface $\mathcal{S}$ that has infinitely large extent in some direction, thus always leaving a worry that an explicit global definition of $\mathcal{S}$ as a closed surface is not given. For completeness, we would like to end with an example that explicitly shows how the surface support $\mathcal{S}$ can be finite.

Recall earlier how we constructed a static color Dirac string on the z-axis. The corresponding worldsheet extended infinitely in the t-z plane, so $\mathcal{S}$ was infinite. Here we will

temper $\mathcal{S}$ onto compact support in two steps. First, we will roll up the string in its spatial directions, yielding a closed circular loop of finite radius. Consequently, $\mathcal{S}$ will be spatially compact. Second, we will pinch off this loop in time by setting the size of this loop to be vanishing except for a finite duration, so $\mathcal{S}$ will be temporally compact as well.

In the first step of this construction, we consider a completely static system in *toroidal coordinates*, which foliate three-dimensional space according to a circular "reference ring" of radius a in the x-y plane. In particular, the coordinates (τ, σ, ϕ) are defined by

$$x = \frac{a \sinh \tau}{\cosh \tau - \cos \sigma} \cos \phi \,, \quad y = \frac{a \sinh \tau}{\cosh \tau - \cos \sigma} \sin \phi \,, \quad z = \frac{a \sin \sigma}{\cosh \tau - \cos \sigma} \,, \tag{2.35}$$

where ϕ is the azimuthal angle in the x-y plane. Surfaces of fixed $\tau \geq 0$ label concentric two-tori which enclose the reference ring, while surfaces of fixed $-\pi < \sigma \leq \pi$ label two-spheres which intersect the reference ring. Also, here we choose the branch cut for σ such that the discontinuity at $\sigma = \pm\pi$ develops on the disc enclosed by the reference ring.

Crucially, we can think of σ as an angular coordinate that winds like a solenoid about the reference ring. Thus we can let the twisted gauge transformation parameter be

$$\Omega = \exp\left(\frac{k}{N} c \sigma \right) , \tag{2.36}$$

which clearly induces a rephasing for any holonomy that wraps the reference ring. The branch cut resides on a volume $\mathcal{V}$ corresponding to the static disc enclosed by the reference ring. Its boundary then defines the surface $\mathcal{S} = \partial \mathcal{V}$, which is the reference ring itself, namely a static loop of radius a in the x-y plane. Furthermore, we see that Ω correctly approaches the identity at spatial infinity, simply because spatial infinity corresponds to $\sigma = 0$ in toroidal coordinates.

In the second step, we allow for the radius of the reference ring to change with time. To allow for this, we define toroidal coordinates for each time slice in which the reference ring has a time-varying radius $a(t)$. For example, let us define $a(t)$ to smoothly increase from and decrease to zero within a time interval $t \in [t_1, t_2]$. With this temporal modification, the surface $\mathcal{S}$ has finite support in both time and space.

Let us end with a final remark. In general, one typically wants to construct a twisted gauge transformation parameter Ω for an arbitrarily shaped surface $\mathcal{S}$ in an arbitrary manifold $\mathcal{M}$ with or without boundaries. How do we know that such an Ω always exists, given some choice of $\mathcal{S}$ and $\mathcal{M}$? In all of the examples above we started with Ω as an input and rather determined the surface $\mathcal{S}$ as an output.

Interestingly, we find that it is always possible to find an Ω with a given $\mathcal{S}$ and $\mathcal{M}$, on account of a closely analogous question in *classical magnetostatics*. That is, deducing Ω from $\mathcal{S}$ and $\mathcal{M}$ is mathematically identical to deducing the static magnetic field and potential of an electric current loop. To see why, imagine we are experimentalists who construct a loop of electrical line current $\mathbf{J}$, built to specification according to some arbitrary contour. On account of Ampère's law, $\nabla \times \mathbf{H} = \mathbf{J}$, we can then deduce the magnetic field $\mathbf{H}$, or even just measure it. In regions away from the current, we can then reconstruct a magnetic scalar potential via $\mathbf{H} = -\nabla \Psi$. If the manifold $\mathcal{M}$ is not closed, then we can

make its boundary $\partial\mathcal{M}$ superconducting to enforce the boundary condition $\mathbf{H}_\perp = 0$, in which case Ψ can be set to a constant over $\partial\mathcal{M}$ that we fix to zero. In this analogy, the electric current $\mathbf{J}$, the magnetic field $\mathbf{H}$, and the magnetic potential Ψ each correspond to the Dirac string defined by $\Omega^{-1}dd\Omega$, the twisted gauge connection $\Omega^{-1}d\Omega$, and the "log" of the twisted gauge parameter Ω, respectively .

2.2 Canonical Formalism

The one-form symmetry of gauge theory can also be understood from the complementary point of view of the Hamiltonian formalism. To this end, we will study Yang-Mills theory on a spacetime described by a product manifold $\mathcal{M} = \mathcal{M}_3 \times \mathbb{R}$ equipped with coordinates $x^\mu = (x^i, x^4)$. In particular, we perform a $3+1$ decomposition in which x^i denotes coordinates of the spatial three-manifold $\mathcal{M}_3$ and $x^4 = t$ defines equal-time slices.

2.2.1 Phase Space

The first-order formulation of Yang-Mills theory is defined in Eq. (2.2). Writing out all indices explicitly, we obtain the Lagrangian,

$$\frac{1}{g^2}\left[\frac{1}{2}B_{a\mu\nu}\Big(\partial_\rho A^a{}_\sigma + \tfrac{1}{2}f^a{}_{bc}A^b{}_\rho A^c{}_\sigma\Big)\epsilon^{\mu\nu\rho\sigma} - \frac{1}{4}B_{a\mu\nu}B^{a\mu\nu}\right], \tag{2.37}$$

where $\epsilon^{\mu\nu\rho\sigma}$ denotes the permutation symbol. Carrying out the $3+1$ decomposition and integrating out B_{ai4}, we immediately see that the dynamical coordinates on the phase space are $A^a{}_i$ together with their canonical conjugates,

$$E^i{}_a = \frac{1}{2}B_{ajk}\,\epsilon^{ijk}\,. \tag{2.38}$$

Their canonical commutation relations given by

$$\begin{aligned}
&[A^a{}_i(x), A^b{}_j(x')] = 0\,,\\
&[E^i{}_a(x), A^b{}_j(x')] = g^2\,\delta^i{}_j\,\delta^b{}_a\,\delta^{(3)}(x-x')\,,\\
&[E^i{}_a(x), E^j{}_b(x')] = 0\,,
\end{aligned} \tag{2.39}$$

where $x, x' \in \mathcal{M}_3$ are points in the spatial manifold. Note that there is no imaginary unit here since we are in Euclidean signature. The phase space is also equipped with the Gauss constraint and a Hamiltonian, the details of which are not important for our purposes.

2.2.2 Ward Identity

In the language of the path integral, a one-form symmetry transformation is implemented through the insertion of a symmetry operator which wraps the line operator. In the operator formalism, however, this corresponds to a conjugation of the latter by the former. To see how this works in detail, consider a line operator $W_\rho(\mathcal{C})$, where $\mathcal{C}$ is restricted to an equal-time slice, say at $t = 0$. As before, we take the symmetry operator $U_\alpha(\mathcal{S})$ to be supported on an exact surface $\mathcal{S}$ with an associated coboundary $\mathcal{V}$, so $\mathcal{S} = \partial\mathcal{V}$.

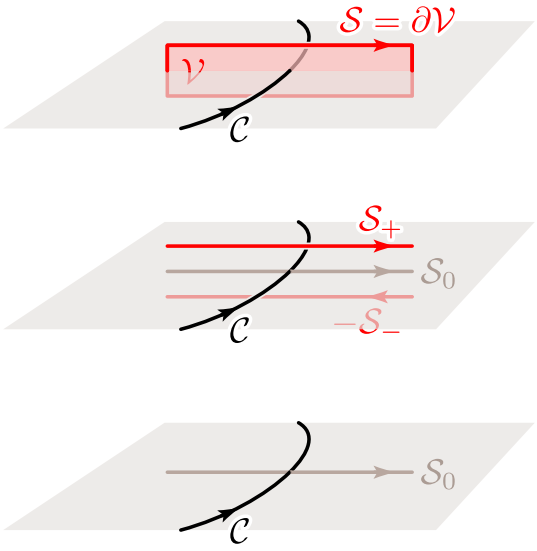


Figure 5. Pancaking the symmetry operator on a purely spatial line operator $\mathcal{C}$. The plane depicts an equal-time slice while time flows upwards. *Top:* The equal-time slice bisects the volume $\mathcal{V}$ by a surface $\mathcal{S}_0$. *Middle:* The surface $\mathcal{S} = \partial\mathcal{V}$ splits into surfaces $\mathcal{S}_+$ and $-\mathcal{S}_-$ at the infinitesimal future and past. *Bottom:* $\mathcal{S}_+$ and $\mathcal{S}_-$ both project down to the surface $\mathcal{S}_0$.

The geometric set-up is depicted in Fig. 5. We assume that the surface $\mathcal{S}$ *links* once with the purely spatial loop $\mathcal{C}$. Consequently, the coboundary $\mathcal{V}$ is *intersected* by the loop $\mathcal{C}$ and *bisected* by the spatial slice. Now imagine continuously squashing or pancaking the coboundary $\mathcal{V}$ along the time direction such that its two-dimensional boundary $\mathcal{S}$ infinitesimally hugs the spatial slice. In this limit, $\mathcal{S} = \mathcal{S}_+ \cup (-\mathcal{S}_-)$ is the union of two disjoint discs $\mathcal{S}_+$ and $-\mathcal{S}_-$ at the infinitesimal future and past across $t = 0$. Once $\mathcal{V}$ has completely collapsed into the spatial slice at $t = 0$, both discs $\mathcal{S}_\pm$ approach the same surface, which we denote by $\mathcal{S}_0$. From Fig. 5, we see that this $\mathcal{S}_0$ will be the intersection between $\mathcal{V}$ and $t = 0$. As a result, the intersection of $\mathcal{C}$ and $\mathcal{V}$ in the four-manifold $\mathcal{M}$ is equivalent to intersection of $\mathcal{C}$ and $\mathcal{S}_0$ in the three-manifold $\mathcal{M}_3$ as the slice $t = 0$. Therefore, we have in general

$$\mathrm{link}(\mathcal{C}, \mathcal{S}) = \mathrm{int}(\mathcal{C}, \mathcal{V}) = \mathrm{int}_3(\mathcal{C}, \mathcal{S}_0)\,, \tag{2.40}$$

where int_3 denotes intersection number in $\mathcal{M}_3$.

Now, we can describe how this pancaking procedure boils down the Ward identity to an equal-time operator equation. Since $\mathcal{S} = \mathcal{S}_+ \cup (-\mathcal{S}_-)$, we see that the symmetry operator factorizes into $U_\alpha(\mathcal{S}) = U_\alpha(\mathcal{S}_+) U_\alpha(-\mathcal{S}_-)$.[13] In turn, the left-hand side of the Ward identity in Eq. (2.11) becomes the time-ordered expression $U_\alpha(\mathcal{S}_+) W_\rho(\mathcal{C}) U_\alpha(-\mathcal{S}_-)$, which in the process of pancaking limits to the equal-time operator product $U_\alpha(\mathcal{S}_0) W_\rho(\mathcal{C}) U_\alpha^{-1}(\mathcal{S}_0)$.

[13] Note that here we have allowed nonclosed surfaces for the support of symmetry operators, which might be a slight abuse of notation.

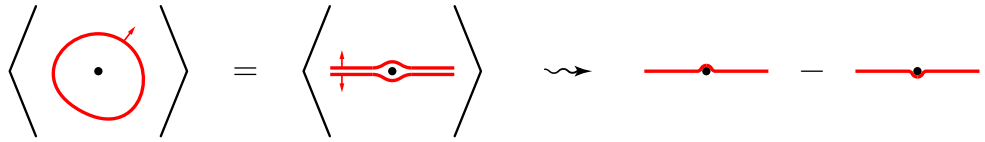


Figure 6. Zero-form version of the pancaking procedure, which the reader may be familiar with. Wrapping an operator with the symmetry operator translates to an equal-time conjugation, where operator ordering traces back to time ordering.

Using Eq. (2.40), we then find that the Ward identity translates to

$$U_\alpha(\mathcal{S}_0)\, W_\rho(\mathcal{C})\, U_\alpha^{-1}(\mathcal{S}_0) \;=\; \rho(\alpha)^{\mathrm{int}_3(\mathcal{C},\mathcal{S}_0)}\, W_\rho(\mathcal{C})\,, \tag{2.41}$$

which is an equal-time operator equation.

Eq. (2.41) is the avatar of the Ward identity in the operator formalism, where all the relevant geometric objects and operations, including the disc $\mathcal{S}_0$ and the closed contour $\mathcal{C}$, are defined within the three-manifold $\mathcal{M}_3$. The key insight here is that time ordering in the path integral formalism turns into operator ordering in the operator formalism.

Finally, it is straightforward to explicitly evaluate Eq. (2.41). Applying the $3+1$ decomposition and using Eq. (2.38), the Hamiltonian formalism avatar of the symmetry operator, supported on a disc $\mathcal{S}_0 \subset \mathcal{M}_3$ in three dimensions, is given as

$$U_\alpha(\mathcal{S}_0) = \exp\left(\frac{\pi}{g^2}\int_{\mathcal{S}_0} dx^j \wedge dx^k\, \epsilon_{ijk}\, E^i{}_a\, \lambda^a\right) \qquad \text{where} \qquad e^{2\pi\lambda} = \alpha\,. \tag{2.42}$$

To compute $U_\alpha(\mathcal{S}_0) W_\rho(\mathcal{C}) U_\alpha^{-1}(\mathcal{S}_0)$, let us first deduce how the conjugation acts on the phase space. While $E^i{}_a(x)$ is left invariant because $[E^i{}_a(x), E^j{}_b(x')] = 0$, the spatial gauge connection $A^a{}_k(x)$ has a nonvanishing commutator with the exponent of Eq. (2.42),

$$\begin{aligned} &\frac{\pi}{g^2}\int_{\mathcal{S}_0} dx'^i \wedge dx'^j\, \lambda^b(x')\, \epsilon_{ijl}\, [E^l{}_b(x'), A^a{}_k(x)] \\ &= 2\pi \int d\sigma_1 d\sigma_2\, \frac{\partial X^i}{\partial\sigma_1}\frac{\partial X^j}{\partial\sigma_2}\, \lambda^a(X)\, \epsilon_{ijk}\, \delta^{(3)}(x-X)\,, \end{aligned} \tag{2.43}$$

where $(\sigma_1,\sigma_2) \mapsto X^i(\sigma_1,\sigma_2)$ is a parameterization of the surface $\mathcal{S}_0$. According to Eq. (B.2), this describes the components of the one-form $2\pi\lambda^a\,\delta_3(\mathcal{S}_0)$, where $\delta_3(\mathcal{S}_0)$ is the Dirac delta one-form of $\mathcal{S}_0$ defined in the spatial three-manifold $\mathcal{M}_3$. Therefore, we can summarize the action of the symmetry transformation on the phase space variables as the following:

$$\begin{aligned} &U_\alpha(\mathcal{S}_0)\, A^a{}_i\, U_\alpha^{-1}(\mathcal{S}_0) = (A^a + 2\pi\lambda^a\,\delta_3(\mathcal{S}_0))_i\,, \\ &U_\alpha(\mathcal{S}_0)\, E^i{}_a\, U_\alpha^{-1}(\mathcal{S}_0) = E^i{}_a\,. \end{aligned} \tag{2.44}$$

As a result, we find that the Wilson loop transforms as

$$U_\alpha(\mathcal{S}_0)\, W_\rho(\mathcal{C})\, U_\alpha^{-1}(\mathcal{S}_0) \;=\; \rho(\alpha)^{\mathrm{int}_3(\mathcal{C},\mathcal{S}_0)}\, W_\rho(\mathcal{C})\,, \tag{2.45}$$

which proves the Hamiltonian counterpart of the Ward identity. We have used the fact that the three-dimensional intersection number is is $\mathrm{int}_3(\mathcal{C},\mathcal{S}_0) = \oint_{\mathcal{C}} \delta_3(\mathcal{S}_0)$.

3 Gravity

Armed with an understanding of higher-form symmetry in gauge theory, we are now equipped to transcribe all of those results to the context of dynamical gravity. As is well-known, the tetradic Palatini formalism is a description of gravity in terms of a gauge theory of Lorentz transformations. Using this framework, we can deduce the higher-form symmetries of gravity by direct analogy. As before, we start with a covariant analysis and then describe the same physics using the canonical formalism.

3.1 Covariant Formalism

Consider dynamical gravity on a four-dimensional manifold $\mathcal{M}$ with Euclidean signature. As noted earlier, we work within the regime of validity of an effective field theory of gravity in which the topology and dimensionality of spacetime do not fluctuate.

Our point of departure is the tetradic Palatini formalism, which formulates gravity as a gauge theory[14] of the four-dimensional Lorentz group G and is inherently first-order. Here we emphasize again that despite our abuse of nomenclature we will consider both Lorentzian and Euclidean signature. The degrees of freedom are a one-form spin connection and one-form tetrad field,

$$\omega^{AB} = \omega^{AB}{}_{\mu}\, dx^{\mu} \qquad \text{and} \qquad e^{A} = e^{A}{}_{\mu}\, dx^{\mu}\,, \tag{3.1}$$

which transform in the adjoint and fundamental representation of G, so the uppercase indices are $A, B, \ldots \in \{1, 2, 3, 4\}$.

In terms of these fields, the action for Palatini gravity is given by the integral over $\mathcal{M}$ of the Lagrangian four-form

$$L = \frac{1}{4g^2}\,\epsilon_{ABCD}\, e^{A} \wedge e^{B} \wedge R^{CD} - \frac{\Lambda}{24g^2}\,\epsilon_{ABCD}\, e^{A} \wedge e^{B} \wedge e^{C} \wedge e^{D}\,, \tag{3.2}$$

where we have defined the Riemann curvature two-form,

$$R^{A}{}_{B} = d\omega^{A}{}_{B} + \omega^{A}{}_{C} \wedge \omega^{C}{}_{B}\,, \tag{3.3}$$

which is simply the field strength for the spin connection. The first term in Eq. (3.2) encodes the Einstein-Hilbert Lagrangian, where we have repackaged Newton's constant into $g = (8\pi G_{\mathrm{N}})^{1/2}$ to draw a closer analogy with gauge theory. The second term in Eq. (3.2) defines the cosmological constant Λ.

It should be reiterated that both the spin connection and tetrad are taken as *independent* degrees of freedom in the tetradic Palatini formulation. To see how this approach reproduces conventional general relativity, let us vary the Lagrangian in Eq. (3.2) with respect to the spin connection and tetrad to obtain their respective equations of motion,

$$D(e^{A} \wedge e^{B}) = 0 \qquad \text{and} \qquad e^{[B} \wedge R^{CD]} = \frac{\Lambda}{3}\, e^{B} \wedge e^{C} \wedge e^{D}\,, \tag{3.4}$$

[14] Here were refer to gauge theory in the restricted sense of a construction based on principal fiber bundles. While diffeomorphisms are of course a redundancy of description, they do not define a gauge theory in this restricted sense because diffeomorphisms inherently shift the base points.

where the square brackets on indices denote antisymmetrization. Here we have defined D as the covariant exterior derivative with respect to the spin connection.

It is not difficult to show that the first equation of motion in Eq. (3.4) is *algebraically* equivalent to vanishing of $De^A = de^A + \omega^A{}_B \wedge e^B$, which is the definition of torsion. This fixes the dynamics to be identically torsion-free. Any coupling of the spin connection to an external source generates nonzero torsion precisely only on the support of that source. Meanwhile, with vanishing torsion the second equation of motion in Eq. (3.4) becomes the Einstein field equations for the associated metric,

$$g_{\mu\nu} = \delta_{AB}\, e^A{}_\mu\, e^B{}_\nu \,. \tag{3.5}$$

As emphasized earlier, we work in an effective field theory description of gravity which describes gravitons propagating over a fixed background. Hence, throughout our analysis the metric and tetrad are implicitly expanded as fluctuations about some choice of background values $\bar{g}_{\mu\nu}$ and $\bar{e}^A{}_\mu$, respectively, though it will usually be simpler to manipulate the full field variables rather than their fluctuations.[15]

Using integration by parts, it is easy to rewrite the Lagrangian in Eq. (3.2) so that it does not contain derivatives of the spin connection. Consequently, the spin connection is an auxiliary field that can be eliminated at tree-level by plugging in the classical solution for $\omega^A{}_B$ in terms of e^A using the torsion-free condition in Eq. (3.4). The resulting Lagrangian, which depends solely on the tetrad, is precisely the usual Einstein-Hilbert Lagrangian. Therefore, tetradic Palatini gravity is classically equivalent to general relativity, provided there are no sources that couple directly to the spin connection so that torsion is vanishing. As noted earlier, quantum equivalence also follows, since the effective field theory of a massless spin two particle is unique, modulo Wilson coefficients.

Next, we would like to recall the well-known fact that tetradic Palatini gravity can be expressed in a way that even more directly parallels the first-order formulation of Yang-Mills theory. This fact will play a crucial role in our identification of the one-form symmetry later on. In particular, let us define the Plebański two-form [67, 68] by

$$B_{AB} = \frac{1}{2}\,\epsilon_{ABCD}\, e^C \wedge e^D \,, \tag{3.6}$$

which is valued in the adjoint of G. Expressed in terms of B_{AB}, the tetradic Palatini Lagrangian in Eq. (3.2) becomes

$$L = \frac{1}{g^2}\, B_a \wedge R^a - \frac{\Lambda}{6g^2}\, B_a \wedge \star B^a \,, \tag{3.7}$$

[15] In any sensible effective field theory, the background spacetime is nondegenerate and thus the background tetrad $\bar{e}^A{}_\mu$ must be nonzero. This means the vacuum breaks diffeomorphism invariance and local Lorentz symmetry down to the diagonal, which naively hinders our analysis. However, our calculation of the Ward identity for the symmetry and line operators utilize the full tetrad field $e^A{}_\mu$, which transforms covariantly, so there is no additional complication. The very same phenomenon occurs in gauge theory, where expanding about a background gauge field $\bar{A}^a{}_\mu$ technically breaks Lorentz invariance and color symmetry down to the diagonal, but of course with no effect on the Ward identities in the theory.

which is by construction identical in form to the Yang-Mills Lagrangian in Eq. (2.2). Here the lowercase indices $a, b, \ldots \in \{1, 2, 3, 4, 5, 6\}$ transform in the adjoint of G, and $\star$ denotes the Hodge star in the *internal* Lorentz space, so when the adjoint indices are all converted to fundamental indices by the Lorentz algebra generator $(t_a)^A{}_B$, we have

$$B_{AB} = \frac{1}{2}\epsilon_{ABCD}\, e^C \wedge e^D \qquad \text{and} \qquad \star B^{AB} = e^A \wedge e^B \,. \tag{3.8}$$

We emphasize that $\star$ should be distinguished from $*$ in Eq. (2.2). Also, we clarify that fundamental indices $A, B, \cdots$ are raised and lowered by the Euclidean flat metric δ_{AB}.

In terms of the Plebański two-form, the Lagrangian in Eq. (3.7), is clearly of the "BF"-type Lagrangian of the form $B_a \wedge F^a + f(B)$ which is familiar from gauge theory. Therefore, we can immediately identify the line operator and the symmetry operator for the one-form symmetry of gravity by essentially copying the formulae from our earlier discussion about Yang-Mills theory.

So what is the one-form symmetry of dynamical gravity? In analog with gauge theory, it is defined by the center of G, which is the four-dimensional Lorentz group. As usual, the center depends crucially on the global structure of G, which is not specified by the Lagrangian in Eq. (3.7). Therefore, it is essential to clarify the global structure of G before we can continue further.

To begin, let us consider the case of Euclidean signature. Given the well-known Lie algebra isomorphism $\mathfrak{so}(4) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2) \cong \mathfrak{spin}(4)$, we can choose G to be either[16]

$$\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2) \qquad \text{or} \qquad \mathrm{SO}(4) \cong \frac{\mathrm{Spin}(4)}{\mathbb{Z}_2} \qquad \text{or} \qquad \frac{\mathrm{SO}(4)}{\mathbb{Z}_2} \cong \frac{\mathrm{Spin}(4)}{\mathbb{Z}_2 \times \mathbb{Z}_2} \,, \tag{3.9}$$

whose center subgroups are given by

$$\mathbb{Z}_2 \times \mathbb{Z}_2 \qquad \text{or} \qquad \mathbb{Z}_2 \qquad \text{or} \qquad \mathbb{1} \,, \tag{3.10}$$

respectively. The one-form charges will be valued in these center subgroups. Note that the zero-form symmetry associated with the center $\mathbb{Z}_2 \times \mathbb{Z}_2$ of $\mathrm{Spin}(4) \cong \mathrm{SU}(2) \times \mathrm{SU}(2)$ acts as parity on chiral and antichiral spinor indices, while that of the center $\mathbb{Z}_2$ of $\mathrm{SO}(4)$ acts as parity on vector indices.

Meanwhile, in Lorentzian signature, possible candidates for the gauge group are

$$\mathrm{Spin}(3,1) \cong \mathrm{SL}(2, \mathbb{C}) \qquad \text{or} \qquad \mathrm{SO}^+(3,1) \cong \frac{\mathrm{Spin}(3,1)}{\mathbb{Z}_2} \,, \tag{3.11}$$

where the latter is the orthochronous Lorentz group and the former is its double cover. The corresponding center subgroups are

$$\mathbb{Z}_2 \qquad \text{or} \qquad \mathbb{1} \,, \tag{3.12}$$

[16] A priori, one can also consider semi-spin groups such as $\mathrm{SemiSpin}(4) \simeq \mathrm{SU}(2) \times \mathrm{SO}(3)$, which is a nonstandard quotient. However, this group does not admit a vector representation so it is not compatible with the tetrad formalism.

respectively, which define allowed one-form symmetry of gravity in Lorentzian signature. Here the zero-form symmetry associated with the former corresponds to net parity on chiral and antichiral spinor indices, also known as fermion parity.

Any choice of G in Eq. (3.9) is viable. For the sake of generality, however, in the remainder of our analysis we will be agnostic and take the gauge group to be some general G with center subgroup $Z(G)$.

3.1.1 Line and Symmetry Operators

We are now equipped to derive explicitly the one-form symmetry of dynamical gravity. As before, we start with identifying the line operator and its symmetry transformation.

Firstly, let us recall the direct parallel between the spin connection in Eq. (3.7) and the gauge connection in Eq. (2.2). As such, it is obvious that the natural line operator in gravity is the spin holonomy,

$$W_\rho(\mathcal{C}) = \mathrm{tr}_\rho \, \mathrm{P}\exp\left(\oint_{\mathcal{C}} \omega\right), \tag{3.13}$$

where $\mathcal{C}$ defines a closed one-dimensional contour and ρ is some spin representation of the Lorentz group G. Far less clear a priori is the identity of the symmetry operator,

$$U_\alpha(\mathcal{S}), \tag{3.14}$$

other than that it should be labeled by a center element $\alpha \in G$ and defined on an exact two-dimensional surface $\mathcal{S} = \partial\mathcal{V}$ that can topologically link with $\mathcal{C}$.

In perfect analogy with the Wilson loop of gauge theory, we expect that the spin holonomy should transform as

$$W_\rho(\mathcal{C}) \;\mapsto\; \rho(\alpha)^{\mathrm{link}(\mathcal{C},\mathcal{S})}\, W_\rho(\mathcal{C}), \tag{3.15}$$

under the one-form symmetry of gravity, and this is indeed the case. As before, $\mathrm{link}(\mathcal{C},\mathcal{S})$ is defined to be the linking number between the contour and surface that define the line and symmetry operators.[17]

Mirroring Eq. (2.7) in gauge theory, the one-form symmetry of gravity is implemented as a transformation of the fields by a closed but not exact form. In the gravitational context, the appropriate map is a local Lorentz transformation that is multivalued. In particular, we consider the case where $\mathcal{S} = \partial\mathcal{V}$ and there is a branch cut on the coboundary $\mathcal{V}$ whose discontinuity is center-valued. Physically, this twisted Lorentz transformation boosts local

[17] As noted previously, we work in an effective field theory of gravity in which the topology and dimensionality of spacetime are robust. Furthermore, the linking number is invariant under any invertible diffeomorphism that is continuously connected to the identity. To see why, consider a putative family of diffeomorphisms labeled by a parameter $\tau \in [0,1]$ such that $\tau = 0$ is the identity and $\tau = 1$ unlinks the surfaces. By continuity, there exists some τ for which the corresponding diffeomorphism results in surfaces which intersect at a point. In this case the diffeomorphism is not invertible, since it maps two points, one on each surface, to a single point.

laboratories in spacetime such that a π rotation of frames is applied after each winding about $\mathcal{S}$. Specifically, the spin connection and the tetrad will transform as

$$\omega^A{}_B \;\mapsto\; (\Omega^{-1})^A{}_C\,\omega^C{}_D\,\Omega^D{}_B + (\Omega^{-1})^A{}_C\,d\Omega^C{}_B \qquad \text{and} \qquad e^A \;\mapsto\; (\Omega^{-1})^A{}_B\,e^B\,, \tag{3.16}$$

from which the transformation of the Plebański two-form is given by

$$B^A{}_B \;\mapsto\; (\Omega^{-1})^A{}_C B^C{}_D\,\Omega^D{}_B\,, \tag{3.17}$$

where Ω is a multivalued zero-form which defines the twisted Lorentz transformation. Like in the case of gauge theory, $d\Omega$ is not exact because Ω carries winding. As before, we also impose a discontinuity condition across the branch cut, given by

$$\lim_{\mathcal{P}_\pm \to \mathcal{P}} \Omega(\mathcal{P}_+)\,\Omega^{-1}(\mathcal{P}_-) \;=\; \alpha \qquad \text{where} \qquad \alpha \in Z(G) \qquad \text{and} \qquad \mathcal{P} \subset \mathcal{V}\,, \tag{3.18}$$

where $\mathcal{P}_+$ and $\mathcal{P}_-$ are infinitesimally split across $\mathcal{V}$. In turn, it follows that the spin holonomy that links once with $\mathcal{S}$ maps to

$$\begin{aligned} W_\rho(\mathcal{C}) \;&\mapsto\; \lim_{\mathcal{C}' \to \mathcal{C}} \mathrm{tr}_\rho \left[\mathrm{P}\exp\left(\int_{\mathcal{C}'} \Omega^{-1}\omega\,\Omega + \Omega^{-1} d\Omega \right) \right], \\ &=\; \lim_{\mathcal{C}' \to \mathcal{C}} \mathrm{tr}_\rho \left[\Omega^{-1}(\mathcal{P}_-)\,\mathrm{P}\exp\left(\int_{\mathcal{C}'} \omega \right) \Omega(\mathcal{P}_+) \right] \\ &=\; \lim_{\mathcal{C}' \to \mathcal{C}} \mathrm{tr}_\rho \left[\alpha\,\mathrm{P}\exp\left(\int_{\mathcal{C}'} \omega \right) \right] \;=\; \rho(\alpha)\,W_\rho(\mathcal{C})\,, \end{aligned} \tag{3.19}$$

which exactly instantiates Eq. (3.15). Thus, we conclude that the twisted local Lorentz transformation in Eq. (3.16) implements the gravitational one-form symmetry.

As in the gauge theory case, we emphasize here that Ω is not a bona fide gauge transformation when the center twist α is nontrivial. In particular, drawing an analogy with Eq. (2.15), we see that the double exterior derivative of Ω is nonzero,

$$(\Omega^{-1})^A{}_C\,dd\Omega^C{}_B = 2\pi\lambda^A{}_B\,\delta(\mathcal{S}) \qquad \text{for any } \lambda^A{}_B \text{ such that} \qquad e^{2\pi\lambda} = \alpha \in Z(G)\,, \tag{3.20}$$

and in fact has nonvanishing support precisely on $\mathcal{S}$.

Before continuing, let us address some possible confusions relating to Lorentz and diffeomorphism invariance. First of all, just like in Yang-Mills theory, we see here that the twisted Lorentz transformation depends on the whole function Ω rather than just α. Naively, this dependence is Lorentz-violating, since Ω defines a trajectory in the space of Lorentz transformations. However, just as before, we can see that this dependence is spurious since different choices for Ω which exhibit the same twist α still act indistinguishably on the spin holonomy, and are thus physically equivalent.

Secondly, in the presence of dynamical gravitation there is a further caveat regarding the *diffeomorphism invariance* of the line operator $W_\rho(\mathcal{C})$ and symmetry operator $U_\alpha(\mathcal{S})$. Although these objects do not carry dangling indices, they do depend on a particular choice of a contour $\mathcal{C}$ and surface $\mathcal{S}$, which define collections of points in spacetime in the very

same way that a local operator $\mathcal{O}(x)$ defines a single point. However, any local or quasi-local object such as a point, curve, or surface in spacetime is famously not diffeomorphism invariant, simply because "x" itself is not diffeomorphism invariant. Note that this annoyance is also implicitly present in any discussion of gauge theory Wilson loops in the presence of gravity, which is central to discussions of swampland conjectures. To address this, one typically appeals to the restoration of diffeomorphism invariance by "gravitationally dressing" [69, 70] the operator in question. A closely related tactic is to define all positions "relationally" in terms of some asymptotic reference, either at spatial infinity or the beginning of time.

However, diffeomorphism invariance is restored here in the very same way as Lorentz invariance. Since the linking number is itself diffeomorphism invariant, it means that operators related to each other by a diffeomorphism are themselves are physically equivalent.

3.1.2 Ward Identity

Next, let us compute the Ward identity associated with the gravitational one-form symmetry. Taking inspiration from Eq. (2.13) in the case of gauge theory, we define the one-form symmetry operator of gravity to be

$$U_\alpha(\mathcal{S}) = \exp\left(\frac{\pi}{g^2}\int_{\mathcal{S}} \lambda^{AB} B_{AB}\right) \qquad \text{where} \qquad e^{2\pi\lambda} = \alpha \in Z(G)\,, \tag{3.21}$$

where λ is a zero-form function which is chosen so that at all points in spacetime it exponentiates to a center element of the Lorentz group. Again, this formula should be understood as a realization of the symmetry operator $U_\alpha(\mathcal{S})$ with a representative Lie algebra element λ for α, as all choices of λ with the same twist α act the same on the spin holonomy and are thus physically equivalent.

Intriguingly, this surface operator $U_\alpha(\mathcal{S})$ literally computes a certain area-like quantity associated with $\mathcal{S}$! In particular, by reverting to tetrad variables and reintroducing Newton's constant, we find that the symmetry operator is

$$U_\alpha(\mathcal{S}) = \exp\left(\frac{1}{4G_{\rm N}}\int_{\mathcal{S}} \frac{1}{2}\star\lambda_{AB}\,(e^A\wedge e^B)\right), \tag{3.22}$$

where $\star\lambda_{AB} = \frac{1}{2}\epsilon_{ABCD}\,\lambda^{CD}$ is the Hodge dual of λ^{AB}. Here we recognize $e^A\wedge e^B$ as the infinitesimal area element in the orthonormal frame, so the exponent computes the area smeared with a reference $\star\lambda_{AB}$, measured in Planck units. Note that $1/2$ is the canonical normalization factor for contracting antisymmetric tensors.

Additionally, it is curious that the exponent in Eq. (3.22), or equivalently in Eq. (3.59), is tantalizingly similar to the area operator in loop quantum gravity which leads to the quantization of tetrahedral volume [71, 72]. The only crucial difference is that here the area is "dotted" with a Lorentz generator λ^{AB} that exponentiates to the center, so in a sense it serves as a discrete and topological reincarnation of the area operator in loop quantum gravity. On the other hand, $U_\alpha(\mathcal{S})$ is also reminiscent of the Bekenstein-Hawking entropy formula [73, 74]. Of course, these could easily be accidents of dimensional analysis

on account of the $1/G_{\rm N}$ normalization in the exponent. But in any case, it would be fascinating to see if any of these superficial similarities carry deeper significance.

Meanwhile, a virtue of the algebra isomorphism $\mathfrak{so}(4) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2)$ is that we always can split the six generators of the Lorentz group G into three chiral and three antichiral SU(2) generators. Assigning dotted and undotted spinor indices to each sector, we see that the spin connection decomposes into self-dual and anti-self-dual components, $\omega^{\dot\alpha\dot\beta}$ and $\omega^{\alpha\beta}$, while the tetrad is $e^{\dot\alpha\alpha}$. Similarly, the Plebański two-form B_a decomposes into $B^{\dot\alpha\dot\beta} = \epsilon_{\alpha\beta}\,(e^{\dot\alpha\alpha} \wedge e^{\dot\beta\beta})$ and $B^{\alpha\beta} = -\tilde\epsilon_{\dot\alpha\dot\beta}\,(e^{\dot\alpha\alpha} \wedge e^{\dot\beta\beta})$. See App. A for more details. Meanwhile, since the two $\mathfrak{su}(2)$ sectors commute, any element of the Lorentz algebra that exponentiates to a center element can be split into self-dual and anti-self-dual generators that separately exponentiate to center elements. Given these facts, we see that the symmetry operator decomposes into more primordial building blocks, which are *chiral* symmetry operators,

$$\begin{aligned} \widetilde U(\mathcal{S}) &= \exp\left(-\frac{1}{4G_{\rm N}}\int_{\mathcal{S}} \tilde\lambda^{\dot\beta}{}_{\dot\alpha}\, e^{\dot\alpha\alpha} \wedge e_{\alpha\dot\beta}\right) \qquad \text{where} \qquad (e^{2\pi\tilde\lambda})^{\dot\alpha}{}_{\dot\beta} = -\delta^{\dot\alpha}{}_{\dot\beta}\,, \\ U(\mathcal{S}) &= \exp\left(-\frac{1}{4G_{\rm N}}\int_{\mathcal{S}} \lambda_\alpha{}^\beta\, e_{\beta\dot\alpha} \wedge e^{\dot\alpha\alpha}\right) \qquad \text{where} \qquad (e^{2\pi\lambda})_\alpha{}^\beta = -\delta_\alpha{}^\beta\,. \end{aligned} \tag{3.23}$$

Here $\tilde\lambda^{\dot\alpha}{}_{\dot\beta}$ and $\lambda_\alpha{}^\beta$ belong to the self-dual and anti-self-dual Lie algebras.

For the case of Euclidean signature with $G = \mathrm{Spin}(4)$, the chiral operators $\widetilde U(\mathcal{S})$ and $U(\mathcal{S})$ are precisely the one-form symmetry operators corresponding to each factor of the center subgroup $Z(G) = \mathbb{Z}_2 \times \mathbb{Z}_2$. Note that, as zero-form symmetries, each factor of the center acts as parity in the numbers of dotted and undotted spinor indices, respectively. In Lorentzian signature, however, the one-form symmetry is only nontrivial if $G = \mathrm{SL}(2,\mathbb{C})$, in which case the center subgroup is $Z(G) = \mathbb{Z}_2$. As a zero-form symmetry this acts as *net* parity on spinor indices. In this case, only the *real* combination of the chiral operators, $\widetilde U(\mathcal{S})\,U(\mathcal{S})$, corresponds to the one-form symmetry operator.

Now let us finally present a derivation of the Ward identity associated with the one-form symmetry of gravity,

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\big\rangle = \rho(\alpha)^{\mathrm{link}(\mathcal{C},\mathcal{S})}\big\langle W_\rho(\mathcal{C})\big\rangle\,. \tag{3.24}$$

Since the tetradic Palatini framework is a first-order formalism, the left-hand side is given by a path integral over all configurations of both the spin connection and the tetrad,

$$\big\langle U_\alpha(\mathcal{S}) W_\rho(\mathcal{C})\big\rangle = \int \mathcal{D}\omega\, \mathcal{D}e\; e^{-S}\, U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\,. \tag{3.25}$$

As before, the symmetry operator $U_\alpha(\mathcal{S})$ merges with e^{-S} to give

$$e^{-S}\, U_\alpha(\mathcal{S}) = \exp\left(-\frac{1}{4g^2}\int_{\mathcal{M}} \epsilon_{ABCD}\, e^A \wedge e^B \wedge (R - \Omega^{-1} d d\Omega)^{CD} + \dots\right), \tag{3.26}$$

in analogy with Eq. (2.18), and where the ellipses denote the cosmological constant term. Meanwhile, we see that the twisted Lorentz transformation in Eq. (3.16) implies

$$R^A{}_B \;\mapsto\; (\Omega^{-1})^A{}_C\, R^C{}_D\, \Omega^D{}_B + (\Omega^{-1})^A{}_C\, d d\Omega^C{}_B\,, \tag{3.27}$$

which is gravitational analog of Eq. (2.14). Applying this to Eq. (3.26), we find

$$e^{-S}\, U_\alpha(\mathcal{S}) \;\mapsto\; e^{-S}\,. \tag{3.28}$$

Together with the transformation of the line operator in Eq. (3.15), Eq. (3.28) sends the left-hand side of the Ward identity in Eq. (3.25) to

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\big\rangle \;\mapsto\; \int \mathcal{D}\omega\, \mathcal{D}e\; e^{-S}\rho(\alpha)^{\text{link}(\mathcal{C},\mathcal{S})}\, W_\rho(\mathcal{C})\,, \tag{3.29}$$

which is precisely its right-hand side. This proves the Ward identity encoding the topological linking of the spin holonomy in Eq. (3.13) with the symmetry operator in Eq. (3.21).

Just like in the case of gauge theory, the one-form symmetry of gravity acts as a twisted Lorentz transformation on *any choice of operator*. Thus, if an operator transforms under the twisted Lorentz transformation as $\mathcal{O} \mapsto \mathcal{O}_\Omega$, then the corresponding Ward identity is $\big\langle U_\alpha(\mathcal{S})\,\mathcal{O}\big\rangle = \big\langle \mathcal{O}_\Omega\big\rangle$ in parallel with Eq. (2.21).

Before continuing, let us highlight an important point: the one-form symmetry of gravity is independent of our choice of formalism. In particular, while our derivations have made elaborate use of tetradic Palatini gravity, our final conclusions remain valid independent of this choice. For example, integrating out the spin connection yields a pure tetrad theory, but this still exhibits the one-form symmetry.

On the other hand, it is natural to ask about gravity in the pure *metric* formulation, where there is no tetrad, spin connection, or local Lorentz symmetry to speak of. In this case one formulates the dynamics only in terms of the metric, which is manifestly invariant under the twisted Lorentz transformation defined in Eq. (3.16), since $g_{\mu\nu} \mapsto g_{\mu\nu}$. Has the one-form symmetry disappeared? For the spinor holonomy, the answer is yes, but for the simple reason that we cannot even write it down since there is no tetrad field to characterize the gravitational interactions of fermions. This would be analogous to studying Yang-Mills theory with the stipulation that we can only ever use adjoint indices, thus precluding the existence of the very fundamental Wilson loops which are charged under the one-form symmetry. On the other hand, the vector holonomy and its one-form symmetry properties should presumably have a pure metric description since spin structure should not be necessary. Note that there has been some interesting recent work constructing *continuous* one-form symmetries of linearized gravity using the metric alone [51–53]. It would be very illuminating to see explicitly how those symmetries relate to the ones derived in this paper.

3.1.3 Chiral Cosmic String

Just as in gauge theory, the one-form symmetry operator in gravity can be interpreted as an insertion of a defect in spacetime. It is then natural to ask, what is the nature of this singular object? Indeed, how would a relativist interpret such a defect?

To answer this question let us consider empty space, as described by a flat metric. Next, we apply the twisted Lorentz transformation in Eq. (3.16), which induces a curvature singularity, $R^A{}_B = (\Omega^{-1}d d\Omega)^A{}_B = 2\pi\lambda^A{}_B\,\delta(\mathcal{S})$, localized on the surface $\mathcal{S}$. To give a

physical interpretation to this twisted geometry, we can straightforwardly reverse engineer the matter source that would directly generate this singularity. To do this we insert $R^A{}_B = 2\pi\lambda^A{}_B\,\delta(\mathcal{S})$ directly into the left-hand side of the Einstein field equations,

$$-\frac{1}{g^2}\star R_{AB}\wedge e^B = \frac{1}{3!}\,|e|\,T^\mu{}_A\,\epsilon_{\mu\nu\rho\sigma}\,dx^\nu\wedge dx^\rho\wedge dx^\sigma\,, \tag{3.30}$$

where $|e|$ denotes the determinant of $e^A{}_\mu$. From the resulting quantity, we then deduce the stress-energy tensor $T^\mu{}_A$ that would be required to generate the corresponding curvature singularity:

$$\begin{aligned} T^\mu{}_\kappa &= -\frac{1}{8G_{\rm N}}\frac{1}{|e|}\star\lambda_{\kappa\nu}\,\delta(\mathcal{S})_{\rho\sigma}\,\epsilon^{\mu\nu\rho\sigma}\,, \\ &= -\frac{1}{4G_{\rm N}}\star\lambda_{\kappa\nu}\int d\sigma_1 d\sigma_2\;\delta^{(4)}(x-X)\left(\frac{\partial X^\mu}{\partial\sigma_1}\frac{\partial X^\nu}{\partial\sigma_2}-\frac{\partial X^\nu}{\partial\sigma_1}\frac{\partial X^\mu}{\partial\sigma_2}\right), \end{aligned} \tag{3.31}$$

where in the last line we have parameterized the surface $\mathcal{S}$ by the function $X^\mu(\sigma_1,\sigma_2)$ in terms of worldsheet coordinates (σ_1,σ_2) and used a formula given in Eq. (B.2). Also, we have freely traded off local Lorentz indices with spacetime indices through the tetrad or its inverse in these final expressions, while $T^\mu{}_A$ in Eq. (3.30) acts as a source for the tetrad $e^A{}_\mu$ in the first-order formulation defined in Eq. (3.7). The stress-energy tensor in Eq. (3.31) is localized along a membrane $\mathcal{S}$ and describes a defect reminiscent of a *cosmic string*, though it is not literally identical to the Nambu-Goto string.

Interestingly, we can also see that the algebraic Bianchi identity, $R_{AB}\wedge e^B = 0$, actually *fails* for this defect configuration, indicating the existence of *magnetic* stress-energy [75]. That is, plugging $R^A{}_B = 2\pi\lambda^A{}_B\,\delta(\mathcal{S})$ into the left-hand side of algebraic Bianchi, we obtain a nonzero expression,

$$-\frac{1}{g^2}\,R_{AB}\wedge e^B = \frac{1}{3!}\,|e|\,T^{\star\mu}{}_A\,\epsilon_{\mu\nu\rho\sigma}\,dx^\nu\wedge dx^\rho\wedge dx^\sigma\,, \tag{3.32}$$

which defines the dual stress-energy tensor $T^{\star\mu}{}_A$. Note that the above equation is manifestly Hodge dual to the Einstein field equations in Eq. (3.30). The fact that the right-hand side of Eq. (3.32) is nonzero is analogous to the failure of the Bianchi identity in Maxwell theory in the presence of a magnetic monopole. For our string geometry, the components of this dual stress-energy tensor are given by

$$T^{\star\mu}{}_\kappa = -\frac{1}{8G_{\rm N}}\frac{1}{|e|}\,\lambda_{\kappa\nu}\,\delta(\mathcal{S})_{\rho\sigma}\,\epsilon^{\mu\nu\rho\sigma}\,, \tag{3.33}$$

describing a line distribution of NUT charge. In the meantime, recall that the Lorentz generator λ that exponentiates to a nontrivial center element has to be either self-dual or anti-self-dual. This implies that the electric and magnetic stress-energy tensors in Eqs. (3.31) and (3.33) are related either as $T^\mu{}_\kappa = T^{\star\mu}{}_\kappa$ or $T^\mu{}_\kappa = -T^{\star\mu}{}_\kappa$, so this string is composed of either self-dual or anti-self-dual matter. Therefore, we conclude that the geometry generated by an insertion of the symmetry operator represents a "*chiral cosmic string*," where self-dual or anti-self-dual stress-energy localizes on a two-dimensional surface $\mathcal{S}$.

Meanwhile, there is another well-known class of line singularities in relativity: the Misner string [75–80], which is famously the line singularity of the Taub-NUT solution [76, 78, 81, 82]. Does the chiral cosmic string carry a Misner string component? The answer turns out to be no. First of all, the Misner string describes a stack of gravitomagnetic *dipoles* [78, 79], and as such it can terminate on a set of gravitomagnetic monopoles, also known as Taub-NUT black holes. However, the magnetic stress-energy found in Eq. (3.33) rather describes a line density of distributed gravitomagnetic *monopole* whose magnitude is equal to the electric stress-energy in Eq. (3.31). Secondly, the Misner string could also be characterized as a localized source of torsion, as it induces a time monodromy [76, 80, 83]. However, it is easy to see that torsion transforms linearly even under twisted local Lorentz transformations, so if torsion was zero in the initial background it stays identically zero after the insertion of the symmetry operator as well. Concretely, one can establish that our string geometry has vanishing torsion by plugging in the transformed tetrad and spin connection in Eq. (3.16) into $De^A = de^A + \omega^A{}_B \wedge e^B$, which also confirms that the failure of algebraic Bianchi is solely due to the multivaluedness of the tetrad, since $dde^A \neq 0$. The absence of torsion is also clear if one recalls the fact that torsion is induced by local source for the spin connection in the tetradic Palatini formalism, while the symmetry operator only couples to the tetrad.[18] For these reasons, we conclude that the singular string geometry generated by the symmetry operator is of a cosmic kind, with no Misner string component.

3.1.4 Linking and Conical Deficit Angle

We have established that the symmetry operator creates a chiral cosmic string singularity that carries both electric and magnetic gravitational charge. What is the meaning of the linking of this string with the spin holonomy? Remarkably, this too has a simple physical interpretation in terms of classical gravitation. To understand why, let us revisit the path integral computation of the Ward identity,

$$\big\langle U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C}) \big\rangle = \int \mathcal{D}\omega\, \mathcal{D}e\; e^{-S}\, U_\alpha(\mathcal{S})\, W_\rho(\mathcal{C})\,, \tag{3.34}$$

but evaluated from using perturbation theory about flat space. The only dynamical degrees of freedom are the tetrad and spin connection, which encode fluctuations of the physical graviton. As we will see, from this viewpoint the linking number is computed by an infinite set of perturbative diagrams whose structure offers some nice physical insights.

To apply perturbation theory, we must interpret $U_\alpha(\mathcal{S})$ and $W_\rho(\mathcal{C})$ as *external sources* for the graviton field. From Eq. (3.13) we see that the spin holonomy $W_\rho(\mathcal{C})$ couples the graviton to the contour $\mathcal{C}$ with a dimensionless coupling strength. That is expected because the holonomy describes the coupling of gravity to spin. Meanwhile, Eq. (3.21) implies that the symmetry operator $U_\alpha(\mathcal{S})$ couples the graviton to the surface $\mathcal{S}$ with a coupling $1/g^2$. Last but not least, from Eq. (3.7), we see that in a normalization where the graviton field is dimensionless, graviton vertices scale as $1/g^2$ and graviton propagators scale as g^2.

[18] For this reason, the fact that the stress-energy tensor in Eq. (3.31) is not symmetric cannot be attributed to nonzero torsion. It rather traces back to the nonvanishing magnetic stress-energy [75].

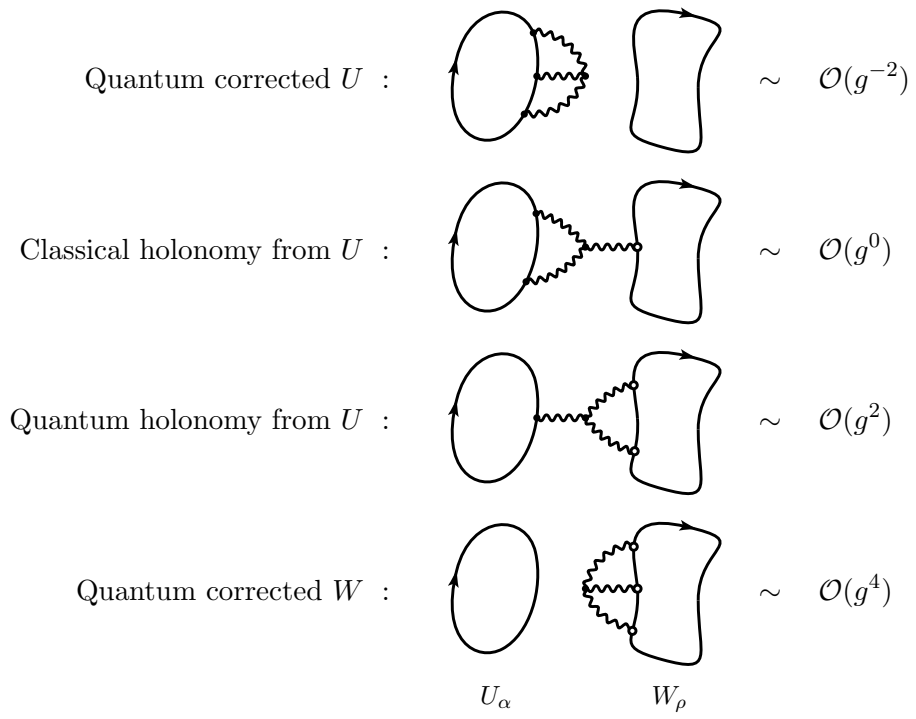


Figure 7. A schematic depiction of Feynman diagrams contributing to the Ward identity at various perturbative orders in the gravitational coupling g. The wavy lines denote gravitons, while the loops of solid lines on the left and right depict the symmetry operator U_α and line operator W_ρ, respectively. The first and fourth rows describe quantum corrections to U_α and W_ρ separately. The second row is a contribution to the classical holonomy measured by W_ρ, treating U_α as a source for the background spacetime. The third row is the one-loop quantum correction to this quantity in that background. Since the linking number is dimensionless and topological, it arises purely from the classical holonomy.

To classify the various contributions in perturbation theory, let us first consider a tree-level n-point correlator of gravitons. This object has $n-2$ vertices and $2n-3$ propagators, so it scales as $g^{2(n-1)}$. Loop-level contributions will be higher order in g^2, so they are subleading. Next, we take this n-point correlator and attach its external legs to the sources $U_\alpha(\mathcal{S})$ and $W_\rho(\mathcal{C})$.

The leading diagrams arise when n external gravitons are connected to $U_\alpha(\mathcal{S})$, yielding n factors of $1/g^2$. This contribution scales as $1/g^2$ and corresponds to the renormalization of $U_\alpha(\mathcal{S})$ coming from graviton loops. Since these contributions do not link with $W_\rho(\mathcal{C})$, they are unrelated to the topological linking number. These diagrams are depicted in the first row of Fig. 7.

Meanwhile, the next-to-leading contributions come from diagrams in which $n-1$ external gravitons are connected to $U_\alpha(\mathcal{S})$, with the last external graviton connected to $W_\rho(\mathcal{C})$ as an insertion of the spin connection. The corresponding diagram scales as a dimensionless constant, since all powers in g^2 exactly cancel. This contribution, resumming up all dia-

grams for all n, is precisely the tree-level one-point function of spin connection computed at all orders in g^2 in the presence of the source $U_\alpha(\mathcal{S})$, otherwise known as the classical spin connection in the corresponding classical gravity problem. This relationship was understood in the seminal work of Ref. [84], which showed how to perturbatively construct the Schwarzschild metric from an analogous set of diagrams.[19] In any case, we can then compute the next-to-leading contribution to the linking number by inserting this one-point function of the spin connection directly into $W_\rho(\mathcal{C})$. See the second row of Fig. 7 for a depiction of these contributions.[20] Note that on account of the exponential in $W_\rho(\mathcal{C})$, it is inserted an infinite number of times. In perturbation theory, such a calculation would be prohibitively hard. But crucially, this procedure is literally exactly equivalent to a calculation of the classical spin holonomy $W_\rho(\mathcal{C})$ evaluated with the spin connection set to its background value sourced by $U_\alpha(\mathcal{S})$.

As is well-known, this classical spin holonomy can be viewed as a geometric phase accounting for the precession of a spinning particle as it circumscribes the contour $\mathcal{C}$. Hence, the classical spin holonomy $W_\rho(\mathcal{C})$ precisely measures a certain *conical deficit angle* induced by $U_\alpha(\mathcal{S})$. Since $W_\rho(\mathcal{C})$ rephases by a center element, this conical deficit angle is quantized, and should be viewed intuitively as a π phase shift. The examples in the subsequent section will verify this.

Of course, there are next-to-next-to-leading contributions and beyond that are ever higher order in g^2. Some of these are just the renormalization of $W_\rho(\mathcal{C})$, which are depicted in the fourth row of Fig. 7. The remaining contributions are shown in the third row of Fig. 7, and correspond to quantum loop corrections to the spin holonomy. However, we can argue a priori that the topological winding number should not receive quantum corrections. This follows from dimensional analysis and topology. Since all quantum corrections enter with additional factors of g^2, the corresponding contributions to the dimensionless linking number must involve some other dimensionful parameter. The only other scales available are the relative distances and sizes associated with $\mathcal{C}$ and $\mathcal{S}$. However, these scales cannot appear, since our results are invariant under topology-preserving deformations. Consequently, all quantum corrections must be power divergent and can be absorbed into counterterms at all orders in perturbation theory. Said another way, the conical deficit angle induced by $U_\alpha(\mathcal{S})$ should not be renormalized.

3.1.5 Explicit Examples

To summarize, we have shown that the gravitational one-form symmetry operator implements a twisted Lorentz transformation, which is in turn equivalent to the insertion of

[19] Recently, this connection between perturbative diagrams and classical dynamics has been used to simplify certain contributions to black hole scattering in a recently proposed effective field theory for extreme mass ratio inspirals [85].

[20] Technically, the Feynman diagram contribution in the second row of Fig. 7 vanishes since the trace in the spin holonomy acts on a single factor of the antisymmetric spin connection, yielding zero. However, there are diagrams with multiple insertions of the one-point function of the spin connection into the spin holonomy which enter at the same order in perturbation theory, and they contribute nontrivially to the linking number.

a chiral cosmic string defect. The symmetry operator can be treated as a matter source which nonlinearly generates the classical one-point function of spin connection, which is then input into the line operator to yield the classical holonomy. Hence, the classical holonomy precisely reproduces the Ward identity for the line and symmetry operators. Let us demonstrate how this works in some simple examples.

A. Symmetry Operator as Chiral Cosmic String in Flat Space

To begin, let us consider an initial background spacetime given by flat Euclidean space $\mathcal{M} = \mathbb{R}^4$, equipped with Cartesian coordinates $x^\mu = (x, y, z, t)$ and the trivial tetrad $\delta^A{}_\mu = \text{diag}(1, 1, 1, 1)$. In this background the classical spin connection coefficients are zero and so the classical holonomy is trivial.

Next, let us induce a twist on this trivial configuration in a way that parallels the example from Yang-Mills theory described in Eq. (2.24). For concreteness, we assume that the Lorentz group is $G = \text{SO}(4)$, so the center one-form symmetry is $Z(G) = \mathbb{Z}_2$, under which vector holonomies are charged. Next, we define a twisted Lorentz transformation that generates a string singularity along the surface at $x = y = 0$, so

$$\Omega = \exp\left(\frac{k}{2}\, c\phi\right) \qquad \text{where} \qquad e^{2\pi c} = \mathbb{1}\,, \tag{3.35}$$

where k is an integer and ϕ is the azimuthal angle such that $\tan\phi = y/x$. For example, the Lorentz reference vector can be chosen to be

$$c^A{}_B = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \tag{3.36}$$

which is shown in the vector representation of the Lorentz group. Then, according to Eqs. (3.16) and (3.27), the spin connection and curvature transform to

$$\begin{aligned} \omega^A{}_B &= (\Omega^{-1} d\Omega)^A{}_B = \frac{k}{2}\, c^A{}_B\, d\phi = \frac{k}{2}\, c^A{}_B\, \frac{x\, dy - y\, dx}{x^2 + y^2} \\ R^A{}_B &= (\Omega^{-1} d d\Omega)^A{}_B = \frac{k}{2}\, c^A{}_B\, dd\phi = \frac{k}{2}\, 2\pi c^A{}_B\, \delta(x)\delta(y)\, dx \wedge dy\,, \end{aligned} \tag{3.37}$$

where the latter exhibits the chiral cosmic string defect whose worldsheet is in the x-y plane. We emphasize here that the orientation of the Lorentz transformation $c^A{}_B$ is completely independent of the direction that the string actually spans in spacetime. In Eq. (3.36), we have arbitrarily chosen $c^A{}_B$ to act on the x-t and y-z planes. This is an essentially random choice—this construction exists for any $c^A{}_B$ that exponentiates properly to the identity, as stipulated in Eq. (3.36). As we saw earlier, while this naively chooses a Lorentz violating reference vector, it is spurious because the reference drops out of the Ward identity for the symmetry operator. Lastly, note that the intrinsic chirality of this construction is evident if we write these expressions in spinor notation, where $c^A{}_B$ splits into $\tilde{c}^{\dot\alpha}{}_{\dot\beta} = -i(\sigma_1)^{\dot\alpha}{}_{\dot\beta}$ and $c_\alpha{}^\beta = 0$. Hence, this defect interacts with antichiral fermions, but not chiral fermions. It is for this reason that we refer to the string itself as chiral.

Plugging in the background spin connection above into the holonomy, we obtain

$$\mathrm{tr}_{\mathrm{vec}}\,\mathrm{P}\exp\left(\frac{k}{2}\,c\int_0^{2\pi}d\phi\right)=4(-1)^k\,, \tag{3.38}$$

where the trace and exponentiation are performed in the vector representation. From the factor of $(-1)^k$, we immediately see that this holonomy flips sign depending on the parity of k, which defines the number of windings.

It is not difficult to understand that this sign factor from the Wilson loop describes a conical deficit angle quantized in the units of π. Suppose a Lorentz vector v^A is initialized to a value v_0^A at a point on $\phi=0$ and then parallel-transported around the string. The spin connection enters into the equation for parallel transport, $\dot{v}^A=-\omega^A{}_{B\rho}\,v^B\,\dot{x}^\rho$, whose solution is $v^A=(\Omega^{-1})^A{}_B\,v_0^B$. That is, the parallel transport of this vector is precisely implemented by a twisted Lorentz transformation. As a result, during a round trip the vector experiences a succession of smooth rotations that accumulates to a net rotation of $e^{\pi kc}=\pm\mathbb{1}$, corresponding to a deficit angle of $k\pi$.[21] This exactly describes how a conical deficit angle is measured in classical gravitation. For even k the holonomy is trivial, while for odd k the vector experiences a net π rotation. Note that this rotation is always *orientation-preserving*, since we consider the proper Lorentz group throughout. Consequently, our results are consistent with [49], which elegantly argues for the impossibility of *orientation-changing* defects. Furthermore, we see that the chiral cosmic string differs from the traditional cosmic string, because the angle deficit is accumulated around an arbitrary axis $c^A{}_B$ which is totally unrelated to the actual orientation of the string in spacetime.

One can repeat this exercise for the case of spinor holonomy, which requires a Lorentz group $G=\mathrm{Spin}(4)$, whose center one-form symmetry is $Z(G)=\mathbb{Z}_2\times\mathbb{Z}_2$. Like before, we apply a twisted Lorentz transformation of the form of Eq. (3.35), except in the spinor representation. In this case, as noted earlier Eq. (3.36) translates to

$$c_\alpha{}^\beta=0\qquad\text{and}\qquad\tilde{c}^{\dot\alpha}{}_{\dot\beta}=-i(\sigma_1)^{\dot\alpha}{}_{\dot\beta}\,. \tag{3.39}$$

It is then easy to compute the holonomies in the chiral and antichiral spinor representations,

$$\begin{aligned}\mathrm{tr}_{\mathrm{sp}}\,\mathrm{P}\exp\left(\frac{k}{2}\,c\int_0^{2\pi}d\phi\right)&=2\\ \mathrm{tr}_{\widetilde{\mathrm{sp}}}\,\mathrm{P}\exp\left(\frac{k}{2}\,\tilde{c}\int_0^{2\pi}d\phi\right)&=2(-1)^k\,,\end{aligned} \tag{3.40}$$

so the antichiral spin holonomy is rephased by $(-1)^k$, while the chiral spin holonomy is invariant. Thus the chiral cosmic string detects the parity of antichiral spinor indices. Like before, these spin holonomies can be deduced from the parallel transport of fermions around a loop, described by $\dot{\psi}_\alpha=-\omega_\alpha{}^\beta{}_\rho\,\psi_\beta\,\dot{x}^\rho$ and $\dot{\tilde{\psi}}^{\dot\alpha}=\tilde{\omega}^{\dot\alpha}{}_{\dot\beta\rho}\tilde{\psi}^{\dot\beta}\,\dot{x}^\rho$.

[21] One might ask whether the same result trivially follows from solving $\dot{v}^\mu=-\Gamma^\mu{}_{\nu\rho}\,v^\nu\,\dot{x}^\rho$, which describes the parallel transport of a vector in terms of spacetime indices. The answer is no, since this equation computes the holonomy of the *Christoffel symbol*, which is a functional of the metric rather than the spin connection. So as an operator in the first-order formalism, the Christoffel holonomy is simply not equal to the spin holonomy defined in Eq. (3.13). There is no contradiction here: to implement the symmetry transformation of the Christoffel holonomy, one should determine the analog of the twisted Lorentz transformation for spacetime indices, which may be possible in a metric-affine formulation of gravity.

B. Symmetry Operator as Cosmic String in Black Hole Geometry

As a more general example, we next consider the one-form symmetry operator in curved spacetime. Conveniently, we can exploit well-known expressions from classical gravity in order to compute at all orders in perturbation theory in the gravitational constant.

Our starting point is an AdS-Schwarzschild background in Boyer-Lindquist coordinates $x^\mu = (r, \theta, \phi, t)$. The line element is given by

$$ds^2 = \frac{1}{f(r)^2}\, dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 + f(r)^2\, dt^2\,, \tag{3.41}$$

where we have denoted

$$f(r) = \sqrt{1 - \frac{2G_{\rm N}M}{r} + \frac{r^2}{l^2}}\,, \tag{3.42}$$

where M is the mass of the black hole and l is the AdS radius.

The goal of this analysis is to compute the classical spin holonomy in the presence or absence of the symmetry operator in order to verify the validity of the Ward identity of the one-form symmetry. To be concrete, let us consider the spin holonomy for a circular loop $\mathcal{C}$ in the plane defined by $\theta = \pi/2$ and with constant radius $r = r_0$:

$$W_{\rm vec}(\mathcal{C}) = {\rm tr}_{\rm vec}\, {\rm P}\exp\left(\int_0^{2\pi} \omega_\phi(r_0, \pi/2, \phi, 0)\, d\phi\right). \tag{3.43}$$

The Ward identity for the gravitational one-form symmetry implies that this spin holonomy should change its value depending on whether or not we insert the symmetry operator.

Firstly, consider a pure AdS-Schwarzschild background in the absence of the symmetry operator. Here the spin connection relevant to the holonomy in Eq. (3.43) is

$$\omega^A{}_{B\phi}(r, \pi/2, \phi, t) = \begin{pmatrix} 0 & 0 & f(r) & 0 \\ 0 & 0 & 0 & 0 \\ -f(r) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.44}$$

This is independent of ϕ, so the path ordering can be dropped. Straightforward calculation shows that the Wilson loop in the vector representation is given by (see also [86, 87])

$$W(\mathcal{C}) = 2 + 2\cos(2\pi f(r_0))\,. \tag{3.45}$$

In fact, the calculation can be done "symbolically" by observing that the spin connection component in Eq. (3.44) splits into self-dual and anti-self-dual parts as

$$\omega^A{}_{B\phi}(r, \pi/2, \phi, t) = \frac{(c_+ + c_-)^A{}_B}{2}\, f(r)\,, \tag{3.46}$$

where we have defined

$$(c_+)^A{}_B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad (c_-)^A{}_B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \tag{3.47}$$

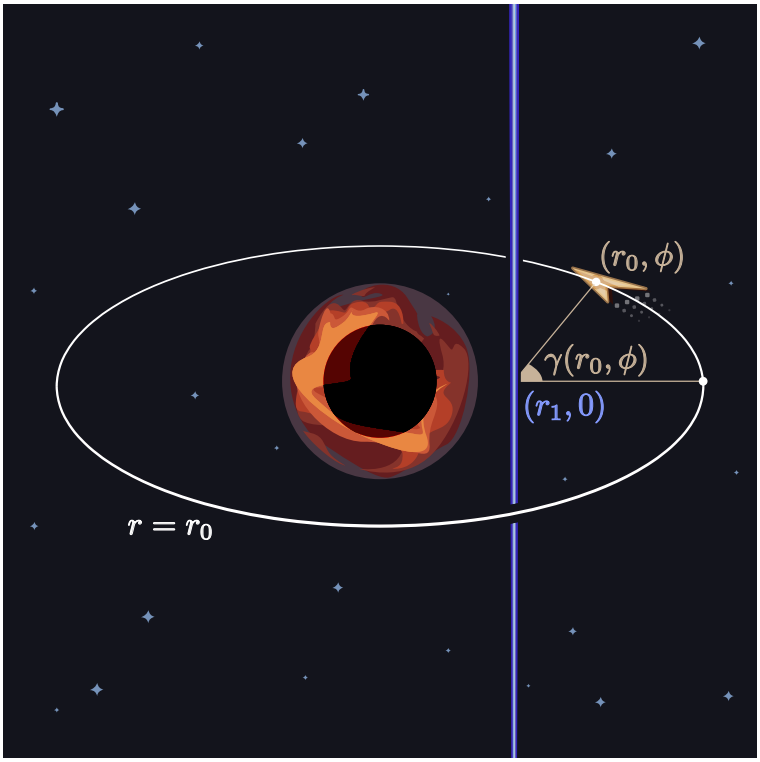


Figure 8. A spaceship is in circular orbit about a black hole adjacent to a chiral cosmic string. On board, a spinning top experiences an additional π rotation of its spin per round trip, as compared to without the string.

The calculation is simplified once one observes that these Lorentz generators commute.

Next, we compute the same spin holonomy but with the insertion of a symmetry operator $U_\alpha(\mathcal{S})$. As depicted in Fig. 8, we take $\mathcal{S}$ to define the worldsheet of a chiral cosmic string spanning the azimuthal direction. Without loss of generality, let us take the string to intersect the plane $\theta = \pi/2$ at $(r, \phi) = (r_1, 0)$. The associated twisted Lorentz transformation is, for example,

$$\Omega = \exp\left(\frac{c}{2}\gamma(r,\phi)\right) \qquad \text{where} \qquad e^{2\pi c} = \mathbb{1}\,, \tag{3.48}$$

which is simply an instance of Eq. (3.35) with $k = 1$ winding. Here $\gamma(r, \phi)$ describes the "apparent" azimuthal angle measured from the string at $(r, \phi) = (r_1, 0)$, so concretely,

$$\tan\gamma(r,\phi) = \frac{r\sin\phi}{r\cos\phi - r_1}\,. \tag{3.49}$$

In principle, we can choose any generator for c to demonstrate the transformation of the Wilson loop. For an explicit check, let us work out a simple case in which the multivalued transformation parameter Ω commutes with the spin connection component in Eq. (3.44). Namely, we take the Lorentz generator in Eq. (3.48) to be the self-dual generator in Eq. (3.47), so $c^A{}_B = (c_+)^A{}_B$. By explicit calculation we find that the spin connection

transforms according to

$$\omega^A{}_{B\phi}(r,\pi/2,\phi,t) = \frac{(c_+ + c_-)^A{}_B}{2} f(r) \;\; \mapsto \;\; \frac{(c_+ + c_-)^A{}_B}{2} f(r) + \frac{(c_+)^A{}_B}{2} \frac{\partial \gamma}{\partial \phi}(r,\phi)\,, \quad (3.50)$$

where we have used that $(c_+)^A{}_B$ commutes with $(c_-)^A{}_B$. By construction the path ordering of this transformed spin connection is trivial and can be dropped. Finally, we find that the transformed spin holonomy is

$$\begin{aligned} W_{\text{vec}}(\mathcal{C}) \;\; &\mapsto \;\; \operatorname{tr} \exp\left(\frac{c_+ + c_-}{2} 2\pi f(r_0) + \frac{c_+}{2} 2\pi n \right), \\ &= 2\cos n\pi + 2\cos\Big(2\pi f(r_0) + n\pi\Big)\,, \\ &= \Big(2 + 2\cos(2\pi f(r_0))\Big) \cdot (-1)^n\,, \\ &= W_{\text{vec}}(\mathcal{C}) \cdot (-1)^n\,, \end{aligned} \quad (3.51)$$

where n is linking number measuring the number of times $\mathcal{C}$ winds about $\mathcal{S}$:

$$n \;=\; \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial \gamma}{\partial \phi}(r,\phi)\, d\phi \;=\; \begin{cases} 1 & \text{if} \quad r_1 < r_0\,, \\ 0 & \text{if} \quad r_1 > r_0\,. \end{cases} \quad (3.52)$$

Note how each term in Eq. (3.45) gets "twisted" to each term in Eq. (3.51), by the addition of $n\pi$ inside the argument of cosines. In conclusion, we find that the twisted local Lorentz transformation, which corresponds to an insertion of the operator $U_\alpha(\mathcal{S})$, flips the sign of the spin holonomy when the contour $\mathcal{C}$ and the surface $\mathcal{S}$ are linked. This verifies the Ward identity for the one-form symmetry in a curved background.

More generally, it is possible to choose the twisted Lorentz transformation such that the twisted spin connection does not commute with itself at different points. In this case the path ordering in the spin holonomy is much more difficult to compute directly and so we resort to numerical methods. For example, let us consider an arbitrary self-dual generator parametrized by a unit three-vector (n_1, n_2, n_3):

$$c^A{}_B \;=\; \begin{pmatrix} 0 & n_3 & -n_2 & n_1 \\ -n_3 & 0 & n_1 & n_2 \\ n_2 & -n_1 & 0 & n_3 \\ -n_1 & -n_2 & -n_3 & 0 \end{pmatrix} \;=\; n_1\,(t_1^+)^A{}_B + n_2\,(t_2^+)^A{}_B + n_3\,(t_3^+)^A{}_B\,. \quad (3.53)$$

In this case the twisted spin connection is given by

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \frac{f(r)}{2} - \begin{pmatrix} \Big(n_2 n_1\, \mathtt{s}^2 - n_3\, \mathtt{sc}\Big)(t_1^+)^A{}_B \\ + \Big(\frac{1}{2} - (1 - {n_2}^2)\, \mathtt{s}^2\Big)(t_2^+)^A{}_B \\ + \Big(n_2 n_3\, \mathtt{s}^2 + n_1\, \mathtt{sc}\Big)(t_3^+)^A{}_B \end{pmatrix} f(r) + \frac{c^A{}_B}{2} \frac{\partial \gamma}{\partial \phi}(r,\phi)\,, \quad (3.54)$$

where $\mathtt{s} = \sin(\gamma(r,\phi)/2)$, $\mathtt{c} = \cos(\gamma(r,\phi)/2)$. By evaluating the path-ordered exponential numerically, we can verify explicitly the final line of Eq. (3.51) for this more general case.

In the above examples we have always assumed that the holonomy contour and the chiral cosmic string are separated from the black hole horizon. Since the divergence of the line element in (3.41) at the horizon is a coordinate singularity, the loop and the string can actually be placed across or inside the horizon. It would be amusing to compute the holonomy in these more exotic configurations using coordinate systems in which the metric is regular on the horizon.

3.2 Canonical Formalism

Last but not least, let us now analyze the one-form symmetry of gravity from the point of view of the Hamiltonian formalism. The relevant framework is the nonchiral or real version of the Ashtekar [88] formulation, as studied in [89–92].

3.2.1 Phase Space

In Eq. (3.7) we described the Lagrangian for tetradic Palatini gravity, expressed in terms of the Plebański two-form field. With all indices written out explicitly, this Lagrangian reads

$$\frac{1}{g^2}\left[\frac{1}{2}B_{a\mu\nu}\Big(\partial_\rho\omega^a{}_\sigma+\tfrac{1}{2}f^a{}_{bc}\omega^b{}_\rho\omega^c{}_\sigma\Big)\epsilon^{\mu\nu\rho\sigma}-\frac{\Lambda}{6}B_{a\mu\nu}\star B^a{}_{\rho\sigma}\,\epsilon^{\mu\nu\rho\sigma}\right]. \tag{3.55}$$

Carrying out the $3+1$ decomposition, we immediately find that the dynamical degrees of freedom coordinatizing the phase space are $\omega^a{}_i$ together with their canonical conjugates,

$$E^i{}_a=\frac{1}{2}B_{ajk}\,\epsilon^{ijk}\,. \tag{3.56}$$

Upon quantization, the canonical commutation relations are then

$$\begin{aligned}
&[\omega^a{}_i(x),\omega^b{}_j(x')]=0\,,\\
&[E^i{}_a(x),\omega^b{}_j(x')]=g^2\,\delta^i{}_j\,\delta^b{}_a\,\delta^{(3)}(x-x')\,,\\
&[E^i{}_a(x),E^j{}_b(x')]=0\,,
\end{aligned} \tag{3.57}$$

where $x,x'\in\mathcal{M}_3$ are points in the spatial manifold. As expected, Eqs. (3.56) and (3.57) exactly mirror Eqs. (2.38) and (2.39) from gauge theory, which is indeed the entire point of the Ashtekar formulation [88, 93].

We restrict our focus to these canonical commutation relations, as they are the only crucial element for proving the one-form symmetry in Hamiltonian framework as we learned in Sec. 2.2. It is known that integrating out the nondynamical degrees of freedom eventually leads to a well-posed constrained Hamiltonian system equipped with several constraints [89–92], such as the Gauss constraints for local Lorentz transformations as well as the Hamiltonian and diffeomorphism constraints familiar from the ADM [94] analysis.[22]

[22] In particular, one can start from the Plebański Lagrangian [67], where the Plebański two-form in Eq. (3.6) is taken as a fundamental degree of freedom rather than a composite field constructed from the tetrad. In this case, the $3+1$ decomposition naturally gives the phase space $(\omega^a{}_i, E^i{}_a)$, where the variable $E^i{}_a$ is not composite.

3.2.2 Ward Identity

Following the logic of gauge theory, we now show how the one-form symmetry of gravity is implemented in the Hamiltonian formulation. Like before, we consider $W_\rho(\mathcal{C})$ oriented in a spatial slice at $t=0$ and linked with $U_\alpha(\mathcal{S})$. Again, we pancake $\mathcal{S}$ onto the spatial slice so that it becomes the disjoint union of two discs, defined by $\mathcal{S} = \mathcal{S}_+ \cup (-\mathcal{S}_-)$. Here $\mathcal{S}_+$ and $-\mathcal{S}_-$ are two halves of a squashed sphere which straddle the spatial slice, as depicted in Fig. 5. These discs are eventually projected down to a surface $\mathcal{S}_0$ in the time slice $t=0$.

The symmetry operator factorizes into $U_\alpha(\mathcal{S}) = U_\alpha(\mathcal{S}_+)U_\alpha(-\mathcal{S}_-)$, so the symmetry transformation becomes the equal-time operator equation,

$$U_\alpha(\mathcal{S}_0)\, W_\rho(\mathcal{C})\, U_\alpha^{-1}(\mathcal{S}_0) \;=\; \rho(\alpha)^{\text{int}_3(\mathcal{C},\mathcal{S}_0)}\, W_\rho(\mathcal{C})\,, \tag{3.58}$$

where $\mathcal{S}_0$ and $\mathcal{C}$ all belong to the spatial three-dimensional manifold $\mathcal{M}_3$. Here $U_\alpha(\mathcal{S}_0)$ is the avatar of the symmetry operator in the Hamiltonian framework, taking a two-dimensional surface $\mathcal{S}_0$ with boundary as its support. Applying the $3+1$ decomposition described in the previous section, the symmetry operator becomes

$$U_\alpha(\mathcal{S}_0) = \exp\left(\frac{\pi}{g^2}\int_{\mathcal{S}_0} dx^j \wedge dx^k\, \epsilon_{ijk}\, E^i{}_a\, \lambda^a\right) \qquad \text{where} \qquad e^{2\pi\lambda} = \alpha\,, \tag{3.59}$$

which exhibits the expected dependence on $E^i{}_a$ as the variable conjugate to the spin connection $\omega^a{}_i$. Using the canonical commutation relations in Eq. (3.57), the left-hand side of Eq. (3.58) can be evaluated as

$$U_\alpha(\mathcal{S}_0)\, W_\rho(\mathcal{C})\, U_\alpha^{-1}(\mathcal{S}_0) \;=\; \text{tr}_\rho\, \text{P}\exp\left(\oint_{\mathcal{C}} \omega + 2\pi\lambda\, \delta_3(\mathcal{S}_0)\right) \;=\; \rho(\alpha)^{\text{int}_3(\mathcal{C},\mathcal{S}_0)}\, W_\rho(\mathcal{C})\,, \tag{3.60}$$

thus establishing the Hamiltonian version of the Ward identity, introduced in Eq. (2.41) and reproduced in Eq. (3.58). Note the near isomorphism between Eqs. (3.59) and (3.60) for gravity and Eqs. (2.42) and (2.45) for gauge theory. Indeed, this parallel between Yang-Mills and gravity has been the central insight in our construction of the gravitational one-form symmetry. It also strongly aligns with the broader philosophy of the Ashtekar formulation, which is that Yang-Mills and gravity are formulated in essentially the same phase space.

We have taken as our starting point the definition of the spin holonomy in Eq. (3.13) and the symmetry operator in Eq. (3.21) and shown how these form the ingredients of a gravitational one-form symmetry. However, it is amusing that the reverse logic is actually possible. Starting from the spin holonomy, we can actually *deduce* the symmetry operator from first principles, as the line and charge operators are necessarily integrals of conjugate phase space variables. Specifically, this is required in order for their commutation relations to yield delta functions that eventually integrate to become field-independent linking numbers. Since $E^i{}_a$ is conjugate to the spin connection $\omega^a{}_i$, we see that Eq. (3.59) and its covariant counterpart were actually inevitable. Namely, this provides an alternative argument for the definition of the symmetry operator in Eq. (3.21).

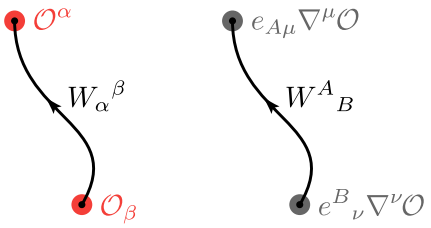


Figure 9. Spinor holonomy is screened by fermions. Vector holonomy is screened by orbital angular momentum.

3.3 Symmetry Breaking

As is well-known, global higher-form symmetries can be broken, either explicitly or spontaneously. In analogy with Yang-Mills theory, the one-form symmetry of gravity is explicitly broken in the presence of matter fields that transform nontrivially under the Lorentz group. For example, if the theory includes a local operator in the spin representation ρ, then it is possible to define a Lorentz invariant line holonomy $W_\rho(\mathcal{C})$ for a contour $\mathcal{C}$ that is not closed, but rather terminates on this local operator. The spin holonomy is then "endable," so it can be unlinked topologically from the symmetry operator $U_\alpha(\mathcal{S})$, and the one-form symmetry is explicitly broken. The physical interpretation of this phenomenon is that the spin holonomy is screened by spinning particles.

Interestingly, this implies that the gravitational one-form symmetry is explicitly broken by particles with spin. For example, if the Lorentz group is $G = \mathrm{Spin}(4)$, then a holonomy in the spinor representation can only end on a local operator $\mathcal{O}_\alpha$ with a free spinor index. Hence, the corresponding holonomy is endable if there exist fermions in the spectrum. The case of $G = \mathrm{SO}(4)$ is more subtle, however. A holonomy in the vector representation of the Lorentz group must end on an operator with a free Lorentz vector index, which can be any operator carrying orbital angular momentum dotted with a tetrad such as $e^A{}_\mu\nabla^\mu\mathcal{O}$.[23] Consequently, the vector holonomy is analogous to the adjoint Wilson loop of gauge theory, which is automatically screened by dynamical gluons. See Fig. 9 for a depiction of this phenomenon.

The above logic has implications for the standard model of physics. Since fermions exist, we know that the Lorentz group is $\mathrm{SL}(2,\mathbb{C})$, which implies that there is a $\mathbb{Z}_2$ one-form gravitational symmetry under which spinor holonomies are charged. Presuming the lightest neutrino is not massless, this one-form symmetry is unbroken below that scale. This implies a new, albeit subtle, exact symmetry in the known laws of physics.

Notably, the explicit breaking of higher-form symmetry is strongly suggested by the

[23] A more subtle question arises in formulations of gravity with differences in field content. For example, in Plebański theory [67], the associated two-form field B is a fundamental degree of freedom. The tetrad is sculpted from B through a constraint [89, 92], while the metric can be expressed in terms of B in closed form [68, 95]. Consequently, there is no field in the pure gravity sector that carries a single vector index, so it is unclear whether the vector spin holonomy is endable in this case.

so-called swampland conjectures. In particular, there is strong evidence that all global symmetries are necessarily broken at some scale in a consistent theory of quantum gravity [96–99]. A well-known avatar of this is the weak gravity conjecture [100, 101], which states that a $U(1)$ gauge theory must exhibit a state whose charge exceeds its mass in Planck units. The weak gravity conjecture quantitatively forbids the strict global symmetry limit of vanishing charge in a $U(1)$ gauge theory.

The swampland conjectures imply that any higher-form symmetry should be either gauged or explicitly broken. In the case of gauge theory coupled to gravity, the latter scenario requires the existence of a tower of charged states, as described by the so-called completeness conjectures [102]. Applying the same logic to dynamical gravity, we expect that something similar applies to the gravitational one-form symmetry. One option is that this symmetry is gauged, for example as would occur if the Lorentz group is $\mathrm{Spin}(4)/\mathbb{Z}_2\times\mathbb{Z}_2$. Alternatively, if the Lorentz group is $\mathrm{Spin}(4)$, then the one-form symmetry is not gauged and must be explicitly broken, thus implying the existence of fermions in the spectrum.

Finally, let us speculate briefly on the possibility of phases in gravity. Taking inspiration from gauge theory, it is natural to wonder whether the expectation value of the spin holonomy is an order parameter for symmetry breaking. In the case of gauge theory, it is well-known that an area versus perimeter law scaling of the Wilson loop expectation value is a diagnostic of whether or not a theory is confining. However, the analogous construction in gravity is far murkier. In particular, as we noted earlier, the diffeomorphism invariance of dynamical gravity suggests that the contour of the spin holonomy—and any contour, actually—must be defined relationally with respect to some invariant boundary data. So to be an order parameter, the spin holonomy must presumably be computed for a contour that circumscribes the boundary.

Even ignoring these subtleties, the fact that the effective field theory description of gravity is intrinsically weakly coupled suggests that confinement is not in play. More generally it is very unclear whether a low-energy effective theory of gravity on a fixed background could even access different phases, or what that would even mean. One speculation is that this might have something to do with degenerate configurations of the metric and their corresponding domain walls [103–106]. Another possibility is that a putative gravitational phase diagram might delineate various choices of compactification or of asymptotic behavior of the metric. Indeed, it is easy to see that the spin holonomy is highly sensitive to the cosmological constant. For these reasons, it would be interesting to explicitly compute the expectation value of the spin holonomy in various examples. A number of existing works have calculated the spin holonomy in various contexts [86, 87, 107–111].

4 Future Directions

In this paper, we have initiated an exploration of generalized symmetry in the context of dynamical gravity. Taking our cues from the one-form symmetries of Yang-Mills theory, we have considered gravity in the tetradic Palatini formalism, which is a gauge theory of the local Lorentz group. We have argued that the gravitational one-form symmetry is defined

by the center of the Lorentz group. The object which is charged under this symmetry is the spin holonomy $W_\rho(\mathcal{C})$. The one-form symmetry transformation is implemented by an operator $U_\alpha(\mathcal{S})$, which has dual interpretations, both as a twisted Lorentz transformation but also as a chiral cosmic string defect carrying both electric and magnetic gravitational charge. The topological linking of the line and symmetry operators corresponds to the measurement of a quantized conical deficit angle by the spin holonomy. In the standard model, this implies the existence of a new symmetry below the mass of the lightest neutrino. The present work leaves numerous avenues for future exploration, which we now describe.

First and foremost are a number of very simple extensions of this work which should be relatively straightforward. These include the question of generalization to higher spacetime dimension, which offers a richer spectrum of one-form symmetry groups. For example, the center group is maximally $Z(\mathrm{Spin}(4)) = \mathbb{Z}_2 \times \mathbb{Z}_2$ in four dimensions, but this grows to $Z(\mathrm{Spin}(6)) = \mathbb{Z}_4$ in six dimensions. Another concrete direction is the inclusion of gravitational higher-curvature corrections. These contributions will clearly preserve the one-form symmetry, whose corresponding symmetry operator will be equal to the surface integral of the canonical conjugate of the curvature in the effective field theory.

Secondly, while the present work has focused on gravitational one-form symmetry of electric type, it is natural to ask whether one can derive an analogous magnetic construction. For gauge theories, the electric and magnetic symmetry operators are straightforwardly related by the spacetime Hodge duality on the field strength [56], which is computed with respect to a *background* volume form. For gravity, the analogous procedure necessarily introduces metric dependence, since the Levi-Civita permutation symbol is a *density*. This additional metric dependence implies that the putative magnetic symmetry operator is no longer the proper integral of a form. It is unclear whether this technical obstruction is insurmountable or merely peculiar. Meanwhile, recent work on cobordism classes in quantum gravity [50] appears to construct certain magnetic gravitational defects. Perhaps a direct link can be drawn between those results and the approach taken here.

A third topic of future study is higher-group symmetry, which describes a certain non-abelian structure built from the fusion of multiple higher-form symmetries [112–115]. A well-studied example of this is axion-Yang-Mills theory, which exhibits a two-form symmetry for the axion, together with the one-form symmetry of the gauge theory [39]. These symmetries fuse to yield a two-group structure, which in fact bounds the scale of axion strings from below by the lightest particle in the fundamental of color. Acquainted with this remarkable fact, it would be interesting to investigate if gravity can also exhibit a higher-group symmetry. It seems quite likely that a similar two-group symmetry will appear in gravity coupled to an axion, in which case we should expect that the axion string scale is bounded by the mass of the lightest fermion.

Fourthly, while our approach of treating gravity as a "BF"-type theory yields a simple route to the one-form symmetry, there is the question of how this framework is explicitly realized in other physically equivalent formulations of gravity [116] in the literature. For example, it is known that in the presence of a nonzero cosmological constant one can integrate out the tetrad altogether, yielding a theory of gravity described purely in terms

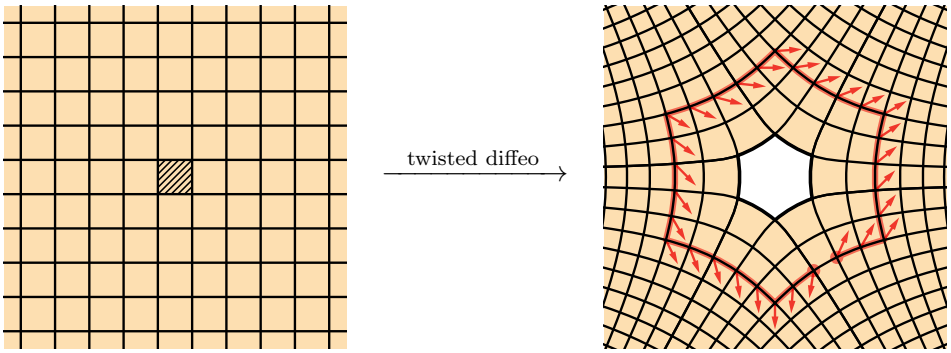
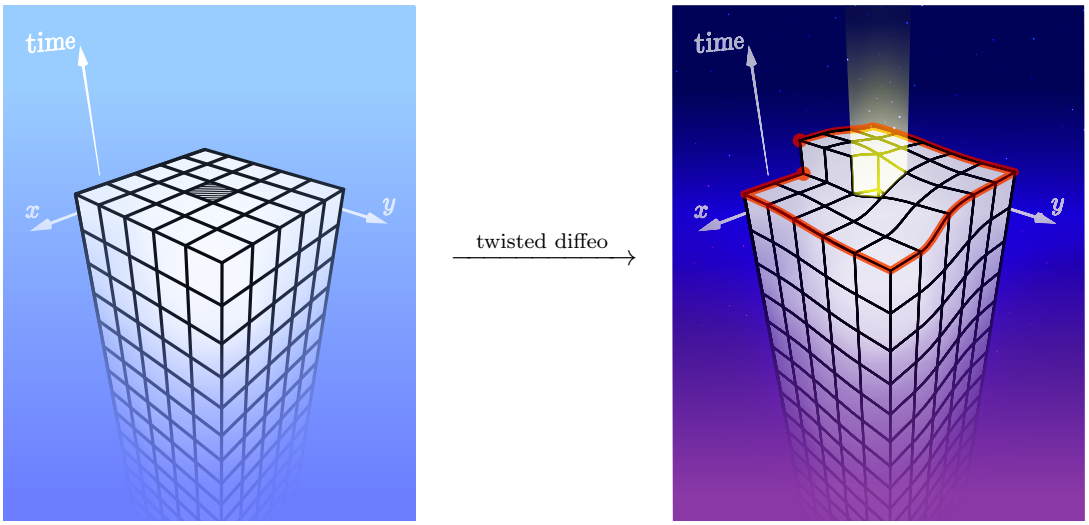


Figure 10. A diffeomorphism with a multivalued Jacobian creates a cosmic string from empty space, analogous to disinclinations in lattice systems.

of the spin connection [117, 118]. It would be interesting to understand the emergence of the gravitational one-form symmetry in this "pure connection" formalism given that all of our results were derived in the presence of a cosmological constant. Another open question is the fate of the one-form symmetry in gravitational formulations with subtly different field content, as referred to in a footnote in Sec. 3.3.

A fifth area of study concerns the question of what topological symmetry can teach us about classical gravitation. As a theory of spacetime geometry, gravity boasts a rich array of classical vacuum solutions, each showcasing distinctive *singularity structures* that are themselves a focal point of study [119]. In particular, it could be very illuminating to initiate a systematic analysis and classification of gravitational singularities from the perspective of topological operators and their algebra. For example, even in the absence of spin holonomy, the the one-form symmetry operator can link with "holes" in spacetime. It would be interesting to study whether there is physical information encoded in such a linking, for example regarding the singularity of a black hole. Moreover, it is conceivable that known spacetime singularities in the literature have an alternative interpretation as symmetry operators, like we discovered for the chiral cosmic string.

Last but not least, it would be interesting to see if the theoretical framework developed in this paper can be applied to lattice systems such as crystals with impurities [120–122]. Indeed, as described in [123–125], the physics of lattice systems has a formulation that is strikingly reminiscent of gravity. In this picture, lattice disinclinations are analogous to cosmic strings, as depicted in Fig. 10. Furthermore, the analog of "translation holonomy" is manifested by the Burgers vector, which measures the net drift in the "lattice frame" per round trip about a defect. Hence, as depicted in Fig. 11, we observe that lattice dislocations are analogous to the Misner string—a Dirac string of time monodromy flux endable on Taub-NUT charges [75–80, 83]. These two types of lattice defects are described in terms of *multivalued* coordinate transformations [123–125], which are the diffeomorphism analogs of the twisted gauge transformations that played such a crucial role in our construction



"Empty Space" "Misner String"

Figure 11. A multivalued diffeomorphism creates a Misner string from empty space, whose time monodromy is analogous to the Burgers vector of screw dislocations in lattice systems.

of a gravitational one-form symmetry. It would be interesting if this convergence between gravity and lattice systems could cross-pollinate new insights across these fields.

Acknowledgements. We are very much grateful to Clay Córdova, Kurt Hinterbichler, Ted Jacobson, Mrunmay Jagadale, Austin Joyce, Anton Kapustin, Jake McNamara, Julio Parra-Martinez, Matthew Reece, and Shu-Heng Shao for many insightful discussions and detailed comments on our draft. C.C., M.D., J.-H.K., V.N., and N.S. are supported by the Department of Energy (Grant No. DE-SC0011632) and by the Walter Burke Institute for Theoretical Physics. J.-H.K. is also supported in part by Ilju Academy and Culture Foundation. I.Z.R. is supported by the Department of Energy (Grant No. DE-FG02-04ER41338 and FG02-06ER41449).

A Notations and Conventions

Our results rely on numerous notational conventions and terminologies. For completeness, let us briefly summarize our nomenclature here.

A. Manifold and Index Conventions

Throughout our paper, the symbol $\mathcal{M}$ denotes the full four-dimensional manifold of space-time. Within it resides zero-, one-, two-, and three-dimensional submanifolds, which we denote by $\mathcal{P}$, $\mathcal{C}$, $\mathcal{S}$, and $\mathcal{V}$, respectively. We often define $\mathcal{S}$ to be exact, so it is the boundary of a corresponding coboundary manifold $\mathcal{V}$ such that $\mathcal{S} = \partial\mathcal{V}$. On the other hand, we take $\mathcal{C}$ to be closed, so it has no boundary and thus $\partial\mathcal{C} = 0$. In the Hamiltonian formalism,

we perform a $3+1$ decomposition which generates quantities associated with a spatial three-dimensional submanifold $\mathcal{M}_3$, like the three-dimensional intersection number int_3.

Spacetime indices are $\mu, \nu, \rho, \sigma, \ldots \in \{1, 2, 3, 4\}$, while spatial indices are $i, j, k, l, \ldots \in \{1, 2, 3\}$. Adjoint indices of the Lorentz group are $a, b, c, d, \ldots \in \{1, 2, 3, 4, 5, 6\}$ while vector indices are $A, B, C, D, \cdots \in \{1, 2, 3, 4\}$. Undotted and dotted spinor indices are $\alpha, \beta, \gamma, \delta, \ldots \in \{0, 1\}$ and $\dot{\alpha}, \dot{\beta}, \dot{\gamma}, \dot{\delta}, \ldots \in \{0, 1\}$, respectively.

All epsilon tensors $\epsilon^{\mu\nu\rho\sigma}$ and $\epsilon_{\mu\nu\rho\sigma}$ are pure permutation symbols valued in $\{+1, -1, 0\}$, sans dressing by metric-dependent determinant factors, so their indices are never partially raised or lowered. In particular, we employ the sign and normalization conventions,

$$\epsilon^{1234} = +1\,, \quad dx^\mu \wedge dx^\nu \wedge dx^\rho \wedge dx^\sigma = \epsilon^{\mu\nu\rho\sigma}\, d^4x\,, \quad \epsilon_{\mu\nu\rho\sigma}\, \epsilon^{\mu\nu\rho\sigma} = +4!\,. \tag{A.1}$$

On the other hand, ϵ^{ABCD} and ϵ_{ABCD} have Lorentz vector indices which are raised and lowered by the flat Euclidean metric $\delta_{AB} = \mathrm{diag}(+1, +1, +1, +1)$.

B. Lorentz Algebra Conventions

Let us now summarize our conventions for the Lorentz Lie algebra. First of all, we employ an anti-Hermitian convention for Lie algebra generators t_a so that $[t_a, t_b] = f^c{}_{ab}\, t_c$ without the imaginary unit. The adjoint and coadjoint actions on Lie algebra elements and their duals act as $(t_a)^i{}_j\, X^a \mapsto (\Omega^{-1})^i{}_k\, (t_a)^k{}_l\, \Omega^l{}_j\, X^a$ and $Y_a\, (t^a)^i{}_j \mapsto Y_a\, (\Omega^{-1})^i{}_k\, (t^a)^k{}_l\, \Omega^l{}_j$, respectively, so that $Y_a\, X^a \propto Y_a\, (t^a t_b)^i{}_i\, X^b$ is invariant. In the main text, we have simply denoted these as $X^a \mapsto (\Omega^{-1} X\, \Omega)^a$ and $Y_a \mapsto (\Omega^{-1} Y\, \Omega)_a$.

The generators of the Lorentz Lie algebra $\mathfrak{so}(4)$ are six anti-symmetric matrices $(t_a)^{AB} = (t_a)^{[AB]}$, while those of the dual Lie algebra $\mathfrak{so}(4)^*$ are $(t^a)_{AB} = (t^a)_{[AB]}$. These are normalized according to

$$\delta^a{}_b = \tfrac{1}{2}\, (t^a)_{AB}\, (t_b)^{AB} \qquad \text{and} \qquad (t_a)^{AB}\, (t^a)_{CD} = 2\, \delta^{[A}{}_C\, \delta^{B]}{}_D\,. \tag{A.2}$$

Accordingly, raised and lowered adjoint indices are related to fundamental indices by

$$\begin{aligned} X^{AB} &= (t_a)^{AB}\, X^a\,, \quad & X^a &= \tfrac{1}{2}\, (t^a)_{AB}\, X^{AB}\,, \\ Y_{AB} &= Y_a\, (t^a)_{AB}\,, \quad & Y_a &= \tfrac{1}{2}\, Y_{AB}\, (t_a)^{AB}\,. \end{aligned} \tag{A.3}$$

Note that the pairing between $\mathfrak{so}(4)$ and $\mathfrak{so}(4)^*$ is given by

$$Y_a\, X^a = \tfrac{1}{2}\, Y_{AB}\, X^{AB}\,. \tag{A.4}$$

Finally, the structure constants are

$$(t_a)^{AB}\, f^a{}_{bc}\, X_1^b\, X_2^c = X_1^{AC}\, \delta_{CD}\, X_2^{DB} - X_2^{AC}\, \delta_{CD}\, X_1^{DB}\,, \tag{A.5}$$

when expressed in the adjoint and fundamental representations.

Interestingly, two metrics can be endowed to the Lie algebra $\mathfrak{so}(4)$. Firstly, we have the usual positive-definite Killing form that universally exists in any dimensions:

$$\delta_{ab} = \tfrac{1}{2}\, \delta_{AC}\, \delta_{BD}\, (t_a)^{AB}\, (t_b)^{CD}\,, \quad \delta^{ab} = \tfrac{1}{2}\, (t^a)_{AB}\, (t^b)_{CD}\, \delta^{AC}\, \delta^{BD}\,, \quad \delta^{ac}\, \delta_{cb} = \delta^a{}_b\,. \tag{A.6}$$

Indices are raised and lowered with this metric, which is consistent with the usual practice of raising and lowering fundamental indices with the flat Euclidean metric.

On the other hand, there is also a metric specific to four dimensions:

$$\star_{ab} = \tfrac{1}{4}\,\epsilon_{ABCD}\,(t_a)^{AB}\,(t_b)^{CD}\,, \quad \star^{ab} = \tfrac{1}{4}\,(t^a)_{AB}\,(t^b)_{CD}\,\epsilon^{ABCD}\,, \quad \star^{ac}\star_{cb} = \delta^a{}_b\,. \tag{A.7}$$

Evidently, this implements the Hodge dual. For instance,

$$\begin{aligned}(\star_{ab}\,X^b)\,(t^a)_{AB} &= \tfrac{1}{2}\,\epsilon_{ABCD}\,X^{CD} = \star X_{AB}\,,\\ (t_a)^{AB}\,(\star^{ab}Y_b) &= \tfrac{1}{2}\,\epsilon^{ABCD}\,Y_{CD} = \star Y^{AB}\,.\end{aligned} \tag{A.8}$$

Note that this metric naturally appears in the Palatini Lagrangian in Eq. (3.7).[24] It has $(3,3)$ split signature, which connects to the decomposition $\mathfrak{so}(4) \cong \mathfrak{su}(2)\oplus\mathfrak{su}(2)$. Note also the identities

$$\star^{ad}\delta_{dc}\star^{ce}\delta_{eb} = \delta^a{}_b\,, \quad f^a{}_{bc}\star^{be}\star^{cf} = f^a{}_{bc}\delta^{be}\delta^{cf}\,, \tag{A.9}$$

which hold because the Hodge star squares to the identity in Euclidean signature.

Next, we discuss the spinor representations. In accordance with the isomorphism $\mathfrak{so}(4) \cong \mathfrak{su}(2)\oplus\mathfrak{su}(2)$, the six Lorentz generators split into two sets of $\mathfrak{su}(2)$ generators, which are symmetric 2×2 matrices normalized according to

$$\begin{aligned}\delta^a{}_b &= (\tilde{t}^a)_{\dot\alpha\dot\beta}\,(\tilde{t}_b)^{\dot\alpha\dot\beta} + (t^a)_{\alpha\beta}\,(t_b)^{\alpha\beta}\,,\\ (\tilde{t}_a)^{\dot\alpha\dot\beta}\,(\tilde{t}^a)_{\dot\gamma\dot\delta} &= \delta^{(\dot\alpha}{}_{\dot\gamma}\,\delta^{\dot\beta)}{}_{\dot\delta}\,, \quad (t_a)^{\alpha\beta}\,(t^a)_{\gamma\delta} = \delta_\gamma{}^{(\alpha}\,\delta_\delta{}^{\beta)}\,.\end{aligned} \tag{A.10}$$

In accordance with the above, adjoint indices are unpacked to spinor indices as

$$\begin{aligned}&\tilde{X}^{\dot\alpha\dot\beta} = (\tilde{t}_a)^{\dot\alpha\dot\beta}X^a\,, \quad X^a = (\tilde{t}^a)_{\dot\alpha\dot\beta}\,\tilde{X}^{\dot\alpha\dot\beta}\,, \quad \tilde{Y}_{\dot\alpha\dot\beta} = Y_a\,(\tilde{t}^a)_{\dot\alpha\dot\beta}\,, \quad Y_a = \tilde{Y}_{\dot\alpha\dot\beta}\,(\tilde{t}_a)^{\dot\alpha\dot\beta}\,,\\ &X^{\alpha\beta} = (t_a)^{\alpha\beta}X^a\,, \quad X^a = (t^a)_{\alpha\beta}\,X^{\alpha\beta}\,, \quad Y_{\alpha\beta} = Y_a\,(t^a)_{\alpha\beta}\,, \quad Y_a = Y_{\alpha\beta}\,(t_a)^{\alpha\beta}\,.\end{aligned} \tag{A.11}$$

For example, it follows that

$$X^a = (\tilde{t}^a)_{\dot\alpha\dot\beta}\,\tilde{X}^{\dot\alpha\dot\beta} + (t^a)_{\alpha\beta}\,X^{\alpha\beta}\,, \quad Y_a\,X^a = \tilde{Y}_{\dot\alpha\dot\beta}\,\tilde{X}^{\dot\alpha\dot\beta} + Y_{\alpha\beta}\,X^{\alpha\beta}\,. \tag{A.12}$$

Also, the structure constants are given as

$$\begin{aligned}(t_a)^{\dot\alpha\dot\beta}\,f^a{}_{bc}\,X_1^b\,X_2^c &= (\tilde{X}_1)^{\dot\alpha\dot\gamma}\,\tilde{\epsilon}_{\dot\gamma\dot\delta}\,(\tilde{X}_2)^{\dot\delta\dot\beta} - (\tilde{X}_2)^{\dot\alpha\dot\gamma}\,\tilde{\epsilon}_{\dot\gamma\dot\delta}\,(\tilde{X}_1)^{\dot\delta\dot\beta}\,,\\ (t_a)^{\alpha\beta}\,f^a{}_{bc}\,X_1^b\,X_2^c &= (X_1)^{\alpha\beta}\,\epsilon_{\beta\delta}\,(X_2)^{\delta\gamma} - (X_2)^{\alpha\beta}\,\epsilon_{\beta\delta}\,(X_1)^{\delta\gamma}\,.\end{aligned} \tag{A.13}$$

The dotted generators describe self-dual (right-handed) rotations while the undotted generators describe anti-self-dual (left-handed) rotations:

$$\star_{ab}\,(\tilde{t}^b)_{\dot\alpha\dot\beta} = +\tilde{\epsilon}_{\dot\alpha\dot\gamma}\tilde{\epsilon}_{\dot\beta\dot\delta}\,(\tilde{t}_a)^{\dot\gamma\dot\delta}\,, \quad \star_{ab}\,(t^b)_{\alpha\beta} = -\epsilon_{\alpha\gamma}\epsilon_{\beta\delta}\,(t_a)^{\gamma\delta}\,. \tag{A.14}$$

[24] In fact, Plebański gravity without the Immirzi constant can be described solely in terms of this split-signature Killing form.

In fact, self-dual and anti-self-dual projectors arise as

$$(\tilde{t}^a)_{\dot{\alpha}\dot{\beta}}\,(\tilde{t}^b)_{\dot{\gamma}\dot{\delta}}\,\tilde{\epsilon}^{\dot{\alpha}\dot{\gamma}}\,\tilde{\epsilon}^{\dot{\beta}\dot{\delta}} = \tfrac{1}{2}\,(\delta^{ab} + \star^{ab})\,, \quad (t^a)_{\alpha\beta}\,(t^b)_{\gamma\delta}\,\epsilon^{\alpha\gamma}\,\epsilon^{\beta\delta} = \tfrac{1}{2}\,(\delta^{ab} - \star^{ab})\,. \tag{A.15}$$

Having stated our conventions abstractly, let us be slightly more concrete and note that the self-dual and anti-self-dual generators are

$$(\tilde{t}^{AB})_{\dot{\alpha}\dot{\beta}} = \tfrac{1}{2}\,(\tilde{\epsilon}\tilde{\sigma}^{[A}\sigma^{B]})_{\dot{\alpha}\dot{\beta}}\,, \quad (t^{AB})_{\alpha\beta} = \tfrac{1}{2}\,(\sigma^{[A}\tilde{\sigma}^{B]}\epsilon)_{\alpha\beta}\,, \tag{A.16}$$

where an explicit representation of the Euclidean sigma matrices is given as

$$v_A\,(\sigma^A)_{\alpha\dot{\alpha}} = \begin{pmatrix} iv_4 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & iv_4 - v_3 \end{pmatrix}, \quad v_A\,(\tilde{\sigma}^A)^{\dot{\alpha}\alpha} = \begin{pmatrix} iv_4 - v_3 & -v_1 + iv_2 \\ -v_1 - iv_2 & iv_4 + v_3 \end{pmatrix}, \tag{A.17}$$

with the convention $\tilde{\epsilon}^{01} = \epsilon^{01} = \tilde{\epsilon}_{10} = \epsilon_{10} = +1$ for the epsilon tensors.

Finally, we end with a demonstration of index conversions in the context of tetradic Palatini gravity. Let the spinor indices be raised and lowered as

$$\psi^\alpha = \epsilon^{\alpha\beta}\psi_\beta\,, \quad \psi_\alpha = \epsilon_{\alpha\beta}\,\psi^\beta\,, \quad \tilde{\psi}^{\dot{\alpha}} = \tilde{\epsilon}^{\dot{\alpha}\dot{\beta}}\tilde{\psi}_{\dot{\beta}}\,, \quad \tilde{\psi}_{\dot{\alpha}} = \tilde{\epsilon}_{\dot{\alpha}\dot{\beta}}\,\tilde{\psi}^{\dot{\beta}}\,. \tag{A.18}$$

The spin connection splits into self-dual and anti-self-dual parts as

$$\omega^a = (\tilde{t}^a)_{\dot{\alpha}\dot{\beta}}\,\tilde{\omega}^{\dot{\alpha}\dot{\beta}} + (t^a)_{\alpha\beta}\,\omega^{\alpha\beta}\,. \tag{A.19}$$

The self-dual and anti-self-dual parts of the field strength are given by

$$\tilde{R}^{\dot{\alpha}}{}_{\dot{\beta}} = d\tilde{\omega}^{\dot{\alpha}}{}_{\dot{\beta}} - \tilde{\omega}^{\dot{\alpha}}{}_{\dot{\gamma}} \wedge \tilde{\omega}^{\dot{\gamma}}{}_{\dot{\beta}}\,, \quad R_\alpha{}^\beta = d\omega_\alpha{}^\beta + \omega_\alpha{}^\gamma \wedge \omega_\gamma{}^\beta\,. \tag{A.20}$$

Meanwhile, it is convenient to convert the vector index of the tetrad to spinor indices with a customary factor of $-1/2$, which originates from the -2 of $(\tilde{\sigma}_A)^{\dot{\alpha}\alpha}\,(\sigma^A)_{\beta\dot{\beta}} = -2\,\delta_\beta{}^\alpha\,\delta^{\dot{\alpha}}{}_{\dot{\beta}}$:

$$e^{\dot{\alpha}\alpha} = -\frac{1}{2}\,(\tilde{\sigma}_A)^{\dot{\alpha}\alpha}\,e^A\,. \tag{A.21}$$

Then the spinor components of the Plebański two-form defined in Eq. (3.6) are given by

$$\tilde{B}^{\dot{\alpha}\dot{\beta}} = \epsilon_{\alpha\beta}\,e^{\dot{\alpha}\alpha} \wedge e^{\dot{\beta}\beta}\,, \quad B^{\alpha\beta} = -\tilde{\epsilon}_{\dot{\alpha}\dot{\beta}}\,e^{\dot{\alpha}\alpha} \wedge e^{\dot{\beta}\beta}\,. \tag{A.22}$$

In turn, the Lagrangian four-form in Eq. (3.7) boils down to

$$-\frac{1}{g^2}\left(e^{\dot{\alpha}\alpha} \wedge e_{\alpha\dot{\beta}} \wedge \tilde{R}^{\dot{\beta}}{}_{\dot{\alpha}} + e_{\beta\dot{\alpha}} \wedge e^{\dot{\alpha}\alpha} \wedge R_\alpha{}^\beta\right) + \frac{\Lambda}{3g^2}\,e_{\alpha\dot{\alpha}} \wedge e^{\dot{\alpha}\beta} \wedge e_{\beta\dot{\beta}} \wedge e^{\dot{\beta}\alpha}\,. \tag{A.23}$$

The above expressions are consistent with the chiral symmetry operators in Eq. (3.23).

B Geometry of Linking

In this appendix, we expand on the technical details of the Dirac delta form and topological numbers in general spacetime dimension. Here we will indicate the dimensionality of each manifold with a subscript. All manifolds will be assumed to be orientable.

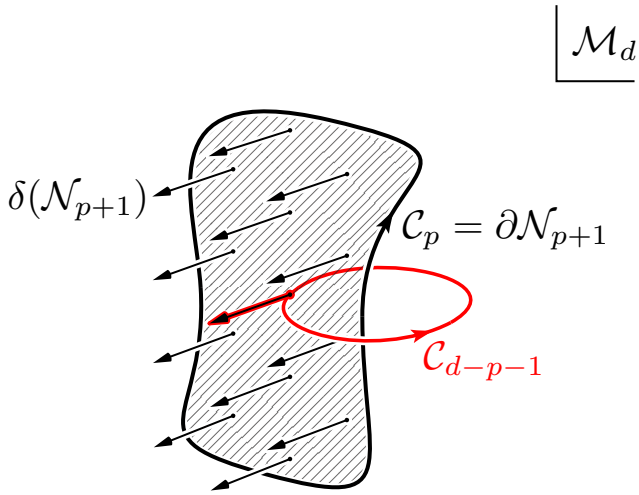


Figure 12. Linking between two submanifolds in d dimensions.

A. Definitions

Consider a p-dimensional submanifold $\mathcal{N}_p$ in a d-dimensional manifold $\mathcal{M}_d$. We define a differential form analog of Dirac's delta distribution as

$$\int_{\mathcal{N}_p} \alpha^{(p)} \;=\; \int_{\mathcal{M}_d} \alpha^{(p)} \wedge \delta(\mathcal{N}_p)\,, \tag{B.1}$$

where $\alpha^{(p)}$ is an arbitrary smooth p-form in $\mathcal{M}_d$. Evidently, $\delta(\mathcal{N}_p)$ is a differential $(d-p)$-form distribution that peaks on the submanifold $\mathcal{N}_p$ while vanishing elsewhere.

More explicitly, suppose $\mathcal{N}_p$ is parameterized by coordinates $\sigma^1, \sigma^2, \cdots, \sigma^p$ as $x^\mu = X^\mu(\sigma)$, where x^μ denote coordinates of $\mathcal{M}_d$. Then $\delta(\mathcal{N}_p)$ is given by

$$\frac{1}{(d-p)!}\left[\int d^p\sigma\; \delta^{(d)}(x - X(\sigma))\, \frac{\partial X^{\lambda_1}}{\partial\sigma^1} \cdots \frac{\partial X^{\lambda_p}}{\partial\sigma^p}\right] \epsilon_{\lambda_1\cdots\lambda_p\mu_1\cdots\mu_{d-p}}\, dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_{d-p}}\,, \tag{B.2}$$

where $\delta^{(d)}(x - X(\sigma))$ is the ordinary Dirac delta function. This is higher-form generalization of Dirac delta function will be referred to as a "*Dirac delta form.*" The ordinary Dirac delta corresponds to $p = 0$. It is also interesting to note that $\delta(\mathcal{N}_d)$ for a top-dimensional submanifold $\mathcal{N}_d$ is its characteristic function, e.g., $\delta(\mathcal{M}_d) = 1$.

In terms of Dirac delta forms, the *intersection number* between two orientable submanifolds $\mathcal{N}_p$ and $\mathcal{N}_{d-p}$ in $\mathcal{M}_d$ can be defined in a coordinate-free fashion:

$$\mathrm{int}(\mathcal{N}_p, \mathcal{N}_{d-p}) \;=\; \int_{\mathcal{M}_d} \delta(\mathcal{N}_p) \wedge \delta(\mathcal{N}_{d-p})\,. \tag{B.3}$$

Note that the integrand $\delta(\mathcal{N}_p) \wedge \delta(\mathcal{N}_{d-p})$ is a top form that localizes at the intersection points. In fact, it could be argued that $\delta(\mathcal{N}_p) \wedge \delta(\mathcal{N}_{d-p}) = \delta(\mathcal{N}_p \cap \mathcal{N}_{d-p})$. It is instructive to check if Eq. (B.3) defines the intersection number in an expected way with specific values

of p and d, such as $p = 0$, or $d = 3$. Also, it is easy to see that

$$\mathrm{int}(\mathcal{N}_p, \mathcal{N}_{d-p}) = (-1)^{p(d-p)}\, \mathrm{int}(\mathcal{N}_{d-p}, \mathcal{N}_p)\,. \tag{B.4}$$

Given Eq. (B.3), we can easily define the *linking number* between two *closed* orientable submanifolds $\mathcal{C}_p$ and $\mathcal{C}_{d-p-1}$ provided that one of them is exact. Without loss of generality, suppose $\mathcal{C}_p = \partial\mathcal{N}_{p+1}$. Then we define

$$\mathrm{link}_*(\partial\mathcal{N}_{p+1}, \mathcal{C}_{d-p-1}) \;=\; \mathrm{int}(\mathcal{N}_{p+1}, \mathcal{C}_{d-p-1}) \;=\; \int_{\mathcal{C}_{d-p-1}} \delta(\mathcal{N}_{p+1})\,, \tag{B.5}$$

which counts how many times $\mathcal{C}_{d-p}$ penetrates the coboundary $\mathcal{N}_{p+1}$ as depicted in Fig. 12.

Note that the above definition necessarily puts the two arguments of link_* in an unequal footing: the first argument must be exact. However, as practiced in the main text, it can be convenient to use a more handy notation for the linking number such that the exact argument can be placed in any slot. With hindsight, we define such a notation "link" to satisfy the following relations:

$$\begin{aligned} \mathrm{link}(\partial\mathcal{N}_{d-p}, \mathcal{C}_p) &= \mathrm{link}_*(\partial\mathcal{N}_{d-p}, \mathcal{C}_p)\,, \\ \mathrm{link}(\mathcal{C}_{d-p-1}, \mathcal{C}_p) &= (-1)^{dp+1}\, \mathrm{link}(\mathcal{C}_p, \mathcal{C}_{d-p-1})\,. \end{aligned} \tag{B.6}$$

The first equation states that link and link_* give the same number when the first argument is exact. The second equation allows the arguments of link to be freely rearranged, given the implicit assumption that one of the arguments is exact. The factor $(-1)^{dp+1}$ is necessitated from consistency with Eq. (B.9). Note also that the following holds:

$$\mathrm{link}(\mathcal{C}_{d-p-1}, \partial\mathcal{N}_{p+1}) = (-1)^{d-p}\, \mathrm{int}(\mathcal{C}_{d-p-1}, \mathcal{N}_{p+1})\,. \tag{B.7}$$

B. Identities

Using either the coordinate-free/axiomatic definition of the Dirac delta form in Eq. (B.1) or the explicit definition in Eq. (B.2), one can derive various identities of Dirac delta forms and also intersection and linking numbers. Some simple ones are

$$d\,\delta(\mathcal{N}_p) = (-1)^p\, \delta(\partial\mathcal{N}_p)\,, \tag{B.8}$$

$$\mathrm{link}(\partial\mathcal{N}_{p+1}, \partial\mathcal{N}_{d-p}) = (-1)^{dp+1}\, \mathrm{link}(\partial\mathcal{N}_{d-p}, \partial\mathcal{N}_{p+1})\,. \tag{B.9}$$

The first identity encodes Stokes theorem, while the second identity describes a duality in the linking number computation.

A few more identities are found in the context of Hamiltonian formalism. Suppose the spacetime is a product manifold $\mathcal{M}_d = \mathcal{X}_{d-1} \times \mathbb{R}$ with coordinates $x^\mu = (x^i, t)$, where the last[25] coordinate $x^{\mu=d} = t$ labels hypersurfaces foliating $\mathcal{M}_d$. Then we consider submanifolds $\mathcal{K}_p$ and $\mathcal{Y}_{d-p-1}$ of $\mathcal{X}_{d-1}$ that can possibly intersect, with the former being closed: $\partial\mathcal{K}_p = 0$. The corresponding relevant objects in spacetime are $\mathcal{C}_p$ and $\mathcal{N}_{d-p}$, defined as the following. First, we consider an embedding of the closed submanifold $\mathcal{K}_p$ in $\mathcal{M}_d$:

$$\mathcal{C}_p = \mathcal{K}_p \times \{t_0\}\,. \tag{B.10}$$

[25] This matches the convention chosen in Eq. (B.1) where the Dirac delta form is appended from the right.

Second, we consider a timelike $(d-p)$-dimensional submanifold $\mathcal{N}_{d-p}$ of $\mathcal{M}_d$ defined as the following, where $\mathcal{I} = [t_-, t_+]$ is an interval in $\mathbb{R}$ such that $t_- < t_0 < t_+$:

$$(-1)^{d-p}\mathcal{N}_{d-p} = \mathcal{Y}_{d-p-1} \times \mathcal{I}\,. \tag{B.11}$$

The customary orientation of $\mathcal{N}_{d-p}$ by a sign factor $(-1)^{d-p}$ stipulates that its boundary is given in the form

$$\partial\mathcal{N}_{d-p} \;=\; \Big(-\mathcal{Y}_{d-p-1} \times \{t_+\}\Big) \cup \Big(\mathcal{Y}_{d-p-1} \times \{t_-\}\Big) \cup \Gamma\,, \tag{B.12}$$

for a timelike submanifold Γ in $\mathcal{M}_d$. Specifically, the consistency between Eqs. (B.11) and (B.12) can be established by considering the Stokes theorem relating $\mathcal{N}_{d-p}$ and $\partial\mathcal{N}_{d-p}$, where a factor of $(-1)^{d-p-1}$ arising from rearranging the order of the time differential dt combines with the minus sign in Eq. (B.12) to give $(-1)^{d-p}$.

With these definitions, it eventually follows that

$$(-1)^{d-p}\,\mathrm{link}(\partial\mathcal{N}_{d-p}, \mathcal{C}_p) \;=\; \mathrm{int}(\mathcal{Y}_{d-p-1} \times \mathcal{I}, \mathcal{C}_p) \;=\; (-1)^p\,\mathrm{int}_{d-1}(\mathcal{Y}_{d-p-1}, \mathcal{K}_p)\,, \tag{B.13}$$

where $\mathrm{int}_{d-1}(\mathcal{Y}_{d-p-1}, \mathcal{K}_p)$ is the intersection number between $\mathcal{Y}_{d-p-1}$ and $\mathcal{K}_p$ in $\mathcal{X}_{d-1}$:

$$\mathrm{int}_{d-1}(\mathcal{Y}_{d-p-1}, \mathcal{K}_p) \;=\; \int_{\mathcal{X}_{d-1}} \delta_{d-1}(\mathcal{Y}_{d-p-1}) \wedge \delta_{d-1}(\mathcal{K}_p)\,. \tag{B.14}$$

To derive Eq. (B.13), one may work in component terms with Eq. (B.2). One will find that the factor of $(-1)^p$ arises from an index rearrangement,

$$\epsilon_{i_1\cdots i_{d-p-1} d\, j_1\cdots j_p} = (-1)^p\,\epsilon_{i_1\cdots i_{d-p-1} j_1\cdots j_p d} = (-1)^p\,\epsilon_{i_1\cdots i_{d-p-1} j_1\cdots j_p}\,, \tag{B.15}$$

where we work with the convention such that $\epsilon_{12\cdots d} = +1$.

c. Example: Abelian p-Form Symmetry in d Dimensions

Having derived various sign factors, we can describe a universal sign convention for higher-form symmetries of general rank p in Euclidean signature spacetimes of general d dimensions. Let us take an abelian BF theory as a prototypical model,[26] whose action reads

$$S \;=\; \int_{\mathcal{M}_d} B^{(d-p-1)} \wedge F^{(p+1)} + f(B^{(d-p-1)})\,, \tag{B.16}$$

where $F^{(p+1)} = dA^{(p)}$. Here the superscripts denote the ranks, while $f(B^{(d-p-1)})$ is a d-form functional of $B^{(d-p-1)}$.

First, let us describe the p-form symmetry in the covariant formalism. For simplicity, consider the Wilson loop for an exact support:

$$W_q(\partial\mathcal{N}_{p+1}) = \exp\left(q \int_{\mathcal{N}_{p+1}} F^{(p+1)}\right). \tag{B.17}$$

[26] For example, it is an amusing check to consider the case of $d=1$, $p=0$ with a Lagrangian one-form $P\,dX - H(P)\,dt$, which describes a point particle. In particular, it readily follows that operators $U_\epsilon([t_1,t_2]) = e^{-\epsilon P(t_2)} e^{\epsilon P(t_1)}$ and $Q(t) = P(t)$ implement translations of X, so for example $e^{-\epsilon P} X\, e^{\epsilon P} = X + \epsilon$.

We want this to be transformed as

$$W_q(\partial\mathcal{N}_{p+1}) \;\mapsto\; \exp\Big(q\varepsilon\,\mathrm{int}(\mathcal{C}_{d-p-1},\mathcal{N}_{p+1})\Big)\,W_q(\partial\mathcal{N}_{p+1})\,, \tag{B.18}$$

when the symmetry operator is inserted along an exact submanifold $\mathcal{C}_{d-p-1} = \partial\mathcal{N}_{d-p}$. Clearly, this can be *undone* by a field redefinition that shifts the field strength as $F^{(p+1)} \mapsto F^{(p+1)} - \varepsilon\,\delta(\mathcal{C}_{d-p-1})$,[27] which, in turn, transforms the action in Eq. (B.16) as

$$-S \;\mapsto\; -S + \varepsilon\int_{\mathcal{M}_d} B^{(d-p-1)}\wedge\delta(\mathcal{C}_{d-p-1})\,, \tag{B.19}$$

from which the symmetry operator is identified:

$$U_\varepsilon(\mathcal{C}_{d-p-1}) = \exp\left(\varepsilon\int_{\mathcal{C}_{d-p-1}} B^{(d-p-1)}\right). \tag{B.20}$$

Namely, the symmetry operator is derived as the term that is generated from the action term in the path integral after a field transformation that pushes the symmetry charge back into the defect operator. Crucially, the BF structure leverages between the both sides of the Ward identity, which reads

$$\big\langle U_\varepsilon(\mathcal{C}_{d-p-1})\,W_q(\partial\mathcal{N}_{p+1})\big\rangle = \exp\Big(q\varepsilon\,\mathrm{int}(\mathcal{C}_{d-p-1},\mathcal{N}_{p+1})\Big)\,\big\langle W_q(\partial\mathcal{N}_{p+1})\big\rangle\,. \tag{B.21}$$

Finally, denoting $\mathcal{C}_p = \partial\mathcal{N}_{p+1}$, we can state the Ward identity in terms of a linking number:

$$\big\langle U_\varepsilon(\mathcal{C}_{d-p-1})\,W_q(\mathcal{C}_p)\big\rangle = \exp\Big((-1)^{d-p}q\varepsilon\,\mathrm{link}(\mathcal{C}_{d-p-1},\mathcal{C}_p)\Big)\,\big\langle W_q(\mathcal{C}_p)\big\rangle\,, \tag{B.22}$$

which follows through using Eq. (B.7) (or Eqs. (B.4), (B.5), and (B.9) with link_*).

Note that the Ward identities in the main text are reproduced when one takes $d=4$ and $p=1$ and makes the following identifications:

$$\mathcal{C}_p \leftrightarrow \mathcal{C}\,,\quad \mathcal{C}_{d-p-1} \leftrightarrow \mathcal{S}\,,\quad \mathcal{N}_{d-p} \leftrightarrow \mathcal{V}\,. \tag{B.23}$$

Next, we can also double check from the angle of Hamiltonian formalism. Eq. (B.12) as a convetion for pancaking the coboundary $\mathcal{N}_{d-p}$ of the symmetry operator down to a submanifold $\mathcal{Y}_{d-p-1}$ in $\mathcal{X}_{d-1}$ stipulates that the Ward identity in Eq. (B.22) boils down to a conjugation of the defect operator as

$$e^{-\varepsilon Q(\mathcal{Y}_{d-p-1})}\,W_q(\mathcal{K}_p)\,e^{\varepsilon Q(\mathcal{Y}_{d-p-1})} = \exp\Big((-1)^p q\varepsilon\,\mathrm{int}_{d-1}(\mathcal{Y}_{d-p-1},\mathcal{K}_p)\Big)\,W_q(\mathcal{C}_p)\,, \tag{B.24}$$

provided that the charge operator is defined as

$$Q(\mathcal{Y}_{d-p-1}) = \int_{\mathcal{Y}_{d-p-1}} B^{(d-p-1)} \tag{B.25}$$

[27] Concretely, from Eq. (B.8) it can be realized as $A^{(p)} \mapsto A^{(p)} - (-1)^{d-p}\,\epsilon\,\delta(\mathcal{N}_{d-p})$, $B^{(d-p-1)} \mapsto B^{(d-p-1)}$.

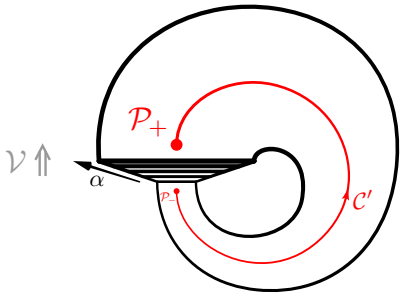


Figure 13. The center periodicity condition on the multivalued group parameter Ω, in terms of the discontinuity $\lim_{\mathcal{P}_\pm \to \mathcal{P}} (\Omega(\mathcal{P}_+)\Omega^{-1}(\mathcal{P}_-)) = \alpha$ across a branch cut $\mathcal{V}$.

as an object that lives in the $d-1$ dimensions. Note that the identity in Eq. (B.13) has been used to deduce Eq. (B.24) from Eq. (B.22). It is straightforward to verify Eq. (B.24) from the canonical commutation relations:

$$\begin{aligned} [A_{i_1\cdots i_p}, B_{k_1\cdots k_{d-p-1}}]\,\epsilon^{k_1\cdots k_{d-p-1}j_1\cdots j_p} &= (d-p-1)!\,p!\,(-1)^p\,\delta^{j_1}{}_{[i_1}\cdots\delta^{j_p}{}_{i_p]}\,, \\ \Longrightarrow \quad [A_{i_1\cdots i_p}, Q(\mathcal{Y}_{d-p-1})] &= (-1)^p\,(\delta_{d-1}(\mathcal{Y}_{d-p-1}))_{i_1\cdots i_p}\,, \end{aligned} \tag{B.26}$$

where the factor $(-1)^p$ precisely arises from the same index rearrangement as in Eq. (B.15), provided the conventions $\epsilon_{12\cdots d} = +1$ and $t = x^d$.

The above discussion corresponds to the calculation carried out in Secs. 2.2.2 and 3.2.2, so one can check consistency by taking $d=4$ and $p=1$. Yet, note that the spatial surface "$\mathcal{S}_0$" there corresponds to $\mathcal{Y}_{d-p-1}$ here with a flip of orientation:

$$\mathcal{K}_p \leftrightarrow \mathcal{C}\,, \quad \mathcal{Y}_{d-p-1} \leftrightarrow -\mathcal{S}_0\,. \tag{B.27}$$

This single flip is because the main text attempts to avoid any possible distractions from convention-dependent minus signs.

D. Proof of Equivalence between Eqs. (2.8) and (2.15)

Before ending, we should show that Eqs. (2.8) and (2.15), used in the main text, are equivalent statements. In d-dimensional spacetime, the claim reads

$$\begin{aligned} &\lim_{\mathcal{P}_\pm \to \mathcal{P}} \Omega(\mathcal{P}_+)\Omega^{-1}(\mathcal{P}_-) = \alpha \in Z(G) \\ &\Longleftrightarrow \quad \exists\,\lambda^a \qquad \text{s.t.} \qquad (\Omega^{-1}d\Omega)^a = \lambda^a\,\delta(\partial\mathcal{V}_{d-1}) \qquad \text{and} \qquad e^{2\pi\lambda} = \alpha^{(-1)^d}\,, \end{aligned} \tag{B.28}$$

where $\mathcal{P}_+$ and $\mathcal{P}_-$ are points infinitesimally deviating from a point on $\mathcal{V}_{d-1}$ from above and below. The notion of above and below is well-defined, as $\mathcal{V}_{d-1}$ is oriented and codimension one. As shown below, the identities stated in Eqs. (B.4) and (B.9) play a role in the proof.

First of all, the fact that Ω as a multivalued function is everywhere smooth off the surface $\partial\mathcal{V}_{d-1}$ implies that $dd\Omega$ localizes on $\partial\mathcal{V}_{d-1}$ and vice versa, which is in turn equivalent to the existence of an algebra-valued zero-form λ^a such that $(\Omega^{-1}d\Omega)^a = 2\pi\lambda^a\,\delta(\partial\mathcal{V}_{d-1})$.

Thus, it remains to verify that the center periodicity condition on Ω is equivalent to $e^{2\pi\lambda} = \alpha^{(-1)^d}$. It suffices to work with an exact one-dimensional loop $\mathcal{C}_1 = \partial\mathcal{N}_2$ in the

vicinity of $\partial\mathcal{V}_{d-1}$. Suppose $\mathrm{int}(\mathcal{V}_{d-1},\partial\mathcal{N}_2)=-1$. Then the periodicity condition implies

$$\lim_{\mathcal{C}_1'\to\mathcal{C}_1}\mathrm{Pexp}\left(\oint_{\mathcal{C}_1'}\Omega^{-1}d\Omega\right)=\lim_{\mathcal{C}_1'\to\mathcal{C}_1}\Omega^{-1}(\mathcal{P}_-)\,\Omega(\mathcal{P}_+)=\alpha\,, \tag{B.29}$$

where $\mathcal{C}_1'$ is the infinitesimal opening of $\mathcal{C}_1$ described before. Meanwhile, the left-hand side can be computed as a surface-ordered integral, thanks to the nonabelian Stokes theorem:

$$\mathrm{Pexp}\left(\oint_{\partial\mathcal{N}_2}\Omega^{-1}d\Omega\right)=\mathrm{Pexp}\left(\int_{\mathcal{N}_2}\Omega^{-1}dd\Omega\right)=\exp\left(\int_{\mathcal{M}_d}\Omega^{-1}dd\Omega\wedge\delta(\mathcal{N}_2)\right). \tag{B.30}$$

Note that the last equality has dropped the surface ordering because the integrand localizes at a single point. Plugging in $(\Omega^{-1}dd\Omega)^a=2\pi\lambda^a\,\delta(\partial\mathcal{V})$, the right-hand side boils down to $\exp(2\pi\lambda\,\mathrm{int}(\partial\mathcal{V}_{d-1},\mathcal{N}_2))$, which equals $\exp((-1)^d 2\pi\lambda)$ given $\mathrm{int}(\mathcal{V}_{d-1},\partial\mathcal{N}_2)=-1$ by Eqs. (B.4) and (B.9). Therefore, demanding that the outcomes of Eqs. (B.29) and (B.30) are the same, we find $\alpha=\exp((-1)^d 2\pi\lambda)$, i.e., $e^{2\pi\lambda}=\alpha^{(-1)^d}$. Lastly, recalling the fact that $\alpha\in Z(G)$, it can be seen that supposing generic curves with arbitrary linking numbers does not impose a further condition on λ^a. This concludes the proof.

To make it certain that all the signs have worked properly, it is instructive to double check by considering the abelian version of the statement:

$$\lim_{\mathcal{P}_\pm\to\mathcal{P}}\chi(\mathcal{P}_+)-\chi(\mathcal{P}_-)=(-1)^d 2\pi\lambda\quad\Longleftrightarrow\quad dd\chi=2\pi\lambda\,\delta(\partial\mathcal{V}_{d-1})\,. \tag{B.31}$$

Here λ is a constant. The periodicity condition in the left-hand side is equivalent to

$$d\chi+(-1)^d 2\pi\lambda\,\delta(\mathcal{V}_{d-1})=df\,, \tag{B.32}$$

for some single-valued zero-form f; i.e., $d\chi$ is "gauge equivalent" to $(-1)^{d-1}2\pi\lambda\,\delta(\mathcal{V}_{d-1})$. Thus it follows that $dd\chi=(-1)^{d-1}2\pi\lambda\,d\delta(\mathcal{V}_{d-1})$, which equals $2\pi\lambda\,\delta(\partial\mathcal{V}_{d-1})$ by Eq. (B.8).

Note that the points $\mathcal{P}_+$ and $\mathcal{P}_-$ in Fig. 2 are correctly deviating from above and below as stated in Eq. (B.28) ($-\mathrm{link}(\mathcal{C},\mathcal{S})=\mathrm{link}(\mathcal{S},\mathcal{C})=\mathrm{int}(\mathcal{V},\mathcal{C})=-1$), but the three-dimensional visualization seemingly appears the opposite due to the relative orientation with the hidden fourth dimension.

References

[1] M. G. Alford and J. March-Russell, "Discrete gauge theories," *Int. J. Mod. Phys. B* **5** (1991) 2641–2674.

[2] M. G. Alford and J. March-Russell, "New order parameters for non-Abelian gauge theories," *Nucl. Phys. B* **369** (1992) 276–298.

[3] M. G. Alford, K.-M. Lee, J. March-Russell, and J. Preskill, "Quantum field theory of non-Abelian strings and vortices," *Nucl. Phys. B* **384** (1992) 251–317, arXiv:hep-th/9112038.

[4] M. Bucher, K.-M. Lee, and J. Preskill, "On detecting discrete Cheshire charge," *Nucl. Phys. B* **386** (1992) 27–42, arXiv:hep-th/9112040.

[5] T. Pantev and E. Sharpe, "Notes on gauging noneffective group actions," arXiv:hep-th/0502027.

[6] T. Pantev and E. Sharpe, "String compactifications on Calabi-Yau stacks," *Nucl. Phys. B* **733** (2006) 233–296, arXiv:hep-th/0502044.

[7] T. Pantev and E. Sharpe, "GLSM's for Gerbes (and other toric stacks)," *Adv. Theor. Math. Phys.* **10** no. 1, (2006) 77–121, arXiv:hep-th/0502053.

[8] S. Hellerman, A. Henriques, T. Pantev, E. Sharpe, and M. Ando, "Cluster decomposition, T-duality, and gerby CFT's," *Adv. Theor. Math. Phys.* **11** no. 5, (2007) 751–818, arXiv:hep-th/0606034.

[9] Z. Nussinov and G. Ortiz, "A symmetry principle for topological quantum order," *Annals Phys.* **324** (2009) 977–1057, arXiv:cond-mat/0702377.

[10] O. Aharony, N. Seiberg, and Y. Tachikawa, "Reading between the lines of four-dimensional gauge theories," *JHEP* **08** (2013) 115, arXiv:1305.0318 [hep-th].

[11] D. Gaiotto, A. Kapustin, N. Seiberg, and B. Willett, "Generalized Global Symmetries," *JHEP* **02** (2015) 172, arXiv:1412.5148 [hep-th].

[12] N. Iqbal and J. McGreevy, "Mean string field theory: Landau-Ginzburg theory for 1-form symmetries," *SciPost Phys.* **13** (2022) 114, arXiv:2106.12610 [hep-th].

[13] X.-G. Wen, "Emergent anomalous higher symmetries from topological order and from dynamical electromagnetic field in condensed matter systems," *Phys. Rev. B* **99** no. 20, (2019) 205139, arXiv:1812.02517 [cond-mat.str-el].

[14] M. A. Levin and X.-G. Wen, "String net condensation: A Physical mechanism for topological phases," *Phys. Rev. B* **71** (2005) 045110, arXiv:cond-mat/0404617.

[15] M. A. Levin and X.-G. Wen, "Colloquium: Photons and electrons as emergent phenomena," *Rev. Mod. Phys.* **77** (2005) 871–879, arXiv:cond-mat/0407140.

[16] M. B. Hastings and X.-G. Wen, "Quasi-adiabatic continuation of quantum states: The Stability of topological ground state degeneracy and emergent gauge invariance," *Phys. Rev. B* **72** (2005) 045141, arXiv:cond-mat/0503554.

[17] H. Shimizu and K. Yonekura, "Anomaly constraints on deconfinement and chiral phase transition," *Phys. Rev. D* **97** no. 10, (2018) 105011, arXiv:1706.06104 [hep-th].

[18] Y. Choi, H. T. Lam, and S.-H. Shao, "Noninvertible Global Symmetries in the Standard Model," *Phys. Rev. Lett.* **129** no. 16, (2022) 161601, arXiv:2205.05086 [hep-th].

[19] C. Córdova and K. Ohmori, "Noninvertible Chiral Symmetry and Exponential Hierarchies," *Phys. Rev. X* **13** no. 1, (2023) 011034, arXiv:2205.06243 [hep-th].

[20] D. Gaiotto, A. Kapustin, Z. Komargodski, and N. Seiberg, "Theta, Time Reversal, and Temperature," *JHEP* **05** (2017) 091, arXiv:1703.00501 [hep-th].

[21] Z. Wan, J. Wang, and Y. Zheng, "New higher anomalies, SU(N) Yang–Mills gauge theory and $\mathbb{CP}^{N-1}$ sigma model," *Annals Phys.* **414** (2020) 168074, arXiv:1812.11968 [hep-th].

[22] L. V. Delacrétaz, D. M. Hofman, and G. Mathys, "Superfluids as Higher-form Anomalies," *SciPost Phys.* **8** (2020) 047, arXiv:1908.06977 [hep-th].

[23] P.-S. Hsin, H. T. Lam, and N. Seiberg, "Comments on One-Form Global Symmetries and Their Gauging in 3d and 4d," *SciPost Phys.* **6** no. 3, (2019) 039, arXiv:1812.04716 [hep-th].

[24] Y. Tanizaki, Y. Kikuchi, T. Misumi, and N. Sakai, "Anomaly matching for the phase diagram of massless $\mathbb{Z}_N$-QCD," *Phys. Rev. D* **97** no. 5, (2018) 054012, arXiv:1711.10487 [hep-th].

[25] C. Córdova and K. Ohmori, "Anomaly Obstructions to Symmetry Preserving Gapped Phases," arXiv:1910.04962 [hep-th].

[26] Z. Wan and J. Wang, "Adjoint QCD_4, Deconfined Critical Phenomena, Symmetry-Enriched Topological Quantum Field Theory, and Higher Symmetry-Extension," *Phys. Rev. D* **99** no. 6, (2019) 065013, arXiv:1812.11955 [hep-th].

[27] C. Córdova, D. S. Freed, H. T. Lam, and N. Seiberg, "Anomalies in the Space of Coupling Constants and Their Dynamical Applications I," *SciPost Phys.* **8** no. 1, (2020) 001, arXiv:1905.09315 [hep-th].

[28] C. Córdova, D. S. Freed, H. T. Lam, and N. Seiberg, "Anomalies in the Space of Coupling Constants and Their Dynamical Applications II," *SciPost Phys.* **8** no. 1, (2020) 002, arXiv:1905.13361 [hep-th].

[29] D. G. Delmastro, J. Gomis, P.-S. Hsin, and Z. Komargodski, "Anomalies and symmetry fractionalization," *SciPost Phys.* **15** no. 3, (2023) 079, arXiv:2206.15118 [hep-th].

[30] A. Kovner and B. Rosenstein, "New look at QED in four-dimensions: The Photon as a Goldstone boson and the topological interpretation of electric charge," *Phys. Rev. D* **49** (1994) 5571–5581, arXiv:hep-th/9210154.

[31] D. M. Hofman and N. Iqbal, "Goldstone modes and photonization for higher form symmetries," *SciPost Phys.* **6** no. 1, (2019) 006, arXiv:1802.09512 [hep-th].

[32] E. Lake, "Higher-form symmetries and spontaneous symmetry breaking," arXiv:1802.07747 [hep-th].

[33] N. Sogabe and N. Yamamoto, "Triangle Anomalies and Nonrelativistic Nambu-Goldstone Modes of Generalized Global Symmetries," *Phys. Rev. D* **99** no. 12, (2019) 125003, arXiv:1903.02846 [hep-th].

[34] I. Garcia Etxebarria and N. Iqbal, "A Goldstone theorem for continuous non-invertible symmetries," *JHEP* **09** (2023) 145, arXiv:2211.09570 [hep-th].

[35] C. Córdova and S. Koren, "Higher Flavor Symmetries in the Standard Model," *Annalen Phys.* **535** no. 8, (2023) 2300031, arXiv:2212.13193 [hep-ph].

[36] C. Córdova, S. Hong, S. Koren, and K. Ohmori, "Neutrino Masses from Generalized Symmetry Breaking," arXiv:2211.07639 [hep-ph].

[37] Y. Hidaka, M. Nitta, and R. Yokokura, "Higher-form symmetries and 3-group in axion electrodynamics," *Phys. Lett. B* **808** (2020) 135672, arXiv:2006.12532 [hep-th].

[38] Y. Hidaka, M. Nitta, and R. Yokokura, "Global 3-group symmetry and 't Hooft anomalies in axion electrodynamics," *JHEP* **01** (2021) 173, arXiv:2009.14368 [hep-th].

[39] T. D. Brennan and C. Córdova, "Axions, higher-groups, and emergent symmetry," *JHEP* **02** (2022) 145, arXiv:2011.09600 [hep-th].

[40] Y. Choi, H. T. Lam, and S.-H. Shao, "Non-invertible Gauss law and axions," *JHEP* **09** (2023) 067, arXiv:2212.04499 [hep-th].

[41] R. Yokokura, "Non-invertible symmetries in axion electrodynamics," arXiv:2212.05001 [hep-th].

[42] T. D. Brennan, S. Hong, and L.-T. Wang, "Coupling a Cosmic String to a TQFT," arXiv:2302.00777 [hep-ph].

[43] Y. Choi, M. Forslund, H. T. Lam, and S.-H. Shao, "Quantization of Axion-Gauge Couplings and Non-Invertible Higher Symmetries," arXiv:2309.03937 [hep-ph].

[44] C. Córdova, S. Hong, and L.-T. Wang, "Axion Domain Walls, Small Instantons, and Non-Invertible Symmetry Breaking," arXiv:2309.05636 [hep-ph].

[45] M. Reece, "Axion-gauge coupling quantization with a twist," *JHEP* **10** (2023) 116, arXiv:2309.03939 [hep-ph].

[46] P. Agrawal and A. Platschorre, "The monodromic axion-photon coupling," *JHEP* **01** (2024) 169, arXiv:2309.03934 [hep-th].

[47] J. F. Donoghue, "Quantum General Relativity and Effective Field Theory," arXiv:2211.09902 [hep-th].

[48] C. P. Burgess, "Quantum gravity in everyday life: General relativity as an effective field theory," *Living Rev. Rel.* **7** (2004) 5–56, arXiv:gr-qc/0311082.

[49] J. McNamara and M. Reece, "Reflections on Parity Breaking," arXiv:2212.00039 [hep-th].

[50] J. McNamara and C. Vafa, "Cobordism Classes and the Swampland," arXiv:1909.10355 [hep-th].

[51] K. Hinterbichler, D. M. Hofman, A. Joyce, and G. Mathys, "Gravity as a gapless phase and biform symmetries," *JHEP* **02** (2023) 151, arXiv:2205.12272 [hep-th].

[52] V. Benedetti, H. Casini, and J. M. Magan, "Generalized symmetries of the graviton," *JHEP* **05** (2022) 045, arXiv:2111.12089 [hep-th].

[53] V. Benedetti, P. Bueno, and J. M. Magan, "Generalized Symmetries for Generalized Gravitons," *Phys. Rev. Lett.* **131** no. 11, (2023) 111603, arXiv:2305.13361 [hep-th].

[54] P. R. S. Gomes, "An introduction to higher-form symmetries," *SciPost Phys. Lect. Notes* **74** (2023) 1, arXiv:2303.01817 [hep-th].

[55] T. D. Brennan and S. Hong, "Introduction to Generalized Global Symmetries in QFT and Particle Physics," arXiv:2306.00912 [hep-ph].

[56] S. Schafer-Nameki, "ICTP Lectures on (Non-)Invertible Generalized Symmetries," arXiv:2305.18296 [hep-th].

[57] J. McGreevy, "Generalized Symmetries in Condensed Matter," arXiv:2204.03045 [cond-mat.str-el].

[58] C. Córdova, T. T. Dumitrescu, K. Intriligator, and S.-H. Shao, "Snowmass White Paper: Generalized Symmetries in Quantum Field Theory and Beyond," in *Snowmass 2021*. 2021. arXiv:2205.09545 [hep-th].

[59] E. Witten, "Wilson Loops, 't Hooft Loops, and 't Hooft's Picture of Confinement," in *Quantum fields and strings: A course for mathematicians Vol.2*, ch. 10, pp. 1255–1261. AMS, 1999.

[60] L. Bhardwaj, L. E. Bottini, L. Fraser-Taliente, L. Gladden, D. S. W. Gould, A. Platschorre, and H. Tillim, "Lectures on generalized symmetries," *Phys. Rept.* **1051** (2024) 1–87, arXiv:2307.07547 [hep-th].

[61] S. Gukov and E. Witten, "Gauge Theory, Ramification, And The Geometric Langlands Program," arXiv:hep-th/0612073.

[62] D. Tong, "Gauge theory," *Lecture notes, DAMTP Cambridge* **10** (2018) .

[63] H. Reinhardt, "On 't Hooft's loop operator," *Phys. Lett. B* **557** (2003) 317–323, arXiv:hep-th/0212264.

[64] M. Engelhardt and H. Reinhardt, "Center projection vortices in continuum Yang-Mills theory," *Nucl. Phys. B* **567** (2000) 249, arXiv:hep-th/9907139.

[65] G. 't Hooft, "On the Phase Transition Towards Permanent Quark Confinement," *Nucl. Phys. B* **138** (1978) 1–25.

[66] C. Córdova, K. Ohmori, and T. Rudelius, "Generalized symmetry breaking scales and weak gravity conjectures," *JHEP* **11** (2022) 154, arXiv:2202.05866 [hep-th].

[67] J. F. Plebański, "On the separation of Einsteinian substructures," *J. Math. Phys.* **18** (1977) 2511–2520.

[68] R. Capovilla, T. Jacobson, J. Dell, and L. J. Mason, "Self-dual 2-forms and gravity," *Class. Quant. Grav.* **8** (1991) 41–57.

[69] W. Donnelly and S. B. Giddings, "Observables, gravitational dressing, and obstructions to locality and subsystems," *Phys. Rev. D* **94** no. 10, (2016) 104038, arXiv:1607.01025 [hep-th].

[70] S. B. Giddings and J. Perkins, "Perturbative quantum evolution of the gravitational state and dressing in general backgrounds," arXiv:2209.06836 [hep-th].

[71] C. Rovelli and L. Smolin, "Discreteness of area and volume in quantum gravity," *Nucl. Phys. B* **442** (1995) 593–622, arXiv:gr-qc/9411005. [Erratum: Nucl.Phys.B 456, 753–754 (1995)].

[72] A. Ashtekar and J. Lewandowski, "Quantum theory of geometry. 1: Area operators," *Class. Quant. Grav.* **14** (1997) A55–A82, arXiv:gr-qc/9602046.

[73] J. D. Bekenstein, "Black holes and entropy," *Phys. Rev. D* **7** (1973) 2333–2346.

[74] S. W. Hawking, "Particle Creation by Black Holes," *Commun. Math. Phys.* **43** (1975) 199–220. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[75] Y. Cho, "Magnetic theory of gravitation," in *Primordial Nucleosynthesis and Evolution of Early Universe*, pp. 203–211. Springer, 1991.

[76] C. W. Misner, "The flatter regions of Newman, Unti and Tamburino's generalized Schwarzschild space," *J. Math. Phys.* **4** (1963) 924–938.

[77] C. W. Misner, "Taub-NUT space as a counterexample to almost anything," *University of Maryland, Technical Report* no. 529, (1965) .

[78] W. B. Bonnor, "A new interpretation of the NUT metric in general relativity," *Math. Proc. Cambridge Phil. Soc.* **66** no. 1, (1969) 145–151.

[79] A. Sackfield, "Physical interpretation of NUT metric," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 70, pp. 89–94, Cambridge University Press. 1971.

[80] J. S. Dowker and J. A. Roche, "The Gravitational analogues of magnetic monopoles," *Proc. Phys. Soc.* **92** (1967) 1–8.

[81] A. H. Taub, "Empty space-times admitting a three parameter group of motions," *Annals Math.* **53** (1951) 472–490.

[82] E. Newman, L. Tamburino, and T. Unti, "Empty space generalization of the Schwarzschild metric," *J. Math. Phys.* **4** (1963) 915.

[83] L. Alfonsi, C. D. White, and S. Wikeley, "Topology and Wilson lines: global aspects of the double copy," *JHEP* **07** (2020) 091, arXiv:2004.07181 [hep-th].

[84] M. J. Duff, "Quantum Tree Graphs and the Schwarzschild Solution," *Phys. Rev. D* **7** (1973) 2317–2326.

[85] C. Cheung, J. Parra-Martinez, I. Z. Rothstein, N. Shah, and J. Wilson-Gerow, "Effective Field Theory for Extreme Mass Ratios," arXiv:2308.14832 [hep-th].

[86] G. Modanese, "Wilson loops in four-dimensional quantum gravity," *Phys. Rev. D* **49** (1994) 6534–6542, arXiv:hep-th/9307148.

[87] J. Fredsted, "Comment on Wilson loops in four-dimensional quantum gravity," *Phys. Rev. D* **64** (2001) 088501.

[88] A. Ashtekar, "New Variables for Classical and Quantum Gravity," *Phys. Rev. Lett.* **57** (1986) 2244–2247.

[89] E. Buffenoir, M. Henneaux, K. Noui, and P. Roche, "Hamiltonian analysis of Plebanski theory," *Class. Quant. Grav.* **21** (2004) 5203–5220, arXiv:gr-qc/0404041.

[90] S. Alexandrov and K. Krasnov, "Hamiltonian Analysis of non-chiral Plebanski Theory and its Generalizations," *Class. Quant. Grav.* **26** (2009) 055005, arXiv:0809.4763 [gr-qc].

[91] M. Celada and M. Montesinos, "Lorentz-covariant Hamiltonian analysis of BF gravity with the Immirzi parameter," *Class. Quant. Grav.* **29** (2012) 205010, arXiv:1209.0396 [gr-qc].

[92] M. Celada, D. González, and M. Montesinos, "BF gravity," *Class. Quant. Grav.* **33** no. 21, (2016) 213001, arXiv:1610.02020 [gr-qc].

[93] A. Ashtekar and M. Varadarajan, "Gravitational Dynamics—A Novel Shift in the Hamiltonian Paradigm," *Universe* **7** no. 1, (2021) 13, arXiv:2012.12094 [gr-qc].

[94] R. L. Arnowitt, S. Deser, and C. W. Misner, "Dynamical Structure and Definition of Energy in General Relativity," *Phys. Rev.* **116** (1959) 1322–1330.

[95] H. Urbantke, "On integrability properties of SU(2) Yang–Mills fields. I. Infinitesimal part," *J. Math. Phys.* **25** no. 7, (1984) 2321–2324.

[96] C. W. Misner and J. A. Wheeler, "Classical physics as geometry: Gravitation, electromagnetism, unquantized charge, and mass as properties of curved empty space," *Annals Phys.* **2** (1957) 525–603.

[97] J. Polchinski, "Monopoles, duality, and string theory," *Int. J. Mod. Phys. A* **19S1** (2004) 145–156, arXiv:hep-th/0304042.

[98] T. Banks and N. Seiberg, "Symmetries and Strings in Field Theory and Gravity," *Phys. Rev. D* **83** (2011) 084019, arXiv:1011.5120 [hep-th].

[99] D. Harlow and H. Ooguri, "Symmetries in quantum field theory and quantum gravity," *Commun. Math. Phys.* **383** no. 3, (2021) 1669–1804, arXiv:1810.05338 [hep-th].

[100] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, "The String Landscape, Black Holes and Gravity as the Weakest Force," *JHEP* **06** (2007) 060, arXiv:hep-th/0601001.

[101] D. Harlow, B. Heidenreich, M. Reece, and T. Rudelius, "Weak gravity conjecture," *Rev. Mod. Phys.* **95** no. 3, (2023) 035003, arXiv:2201.08380 [hep-th].

[102] B. Heidenreich, J. McNamara, M. Montero, M. Reece, T. Rudelius, and I. Valenzuela, "Non-invertible global symmetries and completeness of the spectrum," *JHEP* **09** (2021) 203, arXiv:2104.07036 [hep-th].

[103] T. Jacobson and J. D. Romano, "Degenerate extensions of general relativity," *Class. Quant. Grav.* **9** (1992) L119–L124, arXiv:gr-qc/9207005.

[104] J. D. Romano, "On the constraint algebra of degenerate relativity," *Phys. Rev. D* **48** (1993) 5676–5683, arXiv:gr-qc/9306034.

[105] I. Bengtsson and T. Jacobson, "Degenerate metric phase boundaries," *Class. Quant. Grav.* **14** no. 11, (1997) 3109–3121, arXiv:gr-qc/9706027. [Erratum: Class.Quant.Grav. 15, 3941–3942 (1998)].

[106] G. E. Volovik, "Vierbein walls in condensed matter," *JETP Lett.* **70** (1999) 711–716, arXiv:gr-qc/9911026.

[107] G. Modanese, "Geodesic round trips by parallel transport in quantum gravity," *Phys. Rev. D* **47** (1993) 502–509.

[108] R. Alawadhi, D. S. Berman, C. D. White, and S. Wikeley, "The single copy of the gravitational holonomy," *JHEP* **10** (2021) 229, arXiv:2107.01114 [hep-th].

[109] T. Jacobson and J. D. Romano, "The Spin holonomy group in general relativity," *Commun. Math. Phys.* **155** (1993) 261–276, arXiv:gr-qc/9207006.

[110] J. F. Donoghue, "Is the spin connection confined or condensed?," *Phys. Rev. D* **96** no. 4, (2017) 044003, arXiv:1609.03523 [hep-th].

[111] A. Brandhuber, P. Heslop, A. Nasti, B. Spence, and G. Travaglini, "Four-point Amplitudes in N=8 Supergravity and Wilson Loops," *Nucl. Phys. B* **807** (2009) 290–314, arXiv:0805.2763 [hep-th].

[112] A. Kapustin and R. Thorngren, "Higher Symmetry and Gapped Phases of Gauge Theories," *Prog. Math.* **324** (2017) 177–202, arXiv:1309.4721 [hep-th].

[113] F. Benini, C. Córdova, and P.-S. Hsin, "On 2-Group Global Symmetries and their Anomalies," *JHEP* **03** (2019) 118, arXiv:1803.09336 [hep-th].

[114] C. Córdova, T. T. Dumitrescu, and K. Intriligator, "Exploring 2-Group Global Symmetries," *JHEP* **02** (2019) 184, arXiv:1802.04790 [hep-th].

[115] C. Córdova, T. T. Dumitrescu, and K. Intriligator, "2-Group Global Symmetries and Anomalies in Six-Dimensional Quantum Field Theories," *JHEP* **04** (2021) 252, arXiv:2009.00138 [hep-th].

[116] K. Krasnov, *Formulations of general relativity: Gravity, spinors and differential forms*. Cambridge University Press, 2020.

[117] R. Capovilla, T. Jacobson, and J. Dell, "A pure spin connection formulation of gravity," *Class. Quant. Grav.* **8** (1991) 59–73.

[118] K. Krasnov and E. Mitsou, "Pure Lorentz spin connection theories and uniqueness of general relativity," *Class. Quant. Grav.* **38** no. 20, (2021) 205009, arXiv:2106.05803 [gr-qc].

[119] R. Penrose, "Gravitational collapse and space-time singularities," *Physical Review Letters* **14** no. 3, (1965) 57.

[120] P. M. Chaikin, T. C. Lubensky, and T. A. Witten, *Principles of condensed matter physics*, vol. 10. Cambridge university press Cambridge, 1995.

[121] S. Grozdanov and N. Poovuttikul, "Generalized global symmetries in states with dynamical defects: The case of the transverse sound in field theory and holography," *Phys. Rev. D* **97** no. 10, (2018) 106005, arXiv:1801.03199 [hep-th].

[122] S. D. Pace, "Emergent generalized symmetries in ordered phases," arXiv:2308.05730 [cond-mat.str-el].

[123] P. Fiziev and H. Kleinert, "New action principle for classical particle trajectories in spaces with torsion," *EPL* **35** (1996) 241–246, arXiv:hep-th/9503074.

[124] H. Kleinert and A. Pelster, "Lagrange mechanics in spaces with curvature and torsion," *Gen. Rel. Grav.* **31** (1999) 1439, arXiv:gr-qc/9605028.

[125] H. Kleinert, *Multivalued fields. In condensed matter, electromagnetism, and gravitation*. World Scientific, 2008.

An analytical approximation of the evolution of the primordial curvature perturbation in the ultraslow-roll inflation

Ji-Xiang Zhao[a] **Nan Li**[a]

[a] *Department of Physics, College of Sciences, Northeastern University No. 3-11, Wenhua Road, Shenyang, 110819, China*

E-mail: 2100189@stu.neu.edu.cn, linan@mail.neu.edu.cn

Abstract: Cosmic inflation can enter an ultraslow-roll (USR) stage, if there is a plateau on the inflaton potential. During this stage, the primordial curvature perturbation $\mathcal{R}_k$ and its power spectrum $\mathcal{P}_\mathcal{R}$ can be remarkably enhanced on small scales. In this work, an analytical approximation is provided to systematically study the evolution of $\mathcal{R}_k$ in the USR inflation. We first discuss the asymptotic solutions of the moduli and arguments of $\mathcal{R}_k$ and its time derivative $\mathcal{R}_{k,N}$ on the sub- and super-horizon scales separately and find that all these solutions have simple exponential forms. Then, $\mathcal{R}_k$ on five typical scales are investigated in order. Our analytical approximation predicts that $\mathcal{R}_k$ first revolves around the origin in the complex plane, but if it crosses the horizon around the start of the USR stage, there appears a subsequent linear evolution towards or away from the origin. This behavior naturally explains the shape of $\mathcal{P}_\mathcal{R}$ from the sharp dip to the peak and matches the numerical results perfectly. Moreover, the minimum of $\mathcal{P}_\mathcal{R}$ is exactly proved to be nonvanishing. Our analytical approximation will help the model building in primordial black hole and gravitational wave physics.

arXiv:2403.01416v1 [astro-ph.CO] 3 Mar 2024

Contents

1 Introduction **1**

2 Basic equations **3**
- 2.1 USR inflation 3
- 2.2 $\mathcal{R}_k$ and $\mathcal{P}_\mathcal{R}$ 4
- 2.3 θ_k and φ_k 6

3 Analytical approximation **7**
- 3.1 $k \gg He^N$ 8
- 3.2 $k \ll He^N$ 10

4 Evolution of the primordial curvature perturbation **11**

5 Conclusion **17**

A Derivation of Eq. (2.14) **19**

1 Introduction

The milestone in multi-messenger astronomy is marked by the detection of the gravitational waves from the merger of binary black holes [1]. Some typical characteristics of these black holes, such as their unexpectedly large masses and relatively small spins, are not consistent with usual astrophysical black holes, but are more possible to be of primordial origin [2–4]. Primordial black holes (PBHs) are theorized to form before the Big Bang nucleosynthesis, so they are a very powerful probe in the early Universe [5]. For instance, they could explain the strong absorption trough found in the 21-cm global spectrum [6, 7], and serve as the seeds of the supermassive black holes in galactic centers [8], supported by the recent observation from the James Webb Space Telescope on the massive galaxies at high redshifts [9]. Meanwhile, the first-order scalar perturbations that generate PBHs can also produce the second-order scalar-induced gravitational waves (SIGWs) [10–16], such as the possible nHz gravitational wave background discovered recently [17–27]. More important, PBHs are a natural and promising candidate of dark matter (DM) [28].

Unlike astrophysical black holes, the masses of PBHs can range from the Planck mass to supermassive scale (10^{-38}–10^{10} $M_\odot$, with $M_\odot = 1.989 \times 10^{30}$ kg being the solar mass) [29]. This versatility enables PBHs to offer valuable insights into different cosmic conundrums [30]. The PBHs with mass $M < 5 \times 10^{-19}$ $M_\odot$ have already evaporated because of the Hawking radiation, changing the background intensities of various cosmic rays [31]. However, those with mass $M > 5 \times 10^{-19}$ $M_\odot$ can still exist today, acting as a stable and

pressureless candidate of DM. The PBH abundance $f_{\rm PBH}$ is defined as its proportion in DM today. If $f_{\rm PBH} \gtrsim 0.1$, the PBHs can be considered as an effective candidate of DM; if $f_{\rm PBH} \ll 10^{-3}$, its possibility as DM can be safely excluded in the relevant mass range. Currently, according to various constraints on the upper bounds of $f_{\rm PBH}$ in different mass ranges, there still remains an open mass window from $10^{-17}\ M_\odot$ (the asteroid mass range) to $10^{-13}\ M_\odot$ (the sub-lunar mass range) that is possible to compose all DM with PBHs [28, 29, 31].

The PBH abundance $f_{\rm PBH}$ can be calculated from the power spectrum $\mathcal{P}_\mathcal{R}(k)$ of the primordial curvature perturbation $\mathcal{R}$ [32–34]. On large scales (e.g., the pivot scale $k_* = 0.05\ {\rm Mpc}^{-1}$ in the Planck satellite experiment), $\mathcal{P}_\mathcal{R}$ is precisely measured by the anisotropies of the cosmic microwave background (CMB), with an amplitude of 2.10×10^{-9} [35]. However, to produce abundant PBHs, $\mathcal{R}$ must be large enough, so that $\mathcal{P}_\mathcal{R}$ is significantly enhanced up to $\mathcal{O}(10^{-2})$ on small scales. In the usual single-field slow-roll (SR) inflation models, such a huge enhancement of $\mathcal{P}_\mathcal{R}$ is almost impossible. Therefore, the SR conditions must be violated on small scales, and various ultraslow-roll (USR) inflation models are thus proposed [36–74]. In these models, there is always a plateau or a (near-) inflection point on the inflaton potential, where the inflaton rolls down extremely slowly, amplifying $\mathcal{P}_\mathcal{R}$ accordingly.

Albeit the amplitude and shape of $\mathcal{P}_\mathcal{R}$ can be straightforwardly obtained in various USR inflation models by numerical methods, the physical comprehension behind seems still indistinct. Previously, in Refs. [50, 75, 76], the spectral index of the steepest growth of $\mathcal{P}_\mathcal{R}$ was analytically investigated, and in Ref. [77], the authors raised an approximation method to calculate the Fourier mode $\mathcal{R}_k$ of the primordial curvature perturbation. However, to our knowledge, these analyses are still not sufficient in some aspects. For example, there is always a sharp dip in $\mathcal{P}_\mathcal{R}$ before its steep growth. In some early literature [78–86], the location of the dip was calculated numerically, but the underlying physical explanation is lacking yet. Although the detail of this dip does not affect the PBH abundance or SIGW spectrum, as they are relevant mainly to the peak of $\mathcal{P}_\mathcal{R}$, we still hope to answer whether the minimum of $\mathcal{P}_\mathcal{R}$ is zero analytically.

Consequently, this paper is dedicated to study the evolution of $\mathcal{R}_k$ and the resulting $\mathcal{P}_\mathcal{R}$ from a more theoretical perspective, and is a natural extension of our previous numerical results in Ref. [86]. We will provide an analytical approximation to explore the USR inflation, which will save us away from the mathematical tediousness and extract the physical essence to the most extent. Following our early works in Refs. [57, 58, 72, 86], by considering an antisymmetric perturbation on the background inflaton potential, inflation can be led into the USR stage smoothly. Our main improvements in this work are threefold. First, by simplifying the two parameters ε and η in the SR and USR stages separately, we are able to obtain the analytical solutions of $\mathcal{R}_k$ and its time derivative $\mathcal{R}_{k,N}$. From their asymptotic forms, we achieve $|\mathcal{R}_k|$, $\theta_{k,N}$, and $\varphi_{k,N}$ (with θ_k and φ_k being the arguments of $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$) and find that all these quantities possess concise exponential forms, perfectly in agreement with the numerical results in Ref. [86]. Second, we predict that, besides the revolving evolution of $\mathcal{R}_k$ around the origin in the complex plane when $k \gg He^N$, there appears an interesting linear evolution of $\mathcal{R}_k$ towards or away from the origin if $\mathcal{R}_k$ crosses

the horizon around the start of the USR stage, naturally explaining the sharp dip and the peak in $\mathcal{P}_\mathcal{R}$. Third, we analytically study the minimum of the dip in $\mathcal{P}_\mathcal{R}$ and prove that it cannot reach zero exactly, which is also seldom mentioned before. Altogether, we wish to give a whole picture and thorough understanding of the physical essence in the complicated evolution of $\mathcal{R}_k$ and the relevant $\mathcal{P}_\mathcal{R}$, which will be beneficial to the model building of the USR inflation and help our research on PBH and gravitational wave physics in future.

This paper is organized as follows. The basic equations of motion for $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$ are presented in Sec. 2. Then, we show our analytical approximation in Sec. 3 and discuss the asymptotic solutions of $|\mathcal{R}_k|$, $|\mathcal{R}_k|_{,N}$, $\theta_{k,N}$, and $\varphi_{k,N}$ in the sub- and super-horizon limits, respectively. In Sec. 4, the evolution of $\mathcal{R}_k$ and the resulting $\mathcal{P}_\mathcal{R}$ for five typical scales with different times of horizon-exit are systematically investigated. We conclude in Sec. 5. We work in the natural system of units and set $c = \hbar = k_\mathrm{B} = 1$.

2 Basic equations

In this section, we show the equation of motion for the primordial curvature perturbation $\mathcal{R}_k$ and its relation to the power spectrum $\mathcal{P}_\mathcal{R}(k)$. Two important variables θ_k and φ_k are introduced, with their evolutions calculated in detail.

2.1 USR inflation

We start from the single-field inflation model, with the corresponding action being

$$S = \int \mathrm{d}^4x\,\sqrt{-g}\left[\frac{m_\mathrm{P}^2}{2}R - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right],$$

where $m_\mathrm{P} = 1/\sqrt{8\pi G}$ is the reduced Planck mass, R is the Ricci scalar, ϕ is the inflaton field, and $V(\phi)$ is its potential. The inflaton potential V can be further decomposed into its background $V_\mathrm{b}(\phi)$ and a perturbation $\delta V(\phi)$ on it. In this work, we choose V_b as the Kachru–Kallosh–Linde–Trivedi potential [87],

$$V_\mathrm{b}(\phi) = V_0\frac{\phi^2}{\phi^2 + (m_\mathrm{P}/2)^2}.$$

Meanwhile, we consider a perturbation δV on V_b,

$$\delta V(\phi) = -A\frac{\phi - \phi_0}{1 + (\phi - \phi_0)^2/(2\sigma^2)},$$

where three model parameters A, ϕ_0, and σ characterize the amplitude, position, and width of δV, respectively. (In principle, the specific form of δV is not unique, and other kinds of δV can be found in Refs. [57, 58, 72].) Hence, δV is antisymmetric around ϕ_0, so it can be smoothly connected on V_b on both sides of ϕ_0. In addition, we demand $A = V_{\mathrm{b},\phi}(\phi_0)$ for simplicity, in order to create a perfect plateau around ϕ_0, leading inflation into the USR stage. Following Ref. [58], we choose the model parameters as $V_0/m_\mathrm{P}^4 = 10^{-10}$, $\phi_0/m_\mathrm{P} = 1.31$, and $\sigma/m_\mathrm{P} = 0.0831881$, such that there can be PBHs with mass $M \approx 10^{-17}\ M_\odot$ and abundance $f_\mathrm{PBH} \approx 0.1$. Moreover, the initial conditions for inflation are set to be

$\phi/m_{\rm P} = 3.30$ and $\phi_{,N}/m_{\rm P} = -0.0137$, so as to satisfy the Planck constraints on the power spectrum and the tensor-to-scalar ratio on the CMB pivot scale k_* [35].

In cosmic inflation, it is more convenient to utilize the number of e-folds N as the time variable, defined as $\mathrm{d}N = H\,\mathrm{d}t = \mathrm{d}\ln a$, where t is the cosmic time, $a(t) = e^N$ is the scale factor, and $H(t) = \dot{a}/a$ is the Hubble expansion rate. Furthermore, to characterize the motion of the inflaton, two useful parameters are introduced as

$$\varepsilon = -\frac{\dot{H}}{H^2} = \frac{\phi_{,N}^2}{2m_{\rm P}^2}, \tag{2.1}$$

$$\eta = -\frac{\ddot{\phi}}{H\dot{\phi}} = \frac{\phi_{,N}^2}{2m_{\rm P}^2} - \frac{\phi_{,NN}}{\phi_{,N}}. \tag{2.2}$$

In terms of these parameters, the Friedmann equation for cosmic expansion becomes

$$H^2 = \frac{V}{(3-\varepsilon)m_{\rm P}^2}, \tag{2.3}$$

and the Klein–Gordon equation for the inflaton field can be reexpressed as

$$\phi_{,NN} + (3-\varepsilon)\phi_{,N} + \frac{1}{H^2}V_{,\phi} = 0. \tag{2.4}$$

In the usual SR inflation, both ε and $|\eta|$ are always much smaller than 1 and are thus named as the SR parameters. However, in the USR stage, ε drops almost exponentially and $|\eta|$ may even approach $\mathcal{O}(1)$, so the SR conditions are broken. Thus, the starting and ending points of the USR stage can be determined by $\eta(N_{\rm s}) = \eta(N_{\rm e}) = 0$. With the model parameters listed above, we have $N_{\rm s} = 56.81$ and $N_{\rm e} = 60.93$. The evolutions of ε and $|\eta|$ in both SR and USR stages are shown in Fig. 1. Moreover, the number of e-folds when the relevant scale crosses the horizon is denoted as $N_{\rm out}$ (i.e., when $k = He^{N_{\rm out}}$), which is an important quantity for the subsequent discussions in Sec. 4.

2.2 $\mathcal{R}_k$ and $\mathcal{P}_\mathcal{R}$

Now, we move on to the perturbations on the background spacetime. In Newtonian gauge, the perturbed metric reads

$$\mathrm{d}s^2 = -(1+2\Psi)\,\mathrm{d}t^2 + a^2(1-2\Psi)\delta_{ij}\,\mathrm{d}x^i\mathrm{d}x^j,$$

where Ψ is the scalar perturbation. (Here, we have neglected the vector perturbation, tensor perturbation, and anisotropic stress, as they are irrelevant to PBH formation.) A more convenient and gauge-invariant scalar perturbation is the primordial curvature perturbation $\mathcal{R}$ defined as

$$\mathcal{R} = \Psi + \frac{H}{\dot{\phi}}\,\delta\phi = \Psi + \frac{\delta\phi}{\phi_{,N}},$$

and the equation of motion for its Fourier mode $\mathcal{R}_k$ is the Mukhanov–Sasaki (MS) equation [88, 89],

$$\mathcal{R}_{k,NN} + (3+\varepsilon-2\eta)\mathcal{R}_{k,N} + \left(\frac{k}{He^N}\right)^2\mathcal{R}_k = 0. \tag{2.5}$$

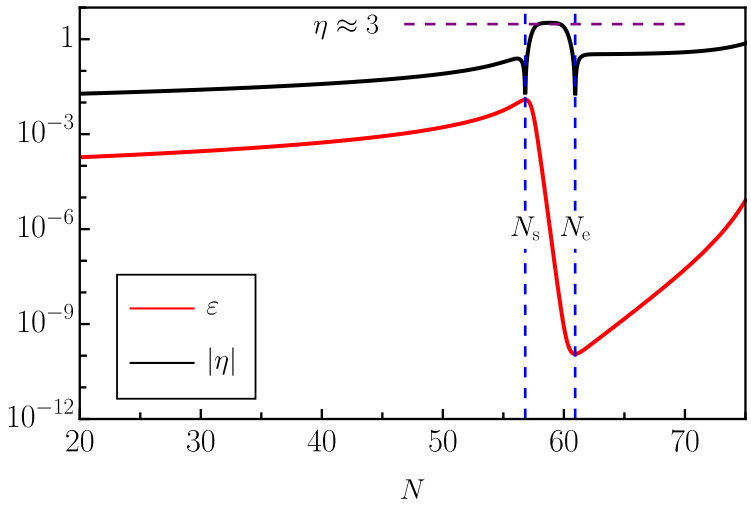


Figure 1. The evolutions of the parameters ε and $|\eta|$ as a function of the number of e-folds N. The starting and ending points of the USR stage are determined by $\eta(N_{\rm s}) = \eta(N_{\rm e}) = 0$, with $N_{\rm s} = 56.81$ and $N_{\rm e} = 60.93$. During the USR stage, ε decreases nearly exponentially, but $|\eta|$ increases significantly and maintains a value around 3 for a sufficiently long period.

The initial conditions for the MS equation can be fixed by the Bunch–Davies vacuum [90]. In the very early Universe, both ϵ and $|\eta|$ are very small, and the asymptotic solution of $\mathcal{R}_k$ is

$$\mathcal{R}_k = \frac{e^{ik/(He^N)}}{z\sqrt{2k}}\left(1+\frac{iHe^N}{k}\right), \tag{2.6}$$

where $z = \phi_{,N}e^N$ (i.e., $z^2 = 2m_{\rm P}^2a^2\varepsilon$). At the initial time, we have $k \gg He^N$, so from Eq. (2.6), we obtain

$$\mathcal{R}_k\Big|_{N\to N_{\rm ini}} = \frac{e^{ik/(He^N)}}{z\sqrt{2k}}, \tag{2.7}$$

$$\mathcal{R}_{k,N}\Big|_{N\to N_{\rm ini}} = -\frac{e^{ik/(He^N)}}{z\sqrt{2k}}\left(\frac{z_{,N}}{z}+\frac{ik}{He^N}\right), \tag{2.8}$$

where $N_{\rm ini}$ is the initial number of e-folds.

Furthermore, the PBH abundance and the SIGW spectrum can be calculated from the power spectrum of $\mathcal{R}$, which corresponds to the two-point correlation function of $\mathcal{R}_k$ in the Fourier space. The statistical information and observational significance of $\mathcal{R}_k$ is encoded in the dimensionless power spectrum $\mathcal{P}_\mathcal{R}(k)$ as

$$\mathcal{P}_\mathcal{R}(k) = \frac{k^3}{2\pi^2}|\mathcal{R}_k|^2\Big|_{k\ll aH}.$$

We should point out that, in the usual SR inflation, $\mathcal{R}_k$ becomes almost frozen when the relevant scale crosses the horizon, and $\mathcal{P}_\mathcal{R}$ can be effectively evaluated at the epoch of

horizon-exit. Nonetheless, in the USR inflation, $\mathcal{R}_k$ can still evolve significantly after the horizon-exit, so $\mathcal{P}_\mathcal{R}$ must be determined at the end of inflation.

Generally speaking, both $\mathcal{R}_k$ and $\mathcal{P}_\mathcal{R}$ can be obtained by numerically solving Eqs. (2.1)–(2.5), but this is always a time-consuming task. Consequently, one of the aims of this work is to explore an appropriate and analytical approximation to calculate $\mathcal{R}_k$ and $\mathcal{P}_\mathcal{R}$. Only in this way can we extract the physical essence in the complicated evolution of $\mathcal{R}_k$ and the characteristic shape of $\mathcal{P}_\mathcal{R}$ in the USR inflation, without getting lost in the tedious calculation program.

2.3 θ_k and φ_k

As the MS equation is complex, it is more convenient to decompose $\mathcal{R}_k$ as

$$\mathcal{R}_k = |\mathcal{R}_k| e^{-i\theta_k}, \tag{2.9}$$

where $|\mathcal{R}_k|$ and θ_k are the modulus and argument of $\mathcal{R}_k$. Another advantage of this decomposition, rather than decomposing $\mathcal{R}_k$ into the real and imaginary parts, is that we will observe interesting revolving and linear evolutions of $\mathcal{R}_k$ in the complex plane, to be shown in Sec. 4.

Substituting Eq. (2.9) into Eq. (2.5), we arrive at the equations of motion for $|\mathcal{R}_k|$ and θ_k,

$$|\mathcal{R}_k|_{,NN} + (3+\varepsilon-2\eta)|\mathcal{R}_k|_{,N} + \left(\frac{k}{He^N}\right)^2 \left[1 - \left(\frac{He^N\theta_{k,N}}{k}\right)^2\right] |\mathcal{R}_k| = 0, \tag{2.10}$$

$$\theta_{k,NN} + \left(3+\varepsilon-2\eta+2\frac{|\mathcal{R}_k|_{,N}}{|\mathcal{R}_k|}\right)\theta_{k,N} = 0. \tag{2.11}$$

Furthermore, from Eq. (2.7), the initial conditions for $|\mathcal{R}_k|$ read

$$|\mathcal{R}_k|\Big|_{N\to N_{\rm ini}} = -\frac{1}{z\sqrt{2k}}, \tag{2.12}$$

$$|\mathcal{R}_k|_{,N}\Big|_{N\to N_{\rm ini}} = \frac{z_{,N}}{z^2\sqrt{2k}}. \tag{2.13}$$

When the inflaton ϕ rolls down from its potential V, we have $z = \phi_{,N}e^N < 0$. Since $|\mathcal{R}_k|$ is positive, we have assigned a negative sign in Eq. (2.12).

Below, for our analytical purpose, we provide two mathematical tricks to simplify the calculations of $|\mathcal{R}_k|$ and θ_k. First, from Eqs. (2.6) and (2.9), we can obtain the evolution of the argument θ_k as [56]

$$\theta_{k,N} = \frac{1}{2z^2He^N|\mathcal{R}_k|^2}, \tag{2.14}$$

and the detailed derivation can be found in A. Hence, $\theta_{k,N}$ is positive-definite, so θ_k is a monotonic function.

Second, from Eq. (2.9), we have $\mathcal{R}_{k,N} = |\mathcal{R}_k|_{,N} e^{-i\theta_k} - i|\mathcal{R}_k|\theta_{k,N} e^{-i\theta_k}$, so the evolution of the modulus $|\mathcal{R}_k|$ is

$$|\mathcal{R}_k|_{,N} = \sqrt{|\mathcal{R}_{k,N}|^2 - (|\mathcal{R}_k|\theta_{k,N})^2}. \tag{2.15}$$

By this means, we are able to achieve $|\mathcal{R}_k|$ and θ_k more quickly, as both Eqs. (2.14) and (2.15) are first-order differential equations, in which $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$ can be easily obtained from the MS equation via the analytical approximation in Sec. 3. In contrast, it is impossible to solve Eq. (2.10) analytically, due to the complicity in the prefactor of $|\mathcal{R}_k|$.

Similarly, we decompose the derivative of $\mathcal{R}_k$ as

$$\mathcal{R}_{k,N} = |\mathcal{R}_{k,N}| e^{-i\varphi_k},$$

where the argument φ_k can be obtained from Eq. (2.9) as

$$\varphi_k = \theta_k + \arctan\left(\frac{|\mathcal{R}_k|}{|\mathcal{R}_k|_{,N}}\theta_{k,N}\right). \tag{2.16}$$

Substituting Eq. (2.14) into Eq. (2.16), we arrive at the evolution of φ_k as

$$\varphi_{k,N} = \frac{1}{2z^2 He^N |\mathcal{R}_k|^2} - \frac{(2z^2 He^N |\mathcal{R}_k||\mathcal{R}_k|_{,N})_{,N}}{(2z^2 He^N |\mathcal{R}_k||\mathcal{R}_k|_{,N})^2 + 1}. \tag{2.17}$$

Different from $\theta_{k,N}$ in Eq. (2.14), $\varphi_{k,N}$ depends not only on $|\mathcal{R}_k|$, but also on $|\mathcal{R}_k|_{,N}$.

In summary, Eqs. (2.9), (2.14), (2.15), and (2.17) summarize the evolutions of the moduli and arguments of $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$, which are the starting point of our following discussions.

3 Analytical approximation

In this section, we propose our analytical approximation to calculate $|\mathcal{R}_k|$ and $|\mathcal{R}_k|_{,N}$ in two different cases: $k \gg He^N$ and $k \ll He^N$, in which the scale of $\mathcal{R}_k$ is sub- and super-horizon, respectively. Next, we bring the results into Eqs. (2.14) and (2.17) to obtain $\theta_{k,N}$ and $\varphi_{k,N}$. Our main results are summarized in Tabs. 1 and 2.

We start our analytical approximation from the two parameters ε and η. First, from Fig. 1, in the SR stage when $N < N_\mathrm{s}$, ε is nearly invariant and η is very small, so we approximately have

$$\varepsilon \approx \varepsilon(N_\mathrm{s}) = \varepsilon_\mathrm{s}, \quad \eta \approx 0.$$

Second, in the USR stage when $N_\mathrm{s} < N < N_\mathrm{e}$, as the inflaton potential is extremely flat, from Eqs. (2.2) and (2.4), we arrive at $\eta \approx 3$. Then, from Eqs. (2.1) and (2.2), η can be reexpressed as $\eta = \varepsilon - \varepsilon_{,N}/(2\varepsilon)$, so we approximately obtain

$$\varepsilon \approx \varepsilon_\mathrm{s} e^{-6(N-N_\mathrm{s})}, \quad \eta \approx 3.$$

Third, when $N > N_{\rm e}$, the inflation dynamics significantly depends on the specific inflaton potential, so it is hard to have a general approximation method any more, and we will not consider it in this work.

Before exploring Eq. (2.5), we still need to make two assumptions. First, η and H are assumed to be constant. Second, ε is neglected when it is added to η. These assumptions are valid in both SR and USR stages, as can be evidently seen from Fig. 1. With these preparations, Eq. (2.5) can be exactly solved as a function of N as

$$\mathcal{R}_k = e^{-\left(\frac{3}{2}-\eta\right)N}\left[AJ_{\frac{3}{2}-\eta}\left(\frac{k}{He^N}\right) + BJ_{-\frac{3}{2}+\eta}\left(\frac{k}{He^N}\right)\right], \tag{3.1}$$

$$\mathcal{R}_{k,N} = e^{-\left(\frac{5}{2}-\eta\right)N}\frac{k}{H}\left[-AJ_{\frac{1}{2}-\eta}\left(\frac{k}{He^N}\right) + BJ_{-\frac{1}{2}+\eta}\left(\frac{k}{He^N}\right)\right], \tag{3.2}$$

where $J_\alpha(x)$ is the Bessel function of the first kind, and the coefficients A and B are the functions of k, H, and η, but are independent of N. Also, we should mention that both A and B are complex numbers, since $\mathcal{R}_k$ itself is complex.

3.1 $k \gg He^N$

We first consider the sub-horizon case with $k \gg He^N$. Under this circumstance, our analytical approximation is relatively simple, because $J_\alpha(x) \sim 1/\sqrt{x}$ when $x = k/(He^N) \gg 1$, irrespective of the value of α. As a result, from Eqs. (3.1) and (3.2), we easily obtain

$$|\mathcal{R}_k| \approx c_1 e^{-(1-\eta)N} \sim \begin{cases} e^{-N} & (\text{when } \eta \approx 0), \\ e^{2N} & (\text{when } \eta \approx 3), \end{cases} \tag{3.3}$$

$$|\mathcal{R}_{k,N}| \approx c_2 e^{-(2-\eta)N} \sim \begin{cases} e^{-2N} & (\text{when } \eta \approx 0), \\ e^{N} & (\text{when } \eta \approx 3), \end{cases} \tag{3.4}$$

where the coefficients c_1 and c_2 can be obtained from A and B, but the explicit expressions are unnecessary here. We observe that the time-dependence of $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$ is rather simple. More interpretation of these results will be presented in Sec. 4.

Now, we discuss the SR and USR stages separately. In the SR case, $\eta \approx 0$, so from Eqs. (3.3) and (3.4), we have

$$|\mathcal{R}_k| \sim e^{-N}, \quad |\mathcal{R}_{k,N}| \sim e^{-2N}. \tag{3.5}$$

Substituting $|\mathcal{R}_k|$ into Eq. (2.14) and considering $\varepsilon \approx \varepsilon_{\rm s}$, we obtain the evolution of θ_k as

$$\theta_{k,N} \sim \frac{1}{e^{2N}\varepsilon_{\rm s} H e^N e^{-2N}} \sim e^{-N}. \tag{3.6}$$

Then, substituting Eqs. (3.5) and (3.6) into Eq. (2.15), we have

$$|\mathcal{R}_k|_{,N} \sim e^{-2N}.$$

For the evolution of φ_k, substituting Eqs. (3.5) and (3.6) into Eq. (2.17), we obtain

$$\varphi_{k,N} \sim \frac{1}{e^{2N}\varepsilon_{\rm s} H e^N e^{-2N}} \sim e^{-N}.$$

Similarly, in the USR case, $\eta \approx 3$, so

$$|\mathcal{R}_k| \sim e^{2N}, \quad |\mathcal{R}_{k,N}| \sim e^{N}.$$

Taking into account $\varepsilon \approx \varepsilon_{\rm s} e^{-6(N-N_{\rm s})}$, we have

$$\theta_{k,N} \sim \frac{1}{e^{2N}\varepsilon_{\rm s} e^{-6(N-N_{\rm s})} H e^{N} e^{4N}} \sim e^{-N}. \tag{3.7}$$

Thus, $\theta_{k,N}$ has the same time-dependence as that in Eq. (3.6). However, this happens only for the sub-horizon case. In the super-horizon case, $\theta_{k,N}$ may have totally different results in the SR and USR cases, to be shown in Sec. 3.2. Furthermore, by the same means, we obtain

$$|\mathcal{R}_k|_{,N} \sim e^{N},$$

and

$$\varphi_{k,N} \sim \frac{1}{e^{2N}\varepsilon_{\rm s} e^{-6(N-N_{\rm s})} H e^{N} e^{4N}} \sim e^{-N}.$$

It is interesting to find that $\varphi_{k,N}$ behaves the same in both SR and USR stages, but this is understandable. In the numerator of the second term in Eq. (2.17), $2z^2 H e^{N} |\mathcal{R}_k||\mathcal{R}_k|_{,N}$ is actually constant no matter $\eta \approx 0$ or 3. Thus, its derivative with respect to N vanishes, so $\varphi_{k,N} \approx \theta_{k,N} \sim e^{-N}$.

All the above results are summarized in Tab. 1, indicating that, when a scale is sub-horizon, $|\mathcal{R}_k|$ is a monotonic function of N. Therefore, $\mathcal{R}_k$ either revolves towards the origin in the complex plane at a decelerated rate in the SR stage, or away from the origin at an accelerated rate in the USR stage. These conclusions will be compared with the numerical results in Sec. 4.[1]

η	$\lvert\mathcal{R}_k\rvert$	$\theta_{k,N}$	$\lvert\mathcal{R}_k\rvert_{,N}$	$\varphi_{k,N}$
0	e^{-N}	e^{-N}	e^{-2N}	e^{-N}
3	e^{2N}	e^{-N}	e^{N}	e^{-N}

Table 1. All the possible results of $|\mathcal{R}_k|$, $\theta_{k,N}$, $|\mathcal{R}_k|_{,N}$, and $\varphi_{k,N}$ in the sub-horizon case with $k \gg He^{N}$ via our analytical approximation. The parameter $\eta \approx 0$ or 3 corresponds to the SR or USR stage, respectively. From the behaviors of $|\mathcal{R}_k|$ and $\theta_{k,N}$, $\mathcal{R}_k$ either revolves towards the origin at a decelerated rate in the SR stage, or away from the origin at an accelerated rate in the USR stage.

[1] Here, two subtle points should be mentioned. First, the results of $|\mathcal{R}_k|_{,N}$ seem inconsistent with those of $|\mathcal{R}_k|$, but we must point out that $|\mathcal{R}_k|_{,N}$ cannot be trivially obtained from the derivative of $|\mathcal{R}_k|$, but must be calculated from Eq. (2.15). Second, one cannot naively expect $|\mathcal{R}_k|_{,N} \sim |\mathcal{R}_{k,N}|$ (although their dependence on N does coincide), but must use Eq. (2.15) in principle. These two points also apply in Sec. 3.2.

3.2 $k \ll He^N$

Next, we turn to the super-horizon case with $k \ll He^N$. The current situation is a little bit complicated than the sub-horizon case, and the essential reason is that, when $x = k/(He^N) \ll 1$, the asymptotic form of the Bessel function $J_\alpha(x) \sim x^\alpha$ depends on α. Consequently, from Eqs. (3.1) and (3.2), we obtain

$$|\mathcal{R}_k| \approx \left|c_3 e^{-(3-2\eta)N} + c_4\right| = \begin{cases} \left|c_3 e^{-3N} + c_4\right| & (\text{when } \eta \approx 0), \\ \left|c_3 e^{3N} + c_4\right| & (\text{when } \eta \approx 3), \end{cases} \tag{3.8}$$

$$|\mathcal{R}_{k,N}| \approx \left|c_5 e^{-(3-2\eta)N} + c_6 e^{-2N}\right| = \begin{cases} \left|c_5 e^{-3N} + c_6 e^{-2N}\right| & (\text{when } \eta \approx 0), \\ \left|c_5 e^{3N} + c_6 e^{-2N}\right| & (\text{when } \eta \approx 3), \end{cases} \tag{3.9}$$

where the coefficients c_3 to c_6 can also be obtained from A and B. Now, the situation becomes more intractable, as there are two competitive terms in both $|\mathcal{R}_k|$ and $|\mathcal{R}_k|_{,N}$. Hence, we must be very careful to determine the leading-order one, and we discuss this issue in detail below.

We start from the SR stage with $\varepsilon \approx \varepsilon_\mathrm{s}$ and $\eta \approx 0$. If the leading-order terms are

$$|\mathcal{R}_k| \approx c_3 e^{-3N}, \quad |\mathcal{R}_{k,N}| \approx c_5 e^{-3N},$$

from Eqs. (2.14), (2.15), and (2.17), by the same method in Sec. 3.1, we obtain

$$\theta_{k,N} \sim e^{3N}, \quad |\mathcal{R}_k|_{,N} \sim e^{-3N}, \quad \varphi_{k,N} \sim e^{3N}.$$

Similarly, if the leading-order terms are

$$|\mathcal{R}_k| \approx c_4, \quad |\mathcal{R}_{k,N}| \approx c_6 e^{-2N},$$

the corresponding results read

$$\theta_{k,N} \sim e^{-3N}, \quad |\mathcal{R}_k|_{,N} \sim e^{-2N}, \quad \varphi_{k,N} \sim e^{-N}.$$

Next, we move on to the USR stage with $\varepsilon \approx \varepsilon_\mathrm{s} e^{-6(N-N_\mathrm{s})}$ and $\eta \approx 3$. If the leading-order terms are

$$|\mathcal{R}_k| \approx c_3 e^{3N}, \quad |\mathcal{R}_{k,N}| \approx c_5 e^{3N},$$

we obtain

$$\theta_{k,N} \sim e^{-3N}, \quad |\mathcal{R}_k|_{,N} \sim e^{3N}, \quad \varphi_{k,N} \sim e^{-3N}.$$

Similarly, if the leading-order terms are

$$|\mathcal{R}_k| \approx c_4, \quad |\mathcal{R}_{k,N}| \approx c_6 e^{-2N},$$

we have

$$\theta_{k,N} \sim e^{3N}, \quad |\mathcal{R}_k|_{,N} \sim e^{-2N}, \quad \varphi_{k,N} \sim e^{-5N}.$$

Again, all the above results are summarized in Tab. 2, according to the increasing rate of $|\mathcal{R}_k|$. Because of the competition between the two terms in Eqs. (3.8) and (3.9), there are altogether four different combinations at present. From the last two rows, we find that, during the USR stage, $\varphi_{k,N}$ decreases dramatically, so φ_k will reach a constant very soon. Under these circumstances, $\mathcal{R}_k$ will have tiny angular velocity in the complex plane, so there will appear a linear evolution of $\mathcal{R}_k$, to be shown in Figs. 4–6.

η	$\lvert\mathcal{R}_k\rvert$	$\theta_{k,N}$	$\lvert\mathcal{R}_k\rvert_{,N}$	$\varphi_{k,N}$
0	e^{-3N}	e^{3N}	e^{-3N}	e^{3N}
0	const.	e^{-3N}	e^{-2N}	e^{-N}
3	const.	e^{3N}	e^{-2N}	e^{-5N}
3	e^{3N}	e^{-3N}	e^{3N}	e^{-3N}

Table 2. Same as Tab. 1, but in the super-horizon case with $k \ll He^N$. The current situation is more complicated, due to the competition between the two terms in Eqs. (3.8) and (3.9), so all the possible four combinations of the leading-order terms are considered.

Last, we briefly comment that our analytical approximation mainly applies to the two limiting cases with $k \gg He^N$ and $k \ll He^N$, but is unfortunately not suitable for the intermediate region around $N \sim N_{\rm out}$ with $k \sim He^N$, because at this time, it is difficult to determine the value of $[k/(He^N)]^2$ in front of $\mathcal{R}_k$ in Eq. (2.5).

4 Evolution of the primordial curvature perturbation

In this section, we utilize the results in Sec. 3 to provide a whole picture of the evolution of the primordial curvature perturbation $\mathcal{R}_k$ in the USR inflation. Meanwhile, we compare our analytical approximation with the numerical results. For such comparison, five typical scales are chosen, for which the numbers of e-folds when they cross the horizon are $N_{\rm out} = 50$, 52.89, 55, 58.84, and 62, respectively. The reason for choosing these numbers is that they correspond to five typical positions in the power spectrum $\mathcal{P}_{\mathcal{R}}(k)$: the nearly scale-invariant region, sharp dip, steep growth, peak, and falling stage, respectively, as shown in Fig. 2.

Below, the evolution of $\mathcal{R}_k$ with different $N_{\rm out}$ is systematically explored in order and compared with the numerical results (i.e., the lower right panels in Figs. 3–7) in our previous work in Ref. [86].

First, we start from the case with $N_{\rm out} = 50$, in which the relevant scale crosses the horizon much earlier than $N_{\rm s}$, and our results are shown in Fig. 3. Before $N_{\rm out}$, we have $k \gg He^N$ and $\eta \approx 0$. Therefore, $|\mathcal{R}_k| \sim e^{-N}$ [or equivalently, $k^3|\mathcal{R}_k|^2/(2\pi^2) \sim e^{-2N}$], $\theta_{k,N} \sim e^{-N}$, and $\varphi_{k,N} \sim e^{-N}$ (the first row in Tab. 1), as depicted in the blue lines in Figs. 3(a)–3(c). However, after $N_{\rm out}$, we should be cautious and distinguish the SR and USR stages. When $N_{\rm out} < N < N_{\rm s}$ (i.e., the SR stage), we have $k \ll He^N$ and $\eta \approx 0$, so $|\mathcal{R}_k| \sim$ const., $\theta_{k,N} \sim e^{-3N}$, and $\varphi_{k,N} \sim e^{-N}$ (the second row in Tab. 2). However, when $N_{\rm s} < N < N_{\rm e}$ (i.e., the USR stage), we have $k \ll He^N$ and $\eta \approx 3$, so $|\mathcal{R}_k| \sim$ const., $\theta_{k,N} \sim e^{3N}$, and $\varphi_{k,N} \sim e^{-5N}$ (the third row in Tab. 2). These behaviors can also be found

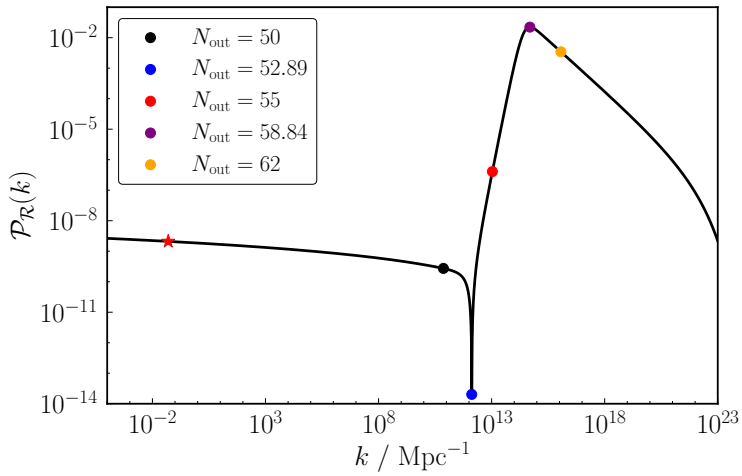


Figure 2. The power spectrum $\mathcal{P}_\mathcal{R}(k)$ in the USR inflation, with the model parameters in Sec. 2.1. On large scales (e.g., the CMB pivot scale $k_* = 0.05$ Mpc^{-1} denoted with an asterisk), $\mathcal{P}_\mathcal{R}$ is nearly scale-invariant with an amplitude of 2.10×10^{-9} [35]. On small scales, $\mathcal{P}_\mathcal{R}$ can be enhanced up to $\mathcal{O}(10^{-2})$ to produce abundant PBHs and SIGWs. Five typical scales with different $N_{\rm out} = 50$, 52.89, 55, 58.84, and 62 are marked as dots with different colors, corresponding to the nearly scale-invariant region, sharp dip, steep growth, peak, and falling stage, respectively. The minimum of $\mathcal{P}_\mathcal{R}$ in the sharp dip is merely 1.61×10^{-14}, but not zero.

in the red and purple lines in Figs. 3(a)–3(c). Altogether, we observe that our analytical approximation fits the numerical results (the dashed lines) perfectly in all these situations. Furthermore, the evolution of $\mathcal{R}_k$ from $N = 47$ to 52 is exhibited in the complex plane in Fig. 3(d), with the difference of the number of e-folds between two adjacent dots set to be $\Delta N = 0.01$. From the results of $|\mathcal{R}_k|$ and $\theta_{k,N}$, it is easy to find that $\mathcal{R}_k$ revolves clockwise towards the origin and eventually stops near it, where $|\mathcal{R}_k|$ tends to be constant. The interval between the dots monotonically decreases, meaning that $\mathcal{R}_k$ decelerates gradually. All these behaviors indicate that the USR stage has negligible effect on such large-scale curvature perturbation, and $|\mathcal{R}_k|$ becomes almost frozen once $N > N_{\rm out}$, which is reflected in the nearly scale-invariant region in $\mathcal{P}_\mathcal{R}$.

Second, we study the scale with $N_{\rm out} = 52.89$, and our results are shown in Fig. 4. In fact, most behaviors of $|\mathcal{R}_k|$, $\theta_{k,N}$, and $\varphi_{k,N}$ [e.g., the slopes of the blue, red, and purple lines in Figs. 4(a)–4(c)] remain the same as those in the previous case. The unique distinction is that, from the first row in Tab. 2, around the end of the USR stage, $|\mathcal{R}_k| \sim e^{-3N}$ decreases dramatically [the green line in Fig. 4(a)], and $\varphi_{k,N}$ is extremely small. Hence, there appears a new linear evolution of $\mathcal{R}_k$ towards but not through the origin subsequent to the revolving process, as shown in Fig. 4(d), with $N = 56.8$ to 62.9. This interesting point naturally explains the sharp dip in $\mathcal{P}_\mathcal{R}$. Moreover, we should emphasize that this dip actually cannot reach zero, consistent with the numerical result [the dashed line in Fig. 4(a)]. If there were $\mathcal{R}_k$ that eventually stops at the origin, we would have $|\mathcal{R}_k| = 0$ and $\theta_{k,N} = 0$ simultaneously. However, from Eq. (2.14), this is strictly forbidden, so the minimum of $\mathcal{P}_\mathcal{R}$ can never be zero, contrary to the naive inference in Ref. [91].

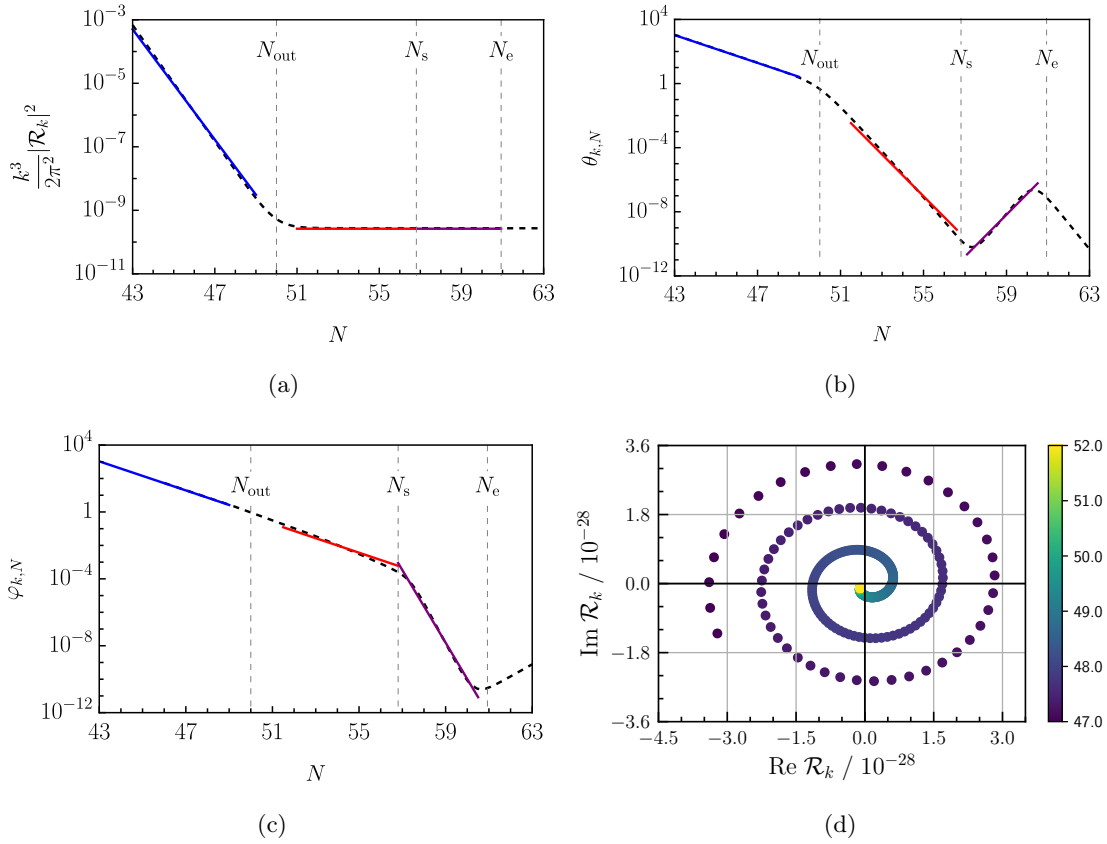


Figure 3. The evolutions of the primordial curvature perturbation $\mathcal{R}_k$ and the two arguments θ_k and φ_k with $N_{\rm out} = 50$. Before $N_{\rm out}$, we have $k \gg He^N$ and $\eta \approx 0$, and $k^3|\mathcal{R}_k|^2/(2\pi^2) \sim e^{-2N}$, $\theta_{k,N} \sim e^{-N}$, and $\varphi_{k,N} \sim e^{-N}$, so the slopes of the blue lines in Figs. 3(a)–3(c) are -2, -1, and -1. Similarly, when $N_{\rm out} < N < N_{\rm s}$, we have $k \ll He^N$ and $\eta \approx 0$, so the slopes of the red lines are 0, -3, and -1. When $N_{\rm s} < N < N_{\rm e}$, we have $k \ll He^N$ and $\eta \approx 3$, so the slopes of the purple lines are 0, 3, and -5. All these values match the numerical results (the dashed lines) very well. Moreover, the evolution of $\mathcal{R}_k$ from $N = 47$ to 52 ($\Delta N = 0.01$) is shown in the complex plane in Fig. 3(d), in which $\mathcal{R}_k$ first revolves around the origin at a decelerated rate and finally stops near it. This means that, for the large-scale curvature perturbation that crosses the horizon much earlier before the USR stage, $|\mathcal{R}_k|$ becomes almost constant once $N > N_{\rm out}$, and the influence from the USR inflation is negligible.

Third, we consider the scale with $N_{\rm out} = 55$, which is before but near $N_{\rm s}$. When $N < N_{\rm out}$, the situation is analogous to the former two cases, as shown in the blue lines in Figs. 5(a)–5(c). However, when $N_{\rm s} < N < N_{\rm e}$, we have $k \ll He^N$ and $\eta \approx 3$, so we have to face the two competitive terms in Eq. (3.8) and determine the leading-order one in advance. As a result, in the early phase of the USR stage, we follow the third row in Tab. 2, so $|\mathcal{R}_k| \sim$ const., $\theta_{k,N} \sim e^{3N}$, and $\varphi_{k,N} \sim e^{-5N}$, corresponding to the red lines. In contrast, in the late phase of the USR stage, we arrive at the last row in Tab. 2, so $|\mathcal{R}_k| \sim e^{3N}$, $\theta_{k,N} \sim e^{-3N}$, and $\varphi_{k,N} \sim e^{-3N}$, as shown with the purple lines. As $|\mathcal{R}_k|$ is significantly amplified and $\varphi_{k,N}$ is tiny simultaneously, $\mathcal{R}_k$ shows a linear evolution away

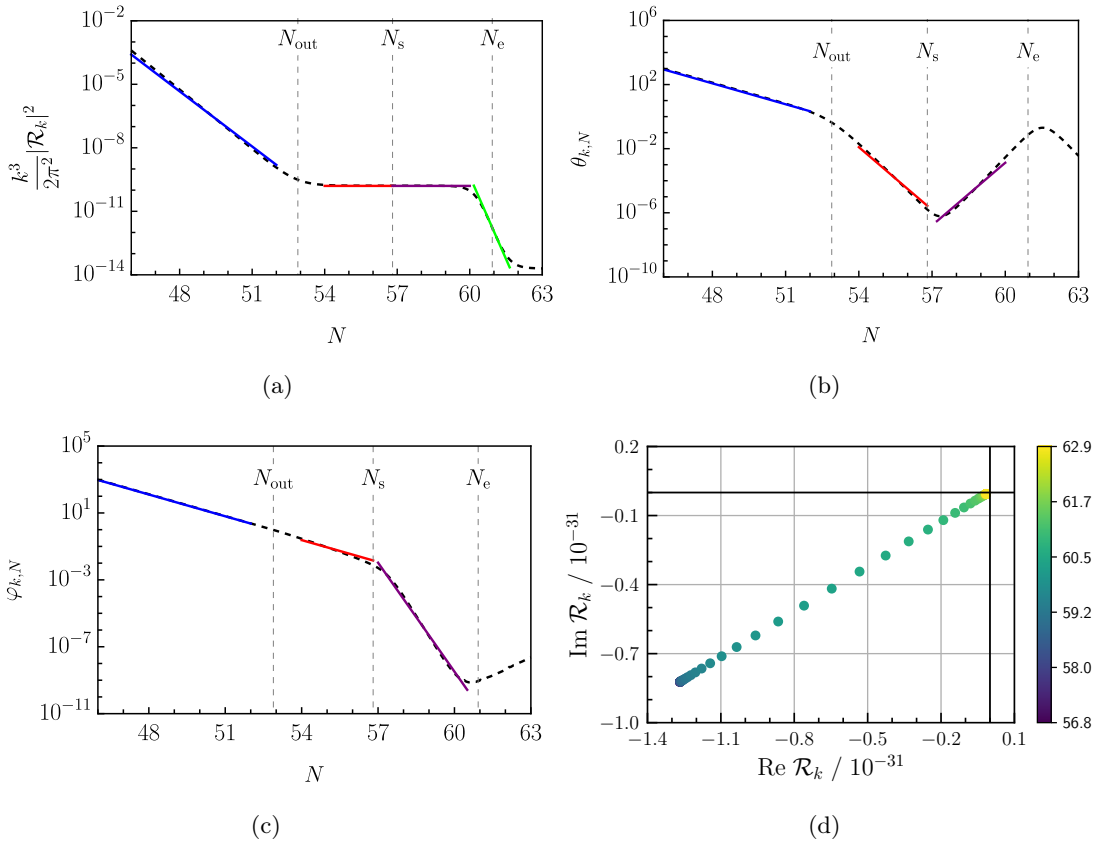


Figure 4. Same as Fig. 3, but with $N_{\rm out} = 52.89$. Before the end of the USR stage, the evolutions of $|\mathcal{R}_k|$, θ_k, and φ_k basically resemble those in Fig. 3. However, around $N_{\rm e}$, $\mathcal{R}_k$ is no longer frozen but decreases dramatically as $|\mathcal{R}_k| \sim e^{-3N}$, and $\varphi_{k,N}$ is extremely small at the same time, so there appears a new linear evolution of $\mathcal{R}_k$ after the revolving process (not shown in the figure) towards but not through origin, as shown in Fig. 4(d) with $N = 56.8$ to 62.9 ($\Delta N = 0.1$). This linear evolution naturally explains the sharp dip in $\mathcal{P}_\mathcal{R}$, and the minimum of the dip is not exactly zero from both theoretical and numerical points of view.

from the origin and finally terminates at a remote point in the complex plane, inducing the steep growth in $\mathcal{P}_\mathcal{R}$, as shown Fig. 5(d) with $N = 52$ to 65. In addition, the competition between the two terms in Eq. (3.8) causes a sharp decrease of $|\mathcal{R}_k|$ at $N \approx 59.15$. From Eq. (2.14), this decrease also explains a spike in $\theta_{k,N}$, as can be seen in Figs. 5(a) and 5(b) (such divergent behaviors are beyond our analytical approximation).

Fourth, the peak of the power spectrum $\mathcal{P}_\mathcal{R}$ corresponds to the scale with $N_{\rm out} = 58.84$, and our analytical approximation is partly suitable to this special case. Now, $N_{\rm out}$ is almost in the middle between $N_{\rm s} = 56.81$ and $N_{\rm e} = 60.93$, but the approximation is inapplicable when $k \sim He^N$. Therefore, we can only find the similar behaviors of $|\mathcal{R}_k|$, $\theta_{k,N}$, and $\varphi_{k,N}$ to those in the previous three cases when $N < N_{\rm s}$ (not $N_{\rm out}$, since it is larger than $N_{\rm s}$ now). During the USR stage, the numerical method is still needed. From the dashed lines in Figs. 6(a) and 6(c), the rapid increase of $|\mathcal{R}_k|$ and decrease of $\varphi_{k,N}$ almost maintain as before in

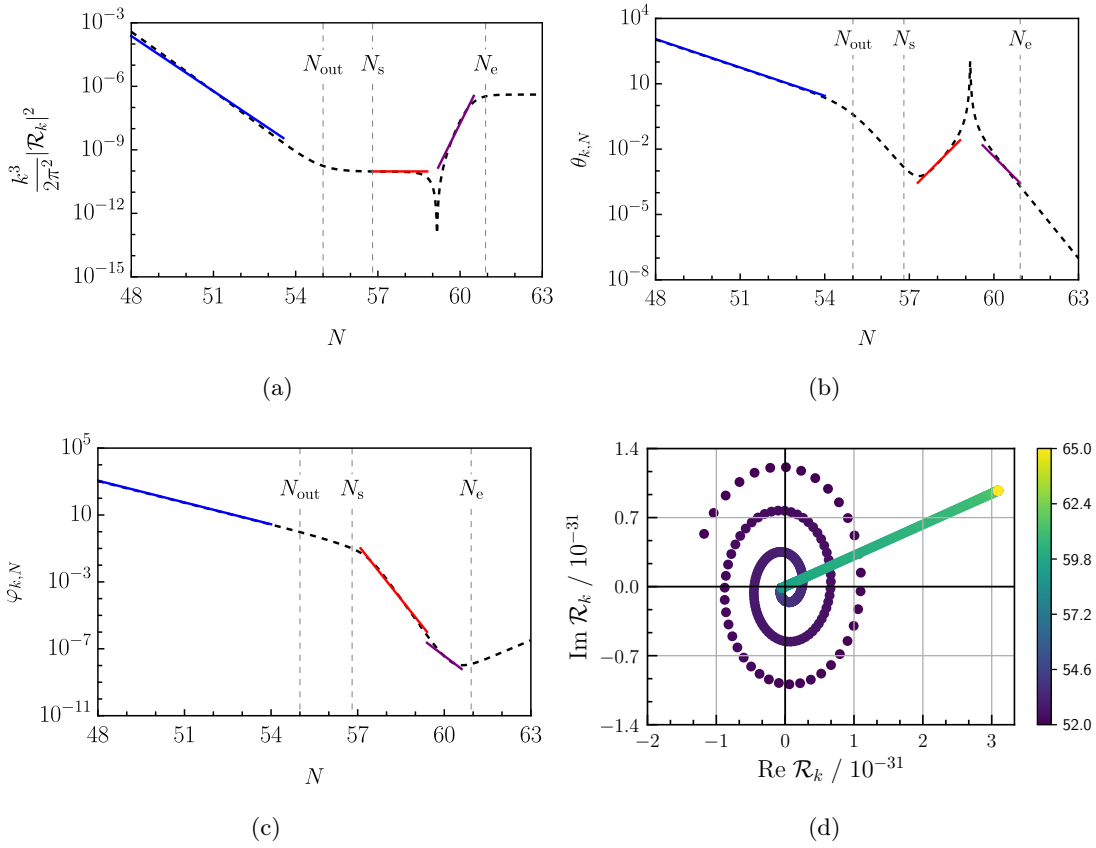


Figure 5. Same as Fig. 3, but with $N_{\rm out} = 55$. When $N < N_{\rm out}$, the evolutions of $|\mathcal{R}_k|$, θ_k, and φ_k are analogous to those in Fig. 3, as shown in the blue lines in Figs. 5(a)–5(c). During the USR stage, we have $k \ll He^N$ and $\eta \approx 3$, but there are two competitive terms in $|\mathcal{R}_k|$ from Eq. (3.8). As a result, in the early phase, we have $|\mathcal{R}_k| \sim$ const., $\theta_{k,N} \sim e^{3N}$, and $\varphi_{k,N} \sim e^{-5N}$, so the slopes of the red lines are 0, 3, and -5. Similarly, in the late phase, the slopes of the purple lines are 6, -3, and -3. Hence, $\mathcal{R}_k$ shows a linear evolution away from the origin and finally stops at a distant point in the complex plane, explaining the steep growth in $\mathcal{P}_\mathcal{R}$, as shown in Fig. 5(d) with $N = 52$ to 65 ($\Delta N = 0.01$). Moreover, at $N \approx 59.15$, a sharp decrease appears in $|\mathcal{R}_k|$ in Fig. 5(a), accompanied by a spike in $\theta_{k,N}$ in Fig. 5(b), as a natural consequence from Eq. (2.14).

Figs. 5(a) and 5(c), so the linear evolution of $\mathcal{R}_k$ still exists, resulting in the peak in $\mathcal{P}_\mathcal{R}$ eventually, as shown in Fig. 6(d) with $N = 58.8$ to 62.

Last, we finish our discussion with $N_{\rm out} = 62$. Because $N_{\rm out} > N_{\rm e}$ at present, we only need to take into account the simple case with $k \gg He^N$. Therefore, in the SR stage, we have $\eta \approx 0$, so $|\mathcal{R}_k| \sim e^{-N}$, $\theta_{k,N} \sim e^{-N}$, and $\varphi_{k,N} \sim e^{-N}$ (the first row in Table 1), as shown in the blue lines in Figs. 7(a)–7(c). Next, during the USR stage, we have $\eta \approx 3$, so $|\mathcal{R}_k| \sim e^{2N}$, $\theta_{k,N} \sim e^{-N}$, and $\varphi_{k,N} \sim e^{-N}$ (the second row in Tab. 1), as shown in the red lines. Since $|\mathcal{R}_k|$ first decreases and then increases in the current situation, $\mathcal{R}_k$ first revolves towards the origin and then away from it, as can be seen in Fig. 7(d) with $N = 59$ to 64. Besides, there is no more linear evolution of $\mathcal{R}_k$, which can also be understood from

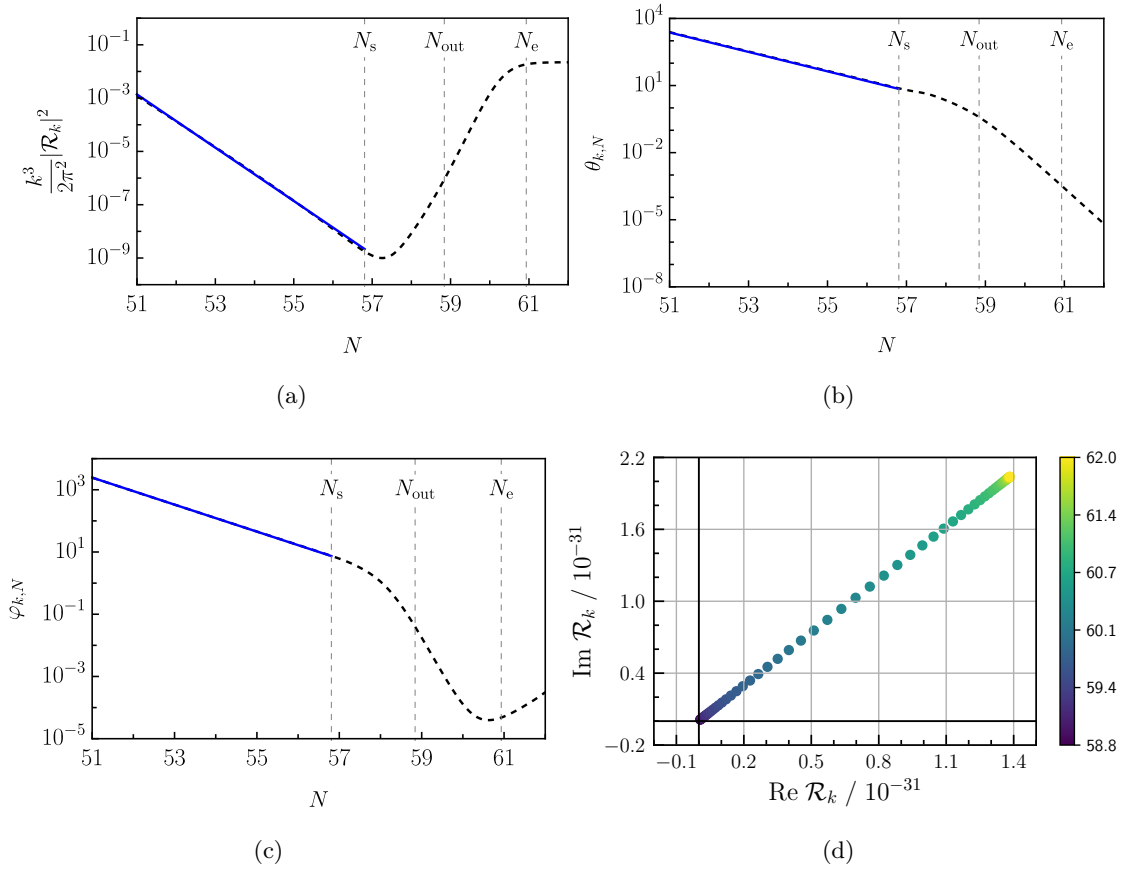


Figure 6. Same as Fig. 3, but with $N_{\rm out} = 58.84$. Here, our analytical approximation applies only to the situation when $N < N_{\rm s}$ with $k \gg He^N$. As shown in the blue lines in Figs. 6(a)–6(c), the evolutions of $|\mathcal{R}_k|$, θ_k, and φ_k are similar to the previous cases. From the dashed lines in Figs. 6(a) and 6(c), the rapid increase of $|\mathcal{R}_k|$ and decrease of $\varphi_{k,N}$ almost maintain as before in Figs. 5(a) and 5(c), so the linear evolution of $\mathcal{R}_k$ still exists. In Fig. 6(d), $\mathcal{R}_k$ linearly moves away from the origin and eventually stops at a rather distant point, resulting in the peak in $\mathcal{P}_\mathcal{R}$ (the previous revolving process is not shown), with $N = 58.8$ to 62 ($\Delta N = 0.05$).

the behavior of φ_k, since $\varphi_{k,N}$ is still large enough in Fig. 7(c).

In summary, we compare our analytical approximation of the evolution of $\mathcal{R}_k$ to the numerical calculation in our previous work in Ref. [86]. Except the epoch around $N_{\rm out}$ with $k \sim He^N$, the theoretical and numerical results match each other very well. We admit that there is indeed some situations where our approximation is insufficient. These intricacies mainly root from the competition between the two terms in Eqs. (3.8) and (3.9), because it is not always easy to determine the leading-order one between them, so numerical calculation is still necessary there. Last, we claim that there are revolving processes of $\mathcal{R}_k$ in all the five typical cases, but the subsequent linear evolutions only exist for the three intermediate scales, and the relevant criterion should be that $|\mathcal{R}_k|$ changes dramatically and $\varphi_{k,N}$ is tiny simultaneously.

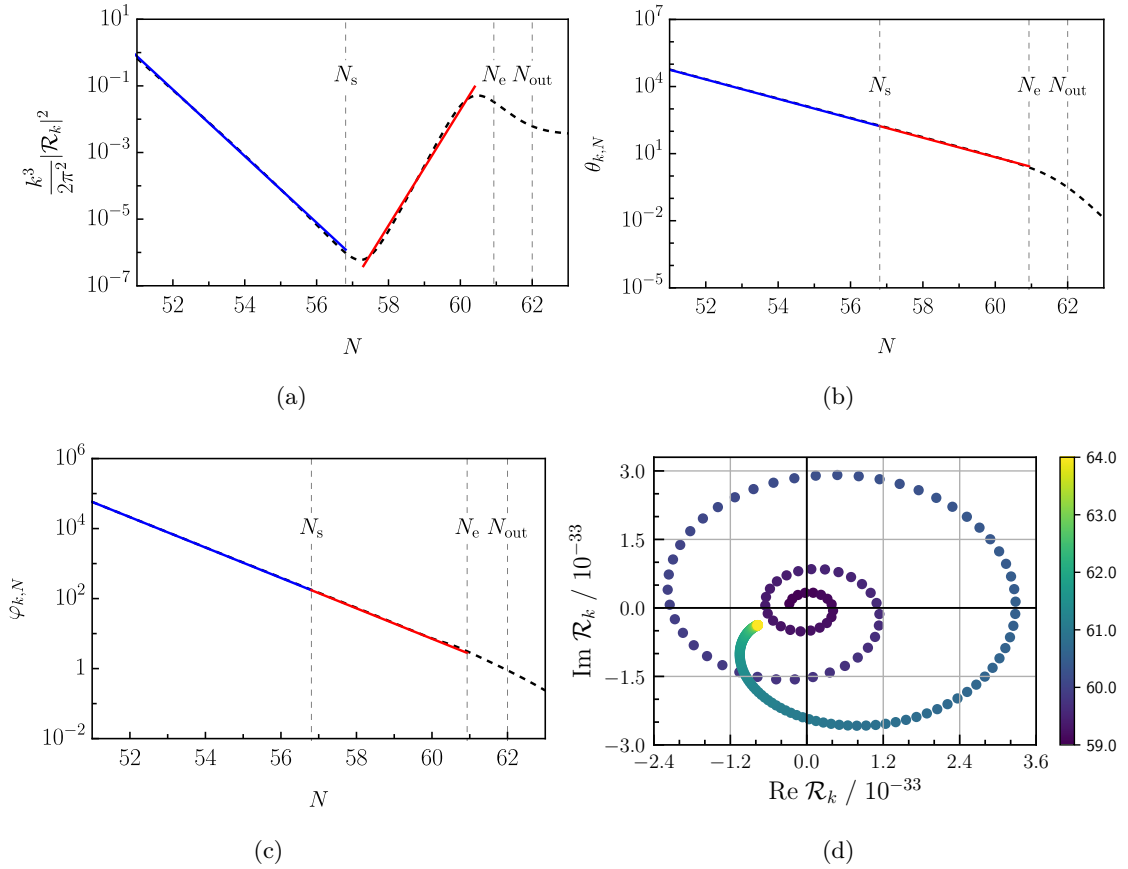


Figure 7. Same as Fig. 3, but with $N_{\rm out} = 62$. Because $N_{\rm out} > N_{\rm e}$, we only need to consider the simple case with $k \gg He^N$. When $N < N_{\rm s}$, we have $\eta \approx 0$, so the slopes of the blue lines in Figs. 7(a)–7(c) are -2, -1, and -1. Next, when $N_{\rm s} < N < N_{\rm e}$, we have $\eta \approx 3$, so the slopes of the red lines are 4, -1, and -1. Moreover, $\mathcal{R}_k$ first revolves towards the origin and then away from it, as can be seen in Fig. 7(d) with $N = 59$ to 64 ($\Delta N = 0.02$). Although $|\mathcal{R}_k|$ changes dramatically around the USR stage, the value of $\varphi_{k,N}$ is still large enough, so $\mathcal{R}_k$ cannot show a linear evolution in the complex plane any more.

5 Conclusion

The USR inflation is receiving increasing research interest in recent years, in which the primordial curvature perturbation $\mathcal{R}_k$ behaves completely different in contrast to that in the SR case. The most remarkable character is that $\mathcal{R}_k$ can still significantly increase after the horizon-exit and thus greatly enhance the power spectrum $\mathcal{P}_\mathcal{R}$ on small scales. As a result, PBHs at certain mass windows can be produced in abundance, serving as an effective candidate of DM. Meanwhile, SIGWs at certain frequencies can also be intense enough to be discovered by current and future gravitational wave detectors.

In this paper, we provide an analytical approximation to systematically investigate the evolution of $\mathcal{R}_k$ and the relevant $\mathcal{P}_\mathcal{R}$ in the USR inflation, which is usually a highly numerical and time-consuming task. The asymptotic solutions of the moduli and arguments of $\mathcal{R}_k$ and $\mathcal{R}_{k,N}$ are obtained in both sub- and super-horizon limits with $k \gg He^N$ and

$k \ll He^N$, perfectly matching the numerical results in our previous work in Ref. [86]. Moreover, five typical scales with different $N_{\rm out}$ are studied in order, corresponding to five different positions in $\mathcal{P}_\mathcal{R}$. In all, we hope to construct a framework to facilitate the analytical calculation of $\mathcal{R}_k$. Our basic conclusions are summarized as follows.

First, we obtain the approximate forms of the parameters ε and η in both SR and USR stages. By this means, all the analytical solutions of $|\mathcal{R}_k|$, $|\mathcal{R}_k|_{,N}$, $\theta_{k,N}$, and $\varphi_{k,N}$ can be expressed in simple exponential forms. However, due to the different asymptotic expansion of the Bessel function, the solutions with $k \ll He^N$ are more complicated than those with $k \gg He^N$, because we have to determine the leading-order term of $\mathcal{R}_k$ in advance. All the possibilities are summarized in Tabs. 1 and 2.

Second, as for the evolution of $\mathcal{R}_k$ for the five typical scales with different $N_{\rm out}$, the numerical results validate our analytical approximation where it is applicable, as shown in Figs. 3–7. The evolution of $\mathcal{R}_k$ can basically be divided into two types: the revolving and linear processes. The former exists in all the five cases. At early times when $k \gg He^N$, in the first four cases with $N_{\rm out} < N_{\rm e}$, we have $|\mathcal{R}_k| \sim e^{-N}$ and $\theta_{k,N} \sim e^{-N}$, so $\mathcal{R}_k$ presents a revolving motion towards the origin in the complex plane with decreasing angular velocity, as shown in Figs. 3(d), 4(d), 5(d), and 6(d). However, in the last case with $N_{\rm out} > N_{\rm e}$, the USR stage causes $|\mathcal{R}_k| \sim e^{2N}$ and $\theta_{k,N} \sim e^{-N}$, so $\mathcal{R}_k$ revolves away from the origin, as shown in Fig. 7(d).

Third, for the three intermediate scales with $N_{\rm out} = 52.89$, 55, and 58.84, besides the revolving motion, a linear evolution of $\mathcal{R}_k$ follows afterward. In these situations, $|\mathcal{R}_k|$ changes violently, but at the same time $\varphi_{k,N}$ is extremely small, so $\mathcal{R}_k$ nearly evolves along a straight line in the complex plane. When $N_{\rm out} = 52.89$, $\mathcal{R}_k$ moves towards but not through the origin in Fig. 4(d), inducing the sharp dip in $\mathcal{P}_\mathcal{R}$ but with a nonvanishing minimum. On the contrary, when $N_{\rm out} = 55$ and 58.84, $\mathcal{R}_k$ departs away from the origin in Figs. 5(d) and 6(d), explaining the steep growth and the peak in $\mathcal{P}_\mathcal{R}$. However, if $N_{\rm out}$ is too small or too big, $\varphi_{k,N}$ remains large enough during the USR stage, so there is only revolving process of $\mathcal{R}_k$ without the subsequent linear evolution.

Altogether, by exploring the evolution of $\mathcal{R}_k$ via our analytical approximation, we wish to provide a whole picture and thorough understanding of the primordial curvature perturbation and the power spectrum in the USR inflation from a theoretical perspective. Our work will be helpful to the model building of the USR inflation and contribute our knowledge of PBH and gravitational wave physics.

Acknowledgments

We are very grateful to G. Franciolini, B.-M. Gu, Y.-C. Liu, S. Pi, D. J. Schwarz, B.-Y. Su, and H.-R. Zhao for fruitful discussions. This work is supported by the Fundamental Research Funds for the Central Universities of China (No. N232410019).

A Derivation of Eq. (2.14)

Below, we show the derivation of Eq. (2.14) in more detail and prove that it is a general result.

For this purpose, it is more convenient to utilize the Mukhanov variable $v_k = \mathcal{R}_k z$. By means of v_k, the MS equation in Eq. (2.5) can be rewritten as

$$v_k'' + \left(k^2 - \frac{z''}{z}\right) v_k = 0, \tag{A.1}$$

where a prime denotes the derivative with respect to the conformal time τ defined as

$$\mathrm{d}\tau = \frac{\mathrm{d}t}{a} = \frac{\mathrm{d}N}{He^N}. \tag{A.2}$$

Next, we decompose v_k into its modulus and argument as

$$v_k = |v_k| e^{-i\theta_k}. \tag{A.3}$$

Since z is real, v_k has the same argument θ_k as that of $\mathcal{R}_k$.

From Eq. (A.3), the first- and second-order derivatives of v_k read

$$\begin{aligned} v_k' &= (|v_k|' - i|v_k|\theta_k') e^{-i\theta_k}, \\ v_k'' &= (|v_k|'' - |v_k|\theta_k'^2 - 2i|v_k|'\theta_k' - i|v_k|\theta_k'') e^{-i\theta_k}. \end{aligned}$$

Substituting these results into Eq. (A.1), it is straightforward to obtain $(|v_k|^2\theta_k')' = 0$, indicating that $|v_k|^2\theta_k'$ is a constant. Equivalently, in terms of $\mathcal{R}_k$ and N, we have

$$|\mathcal{R}_k|^2 z^2 \theta_{k,N} H e^N = \text{const}. \tag{A.4}$$

Here, we should emphasize that Eq. (A.4) is general, as it is a direct result from the MS equation in essential. Consequently, it is valid in both SR and USR stages, and on both sub- and super-horizon scales.

Furthermore, we choose the Bunch–Davies vacuum as a special point to fix the constant in Eq. (A.4). First, we rewrite $\mathcal{R}_k$ in Eq. (2.6) via the conformal time τ as

$$\mathcal{R}_k = \frac{e^{-ik\tau}}{z\sqrt{2k}} \left(1 - \frac{i}{k\tau}\right).$$

Therefore, we have

$$\begin{aligned} |\mathcal{R}_k| &= \sqrt{\frac{1}{2z^2 k}\left[1 + \frac{1}{(k\tau)^2}\right]}, \\ \theta_k &= k\tau + \arctan\frac{1}{k\tau}. \end{aligned}$$

Substituting these results into Eq. (A.4) and taking Eq. (A.2) into account, we can eventually determine the constant in Eq. (A.4) as

$$|\mathcal{R}_k|^2 z^2 \theta_{k,N} H e^N = \frac{1}{2},$$

which is just Eq. (2.14).

References

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Phys. Rev. Lett.* **116** (2016) 061102, arXiv:1602.03837[gr-qc].

[2] S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, *Did LIGO detect dark matter?*, *Phys. Rev. Lett.* **116** (2016) 201301, arXiv:1603.00464[astro-ph].

[3] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, *Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914*, *Phys. Rev. Lett.* **117** (2016) 061101, arXiv:1603.08338[astro-ph].

[4] S. Clesse and J. García-Bellido, *The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO*, *Phys. Dark Univ.* **15** (2017) 142, arXiv:1603.05234[astro-ph].

[5] A. M. Green and B. J. Kavanagh, *Primordial Black Holes as a dark matter candidate*, *J. Phys. G* **48** (2021) 043001, arXiv:2007.10722[astro-ph].

[6] J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, *An absorption profile centred at 78 megahertz in the sky-averaged spectrum*, *Nature* **555** (2018) 67, arXiv:1810.05912[astro-ph].

[7] Z. Zhang, B. Yue, Y. Xu, Y.-Z. Ma, X. Chen, and M. Liu, *Cosmic Radio Background from Primordial Black Holes at Cosmic Dawn*, *Phys. Rev. D* **107** (2023) 083013, arXiv:2303.06616[astro-ph].

[8] S. Clesse and J. García-Bellido, *Massive Primordial Black Holes from Hybrid Inflation as Dark Matter and the seeds of Galaxies*, *Phys. Rev. D* **92** (2015) 023524, arXiv:1501.07565[astro-ph].

[9] M. Castellano *et al.*, *Early results from GLASS-JWST. III: Galaxy candidates at z∼9-15*, *Astrophys. J. Lett.* **938** (2022) L15, arXiv:2207.09436[astro-ph].

[10] R. Saito and J. Yokoyama, *Gravitational wave background as a probe of the primordial black hole abundance*, *Phys. Rev. Lett.* **102** (2009) 161101, arXiv:0812.4339[astro-ph].

[11] G. Domènech, *Scalar Induced Gravitational Waves Review*, *Universe* **7** (2021) 398, arXiv:2109.01398[gr-qc].

[12] W. Ahmed, M. Junaid, and U. Zubair, *Primordial black holes and gravitational waves in hybrid inflation with chaotic potentials*, *Nucl. Phys. B* **984** (2022) 115968, arXiv:2109.14838[astro-ph].

[13] S. Kawai and J. Kim, *Primordial black holes from Gauss-Bonnet-corrected single field inflation*, *Phys. Rev. D* **104** (2021) 083545, arXiv:2108.01340[astro-ph].

[14] J. Lin, S. Gao, Y. Gong, Y. Lu, Z. Wang, and F. Zhang, *Primordial black holes and scalar induced gravitational waves from Higgs inflation with noncanonical kinetic term*, *Phys. Rev. D* **107** (2023) 043517, arXiv:2111.01362[gr-qc].

[15] Z. Yi, Q. Gao, Y. Gong, and Z.-h. Zhu, *Primordial black holes and scalar-induced secondary gravitational waves from inflationary models with a noncanonical kinetic term*, *Phys. Rev. D* **103** (2021) 063534, arXiv:2011.10606[astro-ph].

[16] H. Di and Y. Gong, *Primordial black holes and second order gravitational waves from ultra-slow-roll inflation*, *J. Cosmol. Astropart. Phys.* **07** (2018) 007, arXiv:1707.09578[astro-ph].

[17] G. Agazie *et al.* (NANOGrav Collaboration), *The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background*, *Astrophys. J. Lett.* **951** (2023) L8, arXiv:2306.16213[astro-ph].

[18] J. Antoniadis *et al.*, *The second data release from the European Pulsar Timing Array III. Search for gravitational wave signals*, *Astron. Astrophys.* **678** (2023) A50, arXiv:2306.16214[astro-ph].

[19] D. J. Reardon *et al.*, *Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array*, *Astrophys. J. Lett.* **951** (2023) L6, arXiv:2306.16215[astro-ph].

[20] H. Xu *et al.*, *Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I*, *Res. Astron. Astrophys.* **23** (2023) 075024, arXiv:2306.16216[astro-ph].

[21] A. Afzal *et al.* (NANOGrav Collaboration), *The NANOGrav 15 yr Data Set: Search for Signals from New Physics*, *Astrophys. J. Lett.* **951** (2023) L11, arXiv:2306.16219[astro-ph].

[22] K. Inomata, K. Kohri, and T. Terada, *The Detected Stochastic Gravitational Waves and Subsolar-Mass Primordial Black Holes*, arXiv:2306.17834[astro-ph].

[23] S. Wang, Z.-C. Zhao, J.-P. Li, and Q.-H. Zhu, *Implications of Pulsar Timing Array Data for Scalar-Induced Gravitational Waves and Primordial Black Holes: Primordial Non-Gaussianity $f_{\rm NL}$ Considered*, arXiv:2307.00572[astro-ph].

[24] Z. Yi, Q. Gao, Y. Gong, Y. Wang, and F. Zhang, *Scalar induced gravitational waves in light of Pulsar Timing Array data*, *Sci. China Phys. Mech. Astron.* **66** (2023) 120404, arXiv:2307.02467[astro-ph].

[25] H. Firouzjahi and A. Talebian, *Induced Gravitational Waves from Ultra Slow-Roll Inflation and Pulsar Timing Arrays Observations*, *J. Cosmol. Astropart. Phys.* **10** (2023) 032, arXiv:2307.03164[astro-ph].

[26] S. Balaji, G. Domènech, and G. Franciolini, *Scalar-induced gravitational wave interpretation of PTA data: the role of scalar fluctuation propagation speed*, *J. Cosmol. Astropart. Phys.* **10** (2023) 041, arXiv:2307.08552[astro-ph].

[27] G. Franciolini, A. J. Iovino, V. Vaskonen, and H. Veermäe, *The recent gravitational wave observation by pulsar timing arrays and primordial black holes: the importance of non-gaussianities*, *Phys. Rev. Lett.* **131** (2023) 201401, arXiv:2306.17149[astro-ph].

[28] B. Carr and F. Kühnel, *Primordial Black Holes as Dark Matter: Recent Developments*, *Ann. Rev. Nucl. Part. Sci.* **70** (2020) 355, arXiv:2006.02838[astro-ph].

[29] A. Escrivà, F. Kühnel, and Y. Tada, *Primordial Black Holes*, arXiv:2211.05767[astro-ph].

[30] S. Clesse and J.García-Bellido, *Seven Hints for Primordial Black Hole Dark Matter*, *Phys. Dark Univ.* **22** (2018) 137, arXiv:1711.10458[astro-ph].

[31] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *Constraints on primordial black holes*, *Rept. Prog. Phys.* **84** (2021) 116902, arXiv:2002.12778[astro-ph].

[32] W. H. Press and P. Schechter, *Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation*, *Astrophys. J.* **187** (1974) 425.

[33] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *The Statistics of Peaks of Gaussian Random Fields*, *Astrophys. J.* **304** (1986) 15.

[34] A. M. Green, A. R. Liddle, K. A. Malik, and M. Sasaki, *A new calculation of the mass fraction of primordial black holes*, *Phys. Rev. D* **70** (2004) 041502(R), arXiv:0403181[astro-ph].

[35] N. Aghanim *et al.* (Planck Collaboration), *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, arXiv:1807.06209[astro-ph].

[36] J. García-Bellido and E. R. Morales, *Primordial black holes from single field models of inflation*, *Phys. Dark Univ.* **18** (2017) 47, arXiv:1702.03901[astro-ph].

[37] K. Kannike, L. Marzola, M. Raidal, and H. Veermäe, *Single Field Double Inflation and Primordial Black Holes*, *J. Cosmol. Astropart. Phys.* **09** (2017) 020, arXiv:1705.06225[astro-ph].

[38] C. Germani and T. Prokopec, *On primordial black holes from an inflection point*, *Phys. Dark Univ.* **18** (2017) 6, arXiv:1706.04226[astro-ph].

[39] H. Motohashi and W. Hu, *Primordial Black Holes and Slow-Roll Violation*, *Phys. Rev. D* **96** (2017) 063503, arXiv:1706.06784[astro-ph].

[40] K. Dimopoulos, *Ultra slow-roll inflation demystified*, *Phys. Lett. B* **775** (2017) 262, arXiv:1707.05644[hep-ph].

[41] J. M. Ezquiaga, J. García-Bellido, and E. R. Morales, *Primordial Black Hole production in Critical Higgs Inflation*, *Phys. Lett. B* **776** (2018) 345, arXiv:1705.04861[astro-ph].

[42] G. Ballesteros and M. Taoso, *Primordial black hole dark matter from single field inflation*, *Phys. Rev. D* **97** (2018) 023501, arXiv:1709.05565[hep-ph].

[43] M. Cicoli, V. A. Diaz, and F. G. Pedro, *Primordial Black Holes from String Inflation*, *J. Cosmol. Astropart. Phys.* **06** (2018) 034, arXiv:1803.02837[hep-th].

[44] O. Özsoy, S. Parameswaran, G. Tasinato, and I. Zavala, *Mechanisms for Primordial Black Hole Production in String Theory*, *J. Cosmol. Astropart. Phys.* **07** (2018) 005, arXiv:1803.07626[hep-th].

[45] I. Dalianis, A. Kehagias, and G. Tringas, *Primordial Black Holes from α-attractors*, *J. Cosmol. Astropart. Phys.* **01** (2019) 037, arXiv:1805.09483[astro-ph].

[46] G. Ballesteros, J. B. Jiménez, and M. Pieroni, *Black hole formation from a general quadratic action for inflationary primordial fluctuations*, *J. Cosmol. Astropart. Phys.* **06** (2019) 016, arXiv:1811.03065[astro-ph].

[47] S.-L. Cheng, W. Lee, and K.-W. Ng, *Superhorizon curvature perturbation in ultra-slow-roll inflation*, *Phys. Rev. D* **99** (2019) 063524, arXiv:1811.10108[astro-ph].

[48] N. Bhaumik and R. K. Jain, *Primordial black holes dark matter from inflection point models of inflation and the effects of reheating*, *J. Cosmol. Astropart. Phys.* **01** (2020) 037, arXiv:1907.04125[astro-ph].

[49] R. Mahbub, *Primordial black hole formation in inflationary α-attractor models*, *Phys. Rev. D* **101** (2020) 023533, arXiv:1910.10602[astro-ph].

[50] J. Liu, Z.-K. Guo, and R.-G. Cai, *An analytical approximation of the scalar spectrum in the ultra-slow-roll inflationary models*, *Phys. Rev. D* **101** (2020) 083535, arXiv:2003.02075[astro-ph].

[51] S. S. Mishra and V. Sahni, *Primordial Black Holes from a tiny bump/dip in the Inflaton potential*, *J. Cosmol. Astropart. Phys.* **04** (2020) 007, arXiv:1911.00057[gr-qc].

[52] R.-G. Cai, Z.-K. Guo, J. Liu, L. Liu, and X.-Y. Yang, *Primordial black holes and gravitational waves from parametric amplification of curvature perturbations*, *J. Cosmol. Astropart. Phys.* **06** (2020) 013, arXiv:1912.10437[astro-ph].

[53] D. G. Figueroa, S. Raatikainen, S. Räsänen, and E. Tomberg, *Non-Gaussian tail of the curvature perturbation in stochastic ultra-slow-roll inflation: implications for primordial black hole production*, *Phys. Rev. Lett.* **127** (2021) 101302, arXiv:2012.06551[astro-ph].

[54] H. V. Ragavendra, P. Saha, L. Sriramkumar, and J. Silk, *PBHs and secondary GWs from ultra slow roll and punctuated inflation*, *Phys. Rev. D* **103** (2021) 083510, arXiv:2008.12202[astro-ph].

[55] K.-Y. Choi, S.-b. Kang, and R. N. Raveendran, *Reconstruction of potentials of the hybrid inflation in the light of primordial black hole formation*, *J. Cosmol. Astropart. Phys.* **06** (2021) 054, arXiv:2102.02461[astro-ph].

[56] K. Kefala, G. P. Kodaxis, I. D. Stamou, and N. Tetradis, *Features of the inflaton potential and the power spectrum of cosmological perturbations*, *Phys. Rev. D* **104** (2021) 023506, arXiv:2010.12483[astro-ph].

[57] Y.-C. Liu, Q. Wang, B.-Y. Su, and N. Li, *Primordial black holes from the perturbations in the inflaton potential*, *Phys. Dark Univ.* **34** (2021) 100905.

[58] Q. Wang, Y.-C. Liu, B.-Y. Su, and N. Li, *Primordial black holes from the perturbations in the inflaton potential in peak theory*, *Phys. Rev. D* **104** (2021) 083546, arXiv:2111.10028[astro-ph].

[59] I. Dalianis, G. P. Kodaxis, I. D. Stamou, N. Tetradis, and A. Tsigkas-Kouvelis, *Spectrum oscillations from features in the potential of single-field inflation*, *Phys. Rev. D* **104** (2021) 103510, arXiv:2106.02467[astro-ph].

[60] K. Inomata, E. McDonough, and W. Hu, *Primordial Black Holes Arise When The Inflaton Falls*, *Phys. Rev. D* **104** (2021) 123553, arXiv:2104.03972[astro-ph].

[61] S.-L. Cheng, D.-S. Lee, and K.-W. Ng, *Power spectrum of primordial perturbations during ultra-slow-roll inflation with back reaction effects*, *Phys. Lett. B* **827** (2022) 136956, arXiv:2106.09275[astro-ph].

[62] Y.-P. Wu, E. Pinetti, K. Petraki, and J. Silk, *Baryogenesis from ultra-slow-roll inflation*, *J. High Energy Phys.* **01** (2022) 015, arXiv:2109.00118[astro-ph].

[63] K. Inomata, E. McDonough, and W. Hu, *Amplification of primordial perturbations from the rise or fall of the inflaton*, *J. Cosmol. Astropart. Phys.* **02** (2022) 031, arXiv:2110.14641[astro-ph].

[64] D. G. Figueroa, S. Raatikainen, S. Räsänen, and E. Tomberg, *Implications of stochastic effects for primordial black hole production in ultra-slow-roll inflation*, *J. Cosmol. Astropart. Phys.* **05** (2022) 027, arXiv:2111.07437[astro-ph].

[65] S. Hooshangi, A. Talebian, M. H. Namjoo, and H. Firouzjahi, *Multiple field ultraslow-roll*

inflation: Primordial black holes from straight bulk and distorted boundary, *Phys. Rev. D* **105** (2022) 083525, arXiv:2201.07258[astro-ph].

[66] S. R. Geller, W. Qin, E. McDonough, and D. I. Kaiser, *Primordial black holes from multifield inflation with nonminimal couplings*, *Phys. Rev. D* **106** (2022) 063535, arXiv:2205.04471[hep-th].

[67] G. Franciolini and A. Urbano, *Primordial black hole dark matter from inflation: The reverse engineering approach*, *Phys. Rev. D* **106** (2022) 123519, arXiv:2207.10056[astro-ph].

[68] S. Balaji, H. V. Ragavendra, S. K. Sethi, J. Silk, and L. Sriramkumar, *Observing Nulling of Primordial Correlations via the 21-cm Signal*, *Phys. Rev. Lett.* **129** (2022) 261301, arXiv:2206.06386[astro-ph].

[69] R. N. Raveendran, K. Parattu, and L. Sriramkumar, *Enhanced power on small scales and evolution of quantum state of perturbations in single and two field inflationary models*, *Gen. Relativ. Gravit.* **54** (2022) 91, arXiv:2206.05760[astro-ph].

[70] S. S. Bhatt, S. S. Mishra, S. Basak, and S. N. Sahoo, *Numerical simulations of inflationary dynamics: slow-roll and beyond*, arXiv:2212.00529[gr-qc].

[71] B.-M. Gu, F.-W. Shu, K. Yang, and Y.-P. Zhang, *Primordial black holes from an inflationary potential valley*, *Phys. Rev. D* **107** (2023) 023519, arXiv:2207.09968[astro-ph].

[72] J.-X. Zhao, X.-H. Liu, N. Li, *Primordial black holes and scalar-induced gravitational waves from the perturbations on the inflaton potential in peak theory*, *Phys. Rev. D* **107** (2023) 043515, arXiv:2302.06886[astro-ph].

[73] B. Mu, G. Cheng, J. Liu, and Z.-K. Guo, *Constraints on ultraslow-roll inflation from the third LIGO-Virgo observing run*, *Phys. Rev. D* **107** (2023) 043528, arXiv:2211.05386[astro-ph].

[74] H. V. Ragavendra and L. Sriramkumar, *Observational imprints of enhanced scalar power on small scales in ultra slow roll inflation and associated non-Gaussianities*, *Galaxies* **11** (2023) 34, arXiv:2301.08887[astro-ph].

[75] P. Carrilho, K. A. Malik, and D. J. Mulryne, *Dissecting the growth of the power spectrum for primordial black holes*, *Phys. Rev. D* **100** (2019) 103529, arXiv:1907.05237[astro-ph].

[76] R. Zhai, H. Yu, and P. Wu, *Power spectrum with k^6 growth for primordial black holes*, *Phys. Rev. D* **108** (2023) 043529, arXiv:2308.09286[gr-qc].

[77] G. Franciolini, A. Iovino, Jr., M. Taoso, and A. Urbano, *One loop to rule them all: Perturbativity in the presence of ultra slow-roll dynamics*, arXiv:2305.03491[astro-ph].

[78] C. T. Byrnes, P. S. Cole, and S. P. Patil, *Steepest growth of the power spectrum and primordial black holes*, *J. Cosmol. Astropart. Phys.* **06** (2019) 028, arXiv:1811.11158[astro-ph].

[79] G. Ballesteros, J. Rey, M. Taoso, and A. Urbano, *Stochastic inflationary dynamics beyond slow-roll and consequences for primordial black hole formation*, *J. Cosmol. Astropart. Phys.* **08** (2020) 043, arXiv:2006.14597[astro-ph].

[80] G. Tasinato, *An analytic approach to non-slow-roll inflation*, *Phys. Rev. D* **103** (2021) 023535, arXiv:2012.02518[hep-th].

[81] O. Özsoy and G. Tasinato, *CMB μT cross correlations as a probe of primordial black hole scenarios*, *Phys. Rev. D* **104** (2021) 043526, arXiv:2104.12792[astro-ph].

[82] O. Özsoy and G. Tasinato, *Consistency conditions and primordial black holes in single field inflation*, *Phys. Rev. D* **105** (2022) 023524, arXiv:2111.02432[astro-ph].

[83] P. S. Cole, A. D. Gow, C. T. Byrnes, and S. P. Patil, *Steepest growth re-examined: repercussions for primordial black hole formation*, arXiv:2204.07573[astro-ph].

[84] A. Karam, N. Koivunen, E. Tomberg, V. Vaskonen, and H. Veermäe, *Anatomy of single-field inflationary models for primordial black holes*, *J. Cosmol. Astropart. Phys.* **03** (2023) 013, arXiv:2205.13540[astro-ph].

[85] S. Pi and J. Wang, *Primordial Black Hole Formation in Starobinsky's Linear Potential Model*, *J. Cosmol. Astropart. Phys.* **06** (2023) 018, arXiv:2209.14183[astro-ph].

[86] H.-R. Zhao, Y.-C. Liu, J.-X. Zhao, and N. Li, *The evolution of the primordial curvature perturbation in the ultraslow-roll inflation*, *Eur. Phys. J. C* **83** (2023) 783.

[87] S. Kachru, R. Kallosh, A. D. Linde, and S. P. Trivedi, *de Sitter Vacua in String Theory*, *Phys. Rev. D* **68** (2003) 046005, arXiv:0301240[hep-th].

[88] V. F. Mukhanov, , *Quantum Theory of Gauge Invariant Cosmological Perturbations*, *Sov. Phys. JETP* **68** (1988) 1297.

[89] M. Sasaki, *Gauge Invariant Scalar Perturbations in the New Inflationary Universe*, *Prog. Theor. Phys.* **76** (1986) 1036.

[90] T. S. Bunch and P. C. W. Davies, *Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting*, *Proc. Roy. Soc. Lond. A* **360** (1978) 117.

[91] G. Goswami and T. Souradeep, *Power spectrum nulls due to non-standard inflationary evolution*, *Phys. Rev. D* **83** (2011) 023526, arXiv:1011.4914[astro-ph].

arXiv:2403.01378v1 [hep-lat] 3 Mar 2024

Spontaneous chiral symmetry breaking and mass gap of QCD in finite volume

Xiaolan Meng,[1,2] Bolun Hu,[2] and Yi-Bo Yang[1,2,3,4]

[1] *University of Chinese Academy of Sciences, School of Physical Sciences, Beijing 100049, China*
[2] *CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*
[3] *School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China*
[4] *International Centre for Theoretical Physics Asia-Pacific, Beijing/Hangzhou, China*

(Dated: March 5, 2024)

We present the lattice QCD simulation with the 2+1+1 flavor full QCD ensembles using near-physical quark masses and different spatial sizes L, at $a \sim 0.055$ fm. The results suggest that chiral symmetry is effectively restored at $L \leq 1.0$ fm, while the gap between the nucleon and pion masses remains.

I. INTRODUCTION

Both spontaneous chiral symmetry breaking and confinement are intrinsic features of the strong interaction between quarks and gluons, although their relationship is the subject of debate. Lattice QCD has confirmed the restoration of chiral symmetry and certain aspects of deconfinement at high temperatures [1]. However, there are also studies arguing that the strong interaction between quarks still exists, and therefore, deconfinement is incomplete (e.g., the recent N_f=2 study [2]). These discussions are based on the spatial correlation function of meson interpolators, which differs from the standard hadron mass and would be sensitive to flavors (or other systematics) [3].

All the above lattice QCD calculations focus on the infinite volume limit $m_\pi L \gg 1$, as the finite volume effect on the physical observable can be significant when $m_\pi L$ is not large enough [4]. As evidence, in a Lattice QCD simulation at $a \sim$0.11 fm with $m_\pi L \sim 1$ using near physical light quark mass and $L \sim 1.8$ fm, an effective restoration of chiral symmetry in the mesonic two-point functions was observed [5]. A system with $m_\pi L \leq 1$ is referred to as the ϵ-regime of the chiral perturbative theory (χPT) [4], and reasonable leading order low energy constants of χPT can be extracted using the above lattice calculations [5].

The finite temperature Lattice QCD simulation requires a smaller temporal size to reach higher temperatures. Therefore, one cannot extract the standard ground state hadron mass based on the long-distance temperature correlation function, and the use of the spatial correlation function would be unavoidable. On the other hand, chiral symmetry restoration can also be observed with a long enough temporal size in the ϵ-regime, allowing for reliable extraction of the ground state hadron mass. In this work, we present a preliminary study on the masses of iso-vector mesons and nucleon at different L, using the clover fermion on the HISQ ensembles with near-physical quark masses. We find that the mass gap between the nucleon and pion masses remains even when L is as small as 0.2 fm.

II. NUMERICAL SETUP

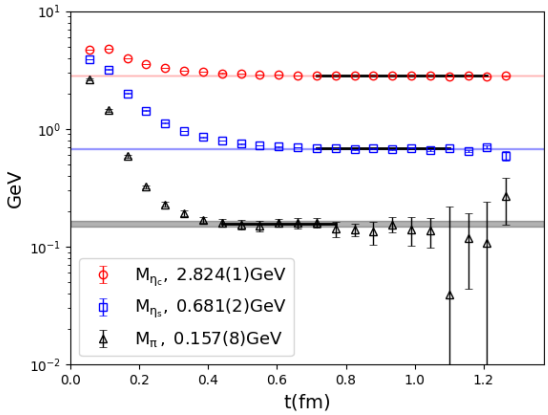


FIG. 1. The pion, η_s η_c effective masses on the 48^4 lattice at $a \sim 0.055$ fm, using the HISQ action.

In order to preserve the chiral symmetry of the light quark in the sea, we chose the HISQ fermion action and one-loop improved Symanzik gauge action used by the MILC collaboration to generate the configurations. The bare coupling and quark masses were interpolated to those at $a = 0.052$ fm based on the parameters used by the MILC configurations [6, 7], as such a lattice spacing is used by the CLQCD ensemble H48P32 [8]. The lattice spacing determined through the Wilson flow [9] is approximately 0.055 fm and close to the target value.

The corresponding effective mass of the pion, η_s, and η_c using this lattice spacing are shown in Fig. 1. Based on the constant fit at a relatively large t, we can extract $m_\pi = 0.157(8)$ GeV, which is quite close to the experimental value 0.135 GeV. The η_s mass $m_{\eta_s} = 0.681(2)$ GeV is also close to the most precise lattice determination of 0.68963(18) MeV [10], and the η_c mass $m_{\eta_c} = 2.824(1)$ GeV is only 5% lower than the experimental value of 2.98 GeV [11].

Since the mixed action effects have been shown to be $\mathcal{O}(a^4)$ with the HISQ fermion sea [12], we use the tadpole

improved clover fermion action with the HYP smeared gauge field for the valence quark, and tune the corresponding pion mass to be about 230 MeV on the 48^4 lattice, which is close to the unitary pion mass. Such a choice is more convenient for investigating the meson and nucleon spectrum without the additional difficulty from the taste-breaking effect of the HISQ action.

TABLE I.

$10/g^2$	$m_l a$	$m_s a$	$m_c a$	V	$n_{\rm cfg}$
6.784	0.000731	0.01975	0.2293	$4^3 \times 96$	238
				$8^3 \times 96$	350
				$8^2 \times 32 \times 96$	100
				$10^3 \times 96$	148
				$12^3 \times 96$	150
				$16^3 \times 96$	146
				$20^3 \times 96$	148
				$24^3 \times 96$	151
				$32^3 \times 96$	98
				$40^3 \times 96$	72
				$40^3 \times 96$	80
				$48^3 \times 48$	100

The ensembles we generated for the this work are summarized in Table I.

III. RESULTS

We calculate the hadron temporal two points correlators including pseudo scalar(P), scalar(S) and proton(N) with clover valence mass $m_u = -0.0414$ at near physical point to extract hadron masses according to the formula

$$\mathcal{C}_{P,S}(t) = \sum_{\vec{x},\vec{y}} \langle \mathcal{O}(\vec{x},t)\mathcal{O}(\vec{y},0)^\dagger \rangle, \tag{1}$$

with P,S interpolators defined as

$$\mathcal{O}_P = \bar{u}\gamma_5 d,\ \mathcal{O}_S = \bar{u}d, \tag{2}$$

and nucleon two-point functions defined as

$$\mathcal{C}_N(t) = \sum_{\vec{x},\vec{y}} \epsilon^{abc}\epsilon^{a'b'c'} \langle \bar{d}^a(\vec{x},t)(C\gamma_5)\bar{u}^{b^T}(\vec{x},t)\bar{u}^c(\vec{x},t) \Gamma_e u^{a'}(\vec{y},0)d^{b'^T}(\vec{y},0)(C\gamma_5)u^{c'} \rangle, \tag{3}$$

with $\Gamma_e = (1+\gamma_4)/2$. When computing quark propagators, Coulomb wall sources with random positions in the temporal direction are used to reduce autocorrelation effects between configurations. The values for the pion masses are extracted from fitting correlators by one-state fit with fitting range selected to ensure acceptable χ^2/d.o.f.$\sim$1,

$$\mathcal{C}_P(t) = A\ \cosh(m_\pi(T-t)) \tag{4}$$

with fitting parameters A, m_π. The result is shown in Fig. 2. We observe clear plateaus in the effective mass

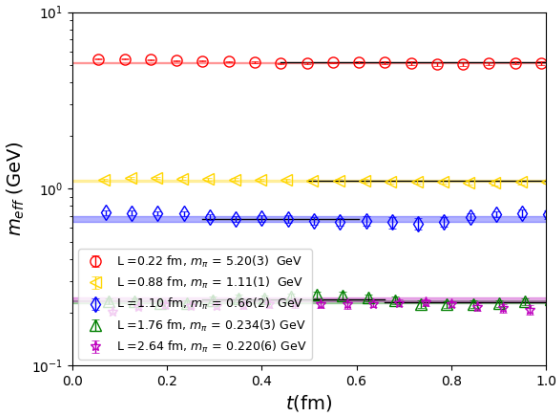


FIG. 2. The non-singlet pion effective masses at different spatial size L.

$m_{\rm eff}(t)$ for all values of L, which suggests that $\mathcal{C}_P(t)$ exhibits exponential decay regardless of the spatial volume when T is large enough. When $L \sim 0.23$ fm, m_π becomes massive around 5 GeV. The m_π rapidly decreases as L increases, and stabilizing at around 230 MeV when $L > 1.8$ fm.

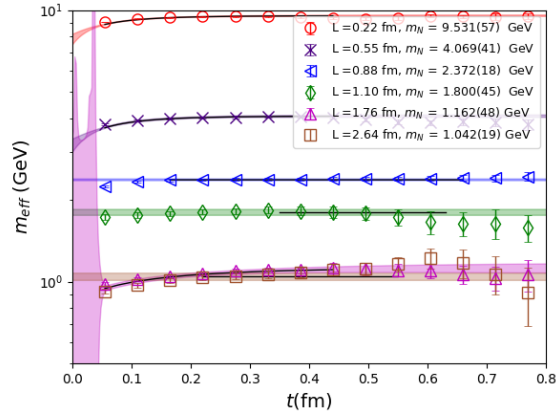


FIG. 3. The nucleon effective masses at different spatial size L.

For proton mass, due to poor signal at large t, we employ a two-state fit for $\mathcal{C}_N(t)$ at relatively smaller t in the ensembles with certain L to eliminate the excited state contamination,

$$\mathcal{C}_N(t) = c_1 e^{-m_N t}(1 + c_2 e^{-\Delta m t}), \tag{5}$$

where m_N, Δm and $c_{1,2}$ are fitting parameters. The remaining cases are fitted with single state($c_2 = 0$). It is revealed that proton masses also follow similar patterns with changes in volume as pion: m_N decreases when spatial extent increases, and tends to about 1 GeV at infinite volume limit. Since pion and nucleon have different num-

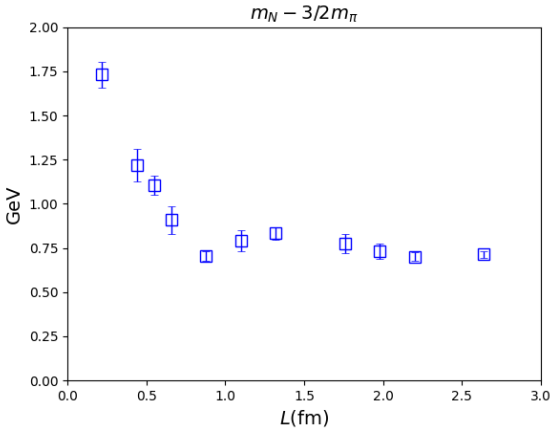


FIG. 4. The mass difference $m_N - 3/2m_\pi$ as function of L.

ber of quarks, we define the combination $m_N - 1.5m_\pi$ to examine the mass gap between proton and pion based on their quark contents. The result in Fig. 4 shows the mass gap between the proton and pion remains above 0.7 GeV and increases even at small volume, which suggests that the QCD confinement remains even at very small L. From another perspective shown in Fig. 5, $m_\pi L$ has a lower limit(around 2) and tends towards $6\sim 2\pi$ at small volumes, which imply the epsilon regime cannot be reached simply by reducing the spatial volume and QCD is not completely dominated by the possible massless pion field at small L. In addition, the variation in the pion and iso-vector scalar meson a_0 two-point function can reflect the nature of chiral symmetry. If chiral symmetry is effectively restored, the quark propagator $\mathcal{S}(x,y) = \psi(x)\bar{\psi}(y)$ on each individual configurations with negligible quark mass will anti-commutate with γ_5, $\{\gamma_5, \mathcal{S}\} \sim 0$. Thus we have

$$\mathcal{C}_P(t) = \sum_{\vec{x},\vec{y}} \langle \mathrm{tr}[\gamma_5 \mathcal{S}(\vec{0},0;\vec{y},t)\gamma_5 \mathcal{S}(\vec{y},t;\vec{0},0)] \rangle = -\mathcal{C}_S(t), \tag{6}$$

and then $\mathcal{C}_P(t)$ and $\mathcal{C}_S(t)$ will be degenerate. The meson two-point functions are shown in Fig. 6 and Fig. 7. When $L < 0.7$ fm, $\mathcal{C}_P(t) = -\mathcal{C}_S(t)$ holds effectively for all time slices except $t = 0$, which may indicate the effectively restoration of chiral symmetry. When L reaches 0.9 fm, $\mathcal{C}_P(t)$ begins to slightly deviate from $-\mathcal{C}_S(t)$ and significantly deviates as L increases to 1.1 fm, including an unexpected sign change in the latter. $\mathcal{C}_S(t)$ returns to negative values at $L \sim 1.8$ fm and show a faster exponential decay at relatively small t which would corresponds to the physical a_0 state in the realistic QCD. $\mathcal{C}_S(t)$ exhibit irregular behaviours as volume increase continually, possibly due to the contamination from the other 0^{++} states.

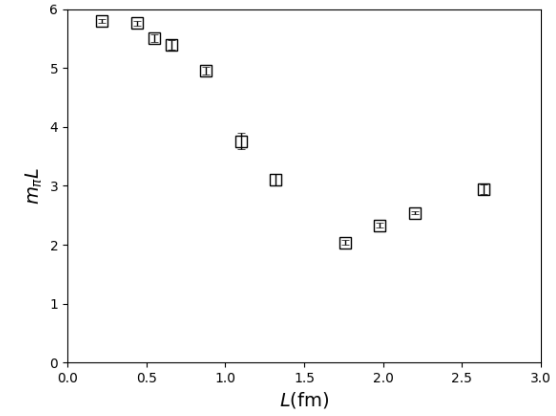


FIG. 5. $m_\pi L$ as the function of L.

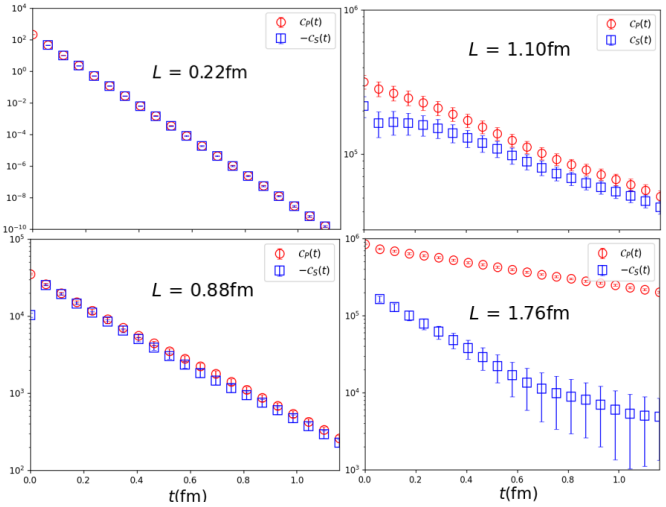


FIG. 6. pion and iso-vector scalar meson a_0 two-point function as function of t, at $L = \{0.2, 0.9, 1.1, 1.8\}$ fm.

The mass splitting between the a_0 and pion also serves as indicator of the chiral symmetry breaking. We determine m_{a_0} using the same formula and fitting patterns as in the case of nucleon. As depicted in Fig. 8 and Fig. 9, while the mass degeneration the two mesons occur naturally of as $\mathcal{C}_P(t) = -\mathcal{C}_S(t)$ at $L < 0.7$ fm, the mass split are still small with $L \in (0.9, 1.5)$ fm, even though obvious difference has been observed in the two point functions above 1.1 fm. m_{a_0} increase to about 0.8 GeV at $L \sim 1.8$ fm if we fit $\mathcal{C}_S(t)$ at small t, and then the uncertainty of $\mathcal{C}_S(t)$ increases rapidly with even larger L. This observation may indicate that the chiral phase transition using L as the order parameter would be a crossover, similar to that using T at finite temperature.

The results of hadron masses are summarized in Fig. 9, showing m_π and m_N stabilization for $L > 1.8$ fm, suggesting minimal finite volume effects, whereas m_{a_0} exhibits fluctuations due to complex 0^+-state contamina-

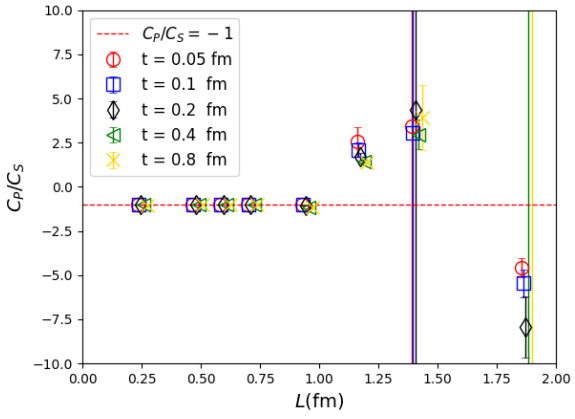


FIG. 7. The chiral symmetry breaking effect $C_P(t)/C_S(t)$ as the function of L, at $t = \{0.05, 0.1, 0.2, 0.4, 0.8\}$ fm.

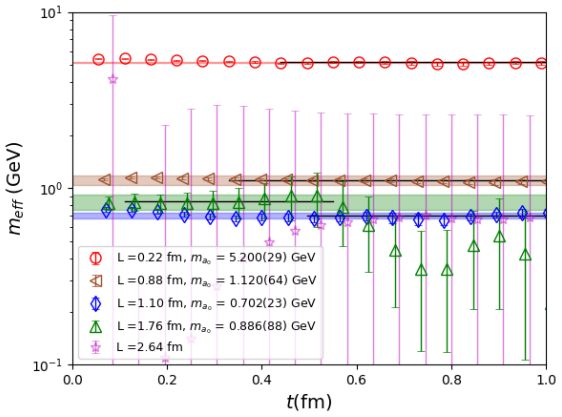


FIG. 8. The a_0 effective masses at different spatial size L.

tion.

We also show an additional case on a $32\times8^2\times96$ lattice for comparison in Fig. 9. The hadron masses m_{π,N,a_0} are quite close to those on the $8^3\times96$ lattice. It suggests that QCD properties would be dominated by the spatial direction with shorter extent in those cases.

IV. SUMMARY

We studied the hadron spectrum in the ϵ-regime with near-physical quark masses with different L, using the clover valence fermion on the HISQ sea configurations at a =0.05 fm. The results show that the pion mass receives a huge $1/L$ correction at small L, leading to an increase in the corresponding $m_\pi L$ when $L \leq 1.2$ fm, where the chiral symmetry breaking between the scalar and pseudoscalar correlation functions is effectively restored. At the same time, the mass gap $m_N - 3m_\pi/2$ is always larger than 700 MeV in the entire $L \in (0.4, 2.5)$

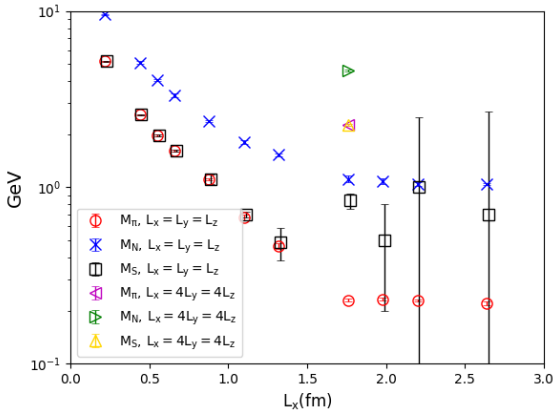


FIG. 9. The non-singlet pion, a_0 and nucleon masses m_π, m_{a0} and m_N (upper panel), and mass difference $m_N - 3/2m_\pi$ on the $L^3 \times 96$ lattice at $a \sim 0.055$ fm

fm range we studied. Such an observation would suggest that the confinement of QCD can exist when the spontaneous chiral symmetry breaking is absent.

Since the valence clover fermion action we used breaks the chiral symmetry additively, it would be non-trivial to further investigate the origin of the restoration of the chiral symmetry and the remaining mass gap. The overlap fermion action would be more suitable for this task, and the study is ongoing. The mixed action effect on the above observations would also be a concern, and further study at multiple lattice spacings is also essential. In addition, the ensemble we used with the largest L gives $m_\pi L \sim 3$, which is still smaller than the usual requirement $m_\pi L \sim 3.5$ which can suppress the finite volume effects efficiently. Thus, simulation with a larger L would be helpful verify the infinite volume limit.

ACKNOWLEDGEMENT

We thank the MILC collaborations for sharing their inputs of the gauge configurations, Carleton Detar, Keh-Fei Liu and Shoji Hashimoto for useful information and discussion. The calculations were performed using the MILC [6, 7] and Chroma software suite [13] with QUDA [14–16] through HIP programming model [17]. The numerical calculation were carried out on the ORISE Supercomputer, and HPC Cluster of ITP-CAS. This work is supported in part by NSFC grants No. 1229060, 1229062, 1229065 and 12047503, the science and education integration young faculty project of University of Chinese Academy of Sciences, the Strategic Priority Research Program of Chinese Academy of Sciences, Grant No. XDB34030303 and YSBR-101, and also a NSFC-DFG joint grant under Grant No. 12061131006 and SCHA 458/22.

[1] A. Bazavov *et al.*, Phys. Rev. D **85**, 054503 (2012), arXiv:1111.1710 [hep-lat].

[2] C. Rohrhofer, Y. Aoki, G. Cossu, H. Fukaya, C. Gattringer, L. Y. Glozman, S. Hashimoto, C. B. Lang, and S. Prelovsek, Phys. Rev. D **100**, 014502 (2019), arXiv:1902.03191 [hep-lat].

[3] T.-W. Chiu, Phys. Rev. D **107**, 114501 (2023), arXiv:2302.06073 [hep-lat].

[4] J. Gasser and H. Leutwyler, Phys. Lett. B **188**, 477 (1987).

[5] H. Fukaya, S. Aoki, S. Hashimoto, T. Kaneko, H. Matsufuru, J. Noaki, K. Ogawa, T. Onogi, and N. Yamada (JLQCD), Phys. Rev. D **77**, 074503 (2008), arXiv:0711.4965 [hep-lat].

[6] A. Bazavov *et al.* (MILC), Phys. Rev. D **82**, 074501 (2010), arXiv:1004.0342 [hep-lat].

[7] A. Bazavov *et al.* (MILC), Phys. Rev. D **87**, 054505 (2013), arXiv:1212.4768 [hep-lat].

[8] Z.-C. Hu, B.-L. Hu, J.-H. Wang, M. Gong, L. Liu, P. Sun, W. Sun, W. Wang, Y.-B. Yang, and D.-J. Zhao, (2023), arXiv:2310.00814 [hep-lat].

[9] M. Lüscher, JHEP **08**, 071 (2010), [Erratum: JHEP 03, 092 (2014)], arXiv:1006.4518 [hep-lat].

[10] S. Borsanyi *et al.*, Nature **593**, 51 (2021), arXiv:2002.12347 [hep-lat].

[11] P. A. Zyla *et al.* (Particle Data Group), PTEP **2020**, 083C01 (2020).

[12] D.-J. Zhao, G. Wang, F. He, L. Jin, P. Sun, Y.-B. Yang, and K. Zhang (χQCD), Phys. Rev. D **107**, L091501 (2023), arXiv:2207.14132 [hep-lat].

[13] R. G. Edwards and B. Joo (SciDAC, LHPC, UKQCD), *Lattice field theory. Proceedings, 22nd International Symposium, Lattice 2004, Batavia, USA, June 21-26, 2004*, Nucl. Phys. Proc. Suppl. **140**, 832 (2005), [,832(2004)], arXiv:hep-lat/0409003 [hep-lat].

[14] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi, Comput. Phys. Commun. **181**, 1517 (2010), arXiv:0911.3191 [hep-lat].

[15] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower, and S. Gottlieb, in *SC11 International Conference for High Performance Computing, Networking, Storage and Analysis Seattle, Washington, November 12-18, 2011* (2011) arXiv:1109.2935 [hep-lat].

[16] M. A. Clark, B. Jo, A. Strelchenko, M. Cheng, A. Gambhir, and R. Brower, (2016), arXiv:1612.07873 [hep-lat].

[17] Y.-J. Bi, Y. Xiao, M. Gong, W.-Y. Guo, P. Sun, S. Xu, and Y.-B. Yang, *Proceedings, 37th International Symposium on Lattice Field Theory (Lattice 2019): Wuhan, China, June 16-22 2019*, PoS **LATTICE2019**, 286 (2020), arXiv:2001.05706 [hep-lat].

PROCEEDINGS OF SCIENCE

Angular analyses of rare decays at the LHC

Biplab Dey, on behalf of the LHCb collaboration[a,*]

[a] *Eotvos University,*
Budapest, Hungary

E-mail: biplab.dey@cern.ch

Loop-suppressed penguin $b \to s$ transitions are sensitive to heavy New Physics particles propagating inside the loops. Thanks to the large sample sizes from the LHC, we are able to perform multidimensional angular analyses that are sensitive to interferences between the Standard Model and New Physics terms. This article surveys the latest results, primarily from LHCb, on $b \to s\mu^+\mu^-$ electroweak and $b \to s\gamma$ radiative penguins.

16th International Conference on Heavy Quarks and Leptons (HQL2023)
28 November-2 December 2023
TIFR, Mumbai, Maharashtra, India

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

arXiv:2403.01143v1 [hep-ex] 2 Mar 2024

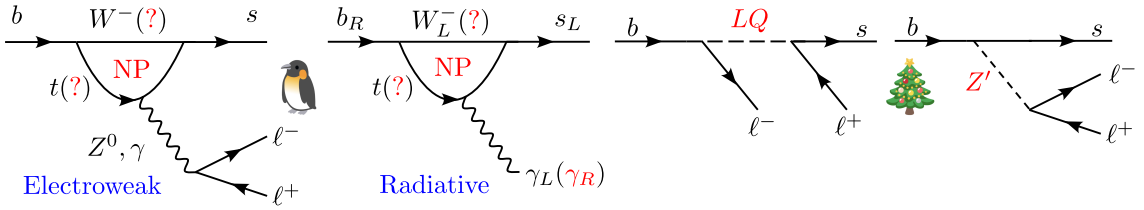


Figure 1: Flavor-changing neutral currents occur only at the loop-level in the SM but can be enhanced by NP effects, both in loop (penguin) and tree-level diagrams.

1. Introduction and theory

In the Standard Model (SM), the flavor-changing neutral current process $b \to s$ is forbidden at the tree-level and proceeds only via loop-suppressed diagrams as shown in Fig. 1. These provide excellent avenues to probe New Physics (NP) contributions that can enter either in loop- or tree-level processes such as via Leptoquarks (LQ) or heavy Z' boson, as shown in Fig. 1. This article focuses on the electroweak (EWP) and radiative (Rad) penguin diagrams, but gluonic penguins can also be an important NP source. A convenient theoretical formalism to study such decays is to regard the SM as a low energy effective field theory containing dimension $d \leq 4$ local operators from renormalizability requirements. Higher dimensional operators can be added with an appropriate cutoff scale Λ, as

$$\mathcal{L}_{\text{eff}}(x) = \mathcal{L}_{\text{SM}}(x) + \sum_{d>4} \frac{C_i}{\Lambda^{d-4}} O_i^{(d)}(x) \tag{1}$$

whereby the NP amplitudes have $(E/\Lambda)^{d-4}$ behavior in the energy, E: divergent at high energies, but suppressed at $E \ll \Lambda$. Most relevant for rare $b \to s$ decays are $d = 6$ operators that yield $\mathcal{A}_{\text{eff}} \sim C^{\text{SM}}/m_W^2 + C^{\text{NP}}/\Lambda_{\text{NP}}^2$. The basis comprises 10 operators [1]: $O_{1,2}$ (4-quark tree), O_{3-6} (4-quark penguins) and O_8 (gluonic penguin) that are suppressed for the EWP/Rad modes. The dominant left-handed contributions are from the electromagnetic dipole and weak vector (axialvector) operators

$$O_{7\gamma} = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu}, \quad O_{9V(10A)} = \frac{e^2}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu(\gamma_5)\ell). \tag{2}$$

The corresponding right-handed (quark side) operators are suppressed in the SM, but can be enhanced in NP scenarios. The dimensionless couplings (Wilson coefficients) associated with the operators in Eq. 2 encode the short distance physics. They are calculated at the m_W scale by integrating out the heavy degrees of freedom from the full theory and evolving to the m_b scale using renormalization group equations. The total amplitudes $\mathcal{A}(i \to f) = C_n(m_b)\langle f|O_n(m_b)|i\rangle_{\text{had}}$ also contains the long-distance physics (QCD/hadronization) which mostly comes from *local* form-factors (FFs) that are computed from lattice QCD and other theory tools, but can get important non-local contributions (rescattering, charm loops) that are hard to estimate theoretically.

Thanks to the large $b\bar{b}$ samples available at the LHC, a comprehensive effort on multidimensional angular analyses in $b \to s\ell^+\ell^-$ and $b \to s\vec{\gamma}$ is ongoing. These offer a rich set of angular observables sensitive to $\Delta C_i \equiv C_i^{\text{SM}} - C_i^{\text{NP}}$. The thrust has been to identify and probe "theoretically clean observables" with reduced dependence on the QCD contributions that often form the

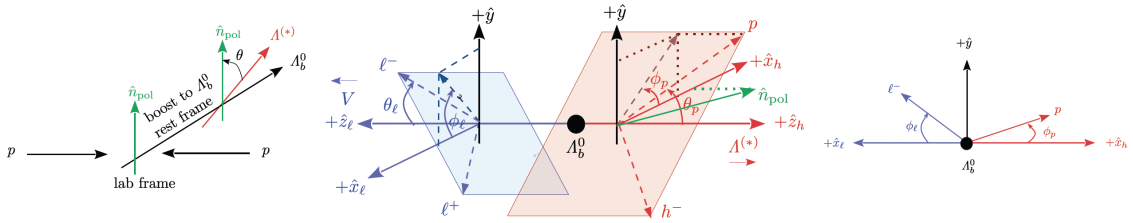


Figure 2: The four angular variables for the decay of a polarized $\vec{\Lambda}_b^0$ decaying to a 4-body final state, $pK^-\ell^+\ell^-$: θ_l (θ_p) is the ℓ^- (proton) helicity angle. The beam direction, Λ^* flight direction and the Λ_b^0 polarization defines a coordinate system relative to which the dilepton and dihadron azimuthal angles can be defined as ϕ_ℓ and ϕ_p.

largest theory uncertainties. The angles for a typical 4-body final state decay is shown in Fig. 2 for $\vec{\Lambda}_b^0 \to pK^-\ell^+\ell^-$; the two other kinematic variables being $q^2 \equiv m^2_{\ell^+\ell^-}$ and $k^2 \equiv m^2_{pK}$. For an unpolarized parent Λ_b^0 (or spin-0 $B_{(s)}$ mesons), ϕ_ℓ is set to 0 and $\chi = \phi_p$ is the single azimuthal angle between the dilepton and dihadron decay planes. Equivalent variables apply for $B \to K\pi\ell^+\ell^-$ and $B_s^0 \to K^+K^-\ell^+\ell^-$.

2. Electroweak penguins

2.1 The golden channel: $B^0 \to K^{*0}\ell^+\ell^-$

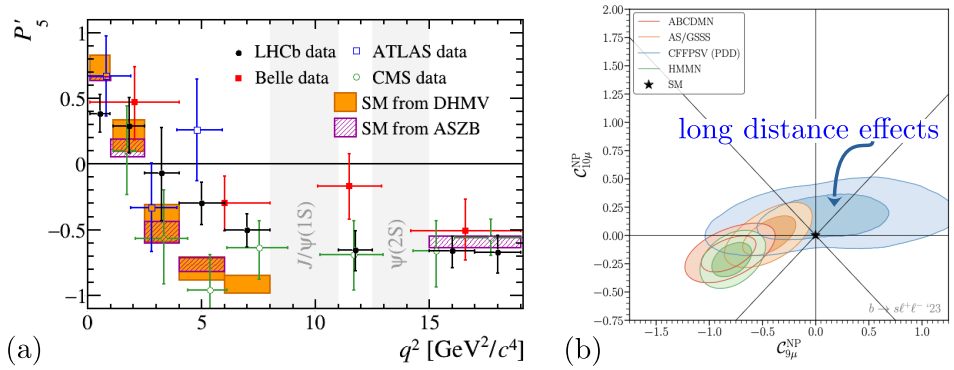


Figure 3: $B^0 \to K^{*0}\mu^+\mu^-$: (a) world-average of the P_5' measurements compared to theory (filled/hatched boxes). The situation is representative after publication of the LHCb Run 1 results. (b) effect on $C_{9,10}^{\rm NP}$ from a recent global fit in Ref. [2].

$B^0 \to K^{*0}\ell^+\ell^-$ has been a gold-plated channel [1] since the B-factory era due to the relatively narrow $K^{*0}(892)$ resonance. Especially in the low q^2 regime, where the recoiling K^{*0} has a large γ-factor in the parent B^0 frame, QCD sum rules on the light-cone (LCSR) affords control over the FF calculations. However, with the BaBar/Belle statistics, only 1-dimensional angular analyses in either the lepton or hadron helicity angles ($\cos\theta_l$ or $\cos\theta_K$) were possible [3]. Full 3-dimensional analysis in $d\theta_l d\theta_K d\chi$ in q^2 bins was possible only with the advent of LHCb. For instance, while the full BaBar dataset had $O(50)$ $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ events, the existing Run1+2 LHCb dataset already includes $O(10^4)$ clean signal events for this muonic mode. On the other hand,

B-factories, including Belle II, have complementary advantages, with better reconstruction for the π^0 isospin modes as well the dielectron channels.

An important result from the muonic analyses at LHCb is tension with SM predictions in the angular observable P_5' [4] (see Fig. 3a). Similar tensions have also been seen in $B^+ \to K^{*+}\mu^+\mu^-$ [5] and $B_s^0 \to \phi\mu^+\mu^-$ [6]. A related tension is observed in the overall branching fractions in several $b \to s\mu^+\mu^-$ modes which tend to consistently lie lower than the SM predictions [7–9]. Competitive results have also come from ATLAS/CMS [10–12] where the advantage is higher overall luminosity, but the disadvantage is the limited B-physics trigger bandwidth and lack of a RICH detector for $K^+/\pi^+/p$ separation. Numerous global fits with different data subsets, statistical methods and theory priors have been performed [2] pointing to a preferred negative $C_9^{\rm NP}$. The major point of contention, however, has been the effect of non-factorizable long-distance contributions due to soft+hard gluons from charm-loops that can mimic NP effects (see Fig. 3b). To constrain the non-factorizable part in a data-driven fashion, LHCb has performed an unbinned angular analysis [13, 14] using the same dataset as in Ref. [5]. The underlying transversity amplitudes are

$$\mathcal{A}^{L,R}_{\lambda=0,\parallel,\perp} = N_\lambda \left\{ \left[(C_9 \pm C_9') \mp (C_{10} \pm C_{10}') \right] \mathcal{F}_\lambda + \frac{2m_b M_B}{q^2} \left[(C_7 \pm C_7') \mathcal{F}_\lambda^T - 16\pi^2 \frac{M_B}{m_b} \mathcal{H}_\lambda \right] \right\}, \quad (3)$$

where $\mathcal{F}_\lambda$ are the usual local FFs (taken from LCSR and lattice QCD) and $\mathcal{H}_\lambda$ are the new non-factorizable part which are extracted from a q^2-dependent parameterization. The values of $C_i^{\rm SM}$ are taken from theory, while allowing for $C_{9,10}^{\rm NP}$ to be floated. The results of the fit are shown in Fig. 4. Good consistency is found in the extracted binned observables compared to Ref. [5]. The overall tension with the SM is reduced to $\sim 1.8\sigma$ in C_9, and $\sim 1.4\sigma$ in global fits. The full Run 1+2 analysis and more precise theory FFs will improve upon these results.

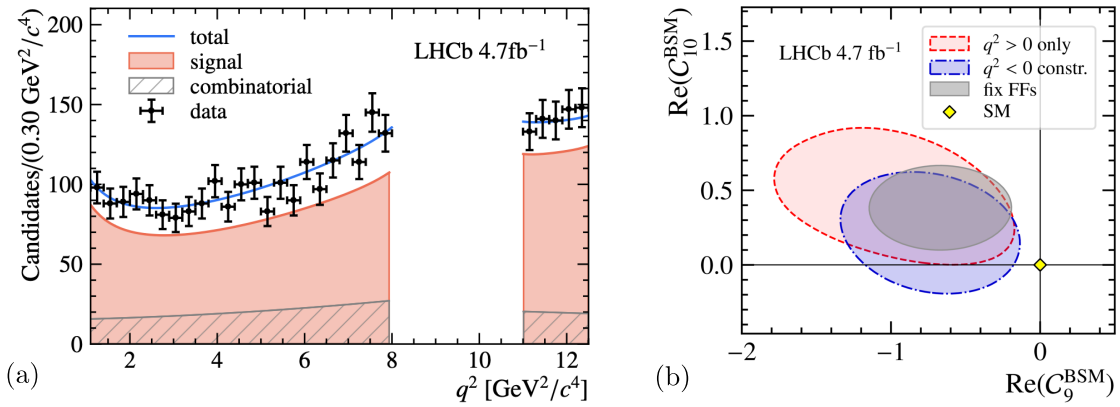


Figure 4: $B^0 \to K^{*0}\mu^+\mu^-$ long-distance contribution fits [13, 14]: (a) q^2 projections; (b) effect on $Re(C_{9,10}^{\rm NP})$.

2.2 Access to tensor states in $B \to X_s\mu^+\mu^-$

While most of the theory and experimental investigations have focused on the ground state vector states $K^*(892)$ and $\phi(1020)$ in B-meson decays, LHCb has also probed the excited $K_{0,2}^*(1430) \to K^+\pi^-$ [15] and $f_2'(1525) \to K^+K^-$ [7], including an angular moments analysis [16] for the former, to separate the S-, P- and D-wave $K\pi$ states. The results are shown in Fig. 5. The theory interpretation however will require reliable FFs for B decays to these excited states.

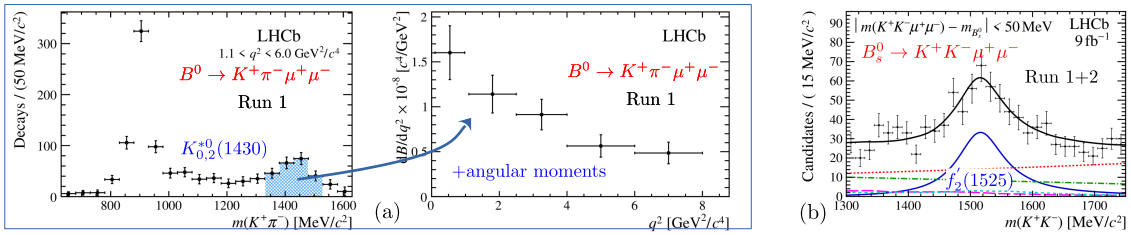


Figure 5: Tensor states in (a) $B^0 \to K^+\pi^-\mu^+\mu^-$, (b) $B^0_s \to K^+K^-\mu^+\mu^-$ decays.

2.3 Angular analysis of $B \to K\mu^+\mu^-$

For the 3-body final state $B \to K\mu^+\mu^-$, only the lepton helicity angle, θ_l, can be defined and the SM predicts an almost pure $\sin^2\theta_l$ distribution save for small effects due to the muon mass. The angular distribution is be sensitive to new scalar and tensor operators via the new terms $A_{\rm FB}$ and $F_{\rm H}$:

$$d\Gamma/d\cos\theta_l \propto \frac{3}{4}(1 - F_{\rm H})\sin^2\theta_\ell + \frac{1}{2}F_{\rm H} + A_{\rm FB}\cos\theta_\ell. \tag{4}$$

Figure 6 shows the results of the fit to Run 1 data for both $B^{+,0}$ [17], subject to the constraints $|A_{\rm FB}| \le F_{\rm H}/2$, $0 \le F_{\rm H} \le 3$, such that the rate in Eq. 4 is positive. The extracted $A_{\rm FB}$ and $F_{\rm H}$ in different q^2 bins are also consistent with SM.

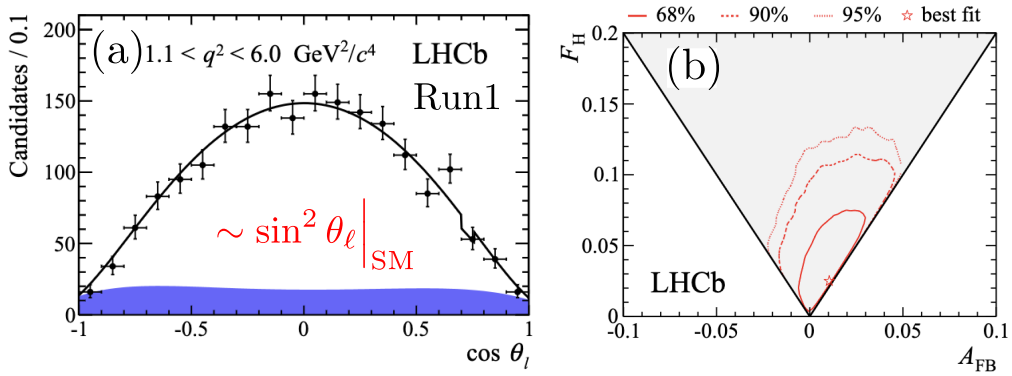


Figure 6: Charged and neutral $B \to K\mu^+\mu^-$ angular analysis [17]: (a) projection in $\cos\theta_l$ showing the expected $\sin^2\theta_l$ distribution. (b) extracted $A_{\rm FB}$ and $F_{\rm H}$ with uncertainty contours.

2.4 $\Lambda^0_{\rm b} \to \Lambda\mu^+\mu^-$ moments analysis

LHCb is unique among the B-factories to have access to all b-hadron species, including large samples of $\Lambda^0_{\rm b}$ baryons. Therefore it is possible to probe $b \to s$ penguin decays in the baryonic sector as well. For $\Lambda^0_{\rm b} \to \Lambda^0$ transition, the narrow Λ^0 state allows lattice QCD calculations [18] for the FFs. For a given value of q^2, the decay rate for polarized $\vec{\Lambda}^0_{\rm b}$ and $\vec{\Lambda}^0$ depends on five angles, $\vec{\Omega} = \{\theta, \theta_l, \phi_\ell, \theta_h, \phi_h\}$ (see Fig. 2) and is expanded in an orthonormal angular basis as

$$\frac{d^5\Gamma(q^2)}{d\vec{\Omega}} = \frac{3}{32\pi^2}\sum_{i=1}^{34} K_i(q^2) f_i(\vec{\Omega}). \tag{5}$$

The K_i moments can be related to more familiar observables such as forward-backward asymmetries in the angles: $A_{\rm FB}^{\ell} = [\frac{3}{2}K_3](\to \cos\theta_l)$, $A_{\rm FB}^{h} = [K_4 + \frac{1}{2}K_5](\to \cos\theta_h)$, $A_{\rm FB}^{\ell h} = [\frac{3}{4}K_6](\to \cos\theta_l \cos\theta_h)$. If the $\Lambda_{\rm b}^0$ is unpolarized, only the first 10 moments are non-zero [19]. Due to the long-lived nature of the Λ^0, its reconstruction in LHCb is somewhat non-trivial. At low q^2, the Run 1 analysis [8] found very few events. Therefore the analysis using Run 1 + partial Run 2 (collected between 2011-16) [20] focused on the high-q^2 region, $q^2 \in [15, 20]$ GeV^2. The results are shown in Fig. 7. The SM predictions are taken from the EOS [19] software package and show a slight tension with the data in the K_6 observable.

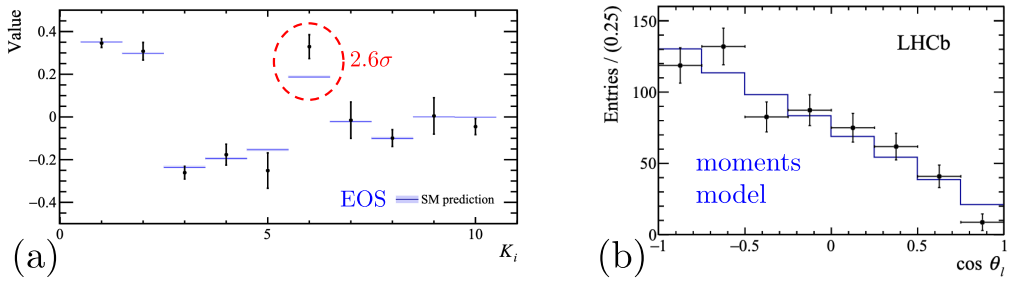


Figure 7: $\Lambda_{\rm b}^0 \to \Lambda^0\mu^+\mu^-$ angular analysis in $q^2 \in [15, 20]$ GeV^2 [20]: (a) first 10 moments from Eq. 5; (b) comparison of the background-subtracted data and angular moments model.

2.5 Differential cross-sections for $\Lambda_{\rm b}^0 \to \Lambda(1520)\mu^+\mu^-$

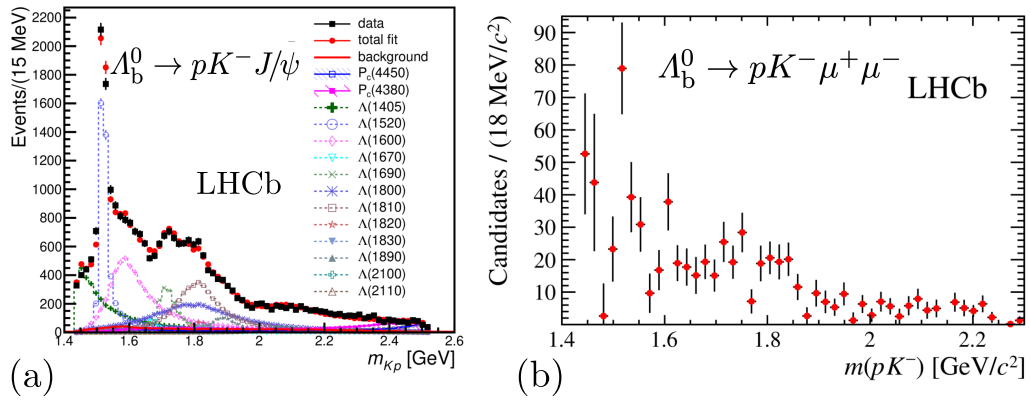


Figure 8: $m(pK^-)$ spectrum from Run 1 $\Lambda_{\rm b}^0 \to pK^-\mu^+\mu^-$ LHCb data: (a) at the J/ψ resonance [21]; (b) in the non-resonant region [22].

The first observation of $\Lambda_{\rm b}^0 \to pK^-\mu^+\mu^-$ decay using LHCb Run 1 data [22] demonstrated a rich $m(pK)$ spectrum. The comparison between the resonant and non-resonant spectra is shown in Fig. 8. Employing the full Run 1+2 datasets, LHCb has examined the $m(pK^-)$ region around the narrow $\Lambda(1520)$ resonance. Integrating over the angles and including the resonances $\Lambda(1405)(\frac{1}{2}^-)$, $\Lambda(1520)(\frac{3}{2}^-)$, $\Lambda(1600)(\frac{1}{2}^+)$ and $\Lambda(1800)(\frac{1}{2}^-)$, 1-dimensional fits are performed in $m(pK^-)$. Differential cross-sections for the $\Lambda_{\rm b}^0 \to \Lambda(1520)\mu^+\mu^-$ decay are also provided [23] that show large discrepancies with theory calculations in the low q^2 region (see Fig. 9).

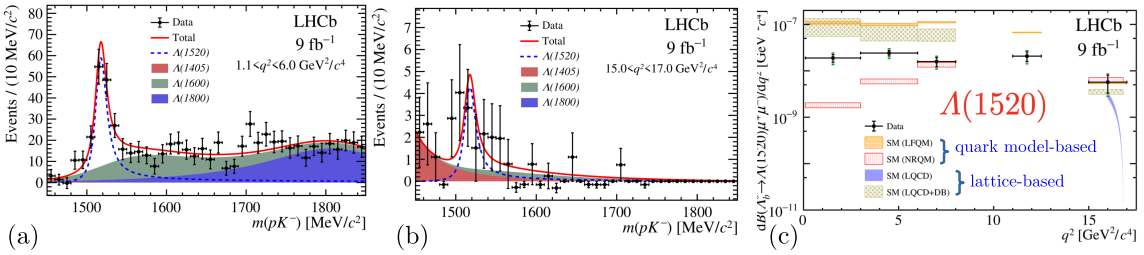


Figure 9: $\Lambda_b^0 \to \Lambda(1520)\mu^+\mu^-$ [23]: (a) low q^2, (b) high q^2, (c) $d\Gamma/dq^2$.

3. Radiative penguins

In the SM, due to the left-handed nature of the weak interaction, the photon from a $b \to s\gamma$ is almost purely left-handed; the right-handed component is suppressed ($C_7' \sim \frac{m_s}{m_b} C_7$) and is a sensitive NP probe. However, one needs angular analyses to extract the interferences. LHCb probes this in several ways: very low q^2 angular analysis of $B^0 \to K^{*0}e^+e^-$ [24] and $B_s^0 \to \phi e^+e^-$; angular analysis of $B^+ \to K^+\pi^+\pi^-\gamma$ [25],; time-dependent CP violation of $B_s^0 \to \phi\gamma$ [26] and $B^0 \to K_S^0\pi^+\pi^-\gamma$; angular analyses of $\Lambda_b^0 \to \Lambda^{0(*)}\gamma$ [27].

3.1 Angular analysis of $\Lambda_b^0 \to \Lambda^0\gamma$

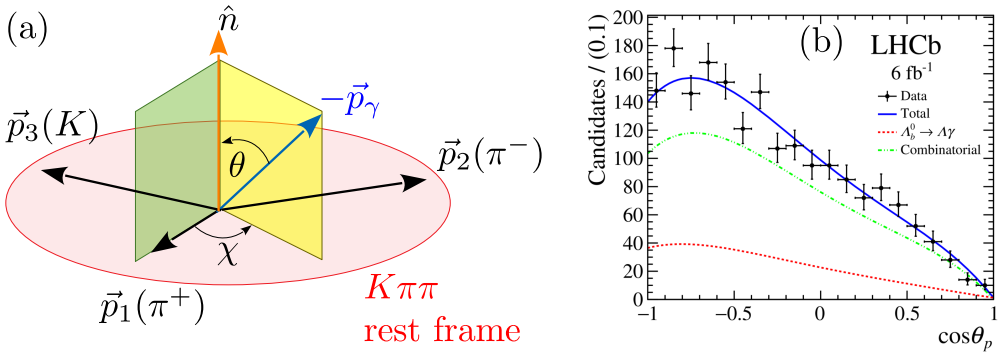


Figure 10: (a) If the hadronic system has a 3-body decay, the normal to the plane defines a preferred direction to extract the photon polarization from the up-down asymmetry in $\cos\theta$ [25]. (b) An exception is the 2-body decay of $\vec{\Lambda} \to p\pi^-$ where the measurable polarization of the Λ^0 provides the preferred direction.

To measure the photon polarization in $H_b \to H_s\vec{\gamma}$, the hadronic system must undergo a 3-body decay, so that the normal to the plane defines a preferred direction and the up-down asymmetry is proportional to the photon polarization, λ_γ. For example, this has been utilized in $B^+ \to K^+\pi^+\pi^-\gamma$ [25] as shown in Fig 10a. In such cases, due to poor knowledge of the resonant structures in the H_s system and thereby the hadronic current $\mathcal{J}_\mu^{\rm had}$, the proportionality factor remains unknown and λ_γ can still not be extracted out. An exception is the 2-body decay of $\vec{\Lambda}^0 \to p\pi^-$, where the self-analyzing nature of the Λ^0 polarization provides the preferred direction. Moreover, the differential rate is $d\Gamma/d\cos\theta_p \propto (1 - \alpha_\Lambda\lambda_\gamma\cos\theta)$ where the Λ^0 decay asymmetry parameter α_Λ is known quite precisely [28] and thereby λ_γ can be extracted from a fit to the $\cos\theta_p$ slope, as shown in Fig 10b. Experimentally, $\Lambda_b^0 \to \Lambda^0\gamma$ is challenging due to lack of a reconstructible secondary

vertex for the Λ_b^0 decay. Employing dedicated reconstruction to reject the high background, the first observation with Run 2 2016 data was reported in Ref. [29] and the polarization measurement with full Run 2 was reported in Ref. [27]. The measured value of $\alpha_\gamma = 0.82^{+0.17}_{-0..26}(\text{stat.})^{+0.04}_{-0.13}(\text{syst.})$ is compatible with the SM expectation of 1 and from global $C_7^{(\prime)}$ fits, reducing a 4-fold ambiguity in the C_7^{NP} phase to a 2-fold ambiguity.

3.2 Angular analysis of $\Lambda_b^0 \to pK^-\gamma$

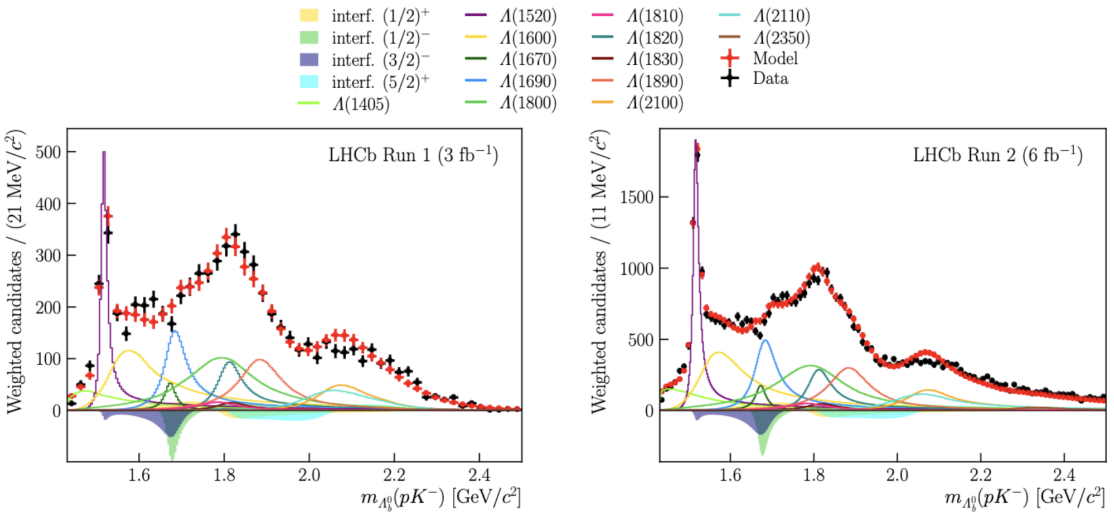


Figure 11: Fit projections for $\Lambda_b^0 \to pK^-\gamma$ in the $m(pK)$ varaiable. The shaded areas depict interferences between Λ^* states with the same spin-paiity.

As for the electroweak penguin case in Sec. 2.5, for the radiative $\Lambda_b^0 \to \Lambda^*(pK^-)\gamma$ decay, the $[pK^-]$ system is dominated by a large number of overlapping resonances. Employing the full Run1+2 data, LHCb has performed a 2-dimensional angular analysis of this mode in $\{m(pK), \cos\theta_p\}$. All well-established Λ^* states [28] are included in an isobar model with Breit-Wigner lineshapes incorporating mass-dependent widths. The only exception is the $\Lambda(1405)$ resonance where a two-pole Flatte-form is used. To reduce the large set of parameters, the maximum orbital angular momentum of the $\Lambda^* \to pK^-$ decay is taken to be $L = 3$. The result from such a *reduced* model is shown in Fig. 11. The fit fractions and interference fit fractions are reported from the amplitude analysis.

4. Summary and outlook

The large existing Run 1+2 data samples at the LHC has already allowed the hitherto (pre-LHC) "rare" $b \to s$ decays to be probed in an unprecedented fashion. This will continue into the LHCb upgrade era, with Run 3 having already commenced. Multi-dimensional angular analyses in both the electroweak and radiative penguin sectors have given rise to surprising tensions. A large and mature angular analysis effort exists at LHCb which will continue to be evolve. The thrust will be to include time-dependent CP violation [30] in more rare decay angular analyses, as well as to probe the further Cabibbo-suppressed $b \to d$ transitions [31, 32].

References

[1] W. Altmannshofer, P. Ball, A. Bharucha, A.J. Buras, D.M. Straub and M. Wick, *Symmetries and Asymmetries of $B \to K^* \mu^+ \mu^-$ Decays in the Standard Model and Beyond*, *JHEP* **01** (2009) 019 [0811.1214].

[2] B. Capdevila, A. Crivellin and J. Matias, *Review of Semileptonic B Anomalies*, *Eur. Phys. J. ST* **1** (2023) 20 [2309.01311].

[3] BaBar collaboration, *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)} l^+ l^-$*, *Phys. Rev. D* **86** (2012) 032012 [1204.3933].

[4] LHCb collaboration, *Measurement of CP-Averaged Observables in the $B^0 \to K^{*0} \mu^+ \mu^-$ Decay*, *Phys. Rev. Lett.* **125** (2020) 011802 [2003.04831].

[5] LHCb collaboration, *Angular Analysis of the $B^+ \to K^{*+} \mu^+ \mu^-$ Decay*, *Phys. Rev. Lett.* **126** (2021) 161802 [2012.13241].

[6] LHCb collaboration, *Angular analysis of the rare decay $B_s^0 \to \phi \mu^+ \mu^-$*, *JHEP* **11** (2021) 043 [2107.13428].

[7] LHCb collaboration, *Branching Fraction Measurements of the Rare $B_s^0 \to \phi \mu^+ \mu^-$ and $B_s^0 \to f_2'(1525) \mu^+ \mu^-$ - Decays*, *Phys. Rev. Lett.* **127** (2021) 151801 [2105.14007].

[8] LHCb collaboration, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda \mu^+ \mu^-$ decays*, *JHEP* **06** (2015) 115 [1503.07138].

[9] LHCb collaboration, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)} \mu^+ \mu^-$ decays*, *JHEP* **06** (2014) 133 [1403.8044].

[10] ATLAS collaboration, *Angular analysis of $B_d^0 \to K^* \mu^+ \mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **10** (2018) 047 [1805.04000].

[11] CMS collaboration, *Measurement of angular parameters from the decay $B^0 \to K^{*0} \mu^+ \mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett. B* **781** (2018) 517 [1710.02846].

[12] CMS collaboration, *Angular analysis of the decay $B^+ \to K^*(892)^+ \mu^+ \mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **04** (2021) 124 [2010.13968].

[13] LHCb collaboration, *Amplitude analysis of the $B^0 \to K^{*0} \mu^+ \mu^-$ decay*, 2312.09115.

[14] LHCb collaboration, *Determination of short- and long-distance contributions in $B^0 \to K^{*0} \mu^+ \mu^-$ decays*, 2312.09102.

[15] LHCb collaboration, *Differential branching fraction and angular moments analysis of the decay $B^0 \to K^+ \pi^- \mu^+ \mu^-$ in the $K_{0,2}^*(1430)^0$ region*, *JHEP* **12** (2016) 065 [1609.04736].

[16] B. Dey, *Angular analyses of exclusive $\overline{B} \to X \ell_1 \ell_2$ with complex helicity amplitudes*, *Phys. Rev. D* **92** (2015) 033013 [1505.02873].

[17] LHCb collaboration, *Angular analysis of charged and neutral $B \to K\mu^+\mu^-$ decays*, *JHEP* **05** (2014) 082 [1403.8045].

[18] W. Detmold and S. Meinel, *$\Lambda_b \to \Lambda\ell^+\ell^-$ form factors, differential branching fraction, and angular observables from lattice QCD with relativistic b quarks*, *Phys. Rev. D* **93** (2016) 074501 [1602.01399].

[19] D. van Dyk et al., *Eos – a hep program for flavor observables*, 2016. 10.5281/zenodo.886055.

[20] LHCb collaboration, *Angular moments of the decay $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ at low hadronic recoil*, *JHEP* **09** (2018) 146 [1808.00264].

[21] LHCb collaboration, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, *Phys. Rev. Lett.* **115** (2015) 072001 [1507.03414].

[22] LHCb collaboration, *Observation of the decay $\Lambda_b^0 \to pK^-\mu^+\mu^-$ and a search for CP violation*, *JHEP* **06** (2017) 108 [1703.00256].

[23] LHCb collaboration, *Measurement of the $\Lambda_b^0 \to \Lambda(1520)\mu^+\mu^-$ differential branching fraction*, *Phys. Rev. Lett.* **131** (2023) 151801 [2302.08262].

[24] LHCb collaboration, *Strong constraints on the $b \to s\gamma$ photon polarisation from $B^0 \to K^{*0}e^+e^-$ decays*, *JHEP* **12** (2020) 081 [2010.06011].

[25] LHCb collaboration, *Observation of Photon Polarization in the $b \to s\gamma$ Transition*, *Phys. Rev. Lett.* **112** (2014) 161801 [1402.6852].

[26] LHCb collaboration, *First experimental study of photon polarization in radiative B_s^0 decays*, *Phys. Rev. Lett.* **118** (2017) 021801 [1609.02032].

[27] LHCb collaboration, *Measurement of the photon polarization in $\Lambda_b^0 \to \Lambda\,\gamma$ decays*, *Phys. Rev. D* **105** (2022) L051104 [2111.10194].

[28] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[29] LHCb collaboration, *First Observation of the Radiative Decay $\Lambda_b^0 \to \Lambda\gamma$*, *Phys. Rev. Lett.* **123** (2019) 031801 [1904.06697].

[30] LHCb collaboration, *Measurement of CP-violating and mixing-induced observables in $B_s^0 \to \phi\gamma$ decays*, *Phys. Rev. Lett.* **123** (2019) 081802 [1905.06284].

[31] LHCb collaboration, *Observation of the suppressed decay $\Lambda_b^0 \to p\pi^-\mu^+\mu^-$*, *JHEP* **04** (2017) 029 [1701.08705].

[32] LHCb collaboration, *First measurement of the differential branching fraction and CP asymmetry of the $B^\pm \to \pi^\pm\mu^+\mu^-$ decay*, *JHEP* **10** (2015) 034 [1509.00414].

arXiv:2403.01040v1 [hep-th] 2 Mar 2024

Bulk-local dS_3 holography: the Matter with $T\bar{T} + \Lambda_2$

Gauri Batra,[a] **G. Bruno De Luca,**[a] **Eva Silverstein,**[a] **Gonzalo Torroba,**[b] **Sungyeon Yang**[a]

[a] *Stanford Institute for Theoretical Physics, 382 Via Pueblo, Stanford, CA 94305*
[b] *Centro Atómico Bariloche and CONICET, Bariloche, RN, Argentina*

Abstract:

We propose an algorithm which builds a concrete dual for large-radius 3d de Sitter with a timelike York boundary for both gravity and bulk effective fields. This generalizes the solvable $T\bar{T} + \Lambda_2$ deformation, whose finite real spectrum accounts for the refined Gibbons-Hawking entropy as a microstate count while reproducing the radial static patch geometry. The required generalization to produce approximately local boundary conditions for bulk quantum fields requires a scheme for defining double-trace operators dual to deformed boundary conditions to realize the finite timelike boundary, valid at finite N. By starting with a small stint of a pure $T\bar{T}$ trajectory, the theory becomes finite, enabling well-defined subtractions to define the double-trace deformation so as to match the large-N prescription of Hartman, Kruthoff, Shaghoulian, and Tajdini to good approximation. We incorporate the matter effecting an uplift from negative to positive cosmological constant, and analyze the effect of matter on the energy spectrum of the theory arising from time-dependent bulk excitations. This validates the cosmic horizon dS_3 microstate count for large-radius dS_3 holography, embedding $T\bar{T}+\Lambda_2$ concretely into a larger theory consistent with bulk locality for matter fields. We comment briefly on potential upgrades to four dimensions and other future directions.

Contents

1 Introduction and high-level summary **1**

2 Comments on the gravitational sector **6**
- 2.1 The dressed classical action for $T\bar{T} + \Lambda_2$ and role of 'boundary gravitons' (extrinsic curvature fluctuations) 6
- 2.2 A single theory captures all the universal energy levels 8

3 Defining the bulk-local deformation as a finite system: type I operator algebra **9**
- 3.1 Large c construction 10
- 3.2 Finite system at finite c 14
- 3.3 Implication for operator algebra 16

4 Uplift and energy continuity **16**

5 Summary of Boundary Theory Prescription **19**

6 Dressed energies capturing time-dependent local bulk matter **22**
- 6.1 $2d$ boundary side derivation of dressed energies in simple cases 22
- 6.2 Bulk counterpart: matter waves and bulk gravity with fixed boundary cylinder 23
- 6.3 First law of thermodynamics for the Cosmic Horizon patch 28

7 $T^2 + J^2 + \Lambda_2$ and a charged static patch horizon **30**

8 Summary and Discussion **32**

1 Introduction and high-level summary

Recently the de Sitter entropy in three dimensional gravity [1, 2],

$$S \simeq \mathcal{A}/4G_N - 3\log(\mathcal{A}/4G_N) \tag{1.1}$$

along with the radial geometry of the static patch has been captured by an explicit boundary dual defined by the solvable $T\bar{T} + \Lambda_2$ deformation [3–5] of a holographic seed CFT. In this formulation, there is a clear explanation for the finiteness of the de Sitter entropy: it is a direct consequence of the finite real spectrum of the $T\bar{T}$-type theory [6–8]. This construction includes a timelike non-asymptotic non-gravitational boundary on the gravity side, in whose presence the thermodynamics matches that of ordinary quantum systems [9, 10].[1] The existence of such boundaries in quantum gravity with

[1] Various subtleties with arbitrary Dirichlet boundary conditions arise in $4d$ General Relativity (see e.g. [11–17]). In this work we will stick to 3 external dimensions in the special case of a cylinder boundary within which we solve the matter plus gravity field equations to ensure a good initial boundary value problem. In Lorentzian signature in 3

matter is a priori an open question, whose resolution may involve direct gravity-side analysis in string theory (e.g. along the lines of [25, 26])[2], or may be established indirectly by formulating a boundary dual matching the low energy bulk theory in a bounded spacetime.

The boundary theory [3] independently computes universal features such as the horizon entropy and geometry, agreeing with that predicted by bulk semiclassical gravity. But it does not accurately capture importand model-dependent details. These include the fine structure of the band of energies corresponding to horizon entropy, the details of the uplift from AdS to dS spacetime, and local bulk matter excitations contained within the bounded patch. These details are subleading in the calculation of the total state count (entropy) and the large-radius geometry, but are important to obtain in a complete theory of de Sitter quantum gravity. It is the goal of this paper to fill in this gap and present a concrete proposal for constructing a boundary theory formulating the large-radius 3d de Sitter static patch with timelike boundary.[3] After developing this theory in the main paper, we will comment on potential generalizations to 4d in the discussion section.

To begin, let us summarize the structure of the universal part of the theory, captured in [3]. A holographic seed CFT of central charge c has two universal bands of energies: the vacuum and a band of energies within order $c^0 = 1$ of $\Delta = c/6$ with a state count obeying the Cardy formula [44, 45] which matches the Bekenstein-Hawking entropy for the black hole of radius equal to the AdS curvature radius. Deforming this theory by the $T\bar{T} + \Lambda_2$ deformation in the way defined in [3], one solves for the spectrum all along the deformation via the method discovered by [6]. For these universal bands, the dressed energy [6] of the solvable model as a function of the deformation parameter (see (4.1) for its explicit form) matches the Brown-York quasilocal energy of the corresponding bounded patches of de Sitter as a function of the boundary size.[4] This precise map includes the contributions of the extrinsic curvature fluctuations known as 'boundary gravitons', as worked out in detail for Dirichlet-bounded AdS_3 in [20].[5] Along the integrable deformation, one can follow the energy levels individually, with the count of states in the real dressed spectrum matching the horizon entropy. The solution of the theory via the method discovered by [6] is valid at finite central charge c.

This universal and solvable deformation captures the gravitational sector of the theory, but not the model-dependent details of bulk matter, including the fine structure in the dressed $\Delta \simeq c/6$ energy

space-time dimensions, there is no pileup of states with large extrinsic curvature (cf [15, 18, 19]), as can be seen by the finiteness of the real spectrum related to the square root appearing in the dressed Hamiltonian (derived explicitly for Dirichlet boundaries in [20, 21]). We comment on this in §2 and on potential generalizations to 4d with conformal boundary conditions [12, 15, 22–24] at the end.

[2]We thank A. Ahmadain, R. Khan, and A. Wall for discussions of this point

[3]This accords with recent work exhibiting simplifications of bounded patch thermodynamics in terms of the quasilocal energy [9, 10] and with many works over the years such as [22, 27–35] analyzing static patch and dS/dS patch holography, including the role of timelike boundaries or observers and the maximal mixing of the correspondinig microstates in global de Sitter [36–40] as well as probes of complexity such as [41]. Moreover, it may connect to approaches with the dual formulated on a spacelike boundary which also admit a formulation in terms of a T^2 deformation [42]. See e.g. [43] for a review of some recent developments.

[4]We should stress that the vacuum and $\Delta \simeq c/6$ energy bands are captured in a single theory (whereas in parts of [3], they were studied separately using different deformation trajectories). They arise together in the same theory once we retain both signs of the square root in the dressed energy formula (4.1) and hence in the Hamiltonian.

[5]In particular, they reproduce on the gravity side a square root formula for energies, matching the $T\bar{T}$ deformation. This cuts off the real spectrum at a finite energy level, giving a finite quantum system which does not contain divergent extrinsic curvature fluctuations. The authors [20] report that this reduces explicitly to [18] in $2d$ (see [19] for another interesting prescription for UV completion of the $2d$ problem). It would be interesting to generalize the analysis of [20] to the $T\bar{T} + \Lambda_2$ phase, but the same general map between gravity equations and the deformation equations indicate a square root cutoff on the spectrum also in this case. We discuss this further in §2.

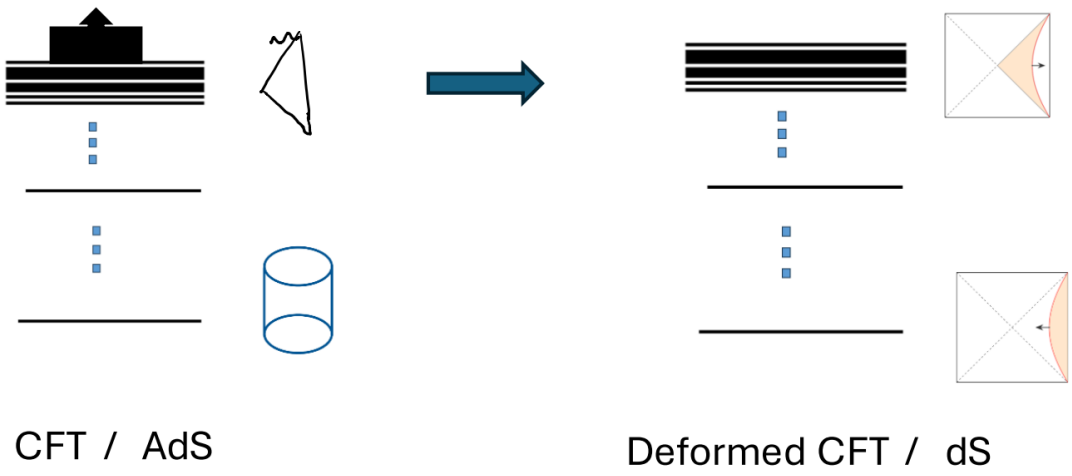


Figure 1: Schematic of the seed spectrum dual to the states of quantum gravity in AdS_3 deforming into the spectrum of the finite quantum theory dual to a bounded patch of dS. Starting from a holographic conformal field theory and deforming according to a specific generalization of the $T\bar{T}$ deformation brings the system to a finite quantum system matching the entropy and quasilocal energy (hence radial geometry) of a bounded patch of de Sitter. Previous work [3] obtained this for the universal gravitational sector which contains the deformed $\Delta \simeq c/6$ band of energies and the vacuum state. These are captured with the solvable $T\bar{T}+\Lambda_2$ deformation whose finite real spectrum explains the finite de Sitter entropy. The present work generalizes the deformation to capture bulk matter, defining a finite quantum system whose leading approximation at large c matches general relativity plus matter with local Dirichlet boundary conditions at the timelike boundary. In this system without fluctuating boundary gravity, the thermodynamic relations take the standard form for an ordinary quantum system [9, 10]. In addition to the depicted states, the theory admits domain wall configurations between AdS and dS.

band. Aside from any energy level which precisely sits at $\Delta = c/6$, states in this band of energies are not exactly continuous in the pure $T\bar{T}+\Lambda_2$ deformation[6]. These discontinuities do not affect the solvability of the theory, as we can follow these subleading jumps in the energy and obtain an explicit finite real spectrum of energies E_n which simply defines the Hamiltonian $H = \sum_{\text{levels } n} E_n |n\rangle\langle n|$ everywhere along the deformation. Happily the universal sector contains the most entropic energy band, whose fine structure is not needed for the state count (entropy) which matches the horizon area of black holes in AdS and the bare cosmic horizon in de Sitter.

The solvable theory captures the first two terms in the generalized entropy

$$\mathcal{S}_{\text{gen}} = \mathcal{A}/4G_N - 3\log(\mathcal{A}/4G_N) + S_{\text{fields}}; \tag{1.2}$$

in terms of the horizon area $\mathcal{A}$, along with the appropriate notion of energy defined in the presence of a timelike boundary by Brown and York [46]. It contains subleading contributions from bulk matter, which also imprints order G_N^0 contributions to the area term (via order G_N back reaction on the area) as well as to the Brown-York energy. With a timelike boundary with frozen gravity, general relativity

[6]This occurs at the step where Λ_2 enters and the branch choice flips in the square root appearing in (4.1). In the present work, full continuity will be restored via matter effects that capture the uplift from AdS to dS.

plus effective quantum field theory (GR + EFT) obeys the standard first law of thermodynamics

$$\delta E_{\text{Brown-York}} = \delta E_{\text{bulk-matter}} + T\delta\mathcal{S} \tag{1.3}$$

discovered in [9, 10]. This thermodynamics emerges from the statistical mechanics of the boundary theory we develop in this work, adding matter to the solvable $T\bar{T} + \Lambda_2$ deformation to produce an ordinary quantum system.

To capture the details of bulk matter fields, relevant both for fine structure of the dominant $\Delta \simeq c/6$ energy band and for less entropic energy bands of the system, we must generalize the formulation of the theory. The reason is well known [47][48]: the solvable $T\bar{T}(+\Lambda_2)$ theory is consistent with a bulk cutoff [49] ([3–5]) for gravity, but does not endow bulk quantum fields with local boundary conditions at the Dirichlet wall [47, 48]. One way to see this intuitively is that multitrace deformations [50, 51] change the boundary conditions of bulk fields [52–54]; the pure $T\bar{T} + \Lambda_2$ deformation does this only for gravity. A generalization to other bulk fields dual to operators $\mathcal{O}$ would require a generalization of the schematic form

$$\partial_\lambda S \sim \int T\bar{T} + \mathcal{O}\mathcal{O} + \Lambda_2 \tag{1.4}$$

along with an explicit uplift from AdS to dS to account for the transition to nonzero Λ_2. This has an immediate difficulty: adding such (generically irrelevant) composite operators leads to ambiguities, suggesting that such a construction would have meaningless output wholly dependent on input assumptions. At infinite c, composite operators factorize, enabling a clear treatment of (1.4) reverse-engineered from the gravity side to ensure local Dirichlet boundary conditions for all fields at the boundary (along the lines of Section 3.3 of [55]). But that as it stands leaves open the question of the existence of a finite-c completion.

In this work, we resolve this, constructing a finite-c completion and incorporating local bulk matter into the duality [3]. To do so, we formulate a finite system whose large central charge ($c \gg 1$) approximation reproduces the deformation required by the large-c factorization results and their implications for the dressed energies. The finiteness of our system is obtained by regulating our construction with an arbitrarily small initial pure $T\bar{T}$ deformation, giving a finite real spectrum. Starting from this finite quantum system, we can define the operators required in the deformation (1.4) by implementing explicit subtractions of singular terms, so as to match the large-c definition of the operators up to small corrections. The theory is defined by this algorithm: at each step of the deformation we update the theory by recomputing $T, \mathcal{O}$ and add the appropriate composite operators (1.4) with the subtractions. In order to capture bulk processes involving not only low energy fields, but also Kaluza-Klein modes and other high energy excitations one needs operators corresponding to each. In our full trajectory, the internal space is never larger than the bounding cylinder, with scale separation in the de Sitter phase arising from the standard 3-term structure of moduli stabilization [31, 56]. On the other hand, to capture only processes accessible at low energies in the bulk, the additional specifications are not needed. In either case, the theory thus defined serves to define the quantum gravity theory, which contains corrections beyond leading order in large c.

GR + EFT is not always valid; we match its predictions when it applies. For example, during the uplift step of the deformation, the low energy GR and EFT description breaks down for the states near $\Delta = c/6$: in the gravity-side description, the boundary skirts the horizon, leading to strong

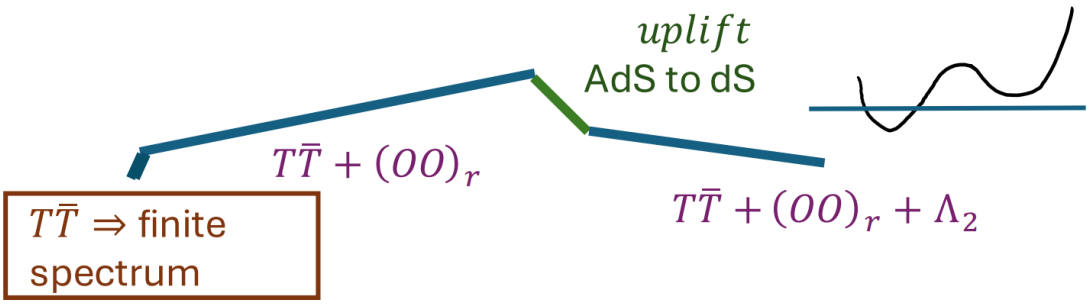


Figure 2: A schematic of our algorithm to define a holographic dual for the dS_3 static patch with timelike non-gravitational boundary. The explicit cut off on the real spectrum from a small initial $T\bar{T}$ deformation enables one to prescribe a concrete subtraction procedure to define the operators $(\mathcal{O}\mathcal{O})_r$ required to endow bulk matter fields with approximately local boundary conditions.

fluctuations.[7] We can view this as a feature rather than a bug: the theory is less constrained at this step. For states well below $\Delta = c/6$, the uplift occurs within the regime of validity of GR + EFT and we define it consistently with that simply by deforming the boundary conditions of an internal scalar to interpolate from the value it takes in an AdS vacuum to a value it takes in a dS vacuum.[8] This restores continuity of the model-dependent low-lying energy levels far from $\Delta \simeq c/6$ along the trajectory.

This subtraction criterion does not specify a unique theory; different versions of the theory which all agree with low energy general relativity and effective field theory in its regime of applicability differ in a way that is subleading at large c. It may be that additional criteria are desirable, beyond simple matching the low energy theory obtained by deforming AdS/CFT dual pairs. We note, however, that the landscape of string vacua leads to a vast multiplicity of internal configurations consistent with similar low-energy physics, including infinite sequences of tractable models [57]. We may view the non-uniqueness of our construction in a similar vein, though the origin of the multiplicity is different.

See Figure 2 and Algorithm 1 for a summary of the defining algorithm for our theory. The prescription agrees with GR + EFT in where they apply (including for the description of the dressed $\Delta \simeq c/6$ states far from the uplift step). The strong quantum gravity effects arising from the details of the boundary, at the uplift for the dressed $\Delta \simeq c/6$ states, and for high energy processes in the bulk are defined by the non-gravitational boundary quantum mechanics theory. This construction is quite formal and rather indirect; it is defined by an algorithmic update rule and does not include a full resummation of the Hamiltonian.[9] But it can be used to address key questions in de Sitter, in particular verifying the microstate-count interpretation of the refined [2] Gibbons-Hawking entropy [1] derived from the solvable $T\bar{T} + \Lambda_2$ core of our more general model.

To use our boundary theory to illustrate the emergence of bulk locality, we generalize the calculation of dressed energy levels to capture time-dependent matter excitations $\langle \mathcal{O}(t) \rangle \sim \sin \omega t$ in

[7] This was studied in [26] for a Neumann boundary condition for compactification degrees of freedom in string/M theory.

[8] Explicit compactifications uplifting AdS/CFT to dS are available in 3 and 4 external dimensions [31, 56].

[9] See e.g. [34, 58]

the leading nontrivial order. In general, this would require generalizing the steps of Smirnov and Zamolodchikov's derivation of the differential equation for dressed energies [6] to incorporate general time dependent excitations. We can more simply analyze situations where the expectation value of the pressure in an energy eigenstate takes a simple form. As we will see, at least in that case this independent boundary-side calculation can be done trivially and matches the physics of the gravity side, including bulk matter with local boundary conditions.

Our analysis of the gravity-side solutions in general illustrate basic differences with respect to the pure gravitational case. In particular it is important to note that the deformation cannot be described as moving a boundary through a pre-existing bulk solution. Rather, the gravity-side counterpart to the dressed energy spectrum arises by keeping track of the solutions as a function of the bounding cylinder size L_c; for each L_c there is a new space of solutions. In addition, the solution space of GR + EFT that we derive below will reveal another significant feature: the effect of matter excitations on the dressed energy formula (4.1) appears multiplying the square root factor in that expression, thus not affecting the reality of the finite spectrum.

Another outcome is the determination that cosmic horizon patch holography has a type I Von Neumann operator algebra. One can think of the boundary here as somewhat analogous to the observer in [39], where the Newton constant was taken to zero at the end, engendering a type II algebra. In our finite quantum system matching gravity-side physics at finite G_N, we obtain type I. We note that the effect of bulk matter on the generalized entropy (1.2) is subleading to the $A/4G_N$ term, contained in model-dependent contributions to the logarithmic term (as computed in [2] but with the Dirichlet boundary conditions on the wall). As stressed above, the $T\bar{T} + \Lambda_2$ deformation captures the area term plus the universal part of the logarithmic correction [3]; in this work we see this survives with matter included in a finite quantum system.

Altogether, this brute force but well-defined prescription provides a holographic dual for the large-radius dS_3 static patch with timelike boundary, with the correct emergent static patch geometry, entropy, and approximately local bulk matter dynamics.[10]

2 Comments on the gravitational sector

Before introducing our framework for local bulk matter, we would like to clarify some aspects of the gravitational sector covered by the solvable $T\bar{T} + \Lambda_2$ deformation in [3, 4].

2.1 The dressed classical action for $T\bar{T} + \Lambda_2$ and role of 'boundary gravitons' (extrinsic curvature fluctuations)

The key relation underlying the duality at the pure gravity level is the following [4, 49]. The $3d$ bulk Einstein equations in the presence of a Dirichlet boundary imply the defining equation of the $T\bar{T} + \Lambda_2$ deformation

$$T_i^i = 4\pi\lambda T\bar{T} - \frac{1}{\pi\lambda}(1-\eta) \Rightarrow \partial_\lambda S = -\int [2\pi T\bar{T} - \frac{1}{2\pi\lambda^2}(1-\eta)] \tag{2.1}$$

[10] It would be interesting to connect it to recent proposals for a concrete dual to dimensionally reduced dS_3 via the SYK model [34, 58].

Algorithm 1 $T\bar{T} + \mathcal{O}O + uplift + \Lambda_2$. Δy: step-size in theory space (default: infinitesimal).

Require: A holographic seed CFT of central charge c living on a cylinder with size L, with its generating functional of correlation functions $\mathcal{W}$ and set of energy eigenstates $\{\varphi_n^{(0)}\}$ with energies $\{E_n^{(0)}\}$, and large-N EFT + GR correlators.
Require: A scalar field Φ with a potential V interpolating from the holographic AdS vacuum at Φ_A to a dS minimum at Φ_S.
Require: y_T (explicit regulator).
$y \leftarrow 0$ ▷ Initialize dimensionless deformation parameter
$y_0 \leftarrow 3/c\pi^2$
function TruncateSpectrum($\{E_n\}, \{\varphi_n\}$)
  for $E_i \in \{E_n\}$ **do**
    if E_i `is complex` **then**
      `remove` (E_i, φ_i) `from` $(\{E_n\}, \{\varphi_n\})$
    end if
  end for
  return $\{E_n\}, \{\varphi_n\}$
end function
$\Phi \leftarrow \Phi_A$
while $y < y_T$ **do**
  $\{E_n\}, \{\varphi_n\} \leftarrow$ TruncateSpectrum($\{E_n\}, \{\varphi_n\}$)
  $T_{ij} \leftarrow \frac{-2}{\sqrt{-\gamma}} \frac{\delta \mathcal{W}}{\delta \gamma^{ij}}$
  $\mathcal{W} \leftarrow \mathcal{W} - L^2 \frac{\pi}{4} \Delta y \int d^2x \sqrt{-\gamma} (T^{ij} T_{ij} - (T_i^i)^2)$
  $y \leftarrow y + \Delta y$
end while
while $y > y_T$ **do**
  if $y = y_0$ **then**
    $\Phi \leftarrow \Phi_S$ ▷ Do the uplift to dS (at $y \sim y_0 \pm \Delta y$ in the discrete version)
  end if
  $\{E_n\}, \{\varphi_n\} \leftarrow$ TruncateSpectrum($\{E_n\}, \{\varphi_n\}$)
  $T_{ij} \leftarrow \frac{-2}{\sqrt{-\gamma}} \frac{\delta \mathcal{W}}{\delta \gamma^{ij}}$
  $\mathcal{O} \leftarrow -\frac{\delta \mathcal{W}}{\delta J}$
  $\langle universal | (\mathcal{O}\mathcal{O})_r \mathcal{O}_1 \dots \mathcal{O}_m | universal \rangle \leftarrow \langle universal | (\mathcal{O}\mathcal{O})_r \mathcal{O}_1 \dots \mathcal{O}_m | universal \rangle_{EFT+GR}$
  $\mathcal{W} \leftarrow \mathcal{W} - L^2 \frac{\pi}{4} \Delta y \int d^2x \sqrt{-\gamma} (T^{ij} T_{ij} - (T_i^i)^2 - \frac{1}{L^4 y^2} (1 + \mathtt{sign}(V(\Phi)) + (\mathcal{O}\mathcal{O})_r)$
  $y \leftarrow y + \Delta y$
end while

where in the gravity-side variables, T is the Brown-York stress tensor [46], and where $T\bar{T} = -\frac{1}{4}\det(T)$. This indicates that pure gravitational configurations in the Dirichlet problem in $3d$ are governed by the $T\bar{T} + \Lambda_2$ deformed theory.

Among the $T\bar{T} + \Lambda_2$ dressed states of the CFT are boundary graviton excitations. Since these figure into some of the analyses of subtleties with Dirichlet boundaries [11, 15] (see also [18, 19] in 2d), it is worth spelling out how they fit in.[11] The boundary gravitons are extrinsic curvature fluctuations arising from deformations of the embedding of the boundary, consistently with the Dirichlet condition that it have fixed intrinsic geometry. This scalar degree of freedom ϕ_g was studied at the linearized

[11] We thank H. Maxfield and D. Stanford for discussions of this point.

level in bounded patches of AdS_3 e.g. in [11] and nonlinearly in [20, 21]. The latter works provide extensive evidence directly on the gravity side that this degree of freedom is governed classically by the Born-Infeld Lagrangian, which is indeed the $T\bar{T}$ dressed theory of a free boson. Generalizing the calculation of the dressed classical action for a free boson to include the $T\bar{T} + \Lambda_2$ deformation, e.g. using the elegant prescription in [59], yields

$$\mathcal{L} = -\frac{1}{\pi\lambda}\left(1 - \sqrt{\eta + \pi\lambda\partial_\mu\phi_g\partial^\mu\phi_g}\right) \tag{2.2}$$

which satisfies the deformation equation (2.1). Moreover, [20] report that the theory they derive reduces explicitly to a version of cut off 2d JT gravity [18] (see [19] for another interesting prescription for UV completion of $2d$ JT gravity with finite renormalized boundary).

Linearized boundary graviton dynamics [11] arises from expanding this action around a suitable background configuration [60]. In the full nonlinear quantum system, since the real spectrum of the $T\bar{T} + \Lambda_2$ deformed theory is finite, and matches to the Dirichlet problem in $3d$ gravity, there is no buildup of arbitrarily strong extrinsic curvature fluctuations.

2.2 A single theory captures all the universal energy levels

Universal states of bounded de Sitter quantum gravity include the cosmic horizon patch and the pole patch, as depicted on the right side of figure 1. In this subsection, we would like to clarify that the version of the $T\bar{T} + \Lambda_2$ trajectory defined in [3] – which we briefly review here – can capture in one theory both the quantum state corresponding to the cosmic horizon patch and the pole patch.

This trajectory is defined by (2.1), proceeding in two segments. First is evolution by pure $T\bar{T}$ with $\eta = 1$ until a value $y = y_0$, after which one turns on the Λ_2 part with $\eta = -1$. (In the present work, we will resolve this step with a matter sector modeling the uplift from AdS to dS.) With $y_0 = 3/c\pi^2$, the resulting dressed energy formula is [3]

$$EL = \mathcal{E} = \frac{1}{\pi y}\left(1 \mp \sqrt{\eta + \frac{y}{y_0}(1-\eta) - 4\pi^2 y\left(\Delta - \frac{c}{12}\right) + 4\pi^4 y^2 J^2}\right) , \tag{2.3}$$

where Δ is the dimension corresponding to each seed CFT energy level that is being dressed.

We have indicated the possibility of states with both signs of the square root here. Indeed, the quasilocal energies of the cosmic horizon patch and the pole patch fit the upper and lower signs of (2.3) with $\Delta = c/6$. We may construct the Hamiltonian $H = \sum_n E_n|n\rangle\langle n|$ keeping states with both signs of the square root. This fits with the gravity side: fixing the proper geometry of the boundary according to the Dirichlet boundary condition does not constrain the extrinsic curvature K, which may take either sign. As explained in [3] and reviewed below, this corresponds to the two branches of the square root in (2.3).

This does not imply that there are equal numbers of states in the two sectors corresponding to the two branches of the square root. In the gravity description, one determines nonsingular solutions obeying the boundary conditions, whose existence and properties can depend on the sign of K. In the dual deformed-CFT description, in constructing the Hamiltonian at each step in the trajectory, the number of states with each sign of the square root can be different. At $y = y_0 = 3/c\pi^2$, where we turn

on Λ_2, this works as follows. Constructing the pole patch from the $-$ sign in (2.3), we include one such state,[12] while for the cosmic horizon patch we keep all of the $\Delta \simeq c/6$ states comprising the horizon microstates. For the latter states, the gravity description at $y = y_0$ consists of a boundary skirting the horizon, for which the AdS BTZ black hole and the dS cosmic horizon are indistinguishable.[13]

So far we have obtained both the cosmic horizon (CH) patch and the pole patch from the $\Delta \simeq c/6$ part of the energy spectrum. We must also address the fate of the CFT vacuum state ($\Delta = 0$) along our trajectory. This yields a state with a gravity-side description involving a domain wall: as y approaches $y_0 = 3/c\pi^2$ where the uplift will occur, the Dirichlet wall on the gravity side still bounds a large region of AdS, distinguishable from dS. Continuing (increasing y) after adding Λ_2 pulls out the boundary, changing the quasilocal energy according to the radial evolution in dS.

3 Defining the bulk-local deformation as a finite system: type I operator algebra

In this section we will build up the definition of our deformed theory. First we will lay it out at large c, refining the treatment in section 3.3 of [55]. Then in §3.2 we will explain how to adjust the construction to obtain a well defined system at finite c.

Our construction will generalize the duality [3] between solvable models and pure 3d gravity quantities. The precise dictionary between bulk and boundary theory parameters derived in [4, 49] and applied in [3] will still apply in our theory.[14] The seed CFT has large but finite central charge c. The deformed CFT lives on a cylinder of fixed spatial size L, evolving along a trajectory parameterized by the dimensionful parameter λ. It is useful to define a dimensionless deformation parameter

$$y = \frac{\lambda}{L^2} = \frac{8G_N\ell}{L_c^2} \tag{3.1}$$

where the last expression uses the dictionary to the gravity side

$$c = \frac{3\ell}{2G_N}, \quad \lambda = 8G_N\ell\frac{L^2}{L_c^2}\,. \tag{3.2}$$

Here ℓ is the curvature radius of the AdS or dS that we start and end with along the trajectory we will define.

Along our trajectory, L, c and G_N, ℓ are fixed on the boundary theory and gravity side respectively. On the boundary theory side λ evolves along the trajectory, and on the gravity side the proper size L_c of the bounding cylinder changes along the trajectory. These agree on the evolution of y (3.1).

[12] The small entropy of the pole patch, whose boundary excludes the cosmic horizon, is independently checked by a different $T\bar{T} + \Lambda_2$ trajectory with $y_0 \to \infty$ [3, 4].

[13] An argument for this persisting in string theory, with its internal degrees of freedom from compactification, appears in [26]

[14] The construction [3] in turn realized and explained the 'numerology' in [61].

3.1 Large c construction

In this subsection, we will derive the necessary generalization of $T\bar{T} + \Lambda_2$ to incorporate local bulk matter, at the large-c level. This by itself is not a complete theory, but it will be a good approximation to the finite-c theory we develop in the subsequent sections.

As described in [62], in the large c limit we fix λc. At a given step in the deformation, we will add contributions including the double trace terms

$$\delta\mathcal{L} = 2\pi\delta\lambda T\bar{T} + f_O(c\lambda)\mathcal{O}\mathcal{O}\,. \tag{3.3}$$

The first coefficient here is of order $1/c$, matching the prescription for multitrace deformations for matrix-like theories introduced in [50]; in our bulk $3d$ context an example of such a seed theory is the CFT obtained from the low energy limit of the D1-D5 system. In general we treat the operators $\mathcal{O}$ dual to bulk matter fields according to their particular representations (which can include for example composites of singlets and fundamentals in addition to traces of products of adjoints and bifundamentals). We will consider to be specific bulk matter fields whose internal dynamics is independent of G_N and will specify below our convention for the definition of these operators.

Following [4, 48, 49, 55] we work from the Einstein equations, including matter stress-energy. The type of matter that is most relevant is (i) bulk gauge field sectors, which are naturally light, and (ii) a sector effecting the uplift from negative to positive bulk cosmological constant. Since in the three dimensional bulk we can dualize vector fields to scalars, we will write everything in terms of scalars.

Let us start by deriving a formula for the deformation including a bulk scalar field with stress-energy tensor $\mathcal{T}$. We will work with a patch of spacetime bounded by a cylinder of proper size L_c. We can coordinatize this with metric

$$ds^2 = dw^2 + g_{ij}(w,x)dx^i dx^j \tag{3.4}$$

where we impose boundary conditions on g_{ij} such that there is a boundary cylinder of size L_c. Without loss of generality we denote its radial position as $w = w_c$. The three-dimensional bulk action is

$$S = \int_{\mathcal{M}} d^3x\,\sqrt{-g}\left[\frac{1}{16\pi G_N}R^{(3)} - \frac{1}{2}(\partial\Phi)^2 - V(\Phi)\right] + \frac{1}{8\pi G_N}\int_{\partial\mathcal{M}} d^2x\,\sqrt{-g}\left(K - \frac{B_{CT}(\Phi)}{\ell}\right)\,, \tag{3.5}$$

with metric signature $(-++)$. For a pure AdS solution, $V = -2/\ell^2$ and $B_{CT} = 1$, but we allow for a more general matter potential. The boundary term proportional to the trace of the extrinsic curvature is needed in order to have a well-defined Dirichlet variational principle; we have also allowed for a boundary conterterm action determined by the function $B_{CT}(\Phi)$.[15]

We denote the bulk matter stress tensor by

$$\mathcal{T}_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}\,, \tag{3.6}$$

[15] There is much freedom in the counterterm action, which must cancel all divergences at the beginning of the trajectory. It is also possible to add counterterms containing derivatives, such as $R^{(2)}$ and $(\partial\phi)^2$, but these will not play an important role in our discussion; so we will not consider them.

and the Einstein tensor

$$E_{\mu\nu} = \frac{1}{8\pi G_N}\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right) . \tag{3.7}$$

The radial-radial Einstein equation $E^w_w = \mathcal{T}^w_w$ combined with the definition of the Brown-York stress energy tensor,

$$T_{ij} = -\frac{2}{\sqrt{-g}}\frac{\delta S_{\text{on-shell}}}{\delta g^{ij}} = -\frac{1}{8\pi G_N}\left(K_{ij} - g_{ij}K + g_{ij}\frac{B_{CT}}{\ell}\right) \tag{3.8}$$

yield

$$T^i_i = \frac{1}{B_{CT}(\Phi)}\left[4\pi G_N \ell\left(T^2_{ij} - (T^i_i)^2\right) + \frac{\ell}{2}(\partial_w\Phi)^2 - \frac{\ell}{2}\partial_i\Phi\partial^i\Phi - \ell V - \frac{1}{8\pi G_N \ell}B_{CT}(\Phi)^2\right] \tag{3.9}$$

at the boundary $w = w_c$. [16]

Next, we apply the E^w_j Einstein equation

$$E^w_j = \frac{1}{8\pi G_N}\nabla^i(K_{ij} - Kg_{ij}) = \mathcal{T}^w_j = \partial_w\Phi\partial_j\Phi\,. \tag{3.10}$$

Using (3.8) to trade K_{ij} for T_{ij}, this becomes

$$\nabla^i T_{ij} = -\left(\partial_w\Phi + \frac{1}{8\pi G\ell}\partial_\Phi B_{CT}(\Phi)\right)\partial_j\Phi\,. \tag{3.11}$$

As expected, x-dependent sources (i.e. $\partial_j\Phi \neq 0$), generically break space-time translations and lead to a non-conserved T_{ij}. Here we will be interested in a conserved energy momentum tensor, obtained for boundary conditions that can be either local Dirichlet ($\Phi = \Phi_0$) or local Neumann-like ($\partial_w\Phi + \partial_\Phi B_{CT}(\Phi) = 0$). In the Dirichlet case, for example, we see that the $-\partial_i\Phi\partial^i\Phi$ term in the trace flow equation vanishes. Let us focus to be specific on the Dirichlet boundary condition for all fields, though we see no obstruction to extending our results readily to other choices of boundary conditions.

The next step is to identify the boundary theory operator corresponding to the bulk scalar field. This proceeds as in e.g. [5, 55]. The natural generalization of the identification of T_{ij} with the extrinsic curvature [46] is the identification of $\mathcal{O}$ with the radial field momentum Π corresponding to Φ:

$$\mathcal{O} = \ell^{-\Delta+3/2}\left(\frac{L_c}{L}\right)^\Delta \Pi = \left(\frac{\lambda c}{12}\right)^{-\Delta/2}\ell^{3/2}\Pi\,. \tag{3.12}$$

In AdS/CFT, this normalization ensures that $\langle\mathcal{O}(x)\mathcal{O}(0)\rangle \sim 1/|x|^{2\Delta}$.[17] Here we have chosen the same definition for the deformed theory, which gives a c-independent 2-point function. We note that this normalization differs from the one used in the double-trace deformation literature, where single-trace operators are N times a trace of products of adjoint fields. These single-trace operators have 2-point

[16] This is essentially equation (A.6) of [4] with η in the Λ_2 term replaced by $\eta \to -V + \frac{1}{2}(-\partial_i\Phi\partial^i\Phi + \Phi'^2)$, which is the combination that appears in the bulk stress-energy tensor component $\mathcal{T}^w_w$. Some terms have different signs from [4], because we define the Brown-York stress tensor as in (3.8).

[17] Intuitively, the power of $(L_c/L)^\Delta$ cancels a power of the radial cutoff that appears in the gravity calculation, while the power of ℓ ensures that the engineering dimensions match.

functions of order N^2, as opposed to the N-independent normalization we use in this work.

Although we have written for brevity a single scalar field and its dual operator, we will be interested in including two matter sectors. First, in our construction we will require a scalar sector encoding the uplift from negative to positive bulk cosmological constant. In §4 we will show that this ensures continuity of the energy levels, including for those below the dominant $\Delta \simeq c/6$ states. Given our choice of normalization with a G_N-independent bulk kinetic term, the $1/G_N$ effects from the uplift sector Φ_u are encoded in a nonlinear potential of the form

$$V(\Phi_u) = \frac{1}{G_N}\tilde{V}(\sqrt{G_N}\Phi_u)\,. \tag{3.13}$$

Second, we will include a probe sector that contributes at $O(G_N^0)$, in order to capture local dynamical bulk fields in the bulk effective field theory. Such fields arise naturally if their mass is protected to be $\ll 1/G_N$ (the bulk Planck scale). Examples include gauge theory sectors. In three dimensions, such gauge fields A can be dualized to scalars ϕ, $dA = *d\phi$. A Dirichlet boundary condition for the scalars corresponds to a Neumann one for the gauge fields and vice versa.[18] The interaction potential in the probe sector can in general be nontrivial; but for that sector we will find it convenient to illustrate our main points in terms of a free bulk scalar field (with the aforementioned uplift sector illustrating nonlinear matter).

In order to compute the conjugate momentum associated to the radial evolution of the scalar field, we note that the counterterm in the action (3.5) can be written as a bulk total derivative,

$$-\frac{1}{8\pi G_N \ell}\int_{\partial\mathcal{M}} d^2x\,\sqrt{-g}\,B_{CT} = -\frac{1}{8\pi G_N \ell}\int_{\mathcal{M}} d^2x dw\sqrt{-g}\left(\frac{1}{2}g^{ij}\partial_w g_{ij}\,B_{CT} + (\partial_\Phi B_{CT})\,\partial_w\Phi\right)\,. \tag{3.14}$$

The first term here, proportional to $\partial_w g_{ij}$, is responsible for the B_{CT} contribution to the stress tensor. The second term, on the other hand, contributes to the scalar canonical momentum,

$$\Pi = \frac{1}{\sqrt{-g}}\frac{\delta S}{\delta(\partial_w\Phi)} = -\partial_w\Phi - \frac{1}{8\pi G\ell}\partial_\Phi B_{CT}\,. \tag{3.15}$$

For our Dirichlet boundary condition, the conjugate momentum Π is free to fluctuate, as befits a dynamical operator in the boundary theory.

Variations of the boundary value $\Phi(w_c) \to \Phi_c + \delta\Phi_c$ of the scalar can be identified with a source $J \propto \delta\Phi_c$. A calculation similar to that in (3.12) gives

$$J = \left(\frac{\lambda c}{12}\right)^{\Delta/2-1}\ell^{1/2}\delta\Phi_c\,. \tag{3.16}$$

The source vanishes, $J = 0$, for our self-contained system at fixed energy at any point along the deformation trajectory including during the uplift. But we keep track of it because it enters into the derivation of the operator $\mathcal{O}$ required to move along the trajectory, adding the new $\mathcal{O}\mathcal{O}$ at each

[18] An advantage of the gauge theory sectors is the simplicity of identifying the operator in a simple way in the boundary dual. It is obtained via the Noether current $\mathcal{J}$ for the corresponding global symmetry in the dual. An ensemble with fluctuating $\pi_\phi = \partial_w\phi + \dots$ as we layed out above corresponds to fixed rather than fluctuating $\mathcal{J}$, so it is obtained by a Legendre transform.

step. In the boundary theory language, the operator $\mathcal{O}\mathcal{O}$ that we need to add is obtained from $\langle \mathcal{O}\dots\mathcal{O}\rangle = \frac{\delta}{\delta J}\dots\frac{\delta}{\delta J}\log Z_{dresssed}$. In these variables, the trace flow equation becomes

$$\begin{aligned} T_i^i \;=\; & \frac{1}{B_{CT}}\left[4\pi G_N \ell\left(T_{ij}^2-(T_i^i)^2\right)+\frac{\ell}{2}\Pi^2\right]+\frac{1}{8\pi G_N}\frac{B'_{CT}}{B_{CT}}\Pi \\ & -\frac{\ell}{2B_{CT}}\partial_i\Phi\partial^i\Phi-\frac{\ell}{B_{CT}}V+\frac{1}{B_{CT}}\left(\frac{\ell}{2}\left(\frac{B'_{CT}}{8\pi G\ell}\right)^2-\frac{B_{CT}^2}{8\pi G\ell}\right). \end{aligned} \tag{3.17}$$

Lets focus on the term proportional to $B'_{CT}\Pi$ in the above expression, which is linear in $\mathcal{O}$. Perturbing the action by $\int d^2x\sqrt{-\gamma}J\mathcal{O}$ (with γ the 2d metric) gives rise at first order in $\mathcal{O}$ to

$$\int d^2x\sqrt{-\gamma}T_i^i \supset -\int d^2x\sqrt{-\gamma}\beta(J)\mathcal{O}\,, \tag{3.18}$$

with $\beta(J)=L_c\partial_{L_c}J$ the QFT beta function for J. Thus, comparing with (3.17), we read off the beta function

$$\beta(J)=L_c\frac{\partial J}{\partial_{L_c}}=-\frac{1}{\pi}\left(\frac{\lambda c}{12}\right)^{\Delta-2}\frac{\partial_J B_{CT}(\Phi_c+J)}{B_{CT}(\Phi_c+J)}\,. \tag{3.19}$$

In an asymptotically AdS region, and choosing B_{CT} as the standard holographic counterterm, this reproduces the correct holographic beta function (see e.g. [63]).

On the other hand, defining $\mathcal{W}[w_c,J]$ as the generating functional of correlation functions in the dual QFT at finite cutoff, we know that

$$\int d^2x\sqrt{-g}T_i^i=L_c\frac{d\mathcal{W}}{dL_c}=L_c\frac{\partial\mathcal{W}}{\partial L_c}+\int d^2x\sqrt{-\gamma}L_c\partial_{L_c}J\frac{\delta\mathcal{W}}{\delta J}=L_c\frac{\partial\mathcal{W}}{\partial L_c}-\int d^2x\sqrt{-\gamma}\beta(J)\mathcal{O} \tag{3.20}$$

where we made use of $\frac{\delta\mathcal{W}}{\delta J}=-\mathcal{O}$. Therefore we conclude that terms proportional to $\beta(J)$ cancel. So far our discussion pertained to any scalar field. We will illustrate our prescription with two scalar sectors: a nonlinear uplift sector scalar Φ_u and a second, Gaussian scalar sector dual to the operator $\mathcal{O}$. Writing $L_c\partial_{L_c}\mathcal{W}=-2\lambda\partial_\lambda\mathcal{W}$, we end up with

$$\begin{aligned} \lambda\frac{\partial\mathcal{W}}{\partial\lambda}=\int d^2x\sqrt{-\gamma}\frac{-1}{B_{CT}(\Phi_{uc}+J_u)}\Bigg[& 2\pi\lambda T\bar{T}+\frac{1}{4}\left(\frac{\lambda c}{12}\right)^{\Delta-1}\left[\mathcal{O}\mathcal{O}+\mathcal{O}_u\mathcal{O}_u\right]-\frac{1}{4}\left(\frac{\lambda c}{12}\right)^{2-\Delta}(\partial_i J)^2 \\ & -\left(\frac{\lambda c}{48}\right)^{1/2}\left(\frac{L_c}{L}\right)^3 V(\Phi_{uc}+J_u)+\frac{1}{4\pi^2\lambda^2}\left(\frac{\lambda c}{12}\right)^{\Delta-1}B'_{CT}(\Phi_{uc}+J_u)-\frac{B_{CT}(\Phi_{uc}+J_u)^2}{2\pi\lambda}\Bigg] \end{aligned} \tag{3.21}$$

where the derivative in B'_{CT} is taken with respect to J_u.[19]

Below we will analyze in more detail the definition of the composite $\mathcal{O}\mathcal{O}$ operator in the field theory side. Besides the semiclassical factorized contribution $\sim(\langle\mathcal{O}\rangle)^2$, there are additional terms controlled by short-distance correlators. Our dual theory will turn out to be a finite quantum-mechanical theory, where these contributions are well-defined. Given this, we will define the $\mathcal{O}\mathcal{O}$ contribution to the trace

[19] The third and the last term in (3.21) contain L_c and Φ_{uc}, but they can be expressed purely in terms of boundary quantities upon substituting explicit expressions for $V(\Phi)$ and $B_{CT}(\Phi)$. For example, in the absence of sources $J_u=0$, when the potential $V(\Phi_{uc})$ is at its AdS/dS minimum, they reduce to the Λ_2 term, $(\eta-B_{CT})/2\pi\lambda$.

flow equation by subtracting these effects; see the discussion below (3.32).

We work with vanishing sources ($J = 0$) to construct a self-contained bounded patch of spacetime. Putting all this together and looking ahead, we will arrive at a definition for the dual theory of the form of an explicitly renormalized version of (1.4). Its existence is aimed first of all to validate the entropy count in [3]. Moreover, in §6 we will show how to use it to compute some simple quantities of interest in the boundary theory and compare them to the gravity side.

3.2 Finite system at finite c

As stressed in §1, the addition of the $\mathcal{OO}$ terms in our defining equation introduces new complications in the finite-c setting of interest. In order to obtain a well-defined theory, we will adjust the above prescription as follows. The resulting theory is not as easily solvable as the pure $T\bar{T} + \Lambda_2$ theory, but it is well defined and we will see below that in simple cases we can generalize the dressed energy calculation to the leading nontrivial order in the effects of matter.

Starting from the seed CFT, we first carry out a pure $T\bar{T}$ deformation to a very small value of the dimensionless deformation parameter $y = \lambda/L^2$; let us denote it $y_T \ll 1$. At this point, the real spectrum is finite [6]. From this point on we embark on the $T\bar{T} + \mathcal{OO} + \Lambda_2$ deformation.

Of course, the reason we include the $\mathcal{OO}$ terms in the first place is to ensure bulk locality at the level of approximation corresponding to low energy bulk EFT. The reason this is consistent with our initial pure $T\bar{T}$ trajectory is as follows. We are free to take y_T small enough that the initial violation of locality introduced by the pure $T\bar{T}$ deformation from y to y_T is only applicable to such high energy excitations that quantum gravity effects enter (for which there is no expectation of locality [64, 65]). Related to this, we stress that only infinite-energy excitations actually reach the original AdS boundary in AdS/CFT. This includes excitations of massless fields for which the wave dynamics beyond the geometric optics approximation is important; a finite-energy excitation does not just follow the null geodesic trajectory (which does reach the boundary). Our finite theory may be taken as the definition of the bulk quantum gravity, which appropriately deviates from EFT at such high energy scales.

Once the theory has a finite-dimensional Hilbert space, the additions along the main trajectory, including the $\mathcal{OO}$ terms, are all well defined: in a basis of energy eigenstates the (unrenormalized) operator is

$$\langle n|\mathcal{OO}|m\rangle = \sum_p \langle n|\mathcal{O}|p\rangle\langle p|\mathcal{O}|m\rangle \tag{3.22}$$

which is unambiguous since the sum over p is finite. In general the expectation value $\langle n|\mathcal{OO}|n\rangle$ does not factorize, in contrast to the special case of $T\bar{T}$ which factorizes up to a total derivative. Moreover, formally subleading contributions in the expansion around large N can diverge as $y_T \to 0$. Below in §5, we will estimate the corrections to factorization and define an appropriate renormalized $\mathcal{OO}$ operator to add to our theory at each step in our algorithm.

After our $T\bar{T}$ deformation up to the small value y_T of the dimensionless deformation parameter

y, the dressed energies take the form

$$\mathcal{E}_T = \frac{1}{\pi y_T}\left(1 - \sqrt{1 - 2\pi y_T \mathcal{E}_{(0)} + 4\pi^4 J^2 y_T^2}\right) \tag{3.23}$$

First, we note that with $J = 0$, the cut off on the real spectrum arises for seed dimensionless energy $\mathcal{E}_{(0)} = 2\pi(\Delta - c/12)$ of a maximal value $\mathcal{E}_{(0),max} = 1/2\pi y_T$. In fact, this scaling with y_T persists in the presence of spin, as we can see in the following way as a consequence of the unitarity bound.

We would like to determine whether there is an upper cutoff on $\varepsilon_{(0)}$ such that for higher energies ε becomes complex. This is controlled by

$$D_0 = 1 - 2\pi\varepsilon_{(0)}\, y_T + 4\pi^4 J^2\, y_T^2\,. \tag{3.24}$$

At a fixed y_T, one could think about trying to increase the spin J such that $D_0 > 0$. However, the unitarity bound in $d = 2$ is

$$J \leq \Delta \;\Rightarrow\; J \leq \frac{1}{2\pi}\left(\varepsilon_{(0)} + \frac{c}{12}\right)\,. \tag{3.25}$$

We are interested in the universal heavy states, which have $\varepsilon_{(0)} \sim c$, and we want to see whether we can make $\varepsilon_{(0)}$ arbitrarily large, at the same time increasing J, such that ε stays real in the window of interest for y_T. So let us set

$$J = \alpha\varepsilon_{(0)}\,, \tag{3.26}$$

and at sufficiently high energies unitarity requires

$$\alpha \leq \frac{1}{2\pi}\,. \tag{3.27}$$

Returning to (6.27d), we have

$$D_0 = 1 - 2\pi y_T\, \varepsilon_{(0)} + 4\pi^4 y_T^2\, \alpha^2\, \varepsilon_{(0)}^2\,. \tag{3.28}$$

So D_0 has two roots as a function of $\varepsilon_{(0)}$,

$$\varepsilon_{(0)}^{\pm} = \frac{1}{4\pi^3\alpha^2 y_T}\left(1 \pm \sqrt{1 - (2\pi\alpha)^2}\right)\,. \tag{3.29}$$

For $\varepsilon \to 0$, $D_0 > 0$. If the two roots are complex, this means that D_0 will always stay positive for all $\varepsilon_{(0)}$, and then we could go to arbitrarily large ε_0 without the dressed energy becoming complex. However, because of the unitarity bound (3.27) on α, we see that $D_0 = 0$ always has two real roots. We conclude that for a given choice of spin compatible with unitarity, we need to impose a cutoff

$$\varepsilon_{(0)} < \varepsilon_{(0)}^{-} = \frac{1}{4\pi^3\alpha^2 y_T}\left(1 - \sqrt{1 - (2\pi\alpha)^2}\right) \equiv \frac{c_\alpha}{y_T} \tag{3.30}$$

so that the dressed energy stays real. From (3.30) we see that including spatial momentum J, the spectrum cuts off at a dimensionless energy $\propto 1/y_T$.

Let us now apply this to the question of the contributions to (3.22). We note that although there are $e^{\mathcal{S}}$ states appearing in the spectral decomposition (3.22), almost all of them have negligible form

factors $\langle n|\mathcal{O}|p\rangle$. The usual UV ambiguities in defining the irrelevant deformation arise from the high momenta entering into the spectral decomposition of $\mathcal{O}\mathcal{O}$, corresponding to a dimensionless energy $\mathcal{E}_p \sim |p|L$. For an operator in the seed CFT of dimension Δ, this goes like

$$\int^{M_*} d^2p(|p|L)^{2\Delta-2} \sim \left(\frac{c_\alpha}{y_T L}\right)^{2\Delta} . \tag{3.31}$$

where M_* is a UV cutoff scale, which for us is proportional to $\frac{1}{y_T L}$ right after the first, pure $T\bar{T}$ segment of our 4-part trajectory. In an energy eigenstate, we obtain for (3.22)

$$\langle n|\mathcal{O}\mathcal{O}|n\rangle = \langle n|\mathcal{O}|n\rangle^2 + o\left(f(\{Q_i\})\left(\frac{c_\alpha}{y_T}\right)^{2\Delta}\right) . \tag{3.32}$$

where $f(\{Q_i\}$ depends on the detailed content of the dual seed CFT (e.g. in the D1-D5 theory, it includes Q_1 and Q_5 dependence).

In [55], at the large-N level (meaning large-c in 3 bulk dimensions), the operator $(\mathcal{O}\mathcal{O})_{c=\infty}$ was defined by subtracting self-contractions, retaining the leading term in the OPE of $\mathcal{O}(x)\mathcal{O}(y)$ as $x \to y$ that is not the identity and that survives in the strict $c = \infty$ theory. Our goal is to approximate this in a finite-c theory. In our explicitly regulated system, the Hilbert space is of course different from the large-c one in [55]. Below in §5, we will spell out the prescription to match correlators in the regime where GR + EFT applies. We can define the operator $\langle n|\mathcal{O}\mathcal{O}|m\rangle$ by hand to match the correlators of GR + EFT to very good approximation at large but finite c. This is a brute force method, but the essential point is that it is well-defined thanks to the power of the $T\bar{T}$ deformation to serve as an explicit regulator by cutting off the real spectrum in a calculable way [6, 8].

Altogether, the operator (3.22) and, given appropriate subtractions, the deformation (1.4), are well defined in the full theory. This will be constructed by hand to complete the duality and microstate count of [3]. The full theory with mattter is not as readily exactly solvable at finite c as in the case of $T\bar{T} + \Lambda_2$. This befits the system of interest, with sufficiently rich dynamics to realize large-radius gravity probed by light quantum fields. Still, below we will find methods to calculate physical quantities of interest, finding agreement on the two sides of the duality.

3.3 Implication for operator algebra

Since we have a finite system, its Von Neumann operator algebra is type I. In a strict $G_N \to 0$ limit, [39] found a type II_1 algebra. Our system has finite c, hence finite G_N, and the algebra is simply type I.

4 Uplift and energy continuity

The transition from $T\bar{T} + (\mathcal{O}\mathcal{O})_r$ to $T\bar{T} + (\mathcal{O}\mathcal{O})_r + \Lambda_2$ involves a particular form of bulk matter. Specifically, we should incorporate a scalar potential that contains holographic AdS and dS minima and take into account the uplift from one to the other during the trajectory. String-theoretic models

uplifting AdS/CFT to dS include [31, 56]. The relevant matter sector for this sector in the gravity-side description is an uplift scalar Φ_u whose potential $V(\Phi_u)$ contains both an AdS minimum at Φ_{uA} and a dS minimum at Φ_{uS}. The uplift step occurs at fixed $y = y_0 = 8\ell G_N/L_{c0}^2 = 3/c\pi^2$. We continuously adjust the Dirichlet boundary condition for the scalar, which is a coupling constant in the boundary theory, from Φ_{uA} to Φ_{uS}. In the gravity-side description, the classical solution space for matter excitations continuously deforms according to this continuous deformation of the scalar boundary condition (which resolves the jump in Λ_2 in [3, 4]). After the uplift step, we proceed with the $T\bar{T} + (\mathcal{O}\mathcal{O})_r + \Lambda_2$ deformation, increasing y from y_0 to capture the bounded cosmic horizon patch.

The detailed uplift was not required for the pure gravity results in [3], but plays a significant role in the presence of matter, leading to continuity of the energy spectrum. We can see this by reviewing the behavior of the various classes of states as we evolve along the $T\bar{T} + \Lambda_2$ trajectory, for which Λ_2 jumps to a nonzero vallue at $y = y_0$. Without the $(\mathcal{O}\mathcal{O})_r$ contribution, the trajectory in [3] does exhibit continuity of the dressed energy along the trajectory for the dressing of the seed states in the energy band $\Delta \sim c/6$ (up to very fine discontinuities that are invisible to the course-grained gravitational description). In the gravity description, this arises from bringing in the boundary with $T\bar{T}$ to the horizon of a BTZ black hole, then bringing it out with $T\bar{T} + \Lambda_2$ to obtain the dS_3 cosmic horizon patch (the static patch, with boundary). At the junction between the two trajectories, at a dimensionless deformation parameter $y_0 = \lambda_0/L^2 = 3/c\pi^2$ the square root term vanishes in the dressed energy

$$\mathcal{E} = \frac{1}{\pi y}\left(1 \mp \sqrt{\eta + \frac{y}{y_0}(1-\eta) - 4\pi^2 y(\Delta - c/12)}\right) \qquad (\text{pure } T\bar{T} + \Lambda_2), \tag{4.1}$$

so that the Λ_2-dependence (in η) drops out for the entropically dominant $\Delta \simeq c/6$ band of energies. For that band of energies, to capture the cosmic horizon patch the branch of the square root is positive in the $T\bar{T} + \Lambda_2$ phase.

The other universal state is the vacuum, which is captured by the original (-) sign of the square root. In particular, as discussed in §2 we stress that once we turn on Λ_2, our spectrum includes both signs of the square root. In gravity-side language, this reflects the fact that the cylinder size L_c is fixed, but the extrinsic curvature is not. It can take both signs; the spectrum includes both universal sectors $\Delta \simeq c/6$ dual classically to the cosmic horizon patch, and the vacuum $\Delta = 0$ dual classically to the pole patch.[20] From the boundary theory perspective, our Hamiltonian $\sum_n E_n|n\rangle\langle n|$ can include both types of levels. In the path integral formalism, a model that illustrates this very simply is the relativistic point particle. Its action $S = -m\int d\tau\sqrt{-\dot{X}^\mu \dot{X}_\mu}$ can be rewritten as a quadratic action in $\dot{X}_\mu$ by introducing a dynamical world-line vielbein [64]. The equation of motion for the vielbein admits two solutions corresponding to both branches of a square root. This structure arises in the complete path integral representations of $T\bar{T} + \Lambda_2$, both using JT gravity [7] and massive gravity [66–68]. We note that before we turn on Λ_2, there is not a role for both signs of the square root. In gravity language, this is because although the extrinsic curvature is not fixed, the + sign corresponds to filling the patch toward the asymptotic boundary, which is not a normalizable excitation.[21]

[20] The T_t^t component of the Brown-York stress tensor is given by $T_t^t = \frac{1}{8\pi G_N}\left(K_t^t - (K - 1/\ell)\right) = \frac{1}{8\pi G_N}\left(\frac{1}{\ell} - K_\theta^\theta\right)$. For a metric of the form $ds^2 = dw^2 - g(w,t)^2dt^2 + r(w,t)^2d\theta^2$, $K_\theta^\theta = \frac{1}{r}\frac{\partial r}{\partial w}$ determines the square root part of T_t^t.

[21] See however interesting works such as [69] [18] which consider both signs in the AdS theory.

More general states with $0 < \Delta \ll c/6$ on the other hand are not captured by the pure gravity sector. They are model dependent; in gravity language, they can include BTZ black holes below the Hawking-Page level, particle states, and quantum field excitations. In the gravity description, we deform from $y = y_T \ll 1$ to $y_0 = 8\ell G_N/L_{c0}^2 = 3/c\pi^2$ by changing the boundary condition governing the bulk solutions, specifically reducing the bounding proper cylinder size L_c to L_{c0}. At this value L_{c0}, the low-lying states do not typically reach the boundary. (If they do so at one time, they bounce off of it back into the bulk, given the $\mathcal{O}\mathcal{O}$ contributions ensuring a local boundary condition.) So (looking at the state at a generic time) the deformation to the matching point y_0 only brings the boundary in to a radial position outside of the black hole horizon or particle position. If we then simply add Λ_2, without an uplift sector, and continue with the trajectory (decreasing L_{c0} with $\eta = -1$ and the opposite branch of the square root) we get a discontinuity. That is, in the absence of an explicit uplift sector, the square root term in the energy (4.1) does not vanish at $y = y_0$ in this case, so reversing its sign implies a discontinuity. As stressed in [3], this is as expected since these low-energy model-dependent states involve the matter contributions, which should therefore be included to restore continuity. It is this point we will establish in this section.

After adding matter contributions and the uplift sector, as discussed above there is a continuous trajectory that combine both branches. This means that both signs are contributing to the path integral. Of course, the nonlinear nature of the equations complicates the solutions matching the boundary value of Φ_{uc} in between its AdS and dS values. An interesting question is the degree of quantum mechanical mixing among semiclassical states with different signs of the extrinsic curvature in the full theory including matter. A quantum mechanical model of such mixing can be given in terms of a potential with two minima, corresponding to the two branches. Small fluctuations around each minimum give rise to dressed energies $E_n^{\pm}$ and eigenstates $|\Psi_n^{\pm}\rangle$. Furthermore, the quantum-mechanical Hamiltonian contains off-diagonal terms that mix both sectors, $\langle\Psi_n^+|H|\Psi_m^-\rangle \neq 0$; these reflect the finite-energy trajectories that connect both types of states.

In addition to the states described above, our theory contains configurations which access the landscape of AdS and dS vacua given the uplift sector potential. These are also continuous along our trajectory. We can construct a domain wall at a fixed time t which interpolates between Φ_A and Φ_S. For these states as well, we would like to see how the deformation including matter, as expressed for example by the trace flow equation (3.9), automathically yields continuity of the states making up the model-dependent sparse light spectrum. In the gravity description, given the uplift potential we can construct a state which at a moment of time symmetry contains a domain wall interpolating between the locally AdS and dS regions. In these variables, there is a manifestly continuous trajectory obtained by moving the Dirichlet condition on Φ to interpolate between these two values. The boundary cylinder grows in the AdS region as we move away from the particle or BTZ horizon, and it shrinks once we continue into the dS region. Therefore it will have a maximum somewhere at which $dg_{\theta\theta}/dw = 0$ as it flips sign. This implies that the extrinsic curvature component K_θ^θ, which determines the square root part of the dressed (=Brown-York) energy formula vanishes, and it connects continuously to the opposite sign of the square root.

5 Summary of Boundary Theory Prescription

We now express the construction of our deformed theory in $2d$ boundary variables and explain how to apply it to derive properties of the emergent geometry. We will focus on new features and issues arising from the addition of matter. In particular we will use the matter at the leading nontrivial order to validate and probe the emergent large-radius de Sitter geometry.

The deformation of our theory was written in (3.21) in Lagrangian language at large c. In order to make contact with the finiteness of the real spectrum at finite c, it is useful to recast this in Hamiltonian language where the path integral for a transition amplitude with sources takes the form

$$\langle \chi(\theta,t)|\chi'(\theta',t')\rangle_{J,K} = \int_{b.c.} D\chi D\pi_\chi \exp\left(i\int d^2x[\dot\chi\pi_\chi - \mathcal{H}(\chi,\pi_\chi) + J_\chi\chi + K_\pi\pi_\chi + J\mathcal{O}]\right) \tag{5.1}$$

Here we use χ to denote the deformed-CFT fields.

We will include an uplift sector denoted with a subscript u here, and a marginal scalar $\mathcal{O}$ with $\Delta = 2$ dual to a gauge field which will realize the required approximate bulk locality. Including the initial stint of pure $T\bar T$ to instate a finite real spectrum as explained in §3.2, and defining $\mathcal{S}_H = \int d^2x\mathcal{H}$ where $\mathcal{H}$ is the Hamiltonian density of the theory, we have

$$\lambda\frac{\partial\mathcal{W}}{\partial\lambda} = -\lambda\frac{\partial\mathcal{S}_\mathcal{H}}{\partial\lambda} = -\int d^2x\sqrt{-\gamma}\,2\pi\,\lambda T\bar T, \qquad y < y_T \tag{5.2}$$

$$\begin{aligned}\lambda\frac{\partial\mathcal{W}}{\partial\lambda} = -\lambda\frac{\partial\mathcal{S}_\mathcal{H}}{\partial\lambda} = \int d^2x\sqrt{-\gamma}\frac{-1}{B_{CT}(\Phi_{uc}+J_u)}\Bigg[2\pi\lambda T\bar T + \frac14\left(\frac{\lambda c}{12}\right)^{\Delta-1}[(\mathcal{O}\mathcal{O})_r + (\mathcal{O}_u\mathcal{O}_u)_r] - \frac14\left(\frac{\lambda c}{12}\right)^{2-\Delta}(\partial_i J)^2 \\ - \left(\frac{\lambda c}{48}\right)^{1/2}\left(\frac{L_c}{L}\right)^3 V(\Phi_{uc}+J_u) + \frac{1}{4\pi^2\lambda^2}\left(\frac{\lambda c}{12}\right)^{\Delta-1}B'_{CT}(\Phi_{uc}+J_u) - \frac{B_{CT}(\Phi_{uc}+J_u)^2}{2\pi\lambda}\Bigg]\,,\ y > y_T\,.\end{aligned} \tag{5.3}$$

where we indicated renormalized operators we will shortly define, and Δ refers to the CFT dimension of the appropriate (uplift or probe) operator. For a given seed theory, the light spectrum may contain multiple such sectors, which enter similarly in the definition of the deformation.

For the first segment (5.2) of our deformation, for a given total spatial momentum p_θ, the Hamiltonian manifestly goes complex outside a finite region of phase space. This is related to the energies going complex as reviewed above in §2.2. It pertains to the dressed classical Hamiltonian as was computed for a free scalar seed theory: see e.g. equations (2.16-17) of [48]. If we restrict the path integral to this region, we respect unitarity. Combining this with the discretenesss of the spectrum, we realize the finite real spectrum of the theory, while retaining use of the tool of continuous path integration. This formalism will be useful for deriving the operators we add along the trajectory.

Once we start adding the $(\mathcal{O}\mathcal{O})_r$ contributions, we will not be able to analytically resum the Hamiltonian in the same way as was done for the solvable $T\bar T + \Lambda_2$ deformation. This complication is as expected: in general energy levels do not evolve independently in the more general case.

However, such a resummation is not needed to define the theory. As in the familiar $T\bar T$ deformation we can think of this as an algorithmic step by step procedure, advancing along the trajectory in the

space of theories according to (5.2)-(5.3) as λ infinitesimally increases by $\Delta\lambda$. In order to determine the operators associated to conserved currents, we may go between the Hamiltonian and Lagrangian formalisms, using the latter to derive the currents and the former to manifest the finiteness of the real energy spectrum. In particular, at each step in the deformation, we recompute $T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta}{\delta g^{\mu\nu}}S_{dressed}$ in order to add $\Delta\lambda T\bar{T}$ to the Hamiltonian density. Similarly, for a sector consisting of a conserved current $\mathcal{J}$ corresponding to a global symmetry $\chi \to \chi + \Delta\chi$, $\mathcal{L}_{dressed} \to \mathcal{L}_{dressed} + \partial_\mu V^\mu$, we recompute $\mathcal{J}^\mu = \frac{\partial\mathcal{L}_{dressed}}{\partial(\partial_\mu\chi)\Delta\chi} - V^\mu$. (We could also use the Noether procedure for the stress-energy tensor itself.) These fields associated with symmetries are naturally light on the gravity side and hence well motivated as appropriate probes of the bulk locality.

In the more general case, the operators $\mathcal{O}$ can be obtained in principle as follows. They satisfy $\langle\mathcal{O}_{I_1}\dots\mathcal{O}_{I_n}\rangle = \frac{\delta}{\delta J^{I_1}}\dots\frac{\delta}{\delta J^{I_n}}\mathcal{W} = -\frac{\delta}{\delta J^{I_1}}\dots\frac{\delta}{\delta J^{I_n}}(\mathcal{S}_H - i\int d^2x J\mathcal{O})$. With the initial pure $T\bar{T}$ trajectory in place (5.2), $\langle\mathcal{O}(x_1)\mathcal{O}(x_2)\prod_k\mathcal{O}_k\rangle$ will be finite as we take $x_1 \to x_2$. This follows from the discreteness of the spectrum and the finite range of real energies in the phase space region of real $\mathcal{H}$. In the finite system, the labeling of operators $\mathcal{O}(x)$ in terms of x will yield redundant operators. Thus the operator $(\mathcal{O}\mathcal{O})_r$ that we deform by is well defined and in principle determined by calculating all of its correlators (or enough for the desired precision) at a given step in the trajectory, comparing it to the results of GR + EFT when available, and explicitly subtracting deviations from the latter in order to determine what to add at the next step.

More specifically, we can arrange this agreement as follows, taking into account the difference in Hilbert spaces between our finite theory and the formal large-N theory [55]. Start from an energy level or band of energies visible in the universal, pure gravity, sector. In the bounded AdS part of the trajectory, this includes states whose gravity description matches the external geometry of the BTZ black holes with masses corresponding to operator dimensions which exist in the seed CFT spectrum, along with the vacuum state matching empty AdS. In the bounded dS phase it includes states matching the pole patch and the cosmic horizon patch. From there we can construct, in our finite theory, correlators

$$\langle universal|(\mathcal{O}\mathcal{O}(x))_r\mathcal{O}(y_1)\dots\mathcal{O}(y_m)|universal\rangle \tag{5.4}$$

which probe matter excitations. These are limited in GR + EFT in several ways. The distance between points on the boundary cannot fall beneath the Planck length, $|y_i - y_j| > G_N, |x - y_i| > G_N$. The lifetime of the theory is uncertain in semiclassical GR + EFT after a timescale $T \sim G_N\exp(S)$ (for entropy S), corresponding to energy resolution bounded by $\Delta E > G_N^{-1}\exp(-S)$. These limitations reduce the number of independent time-ordered correlators to match to be less than order $S\exp(S)$. In fact this is a conservative over-estimate of the count of correlators to match between our theory and GR: GR + EFT is only a good approximation down to length scales $\gg G_N$. Moreover, matter containing too much entropy will condense into a black hole, thus joining the universal sector of states corresponding to pure gravity. Altogether, we have the freedom to tune the e^{2S} matrix elements $\langle n|(\mathcal{O}\mathcal{O})_r|m\rangle$ such that our correlators match those of GR + EFT when it is under control.

This leads to an equation for the deformation of the expectation value of the interaction Hamito-

nian in the nth energy eigenstate as in [6] as

$$\partial_\lambda \langle n|H|n\rangle = \langle n|2\pi T\bar{T}+\frac{1}{4\lambda}\left(\left(\frac{\lambda c}{12}\right)^{\Delta_u-1}(\mathcal{O}_u\mathcal{O}_u)_r+\left(\frac{\lambda c}{12}\right)^{\Delta-1}(\mathcal{O}\mathcal{O})_r\right)|n\rangle-\tilde{V}(\Phi_{uc}(\lambda))-\frac{B_{CT}(\Phi_{uc}(\lambda))^2}{2\pi\lambda^2} \tag{5.5}$$

where again the last two terms account for the uplift explained in the previous section.[22]

For the pure $T\bar{T}+\Lambda_2$ theory (with no $\mathcal{O}\mathcal{O}$ terms and with V and B_{CT} constant), the dressed energies $E_n = \langle n|H|n\rangle$ can be analytically determined [6][3]. This proceeds from either the trace flow equation (2.1) or (5.5) by relating the independent stress-energy components T^0_0, T^1_1, and T^0_1 to the energy, dE/dL, and conserved spatial momentum respectively. Consider the T^θ_θ component of the dressed stress energy tensor (with θ the spatial direction around the cylinder $-dt^2+L^2d\theta^2$ that the boundary theory lives on). For a homogeneous time-independent state this is the pressure $T^\theta_\theta = -dE/dL$. It is useful to work with the dimensionless deformation parameter $y=\lambda/L^2$. With these substitutions, one obtains the differential equation

$$\pi y\mathcal{E}(y)\mathcal{E}'(y)-\mathcal{E}'(y)+\frac{\pi}{2}\mathcal{E}(y)^2=\frac{1-\eta}{2\pi y^2}+2\pi^3J^2 \tag{5.6}$$

in terms of the dimensionless dressed energy.

$$\mathcal{E}=EL\,. \tag{5.7}$$

For a seed CFT,

$$\mathcal{E}|_{y=0,\eta=1}=\mathcal{E}_{CFT}=2\pi\left(\Delta-\frac{c}{12}\right)\,. \tag{5.8}$$

An equivalent way to determine the dressed energies is to apply the trace flow equation

$$T^i_i=4\pi\lambda T\bar{T}+\frac{1}{2}\left(\left(\frac{\lambda c}{12}\right)^{\Delta_u-1}(\mathcal{O}_u\mathcal{O}_u)_r+\left(\frac{\lambda c}{12}\right)^{\Delta-1}(\mathcal{O}\mathcal{O})_r\right)-\Lambda_2+\ldots \tag{5.9}$$

Without matter, one can substitute $T^\theta_\theta=-dE/dL$ and again obtain a differential equation for $E(L)$ [4].

With local bulk matter, we must account for states with more general time and θ dependence. In the general case, $T^\theta_\theta(t,\theta)\neq -dE/dL$ and we have the additional $(\mathcal{O}\mathcal{O})_r\simeq\langle\mathcal{O}(t,\theta)\rangle^2$ terms in the defining equation for the deformation. As stressed above, the procedure is algorithmically well defined, but it is interesting to ask if we can derive a tractable differential equation for the dressed energies (and dressed operators) in the case with approximately local bulk matter. We do not have any reason to expect a simple expression in general, but in the next section we will find this to work out in a simple class of time-dependent states (at fixed energy) which directly probe the approximate locality of the bulk matter. Moreover, in the analysis of the GR + EFT solution space, we will discover that the matter excitations lead to an effect on the energy multiplying the square root term in the dressed energy (2.3), retaining their branch structure dictating their reality properties.

[22] In (5.5), $\tilde{V}$ is the term proportional to the uplift potential in (5.3): $\tilde{V}(\Phi_{uc})=\left(\frac{\lambda_c}{48}\right)^{1/2}\left(\frac{L_c}{L}\right)^3V(\Phi_{uc})$.

6 Dressed energies capturing time-dependent local bulk matter

In this section, we derive the leading effect of matter on simple states. We begin by deriving its structure on the boundary side of the duality. We then derive time-dependent solutions with matter in the classical approximation on the gravity side and comment on the correspondence.

6.1 $2d$ boundary side derivation of dressed energies in simple cases

We can obtain the leading matter contribution to the dressed energies in a simple way in some cases. Suppose we consider a state with no bulk matter excited, and then add energy by exciting the bulk matter. Such states include oscillating solutions for the bulk matter.

In general, we have from the defining equation of the deformation (with $T^t_\theta = 0$ for simplicity)

$$\partial_\lambda \langle n|H|n\rangle = L\langle n|-\frac{\pi}{2}T^t_t T^\theta_\theta + c_{\mathcal{O}}\mathcal{O}\mathcal{O} + \Lambda_2|n\rangle \tag{6.1}$$

where the Hamiltonian $H = -LT^t_t$ and $c_{\mathcal{O}}$ is a constant that may depend on the operator $\mathcal{O}$ as described in §3.1. We are interested in capturing the effect of matter excitations at the leading nontrivial order, denoting as $\Delta T^\mu_\nu, \Delta H$ the new contributions to these quantities generated by the matter excitation.

The trace flow equation is

$$T^t_t + T^\theta_\theta = -\pi\lambda T^t_t T^\theta_\theta - 2\Lambda_2 + 2c_{\mathcal{O}}\mathcal{O}\mathcal{O} \tag{6.2}$$

We can now substitute (6.2) into (6.1), which in general gives

$$\partial_\lambda E = \frac{1}{2\lambda}(LT^t_t + LT^\theta_\theta). \tag{6.3}$$

with time-dependence arising in T^θ_θ in the general case.

Expanding this in the effects of the matter, in the case of a general equation of state $\langle n|\Delta T^\theta_\theta|n\rangle = f(\lambda)\langle n|\Delta T^t_t|n\rangle$, yields the differential equation

$$\partial_\lambda \Delta E = -\frac{(1+f(\lambda))}{2\lambda}\Delta E \tag{6.4}$$

with solution

$$\Delta E = \Delta E(\lambda_*)\frac{\exp\left(-\int_1^\lambda \frac{(1+f(\tilde\lambda))}{2\tilde\lambda}d\tilde\lambda\right)}{\exp\left(-\int_1^{\lambda_*} \frac{(1+f(\tilde\lambda))}{2\tilde\lambda}d\tilde\lambda\right)} \tag{6.5}$$

Thus in this situation with the expectation value of the pressure being proportional to that of the energy, the effect of matter on the energy levels is straightforward to solve for.

In the next section, we will develop the classical matter solutions in our bounded (A)dS patches. For a spatially homogeneous Gaussian matter variable behaving as a harmonic oscillator of a specific

frequency, the expectation value of the pressure in an energy eigenstate is the same as its time average. We will find examples of matter satisfying an equation of state of the above form, with time averaging playing the role of this expectation value.

6.2 Bulk counterpart: matter waves and bulk gravity with fixed boundary cylinder

We next derive the emergent classical solutions describing excited matter in the bulk. This will exhibit important physical effects of the fixed-metric boundary condition with cylinder size L_c. This boundary condition – which freezes the boundary geometry leaving us with a gravity-free dual theory – is distinct from that of global de Sitter. A major implication of the timelike boundary deduced in [9, 10] is the presence of the standard sign in the first law of thermodynamics when written in terms of the boundary energy $E_{\text{Brown-York}}$:

$$\delta E_{\text{Brown-York}} = \delta E_{\text{bulk matter}} + T\delta S_{\text{horizon}} \,. \tag{6.6}$$

In this section, we will illustrate the effect of matter in explicit classical solutions. In an ongoing work we are also studying quantum effects associated to imposing boundary conditions on matter at the timelike boundary. In particular, the resulting negative Casimir energy of the matter in its vacuum state has interesting effects on the causal properties of the spacetime[23]. See [70] for related discussions in dS JT gravity in the absence of timelike boundaries.

Let us consider the purely AdS or dS part of our trajectory (i.e. before or after the uplift steps in our deformation trajectory), and work with the action:

$$S = \frac{1}{16\pi G_N}\int \sqrt{-g}\left(R - 2\frac{s}{\ell^2}\right) - \int \sqrt{-g}\left(\frac{1}{2}(\nabla\phi)^2 + V(\phi)\right) + S_{\text{bry}} \,. \tag{6.7}$$

On the gravity side, we can work at the probe order, where the bulk quantum fields propagate on the un-back-reacted bounded geometry, up to corrections suppressed by powers of G_N times gradients. In order to capture the Brown-York stress-energy

$$T_{ij} = -\frac{1}{8\pi G_N}\left(K_{ij} - \left(K - \frac{1}{\ell}\right)\gamma_{ij}\right) \tag{6.8}$$

in this probe regime, we do need to calculate the order G_N back reaction on the metric. The system of probe scalar and back reacted metric have simple standing wave solutions satisfying fixed boundary conditions. We stress again that in the general time-dependent setting with local bulk matter, Changing L_c is not the same as moving a cylinder within a solution for the original L_c to adopt as a new boundary, since the quantization of modes depends on L_c. In other words, we are not 'moving in the boundary' in a pre-existing solution, but are instead changing its proper size.

We can organize the calculation in an expansion in G_N times the energy scale of the bulk EFT matter, and will work in the leading nontrivial order in this expansion, as follows. We treat the bulk scalar as a probe of the original geometry, but we include the back reaction on gravity necessary to capture the order 1 contributions to both the Brown-York energy and the generalized entropy. That is, we compute the order G_N contribution to the extrinsic curvature K at our boundary, and the order G_N contribution to the horizon area $\mathcal{A}$.

[23] We thank Edgar Shaghoulian for many helpful discussions on this topic.

In this section, we summarize the calculation of the metric back-reaction.[24] We consider for simplicity backreacted configurations that preserve a $U(1)$ symmetry around the θ circle, and we display here an explicit class of solutions on bounded patch that includes the de Sitter observer, imposing Dirichlet conditions at the boundary of this patch. We work with the metric ansatz of the form:

$$ds_3^2 = \frac{4}{(1+s\frac{\rho^2}{\ell^2})^2}\left(-\rho^2 B e^{-2D}dt^2 + B^{-1}d\rho^2 + \frac{\ell^2}{4}\left(1-s\frac{\rho^2}{\ell^2}\right)^2 d\theta^2\right). \tag{6.9}$$

When $B=1$ and $D=0$, this metric corresponds to BTZ, Rindler, and the static patch for $s=-1,0$, and $+1$ respectively. In these coordinates, the pole of the static patch is at $\rho=\ell$, and the horizon is at $\rho=0$. We will study perturbations around the background geometry including those sourced by matter excitations by writing (with $\kappa^2 = 16\pi G_N$)

$$B = 1+\kappa^2\Delta B(t,\rho)\,, \qquad D = 0+\kappa^2\Delta D(t,\rho)\,. \tag{6.10}$$

Before we proceed to the calculations of the back-reactions ΔB and ΔD, we note that the Brown-York energy of the pole patch is given as

$$T_t^t = \frac{-1}{8\pi G_N \ell}\left(1 \pm \frac{2s\rho_c\ell}{\ell^2 - s\rho_c^2}\sqrt{1+\kappa^2\Delta B}\right) \tag{6.11}$$

for a boundary situated at $\rho=\rho_c$. The first factor in the second term corresponds to the square root part of the dressed energy (4.1) in the $T\bar{T}+\Lambda_2$ deformed theory. We can see this as follows. The relationship between the ρ coordinate in (6.9) and the radial r coordinate in the usual BTZ/static patch metric,

$$ds^2_{\text{BTZ/Static}} = -\frac{s(r_h^2-r^2)}{\ell^2}d\tau^2 + \frac{s\ell^2}{r_h^2-r^2}dr^2 + r^2 d\theta^2\,, \tag{6.12}$$

is given as

$$\frac{r}{r_h} = \frac{\ell^2+\rho^2}{\ell^2-\rho^2}, \tag{6.13}$$

and applying the dictionary (3.2) we have, for example for the BTZ black hole at the Hawking-Page level ($r_h=\ell$ and $\Delta = c/6$),

$$\frac{2\rho_c\ell}{\ell^2-s\rho_c^2} \Leftrightarrow \sqrt{1-\frac{4\pi^2\ell^2}{L_c^2}} = \sqrt{1-4\pi^2\frac{8G_N\ell}{L_c^2}\frac{\ell}{8G_N}} \Leftrightarrow \sqrt{1-4\pi^2 y\left(\frac{c}{12}\right)}\,. \tag{6.14}$$

The last expression is the square root in the dressed energy formula for $\Delta = c/6$. Therefore, from (6.11), we see that the effect of the matter multiplies the original square root and does not affect the reality of the spectrum at the level of the classical solutions, as we advertised earlier.

In the following, we will solve the linearized equation in specific situations. Before doing that, we discuss here their general structure. Upon linearization, we have 1 equation for the scalar (E_ϕ) and we organize the Einstein equations as the Hamiltonian constraint ($\mathcal{H}_0 \equiv E_{0,0}$), the 2 momentum constraints ($\mathcal{H}_i \equiv E_{0,i}$), one which is automatically zero in our ansatz because of translation symmetry along the circle, and the remaining Einstein equations $E_{\rho,\rho}$ and $E_{\theta,\theta}$, so that we have 5 total equations.

[24] Our analysis generalizes the linearized part of the analysis of [71] to the case with finite timelike boundary in (A)dS.

The Hamiltonian constraint is preserved by time evolution. Indeed, combining the linearized equations one can see that there exist $c_i(\rho)$ such that

$$\partial_t \mathcal{H}_0 = c_1(\rho)\partial_\rho \mathcal{H}_1 + c_2(\rho)\mathcal{H}_1 + c_3(\rho)E_\phi. \tag{6.15}$$

Thus, we only need to enforce $\mathcal{H}_0$ at the initial time. In addition, for $s \neq 0$, $E_{\theta,\theta}$ is redundant, being a combination of the other other ones, their derivatives, and the equation for the scalar. Summing up, our linearized system is composed by the equation for the scalar (E_ϕ), a constraint equation for the metric initial data ($\mathcal{H}_0$), and three other equations.

6.2.1 Scalar solutions and bakcreactions in the Pole Patch

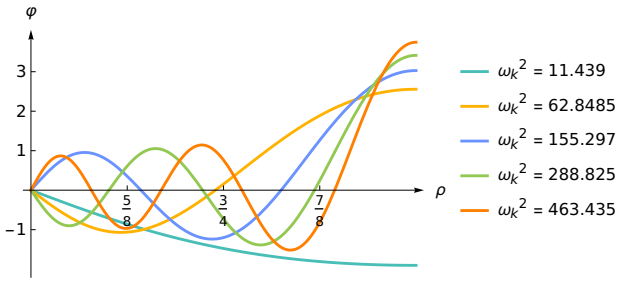
We start by considering fluctuations of the scalar around the background, and we choose $s = +1$ and work with the pole patch $\rho \in [\rho_c, \ell]$ in this section. Due to linearity of the scalar equation, we can decompose a generic scalar fluctuation as a superposition of modes

$$\phi_1(t,\rho) \equiv \sum_k c_k e^{i\omega_k t}\varphi_k(\rho) \tag{6.16}$$

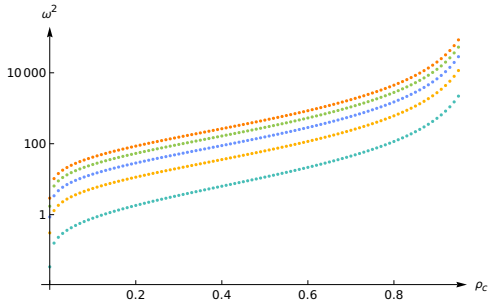
and we impose regularity at the pole by requiring $\varphi'(\ell) = 0$ and the Dirichlet condition at $\rho = \rho_c$ by setting $\varphi_(\rho_c) = 0$. With these conditions, the set of frequencies ω_k^2 is discrete and dependent on the cylinder size $L_c = L(\rho_c)$. We stress here again the difference between our prescription and 'bringing in the boundary' in a pre-existing solution. Fixing the proper size L_c of the cylinder for each point along our trajectory leads to new solutions with different scalar frequencies as we evolve along the trajectory by changing $y = 8\ell G_N/L_c^2$. Explicitly, the equation of motion for a minimally-coupled massless scalar reduces to the eigenvalue problem

$$-\partial_\rho\left(\rho\frac{\rho^2-\ell^2}{\rho^2+\ell^2}\partial_\rho\varphi_k\right) = \omega_k^2\frac{\rho^2-\ell^2}{\rho(\ell^2+\rho^2)}\varphi_k \tag{6.17}$$

with boundary conditions $\partial_\rho\varphi_k(\ell) = 0$ (regularity) and $\varphi_k(\rho_c) = 0$ (Dirichlet). Some examples of solutions are shown in Fig. 3.



(a) First 5 scalar modes for $\rho_c = \frac{1}{2}$



(b) Behavior of the first 5 scalar frequencies as function of ρ_c (which determines L_c in the metric (6.9)).

Figure 3: Example modes solving the scalar equation (6.17) for $\ell = 1$.

We can now turn to the equations for the metric backreaction. We have one constraint equation plus two equations for the metric perturbations ΔB and ΔD. Explicitly, they read

$$\partial_\rho\left(\frac{\rho^2}{(\ell^2+\rho^2)^2}\Delta B\right)=\frac{\left(\ell^2-\rho^2\right)\left(\rho^2\left(\phi_1'\right)^2+\dot{\phi}_1^{\,2}\right)}{8\ell^2\rho\left(\ell^2+\rho^2\right)}=\frac{\mathcal{E}^\phi}{4\ell^3} \tag{6.18a}$$

$$\partial_t\left(\frac{\rho^2}{(\ell^2+\rho^2)^2}\Delta B\right)=\frac{\rho\left(\ell^2-\rho^2\right)\phi_1'\dot{\phi}_1}{4\ell^2\left(\ell^2+\rho^2\right)} \tag{6.18b}$$

$$\partial_\rho\Delta D=\frac{\left(\ell^4-\rho^4\right)\left(\rho^2\left(\phi_1'\right)^2+\dot{\phi}_1^{\,2}\right)}{8\ell^2\rho^3}=\frac{(\ell^2+\rho^2)^2}{4\ell^3\rho^2}\mathcal{E}^\phi\,. \tag{6.18c}$$

Here $\mathcal{E}$ is related to the energy density of the scalar field,

$$\mathcal{E}^\phi=\sqrt{-g}u^\mu u^\nu\mathcal{T}^\phi_{\mu\nu}\,,\qquad \mathcal{T}^\phi_{\mu\nu}\equiv-\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}}=\nabla_\mu\phi\nabla_\nu\phi-\frac{1}{2}g_{\mu\nu}\left(2V+(\nabla\phi)^2\right) \tag{6.19}$$

where u^μ is a unit future-pointing time-like normal vector.

The first equation is the Hamiltonian constraint, and once imposed at $t=0$, it is automathically solved for any t, provided the other equations are solved. To display an explicit backreacted solution, we now choose as initial condition for the scalar $\phi_1(0,\rho)=0$ and we consider single mode solutions. That is, we write the scalar perturbation as

$$\phi_1(t,\rho)\equiv\sin(\omega_k t)\varphi_k(\rho), \tag{6.20}$$

for a given fixed pair (ω_k,φ_k).

Since the appearance of the scalar in the metric equations of motion is quadratic, a single mode will give rise to a piece constant in time and a piece proportional to $\sin^2(\omega_k t)$. We can thus parametrize the backreaction as

$$\Delta B(t,\rho)\equiv\Delta B_0(\rho)+\sin^2(\omega_k t)\Delta B_1(\rho) \tag{6.21}$$

$$\Delta D(t,\rho)\equiv\Delta D_0(\rho)+\sin^2(\omega_k t)\Delta D_1(\rho) \tag{6.22}$$

Plugging in the equations of motion (6.18) we find

$$\Delta B_1=\frac{\left(\ell^4-\rho^4\right)\varphi_k\varphi_k'}{8\ell^2\rho} \tag{6.23}$$

and the system of ODEs

$$\partial_\rho\left(\frac{\rho^2}{(\ell^2+\rho^2)^2}\Delta B_0\right)=\frac{\varphi_k^2\omega_k^2(\ell-\rho)(\ell+\rho)}{8\ell^2\rho\left(\ell^2+\rho^2\right)} \tag{6.24a}$$

$$\Delta D_0'=\frac{\varphi_k^2\omega_k^2\left(\ell^4-\rho^4\right)}{8\ell^2\rho^3} \tag{6.24b}$$

$$\Delta D_1'=-\frac{\left(\ell^4-\rho^4\right)\left(\rho^2\left(\varphi_k'\right)^2-\varphi_k^2\omega_k^2\right)}{8\ell^2\rho^3} \tag{6.24c}$$

When solving these equations, we impose regularity of the solution at the pole

$$\Delta B_0(t,\ell) = \Delta B_1(t,\ell) = 0. \tag{6.25}$$

Finally, to enforce the Dirichlet condition at $\rho = \rho_c$, we also impose

$$\Delta g_{tt}|_{\rho=\rho_c} = 0 \quad \Leftrightarrow \quad \Delta B_0(t,\rho_c) = 2\Delta D_0(t,\rho_c),\ \Delta B_1(t,\rho_c) = 2\Delta D_1(t,\rho_c) \tag{6.26}$$

Given (6.23), the regularity of ΔB_1 at the pole is automatically satisfied if the scalar field is regular. There are three integration constants to fix when solving the ODE's (6.24). The regularity of ΔB_0 at the pole fixes one. Imposing (6.26) fixes the remaining two integration constants. After imposing the regularity condition and the Dirichlet condition for the time-dependent part of Δg_{tt} (setting $\Delta B_1 = 2\Delta D_1$ at $\rho = \rho_c$), we get

$$\Delta B_1 = \frac{\left(\ell^4 - \rho^4\right)\varphi(\rho)\varphi'(\rho)}{8\ell^2\rho} \tag{6.27a}$$

$$\Delta B_0 = \frac{(\ell^2+\rho^2)^2}{\rho^2}\int_\ell^\rho \frac{\omega_k^2\left(\ell^2-\tilde\rho^2\right)\varphi(\tilde\rho)^2}{8\ell^2\tilde\rho\,(\ell^2+\tilde\rho^2)}d\tilde\rho \tag{6.27b}$$

$$\Delta D_1 = \int_{\rho_c}^\rho \frac{(\ell^2-\tilde\rho^2)^2\varphi(\tilde\rho)\varphi'(\tilde\rho)}{4\ell^2\tilde\rho^2}d\tilde\rho + \frac{\left(\ell^4-\rho^4\right)\varphi(\rho)\varphi'(\rho)}{8\ell^2\rho} \tag{6.27c}$$

$$\Delta D_0 = \int_{\rho_c}^\rho \frac{\omega_k^2\varphi(\tilde\rho)^2\left(\ell^4-\tilde\rho^4\right)}{8\ell^2\tilde\rho^3}d\tilde\rho + \gamma_0 \tag{6.27d}$$

as the solutions to the equations (6.24). Fully imposing the Dirichlet condition (6.26) then fixes the remaining integration constant as

$$\gamma_0 = \frac{(\ell^2+\rho_c^2)^2}{2\rho_c^2}\int_\ell^{\rho_c}\frac{\omega_k^2(\ell^2-\tilde\rho^2)\varphi(\tilde\rho)^2}{8\ell^2\tilde\rho(\ell^2+\tilde\rho^2)}d\tilde\rho\,, \tag{6.28}$$

completely determining the solutions. Fig. 4 shows some plots of the metric back-reaction for the first three modes in this case.

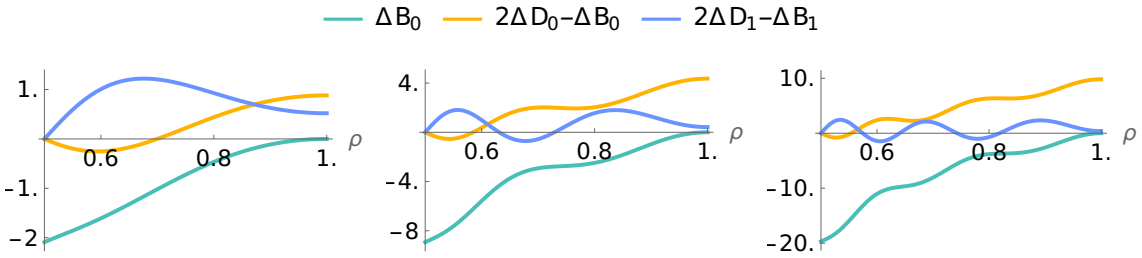


Figure 4: Metric backreaction with Dirichlet boundary conditions for the first three scalar modes ($\rho_c = 0.5, \ell = 1$).

We can compare the time-averaged change in Brown-York pressure and energy on these back-reacted solutions, in order to obtain an equation of state. Explicitly, the change in the (t,t) and (θ,θ)

components of the BY tensor after taking backreaction into account is given as

$$\Delta T^t_t = \frac{-2\rho_c \Delta B(\rho_c)}{\ell^2 - \rho_c^2} \tag{6.29}$$

$$\Delta T^\theta_\theta = \frac{1}{2\ell^2\rho_c}\left((\ell^2-\rho_c^2)\Delta B(\rho_c) + \rho_c(\ell^2+\rho_c^2)(\Delta B'(\rho_c) - 2\Delta D'(\rho_c))\right) . \tag{6.30}$$

Using the Einstein's equations (6.18a) and (6.18c), we have

$$\rho_c(\ell^2+\rho_c^2)\left(\Delta B'(\rho_c) - 2\Delta D'(\rho_c)\right) = -2(\ell^2-\rho_c^2)\Delta B(\rho_c) - \frac{(\ell^2-\rho_c^2)(\ell^2+\rho_c^2)^2}{16\ell^4\rho_c}\phi'(\rho_c)^2 . \tag{6.31}$$

Combining the results, we can relate the order-1 change in the BY pressure and the energy density as

$$\Delta T^\theta_\theta = \frac{(\ell^2-\rho_c^2)^2}{4\ell^2\rho_c^2}\Delta T^t_t - \frac{(\ell^2-\rho_c^2)(\ell^2+\rho_c^2)^2}{16\ell^4\rho_c}\phi'(\rho_c)^2 . \tag{6.32}$$

We consider a single-mode scalar with $\phi = \sin(\omega t)\varphi(\rho)$ and time-average the above equation over a period ($\int_0^{2\pi/\omega} dt$) to get

$$\langle \Delta T^\theta_\theta \rangle = \frac{(\ell^2-\rho_c^2)^2}{4\ell^2\rho_c^2}\langle \Delta T^t_t \rangle - \frac{\pi(\ell^2-\rho_c^2)(\ell^2+\rho_c^2)^2}{16\ell^4\rho_c\omega_k}\varphi'(\rho_c)^2 \tag{6.33}$$

The terms $\langle \Delta T^\theta_\theta \rangle$ and $\langle \Delta T^t_t \rangle$ are proportional to ω, as a consequence of the relevant metric backreaction being proportional to ω^2, as seen from (6.27), and of the time-averaging introducing an ω^{-1} factor. For large ω, the final term in (6.33) is also proportional to ω , by virtue of the fact that $\varphi'(\rho_c) \propto \omega$. This latter statement can be understood from the large frequency behavior of the scalar modes, which for large ω approach the flat space oscillatory behavior $\sim e^{i\omega x}$, or numerically by using the explicit mode solutions of (6.17). This assumes a common normalization of the solutions to (6.17). Regardless of the normalization we can consider ratios, and combining the arguments above we obtain the equation of state of the form

$$\frac{\langle \Delta T^\theta_\theta \rangle}{\langle \Delta T^t_t \rangle} \xrightarrow{k\gg 1} f(\rho_c/\ell) , \tag{6.34}$$

where k denotes the index of the mode, with frequency ω_k. This relation realizes the case discussed above in boundary language in §6.1.

6.3 First law of thermodynamics for the Cosmic Horizon patch

In this section, we study the back-reaction in the cosmic horizon patch and verify that the solutions satisfy the first law of thermodynamics (6.6) in [9, 10] at the probe level. We modify the metric ansatz a little by writing the back-reaction in terms of $f_1(t,\rho), f_2(t,\rho)$ defined as

$$B(t,\rho) = f_1(t,\rho)\left(\frac{(\ell^2+s\rho^2)^2}{4\rho^2\ell^2}\right), \quad D(t,\rho) = \frac{1}{2}\log\left(\frac{f_1(t,\rho)}{f_2(t,\rho)}\right) , \tag{6.35}$$

in terms of which the metric ansatz reads

$$ds^2 = -s\ell^2 f_2 dt^2 + \frac{16\ell^6\rho^2}{(\ell^2+s\rho^2)^4 f_1}d\rho^2 + \ell^2\frac{(\ell^2-s\rho^2)^2}{(\ell^2+s\rho^2)^2}d\theta^2 . \tag{6.36}$$

On the background, we have

$$f_2^{(0)} = 1 - \frac{(\ell^2 - s\rho^2)^2}{(\ell^2 + s\rho^2)^2}, \quad f_1^{(0)} = s f_2^{(0)}\,, \tag{6.37}$$

and the zero-th order solution describes the static patch/BTZ for $s = +1/-1$ as before. To work with the cosmic horizon patch, we set $s = 1$ and $\rho \in [0, \rho_c]$, and impose Dirichlet boundary conditions for the metric back-reaction at $\rho = \rho_c$. We look for linearized solutions of the form

$$f_1 = f_1^{(0)} + \kappa^2 \Delta f_1, \quad f_2 = f_2^{(0)} + \kappa^2 \Delta f_2\,. \tag{6.38}$$

The change in the horizon entropy $\Delta S = \Delta A/4G_N$ is captured by a shift in the position of the horizon from $\rho = 0$ to $\rho = \rho_h$. To find this shift, we need a solution for Δf_1, which can be found by integrating

$$\partial_\rho \Delta f_1 = \frac{(\ell^2 - \rho^2)(\rho^2 \phi'^2 + \dot{\phi}^2)}{2\rho(\ell^2 + \rho^2)} = \frac{\mathcal{E}^\phi}{\ell}\,. \tag{6.39}$$

Vanishing of $g^{\rho\rho}$ at the new horizon ρ_h means that we have

$$f_1(t, \rho_h) = \frac{4\ell^2 \rho_h^2}{(\ell^2 + \rho_h^2)^2} + \kappa^2 \Delta f_1 = 0\,. \tag{6.40}$$

The solution for Δf_1 can be written as

$$\Delta f_1(t, \rho) = \int_{\rho_c}^{\rho} \frac{\mathcal{E}^\phi}{\ell} d\tilde{\rho} + C_{f_1}\,, \tag{6.41}$$

and using this solution, (6.40) becomes

$$\frac{4\ell^2 \rho_h^2}{(\ell^2 + \rho_h^2)^2} + \kappa^2 \int_{\rho_c}^{\rho_h} \frac{\mathcal{E}^\phi}{\ell} d\rho + \kappa^2 C_{f_1} = 0\,. \tag{6.42}$$

Each term in (6.42) can be re-written using

$$T\Delta S_{hor} = \frac{T\Delta \mathcal{A}}{4G_N} = -\frac{\rho_h^2}{4G_N \ell \rho_c} \frac{\ell^2 + \rho_c^2}{\ell^2 + \rho_h^2}, \tag{6.43a}$$

$$\Delta E_{mat} = \frac{\pi(\ell^2 + \rho_c^2)}{\ell \rho_c} \int_{\rho_h}^{\rho_c} \frac{\mathcal{E}^\phi}{\ell} d\rho, \tag{6.43b}$$

$$\Delta E_{BY} = \frac{\pi C_{f_1}(\ell^2 + \rho_c^2)}{\ell \rho_c} \tag{6.43c}$$

to yield

$$\frac{\kappa^2 \ell \rho_c}{\pi(\ell^2 + \rho_c^2)} \left(-\frac{\ell^2\, T\Delta S_{hor}}{\ell^2 + \rho_h^2} - \Delta E_{mat} + \Delta E_{BY} \right) = 0\,. \tag{6.44}$$

The shift in the horizon $\rho_h^2 \sim G_N$ is suppressed in our approximation scheme, and therefore, at the probe level, the above equation implies

$$\Delta E_{BY} = \Delta E_{mat} + T\Delta S_{hor} \tag{6.45}$$

confirming that the first law of thermodynamics (6.6) holds for our solutions, with the positive sign [9, 10].

7 $T^2 + J^2 + \Lambda_2$ and a charged static patch horizon

A special time-independent case of the $T\bar{T} + \mathcal{O}^2 + \Lambda_2$ deformation is the case where $\mathcal{O}$ corresponds to a conserved $U(1)$ current J in the boundary theory. A CFT deformed this way has been discussed previously in [55] as being dual to a $U(1)$ gauge field A_μ in the bulk, with J being the corresponding conserved current. At large c $\langle J^2 \rangle$ factorizes, and one can solve for the deformed energies in closed form. Here we present this solution for $d = 2$. For $d > 2$, the deformed energies are solved for in [55].

The deformation of the Lagrangian is given by

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -2\pi T\bar{T} \mp \frac{\lambda_J}{\lambda} J\bar{J} + \frac{1-\eta}{2\pi\lambda^2}. \tag{7.1}$$

The differential equation for the deformed energies is computed using the Hellman-Feynman theorem:

$$\begin{aligned} \frac{\partial E_n}{\partial \lambda} &= \langle n | \, \partial_\lambda H \, | n \rangle & (7.2)\\ &= L\left(2\pi \langle T\bar{T} \rangle \pm \frac{\lambda_J}{\lambda} \langle J\bar{J} \rangle - \frac{1-\eta}{2\pi\lambda^2}\right) & (7.3)\\ &= -\frac{2\pi}{4}\left(E_n \frac{\partial E_n}{\partial L} + \frac{P_n^2}{L}\right) \pm \frac{Q^2}{\lambda L} - \frac{(1-\eta)L}{2\pi\lambda^2}, & (7.4) \end{aligned}$$

where we used that in Euclidean signature $H_{int} = \int d\phi \mathcal{L}_{int}$, and $\langle J\bar{J} \rangle = \langle J \rangle \langle \bar{J} \rangle = \frac{Q_1 Q_2}{L^2} = \frac{Q^2}{\lambda_J L^2}$ at large c. In terms of the dimensionless energy $\mathcal{E}_n = E_n L$ and $y = \lambda / L^2$, the solutions to the above differential equation are

$$\mathcal{E}_n(y) = \frac{1}{\pi y}\left[1 \pm \sqrt{\eta + C_1 y + 4\pi^4 J_n^2 y^2 \mp 2\pi Q^2 y \log(y)}\right], \tag{7.5}$$

with C_1 a constant. Here, we have the freedom to choose the sign in front of the first square root, and the sign in the square root is fixed by the sign of the $J\bar{J}$ term in the flow equation. The above solution differs from the pure $T\bar{T}$ solution in the extra $y \log(y)$ term inside the square root.

These deformed energies are reproduced via a bulk analysis by considering a 3d Einstein plus Maxwell theory with cosmological constant

$$S = \int d^3x \left[\frac{1}{16\pi G}\left(R + 2\frac{\eta}{\ell^2} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}\right)\right]. \tag{7.6}$$

We consider Dirichlet boundary conditions for the gauge field A_μ, so that the variation of the action with respect to A at the boundary goes to zero

$$\delta_A S_{\text{bdry}} = -\int d^3x \sqrt{-g} \partial_\mu (\delta A_\nu F^{\mu\nu}) = 0, \tag{7.7}$$

eliminating the need to add a boundary term associated to the Maxwell field to the action. The solution to the corresponding equations of motion is the Reissner-Nordstrom black hole in AdS, and the static patch with a charged horizon in dS. The metric is [72]

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2\left(d\phi^2 - \frac{4GJ}{r^2}dt\right)^2, \tag{7.8}$$

where

$$f(r) = -8GM + \eta\frac{r^2}{\ell^2} + \frac{16G^2J^2}{r^2} - 8\pi Gq^2\log\frac{r}{\ell}. \tag{7.9}$$

In $d = 2$ the electric charge renormalizes the mass, and we can write $f(r)$ in terms of the renormalized mass at some scale r_0 [72],

$$f(r) = -8GM(r_0) + \eta\frac{r^2}{\ell^2} + \frac{16G^2J^2}{r^2} - 8\pi Gq^2\log\frac{r}{r_0}, \tag{7.10}$$

$$M(r_0) = M + \pi q^2\log\frac{r_0}{\ell}. \tag{7.11}$$

To derive the deformed energies, we again use the holographic stress tensor

$$T_{ij} = \frac{-1}{8\pi G}\left(K_{ij} - Kg_{ij} + \frac{1}{l}g_{ij}\right) \tag{7.12}$$

to get

$$-T^t_t = \frac{1}{8\pi Gl}\left(1 - \frac{\ell}{r_c}f(r_c)^{1/2}\right) = \frac{1}{8\pi G\ell}\left(1 - \sqrt{\eta - 8GM\frac{\ell^2}{r_c^2} - 8\pi Gq^2\frac{\ell^2}{r_c^2}\log\left(\frac{r_c}{r_0}\right)}\right). \tag{7.13}$$

We note that this energy matches with (7.5) upon applying (3.2) and identifying $\pi^2\ell q^2 = Q^2$ to get

$$-L_c^2T^t_t = \frac{1}{\pi y}\left(1 - \sqrt{\eta - 2\pi y\mathcal{E}_n - 2\pi Q^2 y\log\left(\frac{y_0}{y}\right)}\right) \tag{7.14}$$

where we defined $y_0 = 8G\ell/(2\pi r_0)^2$ so that

$$\log\left(\frac{r_c}{r_0}\right) = \frac{1}{2}\log\left(\frac{L_c^2}{L_0^2}\right) = \frac{1}{2}\log\left(\frac{y_0}{y}\right). \tag{7.15}$$

This evaluates to a result that matches Eq. (7.5):

$$-T^t_t = \frac{1}{\pi\lambda}\left(1 - \sqrt{\eta - \ell M(r_0)\frac{\lambda}{r^2} - \pi q^2\frac{\lambda}{r^2}\log\frac{r}{r_0} + \frac{1}{4}\frac{\lambda^2J^2}{r^4}}\right) \tag{7.16}$$

8 Summary and Discussion

In this work, we have developed a concrete proposal for a holographic dual definition of weakly curved three-dimensional gravity plus matter with positive cosmological constant, in the presence of a non-gravitational timelike boundary. This finite quantum system validates the microstate count obtained from the universal and solvable $T\bar{T} + \Lambda_2$ deformation in [3] (matching [1, 2]) by embedding the latter into a theory which captures effective field theory matter with local boundary conditions.

The algorithm we presented to construct the theory prescribes brute-force tuning steps which are possible thanks to various technical developments in the field. In particular, the utility [73] of specifying theories through a differential equation extends beyond the simple solvable $T\bar{T}$ case [55], but requires extra ingredients to address bulk matter. We can work with an explicitly regulated, finite theory using the cutoff on the real spectrum from pure $T\bar{T}$ as the first, infinitesimal step in our procedure. With this regulator in place, we can tune away divergences in composite operators required to deform bulk matter boundary conditions, matching to their large-N definition (worked out in [55] and generalized here). This enables one to formulate the boundary theory as in algorithm 1. This theory matches, by construction, GR + EFT where that description applies, and defines the quantum gravity theory beyond that (including at the bounding wall, which is a UV-sensitive ingredient). Along the way, we analyzed the solution space of bulk matter interacting with gravity, illustrating their interplay and reproducing the thermodynamic relations [9, 10] appropriate to the dual as an ordinary quantum system: e.g. the first law of thermodynamics takes the standard form (6.6) in terms of the quasilocal (dressed) energy.

A key outcome is, as already emphasized, the confirmation that the Gibbons-Hawking entropy admits a precise microstate count explanation [3], in a finite quantum system with a type I Von Neumann algebra which captures the local dS geometry. Although defined concretely by algorithm 1, we do not have a closed-form expression for the Hamiltonian. Still, we found some observables – such as the effect of matter excitations with a simple equation of state on dressed energies – whose structure is calculable on both sides. More could likely be done along those lines. Moreover, as already stressed, our theory contains a core, universal sector matching de Sitter gravity that is described by the solvable $T\bar{T} + \Lambda_2$ deformation with explicit Hamiltonian (up to matter-dependent fine structure among the cosmic horizon microstates that is not visible in the gravity description). It would be interesting to connect to other proposals under development for finite quantum systems with de Sitter features such as [29, 34, 35, 74], including the method of trading fluxes in explicit compactifications for branes which reveal some of their microscopic content [75], applying that to modern examples such as [31, 56, 76].

In general, the non-gravitational timelike boundary included in this work simplifies holographic duality. It is of course essential to upgrade to four external spacetime dimensions (with uplift examples starting from AdS/CFT available in [56]). There, at least naively we lose one of our essential ingredients: the 2d $T\bar{T}$ prescription for regulating (cutting off) the theory. The obstruction lies in the graviton sector of the 4d gravity theory, which is absent in 3d. This 4d graviton sector is analogous to the bulk matter in 3d. Perhaps it is possible to define a solvable quadratic operator even in 4d which only provides a Dirichlet condition for the non-graviton gravitational degrees of freedom. As in the present work, one would obtain this regulator via a very short initial segment of the trajectory, after which the full quadratic operator [55, 77] would enter (analogously to $(\mathcal{O}\mathcal{O})_r$ in the present work). Also in 4d, the subtleties with Dirichlet boundary conditions may require switching to other options

such as conformal boundary conditions as in [12, 15, 22–24].[25]

Let us finish with one final comment. In addition to simplifying holography, manifolds with boundary are as generic (perhaps more so) than the case without. However, such holes in space might be expected to have been diluted by inflation, similarly to monopoles and other exotica. In any case, independently of theoretical motivations, the observed positive cosmological constant and other data is presumably compatible with the existence of such boundaries at some level, provided they are sufficiently small and of sufficiently low density. It might be interesting to develop signatures – such as novel types of mirrored images that are not caused by gravitational lensing – and use observations to place bounds on the presence of non-gravitational timelike boundaries within our observable horizon (in the spirit of e.g. [78, 79]).

Acknowledgements

We would like to thank D. Freedman and J. Aguilera-Damia for early collaboration on GR + EFT in bounded regions and V. Shyam, R. Soni, B. Banihashemi and E. Shaghoulian for many discussions on aspects of the duality with matter. We are grateful to O. Aharony, A. Ahmadain, D. Anninos, F. Denef, V. Gorbenko, M. Guica, R. Khan, H. Maxfield, G. Pennington, S. Shenker, D. Stanford, L. Susskind, H. Verlinde, and A. Wall for very useful discussions. This research is supported by a Simons Investigator award and National Science Foundation grant PHY-2310429. GT is supported by CONICET (PIP grant 11220200101008CO), ANPCyT (PICT 2018-2517), CNEA, and Instituto Balseiro.

References

[1] G.W. Gibbons and S.W. Hawking, *Cosmological Event Horizons, Thermodynamics, and Particle Creation*, *Phys. Rev. D* **15** (1977) 2738.

[2] D. Anninos, F. Denef, Y.T.A. Law and Z. Sun, *Quantum de Sitter horizon entropy from quasicanonical bulk, edge, sphere and topological string partition functions*, 2009.12464.

[3] E. Coleman, E.A. Mazenc, V. Shyam, E. Silverstein, R.M. Soni, G. Torroba et al., *De Sitter microstates from $T\overline{T}$ + Λ_2 and the Hawking-Page transition*, *JHEP* **07** (2022) 140 [2110.14670].

[4] V. Gorbenko, E. Silverstein and G. Torroba, *dS/dS and $T\overline{T}$*, *JHEP* **03** (2019) 085 [1811.07965].

[5] A. Lewkowycz, J. Liu, E. Silverstein and G. Torroba, *$T\overline{T}$ and EE, with implications for (A)dS subregion encodings*, *JHEP* **04** (2020) 152 [1909.13808].

[6] F.A. Smirnov and A.B. Zamolodchikov, *On space of integrable quantum field theories*, *Nucl. Phys.* **B915** (2017) 363 [1608.05499].

[7] S. Dubovsky, V. Gorbenko and G. Hernández-Chifflet, *$T\overline{T}$ partition function from topological gravity*, *JHEP* **09** (2018) 158 [1805.07386].

[8] A. Cavaglia, S. Negro, I.M. Szecsenyi and R. Tateo, *$T\bar{T}$-deformed 2D Quantum Field Theories*, *JHEP* **10** (2016) 112 [1608.05534].

[9] B. Banihashemi, T. Jacobson, A. Svesko and M. Visser, *The minus sign in the first law of de Sitter horizons*, 2208.11706.

[25] We note, however, that the Dirichlet approach to this would not require the well-definedness of Dirichlet conditions for an arbitrary boundary geometry.

[10] B. Banihashemi and T. Jacobson, *Thermodynamic ensembles with cosmological horizons*, *JHEP* **07** (2022) 042 [2204.05324].

[11] D. Marolf and M. Rangamani, *Causality and the AdS Dirichlet problem*, *JHEP* **04** (2012) 035 [1201.1233].

[12] M.T. Anderson, *On boundary value problems for Einstein metrics*, *Geom. Topol.* **12** (2008) 2009 [math/0612647].

[13] T. Andrade, W.R. Kelly, D. Marolf and J.E. Santos, *On the stability of gravity with Dirichlet walls*, *Class. Quant. Grav.* **32** (2015) 235006 [1504.07580].

[14] Z. An and M.T. Anderson, *The initial boundary value problem and quasi-local Hamiltonians in General Relativity*, 2103.15673.

[15] E. Witten, *A note on boundary conditions in Euclidean gravity*, *Rev. Math. Phys.* **33** (2021) 2140004 [1805.11559].

[16] D. Marolf and J.E. Santos, *Stability of the microcanonical ensemble in Euclidean Quantum Gravity*, *JHEP* **11** (2022) 046 [2202.12360].

[17] D. Marolf and J.E. Santos, *The canonical ensemble reloaded: the complex-stability of Euclidean quantum gravity for black holes in a box*, *JHEP* **08** (2022) 215 [2202.11786].

[18] L.V. Iliesiu, J. Kruthoff, G.J. Turiaci and H. Verlinde, *JT gravity at finite cutoff*, *SciPost Phys.* **9** (2020) 023 [2004.07242].

[19] D. Stanford and Z. Yang, *Finite-cutoff JT gravity and self-avoiding loops*, 2004.08005.

[20] P. Kraus, R. Monten and K. Roumpedakis, *Refining the cutoff 3d gravity/$T\overline{T}$ correspondence*, *JHEP* **10** (2022) 094 [2206.00674].

[21] S. Ebert, E. Hijano, P. Kraus, R. Monten and R.M. Myers, *Field Theory of Interacting Boundary Gravitons*, *SciPost Phys.* **13** (2022) 038 [2201.01780].

[22] D. Anninos, D.A. Galante and C. Maneerat, *Gravitational observatories*, *JHEP* **12** (2023) 024 [2310.08648].

[23] D. Anninos, D.A. Galante and C. Maneerat, *Cosmological Observatories*, 2402.04305.

[24] X. Liu, J.E. Santos and T. Wiseman, *New Well-Posed Boundary Conditions for Semi-Classical Euclidean Gravity*, 2402.04308.

[25] A. Ahmadain and A.C. Wall, *Off-Shell Strings I: S-matrix and Action*, 2211.08607.

[26] E. Silverstein, *Black hole to cosmic horizon microstates in string/M theory: timelike boundaries and internal averaging*, 2212.00588.

[27] D. Anninos, S.A. Hartnoll and D.M. Hofman, *Static Patch Solipsism: Conformal Symmetry of the de Sitter Worldline*, *Class. Quant. Grav.* **29** (2012) 075002 [1109.4942].

[28] D. Anninos, G.S. Ng and A. Strominger, *Future Boundary Conditions in De Sitter Space*, *JHEP* **02** (2012) 032 [1106.1175].

[29] D. Anninos, D.A. Galante and B. Mühlmann, *Finite features of quantum de Sitter space*, *Class. Quant. Grav.* **40** (2023) 025009 [2206.14146].

[30] M. Alishahiha, A. Karch, E. Silverstein and D. Tong, *The dS/dS correspondence*, *AIP Conf. Proc.* **743** (2005) 393 [hep-th/0407125].

[31] X. Dong, B. Horn, E. Silverstein and G. Torroba, *Micromanaging de Sitter holography*, *Class. Quant. Grav.* **27** (2010) 245020 [1005.5403].

[32] E. Shaghoulian and L. Susskind, *Entanglement in De Sitter space*, *JHEP* **08** (2022) 198 [2201.03603].

[33] L. Susskind, *De Sitter Space, Double-Scaled SYK, and the Separation of Scales in the Semiclassical Limit*, 2209.09999.

[34] V. Narovlansky and H. Verlinde, *Double-scaled SYK and de Sitter Holography*, 2310.16994.

[35] A.A. Rahman and L. Susskind, *Infinite Temperature is Not So Infinite: The Many Temperatures of de Sitter Space*, 2401.08555.

[36] T. Banks, *Cosmological breaking of supersymmetry?*, *Int. J. Mod. Phys. A* **16** (2001) 910 [hep-th/0007146].

[37] R. Bousso, *Positive vacuum energy and the N bound*, *JHEP* **11** (2000) 038 [hep-th/0010252].

[38] X. Dong, E. Silverstein and G. Torroba, *De sitter holography and entanglement entropy*, *JHEP* **07** (2018) 050 [1804.08623].

[39] V. Chandrasekaran, R. Longo, G. Penington and E. Witten, *An algebra of observables for de Sitter space*, *JHEP* **02** (2023) 082 [2206.10780].

[40] H. Lin and L. Susskind, *Infinite Temperature's Not So Hot*, 2206.01083.

[41] S. Baiguera, R. Berman, S. Chapman and R.C. Myers, *The cosmological switchback effect*, *JHEP* **07** (2023) 162 [2304.15008].

[42] G. Araujo-Regado, R. Khan and A.C. Wall, *Cauchy slice holography: a new AdS/CFT dictionary*, *JHEP* **03** (2023) 026 [2204.00591].

[43] D.A. Galante, *Modave lectures on de Sitter space & holography*, *PoS* **Modave2022** (2023) 003 [2306.10141].

[44] T. Hartman, C.A. Keller and B. Stoica, *Universal spectrum of 2d conformal field theory in the large c limit*, *Journal of High Energy Physics* **2014** (2014) .

[45] B. Mukhametzhanov and A. Zhiboedov, *Modular invariance, tauberian theorems and microcanonical entropy*, *JHEP* **10** (2019) 261 [1904.06359].

[46] J.D. Brown and J.W. York, Jr., *Quasilocal energy and conserved charges derived from the gravitational action*, *Phys. Rev. D* **47** (1993) 1407 [gr-qc/9209012].

[47] M. Guica and R. Monten, *$T\bar{T}$ and the mirage of a bulk cutoff*, 1906.11251.

[48] P. Kraus, J. Liu and D. Marolf, *Cutoff AdS_3 versus the $T\overline{T}$ deformation*, *JHEP* **07** (2018) 027 [1801.02714].

[49] L. McGough, M. Mezei and H. Verlinde, *Moving the CFT into the bulk with $T\overline{T}$*, *JHEP* **04** (2018) 010 [1611.03470].

[50] O. Aharony, M. Berkooz and E. Silverstein, *Multiple trace operators and nonlocal string theories*, *JHEP* **08** [hep-th/0105309].

[51] O. Aharony, M. Berkooz and E. Silverstein, *Nonlocal string theories on AdS(3) x S**3 and stable nonsupersymmetric backgrounds*, *Phys. Rev. D* **65** (2002) 106007 [hep-th/0112178].

[52] M. Berkooz, A. Sever and A. Shomer, *'Double trace' deformations, boundary conditions and space-time singularities*, *JHEP* **05** (2002) 034 [hep-th/0112264].

[53] E. Witten, *Multitrace operators, boundary conditions, and AdS / CFT correspondence*, hep-th/0112258.

[54] W. Mueck, *An Improved correspondence formula for AdS / CFT with multitrace operators*, *Phys. Lett. B* **531** (2002) 301 [hep-th/0201100].

[55] T. Hartman, J. Kruthoff, E. Shaghoulian and A. Tajdini, *Holography at finite cutoff with a T^2 deformation*, 1807.11401.

[56] G.B. De Luca, E. Silverstein and G. Torroba, *Hyperbolic compactification of M-theory and de Sitter quantum gravity*, *SciPost Phys.* **12** (2022) 083 [2104.13380].

[57] L.V. Delacretaz, V. Gorbenko and L. Senatore, *The Supersymmetric Effective Field Theory of Inflation*, *JHEP* **03** (2017) 063 [1610.04227].

[58] L. Susskind, *Entanglement and Chaos in De Sitter Holography: An SYK Example*, 2109.14104.

[59] G. Bonelli, N. Doroud and M. Zhu, *$T\overline{T}$-deformations in closed form*, *JHEP* **06** (2018) 149 [1804.10967].

[60] P. Cooper, S. Dubovsky and A. Mohsen, *Ultraviolet complete Lorentz-invariant theory with superluminal signal propagation*, *Phys. Rev. D* **89** (2014) 084044 [1312.2021].

[61] R. Bousso, A. Maloney and A. Strominger, *Conformal vacua and entropy in de sitter space*, *Physical Review D* **65** (2002) .

[62] O. Aharony and T. Vaknin, *The TT^* deformation at large central charge*, *JHEP* **05** (2018) 166 [1803.00100].

[63] J. de Boer, E.P. Verlinde and H.L. Verlinde, *On the holographic renormalization group*, *JHEP* **08** (2000) 003 [hep-th/9912012].

[64] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*, Cambridge University Press (2007).

[65] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond*, Cambridge Monographs on Mathematical Physics, Cambridge University Press (12, 2007), 10.1017/CBO9780511618123.

[66] A.J. Tolley, *$T\overline{T}$ deformations, massive gravity and non-critical strings*, *JHEP* **06** (2020) 050 [1911.06142].

[67] E.A. Mazenc, V. Shyam and R.M. Soni, *A $T\overline{T}$ Deformation for Curved Spacetimes from 3d Gravity*, 1912.09179.

[68] G. Torroba, *$T\overline{T}$ + Λ_2 from a 2d gravity path integral*, *JHEP* **01** (2023) 163 [2212.04512].

[69] L. Griguolo, R. Panerai, J. Papalini and D. Seminara, *Nonperturbative effects and resurgence in JT gravity at finite cutoff*, 2106.01375.

[70] A. Levine and E. Shaghoulian, *Encoding beyond cosmological horizons in de sitter jt gravity*, *Journal of High Energy Physics* **2023** (2023) .

[71] P. Bizon and A. Rostworowski, *On weakly turbulent instability of anti-de Sitter space*, *Phys. Rev. Lett.* **107** (2011) 031102 [1104.3702].

[72] C. Martínez, C. Teitelboim and J. Zanelli, *Charged rotating black hole in three spacetime dimensions*, *Physical Review D* **61** (2000) .

[73] A.B. Zamolodchikov, *Expectation value of composite field T anti-T in two-dimensional quantum field theory*, hep-th/0401146.

[74] D. Anninos and B. Mühlmann, *Matrix integrals & finite holography*, *JHEP* **06** (2021) 120 [2012.05224].

[75] E. Silverstein, *AdS and dS entropy from string junctions: or, The Function of junction conjunctions*, in *From Fields to Strings: Circumnavigating Theoretical Physics: A Conference in Tribute to Ian Kogan*, pp. 1848–1863, 8, 2003 [hep-th/0308175].

[76] M. Demirtas, M. Kim, L. McAllister, J. Moritz and A. Rios-Tascon, *Exponentially Small Cosmological Constant in String Theory*, *Phys. Rev. Lett.* **128** (2022) 011602 [2107.09065].

[77] M. Taylor, *TT deformations in general dimensions*, 1805.10287.

[78] J.R. Bond, D. Pogosian and T. Souradeep, *CMB anisotropy in compact hyperbolic universes. 2. COBE maps and limits*, *Phys. Rev. D* **62** (2000) 043006 [astro-ph/9912144].

[79] N.J. Cornish, D.N. Spergel, G.D. Starkman and E. Komatsu, *Constraining the topology of the universe*, *Phys. Rev. Lett.* **92** (2004) 201302 [astro-ph/0310233].

A New Determination of the Millisecond Pulsar Gamma-Ray Luminosity Function and Implications for the Galactic Center Gamma-Ray Excess

Ian Holst[1,2*] and Dan Hooper[1,2,3†]
[1] University of Chicago, Department of Astronomy and Astrophysics, Chicago, Illinois 60637, USA
[2] University of Chicago, Kavli Institute for Cosmological Physics, Chicago, Illinois 60637, USA and
[3] Fermi National Accelerator Laboratory, Theoretical Astrophysics Group, Batavia, Illinois 60510, USA
(Dated: March 5, 2024)

It has been suggested that the Galactic Center Gamma-Ray Excess (GCE) could be produced by a large number of centrally-located millisecond pulsars. The fact that no such pulsar population has been detected implies that these sources must be very faint and very numerous. In this study, we use the contents of Fermi's recently released Third Pulsar Catalog (3PC) to measure the luminosity function of the millisecond pulsars in the Milky Way's Disk. We find that this source population exhibits a luminosity function with a mean gamma-ray luminosity of $\langle L_\gamma \rangle \sim 6{\times}10^{32}$ erg/s (integrated above 0.1 GeV). If the GCE were generated by millisecond pulsars with the same luminosity function, we find that ~ 20 such sources from the Inner Galaxy population should have already been detected by Fermi and included in the 3PC. Given the lack of such observed sources, we exclude the hypothesis that the GCE is generated by pulsars with the same luminosity function as those in the Galactic Disk with a significance of 3.4σ. We conclude that either less than 39% of the GCE is generated by pulsars, or that the millisecond pulsars in the Inner Galaxy are at least 5 times less luminous on average than those found in the Galactic Disk.

arXiv:2403.00978v1 [astro-ph.HE] 1 Mar 2024

I. INTRODUCTION

Data collected by the Fermi Gamma-Ray Space Telescope has been used to identify a bright and statistically significant excess of GeV-scale gamma rays from the direction of the Inner Galaxy, when compared to all existing models of astrophysical diffuse emission and known point sources [1, 2] (for early work, see Refs. [3–11]). The spectrum, angular distribution, and overall normalization of this Galactic Center Gamma-Ray Excess (GCE) are each consistent with those predicted from annihilating dark matter particles.

The leading astrophysical interpretation of the GCE is that it is generated by a large population of unresolved millisecond pulsars (MSPs), distributed in a highly concentrated and approximately spherical pattern throughout the Inner Galaxy [4–6, 8, 12–15]. The primary motivation for this possibility is that the spectral shape of the GCE is similar to that observed from typical MSPs.

MSPs are old neutron stars that were spun-up to rapid rates of rotation through the transfer of angular momentum from a stellar companion [16–20]. In contrast to young pulsars, MSPs have weaker magnetic fields and therefore lose rotational kinetic energy more slowly, often remaining luminous for billions of years. The long lifetimes of MSPs have lead some to speculate that large numbers of such objects could exist in or around the Galactic Center. If a large number of pulsars were present in the Inner Galaxy, however, it would be expected that a fraction of these objects would have already been detected by Fermi as individual gamma-ray point sources (or as individual radio sources). A quantitative assessment of this problem requires a measurement of the MSP luminosity function. With this in mind, several groups have studied known MSP populations in an effort to measure or constrain their gamma-ray luminosity functions [21–26]. Under the assumption that the Inner Galaxy's MSP population has a similar luminosity function to that of the MSPs observed elsewhere, one can use the measured luminosity function of those sources to place constraints on the fraction of the GCE that could originate from unresolved MSPs.

The Fermi Collaboration recently published its Third Catalog of Gamma-Ray Pulsars (3PC) [27]. This catalog includes 138 confirmed (TS > 25) gamma-ray pulsars with millisecond-scale periods ($P < 30$ ms), of which 4 are located in a globular cluster (the remaining 134 are distributed throughout the Galactic Disk, or the "field" of the Milky Way). In addition, the 3PC also contains 45 other gamma-ray sources which are co-located with a known MSP (see Tables 5 and 6 of Ref. [27]); despite the lack of detected gamma-ray pulsations from these sources, they are likely to be MSPs. In this paper, we use the coordinates and gamma-ray fluxes of these 179 field MSPs (see Fig. 1) to measure the gamma-ray luminosity function of this source population. We then use this information to draw conclusions regarding the origin of the GCE. In particular, we find that if the GCE is generated by MSPs, the pulsars responsible for that signal must be significantly less luminous than those found within the Galactic Disk.

* holst@uchicago.edu, https://orcid.org/0000-0003-4256-3680
† dhooper@fnal.gov, http://orcid.org/0000-0001-8837-4127

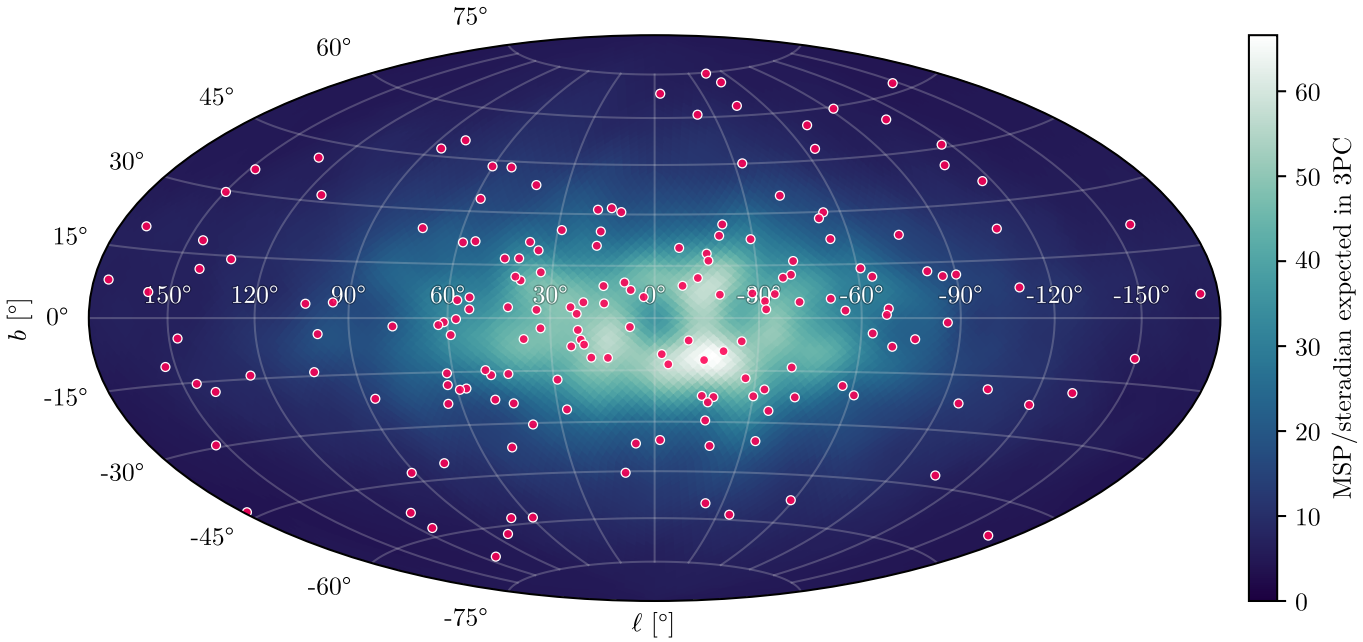


FIG. 1. The red circles represent millisecond pulsars contained within Fermi's Third Pulsar Catalog (3PC), excluding those within globular clusters [27]. This is overlaid on the MSP distribution that we predict should be observed by Fermi for our best-fit model parameters, summed over all flux bins.

II. CONSTRAINING THE LUMINOSITY FUNCTION OF THE MILKY WAY'S MILLISECOND PULSAR POPULATION

We begin by constructing a model for the spatial distribution and luminosity function of the Milky Way's MSP population. For the spatial distribution, we adopt the following parameterization [28]:

$$n_{\rm MSP} \propto e^{-|z|/z_0}\, e^{-r^2/2\sigma_R^2}, \tag{1}$$

where r and z describe the location of a pulsar in cylindrical galactocentric coordinates, and σ_R and z_0 are taken to be free parameters. In these coordinates, the Solar System is located at $r = 8.122\,{\rm kpc}$ [29] and $z = 0.0208\,{\rm kpc}$ [30].

We parameterize the gamma-ray luminosity function of MSPs using a log-normal distribution,

$$\frac{dN}{dL_\gamma} \propto \frac{1}{L_\gamma} \exp\left(-\frac{(\ln L_\gamma - \ln L_0)^2}{2\sigma_L^2}\right), \tag{2}$$

where we take L_0, σ_L, and the normalization $N_{\rm disk}$ (the total number of MSPs in the Galactic Disk) to be free parameters. Note that we will often present our results in terms of the mean gamma-ray luminosity, given by $\langle L_\gamma \rangle = L_0\, e^{\sigma_L^2/2}$. We define a pulsar's gamma-ray luminosity L_γ as the luminosity integrated above 0.1 GeV.

For a given set of model parameters, we can calculate the number of MSPs that should be included in the Fermi 3PC catalog in any direction on the sky:

$$\begin{aligned} N_{\rm 3PC} &= \int dL_\gamma \int dV\, n_{\rm MSP}\, \frac{dN}{dL_\gamma}\, P \\ &= \Delta\Omega \int dL_\gamma \int_{\rm los} dD\, D^2\, n_{\rm MSP}\, \frac{dN}{dL_\gamma}\, P. \end{aligned} \tag{3}$$

The volume integral is performed over a cone of solid angle $\Delta\Omega$, and in the second line this is related to an integral over the line-of-sight distance D, with $dV = \Delta\Omega\, D^2 dD$.

The factor P is the probability that a pulsar with flux $F_\gamma = L_\gamma/4\pi D^2$ in a given direction will be detected by Fermi and included in the 3PC catalog, taking into account Fermi's detection threshold for point sources. This threshold varies significantly with direction, being higher in directions with large backgrounds, such as along the Galactic Plane. To account for this, we use the sensitivity map as given in Fig. 25 of Ref. [27]. As explained in that paper, this map can be rescaled by a numerical constant κ to become a direction-dependent 50% completeness threshold for inclusion in the 3PC catalog. To account for source-to-source variation, we take the detection threshold to be drawn from a log-normal distribution with a width of σ_F, and centered around the reported sensitivity. We fix these parameters to values which allow us to best fit the observed pulsar flux distribution near Fermi's detection threshold, $\kappa = 5.1$ and $\sigma_F = 0.64$. Integrating over the log-normal distribution, we find:

$$P(F_\gamma) = \frac{1}{2}\left[1 + {\rm erf}\left(\frac{\ln F_\gamma - \ln(\kappa\, F_{\rm 3PC})}{\sqrt{2}\sigma_F}\right)\right], \tag{4}$$

where $F_{\rm 3PC}$ is the 3PC sensitivity in a given direction.

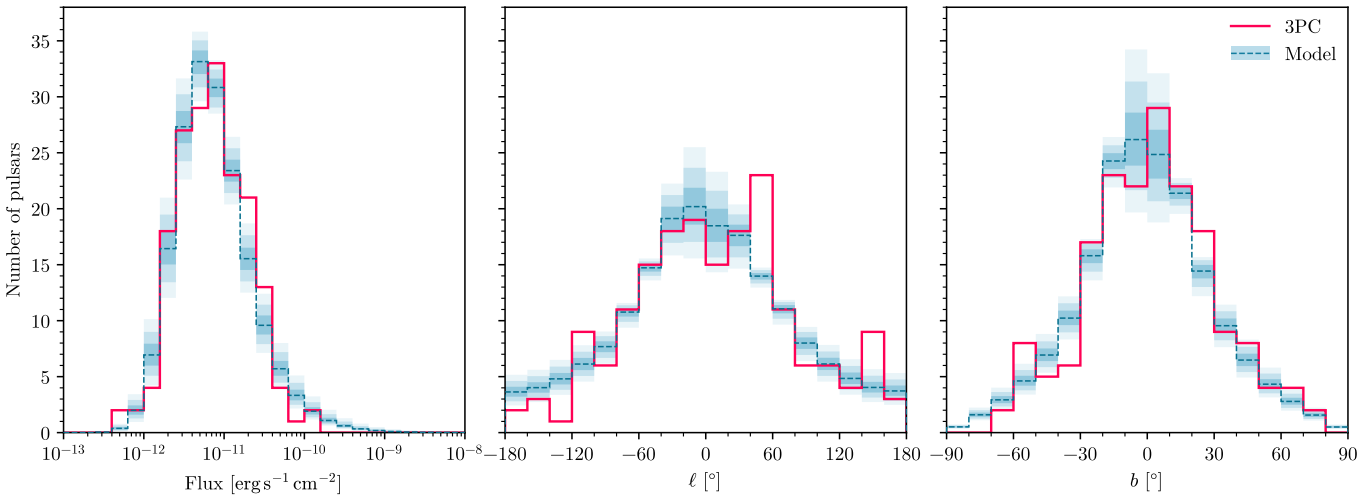


FIG. 2. The number of millisecond pulsars (MSPs) found in the Fermi Third Pulsar Catalog, in flux, longitude, and latitude bins (solid red lines). This is compared to the distributions that are predicted by our model with parameter values which lie within $2\Delta\ln\mathcal{L} = 1$, 4 or 9 of the best fit point (blue dashed line and surrounding bands).

For each set of model parameters, we compute the integral in Eq. (3) in each flux bin (using 5 flux bins per decade) and in each of the 12,288 angular pixels (a HEALPix map with `NSIDE`=32), convolving the result with a 5° radius top-hat function to reduce the impact of small-scale variations in the sensitivity map. To downsample the original 3PC sensitivity map (HEALPix with `NSIDE`=256), for a given pulsar flux F_γ, we calculate the value of P for each of the original 786,432 pixels. We then bin and average over this map to obtain a map with `NSIDE` = 32.

We then compare these predictions to the distribution of MSPs found within the 3PC catalog. The likelihood of some set of model parameters given this data is

$$\mathcal{L} = \prod_i \frac{M_i^{D_i} e^{-M_i}}{D_i!}, \tag{5}$$

where M_i and D_i denote, respectively, the numbers of MSPs predicted by our model and observed by Fermi in the ith bin (of direction and flux).

The best fit to the 3PC data is found for a luminosity function with $\langle L_\gamma \rangle \approx 5.9 \times 10^{32}$ erg/s and $\sigma_L \approx 2.3$ (corresponding to $L_0 \approx 3.9 \times 10^{31}$ erg/s). In Fig. 1, we show the locations of the 179 field MSPs contained within the 3PC overlaid on the distribution predicted for our best-fit model parameters, summed over all flux bins.

There is, however, a significant range of parameter space around this best-fit point that is consistent with the 3PC data. In Fig. 2, we show the distribution of MSPs in flux and angular bins, as found in the 3PC (solid red lines), compared to the distributions predicted for models with parameter values which yield fits which lie within $2\Delta\ln\mathcal{L} = 1$, 4 or 9 of the best fit parameter set (dashed line and shaded blue bands). In Fig. 3, we plot the regions of parameter space that are favored by our fit at the 1σ, 2σ, and 3σ confidence level (precise marginalized values are given in Table I). In the left frame of this figure, we show our results in the $\langle L_\gamma \rangle$-σ_L plane, marginalizing over the values of σ_R and z_0, while in the right frame we show the results in the σ_R-z_0 plane, marginalizing over $\langle L_\gamma \rangle$ and σ_L (in each case, $N_{\rm disk}$ is fixed to yield a total of 179 observed MSPs). Note that for 2 degrees of freedom, these contours correspond to $-2\Delta\ln\mathcal{L} = 2.30$, 6.18, and 11.83 relative to the best-fit point. Also shown in this figure are contours representing different values of $N_{\rm disk}$. In Fig. 4, we plot the MSP gamma-ray luminosity function over the range of parameters which yield $2\Delta\ln\mathcal{L}$ within 1, 2, or 9 of the best-fit point. Note that throughout this study we have restricted our luminosity function to parameter space with $\langle L_\gamma \rangle \geq 10^{30}$erg/s.[1]

III. COMPARISONS WITH PREVIOUS STUDIES

Several groups have attempted to measure the gamma-ray luminosity function of MSPs. In this section, we will discuss these previous results and compare them to the determination of the MSP luminosity function as presented in this study.

The analysis procedure followed in this work most closely resembles that carried out by Hooper &

[1] We can estimate the number of neutron stars in the Milky Way by integrating the Salpeter mass function [31, 32], $n(m) \propto m^{-2.35}$, between 8 and 25 $M_\odot$, finding that only 0.2% of all stars have an initial mass in the range required to form a neutron star. This corresponds to an estimated $\sim 10^9$ neutron stars in the disk of the Milky Way. Given that only a small fraction of neutron stars will go on to become binary pulsars, we consider models with $N_{\rm disk} \gtrsim 10^7$ to be unrealistic.

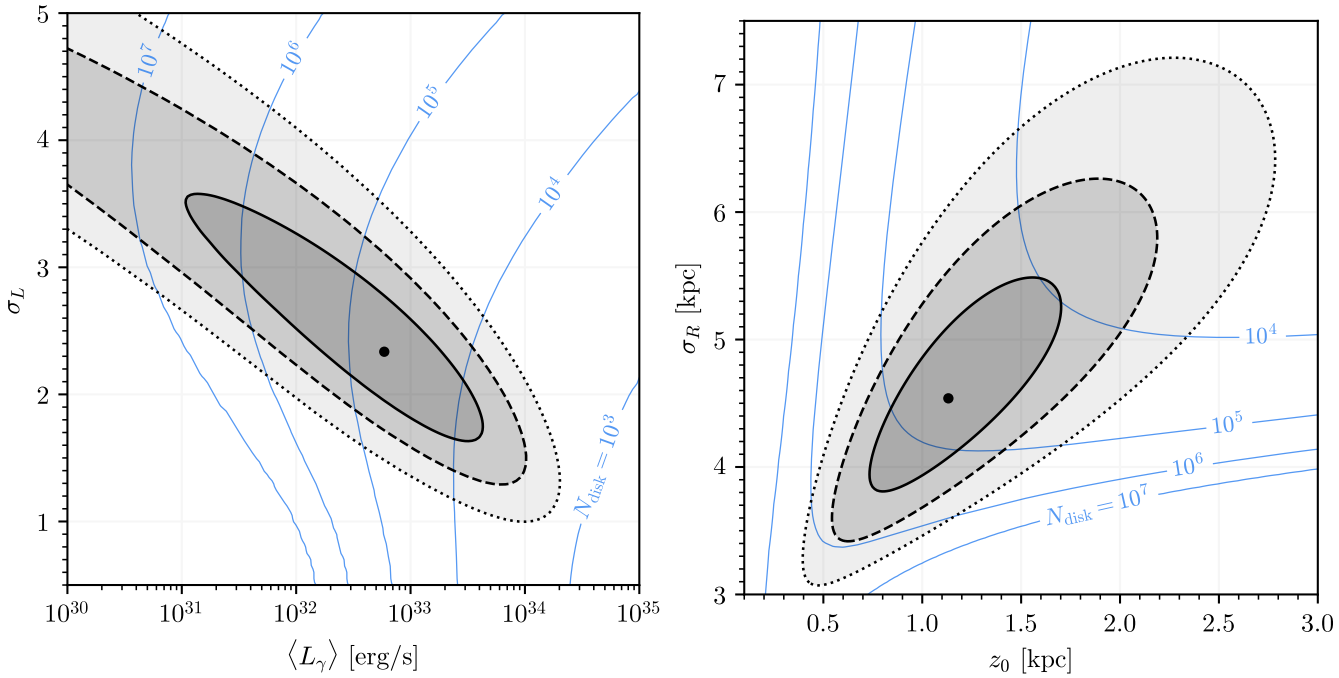


FIG. 3. Our determination of the parameters describing the gamma-ray luminosity function (left) and the spatial distribution (right) of the millisecond pulsar (MSP) population associated with the disk of the Milky Way (as opposed to those contained within globular clusters, or within our galaxy's bulge or central stellar cluster). The black dot denotes the best-fit point, while the solid, dashed, and dotted contours represent the 1σ, 2σ, and 3σ confidence intervals, respectively. In each frame, we have marginalized over the unplotted parameters. The blue contours in each frame indicate the total number of disk MSPs, $N_{\rm disk}$.

Parameter	**Best Fit**	**1σ Range**	**2σ Range**	**3σ Range**	**Unit**
$\langle L_\gamma \rangle$	5.9×10^{32}	$(0.67 - 23) \times 10^{32}$	$(0.01 - 67) \times 10^{32}$	$(0.01 - 150) \times 10^{32}$	erg/s
σ_L	2.3	$1.9 - 3.0$	$1.5 - 4.2$	$1.1 - 5.0$	–
z_0	1.1	$0.87 - 1.5$	$0.65 - 1.9$	$0.47 - 2.5$	kpc
σ_R	4.5	$4.1 - 5.1$	$3.7 - 5.8$	$3.2 - 6.7$	kpc
$N_{\rm disk}$	47,000	11,000 – 470,000	3,800 – 39,000,000	1,900 – 56,000,000	–

TABLE I. The parameter values favored by our analysis for the millisecond pulsar (MSP) luminosity function and spatial distribution.

Mohlabeng in Ref. [21]. Whereas this study utilized a dataset consisting of 179 MSPs in the field of the Milky Way, however, Ref. [21] relied on a significantly smaller sample of 66 MSPs from the 3FGL catalog [33]. As shown in Fig. 5, our results are consistent with those presented in Ref. [21], in that our preferred regions of parameter space overlap substantially.

If we had reliable measurements of the distances to each of the pulsars in our sample, we could directly relate the measured fluxes of these sources to their intrinsic luminosities. Pulsar distance determinations, however, are often subject to significant and difficult to quantify systematic uncertainties. In particular, most known MSPs do not have measured values of their parallax. For the bulk of pulsars, distance estimates are instead derived from measurements of their dispersion measure (*ie.*, the frequency-dependent time delay of a pulsar's radio pulses, which can be used to determine the the column density of free electrons along the pulsar's line-of-sight). When a pulsar's dispersion measure is combined with a model of the free electron distribution, this information can be used to produce a distance estimate. Such estimates can vary significantly, however, depending on the free electron distribution model that is adopted (for example, those of Refs. [34, 35]). Such distance estimates have sometimes been found to disagree with more reliable parallax determinations by factors of up to ~ 3 (see Ref. [27] and references therein).

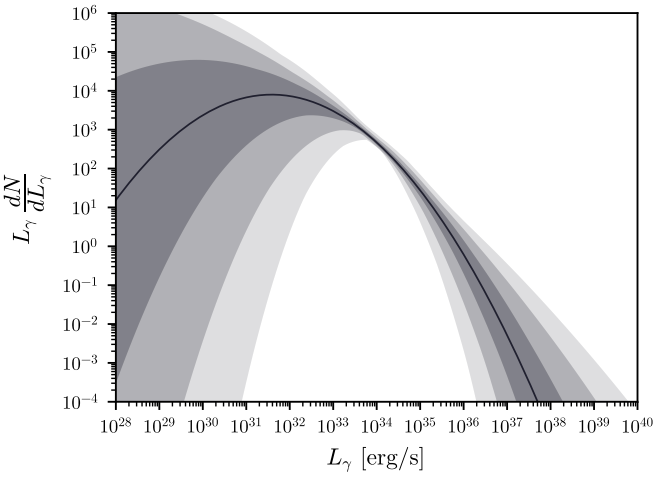


FIG. 4. The range of millisecond pulsar (MSP) gamma-ray luminosity functions that are consistent (at the 1, 2, and 3σ level) with the contents of Fermi's Third Pulsar Catalog (3PC).

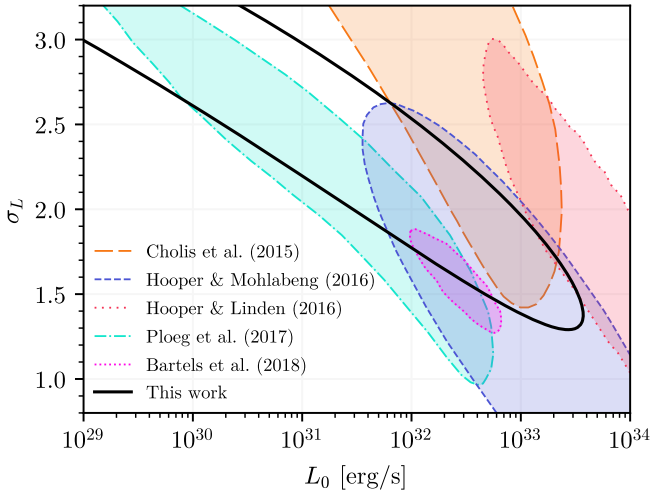


FIG. 5. A comparison of our determination of the millisecond pulsar (MSP) gamma-ray luminosity function (solid black) to those presented in earlier studies [21–24, 26], each at the 2σ confidence level. The luminosity function found here is broadly consistent with these earlier results. Note that the study of Hooper & Linden [23] was based on observations of millisecond pulsars in globular clusters, which could (in principle) feature a different luminosity function than pulsars in the Disk of the Milky Way.

In light of these considerations, we have not used any pulsar distance measurements in our main analysis. In contrast, the earlier studies of Cholis *et al.* [26] and Bartels *et al.* [22] each incorporated distance measurements into their determinations of the MSP luminosity function. By excluding distance information from our analysis, we are able to obtain a robust and reliable determination of the MSP luminosity function. In this sense, the approach taken here should be viewed as conservative.

In Fig. 5, we compare our results to those of several previous studies. Our luminosity function determination is consistent with the results of Hooper & Mohlabeng [21], Cholis *et al.* [26], Bartels *et al.* [22], and Ploeg *et al.* [24].[2] We also show in Fig. 5 the range of luminosity function parameters that were favored in the study of Hooper & Linden [23], which was based on observations of the gamma-ray emission from globular clusters. The difference between our results and those of that study could, in principle, be due to differences between the luminosity functions of these two pulsars populations.

IV. IMPLICATIONS FOR THE GALACTIC CENTER GAMMA-RAY EXCESS

Motivated by the goal of determining whether the GCE could be generated by a population of unresolved MSPs [4–6, 8, 13–15, 36], we adopt the range of luminosity functions identified in Sec. II, and use this information to calculate the number of MSPs in the Inner Galaxy that should have already been detected by Fermi, assuming that pulsars with the same luminosity function as field MSPs are responsible for the observed gamma-ray excess. In this calculation, we adopt a spherically symmetric distribution of Inner Galaxy pulsars, with a number density that scales as $n \propto r^{-2.4}$, as favored by the observed morphology of the GCE [1, 37]. We take this population to extend out to a distance of 3 kpc from the Galactic Center. We then normalize this population such that its total gamma-ray luminosity is equal to $L_\gamma^{\rm tot} = 2.1 \times 10^{37}\,{\rm erg/s}$, chosen to match the observed intensity of the GCE from within the innermost 3 kpc [1].

For this hypothesized pulsar population, we follow the approach described in Sec. II to calculate how many MSPs from the Inner Galaxy population should be contained within Fermi's 3PC catalog. In Fig. 6, we plot contours of the number of MSPs in the Inner Galaxy population that should be included in the 3PC, as a function of $\langle L_\gamma \rangle$ and σ_L. For our best-fit luminosity function parameters ($\langle L_\gamma \rangle = 5.9 \times 10^{32}\,{\rm erg/s}$, $\sigma_L = 2.3$), we find that 36,000 MSPs would be required to generate the total measured intensity of the GCE. On average, 22 of these MSPs should be bright enough to have been included within the 3PC catalog. Moving beyond the best-fit parameters, we find that for parameters within the 1σ (2σ) range of our analysis, the 3PC catalog should contain between $N_{\rm GCE,\,3PC} = 16-28$ $(10-35)$ Inner Galaxy MSPs (see Table II).

This prediction can be compared to the number of MSP candidates that Fermi has actually detected from the direction of the Inner Galaxy. In the 3PC catalog,

[2] As the study of Ploeg *et al.* [25] did not present its findings in terms of a log-normal distribution, their results could not be included in our Fig. 5. Broadly speaking, our results are statistically consistent with those of that study.

Parameter	Best Fit	1σ Range	2σ Range	3σ Range
$N_{\rm GCE,\,tot}$	36,000	9,300 – 320,000	3,200 – 21,000,000	1,400 – 21,000,000
$N_{\rm GCE,\,3PC}$	22	16 – 28	10 – 35	5.6 – 42
$f_{\rm GCE}$	–	< 0.21	< 0.39	< 0.74
$\langle L_\gamma \rangle_{\rm disk} \,/\, \langle L_\gamma \rangle_{\rm IG}$	–	> 14	> 4.6	> 1.6

TABLE II. The total number of millisecond pulsars (MSPs) that would be required to produce the Galactic Center Gamma-Ray Excess, $N_{\rm GCE,\,tot}$, and the number of those pulsars that should be contained in Fermi's Third Pulsar Catalog (3PC), $N_{\rm GCE,\,tot}$, for the luminosity function parameters shown in Fig. 3. As there are only 3 Inner Galaxy pulsars candidates in the 3PC, MSPs can generate the excess gamma-ray emission only if that source population consists of members that are significantly less luminous than those present in the Galactic Plane. The last two rows show constraints on the maximum fraction of the gamma-ray excess that could be generated by MSPs, $f_{\rm GCE}$, assuming a common luminosity function, and the minimum ratio of the average luminosity of disk MSPs to Inner Galaxy MSPs, $\langle L_\gamma \rangle_{\rm disk} \,/\, \langle L_\gamma \rangle_{\rm IG}$.

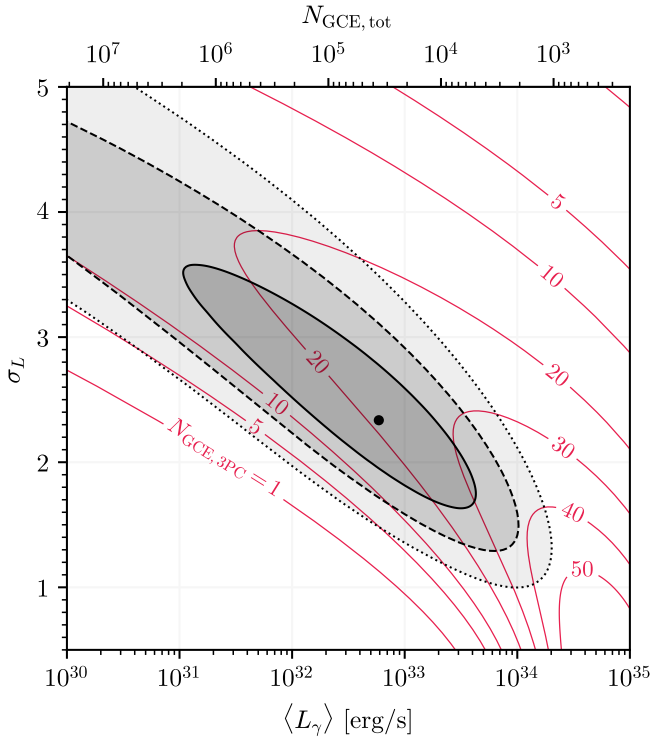


FIG. 6. As in Fig. 3, but with contours showing the number of Inner Galaxy millisecond pulsars (MSPs) that should appear in the Fermi 3PC, assuming that the Galactic Center Gamma-Ray Excess (GCE) is generated by MSPs. The top axis indicates the conversion to the total number of millisecond pulsars that would be required to generate the total luminosity of the GCE. As there are only 3 Inner Galaxy pulsars candidates in the 3PC, MSPs can generate the excess gamma-ray emission only if that source population consists of members that are significantly less luminous than those present in the Galactic Plane.

there are several such candidates with appropriate values of b and l, although most of these have distance measurements which place them much too far from the Galactic Center to be part of any source population that could be responsible for the GCE. There are, however, three known MSPs that could, in principle, be part of a central source population. These are PSR J1649-3012, PSR J1747-4036, and PSR J1833-3840, whose dispersion measures place them at distances of $d < 15.8\,$kpc, $d = 7.14 \pm 2.85\,$kpc, and $d = 4.64 \pm 1.85\,$kpc, respectively [27]. Combining this with the galactic coordinates of these pulsars, we find that they could each potentially be located within 3 kpc of the Galactic Center and be part of an Inner Galaxy population (they could also, of course, be part of the disk population).

One might further wonder whether there could be gamma-ray pulsars in the Inner Galaxy that have been detected as point sources by Fermi, but that have not yet had any pulsations detected, either in the gamma-ray or radio bands. With this in mind, we have reviewed the contents of the Fermi 4FGL catalog, finding no unassociated sources in the direction of the Inner Galaxy that have a pulsar-like spectral shape and that do not exhibit variability.[3] After applying these requirements, the only known gamma-ray sources that could be part of an Inner Galaxy MSP population are those mentioned in the previous paragraph (PSR J1649-3012, PSR J1747-4036, and PSR J1833-3840).

Combining our constraints on the MSP luminosity function with the fact that only three or fewer Inner Galaxy MSPs have been observed, we can exclude the possibility that the entire GCE is generated by a pulsar population with the same luminosity function as the field population at a level of 3.4σ. Alternatively, we can consider hypotheses in which only a fraction of the GCE is generated by MSPs. We constrain that fraction to $f_{\rm GCE} < 0.39$ (at the 95% confidence level; see Table II).

In light of the lack of Inner Galaxy MSP candidates contained in the 3PC catalog, pulsar interpretations of

[3] More precisely, for us to consider an unassociated Fermi source to be a pulsar candidate, we require its variability index to be less than 18.48, and its spectrum to be characterized by an exponentially cutoff power-law (as is the case for the vast majority of known gamma-ray pulsars).

the GCE are in significant tension with the luminosity function that we have derived here. Despite this apparent tension, the results of this study do not strictly rule out the possibility that MSPs could be responsible for the GCE. In particular, it remains possible that the MSPs in the Inner Galaxy could be systematically fainter than those found in the Galactic Disk, or in globular clusters. If the average gamma-ray luminosity of MSPs in the Inner Galaxy were $\gtrsim 5$ times lower than that of the MSPs in the Galactic Disk, this tension would be relieved (at the 95% confidence level; see Table II). Such a scenario would require $N_{\rm GCE,tot} \gtrsim 2 \times 10^5$ MSPs to produce the observed intensity of the GCE.

The most straightforward way to explain how a population of MSPs could be systematically lower in luminosity would be for it to be older (see, for example, Ref. [25]). Pulsars lose rotational kinetic energy through magnetic dipole breaking, causing their spindown power to evolve as follows:

$$\dot{E} \propto \left(1 + \frac{t}{\tau}\right)^2, \tag{6}$$

where the spindown timescale is related to the pulsar's period and magnetic field:

$$\tau \equiv \frac{E}{\dot{E}} = \frac{P}{2\dot{P}} \simeq 4.4\,{\rm Gyr} \times \left(\frac{P}{5\,{\rm ms}}\right)^2 \left(\frac{3 \times 10^8\,{\rm G}}{B}\right)^2. \tag{7}$$

Thus, over the course of several Gyr, the average luminosity of a MSP could potentially fall by a significant factor. Alternatively, it has been suggested that the Inner Galaxy population might include pulsars that were produced through accretion-induced collapse, potentially leading to a population of low-luminosity gamma-ray sources [24, 38].

In previous studies, machine learning techniques have been applied to the Fermi data in an effort to search for evidence of unresolved point sources that might contribute to the GCE [39–41]. The results of these analyses can also be used to place upper limits on the number of gamma-ray bright MSPs that are present in the Inner Galaxy. For example, the analysis of Ref. [39] finds that a significant fraction of the GCE could arise from a population of $\sim 3 \times 10^4$ point sources [39] (corresponding to $\langle L_\gamma \rangle \sim 7 \times 10^{32}$ erg/s). For this value of $\langle L_\gamma \rangle$, we would expect significantly more than three MSP candidates in the 3PC, unless $\sigma_L \lesssim 1.3$, which is well outside of the range favored by our analysis.

Our results and the results of Refs. [39–41] would each appear to be consistent with a scenario in which a relatively small fraction of the GCE ($f_{\rm GCE} \lesssim 0.39$) is generated by MSPs with our best-fit luminosity function. This scenario is also independently supported by the observed numbers of low-mass X-ray binaries in the Inner Galaxy and in globular clusters, which yields an estimate of $f_{\rm GCE} \sim 0.04 - 0.23$ [42].

V. SUMMARY AND CONCLUSIONS

Whether or not millisecond pulsars (MSPs) could potentially generate the observed Galactic Center Gamma-Ray Excess (GCE) depends critically on the luminosity function, spatial distribution, and abundance of that source population. In this study, we have revisited the MSP gamma-ray luminosity function, making use of Fermi's recently released Third Pulsar Catalog (3PC). Comparing the contents of this catalog to the predicted distribution of MSPs, we have obtained a robust determination of the gamma-ray luminosity function for the population of MSPs present in the Milky Way's Disk.

The results presented here have significant implications for the origin of the GCE. In particular, if the Inner Galaxy's MSP population has a luminosity function that is similar to that derived in this study, we predict that ~ 20 of those pulsars should have already been detected by Fermi and be contained within the 3PC catalog. In actuality, there exist only three such pulsar candidates. From these arguments, we exclude the hypothesis that the GCE is generated by pulsars with the same luminosity function as those in the Galactic Disk with a significance of 3.4σ. These results can be used to conclude (at the 95% confidence level) that either less than 39% of the GCE is generated by pulsars, or that the Inner Galaxy MSPs are, on average, at least 5 times less luminous than those found in the Galactic Disk.

Acknowledgments. We would like to thank Tim Linden, David Smith, and Toby Burnett for helpful discussions. IH is supported by generous contributions from Philip Rice. DH is supported by the Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

[1] I. Cholis, Y.-M. Zhong, S. D. McDermott, and J. P. Surdutovich, "Return of the templates: Revisiting the Galactic Center excess with multimessenger observations," *Phys. Rev. D* **105** no. 10, (2022) 103023, arXiv:2112.09706 [astro-ph.HE].

[2] M. Di Mauro, "Characteristics of the Galactic Center excess measured with 11 years of *Fermi*-LAT data," *Phys. Rev. D* **103** no. 6, (2021) 063029, arXiv:2101.04694 [astro-ph.HE].

[3] L. Goodenough and D. Hooper, "Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope," arXiv:0910.2998 [hep-ph].

[4] D. Hooper and L. Goodenough, "Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope," *Phys. Lett. B* **697** (2011) 412–428, arXiv:1010.2752 [hep-ph].

[5] D. Hooper and T. Linden, "On The Origin Of The Gamma Rays From The Galactic Center," *Phys. Rev. D*

84 (2011) 123005, arXiv:1110.0006 [astro-ph.HE].
[6] K. N. Abazajian and M. Kaplinghat, "Detection of a Gamma-Ray Source in the Galactic Center Consistent with Extended Emission from Dark Matter Annihilation and Concentrated Astrophysical Emission," *Phys. Rev. D* **86** (2012) 083511, arXiv:1207.6047 [astro-ph.HE]. [Erratum: Phys.Rev.D 87, 129902 (2013)].
[7] D. Hooper and T. R. Slatyer, "Two Emission Mechanisms in the Fermi Bubbles: A Possible Signal of Annihilating Dark Matter," *Phys. Dark Univ.* **2** (2013) 118–138, arXiv:1302.6589 [astro-ph.HE].
[8] C. Gordon and O. Macias, "Dark Matter and Pulsar Model Constraints from Galactic Center Fermi-LAT Gamma Ray Observations," *Phys. Rev. D* **88** no. 8, (2013) 083521, arXiv:1306.5725 [astro-ph.HE]. [Erratum: Phys.Rev.D 89, 049901 (2014)].
[9] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, "The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter," *Phys. Dark Univ.* **12** (2016) 1–23, arXiv:1402.6703 [astro-ph.HE].
[10] F. Calore, I. Cholis, and C. Weniger, "Background Model Systematics for the Fermi GeV Excess," *JCAP* **03** (2015) 038, arXiv:1409.0042 [astro-ph.CO].
[11] **Fermi-LAT** Collaboration, M. Ajello *et al.*, "Fermi-LAT Observations of High-Energy γ-Ray Emission Toward the Galactic Center," *Astrophys. J.* **819** no. 1, (2016) 44, arXiv:1511.02938 [astro-ph.HE].
[12] I. Cholis, D. Hooper, and T. Linden, "Challenges in Explaining the Galactic Center Gamma-Ray Excess with Millisecond Pulsars," *JCAP* **06** (2015) 043, arXiv:1407.5625 [astro-ph.HE].
[13] Q. Yuan and B. Zhang, "Millisecond pulsar interpretation of the Galactic center gamma-ray excess," *JHEAp* **3-4** (2014) 1–8, arXiv:1404.2318 [astro-ph.HE].
[14] J. Petrović, P. D. Serpico, and G. Zaharijas, "Millisecond pulsars and the Galactic Center gamma-ray excess: the importance of luminosity function and secondary emission," *JCAP* **02** (2015) 023, arXiv:1411.2980 [astro-ph.HE].
[15] T. D. Brandt and B. Kocsis, "Disrupted Globular Clusters Can Explain the Galactic Center Gamma Ray Excess," *Astrophys. J.* **812** no. 1, (2015) 15, arXiv:1507.05616 [astro-ph.HE].
[16] M. A. Alpar, A. F. Cheng, M. A. Ruderman, and J. Shaham, "A new class of radio pulsars," *Nature* **300** no. 5894, (Dec., 1982) 728–730.
[17] E. S. Phinney and S. R. Kulkarni, "Binary and Millisecond Pulsars," *Annu. Rev. Astron. Astrophys.* **32** (Jan., 1994) 591–639.
[18] D. R. Lorimer, "Binary and millisecond pulsars at the new millennium," *Living Rev. Rel.* **4** (2001) 5, arXiv:astro-ph/0104388.
[19] D. R. Lorimer, "Binary and Millisecond Pulsars," *Living Rev. Rel.* **11** (2008) 8, arXiv:0811.0762 [astro-ph].
[20] B. Kiziltan and S. E. Thorsett, "Millisecond Pulsar Ages: Implications of Binary Evolution and a Maximum Spin Limit," *Astrophys. J.* **715** no. 1, (May, 2010) 335–341, arXiv:0909.1562 [astro-ph.GA].
[21] D. Hooper and G. Mohlabeng, "The Gamma-Ray Luminosity Function of Millisecond Pulsars and Implications for the GeV Excess," *JCAP* **03** (2016) 049, arXiv:1512.04966 [astro-ph.HE].
[22] R. T. Bartels, T. D. P. Edwards, and C. Weniger, "Bayesian model comparison and analysis of the Galactic disc population of gamma-ray millisecond pulsars," *Mon. Not. Roy. Astron. Soc.* **481** no. 3, (2018) 3966–3987, arXiv:1805.11097 [astro-ph.HE].
[23] D. Hooper and T. Linden, "The Gamma-Ray Pulsar Population of Globular Clusters: Implications for the GeV Excess," *JCAP* **08** (2016) 018, arXiv:1606.09250 [astro-ph.HE].
[24] H. Ploeg, C. Gordon, R. Crocker, and O. Macias, "Consistency Between the Luminosity Function of Resolved Millisecond Pulsars and the Galactic Center Excess," *JCAP* **08** (2017) 015, arXiv:1705.00806 [astro-ph.HE].
[25] H. Ploeg, C. Gordon, R. Crocker, and O. Macias, "Comparing the Galactic Bulge and Galactic Disk Millisecond Pulsars," *JCAP* **12** (2020) 035, arXiv:2008.10821 [astro-ph.HE]. [Erratum: JCAP 07, E01 (2021)].
[26] I. Cholis, D. Hooper, and T. Linden, "A New Determination of the Spectra and Luminosity Function of Gamma-Ray Millisecond Pulsars," arXiv:1407.5583 [astro-ph.HE].
[27] **Fermi-LAT** Collaboration, D. A. Smith *et al.*, "The Third Fermi Large Area Telescope Catalog of Gamma-Ray Pulsars," *Astrophys. J.* **958** no. 2, (2023) 191, arXiv:2307.11132 [astro-ph.HE].
[28] C.-A. Faucher-Giguère and A. Loeb, "The pulsar contribution to the gamma-ray background," *JCAP* **2010** no. 1, (Jan., 2010) 005, arXiv:0904.3102 [astro-ph.HE].
[29] **GRAVITY** Collaboration, R. Abuter *et al.*, "Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole," *Astron. Astrophys.* **615** (2018) L15, arXiv:1807.09409 [astro-ph.GA].
[30] M. Bennett and J. Bovy, "Vertical waves in the solar neighbourhood in Gaia DR2," *Mon. Not. Roy. Astron. Soc.* **482** no. 1, (Jan., 2019) 1417–1425, arXiv:1809.03507 [astro-ph.GA].
[31] E. E. Salpeter, "The Luminosity Function and Stellar Evolution.," *Astrophys. J.* **121** (Jan., 1955) 161.
[32] G. Chabrier, "Galactic stellar and substellar initial mass function," *Publ. Astron. Soc. Pac.* **115** (2003) 763–796, arXiv:astro-ph/0304382.
[33] **Fermi-LAT** Collaboration, F. Acero *et al.*, "Fermi Large Area Telescope Third Source Catalog," *Astrophys. J. Suppl.* **218** no. 2, (2015) 23, arXiv:1501.02003 [astro-ph.HE].
[34] J. M. Yao, R. N. Manchester, and N. Wang, "A New Electron-density Model for Estimation of Pulsar and FRB Distances," *Astrophys. J.* **835** no. 1, (Jan., 2017) 29, arXiv:1610.09448 [astro-ph.GA].
[35] J. M. Cordes and T. J. W. Lazio, "NE2001. 1. A New model for the galactic distribution of free electrons and its fluctuations," arXiv:astro-ph/0207156.
[36] R. Bartels, S. Krishnamurthy, and C. Weniger, "Strong support for the millisecond pulsar origin of the Galactic center GeV excess," *Phys. Rev. Lett.* **116** no. 5, (2016) 051102, arXiv:1506.05104 [astro-ph.HE].
[37] Y.-M. Zhong and I. Cholis, "Robustness of the Galactic Center Excess Morphology Against Masking,"

arXiv:2401.02481 [astro-ph.HE].

[38] A. Gautam, R. M. Crocker, L. Ferrario, A. J. Ruiter, H. Ploeg, C. Gordon, and O. Macias, "Millisecond pulsars from accretion-induced collapse as the origin of the Galactic Centre gamma-ray excess signal," *Nature Astron.* **6** no. 6, (2022) 703–707, arXiv:2106.00222 [astro-ph.HE].

[39] F. List, N. L. Rodd, and G. F. Lewis, "Extracting the Galactic Center excess' source-count distribution with neural nets," *Phys. Rev. D* **104** no. 12, (2021) 123022, arXiv:2107.09070 [astro-ph.HE].

[40] F. List, N. L. Rodd, G. F. Lewis, and I. Bhat, "The GCE in a New Light: Disentangling the γ-ray Sky with Bayesian Graph Convolutional Neural Networks," *Phys. Rev. Lett.* **125** (2020) 241102, arXiv:2006.12504 [astro-ph.HE].

[41] S. Mishra-Sharma and K. Cranmer, "Neural simulation-based inference approach for characterizing the Galactic Center γ-ray excess," *Phys. Rev. D* **105** no. 6, (2022) 063017, arXiv:2110.06931 [astro-ph.HE].

[42] D. Haggard, C. Heinke, D. Hooper, and T. Linden, "Low Mass X-Ray Binaries in the Inner Galaxy: Implications for Millisecond Pulsars and the GeV Excess," *JCAP* **05** (2017) 056, arXiv:1701.02726 [astro-ph.HE].

ee∈MC: Comments on Asymmetries in QED

Ian M. Nugent*
Victoria, B.C., Canada

Abstract

In the Quantum Electrodynamics process $e^+e^-\to l^+l^-(n\gamma)$, there are two well known angular asymmetries in the $cos(\theta)$ and the $cos(\theta^*)$ distributions. In this paper, the QED angular asymmetry related to the $cos(\theta^*)$ distribution is investigated in terms of the Dirac propagator and the associated boundary conditions from which the Dirac propagator is constructed and the potential implications are examined.

Keywords: Electron-Positron Collider, Tau Lepton, Monte-Carlo Simulation

arXiv:2403.02323v1 [hep-ph] 4 Mar 2024

1 Introduction

In Quantum Electrodynamics (QED) processes, the higher order emission of hard photons directly influences the angular dependence in the differential cross-section. Consequently, for the $e^+e^-\to\mu^+\mu^-(\gamma)$ and $e^+e^-\to\tau^+\tau^-(\gamma)$ processes, this is related to the angular asymmetry between the outgoing charged leptons $cos(\theta)$ [1, 2, 3, 4, 5, 6, 7] and $cos(\theta^*)$ [8], asymmetry in the angle between the outgoing lepton and the radiated hard photon in the center-of-mass frame of the outgoing lepton pair. Higher order contributions, in particular, Feynman Diagrams with k=2, ... internal photon exchanges further modify the $cos(\theta)$ asymmetry [9][1], where the virtual and soft-photon contributions are most significant. This is in contrast to the $cos(\theta^*)$ angular distributions, where the asymmetry originates from the inclusion of Feynman Diagrams that contain n=1 or more real hard photon emissions. In $ee\in MC$ [10], each of the hard matrix elements, $\mathcal{M}_n^k$ for n real hard photon emissions and k internal photon exchanges, are explicitly calculated without approximations from the Gamma-Matrices and Dirac spinors using an object-oriented structure. This allows for the investigation of the angular asymmetry $cos(\theta^*)$ in terms of the treatment of the Dirac propagator and the associated boundary conditions.

2 Overview of ee∈MC Formalism

The ee∈MC [10] Monte-Carlo generator, is a stand-alone software program which contains the random-number generation [11, 12, 13, 14, 15], the phase-space generation [10, 16] based on the [17] algorithm modified to include embedded importance sampling [18, 19] and the theoretical models for the QED processes $e^-e^+\to\mu^+\mu^-(n\gamma)$, $e^-e^+\to\tau^+\tau^-(n\gamma)$, $e^-e^+\to hadrons(n\gamma)$ and τ lepton decays. The cross-section for the QED processes $e^-e^+\to\mu^+\mu^-(n\gamma)$, $e^-e^+\to\tau^+\tau^-(n\gamma)$ and $e^-e^+\to hadrons(n\gamma)$ is constructed within the Yennie-Frautschi-Suura (YFS) Exponentiation Formalism [20] for the infra-red subtraction

$$d\sigma=\frac{\sum_{n=0}^{\infty}\Pi_{i=0}^{n}\Pi_{j=0}^{j<i}F\left(Y_{i,j}^{\mathcal{O}(\alpha)}(P_i^\mu,P_j^\mu)\right)|\sum_{k=1}^{\infty}\bar{\mathcal{M}}_n^k|^2dPS_n}{4(|\vec{P}_{e^-}\ |E_{e^+}+E_{e^-}|\vec{P}_{e^+}|)} \tag{1}$$

where $F(x)$ is some functional form representing the resummation of all permutations for soft or virtual photon exchanges. For the Initial and Final YFS multiplicative subtraction, the function form $F(x)$ is the standard exponential YSF Form-Factor for the Yennie-Frautschi-Suura calculation [20]; the KK2F approximation [21]; and the Sudakov Form-Factor [3]. For the Full LO calculation from [22] applying corrections from [3, 10, 23], the $F(x)$ is the product of the exponential Form-Factor determined from [22] with the Coulomb potential factored out into a separate multiplicative resummation series, the Sommerfeld-Sakharov resummation factor [23]. The hard matrix elements, $\bar{\mathcal{M}}_n^k$, is determined with the spin-average-sum for an arbitrary Initial-State spin configuration [24] and is explicitly determined from the Feynman calculus corresponding to each Feynman Diagrams using an object-orientated representation of the Gamma-Matrices and Dirac Spinors [10]. Ward's Identity [6] is applied to incorporate the renormalization through the running of electromagnetic coupling constant [10, 25, 26]. Details on the simulation of the τ decays can be found in [10, 27].

3 Dirac Propagator Formalism

It is well known that the Dirac propagator can be described in terms of the retarded Green's function for a free particle with retarded boundary conditions [3],

$$G=\imath<0|\{\psi_a(x')\bar{\psi}_b(x)\}|0>\theta(x'_{(0)}-x_{(0)}). \tag{2}$$

From which it follows that:

$$\left(\imath\gamma^\mu\tfrac{\partial}{\partial x'_\mu}-m\right)G(x'-x)=\delta^4(x'-x). \tag{3}$$

*Corresponding Author
Email: inugent.physics@outlook.com

[1] These are the most significant terms in the infinite perturbative series.

Using the Fourier Transform between the momentum and the coordinate space[2, 3, 4], this can trivially be represented as:

$$G(p) = \frac{1}{\not{p}-m} = \frac{\not{p}+m}{p^2-m^2} \tag{4}$$

[2, 3, 4]. It is the retarded Green's function, Equation 4, which is generally used to describe the propagator for each internal fermion line in perturbation theory [2, 3, 4, 5, 6, 7]. In contrast, the Dirac propagator defined with the Feynman Boundary conditions is written as:

$$\begin{aligned} G(x'-x) &= \int \frac{d^4p}{(2\pi)^4} e^{-\imath p\cdot(x'-x)} G(p) \\ &= -\imath \int \frac{d^3p e^{\imath p\cdot(x'-x)}}{(2\pi)^3 2E} \Big(\left(\gamma^0 E - \gamma\cdot\mathbf{p} + m\right) e^{-\imath E(t'-t)}\theta(t'-t) \\ &\quad + \left(-\gamma^0 E - \gamma\cdot\mathbf{p} + m\right) e^{-\imath E(t-t')}\theta(t-t') \Big) \\ &= -\imath \int \frac{d^3p}{(2\pi)^3 2E} \Big(\sum_s u_a^s(p)\bar{u}_b^s e^{-\imath p\cdot(x'-x)}\theta(t'-t) \\ &\quad - \sum_s v_a^s(p)\bar{v}_b^s e^{-\imath p\cdot(x-x')}\theta(t-t') \Big) \\ &= -\imath \int \frac{d^3p}{(2\pi)^3} \Big(\frac{m}{E}\Lambda_+ e^{-\imath p\cdot(x'-x)}\theta(t'-t) \\ &\quad + \frac{m}{E}\Lambda_- e^{-\imath p\cdot(x-x')}\theta(t-t') \Big) \\ &= -\imath < 0|T\{\psi_a(x')\bar{\psi}_b(x)\}|0> . \end{aligned} \tag{5}$$

[2, 3] Λ_+ and Λ_- are the standard positive and negative projection operators for the Dirac spinors,

$$\Lambda_+ = \frac{(\not{p}+m)}{2m}, \quad \Lambda_- = \frac{(-\not{p}+m)}{2m} \tag{6}$$

[2]. This is a time-ordered solution, where the integration contour for $t'-t>0$ is in the lower half of the plane and corresponds to the positive energy solution while the integration contour for $t'-t<0$ is in the upper half plane and corresponds to the negative energy solution. The $\not{p}-m$ and $\not{p}+m$ factors project out the positive and negative contribution to the propagator. Then, given that $(\sum_s u_a^s(p)\bar{u}_b^s = \left(\not{p}+m\right)$ and $\sum_s v_a^s(p)\bar{v}_b^s = \left(\not{p}-m\right)$, and Equation 5 it follows that $G_+ = \frac{\not{p}+m}{p^2-m^2}$ and $G_- = \frac{\not{p}-m}{p^2-m^2}$ for the positive and negative energy states going forward and backward in time respectively. From the derivation of the Feynman calculus, it can be seen that before applying Wick's Theorem to obtain the non-vanishing propagators and vertices, the expectation values in the s-matrix for the perturbative expansion must be time-ordered [2]. Naïvely, this implies that the Dirac propagators should also be time-ordered. Therefore, if one applies the Green's function with Feynman boundary conditions directly to the positive energy fermion states ($u^s(p)$) and negative energy fermion states ($v^s(p)$), taking into account the time-ordered direction of the particle/anti-particle states one obtains interesting results in terms of the angular asymmetries in QED[2,3]. Figure 1 presents radiative Born plus LO cross-section dependence on the angle between the emitted γ and the outgoing lepton (μ^- or τ^-) in the center-of-mass frame for the outgoing lepton-pair, $cos(\theta^*)$. The reported asymmetry in the $cos(\theta^*)$ distribution [8, 28], is reproduced by the Dirac propagator corresponding to the retarded Green's function solution. The asymmetry is most significant at $cos(\theta^*)=\pm 1$. This region is removed in many other MC generators. The $cos(\theta^*)$ angular asymmetry is more strongly peaked in the $e^+e^-\to\mu^+\mu^-(\gamma)$ process while the $e^+e^-\to\tau^+\tau^-(\gamma)$ process is more spread out, a consequence of the larger τ mass. When the time-ordered propagator is applied for the corresponding particle and anti-particle respectively, no asymmetry is observed in the $cos(\theta^*)$ distribution. This suggests that the $cos(\theta^*)$ asymmetry in QED is directly related to the choice of the Dirac propagator and the application in the perturbation theory.

[2] At this stage we would like to remind the reader of the Feynman rules for writing down the fermion line in the QED process $f^+f^-\to f'^+f'^-$. More specifically, going right to left for incoming particles, one starts with the positive energy state going forward in time to the electro-magnetic vertex and then proceeds backwards in time for the anti-fermion line. Similarly, for the outgoing particles one starts with the anti-fermion line going backwards in time to the QED vertex, and then proceeds forward in time for the outgoing fermion line.

[3] This formulation of the Dirac propagator also has implications for the box-diagram terms. We noted that in other processes, for example the K_L K_S mass difference [2], that Feynman boundary conditions for the Dirac propagator are consistent with known results.

4 Conclusion

The $cos(\theta^*)$ asymmetries in the $e^+e^-\to\mu^+\mu^-(\gamma)$ and $e^+e^-\to\tau^+\tau^-(\gamma)$ QED interactions were investigated in terms of the choice of boundary conditions applied in the formulation of the Dirac propagators. The differential cross-section is symmetric in $cos(\theta^*)$ for the Feynman boundary conditions which incorporate the time-ordering of the positive and negative energy Dirac states when applied to the Born and LO simulation, while the retarded Green's function for the Dirac propagator has a clear asymmetry. The asymmetry is most significant at $cos(\theta^*)=\pm 1$, where due to the mass of the τ lepton is more visible away from the angular boundaries for the $e^+e^-\to\tau^+\tau^-(\gamma)$. We argue that applying the Feynman boundary conditions for the Dirac propagator is more consistent with the time-ordering from which the perturbative s-matrix is constructed.

Acknowledgement

GCC Version 4.8.5 was used for compilation and the plots are generated using the external program GNUPlot [29].

References

[1] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak, Heavy Flavour Groups. Precision Electroweak Measurements on the Z Resonance. *Physics Reports*, 427(5):257–454, 2006. doi: 10.1016/j.physrep.2005.12.006.

[2] P. Renton. *Electroweak Interactions: An Introduction to the Physics of Quarks and Leptons.* Cambridge University Press, New York, USA, 5 1990. ISBN 978-0-521-36692-2.

[3] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Addison-Wesley, Reading, USA, 1995. ISBN 978-0-201-50397-5.

[4] F. Halzen and Alan D. Martin. *Quarks And Leptons: An Introductory Course In Modern Particle Physics.* John Wiley & Sons, USA, 1 1984. ISBN 978-0-471-88741-6.

[5] L.H. Ryder. *Quantum Field Theory.* Cambridge University Press, New York, USA, 6 1996. ISBN 978-0-521-47814-4, 978-1-139-63239-3, 978-0-521-23764-2.

[6] Franz Mandl and Graham Shaw. *Quantum Field Theory: Revised Edition.* John Wiley & Sons, Great Britain, 1 1985.

[7] David J. Griffiths. *Introduction to Elementary Particle*. John Wiley & Sons, USA, 1987.

[8] J.P. Lees et al. Measurement of Initial-State/Final-State Radiation Interference in the Processes $e^+e^-\mu^+\mu^-\gamma$ and $e^+e^-\pi^+\pi^-\gamma$. *Phys. Rev. D*, 92:072015, 2015. doi: 10.1103/PhysRevD.92.072015.

[9] F. Campanario, H. Czyż, J. Gluza, M. Gunia, T. Riemann, G. Rodrigo, and V. Yundin. Complete QED NLO Contributions to the Reaction $e^+e^-\to\mu^+\mu^-\gamma$ and Their Implementation in the Event Generator PHOKHARA. *JHEP*, 02:114, 2014. doi: 10.1007/JHEP02(2014)114.

[10] I. M. Nugent. `ee∈MC`: Simulation of $\mathbf{e^+e^-}\to\mu^+\mu^-(\gamma)$ and $\mathbf{e^+e^-}\to\tau^+\tau^-(\gamma)$ Events. 2022. arXiv: 2202.02318 [hep-ph].

[11] M. Matsumoto and T. Nishimura. Mersenne Twister: a 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator. *ACM Trans. Model. Comput. Simul.*, 8:3–30, 1998. doi: 10.1145/272991.272995.

[12] T. Nishimura. Tables of 64-bit Mersenne twisters. *ACM Trans. Model. Comput. Simul.*, 10:348–357, 2000. doi: 10.1145/369534.369540.

[13] S. Harase and T. Kimoto. Implementing 64-bit Maximally Equidistributed F2-Linear Generators with Mersenne Prime Period. *ACM Trans. on Mathematical Software*, 44:1–11, 2018. doi: 10.1145/3159444.

[14] S. Vigna. An Experimental Exploration of Marsaglia's xorshift Generators, Scrambled. 2016. arXiv: 1402.6246.

[15] D. E. Knuth. *The Art of Computer Programming Vol 2 (2nd Ed.)*. Addison-Wesley, Reading, USA, 1981.

[16] I. M. Nugent. `ee∈MC`: Higher Order Radiative Emissions for $\mathbf{e^+e^-}\to\mu^+\mu^-(\mathbf{n}\gamma)$ and $\mathbf{e^+e^-}\to\tau^+\tau^-(\mathbf{n}\gamma)$ Events. 2022. arXiv: 2212.05388 [hep-ph].

[17] E. Byckling and K. Kajantie. n-Particle Phase Space in Terms of Invariant Momentum Transfer. *Nucl. Phys. B*, 9:568–576, 1969. doi: 10.1016/0550-3213(69)90271-5.

[18] Dani Gamerman and Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference 2nd ed.* Chapman & Hall/CRC Taylor & Francis Group, 2006.

[19] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis 3rd Ed.* CRC Press Taylor & Francis Group, 2014.

[20] D. R. Yennie, S. C. Frautschi, and H. Suura. The Infrared Divergence Phenomena and High-Energy Processes. *Annals of Physics*, 13:379–452, 1961. doi: 10.1016/0003-4916(61)90151-8.

[21] S. Jadach, B.F.L. Ward, and Z. Wąs. The Precision Monte Carlo Event Generator KK For Two-Fermion Final States In e^+e^- Collisions. *Comput. Phys. Commun.*, 130:260–325, 2000. doi: 10.1016/S0010-4655(00)00048-5.

[22] J. Schwinger. *Particle, Sources, and Fields Volumes I-III*. Perseus Books Publishing, L.L.C., Reading, Massachusetts, USA, 1998.

[23] B. Smith and M.B Voloshin. $e^+e^-\to\tau^+\tau^-$ at Threshold and Beyond. *Phys. Lett. B*, 324(1):117–120, 1994. doi: 10.1016/0370-2693(94)00095-6.

[24] I. M. Nugent. `ee∈MC`: Arbitrary Initial Spin States for the Production of $\mathbf{e^+e^-}\to\mu^+\mu^-(\gamma)$ and $\mathbf{e^+e^-}\to\tau^+\tau^-(\gamma)$ Events. 2022. arXiv: 2202.02318 [hep-ph].

[25] Friedrich Jegerlehner. *The Anomalous Magnetic Moment of the Muon*. Springer Tracts in Modern Physics 226, Heidelberg, Germany, 2007.

[26] Christian Sturm. Leptonic Contributions to the Effective Electromagnetic Coupling at Four-Loop Order in QED. *Nuclear Physics B*, 874(3):698–719, Sep 2013. ISSN 0550-3213. doi: 10.1016/j.nuclphysb.2013.06.009.

[27] I. M. Nugent. `ee∈MC`: Low energy Mesons and the residual QCD Potential. 2023. arXiv: 2312.01470 [hep-ph].

[28] S. Actis et al. Quest for Precision in Hadronic Cross-Sections at Low Energy: Monte Carlo Tools vs. experimental Data. *The European Physical Journal C*, 66(3-4):585–686, 2010. doi: 10.1140/epjc/s10052-010-1251-4.

[29] Thomas Williams, Colin Kelley, et al. Gnuplot 4.2: An Interactive Plotting Program. `http://gnuplot.sourceforge.net/`, 2007.

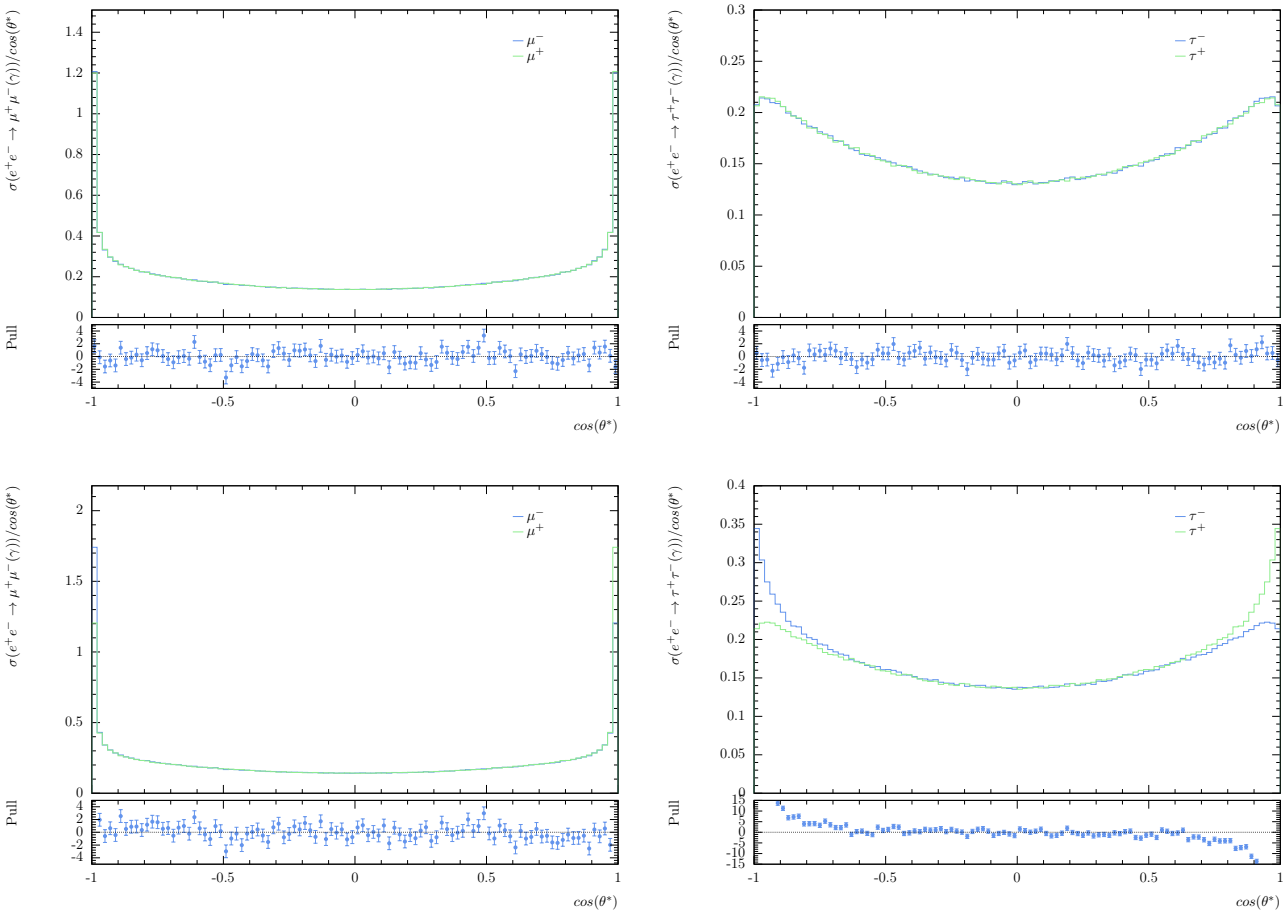


Figure 1: The $cos(\theta^*)$ distribution for Born+LO determined with the Dirac propagator constructed from the Feynman boundary conditions (top) and the retarded Dirac propagator (bottom) for the $e^+e^- \to \mu^+\mu^-(\gamma)$ (left) and $e^+e^- \to \tau^+\tau^-(\gamma)$ (right) QED interactions for a center-of-mass energy of $10.58 GeV/c^2$ and a soft-photon cut-off of E_l=$1 MeV$. The blue line represents the $cos(\theta^*)$ angle between the photon and l^- in the rest frame of the outgoing lepton pair, while the green line represents the $cos(\theta^*)$ angle between the photon and l^+ in the rest frame of the outgoing lepton pair.

arXiv:2403.02306v1 [hep-ph] 4 Mar 2024

On the study of new proxies for second order cumulants of conserved charges in heavy-ion collisions with EPOS4

Johannes Jahan,[1,2,*] Claudia Ratti,[1] Maria Stefaniak,[3,4] and Klaus Werner[2]

[1] *Department of Physics - University of Houston, Houston, TX 77204, USA*
[2] *SUBATECH, University of Nantes - IN2P3/CNRS - IMT Atlantique, Nantes, France*
[3] *Department of Physics - The Ohio State University, Columbus, OH 43210, USA*
[4] *GSI Helmholtz Centre for Heavy Ion Research, 64291 Darmstadt, Germany*

(Dated: March 5, 2024)

Proxies for cumulants of baryon number B, electric charge Q and strangeness S are usually measured in heavy-ion collisions via moments of net-number distribution of given hadronic species. Since these cumulants of conserved charges are expected to be sensitive to the existence of a critical point in the phase diagram of nuclear matter, it is crucial to ensure that the proxies used as substitutes are as close to them as possible. Hence, we use the EPOS4 framework to generate Au+Au collisions at several collision energies of the RHIC Beam Energy Scan. We compute 2$^{\text{nd}}$ order net-cumulants of π, K and p, for which experimental data has been published, as well as the corresponding conserved charge cumulants. We then compare them with proxies, defined in previous lattice QCD and Hadron Resonance Gas model studies, which are shown to reproduce more accurately their associated conserved charge cumulants. We investigate the impact of hadronic re-scatterings occurring in the late evolution of the system on these quantities, as well as the amount of signal actually originating from the bulk medium which endures a phase transition.

I. INTRODUCTION

For several decades now, heavy-ion collisions (HICs) at relativistic energies have been used to study the properties of nuclear matter, and to map the phase diagram of Quantum Chromodynamics (QCD). Since the creation of a deconfined state of quarks and gluons at very high temperature, the Quark-Gluon Plasma (QGP), has been brought to light by several experiments at the Relativistic Heavy-Ion Collider (RHIC) [1–4], understanding the nature of the transition from a hadronic phase to the QGP has been one of the main focuses in the field. On the theoretical side, first-principle calculations from lattice QCD simulations predict a crossover at low baryon chemical potential μ_B [5–7], which seems confirmed by experimental results from RHIC and the Large Hadron Collider (LHC) [8]. Several approaches predict a 1$^{\text{st}}$ order phase transition at high μ_B, which would imply the existence of a critical point on the QCD phase diagram [9–12]. The RHIC Beam Energy Scan (BES) program has thus been set up to scan a large region of the phase diagram at finite μ_B, by colliding $^{197}_{79}$Au nuclei in a wide range of center-of-mass energies. They have collected data at collision energies from $\sqrt{s_{NN}} = 200$ GeV down to 7.7 GeV in collider mode through phases I and II [13, 14], and even down to $\sqrt{s_{NN}} = 3.0$ GeV through the fixed target program [15]. Indeed, one of the main goals of this program is to search for the existence of a 1$^{\text{st}}$ order phase transition, and thus to asses the location of the hypothetical critical point.

To probe the existence of a critical point in the QCD phase diagram, susceptibilities of conserved charges such as baryon number B, electric charge Q and strangeness S are the theoretical quantities of choice, as they diverge in the vicinity of the critical point. They can be calculated as derivatives of the pressure p with respect to the chemical potential of the corresponding conserved charge [16]:

$$\chi^{BQS}_{ijk} = \frac{\partial^{i+j+k}(p/T^4)}{(\partial\hat{\mu}_B)^i(\partial\hat{\mu}_Q)^j(\partial\hat{\mu}_S)^k} \,, \tag{1}$$

where $\hat{\mu}_i = \mu_i/T$. Because the pressure is directly related to the partition function, they quantify how much the latter is modified under variations of the different chemical potentials, and are thus sensitive to radical changes in the state of nuclear matter, *e.g.* the effect of a first order phase transition or a critical point. One can also relate the susceptibilities to net-cumulants of conserved charges. For the case of second order cumulants, that will be at the center of our study, these relationships are:

$$\chi^{XY}_{11} = \frac{\langle N_X N_Y\rangle - \langle N_X\rangle\langle N_Y\rangle}{VT^3} = \frac{\sigma^{XY}_{11}}{VT^3} \,, \tag{2}$$

$$\chi^{X}_{2} = \frac{\langle {N_X}^2\rangle - \langle N_X\rangle^2}{VT^3} = \frac{\sigma^{2}_{X}}{VT^3} \,, \tag{3}$$

where σ^X_2 and σ^{XY}_{11} are respectively variance and covariance of the net-numbers of conserved charges $N_{X/Y}$ [17]. In the context of HICs, the (co)variances of net-hadron multiplicities can be used as probes for conserved charges [18]. So far, because they are the most abundant and easiest species to identify in experiments, net-proton cumulants have been used as proxies for B fluctuations [19], net-kaon cumulants have been used as proxies for S fluctuations [20], and net-multiplicity cumulants of combined identified hadrons have been used as proxies

* jahan.johannes@gmail.com

for Q fluctuations [21]. By convention, the net-numbers N_b of any baryon specie b are defined as the difference between number of baryons and number of anti-baryons $n_b - n_{\overline{b}}$, while for meson species m the net-number N_m is usually defined as the difference between positively charged mesons and negatively charged ones $n_{m^+} - n_{m^-}$.

However, using HICs to probe the phase diagram of nuclear matter comes with a lot of limitations. While on the theoretical side, most predictions are obtained from grand-canonical ensemble calculations, assuming thermal equilibrium and infinite volume, the system created in heavy-ion collisions is very short-lived, finite, small-sized and experiences rapid expansion [22]. Moreover, the volume and temperature are not accessible through measurements, and are changing during the evolution of a single event. It has been thus proposed to use ratios of net-cumulants to compare with theoretical predictions of susceptibilities, in order to get rid of the volume and temperature dependence at leading order [23]. Finally, using particles as experimental probes shows severe limitations, first of all because not all hadrons can be experimentally detected. For example, neutral hadrons can carry e.g. baryon number or strangeness (like neutrons and Λ baryons). Besides, measuring conserved charges through hadron distributions includes additional fluctuations, since particle production is a quantum process, thus probabilistic in essence.

It is hence important to find hadronic proxies that reflect conserved charge fluctuations as accurately as possible through the cumulants of their net-multiplicity distributions. In this paper, we want to address this question for 2$^{\text{nd}}$ order cumulants of conserved charges, by comparing proxies measured by STAR in [21] against new proxies inspired by [24] and with their corresponding conserved charge quantities, measured in Au+Au simulations in the RHIC beam energy range. We will employ the EPOS4 event generator [25], and take advantage of its modular design to compare the several observables of interests as measured through the final-state particle distribution with the same quantities at hadronisation time, to quantify the impact of the hadronic re-scatterings on them. Finally, as EPOS4 relies on a separate treatment of the bulk matter and high-p_T particles to model the presence of a fluid in the system, as an imitation of the QGP, we will check how much of the signal of these cumulant proxies actually originates from the deconfined medium.

In Section II, we will start by describing the different observables we will use in our work, before to introduce EPOS4 in Section III, the event generator used to produce the simulations of Au+Au collisions on which rely our study. Then, after having discussed the details of our analysis and the simulations in Section IV, we will present our results and their interpretation them in Section V, to finally summarise the conclusions of our study and mention potential outlooks in Section VI.

II. PROXIES OF CONSERVED CHARGE CUMULANTS

As mentioned in Section I, one way to probe fluctuations of conserved charges in HICs is to measure fluctuations in the net-distribution of hadronic species used as proxies. However, choosing appropriate proxies is essential to ensure that the observables we study reflect properly the corresponding conserved charge fluctuations. This is the motivation of our study. In this section, we present the commonly used experimental proxies at first, before introducing the new ones we propose because of their more accurate description of conserved charges in the context of HICs. Our new proposed proxies, taken in part from a previous work [24], will then be tested against the ones commonly measured in experiment, as well as the corresponding cumulants of conserved charges, through realistic simulations of HICs collisions generated with EPOS4.

A. STAR proxies

The experimental results on which this study is based are 2$^{\text{nd}}$ order diagonal and off-diagonal cumulants (also referred to as variances and covariances) from net-π, net-K and net-p distribution, published by the STAR collaboration in [21]. They present, for Au+Au collisions at $\sqrt{s_{NN}} = 7.7$, 11.5, 14.5, 19.6, 27, 39, 62.4 and 200 GeV, the different (co)variances:

$$\begin{pmatrix} \sigma^2_\pi & \sigma_{\pi p} & \sigma_{\pi K} \\ " & \sigma^2_p & \sigma_{pK} \\ " & " & \sigma^2_K \end{pmatrix} , \tag{4}$$

within a pseudorapidity window $|\eta| < 0.5$ and for particles with $0.4 < p_T < 1.6$ GeV. Results are notably shown as a function of the centrality class for $0 - 80\%$ centrality (through the mean number of participant nucleons $\langle N_{part} \rangle$) for each collision energy, corrected for centrality bin-width effects (CBWE) [26]. These (co)variances are then used to build proxy ratios for net-charge cumulant ratios, in order to get rid of the explicit dependence on T and V, which are not experimentally accessible event-by-event. The proxy ratios defined by STAR are the following (in parenthesis we indicate the corresponding ratio of conserved charge fluctuations they are meant to represent):

$$C_{Qp} = \frac{\sigma^{11}_{Qp}}{\sigma^2_p} \left(\cong \frac{\chi^{QB}_{11}}{\chi^B_2} \right) \ , \ C_{QK} = \frac{\sigma^{11}_{QK}}{\sigma^2_K} \left(\cong \frac{\chi^{QS}_{11}}{\chi^S_2} \right) ,$$
$$C_{pK} = \frac{\sigma^{11}_{pK}}{\sigma^2_K} \left(\cong \frac{\chi^{BS}_{11}}{\chi^S_2} \right) . \tag{5}$$

These proxy ratios are displayed as functions of $\langle N_{part} \rangle$ too, but also as functions of collision energy for most

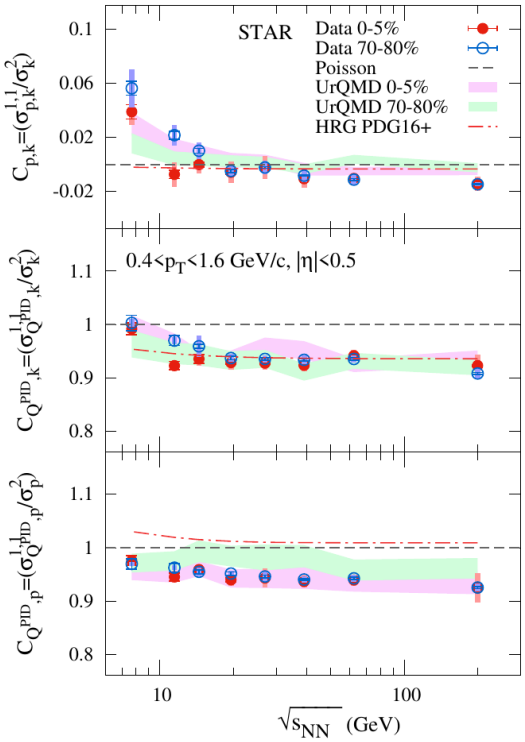


FIG. 1. Energy dependence of the cumulants ratios C_{pK}, C_{QK} for central (0-5%) and peripheral (70-80%) Au-Au collisions, within $|\eta| < 0.5$ and $0.4 < p_T < 1.6$ GeV. Results from STAR have systematic errors indicated by the boxes and statistical ones by bars. Poisson baseline, UrQMD simulations results and calculations using HRG model including the PDG16+ [29] particles set are displayed for comparison [21].

central $(0-5\%)$ and most peripheral $(70-80\%)$ collisions. These last results are of particular importance as they are expected to hint at the existence of the critical point, since decreasing the collision energy increases the baryonic chemical potential of the system created in the collision. Note that any proxy involving Q is calculated through measured net-numbers of identified charged π, K and $p/\overline{p}$, and not directly from unidentified charged particle tracks. This is done to ensure a proper estimation of detection efficiency correction, as pointed out in [27, 28].

One interesting observation from the results shown in Fig. 1 is the change of sign of the C_{pK} ratio. While it shows no centrality dependence for $\sqrt{s_{NN}} \geq 19.6$ GeV, it changes sign at lower collision energies, where it also starts to show a centrality dependence. According to [23], this could be due to a modification of the dynamics in the system, hinting for a change in the phase from which the studied correlations arise. This motivates even more the need to provide accurate proxies for conserved charge fluctuations, as this change of dynamics has also been shown through the observed disappearance of partonic collectivity in Au+Au collisions at $\sqrt{s_{NN}} = 3$ GeV [30].

In this paper, we will systematically compare our results with the data from [21], primarily as a benchmark to see if EPOS4 simulations are able to reproduce experimental results of net-multiplicity (co)variances for light hadrons. This will complement all the results from EPOS4 published so far, in particular the ones on particle production and flow for BES energies in [31, 32]. Moreover, it is also interesting to compare those data with EPOS4 simulations and look for any significant deviation, because the latter can be considered as another non-critical baseline, as will be explained in Sec. III.

B. New improved proxies

The work presented here has also been motivated by a theoretical study of proxies for fluctuations of B, Q and S [24]. In this paper, the HRG model has been used to quantify how much every different hadronic species experimentally measured contributes to the lattice QCD susceptibilities, in the region where both approaches overlap. As HRG relies on a description of QCD matter in terms of a gas of non-interacting hadrons and resonances, it is by construction quite straightforward to express susceptibilities in terms of a sum of hadronic contributions:

$$\chi^{BQS}_{ijk}(T,\vec{\hat{\mu}}) = \sum_R \sum_{h\,\in\,\text{stable}} (P_{R\to h})^l (B_h)^i (Q_h)^j (S_h)^k \times \frac{\partial^l p_h/T^4}{(\partial\hat{\mu}_B)^i(\partial\hat{\mu}_Q)^j(\partial\hat{\mu}_S)^k} \,. \quad (6)$$

In this equation, the $l = i + j + k$ (with $l = 2$ for our study). The R index refers to resonances and h to the stable hadrons these resonances can decay into. $P_{R\to h}$ represents the branching ratio for R decaying into h, and contains implicitly an additional probabilistic fluctuation term arising from resonance decays [33]. B_h, Q_h and S_h are the respective baryon number, electric charge and strangeness of hadron specie h, and p_h the pressure due to hadron specie h. Thanks to the statistical nature of the HRG model, one can easily integrate the particle distributions over only a fraction of the complete phase space, in order to reproduce the finite coverage of detectors in experiments.

By doing so, the authors of [24] have been able to construct new proxies based on the variance of net-distributions of some hadron species, that have been shown to reproduce susceptibility ratios very accurately. Some of these proxy ratios, denoted here as $\tilde{C}_{QS}$ and $\tilde{C}_{BS}(\tilde{C}'_{BS})$ and respectively probing the same quantities as the STAR proxies C_{QK} and C_{pK} from eq. (5), are

defined as:

$$\tilde{C}_{QS} = \frac{1}{2}\cdot\frac{\sigma_K^2}{\sigma_K^2+\sigma_\Lambda^2}\left(\cong\frac{\chi_{11}^{QS}}{\chi_2^S}\right), \quad (7)$$

$$\tilde{C}_{BS} = \frac{\sigma_\Lambda^2}{\sigma_K^2+\sigma_\Lambda^2} \quad / \quad (8)$$

$$\tilde{C}'_{BS} = \frac{\sigma_\Lambda^2+2\sigma_\Xi^2+3\sigma_\Omega^2}{\sigma_\Lambda^2+4\sigma_\Xi^2+9\sigma_\Omega^2+\sigma_K^2}\left(\cong\frac{-\chi_{11}^{BS}}{\chi_2^S}\right). \quad (9)$$

One can observe that these proxy ratios are only based on single hadron specie variances, and do not involve any covariance between different species. This is because cross-correlators between different conserved charges, χ_{11}^{XY}, receive most of their contribution from the variances $\sigma_{h_1}^2$ of single species h_1 carrying both charges X and Y. Only a very small fraction of χ_{11}^{XY} actually comes from covariances $\sigma_{h_1h_2}$ between two different species h_1 and h_2, each of them carrying one of the two conserved charges X and Y. Note also that two ratios, $\tilde{C}_{BS}$ and $\tilde{C}'_{BS}$, have been proposed as proxies for χ_{11}^{BS}/χ_2^S. Both of them are built in the same way, $\tilde{C}'_{BS}$ being an extension of $\tilde{C}_{BS}$ that includes in addition the contribution of multi-strange baryons. Even though this extra contribution in $\tilde{C}'_{BS}$ (represented by the dashed-dotted orange line on Fig. 2) does not improve significantly the situation compared to $\tilde{C}_{BS}$ (shown by the dotted blue line), we will still compute both of them to see how they compare in event generator simulations.

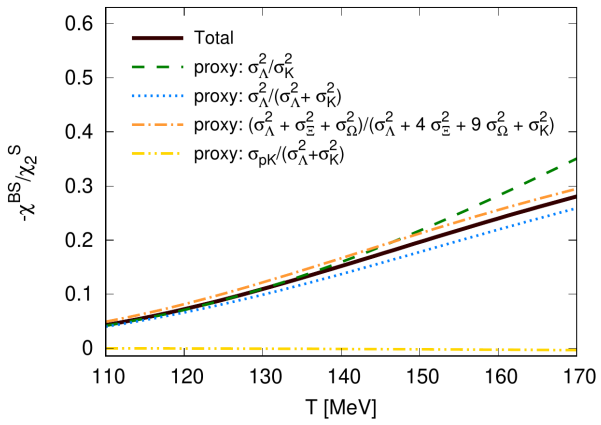


FIG. 2. Comparison of several proxy ratios for BS correlations obtained via HRG model calculations. Both are plotted as a function of T, with their related ratio of exact susceptibilities calculated with lattice QCD [24].

However, no proxy for χ_{11}^{QB}/χ_2^B has been proposed in [24], because authors claim that isospin randomization of the nucleons, due to reactions of the type: [34, 35]

$$p+\pi^{0/-} \leftrightarrow \Delta^{+/0} \leftrightarrow n+\pi^{+/0}$$

happening in the hadronic phase to both baryons and anti-baryons,' and occurring for several cycles, prevents

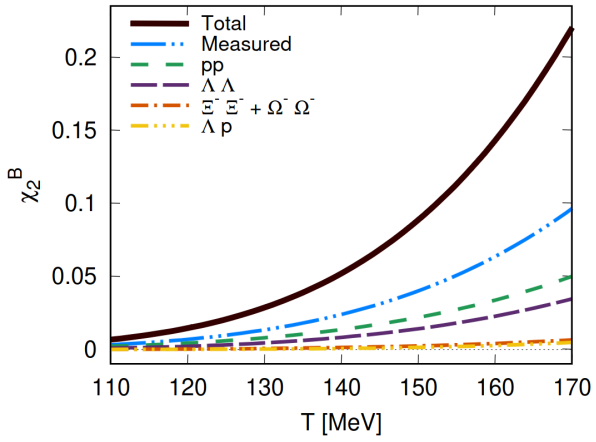


FIG. 3. Breakdown of the different final state hadronic contributions to the diagonal correlators of χ_2^B. The total contribution and the measured part are shown as solid black and dashed-dotted blue lines respectively. The main single contributions from measured hadronic observables are shown with different colored dashed and dashed-dotted lines [24].

from building a good proxy for this ratio. Since STAR measured this quantity through the proxy C_{Qp}, introduced in Eq. (5), we still want to propose here a new proxy. We start with the numerator $\chi_{11}^{QB}(=\chi_{11}^{BQ})$, which cannot be probed by σ_p^2 alone since it is affected by isospin randomization, but instead by $\sigma_{p\pi}+\sigma_p^2$, which is not. The use of this quantity was already attempted in [24], when trying to construct a proxy ratio for χ_{11}^{QB}/χ_2^Q. Next, we need a proxy for the denominator χ_2^B. Because isospin randomization leads to equilibrate the amount of protons and neutrons in the system, one can probe the nucleon variance using the proton one through $\sigma^2_{\text{Net-}N} = 2\times\sigma_p^2$ [24]. We propose to use the net-Λ variance in addition, as it is the second major contributor to χ_2^B as shown in Fig. 3, it is also used in the other proxy ratios (7), (8) and (9), and was already measured experimentally [36]. Therefore, we propose the proxy ratio:

$$\tilde{C}_{QB} = \frac{\sigma_{\pi p}^{11}+\sigma_p^2}{2\sigma_p^2+\sigma_\Lambda^2}\left(\cong\frac{\chi_{11}^{QB}}{\chi_2^B}\right), \quad (10)$$

to probe a similar quantity as the STAR proxy C_{Qp} from eq. (5).

III. EPOS4

Event generators are essential computational tools in modern high-energy physics phenomenology, enabling to bridge the gap between theoretical models and experimental data [37]. In this study, we employ EPOS4.0.0 [25], the latest version of the multi-purpose event generator, released publicly through a dedicated webpage [38]. EPOS4 is built to offer a realistic de-

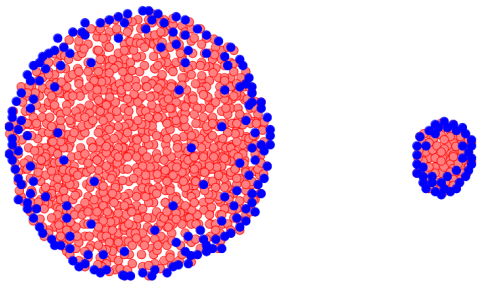


FIG. 4. Sketch of the core-corona separation for a "big" and a "small" system. The dots are pre-hadrons in the transverse plane, with red referring to pre-hadrons from core, blue to pre-hadrons from corona [43].

scription of high-energy hadronic collisions from several GeV to several TeV per-nucleon center-of-mass energy. Testing the proxies for conserved charge fluctuations introduced in Section II aims thus at validating the study using lattice QCD data and HRG simulations, on which these proxies are based. In addition, achieving this study through realistic simulations of HICs with EPOS4 offers a more direct connection with the experimental scenario, and a meaningful comparison with experimental proxies already in use to probe fluctuations.

In EPOS4, the primary interactions are modeled through a multiple-interaction approach based on the S-matrix theory including perturbative QCD calculations and saturation effects [39]. In EPOS4, major improvements have been achieved by ensuring a consistent treatment of saturation and factorization, through the implementation of a dynamical saturation scale within this energy-conserving parallel scattering framework [25, 32]. The treatment of perturbative calculations has also been updated, with a particular care for heavy flavors [40].

The pre-hadrons formed through primary scatterings and remnants excitation are then split into 2 categories, following a so-called "core-corona" procedure, as shown in Fig. 4, based on their density and some energy loss considerations [41, 42]. This procedure, applied at some initial time $\tau_0 = 0.4$ fm/c, has been discussed with many details in the context of EPOS4 in [43]. The hot and dense bulk matter made from low transverse momentum pre-hadrons, the "core", evolves following (3+1)D viscous hydrodynamics equations using vHLLE [44], and employing a simple crossover equation of state (EoS) matched with lattice QCD data [42]. Once the system has expanded and cooled down, it will hadronize when reaching an energy density value $\epsilon_H = 0.57$ GeV/fm^3, based on a micro-canonical procedure [43]. This is particularly important for simulating small systems, where few particles are produced, and hence local energy and charge conservation play an important role. On the other hand, pre-hadrons with high-p_T (sufficient to escape the core) form the "corona", and may then still re-interact with the core hadrons through hadronic scatterings.

Hadronic re-scatterings between all formed hadrons (also referred to as hadronic cascades) are simulated using UrQMD as an after-burner [45, 46]. It models elastic scatterings, string and resonance excitations as well as $2 \rightarrow n$ processes and strangeness exchange reactions based on measured reaction cross-sections, and includes 60 baryonic species + 40 mesonic states.

For the work presented here, we take advantage of the modular construction of EPOS4, since one can access the distribution of particles at hadronization, just before hadrons are fed into UrQMD. This allows us to study the impact of hadronic cascades on fluctuation observables by comparing this distribution with the final-state particle one. Moreover, the core-corona procedure applied in EPOS4 allows to differentiate particles originating from the bulk matter (the "core" in the EPOS4 framework) and the ones originating from hard processes ("corona" hadrons). Hence, one can extract the proportion of the total fluctuation signal measured in the final-state that is actually coming from the core, which contains the potential critical fluctuations.

We want to mention that, among the improvements brought to the EPOS framework in EPOS4, it is now possible to use an EoS including a 1$^{\text{st}}$ order phase transition and a critical point from the 3D Ising model universality class, such as e.g. the one published by the BEST collaboration [47] and its recent improvement [48], to model the hydrodynamical evolution of the core [49]. However, we do not use it for the HIC simulations used in this work, but employ a simple crossover EoS as mentioned previously. The main reason is that our goal is to compare different proxy observables with their corresponding conserved charge quantities, and to study the impact of hadronic cascades on them. We do not aim at investigating the effect of criticality on these observables, since EPOS4 does not propose the adequate framework for such investigation. In fact, it employs an "ordinary" hydrodynamical code that does not incorporate any out-of-equilibrium fluctuations nor charge dissipation relations, which are necessary to study the impact of critical behaviour on the evolution of the system created in HICs. Several works are currently ongoing towards the development of hydrodynamics framework that would properly take fluctuations into account [50, 51], which might eventually be included in EPOS4 in the future. Until then, EPOS4 can be considered as a non-critical baseline, especially when run with the crossover EoS employed here.

IV. ANALYSIS DETAILS

For all the results presented in Section V, particle distributions have been evaluated using the same cuts employed by STAR in [21], to allow direct comparison with

their data. We take into account all particle distributions within a pseudorapidity window $|\eta| < 0.5$ and for particles with $0.4 < p_T < 1.6$ GeV. To correct from feed-down, we chose to consider only particles that have a vertex of origin located at less than 1 cm from the interaction vertex, instead of explicitly neglecting all secondary particles as EPOS4 would allow. This is to stay as close to the experimental analysis as possible, where they correct from contamination of secondary particles by considering only tracks with a distance of closest approach (DCA) to the primary vertex < 1 cm.

$\sqrt{s_{NN}}$ (GeV)	19.6	27	39	62.4	200
N_{evts}	2.5M	2.5M	2.5M	2.0M	1.5M

TABLE I. Total number of individual Au+Au events per collision energy simulated with EPOS4.

In Table I, one can find the number of events simulated with EPOS4 for each of the collision energies shown in this work. Note that we will only show EPOS4 results from $\sqrt{s_{NN}} = 19.6$ GeV and above, although STAR data are available down to $\sqrt{s_{NN}} = 7.7$ GeV. We choose to do so since a previous study of EPOS4 results, in the energy range of the BES, showed that the model is not applicable anymore below such collision energy [31]. In fact, a systematic study of p_T spectra of many light and strange hadrons at several collision energies has shown that EPOS4 fails at reproducing, even qualitatively, the available experimental data below $\sqrt{s_{NN}} = 19.6$ GeV. This is expected, due to current limitations inherent to the approach employed to model the primary interactions. Thus, we will only discuss here results in a collision energy range which is physically relevant to the model. All events used in this study have been generated using the core-corona separation procedure, the so-called x3FF crossover equation of state mentioned in Section III with the vHLLE hydrodynamic model for the evolution of bulk matter, and neglecting any jet-fluid interaction.

Another important aspect when studying cumulants of net-multiplicity distributions in HICs, is the definition of centrality classes, and how we correct cumulants from the CBWE [26]. For all results shown as functions of the number of participating nucleons N_{part}, we compute (co)variances directly as functions of this observable, since it is accessible through our simulations. However, we do not use the actual N_{part} value from EPOS4 primary interaction model, but the Glauber estimated value based on the impact parameter. This way, we ensure the closest apple-to-apple comparison with the N_{part} values estimated for each centrality class in the STAR analysis [21]. For these results, we divide our distributions into 20 bins of width $\Delta N_{part} = 20$, which broadly correspond to centrality classes of $\sim 5\%$. The advantage is that, using directly N_{part} to determine centrality, with such narrow bins in particular, reduces the volume fluctuations within a given centrality class compared with experimental centrality classes based on charged multiplicity. Moreover, a study of the CBWE has shown that second order cumulants should not be affected much by this effect, for centrality classes smaller than 10% [26]. We have successfully tested this in EPOS4 simulations for collisions at $\sqrt{s_{NN}} = 200$ GeV, as shown in Appendix B of [52]. All results displayed as functions of collision energy $\sqrt{s_{NN}}$ for different centrality classes are obtained using the same centrality determination procedure.

The main goal of our study being to test the accuracy of the net-charge proxy cumulants introduced in Section II, we will compare them with the actual net-B, net-Q and net-S cumulants, which are computable from EPOS4 simulations because all particles are accessible. In order to make a relevant comparison with the proxy cumulants, we calculate the net-charge cumulants by considering the charges carried by all particles in the same phase space as the one used for the proxies and the STAR analysis, *i.e.* within a pseudorapidity window $|\eta| < 0.5$ and for particles with $0.4 < p_T < 1.6$ GeV. We apply the same DCA < 1 cm criterion to select particles, in order to correct from feed-down contamination, and consider particles before weak decays, since the latter violates S conservation. Technical details about the way we perform the calculation of net-charge cumulants in the EPOS4 simulated events are provided in Appendix C of [52].

V. RESULTS AND DISCUSSION

In this section, we discuss the results of our study of the proxies for 2$^{\text{nd}}$ order cumulants of conserved charges, obtained through the analyses described in Section IV. We will first show the (co)variances of different identified hadron species at several collision energies $\sqrt{s_{NN}}$, as functions of N_{part}, for final-state particle distributions and also particle distributions just after hadronization, compared with experimental data from STAR [21]. From there, we will build the different proxy ratios, and compare them with their corresponding conserved charge cumulant ratios, as well as with STAR data, still as functions of N_{part} for different collision energies. In order to discuss the results of our study in a more summarized way, we will then show the different ratios as functions of $\sqrt{s_{NN}}$, for the most central collisions. At first, we will show results for both final-state particle distributions and particle distributions at hadronization, to see the impact of hadronic cascades on displayed quantities. Finally, we will show the proxy and corresponding conserved charge ratios for distributions of all particles at hadronization, compared with the same quantities but only from core particles, in order to highlight the contribution of bulk matter to the fluctuation signal measured in the phase space considered in this analysis.

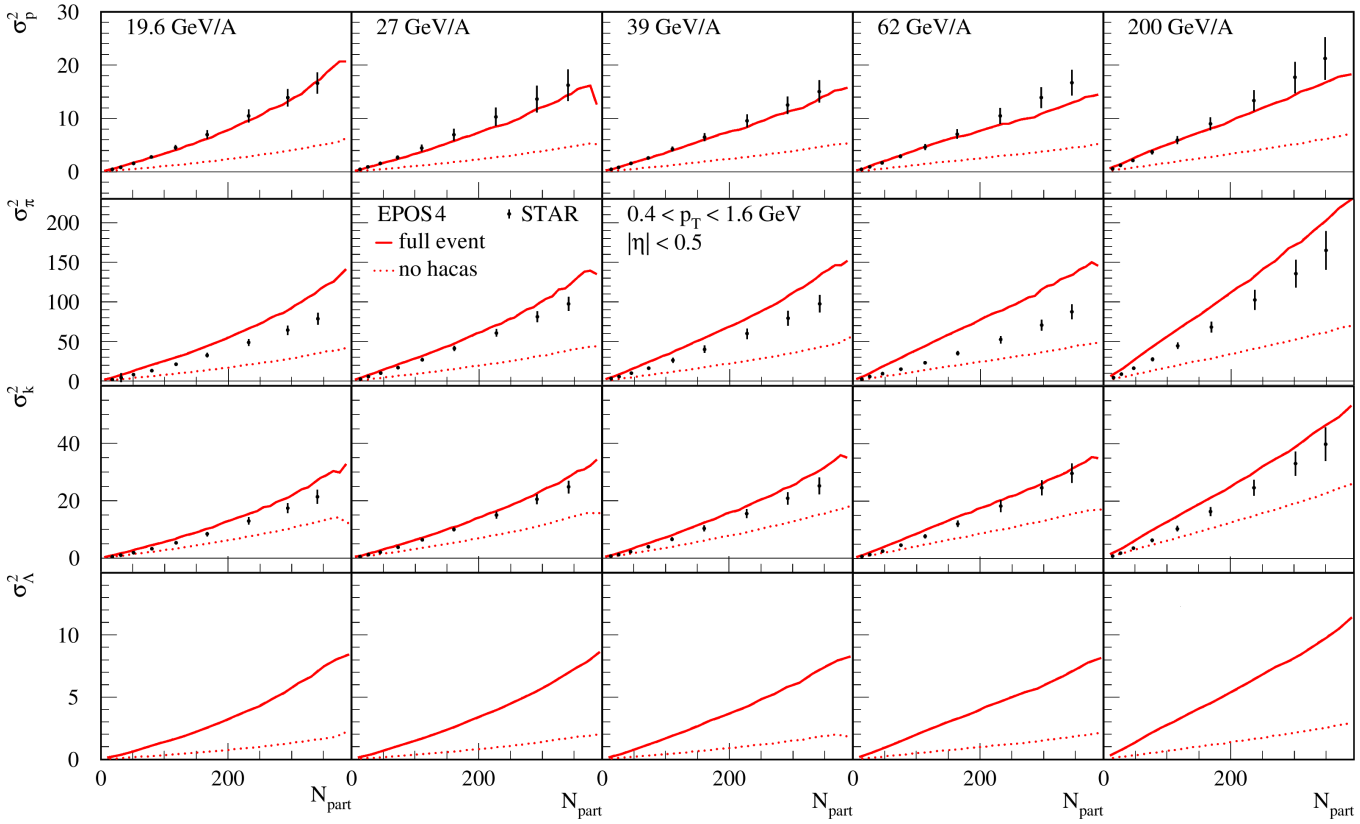


FIG. 5. Net-p, net-π, net-K and net-Λ multiplicity variances vs. N_{part} for different energies, obtained from EPOS4 simulations and compared with STAR results (dots) [21]. The results calculated from full events are represented by full lines, while results calculated before hadronic cascades are represented by dotted lines.

More detailed results on particle production relevant for the system and collision energies studied in this paper, based on the same set of simulated events, can be found in [31, 32].

A. Variances and covariances of identified hadron net-multiplicities

In Fig. 5, we show the centrality dependence (via the number of participants N_{part}) of variances of net-π, net-K, net-p and net-Λ multiplicities at collision energies $\sqrt{s_{NN}} = 19.6$, 27, 39, 62.4 and 200 GeV. There, one can see variances from final-state particle distributions (full lines) and from particle distributions at hadronization (dashed lines), both from EPOS4, compared with STAR results (dots) [21]. We observe that EPOS4 reproduces qualitatively the expected (quasi-)linear increase with decreasing centrality (increasing N_{part}) of the different variances in both distributions. On a quantitative level, EPOS4 results for final-state distributions do not match perfectly the data from STAR, although describing them quite well overall. This is however not of extreme importance in the context of this work, since we only aim at making a quantitative comparative study. Another important aspect to notice in this Figure, is the difference between variances for particle distribution at hadronization and in the final-state. For all hadronic species shown, hadronic cascades increase the variance at all collision energies, by a factor unique to each species which is quite constant regardless of the energy considered. They increase the magnitude of the signal by a factor ~ 2 for σ^2_K to ~ 3 for σ^2_π and σ^2_p, even up to a factor ~ 4 for σ^2_Λ where the effect is the most significant.

In Fig. 6, the same plots are shown for the different covariances $\sigma^{11}_{\pi p}$, $\sigma^{11}_{\pi K}$ and σ^{11}_{pK}. Despite the apparent wiggles observed in the results for covariances, caused by the limited statistics used for this analysis (see Table I), one can conclude similarly to the case of variances that the results from EPOS4 final-state distributions reproduce the qualitative behavior of STAR data, and are reasonably close on a quantitative level. We also note an increase of the absolute amplitude of the signal due to hadronic cascades, when comparing EPOS4 distributions at hadronization (again shown as a dashed line) with EPOS4 distributions of final-state hadrons (full line).

The observed increase of (co)variances can be ex-

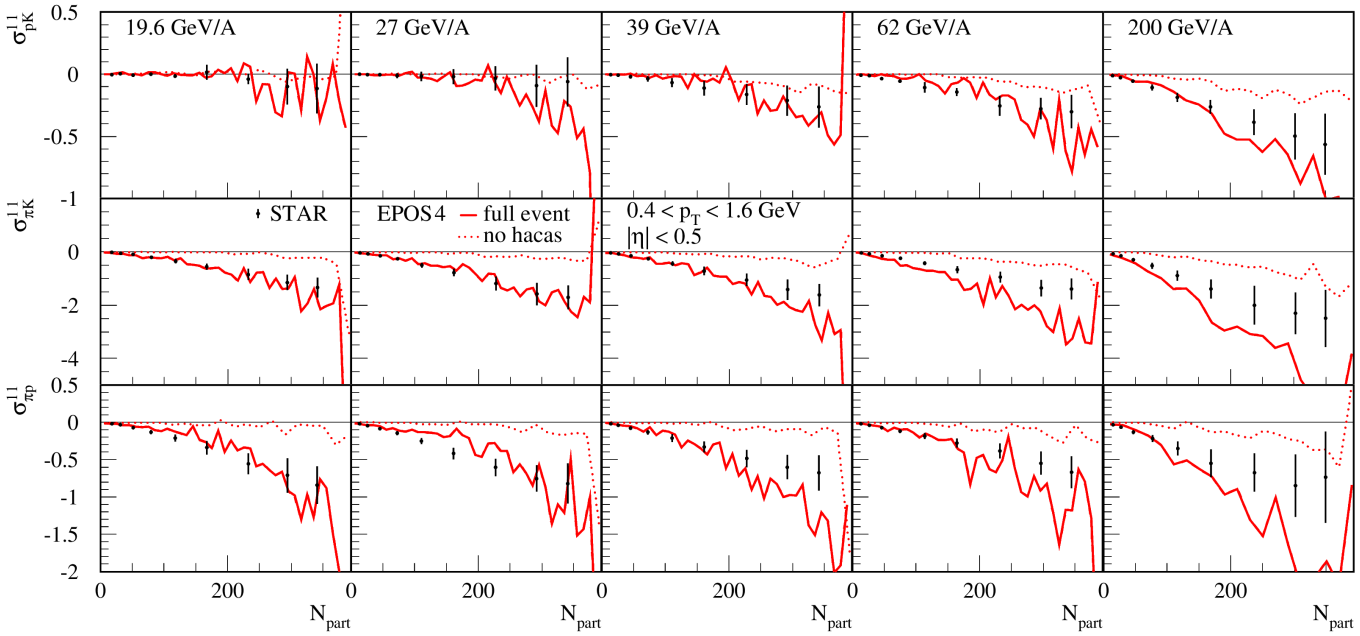


FIG. 6. Covariances of net-p, net-π and net-K multiplicity distributions vs. N_{part} for different energies, obtained from EPOS4 simulations and compared with STAR results (dots) [21]. Results calculated from full events are represented by full lines, while results calculated before hadronic cascades are represented by dotted lines.

plained by the fact that, during hadronic cascades, the numerous inelastic scatterings produce many more of the involved hadronic species, coupled to the successive decays of heavy resonances produced at hadronization. Hence, the successive decays and inelastic scatterings, producing simultaneously several of the hadronic species of interest, lead unavoidably to the amplification of the correlations carried by (co)variances, blurring the actual magnitude of the thermal fluctuations of interest by increasing the signal. These observations support the results of a study led with a hybrid version of UrQMD, which has been coupled to a fluid dynamical simulation of bulk matter [53]. The authors of [53] investigated the time-evolution of the correlation of distributions of hadron families carrying conserved charges, compared to the same distributions at hadronization time. The conclusion was that the final distribution of hadrons observed by experiments is not really related anymore to the one obtained after hadronization, which would contain the potential critical signal from the QCD phase transition.

B. Ratios of B, Q, S cumulants and associated proxies

In Fig. 7, we show the different ratios of 2$^{\text{nd}}$ order conserved charge cumulants with their corresponding proxies, both from EPOS4 and STAR measurements from [21], as functions of N_{part} for different collision energies. The red curves correspond to the STAR proxy ratios C_{Qp}, C_{QK} and C_{pK} measured in EPOS4 simulations (defined in Eq. (5)), to be compared with STAR measurements of the same quantities shown as black dots. For each of these 3 quantities, we also plot in dark blue the corresponding new proposed proxies $\tilde{C}_{QB}$, $\tilde{C}_{QS}$ and $\tilde{C}_{BS}$ + $\tilde{C}'_{BS}$ (in light blue), defined in Eqs. (10), (7) and (8) + (9) respectively. Finally, we display the corresponding ratios of exact cumulants of conserved charges in green, as a reference. Note that we will use this same color code for all figures displaying cumulants ratios.

Let us discuss first the comparison of STAR proxies measured in EPOS4 simulations, with the actual data from [21] for those quantities. Although variances and covariances computed from EPOS4 simulations do not reproduce perfectly the STAR data, as discussed in the previous subsection, we notice in Fig. 7 that the proxy ratios C_{Qp}, C_{QK} and C_{pK} constructed from them do match STAR data point very well on a quantitative level.

Another noticeable feature is the apparent absence of centrality dependence, observed for all ratios displayed in Fig. 7 (apart from a slight decrease of the ratio C_{Qp} with increasing N_{part}, that appears to be more significant with increasing collision energy). For this reason, we will compare from now on the different ratios of conserved charge cumulants with their corresponding proxies as a function of collision energy, for central collisions (0-5% centrality class) only.

Finally, we want to address the case of the ratio C_{BS} ($= \sigma^{11}_{BS}/\sigma^2_S$) and its proxies. As an alternative to the STAR proxy C_{pK}, we have decided to consider two

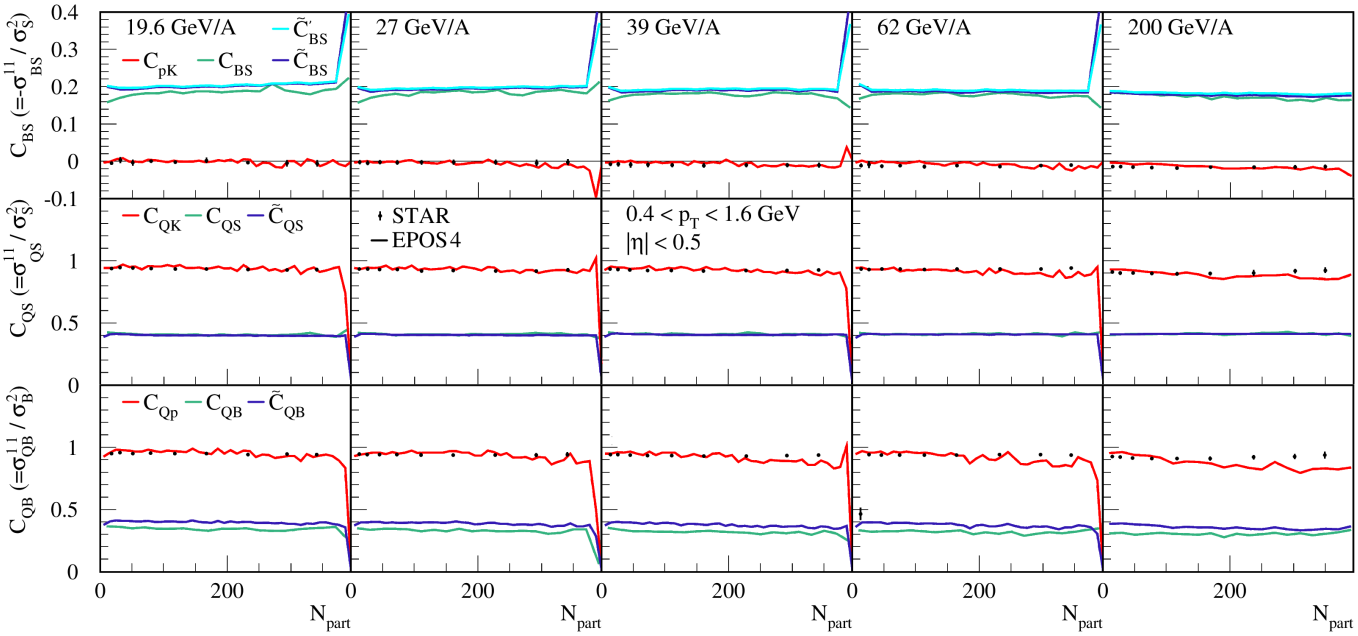


FIG. 7. Ratios of exact B, Q and S conserved charge (co)variances and their proxies displayed as functions of N_{part} for different energies, obtained from EPOS 3.451 simulations (lines) and compared with STAR data (dots) as a reference [21].

proxy ratios from [24], $\tilde{C}_{BS}$ and $\tilde{C}'_{BS}$. While the first one, defined in Eq. (8), relies only on net-Λ variance as a carrier of both strangeness and baryon charge, the second one defined in Eq. (9) also incorporates variances of net-Ξ and net-Ω. However, despite the fact that Ξ and Ω baryons carry respectively a strangeness content of $|S| = 2$ and $|S| = 3$, they are produced at a significantly lower rate compared to Λ baryons in HICs, approximately 1 and 2 orders magnitude less, respectively. This low relative abundance makes their contribution to the C_{BS} proxy ratio almost insignificant in comparison with the net-Λ variance. Looking at the top panels of Fig. 7, showing C_{BS} and related proxies, one can effectively see that $\tilde{C}_{BS}$ and $\tilde{C}'_{BS}$ results completely overlap with over the whole centrality range and all collision energies displayed. Thus, we choose to discard $\tilde{C}'_{BS}$ from the results from now on, and only focus on $\tilde{C}_{BS}$ as a new alternative proxy ratio to C_{BS}.

Coming now to the first main results of our study, we show in Fig. 8 the ratios C_{BS}, C_{QS} and C_{QB} compared with their respective proxy ratios and STAR results, as functions of collision energy $\sqrt{s_{NN}}$ for 0-5% centrality class events. We observe that the STAR proxy ratios from Eq. 5 measured in EPOS4 simulations are very close to the STAR data (as already noticed previously), despite a slight energy dependence not present in the data. Overall, we note that none of the displayed ratios shows any clear collision energy dependence, for the range that is shown here. However, the important result in these plots lies in the comparison between the ratios of conserved charge cumulants (in green), taken as a reference, and their proxies. STAR data are only shown as a benchmark, to see how well EPOS4 can reproduce experimental results. One can observe how remarkably well the new proposed proxies $\tilde{C}_{XY}$ (in blue) are able to reproduce quantitatively the conserved charge cumulant ratios, compared to the STAR proxies C_{QpK} (in red) which show a clear deviation in magnitude. These new proxies are thus shown to be better quantitative probes than the ones traditionally used by STAR, as well as other experiments to probe fluctuations through 2$^{\text{nd}}$ order cumulants of conserved charges in HICs (see [54] for a review of these measurements).

C. Impact of hadronic cascades

In Fig. 9, we show all the different ratios and their corresponding proxies as functions of collision energy for the most central (0-5%) collisions, for final-state particle distributions like in Fig. 8 (full lines), but this time compared with their associated value measured at hadronization (dashed lines). As explained in Section III, the latter can be measured thanks to the modular construction of EPOS4, which allows to access particle distributions just after hadronization of the core, before re-interactions between formed hadrons are simulated (using UrQMD in cascade mode).

One can see that the magnitude of almost all ratios is affected very little by the evolution of the system during hadronic cascades, as results for particle distributions at hadronization are very close to their associated

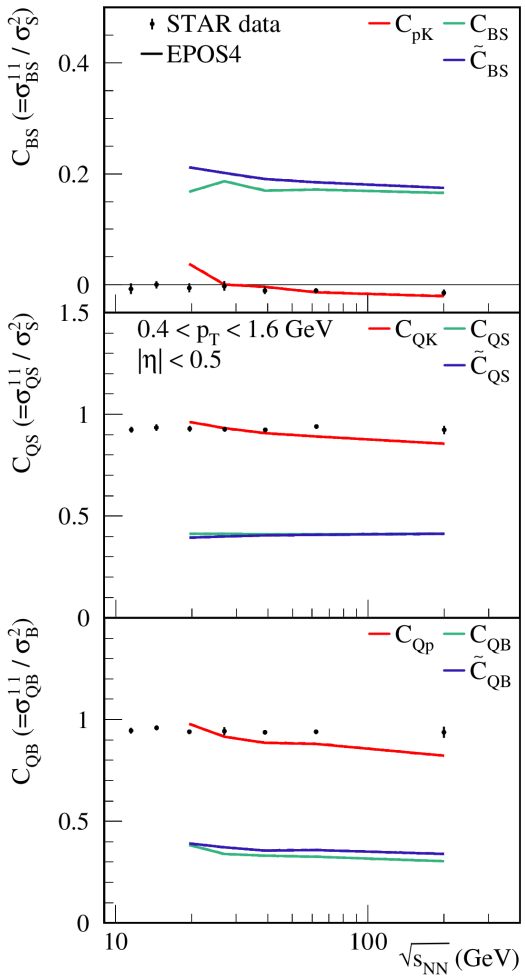


FIG. 8. Ratios of exact B, Q and S conserved charge (co)variances and their proxies, displayed as functions of $\sqrt{s_{NN}}$ for central collisions (0-5% centrality class). Results from EPOS4 simulations for final-state particle distributions (lines) are compared with STAR data (dots) as a reference [21].

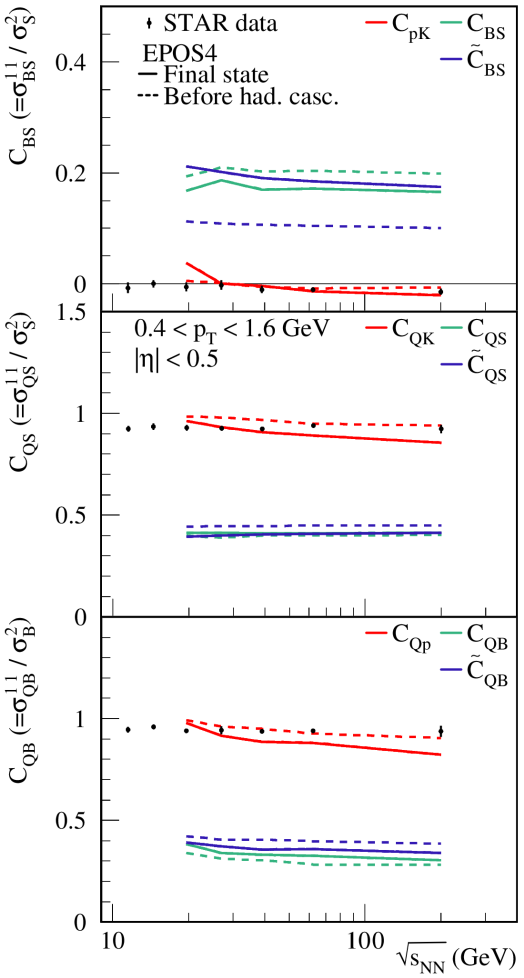


FIG. 9. Ratios of exact B, Q and S conserved charge (co)variances and their proxies plotted as functions of $\sqrt{s_{NN}}$ for central collisions (0-5% centrality class). Results are displayed from particle distribution in the final-state (full lines) and before hadronic cascades (dashed lines), from EPOS4 simulations. STAR data (dots) are shown as a reference [21].

value for finale-state distributions. While this result was expected in the case of ratios of conserved charge cumulants, it demonstrates that most proxies discussed here are somehow stable quantities, in the sense that they conserve their magnitude even after the system has gone through hadronic re-interactions which modify its chemistry. Although we have shown in Subsection V A that the (co)variances used to build the proxy ratios are largely impacted by this phase in the evolution of the system, it appears that the modifications of both their respective numerator and denominator compensate overall to leave most of these ratios almost unaffected.

There is one exception to this picture for the proxy ratio $\tilde{C}_{BS}$, which gets a significant modification through hadronic re-interactions. Following the same logic as in the previous paragraph, we understand this to be due to the fact that $\tilde{C}_{BS}$ is built upon σ^2_Λ and σ^2_K, which are the variances that are affected respectively the most and the least from hadronic cascades. In particular, looking at how $\tilde{C}_{BS}$ is constructed in Eq. (8), we see that the numerator consist of σ^2_Λ alone, which gets a growth of about $\sim$ 4 through the hadronic cascades, while σ^2_K in the denominator is only increased by a factor ~ 2. Even though the latter is added to σ^2_Λ, their relative difference in magnitude makes σ^2_K dominate the denominator, hence explaining the increase of $\tilde{C}_{BS}$ after hadronic re-scatterings. Such an important growth of net-Λ variance, relative to other hadronic species, might come from the fact that Λ baryons receive an important feed-down contribution from resonance decays, which are not included when considering particles at hadronization time, while they are taken into account when looking at particle distributions in the final-state. Moreover, Λ baryons are also less affected by baryon annihila-

tion than protons, as discussed in [55], which leads to a smaller decrease of σ^2_Λ from this process as compared to σ^2_p, since their net-multiplicity will decrease less.

In order to investigate this issue further, we tried to find another proxy ratio for C_{BS} which would be less modified than $\tilde{C}_{BS}$. Thus, we have computed the following ratio:

$$\tilde{C}^*_{BS} = \frac{\sigma^2_\Xi}{\sigma^2_K + 4\sigma^2_\Xi}\,, \tag{11}$$

replacing variance of net-Λ by variance of net-Ξ which, even though produced less abundantly in the system, carries more strangeness and is less impacted by hadronic cascades. This quantity is compared with C_{BS}, $\tilde{C}_{BS}$ and C_{pK} in Fig. 10, shown in solid yellow for final state particle distributions and in dashed yellow for distributions before hadronic cascades. We see that although $\tilde{C}^*_{BS}$ suffers almost no change from the hadronic re-interactions phase, it shows a large discrepancy in magnitude with C_{BS}, because the Ξ baryons carry too small a fraction of baryon-strangeness correlation in the system. For this reason, $\tilde{C}_{BS}$ remains the best proxy, especially since it matches C_{BS} very well qualitatively in the final state, which is the only quantity accessible in experiments.

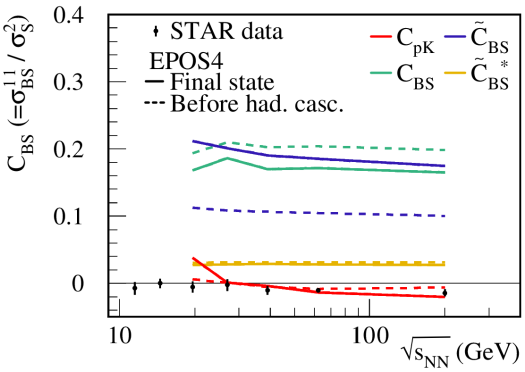


FIG. 10. C_{BS} ratio and different corresponding proxies plotted as functions of $\sqrt{s_{NN}}$ for central collisions (0-5% centrality class). Results are displayed for full event calculations (full lines) and calculations before hadronic cascades (dashed lines) from EPOS4 simulations. STAR data (dots) are shown as a reference [21].

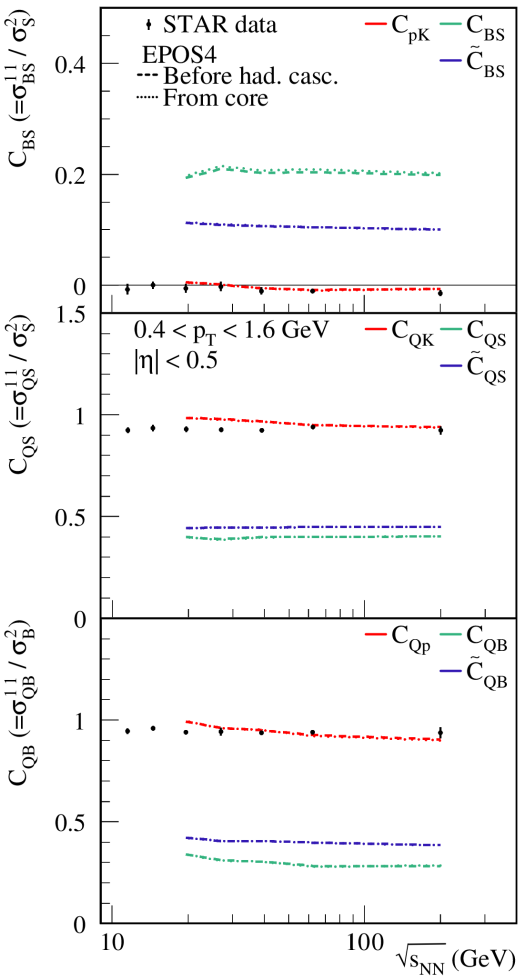


FIG. 11. Core proportion of C_{XY} and enhanced proxies $\tilde{C}_{XY}$ ratios over the total value calculated before hadronic cascades, plotted as a function of $\sqrt{s_{NN}}$ for central collisions (0-5% centrality class). Ratios of exact B, Q and S conserved charge (co)variances and their proxies plotted as functions of $\sqrt{s_{NN}}$ for central collisions (0-5% centrality class). Results are displayed from particle distribution before hadronic cascades (dashed lines) and from core particles only (dotted lines), from EPOS4 simulations. STAR data (dots) are shown as a reference [21].

D. Effective contribution from bulk matter

Finally, we show in Fig. 11 the different ratios from particles at hadronization time (dashed lines), compared with the same quantities computed exclusively from particles originating from the core (dotted lines), *i.e.* the part of the system mimicking the behaviour of the QGP. Like in the previous figures, they are displayed as functions of the collision energy, for the most central (0-5%, left panel) and most peripheral (0-80%, right panel) collisions.

We observe that for every ratio, in both central and peripheral centrality classes, the ratios computed only from core particles at hadronization are overlapping almost perfectly with the ratios computed from all particles at hadronization. This means that the signal measured by the ratios of second order cumulants in the phase space considered here ($|\eta| < 0.5$ and $0.4 < p_T < 1.6$ GeV) is almost exclusively originating from the bulk matter, modeled by the core in EPOS4. Hence, if the system created in such collisions would endure critical behavior, these ratios should indeed be sensitive to it, and carry its signature as long as they are measured in a similar phase

space.

VI. SUMMARY & OUTLOOKS

We discussed the use of new proxies for conserved charge cumulants, which are of primary importance in the study of correlations between conserved charges B, Q and S, and expected to be key observables in the search for the existence of a critical point in the phase diagram of nuclear matter. The new quantities $\tilde{C}_{QS}$ and $\tilde{C}_{BS}$, proposed in [24], as well as $\tilde{C}_{QB}$ which we proposed based on that previous study, are compared to corresponding proxies commonly measured in experiment C_{QK}, C_{pK} and C_{Qp}, as well as their counterparts based on exact conserved charge quantities. We employed EPOS4 to evaluate all these observables through simulations of Au+Au collisions at energies $\sqrt{s_{NN}} = 19.6 - 200$ GeV, where data was also available for comparison in [21], in order to assess the actual performance of the new proxies in the context of a realistic out-of-equilibrium system, close to the one created experimentally.

The results of our study showed that the newly proposed proxies are better at reproducing their corresponding conserved charge cumulant ratios, as compared to the equivalent STAR proxies. We note first of all that, with the only exception of $\tilde{C}_{BQ}$, all these quantities are uniquely built upon variances. This makes sense, since correlations between different conserved charges are naturally carried by some hadron species (e.g. B and Q by the protons, Q and S by kaons or B and S by Λ baryons). Moreover, we observe that the addition of σ^2_Λ alone to the already measured σ^2_π, σ^2_K and σ^2_p is enough to build those new proxies. Luckily, this quantity was already measured (although in a different phase space) by STAR and published in [36].

We compared the different ratios measured from final-state particle distributions, to their value measured from particle distribution at hadronization time. This way, we were able to quantify the effect of hadronic cascades on these quantities. While the ratio of conserved charge cumulants are almost not affected by hadronic rescatterings, as expected, we can also see that $\tilde{C}_{QB}$ and $\tilde{C}_{QS}$ are also left unchanged after the hadronic cascades. For this reason, they represent excellent proxies of their corresponding ratios of conserved charge cumulants C_{QB} and C_{QS}. The only significant deviation between measurement at hadronization and in the final-state is observed for $\tilde{C}_{BS}$. We saw that this discrepancy originates from the fact that σ^2_Λ is more enhanced by hadronic cascades than other net-variances, in particular σ^2_K. Potential reasons to explain this difference have been discussed.

Finally, we investigated how much of the signal, for the phase space used in this analysis, is actually coming from particles originating from the bulk matter of the system. Unsurprisingly, the quasi-totality of the signal effectively originates from the core in EPOS4, meaning that measuring net-cumulants of hadronic species within $|\eta| < 0.5$ and with $0.4 < p_T < 1.6$ GeV actually probes fluctuations coming uniquely from bulk matter. Hence, if the system created in such collisions was to experience the impact of a critical point during its evolution, fluctuation observables should carry its imprint.

It would be interesting to test the ability of the new proxy ratios $\tilde{C}_{QB}$, $\tilde{C}_{QS}$ and $\tilde{C}_{BS}$ to reproduce their corresponding ratios of cumulants of conserved charges at lower collision energies, compared to the ones shown in this work. Most of the studies published in recent years seem to point at a collision-energy range $1 - 20$ GeV/A to observe phenomena which could hint at the existence of a critical point [13]. Hence, using transport models which are suited for modeling the dynamics of HICs at these energies seems highly relevant and timely; see [56] for a recent review of the status of such models.

Investigating with more details the impact of the hadronic re-scatterings on the new proxy ratios, in particular within the previously mentioned collision-energy range, is also important to ensure their robustness through the evolution of the system. Some work in this direction has already been initiated, for instance by using an extended particle list to study the impact of resonances in [57], or by including mean-field potentials to model hadron interactions in a more realistic manner in transport approaches [58].

ACKNOWLEDGMENTS

J.J. thanks Jean-Yves Ollitrault, Marlene Nahrgang, Grégoire Pihan and Voldymyr Vovchenko for the instructive discussions about this work. This material is based upon work supported by the National Science Foundation under grants No. PHY-2208724, PHY-1654219 and PHY-2116686, and within the framework of the MUSES collaboration, under grant number No. OAC- 2103680. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Award Number DE-SC0022023.

[1] I. Arsene *et al.* (BRAHMS), Quark gluon plasma and color glass condensate at RHIC? The Perspective from the BRAHMS experiment, Nucl. Phys. A **757**, 1 (2005), arXiv:nucl-ex/0410020.

[2] B. B. Back *et al.* (PHOBOS), The PHOBOS perspective on discoveries at RHIC, Nucl. Phys. A **757**, 28 (2005), arXiv:nucl-ex/0410022.

[3] J. Adams *et al.* (STAR), Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions, Nucl. Phys. A **757**, 102 (2005), arXiv:nucl-ex/0501009.

[4] K. Adcox *et al.* (PHENIX), Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration, Nucl. Phys. A **757**, 184 (2005), arXiv:nucl-ex/0410003.

[5] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz, and K. K. Szabo, The Order of the quantum chromodynamics transition predicted by the standard model of particle physics, Nature **443**, 675 (2006), arXiv:hep-lat/0611014 [hep-lat].

[6] A. Bazavov *et al.* (HotQCD), Chiral crossover in QCD at zero and non-zero chemical potentials, Phys. Lett. B **795**, 15 (2019), arXiv:1812.08235 [hep-lat].

[7] S. Borsanyi, Z. Fodor, J. N. Guenther, R. Kara, S. D. Katz, P. Parotto, A. Pasztor, C. Ratti, and K. K. Szabo, QCD Crossover at Finite Chemical Potential from Lattice Simulations, Phys. Rev. Lett. **125**, 052001 (2020), arXiv:2002.02821 [hep-lat].

[8] T. Niida and Y. Miake, Signatures of QGP at RHIC and the LHC, AAPPS Bull. **31**, 12 (2021), arXiv:2104.11406 [nucl-ex].

[9] C. S. Fischer, QCD at finite temperature and chemical potential from Dyson–Schwinger equations, Prog. Part. Nucl. Phys. **105**, 1 (2019), arXiv:1810.12938 [hep-ph].

[10] D. Fuseau, T. Steinert, and J. Aichelin, Phase diagram of the Polyakov–Nambu–Jona-Lasinio approach for finite chemical potentials, Phys. Rev. C **101**, 065203 (2020), arXiv:1908.08122 [hep-ph].

[11] W.-j. Fu, J. M. Pawlowski, and F. Rennecke, QCD phase structure at finite temperature and density, Phys. Rev. D **101**, 054032 (2020), arXiv:1909.02991 [hep-ph].

[12] M. Hippert, J. Grefa, T. A. Manning, J. Noronha, J. Noronha-Hostler, I. Portillo Vazquez, C. Ratti, R. Rougemont, and M. Trujillo, Bayesian location of the QCD critical point from a holographic perspective, (2023), arXiv:2309.00579 [nucl-th].

[13] A. Bzdak, S. Esumi, V. Koch, J. Liao, M. Stephanov, and N. Xu, Mapping the Phases of Quantum Chromodynamics with Beam Energy Scan, Phys. Rept. **853**, 1 (2020), arXiv:1906.00936 [nucl-th].

[14] D. Tlusty, The RHIC Beam Energy Scan Phase II: Physics and Upgrades, in *13th Conference on the Intersections of Particle and Nuclear Physics* (2018) arXiv:1810.04767 [nucl-ex].

[15] K. C. Meehan (STAR), Fixed Target Collisions at STAR, Nucl. Phys. A **956**, 878 (2016).

[16] R. Bellwied, S. Borsanyi, Z. Fodor, S. D. Katz, A. Pasztor, C. Ratti, and K. K. Szabo, Fluctuations and correlations in high temperature QCD, Phys. Rev. D **92**, 114505 (2015), arXiv:1507.04627 [hep-lat].

[17] X. Luo and N. Xu, Search for the QCD Critical Point with Fluctuations of Conserved Quantities in Relativistic Heavy-Ion Collisions at RHIC : An Overview, Nucl. Sci. Tech. **28**, 112 (2017), arXiv:1701.02105 [nucl-ex].

[18] M. Stephanov, Non-Gaussian fluctuations near the QCD critical point, Phys. Rev. Lett. **102**, 032301 (2009), arXiv:0809.3450 [hep-ph].

[19] M. Abdallah *et al.* (STAR), Cumulants and correlation functions of net-proton, proton, and antiproton multiplicity distributions in Au+Au collisions at energies available at the BNL Relativistic Heavy Ion Collider, Phys. Rev. C **104**, 024902 (2021), arXiv:2101.12413 [nucl-ex].

[20] L. Adamczyk *et al.* (STAR), Collision Energy Dependence of Moments of Net-Kaon Multiplicity Distributions at RHIC, Phys. Lett. B **785**, 551 (2018), arXiv:1709.00773 [nucl-ex].

[21] J. Adam *et al.* (STAR), Collision-energy dependence of second-order off-diagonal and diagonal cumulants of net-charge, net-proton, and net-kaon multiplicity distributions in Au + Au collisions, Phys. Rev. C **100**, 014902 (2019), [Erratum: Phys.Rev.C 105, 029901 (2022)], arXiv:1903.05370 [nucl-ex].

[22] M. Bluhm *et al.*, Dynamics of critical fluctuations: Theory – phenomenology – heavy-ion collisions, Nucl. Phys. A **1003**, 122016 (2020), arXiv:2001.08831 [nucl-th].

[23] V. Koch, A. Majumder, and J. Randrup, Baryon-strangeness correlations: A Diagnostic of strongly interacting matter, Phys. Rev. Lett. **95**, 182301 (2005), arXiv:nucl-th/0505052.

[24] R. Bellwied, S. Borsanyi, Z. Fodor, J. N. Guenther, J. Noronha-Hostler, P. Parotto, A. Pasztor, C. Ratti, and J. M. Stafford, Off-diagonal correlators of conserved charges from lattice QCD and how to relate them to experiment, Phys. Rev. D **101**, 034506 (2020), arXiv:1910.14592 [hep-lat].

[25] K. Werner, Revealing a deep connection between factorization and saturation: New insight into modeling high-energy proton-proton and nucleus-nucleus scattering in the EPOS4 framework, Phys. Rev. C **108**, 064903 (2023), arXiv:2301.12517 [hep-ph].

[26] N. R. Sahoo, S. De, and T. K. Nayak, Baseline study for higher moments of net-charge distributions at energies available at the BNL Relativistic Heavy Ion Collider, Phys. Rev. C **87**, 044906 (2013), arXiv:1210.7206 [nucl-ex].

[27] V. Vovchenko and V. Koch, Efficiency corrections for factorial moments and cumulants of overlapping sets of particles, Nucl. Phys. A **1010**, 122179 (2021), arXiv:2101.02182 [nucl-th].

[28] A. Chatterjee, T. Nonaka, S. Esumi, and X. Luo, Efficiency corrections for mutually inclusive variables and particle identification effect for mixed-cumulants in heavy-ion collisions, Chin. Phys. C **45**, 104001 (2021), arXiv:2104.08077 [physics.data-an].

[29] P. Alba *et al.*, Constraining the hadronic spectrum through QCD thermodynamics on the lattice, Phys. Rev. D **96**, 034517 (2017), arXiv:1702.01113 [hep-lat].

[30] M. S. Abdallah *et al.* (STAR), Disappearance of partonic collectivity in sNN=3GeV Au+Au collisions at RHIC, Phys. Lett. B **827**, 137003 (2022), arXiv:2108.00908 [nucl-ex].

[31] K. Werner, J. Jahan, I. Karpenko, T. Pierog, M. Stefaniak, and D. Vintache, Heavy ion collisions from $\sqrt{s_{NN}}$ of 62.4 GeV down to 4 GeV in the EPOS4 framework, (2024), arXiv:2401.11275 [hep-ph].

[32] K. Werner, Parallel scattering, saturation, and generalized AGK theorem in the EPOS4 framework, with applications for heavy ion collisions at 5.02 ATeV and at 200 AGeV, (2023), arXiv:2310.09380 [hep-ph].

[33] V. V. Begun, M. I. Gorenstein, M. Hauer, V. P. Konchakovski, and O. S. Zozulya, Multiplicity Fluctuations in Hadron-Resonance Gas, Phys. Rev. C **74**, 044903 (2006), arXiv:nucl-th/0606036.

[34] M. Kitazawa and M. Asakawa, Revealing baryon number fluctuations from proton number fluctuations in relativis-

tic heavy ion collisions, Phys. Rev. C **85**, 021901 (2012), arXiv:1107.2755 [nucl-th].
[35] M. Kitazawa and M. Asakawa, Relation between baryon number fluctuations and experimentally observed proton number fluctuations in relativistic heavy ion collisions, Phys. Rev. C **86**, 024904 (2012), [Erratum: Phys.Rev.C 86, 069902 (2012)], arXiv:1205.3292 [nucl-th].
[36] J. Adam *et al.* (STAR), Beam energy dependence of net-Λ fluctuations measured by the STAR experiment at the BNL Relativistic Heavy Ion Collider, Phys. Rev. C **102**, 024903 (2020), arXiv:2001.06419 [nucl-ex].
[37] J. M. Campbell *et al.*, Event Generators for High-Energy Physics Experiments, in *Snowmass 2021* (2022) arXiv:2203.11110 [hep-ph].
[38] `https://klaus.pages.in2p3.fr/epos4/`.
[39] H. J. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner, Parton based Gribov-Regge theory, Phys. Rept. **350**, 93 (2001), arXiv:hep-ph/0007198 [hep-ph].
[40] K. Werner and B. Guiot, Perturbative QCD concerning light and heavy flavor in the EPOS4 framework, Phys. Rev. C **108**, 034904 (2023), arXiv:2306.02396 [hep-ph].
[41] K. Werner, Core-corona separation in ultra-relativistic heavy ion collisions, Phys. Rev. Lett. **98**, 152301 (2007), arXiv:0704.1270 [nucl-th].
[42] K. Werner, B. Guiot, I. Karpenko, and T. Pierog, Analysing radial flow features in p-Pb and p-p collisions at several TeV by studying identified particle production in EPOS3, Phys. Rev. **C89**, 064903 (2014), arXiv:1312.1233 [nucl-th].
[43] K. Werner, Core-corona procedure and microcanonical hadronization to understand strangeness enhancement in proton-proton and heavy ion collisions in the EPOS4 framework, Phys. Rev. C **109**, 014910 (2024), arXiv:2306.10277 [hep-ph].
[44] I. Karpenko, P. Huovinen, and M. Bleicher, A 3+1 dimensional viscous hydrodynamic code for relativistic heavy ion collisions, Comput. Phys. Commun. **185**, 3016 (2014), arXiv:1312.4160 [nucl-th].
[45] M. Bleicher *et al.*, Relativistic hadron hadron collisions in the ultrarelativistic quantum molecular dynamics model, J. Phys. **G25**, 1859 (1999), arXiv:hep-ph/9909407 [hep-ph].
[46] H. Petersen, J. Steinheimer, G. Burau, M. Bleicher, and H. Stöcker, A Fully Integrated Transport Approach to Heavy Ion Reactions with an Intermediate Hydrodynamic Stage, Phys. Rev. C **78**, 044901 (2008), arXiv:0806.1695 [nucl-th].
[47] P. Parotto, M. Bluhm, D. Mroczek, M. Nahrgang, J. Noronha-Hostler, K. Rajagopal, C. Ratti, T. Schäfer, and M. Stephanov, QCD equation of state matched to lattice data and exhibiting a critical point singularity, Phys. Rev. C **101**, 034901 (2020), arXiv:1805.05249 [hep-ph].
[48] M. Kahangirwe, S. A. Bass, E. Bratkovskaya, J. Jahan, P. Moreau, P. Parotto, D. Price, C. Ratti, O. Soloveva, and M. Stephanov, Finite density QCD equation of state: critical point and lattice-based T'-expansion, (2024), arXiv:2402.08636 [nucl-th].
[49] M. Stefaniak, K. Werner, J. Jahan, and H. P. Zbroszczyk, Equation of state within the EPOS3 model, Phys. Rev. C **108**, 014905 (2023), arXiv:2209.12979 [hep-ph].
[50] M. Stephanov and Y. Yin, Hydrodynamics with parametric slowing down and fluctuations near the critical point, Phys. Rev. D **98**, 036006 (2018), arXiv:1712.10305 [nucl-th].
[51] G. Pihan, M. Bluhm, M. Kitazawa, T. Sami, and M. Nahrgang, Critical net-baryon fluctuations in an expanding system, Phys. Rev. C **107**, 014908 (2023), arXiv:2205.12834 [nucl-th].
[52] J. Jahan, *The Beam Energy Scan under the scope of EPOS 4*, PhD Thesis, Nantes Université (2022).
[53] J. Steinheimer, V. Vovchenko, J. Aichelin, M. Bleicher, and H. Stöcker, Conserved charge fluctuations are not conserved during the hadronic phase, Phys. Lett. B **776**, 32 (2018), arXiv:1608.03737 [nucl-th].
[54] R. Kumar *et al.* (MUSES), Theoretical and Experimental Constraints for the Equation of State of Dense and Hot Matter, (2023), arXiv:2303.17021 [nucl-th].
[55] J. Steinheimer, J. Aichelin, M. Bleicher, and H. Stöcker, Influence of the hadronic phase on observables in ultra-relativistic heavy ion collisions, Phys. Rev. C **95**, 064902 (2017), arXiv:1703.06638 [nucl-th].
[56] H. Wolter *et al.* (TMEP), Transport model comparison studies of intermediate-energy heavy-ion collisions, Prog. Part. Nucl. Phys. **125**, 103962 (2022), arXiv:2202.06672 [nucl-th].
[57] J. Hammelmann, M. Bluhm, M. Nahrang, and H. Elfner, Fate of critical fluctuations in an interacting hadronic medium using maximum entropy distributions, (2023), arXiv:2310.06636 [nucl-th].
[58] A. Sorensen and V. Koch, Phase transitions and critical behavior in hadronic transport with a relativistic density functional equation of state, Phys. Rev. C **104**, 034904 (2021), arXiv:2011.06635 [nucl-th].

PROCEEDINGS OF SCIENCE

Towards a SM prediction for CP violation in charm

Alexander Lenz,[a] **Maria Laura Piscopo**[a,*] **and Aleksey V. Rusov**[a]

[a] *Theoretische Teilchenphysik, Center for Particle Physics Siegen, Physik Department, Universität Siegen, Walter-Flex-Str. 3, 57068 Siegen, Germany*

E-mail: alexander.lenz@uni-siegen.de, maria.piscopo@uni-siegen.de, rusov@physik.uni-siegen.de

We provide an overview of the current experimental and theoretical status of charm CP violation and discuss recent progress in obtaining a Standard Model prediction for $\Delta a_{\rm CP}^{\rm dir}$ using the framework of light-cone sum rules. Furthermore, we present new results for the ratios of the direct CP asymmetries and of the branching fractions for the modes $D^0 \to \pi^+\pi^-$ and $D^0 \to K^+K^-$.

20th International Conference on B-Physics at Frontier Machines (Beauty2023)
3-7 July, 2023
Clermont-Ferrand, France

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

arXiv:2403.02267v1 [hep-ph] 4 Mar 2024

1. Introduction

The charm sector offers a unique system for testing the Standard Model of particle physics (SM), see Ref. [1] for a recent review. The peculiarities of charm are twofold. On the one hand, achieving precise theoretical predictions for charm observables is currently very challenging. This follows from the value of the charm quark mass which lies at the boundary between the heavy and the light quark regimes, such that the typical theoretical methods employed for the study of heavy hadrons might be less suitable or even inapplicable for the description of charmed systems. The behaviour of both the perturbative and power expansions becomes, in fact, a priori questionable as

$$\alpha_s(m_c) \sim 0.35\,, \qquad \frac{\Lambda_{\rm QCD}}{m_c} \sim 0.30\,. \tag{1}$$

On the other hand, charmed hadrons provide essential complementary information with respect to kaon- and b-physics, constituting, for instance, the only system to study meson-mixing in the up-quark sector. Additionally, the sensitivity to potential new physics (NP) contributions is high for charm observables, as pronounced cancellations often affect their SM predictions. The latter follow from the Glashow-Iliopoulos-Maiani (GIM) mechanism due to $m_b, m_s, m_d \ll m_W$, as well as from the size of the relevant elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix i.e. $\lambda_q \equiv V_{cq}^* V_{uq}$, namely

$$\lambda_d = -0.21874 + 2.51 \times 10^{-5} i\,, \qquad \lambda_s = 0.21890 + 0.13 \times 10^{-5} i\,, \tag{2}$$

$$\lambda_b = 6.3 \times 10^{-5} - 1.4 \times 10^{-4} i\,. \tag{3}$$

In particular, having λ_b the biggest relative imaginary part but being much smaller in magnitude compared to $\lambda_{d,s}$, the amount of CP violation in the charm sector is expected to be small in the SM. Testing this result against the experimental data, although theoretically difficult, is clearly a task of primary importance in order to strengthen the current understanding of the SM and search for NP.

2. Experimental status of CP violation in charm

The observation of CP violation in the charm sector was made in 2019 by the LHCb Collaboration [2], by measuring the difference of the time-integrated CP asymmetries in the $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ modes, that is $\Delta A_{\rm CP} \equiv A_{\rm CP}(K^+K^-) - A_{\rm CP}(\pi^+\pi^-)$. The corresponding difference of the direct CP asymmetries in the above channels turned out to be

$$\Delta a_{\rm CP}^{\rm dir}\big|_{\rm exp} = (-15.7 \pm 2.9) \times 10^{-4}\,. \tag{4}$$

Recently, also a measurement of the CP asymmetry in $D^0 \to K^+K^-$ was published by the LHCb Collaboration [3], which yields, when combined with the result in Eq. (4), the following values for the direct CP asymmetries in the two individual modes, namely

$$a_{\rm CP}^{\rm dir}(K^+K^-)\big|_{\rm exp} = (7.7 \pm 5.7) \times 10^{-4}\,, \qquad a_{\rm CP}^{\rm dir}(\pi^+\pi^-)\big|_{\rm exp} = (23.2 \pm 6.1) \times 10^{-4}\,. \tag{5}$$

While the result for $a_{\rm CP}^{\rm dir}(\pi^+\pi^-)$ provides the first evidence for CP violation in a specific D-meson decay, a clear theoretical interpretation of the measurements in Eqs. (4), (5) is currently still missing, particularly as the values of the individual CP asymmetries in Eq. (5) would imply a surprisingly large breaking of the U-spin symmetry [4].

3. Theoretical status of CP violation in charm

Exclusive hadronic decays of charmed hadrons pose significant challenges for robust theoretical predictions and although several studies have been carried out in the literature, no unanimous conclusion has yet been reached on the origin of the experimental value of $\Delta a_{\rm CP}^{\rm dir}$.
Naive estimates, see e.g. Ref. [5], point towards a value of $\Delta a_{\rm CP}^{\rm dir}$ about an order of magnitude smaller than the one in Eq. (4). This result was confirmed in Refs. [6, 7] using the framework of light-cone sum rule (LCSR) [8], and analogous conclusions were also obtained in a recent study of final state interactions [9]. Consequently, following these findings, several investigations of possible NP scenarios have been triggered in the effort to accommodate the experimental data [10].
Furthermore, also SM interpretations of the experimental value of $\Delta A_{\rm CP}$ have been advanced. These include analyses based on U-spin relations, see e.g. Ref. [11], as well as studies of rescattering contributions [12] and of final state interactions [13]. In particular, in Ref. [12], the possibility that nearby resonances, like the $f_0(1710)$ or $f_0(1790)$, could lead to a large enhancement of the SM value of $\Delta A_{\rm CP}$, was pointed out. No sign of this effect, however, has been observed in the analysis of Ref. [9], and the latter work also indicated some inconsistencies in Ref. [13]. Finally, approaches based on topological diagram analyses have also been employed [14], although these often rely on qualitative studies and do not provide a first principle determination.

4. Theory of the decays $D^0 \to \pi^+\pi^-$ and $D^0 \to K^+K^-$

Using the unitarity of the CKM matrix $\lambda_d + \lambda_s + \lambda_b = 0$, the amplitudes for the non-leptonic decays $D^0 \to \pi^+\pi^-$ and $D^0 \to K^+K^-$ can be recast in the form [6]

$$\mathcal{A}(D^0 \to \pi^+\pi^-) = \lambda_d\, \mathcal{A}_{\pi\pi} \left[1 - \frac{\lambda_b}{\lambda_d} \frac{\mathcal{P}_{\pi\pi}}{\mathcal{A}_{\pi\pi}}\right], \tag{6}$$

$$\mathcal{A}(D^0 \to K^+K^-) = \lambda_s\, \mathcal{A}_{KK} \left[1 - \frac{\lambda_b}{\lambda_s} \frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}}\right], \tag{7}$$

by singling out, respectively, the contribution due to the CKM dominant combination $\lambda_{d,s}$ from that of the strongly suppressed factor λ_b, with the definitions

$$\mathcal{A}_{\pi\pi} = \langle \pi^+\pi^- | O^d | D^0 \rangle - \langle \pi^+\pi^- | O^s | D^0 \rangle\,, \tag{8}$$

$$\mathcal{A}_{KK} = \langle K^+K^- | O^s | D^0 \rangle - \langle K^+K^- | O^d | D^0 \rangle\,, \tag{9}$$

and

$$\mathcal{P}_{\pi\pi} = \langle \pi^+\pi^- | O^s | D^0 \rangle\,, \qquad \mathcal{P}_{KK} = \langle K^+K^- | O^d | D^0 \rangle\,. \tag{10}$$

In Eqs. (8) - (10), the notation $O^q \equiv -(G_F/\sqrt{2}) \sum_{i=1,2} C_i O_i^q$ is used, where $O_1^q = (\bar{q}^i \Gamma_\mu c^i)(\bar{u}^j \Gamma^\mu q^j)$ and $O_2^q = (\bar{q}^i \Gamma_\mu c^j)(\bar{u}^j \Gamma^\mu q^i)$, with $q = d, s$, denote the current-current operators in the weak effective Hamiltonian describing the charm-quark transitions $c \to q\bar{q}u$ [15], and $C_{1,2}$ are the corresponding Wilson coefficients. The leading CKM amplitudes $\mathcal{A}_{\pi\pi}$, $\mathcal{A}_{KK}$ in Eqs. (6), (7) receive contributions from color-allowed tree-level, exchange and penguin topologies, whereas only the penguin topology can contribute to $\mathcal{P}_{\pi\pi}$, $\mathcal{P}_{KK}$, cf. Fig. 1.

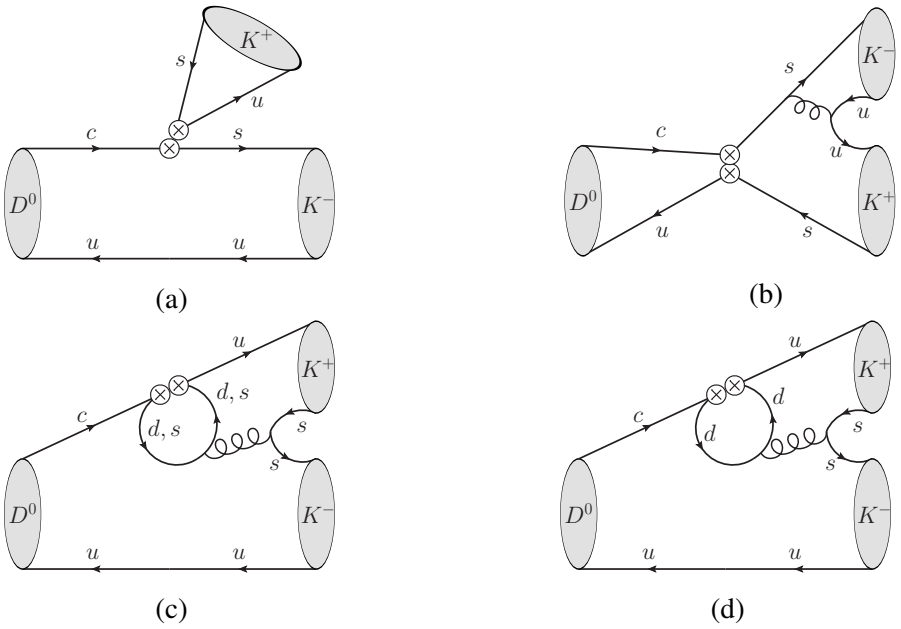


Figure 1: Examples of tree-level (a), exchange (b) and penguin (c) topologies contributing to $\mathcal{A}_{KK}$. Example of penguin topology contributing to $\mathcal{P}_{KK}$ (d). The corresponding diagrams for $\mathcal{A}_{\pi\pi}$ and $\mathcal{P}_{\pi\pi}$ can be obtained replacing $K \to \pi$, $s \leftrightarrow d$.

Considering for simplicity only the decay $D^0 \to K^+K^-$, the corresponding branching fraction reads

$$\mathcal{B}(D^0 \to K^+K^-) \propto |\lambda_s|^2 |\mathcal{A}_{KK}|^2 \left| 1 - \frac{\lambda_b}{\lambda_s} \frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}} \right|^2 , \tag{11}$$

up to phase-space and normalisation factors. Similarly, the direct CP asymmetry, defined as

$$a_{\rm CP}^{\rm dir}(f) \equiv \frac{\Gamma(D^0 \to f) - \Gamma(\overline{D}^0 \to \bar{f})}{\Gamma(D^0 \to f) + \Gamma(\overline{D}^0 \to \bar{f})} , \tag{12}$$

becomes

$$a_{\rm CP}^{\rm dir}(K^+K^-) = - \frac{2 \left| \frac{\lambda_b}{\lambda_s} \right| \sin\gamma \left| \frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}} \right| \sin\phi_{KK}}{1 - 2 \left| \frac{\lambda_b}{\lambda_s} \right| \cos\gamma \left| \frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}} \right| \cos\phi_{KK} + \left| \frac{\lambda_b}{\lambda_s} \right|^2 \left| \frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}} \right|^2} , \tag{13}$$

where we have defined the strong phase difference $\phi_{KK} \equiv \arg(\mathcal{P}_{KK}/\mathcal{A}_{KK})$, and introduced the angle $\gamma \equiv -\arg(\lambda_b/\lambda_s)$. Analogous expressions can be straightforwardly obtained for the mode $D^0 \to \pi^+\pi^-$ by replacing $KK \to \pi\pi$, $\lambda_s \to \lambda_d$ and $\sin\gamma \to -\sin\gamma$ in Eqs. (11), (13).
Taking into account the large hierarchy $\lambda_b/\lambda_{d,s} \ll 1$, it follows that the amplitudes $\mathcal{A}_{\pi\pi}$, $\mathcal{A}_{KK}$ give the dominant contribution to the branching fractions, i.e.

$$\mathcal{B}(D^0 \to \pi^+\pi^-) \sim |\lambda_d|^2 |\mathcal{A}_{\pi\pi}|^2 , \qquad \mathcal{B}(D^0 \to K^+K^-) \sim |\lambda_s|^2 |\mathcal{A}_{KK}|^2 , \tag{14}$$

whereas the direct CP asymmetries are driven only by the ratio of the penguin over the CKM leading amplitudes, that is

$$a_{\rm CP}^{\rm dir}(\pi^+\pi^-) \simeq 2\left|\frac{\lambda_b}{\lambda_d}\right|\sin\gamma\left|\frac{\mathcal{P}_{\pi\pi}}{\mathcal{A}_{\pi\pi}}\right|\sin\phi_{\pi\pi}\,, \qquad a_{\rm CP}^{\rm dir}(K^+K^-) \simeq -2\left|\frac{\lambda_b}{\lambda_s}\right|\sin\gamma\left|\frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}}\right|\sin\phi_{KK}\,. \tag{15}$$

Finally, the above results, together with $|\lambda_d| \simeq |\lambda_s|$, yield the following expression for the difference of direct CP asymmetries $\Delta a_{\rm CP}^{\rm dir}$, namely

$$\Delta a_{\rm CP}^{\rm dir} \simeq -2\left|\frac{\lambda_b}{\lambda_s}\right|\sin\gamma\left(\left|\frac{\mathcal{P}_{KK}}{\mathcal{A}_{KK}}\right|\sin\phi_{KK} + \left|\frac{\mathcal{P}_{\pi\pi}}{\mathcal{A}_{\pi\pi}}\right|\sin\phi_{\pi\pi}\right)\,. \tag{16}$$

5. Determination of $\Delta a_{\rm CP}^{\rm dir}$ within LCSR

A first computation of the penguin amplitudes $\mathcal{P}_{\pi\pi}, \mathcal{P}_{KK}$ was performed in Ref. [6] using the framework of LCSR with, respectively, pion and kaon light-cone distribution amplitudes (LCDAs), and following previous studies for the $B \to \pi\pi$ decay [16]. The values of $|\mathcal{A}_{\pi\pi}|$ and $|\mathcal{A}_{KK}|$ needed to determine the direct CP asymmetries were instead extracted, taking into account the relations in Eq. (14), from the precise experimental data on the branching ratios [17]

$$\mathcal{B}(D^0 \to \pi^+\pi^-)\big|_{\rm exp} = (1.454 \pm 0.024) \times 10^{-3}\,, \tag{17}$$

$$\mathcal{B}(D^0 \to K^+K^-)\big|_{\rm exp} = (4.08 \pm 0.06) \times 10^{-3}\,. \tag{18}$$

The authors of Ref. [6] obtained the following SM bound for the difference of CP asymmetries

$$|\Delta a_{\rm CP}^{\rm dir}|_{\rm SM} \leq 2.3 \times 10^{-4}\,, \tag{19}$$

which is about a factor of 6 lower than the experimental value in Eq. (4). Recently, a study of the leading decay amplitudes $\mathcal{A}_{\pi\pi}$, $\mathcal{A}_{KK}$ has been performed in Ref. [7], where the corresponding tree-level matrix elements have also been determined using LCSR with pion and kaon LCDAs, see also Ref. [18] for more details on the general framework. Specifically, from naive power counting, Eqs. (8), (9) can be expressed as

$$\mathcal{A}_{\pi\pi} = \langle\pi^+\pi^-|O^d|D^0\rangle\Big|_{\rm tree} + O(\alpha_s) + O(1/m_c)\,, \tag{20}$$

$$\mathcal{A}_{KK} = \langle K^+K^-|O^s|D^0\rangle\Big|_{\rm tree} + O(\alpha_s) + O(1/m_c)\,, \tag{21}$$

retaining the dominant contribution due to the tree-level amplitude and neglecting sub-leading diagrams due to both hard and soft QCD corrections. In this approximation it follows that

$$\mathcal{A}_{\pi\pi} \simeq -\frac{G_F}{\sqrt{2}}\left(C_1 + \frac{C_2}{3}\right)\langle\pi^+\pi^-|O_1^d|D^0\rangle\Big|_{\rm tree}\,, \tag{22}$$

$$\mathcal{A}_{KK} \simeq -\frac{G_F}{\sqrt{2}}\left(C_1 + \frac{C_2}{3}\right)\langle K^+K^-|O_1^s|D^0\rangle\Big|_{\rm tree}\,. \tag{23}$$

A first estimate of the matrix elements in Eqs. (22), (23) can be derived using the naive QCD factorisation approximation, which, surprisingly, already yields values for the branching fractions in very good agreement with the experimental data [7]. Furthermore, the computation of the tree-level matrix elements within the framework of LCSR gives [7]

$$\mathcal{B}(D^0 \to \pi^+\pi^-)\big|_{\rm LCSR} = \left(1.40^{+1.53}_{-1.06}\right) \times 10^{-3}\,, \tag{24}$$

$$\mathcal{B}(D^0 \to K^+K^-)\big|_{\rm LCSR} = \left(3.67^{+3.90}_{-2.69}\right) \times 10^{-3}\,, \tag{25}$$

where the central values again agree very well with the data, however, the uncertainties are large and mostly follow from a conservative treatment of missing contributions. Importantly, these results do not indicate any sign of potential large enhancement due to subleading topologies. On the other hand, in the ratio of branching fractions many theoretical uncertainties cancel, leading, after taking into account correlations due to common inputs, to the significantly more precise prediction

$$\left.\frac{\mathcal{B}(D^0 \to K^+K^-)}{\mathcal{B}(D^0 \to \pi^+\pi^-)}\right|_{\rm LCSR} = 2.63 \pm 0.86\,, \tag{26}$$

which perfectly reproduces the observed size of $SU(3)_F$ breaking in the two modes, namely

$$\left.\frac{\mathcal{B}(D^0 \to K^+K^-)}{\mathcal{B}(D^0 \to \pi^+\pi^-)}\right|_{\rm exp} = 2.81 \pm 0.06\,. \tag{27}$$

Combining the LCSR results for the penguin and tree-level amplitudes, as computed in Ref. [6] and Ref. [7], respectively, then yields the following value for the ratio of the two direct CP asymmetries defined in Eq. (15), i.e.

$$\left.\frac{a_{\rm CP}^{\rm dir}(K^+K^-)}{a_{\rm CP}^{\rm dir}(\pi^+\pi^-)}\right|_{\rm LCSR} = -0.65^{+0.09}_{-0.08}\,, \tag{28}$$

where we have used the estimates of the strong phases as obtained in Ref. [6], since the sensitivity to potential missing contributions not yet accounted in the LCSR result is expected to be softened in the ratio $\sin\phi_{KK}/\sin\phi_{\pi\pi}$. The value in Eq. (28) is well consistent with -1, the result that would be obtained in the limit of exact U-spin symmetry, and must be compared with the corresponding experimental ratio

$$\left.\frac{a_{\rm CP}^{\rm dir}(K^+K^-)}{a_{\rm CP}^{\rm dir}(\pi^+\pi^-)}\right|_{\rm exp} = 0.33^{+0.45}_{-0.26}\,. \tag{29}$$

Finally, allowing for arbitrary strong phase differences, that is varying both $\sin\phi_{\pi\pi}$ and $\sin\phi_{KK}$ from -1 to 1, the bound on $\Delta a_{\rm CP}^{\rm dir}$ obtained entirely using LCSR, reads [7]

$$|\Delta a_{\rm CP}^{\rm dir}|_{\rm LCSR} \leq 2.4 \times 10^{-4}\,, \tag{30}$$

in perfect agreement with the result in Eq. (19), and again about a factor of 6 lower than the corresponding measurement.

6. Conclusion

We have briefly described the current experimental and theoretical status of charm CP violation and discussed recent progress obtained in the study of the hadronic decays $D^0 \to K^+K^-$ and $D^0 \to \pi^+\pi^-$ using the framework of LCSR. In particular, we have shown that first steps towards a description of the corresponding branching fractions using this method yields very promising results [1], and that LCSR leads to a bound on the value of $\Delta a_{\rm CP}^{\rm dir}$ in the SM which is about a factor of 6 lower than the experimental data.

7. Acknowledgements

MLP is very grateful to the organisers of *Beauty2023* for the invitation and for creating a lively atmosphere rich of fruitful discussions. Moreover, AL and MLP would like to thank the participants of the recent LHCb school in Meinerzhagen for the interesting discussions and for proposing the computation of some of the observables presented in this manuscript. The work of MLP is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 500314741.

References

[1] A. Lenz and G. Wilkinson, Ann. Rev. Nucl. Part. Sci. **71**, 59-85 (2021).

[2] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **122** (2019) no.21, 211803.

[3] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **131** (2023) no.9, 091802.

[4] S. Schacht, JHEP **03** (2023), 205; R. Bause, H. Gisbert, G. Hiller, T. Höhne, D. F. Litim and T. Steudtner, Phys. Rev. D **108** (2023).

[5] Y. Grossman, A. L. Kagan and Y. Nir, Phys. Rev. D **75** (2007), 036008.

[6] A. Khodjamirian and A. A. Petrov, Phys. Lett. B **774** (2017), 235-242.

[7] A. Lenz, M. L. Piscopo and A. V. Rusov, [arXiv:2312.13245 [hep-ph]].

[8] I. I. Balitsky, V. M. Braun and A. V. Kolesnichenko, Sov. J. Nucl. Phys. **44** (1986), Nucl. Phys. B **312** (1989); V. L. Chernyak and I. R. Zhitnitsky, Nucl. Phys. B **345** (1990).

[9] A. Pich, E. Solomonidi and L. Vale Silva, Phys. Rev. D **108** (2023).

[10] M. Chala, A. Lenz, A. V. Rusov and J. Scholtz, JHEP **07** (2019), 161; A. Dery and Y. Nir, JHEP **12** (2019), 104; L. Calibbi, T. Li, Y. Li and B. Zhu, JHEP **10** (2020), 070; R. Bause, H. Gisbert, M. Golz and G. Hiller, Phys. Rev. D **101** (2020).

[11] Y. Grossman and S. Schacht, JHEP **07** (2019), 020.

[1] See Section 6 of Ref. [7] for a list of future improvements.

[12] S. Schacht and A. Soni, Phys. Lett. B **825** (2022).

[13] I. Bediaga, T. Frederico and P. C. Magalhães, Phys. Rev. Lett. **131** (2023).

[14] H. N. Li, C. D. Lü and F. S. Yu, [arXiv:1903.10638 [hep-ph]]; H. Y. Cheng and C. W. Chiang, Phys. Rev. D **100** (2019); D. Wang, C. P. Jia and F. S. Yu, JHEP **21** (2020), 126.

[15] G. Buchalla, A. J. Buras and M. E. Lautenbacher, Rev. Mod. Phys. **68** (1996).

[16] A. Khodjamirian, Nucl. Phys. B **605** (2001); A. Khodjamirian, T. Mannel and B. Melic, Phys. Lett. B **571** (2003).

[17] R. L. Workman *et al.* [Particle Data Group], PTEP **2022** (2022), 083C01.

[18] M. L. Piscopo and A. V. Rusov, JHEP **10** (2023), 180.

Prepared for submission to JHEP

The U-Matrix geometrical model for multi-particle production in high-energy hadronic collisions

Rami Oueslati [a] **and Adel Trabelsi** [b]

[a] *Space sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, Bât. B5a, 4000 Liège, Belgium*

[b] *Tunis El Manar University, Faculty of Sciences of Tunis, Nuclear Physics and High Energy Research Unit, Tunis, Tunisia*

E-mail: rami.oueslati@uliege.be, adel.trabelsi@cern.ch

Abstract: Inspired by the picture portraying the KNO scaling violation as an extension of the geometrical scaling violation, the current study proposes a phenomenological model for multi-particle production in hadron collisions based on the geometrical approach and using the U-Matrix unitarization scheme of the scattering amplitude. The model has been fine-tuned and all parameters have been derived from optimal fits to various hadronic multiplicity distributions data in $p + p(\bar{p})$ collisions across a broad range of energies. The results have revealed that our model furnishes a reasonable description of diverse multiplicity distributions at various energies. Besides, they have demonstrated a pronounced violation of the geometrical scaling, which eventually resulted in a significant violation of the KNO scaling. The study has also analyzed the higher-order moments of the multiplicity distribution. We have observed an unexpected overestimation of the fluctuations and correlations between final state particles with increasing energy, particularly above LHC energy. It is claimed that this overestimation is due to statistical fluctuations embedded in the U-matrix scheme. The findings of this study have shed light on the key role of the U-matrix scheme in the impact of collision geometry on multi-particle production processes at high energy.

Keywords: QCD Phenomenology; Unitarity constraint; geometrical and KNO scaling violation; particle correlation

arXiv:2403.02263v1 [hep-ph] 4 Mar 2024

Contents

1 Introduction **1**

2 The theoretical framework of the Model **4**

3 Explicit model and data **8**

4 Results and discussion **11**
- 4.1 Geometrical scaling violation 11
- 4.2 Hadronic Multiplicity Distributions 14
- 4.3 KNO scaling violation 17
- 4.4 Hadronic multi-particle correlations 19

5 Conclusions **22**

1 Introduction

Throughout the years, the study of multi-particle production in hadron collisions at high energies has sparked the interest of both theoretical physicists and experimentalists given its significance as it offers valuable insights into the intricate mechanisms underlying the production of particles [1–5].

Similarly, the hadronic multi-particle production is primordial at ultra-high energies. Indeed, it is necessary to have a solid grasp of it so as to properly interpret air-shower cosmic ray observables, which are obtained through a simulation of a wide range of event generators based on Monte Carlo models. These models have been created and adjusted to describe hadronic multi-particle production in man-made accelerator experiments. Despite some minor differences in their fundamental assumptions, the majority of them employ the eikonal as the scheme of the unitarisation of the scattering amplitude. They must also be internally consistent as we depend on them for extrapolating to ultra-high energy scenarios. Nevertheless, modelling hadronic interactions is not devoid of uncertainties, which gives rise to model-dependent outcomes. Among them, one can cite the mass composition problem and the long-standing muon puzzle, which is regarded as one of the most significant issues in hadronic interaction physics [6].

Having said that, advances in the phenomenology and theory of multi-particle production may provide a solution to the aforementioned issues.

The study of multi-particle production may be carried out, particularly, through examining the multiplicity distributions of the produced particles. Their analysis is crucial since the patterns observed in them help reveal the complexity of the collision process, shedding light on the interactions involved in particle creation. In fact, this distribution has been

thoroughly investigated by experimental collaborations at the LHC, including ALICE, ATLAS, CMS, and LHCb [7], to enhance our comprehension of the rudimentary characteristics of strong interactions and the behaviour of matter under extreme circumstances as well as to verify theoretical predictions.

Besides, the study of the higher-order moments of the multiplicity distribution is of paramount importance as they provide insight into particle features in high-energy collisions, such as correlation and collective behaviour. Indeed, the behaviour of these moments has been the subject of various experimental studies over a wide range of collision energies [8, 9], furnishing valuable data which are crucial in constraining theoretical models.

The multiplicity distribution $P_n(s)$ in hadron collisions refers to the distribution of the number of produced particles in a collision event. It is influenced by various factors, such as the colliding hadrons' energy, the collision geometry, and the underlying dynamics of the interaction. More specifically, it is given by the ratio of the topological cross-section σ_n to the inelastic cross-section σ_{in}. The topological cross-section represents the probability or rate of observing a specific configuration of particles with a particular multiplicity value n. It is derived from the scattering amplitudes, which are influenced by the underlying scattering processes and dynamics. The inelastic cross-section, on the other hand, characterizes the probability of observing any final-state interaction or particle production, regardless of the specific configuration or multiplicity. It is closely related to the total cross-section and includes contributions from various interaction channels, including both diffractive and non-diffractive processes.

It is important to note that these cross-sections, and thus the multiplicity distribution P_n, cannot yet be calculated within the framework of quantum chromodynamics (QCD). Therefore, our current understanding of multi-particle production dynamics relies primarily on phenomenological approaches as well as certain underlying theoretical principles, which form the basis for a wide range of models [7]. Needless to say, the continuous enhancement of the theoretical models for particle production is critical so as to maintain consistency and coherence with the foundational principles of the Quantum Field Theory (QFT).

One of the key theoretical tenets in the construction of phenomenological models, such as those based on the geometrical or string approach, is the unitarity constraint [10–13]. These models often utilize the eikonal scheme as a means to unitarize the scattering amplitude and describe inclusive multiplicity distributions. While this scheme provides a reasonable description of certain hadronic observables, there are compelling reasons to explore alternative schemes for unitarizing the scattering amplitude, namely the U-matrix one.

In a recent study [14], it was shown that despite unexpected agreement in the description of the inelastic pp and $p\bar{p}$ cross-sections with the eikonal and U-matrix schemes, a divergence in the individual order-by-order amplitudes can potentially impact the topological cross-section by influencing the relative probabilities of different multiplicity configurations and capturing specific physical processes.

By the same token, in another study [15], it has been found that the U-matrix scheme exhibits a slightly better fit to the single diffractive data at high-energies and a faster growth at ultra-high energies compared to the eikonal scheme. This result implies that

the underlying physics and dynamics of the single diffractive scattering are sensitive to the choice of scheme, especially at ultra-high energy, further reinforcing the significance of considering the U-matrix. Hence, it is worth noting that if the discrepancy in the single diffractive cross-section propagates to the topological cross-section, it can affect the relative probabilities or rates of observing different multiplicity configurations. Moreover, as the inelastic cross-section encompasses both diffractive and non-diffractive processes, the different behaviours in the single diffractive cross-section between unitarization schemes can contribute to variations in the overall inelastic cross-section.

It should be noted that the specific impact on the topological and inelastic cross-sections would depend on the detailed correlations and interplay between different interaction channels, including both diffractive and non-diffractive processes. In fact, when hadrons collide at high energies, several interactions and processes take place, leading to the formation of large numbers of particles with various species. Given the complex nature of these interactions, deciphering the contributions from individual channels and identifying the specific processes behind multi-particle production appears to be a daunting task [16].

On the whole, the inelastic and topological cross-sections are scheme-dependent, implying that the multiplicity distribution P_n is scheme-dependent as well. This will, therefore, impact the description of the multiplicity distribution in terms of its shape, magnitude, or other characteristic features.

An additional reason for considering an alternative to the eikonal unitarization scheme is related to the geometrical scaling, as supported by experimental observations. This scaling is a regularity established by the ISR measurements of the proton-proton and proton-antiproton scattering and refers to the invariability of the ratio between elastic and total cross-sections. Interestingly, experiments conducted at the CERN (SPS) collider have revealed that this regularity is violated when the energy surpasses the ISR energy range [17, 18]. Furthermore, from a theoretical perspective, it has been shown that the violation of the geometrical scaling is more pronounced with the U-matrix scheme than with the eikonal as energy increases [14]. It should be emphasized that this remarkable discrepancy has proved to occur mainly when we venture into the extremely high-energy region, roughly near the Grand Unification Scale. In this regard, one of the preliminary objectives of the present study is to examine this disparity within an accessible energy range.

Another regularity put forth by the ISR measurements is the KNO scaling so named after its proponents Koba, Nielsen, and Olesen (KNO) [19]. It refers to the constancy of normalized moments in the distribution of multiplicities. Practically, the KNO function is often employed for the examination of multiplicity distributions in particle collisions. It is denoted as $\Phi(z)$, where $\langle n \rangle$ stands for the average of the multiplicity distribution, and $z = n/\langle n \rangle$ represents the normalized multiplicity. It is noteworthy that the KNO function $\Phi(z)$ tends to be independent of the collision energy $\sqrt{s}$ within the ISR energy range. However, considerable deviations from the KNO scaling start to occur at higher energies, such as those attained at FNAL and LHC.

As a matter of fact, several phenomenological studies have demonstrated the existence of a connection between the geometrical scaling of the profile function and the KNO scaling

of the multiplicity distribution [20–22]. More precisely, these models have shown that the violation of the geometrical scaling, indicated by an increase in the ratio of elastic to total cross-sections (σ_{el}/σ_{tot}) between ISR and collider energies, is associated with the violation of the KNO scaling across these energy ranges within the eikonal scheme. This raises the question of potential consequences for the violation of the KNO scaling within the context of the U-matrix scheme. It is expected that the latter, with its modifications to the scattering amplitude, may introduce novel dynamics and fluctuations that influence the statistical behaviour of particle production. These effects can in turn influence the universal properties assumed in the KNO scaling and significantly lead to a violation or alteration of the constancy of normalized moments.

In view of all that has been mentioned so far, one can assume that considering an alternative adequate unitarization scheme can considerably change our understanding of the description of multi-particle production. In the present study, we propose a phenomenological model for multi-particle production in hadron-hadron collisions that hinges on both the geometrical approach and the picture depicting the KNO scaling violation as an extension of the geometrical scaling violation, using the U-Matrix scheme. The chief purposes of the study are to examine the geometrical scaling violation within an accessible energy range, describe the hadronic multiplicity distributions in full-phase space over a wide energy range in $p+p(\bar{p})$ collisions and to investigate the KNO scaling violation. It also aims at probing the correlation between the final particles and revealing the physics that underlies it. In particular, it seeks to provide valuable insights into the relationship between the violation of both the geometrical and KNO scaling and the mechanism of particle production within the U-matrix scheme.

The remainder of the paper is organized as follows: In Section II, the theoretical background of the multi-particle production model will be outlined. Section III will focus on the explicit model and the data used. Section IV will present and discuss the results. Finally, Section V will summarize the findings and discuss the limitations and implications of this work so as to guide future research.

2 The theoretical framework of the Model

The multiplicity distribution $P_n(s)$, i.e., the probability of producing n charged particles in an inelastic $p+p(\bar{p})$ collision at the energy $\sqrt{s}$, is given by

$$P_n(s) = \frac{\sigma_n(s)}{\sigma_{in}(s)} \tag{2.1}$$

where $\sigma_n(s)$ and $\sigma_{in}(s)$ are the n-particle topological and inelastic cross-sections, respectively, with $\sum_n \sigma_n(s) = \sigma_{in}(s)$.

Great efforts have been devoted to explaining why the normalized moments in this multiplicity distribution remain constant in the ISR energy range but diverge from this pattern from the LHC energy range. A possible explanation for this phenomenon might be ascribed to the increasing importance of mini-jets production, resulting from both soft

and semi-hard partonic processes, as energy increases. In fact, these mini-jets not only contribute to the rapid growth of hadron-hadron cross-sections, as demonstrated by several geometrical models [14, 15, 23, 24], but may also play a crucial role in the violation of the KNO scaling.

It goes without saying that the geometrical models based on the impact parameter space representation provide a good description of various aspects of hadron collisions at high energies. Technically speaking, by considering the impact parameter, which quantifies the distance between the colliding particles' centres, these models actually furnish a solid framework for understanding the initial stages of collisions, allowing a systematic exploration of the collision geometry, ranging from central (small impact parameters) to peripheral (large impact parameters) collisions. Therefore, this geometrical approach links the collision geometry to the underlying physics processes, enabling us to infer certain properties. We particularly assume that this approach can also shed light on the intricate interplay between collision geometry and particle multiplicity distribution.

In addition, as has been mentioned in the previous section, the KNO scaling violation can be perceived as an extension of the geometrical scaling violation to the multi-particle production process. This extension highlights that the dynamics governing particle production become more complex and energy-dependent than what a purely geometrical approach can capture. It also suggests that additional physical processes, such as parton interactions, collective effects, or fluctuations, play a role in shaping the multiplicity distribution as collision energies increase. As a result, a more accurate description of the geometrical scaling violation is needed, which in turn will provide a better description of the violation extension to multi-particle production. In fact, the violation of geometrical scaling in impact parameter space occurs when the initial assumptions about simple geometrical overlap and scaling behaviour break down due to more intricate particle-particle interactions or energy-dependent effects. Consequently, this concept harmoniously aligns with the utilization of the impact parameter space representation, emphasizing the significance of collision geometry.

Overall, when combined with the geometrical scaling tenets, the geometrical models can be enhanced and thus provide a solid framework for describing the multiplicity distribution and uncovering the underpinning universal features of particle production in high-energy hadron collisions. They also furnish a reliable tool that serves to explain the impact of collision geometry on the multiplicity distribution and to analyze it across a broad range of collision energies and particle species, improving prediction and comparison accuracy in multiple collision systems.

To construct our model highlighting the intrinsic relationship between geometrical scaling and KNO scaling within a geometrical approach, we draw inspiration from [22, 25–27], where the essence of the approach lies in expressing the overall hadronic multiplicity distributions in the inelastic channel through a combination of an elementary distribution and the inelastic overlap function. Thus, this approach establishes a direct connection between the fluctuations in multiplicity and the fluctuations in the impact parameter, driven by the intuitive understanding that collisions occurring at shorter distances tend to be more violent and yield a higher number of fragments, thereby leading to increased

multiplicities. Therefore, in order to establish a reliable geometrical approach to calculate the multiplicity distribution in the impact parameter space representation, we rely on the unitarity condition of the S matrix. This involves examining the elastic scattering amplitude in the impact parameter b space, expressed by the equation:

$$2\,\mathrm{Im}[\Gamma(s,b)] = |\Gamma(s,b)|^2 + G_{in}(s,b), \tag{2.2}$$

where $\Gamma(s,b)$ denotes the profile function, i.e., the elastic hadron scattering amplitude, and $G_{in}(s,b)$ represents the inelastic overlap function. By employing the optical theorem, we can obtain

$$\sigma_{tot}(s) = 2\int d^2b\;\mathrm{Im}[\Gamma(s,b)], \tag{2.3}$$

and recognizing that the function $G_{in}(s,b)$ signifies the absorption probability associated with each b value, we can derive the total inelastic cross-section

$$\sigma_{in}(s) = \int d^2b\;G_{in}(s,b). \tag{2.4}$$

Thus, the unitarity condition (2.2) is equivalent to $\sigma_{tot}(s) = \sigma_{el}(s) + \sigma_{in}(s)$, where $\sigma_{el}(s) = \int d^2b\,|\Gamma(s,b)|^2$.

Various research papers in the field have already explored this phenomenological procedure based on the impact parameter space representation known as the geometrical or string approach [11–13, 28]. In this geometrical approach, wherein incident hadrons are treated as spatially extended objects and their collisions are represented as an ensemble of elementary interactions between quarks and/or gluons, the hadronic multiplicity distribution $P_n(s)$ is constructed based on elementary quantities associated with microscopic processes. It follows from this that the overall hadronic multiplicity distribution is estimated through summing contributions emerging from each impact parameter b of the incident hadronic system.

To do this, the topological cross-section σ_n is decomposed into contributions from each impact parameter b with weight $G_{in}(s,b)$ as follows:

$$\begin{aligned}\sigma_n(s) &\equiv \int d^2b\;\sigma_n(s,b)\\ &= \int d^2b\;G_{in}(s,b)\left[\frac{\sigma_n(s,b)}{\sigma_{in}(s,b)}\right],\end{aligned} \tag{2.5}$$

where $\sigma_{in}(s) \equiv \int d^2b\,\sigma_{in}(s,b) = \int d^2b\,G_{in}(s,b)$.

The quantity enclosed in brackets $\sigma_n(s,b)/\sigma_{in}(s,b) \equiv p_n(s,b)$ can be interpreted as the probability of producing n particles at impact parameter b. It accounts for interactions among the elementary components of the colliding hadrons. Keeping in mind that $p_n(s,b)$ should scale in KNO sense given its elementary structure, the multiplicity distribution $P_n(s)$ (2.1) can be reformulated as :

$$P_n(s) = \frac{1}{\sigma_{in}(s)}\int d^2b\,\frac{G_{in}(s,b)}{\langle n(s,b)\rangle}\,[\langle n(s,b)\rangle p_n(s,b)] \tag{2.6}$$

where $\langle n(s,b)\rangle$ represents the average number of particles produced at b and s.

Let $\Phi(s,z) = \langle n(s)\rangle P_n(s)$ be the overall multiplicity distribution in KNO form, where $z = n(s)/\langle n(s)\rangle$ is the corresponding KNO variable. A multiplicity distribution $\phi(s,\underline{z})$, associated with elementary processes occurring at b and s, can be written in the form $\phi(s,\underline{z}) = \langle n(s,b)\rangle p_n(s,b)$, where $\underline{z} = n(s)/\langle n(s,b)\rangle$. It is worth noting that both distributions adhere to the standard normalizations [22] :

$$\int_0^\infty dz\,\Phi(z) = \int_0^\infty d\underline{z}\,\phi(\underline{z}) = 2 \tag{2.7}$$

and

$$\int_0^\infty dz\,z\,\Phi(z) = \int_0^\infty d\underline{z}\,\underline{z}\,\phi(\underline{z}) = 2. \tag{2.8}$$

In this theoretical framework, the unknown function representing the average number of particles produced at a specific impact parameter b and energy s can be comprehended as the product of two factors: The first factor, $\langle n(s)\rangle$, representing the general behaviour depicts the average number of particles generated in a collision, regardless of the specific impact parameter. As for the second, it is denoted as $f(s,b)$ and perceived as a multiplicity function. It describes the variation of the average number of particles in accordance with the impact parameter b quantifying how collision geometry affects particle production. The mathematical expression for this factorization is:

$$\langle n(s,b)\rangle = \langle n(s)\rangle f(s,b) \tag{2.9}$$

and therefore the expression (2.6) can be rewritten in the KNO form as follows:

$$\Phi(s,z) = \langle n(s)\rangle P_n(s) = \frac{1}{\int d^2b\,G_{in}(s,b)} \int d^2b\,\frac{G_{in}(s,b)}{f(s,b)}\,\phi\left(\frac{z}{f(s,b)}\right), \tag{2.10}$$

and we also have

$$P_n(s) = \frac{1}{\langle n(s)\rangle \int d^2b\,G_{in}(s,b)} \int d^2b\,\frac{G_{in}(s,b)}{f(s,b)}\,\phi^{(1)}\left(\frac{z}{f(s,b)}\right), \tag{2.11}$$

The master formula eq. (2.11) highlights the connection between the multiplicity distribution $P_n(s)$ and the unitarisation scheme of the scattering amplitude which can ultimately be established within the inelastic overlap function, emphasizing that this multiplicity is scheme-dependent as stated in the introduction. The full phase space hadronic multiplicity distribution $P_n(s)$ is hence constructed by summing contributions from parton–parton collisions occurring at each value of b. These interactions give rise to the formation of string-like objects, similar to the string formation described by the Lund model [29]. These strings in turn fragment into hadrons. So, in order to stress the fundamental principle of the

string approach, an index labelling the elementary multiplicity distribution $\phi^{(1)}$ has been introduced in the master formula (2.11). The essential elements in this formula eq. (2.11), namely the inelastic overlap function, the average number of particles produced at a specific impact parameter b and energy s, as well as the elementary multiplicity distribution function, will be explicitly presented in the following section, laying the groundwork for an in-depth analysis of the complex interactions between collision geometry and particle production.

3 Explicit model and data

The explicit model presented in this paper for describing High-energy hadronic scattering is theoretically grounded in the geometrical approach and is based on the Reggeon exchanges picture (see, e.g. [30] and references therein). In this picture, hypothetical exchange particles, known as pomerons, mediate interactions between hadrons and the procedure of obtaining the amplitude of a given hadronic process involves summing over all conceivable ways in which pomerons can be exchanged.

Diagrammatically, every pomeron exchange is represented by a line that connects the incoming and outgoing hadrons. Hence, by summing all possible topologies of these diagrams, the total amplitude is determined. Technically, the single-pomeron exchange amplitude, also known as, the Born term can be parameterized as follows:

$$a(s,t) = g_p^2\, \mathcal{F}_1(t)^2 \left(\frac{s}{s_0}\right)^{\alpha(t)} \xi(t), \tag{3.1}$$

where g_p is the pomeron-proton coupling, $\mathcal{F}_1(t)$ denotes the proton elastic form factor, and $\xi(t)$ stands for the signature factor that is given by:

$$\xi(t) = -e^{-i\pi\alpha(t)/2}, \tag{3.2}$$

and the pomeron trajectory, represented by $\alpha(t)$, approximates a straight line:

$$\alpha(t) = 1 + \epsilon + \alpha'_P t. \tag{3.3}$$

In the impact-parameter space representation, where the Fourier transform of the amplitude $a(s,t)$ rescaled by $2s$ corresponds to a partial wave, we have:

$$\chi(s,\mathbf{b}) = \int \frac{\mathrm{d}^2\mathbf{q}}{(2\pi)^2} \frac{a(s,t)}{2s} \exp(\mathrm{i}\mathbf{q}\cdot\mathbf{b}). \tag{3.4}$$

When we venture into high energies, summing amplitudes may lead to a violation of unitarity, especially in the case of multi-pomeron exchanges. Therefore, in order to make the summed amplitudes comply with the unitarity constraint, unitarization techniques come into play by resuming infinite series or introducing additional terms to modify the amplitude's behaviour. As a matter of fact, there is a plethora of techniques in the literature [31–33] whose shared objective is to come up with a consistent approach to summing over various exchange contributions while making sure that the resulting amplitude $\Gamma(b,s)$,

Scheme	ϵ	α'_P (GeV^{-2})	g_p	t_0 (GeV^2)	χ^2/d.o.f
U-matrix	0.10 ± 0.01	0.37 ± 0.28	7.5 ± 0.8	2.5 ± 0.6	1.436
Eikonal	0.11 ± 0.01	0.31 ± 0.19	7.3 ± 0.9	1.9 ± 0.4	1.442

Table 1: χ^2/d.o.f and best-fit parameters obtained using the eikonal and U-matrix unitarisation schemes.

i.e., the hadronic profile function, satisfies the unitarity condition and may account for all features of interactions in the context of hadron collisions. Among them, we can cite the eikonal scheme, which is one of the most commonly used methods, positing that the profile function is provided by:

$$\Gamma_E(s, \boldsymbol{b}) = \mathrm{i}\left[1 - \mathrm{e}^{\mathrm{i}\chi(s,\boldsymbol{b})}\right] , \tag{3.5}$$

The U-matrix scheme is another illustration, which asserts that:

$$\Gamma_U(s, \boldsymbol{b}) = \frac{\chi(s, \boldsymbol{b})}{1 - \mathrm{i}\,\chi(s, \boldsymbol{b})/2}. \tag{3.6}$$

As previously mentioned in the introduction, we will consider that the hadronic profile function $\Gamma(b, s)$ is given by the U-Matrix form (3.6). This function represents the sum of all n-pomeron exchange contributions obtained from the single-pomeron exchange amplitude which is, in turn, related to the expected number $\chi(b, s)$ of interactions between partons of the incident hadrons.

Using both the eikonal and the U-matrix schemes as well as a dipole-like form factor for the proton, where $\mathcal{F}_1(t) = 1/(1 - t/t_0)^2$, the parameters ϵ and α'_P describing the pomeron trajectory, the coupling constant g_p, and the form-factor scale t_0 are adjusted based on a best fit to up-to-date hadron collider data on total, elastic, and inelastic cross-sections. The values of these parameters are provided in Table 1 [14]. Following this adequate description of the hadronic profile function, we can determine the inelastic overlap function $G_{in}(s, b)$, needed in our explicit model, by using the equation presented in (2.2).

Since the primary objective of this work is to examine the influence of geometrical collisions on multi-particle production processes, we shall propose two hypotheses concerning our choice of the elementary multiplicity distribution. First, it is sufficient for our study to consider that, on average, every string created in parton-parton interactions has the same likelihood of producing a certain number of charged hadrons, which is described by the elementary multiplicity distribution $\phi^{(1)}(z)$, even though the strings created may have different probabilities of turning into a pair of charged hadrons. Second, despite the fact that the parametrization of the elementary distribution $\phi^{(1)}(z)$ is key to capturing the overall shape of the multiplicity distribution, these two do not necessarily have similar shapes. This is mainly because the overall distribution is obtained by summing contributions from elementary processes at different impact parameters, in the context of this superposition model eq. (2.11). That is to say, the peculiar combination of the individual contributions emerging from distinct impact parameters and which may be having differing shapes results in an overall distribution that reflects these contributions' combined effects

and whose characteristics should be represented by their superposition. Hence, we will assume, as a first approximation, that the elementary distribution has the same shape as the overall distribution. More specifically, the choice of the functional form for this elementary distribution is motivated by phenomenological fits to data.

As a matter of fact, the Negative Binomial Distribution (NBD) has proved to provide a good description of the experimental data on the multiplicity distributions in the context of high-energy physics and hadron collisions [7]. Therefore, in the present study, the KNO form of the NBD, also known as the Gamma distribution, is adopted for the elementary multiplicity distribution and given by:

$$\phi^{(1)}(z) = 2\,\frac{K^K}{\Gamma(K)}\,z^{K-1}\,e^{-Kz}, \tag{3.7}$$

where K is a dimensionless parameter.

It should be pointed out that there exists a connection between the average number of particles generated at a specific impact parameter b and energy s, $\langle n(s,b)\rangle$, which determines the unknown multiplicity function $f(s,b)$ by the eq. (2.9), and the Born term in b space $\chi(s,b)$.

This link can be attributed to the various roles that the Born term plays. To begin with, the multiplicity of generated particles in b space and the Born term are interrelated given that the former can be impacted by the effective interaction of partons within the colliding hadrons and the latter provides a measure of this effective interaction since it represents the overlap of the colliding matter distributions. Not to mention, the Born term depends on the impact parameter. As such, it reflects how strong the interaction between colliding hadrons is at distinct impact parameters. Indeed, the impact-parameter-dependent strength of the interaction between hadrons may have an impact on the multiplicity of produced particles. Secondly, owing that the Born term is a crucial parameter in describing the energy dependence of the scattering amplitude and that particle generation is more likely to occur at higher energies, the Born term is inextricably connected to the average particle production. Thirdly, on the one hand, the imaginary part of the Born term is related to inelastic processes in high-energy collisions and, on the other hand, multi-particle production is often associated with inelastic interactions, so relying on the imaginary part of the Born term will reflect the possibility of inelastic scattering and subsequent particle production. Therefore, the multiplicity function $f(s,b)$ is constructed to depend on the imaginary part of the Born term, denoted as $\chi_I(s,b)$.

Therefore, the connection between the multiplicity function $f(s,b)$ and the born term is formally defined as follows:

$$f(s,b) = \beta(s)\,[\chi_I(s,b)]^{2\lambda}, \tag{3.8}$$

where $\beta(s)$ is determined by the normalisation condition (2.8) :

$$\beta(s) = \frac{\int d^2b\, G_{in}(s,b)}{\int d^2b\, G_{in}(s,b)\,[\chi_I(s,b)]^{2\lambda}} \tag{3.9}$$

and the exponent 2λ in eq. (3.8) introduces a power-law dependence on $\chi_I(s,b)$, suggesting a non-linear relationship between the effective overlap and the particle production in the impact-parameter space.

It should be emphasized that according to the geometrical approach, the phenomenological portrayal of hadronic multiplicity distributions is considerably influenced by the values and behaviours of three key inputs: the inelastic overlap function G_{in}, the elementary multiplicity distribution $\phi(z)$, and the λ parameter determining the power-law dependence on the effective overlap $\chi_I(s,b)$. A previous study [11] has shown that changing one of these inputs, while keeping the other two fixed, produces different results across different parameterizations. For instance, if the inputs of the inelastic overlap function are different and λ, as well as $\phi(z)$, are kept constant, the obtained hadronic multiplicity distributions successfully replicate the experimental data. Interestingly, in all cases, the physical picture is that large multiplicities occur for small impact parameters while peripheral collisions (large b) lead to small multiplicities. It is worth noting that all these results, obtained for different inputs of the inelastic overlap function, are generated using the eikonal scheme. However, in the present study, we employ the U-matrix scheme for the reasons mentioned in the previous section. This will eventually allow us to fix our choice of the inelastic overlap function.

In order to present our findings in the following section, it is essential to provide an overview of the model parameters that were determined through data fitting, as well as the specific experimental data employed in our analysis. Since the Born term $\chi_I(s,b)$ is completely determined from the best-fit, Table 1, we see from the master formula for the multiplicity distribution eq. (2.11) that the only free parameters are K, λ, and $\langle n(s,b)\rangle$. Further, for our purposes, it is sufficient to fix the value of K and assume λ and $\langle n(s,b)\rangle$ as the only fitting parameters. Indeed, in [11], it was shown that the choice of $K = 10.775$ by assuming a Gamma distribution gives a good description of the charged multiplicity distributions for e^+e^- annihilation data in a large energy interval. Based on the universality of multiplicities in e^+e^- and $p + p(\bar{p})$ collisions, we adopt this choice.

To determine the values of the parameters λ and $\langle n\rangle$, we fix the dimensionless parameter to $K = 10.775$ and conduct fits to full phase space P_n data in $p + p(\bar{p})$ collision across a wide range of energies, specifically at $\sqrt{s} = 30.4, 44.5, 52.6, 62.2, 300, 546, 1000$, and 1800 GeV [34, 35].

With regards to the fitting process, we utilized the Minuit2 class from ROOT [36] and implemented the MIGRAD algorithm. The primary objective of the fitting procedure was to minimize the χ^2 value and the uncertainties associated with the free parameters were calculated using a 1σ confidence level.

4 Results and discussion

4.1 Geometrical scaling violation

The preliminary objective of this study was to analyze the behaviour of the inelastic overlap function in the impact parameter space as well as the geometrical scaling violation using the U-matrix and the eikonal schemes, eqs. (3.5) and (3.6), in an attempt to offer valuable

insights into the collision geometry and the interaction dynamics of colliding particles. Fig. 1 depicts the predictions for the energy evolution of the inelastic overlap function, eq. (2.2), in the impact parameter space at energies spanning from ISR to LHC levels using both schemes.

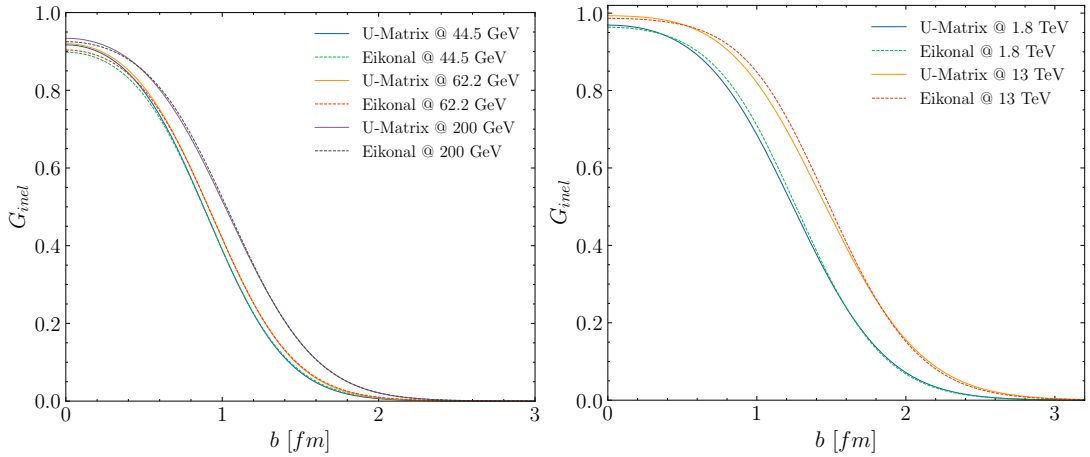


Figure 1: The energy evolution of the inelastic overlap function in the impact parameter space at energies spanning from ISR to LHC levels using both the U-matrix and the eikonal schemes.

Looking at Fig. 1, it is apparent that this function exhibits a generally comparable pattern in impact parameter space across both schemes, with only minor differences. More specifically, it shows that it is predominantly central, indicating that it has a significant contribution at small impact parameters. This entails that most of the inelastic processes are more probably to take place when the colliding particles pass through one another within close proximity. Moreover, this function tends to decline more slowly with the impact parameter as energy rises. This points out that the inelastic processes become less dependent on the specific spatial distance between colliding particles and have a wider range of impact parameters at which they can happen at higher energies.

Furthermore, it can be seen from Fig. 1 that the magnitude of the inelastic overlap function grows with energy at the central impact parameter $b = 0$. This indicates that the higher the energy, the more likely it is that inelastic events will occur at the central impact parameter. Interestingly, as Fig. 1 shows, there is a significant divergence in the magnitude of the inelastic overlap function obtained from the U-matrix and the eikonal schemes. To be more specific, the former yields a greater magnitude than the latter at $b = 0$. It is clear from this difference that the chosen scheme has a profound impact on the inelastic processes, particularly on central collisions which will have distinct characteristics or probabilities.

In the same vein, Fig. 2 depicts the energy evolution of the ratio elastic-to-total cross-section using both schemes. As this Figure demonstrates, in both cases there is an overall increase of this ratio as energy rises from 10 GeV to 10 TeV, but does so in a non-linear

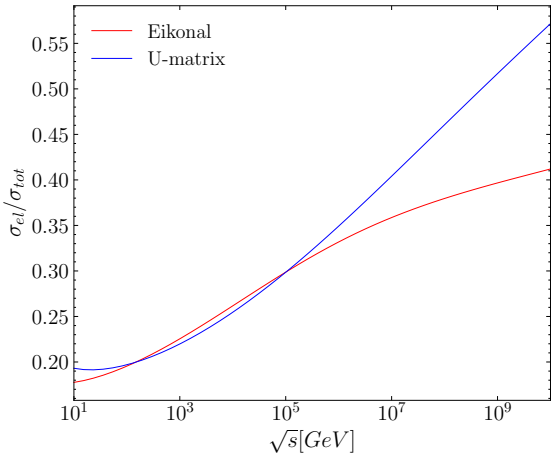


Figure 2: Energy evolution of the ratio elastic-to-total cross-section in both cases, the eikonal and U-matrix schemes

manner, which is indicative of geometrical scaling violation. Interestingly, comparing the behaviour of this ratio in both cases, it is clear that it increases more rapidly in the U-matrix case than in the eikonal one, as energy increases, demonstrating a stronger violation of the geometrical scaling and a more intricate behaviour in the former case.

It is possible to explain the divergence observed in the geometrical scaling violation between the two distinct schemes in terms of the behaviour of the inelastic overlap function in the impact parameter space with respect to energy. To be more precise, this difference lies in the behaviour of the inelastic overlap function at $b = 0$, where we can see from Fig. 1 that it has a magnitude that grows with energy in both cases. However, the U-matrix scheme yields a larger magnitude than the eikonal one. This implies that, in the former case, the inelastic interactions are more prominent, indicating that they have a remarkably and rapidly increasing strength at the central impact parameter with increasing energy, compared to the eikonal case.

It can be argued that the discrepancy found in the behavior of the inelastic overlap function as well as in the observed violation of the geometrical scaling between the two schemes can be ascribed to their differing approaches to the unitarization of the scattering amplitude. Indeed, based upon its thorough treatment of processes and its ability to take into account a wider range of interactions and collision dynamics compared to the Eikonal scheme, the U-matrix scheme is likely to be a more advanced mechanism for dealing with scattering processes and capturing the intricate dynamics of hadronic interactions, which is also supported by other results [14, 15, 24].

On the whole, this result validates the remarkable difference in the geometrical scaling violation between the U-matrix and eikonal schemes. Specifically, it supports the claim that this violation is more noticeable with the former scheme when energy levels rise. Notably, this outcome is in agreement with the theoretical prediction provided in [14] and further highlights that this discrepancy becomes apparent even before reaching the extremely high-energy region. Furthermore, it reinforces our motivation, as stated in the

introductory section, for selecting the U-matrix scheme. In fact, the choice of this scheme, in particular, may contribute to answering the question posed in the introduction regarding its potential implications for KNO scaling violation and unravelling the underlying physics behind the multi-particle production mechanism, as will be elaborated in the forthcoming sections.

4.2 Hadronic Multiplicity Distributions

Using the master formula (2.11), the outcomes of the fitting procedure of the multiplicity distributions data across a wide range of energies are provided in Fig. 3, Fig. 4 and, Table 2, respectively, where the values of the parameters λ and $\langle n(s)\rangle$ obtained in each fitting procedure, along with the corresponding $\beta(s)$, are furnished, as well as the different χ^2/dof values.

$\sqrt{s}$ [GeV]	λ	$\beta(s)$	$\langle n(s)\rangle$	χ^2/DOF
30.4	0.2837 ± 0.0133	1.6014	9.0583 ± 0.1452	1.2714
44.5	0.2720 ± 0.0119	1.6308	10.5900 ± 0.1339	0.6170
52.6	0.2775 ± 0.0104	1.6430	11.2885 ± 0.1299	0.6443
62.2	0.2709 ± 0.0107	1.6547	12.0090 ± 0.1551	1.3303
300	0.3695 ± 0.0113	1.7327	23.4646 ± 0.2915	0.6046
546	0.4618 ± 0.0148	1.7420	27.7185 ± 0.4145	0.3406
1000	0.4230 ± 0.0110	1.7362	36.6360 ± 0.4042	1.5301
1800	0.4836 ± 0.0070	1.7132	42.7217 ± 0.3184	1.2004
7000	0.5936	1.5833	73.7117	–
14000	0.6597	1.4716	96.0705	–

Table 2: Values of the λ parameter and $\langle n(s)\rangle$ resulting from fits to the P_n data. The values of $\beta(s)$ were obtained from eq. (3.9).

According to Fig. 3, Fig. 4, and the different χ^2/dof values, our model gives a reasonable description of the different multiplicity distributions at each energy. Moreover, it allows to predict the multiplicity distribution P_n at LHC energies by evaluating the energy dependence of the parameter λ using an appropriate function $\lambda(s)$.

As can be seen from Fig.5 (left panel), the behaviour of this parameter is energy-dependent and rapidly increases with increasing energy. This energy dependence can be aptly described by the following function:

$$\lambda(s) = a_0\ s^{a_1} \tag{4.1}$$

where the values $a_0 = 0.154$, $a_1 = 0.0762$ were determined by a careful χ^2 analysis. Thus, based on this function, the λ values at 7 and 14 TeV are retrieved and then displayed in Table 2.

Similarly, the estimates for the average multiplicity of hadrons at LHC energies can be derived. As depicted in Figure 5 (right panel), this parameter exhibits a rapid increase with the energy s, and its energy dependence can be consistently described using this function:

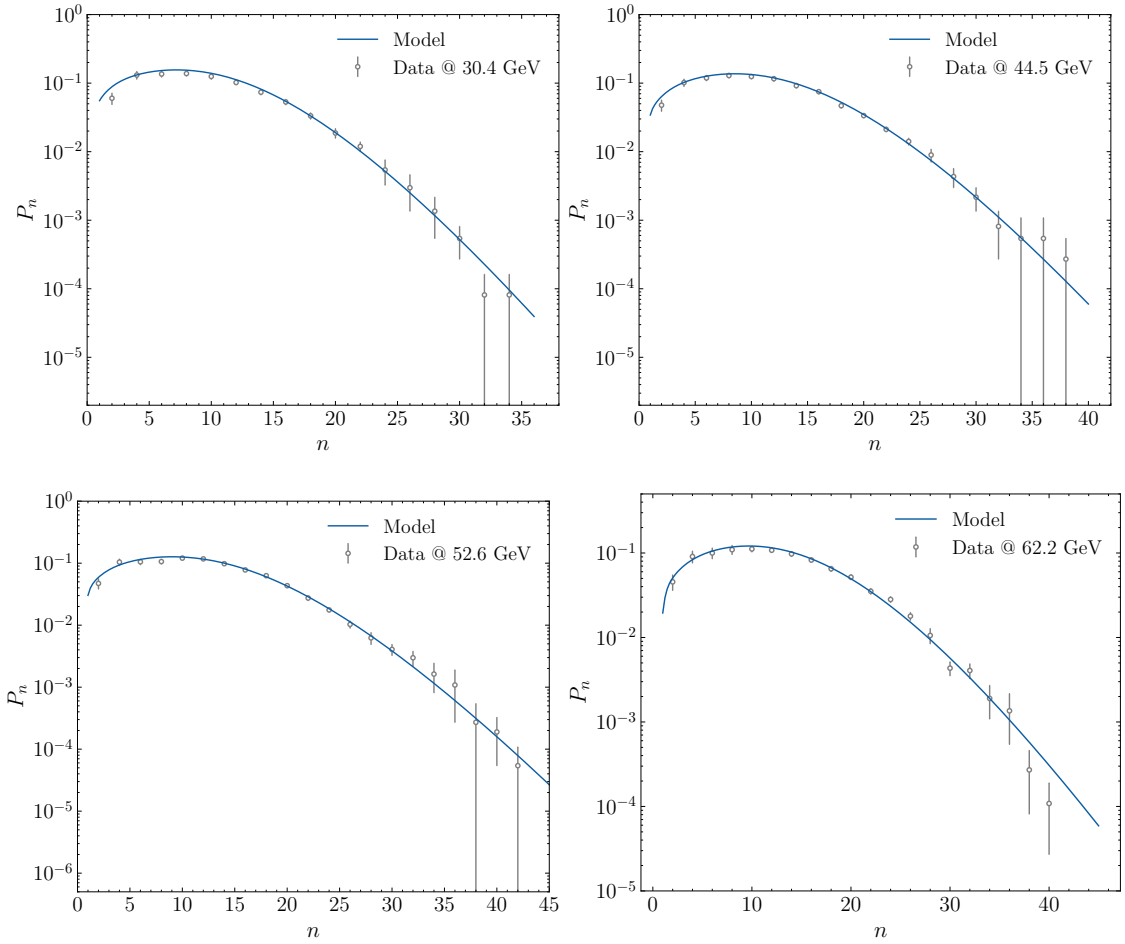


Figure 3: Multiplicity distributions for inelastic pp data at $\sqrt{s} = 30.4, 44.5, 52.6$ and 62.2 GeV compared with theoretical expectations.

$$\langle n(s) \rangle = b_0 \; s^{b_1}, \tag{4.2}$$

where the values of b_0 and b_1 are 2.5 and 0.1911, respectively. These coefficients were determined through a rigorous χ^2 analysis. Hence, by using this dependence, one can determine the values of $\langle n(s) \rangle$ at 7 and 14 TeV, as illustrated in Table 2.

An intriguing aspect of our findings lies in the remarkable accord between our result concerning the energy dependence of the hadron mean multiplicity and that obtained by Troshin and Tyurin with their model for multi-particle production with antishadowing [37], adding to the growing body of evidence that supports the fundamental principles underlying the U-Matrix approach. This alignment is highlighted by the following equation [37] :

$$\langle n(s) \rangle = 2.328 \; s^{0.201}, \tag{4.3}$$

This also implies that this approach is highly predictive and can accurately describe and interpret multi-particle production in different energy regimes. Besides, we should

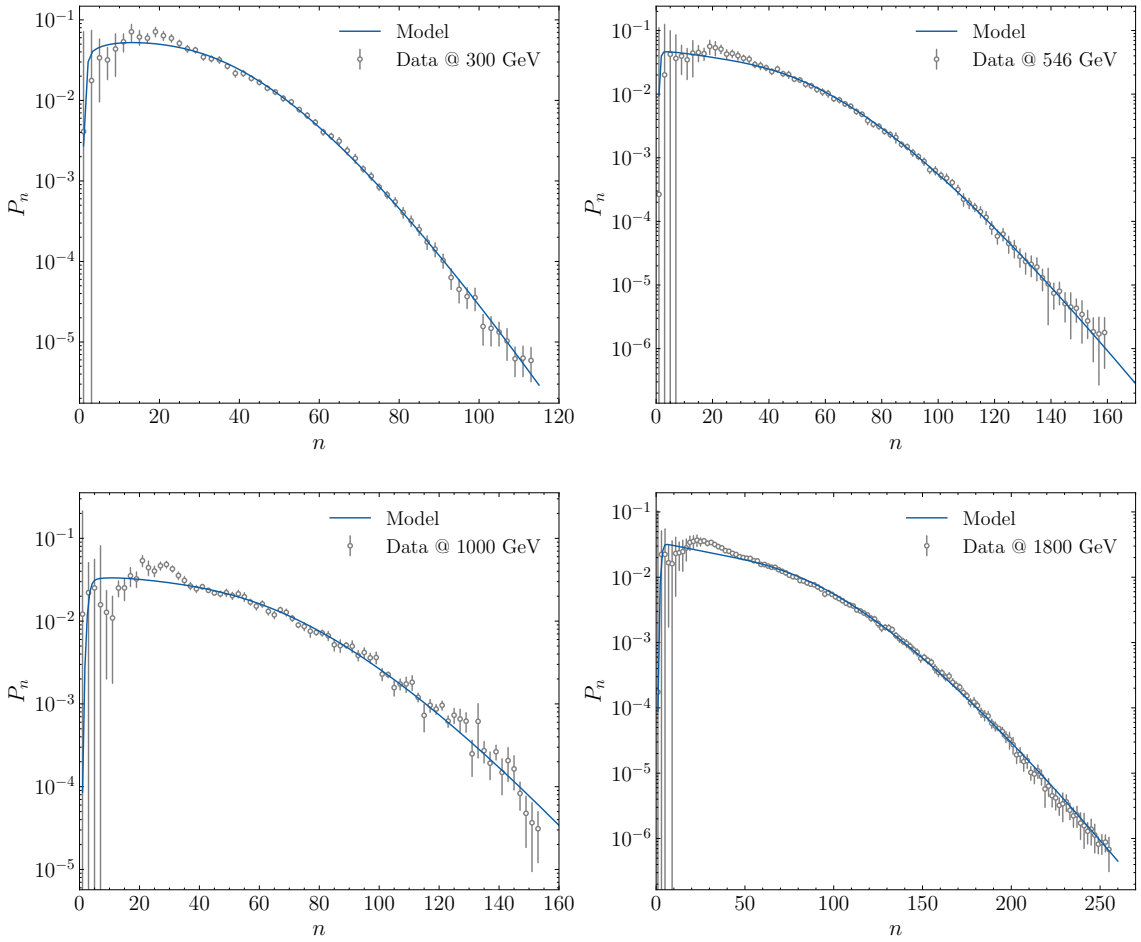


Figure 4: Multiplicity distributions for inelastic $\bar{p}p$ data at $\sqrt{s} = 300, 546, 1000$ and 1800 GeV compared with theoretical expectations.

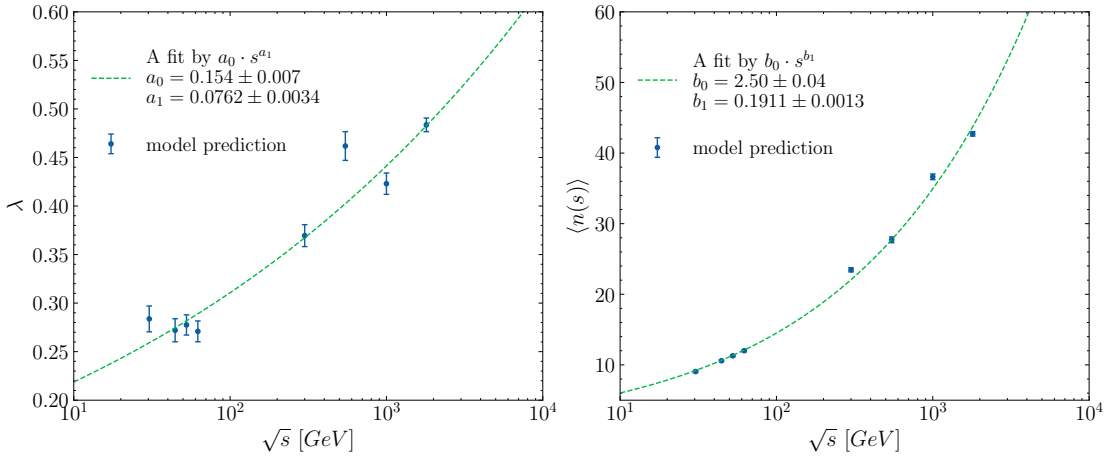


Figure 5: The energy dependence of the parameter λ (left panel) and the mean multiplicity (right panel)

emphasize that the power-law energy dependence of the hadron mean multiplicity is often regarded as a prominent feature observed in different models and consistent with experimental data from heavy ion collisions [38, 39] and this alignment in results further reinforces this assumption.

Having tuned our model with all parameters obtained from the best fits, we can now rely on its potential extrapolations to novel collision energy regimes and investigate various phenomena, such as the KNO scaling violation and the correlation of final state particles, as will be presented in the subsequent sections.

4.3 KNO scaling violation

Using our model, the KNO scaling violation was also examined. The predictions for the full-phase space multiplicity distribution in $p+p(\bar{p})$ collision, in KNO form at various energies, spanning from ISR to LHC ones, are displayed in Fig. 6. Aside from the logarithmic view in the left panel, a linear scale along with a zoom into the low-multiplicity region are provided in the right panel.

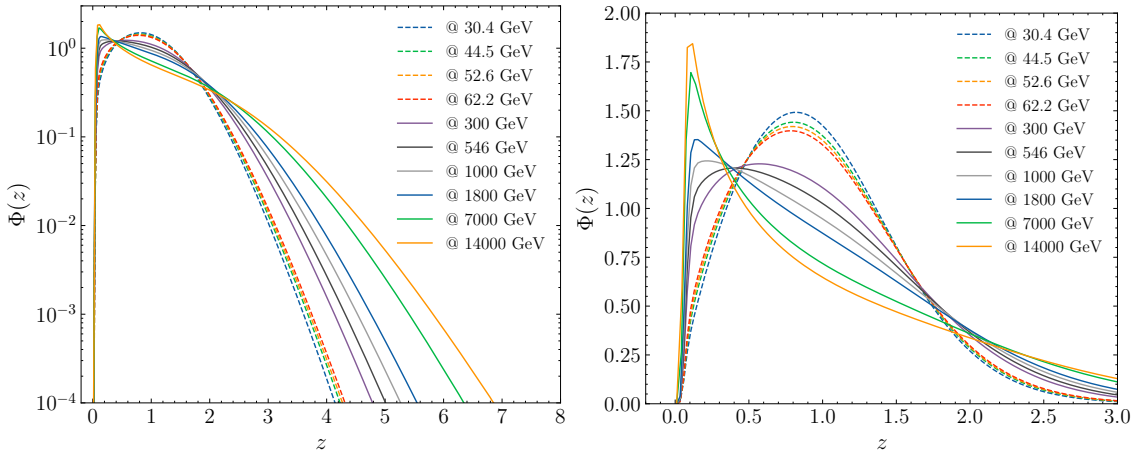


Figure 6: The multiplicity distributions from ISR to LHC energies in full-phase space (left panel). In addition to the logarithmic view, the right panel shows a linear scale and a zoom into the low-multiplicity region.

The left panel of Fig. 6 shows that the high-multiplicity tail rises with increasing energy. At the same time, by looking at the right panel, we can see that the maximum of the distribution shifts towards smaller values of z.

This interpretation demonstrates the dynamic behaviour of the system, especially as the energy level rises, and further validates the violation of the KNO scaling. Interestingly, we can also see that beyond the ISR energy range, the width of the distribution gets larger with increasing energy, which underscores the strong violation of the KNO scaling. It is worth noting that this finding resonates with experimental observations [7].

Most importantly, based on the picture that KNO scaling violation is an extension of geometrical scaling violation, we can also claim that the strong violation of the former

stems from the strong violation of the latter, emphasizing the interconnected nature of these phenomena within the U-matrix representation and stressing the latter's pivotal role in describing collision geometry and the processes of multi-particle production in hadron collisions.

To further illustrate the role of the U-matrix scheme, we examined the average number of particles $\langle n(b,s)\rangle$ as a function of impact parameter b for various collision energies, as its pattern offers insights into the collision geometry and the distribution of particles in the transverse plane. The result is illustrated in Fig. 7 (left panel). Based on this figure, it is clear that, at central collision ($b = 0$), the magnitude of the average number of particles increases with increasing energy. This is quite anticipated since central collisions yield more produced particles than peripheral collisions [40]. This trend causes the tail of the multiplicity distribution to extend to higher values and eventually to a broader distribution as energy increases, indicating the possibility of rare high-multiplicity events.

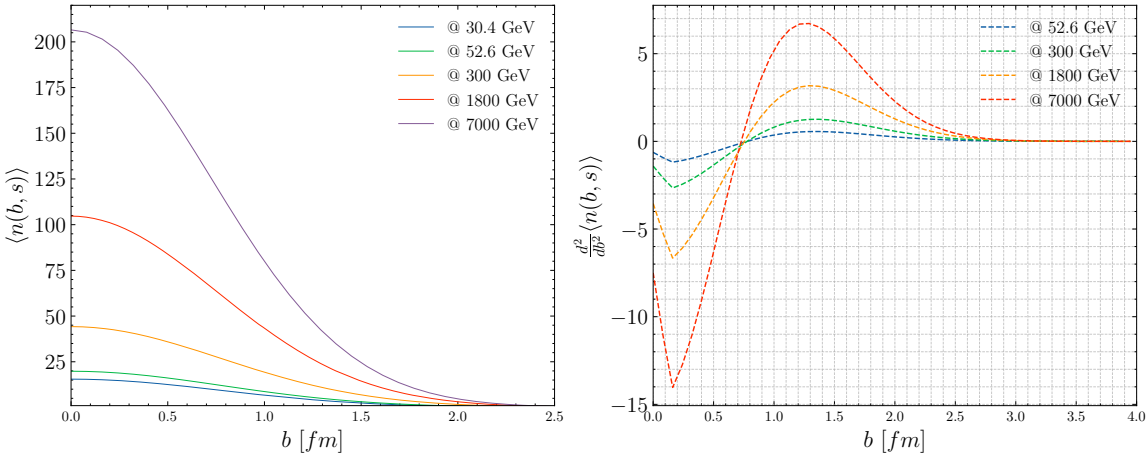


Figure 7: The energy dependence of the average multiplicity in the impact parameter space.

A possible explanation for the broadening of the multiplicity distribution might be related to the effect of using the U-matrix scheme. Indeed, as previously illustrated in Fig. 1, the disparity found between the eikonal and U-matrix schemes is noticeable at the central impact parameter, where the overlap of hadronic matter distributions is greater with the latter scheme. This discrepancy not only influences the inelastic overlap function, and hence the overall magnitude of multiplicity, but it also has a significant impact on the tail of the multiplicity distribution.

In addition, irrespective of the energy level, the average number of particles generally decreases as the impact parameter b increases, which eventually leads to a decrease in the magnitude of the multiplicity distribution. This phenomenon aligns with the notion that larger impact parameters lead to less violent collisions, resulting in events with smaller multiplicities [40].

Furthermore, as Fig. 7 manifests, the average number of particles is energy-dependent,

increasing with higher collision energies at each impact parameter. This behaviour is also expected in high-energy physics experiments, where higher energies often result in increased particle production [7].

Fig. 7 (left panel) shows that the curvature of the average number of particles changes, at a specific impact parameter, regardless of the energy level. More quantitatively, we show the second derivative of the average number of particles as a function of the impact parameter in Fig. 7 (right panel), illustrating a change of sign of this function which demonstrates the existence of an inflection point at exactly 0.7 fm. This change in the curvature could signify a shift in the particle production behaviour. For instance, at this specific impact parameter value, there might be a change in the interaction dynamics or the nature of the collision process. It should be noted that this result is in line with what has been reported in [41], where the inflection point was roughly estimated to be at around 1fm. The fact that the inflection point is specifically at around 1 fm suggests that there is something unique or significant about collisions occurring at this distance. This could be related to the characteristics of the colliding particles or the structure of the hadrons involved. This common behaviour underscores the notion of the critical phenomenon in hadronic interactions [42].

On the whole, the shape of the multiplicity distribution is influenced not only by the impact parameter and collision energy but also by the unitarisation scheme, particularly the U-matrix, which reinforces our claim that this distribution is scheme-dependent, as outlined in the introductory section. Besides, the tail of the multiplicity distribution gives insights into the rare but significant events that constitute the overall dynamics of high-energy collisions and hence highlights the importance of this scheme in multi-particle production processes.

In light of this result, the U-matrix scheme may prove to be a significant alternative for addressing multi-particle production challenges at ultra-high energies, including the muon puzzle in cosmic ray interactions at this energy level. In situations such as these, where extreme conditions and rare events are likely to play a significant role, scheme-dependent effects on multiplicity distribution become relevant.

Having said that, it is vital to consider this scheme in enhancing the existing hadronic interaction models to tackle several lingering issues in high and ultra-high energy physics. However, this is beyond the remit of this study.

4.4 Hadronic multi-particle correlations

Now that we have estimated the multiplicity distribution P_n across various energies, it seems appropriate to characterize it in an attempt to obtain a better understanding of the dynamics of the particle production process in hadron collisions. In order to fulfil this purpose, the normalized ordinary higher-order moments ($q > 2$) of this distribution are analyzed. Technically speaking, P_n's moments of order q are defined as follows:

$$C_q = M_q/M_1^q, \tag{4.4}$$

and

$$M_q = \sum_{n=0}^{\infty} n^q P_n, \tag{4.5}$$

Our results, along with their comparison with the experimental data [1], are given in Table 3 and illustrated in Fig. 8 (left and middle panels).

$\sqrt{s}$ $[GeV]$	C_2	C_3	C_4	C_5
30.4	1.29 ± 0.05 1.28	1.97± 0.09 1.94	3.45 ± 0.21 3.36	6.68 ± 0.52 6.47
44.5	1.28 ± 0.04 1.3	1.95± 0.07 2.01	3.40 ± 0.17 3.55	6.58 ± 0.47 6.99
52.6	1.29 ± 0.03 1.3	1.98± 0.06 2.04	3.48 ± 0.15 3.64	6.81 ± 0.42 7.24
62.2	1.29 ± 0.03 1.31	1.97± 0.06 2.07	3.40 ± 0.14 3.73	6.43 ± 0.33 7.51
300	1.34 ± 0.02 1.41	2.21± 0.04 2.46	4.26 ± 0.07 4.95	9.23 ± 0.17 11.17
546	1.41 ± 0.03 1.46	2.52± 0.05 2.67	5.31 ± 0.10 5.63	12.72 ± 0.24 13.38
1000	1.41 ± 0.02 1.52	2.47± 0.05 2.91	5.11 ± 0.13 6.5	11.87 ± 0.36 16.33
1800	1.47 ± 0.02 1.59	2.78± 0.03 3.21	6.23 ± 0.07 7.57	15.91 ± 0.21 20.18
7000	1.78	4.14	11.36	35.3
8000	1.81	4.26	11.87	37.51
13000	1.89	4.73	14.03	47.19
14000	1.91	4.81	14.41	48.93

Table 3: C_q Moments: experimental data with error bar and theoretical predictions. Data points are from [43]

Based on Fig. 8, we can see that, as energy levels rise, our proposed model predicts a gradual increase in the ordinary higher-order moments, represented by C_2, C_3, C_4, and C_5. Surprisingly, while our predictions match with the data points within the ISR energy range, it is clear that the model overestimates the fluctuations and correlations in the multiplicity distribution with rising energy, notably above LHC energy. In order to further illustrate this overestimation, we computed the f_2 moment (or the two-particle correlation parameter), as a means of examining the correlation between pairs of particles during a collision event, which is defined by the following formula :

$$f_2 = < n(n-1) > - < n >^2 \tag{4.6}$$

[1] see compilation in [43]

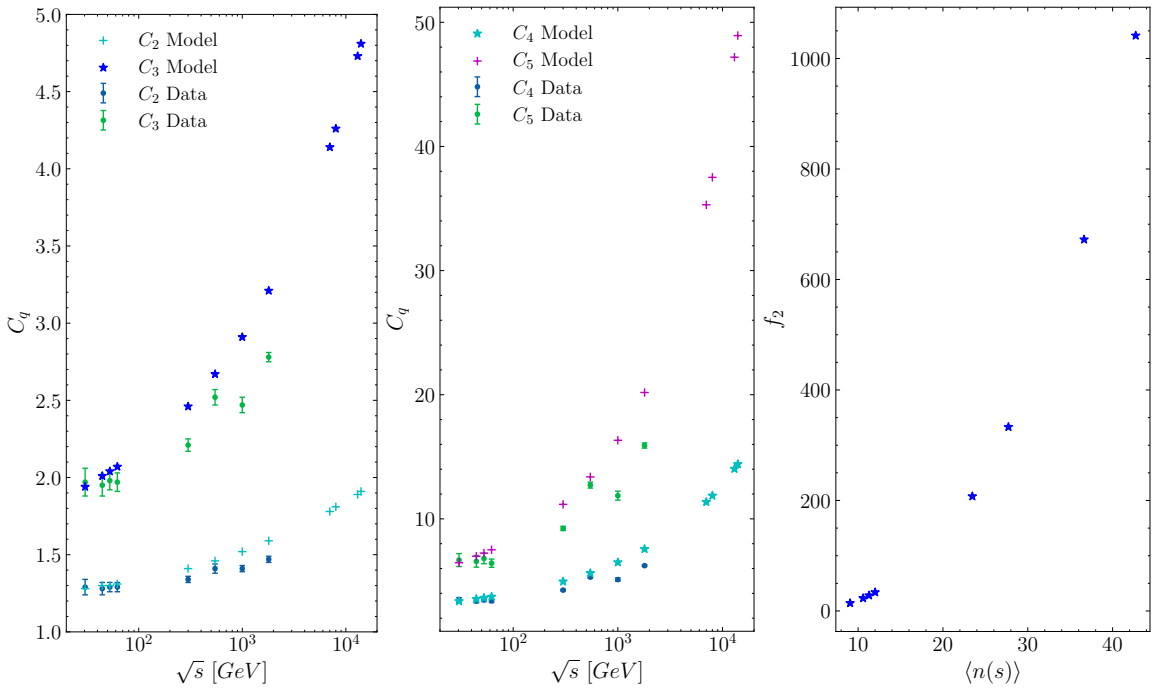


Figure 8: Experimental and Theoretical C_q moments, q = 2, 3 (left panel), q = 4,5 (middle panel). f_2 moment versus the average number of produced particles (right panel).

Interestingly, as illustrated in Fig. 8 (right panel), there is a noteworthy and sudden increase in the two-particle correlation parameter versus the average number of produced particles, indicating the existence of strong correlations among the charged particles. Consequently, we can infer that the model incorporates correlations in the final state, despite being constructed on the basis of independent particle production. So the question that arises is where this correlation emerges from. As the overall hadronic multiplicity distribution is constructed by summing contributions from parton–parton collisions occurring at each impact parameter weighted by the inelastic overlap function, this overestimation of correlation is linked to the weight in this superposition model, and hence to the unitarisation scheme, in comparison with the predictions provided by an eikonal geometrical independent string model [43].

Our model's outcomes of a pronounced KNO scaling violation, together with the unexpected overestimation of the fluctuations and correlations with increasing energy, can potentially be attributed to statistical fluctuations. Hence, we may claim that in this U-matrix representation, pomeron exchange may involve more intricate dynamics, such as collective effects, non-perturbative QCD dynamics, or other interactions, leading to different statistical fluctuations beyond a simple Poissonian eikonal summation [44].

5 Conclusions

The first part of the results' section was concerned with the description of the geometrical scaling violation. In fact, the energy evolution of the elastic-to-total cross-section ratio was investigated using both the eikonal and U-matrix schemes. The results have revealed that when energy rises, this ratio increases non-linearly and more rapidly with the U-matrix scheme than with the eikonal, implying a stronger violation of the geometrical scaling. This pronounced violation was understood in terms of the divergence in the behaviour of the inelastic overlap function, particularly at the central impact parameter, where the magnitude of this function is greater with the U-matrix scheme than the eikonal, regardless of the energy level.

The second part of the results' section was devoted to the hadronic multiplicity distributions. Our model was tuned and all parameters were obtained from the best fits to various hadronic multiplicity distributions data over a wide range of energies. It has been found that the present model provides a reasonable description of the different multiplicity distributions at each energy. Interestingly, our findings about the energy dependence of the hadron mean multiplicity agree well with Troshin and Tyurin's analyses of the multiparticle production in the antishadowing model [37]. This agreement adds to the growing body of evidence supporting the fundamental principles underlying the U-Matrix approach, which ensures reliable extrapolations to novel collision energy regimes.

Based on our model, the KNO scaling violation was investigated as well. Our results, related to the behaviour of the tail, as well as the maximum of the multiplicity distribution, and especially to the strong KNO scaling violation, are in line with the experimental findings [7]. The broadening of the multiplicity distribution has been also confirmed with the behaviour of the particles' distribution in the transverse plane at various collision energies. Besides, the interesting finding of an inflection point in the average number of particles' curve irrespective of the energy level, corroborates with another finding (ref) and highlights the concept of a critical phenomenon in hadronic interactions.

Besides, the normalized ordinary higher-order moments ($q > 2$) of the multiplicity distribution were analyzed. Another surprising result was related to our model's overestimation of the fluctuations and correlations in the multiplicity distribution as energy rises, notably above LHC energy. It is argued that this overestimation, is linked to the weight in this superposition model, and hence to the unitarisation scheme. It should be noted that the shape of the hadron multiplicity distribution is influenced not only by the impact parameter and collision energy but also by the unitarisation scheme, particularly the U-matrix.

On the whole, the results of this study, such as those related to the strong geometrical scaling violation and its resultant pronounced KNO scaling violation, coupled with the overestimation of the charged particles' correlation, can potentially be attributed to statistical fluctuations inherent in the U-matrix scheme. Hence, we may argue that in this U-matrix representation, pomeron exchange may involve more intricate dynamics, leading to different statistical fluctuations beyond a simple Poissonian eikonal summation.

As any research, this study is not without limitations. For instance, the assumption

that each created string has an equal probability of turning into a pair of charged hadrons is acceptable but only as a first approximation. To refine the present model for a more reliable description, we can consider the introduction of charged particle correlations in the final state by incorporating, for example, string fusion, overlapping processes, or other interactions leading to correlated string dynamics as evidenced by theoretical predictions [12] and experimental observations [8]. Additionally, it is important to investigate the impact of the implicit different statistical fluctuations in pomeron exchanges within the U-matrix representation, along with the correlation and/or collective behaviour in the production of final state particles.

The study concludes by proposing the U-matrix scheme as a noteworthy alternative for tackling challenges related to multi-particle production in hadron collisions, especially in scenarios where extreme conditions and rare events play a significant role in high and ultra-high energy physics. It also prompts an inquiry into the fundamental nature of pomeron exchange within the U-matrix scheme in comparison to the eikonal, despite that both schemes verify the unitarity constraint principle.

Acknowledgments

RO would like to thank Jean-René Cudell for his invaluable comments. RO also appreciates the insightful discussions with Paulo Cesar Beggio and Emerson Luna. Special thanks go to the computational resource provided by Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) where a part of the computational work was carried out.

References

[1] L. Van Hove, *Particle production in high energy hadron collisions*, *Physics Reports* **1** (1971) 347.

[2] G. Giacomelli, *Particle production in high energy collisions*, 0901.4924.

[3] F.-H. Liu, K.K. Olimov and A. Beraudo, *Editorial: Particle production and system evolution in collisions from gev to tev*, *Frontiers in Physics* **10** (2022) .

[4] H. Koley, A. Mondal, S. Kar, S. Mukherjee, J. Mondal, A. Deb et al., *Different degrees of complexity in multiparticle production at the LHC energies with the transition from soft to hard processes*, in *65th DAE BRNS Symposium on nuclear physics*, 5, 2022 [2205.04429].

[5] I. Dremin and J. Gary, *Hadron multiplicities*, *Physics Reports* **349** (2001) 301.

[6] R. Alves Batista, J. Biteau, M. Bustamante, K. Dolag, R. Engel, K. Fang et al., *Open questions in cosmic-ray research at ultrahigh energies*, *Frontiers in Astronomy and Space Sciences* **6** (2019) 23.

[7] J.F. Grosse-Oetringhaus and K. Reygers, *Charged-particle multiplicity in proton–proton collisions*, *Journal of Physics G: Nuclear and Particle Physics* **37** (2010) 083001.

[8] CMS collaboration, *Bose-Einstein correlations of charged hadrons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **03** (2020) 014 [1910.08815].

[9] A. Pandav, D. Mallick and B. Mohanty, *Search for the QCD critical point in high energy nuclear collisions*, *Prog. Part. Nucl. Phys.* **125** (2022) 103960 [2203.07817].

[10] C.C. Shih and J.-y. Zhang, *Particle production at high energy. i. geometrical orientated simulation of hadron-hadron collisions below 250 gev/c*, *Phys. Rev. C* **55** (1997) 378.

[11] P.C. Beggio, M.J. Menon and P. Valin, *Scaling violations: Connections between elastic and inelastic hadron scattering in a geometrical approach*, *Phys. Rev. D* **61** (2000) 034015 [hep-ph/9908389].

[12] P.C. Beggio and Y. Hama, *A new scheme for calculation of the multiplicity distributions in hadronic interactions*, *Braz. J. Phys.* **37** (2007) 1164.

[13] P. Beggio and E. Luna, *Cross sections, multiplicity and moment distributions at the lhc*, *Nuclear Physics A* **929** (2014) 230.

[14] A. Bhattacharya, J.-R. Cudell, R. Oueslati and A. Vanthieghem, *Proton inelastic cross section at ultrahigh energies*, *Phys. Rev. D* **103** (2021) L051502 [2012.07970].

[15] A. Vanthieghem, A. Bhattacharya, R. Oueslati and J.-R. Cudell, *Unitarisation dependence of diffractive scattering in light of high-energy collider data*, *JHEP* **09** (2021) 005 [2104.12923].

[16] M.B. Gay Ducati, M.M. Machado and M.V.T. Machado, *Investigating diffractive W production in hadron-hadron collisions at high energies*, *Int. J. Mod. Phys. E* **16** (2007) 2956 [0706.3468].

[17] M. Bozzo, P. Braccini, F. Carbonara, R. Castaldi, F. Cervelli, G. Chiefari et al., *Measurement of the proton-antiproton total and elastic cross sections at the cern sps collider*, *Physics Letters B* **147** (1984) 392.

[18] G. Alner, K. Alpgård, R. Ansorge, B. Åsman, K. Böckmann, C. Booth et al., *Scaling violation favouring high multiplicity events at 540 gev cms energy*, *Physics Letters B* **138** (1984) 304.

[19] Z. Koba, H. Nielsen and P. Olesen, *Scaling of multiplicity distributions in high energy hadron collisions*, *Nuclear Physics B* **40** (1972) 317.

[20] K. Itabashi, *Koba-Nielsen-Olesen Scaling, Geometrical Scaling and Barshay-Yamaguchi Scaling*, *Progress of Theoretical Physics* **54** (1975) 1168.

[21] J. Finkelstein, *AN IMPACT PARAMETER MODEL FOR MULTIPLICITY DISTRIBUTIONS*, *Z. Phys. C* **41** (1988) 167.

[22] C.S. Lam and P.S. Yeung, *Possible connection between KNO and geometrical scaling*, *Physics Letters B* **119** (1982) 445.

[23] D. Fagundes, E. Luna, M. Menon and A. Natale, *Aspects of a dynamical gluon mass approach to elastic hadron scattering at lhc*, *Nuclear Physics A* **886** (2012) 48.

[24] R. Oueslati, *A multi-channel U-Matrix model of hadron interaction at high energy*, *JHEP* **08** (2023) 087 [2305.03424].

[25] T. Chou and C.N. Yang, *Remarks on multiplicity fluctuations and kno scaling in pp collider experiments*, *Physics Letters B* **116** (1982) 301.

[26] S. Barshay, *Geometric derivation of the diffractive multiplicity distribution*, *Phys. Rev. Lett.* **49** (1982) 1609.

[27] C. Kuang-chao, L. Lian-sou and M. Ta-chung, *Koba-nielsen-olesen scaling and production mechanism in high-energy collisions*, *Phys. Rev. D* **28** (1983) 1080.

[28] P.C. Beggio and F.R. Coriolano, *Energy dependence of the inelasticity in $pp/p\bar{p}$ collisions from experimental information on charged particle multiplicity distributions*, *Eur. Phys. J. C* **80** (2020) 437 [2004.06839].

[29] X. Artru and G. Mennessier, *String model and multiproduction*, *Nuclear Physics B* **70** (1974) 93.

[30] A. Donnachie and P.V. Landshoff, *pp and $\bar{p}p$ total cross sections and elastic scattering*, *Phys. Lett. B* **727** (2013) 500 [1309.1292].

[31] O.V. Selyugin, J.R. Cudell and E. Predazzi, *Analytic properties of different unitarization schemes*, *Eur. Phys. J. ST* **162** (2008) 37 [0712.0621].

[32] J.R. Cudell, E. Predazzi and O.V. Selyugin, *New analytic unitarisation schemes*, *Phys. Rev.* **D79** (2009) 034033 [0812.0735].

[33] H.J. Drescher, M. Hladik, S. Ostapchenko, T. Pierog and K. Werner, *Parton based Gribov-Regge theory*, *Phys. Rept.* **350** (2001) 93 [hep-ph/0007198].

[34] Ames-Bologna-CERN-Dortmund-Heidelberg-Warsaw Collaboration collaboration*Phys. Rev. D* **30** (1984) 528.

[35] T. Alexopoulos, E. Anderson, N. Biswas, A. Bujak, D. Carmony, A. Erwin et al., *The role of double parton collisions in soft hadron interactions*, *Physics Letters B* **435** (1998) 453.

[36] M. Hatlo, F. James, P. Mato, L. Moneta, M. Winkler and A. Zsenei, *Developments of mathematical software libraries for the LHC experiments*, *IEEE Trans. Nucl. Sci.* **52** (2005) 2818.

[37] S.M. Troshin and N.E. Tyurin, *Multiparticle production in the model with antishadowing*, *Journal of Physics G: Nuclear and Particle Physics* **29** (2003) 1061.

[38] S. Barshay and G. Kreyerhoff, *Related power law growth of particle multiplicities near mid-rapidity in central Au + Au collisions and in anti $\bar{p}(p) - p$ collisions*, *Nucl. Phys. A* **697** (2002) 563 [hep-ph/0104303].

[39] R. Singh, L. Kumar, P.K. Netrakanti and B. Mohanty, *Selected Experimental Results from Heavy Ion Collisions at LHC*, *Adv. High Energy Phys.* **2013** (2013) 761474 [1304.2969].

[40] H. Yudong, W. Guangjun, M. Jun and T. An, *On the geometrical picture of multiparticle production in high-energy hadron-hadron collisions*, *Europhysics Letters* **9** (1989) 645.

[41] P.C. Beggio, *A multiparton model for p p / p anti-p inelastic scattering*, *Braz. J. Phys.* **38** (2008) 598.

[42] S. Iwao and M. Shako, *Hadron-Hadron Interactions and their Interpretation as Critical Phenomena*, .

[43] P.C. Beggio, *Inelastic cross sections, overlap functions and C_q moments from ISR to LHC energies in proton interactions*, *J. Phys. G* **44** (2017) 025002 [1701.08574].

[44] K.G. Boreskov, A.B. Kaidalov, V.A. Khoze, A.D. Martin and M.G. Ryskin, *The Partonic interpretation of reggeon theory models*, *Eur. Phys. J. C* **44** (2005) 523 [hep-ph/0506211].

arXiv:2403.02237v1 [hep-ph] 4 Mar 2024

Analytic continuations and numerical evaluation of the Appell F_1, F_3, Lauricella $F_D^{(3)}$ and Lauricella-Saran $F_S^{(3)}$ and their Application to Feynman Integrals

Souvik Bera * & Tanay Pathak †

Centre for High Energy Physics, Indian Institute of Science,
Bangalore-560012, Karnataka, India

Abstract

We present our investigation of the study of two variable hypergeometric series, namely Appell F_1 and F_3 series, and obtain a comprehensive list of its analytic continuations enough to cover the whole real (x, y) plane, except on its singular loci. We also derive analytic continuations of their 3-variable generalization, the Lauricella $F_D^{(3)}$ series and the Lauricella-Saran $F_S^{(3)}$ series, leveraging the analytic continuations of F_1 and F_3, which ensures that the whole real (x, y, z) space is covered, except on the singular loci of these functions. While these studies are motivated by the frequent occurrence of these multivariable hypergeometric functions in Feynman integral evaluation, they can also be used whenever they appear in other branches of mathematical physics. To facilitate their practical use, we provide four packages: `AppellF1.wl`, `AppellF3.wl`, `LauricellaFD.wl`, and `LauricellaSaranFS.wl` in MATHEMATICA. These packages are applicable for generic as well as non-generic values of parameters, keeping in mind their utilities in the evaluation of the Feynman integrals. We explicitly present various physical applications of these packages in the context of Feynman integral evaluation and compare the results using other means as well. Various MATHEMATICA notebooks demonstrating different numerical results are also provided along with this paper.

Summary of Programs:

- *Title of Programs*: `AppellF1.wl` , `AppellF3.wl` , `LauricellaFD.wl` , `LauricellaSaranFS.wl` .
- *Licensing provisions*: GNU General Public License v.3.0.
- *Programming language*: Wolfram Mathematica
- *Repository* : https://github.com/souvik5151/Appell_Lauricella_Saran_functions.
- *Nature of problem*: To find and develop a numerically consistent implementation of the analytic continuations of various two and three-variable hypergeometric functions, namely Appell F_1, F_3, Lauricella $F_D^{(3)}$ and Lauricella-Saran $F_S^{(3)}$ that typically appear in the evaluation of Feynman integrals.
- *Solution method*: Use the method of Olsson to find the analytic continuation of these functions and to implement these analytic continuations, following appropriate convention, in MATHEMATICA for consistent numerical evaluation. For the values of the Pochhammer parameters corresponding to the non-generic cases, a proper limiting procedure is implemented internally.

*souvikbera@iisc.ac.in
†tanaypathak@iisc.ac.in

1 Introduction

In this paper, we present our study on the analytic continuations of multivariable hypergeometric functions. In particular we study the two-variable Appell F_1 and F_3 [1] and three-variable Lauricella-Saran $F_D^{(3)}$ and $F_S^{(3)}$ [2,3] series. These investigations are a continuation of various other works carried out for various multivariable hypergeometric series such as Appell F_1 [4, 5], Appell F_2 [6, 7], Appell F_4 [8–10], Horn H_1 and H_5 [11], Kampé de Fériet series [12] and Lauricella series F_D^N [13–16], Srivastava $H_C^{(3)}$ [17] and other multivariable Horn functions [18–20]. Standard references of various properties and expressions related to multivariate hypergeometric functions can be found in [13, 21–25].

The motivation behind studying these particular functions comes from their ubiquitous presence in physics and mathematical physics, notably within the domain of Feynman integral calculus. In particular, it has been shown that almost all the Feynman integral can be written in terms of a general system of a hypergeometric function called the GKZ /$\mathcal{A}$- hypergeometric system [26–32], which is now a very active field of research [33–41]. There also exists computer program that can provide hypergeometric function representation of a given Feynman integral [42].

The hypergeometric functions have a long history of appearing in the evaluation of Feynman integrals. For example, in [43, 44] some general results for massive and massless $N-$point functions are obtained in terms of multivariable hypergeometric functions, including $N-$ variable Lauricella $F_D^{(N)}$. Moreover, the general massive one-loop Feynman integral can be represented as a meromorphic function of space-time dimensions using Gauss hypergeometric function ${}_2F_1$, Appell F_1 and Lauricella-Saran function $F_S^{(3)}$ for self-energy, vertex and box integrals, respectively [45–47]. For the one loop pentagon integral in the Regge limit, the result is expressed in terms of Appell F_4 and some Kampé de Fériet functions [48]. At the two loop level, Appell F_4 and its 3-variable generalization Lauricella $F_C^{(3)}$ appear in the study of sunset integrals [49–51]. Other examples can be found in [52–56]. In fact, the relation between hypergeometric functions and Feynman integrals is bidirectional. A number of new relations are discovered while studying the hypergeometric functions representation of Feynman integrals [57–63]. The application of these functions is also widespread in other fields of physics, such as in conformal field theory [64–66], where they occur quite frequently.

We focus on the evaluation of these multivariable hypergeometric functions outside their defining domain of convergence. In the hypergeometric function representation, the ratio of the scales involved in a Feynman integral appears in the variables of the hypergeometric functions. Thus, these analytic continuations help to evaluate these Feynman integrals for any kinematical values of choice. To derive the analytic continuations of Appell F_1 and F_3 and Lauricella-Saran $F_D^{(3)}$ and $F_S^{(3)}$, we use the method of Olsson [4]. This approach leverages the transformation properties of hypergeometric functions with fewer variables to determine the analytic continuations of the specified hypergeometric function. The method has been extensively used to study and obtain the analytic continuations of two variable Appell F_1 [4] and its $N-$variable generalization, i.e., Lauricella $F_D^{(N)}$ [13], Appell F_4 [8, 9], Appell F_2 [7] and Horn H_1 and H_5 [11]. The analytic continuations of Appell F_1 have been implemented for numerical purposes in `Fortran`, which is called `f1` [67,68]. While Appell F_2 is also implemented as the `AppellF2.wl` [7] package in MATHEMATICA. The method of Olsson itself has also been automated in the form of a package named `Olsson.wl` [69] in MATHEMATICA, which we use heavily in our present study. The package makes, the otherwise laborious and error prone task of deriving analytic continuations easy, quick and error-free.

Our study aims to achieve consistent and efficient evaluation of the derived analytic continuations. Consistency refers to the property that different analytic continuations that converge at the same point yield the same result. This is theoretically expected, but in practice, due to multi-valuedness, different analytic continuations may be evaluated on different sheets and produce different results. We address this problem by following the approach used for `AppellF2.wl` in [7]. Another important aspect of the packages is the evaluation of the analytic continuations for the non-generic values of Pochhammer parameters. Previously, in the `AppellF2.wl` package this issue was handled manually via a proper limiting procedure. While in all the packages that we present in this paper the necessary steps are incorporated internally, which makes them applicable for non-generic values of Pochhammer parameters as well. For the efficient evaluations of the finite series summation, we use techniques to reduce

the number of operations needed to perform the finite summations, as well as internal commands of MATHEMATICA that have a fast and efficient implementation. These techniques effectively reduce the number of operations through certain transformations of summation indices that we discuss in later sections. We observe significant improvement in time and accuracy in the evaluation and results obtained. Using these techniques, the analytic continuations form the core of, the widely used computer program, MATHEMATICA based packages `AppellF1.wl`, `AppellF3.wl`, `LauricellaFD.wl` and `LauricellaSaranFS.wl`, which can evaluate these functions for any values of Pochhammer parameters and real values of its arguments, except the singular loci of associated functions.

The article is organized as follows: The definitions and useful relations of Appell F_1, F_3, Lauricella $F_D^{(3)}$ and Lauricella-Saran $F_S^{(3)}$ are provided in Section 2. The method of Olsson for deriving the analytic continuations of multivariate hypergeometric functions is reviewed with an example of Lauricella $F_S^{(3)}$, in Section 3. The algorithm of these packages is discussed in Section 4, which is followed by a demonstration of the commands of the packages in Section 5. The numerical tests of these packages are presented in Section 6. The tests of these packages for non-generic values of the parameters, motivated by the examples of Feynman integrals, are detailed in Section 7. The summary and conclusions are drawn in Section 8. All the packages and several test files can be found in the GitHub repository.

2 Definitions

In this Section, we provide the definitions and useful expressions related to the functions of our study. We start with double variable Appell functions.

2.1 Appell Functions: F_1 and F_3

The four Appell functions [1] are the two-variable generalizations of the Gauss hypergeometric functions ${}_2F_1(a,b;c|x)$, whose series representation is given by

$$ {}_2F_1(a,b;c|x) = \sum_{m=0}^{\infty} \frac{(a)_m (b)_m}{(c)_m} \frac{x^m}{m!}, \qquad |x|<1 \tag{1} $$

where $(a)_m = \frac{\Gamma(a+m)}{\Gamma(a)}$ is the Pochhammer symbol.

The analytic continuations of Gauss ${}_2F_1(\dots|x)$ around $x=1$ and $x=\infty$ are well-known [22,23]

$$ \begin{aligned} {}_2F_1(a,b,c|x) = {} & \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} {}_2F_1(a,b,a+b-c+1|1-x) \\ & + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} (1-z)^{c-a-b} {}_2F_1(c-a,c-b,c-a-b+1|1-x) \end{aligned} \tag{2} $$

which is valid for $|1-x|<1$ and

$$ \begin{aligned} {}_2F_1(a,b,c|x) = {} & \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)} (-z)^{-a} {}_2F_1\left(a,a-c+1,a-b+1|\frac{1}{x}\right) \\ & + \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)} (-z)^{-b} {}_2F_1\left(b,b-c+1,b-a+1|\frac{1}{x}\right) \end{aligned} \tag{3} $$

which is valid for $\left|\frac{1}{x}\right|<1$.

The four Appell functions are defined by the following series representation

$$ F_1\left(a,b_1,b_2;c|x,y\right) = \sum_{m,n=0}^{\infty} \frac{(a)_{m+n} \left(b_1\right)_m \left(b_2\right)_n}{(c)_{m+n}} \frac{x^m}{m!} \frac{y^n}{n!}, \qquad |x|<1 \wedge |y|<1 \tag{4} $$

$$F_2\left(a,b_1,b_2;c_1,c_2|x,y\right)=\sum_{m,n=0}^{\infty}\frac{(a)_{m+n}\,(b_1)_m\,(b_2)_n}{(c_1)_m\,(c_2)_n}\frac{x^m}{m!}\frac{y^n}{n!},\qquad |x|+|y|<1 \tag{5}$$

$$F_3\left(a_1,a_2,b_1,b_2;c|x,y\right)=\sum_{m,n=0}^{\infty}\frac{(a_1)_m\,(a_2)_n\,(b_1)_m\,(b_2)_n}{(c)_{m+n}}\frac{x^m}{m!}\frac{y^n}{n!},\qquad |x|<1\wedge|y|<1 \tag{6}$$

$$F_4\left(a,b;c_1,c_2|x,y\right)=\sum_{m,n=0}^{\infty}\frac{(a)_{m+n}(b)_{m+n}}{(c_1)_m\,(c_2)_n}\frac{x^m}{m!}\frac{y^n}{n!},\qquad \sqrt{|x|}+\sqrt{|y|}<1 \tag{7}$$

The analysis of three out of the four Appell functions has been done previously. The analytic continuations of Appell F_1 are obtained in [4], while in [8, 9], Appell F_4 is analysed in a similar manner and Appell F_2 is treated in [7]. A package `AppellF2.wl` for the numerical implementation of F_2 in MATHEMATICA is also presented in [7], along with the usual analysis and evaluation of analytic continuations. There already exist a numerical program to evaluate Appell F_1 [67,68]. It should also be noted that all the Appell functions have been implemented in MAPLE [70] and in MATHEMATICA v.13.3 [71] recently. However, the numerical tests performed and presented in Section 6, show that the numerical evaluation of these functions using these implementations is still far from perfect. Therefore, the present analysis is timely and may be useful for improving such implementations further.

Apart from the above series representations, the Appell functions also have Euler integral representations that can be used for analytic continuation purposes. Since our interest lies in the analytic continuations of Appell F_1 and F_3, we only present the integral representations of these, which are as follows.

$$F_1\left(a,b_1,b_2,c|x,y\right)=\frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 t^{a-1}(1-t)^{c-a-1}(1-xt)^{-b_1}(1-yt)^{-b_2}\,\mathrm{d}t,\quad \Re(c)>\Re(a)>0. \tag{8}$$

and,

$$F_3\left(a_1,a_2,b_1,b_2;c|x,y\right)=\frac{\Gamma(c)}{\Gamma(b_1)\Gamma\left(b_2\right)\Gamma\left(c-b_1-b_2\right)}\times\iint u^{b-1}v^{b_2-1}(1-u-v)^{c-b_1-\beta'-1}(1-ux)^{-a_1}(1-vy)^{-a_2}\mathrm{d}u\;\mathrm{d}v, \tag{9}$$

where integration domain is given by: $u\geq 0,\quad v\geq 0,\quad u+v\leq 1,$
and the integral is convergent for: $\Re(b_1)>0,\,\Re\left(b_2\right)>0,\,\Re\left(c-b_1-b_2\right)>0.$

Another important property of interest to us is the knowledge of the singular loci of these hypergeometric functions, where we omit the evaluation of these functions in general. For Appell F_1 the singular loci are found to be [4]

$$x=0\vee y=0\vee x=1\vee y=1\vee x=y \tag{10}$$

Similarly for Appell F_3 they are as follows [72]

$$x=0\vee y=0\vee x=1\vee y=1\vee xy=x+y \tag{11}$$

2.2 Lauricella Function: $F_D^{(N)}$

Lauricella further generalized the four Appell functions to $N-$variables in [2]. In this Section, we provide the definition and some basic properties of one such function of interest to us, i.e., Lauricella $F_D^{(N)}$ which is the generalization of double variable Appell F_1. The Lauricella $F_D^{(N)}$, in $N-$variables, is defined by the following series representation

$$F_D^{(N)}(a,b_1,\cdots,b_N;c|z_1,\cdots,z_N)=\sum_{j_1=0}^{\infty}\cdots\sum_{j_N=0}^{\infty}\frac{(a)_{j_1+\cdots+j_N}(b_1)_{j_1}\cdots(b_N)_{j_N}}{(c)_{j_1+\cdots+j_N}}\frac{z_1^{j_1}\cdots z_N^{j_N}}{j_1!\cdots j_N!} \tag{12}$$

which is valid for $|z_1| < 1 \wedge \cdots \wedge |z_N| < 1$.

The function remarkably has a one-dimensional Euler-type integral representation given by

$$F_D^{(N)}\left(a, b_1, \ldots, b_n, c | x_1, \ldots, x_n\right) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 t^{a-1}(1-t)^{c-a-1}\left(1-x_1 t\right)^{-b_1} \cdots \left(1-x_n t\right)^{-b_n} \mathrm{d}t, . \tag{13}$$

which is convergent for $\Re(c) > \Re(a) > 0$.

For $N = 2$, Eq. (12) reduces to Appell F_1 series

$$F_1(a, b_1, b_2; c | z_1, z_2) = F_D^{(2)}(a, b_1, b_2; c | z_1, z_2) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \frac{(a)_{j_1+j_2} (b_1)_{j_1} (b_2)_{j_2}}{(c)_{j_1+j_2}} \frac{z_1^{j_2} z_2^{j_2}}{j_2! j_2!} \tag{14}$$

For $N = 3$, Eq. (12) reduces to Lauricella $F_D^{(3)}$ series

$$F_D^{(3)}(a, b_1, b_2, b_3; c | z_1, z_2, z_3) = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \sum_{j_3=0}^{\infty} \frac{(a)_{j_1+j_2+j_3} (b_1)_{j_1} (b_2)_{j_2} (b_3)_{j_3}}{(c)_{j_1+j_2+j_3}} \frac{z_1^{j_2} z_2^{j_2} z_3^{j_3}}{j_2! j_2! j_3!} \tag{15}$$

The singular loci of $F_D^{(3)}$ are given by,

$$x = 0 \vee y = 0 \vee z = 0 \vee x = 1 \vee y = 1 \vee z = 1 \vee x = y \vee y = z \vee x = z \tag{16}$$

2.3 Lauricella-Saran function : $F_S^{(3)}$

Another three-variable hypergeometric series of interest to us is the Lauricella-Saran $F_S^{(3)}$ [3]. It is defined as

$$F_S^{(3)} := F_S^{(3)}(a_1, a_2, b_1, b_2, b_3; c | x, y, z) = \sum_{m,n,p=0}^{\infty} \frac{(a_1)_m (a_2)_{n+p} (b_1)_m (b_2)_n (b_3)_p}{(c)_{m+n+p}} \frac{x^m y^n z^p}{m! n! p!} \tag{17}$$

The defining series representation of the $F_S^{(3)}$ is valid in [23]

$$|x| < 1 \wedge |y| < 1 \wedge |z| < 1 \tag{18}$$

The Lauricella-Saran function $F_S^{(3)}$ also has a triple Euler type integral representation as follows

$$F_S^{(3)} = \frac{\Gamma(c)}{\Gamma(b_1)\Gamma(b_2)\Gamma(b_3)\Gamma(c-b_1-b_2-\beta_3)} \times \iiint u^{b_1-1} v^{b_2-1} w^{b_3-1} (1-u-v-w)^{c-b_1-b_2-b_2-1} (1-ux)^{-a_1} (1-vy-wz)^{-a_2} du\, dv\, dw. \tag{19}$$

which is convergent for $\Re(c) > \Re(b_1+b_2+b_3) > 0$, $\Re(c) > \Re(b_1) > 0$, $\Re(b_2) > 0$, $\Re(b_3) > 0$.

The singular loci of the Lauricella-Saran function are found to be [73]

$$x = 0 \vee y = 0 \vee z = 0 \vee x = 1 \vee y = 1 \vee z = 1 \vee y = z \vee x + y = xy \vee x + z = xz \tag{20}$$

We would like to remark that we only list some basic properties of all these functions of interest. For other properties of these functions, such as the system of the partial differential equation they satisfy, symmetry relations, transformation properties, etc, we refer the readers to the standard references [2, 13, 21–25, 74].

3 The Method of Olsson

Appell and Kampé de Feriet [75] have pointed out that analytic continuations of a multivariable hypergeometric function can be derived by leveraging the known analytic continuations of the hypergeometric functions with variables lower in number than that of the former function. In [4], Olsson has obtained the analytic continuations of the double variable Appell F_1 function [1] employing the linear transformation formulae of the Gauss ${}_2F_1$ function. The same technique has been used to find the analytic continuations of the Appell F_2 recently, which are the backbone of the numerical package `AppellF2.wl` [7]. This powerful technique to find the analytic continuations of multivariable hypergeometric function is automated in the MATHEMATICA based package `Olsson.wl` [69], which eases the calculations significantly.

In this Section, to illustrate the method, we find the analytic continuations of the Lauricella-Saran function around $(0,\infty,\infty)$ and (∞,∞,∞) using the method of Olsson.

One can immediately observe the following symmetry of the F_S.

$$F_S^{(3)}(a_1,a_2,b_1,b_2,b_3;c|x,y,z)=F_S^{(3)}(a_1,a_2,b_1,b_3,b_2;c|x,z,y) \tag{21}$$

We observe that summing over one of the summation indices, one can write the summand in terms of double variable Appell F_1 or F_3 function

$$F_S^{(3)}(a_1,a_2,b_1,b_2,b_3;c|x,y,z)=\sum_{m=0}^{\infty}\frac{(a_1)_m\,(b_1)_m}{(c)_m}\frac{x^m}{m!}F_1\left(a_2;b_2,b_3;c+m|y,z\right) \tag{22}$$

$$=\sum_{n=0}^{\infty}\frac{(a_2)_n\,(b_2)_n}{(c)_n}\frac{y^n}{n!}F_3\left(a_1,b_3,b_1,a_2+n;c+n|x,z\right) \tag{23}$$

$$=\sum_{p=0}^{\infty}\frac{(a_2)_p\,(b_3)_p}{(c)_p}\frac{z^p}{p!}F_3\left(a_1,b_2,b_1,a_2+p;c+p|x,y\right) \tag{24}$$

Thus, a good use of the analytic continuations of the Appell functions can be made to find the analytic continuations of $F_S^{(3)}$. To find the analytic continuation of the Lauricella-Saran function around $(0,\infty,\infty)$, we use the analytic continuation of the Appell F_1 around (∞,∞) from [4] (precisely Eq. (22) of the [4]), which is shown below.

$$\begin{aligned}
&F_1(a,b_1,b_2;c|x,y)\\
&=(-x)^{-b_1}(-y)^{b_1-a}\frac{\Gamma(c)\Gamma\left(a-b_1\right)\Gamma\left(-a+b_1+b_2\right)}{\Gamma(a)\Gamma\left(b_2\right)\Gamma(c-a)}\,G_2\left(b_1,a-c+1,a-b_1,-a+b_1+b_2\Big|-\frac{y}{x},-\frac{1}{y}\right)\\
&+(-x)^{-a}\frac{\Gamma(c)\Gamma\left(b_1-a\right)}{\Gamma\left(b_1\right)\Gamma(c-a)}F_1\left(a,a-c+1,b_2;a-b_1+1\Big|\frac{1}{x},\frac{y}{x}\right)\\
&+(-x)^{-b_1}(-y)^{-b_2}\frac{\Gamma(c)\Gamma\left(a-b_1-b_2\right)}{\Gamma(a)\Gamma\left(c-b_1-b_2\right)}F_1\left(b_1+b_2-c+1,b_1,b_2;-a+b_1+b_2+1\Big|\frac{1}{x},\frac{1}{y}\right)
\end{aligned} \tag{25}$$

where G_2 is the one of the Horn's functions. The domain of convergence of the above analytic continuation is given by $\frac{1}{|x|}<1\wedge\frac{1}{|y|}<1\wedge\left|\frac{y}{x}\right|<1$.

Inserting the above expression in Eq. (22) and simplifying the gamma functions and Pochhammer symbols, we obtain the analytic continuation of $F_S^{(3)}$ around $(0,\infty,\infty)$

$$\begin{aligned}
F_S^{(3)}&=(-y)^{-a_2}\frac{\Gamma(c)\Gamma\left(b_2-a_2\right)}{\Gamma\left(b_2\right)\Gamma\left(c-a_2\right)}\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m\,(b_1)_m\,(b_3)_p\,(a_2)_{n+p}\,(-c+a_2+1)_{n-m}}{(a_2-b_2+1)_{n+p}}\frac{(-x)^m y^{-n-p}z^p}{m!n!p!}\\
&+(-y)^{-b_2}(-z)^{b_2-a_2}\frac{\Gamma(c)\Gamma\left(a_2-b_2\right)\Gamma\left(-a_2+b_2+b_3\right)}{\Gamma\left(a_2\right)\Gamma\left(b_3\right)\Gamma\left(c-a_2\right)}\\
&\times\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m\,(b_1)_m\,(b_2)_n\,(a_2-b_2)_{p-n}\,(-c+a_2+1)_{p-m}}{(a_2-b_2-b_3+1)_{p-n}}\frac{(-x)^m y^{-n}z^{n-p}}{m!n!p!}
\end{aligned}$$

$$+(-y)^{-b_2}(-z)^{-b_3}\frac{\Gamma(c)\Gamma(a_2-b_2-b_3)}{\Gamma(a_2)\Gamma(c-b_2-b_3)}$$
$$\times\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_1)_m(b_2)_n(b_3)_p(-c+b_2+b_3+1)_{-m+n+p}}{(-a_2+b_2+b_3+1)_{n+p}}\frac{(-x)^m y^{-n}z^{-p}}{m!n!p!} \tag{26}$$

The domain of convergence can be found using the Horn's theorem [23]. The above analytic continuation is valid in:

$$|x|<1\wedge\frac{1}{|y|}+1<\frac{1}{|x|}\wedge\frac{1}{|y|}<1\wedge\frac{1}{|z|}+1<\frac{1}{|x|}\wedge\frac{1}{|z|}<1\wedge\left|\frac{z}{y}\right|<1 \tag{27}$$

One can immediately find another analytic continuation of the $F_S^{(3)}$ around $(0,\infty,\infty)$ using the symmetry relation (i.e., Eq. (21)), which we do not present here explicitly.

Next, we find the analytic continuation around (∞,∞,∞) by taking each of the three series of Eq. (26), summing over the index m and using the analytic continuation of ${}_2F_1(x)$ around $x=\infty$ (i.e., Eq. (3)). Let us denote the first series of Eq. (26) as S_1

$$\begin{aligned}S_1&=\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_1)_m(b_3)_p(a_2)_{n+p}(-c+a_2+1)_{n-m}}{(a_2-b_2+1)_{n+p}}\frac{(-x)^m y^{-n-p}z^p}{m!n!p!}\\&=\sum_{n,p=0}^{\infty}\frac{(b_3)_p(-c+a_2+1)_n(a_2)_{n+p}}{(a_2-b_2+1)_{n+p}}\frac{y^{-n-p}z^p}{n!p!}\,{}_2F_1\left(a_1,b_1;c-n-a_2|x\right)\end{aligned} \tag{28}$$

In the second equality, the summation over the index m is explicitly taken. As a result, the Gauss ${}_2F_1$ appears in the summand. Now we use Eq. (3) to find the analytic continuation of the series S_1, which is labelled as S_1' below

$$\begin{aligned}S_1'=&(-x)^{-a_1}\frac{\Gamma(b_1-a_1)\Gamma(c-a_2)}{\Gamma(b_1)\Gamma(c-a_1-a_2)}\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_3)_p(a_2)_{n+p}(-c+a_1+a_2+1)_{m+n}}{(a_1-b_1+1)_m(a_2-b_2+1)_{n+p}}\frac{x^{-m}y^{-n-p}z^p}{m!n!p!}\\&+(-x)^{-b_1}\frac{\Gamma(a_1-b_1)\Gamma(c-a_2)}{\Gamma(a_1)\Gamma(c-a_2-b_1)}\sum_{m,n,p=0}^{\infty}\frac{(b_1)_m(b_3)_p(a_2)_{n+p}(-c+a_2+b_1+1)_{m+n}}{(-a_1+b_1+1)_m(a_2-b_2+1)_{n+p}}\frac{x^{-m}y^{-n-p}z^p}{m!n!p!}\end{aligned} \tag{29}$$

Similarly, taking the second series of Eq. (26) (which we denote by S_2) and proceeding as before

$$\begin{aligned}S_2&=\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_1)_m(b_2)_n(a_2-b_2)_{p-n}(-c+a_2+1)_{p-m}}{(a_2-b_2-b_3+1)_{p-n}}\frac{(-x)^m y^{-n}z^{n-p}}{m!n!p!}\\&=\sum_{n,p=0}^{\infty}\frac{(b_2)_n(-c+a_2+1)_p(a_2-b_2)_{p-n}}{(a_2-b_2-b_3+1)_{p-n}}\frac{y^{-n}z^{n-p}}{n!p!}\,{}_2F_1\left(a_1,b_1;c-p-a_2|x\right)\end{aligned} \tag{30}$$

Again using the analytic continuation of ${}_2F_1(\ldots;x)$ around $x=\infty$, we obtain the analytic continuation of S_2 as

$$\begin{aligned}S_2'=&(-x)^{-a_1}\frac{\Gamma(b_1-a_1)\Gamma(c-a_2)}{\Gamma(b_1)\Gamma(c-a_1-a_2)}\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_2)_n(a_2-b_2)_{p-n}(-c+a_1+a_2+1)_{m+p}}{(a_1-b_1+1)_m(a_2-b_2-b_3+1)_{p-n}}\frac{x^{-m}y^{-n}z^{n-p}}{m!n!p!}\\&+(-x)^{-b_1}\frac{\Gamma(a_1-b_1)\Gamma(c-a_2)}{\Gamma(a_1)\Gamma(c-a_2-b_1)}\sum_{m,n,p=0}^{\infty}\frac{(b_1)_m(b_2)_n(a_2-b_2)_{p-n}(-c+a_2+b_1+1)_{m+p}}{(-a_1+b_1+1)_m(a_2-b_2-b_3+1)_{p-n}}\frac{x^{-m}y^{-n}z^{n-p}}{m!n!p!}\end{aligned} \tag{31}$$

Finally, we consider the third series S_3

$$S_3=\sum_{m,n,p=0}^{\infty}\frac{(a_1)_m(b_1)_m(b_2)_n(b_3)_p(-c+b_2+b_3+1)_{-m+n+p}}{(-a_2+b_2+b_3+1)_{n+p}}\frac{(-x)^m y^{-n}z^{-p}}{m!n!p!}$$

$$= \sum_{n,p=0}^{\infty} \frac{(b_2)_n (b_3)_p (-c+b_2+b_3+1)_{n+p}}{(-a_2+b_2+b_3+1)_{n+p}} \frac{y^{-n} z^{-p}}{n!p!} \, {}_2F_1(a_1, b_1; c-n-p-b_2-b_3|x) \tag{32}$$

Using the analytic continuation of ${}_2F_1(\ldots;x)$ around $x=\infty$, we find the analytic continuation of S_3

$$S_3' = (-x)^{-a_1} \frac{\Gamma(b_1-a_1)\Gamma(c-b_2-b_3)}{\Gamma(b_1)\Gamma(c-a_1-b_2-b_3)} \sum_{m,n,p=0}^{\infty} \frac{(a_1)_m (b_2)_n (b_3)_p (-c+a_1+b_2+b_3+1)_{m+n+p}}{(a_1-b_1+1)_m (-a_2+b_2+b_3+1)_{n+p}} \frac{x^{-m} y^{-n} z^{-p}}{m!n!p!}$$
$$+ (-x)^{-b_1} \frac{\Gamma(a_1-b_1)\Gamma(c-b_2-b_3)}{\Gamma(a_1)\Gamma(c-b_1-b_2-b_3)} \sum_{m,n,p=0}^{\infty} \frac{(b_1)_m (b_2)_n (b_3)_p (-c+b_1+b_2+b_3+1)_{m+n+p}}{(-a_1+b_1+1)_m (-a_2+b_2+b_3+1)_{n+p}} \frac{x^{-m} y^{-n} z^{-p}}{m!n!p!} \tag{33}$$

Thus, the analytic continuation of the $F_S^{(3)}$ around (∞,∞,∞) is sum of the expressions Eq. (29), Eq. (33) and Eq. (33) multiplied by appropriate prefactors.

$$\begin{aligned} F_S^{(3)} &= (-y)^{-a_2} \frac{\Gamma(c)\Gamma(b_2-a_2)}{\Gamma(b_2)\Gamma(c-a_2)} \times S_1' \\ &+ (-y)^{-b_2} (-z)^{b_2-a_2} \frac{\Gamma(c)\Gamma(a_2-b_2)\Gamma(-a_2+b_2+b_3)}{\Gamma(a_2)\Gamma(b_3)\Gamma(c-a_2)} \times S_2' \\ &+ (-y)^{-b_2} (-z)^{-b_3} \frac{\Gamma(c)\Gamma(a_2-b_2-b_3)}{\Gamma(a_2)\Gamma(c-b_2-b_3)} \times S_3' \end{aligned} \tag{34}$$

We do not write the expression explicitly due to its long length. The corresponding domain of convergence can be found using the Horn's theorem as

$$\frac{1}{|x|} < 1 \wedge \frac{1}{|x|} + \frac{1}{|y|} < 1 \wedge \frac{1}{|y|} < 1 \wedge \frac{1}{|x|} + \frac{1}{|z|} < 1 \wedge \frac{1}{|z|} < 1 \wedge \left|\frac{z}{y}\right| < 1 \tag{35}$$

Another analytic continuation of the $F_S^{(3)}$ around (∞,∞,∞) can be readily found by employing the symmetry relation Eq. (21).

It is to be noted that, the analytic continuations above are valid for *generic values* of Pochhammer parameters, which means that the parameters do not assume such values that cause the series to terminate or give divergent results. This typically corresponds to the cases when the difference of parameters is not an integer. However, for the non-generic case, when the difference is an integer, these analytic continuations remain valid as long as a proper limiting procedure is applied. Strategies for using the analytic continuations for non-generic cases have also been implemented in all the four packages developed, which is be discussed in Section 7.4.

Due to the large number of analytic continuations derived, it is not possible to present all of them in text form. For the ease of the reader, we have built commands, to be discussed in Section 5, so that one can see the explicit form of the analytic continuations directly from the package. The total number of analytic continuations, including the original series definition of each of the functions, that are stored inside the packages is given below.

- `AppellF1.wl`: 28
- `AppellF3.wl`: 24
- `LauricellaFD.wl`: 96
- `LauricellaSaranFS.wl`: 102

4 Algorithm of the packages

We now discuss the algorithm used for the package in detail. Three primary issues must be addressed to ensure the efficient evaluation of numerical values within the package, which are as follows.

1. *Selection of a suitable analytic continuation*: Although for a point of interest, there are many analytic continuations that can be used for evaluation, not all of them are equally efficient for evaluation purposes because convergence rates vary among them. The analytic continuations, being composed of multiple series, further complicate the matter, as not all series have favourable convergence rates. Hence, a suitable analytic continuation has to be selected for numerical evaluation, taking into account all of these factors.

2. *Strategy to perform the summation efficiently*: With the increase in the fold of the series, the amount of time required to evaluate finite series summation takes an exponentially larger time. Given a $p-$fold series, the number of operations required to perform summation, taking `N` term in all the summation indices, is $\mathcal{O}(\texttt{N}^p)$. Hence, any strategy that can reduce the number of operations is the key to achieving faster evaluations.

3. *Strategy for numerical evaluation with non-generic parameter values*: In general, the analytic continuations obtained using the method of Olsson are only valid for generic values of the Pochhammer parameters. They, however, are also required for non-generic parameters, which are useful for physical applications as well. For such non-generic parameters, proper limiting procedures have to be used such that the divergences amongst all the series appearing in a certain analytic continuation cancels.

We discuss the first two issues below, keeping the last one for Section 7.4.

For the first problem, we follow the strategy utilized in [7], for the package `AppellF2.wl`. It relies on defining a certain 'rate of convergence' for a given analytic continuation and then selecting the one with the fastest rate of convergence. We have used the same strategy for all the four packages.

For the second issue, we use two strategies, which we now discuss in detail for both two and three-variable cases. As a demonstrative example, we consider Appell F_1, though the same strategies apply to others as well. Consider the series representation of F_1 given in Eq. (4). If we sum over index n, we get the following

$$F_1\left(a, b_1, b_2; c|x, y\right) = \sum_{m=0}^{\infty} \frac{(a)_m \left(b_1\right)_m}{(c)_m m!} x^m \, {}_2F_1(a+m, b_2; c+m|y) \tag{36}$$

The strategy now is to use Eq. (36), since the evaluation of F_1, which originally is given by a double summation, has now been reduced to the evaluation of a single fold summation. For such evaluations, we evaluate the ${}_2F_1$ on the right-hand side of Eq. (36) using the internal command of MATHEMATICA, `Hypergeometric2F1`. The time taken for the evaluation using this strategy for various points for Appell F_1 is compared with that of the usual double summation in Table 1.

(x, y)	Eq. (4) (`N` = 100)	Eq. (36) (`N` = 100)	Eq. (4) (`N` = 300)	Eq. (36) (`N` = 300)
$(\frac{1}{10}, \frac{1}{10})$	1.1927075668481758434 t= 0.978377	1.1927075668481758434 t= 0.373886	1.1927075668481758434 t= 27.0116	1.1927075668481758434 t= 0.579974
$(\frac{1}{10}, \frac{9}{10})$	9.6274055869884276609 t= 0.731865	9.6274055869884276609 t= 0.389833	9.6274915622753831775 t= 26.638	9.6274915622753831775 t= 0.398074
$(\frac{9}{10}, \frac{1}{10})$	3.0767526672458511801 t= 1.03145	3.0767528856677694923 t= 0.469786	3.0767528856677694748 t= 26.7015	3.0767528856677694923 t= 1.00073
$(\frac{9}{10}, \frac{9}{10})$	135.11971472599376751 t= 1.05071	135.12030231208480330 t= 0.464883	135.12909322010073107 t= 27.4262	135.12909322010080844 t= 1.18971

Table 1: The values of Appell F_1 for various points are obtained using Eq. (4) and Eq. (36). The Pochhammer parameters are $a = \frac{123}{100}, b_1 = \frac{234}{100}, b_2 = \frac{398}{100}, c = \frac{47}{10}$. The values are displayed with 20 significant digits. `N` is the maximum value of each index for finite summation, and t is the time in seconds for the evaluation in a typical run. The run timing corresponds to the time taken for a typical evaluation in a typical laptop.

From Table 1, it is observed that Eq. (36) outperforms Eq. (4) in both time and convergence. The four points are selected to test the performance for points both inside as well as near the boundaries

of the convergence region, where the convergence is usually slow. For the point $(9/10, 1/10)$, it is observed that Eq. (4) is less accurate than Eq. (36), which has already converged for $\mathtt{N} = 100$. The point $(9/10, 9/10)$ is very close to the boundary (in both directions), and the result is extremely slow to converge even with Eq. (36). The time taken by Eq. (4) is always higher than Eq. (36). It is also noticed that the time taken by Eq. (4) increases exponentially when the finite summation limit goes from $\mathtt{N} = 100$ to $\mathtt{N} = 300$, whereas the increment is not significant with Eq. (36).

We employ a similar strategy for three variable cases as well. Consider Eq. (15), summing over index p we get the following

$$F_D^{(3)}(a, b_1, b_2, b_3; c|x, y, z) = \sum_{m,n=0}^{\infty} \frac{x^m y^n (b_1)_m (b_2)_n (a)_{m+n}}{m! n! (c)_{m+n}} \, {}_2F_1(b_3, a+m+n; c+m+n|z) \tag{37}$$

We now consider the following property of double summation [76]

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a(m, n) = \sum_{m=0}^{\infty} \sum_{n=0}^{m} a(m-n, n) \tag{38}$$

In the finite summations where the left-hand side of Eq. (38) has $\mathtt{N}$ as the summation limit, $\frac{\mathtt{N}^2}{2}$ terms are required to be evaluated on the right-hand side, as opposed to $\mathtt{N}^2$ terms on the left-hand side. Thus, we decrease the number of operators in taking the finite summation by a factor of 2 by rewriting Eq. (37) as

$$F_D^{(3)}(a, b_1, b_2, b_3|c; x, y, z) = \sum_{m=0}^{\infty} \sum_{n=0}^{m} \frac{y^n (a)_m (b_2)_n x^{m-n} (b_1)_{m-n} \, {}_2F_1(b_3, a+m; c+m|z)}{n!(m-n)!(c)_m} \tag{39}$$

The time taken for the evaluation with this strategy for various points for $F_D^{(3)}$ is contrasted with that of the usual double summation in Table 2. We observe similar behaviour as table 1 for table 2 as well.

(x, y, z)	Eq. (15) ($\mathtt{N} = 50$)	Eq. (39) ($\mathtt{N} = 50$)	Eq. (15) ($\mathtt{N} = 100$)	Eq. (39) ($\mathtt{N} = 100$)
$(\frac{1}{10}, \frac{1}{10}, \frac{1}{10})$	1.0732135550913555676 t= 7.26767	1.0732135550913555676 t= 0.869634	1.0732135550913555676 t= 116.874	1.0732135550913555676 t= 1.94078
$(\frac{1}{10}, \frac{1}{10}, \frac{9}{10})$	1.6231078297620145196 t= 8.42558	1.6240002995450544771 t= 2.00745	1.6239968320342584400 t= 124.944	1.6240002995450544771 t= 9.87634
$(\frac{1}{10}, \frac{9}{10}, \frac{1}{10})$	2.0091501375087473590 t= 7.64192	2.0109093110008280114 t= 1.7185	2.0109022514155545663 t= 125.285	2.0109022330996805423 t= 1.87644
$(\frac{9}{10}, \frac{1}{10}, \frac{1}{10})$	2.5054373292884746698 t= 8.04876	2.5054313386996026729 t= 0.870391	2.5085713622543825408 t= 122.186	2.5085713381376995373 t= 1.80294
$(\frac{9}{10}, \frac{9}{10}, \frac{9}{10})$	5.6025209788374342344 t= 8.30426	5.6011675291597760012 t= 2.00511	5.6131499649884968719 t= 122.937	5.6131385069711024026 t= 9.48653

Table 2: Table of values obtained for various points obtained using Eq. (15) and Eq. (39). Pochhammer parameters used are $a = \frac{13}{10}, b_1 = \frac{1}{5}, b_2 = \frac{1}{7}, b_3 = \frac{1}{11}, c = \frac{11}{13}$. Values are shown up to 20 significant digits, $\mathtt{N}$ denotes the maximum value of each of the indices used for finite summation, and t denotes the time taken to complete the evaluation in seconds for a typical run. The run timing corresponds to the time taken for a typical evaluation in a typical laptop.

It is to be noted that out of many choices of choosing a summation index, to sum over, we choose the one such that the simplest of the hypergeometric functions is obtained. The preference is : ${}_2F_1 > {}_3F_2 > \cdots > {}_pF_{p-1}$. In the studies of both two variables and three-variable hypergeometric functions, ${}_3F_2$ and ${}_4F_3$ are rarely obtained, and ${}_5F_4$ or higher order hypergeometric functions are never obtained. This suggests that a particular summed index may be better than the other for numerical evaluation. However, this aspect is not focused on in the present investigation, and it will be a part of future investigation.

5 Demonstration of the packages

We now demonstrate the commands of the packages with examples. There are four commands in the packages `LauricellaFD.wl` and `LauricellaSaranFS.wl` and the `AppellF1.wl` and `AppellF3.wl` contain five commands each, which are enlisted below.

- `F1, F3, FD3, FS3` : Given a point (x, y) for a two variable functions or (x, y, z) for a three variable functions and the values of the parameters, these commands yield the numerical values of the corresponding functions.
- `F1findall, F3findall, FD3findall, FS3findall`: Given a point (x, y) or (x, y, z), these commands give a list of all the analytic continuations, valid at that given point.
- `F1expose, F3expose, FD3expose, FS3expose` : These commands can be used to expose particular analytic continuations that the user wants to see along with its region of convergence.
- `F1evaluate, F3evaluate, FD3evaluate, FS3evaluate`: Given a point (x, y) for two variable functions or (x, y, z) for three variable functions and the numerical values of the parameters, these commands can be used to evaluate the function of interest using a chosen analytic continuation.
- `F1ROC, F3ROC` : These commands are for visual aid to determine where a given point (x, y) lies w.r.t a given analytic continuation.

The packages can be downloaded from the link:

https://github.com/souvik5151/Appell_Lauricella_Saran_functions

Below, we consider the example of the Appell F_1 for the sake of demonstration of these commands. The usage of the commands associated with other functions is analogous. To load the package `AppellF1.wl` , we use the following command,

```
In[1]:= <<AppellF1.wl

        AppellF1 v1.0
        Authors : Souvik Bera & Tanay Pathak
```

The main command of the package is the following:

```
F1[a, b1, b2, c, x, y, precision, sum_limit, verbose-> True]
```

The arguments of the commands are :

- `a, b1, b2, c` are the Pochhammer parameters and `x, y` is the given point (x, y).
- `precision` refers to the precision up to which the result is required to be displayed.
- `sum_limit` refers to the upper limit of the finite summation. In all the future text, we will denote it by `N`.
- `verbose` is an option used to display information related to the computation, such as the serial number of the valid analytic continuations inside the package and their corresponding convergence rates, as we show in the example below. By default, the value of this option is chosen to be `False`.

The command can be used as follows.

```
In[2]:= F1[1.23, 2.34, 3.98, 4.7, 1.9, .9, 5, 100]

Out[2]= 5.6680 + 17.050 I
```

With the value of `verbose` as `True`, we get some more information as follows.

```
In[3]:= F1[1.23, 2.34, 3.98, 4.7, 1.9, .9, 5, 100, verbose-> True]

        valid series : {9,12,20,23}
        convergence rates :{{0.1159109377,23},{0.1159355395,9},{0.2447218050,20},
        {0.6068309198,12}}
        selected series : 23

Out[3]= 5.6680 + 17.050 I
```

The command `F1expose` takes the input in the following way

```
F1expose[series number]
```

The command can be used as follows

```
In[4]:= F1expose[1]

Out[4]= {Abs[x]<1&&Abs[y]<1, x^m y^n Pochhammer[a,m+n]Pochhammer[b1,m]Pochhammer[b2,n] / m!n!Pochhammer[c,m+n]}
```

The command `F1findall` takes the input in the following way

```
F1findall[{x,y}]
```

As an example,

```
In[5]:= F1findall[{1.9, .9}]

Out[5]= {9, 12, 20, 23}
```

The above code implies that the given point `{1.9,.9}` lies inside the region of convergence of analytic continuation number `{9, 12, 20, 23}`. The analytic continuation number corresponds to the number associated with it inside the package internally.

Next, we consider the `F1evaluate` command. It takes the inputs as

```
F1evaluate[#,{a, b1, b2, c, x, y}, precision, sum_limit]
```

For instance, F_1 is evaluated with the analytic continuation labeled as `#` $= 9$, with the same set of Pochhammer parameters and x, y as

```
In[6]:= F1evaluate[9, {1.23, 2.34, 3.98, 4.7, 1.9, .9}, 5, 100]

Out[6]= 5.6680 + 17.050 I
```

We would like to remark that for better efficiency, the internal evaluation of the finite summation, in both `F1` and `F1evaluate` commands (and analogous commands in other packages), is done using `ParalllelSum` command of MATHEMATICA .

All the above commands, which are only demonstrated in the context of the package `AppellF1.wl` , have a similar analogue in the other three packages as well. However, in each of the packages `AppellF1.wl` and `AppellF3.wl` there is one more command which is useful for visualization of the region of convergences. The command is `F1ROC` (and `F3ROC`). It takes the input as

```
F1ROC[{x,y}, #, range]
```

For example,

```
In[7]:= F1ROC[{1.9, .9}, 9, {0, 5}]
```

The output of the above command is the plot shown in Fig.1. The red dot is the point $(x, y) = (1.9, .9)$, and the blue region is the region of convergence of the analytic continuation labeled as `# = 9`. The range of the plot is `range` $= [0, 5]$, which is specified by the user in the last argument of `F1ROC` command.

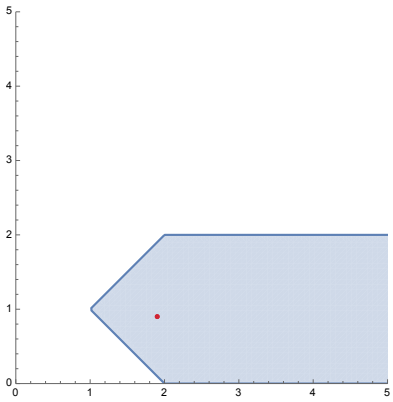


Figure 1: Output of `F1ROC` command.

6 Numerical tests

We perform a variety of numerical tests to check the consistency and accuracy of the packages. A detailed list of these numerical checks is given below.

- Consistency check for generic and non-generic cases. These checks ensure that all the valid analytic continuations for a given point gives the same result.
- Test of comparison with the internal commands of MATHEMATICA and MAPLE.
- Testing against reduction formula.
- Testing against integral representation.
- For the cases of F_1, F_3 and $F_D^{(3)}$, checks are also performed by doing the numerical evaluation of Feynman integrals, in which these functions appear. Feynman integrals are evaluated numerically using FIESTA [77].

All the crosschecks are performed in MATHEMATICA v.13.3 and MAPLE 2021.

In Table 3, we present the comparison of values of Appell F_1 obtained using the package `AppellF1.wl` , inbuilt command `AppellF1` of MATHEMATICA [1] and MAPLE. The Pochhammer parameters and the points (x, y) are chosen randomly in the range $[-5, 5]$. In Table 3, we only display the values obtained for five such points, and a MATHEMATICA notebook containing a larger table of consistency checks is provided in the ancillary file. We observe that there are instances where MAPLE does not give any output. We also observe that there are two instances where results obtained using MATHEMATICA do not match with the results obtained using the package `AppellF1.wl` and MAPLE, whereas the latter two results are consistent. Furthermore, for all of the mentioned points, we also perform internal consistency checks. Given a random point, we evaluate all the valid analytic continuations and found that values obtained using different analytic continuations are consistent with each other.

In Table 4, we present a similar comparison of the values obtained using the package `AppellF3.wl` , and inbuilt command `AppellF3` of MATHEMATICA and MAPLE.

[1] These implementations are available in v13.3 and later versions.

a, b_1, b_2, c, x, y	`AppellF1.wl`	MATHEMATICA	MAPLE
-4.910115524 0.05551341196 1.272258581 1.701265421 1.865847217 -4.593616044	2346.739507 + 0.000021i	2346.739507 +0.000021i	#
0.06933467465 4.486129287 -2.299060382 3.132430057 4.301140034 1.525891559	0.8873796949 $-0.1955529752i$	0.0013530823877 $-0.0002148893964i$	#
1.903029939 1.018031546 -2.525202153 3.573481224 -3.141064731 4.180496216	-0.0677462366 $+1.0368642253i$	-0.0966032877 $+1.0345746851i$	-0.06774623789 $+1.036864225\,\mathrm{I}$
-0.7885795574 2.712909050 -2.552645509 -0.1250091986 0.3202377540 -1.557083143	58.57916734	58.57916734	58.57916733
4.425727083 -4.858260819 4.467162437 0.1571461627 -0.6971595548 -4.125289320	-0.002461165757	-0.01014766698	-0.002461165753

Table 3: Comparison of numerical values of Appell F_1 obtained using the package `AppellF1.wl`, inbuilt command `AppellF1` of MATHEMATICA and MAPLE. The symbol # in the last column means that, MAPLE fails to give any output in 10 minutes of run time.

7 Numerical test with non-generic values of parameters : An application to Feynman integrals

The numerical test of our packages for the non-generic values of Pochhammer parameters is discussed in this Section. It is frequently observed that the evaluation of Feynman integral in hypergeometric function representation produces function with non-generic values of Pochhammer parameters. Thus, our tests of these packages for non-generic values of parameters are motivated by Feynman integrals.

7.1 Passarino-Veltman functions

Consider the one loop N-point scalar Feynman integral, for the case when all external momenta are zero, $p_1 = p_2 = \cdots = p_N = 0$. Each propagator has index ν_i and mass m_i. The corresponding integral is given below.

$$I^N(\{\nu_i\};\{m_i\}) = \int \frac{d^d k}{(k^2 - m_1^2)^{\nu_1} \cdots (k^2 - m_N^2)^{\nu_N}} \tag{40}$$

a, b_1, b_2, c, x, y	Package	MATHEMATICA	MAPLE
-4.627149094 -1.817429336 4.926597166 3.148456748 -2.594781226 -3.573497861 -2.237584115	$-4.00373992374711 \times 10^7$	3.763432588×10^7 $-1.366207152 \times 10^7 i$	-4.003739919×10^7 $+0.002794901917i$
1.838637549 0.4850796858 0.09513160021 -0.7805808901 1.004405568 2.856794703 3.475873889	-0.2910249012 $+0.6554323865i$	-0.2910249012 $+0.6554323865i$	-0.2910249014 $+0.6554323869i$
4.162091546 -2.843597171 1.482790024 -0.6871600251 0.4147393972 0.8642634792 -4.248443060	226772.4287	#	226772.4289
3.498347332 -2.874934844 -2.837329275 -1.228710753 2.683926791 0.3173558206 -0.3309225764	0.04568785423	0.04568785423	0.04568785425
1.041216032 2.834971008 -0.01534212602 4.704941678 0.5710450502 -0.2428307966 -2.054034590	0.01661851674	0.01661851674	0.01661851674

Table 4: Comparison of numerical values obtained using the package `AppellF3.wl` , inbuilt command of `AppellF3` of MATHEMATICA and MAPLE.The symbol # in the 3rd column means that MATHEMATICA fails to give any output inn 10 minutes.

This integral can be written in terms of $(N-1)$ variable Lauricella $F_D^{(N-1)}$ function [43].

$$I^N(\{\nu_i\};\{m_i\}) = \pi^{d/2} i^{-d} (-m_N)^{d/2-\sum_j \nu_j} \frac{\Gamma(\sum_j \nu_j - d/2)}{\Gamma(\sum_j \nu_j)} \times F_D^{(N-1)}\left(\sum_j \nu_j - \frac{d}{2}, \nu_1, \cdots, \nu_{N-1}; \sum_j \nu_j \Big| 1 - \frac{m_1^2}{m_N^2}, \cdots, 1 - \frac{m_{N-1}^2}{m_N^2}\right) \tag{41}$$

Let us now consider some particular situations. When $N = 3$, we have the Passarino-Veltman C_0 function

$$\begin{aligned} C_0 :&= I^3(\{1,1,1\},\{m_1,m_2,m_3\}) \\ &= \frac{1}{2} i^{2\varepsilon+1} \pi^{2-\varepsilon} \Gamma(\varepsilon+1) \left(-m_3^2\right)^{-\varepsilon-1} F_1\left(\varepsilon+1,1,1;3\Big|1-\frac{m_1^2}{m_3^2}, 1-\frac{m_2^2}{m_3^2}\right) \end{aligned} \tag{42}$$

and for $N = 4$,

$$\begin{aligned} D_0 :&= I^3(\{1,1,1,1\},\{m_1,m_2,m_3,m_4\}) \\ &= \frac{1}{6} i^{2\varepsilon+1} \pi^{2-\varepsilon} \Gamma(\varepsilon+2) \left(-m_4^2\right)^{-\varepsilon-2} F_D^{(3)}\left(\varepsilon+2,1,1,1;4\Big|1-\frac{m_1^2}{m_4^2},1-\frac{m_2^2}{m_4^2},1-\frac{m_3^2}{m_4^2}\right) \end{aligned} \tag{43}$$

It is well-known, and in fact obvious, from the expression of Eqs. (42) and (43) that, the 3- and 4-point functions are convergent in $d = 4$ (i.e., $\varepsilon = 0$)

$$C_0(d=4) = -\frac{i\pi^2}{2m_3^2} F_1\left(1,1,1;3\Big|1-\frac{m_1^2}{m_3^2},1-\frac{m_2^2}{m_3^2}\right) \tag{44}$$

valid for $\left|1-\frac{m_1^2}{m_3^2}\right| < 1 \wedge \left|1-\frac{m_2^2}{m_3^2}\right| < 1$, and

$$D_0(d=4) = \frac{i\pi^2}{6m_4^4} F_D^{(3)}\left(2,1,1,1;4\Big|1-\frac{m_1^2}{m_4^2},1-\frac{m_2^2}{m_4^2},1-\frac{m_3^2}{m_4^2}\right) \tag{45}$$

valid for $\left|1-\frac{m_1^2}{m_4^2}\right| < 1 \wedge \left|1-\frac{m_2^2}{m_4^2}\right| < 1 \wedge \left|1-\frac{m_3^2}{m_4^2}\right| < 1$.

It is to be noted that, the ε-expansion of Appell F_1 and Lauricella $F_D^{(3)}$ functions in Eqs (42) and (43) can be found using computer programs such as Xsummer [78], nestedsums [79], `MultiHypExp` [80] and Diogenes [81]. Moreover, there exists reduction formulae of Appell F_1 and Lauricella $F_D^{(3)}$ functions, for integer valued Pochhammer parameters, to simpler functions such as ordinary logarithms. These are available in the literature and can also be obtained using the `ReduceFunction` command of `MultiHypExp` package. We find,

$$F_1(1,1,1;3 \mid x,y) = \frac{2\log(1-x)}{x(x-y)} + \frac{2\log(1-y)}{x-y} - \frac{2\log(1-x)}{x-y} - \frac{2\log(1-y)}{y(x-y)} \tag{46}$$

and

$$\begin{aligned} F_D^{(3)}(2,1,1,1;4 \mid x,y,z) = &\frac{6\log(1-x)}{x(x-y)(x-z)} + \frac{6\log(1-y)}{(x-y)(y-z)} + \frac{6\log(1-z)}{(x-z)(z-y)} \\ &- \frac{6\log(1-x)}{(x-y)(x-z)} - \frac{6\log(1-y)}{y(x-y)(y-z)} - \frac{6\log(1-z)}{z(x-z)(z-y)} \end{aligned} \tag{47}$$

Following the pattern, one can easily generalize the reduction formula for $F_D^{(N)}$.

Clearly, one needs to find analytic continuations of the Appell F_1 and Lauricella $F_D^{(3)}$ to find the numerical values of the corresponding integrals beyond the respective domain of convergences. We use our packages to find the numerical value of these functions and compare with `FIESTA5` [77] and the inbuilt `AppellF1` command of MATHEMATICA and MAPLE. The comparisons are reported in Tables 5 and 6.

7.2 Photon-photon scattering

We now consider another physical example where the Appell F_3 function appears, the case of photon-photon scattering [82]. The loop amplitude for the four-point function can be written as follows

$$A(s,t;m) = \frac{i\pi^2}{6m^4} F_3\left(1,1,1,1;\frac{5}{2}\Big|\frac{s}{4m^2},\frac{t}{4m^2}\right) \tag{48}$$

where s and t are the Mandelstam variables and m is the mass of the propagators.

The right hand side of Eq. (48), is further given by the following reduction formula [82]

$$\begin{aligned} &F_3(1,1,1,1;5/2 \mid x,y) \\ &\quad = \frac{3}{4xy\beta_{xy}} \left\{ 2\ln^2\left(\frac{\beta_{xy}+\beta_x}{\beta_{xy}+\beta_y}\right) + \ln\left(\frac{\beta_{xy}-\beta_x}{\beta_{xy}+\beta_x}\right)\ln\left(\frac{\beta_{xy}-\beta_y}{\beta_{xy}+\beta_y}\right) - \frac{\pi^2}{2} \right. \end{aligned}$$

$\{m_1^2, m_2^2, m_3^2\}$	FIESTA5	Eq. (46)	AppellF1.wl	MATHEMATICA	MAPLE
{1,4,9}	0.124697034815251 $\pm 7.55843197 \times 10^{-7}$	0.124697033602012	0.124697033602012	0.124697033602012	0.124697033602013
{9,4,1}	0.124697034815251 $\pm 7.558431971 \times 10^{-7}$	0.124697033602012	0.124697033602012	0.124697033602012	0.124697033602012
{0,4,9}	0.16218604696104502 $\pm 8.34593085 \times 10^{-7}$	0.162186043243266	0.162186043243266	0.162186043243266	0.162186043243266
$\{\frac{1}{25}, \frac{49}{64}, \frac{1}{169}\}$	3.662930006248521 ± 0.000023800445347	3.66292995929945	3.66292995929945	3.66292995929945	3.66292995929945

Table 5: In this Table, we evaluate the C_0 function in Eq. (44) with four different sets of values of masses using our package `AppellF1.wl` (with `N` = 250) and compare the outputs against the inbuilt command `AppellF1` of MATHEMATICA and MAPLE, reduction formula of Appell F_1 (Eq. (46)) and the results from `FIESTA5`. Note that, to be consistent with the definition of Feynman integrals of `FIESTA5`, we multiply a factor of $-\frac{1}{i\pi^2}$ with Eq. (44).

$m_1^2, m_2^2, m_3^2, m_4^2$	FIESTA5	Eq. (47)	LauricellaFD.wl
{1,4,9,16}	0.004611058631625 $\pm 3.6492129 \times 10^{-8}$	0.00461105850390750	0.00461105850390750
{9,16,1,4}	0.004611058631625 $\pm 3.6492129 \times 10^{-8}$	0.00461105850390750	0.00461105850390750 $+5.150233 \times 10^{-11} i$
{0,4,9,16}	0.006665930571901001 $\pm 3.0497681 \times 10^{-8}$	0.00666593044999165	0.00666593044999165
$\{\frac{1}{100}, \frac{9}{49}, 9, \frac{16}{9}\}$	0.126545889960494 $\pm 2.108113451 \times 10^{-6}$	0.126545880468854	0.126545880468854

Table 6: In this Table, we evaluate the D_0 function in Eq. (45) with four different sets of values of masses using our package `LauricellaFD.wl` (with `N` = 300) and compare the outputs against reduction formula of Lauricella $F_D^{(3)}$ (Eq. (47)) and the results from `FIESTA5`. Note that, to be consistent with the definition of Feynman integrals of `FIESTA5`, we multiply a factor of $\frac{1}{i\pi^2}$ with Eq. (45).

$$+ \sum_{i=x,y} \left[2\,\mathrm{Li}_2\left(\frac{\beta_i - 1}{\beta_{xy} + \beta_i}\right) - 2\,\mathrm{Li}_2\left(-\frac{\beta_{xy} - \beta_i}{\beta_i + 1}\right) - \ln^2\left(\frac{\beta_i + 1}{\beta_{xy} + \beta_i}\right) \right] \Bigg\}, \tag{49}$$

where $\beta_x \equiv \sqrt{1 - \frac{1}{x}}$, $\beta_y \equiv \sqrt{1 - \frac{1}{y}}$, $\beta_{xy} \equiv \sqrt{1 - \frac{1}{x} - \frac{1}{y}}$..

{s,t,m}	FIESTA5	Eq. (49)	AppellF3.wl	MATHEMATICA	MAPLE
{1,1,1}	0.207247806735362 $\pm 2.069729919 \times 10^{-6}$	0.207247795188039	0.207247795188039	0.207247795188039	0.207247795188039
{10,2,1}	0.09094036900538302 $+0.307804156595376i$ $\pm(0.000042877912657$ $+0.0000374042950476i)$	0.090940335591268 $-0.307803818883125i$	0.090940335591268 $-0.307803818883124i$	*	0.0909403355912683 $-0.307803818883126i$
{5,10,1}	0.37601608623542404 $+1.029892882655468i$ $\pm(0.000125266562509$ $+0.000105570122816i)$	0.376014441580313 $-1.029891629753789i$	0.376014441580316 $-1.029891629753789i$	*	0.376014441582954 $-1.029891629756351i$
{11,10,1}	-0.5345491671227911 $-0.339676045083691i$ $\pm(0.000039238248217$ $+0.000033382979669i)$	-0.534549424240674 $+0.339676241448472i$	-0.534549424240674 $+0.339676241448473i$	*	-0.534549424241419 $+0.339676241449579i$

Table 7: The results from the reduction formula (Eq. (49)) are obtained with a small negative imaginary part in order to avoid the evaluation on the branch cuts. The numbers in the 4th column are produced using the package `AppellF3.wl` with 20 precision and with `N` = 200. The '*' in the 5th column denotes that MATHEMATICA is unable to give any output.

In Table 7, we compare the values obtained using the reduction formula of F_3, (i.e., Eq. (49)), the package `AppellF3.wl` and using the inbuilt `AppellF3` commands of MATHEMATICA and MAPLE for four kinematic points. Also, the result of the Feynman integral using `FIESTA5` are provided in

the second column. Note that, there is a mismatch in the sign of the imaginary parts of the numbers from the `FIESTA5` column with the corresponding numbers from the other columns. This is due to the different conventions used to evaluate the functions on the branch cut. From the Feynman's $-i\varepsilon$, we observe that, the amplitude Eq. (48) has to be evaluated as

$$A(s,t;m) = \frac{i\pi^2}{6m^4} F_3\left(1,1,1,1;\frac{5}{2}\middle|\frac{s}{4m^2-i\varepsilon},\frac{t}{4m^2-i\varepsilon}\right) \tag{50}$$

$$= \frac{i\pi^2}{6m^4} F_3\left(1,1,1,1;\frac{5}{2}\middle|\frac{s}{4m^2}+i\varepsilon,\frac{t}{4m^2}+i\varepsilon\right) \tag{51}$$

whereas, in the mathematics literature and in the implementation of the package `AppellF3.wl`, on the branch cut, the Appell F_3 is evaluated as

$$\lim_{\varepsilon\to 0} F_3(a_1,a_2,b_1,b_2;c|x-i\varepsilon,y-i\varepsilon) \tag{52}$$

which is the reason behind the mismatch of the sign of the imaginary part of the numbers from the second column to the others in Table 7. We have also confirmed this observation by evaluating the reduction formula Eq. (49) and the derived analytic continuations using the $+i\varepsilon$ convention and found it to be consistent with `FIESTA5`.

7.3 Lauricella Saran $F_S^{(3)}$

As a simple test of our package, we now consider a reduction formula taken from [80],

$$F_S(1,1,1,1,1;2\mid x,y,z) = -\frac{x\log(1-x)}{(x(y-1)-y)(x(z-1)-z)} + \frac{y\log(1-y)}{(x(y-1)-y)(y-z)} - \frac{z\log(1-z)}{(x(z-1)-z)(y-z)} \tag{53}$$

The comparison is reported in Table 8.

(x,y,z)	Eq. (53)	`LauricellaSaranFS.wl`
-3.400643617 0.7907915243 2.852959631	-0.7166971853 $-0.4747886477i$	-0.7166971853 $-0.4747886477ii$
2.863643067 3.310633871 4.187269367	0.0550038448 $-1.1016637066i$	0.0550038448 $-1.1016637066i$
3.217878146 1.913011642 3.741614131	0.006843860 $-3.882712970i$	0.006843860 $-3.882712970i$
4.149336056 -0.2628593004 1.966914729	0.447598037 $+2.635667205i$	0.447598037 $+2.635667205i$
1.410789108 4.730065747 0.7786036673	0.52671335 $+14.69790065I$	0.52671335 $+14.69790065i$

Table 8: Table of comparison of values of $F_S^{(3)}(1,1,1,1,1;2|x,y,z)$ obtained using `LauricellaSaranFS.wl` (with `N` = 200) are compared with the value obtained using reduction formula Eq.(53).

7.4 A note on evaluation with non-generic parameters

In principle, the analytic continuations derived using the method of Olsson are only valid for generic values of Pochhammer parameters. In other words, we tacitly assume that, the parameters do not take

such values which results either in the termination of the series or yielding divergent results. However, these analytic continuations can also be used for numerical evaluation of functions with non-generic values of parameters as we show below.

For the sake of demonstration, let us take the example of Lauricella $F_D^{(3)}(2,1,1,1;4;x,y,z)$ with the value of arguments corresponding to the second row of Table 6.

$$(x,y,z)=\left(1-\frac{m_1^2}{m_4^2},1-\frac{m_2^2}{m_4^2},1-\frac{m_3^2}{m_4^2}\right)=\left(-\frac{5}{4},-3,\frac{3}{4}\right)$$

We also consider one particular analytic continuation (say, # 92) which is valid at our point of interest and given by

$$\begin{aligned}
\text{FD3\#92} &= (1-x)^{-b_1}(1-y)^{-b_2}(1-z)^{-b_3}\left(-\frac{z-1}{z}\right)^{b_3}\frac{\Gamma(c)\Gamma\left(a-b_2\right)\Gamma\left(-a+c-b_3\right)}{\Gamma(a)\Gamma(c-a)\Gamma\left(c-b_2-b_3\right)}\\
&\times\sum_{m,n,p=0}^{\infty}\frac{(b_1)_m\,(b_2)_n\,(b_3)_p\,(-a+c-b_3)_{m+n-p}}{m!n!p!\,(-a+b_2+1)_n\,(c-b_2-b_3)_{m-p}}\left(\frac{x}{x-1}\right)^m\left(\frac{1}{1-y}\right)^n\left(\frac{z-1}{z}\right)^p\\
&+(1-x)^{-b_1}(1-y)^{-b_2}(1-z)^{-b_3}\Gamma(c)\left(-\frac{z-1}{z}\right)^{c-a}\frac{\Gamma\left(a-c+b_3\right)}{\Gamma(a)\Gamma\left(b_3\right)}\\
&\times\sum_{m,n,p=0}^{\infty}\frac{(b_1)_m\,(b_2)_n\,(-a+b_2+1)_{n+p}\,(c-a)_{m+n+p}}{m!n!p!\,(-a+b_2+1)_n\,(-a+c-b_3+1)_{m+n+p}}\left(\frac{x(z-1)}{(x-1)z}\right)^m\left(\frac{z-1}{(y-1)z}\right)^n\left(\frac{z-1}{z}\right)^p\\
&+\frac{(1-y)^{-a}(1-x)^{-b_1}(1-z)^{-b_3}\left(-\frac{z-1}{z}\right)^{b_3}\Gamma(c)\Gamma\left(b_2-a\right)}{\Gamma\left(b_2\right)\Gamma(c-a)}\\
&\times\sum_{m,n,p=0}^{\infty}\frac{(a)_n\,(b_1)_m\,(b_3)_p\,(c-b_2-b_3)_{m+n-p}}{m!n!p!\,(a-b_2+1)_n\,(c-b_2-b_3)_{m-p}}\left(\frac{x}{x-1}\right)^m\left(\frac{1}{1-y}\right)^n\left(\frac{z-1}{z}\right)^p\\
&=\sum_{i=1,2,3}\text{prefactor}_i\times\text{series}_i
\end{aligned}$$

We observe that, for the particular values of parameters

$$a=2,b_1=1,b_2=1,b_3=1,c=4 \tag{54}$$

the prefactor_2 and prefactor_3 are divergent and prefactor_1 is finite. Indeed, by setting

$$a=2,b_1=1,b_2=1+\varepsilon,b_3=1,c=4-\varepsilon \tag{55}$$

we find that,

$$\text{prefactor}_2=\frac{1}{\varepsilon}\frac{6(z-1)}{(x-1)(y-1)z^2}+O(\varepsilon^0)\quad\text{and}\quad\text{series}_2=O(\varepsilon^0) \tag{56}$$

$$\text{prefactor}_3=\frac{1}{\varepsilon}\frac{6}{(x-1)(y-1)^2z}+O(\varepsilon^0)\quad\text{and}\quad\text{series}_3=O(\varepsilon^0) \tag{57}$$

From the definition, one expects that $F_D^{(3)}(2,1,1,1;4|x,y,z)$ to be finite. Thus,

$$\sum_{i=2,3}\text{prefactor}_i\times\text{series}_i=\text{finite} \tag{58}$$

The divergent parts coming from two different terms cancel out eventually. Numerically, to get an accurate cancellation, the series have to be summed over large number of terms `N`.

Inside the packages, when dealing with non-generic parameter values, the analytic continuations are evaluated with a shift of $q\varepsilon$ (where q is a rational number and ε being a very small imaginary number), to the given values of parameters. In such cases, to get more accurate result, the users are advised to take the summation with higher number of terms. The impact of the variation of the number of terms in the summation on the evaluation of $F_D^{(3)}(2,1,1,1;4|x,y,z)$ is elucidated in Table 9, showcasing how the result is contingent upon the number of terms employed for summation.

N	Results from LauricellaFD.wl	Expected value from Eq. (47)
10	$0.40027651208478822796 + 8.4411689391473037971 \times 10^{18} i$	
50	$0.44265212069383465865 + 1.4798709714436027997 \times 10^{15} i$	
100	$0.44266161616971901410 + 2.9853737369244941467 \times 10^{10} i$	
150	$0.44266161637511554096 + 602245.50120234373334 i$	
200	$0.44266161637511983785 + 12.149220689939886743 i$	
250	$0.44266161637511983794 + 0.00024508869402624438466 i$	
300	$0.44266161637511983794 + 4.9442239525071142523 \times 10^{-9} i$	
350	0.44266161637511983794	
400	0.44266161637511983794	0.44266161637511983794

Table 9: The variation of the results with varying numbers of terms employed for summation is shown. The command `FD3evaluate[92,{2,1,1,1,4,-(5/4),-3,3/4},20, N]` is used to generate the values of $F_D^{(3)}$ using the analytic continuation numbered #92 with 20 precision and shown in the second column. The expected value in the last column is obtained using the reduction formula Eq. (47).

8 Summary & Outlook

A detailed investigation of various two and three-variable hypergeometric series, viz. Appell F_1 and F_3, Lauricella $F_D^{(3)}$ and Lauricella-Saran $F_S^{(3)}$, is carried out in this work, and their analytic continuations are obtained using the method of Olsson. Four packages, named `AppellF1.wl`, `AppellF3.wl`, `LauricellaFD.wl` and `LauricellaSaranFS.wl`, are developed in MATHEMATICA for efficient evaluations using all of these analytic continuations. Techniques for fast numerical evaluation of the finite summation of series are also employed. The time and the accuracy obtained using these techniques are contrasted with the usual 'brute force' summation, and we show that these techniques result in significant improvement in both time and accuracy. The packages are built in a way that the evaluation can be carried out for both generic and non-generic cases. A comparison of the results using our packages with publicly available implementation in MATHEMATICA and MAPLE is also attempted, wherever possible. We find that our packages offer better reliability than these implementations. A variety of numerical checks are also performed to check the consistency of the package for both generic and non-generic cases. These non-generic cases have further physical applications in Feynman integral calculus, which are discussed in detail.

In the present version, the packages are applicable to the real values of their arguments. However, the analytic continuations derived using the method of Olsson are valid for complex valued arguments as well. We plan to study in this direction in the near future and build programs that can evaluate these functions for complex-valued arguments.

Acknowledgements

We thank Prof. B. Ananthanarayan for reading the manuscript carefully and providing valuable suggestions and useful comments. We also thank Sudeepan Datta for technical help with some cross-checks. SB extends sincere gratitude to Prof. Thomas Gehrmann for the hospitality at University of Zurich, where a part of this work was carried out. SB thanks Prof. Narayan Rana and NISER Bhubaneshwar for their hospitality at the time of completion of this manuscript. TP is grateful to the long-term workshop YITP-T-23-01 held at YITP, Kyoto University, where a part of this work was done.

References

[1] P. Appell. Sur les séries hypergéométriques de deux variables et sur dés équations différentielles linéaires aux dérivés partielles. *C. R. Acad. Sci., Paris*, 90:296–299, 731–735, 1880.

[2] Giuseppe Lauricella. Sulle funzioni ipergeometriche a piu variabili. *Rendiconti del Circolo Matematico di Palermo*, 7:111–158, 1893.

[3] Shanti Saran. Transformations of certain hypergeometric functions of three variables. *Acta Mathematica*, 93:293–312, 1955.

[4] Olsson, Per O. M. . Integration of the Partial Differential Equations for the Hypergeometric Functions F_1 and F_D of Two and More Variables. *Journal of Mathematical Physics*, 5(3):420–430, 1964.

[5] S. I. Bezrodnykh. Analytic continuation of the Appell function F_1 and integration of the associated system of equations in the logarithmic case. *Computational Mathematics and Mathematical Physics*, 57:559–589, 2017.

[6] Olsson, P. O. M. . On the integration of the differential equations of five-parametric double-hypergeometric functions of second order. *Journal of Mathematical Physics*, 18(6):1285–1294, 1977.

[7] B. Ananthanarayan, Souvik Bera, S. Friot, O. Marichev, and Tanay Pathak. On the evaluation of the Appell F_2 double hypergeometric function. *Comput. Phys. Commun.*, 284:108589, 2023.

[8] H Exton. On the system of partial differential equations associated with Appell's function F_4. *Journal of Physics A: Mathematical and General*, 28(3):631–641, feb 1995.

[9] Markus Huber. Infrared behavior of vertex functions in d-dimensional Yang-Mills theory. Other thesis, 2007.

[10] B. Ananthanarayan, Samuel Friot, Shayan Ghosh, and Anthony Hurier. New analytic continuations for the Appell F_4 series from quadratic transformations of the Gauss ${}_2F_1$ function. 5 2020.

[11] Souvik Bera and Tanay Pathak. Analytic continuations of the Horn H_1 and H_5 functions. *The European Physical Journal Special Topics*, pages 1–24, 2023.

[12] S. I. Bezrodnykh. Analytic continuation of the Kampé de Fériet function and the general double Horn series. *Integral Transforms and Special Functions*, 33(11):908–928, 2022.

[13] H Exton. *Multiple hypergeometric functions and applications*. 1976.

[14] S.I. Bezrodnykh. On the Analytic Continuation of the Lauricella Function $F_D^{(N)}$. *Matematicheskie Zametki*, 100(2):296–302, 2016.

[15] S. I. Bezrodnykh. Analytic continuation of the Lauricella function $F_D^{(N)}$ with arbitrary number of variables. *Integral Transforms and Special Functions*, 29(1):21–42, 2018.

[16] S.I. Bezrodnykh. Analytic continuation of lauricella's function $F_D^{(N)}$ for large in modulo variables near hyperplanes $\{z_j = z_l\}$. *Integral Transforms and Special Functions*, 33(4):276–291, 2022.

[17] S. Friot and G. Suchet-Bernard. Linear transformations of Srivastava's H_C triple hypergeometric function. 5 2022.

[18] S.I. Bezrodnykh. Formulas for analytic continuation of horn functions of two variables. *Computational Mathematics and Mathematical Physics*, 62(6):884–903, 2022.

[19] S.I. Bezrodnykh. Analytic continuation of the Kampé de Fériet function and the general double Horn series. *Integral Transforms and Special Functions*, 33(11):908–928, 2022.

[20] S.I. Bezrodnykh. Analytic continuation of the Horn Hypergeometric series with an arbitrary number of variables. *Integral Transforms and Special Functions*, 31(10):788–803, 2020.

[21] LJ Slater. *Generalized Hypergeometric Functions*. 1966.

[22] H. Bateman and A. Erdélyi. *Higher Transcendental Functions*. Number v. 1 in Higher Transcendental Functions. Dover Publications, 2006.

[23] H. M. Srivastava and P. W. Karlsson. *Multiple gaussian hypergeometric series*. 1985.

[24] Wilfrid Norman Bailey. *Generalized hypergeometric series*. 1935.

[25] Michael J Schlosser. Multiple hypergeometric series: Appell series and beyond. *Computer algebra in quantum field theory: integration, summation and special functions*, pages 305–324, 2013.

[26] Izrail Moiseevich Gel'fand, Mark Iosifovich Graev, and Vladimir Solomonovich Retakh. General Hypergeometric systems of equations and series of Hypergeometric type. *Russian Mathematical Surveys*, 47(4):1, 1992.

[27] PE Newstead. Discriminants, Resultants and Multidimensional Determinants, by I.M. Gelfand, M.M. Kapranov and A. Zelevinsky.(Birkhäuser). *The Mathematical Gazette*, 79(485):439–440, 1995.

[28] Izrail Moiseevich Gel'fand, Mark Iosifovich Graev, and Andrei Vladlenovich Zelevinskii. Holonomic systems of equations and series of Hypergeometric type. In *Doklady Akademii Nauk*, volume 295, pages 14–19. Russian Academy of Sciences, 1987.

[29] I.M. Gel'fand, M.M. Kapranov, and AV Zelevinsky. Hypergeometric functions, toric varieties and newton polyhedra. In *ICM-90 Satellite Conference Proceedings: Special Functions*, pages 104–121. Springer, 1991.

[30] Izrail Moiseevich Gelfand, Mikhail M Kapranov, and Andrei V Zelevinsky. Generalized Euler integrals and A-hypergeometric functions. *Advances in Mathematics*, 84(2):255–271, 1990.

[31] Leonardo de la Cruz. Feynman integrals as A-hypergeometric functions. *JHEP*, 12:123, 2019.

[32] René Pascal Klausen. Hypergeometric Series Representations of Feynman Integrals by GKZ Hypergeometric Systems. *JHEP*, 04:121, 2020.

[33] Tai-Fu Feng, Chao-Hsi Chang, Jian-Bin Chen, and Hai-Bin Zhang. GKZ-hypergeometric systems for Feynman integrals. *Nucl. Phys. B*, 953:114952, 2020.

[34] René Pascal Klausen. Kinematic singularities of Feynman integrals and principal A-determinants. *JHEP*, 02:004, 2022.

[35] Tai-Fu Feng, Hai-Bin Zhang, and Chao-Hsi Chang. Feynman integrals of Grassmannians. *Phys. Rev. D*, 106(11):116025, 2022.

[36] Daniele Agostini, Claudia Fevola, Anna-Laura Sattelberger, and Simon Telen. Vector Spaces of Generalized Euler Integrals. 8 2022.

[37] Vsevolod Chestnov, Federico Gasparotto, Manoj K. Mandal, Pierpaolo Mastrolia, Saiei J. Matsubara-Heo, Henrik J. Munch, and Nobuki Takayama. Macaulay matrix for Feynman integrals: linear relations and intersection numbers. *JHEP*, 09:187, 2022.

[38] Tai-Fu Feng, Hai-Bin Zhang, Yan-Qing Dong, and Yang Zhou. GKZ-system of the 2-loop self energy with 4 propagators. *Eur. Phys. J. C*, 83(4):314, 2023.

[39] Hai-Bin Zhang and Tai-Fu Feng. GKZ hypergeometric systems of the three-loop vacuum Feynman integrals. *JHEP*, 05:075, 2023.

[40] Vsevolod Chestnov, Saiei J. Matsubara-Heo, Henrik J. Munch, and Nobuki Takayama. Restrictions of Pfaffian systems for Feynman integrals. *JHEP*, 11:202, 2023.

[41] Saiei-Jaeyeong Matsubara-Heo, Sebastian Mizera, and Simon Telen. Four lectures on Euler integrals. *SciPost Phys. Lect. Notes*, 75:1, 2023.

[42] B. Ananthanarayan, Sumit Banik, Souvik Bera, and Sudeepan Datta. FeynGKZ: A Mathematica package for solving Feynman integrals using GKZ hypergeometric systems. *Comput. Phys. Commun.*, 287:108699, 2023.

[43] Andrei I. Davydychev. General results for massive $N-$ point Feynman diagrams with different masses. *J. Math. Phys.*, 33:358–369, 1992.

[44] Andrei I. Davydychev. Some exact results for $N-$point massive Feynman integrals. *J. Math. Phys.*, 32:1052–1060, 1991.

[45] O. V. Tarasov. Application and explicit solution of recurrence relations with respect to space-time dimension. *Nucl. Phys. B Proc. Suppl.*, 89:237–245, 2000.

[46] J. Fleischer, F. Jegerlehner, and O. V. Tarasov. A New hypergeometric representation of one loop scalar integrals in d dimensions. *Nucl. Phys. B*, 672:303–328, 2003.

[47] Khiem Hong Phan and Tord Riemann. Scalar 1-loop Feynman integrals as meromorphic functions in space-time dimension d. *Phys. Lett. B*, 791:257–264, 2019.

[48] Vittorio Del Duca, Claude Duhr, E. W. Nigel Glover, and Vladimir A. Smirnov. The One-loop pentagon to higher orders in epsilon. *JHEP*, 01:042, 2010.

[49] Frits A. Berends, M. Buza, M. Bohm, and R. Scharf. Closed expressions for specific massive multiloop selfenergy integrals. *Z. Phys. C*, 63:227–234, 1994.

[50] B. Ananthanarayan, Samuel Friot, and Shayan Ghosh. New Series Representations for the Two-Loop Massive Sunset Diagram. *Eur. Phys. J. C*, 80(7):606, 2020.

[51] B. Ananthanarayan, Samuel Friot, Shayan Ghosh, and Anthony Hurier. New analytic continuations for the Appell F_4 series from quadratic transformations of the Gauss ${}_2F_1$ function. 5 2020.

[52] O. V. Tarasov. Hypergeometric representation of the two-loop equal mass sunrise diagram. *Phys. Lett. B*, 638:195–201, 2006.

[53] Tai-Fu Feng, Chao-Hsi Chang, Jian-Bin Chen, and Hai-Bin Zhang. The system of partial differential equations for the C_0 function. *Nucl. Phys. B*, 940:130–189, 2019.

[54] Zhi-Hua Gu and Hai-Bin Zhang. Three-loop vacuum integral with four-propagators using hypergeometry. *Chin. Phys. C*, 43(8):083102, 2019.

[55] Zhi-Hua Gu, Hai-Bin Zhang, and Tai-Fu Feng. Hypergeometric expression for a three-loop vacuum integral. *Int. J. Mod. Phys. A*, 35(19):2050089, 2020.

[56] Tai-Fu Feng, Chao-Hsi Chang, Jian-Bin Chen, Zhi-Hua Gu, and Hai-Bin Zhang. Evaluating Feynman integrals by the hypergeometry. *Nucl. Phys. B*, 927:516–549, 2018.

[57] J. Fleischer, F. Jegerlehner, and O. V. Tarasov. Algebraic reduction of one loop Feynman graph amplitudes. *Nucl. Phys. B*, 566:423–440, 2000.

[58] O. V. Tarasov. Derivation of Functional Equations for Feynman Integrals from Algebraic Relations. *JHEP*, 11:038, 2017.

[59] Bernd A. Kniehl and Oleg V. Tarasov. Finding new relationships between hypergeometric functions by evaluating Feynman integrals. *Nucl. Phys. B*, 854:841–852, 2012.

[60] O. V. Tarasov. Functional reduction of one-loop Feynman integrals with arbitrary masses. *JHEP*, 06:155, 2022.

[61] J. Blümlein, M. Saragnese, and C. Schneider. Hypergeometric Structures in Feynman Integrals. 11 2021.

[62] B. Ananthanarayan, Souvik Bera, and Tanay Pathak. AlgRel.wl: Algebraic relations for the product of propagators in Feynman integrals. *Nucl. Phys. B*, 995:116345, 2023.

[63] M. A. Shpot. A Massive Feynman integral and some reduction relations for Appell functions. *J. Math. Phys.*, 48:123512, 2007.

[64] Wei Fan. Celestial conformal blocks of massless scalars and analytic continuation of the Appell function F_1. *JHEP*, 01:145, 2024.

[65] Nikolas Akerblom and Michael Flohr. Explicit formulas for the scalar modes in Seiberg-Witten theory with an application to the Argyres-Douglas point. *JHEP*, 02:057, 2005.

[66] Isachenkov, Mikhail. Conformal field theory and functions of hypergeometric type. Technical report, Deutsches Elektronen-Synchrotron (DESY), 2016.

[67] F. D. Colavecchia, G. Gasaneo, and J. E. Miraglia. Numerical evaluation of Appell's F_1 hypergeometric function. *Computer Physics Communications*, 138(1):29–43, 2001.

[68] F. D. Colavecchia and G. Gasaneo. F1: A code to compute Appell's F_1 hypergeometric function. *Computer Physics Communications*, 157(1):32–38, 2004.

[69] B. Ananthanarayan, Souvik Bera, S. Friot, and Tanay Pathak. Olsson.wl : a *Mathematica* package for the computation of linear transformations of multivariable hypergeometric functions. 12 2021.

[70] Maplesoft a division of Waterloo Maple Inc. Maple 2022. Software, 2022.

[71] Wolfram Research, Inc. Mathematica, Version 13.3. Champaign, IL, 2023.

[72] A Erdélyi. The analytic theory of systems of partial differential equations. *Bulletin of the American Mathematical Society*, 57(5):339–353, 1951.

[73] Vladimir V. Bytev, Mikhail Yu. Kalmykov, and Sven-Olaf Moch. HYPERgeometric functions DIfferential REduction (HYPERDIRE): MATHEMATICA based packages for differential reduction of generalized hypergeometric functions: F_D and F_S Horn-type hypergeometric functions of three variables. *Comput. Phys. Commun.*, 185:3041–3058, 2014.

[74] A. Erdélyi. Xxxix.—Transformations of hypergeometric functions of two variables. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, 62(3):378–385, 1948.

[75] P. J. Appell and Kampé de Fériet. Fonctions Hypergéométriques et Hypersphériques. *Polynômes d'Hermite*, 1926.

[76] George B Arfken, Hans J Weber, and Frank E Harris. *Mathematical methods for physicists: a comprehensive guide*. Academic press, 2011.

[77] A. V. Smirnov, N. D. Shapurov, and L. I. Vysotsky. FIESTA5: Numerical high-performance Feynman integral evaluation. *Comput. Phys. Commun.*, 277:108386, 2022.

[78] S. Moch and P. Uwer. XSummer: Transcendental functions and symbolic summation in form. *Comput. Phys. Commun.*, 174:759–770, 2006.

[79] Andrew J. McLeod, Henrik J. Munch, Georgios Papathanasiou, and Matt von Hippel. A Novel Algorithm for Nested Summation and Hypergeometric Expansions. *JHEP*, 11:122, 2020.

[80] Souvik Bera. MultiHypExp: A Mathematica package for expanding multivariate hypergeometric functions in terms of multiple polylogarithms. *Comput. Phys. Commun.*, 297:109060, 2024.

[81] M. A. Bezuglov and A. I. Onishchenko. Expansion of hypergeometric functions in terms of polylogarithms with nontrivial variable change. 12 2023.

[82] Andrei I. Davydychev. Standard and hypergeometric representations for loop diagrams and the photon-photon scattering, 5 1993.

Perturbative QCD meets phase quenching: The pressure of cold Quark Matter

Pablo Navarrete,[1,*] Risto Paatelainen,[1,†] and Kaapo Seppänen[1,‡]

[1] *Department of Physics and Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*

Nonperturbative inequalities constrain the pressure of Quantum Chromodynamics (QCD) with its phase-quenched version, a Sign-Problem-free theory amenable to lattice treatment. In the perturbative regime with a small QCD coupling constant α_s, one of these inequalities manifests as an $O(\alpha_s^3)$ difference between the phase-quenched and QCD pressures at large baryon chemical potential μ_B. In this work, we generalize state-of-the-art algorithmic techniques in vacuum quantum field theory to address large-scale multiloop computations at finite chemical potential. Using this approach, we evaluate this $O(\alpha_s^3)$ difference and show that it is a gauge-independent and small positive number compared to the known perturbative coefficients at this order. This implies that phase-quenched lattice simulations can provide a complementary nonperturbative method for accurately determining the pressure of cold and dense quark matter at $O(\alpha_s^3)$.

arXiv:2403.02180v1 [hep-ph] 4 Mar 2024

Introduction.—Determining the bulk thermodynamic properties of strongly interacting matter at the highest temperatures and densities reached in Nature is one of the major challenges in theoretical physics. At high temperatures and small baryon densities, experiments conducted with ultrarelativistic heavy-ion collisions (see e.g. [1] for a review) and theoretical work using nonperturbative lattice simulations [2, 3] have been crucial in understanding the equilibrium properties of Quantum Chromodynamic (QCD) matter. Simultaneously to these developments, the past decade has witnessed rapid progress in the study of neutron stars (NSs). Due to the tight link between NS-matter microphysics and the macroscopic properties of the stars, NSs have become unique cosmic laboratories for studying dense QCD matter (see e.g. [4] for a review). What makes them particularly fascinating in the context of strongly interacting physics is that, unlike any other accessible system in the Universe, the maximal densities reached inside the cores of massive NSs may exceed the limit where nuclear matter undergoes a transition into deconfined cold quark matter (QM) [5, 6].

To describe the thermodynamic properties of dense QCD matter in equilibrium, one needs information about the equation of state (EOS), which relates the pressure p to energy density. Ideally, the EOS should be determined using nonperturbative lattice field theory tools that have provided accurate results in collider environments. Unfortunately, the standard Monte Carlo (MC) techniques fail at finite baryon density due to the Sign Problem [7–11], which prevents lattice QCD simulations in the region of low temperature T and large baryon chemical potential μ_B. Consequently, the only practical methods available to reliably describe the EOS of cold and dense QCD matter are limited to Chiral Effective Theory, applicable at densities around the nuclear matter saturation density [12–15], and weak-coupling perturbative QCD within high-density QM [16–22]. Several studies [5, 6, 23–28] have shown that systematic interpolations between low- and high-density regions, combined with information from astrophysical NS observations [29–34], can effectively constrain the behavior of the EOS across all densities, provided that accurate asymptotic limits are established at both ends.

Additionally, further bounds for the dense QCD matter EOS can be obtained using phase-quenched (PQ) lattice QCD [35]. In the phase-quenched approximation, the fermion determinant in the QCD partition function for a quark with flavor f, nonzero mass m_f, and chemical potential μ_f is replaced by its magnitude

$$\begin{aligned}\det(\mathcal{D}(\mu_f)) &\to |\det(\mathcal{D}(\mu_f))| \\ &= \sqrt{\det(\mathcal{D}(\mu_f))\det(\mathcal{D}(-\mu_f))},\end{aligned} \tag{1}$$

where the Dirac operator $\mathcal{D}(\mu_f) \equiv \not{D} + m_f + \mu_f\gamma^0$, and the last equality follows from its γ^5-hermiticity. Here, each phase-quenched fermion corresponds to two copies of "half-species", appearing in pairs with equal mass but opposite chemical potential. Although this formulation is unphysical, it can constrain the physically more interesting scenarios involving finite μ_B.

For example, consider QCD with two degenerate flavors of light up and down quarks, and a baryon chemical potential $\mu_u = \mu_d = \mu_\mathrm{B}/3$. In such a scenario, it follows from Eq. (1) that the phase-quenched partition function is equivalent to the partition function of QCD with a nonzero isospin chemical potential μ_I, where $\mu_u = -\mu_d = \mu_\mathrm{I}/2$. Using this observation and nonperturbative QCD inequalities [36], Cohen showed in [37] that the phase-quenched pressure of isospin matter p_PQ serves as a strict upper bound on the QCD pressure of baryonic matter p, i.e. $p_\mathrm{PQ}(3\mu_\mathrm{I}/2) \geq p(\mu_\mathrm{B})$.

In NS interiors, the energy density may be sufficiently high to give rise to the onset of strange quarks, either in the form of hyperons or three-flavor QM. Chemical equilibrium under weak interactions (beta-equilibrium) with a nonzero strange quark mass requires $\mu_d = \mu_s$ and $\mu_d > \mu_u$. In this context, Fujimoto and Reddy demonstrated in [38] that a similar bound to the two-flavor case can be derived for the pressure at nonzero μ_B and

$\mu_{\rm I}$, utilizing additional nonperturbative QCD inequalities derived in [39]. This inequality necessitates $\mu_{\rm I}$ as a function of $\mu_{\rm B}$, which remains unknown unless certain model-dependent assumptions are made.

In [40], Moore and Gorda recently noted that the phase-quenched formulation applies to any arbitrary linear combination of chemical potentials, not limited to $\mu_{\rm I}$. They suggested that conducting phase-quenched lattice simulations with three-flavor QM could be the most effective method to constrain the NS EOS. Although this phase-quenched configuration of quark chemical potentials represents a completely unphysical system, it holds the potential to provide the most stringent bounds on the physical NS EOS. In particular, at high densities the relative difference between the phase-quenched pressure and the QCD pressure $p_{\rm PQ} - p$ is perturbatively order α_s^3 in the QCD coupling constant. In any μ-region where this leading $O(\alpha_s^3)$ correction is small, it can be combined with $p_{\rm PQ}$ to provide an improved estimate of p. This estimation would be perturbatively complete up to higher-order corrections in α_s and nonperturbative effects such as pairing gaps [41–44].

In this Letter, we explicitly evaluate this leading perturbative $O(\alpha_s^3)$ correction. We show that this correction, originating from a single four-loop diagram with two fermion loops connected by three gluons, is a gauge-independent and small positive number compared to the known perturbative coefficients at this order [22]. Hence, phase-quenched lattice simulations can offer a complementary nonperturbative method for accurately determining the pressure of cold and dense quark matter at $O(\alpha_s^3)$. Our diagrammatic evaluation is achieved by generalizing state-of-the-art algorithmic techniques in vacuum quantum field theory to tackle large-scale multiloop computations at finite chemical potential. This includes an analytical computation of the fermionic and bosonic momentum integrals for the zero-components, followed by the numerical MC evaluation of the remaining multi-dimensional spatial momentum integrals. This work represents the first-ever computation of a two-particle-irreducible (2PI) four-loop diagram at finite μ, paving the way toward large-scale automatization of multi-loop computations in the context of field theory at finite density.

Setup of the calculation.—In the context of cold and dense QM with N_f quark flavors, the pressure is given in the grand canonical ensemble by $p = -\Omega$, with Euclidean partition function

$$\mathcal{Z}(\{\mu_j\}) = \int \mathcal{D}A\, e^{-S[A]} \prod_{j=1}^{N_f} \det(\mathcal{D}(\mu_j)). \tag{2}$$

The perturbative expansion of the pressure then follows as the sum of all connected vacuum Feynman diagrams. The phase-quenched version of this theory makes use of the replacement in Eq. (1), which can be integrated via

$$\sqrt{\det \mathcal{D}(\pm\mu_j)} = \exp\left\{\frac{1}{2}\mathrm{tr}\,\log \mathcal{D}(\pm\mu_j)\right\}. \tag{3}$$

The corresponding Feynman rule derived from here amounts to averaging over the sign of each of the quarks' chemical potentials. To see the consequences of this averaging, consider a vacuum-type Feynman diagram $\mathcal{G}(\{\mu_j\})$ in QCD with fermion momenta $P = (P^0 + i\mu_j, \vec{p})$ and bosonic momenta $Q = (Q^0, \vec{q})$. Upon the replacement $\mu_j \to -\mu_j$ for all quarks in $\mathcal{G}$ (i.e. take the complex conjugate), one can make use of the integral's symmetry under flipping the signs of all the fermionic and bosonic loop momenta and the Euclidean QCD Feynman rules to obtain

$$\mathcal{G}(\{-\mu_j\}) = (-1)^{v_q+v_3}\mathcal{G}(\{\mu_j\}) = (-1)^{v-v_4}\mathcal{G}(\{\mu_j\}), \tag{4}$$

with v, v_q, v_3, and v_4 the number of total, quark-gluon, three-gluon, and four-gluon vertices, respectively. Now, a vacuum-type graph in QCD can only have an odd number of total vertices if it contains an odd number of vertices of valency 4, i.e. four-gluon interactions (see e.g. [45]). Hence, the exponent $v - v_4$ in Eq. (4) is an even number, leading to $\mathcal{G}(\{-\mu_j\}) = \mathcal{G}(\{\mu_j\})$. This implies two things:

1. Vacuum Feynman diagrams are real numbers,
2. Phase-quenched Feynman rules are equivalent to QCD for diagrams with a single quark loop, vanishing in the difference $p_{PQ}(\{\mu_j\}) - p(\{\mu_j\})$.

Therefore, one expects possible deviations for diagrams containing at least two quark loops.

Let us consider a generic diagram with two quark loops carrying chemical potentials μ_1 and μ_2, written as

$$\mathcal{G}^\xi(\mu_1,\mu_2) = \int_{Q_b} \mathcal{N}^\xi(Q_b)\Gamma^{(n)}(Q_b,\mu_1)\Gamma^{(m)}(Q_b,\mu_2), \tag{5}$$

where Q_b denotes the set of bosonic loop momenta running through the diagram, and $\mathcal{N}^\xi(Q_b)$ contains the color traces and the corresponding bosonic Feynman rules, for general covariant gauges parametrized by ξ; $\Gamma^{(n)}$ is an insertion of a gluon n-point function with one quark loop (suppressing Lorentz and color indices). From the phase-quenched Feynman rules and the reality of vacuum diagrams, it then follows that the contribution of $\mathcal{G}$ to $p_{\rm PQ}(\{\mu_j\}) - p(\{\mu_j\})$ is

$$\begin{aligned} p_{\rm PQ} - p\Big|_{\mathcal{G}} &= -\frac{1}{2}\left[\mathcal{G}^\xi(\mu_1,\mu_2) - \mathcal{G}^\xi(\mu_1,-\mu_2)\right] \\ &= -i\int_{Q_b} \mathcal{N}^\xi(Q_b)\Gamma^{(n)}(Q_b,\mu_1)\mathrm{Im}\,\Gamma^{(m)}(Q_b,\mu_2) \\ &= \int_{Q_b} \mathcal{N}^\xi(Q_b)\mathrm{Im}\,\Gamma^{(n)}(Q_b,\mu_1)\mathrm{Im}\,\Gamma^{(m)}(Q_b,\mu_2). \end{aligned} \tag{6}$$

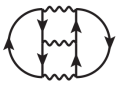
Here, we made use of the fact that bosons do not carry chemical potentials, implying $\mathcal{N}$ is real. One can readily see that the one-loop gluon two-point function, defined as

$$\Gamma^{(2)}_{\alpha\beta}(Q,\mu_j) = \int_{\{P\}_j} \frac{\mathrm{tr}\, \not{P}\gamma_\alpha(\not{P}-\not{Q})\gamma_\beta}{P^2(P-Q)^2}, \tag{7}$$

is a real number, as complex conjugation is equivalent to the shift $P \to Q-P$, which is a symmetry of the integral. Here, $\int_{\{P\}_j} f(P^0,\vec{p}) \equiv \int \frac{\mathrm{d}^4P}{(2\pi)^4} f(P^0+i\mu_j,\vec{p})$ denotes the 4-dimensional fermionic loop momentum integral.

It turns out this argument no longer applies to the three-point function, as this integral has no symmetries of this type. One then expects that the leading order contribution to $p_{\rm PQ}-p$ is a diagram composed of two quark loops connected by three gluon lines, entering at four loops: "the Bugblatter" [46]. In fact, there are two diagrams of this type with differing relative fermion flows, and their explicit expressions using the QCD Feynman rules, up to the terms proportional to ξ, are as follows:

$$= \frac{g_s^6}{6}\mathrm{tr}\,(T^aT^bT^c)\mathrm{tr}\,(T^aT^bT^c)\sum_{i,j=1}^{N_f}\left\{\int_{Q\{P\}_i\{R\}_j^* S} \frac{-\mathrm{tr}\,[\not{P}\gamma^\mu(\not{P}-\not{S})\gamma^\nu(\not{Q}-\not{P})\gamma^\sigma]\,\mathrm{tr}\,[\not{R}\gamma^\mu(\not{R}-\not{S})\gamma^\nu(\not{Q}-\not{R})\gamma^\sigma]}{Q^2P^2R^2S^2(P-S)^2(R-S)^2(Q-P)^2(Q-R)^2(Q-S)^2}\right.$$
$$\left. + \xi\text{ -dependent terms}\right\},$$



$$= \frac{g_s^6}{6}\mathrm{tr}\,(T^aT^bT^c)\mathrm{tr}\,(T^aT^cT^b)\sum_{i,j=1}^{N_f}\left\{\int_{Q\{P\}_i\{R\}_j S} \frac{+\mathrm{tr}\,[\not{P}\gamma^\mu(\not{P}-\not{S})\gamma^\nu(\not{Q}-\not{P})\gamma^\sigma]\,\mathrm{tr}\,[\not{R}\gamma^\sigma(\not{Q}-\not{R})\gamma^\nu(\not{R}-\not{S})\gamma^\mu]}{Q^2P^2R^2S^2(P-S)^2(R-S)^2(Q-P)^2(Q-R)^2(Q-S)^2}\right.$$
$$\left. + \xi\text{ -dependent terms}\right\}. \tag{8}$$

Here, we sum over all N_f flavors running through each quark loop. Denoting the integral expressions inside the curly brackets for each diagram by $B_1^\xi(\mu_i,\mu_j)$ and $B_2^\xi(\mu_i,\mu_j)$, respectively, the Dirac traces can be related through charge conjugation, giving

$$B_2^\xi(\mu_i,\mu_j) = -B_1^\xi(\mu_i,-\mu_j). \tag{9}$$

This is just a manifestation of Furry's theorem, as pointed out in [40].

Moreover, in terms of the SU(N_c) structure constants f^{abc} and the totally symmetric group invariant d^{abc}, the color traces entering Eq. (8) are related through complex conjugation:

$$\mathrm{tr}\,(T^aT^bT^c) = \frac{1}{4}(d^{abc}+if^{abc}) = \mathrm{tr}\,(T^aT^cT^b)^*. \tag{10}$$

With these relations at hand, we proceed to construct $p_{\rm PQ}-p$ as shown above, by averaging over the N_f chemical potentials running in the loops, for both diagrams. As instructed in the NS setup, we specialize to the case of asymptotically dense three-flavor ($N_f=3$) unpaired QM in beta-equilibrium, where quarks can be considered massless and quark chemical potentials are equal. We set $\mu_i = \mu_{\rm B}/3$ for all quarks, obtaining altogether

$$\begin{aligned} p_{\rm PQ}(\mu_{\rm B}) - p(\mu_{\rm B}) = {}& \frac{-3^2}{6\times 3^4}\frac{d^{abc}d^{abc}}{16}g_s^6 \\ & \times\left[B_1^\xi(\mu_{\rm B},\mu_{\rm B}) - B_1^\xi(\mu_{\rm B},-\mu_{\rm B})\right] \\ & + O(g_s^8), \end{aligned} \tag{11}$$

where an additional factor of 3^4 appears through the scaling of B_1^ξ upon replacing μ_i^4 by $\mu_{\rm B}^4$, and $d^{abc}d^{abc} = (N_c^2-1)(N_c^2-4)/N_c$. This expression satisfies several properties, as we demonstrate in the following.

Firstly, expanding the diagrams in general covariant gauge in terms of scalarized integrals, one obtains a quadratic polynomial in the gauge parameter ξ. At this point, one can already appreciate the complexity of these diagrams, as they contain all nine non-factorized planar QCD vacuum topologies at four loops [47]. Upon exploiting linear relations among these integrals, in the form of momentum shifts stemming from the graphs' internal symmetries [48, 49], we explicitly checked that the gauge parameter cancels out at the integrand level in the sum entering Eq. (11), rendering this combination gauge invariant. Consequently, in the following, we set $\xi=0$ and drop the superscript ξ.

Secondly, we can prove that this expression is automatically positive, as expected from the nonperturbative inequality $p_{\rm PQ} \geq p$. In effect, the prefactor in Eq. (11) is nonpositive for all N_c, so it suffices to show that the

difference inside the square brackets is negative, a combination being precisely of the form encountered in Eq. (6). As bosonic propagators are real and positive in the whole region of integration, one has

$$B_1(\mu_{\rm B},\mu_{\rm B}) - B_1(\mu_{\rm B},-\mu_{\rm B}) = -2\int_{QS} \frac{\left[\mathrm{Im}\,\Gamma^{(3)}_{\mu\nu\sigma}(Q,S,\mu_{\rm B})\right]^2}{Q^2S^2(Q-S)^2} \leq 0, \tag{12}$$

proving that the leading-order contribution to $p_{\rm PQ}-p$ is positive.

Thirdly, we can argue that it is infrared finite. Indeed, the infrared physics of this diagram is properly captured by the hard-thermal-loop (HTL) limit of the three-gluon vertex function, whose color structure is $\Gamma^{(3)}_{\rm HTL} \sim f^{abc}$ [20]. As only d^{abc}'s enter Eq. (11), this expression must be infrared finite (see also [40]), further implying that non-analytic logarithms in the coupling arising from the HTL resummation can potentially enter only at order $O(\alpha_s^4 \ln^n \alpha_s)$.

Finally, we argue that Eq. (12) is ultraviolet finite. One might expect this based on the observation that there are no lower loop diagrams with the color factor $d^{abc}d^{abc}$ that could renormalize Eq. (11). Indeed, this combination picks up the imaginary part of the vertex function, a quantity that vanishes in the ultraviolet due to this limit being pure vacuum (i.e. $i\mu \to 0$). Thus, only the $f^{abc}f^{abc}$ piece of the Bugblatter diagrams entering full QCD, containing the real part of the vertex function, requires renormalization.

Numerical evaluation.—Let us then concentrate on the evaluation of the finite four-loop momentum integrals displayed in Eq. (12). In modern quantum field theory, multi-loop diagrams are often first expanded in terms of scalarized integrals, which are then systematically reduced to a basis of so-called master integrals through large-scale automated algorithmic procedures [50, 51]. In this context, techniques for computing master integrals in dimensional regularization are highly developed [52]. However, in a thermal setup (finite μ or T) we break Lorentz symmetry, limiting the usage of the traditional methods that have proven successful in the vacuum setting. Despite this, all the relevant thermal cases up to the three-loop level have been computed on a case-by-case basis, often through brute-force calculation [16, 53]. However, only a few of the simplest four-loop integrals are known [54–56].

Recent developments in high-performance numerical evaluation of vacuum integrals directly in momentum space, such as the Loop-Tree Duality (LTD) (see e.g. [57, 58]), offer a promising alternative avenue to multi-loop integration. The defining feature these techniques capitalize on is the analytical computation of the temporal momentum integrals via the residue theorem. This approach is particularly suitable for finite-density calculations due to the effect of the chemical potentials being entirely encoded in the temporal momentum components.

Inspired by the LTD and related ideas, we put forward a novel generalization of these approaches to finite-density field theory for evaluating our Bugblatter diagrams. Specifically, we implement numerical MC integration routines for the spatial momentum integrals, preceded by an analytical computation of zero-component residues. This is realized by restructuring the integrand into a convenient representation, allowing for the removal of local spurious divergences that hinder direct numerical integration.

In the following, we generalize the derivation of the integral representation presented in [59] to finite μ, employing the notation introduced therein. Let $\mathbf{x} = (x_i)_{i\in\{1,\dots,n\}}$ denote a vector with n components, $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^n x_i y_i$ the standard scalar product of two vectors $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{x}\odot\mathbf{y} = (x_i y_i)_{i\in\{1,\dots,n\}}$ component-wise multiplication of the vectors. Now, a vacuum-type l-loop integral with n single-power propagators at finite μ, stripped of its spatial loop integrations over $\vec{p}_i$, can be written as

$$I \equiv \int \prod_{i=1}^{l} \frac{\mathrm{d}P_i^0}{2\pi} \frac{\mathcal{N}(\mathbf{Q}^0)}{\prod_{j=1}^n \left[\left(Q_j^0\right)^2 + E_j^2\right]}, \tag{13}$$

where the numerator $\mathcal{N}$ is a regular function of the temporal propagator momenta $\mathbf{Q}^0 = \sum_{i=1}^l \mathbf{S}_i P_i^0 + i\boldsymbol{\tau}\mu$, the vectors $\mathbf{S}_i = (S_{1i},\dots,S_{ni})$ with $S_{ji} \in \{\pm 1, 0\}$ fix the loop momentum basis, and the vector $\boldsymbol{\tau} = \sum_{i=1}^l \mathbf{S}_i \tilde{\tau}_i \in \{\pm 1, 0\}^n$ with $\tilde{\tau}_i \in \mathbb{Z}$ determines the fermion signature of each propagator. The propagator energies are given by $E_j^2 = |\vec{q}_j|^2 + m_j^2$, where the spatial loop momenta are $\vec{\mathbf{q}} = \sum_{i=1}^l \mathbf{S}_i \vec{p}_i$ and the masses are denoted by m_j.

Next, we apply the residue theorem to perform the temporal integrals over P_i^0 in Eq. (13), resulting in

$$I = \sum_{\substack{\boldsymbol{\rho}\in\{\pm1\}^n \\ \boldsymbol{\sigma}\in\{\pm1\}^n}} \mathcal{N}(i\boldsymbol{\sigma}\odot\boldsymbol{\rho}\odot\mathbf{E}) \prod_{j=1}^n \frac{\Theta(\rho_j E_j - \sigma_j\tau_j\mu)}{2\rho_j E_j} \times \int_{\mathbb{R}_+^n} \mathrm{d}\boldsymbol{\alpha}\, \mathrm{e}^{-\boldsymbol{\alpha}\cdot(\boldsymbol{\rho}\odot\mathbf{E})} \prod_{i=1}^l \delta\big(\boldsymbol{\alpha}\cdot(\boldsymbol{\sigma}\odot\mathbf{S}_i)\big). \tag{14}$$

In contrast to the vacuum case, this generalization involves step functions that include μ, along with an additional sum over the sign vectors $\boldsymbol{\rho}$. However, the integral on the second line of Eq. (14), representing a Laplace transform of a non-simplicial convex cone, is an object already encountered in the vacuum computation. We perform this integral using the algorithmic procedure described in [59], leading to a sum of rational functions of linear combinations of the energies E_i. The resulting representation for the spatial integral is free of spurious divergences, making it particularly suitable for direct numerical integration.

Nonetheless, the expression resulting from Eq. (14) often contains integrable singularities that may deteriorate the MC integration. To mitigate the issue, we have implemented a multi-channeling technique, as presented in [60]. This procedure flattens some of the integrable singularities by expressing the integrand as a sum of multiple terms (i.e. channels), which are then integrated separately. Additionally, we have found that using small but nonzero masses in the energies E_j improves the stability of the integration. Selecting $m_j = \lambda\mu_{\rm B}$ with $\lambda = 10^{-3}$, we have verified the insensitivity of our results under varying $\lambda \in [10^{-4}, 10^{-2}]$.

Furthermore, it is important to emphasize that the above procedure is not limited to the current finite four-loop diagrams alone; it is equally applicable to diagrams of various topologies and loop orders. For example, we checked that our approach correctly reproduces known results at two- and three-loop levels. These cases include the subtraction of ultraviolet divergences, a topic we will detail in an upcoming publication.

Results and discussion.—We are now prepared to evaluate the new coefficient in Eq. (12) using the integrand representation generated by Eq. (14). With a suitable choice of coordinates, this leads to a ten-dimensional integral over the spatial momenta. We compute this integral numerically employing the VEGAS5.6 integration routine [61], distributing roughly 10^{13} MC samples across 75 channels (see above). We determine the new coefficient to be

$$B_1(\mu_{\rm B},\mu_{\rm B}) - B_1(\mu_{\rm B},-\mu_{\rm B}) = -3.3278(14)\cdot 10^{-6}\mu_{\rm B}^4, \quad (15)$$

where the MC error in the obtained numerical result is small. By utilizing Eq. (11) and setting $N_c = 3$, we then find that the relative difference $p_{\rm PQ} - p$ in the perturbative region is

$$p_{\rm PQ}(\mu_{\rm B}) - p(\mu_{\rm B}) = 3.3681(14)\cdot\left(\frac{\alpha_s(\overline{\Lambda})}{\pi}\right)^3 p_{\rm free}, \quad (16)$$

where $\alpha_s = \alpha_s(\overline{\Lambda})$ represents the renormalized strong coupling constant in the $\overline{\rm MS}$ scheme at the renormalization scale $\overline{\Lambda}$, and we have written our result in terms of the pressure of a free Fermi gas of quarks in beta equilibrium, denoted as $p_{\rm free} = 3(\mu_{\rm B}/3)^4/4\pi^2$.

In Fig. 1, we display the relative difference $p_{\rm PQ} - p$ normalized by the free quark pressure as a function of $\mu_{\rm B}$. We employ the known three-loop running for $\alpha_s(\overline{\Lambda})$, set $\overline{\Lambda} = 2X\mu_{\rm B}/3$, and estimate the uncertainty arising from truncating the perturbative series by varying X in the range $[1/2, 2]$. It is observed that in the perturbative region $\mu_{\rm B} \gtrsim 2.5$ GeV, corresponding to $\alpha_s \lesssim 0.5$, the difference $p_{\rm PQ} - p$ is very small. This implies that phase-quenched lattice simulations provide an accurate approximation for the dense QM pressure in the perturbative regime.

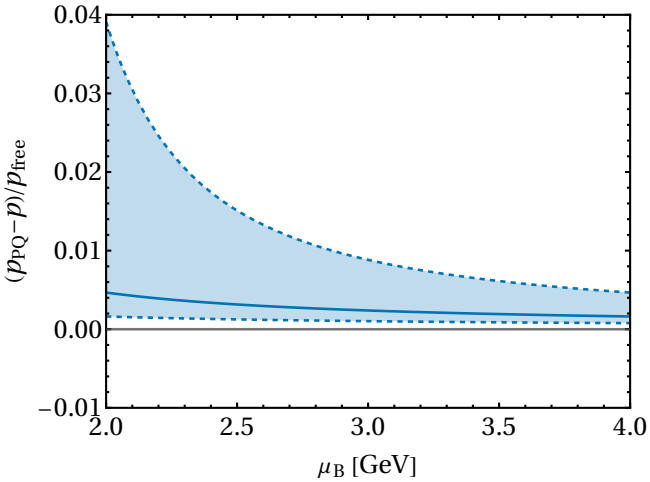


FIG. 1. The leading-order difference between the pressure of three-flavor dense QM and its phase-quenched version, normalized by the free pressure, as a function of the baryon chemical potential. The shaded region shows the scale variation of $X \in [1/2, 2]$, while the solid curve is the central value $X = 1$.

A few comments on the significance of our results are in order. First, we have generalized novel algorithmic techniques in vacuum quantum field theory to address large-scale multiloop computations at finite chemical potential, allowing us to perform the first-ever finite-μ computation of a 2PI four-loop diagram. We stress that our approach is not limited to the computation presented in this Letter, but it is equally applicable to diagrams of various topologies at the four-loop level. Second, our numerical evaluation of the leading $O(\alpha_s^3)$ perturbative correction to the difference $p_{\rm PQ}(\mu_{\rm B}) - p(\mu_{\rm B})$ demonstrates that Sign-Problem-free lattice simulations can provide a complementary method for accurately evaluating the pressure of cold and dense QM at $O(\alpha_s^3)$. Finally, we present an estimate for the necessary accuracy of phase-quenched lattice simulations to effectively constrain the EOS of QM at high density. Drawing upon the latest state-of-the-art perturbative findings (see the right panel in Fig. 2 of [22]), we approximate that the uncertainty in phase-quenched lattice simulations should be less than 25% for $\mu_{\rm B} \geq 2.2$ GeV.

Acknowledgments.—We thank Aleksi Kurkela, Heikki Mäntysaari, York Schröder, Saga Säppi, and Aleksi Vuorinen for their helpful comments and suggestions. P.N., R.P., and K.S. have been supported by the Academy of Finland grants no. 347499 and 353772. P.N. is also supported by the Doctoral School of the University of Helsinki, and K.S. is additionally supported by the Finnish Cultural Foundation. We also wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

* pablo.navarrete@helsinki.fi
† risto.paatelainen@helsinki.fi
‡ kaapo.seppanen@helsinki.fi

[1] W. Busza, K. Rajagopal, and W. van der Schee, Heavy Ion Collisions: The Big Picture, and the Big Questions, Ann. Rev. Nucl. Part. Sci. **68**, 339 (2018), arXiv:1802.04801 [hep-ph].
[2] J. N. Guenther, Overview of the QCD phase diagram: Recent progress from the lattice, Eur. Phys. J. A **57**, 136 (2021), arXiv:2010.15503 [hep-lat].
[3] C. Ratti, Lattice QCD and heavy ion collisions: a review of recent progress, Rept. Prog. Phys. **81**, 084301 (2018), arXiv:1804.07810 [hep-lat].
[4] G. Baym, T. Hatsuda, T. Kojo, P. D. Powell, Y. Song, and T. Takatsuka, From hadrons to quarks in neutron stars: a review, Rept. Prog. Phys. **81**, 056902 (2018), arXiv:1707.04966 [astro-ph.HE].
[5] E. Annala, T. Gorda, A. Kurkela, J. Nättilä, and A. Vuorinen, Evidence for quark-matter cores in massive neutron stars, Nature Phys. **16**, 907 (2020), arXiv:1903.09121 [astro-ph.HE].
[6] E. Annala, T. Gorda, J. Hirvonen, O. Komoltsev, A. Kurkela, J. Nättilä, and A. Vuorinen, Strongly interacting matter exhibits deconfined behavior in massive neutron stars, Nature Commun. **14**, 8451 (2023), arXiv:2303.11356 [astro-ph.HE].
[7] P. Hasenfratz and F. Karsch, Chemical Potential on the Lattice, Phys. Lett. B **125**, 308 (1983).
[8] F. Karsch and H. W. Wyld, Complex Langevin Simulation of the SU(3) Spin Model With Nonzero Chemical Potential, Phys. Rev. Lett. **55**, 2242 (1985).
[9] I. Barbour, N.-E. Behilil, E. Dagotto, F. Karsch, A. Moreo, M. Stone, and H. W. Wyld, Problems with Finite Density Simulations of Lattice QCD, Nucl. Phys. B **275**, 296 (1986).
[10] P. de Forcrand, Simulating QCD at finite density, PoS **LAT2009**, 010 (2009), arXiv:1005.0539 [hep-lat].
[11] K. Nagata, Finite-density lattice QCD and sign problem: Current status and open problems, Prog. Part. Nucl. Phys. **127**, 103991 (2022), arXiv:2108.12423 [hep-lat].
[12] I. Tews, T. Krüger, K. Hebeler, and A. Schwenk, Neutron matter at next-to-next-to-next-to-leading order in chiral effective field theory, Phys. Rev. Lett. **110**, 032504 (2013), arXiv:1206.0025 [nucl-th].
[13] J. E. Lynn, I. Tews, J. Carlson, S. Gandolfi, A. Gezerlis, K. E. Schmidt, and A. Schwenk, Chiral Three-Nucleon Interactions in Light Nuclei, Neutron-α Scattering, and Neutron Matter, Phys. Rev. Lett. **116**, 062501 (2016), arXiv:1509.03470 [nucl-th].
[14] C. Drischler, K. Hebeler, and A. Schwenk, Chiral interactions up to next-to-next-to-next-to-leading order and nuclear saturation, Phys. Rev. Lett. **122**, 042501 (2019), arXiv:1710.08220 [nucl-th].
[15] C. Drischler, R. J. Furnstahl, J. A. Melendez, and D. R. Phillips, How Well Do We Know the Neutron-Matter Equation of State at the Densities Inside Neutron Stars? A Bayesian Approach with Correlated Uncertainties, Phys. Rev. Lett. **125**, 202702 (2020), arXiv:2004.07232 [nucl-th].
[16] A. Kurkela, P. Romatschke, and A. Vuorinen, Cold Quark Matter, Phys. Rev. D **81**, 105021 (2010), arXiv:0912.1856 [hep-ph].
[17] A. Kurkela and A. Vuorinen, Cool quark matter, Phys. Rev. Lett. **117**, 042501 (2016), arXiv:1603.00750 [hep-ph].
[18] T. Gorda, A. Kurkela, P. Romatschke, S. Säppi, and A. Vuorinen, Next-to-Next-to-Next-to-Leading Order Pressure of Cold Quark Matter: Leading Logarithm, Phys. Rev. Lett. **121**, 202701 (2018), arXiv:1807.04120 [hep-ph].
[19] T. Gorda, A. Kurkela, R. Paatelainen, S. Säppi, and A. Vuorinen, Soft Interactions in Cold Quark Matter, Phys. Rev. Lett. **127**, 162003 (2021), arXiv:2103.05658 [hep-ph].
[20] T. Gorda, A. Kurkela, R. Paatelainen, S. Säppi, and A. Vuorinen, Cold quark matter at N3LO: Soft contributions, Phys. Rev. D **104**, 074015 (2021), arXiv:2103.07427 [hep-ph].
[21] T. Gorda and S. Säppi, Cool quark matter with perturbative quark masses, Phys. Rev. D **105**, 114005 (2022), arXiv:2112.11472 [hep-ph].
[22] T. Gorda, R. Paatelainen, S. Säppi, and K. Seppänen, Equation of State of Cold Quark Matter to $O(\alpha_s^3 \ln \alpha_s)$, Phys. Rev. Lett. **131**, 181902 (2023), arXiv:2307.08734 [hep-ph].
[23] A. Kurkela, E. S. Fraga, J. Schaffner-Bielich, and A. Vuorinen, Constraining neutron star matter with Quantum Chromodynamics, Astrophys. J. **789**, 127 (2014), arXiv:1402.6618 [astro-ph.HE].
[24] E. Annala, T. Gorda, A. Kurkela, and A. Vuorinen, Gravitational-wave constraints on the neutron-star-matter Equation of State, Phys. Rev. Lett. **120**, 172703 (2018), arXiv:1711.02644 [astro-ph.HE].
[25] O. Komoltsev and A. Kurkela, How Perturbative QCD Constrains the Equation of State at Neutron-Star Densities, Phys. Rev. Lett. **128**, 202701 (2022), arXiv:2111.05350 [nucl-th].
[26] R. Somasundaram, I. Tews, and J. Margueron, Perturbative QCD and the neutron star equation of state, Phys. Rev. C **107**, L052801 (2023), arXiv:2204.14039 [nucl-th].
[27] O. Komoltsev, R. Somasundaram, T. Gorda, A. Kurkela, J. Margueron, and I. Tews, Equation of state at neutron-star densities and beyond from perturbative QCD, (2023), arXiv:2312.14127 [nucl-th].
[28] T. Gorda, O. Komoltsev, A. Kurkela, and A. Mazeliauskas, Bayesian uncertainty quantification of perturbative QCD input to the neutron-star equation of state, JHEP **06**, 002, arXiv:2303.02175 [hep-ph].
[29] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
[30] B. P. Abbott *et al.* (LIGO Scientific, Virgo), GW170817: Measurements of neutron star radii and equation of state, Phys. Rev. Lett. **121**, 161101 (2018), arXiv:1805.11581 [gr-qc].
[31] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Properties of the binary neutron star merger GW170817, Phys. Rev. X **9**, 011001 (2019), arXiv:1805.11579 [gr-qc].
[32] H. T. Cromartie *et al.* (NANOGrav), Relativistic Shapiro delay measurements of an extremely massive millisecond pulsar, Nature Astron. **4**, 72 (2019), arXiv:1904.06759 [astro-ph.HE].
[33] J. Nättilä, M. C. Miller, A. W. Steiner, J. J. E. Kajava, V. F. Suleimanov, and J. Poutanen, Neutron star mass

and radius measurements from atmospheric model fits to X-ray burst cooling tail spectra, Astron. Astrophys. **608**, A31 (2017), arXiv:1709.09120 [astro-ph.HE].
[34] M. C. Miller *et al.*, PSR J0030+0451 Mass and Radius from *NICER* Data and Implications for the Properties of Neutron Star Matter, Astrophys. J. Lett. **887**, L24 (2019), arXiv:1912.05705 [astro-ph.HE].
[35] J. B. Kogut and D. K. Sinclair, Lattice QCD at finite temperature and density in the phase-quenched approximation, Phys. Rev. D **77**, 114503 (2008), arXiv:0712.2625 [hep-lat].
[36] C. Vafa and E. Witten, Parity Conservation in QCD, Phys. Rev. Lett. **53**, 535 (1984).
[37] T. D. Cohen, QCD inequalities for the nucleon mass and the free energy of baryonic matter, Phys. Rev. Lett. **91**, 032002 (2003), arXiv:hep-ph/0304024.
[38] Y. Fujimoto and S. Reddy, Bounds on the equation of state from QCD inequalities and lattice QCD, Phys. Rev. D **109**, 014020 (2024), arXiv:2310.09427 [nucl-th].
[39] D. Lee, Pressure inequalities for nuclear and neutron matter, Phys. Rev. C **71**, 044001 (2005), arXiv:nucl-th/0407101.
[40] G. D. Moore and T. Gorda, Bounding the QCD Equation of State with the Lattice, JHEP **12**, 133, arXiv:2309.15149 [nucl-th].
[41] D. T. Son, Superconductivity by long range color magnetic interaction in high density quark matter, Phys. Rev. D **59**, 094019 (1999), arXiv:hep-ph/9812287.
[42] H. Malekzadeh and D. H. Rischke, Gluon self-energy in the color-flavor-locked phase, Phys. Rev. D **73**, 114006 (2006), arXiv:hep-ph/0602082.
[43] M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, Color superconductivity in dense quark matter, Rev. Mod. Phys. **80**, 1455 (2008), arXiv:0709.4635 [hep-ph].
[44] Y. Fujimoto, Enhanced contribution of pairing gap to the QCD equation of state at large isospin chemical potential, (2023), arXiv:2312.11443 [hep-ph].
[45] S. Weinzierl, *Feynman Integrals* (Springer Cham, 2022).
[46] J. P. Blaizot, E. Iancu, and A. Rebhan, Quark number susceptibilities from HTL resummed thermodynamics, Phys. Lett. B **523**, 143 (2001), arXiv:hep-ph/0110369.
[47] P. Navarrete and Y. Schröder, Tackling the infamous g^6 term of the QCD pressure, PoS **LL2022**, 014 (2022), arXiv:2207.10151 [hep-ph].
[48] P. Navarrete and Y. Schröder, unpublished.
[49] A. Kärkkäinen, P. Navarrete, M. Nurmela, R. Paatelainen, and A. Vuorinen, unpublished.
[50] K. G. Chetyrkin and F. V. Tkachov, Integration by parts: The algorithm to calculate β-functions in 4 loops, Nucl. Phys. B **192**, 159 (1981).
[51] S. Laporta, High-precision calculation of multiloop Feynman integrals by difference equations, Int. J. Mod. Phys. A **15**, 5087 (2000), arXiv:hep-ph/0102033.
[52] J. M. Henn, Multiloop integrals in dimensional regularization made simple, Phys. Rev. Lett. **110**, 251601 (2013), arXiv:1304.1806 [hep-th].
[53] Y. Schroder, A fresh look on three-loop sum-integrals, JHEP **08**, 095, arXiv:1207.5666 [hep-ph].
[54] A. Gynther, M. Laine, Y. Schroder, C. Torrero, and A. Vuorinen, Four-loop pressure of massless O(N) scalar field theory, JHEP **04**, 094, arXiv:hep-ph/0703307.
[55] A. Gynther, A. Kurkela, and A. Vuorinen, The N(f)**3 g**6 term in the pressure of hot QCD, Phys. Rev. D **80**, 096002 (2009), arXiv:0909.3521 [hep-ph].
[56] T. Gorda, A. Kurkela, J. Österman, R. Paatelainen, S. Säppi, P. Schicho, K. Seppänen, and A. Vuorinen, Degenerate fermionic matter at N3LO: Quantum electrodynamics, Phys. Rev. D **107**, L031501 (2023), arXiv:2204.11893 [hep-ph].
[57] S. Catani, T. Gleisberg, F. Krauss, G. Rodrigo, and J.-C. Winter, From loops to trees by-passing Feynman's theorem, JHEP **09**, 065, arXiv:0804.3170 [hep-ph].
[58] Z. Capatti, V. Hirschi, D. Kermanschah, and B. Ruijl, Loop-Tree Duality for Multiloop Numerical Integration, Phys. Rev. Lett. **123**, 151602 (2019), arXiv:1906.06138 [hep-ph].
[59] Z. Capatti, Exposing the threshold structure of loop integrals, Phys. Rev. D **107**, L051902 (2023), arXiv:2211.09653 [hep-th].
[60] Z. Capatti, V. Hirschi, D. Kermanschah, A. Pelloni, and B. Ruijl, Numerical Loop-Tree Duality: contour deformation and subtraction, JHEP **04**, 096, arXiv:1912.09291 [hep-ph].
[61] `https://github.com/gplepage/vegas`.

QCD Sum Rule study of $J^P = 1^{\pm}$ exotic states with two heavy quarks in the molecular picture

T.M. Aliev [1,*] S. Bilmis [1,2,†] and M. Savcı [1,‡]

[1] *Department of Physics, Middle East Technical University, Ankara, 06800, Turkey*

[2] *TUBITAK ULAKBIM, Ankara, 06510, Turkey*

(Dated: March 5, 2024)

Abstract

In this study, the spectroscopic parameters of the exotic molecular states composed of mesons containing two heavy quarks (scalar - axial and pseudoscalar - axial meson combinations) are investigated within QCD sum rules. Our findings reveal that the molecular states containing charm quarks do not form bound states, whereas the states with b-quarks can form the exotic molecular states. This observation has significant implications for understanding the structure of these exotic states.

arXiv:2403.02147v1 [hep-ph] 4 Mar 2024

* taliev@metu.edu.tr

† sbilmis@metu.edu.tr

‡ savci@metu.edu.tr

I. INTRODUCTION

Numerous exotic hadronic states have been observed [1–5]. Among these states, the exotic doubly charmed T_{cc} state receives particular interest. This state was observed at LHCb near the D^0D^{*+} threshold in $D^0D^0\pi^+$ mass distribution with mass difference $\delta m = m_{T_{cc}} - (m_{D^{*+}} + m_{D^0}) = (-273 \pm 65 \pm 5^{+11}_{-14})$ keV and decay width $\Gamma = (410 \pm 165 \pm 43^{+18}_{-98})$ keV [6]. Further analysis reveals that this state has the quantum numbers $J^P = 1^+$. Although, this state had been already predicted theoretically in [7, 8], this was the first observation of tetraquark with doubly charmed quarks. This discovery triggered intensive theoretical studies [3–22]. See the review articles [23–28] for further details on the topic.

In addition to the exotic state with doubly c-quarks, the quark model also predicts the existence of states with two bb and bc heavy quarks. The states with doubly heavy quarks are usually described as tetraquark or molecular states. In the molecular picture, two quark and two anti-quark can form two color singlet mesons by exchanging light mesons. With this prediction, we can further explore the characteristics of these states. To better understand the characteristics of these states, high-precision experimental data and refined theoretical calculations are needed near $DD^*(s)$, $BB^*(s)$ thresholds.

In the present work, we study the exotic molecular states composed of scalar - axial mesons as well as pseudoscalar - axial mesons containing two heavy quarks with quantum numbers $J^P = 1^\pm$ in the framework of the QCD sum rules [19]. The paper is organized as follows. In Section II, we calculate the correlation functions of the exotic states with quantum numbers $J^P = 1^\pm$ within the molecular picture and derive the formulas for the mass and residues of these states. Section III presents the numerical analysis of the sum rules obtained for the masses and residues and the final section contains our conclusion.

II. SUM RULES FOR THE EXOTIC $J^P = 1^\pm$ STATES IN MOLECULAR PICTURE

In this section, we derive the mass sum rules for the exotic states with quantum numbers $J^P = 1^\pm$ within the molecular picture. The interpolating current for these states can be written as,

$$j_\mu = \left(\bar{Q}_1^a \Gamma_1 q^a\right) \left(\bar{Q}_2^b \Gamma_{2\mu} q^b\right) \ , \tag{1}$$

where q_1, q_2 and Q represent light u, d, s and heavy b or c quarks, respectively. The sum rules for the interpolating current with $\Gamma_1 = i\gamma_5$ and $\Gamma_{2\mu} = \gamma_\mu$ were studied in [11]. However, the interpolating current with $J^P = 1^+$ can also be chosen as $\Gamma_1 = 1$, $\Gamma_{2\mu} = \gamma_\mu\gamma_5$ and the state with $J^P = 1^-$ can be

represented by $\Gamma_1 = \gamma_5$, and $\Gamma_{2\mu} = \gamma_\mu\gamma_5$ or $\Gamma_1 = 1$. These cases are denoted as I and II, respectively. In the present work, we try to answer the following question: if these molecular structures were realized in nature what would be their masses and residues?

In order to determine the mass sum rules, the following correlation function is considered,

$$\begin{aligned}\Pi^{(i)}_{\mu\nu}&=i\int d^4x\, e^{ipx}\left\langle 0\left|T\left\{j_\mu(x)j^+_\nu(0)\right\}\right|0\right\rangle \ , \\ &=\left(-g_{\mu\nu}+\frac{p_\mu p_\nu}{p^2}\right)\Pi^{(i)}_1(p^2)+\frac{p_\mu p_\nu}{p^2}\Pi^{(i)}_0(p^2)\ ,\end{aligned} \tag{2}$$

where i corresponds to the relevant currents.

The polarization functions, $\Pi^{(i)}_0(p^2)$ and $\Pi^{(i)}_1(p^2)$ correspond to the spin-0 and spin-1 intermediate states, respectively. In further discussions, we will focus on the structure $-g_{\mu\nu}+\frac{p_\mu p_\nu}{p^2}$ since it only contains the contribution of the spin-1 particles.

The sum rules for the relevant physical quantity can be obtained by calculating the correlation function in terms of the hadrons and quark-gluon degrees of freedom in the deep Euclidean region $p^2 \to -\infty$ by using the operator product expansion (OPE). Then, by matching both representations and performing Borel transformation over the variable $-p^2$, the desired sum rules are obtained.

At the hadronic level, the correlation function can be calculated with the help of the dispersion relation for the invariant function $\Pi^{(i)}_1$,

$$\Pi^{(i)}_1=\frac{(p^2)^\alpha}{\pi}\int_{(2m_Q+m_{q_1}+m_{q_2})^2}\frac{\mathrm{Im}\Pi^{(i)}_1(s)}{(s^2)^\alpha(s-p^2-i\epsilon)}+\text{subtraction terms}\ . \tag{3}$$

In the sum rules method, $\mathrm{Im}\Pi^{(i)}_1(s)$ is defined as the spectral density function,

$$\rho^{(i)}(s)=\frac{1}{\pi}\mathrm{Im}\Pi^{(i)}_1=\lambda_i^2\delta(s-m_i^2)+\text{continuum}+\text{higher states contributions}\ , \tag{4}$$

where m_i is the lowest-lying resonance. The decay constant λ_i is defined in the standard way as,

$$\left\langle 0\left|j^{(i)}_\mu\right|T_{QQ}(p)\right\rangle=\lambda_i\epsilon_\mu\ , \tag{5}$$

with the polarization vector ϵ_μ of $J^P=1^-(1^+)$ states. Using Eq. (4), we obtain the following expression for the correlation function from hadronic side for the structure $\left(-g_{\mu\nu}+\frac{p_\mu p_\nu}{p^2}\right)$,

$$\Pi^{(i)}_1=\frac{\lambda_i^2}{m_i^2-p^2}\ . \tag{6}$$

Moreover, the correlation function $\Pi_1^{(i)}$ can be calculated from the QCD side with the help of the OPE in deep Euclidean domain $p^2 \to -\infty$ and we get,

$$\Pi_{\mu\nu}=i\int d^4x\, e^{ipx}\Big\{\mathrm{Tr}\Big[\Gamma_1 S_Q^{aa_1}(x)\widetilde{\Gamma}_1 S_{q_1}^{a_1a}(-x)\Big]\times\mathrm{Tr}\Big[\Gamma_2 S_Q^{bb_1}(x)\widetilde{\Gamma}_2 S_{q_2}^{b_1b}(-x)\Big]$$
$$-\mathrm{Tr}\Big[S_Q^{ba_1}(x)\widetilde{\Gamma}_1 S_{q_1}^{a_1a}(-x)\Gamma_1 S_Q^{ab_1}(x)\widetilde{\Gamma}_2 S_{q_2}^{b_1b}(-x)\Gamma_2\Big]\,, \tag{7}$$

in terms of the light S_{q_i} and heavy S_Q quark propagators. Hence, to calculate the correlation function from the QCD side, we need the explicit expressions of the light and heavy quark propagators.

The light quark propagator in x-representation, up to linear order in light quark mass is given as,

$$S_q^{ab}(x)=\frac{i\not{x}}{2\pi^2x^4}\delta^{ab}-\frac{m_q}{4\pi^2x^2}\delta^{ab}-\frac{\langle\bar{q}q\rangle}{12}\left(1-i\frac{m_q}{4}\not{x}\right)\delta^{ab}-\frac{x^2}{192}m_0^2\langle\bar{q}q\rangle\left(1-i\frac{m_q}{6}\not{x}\right)\delta^{ab}$$
$$+\frac{i}{32\pi^2x^2}g_sG_{\mu\nu}^{ab}(\sigma^{\mu\nu}\not{x}+\not{x}\sigma^{\mu\nu})-\frac{1}{3^3\,2^{10}}\langle\bar{q}q\rangle\langle g_s^2G^2\rangle x^4\delta^{ab}+\cdots\,, \tag{8}$$

where $\langle\bar{q}q\rangle$ is the light quark condensate, $G_{\mu\nu}$ is the gluon field strange tensor. Besides, the heavy quark propagator in x–representation is given as (see for example [11]),

$$S_Q(x)^{ab}=\frac{m_Q^2\delta^{ab}}{(2\pi)^2}\left[i\not{x}\frac{K_2(m_Q\sqrt{-x^2})}{(\sqrt{-x^2})^2}+\frac{K_1(m_Q\sqrt{-x^2})}{\sqrt{-x^2}}\right]$$
$$-\frac{m_Qg_sG_{\mu\nu}^{ab}}{8(2\pi)^2}\left[i(\sigma^{\mu\nu}\not{x}+\not{x}\sigma^{\mu\nu})\frac{K_1(m_Q\sqrt{-x^2})}{\sqrt{-x^2}}+2\sigma^{\mu\nu}K_0(m_Q\sqrt{-x^2})\right]$$
$$-\frac{\langle g_s^2G^2\rangle\delta^{ab}}{(3^2\,2^8\pi)^2}\left[(im_Q\not{x}-6)(-x^2)\frac{K_1(m_Q\sqrt{-x^2})}{\sqrt{-x^2}}+m_Qx^4\frac{K_2(m_Q\sqrt{-x^2})}{(\sqrt{-x^2})^2}\right]\,, \tag{9}$$

where K_0, K_1 and K_2 are the modified Bessel functions of the second kind. Using the explicit expressions of the light and heavy quark propagators, we obtain the spectral density $\rho^{(i)}$ after standard but tedious calculations, as given in Appendix A.

By equating the coefficients of the corresponding Lorentz structures of the correlation function from both representations, we obtain the sum rules of the relevant physical quantities. As a final step, to suppress the contributions of higher and continuum, we perform Borel transformation with respect to $-p^2$

$$\lambda^2e^{-m^2/M^2}=\int_{s_{min}}^{s_0}ds\,e^{-s/M^2}\rho(s)\,, \tag{10}$$

where s_0 is the continuum threshold, M^2 is the Borel mass parameter, and $s_{min}=(m_{Q_1}+m_{Q_2}+m_{q_1}+m_{q_2})^2$. Note that the mass sum rules for T_{QQ} can easily be obtained by taking derivative with

respect to inverse Borel mass parameter, from which we get,

$$m^2 = \frac{\int_{s_{min}}^{s_0} ds\, s e^{-s/M^2} \rho(s)}{\int_{s_{min}}^{s_0} ds\, e^{-s/M^2} \rho(s)} \ . \tag{11}$$

Having determined the mass of the considered states, the residues can be found using Eq. 10.

III. NUMERICAL ANALYSIS

In this section, we perform numerical analysis on the sum rules for the tetraquark states with doubly heavy quarks with $J^P = 1^{\pm}$ that we obtained in the previous section. The sum rules contain many input parameters, such as quark and gluon condensates of appropriate dimensions and masses of the quarks. The numerical values of these input parameters are listed in Table I. For the heavy quark masses, we use their values in the $\overline{MS}$ scheme.

Π_i	Structures
$\overline{m}_s(2\text{ GeV})$	$93.4^{+8.6}_{-3.4}$ MeV [29]
$\overline{m}_b(\overline{m}_b)$	$4.18^{+0.03}_{-0.02}$ GeV [29]
$\overline{m}_c(\overline{m}_c)$	(1.27 ± 0.02) GeV [29]
$\langle\bar{q}q\rangle$ (1 GeV)	$(-1.65 \pm 0.15) \times 10^{-2}$ GeV^3 [30]
$\langle\bar{s}s\rangle$	$(0.8 \pm 0.2)\langle\bar{q}q\rangle$ GeV^3 [30]
m_0^2	(0.8 ± 0.2) GeV^2 [30]
$\langle\frac{\alpha_s G^2}{\pi}\rangle$	(0.012 ± 0.006) GeV^4 [31]

TABLE I. The numerical values of the constant parameters used in our calculations are listed.

In addition to these input parameters, the sum rules contain also two auxiliary parameters, namely, Borel mass M^2 and the continuum threshold s_0. To decide the working regions of s_0 and M^2, it is important to examine how the mass of the tetraquark state will vary with changes in parameter values. The dependency should be minimal. Moreover, following requirements should be met for an acceptable Borel mass parameter region:

a) The upper bound of M^2 is obtained from the requirement that the continuum and higher states contributions constitute a maximum of 40% of the total result, i.e., the pole contribution (PC) should dominate. This is determined by the following ratio,

$$PC = \frac{\int_{s_{min}}^{s_{max}} ds\, \rho(s) e^{-s/M^2}}{\int_{s_{min}}^{\infty} ds\, \rho(s) e^{-s/M^2}} \ .$$

b) The minimum value of M^2 is established by the requirement that the OPE should be convergent. We must ensure that any condensate with the highest number of dimensions contributes less than 5% to the total result.

c) The continuum threshold parameter s_0 can be determined by requiring that the dependency of the mass of the considered state exhibit minimum variation for the Borel mass window that satisfies the conditions mentioned above.

To ensure that all the requirements are met, numerical analysis are performed. And the working regions of the parameters M^2 and s_0 are identified which are presented in Table II.

		M^2 (GeV^2)	s_0 (GeV^2)
$J^P = 1^-$(I)	BB_1	$9-12$	130 ± 2
	BB_{1s}	$9-12$	132 ± 2
	B_sB_1	$9-12$	132 ± 2
	B_sB_{1s}	$9-12$	134 ± 2
$J^P = 1^-$(II)	B_0B^*	$9-12$	130 ± 2
	$B_0B_s^*$	$9-12$	132 ± 2
	$B_{0s}B^*$	$9-12$	132 ± 2
	$B_{0s}B_s^*$	$9-12$	134 ± 2
$J^P = 1^+$	B_0B_1	$9-12$	134 ± 2
	B_0B_{1s}	$9-12$	136 ± 2
	$B_{0s}B_1$	$9-12$	136 ± 2
	$B_{0s}B_{1s}$	$9-12$	140 ± 2

TABLE II. The working regions of M^2 and s_0 for the considered molecular states.

As an example, Figs. 1 and 2 show the dependency of the masses and residues of BB_1 and B_0B_1 states on M^2, at the fixed values of s_0 for $J^P = 1^-$ and $J^P = 1^+$ states, respectively. These figures indicate that when M^2 varies in the domain 10 GeV^2 $\leq$ M^2 $\leq$ 12 GeV^2 (after M^2 = 12 GeV^2 the results practically do not change) the masses of the BB_1 and B_0B_1 states exhibit quite good stability.

Performing similar analysis, we obtained the masses and the residues of the other possible molecular states, the results of which are given in Table III.

Similar analysis are also performed for doubly charmed systems. Our numerical calculations show that for the Borel Mass parameter region in which the pole contribution is greater than 1/2, we can not find the stable plateau for the mass of the doubly charmed states. Therefore, we infer that mesons containing two charm quarks do not form bound states in molecular picture considered.

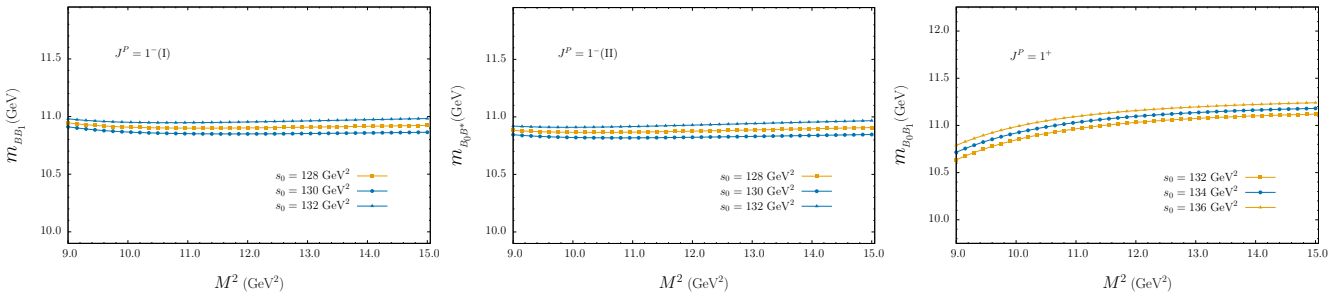


FIG. 1. Dependence of the mass of the BB_1 B_0B^*, and B_0B_1 molecular states on M^2 at fixed values of the continuum threshold s_0

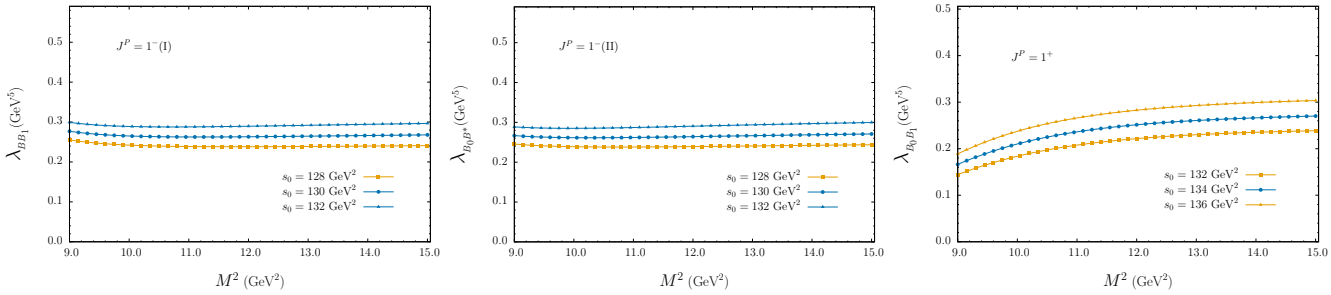


FIG. 2. Dependence of the residue of the BB_1, B_0B^*, and B_0B_1 molecular states on M^2 at fixed values of the continuum threshold s_0

This phenomenon has also been noted in several works [32–37]. The large difference in QQ binding energy for the c and b quarks can explain this outcome. It should be noted that the doubly charmed tetraquark states with $J^P = 1^+$ (scalar-axial current) was considered in [9] with slightly different current and they predicted the mass of these states on the contrary to our result. The main difference

		MASS (GeV)	RESIDUE (GeV^5)
$J^P = 1^-$(I)	BB_1	10.90 ± 0.03	0.26 ± 0.02
	BB_{1s}	10.94 ± 0.03	0.27 ± 0.04
	B_sB_1	10.94 ± 0.03	0.27 ± 0.04
	B_sB_{1s}	11.10 ± 0.02	0.30 ± 0.02
$J^P = 1^-$(II)	B_0B^*	10.92 ± 0.03	0.27 ± 0.02
	$B_sB_s^*$	10.94 ± 0.03	0.28 ± 0.02
	$B_{0s}B^*$	10.96 ± 0.03	0.28 ± 0.02
	$B_{0s}B_s^*$	11.10 ± 0.04	0.32 ± 0.02
$J^P = 1^+$	B_0B_1	11.00 ± 0.10	0.20 ± 0.02
	B_0B_{1s}	11.15 ± 0.05	0.24 ± 0.03
	$B_{0s}B_1$	11.15 ± 0.05	0.24 ± 0.03
	$B_{0s}B_1$	11.25 ± 0.05	0.28 ± 0.02

TABLE III. Numerical results for the masses and residues of the $B_{(s)}B_{1(s)}$ and $B_{0(s)}B_{1(s)}$ molecular systems

is that in [9], the input parameters of the QCD parameters are used at a different μ scale, however, in our case traditional $\mu = 1$ GeV scale is used. These studies can be very helpful in putting together the pieces of the puzzle concerning the composition of these tetraquark states.

For a final remark, we would like to note that the intermediate two meson states can also contribute to the correlation functions under consideration. However, it is shown in [38, 39] that the contributions to the correlation functions coming from these two meson states are quite small, and therefore can safely be neglected. For this reason, these contributions are not taken into account in this study.

CONCLUSION

In this work, we study the spectroscopic parameters, namely masses and residues of the potential exotic states with the quantum numbers $J^P = 1^{\pm}$, containing two heavy quarks in the molecular picture within QCD sum rule. Our findings revealed that no bound states can be formed for the T_{cc} states with quantum numbers $J^P = 1^{\pm}$ for the considered currents. On the other hand, for the T_{bb} states, our calculations demonstrated that bound states can indeed exist within both considered molecular frameworks. This result significantly contributes to the ongoing discussions in the literature regarding the inner nature of exotic states. Furthermore, the obtained mass values for $J^P = 1^+$ and $J^P = 1^-$ exotic states within the molecular picture show a difference of ~ 200 MeV. This result can be checked in future experiments to establish the "right" picture.

Even though it is challenging to determine the states near $BB^*(s)$ and $DD^*(s)$ thresholds, the results obtained for the masses of the tetraquark states can be considered in future experiments looking for exotic hadrons.

ACKNOWLEDGMENT

The authors thank Qi Xin for communication.

Appendix A

In this Appendix, we present the expressions of the spectral density $\rho(s)$ for the $J^P = 1^-$(I), $J^P = 1^-$(II), and $J^P = 1^+$ tetraquark systems. Upper (lower) sign corresponds to 1^-(I) (1^+) quantum number. With the replacement of $m_u \to -m_u$ and $\langle \bar{u}u \rangle = -\langle \bar{u}u \rangle$ for the $J^P = 1^-$(I) case,

one can obtain the spectral density for the $J^P = 1^-$(II) case.

$$\rho^{(pert)}(s)=\pm\frac{1}{(3\times 2^{14})\pi^6}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha^3}\int_{\beta_{min}}^{\beta_{max}}\frac{d\beta}{\beta^3}\Big[(\alpha+\beta)m_b^2-\alpha\beta s\Big]^2$$
$$\times\Big\{(1-\beta)\beta m_b^3\Big([8+29\beta(1+\beta)]m_b\mp 120\beta[(1+\beta)m_d\mp 2m_u)\Big)$$
$$-\alpha^4(29m_b^2-33\beta s)(m_b^2-\beta s)+2\alpha m_b\Big\{m_b\Big[(4+21\beta-58\beta^3)m_b^2$$
$$\mp 12\beta(33-32\beta)m_dm_u\mp 60\beta m_b(m_d-3\beta^2m_d\mp 2m_u\pm 4\beta m_u)\Big]$$
$$-(1-\beta)\beta\Big[(4+31\beta(1+\beta))m_b\mp 60\beta(m_d+\beta m_d\mp 2m_u)\Big]s\Big\}$$
$$-2\alpha^3\beta(58m_b^4\mp 60m_b^3m_d-93\beta m_b^2s\pm 60\beta m_bm_ds+33\beta^2s^2)$$
$$+3\alpha^2\Big\{(7-58\beta^2)m_b^4+40\beta m_b^3(\pm 3\beta m_d-2m_u)\mp 80\beta^2m_b(\beta m_d\mp m_u)s$$
$$\pm 2\beta m_b^2(128m_dm_u\mp 9s\pm 31\beta^2s)+\beta^2s\Big[\pm 8m_dm_u+11(1-\beta^2)s\Big]\Big\}\Big\}$$

$$\rho^{(\langle\bar dd\rangle\langle\bar uu\rangle)}(s)=-\frac{\langle\bar dd\rangle\langle\bar uu\rangle}{(3\times 2^8)\pi^2}\int_{\alpha_{min}}^{\alpha_{max}}d\alpha\Big\{4m_b\Big[13m_b-5(\pm 2m_d-m_u)\Big]$$
$$+\alpha\Big[12m_0^2+8m_b(m_b\pm 5m_d)-m_u(20m_b\pm 33m_d)-12s\Big]$$
$$-3\alpha^2(4m_0^2\mp 11m_dm_u-4s)\Big\}$$

$$\rho^{(m_0^2\langle\bar dd\rangle\langle g^2G^2\rangle)}(s)=\frac{5(m_0^2\langle\bar dd\rangle\langle g^2G^2\rangle)}{(3\times 2^{12})m_b\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}d\alpha(1-\alpha)^2$$

$$\rho^{(m_0^2\langle\bar uu\rangle\langle g^2G^2\rangle)}(s)=\mp\frac{5(m_0^2\langle\bar uu\rangle\langle g^2G^2\rangle)}{(3\times 2^{11})m_b\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}d\alpha(1-\alpha)^2$$

$$\rho^{(m_0^2\langle\bar dd\rangle)}(s)=\mp\frac{m_0^2\langle\bar dd\rangle}{(3\times 2^{10})\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\Big\{m_b\Big[\pm 30m_b^2-20m_dm_u(1-\alpha)\alpha$$
$$-\alpha m_b(12m_d+22\alpha m_d\mp 39m_u\mp 6\alpha m_u)\Big]$$
$$-(1-\alpha)\alpha(\pm 30m_b-22\alpha m_d\pm 9\alpha m_u)s\Big\}$$
$$+\frac{m_0^2\langle\bar dd\rangle}{(3\times 2^{10})\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\int_{\beta_{min}}^{\beta_{max}}d\beta\Big\{m_b^2\Big[30(\alpha+\beta)m_b\mp\alpha m_d\Big]-30m_b\alpha\beta s\Big\}$$

$$\rho^{(m_0^2\langle\bar uu\rangle)}(s)=\pm\frac{m_0^2\langle\bar uu\rangle}{(3\times 2^{10})\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\Big\{m_b\Big[30m_b^2\mp 10m_dm_u(1-\alpha)\alpha$$
$$+\alpha m_b(12m_u+22\alpha m_u\mp 39m_d\mp 6\alpha m_d)\Big]$$
$$-(1-\alpha)\alpha(30m_b\mp 9\alpha m_d+22\alpha m_u)s\Big\}$$
$$\mp\frac{m_b^2m_um_0^2\langle\bar uu\rangle}{(3\times 2^{10})\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}d\alpha\int_{\beta_{min}}^{\beta_{max}}d\beta$$

$$\rho^{(\langle\bar{d}d\rangle\langle g^2G^2\rangle)}(s)=-\frac{\langle\bar{d}d\rangle\langle g^2G^2\rangle}{(3^2\times 2^{13})m_b\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\Big\{4\big[3-\alpha(193-15\alpha)\big]m_b^2$$
$$+\alpha m_b\big[(\pm 763\mp 760\alpha)m_d-2(61+74\alpha+6\alpha^2)m_u\big]$$
$$\pm 40(1-\alpha)\alpha\big[3(1-\alpha)m_dm_u+(\pm 3\mp 2\alpha)s\big]\Big\}$$
$$+\frac{\langle\bar{d}d\rangle\langle g^2G^2\rangle}{(3^2\times 2^{13})m_b\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\int_{\beta_{min}}^{\beta_{max}}d\beta\Big\{4\big[8\alpha-9(3+2\beta)\big]m_b^2$$
$$\pm 3\alpha m_bm_d+120\alpha\beta s\Big\}$$

$$\rho^{(\langle\bar{u}u\rangle\langle g^2G^2\rangle)}(s)=-\frac{\langle\bar{u}u\rangle\langle g^2G^2\rangle}{(3^2\times 2^{13})m_b\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\Big\{\pm 4\big[21+2\alpha(17-5\alpha)\big]m_b^2$$
$$-\alpha m_b\big[2(61+74\alpha+6\alpha^2)m_d\mp(133-130\alpha)m_u\big]$$
$$-60(1-\alpha)^2\alpha(m_dm_u\pm 2s)\Big\}$$
$$\pm\frac{\langle\bar{u}u\rangle\langle g^2G^2\rangle}{(3\times 2^{13})\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\int_{\beta_{min}}^{\beta_{max}}d\beta(12m_b+\alpha m_u)$$

$$\rho^{(\langle g^2G^2\rangle^2)}(s)=\mp\frac{\langle g^2G^2\rangle^2}{(3^3\times 2^{18})m_b^2\pi^6}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\Big\{\big[192-\alpha(4143+584\alpha)\big]m_b^2$$
$$\mp 132\alpha\big[1-2(3-\alpha)\alpha\big]m_dm_u\pm 20\alpha m_b\big[(31+74\alpha)m_d\mp 11(m_u+2\alpha m_u)\big]$$
$$-8(3-\alpha)(1-\alpha)\alpha s\Big\}$$
$$+\frac{\langle g^2G^2\rangle^2}{(3^3\times 2^{17})m_b^2\pi^6}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha}\int_{\beta_{min}}^{\beta_{max}}\frac{d\beta}{\beta}\Big\{(\pm)m_b\big[18\alpha^2m_b+18(22-9\beta)\beta m_b$$
$$+5\alpha\beta(37m_b\mp 24m_d)\big]\mp 6\alpha\beta(3\alpha+5\beta)s\Big\}$$

$$\rho^{(\langle g^2G^2\rangle)}(s)=\pm\frac{\langle g^2G^2\rangle}{(3\times 2^{15})m_b\pi^6}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha(1-\alpha)}\Big\{\big[m_b^2-(1-\alpha)\alpha s\big]$$
$$\times\Big(m_b\big[33m_b^2\mp 80(1-\alpha)m_dm_u-20m_b(m_d\mp m_u)\big]$$
$$-(1-\alpha)\alpha\big[33m_b\mp 20(m_d\mp m_u)\big]s\Big)\Big\}$$
$$-\frac{\langle g^2G^2\rangle}{(3\times 2^{16})m_b\pi^6}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha^2}\int_{\beta_{min}}^{\beta_{max}}\frac{d\beta}{\beta^2}\Big\{\pm\beta m_b^4\Big(\big\{108+\beta\big[219+\beta(360+233\beta)\big]\big\}m_b$$
$$+24\beta\big\{\big[6+4(1-\beta)\beta\big]m_d\mp(3-5\beta)m_u\big\}\Big)$$
$$\mp 3\alpha^4m_b(m_b^2-\beta s)^2\pm 2\alpha^3(m_b^2-\beta s)\big\{2m_b^2\big[(27+47\beta)m_b\pm 18\beta m_d\big]$$
$$\mp\beta^2(\pm 61m_b-80m_d)s\big\}$$
$$+2m_b^2\alpha\Big(m_b\Big[\pm 3\big\{18+\beta\big[1+2\beta(82+55\beta)\big]\big\}m_b^2-6\beta(12+35\beta)m_dm_u$$

$$
\begin{aligned}
&+4\beta m_b\big\{3\big[6-5\beta(14+\beta)\big]m_d\mp(9-94\beta)m_u\big\}\Big]\\
&\mp\beta\Big[\big\{54+\beta\big[111+5\beta(114+37\beta)\big]\big\}m_b\\
&-12\beta\big\{\big[6+\beta(4-9\beta)\big]m_d\mp(3-10\beta)m_u\big\}\Big]s\Big)\\
&-\alpha^2\Big[m_b^3\Big((\pm)3(71-244\beta-206\beta^2)m_b^2+8\beta m_b\big[6(37-\beta)m_d\mp 79m_u\big]+420\beta m_d m_u\Big)\\
&\pm 2\beta m_b\Big(m_b\big\{\big[105-\beta(702+343\beta)\big]m_b\pm 8\beta(111+19\beta)m_d\big\}\\
&-94\beta(4m_b\mp 3m_d)m_u\Big)s\pm\beta^2\big\{\big[3+\beta(780+137\beta)\big]m_b-120\beta(\pm\beta m_d-m_u)\big\}s^2\Big]\Big\}
\end{aligned}
$$

$$
\begin{aligned}
\rho^{(\langle\bar{d}d\rangle)}(s)=&\mp\frac{\langle\bar{d}d\rangle}{2^{10}\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha(1-\alpha)}\big[m_b^2-(1-\alpha)\alpha s\big]\Big\{m_b\big\{11m_b m_d\\
&\mp 2\big[m_b\mp 10(1-\alpha)m_d\big]m_u\big\}-(1-\alpha)\alpha(11m_d\mp 2m_u)s\Big\}\\
&-\frac{\langle\bar{d}d\rangle}{2^{10}\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha^2}\int_{\beta_{min}}^{\beta_{max}}\frac{d\beta}{\beta}\big[(\alpha+\beta)m_b^2-\alpha\beta s\big]\Big\{m_b^2\big\{(\alpha+\beta)\\
&\times\big[20(\alpha+\beta)m_b\pm 9\alpha m_d\big]+44\alpha m_u\big\}\mp\alpha\beta\big[(\pm)20(\alpha+\beta)m_b+11\alpha m_d\big]s\Big\}
\end{aligned}
$$

$$
\begin{aligned}
\rho^{(\langle\bar{u}u\rangle)}(s)=&-\frac{\langle\bar{u}u\rangle}{2^{10}\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha(1-\alpha)}\big[m_b^2-(1-\alpha)\alpha s\big]\\
&\times\big[20(1-\alpha)m_b m_d m_u+m_b^2(2m_d\mp 11m_u)-(1-\alpha)\alpha(2m_d\mp 11m_u)s\big]\\
&\pm\frac{\langle\bar{u}u\rangle}{2^{10}\pi^4}\int_{\alpha_{min}}^{\alpha_{max}}\frac{d\alpha}{\alpha^2}\int_{\beta_{min}}^{\beta_{max}}\frac{d\beta}{\beta}\big[(\alpha+\beta)m_b^2-\alpha\beta s\big]\\
&\times\Big\{m_b\big\{\pm 4m_b\big[\pm 5(\alpha+\beta)m_b-11\alpha m_d\big]-\alpha\big[9(\alpha+\beta)m_b\mp 20\beta m_d\big]m_u\big\}\\
&-\alpha\beta(20m_b-11\alpha m_u)s\Big\}
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{min}&=\frac{s-\sqrt{s(s-4m_b^2)}}{2s}\ ,\\
\alpha_{max}&=\frac{s+\sqrt{s(s-4m_b^2)}}{2s}\ ,\\
\beta_{min}&=\frac{m_b^2\alpha}{s\alpha-m_b^2}\ ,\\
\beta_{max}&=1-\alpha\ .
\end{aligned}
$$

The results for the spectral densities containing strange quarks can also be obtained from the presented results by replacing the mass(es) and condensate(s) of the appropriate light quark(s) with those of the s-quark.

[1] **LHCb** Collaboration, R. Aaij *et al.*, "*Observation of an exotic narrow doubly charmed tetraquark*," Nature Phys. **18** no. 7, (2022) 751–754, [2109.01038].

[2] **Belle** Collaboration, S. K. Choi *et al.*, "*Observation of a narrow charmonium-like state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays*," Phys. Rev. Lett. **91** (2003) 262001, [hep-ex/0309032].

[3] **LHCb** Collaboration, R. Aaij *et al.*, "*Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*," Phys. Rev. Lett. **115** (2015) 072001, [1507.03414].

[4] **D0** Collaboration, V. M. Abazov *et al.*, "*Evidence for a $B_s^0\pi^{\pm}$ state*," Phys. Rev. Lett. **117** no. 2, (2016) 022003, [1602.07588].

[5] **LHCb** Collaboration, R. Aaij *et al.*, "*Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$*," Phys. Rev. Lett. **122** no. 22, (2019) 222001, [1904.03947].

[6] **LHCb** Collaboration, R. Aaij *et al.*, "*Study of the doubly charmed tetraquark T_{cc}^+*," [2109.01056].

[7] J. l. Ballot and J. M. Richard, "*FOUR QUARK STATES IN ADDITIVE POTENTIALS*," Phys. Lett. B **123** (1983) 449–451.

[8] S. Zouzou, B. Silvestre-Brac, C. Gignoux, and J. M. Richard, "*FOUR QUARK BOUND STATES*," Z. Phys. C **30** (1986) 457.

[9] Q. Xin and Z.-G. Wang, "*Analysis of the doubly-charmed tetraquark molecular states with the QCD sum rules*," Eur. Phys. J. A **58** no. 6, (2022) 110, [2108.12597].

[10] S. S. Agaev, K. Azizi, and H. Sundu, "*Newly observed exotic doubly charmed meson T_{cc}^+*," Nucl. Phys. B **975** (2022) 115650, [2108.00188].

[11] T. M. Aliev, S. Bilmis, and M. Savci, "*Determination of the spectroscopic parameters of beauty-partners of T_{cc} from QCD*," [2111.01081].

[12] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, "*QCD and Resonance Physics. Theoretical Foundations*," Nucl. Phys. B **147** (1979) 385–447.

[13] E. Braaten, L.-P. He, and A. Mohapatra, "*Masses of doubly heavy tetraquarks with error bars*," Phys. Rev. D **103** no. 1, (2021) 016001, [2006.08650].

[14] Q. Meng, E. Hiyama, A. Hosaka, M. Oka, P. Gubler, K. U. Can, T. T. Takahashi, and H. S. Zong, "*Stable double-heavy tetraquarks: spectrum and structure*," Phys. Lett. B **814** (2021) 136095, [2009.14493].

[15] J.-B. Cheng, S.-Y. Li, Y.-R. Liu, Z.-G. Si, and T. Yao, "*Double-heavy tetraquark states with heavy diquark-antiquark symmetry*," Chin. Phys. C **45** no. 4, (2021) 043102, [2008.00737].

[16] J. M. Dias, S. Narison, F. S. Navarra, M. Nielsen, and J. M. Richard, "*Relation between $T_{cc,bb}$ and*

$X_{c,b}$ *from QCD*," Phys. Lett. B **703** (2011) 274–280, [1105.5630].

[17] D. Gao, D. Jia, Y.-J. Sun, Z. Zhang, W.-N. Liu, and Q. Mei, "*Masses of doubly heavy tetraquark states with isospin* $= \frac{1}{2}$ *and 1 and spin-parity* $1^{+\pm}$," [2007.15213].

[18] H. Ren, F. Wu, and R. Zhu, "*Hadronic molecule interpretation of* T_{cc}^{+} *and its beauty-partners*," [2109.02531].

[19] B. L. Ioffe and A. V. Smilga, "*Nucleon Magnetic Moments and Magnetic Properties of Vacuum in QCD*," Nucl. Phys. B **232** (1984) 109–142.

[20] C. B. Chiu, J. Pasupathy, and S. L. Wilson, "*The Gluon Field Contribution in QCD Sum Rules for the Magnetic Moments of the Nucleons*," Phys. Rev. D **36** (1987) 1451.

[21] F. S. Navarra, M. Nielsen, and S. H. Lee, "*QCD sum rules study of QQ - anti-u anti-d mesons*," Phys. Lett. B **649** (2007) 166–172, [hep-ph/0703071].

[22] M.-L. Du, W. Chen, X.-L. Chen, and S.-L. Zhu, "*Exotic* $QQ\bar{q}\bar{q}$*,* $QQ\bar{q}\bar{s}$ *and* $QQ\bar{s}\bar{s}$ *states*," Phys. Rev. D **87** no. 1, (2013) 014003, [1209.5134].

[23] N. Brambilla, S. Eidelman, C. Hanhart, A. Nefediev, C.-P. Shen, C. E. Thomas, A. Vairo, and C.-Z. Yuan, "*The* XYZ *states: experimental and theoretical status and perspectives*," Phys. Rept. **873** (2020) 1–154, [1907.07583].

[24] H.-X. Chen, W. Chen, X. Liu, Y.-R. Liu, and S.-L. Zhu, "*An updated review of the new hadron states*," Rept. Prog. Phys. **86** no. 2, (2023) 026201, [2204.02649].

[25] Y.-R. Liu, H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, "*Pentaquark and Tetraquark states*," Prog. Part. Nucl. Phys. **107** (2019) 237–320, [1903.11976].

[26] H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, "*The hidden-charm pentaquark and tetraquark states*," Phys. Rept. **639** (2016) 1–121, [1601.02092].

[27] P. Bicudo, "*Tetraquarks and pentaquarks in lattice QCD with light and heavy quarks*," Phys. Rept. **1039** (2023) 1–49, [2212.07793].

[28] H.-X. Chen, W. Chen, X. Liu, Y.-R. Liu, and S.-L. Zhu, "*A review of the open charm and open bottom systems*," Rept. Prog. Phys. **80** no. 7, (2017) 076201, [1609.08928].

[29] **Particle Data Group** Collaboration, R. L. Workman *et al.*, "*Review of Particle Physics*," PTEP **2022** (2022) 083C01.

[30] B. L. Ioffe, "*QCD at low energies*," Prog. Part. Nucl. Phys. **56** (2006) 232–277, [hep-ph/0502148].

[31] B. L. Ioffe, "*Condensates in quantum chromodynamics*," Phys. Atom. Nucl. **66** (2003) 30–43, [hep-ph/0207191]. [Yad. Fiz.66,32(2003)].

[32] M. Karliner and J. L. Rosner, "*Discovery of doubly-charmed* Ξ_{cc} *baryon implies a stable* $(bb\bar{u}\bar{d})$ *tetraquark*," Phys. Rev. Lett. **119** no. 20, (2017) 202001, [1707.07666].

[33] W. Park, S. Noh, and S. H. Lee, "*Masses of the doubly heavy tetraquarks in a constituent quark model*," Nucl. Phys. A **983** (2019) 1–19, [1809.05257].

[34] Q.-F. Lü, D.-Y. Chen, and Y.-B. Dong, "*Masses of doubly heavy tetraquarks $T_{QQ'}$ in a relativized quark model*," Phys. Rev. D **102** (Aug, 2020) 034012. https://link.aps.org/doi/10.1103/PhysRevD.102.034012.

[35] W.-X. Zhang, H. Xu, and D. Jia, "*Masses and magnetic moments of hadrons with one and two open heavy quarks: Heavy baryons and tetraquarks*," Phys. Rev. D **104** no. 11, (2021) 114011, [2109.07040].

[36] S. Noh, W. Park, and S. H. Lee, "*The Doubly-heavy Tetraquarks ($qq'\bar{Q}\bar{Q}'$) in a Constituent Quark Model with a Complete Set of Harmonic Oscillator Bases*," Phys. Rev. D **103** (2021) 114009, [2102.09614].

[37] M. Sakai and Y. Yamaguchi, "*Analysis of T_{cc} and T_{bb} based on the hadronic molecular model and their spin multiplets*," [2312.08663].

[38] R. M. Albuquerque, S. Narison, and D. Rabetiarivony, "*Z_c-like spectra from QCD Laplace sum rules at NLO*," Phys. Rev. D **103** no. 7, (2021) 074015, [2101.07281].

[39] Z.-G. Wang, "*Landau equation and QCD sum rules for the tetraquark molecular states*," Phys. Rev. D **101** no. 7, (2020) 074011, [2001.04095].

The Z_b states as the mixture of the molecular and diquark-anti-diquark components within the effective field theory

Wei He[1,2],* De-Shun Zhang[1,2],† and Zhi-Feng Sun[1,2,3,4‡]
[1] *School of Physical Science and Technology, Lanzhou University, Lanzhou 730000, China*
[2] *Research Center for Hadron and CSR Physics, Lanzhou University and Institute of Modern Physics of CAS, Lanzhou 730000, China*
[3] *Lanzhou Center for Theoretical Physics, Key Laboratory of Theoretical Physics of Gansu Province, and Key Laboratory of Quantum Theory and Applications of the Ministry of Education, Lanzhou University, Lanzhou, 730000, China*
[4] *Frontiers Science Center for Rare Isotopes, Lanzhou University, Lanzhou, Gansu 730000, China*
(Dated: March 5, 2024)

In this study, we reconsider the states $Z_b(10610)$ and $Z_b(10650)$ by investigating the presence of diquark-anti-diquark components as well as the hadronic molecule components in the framework of effective field theory. The different masses of pseudoscalar mesons such as π^0, η_8, and η_0, as well as vector mesons like ρ^0 and ω violate the OZI rule that is well depicted under the $[U(3)_L \otimes U(3)_R]_{global} \otimes [U(3)_V]_{local}$ symmetry. To account for the contribution of intermediate bosons of heavy masses within the OBE model, we introduce an exponential form factor instead of the commonly used monopole form factor in the past. By solving the coupled-channel Schrödinger equation with the Gaussian expansion method, our numerical results indicate that the $Z_b(10610)$ and $Z_b(10650)$ states can be explained as hadronic molecules slightly mixing with diquark-anti-diquark states.

PACS numbers:

arXiv:2403.02099v1 [hep-ph] 4 Mar 2024

I. INTRODUCTION

In the past decades, a series of quarkonium-like states were discovered. In the $b\bar{b}$ sector, Belle Collaboration reported two charged bottomonium-like states which are known as $Z_b(10610)$ and $Z_b(10650)$ [1] in 2011. Both of them were observed in $\Upsilon(5S) \to \pi^{\pm} h_b(mP)$ $(m = 1, 2)$ and $\Upsilon(5S) \to \pi^{\pm}\Upsilon(nS)$ $(n = 1, 2, 3)$, respectively. Later, Belle confirmed their observations [2, 3]. The next year of the first discovery, the neutral state $Z_b^0(10610)$ was found in the $\Upsilon(5S) \to \Upsilon(2S, 3S)\pi^0\pi^0$ decay [4]. The masses and widths of these states listed in PDG (Particle Data Group) are shown below

$$\begin{aligned} M_{Z_b^{\pm}} &= 10607.2 \pm 2.0 \text{ MeV}, \quad \Gamma_{Z_b^{\pm}} = 18.4 \pm 2.4 \text{ MeV}, \\ M_{Z_b^0} &= 10609 \pm 4.0 \pm 4 \text{ MeV}, \\ M_{Z_b'^{\pm}} &= 10652.2 \pm 1.5 \text{ MeV}, \quad \Gamma_{Z_b'^{\pm}} = 11.5 \pm 2.2 \text{ MeV} \end{aligned}$$

with the quantum numbers $I^G(J^P) = 1^+(1^+)$. For simplicity, here we label the two states $Z_b(10610)$ and $Z_b(10650)$ by Z_b and Z_b', respectively.

Theoretical research had already been performed before the observations of the Z_b states. The authors in Ref. [5, 6] indicated that there may exist a loosely bound S-wave $B\bar{B}^*/B^*\bar{B}$ molecular state.

After the observation, the explanations of the nature of the Z_b states were proposed through different assumptions and theoretical methods. Since the masses of the $Z_b(10610)$ and the $Z_b(10650)$ are close to the $B\bar{B}^*$ and $B^*\bar{B}^*$ thresholds, they are good candidates of $B\bar{B}^*$ and $B^*\bar{B}^*$ molecular states [7–40]. However, tetraquark interpretations including diquark-anti-diquark explanation can not be ruled out [17, 37, 41–47]. As a consequence, in this work we study these two states in the picture of the mixture of molecular and diquark-anti-diquark components, which is used to investigate the nature of the Z_{cs} states observed by LHCb in our previous work [48].

The hadronic molecule has been proposed based on the study of deuteron composed of a proton and a neutron. And this kind of topic has been widely discussed [49–52] within different methods, especially after the observation of $X(3872)$ in 2003 [53]. On the other hand, the concept of the diquark-anti-diquark state was proposed for the first time by Maiani *et al.* [54, 55] following the revitalization of interest on the σ meson. In this work, we use both the molecular state and the diquark-anti-diquark state, and make the calculation in the framework of the effective field theory. In this way, the mesons and the diquarks are viewed as point-like particles, which finally form the color singlet system. The forces between these clusters are provided by exchanging pseudoscalar and vector mesons as well as scalar and axial-vector diquarks.

In order to calculate the effective potentials in the coordinate space, form factors for each vertex are needed, such that the high-momentum contributions are suppressed. In the works related to the One-Boson-Exchange model in the past, the monopole form factor is intro-

*hewei1999@outlook.com
†220220940071@lzu.edu.cn
‡Corresponding Author: sunzf@lzu.edu.cn

duced (see the review [49]). However, in some of our cases, since the exchanged particles' masses are not so small, the monopole form factor does not work very well for it suppressing the corresponding potentials by its numerator. So we introduce the exponentially parameterized form factor and obtain the analytical expressions of the potential in the coordinate space. Considering both the S- and D-wave contributions, we solve the coupled channel Schrödinger equation to see the existence of the composite particles.

The structure of this paper is organized as follows. After the introduction, the theoretical framework is presented in Sec. II. The results and discussion are shown in Sec. III. Finally, a brief summary is given in Sec. IV.

II. FORMALISM

A. Wave functions

We give here the flavour wave functions of the negative and neutral $B\bar{B}^*/B^*\bar{B}$ and $B^*\bar{B}^*$ systems constructed in Ref. [12]

$$\begin{aligned}
|Z^-_{B\bar{B}^*/B^*\bar{B}}\rangle =&\frac{1}{\sqrt{2}}(|B^{*-}B^0\rangle + c\,|B^-B^{*0}\rangle),\\
|Z^0_{B\bar{B}^*/B^*\bar{B}}\rangle =&\frac{1}{2}[|B^{*+}B^-\rangle - |B^{*0}\bar{B}^0\rangle + c(|B^+B^{*-}\rangle\\
&- |B^0\bar{B}^{*0}\rangle)],\\
|Z^-_{B^*\bar{B}^*}\rangle =&|B^{*-}B^{*0}\rangle\,,\\
|Z^0_{B^*\bar{B}^*}\rangle =&\frac{1}{\sqrt{2}}(|B^{*+}B^{*-}\rangle - |B^{*0}\bar{B^{*0}}\rangle).
\end{aligned} \tag{1}$$

The flavour wave functions of diquark and anti-diquark systems are constructed in analogy to the meson-meson systems by swapping b quark and $\bar{b}$ quark,

$$\begin{aligned}
|Z^-_{S\bar{A}/A\bar{S}}\rangle =& \frac{1}{\sqrt{2}}(|\bar{A}_{bu}S_{bd}\rangle + c\,|\bar{S}_{bu}A_{bd}\rangle),\\
|Z^0_{S\bar{A}/A\bar{S}}\rangle =& \frac{1}{2}[|A_{bu}\bar{S}_{bu}\rangle - |A_{bd}\bar{S}_{bd}\rangle + c(|S_{bu}\bar{A}_{bu}\rangle - |S_{bd}\bar{A}_{bd}\rangle)],\\
|Z^-_{A\bar{A}}\rangle =& |A_{bd}\bar{A}_{bu}\rangle\,,\\
|Z^0_{A\bar{A}}\rangle =& \frac{1}{\sqrt{2}}(|A_{bu}\bar{A}_{bu}\rangle - |A_{bd}\bar{A}_{bd}\rangle).
\end{aligned} \tag{2}$$

The value of c depends on the G-parity, i.e., $c = \pm 1$ corresponds to $G = \pm 1$. Here we only pay attention to the situation of $c = +1$, since the G-parity of the considered systems are all $+1$.

In this work, both S- and D-wave interactions between the composed particles are considered. In general, the Z_b and Z_b' states can be expressed as

$$|Z_b\rangle = \begin{pmatrix} Z^-_{B\bar{B}^*/B^*\bar{B}}(^3S_1)\\ Z^-_{B\bar{B}^*/B^*\bar{B}}(^3D_1)\\ Z^-_{B^*\bar{B}^*}(^3S_1)\\ Z^-_{B^*\bar{B}^*}(^3D_1)\\ Z^-_{S\bar{A}/A\bar{S}}(^3S_1)\\ Z^-_{S\bar{A}/A\bar{S}}(^3D_1)\\ Z^-_{A\bar{A}}(^3S_1)\\ Z^-_{A\bar{A}}(^3D_1)\end{pmatrix},\ |Z_b'\rangle = \begin{pmatrix} Z^-_{B^*\bar{B}^*}(^3S_1)\\ Z^-_{B^*\bar{B}^*}(^3D_1)\\ Z^-_{S\bar{A}/A\bar{S}}(^3S_1)\\ Z^-_{S\bar{A}/A\bar{S}}(^3D_1)\\ Z^-_{A\bar{A}}(^3S_1)\\ Z^-_{A\bar{A}}(^3D_1)\end{pmatrix}. \tag{3}$$

Note that for Z_b', $|^5D_1\rangle$ state is forbidden due to its G-even parity.

B. The Effective Lagrangians and Coupling Constants

Next we introduce the interactions of meson-meson and diquark-anti-diquark by constructing the corresponding Lagrangians.

As we all know, each quark is a color triplet resulting in a diquark being a color antitriplet or sextet. The interaction between the two quarks of an antitriplet is attractive, while that of a sextet is repulsive. Consequently, we only consider the effective Lagrangian containing antitriplet. The diquark fields are depicted as

$$S^a = \begin{pmatrix} 0 & S_{ud} & S_{us}\\ -S_{ud} & 0 & S_{ds}\\ -S_{us} & -S_{ds} & 0\end{pmatrix}^a, \tag{4}$$

$$A^a_\mu = \begin{pmatrix} A_{uu} & \frac{1}{\sqrt{2}}A_{ud} & \frac{1}{\sqrt{2}}A_{us}\\ \frac{1}{\sqrt{2}}A_{ud} & A_{dd} & \frac{1}{\sqrt{2}}A_{ds}\\ \frac{1}{\sqrt{2}}A_{us} & \frac{1}{\sqrt{2}}A_{ds} & A_{ss}\end{pmatrix}^a_\mu, \tag{5}$$

$$S^a_b = \begin{pmatrix} S_{bu} & S_{bd} & S_{bs}\end{pmatrix}^a, \tag{6}$$

$$A^a_{b\mu} = \begin{pmatrix} A_{bu} & A_{bd} & A_{bs}\end{pmatrix}^a_\mu, \tag{7}$$

where S^a is the light scalar diquark, A^a_μ the light axial vector diquark, S_b the bottomed scalar diquark, and $A^a_{b\mu}$ the bottomed axial vector diquark. The superscript $a = 1, 2, 3$ is the color index. The meson fields read

$$\Phi = \begin{pmatrix} \frac{\sqrt{3}\pi^0+\eta_8+\sqrt{2}\eta_0}{\sqrt{3}} & \sqrt{2}\pi^+ & \sqrt{2}K^+\\ \sqrt{2}\pi^- & \frac{-\sqrt{3}\pi^0+\eta_8+\sqrt{2}\eta_0}{\sqrt{3}} & \sqrt{2}K^0\\ \sqrt{2}K^- & \sqrt{2}\bar{K}^0 & \frac{-2\eta_8+\sqrt{2}\eta_0}{\sqrt{3}}\end{pmatrix}, \tag{8}$$

$$V_\mu = \frac{g_V}{\sqrt{2}}\begin{pmatrix} \frac{1}{\sqrt{2}}(\rho^0+\omega) & \rho^+ & K^{*+}\\ \rho^- & -\frac{1}{\sqrt{2}}(\rho^0-\omega) & K^{*0}\\ K^{*-} & \bar{K}^{*0} & \phi\end{pmatrix}_\mu, \tag{9}$$

$$P = (B^-, \bar{B}^0, \bar{B}^0_s), \tag{10}$$

$$P^*_\tau = (B^{*-}, \bar{B}^{*0}, \bar{B}^{*0}_s)_\tau \tag{11}$$

with Φ and V_μ the light pseudoscalar and vector, P and P^*_τ the bottomed pseudoscalar and vector.

Considering the $[U(3)_L \otimes U(3)_R]_{global} \otimes [U(3)]_{local}$ symmetry, parity and charge conjugation, the Lagrangians containing mesons and diquarks are shown as follows

$$\begin{aligned}
\mathcal{L}_1 &= -\frac{\beta}{m_P}(iP\hat{\alpha}_{\parallel\mu}D^\mu P^\dagger + h.c.) \\
&\quad -2g(iP\hat{\alpha}_{\perp\mu}P^{*\mu\dagger} + h.c.) \\
&\quad -\frac{g}{m_{P^*}}(\xi^{\mu\nu\alpha\beta}P^*_\nu\hat{\alpha}_{\perp\alpha}D_\mu P^{*\dagger}_\beta + h.c.) \\
&\quad +\frac{\beta}{m_{P^*}}(iP^*_\nu\hat{\alpha}^\mu_\parallel D_\mu P^{*\nu\dagger} + h.c.), \qquad (12) \\
\mathcal{L}_2 &= e_1(iPD_\mu S^a A^{a\mu\dagger}_b - iA^{a\mu}_b D_\mu S^{a\dagger}P^\dagger) \\
&\quad +e_2(iPA^a_\mu D^\mu S^{a\dagger}_b - iD^\mu S^a_b A^{a\dagger}_\mu P^\dagger) \\
&\quad +e_3(\epsilon^{\mu\nu\alpha\beta}PA^a_{\mu\nu}A^{a\dagger}_{b\alpha\beta} + \epsilon^{\mu\nu\alpha\beta}A^a_{b\alpha\beta}A^{a\dagger}_{\mu\nu}P^\dagger) \\
&\quad +e_4(iP^*_\mu D^\mu S^a S^{a\dagger}_b - iS^a_b D^\mu S^{a\dagger}P^{*\dagger}_\mu) \\
&\quad +e_5(\epsilon^{\mu\nu\alpha\beta}P^*_\mu D_\nu S^a A^{a\dagger}_{b\alpha\beta} + \epsilon^{\mu\nu\alpha\beta}A^a_{b\alpha\beta}D_\nu S^{a\dagger}P^{*\dagger}_\mu) \\
&\quad +e_6(\epsilon^{\mu\nu\alpha\beta}P^*_\mu A^a_{\nu\alpha}D_\beta S^{a\dagger}_b + \epsilon^{\mu\nu\alpha\beta}D_\beta S^a_b A^{a\dagger}_{\nu\alpha}P^{*\dagger}_\mu) \\
&\quad +e_7(iP^*_\mu A^{a\mu\nu}A^{a\dagger}_{b\nu} - iA^a_{b\nu}A^{a\mu\nu\dagger}P^{*\dagger}_\mu) \\
&\quad +e_8(iP^*_\mu A^a_\nu A^{a\mu\nu\dagger}_b - iA^{a\mu\nu}_b A^{a\dagger}_\nu P^{*\dagger}_\mu) \\
&\quad +e_9(iP^*_{\mu\nu}A^{a\mu}A^{a\nu\dagger}_b - iA^{a\nu}_b A^{a\mu\dagger}P^{*\dagger}_{\mu\nu}), \qquad (13) \\
\mathcal{L}_3 &= h_1(iS^a_b\hat{\alpha}^{\mu T}_\parallel D_\mu S^{a\dagger}_b - iD_\mu S^a_b\hat{\alpha}^{\mu T}_\parallel S^{a\dagger}_b) \\
&\quad +h_2(\xi^{\mu\nu\alpha\beta}A^a_{b\mu\nu}\hat{\alpha}^{\mu T}_{\parallel\alpha}D_\beta S^{a\dagger}_b + \xi^{\mu\nu\alpha\beta}D_\beta S^a_b\hat{\alpha}^{\mu T}_{\parallel\alpha}A^{a\dagger}_{b\mu\nu}) \\
&\quad +h_3(iA^a_{b\mu}\hat{\alpha}^{\mu T}_\perp S^{a\dagger}_b - iS^a_b\hat{\alpha}^{\mu T}_\perp A^{a\dagger}_{b\mu}) \\
&\quad +h_4(iA^a_{b\mu}\hat{\alpha}^T_{\parallel\nu}A^{a\mu\nu\dagger}_b - iA^{a\mu\nu}_b\hat{\alpha}^T_{\parallel\nu}A^{a\dagger}_{b\mu}) \\
&\quad +h_5(\xi^{\mu\nu\alpha\beta}A^a_{b\mu}\hat{\alpha}^T_{\perp\nu}A^{a\dagger}_{b\alpha\beta} \\
&\quad +\xi^{\mu\nu\alpha\beta}A^a_{b\alpha\beta}\hat{\alpha}^T_{\perp\nu}A^{a\dagger}_{b\alpha\mu}), \qquad (14)
\end{aligned}$$

where

$$\begin{aligned}
D_\mu P &= \partial_\mu P + iP\alpha^\dagger_{\parallel\mu} = \partial_\mu P + iP\alpha_{\parallel\mu}, \qquad (15) \\
D_\mu P^*_\tau &= \partial_\mu P^*_\tau + iP^*_\tau\alpha^\dagger_{\parallel\mu} = \partial_\mu P^*_\tau + iP^*_\tau\alpha_{\parallel\mu}, \qquad (16) \\
\alpha_{\perp\mu} &= (\partial_\mu\xi_R\xi^\dagger_R - \partial_\mu\xi_L\xi^\dagger_L)/(2i), \qquad (17) \\
\alpha_{\parallel\mu} &= (\partial_\mu\xi_R\xi^\dagger_R + \partial_\mu\xi_L\xi^\dagger_L)/(2i), \qquad (18) \\
\hat{\alpha}_{\perp\mu} &= (D_\mu\xi_R\xi^\dagger_R - D_\mu\xi_L\xi^\dagger_L)/(2i), \qquad (19) \\
\hat{\alpha}_{\parallel\mu} &= (D_\mu\xi_R\xi^\dagger_R + D_\mu\xi_L\xi^\dagger_L)/(2i), \qquad (20) \\
\xi_L &= e^{i\sigma/F_\sigma}e^{-i\Phi/(2F_\pi)}, \qquad (21) \\
\xi_R &= e^{i\sigma/F_\sigma}e^{i\Phi/(2F_\pi)}, \qquad (22) \\
A^a_{\mu\nu} &= D_\mu A^a_\nu - D_\nu A^a_\mu, \qquad (23) \\
A^a_{b\mu\nu} &= D_\mu A^a_{b\nu} - D_\nu A^a_{b\mu}, \qquad (24) \\
D_\mu A^a_\nu &= \partial_\mu A^a_\nu - iV_\mu A^a_\nu - iA^a_\nu V^T_\mu, \qquad (25) \\
D_\mu S^a &= \partial_\mu S^a - iV_\mu S^a - iS^a V^T_\mu, \qquad (26) \\
D_\mu A^a_{b\nu} &= \partial_\mu A^a_{b\nu} - iA^a_{b\nu}\alpha^T_{\parallel\mu}, \qquad (27) \\
D_\mu S^a_b &= \partial_\mu S^a_b - iS^a_b\alpha^T_{\parallel\mu}. \qquad (28)
\end{aligned}$$

In Eqs. (13) and (14), the Einstein summation convention is used, i.e., the repeated superscripts "a" mean the summation over them. For $\mathcal{L}_1$, the constants $\beta = 0.9$ and $g = 0.59$. For $\mathcal{L}_2$ and $\mathcal{L}_3$, there are two set of coupling constants e_i $(i = 1, 2, ..., 9)$ and h_j $(j = 1, 2, ..., 5)$ whose values are still unknown. In this work, we naively use the $^3P_0$ model to determine them. Their values are listed in Table I. Note that we can not fix the sign of e_3, e_5 and e_6 because the relative phase between the amplitudes obtained from the Lagrangian and $^3P_0$ model can not be determined. For e_7, e_8 and e_9, we use the phase in Ref. [48] which explains the Z_{cs} well and get the values of them. In Eqs. (21) and (22), we choose $\sigma = 0$ according to the unitary gauge [56].

TABLE I: The values of the low energy constants in the Lagrangians containing diquarks.

e_1 (GeV^{-1})	e_2 (GeV^{-1})	e_3 (GeV^{-2})	e_4 (GeV^{-1})	e_5 (GeV^{-2})
-6.353	1.657	$\pm$0.555	4.885	$\pm$0.566
e_6 (GeV^{-2})	e_7 (GeV^{-1})	e_8 (GeV^{-1})	e_9 (GeV^{-1})	
$\pm$1.005	-0.909	-13.348	11.530	
h_1 (GeV^{-1})	h_2 (GeV^{-1})	h_3 (GeV^{-1})	h_4 (GeV^{-1})	h_5 (GeV^{-1})
0.084	-0.130	0.266	1.457	0.011

C. Effective Potentials With The Exponential Form Factor

Making use of the Breit approximation, we obtain the effective potentials in the momentum space

$$\mathcal{V}^{H_1H_2\to H_3H_4}(\boldsymbol{q}) = \frac{\mathcal{M}^{H_1H_2\to H_3H_4}(\boldsymbol{q})}{\sqrt{\prod_i 2m_i\prod_f 2m_f}} \qquad (29)$$

where m_i $(i = 1, 2, 3, 4)$ denotes the mass of the particle labeled by i. By performing the Fourier transformation, we get the effective potential in the coordinate space

$$\mathcal{V}^{H_1H_2\to H_3H_4}(\boldsymbol{r}) = \int\frac{d^3\boldsymbol{q}}{(2\pi)^3}e^{i\boldsymbol{q}\cdot\boldsymbol{r}}\mathcal{V}^{H_1H_2\to H_3H_4}(\boldsymbol{q})F^2(q^2). \qquad (30)$$

Here, $F(\vec{q}^2)$ is the form factor which suppresses the contribution of high momenta, i.e., small distance. And the presence of such a form factor is dictated by the extended (quark) structure of the hadrons. In this work, we adopt the exponentially parameterized form factor

$$F(q^2) = e^{q^2/\Lambda^2} = e^{(q_0^2-\vec{q}^2)/\Lambda^2} \qquad (31)$$

with Λ the cut-off.

Another option of the form factor is monopole expression, i.e.,

$$F_M(q^2) = \frac{\Lambda^2 - m_E^2}{\Lambda^2 - q^2} \qquad (32)$$

with m_E the mass of the exchanged particle. If the exchanged meson's mass is large, for instance in the case of ϕ- or diquark-exchange, $F_M(q^2)$ is highly suppressed by the numerator $\Lambda^2 - m_E^2$, which would lead to unreasonable results. Consequently, we choose the exponentially parameterized form factor in this work.

In Eqs. (33-51), we list the specific expressions of the non-zero effective subpotentials which are isospin-independent:

$$V_v^{\bar{B}^*B\to\bar{B}^*B/\bar{B}B^*\to\bar{B}B^*} = C_v\frac{(\beta g_V)^2 m_{B^*}m_B}{4M_P M_{P^*}}(\boldsymbol{\epsilon_{1/2}}\cdot\boldsymbol{\epsilon^\dagger_{3/4}}) \times Y(\Lambda,m_v,r), \quad (33)$$

$$V_p^{\bar{B}B^*\to\bar{B}^*B} = C_p\left(\frac{g}{2F_\pi}\right)^2\Big[(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_3})Z(\Lambda,\tilde{m}_p,r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_2},\boldsymbol{\epsilon^\dagger_3})T(\Lambda,\tilde{m}_p,r)\Big], \quad (34)$$

$$V_p^{\bar{B}^*B/\bar{B}B^*\to\bar{B}^*B^*} = \delta_{B\bar{B}^*/B^*\bar{B}}C_p\frac{g^2 m_{B^*}}{4M_{P^*}F_\pi^2} \times\Big[\boldsymbol{\epsilon^\dagger_{4/3}}\cdot(\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{3/4}})Z(\Lambda,\tilde{m}_p,r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_{4/3}},\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{3/4}}) \times T(\Lambda,\tilde{m}_p,r)\Big], \quad (35)$$

$$V_{S_{ud}}^{\bar{B}^*B\to S\bar{A}/\bar{B}B^*\to A\bar{S}} = -\frac{\sqrt{3}}{12}e_1e_4\Big[(\boldsymbol{\epsilon_{1/2}}\cdot\boldsymbol{\epsilon^\dagger_{4/3}})Z(\Lambda,\tilde{m}_{S_{ud}},r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon^\dagger_{4/3}})T(\Lambda,\tilde{m}_{S_{ud}},r)\Big], \quad (36)$$

$$V_{A_{ud}}^{\bar{B}^*B\to S\bar{A}/\bar{B}B^*\to A\bar{S}} = \sqrt{3}e_3e_6m_{S_{bq}}m_{A_{bq}}\Big[\frac{2}{3}(\boldsymbol{\epsilon_{1/2}}\cdot\boldsymbol{\epsilon^\dagger_{4/3}}) \times Z(\Lambda,\tilde{m}_{A_{ud}},r) - \frac{1}{3}S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon^\dagger_{4/3}}) \times T(\Lambda,\tilde{m}_{A_{ud}},r)\Big], \quad (37)$$

$$V_{A_{ud}}^{\bar{B}^*B\to A\bar{S}/\bar{B}B^*\to S\bar{A}} = -\frac{\sqrt{3}}{8}e_2(e_8m_{A_{bq}}+e_9m_{B^*})m_{S_{bq}} \times\left(\frac{\tilde{m}_{A_{ud}}}{m_{A_{ud}}}\right)^2(\boldsymbol{\epsilon_{1/2}}\cdot\boldsymbol{\epsilon^\dagger_{3/4}}) \times Y(\Lambda,\tilde{m}_{A_{ud}},r), \quad (38)$$

$$V_{S_{ud}}^{\bar{B}^*B/\bar{B}B^*\to A\bar{A}} = \delta_{B\bar{B}^*/B^*\bar{B}}\frac{1}{2\sqrt{3}}e_1e_5m_{A_{bq}} \times\Big[\boldsymbol{\epsilon^\dagger_{4/3}}\cdot(\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{3/4}})Z(\Lambda,\tilde{m}_{S_{ud}},r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_{4/3}},\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{3/4}}) \times T(\Lambda,\tilde{m}_{S_{ud}},r)\Big], \quad (39)$$

$$V_{A_{ud}}^{\bar{B}^*B/\bar{B}B^*\to A\bar{A}} = \delta_{B\bar{B}^*/B^*\bar{B}}\sqrt{\frac{1}{3}}e_3e_7m_{A_{bq}} \times\Big[\boldsymbol{\epsilon_{1/2}}\cdot(\boldsymbol{\epsilon^\dagger_{3/4}}\times\boldsymbol{\epsilon^\dagger_{4/3}})Z(\Lambda,\tilde{m}_{A_{ud}},r) -\frac{1}{2}S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_{3/4}},\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{4/3}}) \times T(\Lambda,\tilde{m}_{A_{ud}},r) +\frac{1}{2}S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon^\dagger_{3/4}}\times\boldsymbol{\epsilon^\dagger_{4/3}}) \times T(\Lambda,\tilde{m}_{A_{ud}},r)\Big], \quad (40)$$

$$V_v^{\bar{B}^*B^*\to\bar{B}^*B^*} = C_v\left(\frac{\beta g_V m_{B^*}}{2M_{P^*}}\right)^2(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon^\dagger_3})(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_4}) \times Y(\Lambda,m_v,r), \quad (41)$$

$$V_p^{\bar{B}^*B^*\to\bar{B}^*B^*} = C_p\left(\frac{gm_{B^*}}{2m_{P^*F_\pi}}\right)^2\Big[(\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3})\cdot(\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4}) \times Z(\Lambda,m_p,r)+S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3},\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4}) \times T(\Lambda,m_p,r)\Big], \quad (42)$$

$$V_{S_{ud}}^{\bar{B}^*B^*\to S\bar{A}/\bar{S}A} = -\frac{\delta_{S\bar{A}/A\bar{S}}}{2\sqrt{3}}e_4e_5m_{A_{bq}}\Big[\boldsymbol{\epsilon_{1/2}}\cdot(\boldsymbol{\epsilon_{2/1}}\times\boldsymbol{\epsilon^\dagger_{4/3}}) \times Z(\Lambda,\tilde{m}_{S_{ud}},r)+S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon_{2/1}}\times\boldsymbol{\epsilon^\dagger_{4/3}}) \times T(\Lambda,\tilde{m}_{S_{ud}},r), \quad (43)$$

$$V_{A_{ud}}^{\bar{B}^*B^*\to S\bar{A}/A\bar{S}} = -\delta_{S\bar{A}/A\bar{S}}\frac{1}{2\sqrt{3}}e_6e_7m_{S_{bq}} \times\Big[\boldsymbol{\epsilon_{2/1}}\cdot(\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{4/3}})Z(\Lambda,\tilde{m}_{A_{ud}},r) +\frac{1}{2}S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{2/1}},\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon^\dagger_{4/3}})T(\Lambda,\tilde{m}_{A_{ud}},r) -\frac{1}{2}S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_{4/3}},\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon_{2/1}})T(\Lambda,\tilde{m}_{A_{ud}},r)\Big], \quad (44)$$

$$V_{S_{ud}}^{\bar{B}^*B^*\to A\bar{A}} = \frac{e_5^2m_{A_{bq}}^2}{\sqrt{3}}\Big[(\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3})\cdot(\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4})Z(\Lambda,\tilde{m}_{S_{ud}},r) +S(\boldsymbol{r},\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3},\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4})T(\Lambda,\tilde{m}_{S_{ud}},r)\Big], \quad (45)$$

$$V_{A_{ud}}^{\bar{B}^*B^*\to A\bar{A}} = -\frac{\sqrt{3}e_7^2}{24}\Big[2(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_3})(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon^\dagger_4})Z(\Lambda,\tilde{m}_{A_{ud}},r) -2(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon_2})(\boldsymbol{\epsilon^\dagger_3}\cdot\boldsymbol{\epsilon^\dagger_4})Z(\Lambda,\tilde{m}_{A_{ud}},r) +(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_3})S(\boldsymbol{r},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon^\dagger_4})T(\Lambda,\tilde{m}_{A_{ud}},r) -(\boldsymbol{\epsilon^\dagger_3}\cdot\boldsymbol{\epsilon^\dagger_4})S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon_2})T(\Lambda,\tilde{m}_{A_{ud}},r) +(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon^\dagger_4})S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_2},\boldsymbol{\epsilon^\dagger_3})T(\Lambda,\tilde{m}_{A_{ud}},r) -(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon_2})S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_3},\boldsymbol{\epsilon^\dagger_4})T(\Lambda,\tilde{m}_{A_{ud}},r)\Big] -\frac{\sqrt{3}}{8}(e_8m_{A_{bq}}+e_9m_{B^*})^2\left(\frac{\tilde{m}_{A_{ud}}}{m_{A_{ud}}}\right)^2 \times(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon^\dagger_3})(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_4})Y(\Lambda,\tilde{m}_{A_{ud}},r), \quad (46)$$

$$V_v^{S\bar{A}\to S\bar{A}/A\bar{S}\to A\bar{S}} = C_v\left(\frac{g_V}{2}\right)^2h_1h_4m_{A_{bq}}m_{S_{bq}}(\boldsymbol{\epsilon_{2/1}}\cdot\boldsymbol{\epsilon^\dagger_{4/3}}) \times Y(\Lambda,m_v,r), \quad (47)$$

$$V_p^{S\bar{A}\to A\bar{S}/A\bar{S}\to S\bar{A}} = C_p\left(\frac{h_3}{4F_\pi}\right)^2\Big[(\boldsymbol{\epsilon_{1/2}}\cdot\boldsymbol{\epsilon^\dagger_{4/3}})Z(\Lambda,\tilde{m}_p,r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon^\dagger_{4/3}})T(\Lambda,\tilde{m}_p,r)\Big], \quad (48)$$

$$V_p^{S\bar{A}/A\bar{S}\to A\bar{A}} = \delta_{S\bar{A}/A\bar{S}}C_p\frac{h_3h_5}{4F_\pi^2}m_{A_{bq}} \times\Big[\boldsymbol{\epsilon^\dagger_{3/4}}\cdot(\boldsymbol{\epsilon_{2/1}}\times\boldsymbol{\epsilon^\dagger_{4/3}})Z(\Lambda,\tilde{m}_p,r) +S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon^\dagger_{3/4}},\boldsymbol{\epsilon_{2/1}}\times\boldsymbol{\epsilon^\dagger_{4/3}})T(\Lambda,\tilde{m}_p,r)\Big], \quad (49)$$

$$V_v^{A\bar{A}\to A\bar{A}} = C_v\frac{g_V^2h_4^2m_{A_{bq}}^2}{4}(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon^\dagger_3})(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon^\dagger_4})Y(\Lambda,m_v,r), \quad (50)$$

$$V_p^{A\bar{A}\to A\bar{A}} = C_p\left(\frac{h_5m_{A_{bq}}}{F_\pi}\right)^2\Big[(\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3})\cdot(\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4}) \times Z(\Lambda,m_p,r)+S(\boldsymbol{r},\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon^\dagger_3},\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon^\dagger_4}) \times T(\Lambda,m_p,r)\Big]. \quad (51)$$

In the above equations, the subscript p denotes the light pseudoscalar mesons π, η_8 and η_0, and v denotes the light vectors ρ and ω. The coefficients $C_{\pi^0}=\frac{1}{3}, C_{\eta_8}=-\frac{1}{9}, C_{\eta_0}=-\frac{2}{9}$, $C_\rho=1$, $C_\omega=-1$, $\delta_{B\bar{B}^*}=1$, $\delta_{B^*\bar{B}}=-1$, $\delta_{S\bar{A}}=1$, $\delta_{A\bar{S}}=-1$. And $S(\hat{\boldsymbol{r}},\boldsymbol{a},\boldsymbol{b})=3(\hat{\boldsymbol{r}}\cdot\boldsymbol{a})(\hat{\boldsymbol{r}}\cdot\boldsymbol{b})-\boldsymbol{a}\cdot\boldsymbol{b}$, $q_0=\frac{m_2^2-m_1^2+m_3^2-m_4^2}{2(m_3+m_4)}$, $\tilde{m}_E^2=m_E^2-q_0^2$ with E the exchanged particle. The functions $Y(\Lambda,m,r)$, $Z(\Lambda,m,r)$ and $T(\Lambda,m,r)$ are defined as

$$\begin{aligned}
Y(\Lambda,m,r) &= -\frac{e^{2q_0^2/\Lambda^2}}{(2\pi)^2 r}\frac{\partial}{\partial r}\Bigg\{e^{2m^2/\Lambda^2}\frac{\pi}{2m}\Bigg[e^{mr}\\
&\quad +e^{-mr}-e^{mr}\mathrm{erf}\left(\frac{\Lambda r}{2\sqrt{2}}+\frac{\sqrt{2}m}{\Lambda}\right)\\
&\quad +e^{-mr}\mathrm{erf}\left(\frac{\Lambda r}{2\sqrt{2}}-\frac{\sqrt{2}m}{\Lambda}\right)\Bigg]\Bigg\},\\
Z(\Lambda,m,r) &= \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}Y(\Lambda,m,r),\\
T(\Lambda,m,r) &= r\frac{\partial}{\partial r}\frac{1}{r}\frac{\partial}{\partial r}Y(\Lambda,m,r).
\end{aligned}$$

Taking into account S- and D-wave functions, the products of the polarization vectors in the subpotentials are presented below

$$\left.\begin{array}{c}
\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger\\
\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger\\
(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
-(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)\\
(\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)\cdot(\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)
\end{array}\right\}\longrightarrow\begin{pmatrix}1&0\\0&1\end{pmatrix},$$

$$\left.\begin{array}{c}
S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_2},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)\\
S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1}\times\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger,\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
2(\boldsymbol{\epsilon_2}\cdot\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
2(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_2},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)
\end{array}\right\}\longrightarrow\begin{pmatrix}0&-\sqrt{2}\\-\sqrt{2}&1\end{pmatrix},$$

$$\left.\begin{array}{c}
\boldsymbol{\epsilon_{1/2}}\cdot(\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger\times\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger)\\
\boldsymbol{\epsilon_1}\cdot(\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger)
\end{array}\right\}\longrightarrow\begin{pmatrix}i\sqrt{2}&0\\0&i\sqrt{2}\end{pmatrix},$$

$$\left.\begin{array}{c}
2S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger,\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger\times\boldsymbol{\epsilon_{1/2}})\\
2S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger,\boldsymbol{\epsilon_{1/2}}\times\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
2S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger)\\
2S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_2},\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger\times\boldsymbol{\epsilon_1})\\
S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_{1/2}},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger\times\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)\\
S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon}_{\boldsymbol{3/4}}^\dagger,\boldsymbol{\epsilon_2}\times\boldsymbol{\epsilon_1})
\end{array}\right\}\longrightarrow\begin{pmatrix}0&2i\\2i&-i\sqrt{2}\end{pmatrix},$$

$$\left.\begin{array}{c}
\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon_2},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger\cdot\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger\\
(\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger\cdot\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon_1},\boldsymbol{\epsilon_2})\\
(\boldsymbol{\epsilon_1}\cdot\boldsymbol{\epsilon_2})S(\hat{\boldsymbol{r}},\boldsymbol{\epsilon}_{\boldsymbol{3}}^\dagger,\boldsymbol{\epsilon}_{\boldsymbol{4}}^\dagger)
\end{array}\right\}\longrightarrow\begin{pmatrix}0&0\\0&0\end{pmatrix}.$$

Hereafter, we label the channels $B\bar{B}^*/B^*\bar{B}$, $B^*\bar{B}^*$, $S\bar{A}/A\bar{S}$ and $A\bar{A}$ by CH_1, CH_2, CH_3 and CH_4, respectively. The elements of the total potential matrix $\hat{V}$ is

$$\begin{aligned}
V^{CH_1\to CH_1} &= \frac{1}{2}\Big(V_{\rho,\omega}^{\bar{B}B^*\to\bar{B}B^*}+2V_{\pi,\eta_8,\eta_0}^{\bar{B}B^*\to\bar{B}^*B}\\
&\quad +V_{\rho,\omega}^{\bar{B}^*B\to\bar{B}^*B}\Big), &(52)\\
V^{CH_1\to CH_2} &= \frac{1}{\sqrt{2}}\Big(V_{\pi,\eta_8,\eta_0}^{\bar{B}^*B\to\bar{B}^*B^*}+V_{\pi,\eta_8,\eta_0}^{\bar{B}B^*\to\bar{B}^*B^*}\Big), &(53)\\
V^{CH_1\to CH_3} &= \frac{1}{2}\Big(V_{A_{ud}}^{\bar{B}^*B\to A\bar{S}}+V_{S_{ud},A_{ud}}^{\bar{B}B^*\to A\bar{S}}\\
&\quad +V_{S_{ud},A_{ud}}^{\bar{B}^*B\to S\bar{A}}+V_{A_{ud}}^{\bar{B}^*B\to A\bar{S}}\Big), &(54)\\
V^{CH_1\to CH_4} &= \frac{1}{\sqrt{2}}\Big(V_{S_{ud},A_{ud}}^{\bar{B}^*B\to A\bar{A}}+V_{S_{ud},A_{ud}}^{\bar{B}B^*\to A\bar{A}}\Big), &(55)\\
V^{CH_2\to CH_2} &= V_{\rho,\omega}^{\bar{B}^*B^*\to\bar{B}^*B^*}+V_{\pi,\eta_8,\eta_0}^{\bar{B}^*B^*\to\bar{B}^*B^*}, &(56)\\
V^{CH_2\to CH_3} &= \frac{1}{\sqrt{2}}\Big(V_{S_{ud},A_{ud}}^{\bar{B}^*B^*\to A\bar{S}}+V_{S_{ud},A_{ud}}^{\bar{B}^*B^*\to S\bar{A}}\Big), &(57)\\
V^{CH_2\to CH_4} &= V_{S_{ud},A_{ud}}^{\bar{B}^*B^*\to A\bar{A}}, &(58)\\
V^{CH_3\to CH_3} &= \frac{1}{2}\Big(V_{\rho,\omega}^{S\bar{A}\to S\bar{A}}+2V_{\pi,\eta_8,\eta_0}^{S\bar{A}\to A\bar{S}}\\
&\quad +V_{\rho,\omega}^{A\bar{S}\to A\bar{S}}\Big), &(59)\\
V^{CH_3\to CH_4} &= \frac{1}{\sqrt{2}}\Big(V_{\pi,\eta_8,\eta_0}^{S\bar{A}\to A\bar{A}}+V_{\pi,\eta_8,\eta_0}^{A\bar{S}\to A\bar{A}}\Big), &(60)\\
V^{CH_4\to CH_4} &= V_{\rho,\omega}^{A\bar{A}\to A\bar{A}}+V_{\pi,\eta_8,\eta_0}^{A\bar{A}\to A\bar{A}}. &(61)
\end{aligned}$$

The subscripts mean the following summation

$$V_{E_1,E_2,\cdots}^{CH_i\to CH_j}=\sum_{a=E_1,E_2,\cdots}V_a^{CH_i\to CH_j}. \quad (62)$$

III. RESULTS AND DISCUSSION

With the preparation above, adopting the Gaussian expansion method (GEM) [57], we solve the coupled channel Schrödinger equation to find the bound state solutions,

$$\left(\hat{K}+\hat{M}+\hat{V}\right)\Psi=E\Psi. \quad (63)$$

Here, $\hat{K}=\mathrm{diag}(-\frac{\Delta}{2\mu_1},-\frac{\Delta}{2\mu_2},\cdots)$, $\hat{M}=\mathrm{diag}(0,M_2-M_1,M_3-M_1,\cdots)$. In the central force field problem, the system does not depend on the azimuth and polar angle, so the operator $\Delta=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)$. The coupled channel Schrödinger equation Eq. (63) is symmetric under the following transformation

$$\begin{aligned}
&U\left(\hat{K}+\hat{M}+\hat{V}\right)\Psi=UE\Psi, &(64)\\
\Rightarrow\ &U\left(\hat{K}+\hat{M}+\hat{V}\right)U^{-1}U\Psi=EU\Psi, &(65)\\
\Rightarrow\ &\left(\hat{K}+\hat{M}+U\hat{V}U^{-1}\right)\tilde{\Psi}=E\tilde{\Psi}, &(66)
\end{aligned}$$

where $U=\mathrm{diag}(e^{i\theta_1},e^{i\theta_2},\cdots)$, $\tilde{\Psi}=U\Psi$. Here the word "symmetric" means the energy E does not change under the transformation of the Schrödinger equation. All the U form a reducible Lie group $U(1)\otimes U(1)\otimes\cdots\otimes U(1)$.

TABLE II: The diquark masses we use in effective potentials from Ref. [58]. q indicates u or d quark.

$m_{S_{bq}}$ (GeV)	$m_{A_{bq}}$ (GeV)	$m_{S_{qq}}$ (GeV)	$m_{A_{qq}}$ (GeV)
5.451	5.465	0.691	0.840

In this work, some of the off-diagonal elements of the potential matrix are imaginary, which is not convenient for solving the Schrödinger equation. If we perform the transformation mentioned above, the problem is solved which means that the potential matrix becomes real. The corresponding parameters are chosen as $\theta_1 = \frac{\pi}{2}$, $\theta_2 = 0$, $\theta_3 = \frac{\pi}{2}$, $\theta_4 = 0$, $\cdots$.

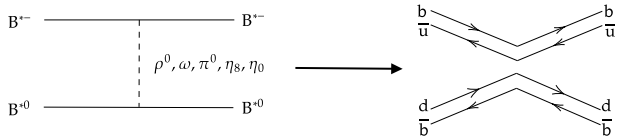


FIG. 1: The diagram of $B^{*-}B^{*0} \to B^{*-}B^{*0}$ on quark and hadron levels.

It is also interesting to discuss the OZI suppressed processes on both the quark and hadron levels. We take the $B^{*-}\bar{B}^{*0} \to B^{*-}\bar{B}^{*0}$ process as an example. On the quark level, we can see from Fig. 1 that the diagram is non-connected, which means it is OZI suppressed. On the hadron level, the exchanged particle could be the pseudoscalar mesons π, η_8, η_0 and the vector mesons ρ, ω. Since the masses of ρ and ω are almost close to each other, their contributions to the total potential cancel out, which can be seen from Eqs. (41) and (56). Under the $[U(3)_L \otimes U(3)_R]_{global} \otimes [U(3)]_{local}$ symmetry, the pseudoscalar mesons appearing in Φ (see Eq. (8)) have the same mass. In this case, The contributions of π, η_8 and η_0 to the total potential also cancel out. This conclusion coincides with the OZI rule. However, the real masses of π, η_8 and η_0 are different, so the total potential is non-zero which depicts the phenomenon of the violation of the OZI rule.

As we mentioned above, in this work, we use the exponentially parameterized form factor in our calculation. For comparison, we take the $B^*\bar{B}^* \to B^*\bar{B}^*$ process as an example, and plot in Fig. 2 the potentials with both the exponential form factor and the monopole one. We notice that

- $V_{\rho^0}^{B^*\bar{B}^* \to B^*\bar{B}^*}$ and $V_{\omega}^{B^*\bar{B}^* \to B^*\bar{B}^*}$ have almost the same absolute value, but different signs, i.e., the contribution of $\rho-$ and $\omega-$exchange is approximately zero. So $V^{B^*\bar{B}^* \to B^*\bar{B}^*}$ is mainly contributed by $\pi-$, η_8- and η_0-exchange;

- the monopole form factor suppresses the contribution from the exchanged particle with heavier mass, which can be clearly seen from the situation of η_0- exchange, since $m_{\eta_0} \simeq 0.96$ GeV;

- in the long interaction range, the pion-exchange subpotential is dominant with both monopole and exponential form factors;

- in the short and medium range, the vector-exchange subpotentials are compatibly larger than that of the pion-exchange with exponential form factor. However, for the monopole form factor, the pion-exchange contribution is still dominant.

The values of the masses of diquarks taken from Ref. [58] are listed in Table II, and the meson-masses are taken from PDG. Then we solve the Schrödinger equation, and the numerical results are presented in Table III. For $B\bar{B}^*/B^*\bar{B}^*/S\bar{A}/A\bar{A}$ and $B^*\bar{B}^*/S\bar{A}/A\bar{A}$ systems, loosely bound states exist when the cut-off is reasonably chosen as $\Lambda \sim 1$ GeV. If the value of Λ increases, the binding energies increase as well while the root-mean-square radii decrease. For both of these two systems, the D-wave contribution is much smaller than that of the S-wave. Besides, the meson-meson component is much larger than the diquark-antidiquark component. That is to say, both $Z_b(10610)$ and $Z_b(10650)$ can be explained as hadronic molecules mixing with a little diquark-antidiquark components.

IV. SUMMARY

In this work, we extend the $[U(3)_L \otimes U(3)_R]_{global} \otimes [U(3)]_{local}$ symmetry to the diquark sector and construct the Lagrangians containing diquarks. In this way, we can introduce the diquark-exchange interactions and study the nature of the Z_b states as the mixture of hadronic molecule and diquark-antidiquark state. In order to make the small distance interaction suppressed, we introduce the exponential form factor instead of the monopole form factor for each vertex, and calculate the effective potentials in the coordinate space analytically. We see that in the short and medium range, $\rho-$ or $omega-$ exchange contribution is larger than that of pion-exchange, while the pion-exchange is dominant in the long range. Besides, $B^{(*)-}\bar{B}^{(*)0} \to B^{(*)-}\bar{B}^{(*)0}$ processes are OZI suppressed which can be easily seen on the quark level. This is well depicted under the $[U(3)_L \otimes U(3)_R]_{global} \otimes [U(3)]_{local}$ symmetry, since the $\pi-$, η_8-, η_0-, $\rho-$ and $\omega-$exchange contributions cancel out. If taking the real masses of the

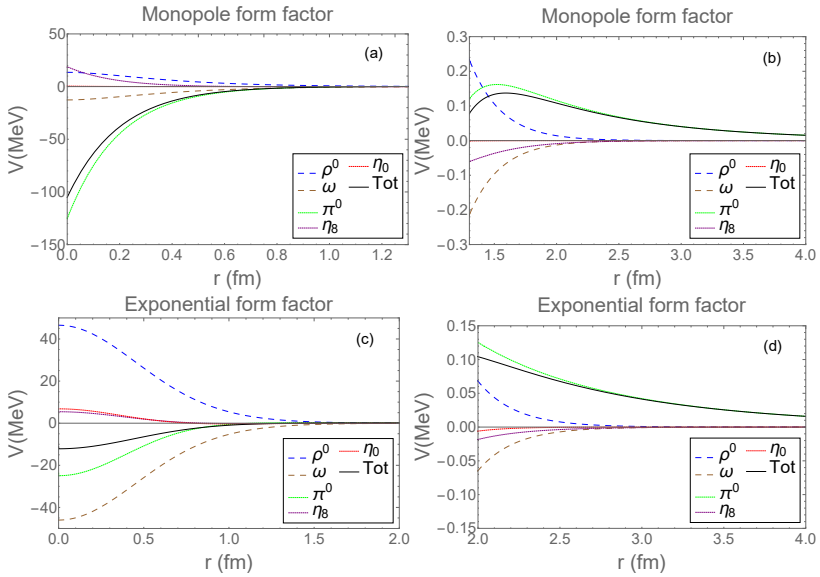


FIG. 2: (Color online) The effective potential for S-wave $B^*\bar{B}^*$ to S-wave $B^*\bar{B}^*$ with different form factor.

TABLE III: The obtained bound state solutions (binding energy E and root-mean-square radius $r_{\rm RMS}$) for the charged Z_b system.

State	Λ(GeV)	E(MeV)	$r_{\rm RMS}$(fm)	$P(M/D)(\%)$	$P(^3\mathbb{S}_1/^3\mathbb{D}_1))(\%)$
$Z_b(10610)$	1.0	-1.83	1.64	97.20/2.80	99.84/0.16
	1.1	-9.96	0.82	92.60/7.40	99.82/0.18
	1.2	-26.14	0.57	87.48/12.52	99.68/0.32
$Z_b(10650)$	1.0	-2.60	1.45	96.37/3.63	99.66/0.34
	1.1	-9.01	0.88	92.62/7.38	99.66/0.34
	1.2	-19.82	0.66	88.55/11.45	99.64/0.36

exchanged particles which are different, the potential is non-zero corresponding to OZI violation.

Taking into account S- and D-wave contributions, we solve the coupled channel Schrödinger equation which obeys the $U(1)\otimes U(1)\otimes ...\otimes U(1)$ symmetry. The numerical results show that both $Z_b(10610)$ and $Z_b(10650)$ can be explained as hadronic molecules slightly mixing with diquark-antidiquark states. If the cut-off is typically taken as $\Lambda = 1$ GeV for both the $Z_b(10610)$ and the $Z_b(10650)$ states, the molecular components are about 97% compared to around 3% diquark-antidiquark state components. Besides, the S-wave contribution is much larger than that of the D-wave. Our work can help to understand the nature of the Z_b states.

Acknowledgments

We would like to thank Zhi-Peng Wang for very useful discussion. This project is supported by the Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2022-sp02, the National Natural Science Foundation of China (NSFC) under Grants No. 11965016, 11705069 and 12247101, and the National Key Research and Development Program of China under Contract No. 2020YFA0406400.

[1] A. Bondar et al. Observation of two charged bottomonium-like resonances in Y(5S) decays. *Phys. Rev. Lett.*, 108:122001, 2012.

[2] A. Garmash et al. Amplitude analysis of $e^+e^- \to \Upsilon(nS)\pi^+\pi^-$ at $\sqrt{s} = 10.865$GeV. *Phys. Rev. D*, 91(7):072003, 2015.

[3] A. Garmash et al. Observation of Zb(10610) and Zb(10650) Decaying to B Mesons. *Phys. Rev. Lett.*, 116(21):212001, 2016.

[4] I. Adachi et al. Evidence for a Zb0(10610) in Dalitz anal-

ysis of Y(5S) -> Y(nS) pi0 pi0. 7 2012.

[5] Yan-Rui Liu, Xiang Liu, Wei-Zhen Deng, and Shi-Lin Zhu. Is $X(3872)$ Really a Molecular State? *Eur. Phys. J. C*, 56:63–73, 2008.

[6] Xiang Liu, Zhi-Gang Luo, Yan-Rui Liu, and Shi-Lin Zhu. X(3872) and Other Possible Heavy Molecular States. *Eur. Phys. J. C*, 61:411–428, 2009.

[7] A. E. Bondar, A. Garmash, A. I. Milstein, R. Mizuk, and M. B. Voloshin. Heavy quark spin structure in Z_b resonances. *Phys. Rev. D*, 84:054010, 2011.

[8] Martin Cleven, Feng-Kun Guo, Christoph Hanhart, and Ulf-G. Meissner. Bound state nature of the exotic Z_b states. *Eur. Phys. J. A*, 47:120, 2011.

[9] J. Nieves and M. Pavon Valderrama. Deriving the existence of $B\bar{B}^*$ bound states from the X(3872) and Heavy Quark Symmetry. *Phys. Rev. D*, 84:056015, 2011.

[10] Jian-Rong Zhang, Ming Zhong, and Ming-Qiu Huang. Could $Z_b(10610)$ be a $B^*\bar{B}$ molecular state? *Phys. Lett. B*, 704:312–315, 2011.

[11] Youchang Yang, Jialun Ping, Chengrong Deng, and Hong-Shi Zong. Possible interpretation of the $Z_b(10610)$ and $Z_b(10650)$ in a chiral quark model. *J. Phys. G*, 39:105001, 2012.

[12] Zhi-Feng Sun, Jun He, Xiang Liu, Zhi-Gang Luo, and Shi-Lin Zhu. $Z_b(10610)^{\pm}$ and $Z_b(10650)^{\pm}$ as the $B^*\bar{B}$ and $B^*\bar{B}^*$ molecular states. *Phys. Rev. D*, 84:054002, 2011.

[13] Shunsuke Ohkoda, Yasuhiro Yamaguchi, Shigehiro Yasui, Kazutaka Sudoh, and Atsushi Hosaka. Exotic Mesons with Hidden Bottom near Thresholds. *Phys. Rev. D*, 86:014004, 2012.

[14] M. T. Li, W. L. Wang, Y. B. Dong, and Z. Y. Zhang. $Z_b(10650)$ and $Z_b(10610)$ States in A Chiral Quark Model. *J. Phys. G*, 40:015003, 2013.

[15] J. M. Dias, F. Aceti, and E. Oset. Study of $B\bar{B}^*$ and $B^*\bar{B}^*$ interactions in $I = 1$ and relationship to the $Z_b(10610)$, $Z_b(10650)$ states. *Phys. Rev. D*, 91(7):076001, 2015.

[16] Q. Wang, V. Baru, A. A. Filin, C. Hanhart, A. V. Nefediev, and J. L Wynen. Line shapes of the $Z_b(10610)$ and $Z_b(10650)$ in the elastic and inelastic channels revisited. *Phys. Rev. D*, 98(7):074023, 2018.

[17] Chun-Yu Cui, Yong-Lu Liu, and Ming-Qiu Huang. Investigating different structures of the $Z_b(10610)$ and $Z_b(10650)$. *Phys. Rev. D*, 85:074014, 2012.

[18] M. B. Voloshin. Radiative transitions from Upsilon(5S) to molecular bottomonium. *Phys. Rev. D*, 84:031502, 2011.

[19] Martin Cleven, Qian Wang, Feng-Kun Guo, Christoph Hanhart, Ulf-G. Meissner, and Qiang Zhao. Confirming the molecular nature of the $Z_b(10610)$ and the $Z_b(10650)$. *Phys. Rev. D*, 87(7):074006, 2013.

[20] Thomas Mehen and Joshua W. Powell. Heavy Quark Symmetry Predictions for Weakly Bound B-Meson Molecules. *Phys. Rev. D*, 84:114013, 2011.

[21] Thomas Mehen and Josh Powell. Line shapes in $\Upsilon(5S) \to B^{(*)}\bar{B}^*\pi$ with $Z(10610)$ and $Z(10650)$ using effective field theory. *Phys. Rev. D*, 88(3):034017, 2013.

[22] Zhi-Gang Wang and Tao Huang. Possible assignments of the $X(3872)$, $Z_c(3900)$ and $Z_b(10610)$ as axial-vector molecular states. *Eur. Phys. J. C*, 74(5):2891, 2014.

[23] Zhi-Gang Wang. Reanalysis of the $Y(3940)$, $Y(4140)$, $Z_c(4020)$, $Z_c(4025)$ and $Z_b(10650)$ as molecular states with QCD sum rules. *Eur. Phys. J. C*, 74(7):2963, 2014.

[24] Yubing Dong, Amand Faessler, Thomas Gutsche, and Valery E. Lyubovitskij. Decays of Zb(+) and Zb'(+) as Hadronic Molecules. *J. Phys. G*, 40:015002, 2013.

[25] Wei Chen, T. G. Steele, Hua-Xing Chen, and Shi-Lin Zhu. Mass spectra of Zc and Zb exotic states as hadron molecules. *Phys. Rev. D*, 92(5):054002, 2015.

[26] Fabian Goerke, Thomas Gutsche, Mikhail A. Ivanov, Jürgen G. Körner, and Valery E. Lyubovitskij. $Z_b(10610)$ and $Z_b'(10650)$ decays in a covariant quark model. *Phys. Rev. D*, 96(5):054028, 2017.

[27] Martin Cleven. *Systematic Study of Hadronic Molecules in the Heavy-Quark Sector*. PhD thesis, Bonn U., 2013.

[28] Shunsuke Ohkoda, Shigehiro Yasui, and Atsushi Hosaka. Decays of $Z_b \to \Upsilon\pi$ via triangle diagrams in heavy meson molecules. *Phys. Rev. D*, 89(7):074029, 2014.

[29] Xin Li and M. B. Voloshin. $Z_b(10610)$ and $Z_b(10650)$ decays to bottomonium plus pion. *Phys. Rev. D*, 86:077502, 2012.

[30] Thomas Gutsche, Valery E. Lyubovitskij, and Ivan Schmidt. Tetraquarks in holographic QCD. *Phys. Rev. D*, 96(3):034030, 2017.

[31] You-Chang Yang, Zhi-Yun Tan, Hong-Shi Zong, and Jialun Ping. Dynamical study of S-wave $\bar{Q}Q\bar{q}q$ system. *Few Body Syst.*, 60(1):9, 2019.

[32] Lu Zhang, Xian-Wei Kang, and Xin-Heng Guo. Composite nature of Z_b states from data analysis. *Eur. Phys. J. C*, 82(4):375, 2022.

[33] Guang-Juan Wang, Xiao-Hai Liu, Li Ma, Xiang Liu, Xiao-Lin Chen, Wei-Zhen Deng, and Shi-Lin Zhu. The strong decay patterns of Z_c and Z_b states in the relativized quark model. *Eur. Phys. J. C*, 79(7):567, 2019.

[34] Shunsuke. Ohkoda, Yasuhiro. Yamaguchi, Shigehiro. Yasui, and Atsushi. Hosaka. Decays and productions via bottomonium for Z_b resonances and other $B\bar{B}$ molecules. *Phys. Rev. D*, 86:117502, 2012.

[35] Feng-Kun Guo, Carlos Hidalgo-Duque, Juan Nieves, and Manuel Pavon Valderrama. Consequences of Heavy Quark Symmetries for Hadronic Molecules. *Phys. Rev. D*, 88:054007, 2013.

[36] V. Baru, E. Epelbaum, A. A. Filin, C. Hanhart, R. V. Mizuk, A. V. Nefediev, and S. Ropertz. Insights into $Z_b(10610)$ and $Z_b(10650)$ from dipion transitions from $\Upsilon(10860)$. *Phys. Rev. D*, 103(3):034016, 2021.

[37] U. Ozdem and K. Azizi. Magnetic dipole moment of $Z_b(10610)$ in light-cone QCD. *Phys. Rev. D*, 97(1):014010, 2018.

[38] Qi Wu, Yuanxin Zheng, Shidong Liu, and Gang Li. Investigations on the light hadron decays of Zb(10610) and Zb(10650). *Phys. Rev. D*, 107(3):034028, 2023.

[39] Jian-Bo Cheng, Bo-Lin Huang, Zi-Yang Lin, and Shi-Lin Zhu. Z_{cs}, Z_c and Z_b states under the complex scaling method. *Eur. Phys. J. C*, 83(11):1071, 2023.

[40] M. B. Voloshin. Mixing model for bottomoniumlike Z_b resonances. *Phys. Rev. D*, 96(9):094024, 2017.

[41] Ahmed Ali, Christian Hambrock, and Wei Wang. Tetraquark Interpretation of the Charged Bottomonium-like states $Z_b^{+-}(10610)$ and $Z_b^{+-}(10650)$ and Implications. *Phys. Rev. D*, 85:054011, 2012.

[42] Angelo Esposito, Andrea L. Guerrieri, Fulvio Piccinini, Alessandro Pilloni, and Antonio D. Polosa. Four-Quark Hadrons: an Updated Review. *Int. J. Mod. Phys. A*, 30:1530002, 2015.

[43] L. Maiani, A. D. Polosa, and V. Riquer. A Theory of X and Z Multiquark Resonances. *Phys. Lett. B*, 778:247–251, 2018.

[44] S. S. Agaev, K. Azizi, and H. Sundu. Spectroscopic parameters and decays of the resonance $Z_b(10610)$. *Eur. Phys. J. C*, 77(12):836, 2017.

[45] Zhi-Gang Wang and Tao Huang. The $Z_b(10610)$ and $Z_b(10650)$ as axial-vector tetraquark states in the QCD sum rules. *Nucl. Phys. A*, 930:63–85, 2014.

[46] A. Ali, L. Maiani, A. D. Polosa, and V. Riquer. Hidden-Beauty Charged Tetraquarks and Heavy Quark Spin Conservation. *Phys. Rev. D*, 91(1):017502, 2015.

[47] Hong-Wei Ke, Xue-Qian Li, Yan-Liang Shi, Guo-Li Wang, and Xu-Hao Yuan. Is $Z_b(10610)$ a Molecular State? *JHEP*, 04:056, 2012.

[48] Ze-Hua Cao, Wei He, and Zhi-Feng Sun. Zcs states and the mixture of hadronic molecule and diquark-anti-diquark components within effective field theory. *Phys. Rev. D*, 107(1):014017, 2023.

[49] Hua-Xing Chen, Wei Chen, Xiang Liu, and Shi-Lin Zhu. The hidden-charm pentaquark and tetraquark states. *Phys. Rept.*, 639:1–121, 2016.

[50] Hua-Xing Chen, Wei Chen, Xiang Liu, Yan-Rui Liu, and Shi-Lin Zhu. An updated review of the new hadron states. *Rept. Prog. Phys.*, 86(2):026201, 2023.

[51] Xiang-Kun Dong, Feng-Kun Guo, and Bing-Song Zou. A survey of heavy–heavy hadronic molecules. *Commun. Theor. Phys.*, 73(12):125201, 2021.

[52] Lu Meng, Bo Wang, Guang-Juan Wang, and Shi-Lin Zhu. Chiral perturbation theory for heavy hadrons and chiral effective field theory for heavy hadronic molecules. *Phys. Rept.*, 1019:1–149, 2023.

[53] S. K. Choi et al. Observation of a narrow charmonium-like state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays. *Phys. Rev. Lett.*, 91:262001, 2003.

[54] L. Maiani, F. Piccinini, A. D. Polosa, and V. Riquer. A New look at scalar mesons. *Phys. Rev. Lett.*, 93:212002, 2004.

[55] G. 't Hooft, G. Isidori, L. Maiani, A. D. Polosa, and V. Riquer. A Theory of Scalar Mesons. *Phys. Lett. B*, 662:424–430, 2008.

[56] Masayasu Harada and Koichi Yamawaki. Hidden local symmetry at loop: A New perspective of composite gauge boson and chiral phase transition. *Phys. Rept.*, 381:1–233, 2003.

[57] E. Hiyama, Y. Kino, and M. Kamimura. Gaussian expansion method for few-body systems. *Prog. Part. Nucl. Phys.*, 51:223–307, 2003.

[58] J. Ferretti. Effective Degrees of Freedom in Baryon and Meson Spectroscopy. *Few Body Syst.*, 60(1):17, 2019.

Azimuthal Angular Correlation of J/ψ Plus Jet Production at the EIC

Luca Maxia[1, *] and Feng Yuan[2, †]

[1] *Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

[2] *Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

By investigating the soft gluon radiation in the J/ψ plus jet photoproduction at the electron-ion collider (EIC), we demonstrate that the azimuthal angular correlations between the leading jet and heavy quarkonium provide a unique probe to the production mechanism of the latter. In particular, a significant $\cos(\phi)$ asymmetry is found for the color-singlet channel, whereas it vanishes or has an opposite sign for color-octet production, depending on the jet transverse momentum. Numerical results of $\cos(\phi)$ and $\cos(2\phi)$ asymmetries employing both the color-singlet model and the nonrelativistic QCD approach are presented for typical kinematics at the future EIC.

arXiv:2403.02097v1 [hep-ph] 4 Mar 2024

Introduction. In recent years, heavy quarkonium production in various inclusive processes has attracted great interest as a way to probe gluon distributions both in initial (nucleon tomography) and final (fragmentation functions) states [1–17]. Among them, Refs. [10, 12] have studied the azimuthal angular correlation in semi-inclusive DIS between J/ψ and leading jet to probe the so-called linearly polarized gluon distribution. In this paper, we will investigate the dominant contributions from the soft gluon radiations and demonstrate that azimuthal correlations can also provide a unique opportunity to disentangle between the color-singlet (CS) and color-octet (CO) mechanisms.

In the nonrelativistic QCD (NRQCD) [18] approach, the heavy-quark pair forms a Fock state specified by $n = {}^{2S+1}L_J^{[c]}$, with S denoting its spin, L the orbital angular momentum, J the total angular momentum and c its color. Note that, within this framework, the pair can couple either as a CS or CO state. Therefore, comprehending the significance of the CS and CO contributions is crucial. Although great progress has been made in understanding heavy quarkonium production in hadronic collisions (for recent reviews see [19, 20]), challenges remain to phenomenologically describe quarkonium formation and, in particular, to distinguish between these two mechanisms.

In this regard, we will demonstrate how azimuthal angular correlations in J/ψ plus jet photoproduction at the electron-ion collider (EIC) offer a unique probe of the underlying production mechanism. In particular, we will show that these correlations significantly differ between the CS and CO channels. We will focus on the correlation kinematics, i.e., the transverse momentum of individual particles is much larger than the total transverse momentum. Therefore, by combining the transverse momenta of the J/ψ and the jet, $k_{\psi\perp}$ and $k_{j\perp}$ respectively, we can identify two scales. The first one is given by $\vec{P}_\perp = \frac{\vec{k}_{\psi\perp} - \vec{k}_{j\perp}}{2}$, while the second by $\vec{q}_\perp = \vec{k}_{\psi\perp} + \vec{k}_{j\perp}$, with $|\vec{q}_\perp| \ll |\vec{P}_\perp|$. Hence, according to this limit, the heavy quarkonium and jet are mainly produced back-to-back in the transverse plane (see Fig. 1). An imbalance between the two final-state particles with nonzero $|\vec{q}_\perp|$ can be generated by high-order perturbative corrections and from the intrinsic transverse momentum of the incoming parton. We identify this imbalance with the angle ϕ, namely the difference between the azimuthal angles of $\vec{q}_\perp$ and $\vec{k}_{\psi\perp}$, where we can approximate the latter as $\vec{k}_{\psi\perp} \approx \vec{P}_\perp$ within the correlation limit.

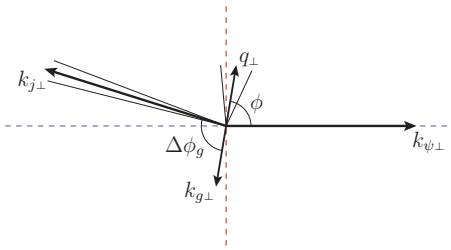


FIG. 1. *Kinematic correlation between the leading jet and heavy quarkonium as viewed in the transverse plane. Here $q_\perp$ (the total outgoing transverse momentum) is small compared to individual transverse momenta.*

Moreover, we remark that in this limit such azimuthal imbalance is mostly generated from the soft/collinear gluon radiation from perturbative diagrams (see for instance Ref. [21, 22]). This contribution, denoted by $k_{g\perp}$ in Fig. 1, tends to align with the jet direction at low $q_\perp$, which leads to significant $\cos(n\phi)$ asymmetries. Detailed examples have been shown for vector boson (photon/Z/Higgs) plus jet production in pp collisions [23, 24] and for lepton plus jet [21] and dijet [22] productions in ep collisions. Nonetheless, this feature is also particularly interesting to investigate the production mechanism of quarkonia.

With this paper, we suggest exploiting azimuthal angular distributions, especially the $\cos(\phi)$ and $\cos(2\phi)$, to unravel the production mechanism. Such findings can

* l.maxia@rug.nl
† fyuan@lbl.gov

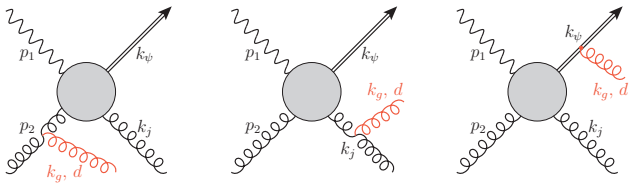


FIG. 2. *Soft gluon radiation in the J/ψ plus jet photoproduction process. All three diagrams contribute to the color-octet channel, whilst only the first two on the left are relevant for the color-singlet one.*

then be applied to electroproduction to investigate gluon distributions (e.g. linearly polarized gluons) in nucleons and nuclei.

Soft gluon radiation at one loop. In this section we discuss the implication of azimuthal correlation between the J/ψ and jet for photoproduction at the future EIC, $\gamma p \to J/\psi + jet + X$. The leading-order (LO) NRQCD contribution comes from the photon-gluon fusion partonic process, namely

$$\gamma(p_1) + g(p_2) \to [Q\bar{Q}]^{(1,8)}(k_\psi) + g(k_j) \ , \tag{1}$$

where the momentum of each particle is given in the parenthesis. Moreover, the heavy-quark pair $Q\bar{Q}$ can be either in a CS (labeled with (1)) or CO (labeled with (8)) configuration. At this order, the J/ψ and jet are back-to-back in the transverse plane, so that $q_\perp = 0$. However, at higher orders, small nonzero $q_\perp$ originates from parton intrinsic transverse momenta and soft gluon radiation. In the following, we will derive the LO soft gluon radiation contribution and the associated azimuthal angular asymmetries (Fig. 2), whereas the collinear gluon radiation can be factorized into the TMD gluon distributions.

The major difference between CS and CO channels is that the soft gluon radiation associated with the heavy quark pair only contributes to the latter. This occurs due to cancellations between the emissions from the heavy quark and antiquark when the pair is in a CS state. This difference has significant implications for the azimuthal asymmetries, as we will discuss in the following.

By adding Fig. 2a and Fig. 2b we obtain the amplitude squared (averaged over the color and spin of incoming particles) for the soft gluon radiation in the CS channel:

$$\overline{|\mathcal{A}_1^{(1)}|^2} = g_s^2 C_A \, S_g(p_2, k_j) \, \overline{|\mathcal{A}_0^{(1)}|^2} \ , \tag{2}$$

where $A_0^{(1)}$ is the LO amplitude and $S_g(v_a, v_b)$ is a shorthand notation for

$$S_g(v_a, v_b) = \frac{2\,(v_a \cdot v_b)}{(v_a \cdot k_g)(v_b \cdot k_g)} \ . \tag{3}$$

Integration over the phase space of the emitted soft gluon follows that in Ref. [21, 22, 25–28], leading to

$$\begin{aligned} &\int \frac{\mathrm{d}^3 k_g}{(2\pi)^3 2E_{k_g}} \overline{|\mathcal{A}_1^{(1)}|^2} \, \delta^{(2)}(q_\perp + k_{g\perp}) \\ &= \frac{\alpha_s C_A}{2\pi^2 |\vec{q}_\perp|^2} \overline{|\mathcal{A}_0^{(1)}|^2} \left[\ln \frac{\hat{s}}{|\vec{q}_\perp|^2} + \ln \frac{\hat{t}}{\hat{u}} + I_j(R, \phi) \right] \ , \end{aligned} \tag{4}$$

where $\hat{s} = (p_1 + p_2)^2$, $\hat{t} = (p_2 - k_j)^2$ and $\hat{u} = (p_1 - k_j)^2$. The first term in the bracket is the double logarithm (in b_T-space) that provides the dominant behavior at low $q_\perp$. The second logarithm, $\ln(\hat{t}/\hat{u})$, depends on the jet rapidity. The third term, I_j, is one of the subjects of this work, being the azimuthal distribution that arises from the soft gluon radiation. As a result of the removal of collinear divergences already included within the jet function, I_j depends on the jet size R. To further investigate this distribution, we expand it in a Fourier series according to

$$I_j(R, \phi) = C_0^{(j)}(R) + 2 \sum_{n=1}^{\infty} C_n^{(j)}(R) \, \cos(n\phi) \ . \tag{5}$$

In line with previous works, we find that the soft gluon radiation associated with the jet leads to a dominant $\cos(\phi)$ asymmetry, with the $\cos(2\phi)$ one being much smaller.

At variance with the CS case, when the pair forms a CO state all diagrams in Fig. 2 are relevant. Independently from the color structures of the LO partonic process, one gluon radiation leads to the same contribution. While it is straightforward to prove that the first two diagrams do not depend on the Fock state, it is worthwhile to provide further elaboration on the last one (Fig. 2c), which corresponds to soft gluon radiation from the heavy-quark pair itself. Within NRQCD, the total soft gluon radiation from the quark and antiquark lines is given by $g_s \frac{k_\psi^\mu}{k_g \cdot k_\psi + i\epsilon} \left(\Phi_{ij} T^d - T^d \Phi_{ij} \right) \Gamma$ [13, 29], where Φ represents the wave function of the pair, i (j) and d are respectively the color indices for the quark (antiquark) and the radiated soft gluon, μ is connected to the soft gluon polarization vector, and Γ stands for other hard parts coming from the scattering amplitude. The minus sign in the second term comes from the interaction with the antiquark which is opposite to that with the quark. When the pair forms a CO state, the color matrix of the wave function can be parameterized as $\Phi_{ij}^{(8,c)} \propto T_{ij}^c$, where $c = 1, \cdots, 8$ represents the color index of the pair. The sum of the two diagrams will then be proportional to $g_s \frac{k_\psi^\mu}{k_\psi \cdot k_g + i\epsilon} (-i f_{c'cd}) \Phi^{(8,c)} \Gamma^{c'}$, where we emphasized that the pair coupled with the rest of the diagram with a different color index c' before the soft gluon emission. Note that we have not taken any assumption on the spin and angular configuration entering in Φ, which implies that the result is independent of those quantum numbers. Moreover, the found result is similar to that for soft gluon radiation from a gluon jet in the final state [26]. The only difference originates from the gluon on-shell condition, $k^2 = 0$ for jets and $k^2 = M_V^2 \approx 4M_Q^2$ for quarkonia.

By including the soft gluon radiations from the initial and final state gluons (Figs. 2a and 2b), the averaged CO amplitude squared is summarized as follows

$$|\overline{\mathcal{A}_1^{(8)}}|^2 = g_s^2 C_A \Big[S_g(p_2,k_j) + \frac{1}{2}\Big(S_g(p_2,k_\psi) - S_g(k_\psi,k_\psi) + S_g(k_j,k_\psi) - S_g(p_2,k_j)\Big)\Big] |\overline{\mathcal{A}_0^{g,(8)}}|^2 \,, \tag{6}$$

where $\mathcal{A}_0^{g,(8)}$ represents the LO (gluon) amplitude. In principle, $|\overline{\mathcal{A}_1^{(8)}}|^2$ depends on quarks too. However, we will neglect the quark contribution in the following, since it is reasonable to expect that it will be suppressed compared to the gluon one. Note that the first term of Eq. (6) is equivalent to that in Eq. (2) for the CS case and, therefore, is a contribution purely driven by the gluon jet. Consequently, all the terms in the round brackets can be associated with the J/ψ.

From Eq. (6), the integration over the phase space of the emitted soft gluon gives

$$\begin{aligned}
&\int \frac{\mathrm{d}^3 k_g}{(2\pi)^3 2E_{k_g}} |\overline{\mathcal{A}_1^{(8)}}|^2 \, \delta^{(2)}(q_\perp + k_{g\perp}) \\
&= \frac{\alpha_s C_A}{2\pi^2 |\vec{q}_\perp|^2} |\overline{\mathcal{A}_0^{g,(8)}}|^2 \Bigg[\ln\frac{\hat{s}}{|\vec{q}_\perp|^2} + \frac{1}{2}\ln\frac{1-M_\psi^2/\hat{u}}{1-M_\psi^2/\hat{t}} \\
&+ I_j(R,\phi) + I_\psi(m_{\psi\perp},\phi) + \frac{1}{2} I_{\psi\text{-}j}(m_{\psi\perp},\Delta y, 2\phi) \\
&- \frac{1}{2} I_\psi^{\mathrm{jet}}(R, m_{\psi\perp}, \Delta y, \phi)\Bigg] \,,
\end{aligned} \tag{7}$$

where we have introduced $m_{\psi\perp} = M_\psi/k_{\psi\perp}$, while $\Delta y = y_\psi - y_j$ is the rapidity difference between the J/ψ (y_ψ) and the leading jet (y_j). The first term of Eq. (7) corresponds to the leading, double logarithmic behavior at low $q_\perp$, which is the same as the CS case. This implies that the soft gluon emission from the (massive) quarkonium does not provide double logarithms, a conclusion in line with other works [13, 30–34]. The second term comes from the sum of rapidities $y_j + y_\psi$, which differs from zero due to the presence of the J/ψ mass. The third term, I_j, is the same azimuthal dependence associated with the jet in Eq. (4), and its Fourier expansion follows Eq. (5). The last three azimuthal distributions, I_ψ, $I_{\psi\text{-}j}$ and I_ψ^{jet}, are the novel terms due to the production of a CO state. Their definitions and derivations are in the supplemental material. Again, we expand them in terms of $\cos(n\phi)$ harmonics according to

$$I_{[\psi]}(K,\phi) = C_0^{([\psi])}(K) + 2\sum_{n=1}^{\infty} C_n^{([\psi])}(K)\,\cos(n\phi) \,, \tag{8}$$

where $I_{[\psi]}$ is either I_ψ, $I_{\psi\text{-}j}$ or I_ψ^{jet}, and K is a short-hand notation for the dependence on the kinematical variables of each distribution.

Mechanism	$C_0^{(c)}$	$C_1^{(c)}$	$C_2^{(c)}$
CS	0.89	2.61	0.95
CO ($m_{\psi\perp} = 0.26$)	2.46	−0.02	1.72
CO ($m_{\psi\perp} = 0.1$)	3.40	−1.80	3.28

TABLE I. *First coefficients of the azimuthal correlation Fourier expansions for the CS and CO mechanisms with $R = 0.4$. Note that the coefficients within the CS channel are independent of $k_{j\perp}$.*

Summarizing the above results, the differential cross section including the soft gluon radiation in the correlation limit is given by

$$\frac{d^4\sigma^{(1,8)}}{d\Omega} = \sigma_0^{(1,8)}\, x_g\, f_g(x_g)\, \frac{\alpha_s C_A}{2\pi^2|\vec{q}_\perp|^2}\Bigg[\ln\frac{\hat{s}}{|\vec{q}_\perp|^2} + 2\sum_{n=0}^{\infty} C_n^{(1,8)}(R, m_{\psi\perp}, \Delta y = 0)\,\cos(n\phi)\Bigg] \,, \tag{9}$$

where $\sigma_0^{(1,8)}$ represents the LO cross section [35], $\mathrm{d}\Omega \equiv \mathrm{d}y_j \mathrm{d}y_\psi \mathrm{d}^2\vec{P}_\perp \mathrm{d}^2\vec{q}_\perp$ the phase space, and $f_g(x_g)$ is the gluon distribution. Moreover, we are considering the case where $y_j = y_\psi = 0$ for simplicity, since this scenario does not present additional, non-divergent logarithms. Hence, in Eq. (9) we have that the first term corresponds to the leading double logarithm, while the second line stands for the single logarithms. Since the latter carries the azimuthal dependence, we have already expanded it into harmonics, with the coefficients $C_n^{(1,8)}$ connected to those in the Fourier series of each distribution. Therefore they depend on the jet size R, the heavy quarkonium mass via $m_{\psi\perp} = M_\psi/P_\perp$ and the rapidity difference $\Delta y = y_\psi - y_j$. Most importantly, these coefficients depend crucially on the production channel considered. To illustrate these differences, we consider typical kinematics for future EIC measurements. We take $R = 0.4$ and $m_{\psi\perp} = 0.26,\ 0.1$, that correspond to $k_{j\perp} \approx 12,\ 30$ GeV. Within these choices, the first three coefficients of the Fourier expansion in Eq. (9) are given in Table I, displaying a clear difference between CS and CO channels. In particular, the coefficient $C_1^{(1)}$ is significantly positive due to the soft gluon radiation associated with the jet in the final state. However, the soft gluon radiation associated with heavy quark pair has an opposite sign. Thus, a proper choice of $m_{\psi\perp}$ can reduce the magnitude, or even change the sign, of $C_1^{(8)}$ when compared to $C_1^{(1)}$, while on the other hand $C_2^{(8)}$ and $C_2^{(1)}$ stay of more or less the same order. For more details on the comparison between the CS and CO channels we refer the reader to the supplemental materials.

All order resummation and predictions for the EIC. All order resummation is needed to make reliable

predictions for the soft gluon radiation contributions. Following the standard TMD framework, we have

$$\frac{d^4\sigma^{(1,8)}}{d\Omega} = \sigma_0^{(1,8)} \int \frac{|\vec{b}_\perp|\, \mathrm{d}|\vec{b}_\perp|}{(2\pi)} \Big[J_0(|\vec{b}_\perp||\vec{q}_\perp|)\, \widetilde{W}_0^{(1,8)}(|\vec{b}_\perp|) + 2\, J_n(|\vec{b}_\perp||\vec{q}_\perp|)\, \widetilde{W}_n^{(1,8)}(|\vec{b}_\perp|)\, \cos(n\phi) \Big] \,, \quad (10)$$

where

$$\widetilde{W}_0^{(1,8)}(|\vec{b}_\perp|) = x_g\, f_g(x_g,\mu_b)\, \mathrm{e}^{-S^{(1,8)}(P_\perp,b_\perp)} \,, \quad (11)$$

and

$$\widetilde{W}_n^{(1,8)}(|\vec{b}_\perp|) = \frac{C_A \alpha_s}{n\pi}\, C_n^{(1,8)}\, \widetilde{W}_0^{(1,8)}(|\vec{b}_\perp|) \,. \quad (12)$$

Note that in Eq. (10) we have already included higher-order double logarithmic corrections present also for the angular dependent term [23, 24]. The Sudakov form factor $S(P_\perp, b_\perp)$ is separated into perturbative and non-perturbative parts: $S(P_\perp,b_\perp) = S_{\rm pert.}(P_\perp,b_\perp) + S_{\rm NP}(P_\perp,b_\perp)$. The perturbative part at one loop is defined as

$$S_{\rm pert.}^{(1,8)} = \int_{\mu_{b_*}^2}^{\hat{s}} \frac{d\mu^2}{\mu^2} \frac{\alpha_s C_A}{2\pi} \left[\ln \frac{\hat{s}}{\mu^2} - 2\,\beta_0 + 2\, C_0^{(1,8)} \right] \,, \quad (13)$$

where $\beta_0 = 11/12 - N_f/18$ and $\mu_b = b_0/b_\perp$ with $b_0 = 2\,\mathrm{e}^{\Gamma_E}$ and Γ_E being the Euler's constant. Moreover, in Eq. (13) we have already introduced the b_*-prescription [36], where $b_* = b_\perp/\sqrt{1+(b_\perp/b_{\rm max})^2}$ with $b_{\rm max} = 1.5$ GeV. For the non-perturbative part, we have contributions driven by the incoming gluon and the outgoing jet and quarkonium. For the first one, we employ the non-perturbative Sudakov found for TMD quark distributions in Refs. [37, 38] with the appropriate Casimir scaling

$$S_{\rm NP}^g = \frac{C_A}{C_F} \left[0.106\, b_\perp^2 + 0.42 \ln \frac{P_\perp}{Q_0} \ln \frac{b_\perp}{b_*} \right] \,, \quad (14)$$

where $Q_0^2 = 2.4$ GeV^2. For the others, we assume that the non-perturbative contribution associated with the jet is given by $S_{\rm NP}^{\rm jet} = g_\Lambda^{\rm jet}\, b_\perp^2$, and the J/ψ one by $S_{\rm NP}^\psi = g_\Lambda^\psi\, b_\perp^2$. Overall, we employ $S_{\rm NP}^{(1)} = S_{\rm NP}^g + S_{\rm NP}^{\rm jet}$ for the CS mechanism and $S_{\rm NP}^{(8)} = S_{\rm NP}^g + S_{\rm NP}^{\rm jet} + S_{\rm NP}^\psi$ for the CO channel, with $g_\Lambda^{\rm jet} = g_\Lambda^\psi = 0.225$ GeV^2, a choice in line with [21, 22]. We also checked that the final azimuthal asymmetries do not depend significantly on these parameters, as expected.

From the above expressions, we find that the azimuthal asymmetries of $\cos(\phi)$ and $\cos(2\phi)$ are linearly proportional to the respective $C_{1,2}$ coefficients which are different between the CS and CO channels. Therefore, they can be used to probe these two production mechanisms.

As an example, in Fig. 3, we show the numerical results for $\langle\cos(\phi)\rangle$ and $\langle\cos(2\phi)\rangle$ as functions of $q_\perp$ for kinematics accessible at the EIC, with $\sqrt{s}_{\gamma p} = 100$ GeV, $R = 0.4$ and $|\vec{k}_{j\perp}| = 12,\ 30$ GeV. For the parton distribution function we have employed the MSHT20 LO set [39]. Predictions are shown for both the color-singlet model (CSM) and NRQCD approach, with the latter mainly depending on the relative fraction of the CS and total CO contributions. Indeed, note that NRQCD predictions are not strongly dependent on the individual CO long-distance matrix elements (LDMEs), as can be understood from the bands in Fig. 3 which are constructed by combining several central values of the associated LDMEs from different global analyses [40–44].

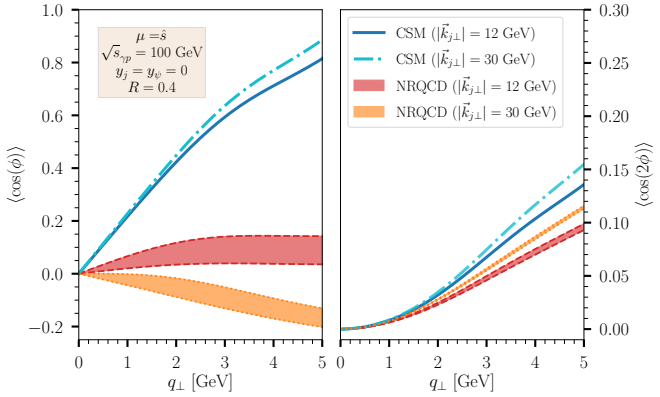


FIG. 3. *Averaged resummed azimuthal asymmetries for J/ψ plus jet photoproduction at $\sqrt{s}_{\gamma p} = 100$ GeV. Solid blue and dash-dotted cyan lines are the CSM predictions at $|\vec{k}_{j\perp}| = 12$ GeV and $|\vec{k}_{j\perp}| = 30$ GeV, whereas for the NRQCD approach we have the dashed red and dotted orange lines, respectively. Bands are obtained by combining the results from different LDMEs central values. Jet size is $R = 0.4$ in all cases.*

As expected, the resummed form of $\langle\cos(n\phi)\rangle$ is exactly 0 at $q_\perp = 0$, and at small $q_\perp$ they scale as $\langle\cos(n\phi)\rangle \propto q_\perp^n$ [24]. At large $q_\perp$, we expect they will decrease as functions of $q_\perp$, where the hard gluon radiation dominates and there is no preferred direction. This will modify the behavior of both $\langle\cos(\phi)\rangle$ and $\langle\cos(2\phi)\rangle$ asymmetries at large $q_\perp$ in Fig. 3. Meanwhile, this figure shows a distinct behavior between $\langle\cos(\phi)\rangle$ and $\langle\cos(2\phi)\rangle$ predictions. Indeed, the former highly depends on the model considered, whereas the CSM and NRQCD outcomes are comparable, and in general non-negligible, for $\langle\cos(2\phi)\rangle$. As demonstrated, such differences offer a unique opportunity to determine the underlying quarkonium production mechanism.

Conclusions. In summary, we have demonstrated that the soft gluon radiation leads to significantly different azimuthal angular correlations between the CS and CO mechanisms in the J/ψ plus jet photoproduction process at the future EIC. This directly affects azimuthal asymmetry predictions within the CSM and the NRQCD approach, where the latter presents a soft dependence on the LDME. Thus, we consider these observables as the stems to disentangle these two production mechanisms. We also expect that similar conclusions can be drawn for

other experiments as well, like in hadronic processes at the LHC. Moreover, $\cos(2\phi)$ has been proposed to study the linearly polarized gluon distribution in the electroproduction of J/ψ plus jet at the EIC. We expect that the soft gluon radiation will lead to sizable contributions to $\cos(2\phi)$ asymmetry in this process as well, with a behavior similar to that shown in Fig. 3 for photoproduction. Understanding the soft gluon radiation will be a crucial step to unambiguously determine the gluon tomography of linearly polarized distribution from this measurement. We will explore all these physics, including higher-order perturbative and power corrections, in the future.

ACKNOWLEDGMENTS

We thank Prof. D. Boer for the valuable discussions. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under contract numbers DE-AC02-05CH11231.

[1] R. M. Godbole, A. Misra, A. Mukherjee, and V. S. Rawoot, Phys. Rev. D **85**, 094013 (2012), arXiv:1201.1066 [hep-ph].
[2] D. Boer and C. Pisano, Phys. Rev. D **86**, 094007 (2012), arXiv:1208.3642 [hep-ph].
[3] R. M. Godbole, A. Misra, A. Mukherjee, and V. S. Rawoot, Phys. Rev. D **88**, 014029 (2013), arXiv:1304.2584 [hep-ph].
[4] W. J. den Dunnen, J. P. Lansberg, C. Pisano, and M. Schlegel, Phys. Rev. Lett. **112**, 212001 (2014), arXiv:1401.7611 [hep-ph].
[5] A. Mukherjee and S. Rajesh, Phys. Rev. D **93**, 054018 (2016), arXiv:1511.04319 [hep-ph].
[6] A. Mukherjee and S. Rajesh, Phys. Rev. D **95**, 034039 (2017), arXiv:1611.05974 [hep-ph].
[7] A. Mukherjee and S. Rajesh, Eur. Phys. J. C **77**, 854 (2017), arXiv:1609.05596 [hep-ph].
[8] S. Rajesh, R. Kishore, and A. Mukherjee, Phys. Rev. D **98**, 014007 (2018), arXiv:1802.10359 [hep-ph].
[9] F. Scarpa, D. Boer, M. G. Echevarria, J.-P. Lansberg, C. Pisano, and M. Schlegel, Eur. Phys. J. C **80**, 87 (2020), arXiv:1909.05769 [hep-ph].
[10] U. D'Alesio, F. Murgia, C. Pisano, and P. Taels, Phys. Rev. D **100**, 094016 (2019), arXiv:1908.00446 [hep-ph].
[11] R. Kishore, A. Mukherjee, and M. Siddiqah, Phys. Rev. D **104**, 094015 (2021), arXiv:2103.09070 [hep-ph].
[12] R. Kishore, A. Mukherjee, A. Pawar, and M. Siddiqah, Phys. Rev. D **106**, 034009 (2022), arXiv:2203.13516 [hep-ph].
[13] D. Boer, J. Bor, L. Maxia, C. Pisano, and F. Yuan, JHEP **08**, 105 (2023), arXiv:2304.09473 [hep-ph].
[14] M. Copeland, S. Fleming, R. Gupta, R. Hodges, and T. Mehen, (2023), arXiv:2308.08605 [hep-ph].
[15] M. G. Echevarria, S. F. Romera, and I. Scimemi, (2023), arXiv:2308.12356 [hep-ph].
[16] M. Copeland, S. Fleming, R. Gupta, R. Hodges, and T. Mehen, (2023), arXiv:2310.13737 [hep-ph].
[17] Z.-B. Kang, E. Li, and F. Salazar, (2023), arXiv:2310.12102 [hep-ph].
[18] G. T. Bodwin, E. Braaten, and G. P. Lepage, Phys. Rev. D **51**, 1125 (1995), [Erratum: Phys.Rev.D 55, 5853 (1997)], arXiv:hep-ph/9407339.
[19] N. Brambilla *et al.*, Eur. Phys. J. C **71**, 1534 (2011), arXiv:1010.5827 [hep-ph].
[20] J.-P. Lansberg, Phys. Rept. **889**, 1 (2020), arXiv:1903.09185 [hep-ph].
[21] Y. Hatta, B.-W. Xiao, F. Yuan, and J. Zhou, Phys. Rev. Lett. **126**, 142001 (2021), arXiv:2010.10774 [hep-ph].
[22] Y. Hatta, B.-W. Xiao, F. Yuan, and J. Zhou, Phys. Rev. D **104**, 054037 (2021), arXiv:2106.05307 [hep-ph].
[23] S. Catani, M. Grazzini, and A. Torre, Nucl. Phys. B **890**, 518 (2014), arXiv:1408.4564 [hep-ph].
[24] S. Catani, M. Grazzini, and H. Sargsyan, JHEP **06**, 017 (2017), arXiv:1703.08468 [hep-ph].
[25] P. Sun, C. P. Yuan, and F. Yuan, Phys. Rev. Lett. **113**, 232001 (2014), arXiv:1405.1105 [hep-ph].
[26] P. Sun, C. P. Yuan, and F. Yuan, Phys. Rev. D **92**, 094007 (2015), arXiv:1506.06170 [hep-ph].
[27] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, Phys. Rev. Lett. **122**, 192003 (2019), arXiv:1812.08077 [hep-ph].
[28] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, Phys. Rev. D **102**, 094022 (2020), arXiv:2007.12866 [hep-ph].
[29] F. Yuan, Phys. Rev. D **78**, 014024 (2008), arXiv:0801.4357 [hep-ph].
[30] P. Sun, C. P. Yuan, and F. Yuan, Phys. Rev. D **88**, 054008 (2013), arXiv:1210.3432 [hep-ph].
[31] H. X. Zhu, C. S. Li, H. T. Li, D. Y. Shao, and L. L. Yang, Phys. Rev. Lett. **110**, 082001 (2013), arXiv:1208.5774 [hep-ph].
[32] R. Zhu, P. Sun, and F. Yuan, Phys. Lett. B **727**, 474 (2013), arXiv:1309.0780 [hep-ph].
[33] M. G. Echevarria, JHEP **10**, 144 (2019), arXiv:1907.06494 [hep-ph].
[34] S. Fleming, Y. Makris, and T. Mehen, JHEP **04**, 122 (2020), arXiv:1910.03586 [hep-ph].
[35] P. Ko, J. Lee, and H. S. Song, Phys. Rev. D **54**, 4312 (1996), [Erratum: Phys.Rev.D 60, 119902 (1999)], arXiv:hep-ph/9602223.
[36] J. C. Collins, D. E. Soper, and G. F. Sterman, Nucl. Phys. B **250**, 199 (1985).
[37] P. Sun, J. Isaacson, C. P. Yuan, and F. Yuan, Int. J. Mod. Phys. A **33**, 1841006 (2018), arXiv:1406.3073 [hep-ph].
[38] A. Prokudin, P. Sun, and F. Yuan, Phys. Lett. B **750**, 533 (2015), arXiv:1505.05588 [hep-ph].
[39] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021), arXiv:2012.04684 [hep-ph].
[40] M. Butenschoen and B. A. Kniehl, Phys. Rev. D **84**, 051501 (2011), arXiv:1105.0820 [hep-ph].
[41] K.-T. Chao, Y.-Q. Ma, H.-S. Shao, K. Wang, and Y.-J. Zhang, Phys. Rev. Lett. **108**, 242004 (2012), arXiv:1201.2675 [hep-ph].
[42] B. Gong, L.-P. Wan, J.-X. Wang, and H.-F. Zhang, Phys. Rev. Lett. **110**, 042002 (2013), arXiv:1205.6682 [hep-ph].

[43] R. Sharma and I. Vitev, Phys. Rev. C **87**, 044905 (2013), arXiv:1203.0329 [hep-ph].
[44] N. Brambilla, H. S. Chung, A. Vairo, and X.-P. Wang, Phys. Rev. D **105**, L111503 (2022), arXiv:2203.07778 [hep-ph].

SUPPLEMENTAL MATERIAL

As the supplemental material of the paper, we provide all the technical details below.

I. EVALUATION OF SOFT GLUON RADIATIONS

In this section, we provide detailed derivations of the soft gluon radiation contributions obtained by integrating the relevant rapidity phase space. This requires discussing four functions of the kind presented in Eq. (3): $S_g(p_2, k_j)$, $S_g(p_2, k_\psi)$, $S_g(k_j, k_\psi)$, and $S_g(k_\psi, k_\psi)$.

Integral of $S_g(p_2, k_j)$. We start with the first one on the list, which is the only one contributing to the CS channel. Although its integral was relevant, and therefore computed, in other works [21, 22, 25–28], for completeness we report the calculation in the following. To better analyze the physical content of this derivation, we can divide $S_g(p_2, k_2)$ into three contributions as follows

$$\begin{aligned} S_g(p_2, k_j) &= \frac{2\,(p_2 \cdot k_j)}{(p_2 \cdot k_g)(k_j \cdot k_g)} \\ &= \frac{2}{|\vec{k}_{g\perp}|^2} \frac{e^{\Delta y_g}}{\cosh(\Delta y_g) - \cos(\Delta\phi_g)} \\ &\approx \frac{2}{|\vec{k}_{g\perp}|^2} \left(1 + \frac{\sinh(\Delta y_g)}{\cosh(\Delta y_g) - \cos(\phi)} \right. \\ &\quad \left. + \frac{\cos(\phi)}{\cosh(\Delta y_g) - \cos(\phi)} \right) , \end{aligned} \tag{15}$$

where we have defined $\Delta y_g = y_g - y_j$, namely the difference between the emitted soft gluon and jet rapidities. The first term in the bracket of Eq. (15) leads to the double-logarithm. The second one, being an odd function of Δy_g, receives contributions only from the boundaries of the integration region, which causes the presence of the additional logarithmic term, $\ln(\hat{t}/\hat{u})$, in Eq. (4). The last term in Eq. (15) contains the jet contribution to azimuthal angular asymmetries.

In particular, this distribution is described by the function

$$\begin{aligned} I_j(R, \phi) &= \int \mathrm{d}\Delta y_g \, \frac{\cos(\phi)}{\cosh(\Delta y_g) - \cos(\phi)} \\ &\quad - \frac{|\vec{k}_{g\perp}|^2}{2} \int \mathrm{d}\Delta y_g \, S_g(p_2, k_j)\, \Theta(\Delta_{k_g k_j} < R^2) \\ &= \int \mathrm{d}\Delta y_g \, \frac{\cos(\phi)}{\cosh(\Delta y_g) - \cos(\phi)} \Theta(\Delta_{k_j k_g} > R^2) \\ &\quad - 2\sqrt{R^2 + \phi^2} \, , \end{aligned} \tag{16}$$

where $\Theta(\Delta_{k_j k_g} \lessgtr R^2)$ implies that the integration is restricted inside ($<$) or outside ($>$) the rapidity region occupied by the jet cone with size R, namely

$$|\Delta y_g| > \sqrt{R^2 + \phi^2} \, . \tag{17}$$

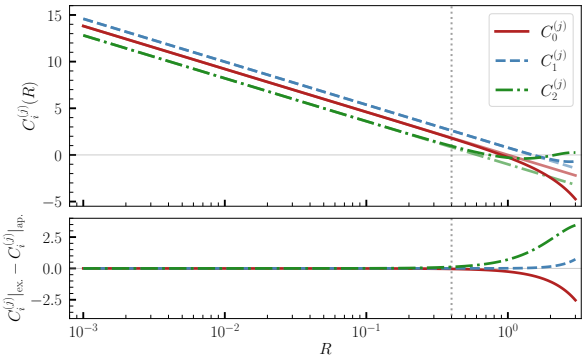


FIG. 4. *Dependence of the first coefficients of Eq. (5) with respect to R. In the upper panel we show both the exact result obtained from a numerical computation (full color) and the approximated one given in Eq. (18) (softer color). In the lower panel, we present the difference between the two. The vertical dotted line corresponds to $R = 0.4$, beyond which the approximation fails.*

For a general R, the Fourier expansion of $I_j(R, \phi)$ (Eq. (5)) is manageable only via computational methods. However, the analytical evaluation of this expansion within the small-R limit is possible. Hence we find

$$\begin{aligned} I_j(R, \phi) &= \ln\frac{1}{R^2} + 2\cos(\phi)\left(\ln\frac{1}{R^2} + 2\ln(4) - 2\right) \\ &\quad + 2\cos(2\phi)\left(\ln\frac{1}{R^2} - 1\right) + \cdots \, , \end{aligned} \tag{18}$$

from which it is evident that there is a non-negligible $\cos(\phi)$ azimuthal angular dependence in $I_j(R, \phi)$.

In Fig. 4, we show how the first three coefficients of the Fourier expansion behave with respect to R. From this figure, we conclude that the approximation used in Eq. (18) is reasonably adequate for $R < 0.4$.

Integral of $S_g(p_2, k_\psi)$. Similarly to $S_g(p_2, k_j)$, we divide $S_g(p_2, k_\psi)$ in the following three terms

$$\begin{aligned} S_g(p_2, k_\psi) &= \frac{2\,(p_2 \cdot k_\psi)}{(p_2 \cdot k_g)(k_\psi \cdot k_g)} \\ &= \frac{2}{|\vec{k}_{g\perp}|^2} \frac{\sqrt{1 + m_{\psi\perp}^2}\, e^{\Delta y_{g\psi}}}{\sqrt{1 + m_{\psi\perp}^2}\cosh(\Delta y_{g\psi}) - \cos(\Delta\phi_{g\psi})} \\ &\approx \frac{2}{|\vec{k}_{g\perp}|^2} \left(1 + \frac{\sqrt{1 + m_{\psi\perp}^2}\sinh(\Delta y_{g\psi})}{\sqrt{1 + m_{\psi\perp}^2}\cosh(\Delta y_{g\psi}) + \cos(\phi)} \right. \\ &\quad \left. - \frac{\cos(\phi)}{\sqrt{1 + m_{\psi\perp}^2}\cosh(\Delta y_{g\psi}) + \cos(\phi)} \right) , \end{aligned} \tag{19}$$

where we have defined $\Delta y_{g\psi} = y_g - y_\psi$, namely the difference between the emitted soft gluon and J/ψ rapidities, and $\Delta\phi_{g\psi} = \phi_g - \phi_\psi = \phi_g$ in a frame where $\phi_\psi = 0$. As for Eq. (15), the first term in the bracket of Eq. (19) leads to the double-logarithm while the second one receives

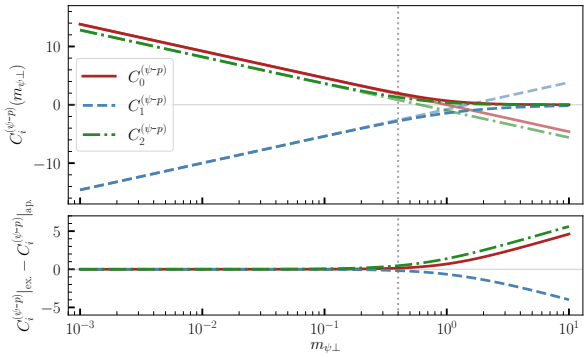


FIG. 5. *Dependence of the first coefficients of Eq. (22) with respect to $m_{\psi\perp}$. Vertical dotted line corresponds to $m_{\psi\perp} = 0.4$. Panels follow the same logic as Fig. 4.*

contributions only from the boundaries of the integration region, providing the logarithms: $\ln\frac{\hat{u}}{\hat{t}} + \ln\frac{1-M_\psi^2/\hat{u}}{1-M_\psi^2/\hat{t}}$. These, once combined with the other analogous logarithms arising from $S_g(p_2,k_j)$, cause the presence of the second term in Eq. (7). The last term in Eq. (19) contains part of the J/ψ contribution to azimuthal angular asymmetries.

We identify this angular distribution as

$$I_{\psi\text{-}p}(m_{\psi\perp},\phi) = \int \mathrm{d}\Delta y_{g\psi}\left[-\frac{\cos(\phi)}{\sqrt{1+m_{\psi\perp}^2}\,\cosh(\Delta y_{g\psi})+\cos(\phi)}\right] \tag{20}$$

Note that the presence of the mass in the denominator acts as a regulator, and therefore $I_{\psi\text{-}p}$ is continuous for all values of ϕ. Moreover, when evaluated within the jet rapidity region, Eq. (19) leads to

$$I^{\text{jet}}_{\psi\text{-}p}(R,m_{\psi\perp},\Delta y,\phi) = \frac{|\vec{k}_{g\perp}|^2}{2}\int \mathrm{d}\Delta y_g\, S_g(p_2,k_\psi)\,\Theta(\Delta_{k_g k_j}<R^2)\ , \tag{21}$$

which not only depends on both parameters R and $m_{\psi\perp}$, but on the rapidity difference $\Delta y = y_\psi - y_j$ too. The distribution in Eq. (20) can be expanded according to Eq. (8), which leads to

$$I_{\psi\text{-}p}(m_{\psi\perp},\phi) = C_0^{(\psi\text{-}p)}(m_{\psi\perp}) + 2\sum_{n=1}^{\infty} C_n^{(\psi\text{-}p)}(m_{\psi\perp})\cos(n\phi)\ . \tag{22}$$

These coefficients can be analytically evaluated only in the small-$m_{\psi\perp}$ limit, for which

$$I_{\psi\text{-}p}(m_{\psi\perp},\phi) = \ln\frac{1}{m_{\psi\perp}^2} - 2\cos(\phi)\left(\ln\frac{1}{m_{\psi\perp}^2}+2\ln(4)-2\right) + 2\cos(2\phi)\left(\ln\frac{1}{m_{\psi\perp}^2}-1\right)+\cdots\ . \tag{23}$$

Fig. 5 shows the dependence of the first three coefficients with respect to $m_{\psi\perp}$, together with the reliability of the approximation introduced in Eq. (23).

Integral of $S_g(k_j,k_\psi)$. Compared to the previous functions, deriving the azimuthal distribution arising from $S_g(k_j,k_\psi)$ requires some extra care. Firstly, we recast the function as follows

$$\begin{aligned} S_g(k_j,k_\psi) &= \frac{2\,(k_j\cdot k_\psi)}{(k_j\cdot k_g)(k_\psi\cdot k_g)} \approx \frac{2}{|\vec{k}_{g\perp}|^2}\frac{\sqrt{1+m_{\psi\perp}^2}\,\cosh(\Delta y)+1}{\left[\cosh(\Delta y_g)-\cos(\phi)\right]\left[\sqrt{1+m_{\psi\perp}^2}\,\cosh(\Delta y_{g\psi})+\cos(\phi)\right]} \\ &= \frac{2}{|\vec{k}_{g\perp}|^2}\left(\frac{\cos(\phi)}{\cosh(\Delta y_g)-\cos(\phi)} - \frac{\cos(\phi)}{\sqrt{1+m_{\psi\perp}^2}\,\cosh(\Delta y_{g\psi})+\cos(\phi)} + \widehat{S}_g(k_j,k_\psi)\right)\ , \end{aligned} \tag{24}$$

where we remark that $\Delta y = y_\psi - y_j$, $\Delta y_g = y_g - y_j$ and $\Delta y_{g\psi} = y_j - y_\psi$, with y_j, y_ψ and y_g being respectively the jet, J/ψ and emitted soft gluon rapidities. The first two terms in the last line of Eq. (24) coincide with the last terms of Eqs. (15) and (19), respectively, and thus remove the double counting in the azimuthal dependences. The last term, which is explicitly given by

$$\widehat{S}_g(k_j,k_\psi) = \frac{1+\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y)}{\cosh(\Delta y_g)+\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})}\left(\frac{1}{\cosh(\Delta y_g)-\cos(\phi)}+\frac{1}{\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})+\cos(\phi)}\right)$$
$$-\frac{\left(\cosh(\Delta y_g)+\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})\right)\cos(\phi)}{\cosh(\Delta y_g)+\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})}\left(\frac{1}{\cosh(\Delta y_g)-\cos(\phi)}-\frac{1}{\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})+\cos(\phi)}\right), \tag{25}$$

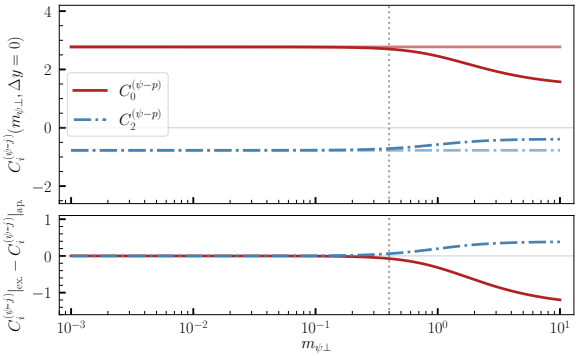


FIG. 6. *Dependence of the first coefficients of Eq. (29) with respect to $m_{\psi\perp}$ and for $\Delta y = 0$. The coefficient C_1, being zero, is not shown here. Vertical dotted line corresponds to $m_{\psi\perp} = 0.4$. Panels follow the same logic as Fig. 4.*

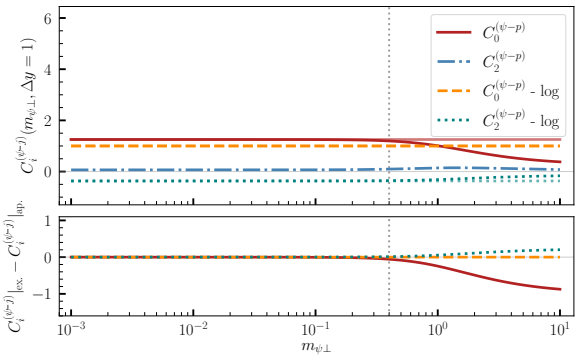


FIG. 7. *Same as Fig. 6 but for $\Delta y = 1$. At variance with the previous figure, we have that the coefficients include extra logarithms (see Eq. (30)), which are shown separately.*

produces the unique azimuthal distribution of $S_g(k_j,k_\psi)$.

Therefore, we identify the azimuthal distribution driven by the first two terms as

$$I_{\psi\text{-}j}(R,m_{\psi\perp},\phi) = \int \mathrm{d}y_g \left(\frac{\cos(\phi)}{\cosh(\Delta y_g)-\cos(\phi)} - \frac{\cos(\phi)}{\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})+\cos(\phi)}\right), \tag{26}$$

which is independent of Δy and part of I_ψ in Eq. (7) (see also the end of this section), while the last term gives

$$I_{\psi\text{-}j}(m_{\psi\perp},\Delta y,2\phi) = \int \mathrm{d}y_g\, \widehat{S}_g(k_j,k_\psi)\,, \tag{27}$$

which is exactly the same of Eq. (7). Moreover, $S_g(k_j,k_\psi)$ provides another azimuthal distribution when evaluated within the jet region

$$I^{\rm jet}_{\psi\text{-}j}(R,m_{\psi\perp},\Delta y,\phi) = \frac{|\vec{k}_{g\perp}|^2}{2}\int \mathrm{d}\Delta y_g\, S_g(k_j,k_\psi)\,\Theta(\Delta_{k_jk_\psi}<R^2)\,. \tag{28}$$

Among these three distribution, Eq. (27) is the most interesting. Note that this distribution only contributes to the even modes of the Fourier expansion and does not depend on the jet variable R, since the integration of $\widehat{S}_g(k_j,k_\psi)$ is continuous at $\phi = 0$. However, it presents an additional dependence on the rapidity difference Δy, which affects its harmonic expansion coefficients, given by

$$I_{\psi\text{-}j}(m_{\psi\perp},\Delta y,2\phi) = C_0^{(\psi\text{-}j)}(m_{\psi\perp},\Delta y) + 2\sum_{n=1}^{\infty} C_{2n}^{(\psi\text{-}p)}(m_{\psi\perp},\Delta y)\,\cos(2n\phi)\,. \tag{29}$$

In particular, depending on the value of Δy we can have additional logarithms of $\hat{u}/\hat{t}$ within the coefficients. Moreover, the closer the two outgoing particles are to the production axis (namely $|\Delta y|\to\infty$), the less relevant the angular distribution in $I_{\psi\text{-}j}(m_{\psi\perp},\Delta y,2\phi)$ becomes, with its sole contribution being restricted to the logarithm of $\hat{u}/\hat{t}$ in the constant term. To see this effect, we consider the analytical expansion in the small-$m_{\psi\perp}$ limit for two values of Δy, namely $\Delta y = 0$ and $\Delta y = 1$:

$$
\begin{aligned}
I_{\psi\text{-}j}(m_{\psi\perp},\Delta y=0,2\phi) &= 2\ln(4)-4\cos(2\phi)\,(\ln(4)-1)+\cdots\,,\\
I_{\psi\text{-}j}(m_{\psi\perp},\Delta y=1,2\phi) &= 2\left[2\Big(\ln(1+\mathrm{e})-1\Big)-\ln\frac{\hat{u}}{\hat{t}}-\frac{1}{2}\ln\frac{1-M_\psi^2/\hat{u}}{1-M_\psi^2/\hat{t}}\right]\\
&\quad-\frac{4}{\mathrm{e}}\cos(2\phi)\left[(1+\mathrm{e}^2)\Big(\ln(1+\mathrm{e})-1\Big)+\mathrm{e}-\ln\frac{\hat{u}}{\hat{t}}-\frac{1}{2}\ln\frac{1-M_\psi^2/\hat{u}}{1-M_\psi^2/\hat{t}}\right]+\ldots\,.
\end{aligned}
\tag{30}
$$

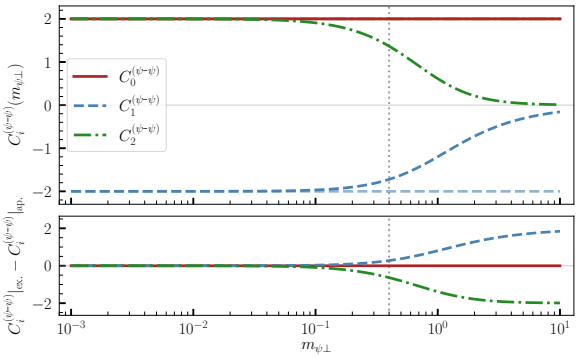


FIG. 8. *Dependence of the first coefficients of Eq. (34) with respect to $m_{\psi\perp}$. Note that the coefficient C_0 does not vary with $m_{\psi\perp}$. Vertical dotted line corresponds to $m_{\psi\perp}=0.4$. Panels follow the same logic as Fig. 4.*

The complete dependence of the first coefficients of Eq. (29) for the same values of Δy is shown in Figs. 6 and 7.

Integral of $S_g(k_\psi,k_\psi)$. The function $S_g(k_\psi,k_\psi)$, non-zero only for massive particles, is given by

$$
\begin{aligned}
&S_g(k_\psi,k_\psi)\\
&\quad\approx\frac{2}{|\vec{k}_{g\perp}^{\,2}|^2}\frac{m_{\psi\perp}^2}{\left(\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})+\cos(\phi)\right)^2}\,,
\end{aligned}
\tag{31}
$$

where we recall that $m_{\psi\perp}=M_\psi/|\vec{P}_\perp|$ with M_ψ being the J/ψ mass, and $\Delta y_{g\psi}$ is the rapidity difference between the soft gluon and the quarkonium.

The azimuthal distribution arising from $S_g(k_\psi,k_\psi)$ is identify by

$$
\begin{aligned}
&I_{\psi\text{-}\psi}(m_{\psi\perp},\phi)\\
&\quad=\int \mathrm{d}\Delta y_{g\psi}\,\frac{m_{\psi\perp}^2}{\left(\sqrt{1+m_{\psi\perp}^2}\cosh(\Delta y_{g\psi})+\cos(\phi)\right)^2}\,,
\end{aligned}
\tag{32}
$$

Moreover, when evaluated within the jet region, $S_g(k_\psi,k_\psi)$ generates

$$
\begin{aligned}
&I^{\rm jet}_{\psi\text{-}\psi}(R,m_{\psi\perp},\Delta y,\phi)\\
&\quad=\frac{|\vec{k}_{g\perp}|^2}{2}\int \mathrm{d}\Delta y_g\,S_g(k_\psi,k_\psi)\,\Theta(\Delta_{k_gk_j}<R^2)\,.
\end{aligned}
\tag{33}
$$

The harmonic expansion of Eq. (32), following that of Eq. (8), is given by

$$
\begin{aligned}
&I_{\psi\text{-}\psi}(m_{\psi\perp},\phi)\\
&\quad=C_0^{(\psi\text{-}\psi)}(m_{\psi\perp})+2\sum_{n=1}^{\infty}C_n^{(\psi\text{-}\psi)}(m_{\psi\perp})\,\cos(n\phi)\,.
\end{aligned}
\tag{34}
$$

It is interesting to notice that these coefficients are non-zero for all $m_{\psi\perp}$. In particular, we have residual contributions in the small-$m_{\psi\perp}$ limit, with $C_0^{(\psi\text{-}\psi)}$, $C_2^{(\psi\text{-}\psi)}\to 2$ and $C_1^{(\psi\text{-}\psi)}\to -2$, which is expected due to the singular behaviour of $I_{\psi\text{-}\psi}$ at $\phi=\pi$.

For completeness, Fig. 8 shows the exact dependence of the coefficients on $m_{\psi\perp}$. From this figure we understand that $C_0^{(\psi\text{-}\psi)}$ is independent of $m_{\perp\psi}$, whereas $C_1^{(\psi\text{-}\psi)}$ and $C_2^{(\psi\text{-}\psi)}$ are negligible when $m_{\perp\psi}\approx 1$.

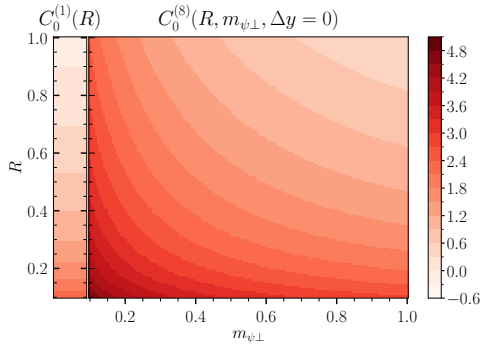
Overall azimuthal distribution. By combining all the soft gluon emission evaluated above, one gets the differential cross section as a function of $q_\perp$ given in Eq. (9), where the expansion in terms of $\cos(n\phi)$ harmonics has already being performed. The coefficients of this expansion depend on the production mechanism considered. For the CS channel, we have that $C_n^{(1)}$ are directly related to the Fourier expansion of I_j in Eq. (5). On the other hand, in the CO channel the coefficients $C_n^{(8)}$ are driven by all distributions in Eq. (7). More specifically, we have:

- $I_\psi(m_{\psi\perp},\phi)=I_{\psi\text{-}p}(m_{\psi\perp},\phi)-\frac{1}{2}I_{\psi\text{-}\psi}(m_{\psi\perp},\phi)$;
- $I_{\psi\text{-}j}(R,m_{\psi\perp},\Delta y,2\phi)$ is given in Eq. (27);
- $I^{\rm jet}_\psi=I^{\rm jet}_{\psi\text{-}p}+I^{\rm jet}_{\psi\text{-}j}-I^{\rm jet}_{\psi\text{-}\psi}-I^{\rm jet}_j(R,\phi)$, where $I^{\rm jet}_j$ is the contribution of I_j inside the jet region (namely the term in the second line of Eq. (16)).

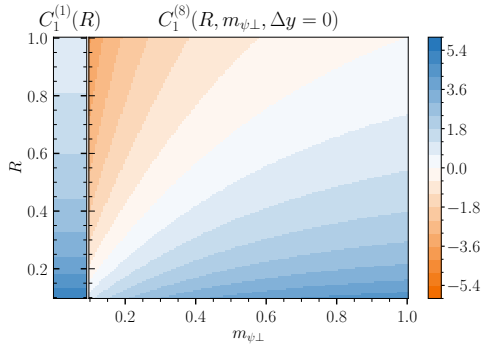
In Fig. 9 we present the dependence of the coefficients with respect to R and $m_{\psi\perp}$ for $\Delta y=0$. From a direct comparison between the CS and CO channels, it is evident that predictions that include the CO mechanism are significantly different from those exclusively driven by the CS one.

II. AZIMUTHAL ASYMMETRIES IN THE CS AND CO CHANNELS SEPARATELY

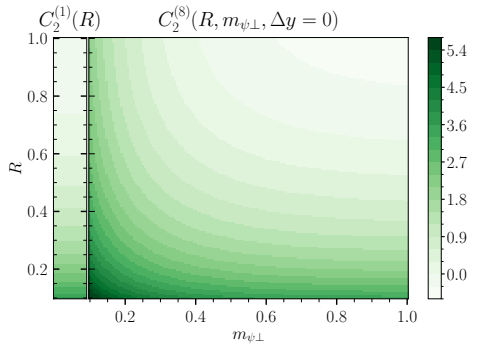
In this section we present direct comparisons between the CS and CO mechanisms, providing more insights on



(a)



(b)



(c)

FIG. 9. *Dependence of the first coefficients on the jet size R and the variable $m_{\psi\perp}$ in the CO channel for transverse production ($\Delta y = 0$). A comparison with the CS channel, for which coefficients solely depend on R, is given by the column on the left.*

the results shown in Fig. 3.

In Fig. 10 we present the normalized differential cross section within both the CS and CO channels and for the same kinematics considered for Fig. 3. The CO channel has a much wider distribution as compared to the CS one as a result of the final state gluon radiation associated with the heavy quark pair, which leads to a significant difference in the associated $C_0^{(c)}$ coefficients (see Table I). Note that, although the normalized distributions are similar for the two values of $|\vec{k}_{j\perp}|$ taken, the differential cross sections decrease significantly (approximately by an order of magnitude). It is worth mentioning that Fig. 10 is obtained for $y_j = y_\psi = 0$ and different rapidities values can slightly modify the picture, e.g. less broadened distributions for $y_j = 1$. Nonetheless, other choices of y_j and y_ψ do not spoil the main conclusion of Fig. 10, namely that we identify different shapes of the normalized differential cross sections within the CS and CO mechanisms.

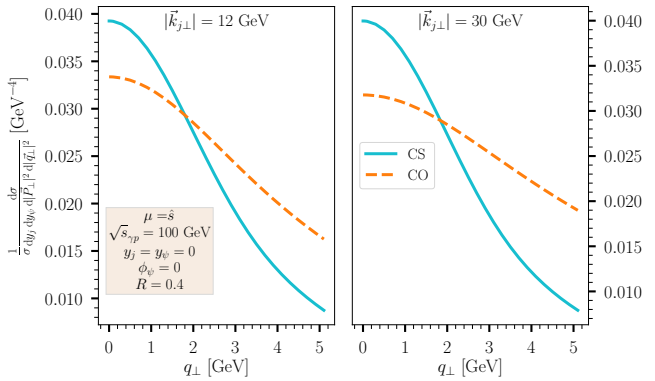
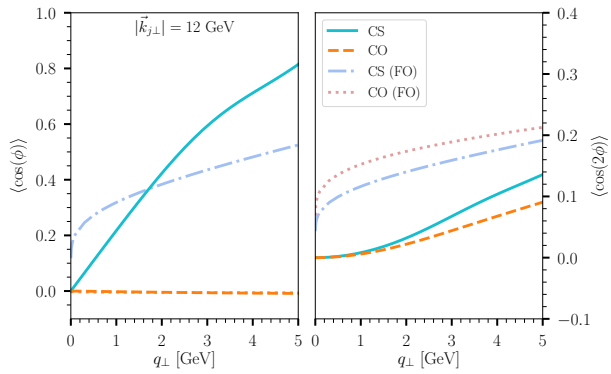
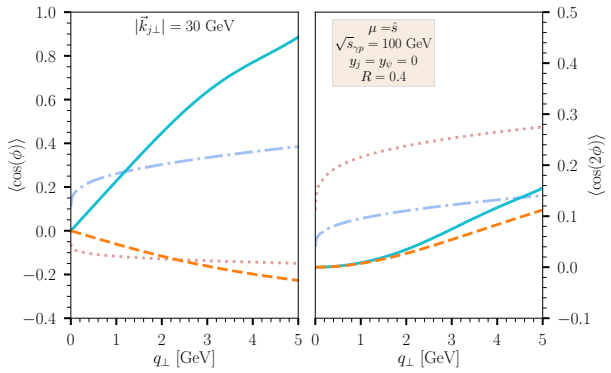


FIG. 10. *Normalized differential cross section of isotropic J/ψ plus jet photoproduction at $\sqrt{s}_{\gamma p} = 100$ GeV in a frame where the azimuthal angle of $\vec{P}_\perp$ is zero and $y_j = y_\psi = 0$. We considered two values of $|\vec{k}_{j\perp}|$, while $R = 0.4$. Solid cyan lines are obtained employing the CS channel, whereas the dashed orange lines the CO one.*

Moving to the asymmetries, in Fig. 11 we present, as functions of $q_\perp$ and for the same kinematic choices of Fig. 3, $\langle\cos(\phi)\rangle$ and $\langle\cos(2\phi)\rangle$ predictions in the CS and CO channels. For comparison, we also show the fixed-order results. In this figure, the behaviors of the resummed asymmetries within the CS and CO mechanisms are analogous to those of the CSM and NRQCD in Fig. 3. Thus, we may conclude that the NRQCD outcomes are mostly driven by the CO channel, at least for the kinematics considered.



(a)



(b)

FIG. 11. *Averaged azimuthal asymmetries for J/ψ plus jet photoproduction at $\sqrt{s}_{\gamma p} = 100$ GeV obtained within the CS (solid cyan line) and CO (dashed orange line) mechanisms. The dash-dotted blue and dotted red lines are the fixed-order results for the CS and CO channels, respectively. The curves in the top panel (11a) are evaluated at $|\vec{k}_{j\perp}| = 12$ GeV, whereas those in the bottom panel (11b) at $|\vec{k}_{j\perp}| = 30$ GeV. Jet size is $R = 0.4$ in both panels.*

Prepared for submission to JHEP

Constraints from the Neutron EDM on Subleading Effective Operators for Direct Dark Matter Searches

Manuel Drees,[a] **Rahul Mehra**[a,1]

[a] *Bethe Center for Theoretical Physics and Physikalisches Institut, Universität Bonn, Nussallee 12, D-53115 Bonn, Germany*

E-mail: drees@th.physik.uni-bonn.de, rmehra@physik.uni-bonn.de

Abstract: Interactions between Dark Matter (DM) and nucleons relevant for direct search experiments can be organised in a model independent manner using a Galiliean invariant, non–relativistic effective field theory (NREFT). Here one expands the interactions in powers of the momentum transfer $\vec{q}$ and DM velocity $\vec{v}$. This approach generates many operators. The potentially most important subleading operators are odd under T, and can thus only be present in a theory with CP violating interactions. We consider two such operators, called $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ in the literature, in simplified models with neutral spin-0 mediators; the couplings are chosen such that the coefficient of the leading spin independent (SI) operator, which survives for $\vec{v} \to 0$, vanishes at tree level. However, it is generically induced at the next order in perturbation theory. We perform a numerical comparison of the number of scattering events between interactions involving the $T-$odd operators and the corresponding loop induced SI contributions. We find that for "maximal" CP violation the former can dominate over the latter. However, in two of the three models we consider, an electric dipole moment of the neutron (nEDM) is induced at two–loop order. We find that the experimental bound on the nEDM typically leads to undetectably small rates induced by $\mathcal{O}_{10}$. On the other hand, the model leading to a nonvanishing coefficient of $\mathcal{O}_{11}$ does not induce an nEDM.

arXiv:2403.02083v1 [hep-ph] 4 Mar 2024

[1]Corresponding author.

Contents

1 Introduction **1**

2 NREFT and Simplified Models **4**
2.1 Model I 6
2.2 Model II 10
2.2.1 Model IIa 11
2.2.2 Model IIb 14
2.3 The Neutron EDM 16

3 Results and Discussions **18**

4 Summary and Conclusions **29**

A 1-loop calculations **32**
A.1 Model I Matrix Element 32
A.1.1 Model I Loop Functions 34
A.2 Model IIa and IIb Matrix Elements 35
A.2.1 Model IIa and IIb Loop Functions 39

B 2–Loop Calculations **42**

1 Introduction

The search for non–gravitational interactions of Dark Matter (DM) has not yielded a convincing signal so far. This is in stark contrast with the accumulated evidence for its gravitational interactions across a broad range of astrophysical length scales. Direct searches aim to detect non–gravitational interactions by observing the recoil of a target nucleus after an incoming DM particle has scattered off it. Terrestrial experiments using this search principle have led to tight constraints on the DM mass and cross-section parameter space [1–14]. For the past few decades such experimental efforts have focused on particle DM with weak scale interactions termed as Weakly Interacting Massive Particles (WIMPs). WIMPs can be produced in the early universe through freeze–out from the thermal plasma, which yields the observed DM relic abundance [15] for very roughly electroweak strength effective couplings. This has made WIMPs a particularly compelling category of DM candidates.
In order to interpret the results of direct search experiments one has to make assumptions regarding the nature of the interaction between DM and nucleons. Traditionally it was assumed that DM–nucleon interactions are dominated by just two operators describing spin independent (SI) and spin dependent (SD) interactions in the limit of zero WIMP velocity

[16]. These SI and SD operators are the leading terms of an EFT description, which is an expansion in powers of small parameters such as the DM velocity v and three momentum transfer scaled by the nucleon mass $\vec{q}/m_N$. Since $v/c \lesssim \mathcal{O}(10^{-3})$ in the solar neighborhood, the momentum transfer is restricted to $|\vec{q}| \lesssim \mathcal{O}(100 \text{ MeV})$. Although the momentum exchange can be substantial on nuclear physics scales, it is far below the electroweak scale, and also below the range of WIMP masses most direct search experiments can probe. Hence, it is reasonable to expect that the canonical SI and SD interactions (zeroth order terms) dominate the EFT expansion.

A detailed non–relativistic effective field theory (NREFT) description of elastic DM–nucleus scattering retains NLO and NNLO terms by modeling the nucleus as a spatially extended composite particle with spin and charge densities [17–20]. For WIMPs of spins 0 or 1/2, truncating the EFT expansion at second order generates a total of 14 operators.[1] Since these higher order terms vanish for $v \to 0$, they lead to a spectrum of recoil energies quite different from the usual quasi–exponential fall–off, often preferring energies larger than the typical range of values implemented for SI and SD searches in experiments [23]. Multiple experiments have now extended their recoil energy search window in order to optimize the search strategy in the NREFT framework [24–33].

The NREFT contains, at least, 28 free parameters when the operators for neutrons and protons are counted separately. Probing this large parameter space is an arduous task. A number of global analyses using data from current and planned direct detection experiments have placed upper limits on the coupling strengths in this multi–dimensional parameter space [34–40]. A common conclusion drawn from these global analyses is that experiments are nearly as sensitive to some momentum– or velocity–dependent operators that are odd under P (parity) and T (time reversal) transformations as they are to the leading SD operator. These subleading operators have been scrutinized less in the literature. It is therefore interesting to study the phenomenology of models where they are generated in the non–relativistic (NR) limit, and to understand when these operators can dominate over the traditional SI and SD interactions.

The CPT theorem [41, 42] implies that every $T-$odd NREFT operator must arise from a $CP-$violating (CPV) quantum field theory, where C refers to charge conjugation. However, there are extremely stringent experimental constraints on CP violation [43], which can be used to place constraints on such NREFT operators. The electric dipole moment of the neutron (nEDM) is one such observable. The sensitivity of current experiments [44–46] is several orders of magnitude above the prediction of the standard model of particle physics (SM), but they provide tight constraints on flavor diagonal CP violation in extensions of the SM. The experimental upper limit on the nEDM is

$$|d_n| < 1.8 \times 10^{-26} \text{ e} \cdot \text{cm} \quad (90 \text{ \% C.L.}) \,. \tag{1.1}$$

There have been a number of articles linking extensions of the SM to low energy NREFT operators [47–49]. In particular, Ref. [47] lists a set of simplified models for scalar, spinorial and vector DM candidates and derives the full set of NREFT operators in terms of the

[1] The set of NREFT operators is substantially enlarged for WIMPs of higher spin [21, 22].

parameters for each simplified model. However, it is worth noting that most extensions of the SM only generate a small subset of the NREFT operators. Moreover, in most cases the leading order operators describing the SI and SD interactions are generated as well, and will then typically dominate. However, it is conceivable that the standard operators are strongly suppressed, in which case formally subleading operators actually provide the dominant contribution to scattering.

In Ref. [50], we considered simplified models with charged mediators, which are exchanged in the $s-$channel in DM–nucleon scattering. We found that suppressing the leading SI operator required finetuning of couplings; moreover, the bound on the nEDM, which is generated at one–loop in these models, implies that the subleading operators lead to undetectably small DM scattering rates. The $P-$odd, $T-$odd operators can thus be neglected in such scenarios.

In this article, we consider models that augment the SM by a WIMP candidate and a mediator particle which does not carry electric or color charge; DM–nucleon scattering then proceeds via $t-$channel diagrams. If one allows all couplings that respect the $SU(3)_C \times U(1)_{\text{em}}$ gauge symmetry of the simplified Lagrangian, the leading SI term will be generated at tree level. However, the coefficient of this operator can be "switched off" by setting relevant coupling(s) to zero. This is an ad hoc choice, which cannot be justified by any symmetry. In this case, at the lowest order in perturbation theory, these models generate the $P-$odd, $T-$odd operators $\mathcal{O}_{10}$ or $\mathcal{O}_{11}$ without giving rise to the leading order operators. However, since the vanishing of the coefficient of the leading SI operator is not enforced by a symmetry, it will usually be generated at the next order in perturbation theory. Since now the contribution from the leading operator to DM–nucleon scattering is loop suppressed[2] while the contributions from $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ are suppressed by powers of v or $|\vec{q}|/m_N$, it is not a priori obvious which contribution is more important. We therefore numerically compare these contributions to the total number of events for a Xenon target. We find that for large regions in the parameter space of these models, the subleading $P-$odd, $T-$odd operators actually dominate over the SI term if the latter is purely loop–induced.

However, in models where tree level interactions generate only $\mathcal{O}_{10}$, we find that an nEDM is generically induced at two–loop order. The resulting upper bound on the couplings then again leads to unobservably small scattering rates, so that contributions from $\mathcal{O}_{10}$ to DM–nucleon scattering can be neglected. On the other hand, in the model where $\mathcal{O}_{11}$ is generated at tree level, the CP violation is restricted to the dark sector, and no nEDM is generated (apart from the tiny SM contribution). This case then provides an example of an NLO operator dominating over the traditional SI operator in DM–nucleon scattering, at least at the level of a simplified model.

The remainder of this article is organized as follows. In Section 2, we provide a brief introduction to the NREFT formalism, and introduce three CP violating simplified models

[2]For a purely pseudoscalar mediator, loop–induced contributions to $\mathcal{O}_1$ typically dominate the DM scattering rate [51–54]. However, the mediator has to have both scalar and pseudoscalar couplings in order to generate $\mathcal{O}_{10}$ or $\mathcal{O}_{11}$. A light mediator with general CP phases was considered in ref.[55], but constraints on these phases from EDMs were not considered, minimal flavor violation was assumed, and a trilinear coupling of the mediator to the SM Higgs boson was introduced.

yielding the $P-$ and $T-$odd NREFT operators. We also compute the loop diagrams for WIMP–nucleon scattering that give rise to the leading SI operator in the non–relativistic limit. We then match the different scattering contributions to the corresponding four–field effective operators and finally match these to the set of NREFT operators. We also compute the two–loop Feynman diagrams that induce a nEDM and discuss the implications for the corresponding $P-, T-$odd operator in the NREFT. In Section 3 we compute the number of events for a Xenon target for the three simplified models. We discuss our main numerical results, comparing the contributions of the NREFT operators to elastic WIMP–nucleon scattering. We conclude in Section 4. Details of our loop computations are given in Appendices A and B.

2 NREFT and Simplified Models

A non–relativistic effective field theory (NREFT) of elastic scattering between DM and nuclei exhaustively categorizes the possible interactions involved in direct searches. An incoming DM particle striking a target nucleus on Earth is quite slow in the detector rest frame, $v/c \sim \mathcal{O}(10^{-3})$, and therefore a non–relativistic EFT can be used to describe the scattering. This simplifies the nuclear physics required to compute the scattering rate, which is nevertheless rather nontrivial [18]. Traditionally only the leading terms were kept, which remain finite as $v \to 0$ [16]. The first and second order terms in DM velocity v and momentum transfer $\vec{q}/m_N$ (in units of the nucleon mass) were considered only relatively recently [18–20, 49, 56]. In the following we briefly summarize the salient points.

In this NREFT elastic DM–nucleon scattering is described using a basis of operators constructed from the following Hermitian quantities invariant under Galilean transformations:

$$i\vec{q}, \quad \vec{v}^{\,\perp} \equiv \vec{v} + \frac{\vec{q}}{2\mu_N}, \quad \vec{S}_N, \quad \vec{S}_\chi\,. \tag{2.1}$$

Here $\mu_N = m_N m_{\rm DM}/(m_N + m_{\rm DM})$ is the reduced mass of the DM–nucleon system. Energy conservation implies that the transverse velocity $\vec{v}^{\perp}$ is orthogonal to the momentum transfer $\vec{q}$. $\vec{S}_N$ and $\vec{S}_\chi$ are the spin of the nucleon and the WIMP χ; of course, the latter may be zero. Using these four building blocks, and only imposing Galilean invariance, one obtains a set of linearly independent operators $\mathcal{O}_i$ when the EFT is truncated at second order in the expansion parameter $\vec{q}/m_N$. Table 1 lists this set of 14 operators; the operators $\mathcal{O}_1$ and $\mathcal{O}_4$ describe the traditional leading SI and SD interaction, respectively. Note that the coefficients of these operators are in general different for neutrons and protons.

No invariance under any discrete symmetry was imposed in the construction of the NREFT operators. Their behavior under discrete transformations, in particular parity (P) and time reversal (T), can thus be used to classify the operators. A parity transformation corresponds to $\vec{q} \to -\vec{q}$, $\vec{v}^{\perp} \to -\vec{v}^{\perp}$, while $\vec{S}_N$ and $\vec{S}_\chi$, being pseudovectors, remain unchanged. On the other hand, under time reversal, all four vectors listed in (2.1) change sign, and in addition $i \to -i$ (i.e. the Hermitian operator $i\vec{q}$ remains unchanged). This leads to the following

$$\mathcal{O}_1 = 1_\chi 1_N; \qquad \mathcal{O}_6 = \left(\frac{\vec{q}}{m_N}\cdot\vec{S}_N\right)\left(\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi\right); \qquad \mathcal{O}_{10} = i\frac{\vec{q}}{m_N}\cdot\vec{S}_N;$$
$$\mathcal{O}_3 = i\vec{S}_N\cdot\left(\frac{\vec{q}}{m_N}\times\vec{v}^\perp\right); \qquad \mathcal{O}_7 = \vec{S}_N\cdot\vec{v}^\perp; \qquad \mathcal{O}_{11} = i\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi;$$
$$\mathcal{O}_4 = \vec{S}_\chi\cdot\vec{S}_N; \qquad \mathcal{O}_8 = \vec{S}_\chi\cdot\vec{v}^\perp; \qquad \mathcal{O}_{12} = \vec{S}_\chi\cdot(\vec{S}_N\times\vec{v}^\perp);$$
$$\mathcal{O}_5 = i\vec{S}_\chi\cdot\left(\frac{\vec{q}}{m_N}\times\vec{v}^\perp\right); \qquad \mathcal{O}_9 = i\vec{S}_\chi\cdot\left(\vec{S}_N\times\frac{\vec{q}}{m_N}\right); \qquad \mathcal{O}_{13} = i(\vec{S}_\chi\cdot\vec{v}^\perp)\left(\frac{\vec{q}}{m_N}\cdot\vec{S}_N\right);$$
$$\mathcal{O}_{14} = i(\vec{S}_N\cdot\vec{v}^\perp)\left(\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi\right)$$

Table 1: List of operators in the NREFT for elastic WIMP–nucleon scattering. We adopt the conventions of [20] by defining the operators normalized by the nucleon mass m_N in order to have a dimensionless basis. We omit the invariant $\mathcal{O}_2 = v_\perp^2$ because it is a second order correction to the SI operator $\mathcal{O}_1$, as well as $\mathcal{O}_{15} = -\left(\vec{S}_\chi\cdot\frac{\vec{q}}{m_N}\right)\left((\vec{S}_N\times\vec{v}^\perp)\cdot\frac{\vec{q}}{m_N}\right)$ since it generates a cross section of order v_T^6, which is N^{3}LO.

classification:

$$\begin{aligned}\mathcal{O}_1, \mathcal{O}_3, \mathcal{O}_4, \mathcal{O}_5, \mathcal{O}_6: &\quad P-\text{even and } T-\text{even},\\ \mathcal{O}_7, \mathcal{O}_8, \mathcal{O}_9: &\quad P-\text{odd and } T-\text{even},\\ \mathcal{O}_{13}, \mathcal{O}_{14}: &\quad P-\text{even and } T-\text{odd},\\ \mathcal{O}_{10}, \mathcal{O}_{11}, \mathcal{O}_{12}: &\quad P-\text{odd and } T-\text{odd}.\end{aligned}$$

The P and T quantum numbers of the NREFT operators must match those of the relativistic operators generating them. In particular, the CPT theorem stipulates that only a CP violating quantum field theory with a DM candidate can yield any of the $T-$odd operators. Out of those, the operators $\mathcal{O}_{12}$, $\mathcal{O}_{13}$ and $\mathcal{O}_{14}$ are not generated in scenarios with $t-$channel mediator with spin ≤ 1;[3] the latter two operators anyway contribute little to the scattering cross section if all operators have coefficients of similar size [39]. Therefore, we focus on $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ in the following.

Connecting a relativistic model for DM leading to specific DM–quark and/or DM–gluon interactions with the NREFT in general involves two steps. First, one integrates out the heavy mediator(s), where "heavy" here refers to all mediators ϕ with mass m_ϕ well above 100 MeV, which is the maximal three–momentum exchange in DM–nucleus scattering.[4] This results in relativistic but non–renormalizable four–field DM–quark and/or DM–gluon operators. In the second step one takes the non–relativistic limit of these four–field operators and matches them onto the NREFT operators listed in Table 1. This leads to an effective Lagrangian containing some (or all) of these operators, with coefficients determined by the couplings and masses of the original relativistic theory. The computation of DM–nucleus

[3] $\mathcal{O}_{12}$ can be generated in a model with a spin$-1/2$ WIMP and spin-0 $s-$channel mediator, but the constraint from the nEDM forces this contribution to be negligible [50].

[4] In order to include light mediators in the NREFT the differential cross section should be multiplied with $[m_\phi^2/(m_\phi^2+\vec{q}^2)]^2$; this will lead to a softening of the recoil spectrum for $\vec{q}^2 \gtrsim m_\phi^2$.

scattering rates from these coefficients involves numerous nuclear "response functions"; we refer to ref.[20], whose expressions we used in our own numerical code.
As already mentioned, the NREFT is an expansion in powers of v or $\vec{q}$. If all coefficients in the NREFT are of comparable magnitude, the total scattering rate typically receives the largest contribution from the SI operator $\mathcal{O}_1$, which is of zeroth order in the expansion. Moreover, this contribution is enhanced by A^2, where A is the nucleon number. For large momentum exchange, $|\vec{q}| \gtrsim 1/r_N$ where r_N is the radius of the target nucleus, the rate is somewhat suppressed by a form factor, but even without form factor this contribution to the scattering rate peaks at $|\vec{q}| \to 0$.
$\mathcal{O}_4$ is also of zeroth order in the expansion. However, since the spins of the nucleons largely cancel in any given nucleus, there is no A^2 enhancement; in fact, $\mathcal{O}_4$ does not contribute at all if the target nucleus has no spin.
As already noted, for sufficiently heavy DM particle (and target nucleus) the three–momentum exchange can reach $|\vec{q}| \sim 100$ MeV. In this high momentum exchange region of phase space the operators that are linear in $\vec{q}$ and independent of $\vec{v}^{\perp}$ are therefore only suppressed by a factor ~ 0.1; these are the operators $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ which are the focus of our study.[5] The contribution from $\mathcal{O}_{11}$ is A^2 enhanced, up to a form factor, but requires the DM particle to carry spin. The contribution from $\mathcal{O}_{10}$ suffers similar cancellations as that from $\mathcal{O}_4$, but can survive even for scalar DM particle.
Naively the contributions of operators involving $\vec{v}^{\perp}$ should be suppressed by a factor $v^2 \sim 10^{-6}$; indeed, compared to the contribution from $\mathcal{O}_1$ this is almost true. However, the richer structures can lead to nuclear response $\propto \vec{S}_N \cdot \vec{L}_N$ [20], $\vec{L}_N$ being the orbital angular momentum of a given nucleon; in this product the contributions of paired nucleons do *not* cancel. As a result, for heavy target nuclei and similar coefficients the contribution from $\mathcal{O}_{12}$ often exceeds that from $\mathcal{O}_4$ [39]; however, as already noted $\mathcal{O}_{12}$ is not generated in the models with neutral mediator that we consider in this article. Similarly, for equal coefficients $\mathcal{O}_3$ typically contributes almost as much as $\mathcal{O}_{10}$ does [39], but $\mathcal{O}_3$ is not generated in leading order when starting from a relativistic theory [47]. $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ are therefore the potentially most important higher order NREFT operators that can be generated from a relativistic QFT.
After these preliminaries, we are ready to introduce the simplified models we consider in this analysis. Recall that, in the spirit of ref.[47], we only impose invariance under $SU(3)_C \times U(1)_{\rm em}$, not under the complete gauge group of the SM.

2.1 Model I

Model I contains a complex spin–zero WIMP S and a real spin–zero mediator ϕ; both are gauge singlets. We assume that the WIMP is odd and the mediator and all the SM particles are even under a new discrete symmetry $\mathbb{Z}_2$. This forbids dark matter decay, but allows ϕ to couple to both S and to SM quarks q. The most general gauge invariant Lagrangian

[5] However, simplified models yielding $\mathcal{O}_{11}$ in the non–relativistic limit always seem to result in the coefficient c_{11} containing an extra factor $m_N/m_{\rm DM}$ [47–50]; see also eq.(2.21) below. A discussion assuming $m_{\rm DM}$ independent Wilson coefficients therefore overestimates the importance of $\mathcal{O}_{11}$. Even with this caveat, $\mathcal{O}_{11}$ remains the potentially most important non–leading NREFT operator for $m_{\rm DM} \lesssim 100$ GeV.

respecting the new $\mathbb{Z}_2$ symmetry and keeping the real and imaginary parts of S degenerate thus is:[6]

$$\mathcal{L}^{\rm I} = (\partial_\mu S)^\dagger\,(\partial^\mu S) - m_S^2\,S^\dagger S - \frac{\lambda_S}{2}\,(S^\dagger\,S)^2 + \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - \frac{1}{2}m_\phi^2\,\phi^2 - \frac{m_\phi\,\mu_1}{3}\,\phi^3 - \frac{\mu_2}{4}\,\phi^4$$
$$- g_1\,m_S\,S^\dagger S\,\phi - \frac{g_2}{2}\,S^\dagger S\,\phi^2 - h_1^{ij}\,\phi\,\bar{q}_i q_j - i h_2^{ij}\,\phi\,\bar{q}_i\gamma^5 q_j\,. \tag{2.2}$$

$U(1)_{\rm em}$ invariance implies that the mediator ϕ can only couple to quarks with identical electric charge. Hence the quark flavor indices i and j in the Yukawa coupling matrices are restricted to the same quark type. However, non–vanishing couplings of the mediator to quarks of different generations generate flavor changing neutral currents (FCNC) processes at tree level. Experimental constraints arising from meson mixing along with rare flavor changing decays severely limit these flavor non–diagonal couplings. Therefore we assume that the Yukawa coupling matrices in eq.(2.2) are diagonal in flavor space, in particular $h_{1,2}^q \equiv h_{1,2}^{ii}$ where q denotes the quark flavor the mediator couples to.
In eq.(2.2) we have allowed both scalar and pseudoscalar couplings of ϕ to the SM quarks. However, if $g_1 h_1^q \neq 0$ there will be tree level contributions to $\mathcal{O}_1$, which will then completely dominate WIMP–nucleus scattering, in which case the nonleading operators in the NREFT would not need to be considered. Moreover, $h_1^q h_2^q \neq 0$ for any quark q would yield an electric dipole moment of that quark at one–loop order, leading to very strong constraints on the parameters of the model. We therefore impose

$$h_1^q = 0\;\forall q\,. \tag{2.3}$$

We emphasize that this is completely ad hoc but necessary for considering the non-leading operators in the NREFT. In particular, it cannot be justified by any symmetry. In contrast, the requirement $h_2^q = 0$, which would also remove the EDMs but allow $\mathcal{O}_1$, could be justified by demanding CP conservation. Nevertheless "switching off" $\mathcal{O}_1$ via eq.(2.3) is still less inelegant than requiring specific relations between non–zero couplings, as we had to do in our earlier analysis of models with charged $s-$channel mediators [50]. On the other hand, since eq.(2.3) cannot be enforced by a symmetry, we can already anticipate that higher order contributions will generate an effective h_1^q or, more generally, lead to nonvanishing $\mathcal{O}_1$.
The matrix element for tree level $t-$channel scattering $S(p_S) + q(p_q) \to S(p'_S) + q(p'_q)$ is given by

$$\mathcal{M}^{\rm I}_{Sq\to Sq} = -\frac{h_2^q\,g_1\,m_S}{q^2 - m_\phi^2}\,\bar{u}(p'_q)\,i\gamma^5\,u(p_q)\,. \tag{2.4}$$

For $\left|q^2\right| \ll m_\phi^2$ we can ignore the q^2 term in the ϕ propagator. The matrix element can then be matched on to a relativistic effective operator $S^\dagger S\;\bar{q}\,i\gamma^5\,q$, which reduces to the operator $\mathcal{O}_{10}$ in the non–relativistic limit.
For Model I, at the lowest order in perturbation theory, thus only the operator $\mathcal{O}_{10}$ contributes to DM–nucleus scattering. However, at the next–to–leading order the one–loop

[6] Essentially identical results can be derived for a real spin–zero WIMP.

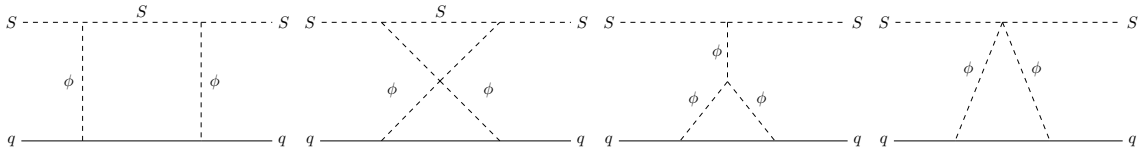


Figure 1: One–loop box and triangle Feynman diagrams in Model I which give contributions to $\mathcal{O}_1$.

Feynman diagrams shown in Fig. 1 can induce contributions from the operator $\mathcal{O}_1$. We use the Dirac equation and four–momentum conservation to write the resulting matrix elements in a form that is symmetric in DM momenta; see Appendix A.1 for details. The relativistic effective Lagrangian for Model I derived in this manner can be written as

$$\mathcal{L}_{\text{eff}}^{\text{I}} \supset c_{1,S}^{q,d5} S^\dagger S \; \bar{q} q \; + \; c_{10}^{q,d5} S^\dagger S \; \bar{q} \, i\gamma^5 \, q \; + \; c_{1,V}^{q,d6} i \left(S^\dagger \overleftrightarrow{\partial_\mu} S \right) \; \bar{q} \gamma^\mu q \, . \tag{2.5}$$

The Hermitean derivative on the complex scalars is defined as $iS^\dagger \overleftrightarrow{\partial_\mu} S \equiv \frac{i}{2}(S^\dagger \partial_\mu S - S \partial_\mu S^\dagger)$. The subscripts i on the quark-level Wilson coefficients $c_i^{q,dj}$ $(j = 5, 6)$ denote the NREFT operator that the corresponding relativistic effective operator reduces; in case of $\mathcal{O}_1$ we have distinguished the coefficient $c_{1,S}$ of the product of two scalar currents from $c_{1,V}$ which multiplies the product of two vector currents. Finally, the superscripts $d5$ and $d6$ refer to field operators with mass dimension 5 and 6, respectively.

Table 2 describes the matching of the relativistic effective operators onto the NREFT operators in terms of the parameters of Model I. The Wilson coefficients for the dimension–5 operators $(S^\dagger S)(\bar{q}q)$ and $(S^\dagger S)(\bar{q} i\gamma^5 q)$ have been divided by a factor of m_S, i.e. these are coefficients of the dimension–6 operators $m_S (S^\dagger S)(\bar{q}q)$ and $m_S (S^\dagger S)(\bar{q} i\gamma^5 q)$; this ensures that the expressions for all DM–nucleon cross sections contain the same factor $\frac{\mu_{\chi N}^2}{\pi} (c_i^N)^2$ irrespective of the mass dimension of the relativistic operator involved. The loop functions $M_i(r, s)$ and $L_i(r)$ appearing in the box and triangle diagrams have been expressed as functions of dimensionless parameters $r \equiv m_q/m_\phi$ and $s \equiv m_S/m_\phi$. Analytical expressions for these loop functions for Model I, and their $m_q \to 0$ limits, can be found in Appendix A.1.1. The dimension–5 scalar–scalar operator $(S^\dagger S)(\bar{q}q)$ as well as the dimension–6 vector–vector operator $i \left(S^\dagger \overleftrightarrow{\partial_\mu} S \right) \; \bar{q} \gamma^\mu q$ both reduce to the leading order SI operator $\mathcal{O}_1$ in the non–relativistic limit. Only the box diagrams give rise to both vector–vector and scalar–scalar relativistic operators, whereas the two triangle diagrams yield only the scalar–scalar operator. The first triangle diagram could be described in terms of an effective, loop–generated coupling h_1^q, thereby confirming our expectation that the choice $h_1^q = 0$ is not technically natural; however, all four diagrams shown in Fig. 1 contribute at the same order in perturbation theory, and should therefore be included in a full NLO treatment. The loop–generated Wilson coefficients are all suppressed by the loop factor $1/16\pi^2$ compared to the tree level diagram. The loop functions M_i have mass dimension -4 while L_1 has mass dimension -2, hence the quark–level Wilson coefficients c_i^q all have the same mass dimension -2. Since the scalar–scalar operator violates chirality, the contributions to $c_{1,S}^{q,d5}$ are all $\propto m_q$.

	$S^\dagger\Gamma_S S\ \bar{q}\,\Gamma_q\, q$		c_i^q
Tree	$c_{10}^{q,d5} S^\dagger S\ \bar{q}\, i\gamma^5\, q$	$\longrightarrow$	$\dfrac{h_2^q\, g_1}{m_\phi^2}$
Box	$c_{1,V,B}^{q,d6} i\left(S^\dagger \overleftrightarrow{\partial_\mu} S\right)\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$\dfrac{(h_2^q)^2\, g_1^2\, m_S^2}{16\pi^2}\, M_1$
	$c_{1,S,B}^{q,d5} S^\dagger S\ \bar{q}q$	$\longrightarrow$	$\dfrac{(h_2^q)^2\, g_1^2\, m_S\, m_q}{16\pi^2}\, M_2$
Crossed	$c_{1,V,C}^{q,d6} i\left(S^\dagger \overleftrightarrow{\partial_\mu} S\right)\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$-\dfrac{(h_2^q)^2\, g_1^2\, m_S^2}{16\pi^2}\, M_3$
Box	$c_{1,S,C}^{q,d5} S^\dagger S\ \bar{q}q$	$\longrightarrow$	$\dfrac{(h_2^q)^2\, g_1^2\, m_S\, m_q}{16\pi^2}\, M_4$
Triangle 1	$c_{1,S,T1}^{q,d5} S^\dagger S\ \bar{q}q$	$\longrightarrow$	$\dfrac{(h_2^q)^2\, g_1\, \mu_1}{16\pi^2}\dfrac{m_q}{m_\phi}\, L_1$
Triangle 2	$c_{1,S,T2}^{q,d5} S^\dagger S\ \bar{q}q$	$\longrightarrow$	$-\dfrac{(h_2^q)^2\, g_2}{16\pi^2}\dfrac{m_q}{2m_S}\, L_1$

Table 2: Non–relativistic reduction of relativistic effective operators in Model I. The middle column gives the relativistic four–field operators that appear in the matrix element for WIMP–nucleon scattering. The right column gives the corresponding quark–level Wilson coefficient c_i^q of the relevant NREFT operators, where we have suppressed the dependence of the loop functions on dimensionless parameters, i.e. $M_i \equiv M_i(r,s)$ and $L_1 \equiv L_1(r)$ with $r \equiv m_q/m_\phi$ and $s \equiv m_S/m_\phi$. The first line contributes to the coefficient of $\mathcal{O}_{10}$, all other lines describe contributions $\propto \mathcal{O}_1$.

The quark bilinears in eq.(2.5) and Table 2 must be promoted to nucleon bilinears in order to describe DM–nucleon scattering. The quark–level Wilson coefficients therefore have to be combined with corresponding nucleon embedding factors in order to derive the Wilson coefficients at the nucleonic level [18, 57]. Including contributions from the box and the triangle diagrams, the nucleonic Wilson coefficient c_1^N of the NREFT operator $\mathcal{O}_1$ becomes:

$$
\begin{aligned}
c_1^N|^{\mathrm{I}} = \frac{1}{16\,\pi^2}\Big[\, g_1^2\, m_S^2 \Big\{ \sum_{u,d} (h_2^q)^2 \mathcal{N}_q^N \big(M_1(r_q,s) - M_3(r_q,s)\big) \Big\} \\
+ m_N \left(\sum_{u,d,s} (h_2^q)^2 f_{Tq}^N + \frac{2}{27} f_{TG}^N \sum_{c,b,t} (h_2^q)^2 \right) \\
\times \Big\{ g_1^2 m_S \Big(M_2(r_q,s) + M_4(r_q,s) \Big) + \left(\frac{g_1 \mu_1}{m_\phi} - \frac{g_2}{2m_S} \right) L_1(r_q) \Big\} \Big]\,. \qquad (2.6)
\end{aligned}
$$

Promoting the vector quark bilinear $\mathcal{N}_q^N \equiv \langle \bar{N} |\, \bar{q}\gamma^\mu q\, | N \rangle$ to the vector nucleon bilinear yields the number of valence quarks of flavor q in the nucleon N, i.e. $\mathcal{N}_u^p = \mathcal{N}_d^n = 2$, $\mathcal{N}_d^p = \mathcal{N}_u^n = 1$. The contribution of light quarks to the nucleon mass (scalar nucleon bilinear) is

given by

$$q = u, d, s : \left\langle \bar{N} \right| m_q \, \bar{q} q \left| N \right\rangle = m_N f_{Tq}^N \, . \tag{2.7}$$

The heavy quarks contribute to the nucleon mass via the trace anomaly of the energy–momentum tensor [58]:

$$q = c, b, t : \left\langle \bar{N} \right| m_q \, \bar{q} q \left| N \right\rangle = \frac{2}{27} m_N f_{TG}^N = \frac{2}{27} m_N \Big(1 - \sum_{q'=u,d,s} f_{Tq'}^N \Big) \, . \tag{2.8}$$

At first sight one might think that the contribution from the product of two vector currents dominates, since the corresponding nucleonic matrix elements $\mathcal{N}_q^N$ are large. However, in the limit $r_q \to 0$ the difference $M_1(r_q, s) - M_3(r_q, s)$ vanishes $\propto r_q^2$. In contrast, the sum $M_2(r_q, s) + M_4(r_q, s) \to 2M_2(0, s)$ remains finite for massless quarks. As a result, the first term in eq.(2.6) typically contributes much less than the remaining terms, which originate from the product of scalar currents.

For the tree–level scattering contribution reducing to $\mathcal{O}_{10}$, the nucleonic Wilson coefficient in the limit of vanishing momentum transfer is [18]

$$c_{10}^N|^{\rm I} = \frac{g_1}{m_\phi^2} \Big(\sum_{u,d,s} h_2^q \, \Delta\tilde{q}^N - \Delta\tilde{G}^N \sum_{c,b,t} \frac{h_2^q}{m_q} \Big) \tag{2.9}$$

where $\Delta\tilde{q}^N$ and $\Delta\tilde{G}^N$ are the light and heavy quark contributions to the nucleon level pseudoscalar bilinear, respectively. The latter hadronic matrix element is due to the QCD chiral anomaly:

$$q = c, b, t : \left\langle \bar{N} \right| \partial_\mu (\bar{q} \gamma^\mu \gamma^5 q) \left| N \right\rangle = 2 m_q \left\langle \bar{N} \right| \bar{q} i \gamma^5 q \left| N \right\rangle + \frac{\alpha_s}{4\pi} \left\langle \bar{N} \right| G^{a\mu\nu} \tilde{G}^a_{\mu\nu} \left| N \right\rangle \, , \tag{2.10}$$

where G and $\tilde{G}$ are the gluonic field strength tensor and its dual, respectively. The left–hand side can be set to zero since heavy quarks have no significant dynamics in the nucleon. Throughout our calculations, we take the numerical values of the coefficients that appear when quark bilinears are promoted to nucleon bilinears as given in the Appendix of Ref. [47].

2.2 Model II

In Model II we replace the complex scalar WIMP by a spin$-1/2$ gauge singlet Dirac fermion χ, again using a real spin–zero mediator ϕ. The couplings of the SM quarks q to the mediator are as in Model I. The WIMP χ can also couple to the mediator ϕ via a scalar (λ_1) and a pseudoscalar coupling (λ_2). The renormalizable $SU(3)_C \times U(1)_{\rm em}$ invariant Lagrangian is thus given by

$$\begin{aligned} \mathcal{L}^{\rm IIa} = & \, i \bar{\chi} \not{D} \chi - m_\chi \, \bar{\chi} \chi + \frac{1}{2} \partial_\mu \phi \, \partial^\mu \phi - \frac{1}{2} m_\phi^2 \, \phi^2 + \frac{m_\phi \, \mu_1}{3} \, \phi^3 - \frac{\mu_2}{4} \, \phi^4 \\ & - \lambda_1 \, \phi \, \bar{\chi} \chi - i \, \lambda_2 \, \phi \, \bar{\chi} \gamma^5 \chi - h_1^{ij} \, \phi \, \bar{q}_i q_j - i h_2^{ij} \, \phi \, \bar{q}_i \gamma^5 q_j \, . \end{aligned} \tag{2.11}$$

We assume that the DM particle is odd and the SM particles along with the mediator are even under a new discrete $\mathbb{Z}_2$ symmetry in order to prevent DM decay. In order to avoid

potentially very large new contributions to FCNC processes we again take flavor diagonal quark couplings, i.e. $h_1^q \equiv h_1^{ij}\delta_{ij}$ and $h_2^q \equiv h_2^{ij}\delta_{ij}$.
If all Yukawa couplings (of a given flavor) appearing in the Lagrangian (2.11) are of similar magnitude, WIMP–nucleus scattering on heavy target nuclei will be completely dominated by a tree level contribution from $\mathcal{O}_1$, with coefficient $\propto \lambda_1 h_1^q$. We thus have to set at least one of these couplings to zero. On the other hand, setting both of them to zero would also "switch off" the operators $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ which we seek to generate. We thus consider two variants of Model II:

$$\begin{aligned} &\text{Model IIa} : h_1^q = 0 \;\forall q\,; \\ &\text{Model IIb} : \lambda_1 = h_2^q = 0 \;\forall q\,. \end{aligned} \tag{2.12}$$

Setting $h_1^q \cdot h_2^q = 0$ again ensures that no electric dipole moments are generated at one–loop level. As before, neither of these choices is protected by a symmetry, i.e. they are not technically natural. We therefore again expect that $\mathcal{O}_1$ will be generated by radiative corrections. On the other hand, CP is violated if for some flavor q,

$$\lambda_1 h_2^q \neq 0 \quad \text{or} \quad \lambda_2 h_1^q \neq 0\,. \tag{2.13}$$

We will discuss both variants of Model II in turn.

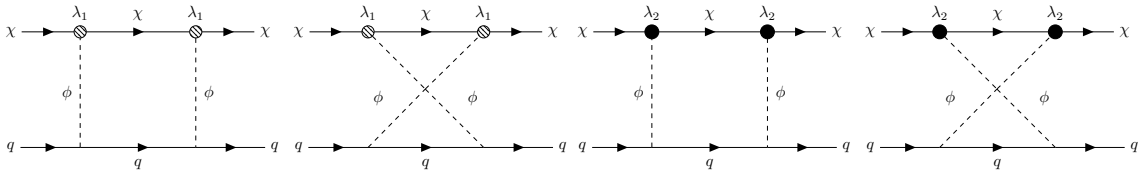
2.2.1 Model IIa

The matrix element for tree–level DM–quark scattering, $\chi(p_\chi) + q(p_q) \to \chi(p'_\chi) + q(p'_q)$, via ϕ exchange in the $t-$channel is given by

$$\mathcal{M}^{\text{IIa}}_{\chi q\to\chi q} = -\frac{h_2^q}{q^2 - m_\phi^2}\; \bar{u}(p'_\chi)(\lambda_1 + i\gamma^5\lambda_2)u(p_\chi)\, \bar{u}(p'_q) i\gamma^5 u(p_q)\;. \tag{2.14}$$

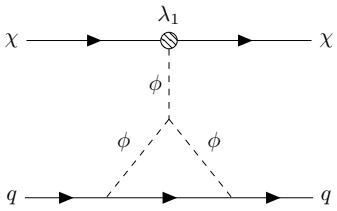
Taking the vanishing momentum transfer limit, $q^2 \to 0$, in the ϕ propagator, the matrix element can be matched onto the relativistic effective operators $(\bar{\chi}\chi)\left(\bar{q}\, i\gamma^5\, q\right)$ and $\left(\bar{\chi}\, i\gamma^5\, \chi\right)\left(\bar{q}\, i\gamma^5\, q\right)$, which reduce to the operators $\mathcal{O}_{10}$ and $\mathcal{O}_6$, respectively, in the non–relativistic limit. Recall, however, from Table 1 that $\mathcal{O}_6$ is doubly suppressed by momentum transfer $\vec{q}$, therefore generating a cross section suppressed by $\mathcal{O}(\vec{q}^{\,2}/m_N^2) \sim 10^{-2} - 10^{-3}$ relative to the contribution from $\mathcal{O}_{10}$. Hence, at the leading order in perturbation theory, $\mathcal{O}_{10}$ dominates the scattering matrix element in Model IIa.
Although the operator $\mathcal{O}_1$ is not generated from leading order Feynman diagrams, one–loop corrections to DM–nucleus scattering again include contributions that reduce to $\mathcal{O}_1$ in the low energy limit; the corresponding Feynman diagrams are shown in Fig. 2.
In this model, there are two different box (and crossed box) diagrams: one involves the scalar coupling λ_1 and the other involves the pseudoscalar coupling λ_2. The triangle diagram involves the scalar coupling λ_1 between the DM and the mediator,[7] as well as the cubic self–interaction of the mediator μ_1. All diagrams contain two vertices with the pseudoscalar coupling h_2^q on the quark line.

[7] There is also a triangle diagram involving λ_2, as well as (crossed) box diagrams with one λ_1 vertex and one λ_2 vertex; however, these diagrams do not contribute to $\mathcal{O}_1$. Moreover, the diagrams shown in Fig. 2 also generate other operators in the non–relativistic limit. We ignore these contributions, since they are of higher order in both the loop and NREFT expansions.



(a) Box and Crossed Box with λ_1

(b) Box and Crossed Box with λ_2



(c) Triangle

Figure 2: One–loop box and triangle Feynman diagrams in Model IIa which contribute to $\mathcal{O}_1$. The shaded blob denotes the scalar DM–mediator coupling λ_1 while the dark blob denotes the pseudoscalar DM–mediator coupling λ_2.

In order to simplify the loop calculation, we have used the Dirac equation as well as 4−momentum conservation, and took the limit of vanishing momentum transfer. Further details can be found in Appendix A.2. Finally, we find the following effective Lagrangian for Model IIa:

$$\mathcal{L}_{\text{eff}}^{\text{IIa}} \supset c_{10}^{q,d6}\,(\bar{\chi}\chi)\left(\bar{q}i\gamma^5 q\right) + c_{6}^{q,d6}\left(\bar{\chi}i\gamma^5\chi\right)\left(\bar{q}i\gamma^5 q\right) + c_{1,S}^{q,d6}\,(\bar{\chi}\chi)\,(\bar{q}q) + c_{1,V}^{q,d6}\,(\bar{\chi}\gamma_\mu\chi)\,(\bar{q}\gamma^\mu q)\ ; \tag{2.15}$$

the notation is as in (2.5). For completeness we have retained $\left(\bar{\chi}\, i\gamma^5\,\chi\right)\left(\bar{q}\, i\gamma^5\, q\right)$ in the set of effective operators, although it can safely be neglected as reasoned earlier. Table 3 displays the list of relativistic effective operators and their matching to the NREFT operators in terms of the parameters of Model IIa. The quark–level Wilson coefficients contain loop functions $N_k(r,s),\ k=1,\ldots,8$, $P_l(r,s),\ l=1,\ldots,4$ and $R_1(r)$, which have been expressed as functions of dimensionless parameters $r \equiv m_q/m_\phi$ and $s \equiv m_\chi/m_\phi$. Analytic expressions for these loop functions can be found in Appendix A.2.1.

Collecting the results of Table 3, and embedding the quark bilinears in nucleonic matrix

	$\bar{\chi}\Gamma_\chi\chi\ \bar{q}\,\Gamma_q\,q$		c_i^q
Tree	$c_{10}^{q,d6}\bar{\chi}\chi\ \bar{q}i\gamma^5 q$	$\longrightarrow$	$\dfrac{h_2^q\,\lambda_1}{m_\phi^2}$
Box$\vert_{\lambda_1}$	$c_{1,V,B1}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$-\dfrac{\lambda_1^2\,(h_2^q)^2}{16\pi^2}\,N_1$
	$c_{1,S,B1}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$-\dfrac{\lambda_1^2\,(h_2^q)^2}{16\pi^2}\,\big(2m_\chi m_q(N_2-P_2)+m_\chi^2(N_3-2P_1)+m_q^2N_4\big)$
Crossed	$c_{1,V,C1}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$\dfrac{\lambda_1^2\,(h_2^q)^2}{16\pi^2}\,N_5$
Box$\vert_{\lambda_1}$	$c_{1,S,C1}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$\dfrac{\lambda_1^2\,(h_2^q)^2}{16\pi^2}\,\big(2m_\chi m_q(N_6-P_4)+m_\chi^2(N_7-2P_3)+m_q^2N_8\big)$
Box$\vert_{\lambda_2}$	$c_{1,V,B2}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$-\dfrac{\lambda_2^2\,(h_2^q)^2}{16\pi^2}\,N_1$
	$c_{1,S,B2}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$-\dfrac{\lambda_2^2\,(h_2^q)^2}{16\pi^2}\,\big(2m_\chi m_q N_2+m_\chi^2N_3+m_q^2N_4\big)$
Crossed	$c_{1,V,C2}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$\dfrac{\lambda_2^2\,(h_2^q)^2}{16\pi^2}\,N_5$
Box$\vert_{\lambda_2}$	$c_{1,S,C2}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$\dfrac{\lambda_2^2\,(h_2^q)^2}{16\pi^2}\,\big(2m_\chi m_q N_6+m_\chi^2N_7+m_q^2N_8\big)$
Triangle	$c_{1,S,T}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$-\dfrac{\lambda_1\,\mu_1\,(h_2^q)^2}{16\pi^2}\,\dfrac{m_q}{m_\phi}R_1$

Table 3: Non–relativistic reduction of effective operators in Model IIa. In the expression for the quark–level Wilson coefficients, we have again suppressed the dependence of the loop functions on dimensionless parameters, i.e. $N_k \equiv N_k(r,s)$, $P_l \equiv P_l(r,s)$ and $R_1 \equiv R_1(r)$, where $r \equiv m_q/m_\phi$ and $s \equiv m_\chi/m_\phi$. The first line contributes to the coefficient of $\mathcal{O}_{10}$, all other lines describe contributions $\propto \mathcal{O}_1$.

elements, the Wilson coefficient c_1^N of the NREFT operator $\mathcal{O}_1$ is given by

$$
\begin{aligned}
c_1^N|^{\rm IIa} = -\frac{1}{16\pi^2}\Bigg[& (\lambda_1^2+\lambda_2^2)\Big(\sum_{q=u,d}(h_2^q)^2\mathcal{N}_q^N(N_1(r,s)-N_5(r,s))\Big) \\
& + m_N\Big(\sum_{q=u,d,s}(h_2^q)^2 f_{Tq}^N + \frac{2}{27}f_{TG}^N\sum_{q=c,b,t}(h_2^q)^2\Big) \\
& \times\Big\{(\lambda_1^2+\lambda_2^2)\Big(2m_\chi\big(N_2(r_q,s)-N_6(r_q,s)\big)+\frac{m_\chi^2}{m_q}\big(N_3(r_q,s)-N_7(r_q,s)\big) \\
& + m_q\big(N_4(r_q,s)-N_8(r_q,s)\big)\Big) \\
& - 2m_\chi\lambda_1^2\Big(P_2(r_q,s)-P_4(r_q,s)+\frac{m_\chi}{m_q}\big(P_1(r_q,s)-P_3(r_q,s)\big)\Big) \\
& + \frac{\lambda_1\,\mu_1}{m_\phi}R_1(r_q)\Big\}\Bigg]\,. \qquad (2.16)
\end{aligned}
$$

The nuclear bilinear coefficients $\mathcal{N}_q^N$, f_{Nq}^T and f_{NG}^T are the same as in eq.(2.6). At first sight $c_1^N|^{\rm IIa}$ appears to be singular in the limit $m_q \to 0$. Note that $f_{N_q}^T$ contains a factor of m_q, see eq.(2.7); hence contributions to $c_{1,S}^{q,d6}$ in Table 3 without an explicit factor m_q appear $\propto 1/m_q$ in eq.(2.16). However, they get multiplied with the differences of loop functions $N_3(r_q,s)-N_7(r_q,s)$ or $P_1(r_q,s)-P_3(r_q,s)$, which scale like r_q for $r_q \to 0$; these contributions therefore approach a finite value as $m_q \to 0$. In contrast, the difference $N_1(r_q,s)-N_5(r_q,s)$, which appears in the contribution from the product of two vector currents, vanishes for $r_q \to 0$. As a result, the dominant contributions to $c_1^N|^{\rm IIa}$ also originate from the product of two scalar currents, as for Model I. Within this category, the contribution $\propto N_4 - N_8$ vanishes as $r_q \to 0$ and is therefore negligible for generation–independent couplings h_2^q, but all other terms are very roughly comparable and must be taken into account.
The Wilson coefficient of the operator $\mathcal{O}_{10}$ arising from the $t-$channel tree–level scattering diagram is

$$c_{10}^N|^{\rm IIa} = \frac{\lambda_1}{m_\phi^2}\left(\sum_{q=u,d,s} h_2^q\,\Delta\tilde{q}^N - \Delta\tilde{G}^N \sum_{q=c,b,t}\frac{h_2^q}{m_q}\right). \tag{2.17}$$

2.2.2 Model IIb

We next turn to Model IIb. It is also described by the Lagrangian of eq.(2.11), but we now postulate purely scalar Yukawa couplings h_1^q on the quark side and a purely pseudoscalar coupling λ_2 on the dark matter side, see eq.(2.12). The tree–level matrix element for DM–quark scattering, $\chi(p_\chi) + q(p_q) \to \chi(p'_\chi) + q(p'_q)$, proceeding via $t-$channel ϕ exchange is then:

$$\mathcal{M}_{\chi q\to\chi q}^{\rm IIb} = -\frac{h_1^q\,\lambda_2}{q^2-m_\phi^2}\,\bar{u}(p'_\chi)i\gamma^5 u(p_\chi)\,\bar{u}(p'_q)u(p_q)\,. \tag{2.18}$$

This matrix element matches onto the dimension-6 effective operator $\left(\bar{\chi}i\gamma^5\chi\right)(\bar{q}q)$, which reduces to the momentum–suppressed SI NREFT operator $\mathcal{O}_{11}$ in the non–relativistic limit. Once again, the choice $\lambda_1 = 0$, which ensures the absence of $\mathcal{O}_1$ at tree–level, is not protected by any symmetry. We therefore again expect contributions $\propto \mathcal{O}_1$ to be generated at the next order in perturbation theory. The relevant Feynman diagrams are shown in Fig. 3. Each diagram involves two factors of the pseudoscalar DM–mediator coupling λ_2; the two γ_5 factors multiply to unity. The triangle and (crossed) box diagrams in addition involve one or two factors of the scalar DM–quark coupling h_1^q, respectively; the former also involves the cubic self–coupling μ_1 of the mediator. The triangle diagram can be interpreted as generating a scalar DM–mediator coupling. Explicit expressions for the resulting amplitudes are provided in Appendix A.2.
When the mediator ϕ is integrated out, the following relativistic effective Lagrangian describes DM–quark scattering:

$$\mathcal{L}_{\rm eff}^{\rm IIb} \supset c_{11}^{q,d6}\left(\bar{\chi}i\gamma^5\chi\right)(\bar{q}q) + c_{1,S}^{q,d6}\left(\bar{\chi}\chi\right)(\bar{q}q) + c_{1,V}^{q,d6}\left(\bar{\chi}\gamma_\mu\chi\right)(\bar{q}\gamma^\mu q)\,. \tag{2.19}$$

Table 4 displays the quark–level Wilson coefficients in terms of the parameters of Model IIb. The loop functions N_k and P_l are identical to the ones appearing in Model IIa. S_1 is

	$\bar{\chi}\Gamma_\chi\chi\ \bar{N}\,\Gamma_N\,N$		c_i^q
Tree	$c_{11}^{q,d6}\bar{\chi}i\gamma^5\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$\dfrac{h_1^q\lambda_2 m_N}{m_\phi^2 m_\chi}$
Box	$c_{1,V,B}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$-\dfrac{\lambda_2^2\,h_1^{q\,2}}{16\pi^2}\,N_1$
	$c_{1,S,B}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$-\dfrac{\lambda_2^2\,h_1^{q\,2}}{16\pi^2}\,\big(2m_\chi m_q(N_2+P_1)+m_\chi^2 N_3+m_q^2(N_4+2P_2)\big)$
Crossed Box	$c_{1,V,C}^{q,d6}\bar{\chi}\gamma_\mu\chi\ \bar{q}\gamma^\mu q$	$\longrightarrow$	$\dfrac{\lambda_2^2\,h_1^{q\,2}}{16\pi^2}\,N_5$
	$c_{1,S,C}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$\dfrac{\lambda_2^2\,h_1^{q\,2}}{16\pi^2}\,\big(2m_\chi m_q(N_6-P_3)+m_\chi^2 N_7+m_q^2(N_8-2P_4)\big)$
Triangle	$c_{1,S,T}^{q,d6}\bar{\chi}\chi\ \bar{q}q$	$\longrightarrow$	$-\dfrac{h_1^q\,\mu_1\,\lambda_2^2}{16\pi^2}\,\dfrac{m_\chi}{m_\phi}S_1$

Table 4: Non–relativistic reduction of effective operators in Model IIb. The arguments of the loop functions N_k, P_l and S_1 have again been suppressed; explicit expressions for these functions can be found in Appendix A.2.1.

the only loop function not defined previously, and its analytic expression along with that of the others can be found in Appendix A.2.1.

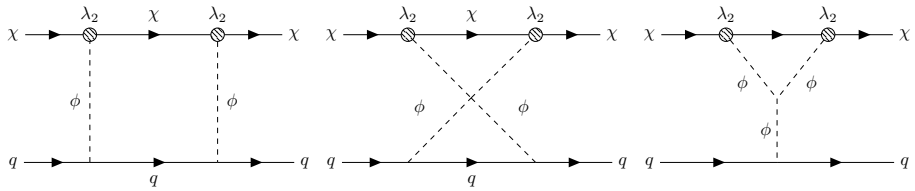


Figure 3: One–loop (crossed) box and triangle Feynman diagrams in Model IIb which contribute to $\mathcal{O}_1$ in the non–relativistic limit.

Collecting the results of Table 4, and inserting the appropriate hadronic coefficients for the nucleonic matrix elements of the quark bilinears, the Wilson coefficient of the operator $\mathcal{O}_1$

is:

$$
\begin{aligned}
c_1^N|^{\rm IIb} = -\frac{\lambda_2^2}{16\pi^2}\Bigg\{ & \sum_{q=u,d} \mathcal{N}_q^N (h_1^q)^2 \Big(N_1(r_q,s) - N_5(r_q,s) \Big) \\
& + m_N \Big(\sum_{q=u,d,s} f_{Tq}^N + \frac{2}{27} f_{TG}^N \sum_{q=c,b,t} \Big) \\
& \times \Bigg[(h_1^q)^2 \Bigg(2m_\chi \Big(N_2(r_q,s) - N_6(r_q,s) + P_1(r_q,s) + P_3(r_q,s) \Big) \\
& \qquad + m_q \Big(N_4(r_q,s) - N_8(r_q,s) + 2\Big(P_2(r_q,s) + P_4(r_q,s) \Big) \Big) \\
& \qquad + \frac{m_\chi^2}{m_q} \Big(N_3(r_q,s) - N_7(r_q,s) \Big) \Bigg) \\
& + \mu_1 h_1^q \frac{m_\chi}{m_\phi} \frac{S_1(s)}{m_q} \Bigg] \Bigg\} \; ;
\end{aligned}
\tag{2.20}
$$

recall that $r_q = m_q/m_\phi$ and $s = m_\chi/m_\phi$. As in case of Model IIa, the combinations of loop functions multiplying $\mathcal{N}_q^N$ and m_χ^2/m_q vanish for $m_q \to 0$; the term $\propto m_\chi^2/m_q$ thus yields a finite result in this limit. However, the very last term in eq.(2.20) also contains an explicit $1/m_q$ factor; the loop function appearing in this term does not depend on m_q at all. Here the required chirality breaking on the quark line is due to the *single* factor of h_1^q. We therefore expect this term to dominate, unless the trilinear scalar coupling μ_1 is for some reason very small.

From eq.(2.18) the tree–level contribution to the Wilson coefficient of $\mathcal{O}_{11}$ is:

$$
c_{11}^N|^{\rm IIb} = \frac{\lambda_2}{m_\phi^2} \frac{m_N^2}{m_\chi} \Big(\sum_{q=u,d,s} \frac{h_1^q f_{Tq}^N}{m_q} + \frac{2}{27} f_{TG}^N \sum_{q=c,b,t} \frac{h_1^q}{m_q} \Big) = \frac{\lambda_2}{m_\phi^2} \frac{m_N^2}{m_\chi} \tilde{f}^N \; . \tag{2.21}
$$

Here we have defined $\tilde{f}^N \equiv \Big(\sum_{q=u,d,s} \frac{h_1^q f_{Tq}^N}{m_q} + \frac{2}{27} f_{TG}^N \sum_{q=c,b,t} \frac{h_1^q}{m_q} \Big)$, and as usual neglected the q^2 term in the ϕ propagator. As in the last term in eq.(2.20) the required chirality breaking on the quark line is provided by h_1^q. Since the hadronic matrix elements f_{Tq}^N have been defined including an explicit factor of m_q, the contributions in $\tilde{f}^N$ scale $\propto 1/m_q$, leading to a large enhancement of the contribution of light quarks. On the other hand, an extra factor m_N/m_χ appears since the pseudoscalar DM current is $\propto |\vec{q}|/m_\chi = (|\vec{q}|/m_N) \times (m_N/m_\chi)$. As already noted in the discussion of Table 1, this appears to be quite generic [47–50].

2.3 The Neutron EDM

In all three cases we considered, the tree–level contribution to the Wilson coefficient of $\mathcal{O}_{10}$ or $\mathcal{O}_{11}$ is nonzero only if CP is violated: in Model I one needs $g_1 h_2^q \neq 0$, where g_1 is a scalar coupling while h_2^q is a pseudoscalar coupling; in Model IIa, $\lambda_1 h_2^q \neq 0$ is required, where λ_1 is a scalar coupling; and in Model IIb, $c_{11}^{d6} \propto \lambda_2 h_1^q$, where λ_2 is pseudoscalar but h_1^q is scalar. Since we assume the quark Yukawa couplings to be flavor diagonal, the most sensitive probe

of CP violation is the EDM of the neutron (or of heavier nuclei). However, CP violation is a necessary condition for the generation of an EDM; it is not by itself sufficient. In the case at hand, since we set $h_1^q = 0$ in Models I and IIa and $h_2^q = 0$ in Model IIb, there is no one–loop contribution to the neutron EDM, in contrast to the models with charged mediator we considered in ref.[50]. In case of Model IIb, as far as the one–loop diagram with a quark and a mediator ϕ in the loop is concerned, the mediator can consistently be defined as being $CP-$even, since only the coupling h_1^q appears in the diagram. Similarly, in Model I and Model IIa, in the one–loop diagram with a quark–mediator loop the mediator can consistently be defined as being $CP-$odd, since only the coupling h_2^q appears.
Turning to higher loops, a $CP-$even mediator can have any self coupling. Hence in Model IIb embellishing the one–loop diagram with additional scalar vertices does not lead to CP violation. In fact, in this model the only coupling that is not consistent with interpreting ϕ to be $CP-$even is λ_2, which couples ϕ to the DM particle. However, this coupling cannot contribute to electric dipole moments. It could appear in any diagram where the only external particles are a through–going quark and a photon (or gluon) only via a closed χ loop, which will either vanish (if an odd number of ϕ legs is attached to it), or simply renormalize a $CP-$even quantity like the ϕ $2-$point function. We thus conclude that in Model IIb, *no* new contributions to the electric dipole moments of SM particles are generated. The conceptually easiest way to prove the existence of CP violation in this model is via χq scattering; in case of $2 \to 2$ scattering, spin observables would have to be included in the construction of a $CP-$odd quantity. While conceptually straightforward, experimentally this seems prohibitively difficult; certainly there are no current experimental constraints from such experiments.
The situation is very different in Models I and IIa, where in the relevant one–loop diagram the mediator behaves like a pseudoscalar. The reason is that a pseudoscalar cannot have a ϕ^3 coupling. Hence, two–loop diagrams containing both the (pseudoscalar) Yukawa coupling h_2^q and the (scalar) trilinear coupling μ_1, see Fig. 4, can be expected to generate a nonvanishing (chromo–)EDM for quark q. It should be noted that here μ_1 is relevant, not the couplings λ_1 or g_1 appearing in the coefficient of $\mathcal{O}_{10}$. However, a theory with $\mu_1 = 0$ but $\lambda_1 \neq 0$ is, strictly speaking, not renormalizable, since a triangle diagram with χ in the loop will generate a *divergent* contribution to $\mu_1 m_\phi \propto m_\chi \lambda_1^3$; similarly, in Model I there are divergent one–loop contributions $\propto g_1 g_2 m_S$ to the ϕ^3 vertex. We will come back to this point later. Because only the trilinear self–coupling of the mediator and its pseudoscalar Yukawa coupling to quarks are relevant here, the calculation of the quark (chromo) EDM is exactly the same in these two models.
The quark EDM d_q is calculated as the coefficient of a dimension-5 $P-$ and $T-$odd interaction term $(-i/2)\,\bar{q}\,\sigma_{\mu\nu}\gamma_5\,q\,F^{\mu\nu}$ at zero momentum transfer. The same quark radiating a gluon instead a photon leads to non–vanishing chromo EDM. These are calculated similar to quark EDMs, by finding the coefficient of $(-i/2)\,\bar{q}\,\sigma_{\mu\nu}t_a\gamma_5\,q\,G_a^{\mu\nu}$ at zero momentum transfer.
The EDM operator breaks chirality, hence d_q is proportional to an odd number of chirality flips. These can come either from fermion masses or from Yukawa couplings in the relevant Feynman diagrams. The diagrams shown in Fig. 4 contain three Yukawa couplings, hence

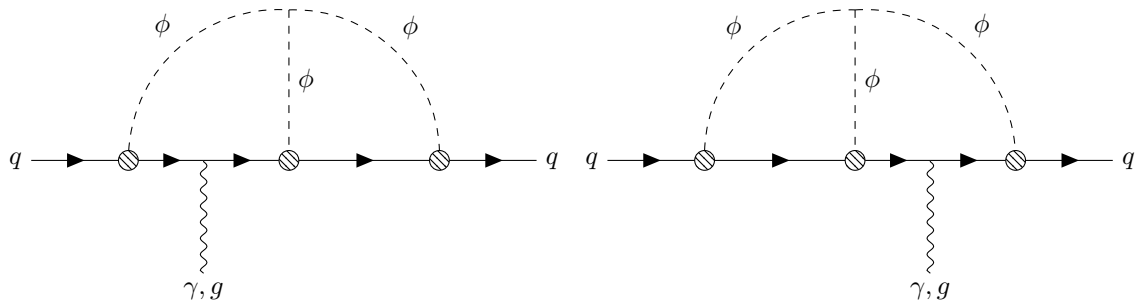


Figure 4: Two–loop Feynman diagrams for quark EDMs and color–EDMs in Model I and IIa. The blobs here indicate the insertion of the mediator pseudoscalar Yukawa coupling.

they can contribute even for $m_q \to 0$. We provide the details of the computation of these two–loop diagrams in Appendix B. We calculate the quark EDM to be

$$d_q = \frac{2e\,Q_q\,(h_2^q)^3}{(16\pi^2)^2}\,\mu_1\,m_\phi \lim_{q^2\to 0}\Big([\mathcal{X}]+[\mathcal{Y}]\Big)\,, \tag{2.22}$$

where Q_q is the electric charge of quark q; $[\mathcal{X}]$ and $[\mathcal{Y}]$ are loop functions expressed as five dimensional integrals over five Feynman parameters given in eqs.(B.19) and (B.20). The color–EDM $\tilde{d}_q$ can be obtained by replacing the external photon with a gluon. Hence $\tilde{d}_q$ can be obtained from eq.(2.22) by replacing eQ_q with the strong coupling g_s.
In order to calculate the value of the nEDM from d_q and $\tilde{d}_q$, we use

$$d_n = g_T^u d_u + g_T^d d_d + g_T^s d_s + 1.1\,e\,(0.5\,\tilde{d}_u + \tilde{d}_d)\,. \tag{2.23}$$

Here the tensor charges $g_T^u = -0.233(28)$, $g_T^d = 0.774(66)$ and $g_T^s = 0.009(8)$ have been calculated using lattice QCD [59, 60] (see also Refs.[61–63]) at a renormalization scale of 2 GeV. We are not aware of a reliable lattice computation of the contribution of the chromo-EDMs to d_n; we therefore employ a computation using QCD sum rules, again evaluated at a renormalization scale of 2 GeV [64], although there is an $\mathcal{O}(50\%)$ uncertainty in these results [65, 66]. In this case the uncertainty in the coefficients in Eq.(2.23) might shift the boundary of the excluded region slightly, without affecting our results qualitatively. Since in our numerical analyses we will assume flavor universal quark Yukawa couplings, the contributions from the (chromo–)EDMs of heavy quarks to the EDM of the neutron can safely be neglected; in fact, since $g_T^s \ll g_T^d$ already the contribution from the strange quark is essentially negligible for equal couplings.

3 Results and Discussions

In the previous section, we described the different contributions to DM–nucleus scattering in Models I, IIa and IIb. These simplified models were designed such there are no tree–level contributions to the NREFT operator $\mathcal{O}_1$; instead, at tree–level only the $P-$ and $T-$odd operators $\mathcal{O}_{10}$ (in Model I and IIa) or $\mathcal{O}_{11}$ (in Model IIb) were generated. Recall that $\mathcal{O}_1$ often gives the dominant contribution to DM–nucleus scattering: it appears in

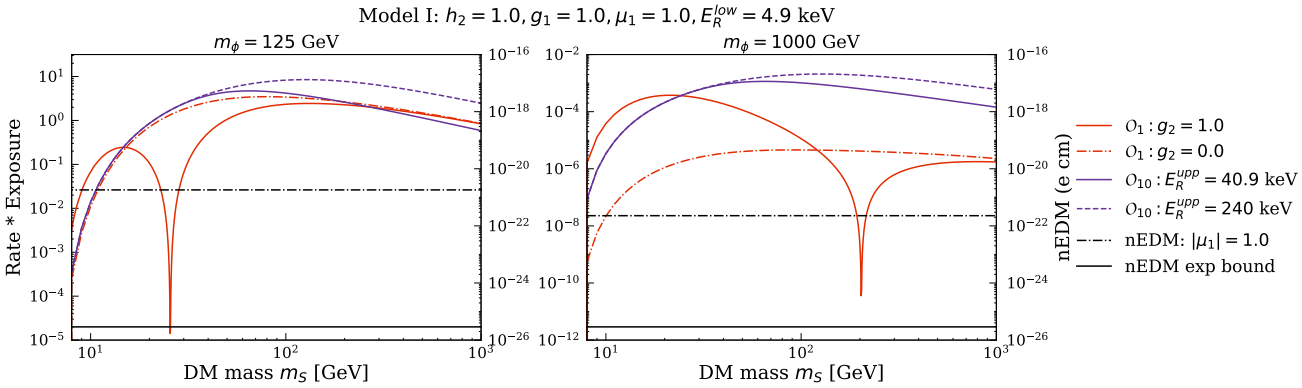


Figure 5: The left $y-$axis shows the total number of scattering events in Model I with $\mu_1 = 1.0$ at XENON1T with a run time exposure of 1 tonne–year as a function of the DM mass m_S for two values of the mediator mass, $m_\phi = 125$ GeV (left frame) and $m_\phi = 1000$ GeV (right frame). The **violet** and **red** curves show the number of events due to $\mathcal{O}_{10}$ and $\mathcal{O}_1$, respectively. The **black** lines refer to the right $y-$axis; the **solid black** line is the experimental upper limit on nEDM at 90% C.L., while the **dashed black** line is the predicted nEDM for the given choice of parameters.

leading order in the NREFT expansion, and its contribution is coherently enhanced by A^2 if $c_1^n \simeq c_1^p$, where c_1^n and c_1^p are the Wilson coefficients accompanying $\mathcal{O}_1$ for neutron and proton respectively. However, in the models considered here, $\mathcal{O}_1$ arises from box and triangle diagrams appearing at next–to–leading order in perturbation theory. Therefore, it is not obvious *a priori* which operator provides the dominant contribution to DM–nucleus scattering.

In this section we compare the contributions from $\mathcal{O}_1$ quantitatively to the contributions from $\mathcal{O}_{10}$ or $\mathcal{O}_{11}$. To that end, we compute the number of events due to these operators for a recent XENON1T run [2] for a variety of benchmark model parameters. Here we use the explicit expressions for the loop functions contributing to the Wilson coefficient of $\mathcal{O}_1$ given in Appendix A. It should be noted that the contribution from $\mathcal{O}_1$ does not interfere with those from $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$, due to the different CP properties of these operators. Hence the total scattering rate is simply given by the sum of these contributions.[8]

Moreover, for Models I and IIa we check numerically if the nEDM surpasses its experimental upper bound [44–46]. We perform the integration of the loop functions in eq.(2.22) using the Monte Carlo integration routine `SUAVE` of the `CUBA` numerical library [67]. The absolute numerical error reported by the routine is around $\mathcal{O}(10^{-8})$ for the two values of the mediator mass m_ϕ we consider.

In order to compute the number of events for the 2018 run of XENON1T [2], we integrate the differential event rate, computed using the the Mathematica code `dmformfactor`, which is based on the formalism of ref.[20], over the recoil energy E_R from 4.9 keV to 40.9 keV,

[8] Recall that we only extract the contribution to $\mathcal{O}_1$ from our loop diagrams. These diagrams will also contribute to additional NREFT operators; however, those contributions will be of higher order both in the loop and NREFT expansions, and can thus safely be neglected.

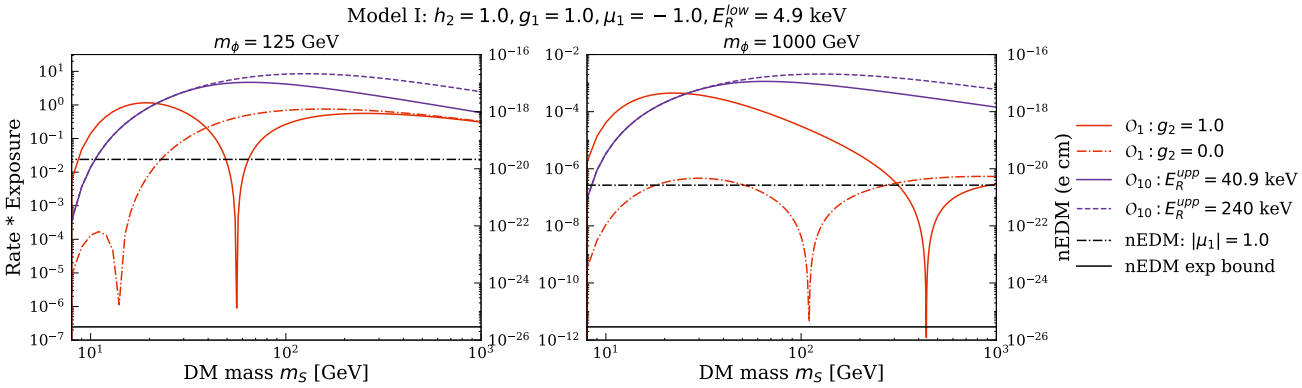


Figure 6: As in Fig. 5, except for $\mu_1 = -1.0$.

in accordance with the recoil energy region of interest (ROI) used in that run. In addition, we include results (shown as dashed curves) where the upper end of integration is set at $E_R^{\rm upp} = 240$ keV, as considered in a dedicated effective field theory search by the XENON experiment. We multiply the integrated rate with the 2018 XENON1T runtime exposure of 278.8 days $\times$ 1.3 tonne = 1.0 tonne-yr. We have further assumed a standard isothermal DM halo with $\rho_\chi = 0.3$ GeVcm^{-3}, $v_0 = 220$ km/s, $v_e = 232$ km/s and $v_{\rm esc} = 544$ km/s. We calculate our event rates using a weighted sum over the different isotopes occurring in naturally abundant Xenon.

We choose the following sets of parameter values for Model I: $h_2^q \equiv h_2 = g_1 = 1.0$ with $m_\phi = 125$ GeV or $m_\phi = 1000$ GeV. Note that we choose a rather large and flavor universal Yukawa coupling of the quarks. For the given couplings, the former (smaller) mediator mass value leads to a DM–nucleon cross section from the operator $\mathcal{O}_{10}$ that lies just below the sensitivity of the 2018 XENON1T run in the $\sigma_{\chi p} - m_S$ exclusion plane. The latter (larger) mediator mass corresponds to a DM–nucleon cross section from the operator $\mathcal{O}_{10}$ that lies just above the irreducible background from coherent neutrino–nucleus scattering (the so–called "neutrino floor"); it will be very difficult to probe even smaller cross sections in direct search experiments.

Figures 5 and 6 depict the resulting number of events from the operators $\mathcal{O}_{10}$ and $\mathcal{O}_1$ as a function of the DM mass m_S in Model I, for different values of the trilinear mediator coupling μ_1. Each figure contains two frames, corresponding to the two values of m_ϕ we consider.

Since the tree–level contribution from $\mathcal{O}_{10}$ does not depend on μ_1, it remains the same in both figures. We note that this contribution is enhanced by up to an order of magnitude once the DM mass m_S exceeds the mass of the target nuclei ($\simeq 125$ GeV for Xenon) if the upper end of the ROI is increased from 40.9 keV to 240 keV. This is not surprising, since $\mathcal{O}_{10}$ predicts a recoil energy spectrum that peaks at sizable energy, due to the explicit factor of $\vec{q}$ that appears in the definition of this operator. Note that at $m_S < m_{\rm Xe}$, the maximal recoil energy is $\mathcal{O}(v^2 m_S^2/m_{\rm Xe})$ rather than $\mathcal{O}(v^2 m_{\rm Xe})$ for $m_S > m_{\rm Xe}$; hence this enhancement only occurs at larger DM masses.

Figure 5 is for $\mu_1 = 1.0$. We see that the contribution from $\mathcal{O}_{10}$ generally exceeds that

from $\mathcal{O}_1$, except at low DM mass where the low maximal recoil energy leads to low values of $|\vec{q}|$. We also note that for $g_2 = 0$, i.e. in the absence of the last diagram in Fig. 1, the $\mathcal{O}_1$ contribution, shown by the dashed red lines, drops faster with increasing m_ϕ than the (tree–level) contribution from $\mathcal{O}_{10}$ does. For a positive coupling $g_2 = 1.0$ the contributions from the (crossed) box and g_2 triangle diagrams have opposite signs, leading to a vanishing $\mathcal{O}_1$ contribution for $m_S \simeq 0.2 m_\phi$.

In order to understand the comparison of the tree–level $\mathcal{O}_{10}$ contributions with the loop suppressed $\mathcal{O}_1$ contributions semi–quantitatively, we estimate the ratio of the number of events from the two contributions $N_{\mathcal{O}_1}/N_{\mathcal{O}_{10}}$ as:

$$\left.\frac{N_{\mathcal{O}_1}}{N_{\mathcal{O}_{10}}}\right|_{\rm I} = \left.\frac{R_{\mathcal{O}_1}}{R_{\mathcal{O}_{10}}}\right|_{\rm I} \sim \frac{(c_1^N|^{\rm I})^2}{(c_{10}^N|^{\rm I})^2}\frac{A^2}{\langle S_{\rm Xe}\rangle^2}\,\frac{m_N^2}{\vec{q}^{\,2}}\,. \tag{3.1}$$

Here $R_{\mathcal{O}_i}$ denotes the scattering rate due to operator $\mathcal{O}_i$ integrated over the recoil energy window. The factor $(\vec{q}^{\,2}/m_N^2)^{-1}$ is due to the momentum suppression of $\mathcal{O}_{10}$. In this estimate, we have assumed the nuclear response functions to be independent of the recoil energy, and approximated the ratio of the nuclear response functions as $A^2/\langle S_N\rangle^2$ since the SI response (from $\mathcal{O}_1$) is coherently enhanced by A^2 while the SD response (from $\mathcal{O}_{10}$) is suppressed by the spin expectation value squared $\langle S_{\rm Xe}\rangle^2$. We estimate $c_1^N|^{\rm I}$ using the vanishing quark mass limit of eq.(2.6):

$$c_1^N|^{\rm I}_{m_q\to 0} = \frac{h_2^2}{16\pi^2} m_N f_T^N \left\{ g_1^2 m_S (M_2 + M_4)|_{r\to 0} + \left(\frac{g_1\mu_1}{m_\phi} - \frac{g_2}{2m_S}\right) L_1(r)|_{r\to 0}\right\}\,. \tag{3.2}$$

For small $s = m_S/m_\phi$ and $g_2 = 1.0$, the last (triangle) term in eq.(3.2) dominates:

$$c_1^N|^{\rm I} \approx \frac{h_2^2}{16\pi^2}\frac{g_2}{4} m_N f^N \frac{1}{m_S}\frac{1}{m_\phi^2}\,, \tag{3.3}$$

where we used $L_1(0) = -1/2m_\phi^2$ and defined $f^N \equiv \left(\sum_{u,d,s} f_{Tq}^N + \frac{2}{27} f_{TG}^N \sum_{c,b,t}\right)$. Inserting eq.(3.3) and eq.(2.9) for $c_{10}^N|^{\rm I}$ in eq. (3.1), we obtain for $m_S < m_\phi = 1000$ GeV:

$$\left.\frac{N_{\mathcal{O}_1}}{N_{\mathcal{O}_{10}}}\right|_{\rm I} \approx \left(\frac{h_2}{16\pi^2}\frac{g_2}{4g_1}\frac{f_T^N}{\tilde{\Delta}^N}\frac{m_N}{m_S}\right)^2 \frac{A}{\langle S_N\rangle^2}\frac{m_N}{2E_R}\,, \tag{3.4}$$

where we defined $\tilde{\Delta}^N = \left(\sum_{q=u,d,s}\Delta\tilde{q}^N - \Delta\tilde{G}^N \sum_{q=c,b,t}\frac{1}{m_q}\right)$ and used $m_{\rm Xe} = A m_N$. Plugging in $h_2 = g_1 = g_2 = 1.0$, $E_R = 30$ keV for Xenon ($A = 131$) at $m_S = 30$ GeV results in a ratio of 0.8, which agrees well with the solid red curve for $\mathcal{O}_1$ being marginally below the purple curve for $\mathcal{O}_{10}$ in the right frame of Figure 5.

For $m_\phi = 125$ GeV and not too large m_S $c_1^N|^{\rm I}$ is instead dominated by the scattering contribution from the g_1 triangle diagram since $g_1\mu_1/m_\phi$ is less suppressed than $g_2/2m_S$. In that case we should replace $g_2/(2m_S)$ by $g_1\mu_1/m_\phi$ in eq.(3.4). Again using $E_R = 30$ keV and $h_2 = g_1 = \mu_1 = 1.0$, the estimated contribution from $\mathcal{O}_1$ is then 0.2 times that from $\mathcal{O}_{10}$. This is in good agreement with the left panel of Figure 5 for $m_S \geq 100$ GeV, above the accidental cancellation around $m_S \simeq 30$ GeV.

For $m_S > m_\phi$ and unit couplings, the box diagrams instead dominate, in which case

$$c_1^N|^{\rm I} \approx \frac{h_2^2 g_1^2}{16\pi^2} f_T^N m_N m_S \, (M_2 + M_4)|_{r\to 0} \ . \tag{3.5}$$

The ratio of the number of scattering events then becomes

$$\left.\frac{N_{\mathcal{O}_1}}{N_{\mathcal{O}_{10}}}\right|_{\rm I} \approx \left(\frac{h_2 g_1}{16\pi^2}\frac{f_T^N}{\tilde{\Delta}^N} m_N \, m_S \, m_\phi^2 \, (M_2 + M_4)\,|_{r\to 0}\right)^2 \frac{A}{\langle S_N\rangle^2}\frac{m_N}{2E_R} \ . \tag{3.6}$$

Using $E_R = 30$ keV and $h_2 = g_1 = \mu_1 = 1.0$, for $m_S = 1000$ GeV and $m_\phi = 125$ GeV the estimated number of events for $\mathcal{O}_1$ is 2.8 times that for $\mathcal{O}_{10}$. The left panel of Fig. 5 shows that this is a slight overestimate even if the smaller ROI is used. For the larger maximal recoil energy, $E_R^{\rm upp} = 240$ keV, $\mathcal{O}_{10}$ still contributes roughly three times more scattering events than $\mathcal{O}_1$. Using the same couplings, for $m_S = m_\phi = 1000$ GeV, the estimated contribution from $\mathcal{O}_1$ is smaller by a factor 2.5×10^{-2} relative to that from $\mathcal{O}_{10}$. This rough estimate matches the ratio between the $\mathcal{O}_1$ and $\mathcal{O}_{10}$ contributions in the right frame of Figure 5.

In Fig. 6 we have changed the sign of μ_1 while keeping g_1 and g_2 positive (or zero). Both triangle diagrams therefore now contribute with opposite signs than the (crossed) box diagrams. As a result, the coefficient of $\mathcal{O}_1$ now vanishes at $m_S \simeq 0.5 m_\phi$ for $g_2 = 1.0$, and at $m_S \simeq 0.1 m_\phi$ for $g_2 = 0$.

As before, for $g_2 = 1.0$ and small DM masses the Wilson coefficient of $\mathcal{O}_1$ can be approximated by the triangle diagram involving the g_2 coupling. From eq.(3.4) for $m_S = 100$ GeV, $m_\phi = 1000$ GeV and $h_2 = g_1 = 1.0$ we estimate 7.2×10^{-2} as ratio of the $\mathcal{O}_1$ and $\mathcal{O}_{10}$ contributions, in rough agreement with the right panel of Fig. 6. For $m_S > m_\phi$, the box diagrams again dominate. Since they do not depend on μ_1 our earlier estimates still apply. However, since the triangle diagrams aren't quite negligible even at $m_S = 8 m_\phi$, the largest ratio covered in Fig. 6, this somewhat overestimates the importance of $\mathcal{O}_1$.

Before turning to the neutron EDM, we discuss results for DM–Xenon scattering for Model IIa. We again chose a flavor–universal Yukawa coupling $h_2^q \equiv h_2 = 1$, and the same two values of mediator mass as for Model I. The coupling λ_1 appearing in c_{10} is set to 1, and we show results for the pseudoscalar mediator–WIMP coupling $\lambda_2 = 1$ or 0.

Figure 7 shows the number of scattering events from $\mathcal{O}_{10}$ and $\mathcal{O}_1$ for $\mu_1 = 1.0$. We again find that for the large mediator mass ($m_\phi = 1000$ GeV, right frame) the tree–level contribution from $\mathcal{O}_{10}$ dominates by approximately two orders of magnitude for the entire range of m_χ shown.[9] On the other hand, for $m_\phi = 125$ GeV the two contributions are roughly comparable, with $\mathcal{O}_1$ dominating at small WIMP masses and $\mathcal{O}_{10}$ dominating for larger masses, in particular if the upper cut on the recoil energy is relaxed to 240 keV.

The main qualitative difference to the results of Model I is that c_1 remains nonzero over the entire range of DM mass shown. The terms from the scalar–scalar current $\propto \lambda_1^2 + \lambda_2^2$ in eq.(2.16), third line, contribute with opposite sign from those $\propto \lambda_1^2$ (in the fifth line), which have the same sign as the contribution from the triangle diagram (the last term).

[9] Due to the lower cut on the recoil energy and the upper limit on the WIMP velocity related to the galactic escape velocity, the entire scattering rate vanishes for WIMP masses below 5 GeV.

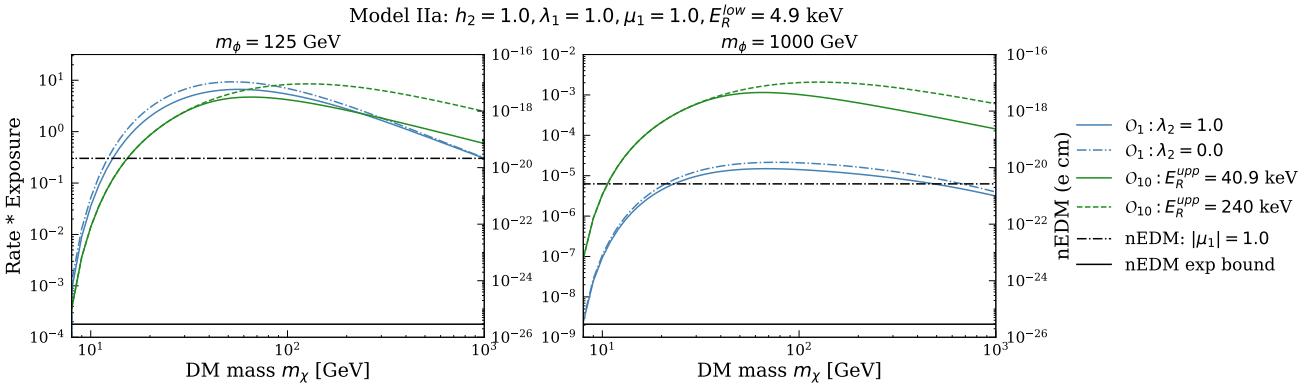


Figure 7: The left $y-$axes show the total number of scattering events in Model IIa with $\mu_1 = 1.0$ at XENON1T with a runtime exposure of 1 tonne-year as a function of the DM mass m_χ for two values of the mediator mass, $m_\phi = 125$ GeV (left frame) and $m_\phi = 1000$ GeV (right frame). The **green** and **blue** curves show the number of events due to $\mathcal{O}_{10}$ and $\mathcal{O}_1$ respectively. The **black** lines refer to the right $y-$axes; **solid black** line is the experimental upper limit on the nEDM, and the **dashed black** line is value of nEDM predicted by Model IIa.

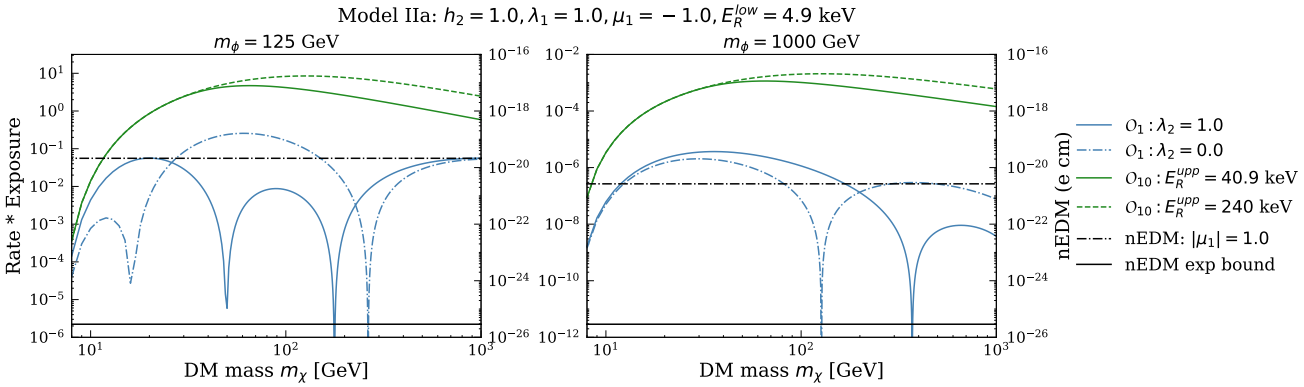


Figure 8: As in Fig.7, but for $\mu_1 = -1.0$.

This last (triangle) term dominates both for $m_\chi \ll m_\phi$ and for $m_\chi \gg m_\phi$, but is slightly smaller than the total contribution from box diagrams for $m_\chi \sim m_\phi$. As a result, when the (relative) sign between these contributions is flipped by choosing $\mu_1 = -1$, as in Fig. 8, c_1 vanishes at the two values of $s = m_\chi/m_\phi$ where the total box contribution has the same magnitude as the contribution from the triangle diagram. Since this cancellation happens at fixed values of s, it occurs at larger m_χ when m_ϕ is increased (right frame).

The upshot of this discussion is that both in Model I and in Model IIa the contribution from the non–leading NREFT operator $\mathcal{O}_{10}$ can indeed dominate the loop suppressed contribution from the leading operator $\mathcal{O}_1$, especially for large mediator masses. However, we saw at the end of the previous Chapter that in these models an nEDM is produced by two–loop diagrams involving a quark and the mediator ϕ, see Fig. 4. Its 90% C.L. experimental upper

bound is depicted by the solid black curves in Figs 5 to 8 which refer to the right $y-$axes, while the nEDM predicted by the models for the given choice of parameters is shown by the dashed black curves. We estimate the numerical value of the nEDM from the dominant contribution from the down quark cEDM $\tilde{d}_d$. Using eqs.(2.22) and (2.23), we obtain for $h_2 = \mu_1 = 1.0$:

$$|d_n| \approx \frac{2\, g_S}{(16\pi^2)^2} \left(\frac{m_\phi}{\text{GeV}}\right) \frac{5.28 \times 10^{-2}}{(m_\phi/\text{GeV})^2}\, \text{e} \cdot \text{GeV}^{-1} \approx \frac{1.58 \times 10^{-19}}{(m_\phi/\text{GeV})}\, \text{e} \cdot \text{ cm}\,, \tag{3.7}$$

where we used $g_S = 1.9$ for the strong coupling at scale of a few GeV. The loop integrals in eq.(2.22) only depend on the quark mass m_q and the mediator mass m_ϕ. On dimensional grounds, they can be written as $1/m_\phi^2$ times a function of the dimensionless ratio m_q/m_ϕ. For the down quark we can safely set $m_q \to 0$ for the values of m_ϕ we consider; the loop integral then evaluates to $5.28\times10^{-2}/m_\phi^2$. Eq.(3.7) therefore predicts $|d_n| \sim \mathcal{O}(10^{-21})\ e\cdot$ cm for $m_\phi = 125$ GeV. Increasing m_ϕ to 1 TeV reduces the nEDM value by a factor of 8; recall that we parameterize the trilinear scalar coupling as $\mu_1 m_\phi$, so for fixed $|\mu_1|$, $d_n \propto 1/m_\phi$. This agrees with the dashed black lines in the panels of Figs. 5 to 8.

Evidently for this set of couplings and $m_\phi = 125$ GeV, the predicted nEDM exceeds the upper bound by about 5 orders of magnitude; increasing m_ϕ to 1 TeV still leads to a discrepancy by about 4 orders of magnitude. Clearly these sets of parameters, which led to in principle observable effects from $\mathcal{O}_{10}$, are not realistic.

We saw in eq.(2.22) that the predicted nEDM scales like $h_2^3\,\mu_1$. In order to suppress the produced nEDM below the upper limit, one thus has to reduce this product of couplings by $\mathcal{O}(10^{-4} - 10^{-5})$. This could be achieved by reducing the Yukawa coupling h_2 by a factor of at least 30 while keeping $|\mu_1| = 1.0$ the same. However, this would reduce the contribution from $\mathcal{O}_{10}$ to the scattering rate by a factor of 10^3; the contribution from $\mathcal{O}_1$ would even be reduced by a factor of 10^6. The resulting cross sections lie below their corresponding neutrino floor(s); this part of parameter space of Models I and IIa is thus beyond the sensitivity of direct search experiments.

Alternatively one can reduce μ_1 by a factor $\geq 10^4$. This effectively removes one of the triangle diagrams contributing to $\mathcal{O}_1$, but the contribution from $\mathcal{O}_{10}$ as well as the box and – for Model I – the second triangle diagram for $\mathcal{O}_1$ remain unchanged. The contribution from $\mathcal{O}_{10}$ would then still dominate over that $\mathcal{O}_1$, except for small m_S as discussed above. However, setting $\mu_1 = 0$ does not increase the symmetry of the theory. Therefore setting μ_1 to zero only suppresses it at the lowest order in perturbation theory. Radiative corrections will in general induce a non–zero value of this cubic self–coupling.

As shown in Fig. 9, this happens at one–loop level in both Model I and Model IIa, via triangle diagrams with the DM particle running in the loop. Crucially, in both cases the DM–mediator coupling used here is the same coupling that appears in c_{10}, i.e. one cannot "switch off" these triangle diagrams without simultaneously setting the contribution from $\mathcal{O}_{10}$ to zero. In case of Model IIa, the diagram shown in Fig. 9 is (logarithmically) divergent, i.e. setting $\mu_1 = 0$ leads, strictly speaking, to a non–renormalizable theory. The triangle diagram shown for Model I is convergent. If the quartic coupling g_2 is nonzero, there is also

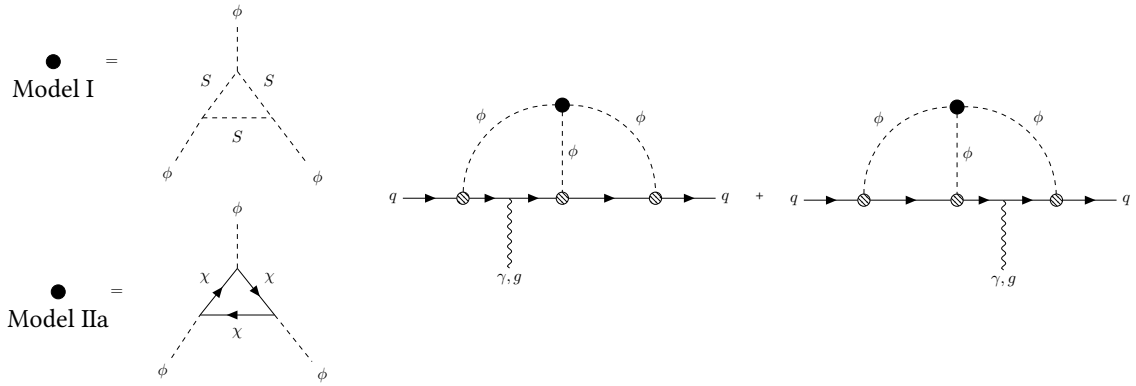


Figure 9: One–loop radiative corrections to the three–point function of the mediator $\langle\phi^3\rangle$ given by triangle diagrams in Model I and Model IIa; such diagrams induce a non–vanishing effective μ_1 even if this trilinear coupling is initially set to zero. The quark EDM and cEDMs are then induced at the three–loop level.

a divergent "bubble" diagram involving one g_2 vertex and one g_1 vertex, but g_2 does not contribute to c_{10} at tree–level, so in the following discussion we will set $g_2 = 0$.
Inserting these one–loop triangle diagrams into the upper vertex of the EDM diagrams results in three–loop diagrams. Rather than performing the challenging full three–loop calculation, we use our earlier two–loop result for d_q and insert a lower bound on μ_1 due to the triangle diagrams:

$$\text{Model I:} \quad \mu_1 \gtrsim \frac{g_1^3}{16\,\pi^2}\,\frac{m_S}{2m_\phi}\,, \tag{3.8}$$

$$\text{Model IIa:} \quad \mu_1 \gtrsim \frac{\lambda_1^3}{16\,\pi^2}\,\frac{2m_\chi}{m_\phi}\,. \tag{3.9}$$

The relative factor of 4 accounts for the four degrees of freedom running in the loop in Model IIa; computationally it results from the Dirac trace appearing in the evaluation of the fermionic triangle diagram. In the next step, we convert this lower bound on μ_1 into a lower bound on the quark EDM (and cEDM) using eq.(2.22):

$$\text{Model I:} \quad d_q \gtrsim \frac{e\,Q_q}{(16\pi^2)^3}\,(h_2\,g_1)^3\,m_S\,\lim_{q^2\to 0}\Big([\mathcal{X}]+[\mathcal{Y}]\Big)\,; \tag{3.10}$$

$$\text{Model IIa:} \quad d_q \gtrsim 4\,\frac{e\,Q_q}{(16\pi^2)^3}\,(h_2\,\lambda_1)^3\,m_\chi\,\lim_{q^2\to 0}\Big([\mathcal{X}]+[\mathcal{Y}]\Big)\,. \tag{3.11}$$

This leads directly to a lower bound on the nEDM, which can be translated into upper bounds on $\left|h_2^3\,g_1^3\right|$ and $\left|h_2^3\,\lambda_1^3\right|$ in Model I and Model IIa, respectively, by requiring our theoretical lower bound on d_n not exceed the stringent experimental upper bound.
Note that the products of couplings which are bounded by the nEDM also appear in the coefficient c_{10}, see eqs.(2.9) and (2.17), although with different powers: the bound on the nEDM scales cubically with the product of couplings whereas c_{10} scales linearly, i.e. the contribution from $\mathcal{O}_{10}$ to the DM scattering rate scales quadratically. The same is true for

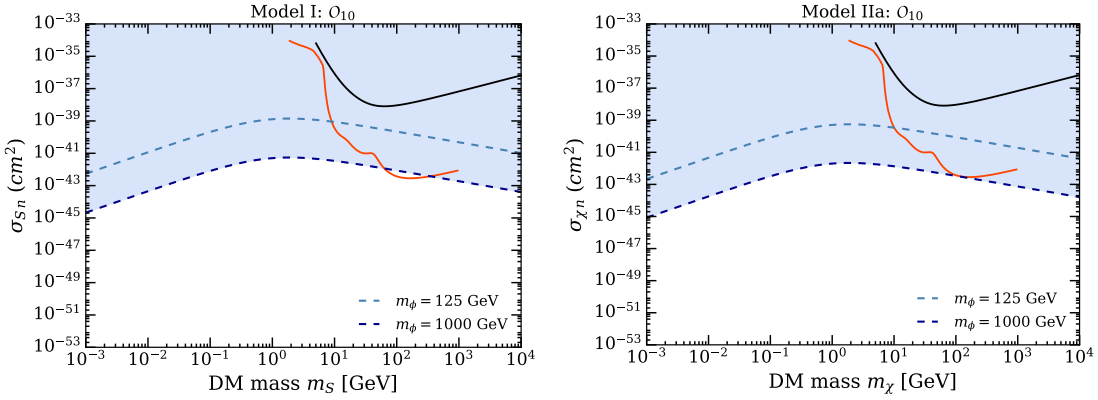


Figure 10: The **black** curves show the XENON1T upper bound on the DM–nucleon cross section from $\mathcal{O}_{10}$ in Model I (left frame) and Model IIa (right frame) as a function of DM mass. The indirect constraint from the nEDM is shown by the **dashed light blue** curve for $m_\phi = 125$ GeV and the **dashed dark blue** curve for $m_\phi = 1000$ GeV; the shaded region is thus excluded by the nEDM bound for $m_\phi = 1000$ GeV. The **red** curve denotes the neutrino floor for $\mathcal{O}_{10}$ and has been taken from Ref. [68].

the "DM–nucleon scattering cross section" due to $\mathcal{O}_{10}$, defined by:

$$\sigma_{\mathcal{O}_{10}}|^{\rm I} = \frac{3\mu_{S\,N}^2}{\pi}\left(c_{10}^N|^{\rm I}\right)^2 \quad \text{and} \quad \sigma_{\mathcal{O}_{10}}|^{\rm IIa} = \frac{3\mu_{\chi\,N}^2}{\pi}\left(c_{10}^N|^{\rm IIa}\right)^2 \ . \tag{3.12}$$

The upper bounds on $|h_2\, g_1|$ and $|h_2\, \lambda_1|$ from the nEDM will thus lead to upper bounds on this cross section.

This is illustrated in Figs. 10, which show excluded regions in the plane spanned by $\sigma_{\mathcal{O}_{10}}$ and the DM mass, for Model I (left) and Model IIa (right). The dashed blue curves depict the nEDM–derived 90% c.l. upper limit on the cross section for our two standard choices of the mediator mass m_ϕ. This bound becomes weakest at DM mass $m_{\rm DM} \simeq 1$ GeV, i.e. close to the nucleon mass; here $m_{\rm DM} = m_S$ (m_χ) for Model I (Model IIa). For fixed couplings our estimated d_n of eq.(3.10) increases linearly with $m_{\rm DM}$, while $\sigma_{\mathcal{O}_{10}} \propto m_{\rm DM}^2$ for $m_{\rm DM} \ll 1$ GeV but becomes independent of it for large DM mass. As a result, taking into account the different powers of couplings involved in the two quantities, the d_n-derived upper bound on $\sigma_{\mathcal{O}_{10}}$ grows like $m_{\rm DM}^{4/3}$ for $m_{\rm DM} \ll 1$ GeV, but declines like $m_{\rm DM}^{-2/3}$ for $m_{\rm DM} \gg 1$ GeV. Moreover, $\sigma_{\mathcal{O}_{10}} \propto m_\phi^{-4}$ while our estimated bound on $d_n \propto m_\phi^{-2}$; the d_n-derived upper bound on $\sigma_{\mathcal{O}_{10}}$ therefore scales $\propto m_\phi^{-8/3}$ for all DM masses, i.e. the bound becomes more stringent for larger mediator mass. Finally, the bound is stronger in Model IIa by a factor of $4^{2/3} \simeq 2.5$ due to the relative factor of 4 between the radiatively generated value of μ_1 given in eqs.(3.8).

The red curves in Fig. 10 show the irreducible background level from coherent neutrino–nucleus scattering ("neutrino floor") as estimated in Ref. [68]. We see that the indirect constraint is around five orders of magnitude below the present XENON1T sensitivity, and for most DM masses well below the neutrino floor for $m_\phi = 1000$ GeV. For this value of the mediator mass the current constraints on d_n therefore imply that the interactions due

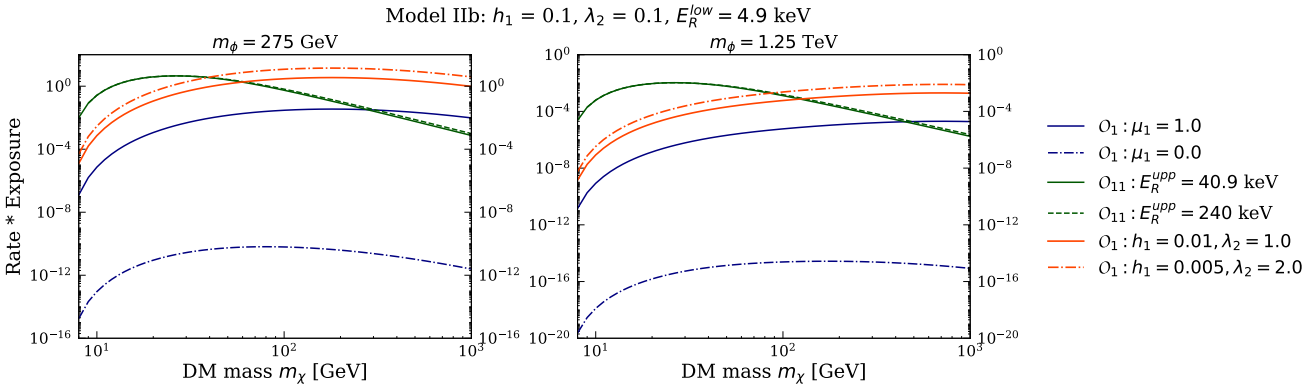


Figure 11: Total number of scattering events in Model IIb at XENON1T with a runtime exposure of 1 tonne–year as a function of the DM mass m_χ for $h_1 = \lambda_2 = 0.1$, with $m_\phi = 275$ GeV (left) and $m_\phi = 1.25$ TeV (right). The **green** curves show the number of events due to $\mathcal{O}_{11}$ and the **blue** curves show the number of events due to $\mathcal{O}_1$. The **red** curves also show contributions from $\mathcal{O}_1$ with the same product $h_1\lambda_2$, and hence the same contribution from $\mathcal{O}_{11}$, but larger λ_2.

to $\mathcal{O}_{10}$ are essentially unobservable; recall that for this large mediator mass, $\mathcal{O}_1$ contributes even less to the scattering rate.

For $m_\phi = 125$ GeV, the nEDM constraint still lies well below the current sensitivity. On the other hand, for $m_{\rm DM} \gtrsim 7$ GeV it is up to two orders of magnitude above the neutrino floor. Our analysis can therefore not completely exclude the possibility that future Xenon experiments might become sensitive to contributions from $\mathcal{O}_{10}$, if the upper bound on d_n remains unchanged. It should be noted, however, that saturating the bound on d_n requires relatively large couplings. Setting $g_1 = \lambda_1 = 1$ and $m_\phi = 125$ GeV, the estimate (3.10) is saturated for Yukawa coupling $h_2 \simeq 0.13[m_\phi/(\kappa m_{\rm DM})]^{1/3}$, where $\kappa = 1/2$ (2) in Model I (Model IIa); reducing the size of the DM–mediator coupling would require even larger h_2. Given our assumption of flavor–universal Yukawa couplings, experiments at LEP, the Tevatron and the LHC should be able to set quite stringent bounds on h_2 for $m_\phi \lesssim 125$ GeV.

We finally discuss numerical results for Model IIb. We recall that this model had been constructed to generate the operator $\mathcal{O}_{11}$ at tree–level, which is independent of the spin of the target nucleus. This required a scalar Yukawa coupling of the mediator to quarks, and a pseudoscalar coupling to the DM fermion. While this assignment quite manifestly again violates CP, it does not generate new contributions to the neutron EDM. However, at one–loop contributions to the leading spin–independent operator $\mathcal{O}_1$ are generated also in this model.

In Figs. 11 and 12, we display the number of scattering events due to $\mathcal{O}_1$ and $\mathcal{O}_{11}$ for 1.0 tonne–year exposure of the XENON1T experiment as a function of the DM mass m_χ. In each figure the mediator mass in the left frame is chosen such that the contribution from $\mathcal{O}_{11}$ saturates the XENON1T constraint for $m_\chi \simeq 30$ GeV, while in the right frame this contribution is barely above the neutrino floor for 10 GeV $\lesssim m_\chi \lesssim 50$ GeV (and slightly below it for larger m_χ). We again assume flavor–universal Yukawa couplings $h_1^q \equiv h_1$. The

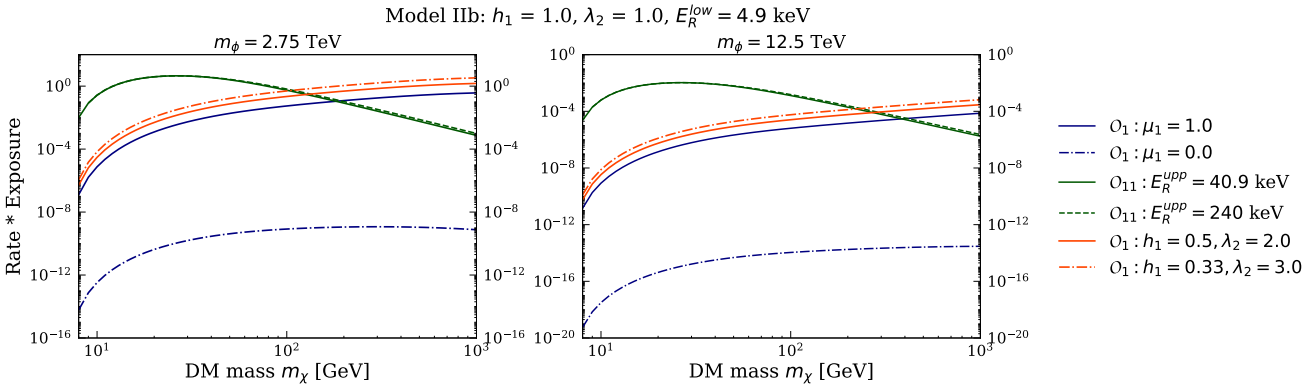


Figure 12: As in Fig. 11, but with $\lambda_2 = h_1 = 1$ and correspondingly larger mediator masses: $m_\phi = 2.75$ TeV (left) and $m_\phi = 12.5$ TeV (right).

event rates due to $\mathcal{O}_{11}$ are shown by the green lines, while the blue curves show the event rate due to $\mathcal{O}_1$ for $h_1 = \lambda_2 = 0.1$, with ($\mu_1 = 1$, solid) or without ($\mu_1 = 0$, dashed) the triangle diagram. The red curves also show contribution due to $\mathcal{O}_1$ with $\mu_1 = 1$, but for $\lambda_2 > h_1$, keeping the product $\lambda_2 \cdot h_1$, and hence the contribution from $\mathcal{O}_{11}$, constant.
Evidently the loop–induced contribution from $\mathcal{O}_1$ can only be competitive if the triangle diagram is not suppressed. This is in accord with our discussion of eq.(2.20), which showed that this contribution is expected to dominate if μ_1 is sizable. We reiterate that there are logarithmically divergent one–loop contributions of order $h_1^3 m_q/(16\pi^2 m_\phi)$ to this coupling from quark triangle diagrams, hence there is no reason to assume that μ_1 is very small.
We also see that in all cases the tree–level contribution from $\mathcal{O}_{11}$ drops quickly for $m_\chi \gtrsim$ 50 GeV. This is partly due to the reduced flux of DM particles, which scales $\propto 1/m_\chi$, but mostly because $c_{11} \propto 1/m_\chi$, as shown in eq.(2.21). In contrast, the loop–induced contribution from $\mathcal{O}_1$ in many cases keeps increasing with increasing m_χ over the entire range shown. This is because for $m_\chi \ll m_\phi$, the triangle loop function satisfies $sS_1(s) \simeq m_\chi/(2m_\phi^3)$, hence this contribution to the scattering cross section scales $\propto m_\chi^2$ for $m_\chi \ll m_\phi$. As a result, for $\mu_1 = 1$ we always find that the contribution from $\mathcal{O}_{11}$ dominates for small m_χ, while that from $\mathcal{O}_1$ is dominant at large m_χ; the DM mass where the two contributions are equal depends on the values of the other parameters. The behavior of the loop function S_1 also explains why the contribution due to $\mathcal{O}_1$ drops faster with increasing m_ϕ than that due to $\mathcal{O}_{11}$, as long as $m_\phi > m_\chi$.
We note that the number of events from $\mathcal{O}_{11}$ remains essentially unchanged when the maximal recoil energy is increased from 40.9 keV to 240 keV. As in case of $\mathcal{O}_{10}$, the scattering rate due to $\mathcal{O}_{11}$ is peaked at non–zero recoil energies, due to the $\vec{q}$ factor in the definition of this operator. However, unlike $\mathcal{O}_{10}$, $\mathcal{O}_{11}$ leads to coherent scattering on the entire nucleus. The rate is thus proportional to the square of the spin–independent elastic form factor of Xenon, which is quite soft. It is this coherent enhancement which leads to a much larger scattering rate from $\mathcal{O}_{11}$ than from $\mathcal{O}_{10}$, for similar Wilson coefficients. However, the product of recoil energy E_R (which is $\propto \vec{q}^2$) and squared form factor already peaks at $E_R \simeq 20$ keV; the convolution with the DM velocity distribution further suppresses

the rate at large E_R.
Semi–quantitatively, the ratio of the two contributions can be estimated as follows:

$$\left.\frac{N_{\mathcal{O}_1}}{N_{\mathcal{O}_{11}}}\right|_{\rm IIb} = \left.\frac{R_{\mathcal{O}_1}}{R_{\mathcal{O}_{11}}}\right|_{\rm IIb} \sim \frac{(c_1^N|^{\rm IIb})^2}{(c_{11}^N|^{\rm IIb})^2}\,\frac{m_N^2}{\vec{q}^{\,2}}\,. \tag{3.13}$$

The nuclear response is the same for both contributions and therefore does not appear in the ratio. The factor $(q^2/m_N^2)^{-1}$ is due to the momentum transfer dependence of $\mathcal{O}_{11}$. We only retain the leading triangle contribution to $c_1^N|^{\rm IIb}$, see eq.(2.20),

$$c_1^N|^{\rm IIb} \approx -\frac{\lambda_2^2}{16\,\pi^2}\,\mu_1\,m_N\,\tilde{f}^N\,\frac{m_\chi}{m_\phi}\,S_1\,. \tag{3.14}$$

Using $c_{11}^N|^{\rm IIb}$ from eq.(2.21), the ratio of events is thus given by

$$\left.\frac{N_{\mathcal{O}_1}}{N_{\mathcal{O}_{11}}}\right|_{\rm IIb} \approx \left(\frac{\lambda_2\,\mu_1}{16\,\pi^2}\frac{m_\chi^2 m_\phi}{m_N}S_1(s)\right)^2\,\frac{m_N}{2AE_R}\,. \tag{3.15}$$

For example, for $\lambda_2 = \mu_1 = 1.0$, $m_\phi = 2.75$ TeV and $m_\chi = 200$ GeV, and taking $E_R = 20$ keV as typical recoil energy in order to account for the soft form factor for coherent scattering, gives 0.45 for the ratio of event numbers, in rough agreement with the results shown in the left frame of Fig. 12. For the same couplings but increasing m_ϕ to 12.5 TeV, eq.(3.15) predicts equal event rates for $m_\chi = 520$ GeV, quite close to the intersection point between the green and blue lines in the right frame of Fig. 12. Moreover, eq.(3.15) also explains why increasing λ_2 while keeping $\lambda_2 \cdot h_1$ fixed (red curves) increases the contribution from $\mathcal{O}_1$. In fact, for $m_\phi \lesssim 1$ TeV reducing h_1 should help to avoid possible constraints on the model from searches at the LHC.

4 Summary and Conclusions

In this paper we explored the detection prospects involving $P-$ and $T-$odd operators arising in the NREFT formalism of WIMP–nucleon scattering. These operators appear at next–to–leading order in an expansion in WIMP velocity v and momentum transfer $\vec{q}$. Since these quantities accompanying the $P-$ and $T-$odd operators are quite small, these operators are expected to be insignificant relative to the leading order operators *if* the corresponding Wilson coefficients are of similar magnitude. Hence the additional operators can make significant contributions only when the coefficient of the leading spin–independent operator $\mathcal{O}_1$ is strongly suppressed or vanishes entirely. This typically requires ad hoc choices of couplings in a relativistic QFT, i.e. one can generally not find a symmetry that suppresses the contribution from $\mathcal{O}_1$ without also suppressing the Wilson coefficients of the additional operators.
Crucially, the $P-$ and $T-$odd NREFT operators can only occur in the low energy limit of a QFT that violates the CP symmetry. This can lead to stringent constraints on the theory, in particular from electric dipole moments.
We addressed these concerns in the framework of three simplified models with uncharged $t-$channel mediators, taken from ref. [47]. These models extend the SM by a real scalar

mediator particle ϕ which does not carry electric or color charge, and a DM particle which has spin 0 (Model I) or spin 1/2 (Model IIa and IIb). These models can generate flavor changing neutral currents already at tree–level unless the Yukawa couplings of the mediator are diagonal in the quark mass basis. This can easily be ensured if these new couplings are flavor–universal, which we therefore assumed in our numerical examples.

The couplings in these models are chosen such that at the lowest order in perturbation theory, only the $P-$ and $T-$odd operators $\mathcal{O}_{10}$ and $\mathcal{O}_{11}$ arise in the non–relativistic limit. In particular, ϕ must not have scalar couplings to both quarks and the DM particle. In Models I and IIa, the quark couplings are pseudoscalar while the DM couplings are scalar, while in Model IIb the quark couplings are scalar but the DM couplings are pseudoscalar. Note that there is no symmetry that forbids scalar quark couplings in Models I and IIa, or scalar DM couplings in Model IIb. It is therefore not surprising that at the next order in perturbation theory, one–loop box and triangle diagrams do induce the canonical SI interactions described by the operator $\mathcal{O}_1$ in these models. We compared the tree–level interactions giving rise to $\mathcal{O}_{10}$ or $\mathcal{O}_{11}$ with the one–loop suppressed interactions yielding $\mathcal{O}_1$. To that end we computed the total number of events due to the two types of interactions for a Xenon target.

In the case of Model I and Model IIa and assuming large couplings in order to generate detectable event rates, we found that the contributions from $\mathcal{O}_{10}$ can be roughly comparable to those from $\mathcal{O}_1$ for mediator mass $m_\phi = 125$ GeV, but for heavier mediator the contributions from $\mathcal{O}_{10}$ clearly dominates. It thus appears as if $\mathcal{O}_{10}$ could indeed be the most relevant NREFT operator in these models.

However, we pointed out that the quark–mediator interactions in both models produce two–loop contributions to d_n, the electric dipole moment of the neutron (nEDM). These contributions scale linearly with the cubic self–interaction of the mediator μ_1. For parameter choices that lead to detectable event rates from $\mathcal{O}_{10}$ and μ_1 of order unity (in units of m_ϕ), the predicted d_n is several orders of magnitude larger than the upper limit reported by experiments. Even if we set $\mu_1 = 0$, non–vanishing trilinear self–interactions are generated at one–loop level by couplings that also appear in the Wilson coefficient of $\mathcal{O}_{10}$. Estimating a lower bound on μ_1 from these loop diagrams, we find that the resulting upper bound on the WIMP–nucleus scattering rate is still well below current sensitivity, and often even below the irreducible background ("neutrino floor"), especially for large m_ϕ where $\mathcal{O}_{10}$ potentially dominates the scattering rate. As in case of models with charged, $s-$channel mediator, where d_n is generated already at one–loop level [50], it is thus essential to consider the d_n constraint when considering prospects for detecting WIMP–nucleus scattering due to $\mathcal{O}_{10}$.

In Model IIb, which generates the spin–independent NREFT operator $\mathcal{O}_{11}$, we again found that the loop–induced contributions from $\mathcal{O}_1$ can be larger than the tree–level contribution due to $\mathcal{O}_{11}$, *if* the trilinear scalar coupling μ_1 is not suppressed. We emphasized that in this model quark loops generate a logarithmically divergent one–loop contribution to μ_1. For $\mu_1 = 1$, $\mathcal{O}_1$ typically dominates for large DM masses, while $\mathcal{O}_{11}$ is dominant for small masses, the cross–over point depending on the values of the other parameters. In this model no new contributions to d_n are generated, and therefore this operator is not subject to the stringent constraints of the neutron EDM, even though it is also $P-$ and $T-$odd.

At least at the level of a simplified model one can therefore engineer a scenario where the non–leading operator $\mathcal{O}_{11}$ dominates the WIMP–nucleon scattering rate.

However, it is by no means clear whether this remains true in the framework of UV–complete theories. Recall that we only require our Lagrangians to be invariant under $SU(3)_C \times U(1)_{\rm em}$; we did not enforce invariance under the electroweak gauge symmetry. In particular, the $\phi\bar{q}q$ couplings are not $SU(2) \times U(1)_Y$ invariant if ϕ is a singlet. The simplest choice would be to identify ϕ with the Higgs boson of the SM, which resides in a doublet of $SU(2)$. However, the $\phi\bar{q}q$ couplings are then known to be very small. Moreover, the WIMP can then also not be a gauge singlet, and would thus have additional (gauge) interactions leading to additional constraints as well as new contributions to WIMP–nucleon scattering. Alternatively one can couple a singlet WIMP to a singlet scalar which mixes with the SM Higgs boson; however, in such a scenario the WIMP–nucleon scattering rate would be suppressed even further by the mediator–Higgs mixing angle, which has to be rather small in order not to distort the properties of the physical 125 GeV particle too much.

We remind the reader that we assumed universal flavor–diagonal couplings of the mediator to the quarks. This simplifies the model building, since these couplings are then flavor–diagonal in any basis. On the other hand, it might appear more natural to assume that the new couplings increase with increasing quark mass, just as the Yukawa couplings of the SM do. In this case tree–level FCNC are avoided if the matrices of new Yukawa couplings commute with the Yukawa coupling matrices of the SM; from the model building point of view it is not clear why this should be the case. Moreover, keeping the couplings to the top quark $\lesssim 1$ would then require very small couplings to the first generation. In such a scenario the contributions from heavy quarks might well dominate the DM scattering cross section, as well as – in Models I and IIa – the electric dipole moment of the neutron. In order to yield detectable event rates, the couplings to third generation quarks would have to be larger, and/or the mediator lighter, than in our numerical examples. It seems rather unlikely to us that this would lead to qualitatively different conclusions in Models I and IIa with unsuppressed trilinear self–coupling of the mediator, given the very large discrepancy between the predicted nEDM and its experimental upper bound that we found in our numerical examples; however, we have not performed an explicit computation to check this.[10]

In summary, the results presented in this paper as well as ref.[50] strongly indicate that the current experimental upper bound on the electric dipole moment of the neutron excludes the possibility that the operator $\mathcal{O}_{10}$ can make contributions to WIMP–nucleus scattering to which current or near–future experiments are sensitive. While no such strong statement can be made for the operator $\mathcal{O}_{11}$ generated by the exchange of a neutral mediator in the $t-$channel, it is currently unclear whether such a model can be constructed that respects the full gauge symmetry of the SM and leads to detectable WIMP–nucleon scattering being dominated by $\mathcal{O}_{11}$.

[10] The cEDMs of heavy quarks generate light quark cEDMs, as well as an electron EDM, at the three–loop level [69]. Moreover, the contribution of the gluonic Weinberg operator, which will generically be generated at three–loop level in Models I and IIa, would probably also have to be taken into account [70].

Acknowledgments

We acknowledge the use of TikZ–Feynman [71] for creating Feynman diagrams and Package–X [72] for verifying parts of our loop computations. We thank Claude Duhr for discussions about the two loop diagram. RM was partially supported by the Bonn Cologne Graduate School of Physics and Astronomy.

A 1-loop calculations

In this appendix we provide details of the calculations of the one–loop box and triangle diagrams appearing in Model I, IIa and IIb.

A.1 Model I Matrix Element

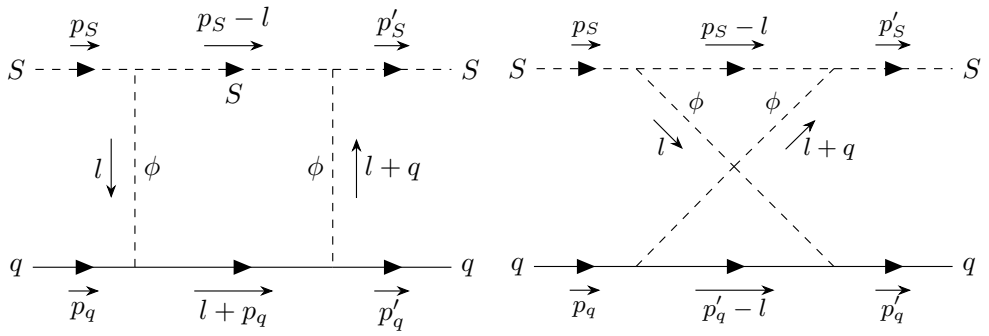


Figure 13: One–loop box and crossed box diagrams contributing to $\mathcal{O}_1$ in Model I.

The box diagram shown in the left Fig. 13 gives the following contribution to the matrix element for DM–quark scattering:

$$i\mathcal{M}_1^{\mathrm{I}} = (-ig_1 m_S)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{(p_S - l)^2 - m_S^2} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p_q + l)^2 - m_q^2} \times \bar{u}(p'_q)\, h_2^q \gamma^5 \,(\not{p}_q + \not{l} + m_q)\, h_2^q \gamma^5\, u(p_q)\,; \tag{A.1}$$

the crossed box diagram shown in the right Fig. 13 contributes:

$$i\mathcal{M}_2^{\mathrm{I}} = (-ig_1 m_S)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{(p_S - l)^2 - m_S^2} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p'_q - l)^2 - m_q^2} \times \bar{u}(p'_q) h_2^q \gamma^5 \,(\not{p}'_q - \not{l} + m_q) h_2^q \gamma^5\, u(p_q)\,. \tag{A.2}$$

After simplifying the numerator by commuting the two γ^5 matrices and using the Dirac equation, we obtain for the two diagrams:

$$i\mathcal{M}_1^{\mathrm{I}} = g_1^2\, (h_2^q)^2\, m_S^2\, \mathrm{M}_{1,\mu}^{\mathrm{I}}\, \bar{u}(p'_q)\gamma^\mu u(p_q) \quad \text{and} \tag{A.3}$$

$$i\mathcal{M}_2^{\mathrm{I}} = -\, g_1^2\, (h_2^q)^2\, m_S^2\, \mathrm{M}_{2,\mu}^{\mathrm{I}}\, \bar{u}(p'_q)\gamma^\mu u(p_q)\,. \tag{A.4}$$

Here $\mathrm{M}^{\mathrm{I}}_{1,\mu}$ and $\mathrm{M}^{\mathrm{I}}_{2,\mu}$ are loop integrals:

$$\mathrm{M}^{\mathrm{I}}_{1,\mu} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu}{[(p_S - l)^2 - m_S^2]\,[l^2 - m_\phi^2]\,[(l+q)^2 - m_\phi^2]\,[(p_q + l)^2 - m_q^2]}\,, \tag{A.5}$$

$$\mathrm{M}^{\mathrm{I}}_{2,\mu} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu}{[(p_S - l)^2 - m_S^2]\,[l^2 - m_\phi^2]\,[(l+q)^2 - m_\phi^2]\,[(p'_q - l)^2 - m_q^2]}\,. \tag{A.6}$$

After Feynman parametrization, these loop integrals can be expressed in the vanishing momentum transfer limit ($q \to 0$, i.e. $p_S \to p'_S$ and $p_q \to p'_q$) in terms of loop functions $M_i(r = m_q/m_\phi, s = m_S/m_\phi)$, $i = 1, \ldots, 4$, which are given in Appendix A.1.1. The contributions from the box diagram and crossed box diagrams can then finally be written as:

$$\mathcal{M}^{\mathrm{I}}_1 = \frac{(h_2^q)^2\, g_1^2\, m_S^2}{16\pi^2} \left[\left(\frac{p_{S,\mu} + p'_{S,\mu}}{2} \right) M_1\, \bar{u}(p'_q)\gamma^\mu u(p_q) + m_q\, M_2\, \bar{u}(p'_q)\, u(p_q) \right]\,, \tag{A.7}$$

$$\mathcal{M}^{\mathrm{I}}_2 = -\frac{(h_2^q)^2\, g_1^2\, m_S^2}{16\pi^2} \left[\left(\frac{p_{S,\mu} + p'_{S,\mu}}{2} \right) M_3\, \bar{u}(p'_q)\gamma^\mu u(p_q) - m_q\, M_4\, \bar{u}(p'_q)\, u(p_q) \right]\,. \tag{A.8}$$

Even though $p_S = p'_S$ for $q = 0$ we've written these contributions in terms of the symmetric sum $p_S + p'_S$, which facilitates matching onto the effective Lagrangian of eq.(2.5) after integrating out the mediator.

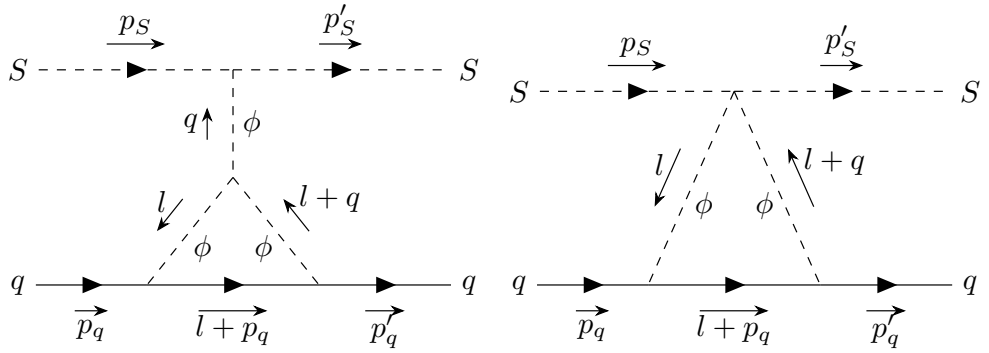


Figure 14: One–loop triangle diagrams contributing to $\mathcal{O}_1$ in Model I.

Fig. 14 shows contributions to WIMP–quark scattering in Model I from triangle diagrams. The matrix element for the left diagram reads:

$$i\mathcal{M}^{\mathrm{I}}_{\triangle_1} = (-im_\phi\,\mu_1)\,(-ig_1 m_S)\frac{i}{q^2 - m_\phi^2} \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p_q + l)^2 - m_q^2} \times \bar{u}(p'_q)\, h_2^q \gamma^5\, (\not{p}_q + \not{l} + m_q)\, h_2^q \gamma^5\, u(p_q)\,; \tag{A.9}$$

the right triangle diagram contributes:

$$i\mathcal{M}^{\mathrm{I}}_{\triangle_2} = \left(\frac{-i\, g_2}{2} \right) \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p_q + l)^2 - m_q^2} \times \bar{u}(p'_q)\, h_2^q \gamma^5\, (\not{p}_q + \not{l} + m_q)\, h_2^q \gamma^5\, u(p_q)\,. \tag{A.10}$$

Again commuting the two γ^5 matrices and using the Dirac equation, we obtain for these two diagrams:

$$i\mathcal{M}^{\mathrm{I}}_{\triangle_1} = g_1\mu_1(h_2^q)^2\,\frac{m_S\,m_\phi}{q^2-m_\phi^2}\,\mathrm{M}^{\mathrm{I}}_{3,\mu}\,\bar{u}(p_q')\gamma^\mu u(p_q) \quad \text{and} \tag{A.11}$$

$$i\mathcal{M}^{\mathrm{I}}_{\triangle_2} = \frac{g_2}{2}(h_2^q)^2\,\mathrm{M}^{\mathrm{I}}_{3,\mu}\,\bar{u}(p_q')\gamma^\mu u(p_q)\;. \tag{A.12}$$

The loop integral $\mathrm{M}^{\mathrm{I}}_{3,\mu}$ is:

$$\mathrm{M}^{\mathrm{I}}_{3,\mu} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4}\,\frac{l_\mu}{[l^2-m_\phi^2]\,[(l+q)^2-m_\phi^2]\,[(p_q+l)^2-m_q^2]}\;. \tag{A.13}$$

In the limit $q \to 0$ it can be expressed in terms of the loop function L_1, whose analytic expression is given in Appendix A.1.1. The contribution of the two triangle diagrams to the matrix element can then finally be written as

$$\mathcal{M}^{\mathrm{I}}_{\triangle_1} = \frac{g_1\mu_1(h_2^q)^2}{16\,\pi^2}\,\frac{m_q m_S}{m_\phi}\,L_1(m_q^2,m_\phi^2)\;\;\bar{u}(p_q')\,u(p_q)\;, \tag{A.14}$$

$$\mathcal{M}^{\mathrm{I}}_{\triangle_2} = -\frac{g_2(h_2^q)^2}{16\,\pi^2}\,\frac{m_q}{2}\,L_1(m_q^2,m_\phi^2)\;\;\bar{u}(p_q')\,u(p_q)\;. \tag{A.15}$$

A.1.1 Model I Loop Functions

We first define the function $\mathrm{L}(x)$ of the real variable x as:

$$\mathrm{L}(x) = \begin{cases} \sqrt{1-4x^2}\cdot\ln\left(\frac{1+\sqrt{1-4x^2}}{2|x|}\right) & \text{for}\quad |x|\le 0.5 \\ -\sqrt{4x^2-1}\cdot\arctan\left(\sqrt{4x^2-1}\right) & \text{for}\quad |x|\ge 0.5 \end{cases}\;. \tag{A.16}$$

In terms of this function, the loop functions $M_i,\ i = 1, \ldots, 4$ and L_1 can be written as:

$$M_1 = \frac{1}{3\,m_\phi^4}\left[\frac{1}{s\,(r+s)} - \frac{(1+2r^2)}{2\,r^2\,(r+s)^2}\,\mathrm{L}(r) + \frac{s-2r}{4\,r^2\,s^3}\,\ln\!\left(1/s^2\right)\right.$$
$$\left. - \frac{\ln\!\left(r^2/s^2\right)}{4\,r^2\,(r+s)^2} + \frac{2r+3s-4rs^2-6s^3+8rs^4}{2\,s^3\,(r+s)^2\,(1-4s^2)}\,\mathrm{L}(s)\right]\,; \tag{A.17}$$

$$M_2 = \frac{1}{3\,m_\phi^4}\left[-\frac{1}{r\,(r+s)} + \frac{1+2s^2}{2s^2\,(r+s)^2}\,\mathrm{L}(s) - \frac{r-2s}{4\,r^3\,s^2}\,\ln\!\left(1/s^2\right)\right.$$
$$\left. - \frac{(3r+2s)\,\ln\!\left(r^2/s^2\right)}{4\,r^3\,(r+s)^2} + \frac{2s+3r-4r^2s-6r^3+8r^4s}{2\,r^3\,(4r^2-1)\,(r+s)^2}\,\mathrm{L}(r)\right]\,; \tag{A.18}$$

$$M_3 = \frac{1}{3\,m_\phi^4}\left[\frac{1}{s\,(s-r)} - \frac{1+2r^2}{2\,r^2\,(r-s)^2}\,\mathrm{L}(r) + \frac{s+2r}{4\,r^2\,s^3}\,\ln\!\left(1/s^2\right)\right.$$
$$\left. - \frac{\ln\!\left(r^2/s^2\right)}{4\,r^2\,(r-s)^2} - \frac{2r-3s-4rs^2+6s^3+8rs^4}{2\,s^3\,(r-s)^2\,(1-4s^2)}\,\mathrm{L}(s)\right]\,; \tag{A.19}$$

$$M_4 = \frac{1}{3\,m_\phi^4}\left[-\frac{1}{r\,(r-s)} + \frac{1+2s^2}{2s^2\,(r-s)^2}\,\mathrm{L}(s) - \frac{r+2s}{4\,r^3\,s^2}\,\ln\!\left(1/s^2\right)\right.$$
$$\left. - \frac{(3r-2s)\,\ln\!\left(r^2/s^2\right)}{4\,r^3\,(r-s)^2} + \frac{2s-3r-4r^2s+6r^3+8r^4s}{2\,r^3\,(1-4r^2)\,(r-s)^2}\,\mathrm{L}(r)\right]\,; \tag{A.20}$$

$$L_1 = \frac{1}{m_\phi^2}\left[\frac{1}{r^2} + \frac{r^2-1}{2r^4}\,\ln\!\left(1/r^2\right) + \frac{1-3r^2}{r^4\,(1-4r^2)}\,\mathrm{L}(r)\right]\,. \tag{A.21}$$

We note that $M_3(r,s) = M_1(-r,s) = M_1(r,-s)$ and $M_4(r,s) = M_2(-r,s) = M_2(r,-s)$, i.e. the functions M_i remain invariant when both arguments change their sign. Of course, in our application only positive arguments are physical, since $r = m_q/m_\phi$ and $s = m_S/m_\phi$. In our examples we assume flavor–universal Yukawa couplings. In this case the contribution from the light quarks will dominate the WIMP–nucleon scattering matrix elements. We therefore also give the **massless quark limits of the loop functions**:

$$M_1, M_3 \xrightarrow{r\to 0} \frac{1}{2m_\phi^4}\frac{1}{s^2}\left[1 - \frac{1}{2s^2}\ln\left(\frac{1}{s^2}\right) + \frac{1-2s^2}{s^2\,(1-4s^2)}\mathrm{L}(s)\right]\,; \tag{A.22}$$

$$M_2, M_4 \xrightarrow{r\to 0} \frac{1}{6m_\phi^4}\frac{1}{s^2}\left[1 - \frac{1}{2s^2}\ln\left(\frac{1}{s^2}\right) + \frac{1+2s^2}{s^2}\mathrm{L}(s)\right]\,; \tag{A.23}$$

$$L_1 \xrightarrow{r\to 0} -\frac{1}{2m_\phi^2}\,. \tag{A.24}$$

A.2 Model IIa and IIb Matrix Elements

We now turn to the models with fermionic WIMP χ. We again begin with the box and crossed box diagrams shown in Fig. 15. The contribution to the matrix element for DM–quark scattering from the box diagram in Model IIa, involving the scalar DM–mediator

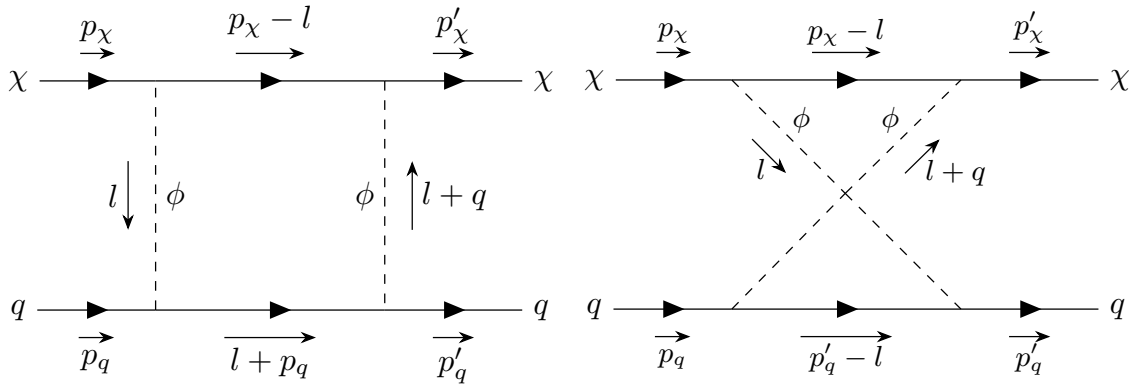


Figure 15: One–loop box diagrams contributing to $\mathcal{O}_1$ in Model IIa or Model IIb.

coupling λ_1, is

$$i\mathcal{M}_1^{\rm IIa} = (-i\lambda_1)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{(p_\chi - l)^2 - m_\chi^2} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p_q + l)^2 - m_q^2} \times \bar{u}(p'_\chi)(\not{p}_\chi - \not{l} + m_\chi) u(p_\chi) \bar{u}(p'_q) h_2^q \gamma^5 (\not{p}_q + \not{l} + m_q) h_2^q \gamma^5 u(p_q) \; , \quad \text{(A.25)}$$

the crossed box diagram contributes:

$$i\mathcal{M}_2^{\rm IIa} = (-i\lambda_1)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{i}{(p_\chi - l)^2 - m_\chi^2} \frac{i}{l^2 - m_\phi^2} \frac{i}{(l+q)^2 - m_\phi^2} \frac{i}{(p'_q - l)^2 - m_q^2} \times \bar{u}(p'_\chi)(\not{p}_\chi - \not{l} + m_\chi) u(p_\chi) \bar{u}(p'_q) h_2^q \gamma^5 (\not{p}'_q - \not{l} + m_q) h_2^q \gamma^5 u(p_q) \; . \quad \text{(A.26)}$$

After simplifying the numerator by commuting the two γ^5 matrices and using the Dirac equation for the external quarks and WIMPs, we obtain for the two diagrams:

$$i\mathcal{M}_1^{\rm IIa} = -\lambda_1^2 (h_2^q)^2 \Big(\mathrm{N}_{1,\mu\nu}^{\rm IIa} \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right] \left[\bar{u}(p'_\chi)\gamma^\nu u(p_\chi)\right] - 2 m_\chi \mathrm{N}_{2,\mu}^{\rm IIa} \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right] \left[\bar{u}(p'_\chi) u(p_\chi)\right] \Big) \; , \quad \text{(A.27)}$$

$$i\mathcal{M}_2^{\rm IIa} = \lambda_1^2 (h_2^q)^2 \Big(\mathrm{N}_{3,\mu\nu}^{\rm IIa} \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right] \left[\bar{u}(p'_\chi)\gamma^\nu u(p_\chi)\right] - 2 m_\chi \mathrm{N}_{4,\mu}^{\rm IIa} \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right] \left[\bar{u}(p'_\chi) u(p_\chi)\right] \Big) \; . \quad \text{(A.28)}$$

Here $\mathrm{N}_{1,\mu\nu}^{\rm IIa}$, $\mathrm{N}_{3,\mu\nu}^{\rm IIa}$, $\mathrm{N}_{3,\mu}^{\rm IIa}$ and $\mathrm{N}_{4,\mu}^{\rm IIa}$ are loop integrals:

$$\mathrm{N}_{1,\mu\nu}^{\rm IIa} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu \, l_\nu}{[(p_\chi - l)^2 - m_\chi^2] \, [l^2 - m_\phi^2] \, [(l+q)^2 - m_\phi^2] \, [(p_q + l)^2 - m_q^2]} \; , \quad \text{(A.29)}$$

$$\mathrm{N}_{2,\mu}^{\rm IIa} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu}{[(p_\chi - l)^2 - m_\chi^2] \, [l^2 - m_\phi^2] \, [(l+q)^2 - m_\phi^2] \, [(p_q + l)^2 - m_q^2]} \; , \quad \text{(A.30)}$$

$$\mathrm{N}_{3,\mu\nu}^{\rm IIa} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu \, l_\nu}{[(p_\chi - l)^2 - m_\chi^2] \, [l^2 - m_\phi^2] \, [(l+q)^2 - m_\phi^2] \, [(p'_q - l)^2 - m_q^2]} \; , \quad \text{(A.31)}$$

$$\mathrm{N}_{4,\mu}^{\rm IIa} = \int \frac{\mathrm{d}^4 l}{(2\pi)^4} \frac{l_\mu}{[(p_\chi - l)^2 - m_\chi^2] \, [l^2 - m_\phi^2] \, [(l+q)^2 - m_\phi^2] \, [(p'_q - l)^2 - m_q^2]} \; . \quad \text{(A.32)}$$

After Feynman parametrization and taking the limit $q \to 0$, the loop integrals in eqs.(A.29) to (A.32) can be expressed in terms of loop functions $N_i,\ i = 1, \ldots, 8$ and $P_j,\ j = 1, \ldots, 4$ whose analytic expressions can be found in Appendix A.2.1. Both sets of functions depend on $r = m_q/m_\phi$ and $s = m_\chi/m_\phi$. The contributions from the box and crossed box diagrams can then be written as:

$$\begin{aligned}\mathcal{M}_1^{\mathrm{IIa}} = -\frac{\lambda_1^2\,(h_2^q)^2}{16\,\pi^2}\Big\{ & N_1\ \left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right] \\ & + \Big(2m_\chi m_q\,(N_2 - P_2) + m_\chi^2\,(N_3 - 2P_1) + m_q^2\,N_4\Big) \\ & \times \left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,u(p_q)\right]\Big\}\,; \end{aligned} \tag{A.33}$$

$$\begin{aligned}\mathcal{M}_2^{\mathrm{IIa}} = \frac{\lambda_1^2\,(h_2^q)^2}{16\,\pi^2}\Big\{ & N_5\ \left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right] \\ & + \Big(2m_\chi m_q\,(N_6 - P_4) + m_\chi^2\,(N_7 - 2P_3) + m_q^2\,N_8\Big) \\ & \times \left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,u(p_q)\right]\Big\}\,. \end{aligned} \tag{A.34}$$

The diagrams of Fig. 15 also contribute when both DM–mediator couplings are λ_2. The contribution from the box diagram reads:

$$\begin{aligned} i\mathcal{M}_3^{\mathrm{IIa}} = \lambda_2^2 (h_2^q)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} & \frac{i}{(p_\chi - l)^2 - m_\chi^2}\,\frac{i}{l^2 - m_\phi^2}\,\frac{i}{(l+q)^2 - m_\phi^2}\,\frac{i}{(p_q + l)^2 - m_q^2} \\ & \times \bar{u}(p'_\chi)\gamma^5(\not{p}_\chi - \not{l} + m_\chi)\gamma^5 u(p_\chi)\bar{u}(p'_q)\gamma^5(\not{p}_q + \not{l} + m_q)\gamma^5 u(p_q)\,. \end{aligned} \tag{A.35}$$

The crossed box contributes:

$$\begin{aligned} i\mathcal{M}_4^{\mathrm{IIa}} = \lambda_2^2 (h_2^q)^2 \int \frac{\mathrm{d}^4 l}{(2\pi)^4} & \frac{i}{(p_\chi - l)^2 - m_\chi^2}\,\frac{i}{l^2 - m_\phi^2}\,\frac{i}{(l+q)^2 - m_\phi^2}\,\frac{i}{(p'_q - l)^2 - m_q^2} \\ & \times \bar{u}(p'_\chi)\gamma^5(\not{p}_\chi - \not{l} + m_\chi)\gamma^5 u(p_\chi)\bar{u}(p'_q)\gamma^5(\not{p}'_q - \not{l} + m_q)\gamma^5 u(p_q)\,. \end{aligned} \tag{A.36}$$

After commuting the two γ^5 matrices and using the Dirac equation, these simplify to

$$i\mathcal{M}_3^{\mathrm{IIa}} = -\lambda_2^2\,(h_2^q)^2\ \mathrm{N}_{1,\mu\nu}^{\mathrm{IIa}}\ \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right]\left[\bar{u}(p'_\chi)\gamma^\nu u(p_\chi)\right]\,; \tag{A.37}$$

$$i\mathcal{M}_4^{\mathrm{IIa}} = \lambda_2^2\,(h_2^q)^2\ \mathrm{N}_{3,\mu\nu}^{\mathrm{IIa}}\ \left[\bar{u}(p'_q)\gamma^\mu u(p_q)\right]\left[\bar{u}(p'_\chi)\gamma^\nu u(p_\chi)\right]\,, \tag{A.38}$$

where $\mathrm{N}_{1,\mu\nu}^{\mathrm{IIa}}$ and $\mathrm{N}_{3,\mu\nu}^{\mathrm{IIa}}$ are loop integrals defined in eqs.(A.29) and (A.31). In the $q \to 0$ limit we finally obtain:

$$\begin{aligned}\mathcal{M}_3^{\mathrm{IIa}} = -\frac{\lambda_2^2\,(h_2^q)^2}{16\,\pi^2}\Big\{ & N_1\ \left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right] \\ & + (2m_\chi m_q\,N_2 + m_\chi^2\,N_3 + m_q^2\,N_4)\ \left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,u(p_q)\right]\Big\}\,; \end{aligned} \tag{A.39}$$

$$\begin{aligned}\mathcal{M}_4^{\mathrm{IIa}} = \frac{\lambda_2^2\,(h_2^q)^2}{16\,\pi^2}\Big\{ & N_5\ \left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right] \\ & + (2\,m_\chi\,m_q\,N_6 + m_\chi^2\,N_7 + m_q^2\,N_8)\ \left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\ \left[\bar{u}(p'_q)\,u(p_q)\right]\Big\}\,. \end{aligned} \tag{A.40}$$

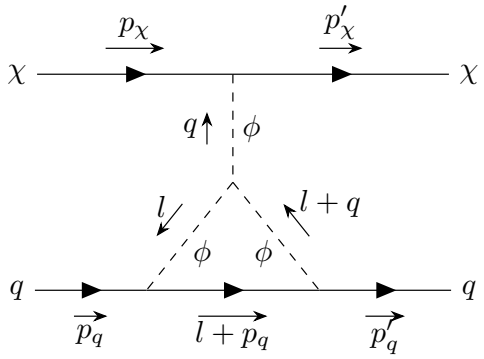


Figure 16: One–loop triangle diagram contributing to $\mathcal{O}_1$ in Model IIa involving the couplings λ_1, μ_1 and h_2^q.

The functions N_i already appeared in the contributions $\propto \lambda_1^2$; they are defined in Appendix A.2.1. This completes the contribution from the diagrams of Fig 15, since diagrams involving one scalar and one pseudoscalar coupling on the WIMP line do not contribute to $\mathcal{O}_1$.

There is only one triangle diagram contributing to $\mathcal{O}_1$ at one–loop order in Model IIa, as shown in Fig. 16; note that only the diagram involving the scalar DM–mediator coupling λ_1 contributes to $\mathcal{O}_1$. Its contribution to the DM–quark scattering matrix element is given by:

$$i\mathcal{M}_{\triangle}^{\mathrm{IIa}} = \frac{i(-i\lambda_1)(-im_\phi\mu_1)}{q^2-m_\phi^2}\int \frac{\mathrm{d}^4 l}{(2\pi)^4}\frac{i}{l^2-m_\phi^2}\frac{i}{(l+q)^2-m_\phi^2}\frac{i}{(l+q)^2-m_\phi^2} \tag{A.41}$$

$$\times\left[\bar{u}(p'_\chi)\,u(p_\chi)\right]\bar{u}(p'_q)\,h_2^q\gamma^5\left(\not{p}_q+\not{l}+m_q\right)h_2^q\gamma^5\,u(p_q)\,. \tag{A.42}$$

After simplifying the numerator as before, we get:

$$\mathcal{M}_{\triangle}^{\mathrm{IIa}} = -\frac{\lambda_1\,(h_2^q)^2\,\mu_1}{16\,\pi^2}\left(\frac{m_q}{m_\phi}\right)R_1\left[\bar{u}(p'_\chi)\,u(p_\chi)\right]\left[\bar{u}(p'_q)\,u(p_q)\right]\,. \tag{A.43}$$

Here a new loop function R_1 appears, which is also defined in Appendix A.2.1.

The calculations for Model IIb are very similar. The (crossed) box diagrams look exactly the same as in Model IIa, but now we have scalar Yukawa couplings on the quark line and pseudoscalar couplings on the WIMP line. We only list the final results. For the box and crossed box we obtain:

$$\mathcal{M}_1^{\mathrm{IIb}} = -\frac{\lambda_2^2\,(h_1^q)^2}{16\,\pi^2}\Big[N_1\,\left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\,\left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right]$$

$$+\left(2m_\chi m_q\,(N_2+P_1)+m_\chi^2\,N_3+m_q^2\,(N_4+2P_2)\right)$$

$$\times\left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\,\left[\bar{u}(p'_q)\,u(p_q)\right]\Big] \tag{A.44}$$

$$\mathcal{M}_2^{\mathrm{IIb}} = \frac{\lambda_2^2\,(h_1^q)^2}{16\,\pi^2}\Big[N_5\,\left[\bar{u}(p'_\chi)\,\gamma^\mu\,\bar{u}(p_\chi)\right]\,\left[\bar{u}(p'_q)\,\gamma_\mu\,u(p_q)\right]$$

$$+\left(2\,m_\chi\,m_q\,(N_6-P_3)+m_\chi^2\,N_7+m_q^2\,(N_8-2P_4)\right)$$

$$\times\left[\bar{u}(p'_\chi)\,\bar{u}(p_\chi)\right]\,\left[\bar{u}(p'_q)\,u(p_q)\right]\Big]\,. \tag{A.45}$$

The functions N_i and P_k are the same as in Model IIa, and are defined in Appendix A.2.1. The Model IIb triangle diagram shown in Fig. 3 yields:

$$\mathcal{M}_{\triangle}^{\rm IIb} = -\frac{h_1^q\,\lambda_2^2\,\mu_1}{16\,\pi^2}\left(\frac{m_\chi}{m_\phi}\right) S_1\left[\bar{u}(p'_\chi)\,u(p_\chi)\right]\left[\bar{u}(p'_q)\,u(p_q)\right]\ . \tag{A.46}$$

Here a new loop function S_1 appears; it is also defined in Appendix A.2.1.

A.2.1 Model IIa and IIb Loop Functions

These functions are again expressed in terms of the function $\mathrm{L}(x)$ defined in eq.(A.16).

$$N_1 = -\frac{1}{6\,m_\phi^2}\left[\frac{1}{rs} - \frac{r^2-1}{r^3(r+s)}\mathrm{L}(r) + \frac{rs-s^2+r^2(3s^2-1)}{2\,r^3s^3}\ln\left(\frac{1}{s^2}\right) + \frac{1-3r^2}{2\,r^3(r+s)}\ln\left(\frac{r^2}{s^2}\right) - \frac{s^2-1}{s^3(r+s)}\mathrm{L}(s)\right]; \tag{A.47}$$

$$N_2 = \frac{1}{30\,m_\phi^4}\left[\frac{3r^2+2rs+3s^2-8r^2s^2}{r^2s^2(r+s)^2} + \frac{5r-5r^3+3s+r^2s+8r^4s}{r^4(r+s)^3}\mathrm{L}(r) + \frac{-3r^2+4rs-3s^2+5r^2s^2}{2r^4s^4}\ln\left(\frac{1}{s^2}\right) + \frac{5r-15r^3+3s-5r^2s}{2r^4(r+s)^3}\ln\left(\frac{r^2}{s^2}\right) + \frac{3r+5s+rs^2-5s^3+8rs^4}{s^4(r+s)^3}\mathrm{L}(s)\right]; \tag{A.48}$$

$$N_3 = \frac{1}{15\,m_\phi^4}\left[\frac{6r^2+9rs+s^2+4r^2s^2}{rs^3(r+s)^2} + \frac{1-3r^2-4r^4}{r^3(r+s)^3}\mathrm{L}(r) + \frac{-6r^2+3rs-s^2+5r^2s^2}{2r^3s^5}\ln\left(\frac{1}{s^2}\right) + \frac{1-5r^2}{2r^3(r+s)^3}\ln\left(\frac{r^2}{s^2}\right) + \frac{6r^2+15rs+10s^2-17r^2s^2-45rs^3-35s^4-2r^2s^4+10s^6+16r^2s^6}{s^5(r+s)^3(1-4s^2)}\mathrm{L}(s)\right]; \tag{A.49}$$

$$N_4 = \frac{1}{15\,m_\phi^4}\left[\frac{r^2+9rs+6s^2+4r^2s^2}{r^3s(r+s)^2} + \frac{-r^2+3rs-6s^2+5r^2s^2}{2r^5s^3}\ln\left(\frac{1}{r^2}\right) + \frac{10r^2-35r^4+10r^6+15rs-45r^3s+6s^2-17r^2s^2-2r^4s^2+16r^6s^2}{r^5(1-4r^2)(r+s)^3}\mathrm{L}(r) + \frac{-1+5s^2}{2s^3(r+s)^3}\ln\left(\frac{r^2}{s^2}\right) + \frac{1-3s^2-4s^4}{s^3(r+s)^3}\mathrm{L}(s)\right]; \tag{A.50}$$

$$N_5 = \frac{1}{6\,m_\phi^2}\left[\frac{1}{rs} - \frac{1-r^2}{r^3(r-s)}\mathrm{L}(r) + \frac{-r^2-rs-s^2+3r^2s^2}{2\,r^3s^3}\ln\left(\frac{1}{s^2}\right) + \frac{-1+3r^2}{2\,r^3(r-s)}\ln\left(\frac{r^2}{s^2}\right) - \frac{1-s^2}{s^3(-r+s)}\mathrm{L}(s)\right]; \tag{A.51}$$

$$N_6 = \frac{1}{30\,m_\phi^4}\Bigg[-\frac{3r^2-2rs+3s^2-8r^2s^2}{r^2s^2(r-s)^2} - \frac{5r-5r^3-3s-r^2s-8r^4s}{r^4(r-s)^3}\mathrm{L}(r) + \frac{3r^2+4rs+3s^2-5r^2s^2}{2r^4s^4}\ln\Big(\frac{1}{s^2}\Big) + \frac{-5r+15r^3+3s-5r^2s}{2r^4(r-s)^3}\ln\Big(\frac{r^2}{s^2}\Big) - \frac{3r-5s+rs^2+5s^3+8rs^4}{s^4(r-s)^3}\mathrm{L}(s)\Bigg]\,; \tag{A.52}$$

$$N_7 = \frac{1}{15\,m_\phi^4}\Bigg[-\frac{6r^2-9rs+s^2+4r^2s^2}{rs^3(r-s)^2} + \frac{1-3r^2-4r^4}{r^3(r-s)^3}\mathrm{L}(r) + \frac{6r^2+3rs+s^2-5r^2s^2}{2r^3s^5}\ln\Big(\frac{1}{s^2}\Big) + \frac{1-5r^2}{2r^3(r-s)^3}\ln\Big(\frac{r^2}{s^2}\Big) - \frac{6r^2-15rs+10s^2-17r^2s^2+45rs^3-35s^4-2r^2s^4+10s^6+16r^2s^6}{s^5(r-s)^3(1-4s^2)}\mathrm{L}(s)\Bigg]\,; \tag{A.53}$$

$$N_8 = \frac{1}{15\,m_\phi^4}\Bigg[-\frac{r^2-9rs+6s^2+4r^2s^2}{r^3s(r-s)^2} + \frac{r^2+3rs+6s^2-5r^2s^2}{2r^5s^3}\ln\Big(\frac{1}{r^2}\Big) + \frac{10r^2-35r^4+10r^6+15rs+45r^3s+6s^2-17r^2s^2-2r^4s^2+16r^6s^2}{r^5(1-4r^2)(r-s)^3}\mathrm{L}(r) + \frac{-1+5s^2}{2s^3(-r+s)^3}\ln\Big(\frac{r^2}{s^2}\Big) + \frac{1-3s^2-4s^4}{s^3(-r+s)^3}\mathrm{L}(s)\Bigg]\,; \tag{A.54}$$

$$P_1 = \frac{1}{3\,m_\phi^4}\Bigg[\frac{1}{s(r+s)} - \frac{1+2r^2}{2r^2(r+s)^2}\mathrm{L}(r) + \frac{-2r+s}{4r^2s^3}\ln\Big(\frac{1}{s^2}\Big) - \frac{1}{4r^2(r+s)^2}\ln\Big(\frac{r^2}{s^2}\Big) + \frac{2r+3s-4rs^2-6s^3+8rs^4}{2s^3(r+s)^2(1-4s^2)}\mathrm{L}(s)\Bigg]\,; \tag{A.55}$$

$$P_2 = \frac{1}{3\,m_\phi^4}\Bigg[-\frac{1}{r(r+s)} + \frac{-r+2s}{4r^3s^2}\ln\Big(\frac{1}{r^2}\Big) + \frac{1+2s^2}{2s^2(r+s)^2}\mathrm{L}(s) - \frac{1}{4s^2(r+s)^2}\ln\Big(\frac{r^2}{s^2}\Big) - \frac{3r-6r^3+2s-4r^2s+8r^4s}{2r^3(r+s)^2(1-4r^2)}\mathrm{L}(r)\Bigg]\,; \tag{A.56}$$

$$P_3 = \frac{1}{3\,m_\phi^4}\Bigg[-\frac{1}{s(r-s)} - \frac{1+2r^2}{2r^2(r-s)^2}\mathrm{L}(r) + \frac{2r+s}{4r^2s^3}\ln\Big(\frac{1}{s^2}\Big) - \frac{1}{4r^2(r-s)^2}\ln\Big(\frac{r^2}{s^2}\Big) - \frac{2r-3s-4rs^2+6s^3+8rs^4}{2s^3(r-s)^2(1-4s^2)}\mathrm{L}(s)\Bigg]\,; \tag{A.57}$$

$$P_4 = \frac{1}{3\,m_\phi^4}\Bigg[\frac{1}{r(r-s)} + \frac{r+2s}{4r^3s^2}\ln\left(\frac{1}{r^2}\right) - \frac{1+2s^2}{2s^2(-r+s)^2}\mathrm{L}(s) + \frac{1}{4s^2(r-s)^2}\ln\left(\frac{r^2}{s^2}\right) + \frac{3r-6r^3-2s+4r^2s-8r^4s}{2r^3(r-s)^2(1-4r^2)}\mathrm{L}(r)\Bigg]\,; \tag{A.58}$$

$$R_1 = \frac{1}{m_\phi^2}\left[-\frac{1}{r^2} - \frac{-1+r^2}{2r^4}\ln\left(\frac{1}{r^2}\right) - \frac{1-3r^2}{r^4\,(1-4r^2)}\,\mathrm{L}(r)\right]\,; \tag{A.59}$$

$$S_1 = \frac{1}{m_\phi^2}\left[-\frac{1}{s^2} - \frac{-1+s^2}{2s^4}\ln\left(\frac{1}{s^2}\right) - \frac{1-3s^2}{s^4\,(1-4s^2)}\,\mathrm{L}(s)\right]\,. \tag{A.60}$$

The functions appearing in the evaluation of the (crossed) box diagrams are again invariant under a simultaneous sign change of both arguments. In addition, they are pairwise related: $N_5(r,s) = N_1(-r,s)$, $N_6(r,s) = -N_2(-r,s)$, $N_7(r,s) = N_3(-r,s)$, $N_8(r,s) = N_4(-r,s)$, $P_3(r,s) = P_1(-r,s)$ and $P_4(r,s) = -P_2(-r,s)$.
We also again provide the **massless quark limits of the loop functions**. Since N_3, N_7, P_1 and P_3 get multiplied with $1/m_q$, we keep terms up to linear in r in these functions:

$$P_1, P_3 \xrightarrow{r\to 0} \frac{1}{2m_\phi^4}\left[\frac{1}{s^2} - \frac{1}{2s^4}\ln\left(\frac{1}{s^2}\right) + \frac{1-2s^2}{s^4}\mathrm{L}(s)\right] \pm \frac{r}{3sm_\phi^4}\Big[-\frac{2}{s^2} + \frac{1}{s^4}\ln\left(\frac{1}{s^2}\right) + \frac{4s^4+4s^2-2}{s^4(1-4s^2)}\mathrm{L}(s)\Big]\,; \tag{A.61}$$

$$P_2, -P_4 \xrightarrow{r\to 0} \frac{1}{6m_\phi^4}\left[\frac{1}{s^2} - \frac{1}{2s^4}\ln\left(\frac{1}{s^2}\right) + \frac{1+2s^2}{s^4}\mathrm{L}(s)\right]\,; \tag{A.62}$$

$$N_1,\ N_5 \xrightarrow{r\to 0} -\frac{1}{6m_\phi^4}\Big[\frac{1}{s^2} - \frac{1-3s^2}{2s^4}\ln\left(\frac{1}{s^2}\right) + \frac{1-s^2}{s^4}\mathrm{L}(s)\Big]\,; \tag{A.63}$$

$$N_2, -N_6 \xrightarrow{r\to 0} \frac{1}{6m_\phi^4}\Big[\frac{1}{s^4} - \frac{3}{2s^2} + \frac{-1+3s^2}{2s^6}\ln\left(\frac{1}{s^2}\right) + \frac{1-s^2}{s^6}\mathrm{L}(s)\Big]\,; \tag{A.64}$$

$$N_3,\ N_7 \xrightarrow{r\to 0} \frac{1}{3m_\phi^4}\Big[\frac{2}{s^4} + \frac{3s^2-2}{2s^6}\ln\left(\frac{1}{s^2}\right) + \frac{2-7s^2+2s^4}{s^6\,(1-4s^2)}\mathrm{L}(s)\Big] \mp \frac{r}{2s^3m_\phi^4}\Big[1 - \frac{2}{s^2} - \frac{2s^2-1}{s^4}\ln\left(\frac{1}{s^2}\right) - \frac{2}{s^4}\left(\frac{1-4s^2+2s^4}{1-4s^2}\right)\mathrm{L}(s)\Big]\,; \tag{A.65}$$

$$N_4,\ N_8 \xrightarrow{r\to 0} \frac{1}{15m_\phi^4}\Big[\frac{1}{s^4} - \frac{7}{2s^2} - \frac{1-5s^2}{2s^6}\ln\left(\frac{1}{s^2}\right) + \frac{1-3s^2-4s^4}{s^6}\mathrm{L}(s)\Big]\,; \tag{A.66}$$

$$R_1 \xrightarrow{r\to 0} \frac{1}{2m_\phi^2}\,. \tag{A.67}$$

In eq.(A.61) the + sign in the second line refers to P_1 and the − sign to P_3; similarly, in the second line of eq.(A.65) the + sign refers to N_3 and the − sign to N_7.

B 2–Loop Calculations

In this appendix we provide details of the calculation of the two–loop diagrams contributing to the neutron EDM.

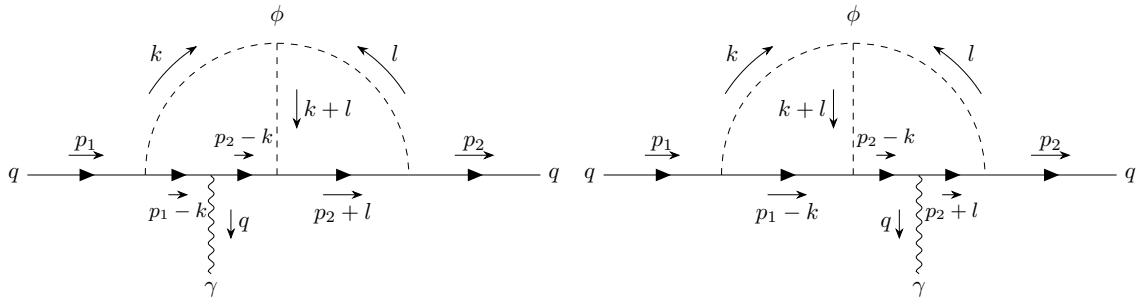


Figure 17: Two–loop Feynman diagrams generating quark EDMs and color–EDMs in Model I and IIa.

The matrix elements for diagram 1 (left) and diagram 2 (right) of Fig.17 generating EDMs in Model I and IIa can be written as:

$$i\mathcal{M}_1 = -\frac{i}{(16\,\pi^2)^2}\,2\,eQ_q\,(h_2^q)^3\,\mu_1\,m_\phi\,\epsilon_\mu^\star(q)\left[u(p_2)\,\mathcal{X}^\mu\,\gamma^5\,u(p_1)\right]\,; \tag{B.1}$$

$$i\mathcal{M}_2 = -\frac{i}{(16\,\pi^2)^2}\,2\,eQ_q\,(h_2^q)^3\,\mu_1\,m_\phi\,\epsilon_\mu^\star(q)\left[u(p_2)\,\mathcal{Y}^\mu\,\gamma^5\,u(p_1)\right]\,. \tag{B.2}$$

Here Q_q is the electric charge of the quark and the loop functions $\mathcal{X}^\mu = \sum_i^4 \mathcal{X}_i^\mu$ and $\mathcal{Y}^\mu = \sum_i^4 \mathcal{Y}_i^\mu$ are described below.

$$\mathcal{X}_1^\mu = \int_X \frac{-2x_1}{\xi^4}\,p_2^\mu(t_3+x_1z_1t_1)\left[\frac{-3}{\Delta'}+\frac{1}{\xi^2}\,\frac{2(p_1\cdot p_2)t_2(t_3+x_1z_1t_1)+m_q^2\,(t_2^2+(t_3+x_1z_1t_1)^2)}{\Delta'^2}\right]; \tag{B.3}$$

$$\mathcal{X}_2^\mu = \int_X \frac{2x_1}{\xi^2}\,p_2^\mu\left[\frac{-1}{\Delta'}+\frac{2}{\xi^2}\,\frac{\left(t_2\,(p_1\cdot p_2)+m_q^2\,(t_3+x_1z_1t_1)\right)(t_3+x_1z_1t_1)}{\Delta'^2}\right]; \tag{B.4}$$

$$\mathcal{X}_3^\mu = \int_X \frac{-2z_1m_q^2}{\xi^4}\,\frac{(t_3+x_1z_1t_1-t_2)}{\Delta'^2}\,\left(t_2\,p_1^\mu+(t_3+x_1z_1t_1)\,p_2^\mu\right); \tag{B.5}$$

$$\mathcal{X}_4^\mu = \int_X \frac{4z_1m_q^2}{\xi^2}\,\frac{(t_3+x_1z_1t_1)}{\Delta'^2}\,p_2^\mu\,. \tag{B.6}$$

Here $t_{1,2,3}$, x_1 and z_1 are the five Feynman parameters required to describe diagram 1. The five dimensional measure $\int_X$ is

$$\int_X \equiv \int_0^1 dt_1 \int_0^{1-t_1} dt_2 \int_0^{1-t_1-t_2} dt_3 \int_0^1 dx_1 \int_0^{1-x_1} dz_1, \quad \xi = \left[1-t_1(1-x_1(1-x_1))\right]^{1/2}. \tag{B.7}$$

Finally, the denominator Δ' is given by:

$$\Delta' = -m_\phi^2(z_1t_1+t_2+t_3-1)+z_1^2m_q^2t_1+\left(\frac{x_1z_1p_2t_1+p_1t_2+p_2t_3}{\xi}\right)^2. \tag{B.8}$$

The loop functions $\mathcal{Y}_i^\mu$ for the second diagram are obtained by the following replacements of Feynman parameters in the $\mathcal{X}_i^\mu$ in eqs.(B.3) to (B.6): $t_1 \to u_1$, $t_2 \to u_3$, $t_3 \to u_2$, $x_1 \to x_2$ and $z_1 \to z_2$, as well as the replacement $p_2^\mu \leftrightarrow p_1^\mu$. This yields:

$$\mathcal{Y}_1^\mu = \int_Y \frac{-2x_2}{\chi^4} p_1^\mu \left(u_2 + x_2 z_2 u_1\right) \left[\frac{-3}{\Delta''} + \frac{1}{\chi^2} \frac{2(p_1 \cdot p_2) u_3 (u_2 + x_2 z_2 u_1) + m_q^2 (u_3^2 + (u_2 + x_2 z_2 u_1)^2)}{\Delta''^2}\right]; \tag{B.9}$$

$$\mathcal{Y}_2^\mu = \int_Y \frac{2x_2}{\chi^2}\, p_1^\mu \left[\frac{-1}{\Delta''} + \frac{2}{\chi^2} \frac{\left(u_3 \,(p_1 \cdot p_2) + m_q^2 \,(u_2 + x_2 z_2 u_1)\right)(u_2 + x_2 z_2 u_1)}{\Delta'^2}\right]; \tag{B.10}$$

$$\mathcal{Y}_3^\mu = \int_Y \frac{2 z_2 m_q^2}{\chi^4} \frac{(u_3 - x_2 z_2 u_1 - u_2)}{\Delta''^2} \left(u_3\, p_2^\mu + (u_2 + x_2 z_2 u_1)\, p_1^\mu\right); \tag{B.11}$$

$$\mathcal{Y}_4^\mu = \int_Y \frac{4 z_2 m_q^2}{\chi^2} \frac{(u_2 + x_2 z_2 u_1)}{\Delta''^2}\, p_1^\mu\,. \tag{B.12}$$

The integration measure $\int_Y$ is

$$\int_Y \equiv \int_0^1 du_1 \int_0^{1-u_1} du_2 \int_0^{1-u_1-u_2} du_3 \int_0^1 dx_2 \int_0^{1-x_2} dz_2\,,\ \chi = \left[1 - u_1(1 - x_2(1 - x_2))\right]^{1/2}\,, \tag{B.13}$$

and the denominator Δ'' is given by

$$\Delta'' = -m_\phi^2 (z_2 u_1 + u_2 + u_3 - 1) + z_2^2 m_q^2 u_1 + \left(\frac{x_2 z_2 p_1 u_1 + p_1 u_2 + p_2 u_3}{\chi}\right)^2. \tag{B.14}$$

The qEDM is given by the coefficient of the dimension-5 $CP-$odd term $\bar{u}(p_2) i\sigma_{\mu\nu} q^\nu \gamma^5 u(p_1)$ in the limit of vanishing momentum transfer $q^2 \to 0$. The γ^5-version of the Gordon identity

$$\bar{u}\,(p_2)(p_1 + p_2)^\mu \gamma^5\, u(p_1) = \bar{u}(p_2)\, i\sigma^{\mu\nu} q_\nu \gamma^5\, u(p_1) \tag{B.15}$$

converts the matrix elements (B.1) into a suitable form to extract the qEDM. To that end the loop functions X_i^μ and Y_i^μ need to be transformed such that they are symmetric in the external quark momenta p_1 and p_2:

$$\mathcal{X}^\mu = [\mathcal{X}]\,(p_1 + p_2)^\mu + \{\mathcal{X}\}\, q^\mu, \tag{B.16}$$

$$\mathcal{Y}^\mu = [\mathcal{Y}]\,(p_1 + p_2)^\mu + \{\mathcal{Y}\}\, q^\mu. \tag{B.17}$$

The parts of the loop functions proportional to q^μ, denoted by $\{\mathcal{X}\}$ and $\{\mathcal{Y}\}$ respectively, can be ignored once the external spinors are taken into account by virtue of the Ward identity. The qEDM generated by the two diagrams is finally given by

$$d_q = \frac{2\, eQ_q\, (h_2^q)^3\, \mu_1\, m_\phi}{(16\,\pi^2)^2} \lim_{q^2 \to 0} \Big([\mathcal{X}] + [\mathcal{Y}]\Big)\,, \tag{B.18}$$

where

$$\lim_{q^2\to 0}[\mathcal{X}] = \frac{1}{2}\int_X -\frac{2z_1 m_q^2}{\xi^4}\frac{(t_3+x_1z_1t_1)^2-t_2^2}{(\Delta'|_{q^2\to 0})^2} + \frac{4z_1m_q^2}{\xi^2}\frac{t_3+x_1z_1t_1}{(\Delta'|_{q^2\to 0})^2} - \frac{2x_1}{\xi^4}(t_3+x_1z_1t_1)\left[\frac{-3}{\Delta'|_{q^2\to 0}} + \frac{m_q^2}{\xi^2}\frac{(t_3+t_2+x_1z_1t_1)^2}{(\Delta'|_{q^2\to 0})^2}\right] + \frac{2x_1}{\xi^2}\left[\frac{-1}{\Delta'|_{q^2\to 0}} + \frac{2m_q^2}{\xi^2}\frac{(t_3+t_2+x_1z_1t_1)(t_3+x_1z_1t_1)}{(\Delta'|_{q^2\to 0})^2}\right]; \tag{B.19}$$

$$\lim_{q^2\to 0}[\mathcal{Y}] = \frac{1}{2}\int_Y -\frac{2z_2 m_q^2}{\chi^4}\frac{(u_2+x_2z_2u_1)^2-u_3^2}{(\Delta''|_{q^2\to 0})^2} + \frac{4z_2m_q^2}{\chi^2}\frac{u_2+x_2z_2u_1}{(\Delta''|_{q^2\to 0})^2} - \frac{2x_2}{\chi^4}(u_2+x_2z_2u_1)\left[\frac{-3}{\Delta''|_{q^2\to 0}} + \frac{m_q^2}{\chi^2}\frac{(u_3+u_2+x_2z_2u_1)^2}{(\Delta''|_{q^2\to 0})^2}\right] + \frac{2x_2}{\chi^2}\left[\frac{-1}{\Delta''|_{q^2\to 0}} + \frac{2m_q^2}{\chi^2}\frac{(u_3+u_2+x_2z_2u_1)(u_2+x_2z_2u_1)}{(\Delta''|_{q^2\to 0})^2}\right]. \tag{B.20}$$

The limit $q^2 \to 0$ implies $p_1 \cdot p_2 \to m_q^2$, hence the denominators simplify to

$$\Delta'|_{q^2\to 0} \equiv \lim_{q^2\to 0}\Delta' = -m_\phi^2(z_1t_1+t_2+t_3-1) + m_q^2\left(z_1^2t_1 + \frac{(t_2+t_3+x_1z_1t_1)^2}{\xi^2}\right); \tag{B.21}$$

$$\Delta''|_{q^2\to 0} \equiv \lim_{q^2\to 0}\Delta'' = -m_\phi^2(z_2u_1+u_2+u_3-1) + m_q^2\left(z_2^2u_1 + \frac{(u_3+u_2+x_2z_2u_1)^2}{\chi^2}\right). \tag{B.22}$$

References

[1] XENON collaboration, *Excess electronic recoil events in XENON1T*, *Phys. Rev. D* **102** (2020) 072004 [2006.09721].

[2] XENON collaboration, *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, *Phys. Rev. Lett.* **121** (2018) 111302 [1805.12562].

[3] PandaX-4T collaboration, *Dark Matter Search Results from the PandaX-4T Commissioning Run*, *Phys. Rev. Lett.* **127** (2021) 261802 [2107.13438].

[4] LUX collaboration, *Results from a search for dark matter in the complete LUX exposure*, *Phys. Rev. Lett.* **118** (2017) 021303 [1608.07648].

[5] LUX collaboration, *Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure*, *Phys. Rev. Lett.* **118** (2017) 251302 [1705.03380].

[6] LZ collaboration, *First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment*, *Phys. Rev. Lett.* **131** (2023) 041002 [2207.03764].

[7] DEAP collaboration, *Search for dark matter with a 231-day exposure of liquid argon using DEAP-3600 at SNOLAB*, *Phys. Rev. D* **100** (2019) 022004 [1902.04048].

[8] SuperCDMS collaboration, *Results from the Super Cryogenic Dark Matter Search Experiment at Soudan*, *Phys. Rev. Lett.* **120** (2018) 061802 [1708.08869].

[9] DarkSide collaboration, *DarkSide-50 532-day Dark Matter Search with Low-Radioactivity Argon*, *Phys. Rev. D* **98** (2018) 102006 [1802.07198].

[10] CRESST collaboration, *First results from the CRESST-III low-mass dark matter program*, *Phys. Rev. D* **100** (2019) 102002 [1904.00498].

[11] NEWS-G collaboration, *First results from the NEWS-G direct dark matter search experiment at the LSM*, *Astropart. Phys.* **97** (2018) 54 [1706.04934].

[12] DAMIC collaboration, *Results on low-mass weakly interacting massive particles from a 11 kg-day target exposure of DAMIC at SNOLAB*, *Phys. Rev. Lett.* **125** (2020) 241803 [2007.15622].

[13] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of the PICO-60 C_3F_8 Bubble Chamber*, *Phys. Rev. D* **100** (2019) 022001 [1902.04031].

[14] LZ collaboration, *First Constraints on WIMP-Nucleon Effective Field Theory Couplings in an Extended Energy Region From LUX-ZEPLIN*, 2312.02030.

[15] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[16] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279 [hep-ph/0404175].

[17] J. Fan, M. Reece and L.-T. Wang, *Non-relativistic effective theory of dark matter direct detection*, *JCAP* **11** (2010) 042 [1008.1591].

[18] A.L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers and Y. Xu, *The Effective Field Theory of Dark Matter Direct Detection*, *JCAP* **02** (2013) 004 [1203.3542].

[19] A.L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers and Y. Xu, *Model Independent Direct Detection Analyses*, 1211.2818.

[20] N. Anand, A.L. Fitzpatrick and W.C. Haxton, *Weakly interacting massive particle-nucleus elastic scattering response*, *Phys. Rev. C* **89** (2014) 065501 [1308.6288].

[21] R. Catena, K. Fridell and M.B. Krauss, *Non-relativistic Effective Interactions of Spin 1 Dark Matter*, *JHEP* **08** (2019) 030 [1907.02910].

[22] P. Gondolo, S. Kang, S. Scopel and G. Tomar, *Effective theory of nuclear scattering for a WIMP of arbitrary spin*, *Phys. Rev. D* **104** (2021) 063017 [2008.05120].

[23] N. Bozorgnia, D.G. Cerdeño, A. Cheek and B. Penning, *Opening the energy window on direct dark matter detection*, *JCAP* **12** (2018) 013 [1810.05576].

[24] XENON collaboration, *Effective field theory search for high-energy nuclear recoils using the XENON100 dark matter detector*, *Phys. Rev. D* **96** (2017) 042004 [1705.02614].

[25] PandaX-II collaboration, *PandaX-II Constraints on Spin-Dependent WIMP-Nucleon Effective Interactions*, *Phys. Lett. B* **792** (2019) 193 [1807.01936].

[26] SuperCDMS collaboration, *Dark matter effective field theory scattering in direct detection experiments*, *Phys. Rev. D* **91** (2015) 092004 [1503.03379].

[27] CRESST collaboration, *Limits on Dark Matter Effective Field Theory Parameters with CRESST-II*, *Eur. Phys. J. C* **79** (2019) 43 [1809.03753].

[28] DarkSide-50 collaboration, *Effective field theory interactions for liquid argon target in DarkSide-50 experiment*, *Phys. Rev. D* **101** (2020) 062002 [2002.07794].

[29] LUX collaboration, *Effective field theory analysis of the first LUX dark matter search*, *Phys. Rev. D* **103** (2021) 122005 [2003.11141].

[30] DEAP collaboration, *Constraints on dark matter-nucleon effective couplings in the presence of kinematically distinct halo substructures using the DEAP-3600 detector*, *Phys. Rev. D* **102** (2020) 082001 [2005.14667].

[31] CDEX collaboration, *First experimental constraints on WIMP couplings in the effective field theory framework from CDEX*, *Sci. China Phys. Mech. Astron.* **64** (2021) 281011 [2007.15555].

[32] LUX collaboration, *Constraints on effective field theory couplings using 311.2 days of LUX data*, *Phys. Rev. D* **104** (2021) 062005 [2102.06998].

[33] SuperCDMS collaboration, *Effective Field Theory Analysis of CDMSlite Run 2 Data*, 2205.11683.

[34] R. Catena and P. Gondolo, *Global fits of the dark matter-nucleon effective interactions*, *JCAP* **09** (2014) 045 [1405.2637].

[35] R. Catena, *Prospects for direct detection of dark matter in an effective theory approach*, *JCAP* **07** (2014) 055 [1406.0524].

[36] R. Catena and P. Gondolo, *Global limits and interference patterns in dark matter direct detection*, *JCAP* **08** (2015) 022 [1504.06554].

[37] H. Rogers, D.G. Cerdeno, P. Cushman, F. Livet and V. Mandic, *Multidimensional effective field theory analysis for direct detection of dark matter*, *Phys. Rev. D* **95** (2017) 082003 [1612.09038].

[38] Z. Liu, Y. Su, Y.-L. Sming Tsai, B. Yu and Q. Yuan, *A combined analysis of PandaX, LUX, and XENON1T experiments within the framework of dark matter effective theory*, *JHEP* **11** (2017) 024 [1708.04630].

[39] S. Kang, S. Scopel, G. Tomar and J.-H. Yoon, *Present and projected sensitivities of Dark Matter direct detection experiments to effective WIMP-nucleus couplings*, *Astropart. Phys.* **109** (2019) 50 [1805.06113].

[40] N.P. Avis Kozar, P. Scott and A.C. Vincent, *A Global Fit of Non-Relativistic Effective Dark Matter Operators Including Solar Neutrinos*, 2310.15392.

[41] G. Lüders, *On the equivalence of invariance under time reversal and under particle-antiparticle conjugation for relativistic field theories*, *Dan. Mat. Fys. Medd.* **28** (1954) 1.

[42] W. Pauli, *Exclusion Principle, Lorentz Group, and reversal of space-time and charge*, *Niels Bohr and the Development of Physics, W. Pauli (ed.) New York: Pergamon* (1955) .

[43] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[44] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, *Phys. Rev. Lett.* **124** (2020) 081803 [2001.11966].

[45] J.M. Pendlebury et al., *Revised experimental upper limit on the electric dipole moment of the neutron*, *Phys. Rev. D* **92** (2015) 092003 [1509.04411].

[46] M. Burghoff et al., *An Improved Search for the Neutron Electric Dipole Moment*, in *Meeting of the APS Division of Particles and Fields*, 10, 2011 [1110.1505].

[47] J.B. Dent, L.M. Krauss, J.L. Newstead and S. Sabharwal, *General analysis of direct dark matter detection: From microphysics to observational signatures*, *Phys. Rev. D* **92** (2015) 063515 [1505.03117].

[48] F. Bishara, J. Brod, B. Grinstein and J. Zupan, *From quarks to nucleons in dark matter direct detection*, *JHEP* **11** (2017) 059 [1707.06998].

[49] E. Del Nobile, *Complete Lorentz-to-Galileo dictionary for direct dark matter detection*, *Phys. Rev. D* **98** (2018) 123003 [1806.01291].

[50] M. Drees and R. Mehra, *Neutron EDM constrains direct dark matter detection prospects*, *Phys. Lett. B* **799** (2019) 135039 [1907.10075].

[51] G. Arcadi, M. Lindner, F.S. Queiroz, W. Rodejohann and S. Vogl, *Pseudoscalar Mediators: A WIMP model at the Neutrino Floor*, *JCAP* **03** (2018) 042 [1711.02110].

[52] N.F. Bell, G. Busoni and I.W. Sanderson, *Loop Effects in Direct Detection*, *JCAP* **08** (2018) 017 [1803.01574].

[53] T. Li, *Revisiting the direct detection of dark matter in simplified models*, *Phys. Lett. B* **782** (2018) 497 [1804.02120].

[54] T. Abe, M. Fujiwara and J. Hisano, *Loop corrections to dark matter direct detection in a pseudoscalar mediator dark matter model*, *JHEP* **02** (2019) 028 [1810.01039].

[55] F. Ertas and F. Kahlhoefer, *Loop-induced direct detection signatures from CP-violating scalar mediators*, *JHEP* **06** (2019) 052 [1902.11070].

[56] E. Del Nobile, *The Theory of Direct Dark Matter Detection: A Guide to Computations*, 2104.12785.

[57] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195 [hep-ph/9506380].

[58] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Remarks on Higgs Boson Interactions with Nucleons*, *Phys. Lett. B* **78** (1978) 443.

[59] T. Bhattacharya, V. Cirigliano, R. Gupta, H.-W. Lin and B. Yoon, *Neutron Electric Dipole Moment and Tensor Charges from Lattice QCD*, *Phys. Rev. Lett.* **115** (2015) 212002 [1506.04196].

[60] PNDME collaboration, *Iso-vector and Iso-scalar Tensor Charges of the Nucleon from Lattice QCD*, *Phys. Rev. D* **92** (2015) 094511 [1506.06411].

[61] R. Gupta, B. Yoon, T. Bhattacharya, V. Cirigliano, Y.-C. Jang and H.-W. Lin, *Flavor diagonal tensor charges of the nucleon from (2+1+1)-flavor lattice QCD*, *Phys. Rev. D* **98** (2018) 091501 [1808.07597].

[62] JLQCD collaboration, *Nucleon charges with dynamical overlap fermions*, *Phys. Rev. D* **98** (2018) 054516 [1805.10507].

[63] C. Alexandrou et al., *Nucleon scalar and tensor charges using lattice QCD simulations at the physical value of the pion mass*, *Phys. Rev. D* **95** (2017) 114514 [1703.08788].

[64] M. Pospelov and A. Ritz, *Neutron EDM from electric and chromoelectric dipole moments of quarks*, *Phys. Rev. D* **63** (2001) 073015 [hep-ph/0010037].

[65] J. Hisano, J.Y. Lee, N. Nagata and Y. Shimizu, *Reevaluation of Neutron Electric Dipole Moment with QCD Sum Rules*, *Phys. Rev. D* **85** (2012) 114044 [1204.2653].

[66] K. Fuyuto, J. Hisano and N. Nagata, *Neutron electric dipole moment induced by strangeness revisited*, *Phys. Rev. D* **87** (2013) 054018 [1211.5228].

[67] T. Hahn, *CUBA: A Library for multidimensional numerical integration*, *Comput. Phys. Commun.* **168** (2005) 78 [hep-ph/0404043].

[68] J.B. Dent, B. Dutta, J.L. Newstead and L.E. Strigari, *Effective field theory treatment of the neutrino background in direct dark matter detection experiments*, *Phys. Rev. D* **93** (2016) 075018 [1602.05300].

[69] Y. Ema, T. Gao and M. Pospelov, *Reevaluation of heavy-fermion-induced electron EDM at three loops*, *Phys. Lett. B* **835** (2022) 137496 [2207.01679].

[70] N. Yamanaka and E. Hiyama, *Weinberg operator contribution to the nucleon electric dipole moment in the quark model*, *Phys. Rev. D* **103** (2021) 035023 [2011.02531].

[71] J. Ellis, *TikZ-Feynman: Feynman diagrams with TikZ*, *Comput. Phys. Commun.* **210** (2017) 103 [1601.05437].

[72] H.H. Patel, *Package-X 2.0: A Mathematica package for the analytic calculation of one-loop integrals*, *Comput. Phys. Commun.* **218** (2017) 66 [1612.00009].

A Global View of the EDM Landscape

Skyler Degenkolb[1], Nina Elmer[2],
Tanmoy Modak[2], Margarete Mühlleitner[3], and Tilman Plehn[2,4]

1 Physikalisches Institut, Universität Heidelberg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Institute for Theoretical Physics, Karlsruhe Institute of Technology, Karlsruhe, Germany
4 Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, Germany

arXiv:2403.02052v1 [hep-ph] 4 Mar 2024

Abstract

Permanent electric dipole moments (EDMs) are sensitive probes of the symmetry structure of elementary particles, which in turn is closely tied to the baryon asymmetry in the universe. A meaningful interpretation framework for EDM measurements has to be based on effective quantum field theory. We interpret the measurements performed to date in terms of a hadronic-scale Lagrangian, using the SFitter global analysis framework. We find that part of this Lagrangian is constrained very well, while some of the parameters suffer from too few high-precision measurements. Theory uncertainties lead to weaker model constraints, but can be controlled within the global analysis.

Contents

1 Introduction **2**

2 EDM Lagrangian **3**
2.1 Weak-scale Lagrangian 3
2.2 Hadronic-scale Lagrangian 4
2.3 Matched Lagrangians for semileptonic interactions 5

3 EDM Measurements **8**
3.1 Nucleons 8
3.2 Paramagnetic systems 9
3.3 Diamagnetic systems 12
3.4 Single parameter ranges 14

4 Global analysis **15**
4.1 SFitter framework 15
4.2 Well-constrained model sub-space 16
4.3 Hadronic parameters from diamagnetic systems 17
4.4 Poorly constrained model parameters 18
4.5 Theory uncertainties 20

5 Outlook **22**

References **24**

1 Introduction

While the Standard Model (SM) is structurally complete, it fails to explain key observations and therefore fails to qualify as a complete theory of elementary particles. The two leading shortcomings are a missing dark matter agent and a missing explanation of the baryon asymmetry in the Universe. For the latter, the Sakharov conditions [1] tell us precisely which structures would be required: (i) C- and CP-violation, (ii) baryon number violation, and (iii) a deviation from thermal equilibrium. The first condition is especially interesting, because we can read it off the fundamental Lagrangian and build, for instance, models for electroweak baryogenesis around it [2–5]. In the SM, CP symmetry is violated through the fermion mixing among three generations and through the adjoint gluon field strength. Measurements of the neutron EDM, consistent with zero, clearly show that CP violation in QCD is too small to explain the observed baryon asymmetry [6–11]. Physics beyond the Standard Model (BSM), explaining the baryon asymmetry, should then violate CP to a greater extent and with observable consequences.

Over recent years, EDM measurements have been performed on a wide range of particles, atoms, and molecules [12–15]. None of them has been able to confirm a signal for CP violation. To judge their combined impact on BSM physics, we need to combine them in a consistent framework. Such a global analysis must start from a Lagrangian and express the experimental limits on all EDMs in terms of its fundamental parameters. One choice of Lagrangian is the hadronic-scale Lagrangian, describing the interactions of nucleons, pions, and electrons at the GeV scale [12, 15–17]. Alternatively, we can combine the EDM measurements at the weak scale, using an effective extension of the renormalizable SM Lagrangian [13, 17–22]. Going beyond effective field theories (EFTs), an ultimate link to the cosmological motivation requires a UV-complete extension of the SM at the weak scale, for instance leptoquark models [17], extended Higgs sectors [23–25], left-right symmetry [26], or supersymmetry [27–32]. Technically, all global analyses [33] face similar challenges which we will tackle in this paper for the hadronic-scale Lagrangian.

We perform a global EDM analysis with the SFitter analysis tool [34–38], with its focus on the comprehensive treatment of uncertainties. Our analysis will provide state-of-the-art limits on the multi-dimensional model parameter space, and it will allow us to judge the impact of new or proposed measurements and to identify shortcomings in relating measurements to fundamental parameters[1], while remaining easy to repeat or adapt in response to new theoretical or experimental inputs. For this purpose it is crucial that we interpret all measurements in the same fundamental physics framework and include all uncertainties, including theory uncertainties, even though they lack a statistical interpretation [39].

We start by introducing the consistent hadronic-scale Lagrangian with properly chosen Wilson coefficients for our global analysis in Sec. 2. We then use this Lagrangian to provide predictions for the measured EDMs, as detailed in Sec. 3. We start our global analysis without theory uncertainties in Sec. 4, to understand the relations among different EDM measurements in terms of the hadronic-scale Lagrangian. To extract correlations and limits on single model parameters we employ a profile likelihood. We find that the current EDM measurements define a subspace of well-constrained model parameters and an orthogonal subspace of poorly constrained parameters with narrow correlation patterns. Adding theory uncertainties on the relations between Lagrangian parameters and observables in Sec. 4.5 degrades the interpretation in terms of fundamental physics. We emphasize that this degradation does not cut into the discovery potential of EDM measurements, as probes of fundamental symmetries of elementary particles, but it hampers their interpretation as limits in fundamental physics.

[1] Only a proper fundamental physics interpretation can make full use of experimental limits.

2 EDM Lagrangian

Because CP violation is motivated by cosmology and can be related to physics beyond the SM at and above the weak scale, we start by introducing it into the weak-scale Lagrangian in Sec. 2.1, relate it to the GeV-scale Lagrangian in Sec. 2.2, and use simple matching arguments to simplify this hadronic-scale Lagrangian which we use as the interpretation framework for our global analysis in Sec. 2.3. A detailed analysis of CP-violating new physics at the electroweak scale includes a renormalization group evolution from the GeV-scale to the electroweak scale and will not be part of this first study. Instead, we will focus here on a global analysis at the hadronic scale and the role of correlations and theory uncertainties.

2.1 Weak-scale Lagrangian

The operators generating CP violation within and beyond the SM-Lagrangian, neglecting CP violation in the neutrino sector [13], appear in the Lagrangian

$$\mathcal{L}_{\text{CPV}} = \mathcal{L}_{\text{CKM}} + \mathcal{L}_{\bar{\theta}} + \mathcal{L}_{\text{dipole}} + \mathcal{L}_{\text{Weinberg}} + \mathcal{L}_{\text{EFT}} \,. \tag{1}$$

The first term represents CP violation at mass dimension four, from the complex phases in the CKM matrix. The second arises from the gluon field strength, also at dimension four,

$$\mathcal{L}_{\bar{\theta}} = \frac{g_3^2}{32\pi^2}\bar{\theta}\,\text{Tr}(G^{\mu\nu}\tilde{G}_{\mu\nu})\,, \tag{2}$$

where g_3 is the strong coupling, $G^{\mu\nu}$ is the gluon field strength, $\tilde{G}^{\mu\nu} = \epsilon^{\mu\nu\lambda\sigma}G_{\lambda\sigma}/2$ is its dual, and $\bar{\theta}$ is the re-scaled CP-violating parameter in QCD. The bar notation indicates that corrections from the quark mass matrix are included. In principle $\bar{\theta}$ can also be included as a model parameter, for instance in testing a specific BSM model [26]. However, we take the view that the neutron EDM experimentally constrains $\bar{\theta}$ to have such a small value that this fine-tuning problem requires a proper explanation. This means there is little to be learned from including $\bar{\theta}$ as a model parameter in our global analysis.

The other three contributions in Eq.(1) are higher-dimensional and not part of a renormalizable extension of the SM-Lagrangian. Electric dipole moments of fermions d_f^E, and chromo-electric dipole moments of quarks d_q^C, appear at mass dimension five:

$$\mathcal{L}_{\text{dipole}} = -\frac{i}{2}F^{\mu\nu}\sum_{f=q,\ell} d_f^E\,\left(\bar{f}\,\sigma_{\mu\nu}\gamma_5 f\right) - \frac{i}{2}g_3 G^a_{\mu\nu}\sum_{f=q} d_q^C\,\left(\bar{q}\sigma^{\mu\nu}\gamma_5 T^a q\right)\,. \tag{3}$$

where the indices q and ℓ denote quarks and leptons of all three generations. The electromagnetic field strength is $F^{\mu\nu}$. We chose the metric convention $(1,-\mathbb{1})$ with $\gamma_5 = -i\gamma_0\gamma_1\gamma_2\gamma_3$. The fermion spins are $\sigma_{\mu\nu} = i\left[\gamma_\mu,\gamma_\nu\right]/2$, and T^a are the $SU(3)$ generators. The Weinberg operator is again built out of the gluon field strength and introduces the gluonic chromo-electric dipole moment d^G,

$$\mathcal{L}_{\text{Weinberg}} = \frac{1}{3}d^G\,f_{abc}G^a_{\mu\nu}\tilde{G}^{b\,\nu\rho}G^{c\,\mu}_{\rho}\,. \tag{4}$$

Additional CP-violation occurs at mass dimension six and higher, generated at a new physics scale larger than the vacuum expectation value, $\Lambda > v$,

$$\mathcal{L}_{\text{EFT}} = \sum_i \frac{C_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \mathcal{O}(\Lambda^{-3})\,. \tag{5}$$

Some relevant dimension-6 operators that generate EDMs include the semileptonic and hadronic 4-fermion operators

$$\begin{aligned}\mathscr{L}_{\rm EFT} \supset & C_{\ell eqd}\,\left(\bar{L}^j e_R\right)\left(\bar{d}_R Q_j\right) + C^{(1)}_{\ell equ}\,\left(\bar{L}^j e_R\right)\epsilon_{jk}\left(\bar{Q}^k u_R\right) + C^{(3)}_{\ell equ}\,\left(\bar{L}^j \sigma_{\mu\nu} e_R\right)\epsilon_{jk}\left(\bar{Q}^k \sigma_{\mu\nu} u_R\right) \\ & + C^{(1)}_{quqd}\,\left(\bar{Q}^j u_R\right)\epsilon_{jk}\left(\bar{Q}^k d_R\right) + C^{(8)}_{quqd}\,\left(\bar{Q}^j T^a u_R\right)\epsilon_{jk}\left(\bar{Q}^k T^a d_R\right) + \text{h.c.} \end{aligned} \tag{6}$$

The dipole moments, the Weinberg operator, and the additional 4-fermion interactions can serve as the basis for an EDM analyses in the SMEFT framework [12, 13, 17–23, 28–30]. As always, the set of higher-dimensional SMEFT operators that turn out to be most relevant depends on the high-scale BSM model that the SMEFT represents. For instance, in supersymmetric models there are no contributions from $\mathcal{O}^{(8)}_{quqd}$ and $\mathcal{O}^{(3)}_{\ell equ}$ at tree level, and the relative size of down-type and up-type quark couplings is affected by a potentially large $\tan\beta$ enhancement.

2.2 Hadronic-scale Lagrangian

The challenge with EDMs in view of the Lagrangian of Eq.(1) is that they are measured far below the electroweak scale, where the propagating degrees of freedom are leptons, non-relativistic nucleons $N = (p, n)^T$ with average mass m_N, and pions $\vec{\pi} = (\pi^+, \pi^0, \pi^-)^T$ [12, 15–17, 28, 30].

When we evolve our EFT to the experimentally relevant GeV scale, only the lepton EDMs $d^E_\ell \equiv d_\ell$ in Eq.(3) of the electroweak Lagrangian in Eq.(1) remain unchanged. While for the weak-scale Lagrangian the relation between the three charged leptons, for instance the scaling with the lepton mass, raises interesting questions, we factorize the muon and tau EDMs from the hadronic-scale Lagrangian. They can be included in the same framework, but given the systems for which experimental limits are available today, the indirect constraints on EDMs of heavy leptons are many orders of magnitude weaker than direct ones and there is little interplay among the relevant model parameters.

We split the hadronic-scale Lagrangian describing EDMs at the experimentally relevant GeV scale into

$$\mathscr{L}_{\rm had} \supset \mathscr{L}_{N,\rm sr} + \mathscr{L}_{\pi N} + \mathscr{L}_{eN} - \frac{i}{2} F^{\mu\nu} d_e\, \bar{e}\sigma_{\mu\nu}\gamma_5 e\,. \tag{7}$$

The dipole moments of the nucleons now read

$$\mathscr{L}_{N,\rm sr} = -2\bar{N}\left[d^{\rm sr}_p \frac{1+\tau_3}{2} + d^{\rm sr}_n \frac{1-\tau_3}{2}\right] S_\mu N v_\nu F^{\mu\nu}\,, \tag{8}$$

where S_μ and v_μ are the spin and velocity of the nucleon. The isoscalar and isovector contributions define $d^{\rm sr}_N$ as the Lagrangian parameters for the short-range contributions to the proton and neutron EDMs.

Next come the interactions of pions and nucleons,

$$\begin{aligned}\mathscr{L}_{\pi N} = & \bar{N}\Big[g^{(0)}_\pi \vec{\tau}\cdot\vec{\pi} + g^{(1)}_\pi \pi^0 + g^{(2)}_\pi\left(3\tau_3\pi^0 - \vec{\tau}\cdot\vec{\pi}\right)\Big]N \\ & + C_1\,\left(\bar{N}N\right)\,\partial_\mu\left(\bar{N}S^\mu\bar{N}\right) + C_2\,\left(\bar{N}\vec{\tau}N\right)\cdot\partial_\mu\left(\bar{N}S^\mu\bar{N}\vec{\tau}\right) + \cdots \end{aligned} \tag{9}$$

where τ are the Pauli matrices and we neglect, for example, interactions with more than one pion. The contribution involving $g^{(2)}_\pi$ is suppressed relative to $g^{(0,1)}_\pi$ by one order in the chiral expansion [40, 41], but can be taken into account in similar fashion. In our parameterization these interactions contribute, for instance, to the nucleon EDMs through calculable pion

loops [16]. Naive dimensional analysis [42] suggests that short-range nucleon interactions enter only at NNLO [43] in the chiral expansion [44] and can be neglected [41]. In Eq.(9) this applies to all interactions in the second line. One caveat is that a consistent treatment of long-range effects may require additional short-distance counter terms [43] that appear as effective short-range nucleon-nucleon forces. These are not presently taken into account.

Finally, the higher-dimensional operators in Eq.(6) induce effective interactions that can be organized according to their tensor structure, isospin character, and their dependence on the electron and nucleon fields and spins [15,17]:

$$\begin{aligned}\mathscr{L}_{eN} =& -\frac{G_F}{\sqrt{2}}\,(\bar{e}i\gamma_5 e)\,\bar{N}\left(C_S^{(0)}+C_S^{(1)}\tau_3\right)N \\ &+\frac{8G_F}{\sqrt{2}}\,v_\nu\,(\bar{e}\sigma^{\mu\nu}e)\,\bar{N}\left(C_T^{(0)}+C_T^{(1)}\tau_3\right)S_\mu N \\ &-\frac{G_F}{\sqrt{2}}\,(\bar{e}e)\,\frac{\partial^\mu}{m_N}\left[\bar{N}\left(C_P^{(0)}+C_P^{(1)}\tau_3\right)S_\mu N\right]. \end{aligned} \tag{10}$$

In a heavy baryon expansion, the last line can be dropped at leading order [13]. However, we retain all three terms since (1) a pion pole enhancement of the isovector contribution somewhat offsets this hierarchy and (2) contributions of heavy quarks can also render it relevant for some new physics models. At this stage, the independent Lagrangian parameters for our global EDM analysis at the hadronic scale are:

$$\left\{\, d_e, C_S^{(0,1)}, C_T^{(0,1)}, C_P^{(0,1)}, g_\pi^{(0)}, g_\pi^{(1)}, d_{n,p}^{\rm sr} \,\right\}. \tag{11}$$

2.3 Matched Lagrangians for semileptonic interactions

The set of model parameters defined in Eq.(11) can be further simplified by matching the semileptonic part of the hadronic-scale Lagrangian of Eq.(10) to the corresponding weak-scale 4-fermion interactions of Eq.(6), both evaluated for external nucleons. The light quark content in the nucleons is related to the nucleon Lagrangian as

$$\begin{aligned} g_S^{(0)}\,\bar{\psi}_N\psi_N &= \frac{1}{2}\langle N\,|\bar{u}u+\bar{d}d|\,N\rangle \\ g_S^{(1)}\,\bar{\psi}_N\tau_3\psi_N &= \frac{1}{2}\langle N\,|\bar{u}u-\bar{d}d|\,N\rangle \\ g_T^{(0)}\,\bar{\psi}_N\sigma_{\mu\nu}\psi_N &= \frac{1}{2}\langle N\,|\bar{u}\sigma_{\mu\nu}u+\bar{d}\sigma_{\mu\nu}d|\,N\rangle \\ g_T^{(1)}\,\bar{\psi}_N\sigma_{\mu\nu}\tau_3\psi_N &= \frac{1}{2}\langle N\,|\bar{u}\sigma_{\mu\nu}u-\bar{d}\sigma_{\mu\nu}d|\,N\rangle \\ g_P^{(0)}\,\bar{\psi}_N\gamma_5\psi_N &= \frac{1}{2}\langle N\,|\bar{u}\gamma_5 u+\bar{d}\gamma_5 d|\,N\rangle \\ g_P^{(1)}\,\bar{\psi}_N\gamma_5\tau_3\psi_N &= \frac{1}{2}\langle N\,|\bar{u}\gamma_5 u-\bar{d}\gamma_5 d|\,N\rangle, \end{aligned} \tag{12}$$

which relations define the scalar, tensor, and pseudoscalar nucleon form factors $g_{S,T,P}^{(0,1)}$. Using them, the hadronic-scale Wilson coefficients in Eq.(10) can be matched to the SMEFT Wilson

coefficients in Eq.(6) as [13, 15],

$$\begin{aligned}
C_S^{(0)} &= -g_S^{(0)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell edq} - C_{\ell equ}^{(1)}\right) & C_S^{(1)} &= g_S^{(1)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell edq} + C_{\ell equ}^{(1)}\right) \\
C_T^{(0)} &= -g_T^{(0)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell equ}^{(3)}\right) & C_T^{(1)} &= -g_T^{(1)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell equ}^{(3)}\right) \\
C_P^{(0)} &= g_P^{(0)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell edq} + C_{\ell equ}^{(1)}\right) & C_P^{(1)} &= -g_P^{(1)} \frac{v^2}{\Lambda^2} \operatorname{Im}\left(C_{\ell edq} - C_{\ell equ}^{(1)}\right) .
\end{aligned} \tag{13}$$

Here the six couplings $C_{S,T,P}^{(0,1)}$ are expressed in terms of only three SMEFT Wilson coefficients, implying

$$\frac{C_P^{(0)}}{g_P^{(0)}} = \frac{C_S^{(1)}}{g_S^{(1)}} \qquad \frac{C_T^{(0)}}{g_T^{(0)}} = \frac{C_T^{(1)}}{g_T^{(1)}} \qquad \frac{C_S^{(0)}}{g_S^{(0)}} = \frac{C_P^{(1)}}{g_P^{(1)}} . \tag{14}$$

Only three independent semileptonic parameters actually enter the hadronic-scale global analysis. We choose them as $C_{S,T,P}^{(0)}$ and combine them using the known ratios of hadronic matrix elements to construct the full Lagrangian. In addition to the light quarks described by Eq.(13), the relations in Eq.(14) also have to include the contributions of heavy quarks. These contributions are contained in the nucleon form factors, renormalized at an appropriate mass scale and accounting for the corresponding anomaly relations.

Note that in this way our global analysis preserves the full dependence on $C_{S,T,P}^{(0,1)}$, although only $C_{S,T,P}^{(0)}$ appear as model parameters. This remains the case regardless of isospin-violating effects, and also when including experimental systems for which the coefficients of $C_S^{(1)}$ differ significantly, as discussed below.

The implementation of Eq.(13) for $C_S^{(0,1)}$ is particularly straightforward, because for each experimental system the effective parameter that combines the isoscalar and isovector terms is independent of the nuclear spin,

$$\begin{aligned}
C_S &= C_S^{(0)} + \frac{Z-N}{Z+N} C_S^{(1)} \\
&= C_S^{(0)} + \frac{Z-N}{Z+N} \frac{g_S^{(1)}}{g_P^{(0)}} C_P^{(0)} \qquad \text{with} \qquad \frac{g_S^{(1)}}{g_P^{(0)}} \approx 0.1 .
\end{aligned} \tag{15}$$

In the second step we replace $C_S^{(1)}$ with $C_P^{(0)}$, reflecting Eq.(13). Since $g_S^{(1)}$ is already suppressed relative to $g_S^{(0)}$ by the small isospin violation of the nucleon matrix element, one could argue that $C_S^{(1)} \ll C_S^{(0)}$. Moreover, in the heavy nuclei of all atomic and molecular systems for which EDMs have been measured so far, the isoscalar and isovector contributions occur in approximately the same ratio, $(Z-N)/(Z+N) \approx -0.2$, so the effective parameter C_S is approximately system-independent. As noted above, we do not rely on these assumptions.

Next, we relate the pseudoscalar and tensor semileptonic interactions in a similar fashion. We start with the linear combinations for nucleons, $C_{P,T}^{(n,p)} = C_{P,T}^{(0)} \mp C_{P,T}^{(1)}$, taking the particle physics convention $\tau_3 |n\rangle = -|n\rangle$). From those, the coefficients for a given nucleus can be constructed according to the sum over spins of the constituent nucleons, where $\langle \sigma_{p,n} \rangle$ represents the average neutron or proton spin, evaluated for the measured nuclear state [45, 46].

$$C_{P,T} = \frac{C_{P,T}^{(n)} \langle \sigma_n \rangle + C_{P,T}^{(p)} \langle \sigma_p \rangle}{\langle \sigma_n \rangle + \langle \sigma_p \rangle} . \tag{16}$$

For C_T we can see from Eq.(13) that the isoscalar and isovector couplings differ only through the corresponding nucleon form factors. These are calculated with small theoretical uncertainties in lattice QCD [47], allowing us to write

$$C_T = \left(1 - \frac{g_T^{(1)}}{g_T^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle}\right) C_T^{(0)} \qquad \text{with} \qquad \frac{g_T^{(1)}}{g_T^{(0)}} \approx 1.7\,. \tag{17}$$

Similarly, for C_P it can be shown that

$$C_P = C_P^{(0)} - \frac{g_P^{(1)}}{g_S^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle} C_S^{(0)} \qquad \text{with} \qquad \frac{g_P^{(1)}}{g_S^{(0)}} \approx 20.2\,. \tag{18}$$

For the derivation of $g_P^{(1)}$ in this ratio we follow Ref. [13]. We consider only the first generation as relevant light quarks, such that $g_P^{(1)}$ is dominated by the π-pole contribution

$$g_P^{(1)} = \frac{g_A \bar{m}_N}{\bar{m}} \frac{m_\pi^2}{m_\pi^2 - q^2} + \text{heavy quarks} \qquad \bar{m} = \frac{m_u + m_d}{2}\,. \tag{19}$$

Here $\bar{m}_N \approx 940$ MeV is the average nucleon mass, $\bar{m}$ the average light quark mass, and g_A is the (isovector) axial vector coupling [48]. The coupling $g_P^{(0)}$, appearing in Eq.(15), involves the isoscalar axial coupling $g_A^{(0)}$, obtained from the sum rather than the difference of the light quark axial charges. We extend this by including a light s-quark, where the π-pole dominance is replaced by an octet η-pole with an appropriately modified average light quark mass m^*

$$g_P^{(0)} = \frac{g_A^{(0)} \bar{m}_N}{m^*} \frac{m_\eta^2}{m_\eta^2 - q^2} + \text{heavy quarks} \qquad m^* = \frac{m_u + m_d + 2m_s}{3}\,. \tag{20}$$

Note that heavy quark contributions enter differently in $g_P^{(0)}$ and $g_P^{(1)}$, and can be derived using the $U(1)_A$ axial anomaly together with the divergence of the anomaly-free axial current $J_{\mu 5}^q = \bar{q}\gamma_\mu\gamma_5 q$ for all quarks q [49–52]. These represent a relatively minor contribution to $g_P^{(1)}$, but due to suppression of the η-pole relative to the π-pole by the factor $m^*/\bar{m}$, contribute to $g_P^{(0)}$ at approximately the same level as the light quarks.

Our final, simplifying assumption is not strictly needed, but is effective in reducing the number of model parameters by one,

$$d_p^{\text{sr}} \approx -d_n^{\text{sr}}\,. \tag{21}$$

The nuclei of diamagnetic systems, which apart from the neutron itself provide the strongest constraints on nucleon EDMs, are typically dominated either by a valence proton or a valence neutron. As for the anomalous magnetic moment, the short-range nucleon EDMs are assumed to be dominated by the isovector contribution. This assumption can be relaxed for the purposes of a rigorous global analysis, provided improved theory uncertainties concerning the contributions of all nucleons to the overall EDM of each measured system. Of the present experimental limits, d_p^{sr} is constrained essentially only by the TlF measurement.

With these simplifications, the set of low-energy parameters given in Eq.(11) reduces to

$$c_j \in \left\{ d_e, C_S^{(0)}, C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\text{sr}} \right\}. \tag{22}$$

These Lagrangian parameters define the model parameters for our global EDM analysis.

3 EDM Measurements

In terms of the Lagrangian parameters in Eq.(22) we can predict the measured EDMs d_i as linear combinations with system-specific coefficients α_{i,c_j},

$$d_i = \sum_{c_j} \alpha_{i,c_j} c_j \,. \tag{23}$$

The measurements we analyze are listed in Tab. 1 and discussed below. Unless otherwise indicated, the isotopes and charge states that we discuss are those given in Tab. 1. The isotopes that are relevant for these molecular systems are ^{180}Hf, ^{232}Th, ^{174}Yb, ^{205}Tl, ^{16}O, and ^{19}F. We neglect constraints from the comparatively weaker experimental bounds from ^{85}Rb [53, 54], Xe^m [55], PbO [56], $Eu_{0.5}Ba_{0.5}TiO_3$ [57], and the Λ hyperon [58]. The experimental bounds for the μ and τ leptons constrain the corresponding Lagrangian parameters and factorize from the hadronic-scale Lagrangian. We do not include them in our first global analysis.

As discussed in the last section, the α-values for all $C_{S,P,T}^{(0,1)}$ can be extracted from the corresponding values for the basis given in Eq.(22). For this we follow Eqs. (15)-(18) and find for example

$$\begin{aligned}
\alpha_{C_S^{(0)}} &= \alpha_{C_S} - \alpha_{C_P} \frac{g_P^{(1)}}{g_S^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle} \\
\alpha_{C_P^{(0)}} &= \alpha_{C_P} + \alpha_{C_S} \frac{g_S^{(1)}}{g_P^{(0)}} \frac{Z-N}{Z+N} \\
\alpha_{C_T^{(0)}} &= \left(1 - \frac{g_T^{(1)}}{g_T^{(0)}} \frac{\langle\sigma_n\rangle - \langle\sigma_p\rangle}{\langle\sigma_n\rangle + \langle\sigma_p\rangle}\right) \alpha_{C_T} \,,
\end{aligned} \tag{24}$$

with the $\alpha_{C_{S,P,T}}$ given in Tab. 2.

3.1 Nucleons

We start with the simplest hadronic systems for which we can measure EDMs, the nucleons. Their EDMs can be written directly as a Lagrangian term

$$-\frac{i}{2} F^{\mu\nu} d_N \left(\bar{N}\sigma_{\mu\nu}\gamma_5 N\right) \,, \tag{25}$$

but can also be described by chiral perturbation theory, based on the hadronic-scale Lagrangian. In that case we consider the nucleon EDMs as observables that receive contributions from: short-range contributions $d_N^{\rm sr}$, NNLO pion-loop contributions, and potential direct contributions to Eq.(25) within and beyond the SM [73],

$$\begin{aligned}
d_n &= d_n^{\rm sr} - \frac{e g_A}{8\pi^2 F_\pi}\left[g_\pi^{(0)}\left(\ln\frac{m_\pi^2}{m_N^2} - \frac{\pi m_\pi}{2m_N}\right) - \frac{g_\pi^{(1)}}{4}(\kappa_0 - \kappa_1)\frac{m_\pi^2}{m_N^2}\ln\frac{m_\pi^2}{m_N^2}\right] \\
d_p &= d_p^{\rm sr} + \frac{e g_A}{8\pi^2 F_\pi}\left[g_\pi^{(0)}\left(\ln\frac{m_\pi^2}{m_N^2} - \frac{2\pi m_\pi}{m_N}\right) - \frac{g_\pi^{(1)}}{4}\left(\frac{2\pi m_\pi}{m_N} + \left(\frac{5}{2} + \kappa_0 + \kappa_1\right)\frac{m_\pi^2}{m_N^2}\ln\frac{m_\pi^2}{m_N^2}\right)\right] ,
\end{aligned} \tag{26}$$

where $F_\pi = 92$ MeV is the pion decay constant [74], $m_\pi = 139$ MeV, $m_N = 940$ MeV, $g_A \approx 1.27$ is the nucleon isovector axial charge, and the isoscalar and isovector nucleon anomalous magnetic moments are $\kappa_0 = -0.12$ and $\kappa_1 = 3.7$. We set the renormalization scale to the nucleon

System i	Measured d_i [e cm]	Upper limit on $\lvert d_i\rvert$ [e cm]	Reference
n	$(0.0 \pm 1.1_{\text{stat}} \pm 0.2_{\text{syst}}) \cdot 10^{-26}$	$2.2 \cdot 10^{-26}$	[59]
^{205}Tl	$(-4.0 \pm 4.3) \cdot 10^{-25}$	$1.1 \cdot 10^{-24}$	[60]
^{133}Cs	$(-1.8 \pm 6.7_{\text{stat}} \pm 1.8_{\text{syst}}) \cdot 10^{-24}$	$1.4 \cdot 10^{-23}$	[61]
HfF$^+$	$(-1.3 \pm 2.0_{\text{stat}} \pm 0.6_{\text{syst}}) \cdot 10^{-30}$	$4.8 \cdot 10^{-30}$	[62]
ThO	$(4.3 \pm 3.1_{\text{stat}} \pm 2.6_{\text{syst}}) \cdot 10^{-30}$	$1.1 \cdot 10^{-29}$	[63]
YbF	$(-2.4 \pm 5.7_{\text{stat}} \pm 1.5_{\text{syst}}) \cdot 10^{-28}$	$1.2 \cdot 10^{-27}$	[64]
^{199}Hg	$(2.20 \pm 2.75_{\text{stat}} \pm 1.48_{\text{syst}}) \cdot 10^{-30}$	$7.4 \cdot 10^{-30}$	[65, 66]
^{129}Xe	$(-1.76 \pm 1.82) \cdot 10^{-28}$	$4.8 \cdot 10^{-28}$	[67, 68]
^{171}Yb	$(-6.8 \pm 5.1_{\text{stat}} \pm 1.2_{\text{syst}}) \cdot 10^{-27}$	$1.5 \cdot 10^{-26}$	[69]
^{225}Ra	$(4 \pm 6_{\text{stat}} \pm 0.2_{\text{syst}}) \cdot 10^{-24}$	$1.4 \cdot 10^{-23}$	[70]
TlF	$(-1.7 \pm 2.9) \cdot 10^{-23}$	$6.5 \cdot 10^{-23}$	[71]
	Measured ω_i [mrad/s]	Rescaling factor x_i for d_i	Reference
HfF$^+$	$(-0.0459 \pm 0.0716_{\text{stat}} \pm 0.0217_{\text{syst}})^\dagger$	0.999	[62]
ThO	$(-0.510 \pm 0.373_{\text{stat}} \pm 0.310_{\text{syst}})$	0.982	[63]
YbF	$(5.30 \pm 12.60_{\text{stat}} \pm 3.30_{\text{syst}})$	1.12	[64]

Table 1: Measured EDM values and 95%CL ranges. For ^{129}Xe we combine two independent results with similar precision, using inverse-variance weighting. For the paramagnetic molecules, we also provide the measured angular frequencies and the rescaling factor which allows us to use $x_i d_i$ for each experimentally reported d_i. For the definition of x_i, see text. †The frequency for HfF$^+$ is scaled by a factor of 2 relative to Ref. [62], to consistently use Eq.(29) for all systems.

mass, and the split between proton and neutron mass leads to a higher-order effect that is presently negligible in relation to other uncertainties. In terms of weak-scale parameters, finite values of $g_\pi^{(0,1)}$ can be related to a CKM phase or $\bar{\theta}$, but also to hadronic 4-fermion operators. The input from the presently most sensitive neutron EDM measurement to our global analysis is given in Tab. 1.

3.2 Paramagnetic systems

Paramagnetic atoms and molecules are primarily sensitive to the electron EDM and the scalar electron-nucleon couplings $C_S^{(0,1)}$ of Eq.(10). It is common to distinguish the case of paramagnetic atoms with the atomic EDMs

$$d_i = \alpha_{i,d_e} d_e + \alpha_{i,C_S} C_S + \sum_{c_j} \alpha_{i,c_j} c_j \qquad \text{for} \qquad i \in \{\text{Tl}, \text{Cs}\} \tag{27}$$

from that of paramagnetic molecules, although in both cases the terms involving the Lagrangian parameters d_e and C_S dominate. The experimentally relevant quantity is a phase difference accumulated over some measurement time, interpreted as a frequency. Measurements with paramagnetic molecules are often reported as a parity- and time-reversal-violating frequency shift ω_i,

$$\omega_i = \eta_{i,d_e}^{(m)} d_e + k_{i,C_S}^{(m)} C_S + \sum_{c_j} \alpha_{i,c_j}^{(m)} c_j \qquad \text{for} \qquad i \in \{\text{HfF}^+, \text{ThO}, \text{YbF}\}\,. \tag{28}$$

The superscript (m) indicates the molecular systems. We also give these measured angular frequencies in Tab. 1. This emphasizes that ω_i does not scale in a simple way with the experimentally applied electric field; it rather depends on the molecular structure via a so-called

System i	$\langle\sigma_n\rangle$	$\langle\sigma_p\rangle$	$\langle\sigma_z\rangle^{(0)}$	α_{i,C_S} [e cm]	α_{i,C_P} [e cm]	α_{i,C_T} [e cm]
Tl	0.274	0.726	1	$-6.77\cdot10^{-18}$	$1.5\cdot10^{-23}$	$5\cdot10^{-21}$
Cs	−0.206	−0.572	−0.778	$7.8\cdot10^{-19}$	$2.2\cdot10^{-23}$	$9.2\cdot10^{-21}$
^{199}Hg	−0.302	−0.032	−0.334	$-2.8\cdot10^{-22}$	$6\cdot10^{-23}$	$1.7\cdot10^{-20}$
^{129}Xe	0.73	0.27	1	$-6.28\cdot10^{-23}$	$1.6\cdot10^{-23}$	$5.7\cdot10^{-21}$
^{171}Yb	−0.3	−0.034	−0.334	$-2.68\cdot10^{-22}$	$4.01\cdot10^{-23}$	$1.24\cdot10^{-20}$
^{225}Ra	0.72	0.28	1	$2.9\cdot10^{-21}$	$-6.4\cdot10^{-22}$	$-1.8\cdot10^{-19}$
TlF	0.274	0.726	1	$2.9\cdot10^{-18}$	$3\cdot10^{-19}$	$2.7\cdot10^{-16}$

Table 2: Effective parameters used as input to our global analysis, as summarized in Tab. 3. Within a shell model of the nucleus, the quantity $\langle\sigma_z\rangle^{(0)} = \langle\sigma_n\rangle + \langle\sigma_p\rangle$ is the isoscalar sum of neutron and proton spin projections. (Note that the shell model is not expected to be reliable for the deformed nuclei ^{171}Yb and ^{225}Ra.) The spin fractions contributing to semileptonic coefficients in TlF only take into account the ^{205}Tl nucleus, though see Ref. [72] for some consideration of contributions from the ^{19}F nucleus.

effective electric field $E_{\rm eff}$ that saturates when the molecule is polarized. Note that α_{i,d_e} is dimensionless, while $\eta^{(m)}_{i,d_e}$ is not; similarly α_{i,C_S} has the units of an EDM, while $k^{(m)}_{i,C_S}$ has units of angular frequency.

The remaining contributions to paramagnetic molecule EDMs have weak dependences on other low-energy constants, and we note that in particular the EDMs of the μ and τ leptons have been indirectly constrained via the experimental limits from ThO and Hg [91]. However, we are not aware of any established values for the coefficients of semileptonic or hadronic parameters. We thus take $\alpha^{(m)}_{i,c_j} = 0$ in Eq.(28), while for the paramagnetic atoms we obtain values for semileptonic coefficients either from the literature or by scaling arguments, as reported in Tabs. 2 and 3.

Now starting with the truncated version of Eq.(28), we adopt the convention that

$$\begin{aligned}\omega_i &= \eta^{(m)}_{i,d_e} d_e + k^{(m)}_{i,C_S} C_S \\ &= -\frac{E_{{\rm eff},i}}{\hbar} d_e + \frac{W_i}{\hbar} C_S \,,\end{aligned} \tag{29}$$

where all signs, g-factors, spin magnitudes, etc. are absorbed into the coefficients on the right-hand side. It is also common to refer to $\mathcal{E}_{\rm eff}$ as the effective electric field, where

$$E_{\rm eff} = \mathcal{E}_{\rm eff}\,{\rm sgn}(\vec{J}\cdot\hat{n})\,\langle\hat{n}\cdot\hat{z}\rangle\,, \tag{30}$$

and $\vec{J}$ is the total electronic angular momentum, $\hat{n}$ is the direction of the internuclear axis, and $\hat{z}$ is the direction of the externally applied electric field. Here $\Omega = \vec{J}\cdot\hat{n}$ is a quantum number used in molecular term symbols, and $\langle\hat{n}\cdot\hat{z}\rangle$ indicates the degree to which the molecule is electrically polarized by the applied electric field. Some of the conventions used for the orientation of the internuclear axis, the internal electric field, etc. are summarized in Appendix A.3 of Ref. [92].

The truncated expression for the frequency shift thus provides us with physical expressions for the two constants,

$$\eta^{(m)}_{i,d_e} = -\frac{E_{{\rm eff},i}}{\hbar} \qquad \text{and} \qquad k^{(m)}_{i,C_S} = \frac{W_i}{\hbar}\,. \tag{31}$$

System i	α_{i,d_e}	$\alpha_{i,C_S^{(0)}}$ [e cm]	$\alpha_{i,C_P^{(0)}}$ [e cm]	$\alpha_{i,C_T^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(0)}}$ [e cm]	$\alpha_{i,g_\pi^{(1)}}$ [e cm]	$\alpha_{i,d_n^{\rm sr}}$	$\alpha_{i,d_p^{\rm sr}}$
n	–	–	–	–	$1.38^{\pm 0.02}\cdot 10^{-14}$	$2.73^{\pm 0.02}\cdot 10^{-16}$	1	−1
^{205}Tl	$-558^{\pm 28}$ [75]	$-6.77^{\pm 0.34}\cdot 10^{-18}$	$1.5^{+2}_{-0.7}\cdot 10^{-19}$	$8.8^{\pm 0.9}\cdot 10^{-21}$	n/a	n/a	n/a	n/a
^{133}Cs	$123^{\pm 4}$	$7.80^{+0.2}_{-0.8}\cdot 10^{-19}$	$-1.4^{+0.8}_{-2}\cdot 10^{-20}$	$1.7^{\pm 0.2}\cdot 10^{-20}$	–	–	–	–
^{199}Hg	$-0.012^{+0.0094}_{-0.002}$ [33,76]	$-1.26^{+0.7}_{-1.2}\cdot 10^{-21}$	$6.6^{+1.2}_{-0.3}\cdot 10^{-23}$	$-6.4^{+3}_{-4}\cdot 10^{-21}$	$-1.18^{+0.19}_{-2.62}\cdot 10^{-17}$	$1.6^{+0}_{-6.5}\cdot 10^{-17}$	$-1.56^{\pm 0.39}\cdot 10^{-4}$	$-1.56^{\pm 0.39}\cdot 10^{-5}$
^{129}Xe	$-8^{+0}_{-8}\cdot 10^{-4}$ [33,77]	$-2.1^{+1.2}_{-2.5}\cdot 10^{-22}$	$1.7^{+0.5}_{-0.4}\cdot 10^{-23}$	$1.24^{+0.78}_{-0.61}\cdot 10^{-21}$	$-0.4^{+1.2}_{-23}\cdot 10^{-19}$	$-2.2^{+1.1}_{-17}\cdot 10^{-19}$	$1.7^{+0.7}_{-0}\cdot 10^{-5}$	$3.51^{\pm 0.88}\cdot 10^{-6}$
^{171}Yb	$(-0.012^{+0.01145}_{-0.002})$ [78]	$-9.1^{+5}_{-11}\cdot 10^{-22}$	$4.5^{+1.8}_{-1.1}\cdot 10^{-23}$	$-4.4^{+2.2}_{-2.9}\cdot 10^{-21}$	$-9.5^{\pm 2.4}\cdot 10^{-18}$	$1.3^{\pm 0.33}\cdot 10^{-17}$	$-1.13^{\pm 0.28}\cdot 10^{-4}$	$-1.13^{\pm 0.28}\cdot 10^{-5}$
^{225}Ra	$-0.054^{\pm 0.002}$ [33]	$8.6^{+9.5}_{-4.5}\cdot 10^{-21}$	$-7.0^{+1.7}_{-2.6}\cdot 10^{-22}$	$-4.5^{+2.0}_{-2.5}\cdot 10^{-20}$	$1.7^{+5.2}_{-0.8}\cdot 10^{-15}$	$-6.9^{+3.1}_{-21}\cdot 10^{-15}$	$-5.36^{\pm 1.34}\cdot 10^{-4}$	$-1.11^{\pm 0.28}\cdot 10^{-4}$
TlF	$81^{\pm 20}$ [51,71]	$5.6^{+4.9}_{-2.5}\cdot 10^{-18}$	$2.4^{+1.0}_{-1.9}\cdot 10^{-19}$	$4.8^{+1.2}_{-1.1}\cdot 10^{-16}$	$1.9^{+0.1}_{-1.4}\cdot 10^{-14}$	$-1.6^{\pm 0.4}\cdot 10^{-13}$	$-9.47^{\pm 2.37}\cdot 10^{-2}$	$-4.59^{\pm 1.15}\cdot 10^{-1}$
HfF^+	1	$9.17^{\pm 0.06}\cdot 10^{-21}$	–	–	–	–	–	–
ThO	1	$1.51^{+0}_{-0.2}\cdot 10^{-20}$	–	–	–	–	–	–
YbF	1	$8.99^{\pm 0.70}\cdot 10^{-21}$	–	–	–	–	–	–
	$\eta_{i,d_e}^{(m)}\left[\frac{\rm mrad}{\rm s\ e\ cm}\right]$	$k_{i,C_S}^{(m)}\left[\frac{\rm mrad}{\rm s}\right]$	α_{i,C_P}	α_{i,C_T}	$\alpha_{i,g_\pi^{(0)}}$	$\alpha_{i,g_\pi^{(1)}}$	$\alpha_{i,d_n^{\rm sr}}$	$\alpha_{i,d_p^{\rm sr}}$
HfF^+	$3.49^{\pm 0.14}\cdot 10^{28}$ [76,79–82]	$3.2^{+0.1}_{-0.2}\cdot 10^{8}$ [76,79,80]	–	–	–	–	–	–
ThO	$-1.21^{+0.05}_{-0.39}\cdot 10^{29}$ [76,83–85]†	$-1.82^{+0.42}_{-0.27}\cdot 10^{9}$ [76,83,85–87]†	–	–	–	–	–	–
YbF	$-1.96^{\pm 0.15}\cdot 10^{28}$ [76,86–89]	$-1.76^{\pm 0.2}\cdot 10^{8}$ [76,86–88]	–	–	–	–	–	–

Table 3: Central values for the α-parameters defined in Eq.(23). Here n/a means that we do not know of a reliable prediction, and – means that we neglect the dependence in our global analysis. †There appears to be an overall sign issue in the coefficients reported for ThO in Table 4 of Ref. [76]. For the coefficient of d_e in ^{171}Yb, we take the value from the structurally similar system ^{199}Hg, since no complete dedicated calculation is presently available (the result of [78] is taken into account via the upper error bound). The impact of this choice on the global analysis is negligible. The values for Tl and Cs of $\alpha_{i,C_T^{(0)}}$ are estimated by simple analytical calculations [90], and the uncertainties quoted here are estimated as approximately twice those arising from the relevant hadronic matrix elements.

The point is that this frequency measurement cannot be directly translated into a permanent molecular dipole moment, since the molecules are substantially or entirely electrically polarized during the measurement. The linear energy shift due to the lab-frame electric field saturates as $\langle \hat{n} \cdot \hat{z} \rangle \to 1$, and the limiting dependence on the external fields is that of an induced dipole moment. To express the frequency difference ω_i in terms of a physically meaningful permanent electric dipole moment, we define the molecule's d_i in relation to the effective electric field,

$$-E_{\mathrm{eff},i} d_i \equiv -E_{\mathrm{eff},i} d_e + W_i C_S \,. \tag{32}$$

We can use this form to define, in analogy to Eq.(27),

$$d_i = d_e + \alpha_{i,C_S} C_S \qquad \Leftrightarrow \qquad \alpha_{i,d_e} = 1 \quad \text{and} \quad \alpha_{i,C_S} = -\frac{W_i}{E_{\mathrm{eff},i}} = \frac{k^{(m)}_{i,C_S}}{\eta^{(m)}_{i,d_e}} \,. \tag{33}$$

For our analysis we re-cast the limits from paramagnetic molecules in units of $e$cm, such that the coefficients for all systems can be expressed in the same units, with the side effect that $\alpha_{i,d_e} = 1$ for all paramagnetic molecules. This approach also serves two additional purposes: (i) Many different conventions are in use for these coefficients, sometimes using the same symbols for different quantities. Since measured quantities are related in publications to the electron EDM, our choice makes comparisons across different works relatively straightforward. Of course, all quantities must be ultimately connected to an experimentally measured phase; (ii) As first pointed out in Refs. [86,87], while there is considerable variation in the literature values for $\eta^{(m)}_{i,d_e}$ and $k^{(m)}_{i,C_S}$ in a given system, there is much less variation in their ratio. Dividing Eq.(28) by $\eta^{(m)}_{i,d_e}$ should pass this uncertainty on to all semileptonic and hadronic coefficients.

Note that $\eta^{(m)}_{i,d_e}$ is used both to obtain sole-source limits on d_e from the experimentally measured ω_i, and to convert the experimentally measured frequencies into EDM units for Tab. 1. The experimental EDM limits for paramagnetic molecules in Tab. 1 are rescaled by the indicated factor x_i, which differs from unity when updated values for $\eta^{(m)}_{i,d_e}$ differ from the cited publication. This is typically the case when improved molecular structure calculations have become available since the experimental limit was published.

Recommended values for many of the α_{i,c_j} are given in Tables III-V of Ref. [15] and in Table 4 of Ref. [76], including in many cases the ranges corresponding to theory uncertainties (or at least different reported values). We give our choices for α_{i,c_j} in Tab. 3.

3.3 Diamagnetic systems

In contrast to the paramagnetic systems, where sub-leading contributions are largely neglected due to lack of theory inputs, the small contributions of d_e and $C_S^{(0)}$ are taken into account for all diamagnetic systems. These are typically smaller contributions to the observable EDM in cases where all electron spins are paired, with the main contributions coming rather from nucleon EDMs, nuclear forces mediated by pion exchange with strengths $g_\pi^{(0,1,2)}$, and the nuclear-spin-dependent semileptonic interactions $C_T^{(0)}$ (and possibly $C_P^{(0)}$). The experimental precision, especially of Hg, is nevertheless high enough to contribute meaningful constraining power for d_e and $C_S^{(0)}$.

Unfortunately it appears that arguments [76] to the effect that diamagnetic systems add significant complementary constraining power in the $d_e - C_S$ subspace due to a different sign of the ratio $\alpha_{i,d_e}/\alpha_{i,C_S}$ in comparison to paramagnetic systems, do not in fact hold for Hg or

most other systems in Tab. 1. On the other hand, recent calculations [33] indicate that α_{i,d_e} and α_{i,C_S} in some highly correlated systems may indeed have different sign, thus reviving this possibility. At present, and as can be seen from Tab. 3, Ra is the only measured system where this appears to be the case. It may thus represent a special case among diamagnetic systems, in that reducing the error of experimental and theory inputs could have significant impact on the $d_e - C_S^{(0)}$ subspace that is complementary to the dominating paramagnetic constraints.

The contributions of nuclear forces and nucleon EDMs are frequently interpreted via the Schiff moment of a given nucleus [15], which is more easily related to nuclear structure parameters [90]. In terms of our model coefficients, the Schiff moment S_i of system i and a system-specific coefficient $k_{i,S}$ can be expressed related to the corresponding EDM via

$$\begin{aligned} k_{i,S} S_i &= \sum_{c_j \in \{d_{n,p}^{\mathrm{sr}}, g_\pi^{(0,1,2)}\}} \alpha_{i,c_j} c_j \\ &= k_{i,S}\left[s_{i,n} d_n + s_{i,p} d_p + \frac{m_N g_A}{F_\pi}\left(a_{i,0} g_\pi^{(0)} + a_{i,1} g_\pi^{(1)} + a_{i,2} g_\pi^{(2)} \right) \right] \\ &\approx k_{i,S}\left[s_{i,n} d_n^{\mathrm{sr}} + s_{i,p} d_p^{\mathrm{sr}} + \frac{m_N g_A}{F_\pi}\left(\tilde{a}_{i,0} g_\pi^{(0)} + \tilde{a}_{i,1} g_\pi^{(1)} \right) \right], \end{aligned} \tag{34}$$

where the coefficients $s_{i,N}$ $(N = n, p)$ indicate the contributions from EDMs of unpaired nucleons and the coefficients $a_{i,m}$ $(m = 0, 1, 2)$ parameterize the strength of CP-violating pion exchange, organized by isospin, for the nucleus of system i. Between the second and third lines we drop the $g_\pi^{(2)}$ term as discussed above, and use Eq.(26) to rewrite the nucleon EDMs themselves and absorb their contributions from $g_\pi^{(0)}$ and $g_\pi^{(1)}$ into the implicitly re-defined coefficients $\tilde{a}_{i,m}$.

The coefficient $k_{i,S}$ is calculated for many systems of interest and can, in principle, be large in heavy and especially deformed nuclei such as ^{225}Ra. This corresponds to a nuclear-structure induced enhancement of the observable EDM, which applies to both the nucleon EDMs and the pion-exchange forces. This implies that among the potentially leading contributions from the six hadronic-scale model parameters $C_{P,T}^{(0)}$, $d_{n,p}^{\mathrm{sr}}$, and $g_\pi^{(0,1)}$, very different weights can be expected according to the electronic and nuclear structure of the various diamagnetic systems. Inspecting Tab. 3 reveals that to a limited extent, this is indeed already the case. However, this complementarity is not yet fully exploited.

Despite the pion pole enhancement, $C_P^{(0)}$ appears suppressed relative to $C_T^{(0)}$ in all measured systems. The coefficients of $g_\pi^{(0,1)}$ are typically of comparable size for a given system, although possibly different in sign. Unfortunately the coefficients of $d_{n,p}$ are not well known for most nuclei, although in principle these can be calculated. For the pion-exchange forces, even in nuclei that have been the object of many studies, the theory uncertainties are large. This is especially true for soft nuclei such as our Hg and Xe, in which non-static deformations can present special challenges.

For nuclei with spin $I > 1/2$, a CP-violating nuclear magnetic quadrupole moment can in principle exist in analogy to a Schiff-moment-induced EDM and be analyzed within a common global analysis. At present among the experimental systems which have already been measured, Cs is the one where this effect would be expected to be most relevant.

As part of a global analysis, a large number of experimental measurements from complementary diamagnetic systems can disentangle contributions from model parameters that are relevant at some level for all of them. In this sense the role of d_e and $C_S^{(0)}$ takes on a new importance, not only to further constrain these parameters themselves, but also as an additional contribution to the EDM that brings along uncertainties which dilute the constraining

power for other model parameters. Diamagnetic molecular systems such as TlF, or molecular systems containing Schiff-enhanced nuclides, introduce complementary constraining power to our global analysis.

We finally note that diamagnetic systems unite theory inputs from different communities. Different conventions for assigning a negative isospin projection in the nucleon doublet have to be carefully noted when combining calculated coefficients from different sources, especially when we include semileptonic interactions together with nuclear forces. Reference [93] makes an effort to disambiguate a part of this issue for $\alpha_{g_\pi^{(0,1,2)},^{129}Xe}$.

It is only possible to establish meaningful constraints in a context where the notation and conventions are clear, and where the uncertainties associated with theory inputs have been clearly assessed.

3.4 Single parameter ranges

In this section we provide toy ranges of individual model parameters from the set

$$\left\{ d_e, C_S^{(0)}, C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr} \right\}, \tag{35}$$

using the set of EDM measurements one by one. To extract these single parameter ranges, we utilize Eq.(27) and include all EDM measurements listed in Tab. 1 and the model dependence from Tab. 3. For the likelihood we assume a Gaussian form, which means that all limits are quoted as symmetric one-sigma error bars around the central, best-fit value. While these single-parameter limits allow us to compare the reach of different measurements for CP violation as a whole, they do not give the allowed ranges of the individual model parameters. The reason is that contributions from different model parameters to the same measurement can cancel. Because an EFT builds on the assumption that many higher-dimensional operators are induced by a given physics model, the single-parameter limits are overly optimistic.

Using the single-parameter constraints in Tab. 4 we can compare the impact of different EDM measurements on a given model parameter, with the caveat that multi-dimensional correlations might change that picture. Starting with the electron EDM d_e, the paramagnetic molecules HfF^+ and ThO provide the strongest constraints, while the paramagnetic YbF molecule has a similar constraining power as the diamagnetic Hg. The same two paramagnetic molecules lead the constraining power for the scalar electron-nucleon coupling $C_S^{(0)}$, again followed by Hg, YbF, and Tl.

The pseudoscalar and tensor electron-nucleon couplings $C_{P,T}^{(0)}$ can be probed by the paramagnetic atoms and the diamagnetic systems. This is done at present by far most efficiently with Hg and then, with much reduced constraining power, by Tl, Xe, and TlF. It is difficult to estimate the impact of these measurements on the combination of $C_P^{(0)}$ and $C_T^{(0)}$, and we will see in the next section how this more complex dependence affects the full global analysis.

The three hadronic parameters, $g_\pi^{(0,1)}$ and $d_n^{\rm sr}$, are most strongly constrained by the neutron EDM measurement and again Hg, suggesting that there will be significant correlations between the leptonic and hadronic model parameters in the global analysis. The other diamagnetic systems lead to much weaker limits, but will still be needed to constrain the 3-dimensional hadronic model space. Note that we are fixing the relation between the short-range neutron and proton parameters in the Lagrangian, Eq.(21), but from Tab. 3 we know that different measurements probe different admixtures of these two parameters.

System i	$d_e\,[e\,\mathrm{cm}]$	$C_S^{(0)}$	$C_P^{(0)}$	$C_T^{(0)}$
Tl	$(7.2\pm7.7)\cdot10^{-28}$	$(5.9\pm6.4)\cdot10^{-8}$	$(-2.7\pm3.0)\cdot10^{-6}$	$(-4.5\pm4.9)\cdot10^{-5}$
Cs	$(-1.4\pm5.6)\cdot10^{-26}$	$(-2.3\pm8.9)\cdot10^{-6}$	$(1.3\pm5)\cdot10^{-4}$	$(-1.1\pm4.2)\cdot10^{-4}$
^{199}Hg	$(-1.8\pm2.6)\cdot10^{-28}$	$(-1.7\pm2.5)\cdot10^{-9}$	$(3.4\pm4.8)\cdot10^{-8}$	$(-3.4\pm4.9)\cdot10^{-10}$
^{129}Xe	$(2.2\pm2.3)\cdot10^{-25}$	$(8.3\pm8.7)\cdot10^{-7}$	$(-1.0\pm1.1)\cdot10^{-5}$	$(-1.4\pm1.5)\cdot10^{-7}$
^{171}Yb	$(5.7\pm4.4)\cdot10^{-25}$	$(7.5\pm5.7)\cdot10^{-6}$	$(-1.5\pm1.2)\cdot10^{-4}$	$(1.6\pm1.2)\cdot10^{-6}$
^{225}Ra	$(-7.4\pm1.1)\cdot10^{-23}$	$(4.7\pm7)\cdot10^{-4}$	$(-5.7\pm8.5)\cdot10^{-3}$	$(-8.9\pm13)\cdot10^{-5}$
TlF	$(-2.1\pm3.6)\cdot10^{-25}$	$(-3.0\pm5.1)\cdot10^{-6}$	$(-7.1\pm12)\cdot10^{-5}$	$(-3.6\pm6.1)\cdot10^{-8}$
HfF^+	$(-1.3\pm2.1)\cdot10^{-30}$	$(-1.4\pm2.3)\cdot10^{-10}$		
ThO	$(4.3\pm4.1)\cdot10^{-30}$	$(2.9\pm2.7)\cdot10^{-10}$		
YbF	$(-2.4\pm5.9)\cdot10^{-28}$	$(-2.7\pm6.6)\cdot10^{-8}$		
	$g_\pi^{(0)}$	$g_\pi^{(1)}$	d_n^{sr}	d_p^{sr}
n	$(0\pm8.1)\cdot10^{-13}$	$(0\pm4.1)\cdot10^{-11}$	$(0\pm1.1)\cdot10^{-26}$	$(0\pm1.1)\cdot10^{-26}$
^{199}Hg	$(-1.9\pm2.7)\cdot10^{-13}$	$(1.4\pm2.0)\cdot10^{-13}$	$(-1.4\pm2.0)\cdot10^{-26}$	$(-1.4\pm2.0)\cdot10^{-25}$
^{129}Xe	$(4.4\pm4.6)\cdot10^{-9}$	$(8\pm8.3)\cdot10^{-10}$	$(-1.0\pm1.1)\cdot10^{-23}$	$(-5.0\pm5.2)\cdot10^{-23}$
^{171}Yb	$(7.2\pm5.5)\cdot10^{-10}$	$(-5.2\pm4.0)\cdot10^{-10}$	$(6.0\pm4.6)\cdot10^{-23}$	$(6.0\pm4.6)\cdot10^{-22}$
^{225}Ra	$(2.4\pm3.5)\cdot10^{-9}$	$(-5.8\pm8.7)\cdot10^{-10}$	$(-7.5\pm11)\cdot10^{-21}$	$(-3.6\pm5.4)\cdot10^{-20}$
TlF	$(-9.0\pm15)\cdot10^{-10}$	$(1.1\pm1.8)\cdot10^{-10}$	$(1.8\pm3.1)\cdot10^{-22}$	$(3.7\pm6.3)\cdot10^{-23}$

Table 4: Single-parameter ranges allowed by each of the EDM measurements given in Tab. 1.

4 Global analysis

To combine the available EDM measurements and analyze them in terms of the hadronic-scale Lagrangian and its parameters, Eq.(22), we use the established SFitter analysis tool. It constructs a global likelihood with a comprehensive uncertainty treatment and analyses it in terms of high-dimensional correlations. Lower-dimensional and one-dimensional likelihoods for the individual model parameters can be derived by profiling or marginalization, depending on the preferred statistical framework. If we assume that experimental uncertainties are Gaussian, profiling and marginalization have to lead to identical results. For the theory uncertainties, discussed in Sec. 4.5, the difference between the two approaches makes a formal, but not significant difference.

4.1 SFitter framework

SFitter [34–36] has been developed for global analyses of LHC measurements in the context of BSM physics [94] and Higgs and top properties [38,95–97], including comprehensive studies of the connection between EFTs and UV-completions of the SM [98,99]. It has a focus on its uncertainty treatment, including theory uncertainties [39], the connection between Bayesian and frequentist approaches [35,37], and published experimental likelihoods [38].

Our first SFitter analysis of EDMs relates the 11 measurements from Tab. 1 to the seven model parameters in Eq.(22). In general, SFitter includes statistical, systematic, and theory uncertainties. At the heart of SFitter is the fully exclusive likelihood as a function of model and nuisance parameters. All measurements are described as uncorrelated, with the individual statistical uncertainties described by Poisson or Gaussian likelihood. Statistical uncertainties are, usually, uncorrelated as well and described by a Poisson distribution, turning into a Gaussian for high statistics. Experimental systematics are assumed to have a Gaussian shape, but

can be described by any nuisance parameter. This Gaussian shape is justified for parameters which are measured elsewhere. We use it for all experimental uncertainties, assuming that the measurements are at least as much dominated by statistical uncertainty, with error distribution appropriate for count-rate limited frequency measurements.

For the EDM analysis the situation is then relatively simple. First, from Eq.(23) we know that all observables depend on the model parameters linearly. Second, we can combine the statistical and systematic experimental uncertainties into the symmetric Gaussian error bars given in Tab. 1. Finally, we do not have to consider nuisance parameters, if we assume that the likelihood has a Gaussian form for each independent measurement. This Gaussian assumption also implies that for uncorrelated uncertainties a profile likelihood and a Bayesian marginalization will give the same result.

Theory uncertainties have no well-defined likelihood shape, and no maximum, but they can be thought of as a range [39]. A flat theory uncertainty is not parametrization-invariant, as one would have expected from a fixed range, but without a preferred central value we consider it conservative. In SFitter it is implemented through an allowed shift of the prediction at no cost in the likelihoods. For the EDM analysis, theory uncertainties affect the α-values with their central values given in Tab. 3.

To construct the exclusive likelihood, SFitter evaluates EDM predictions over the entire model parameter space. It uses a Markov chain to encode the likelihood in the distribution of points covering the model space. A helpful aspect, common to many BSM analyses, is that we can safely assume the global minimum of the likelihood to be at the SM parameter point. To remove nuisance parameters or to extract limits on a reduced number of model parameters, SFitter can employ a profile likelihood or a Bayesian marginalization [35, 37]. These two methods give different results, with the exception of uncorrelated Gaussians. Profiling over flat theory uncertainties and Gaussian experimental uncertainties leads to the RFit [100] prescription, profiling over two parameters with flat likelihood leads to linearly added uncertainties even for uncorrelated parameters.

While Eq.(23) suggest a homogeneous set of model parameters, the typical sizes of the model parameters in Eq.(22) and the α-values in Tab. 3 can be extremely different. For numerical reasons, we internally re-scale each model parameter and each α-value such that all model parameters are evaluated with similar size. Concretely, this means rescaling d_e by a factor 10^{29}, $C_S^{(0)}$ by 10^9, $g_\pi^{(1)}$ and $g_\pi^{(0)}$ by 10^{10}, $C_T^{(0)}$ by 10^8, $C_P^{(0)}$ by 10^6, and $d_{n/p}^{\text{sr}}$ by 10^{23}. These rescalings are also reflected in the way we present our results.

4.2 Well-constrained model sub-space

As a starting point of the global analysis and to understand the main features, we consider a sub-space of well-constrained parameters. Following our discussion in Sec. 3.4 we expect the semileptonic parameters d_e and $C_S^{(0)}$ to be constrained well by the paramagnetic molecules HfF^+ and ThO. Similarly, the hadronic parameters $g_\pi^{(0)}$ and d_n^{sr} are strongly constrained by the neutron and Hg EDMs. This means the model sub-space

$$\left\{ d_e, C_S, g_\pi^{(0)}, d_n^{\text{sr}} \right\} \tag{36}$$

should be constrained well by the full set of measurements given in Tab. 1.

The interesting question is how the constraints on these four model parameters are correlated. In Fig. 1 we show these correlations extracted as 2-dimensional profile likelihoods from the fully exclusive, 4-dimensional likelihood. Three structural aspects stick out: (i) a strong

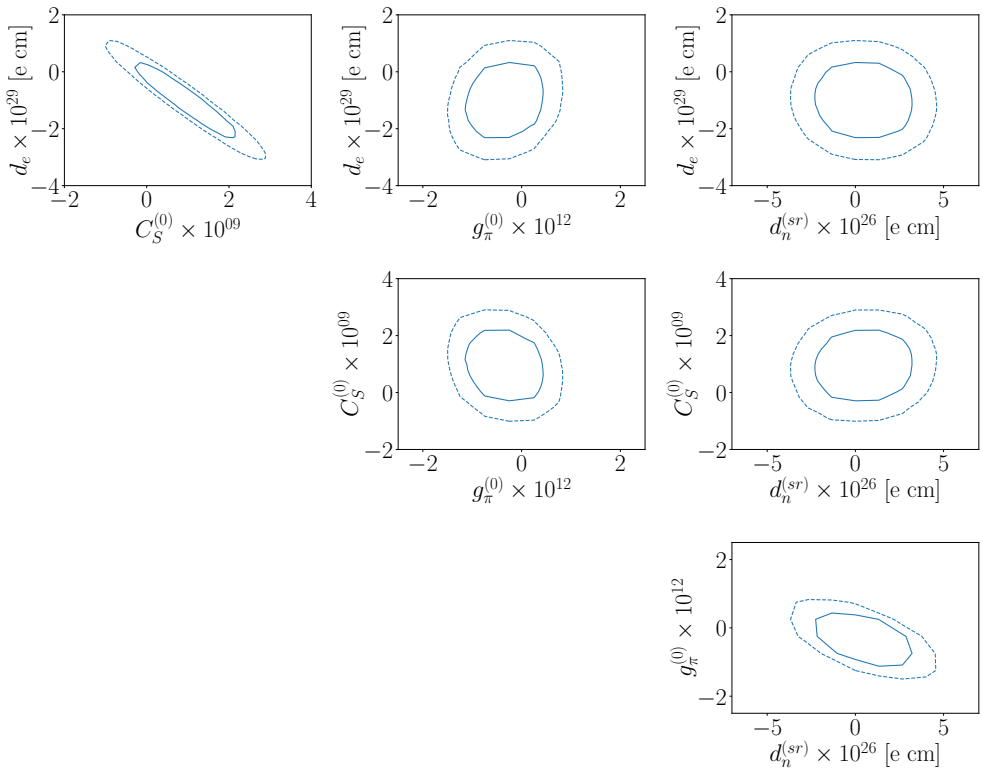


Figure 1: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n^{\rm sr}\}$, based on all EDM measurements but neglecting theory uncertainties. The ellipses indicate 68% and 95% CL.

anti-correlation between d_e and $C_S^{(0)}$; (ii) a strong anti-correlation between $g_\pi^{(0)}$ and $d_n^{\rm sr}$; and (iii) essentially no correlations between the $\{d_e, C_S^{(0)}\}$ and $\{g_\pi^{(0)}, d_n^{\rm sr}\}$ parameter subsets.

Within the leptonic sector, the strong correlation between d_e and $C_S^{(0)}$ and its independence from the remaining parameter space is expected to remain for the full global analysis. The reason is that it is induced by the strongest measurements of HfF^+ and ThO, and according to our parameterization as shown in Tabs. 3 and 4 those two measurements are not affected by any other model parameter. This means the upper-left panel of Fig. 1 factorizes from our global EDM analysis, and we can consider the remaining model parameters separately and without the HfF^+ and ThO measurements.

For the hadronic parameters the situation is different. Again the neutron and Hg measurements are three orders more constraining than the other measurements. However, as the two model parameters we could as well have chosen $g_\pi^{(0)}$ vs $g_\pi^{(1)}$, without any change in the conclusion. This means that we have to expand the hadronic parameter space next, to see what patterns emerge. The parameters d_e and $C_S^{(0)}$ will be kept factorized for the rest of our global analysis.

4.3 Hadronic parameters from diamagnetic systems

For the purely hadronic sector, we define a second simplified model parameter space,

$$\left\{ g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr} \right\} . \tag{37}$$

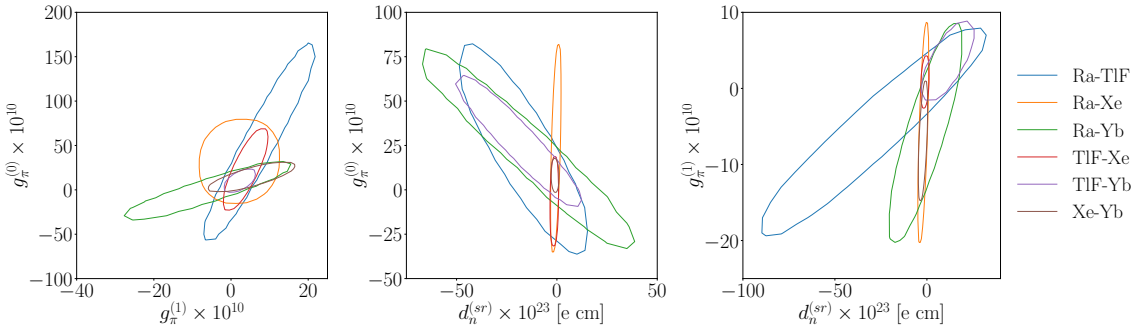


Figure 2: Correlations from three 2-dimensional analyses in the $\{g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$ parameter space, each based on a different pair of diamagnetic EDM measurements, as indicated by the color. The ellipses indicate 68% CL, neglecting theory uncertainties.

All three parameters are constrained by the neutron and diamagnetic EDMs, while we know from the above discussion that the constraints from diamagnetic systems on d_e and $C_S^{(0)}$ are weaker than those from their paramagnetic counterparts.

From Tab. 4 we see that the neutron and and Hg measurements strongly constrain two of the three hadronic model parameters in Eq. (37). To understand the correlations structure of the remaining diamagnetic measurements we show the correlated constraints of the six possible pairs of diamagnetic measurements on the different 2-dimensional sub-spaces of Eq. (37).

Starting with the left panel of Fig. 2, the different pairs of measurements constrain the $g_\pi^{(0)}$ vs. $g_\pi^{(1)}$ plane with different correlation patterns, implying that a global analysis will constrain the 3-dimensional hadronic subspace much better than any single pair of measurements. In the center panel, the situation changes when we look at the correlation with the neutron EDM-parameter $d_n^{\rm sr}$. Three combinations are aligned to similar negative correlations between $d_n^{\rm sr}$ and $g_\pi^{(0)}$. In the right panel, the situation is similar for $g_\pi^{(1)}$, with a positive correlation and less striking. In both cases the exception are the combinations of Xe with Ra, TlF, and Yb, which constrain $d_n^{\rm sr}$ extremely well and without any correlation with $g_\pi^{(0,1)}$. While this sounds like an advantage, we remind ourselves that from Tab. 4 and Fig. 1 we know that those limits on $d_n^{\rm sr}$ are three orders of magnitude weaker than what can be expected from including the neutron and and Hg measurements.

Altogether, Fig. 2 confirms that the constraints of the sub-leading four diamagnetic measurements on the 3-dimensional hadronic parameter space of Eq. (37) are correlated in a non-trivial manner. Evaluating these correlations requires a global analysis of the, formally, over-constraining set of diamagnetic measurements.

4.4 Poorly constrained model parameters

Finally, we combine the effects of all remaining parameters,

$$\left\{ C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr} \right\}, \tag{38}$$

ignoring the ThO and HfF^+ measurements, which constrain the factorized $d_e - C_S^{(0)}$ subspace, and also ignoring the neutron and Hg measurements. The latter constrain the above parameters, but because they are much stronger than all other measurements, they will induce narrow correlations in the allowed 5-dimensional parameter space.

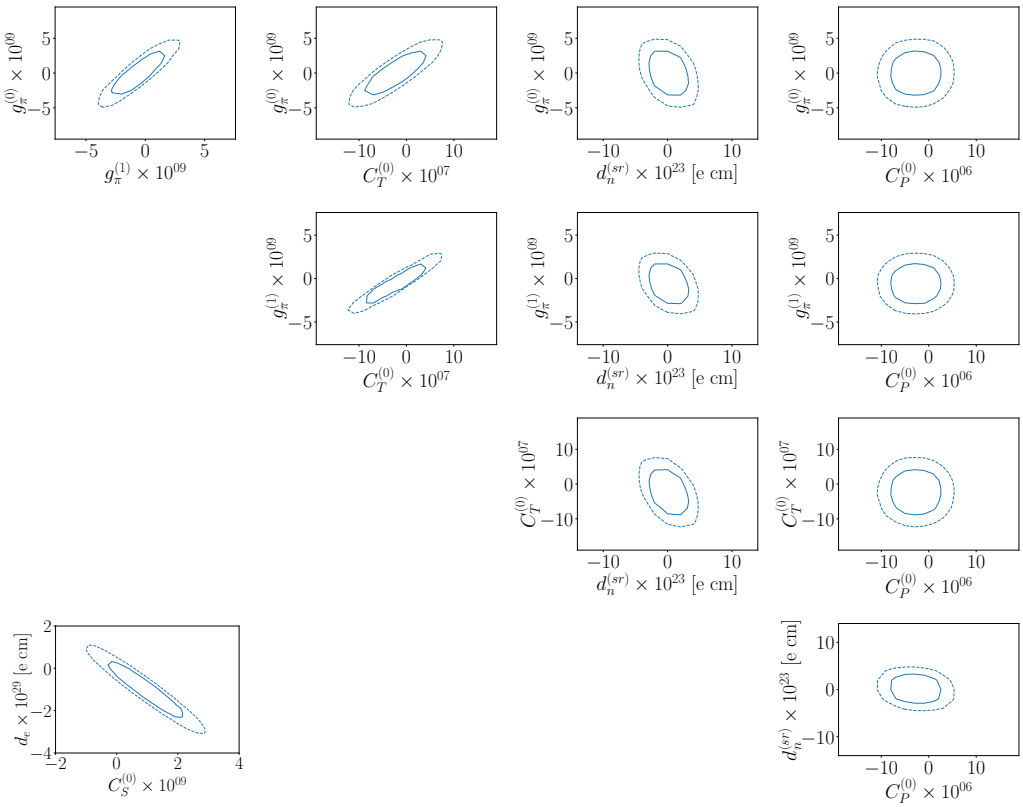


Figure 3: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$ and the factorized $d_e - C_S^{(0)}$ plane from Fig. 1. We ignore the neutron and Hg measurements, which induce narrow correlation patterns in the 5-dimensional parameters space and do not affect the profiled 2-dimensional correlations. The ellipses indicate 68% and 95% CL, neglecting theory uncertainties.

Narrow correlations in a higher-dimensional parameter space vanish when we profile the likelihood onto 2-dimensional correlations or even single model parameters. As an example, consider a 2-dimensional parameter space constrained by one strong and one weak measurement. The strong measurement leads to a narrow correlation between the two parameters. When we extract the profile likelihood for one model parameter we can adjust the other model parameter such that the two parameters trace this narrow correlation. This way, the entire length of the correlation pattern gets projected onto the 1-dimensional profile likelihoods. The weak measurement dominates the individual profile likelihoods, while the strong measurement allows us to link the second model parameter from a first model parameter precisely.

In our case this implies that as long as the correlations from the neutron and Hg measurements cross the entire parameter space, we can ignore these two measurements and their correlation patterns in the following discussion of 2-dimensional correlations and single-parameter profile likelihoods. Losing the best few measurements contributing to the global analysis is an unfortunate effect of the conservative profile likelihood approach, but for standard error ellipses it should also exist for Bayesian marginalization. The main difference is that this kind of effect is numerically extremely challenging to compute using marginalization, while it is, essentially, trivial for the profile likelihood.

Given these considerations, we are left with five model parameters, constrained by seven measurements of similar constraining power. We expect correlation patterns in the allowed

parameter space, but no flat directions. Indeed, in Fig. 3 we see that all Lagrangian parameters are nicely constrained. The allowed range, for instance for $g_\pi^{(0)}$ is of the order 10^{-9}. This can be compared to the constraints from Fig. 1, of the order 10^{-12}. The same hierarchy of measurements can be observed for $g_\pi^{(1)}$ and for $d_n^{\rm sr}$, as confirmed by Tab. 4. This means that the 5-dimensional allowed parameter space illustrated by Fig. 3 is crossed by two correlation patterns, roughly three orders of magnitude more narrow than the full parameter space. We emphasize that explaining this extremely narrow correlation poses a fine-tuning problem in model parameter space, which the profile likelihood does not address.

4.5 Theory uncertainties

Theory uncertainties always appear when we use quantum field theory to predict observables, like EDMs, from Lagrangian parameters: no calculation method is arbitrarily precise, and a variety of systematic errors can effect accuracy. While there is some hope that we can estimate and control uncertainties for small expansion parameters, the uncertainties associated with QCD observables at low energies (whether from lattice calculations or sum-rule estimates) are much more tricky to estimate. The precision of nuclear physics calculations, and their links to effective quantum field theory, are also challenging to quantify. On the other hand we have to estimate all of these uncertainties, and we can only ignore them after having shown that they are significantly smaller than the experimental uncertainties of the associated measurements.

For our global EDM analysis, theory uncertainties affect the α_{i,c_j} in Eq.(23). The estimated

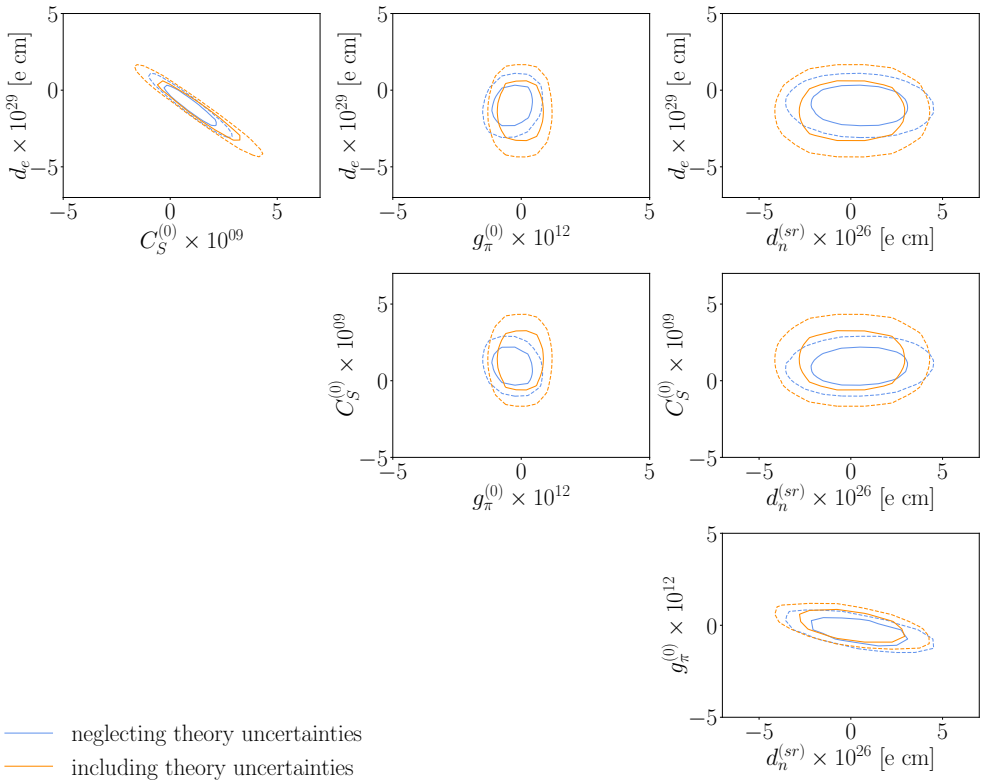


Figure 4: Correlations from the 4-dimensional analysis of $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n^{\rm sr}\}$. The orange curves show the effect of theory uncertainties on the results of Fig. 1. The ellipses indicate 68% and 95% CL.

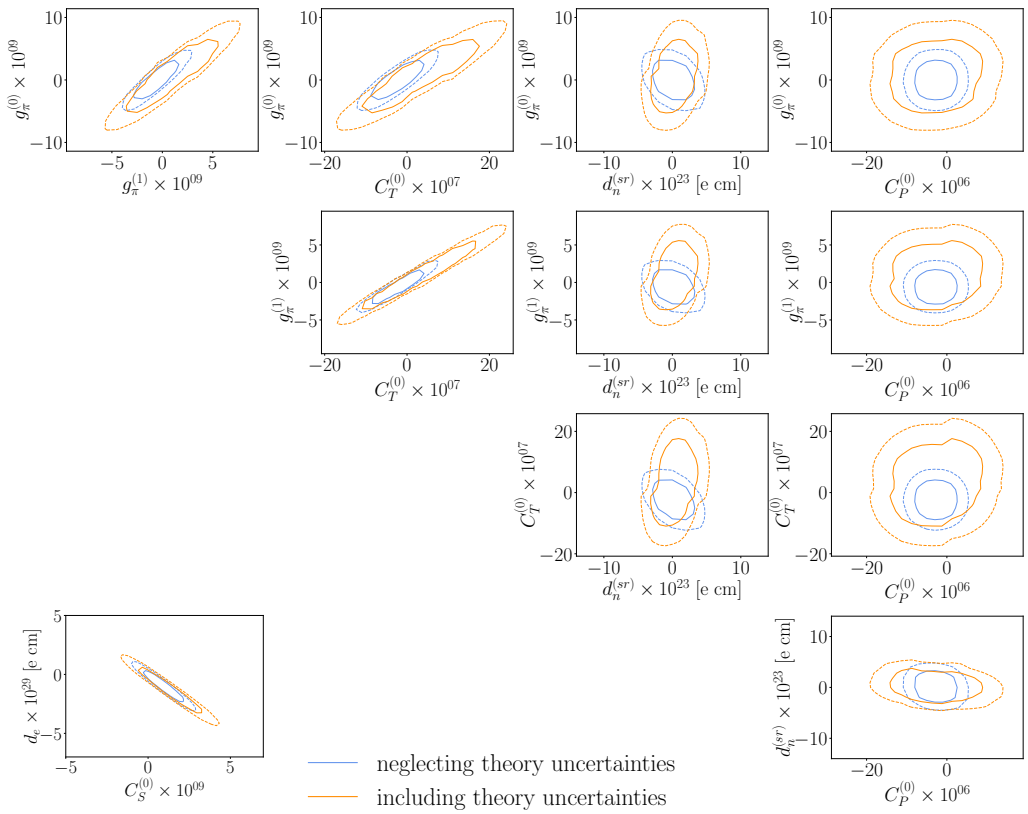


Figure 5: Correlations from the 5-dimensional analysis of $\{C_T^{(0)}, C_P^{(0)}, g_\pi^{(0)}, g_\pi^{(1)}, d_n^{\rm sr}\}$, and the factorized $d_e - C_S^{(0)}$ plane from Fig. 4. The orange curves show the effect of theory uncertainties on the results of Fig 3. The ellipses indicate 68% and 95% CL.

allowable ranges of each coefficient are given in Tab. 3. We include them in SFitter as uncorrelated theory uncertainties, an assumption that can be modified if necessary. Because of the flat likelihood as a function of the theory nuisance parameter, the profile likelihood approach leads to the theory uncertainties adding linearly, weighted by the respective model parameter. Profiling over independent α-ranges simplifies the numerical evaluation in two ways: first, any parameter–observable pair for which α is compatible with zero will effectively be removed from the global analysis, because the optimal choice of α will remove all contributions from the corresponding model parameter; second, even if we cannot choose α such that measurement and prediction agree, we can choose it to maximize the likelihood and to minimize the impact of the measurement, which means we choose the smallest allowed absolute value of α.

As in Sec. 4.2 we start with the well-constrained 4-dimensional subspace $\{d_e, C_S^{(0)}, g_\pi^{(0)}, d_n^{\rm sr}\}$. Again, d_e and $C_S^{(0)}$ are constrained by the paramagnetic molecules HfF^+ and ThO, just as without theory uncertainties. From Tab. 3 we see that we can ignore the theory uncertainty in relating the electron EDM parameter d_e to these systems. In the hadronic sector, the theory uncertainties affecting the relation of $g_\pi^{(0)}$ and $d_n^{\rm sr}$ to the neutron and Hg EDMs is reasonably small, although for the neutron this situation should not be over-interpreted. The quoted uncertainty arises from propagating ranges for the constants within Eq.(26), and not from careful evaluations of the chiral expansion itself. In Fig. 4 we show the numerical impact of the theory uncertainties on the 2-dimensional correlations. The slightly stronger HfF^+ measurement, which determines the width of the correlation pattern, is only minimally affected by the theory uncertainties, while the larger theory uncertainties on the ThO measurement extend the

length of the ellipse visibly.

Moving on, we can now look at the effect of the theory uncertainties on the less-constrained hadronic sector for diamagnetic systems, discussed in Sec. 4.3. Here, Tab. 3 shows sizeable, order-one theory uncertainties. In addition, some of the α-values include an allowed zero value when we include theory uncertainties. Specifically, $g_\pi^{(0)}$ will no longer be constrained by the Xe measurement and $g_\pi^{(1)}$ will lose the Hg constraint. In addition, some of the theory uncertainties in the diamagnetic sector are large, so we expect sizeable impact.

As for the case without theory uncertainties, the neutron and Hg limits are much more constraining than the other measurements of the hadronic sector. Following the argument given in Sec. 4.4, this means that 2-dimensional correlations and single-parameter limits extracted by profiling the likelihoods will not be impacted by these strong measurements. The shift in the constrained 2-dimensional correlations are shown in Fig. 5. In comparison their effect on the factorized parameters d_e and $C_S^{(0)}$ (copied from Fig. 4) is mild, and the constraints on the hadronic model parameters are consistently weaker. The only exception is d_n^{sr}, since as we know from Sec. 4.3 the measurements are hardly correlated and the theory uncertainties are smaller than order-one.

5 Outlook

EDMs are extremely sensitive, targeted probes of one of the most important symmetries of elementary particles, directly related to the observed baryon asymmetry in the Universe. The number of EDM measurements in very different systems has grown rapidly in recent years, leading to the question of how the different measurements can contribute to constraining and understanding CP violation in terms of a fundamental Lagrangian.

We can choose different Lagrangians to answer this question, starting with UV-complete models versus EFTs. In the absence of a specific hint for BSM physics we choose an EFT description. Next, we have a choice of different energy scales with different degrees of freedom. For our first SFitter analysis we rely on the hadronic-scale Lagrangian, valid at the GeV scale and describing the interactions of electrons and nucleons. After relating the hadronic-scale Lagrangian to its weak-scale SMEFT counterpart we want to constrain the seven Lagrangian parameters given in Fig. 6 through 11 independent measurements given in Tab. 1.

As a toy analysis we look at single-parameter constraints from individual EDM measurements. These limits all driven by the same small set of highly constraining measurements, like the paramagnetic molecular ion HfF^+, the neutron EDM, or the diamagnetic atom Hg. The extremely strong constraints indicated in Fig. 6 do not allow for a cancellation of contributions from two model parameters to a given measurement, at the price of creating a fine tuning problem. While the single-parameter estimates indicate the strength of an experiment looking for a sign of CP violation, they should not be confused with a measurement of a given parameter.

For our global analysis we use SFitter, with its focus on the statistical interpretation and a comprehensive uncertainty treatment. First, we ignore all theory uncertainties and only consider experimental uncertainties as uncorrelated and Gaussian (combining the statistical and systematic uncertainties reported in the respective papers). In this case Bayesian marginalization and profile likelihood give the same result.

In Sec. 4.2 we find that a small set of powerful measurements constrains the electron-hadron interactions as well as a subset of the hadronic sector. The correlated limits on the

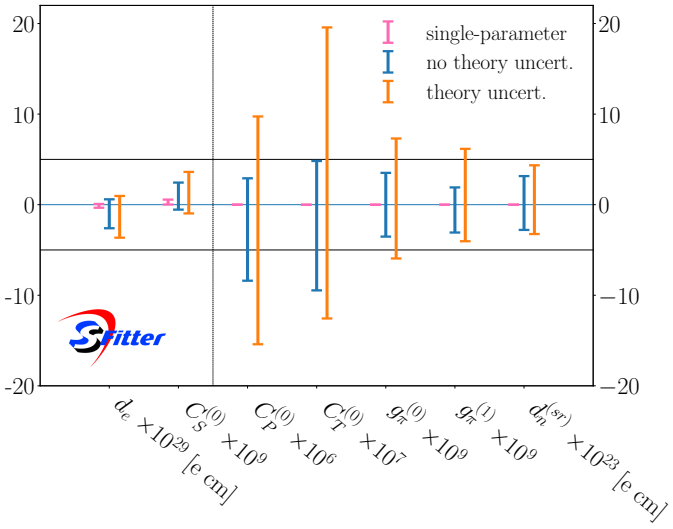


Figure 6: Constraints from the global EDM analysis on the parameters of the hadronic-scale Lagrangian. We show (i) hugely over-constrained single-parameter ranges allowed by the best available measurement; (ii) over-optimistic allowed ranges for profiled single parameters, ignoring theory uncertainties; (iii) allowed ranges for profiled single parameters including experimental and theory uncertainties.

electron EDM parameter d_e and the scalar coupling $C_S^{(0)}$ are especially strong and factorize from the hadronic sector. Next, we find in Sec. 4.3 that the constraints on the hadronic sector from the diamagnetic systems show rich correlations, motivating our global analysis described in Sec. 4.4. For the hadronic parameters, the narrow correlations from the strong neutron and Hg constraints do not appear in profiled 2-dimensional correlations or single-parameter limits. As a result, the hadronic model parameters are constrained much worse than the single-parameter results suggest, as can be in Fig. 6.

Finally, we show the same limits including theory uncertainties. Such theory uncertainties always appear when we relate measurements to fundamental Lagrangian parameters. We assume a flat likelihood within allowed ranges of the factors relating the Lagrangian parameters to the EDM predictions. While the impact of the theory uncertainties on the factorized semileptonic sector is relatively mild, the correlated analysis of the hadronic parameters leads to a further weakening of the constraints on the individual model parameters.

Acknowledgments

We gratefully acknowledge extremely helpful discussions and correspondence with Tim Chupp, Vincenzo Cirigliano, Jordy de Vries, Victor Flambaum, Timo Fleig, Maxim Pospelov, and Adam Ritz. We also thank Duarte Azevedo for his contributions at an early phase of this project. We would like to thank the ECT* Trento for the hospitality while submitting this work and presenting the results for the time.

NE is funded by the Heidelberg IMPRS *Precision Tests of Fundamental Symmetries*. This research is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 – TRR 257: *Particle Physics Phenomenology after the Higgs Discovery* and through Germany's Excellence Strategy EXC 2181/1 – 390900948 (the *Heidelberg STRUCTURES Excellence Cluster*). Furthermore, we acknowledge support by the state

of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 39/963-1 FUGG (bwForCluster NEMO).

References

[1] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32.

[2] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe*, Phys. Lett. B **155** (1985) 36.

[3] M. E. Shaposhnikov, *Baryon Asymmetry of the Universe in Standard Electroweak Theory*, Nucl. Phys. B **287** (1987) 757.

[4] A. E. Nelson, D. B. Kaplan, and A. G. Cohen, *Why there is something rather than nothing: Matter from weak interactions*, Nucl. Phys. B **373** (1992) 453.

[5] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak baryogenesis*, New J. Phys. **14** (2012) 125003, arXiv:1206.2942 [hep-ph].

[6] A. G. Cohen, D. B. Kaplan, and A. E. Nelson, *Progress in electroweak baryogenesis*, Ann. Rev. Nucl. Part. Sci. **43** (1993) 27, arXiv:hep-ph/9302210.

[7] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry*, Mod. Phys. Lett. A **9** (1994) 795, arXiv:hep-ph/9312215.

[8] P. Huet and E. Sather, *Electroweak baryogenesis and standard model CP violation*, Phys. Rev. D **51** (1995) 379, arXiv:hep-ph/9404302.

[9] M. B. Gavela, M. Lozano, J. Orloff, and O. Pene, *Standard model CP violation and baryon asymmetry. Part 1: Zero temperature*, Nucl. Phys. B **430** (1994) 345, arXiv:hep-ph/9406288.

[10] M. B. Gavela, P. Hernandez, J. Orloff, O. Pene, and C. Quimbay, *Standard model CP violation and baryon asymmetry. Part 2: Finite temperature*, Nucl. Phys. B **430** (1994) 382, arXiv:hep-ph/9406289.

[11] A. Riotto and M. Trodden, *Recent progress in baryogenesis*, Ann. Rev. Nucl. Part. Sci. **49** (1999) 35, arXiv:hep-ph/9901362.

[12] M. Pospelov and A. Ritz, *Electric dipole moments as probes of new physics*, Annals Phys. **318** (2005) 119, arXiv:hep-ph/0504231.

[13] J. Engel, M. J. Ramsey-Musolf, and U. van Kolck, *Electric Dipole Moments of Nucleons, Nuclei, and Atoms: The Standard Model and Beyond*, Prog. Part. Nucl. Phys. **71** (2013) 21, arXiv:1303.2371 [nucl-th].

[14] N. Yamanaka, B. K. Sahoo, N. Yoshinaga, T. Sato, K. Asahi, and B. P. Das, *Probing exotic phenomena at the interface of nuclear and particle physics with the electric dipole moments of diamagnetic atoms: A unique window to hadronic and semi-leptonic CP violation*, Eur. Phys. J. A **53** (2017) 3, 54, arXiv:1703.01570 [hep-ph].

[15] T. Chupp, P. Fierlinger, M. Ramsey-Musolf, and J. Singh, *Electric dipole moments of atoms, molecules, nuclei, and particles*, Rev. Mod. Phys. **91** (2019) 1, 015001, arXiv:1710.02504 [physics.atom-ph].

[16] E. Mereghetti, J. de Vries, W. H. Hockings, C. M. Maekawa, and U. van Kolck, *The Electric Dipole Form Factor of the Nucleon in Chiral Perturbation Theory to Sub-leading Order*, Phys. Lett. B **696** (2011) 97, arXiv:1010.4078 [hep-ph].

[17] W. Dekens, J. de Vries, M. Jung, and K. K. Vos, *The phenomenology of electric dipole moments in models of scalar leptoquarks*, JHEP **01** (2019) 069, arXiv:1809.09114 [hep-ph].

[18] D. McKeen, M. Pospelov, and A. Ritz, *Modified Higgs branching ratios versus CP and lepton flavor violation*, Phys. Rev. D **86** (2012) 113004, arXiv:1208.4597 [hep-ph].

[19] S. Dwivedi, D. K. Ghosh, B. Mukhopadhyaya, and A. Shivaji, *Constraints on CP-violating gauge-Higgs operators*, Phys. Rev. D **92** (2015) 9, 095015, arXiv:1505.05844 [hep-ph].

[20] Y. T. Chien, V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Direct and indirect constraints on CP-violating Higgs-quark and Higgs-gluon interactions*, JHEP **02** (2016) 011, arXiv:1510.00725 [hep-ph].

[21] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Is there room for CP violation in the top-Higgs sector?*, Phys. Rev. D **94** (2016) 1, 016002, arXiv:1603.03049 [hep-ph].

[22] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Constraining the top-Higgs sector of the Standard Model Effective Field Theory*, Phys. Rev. D **94** (2016) 3, 034031, arXiv:1605.04311 [hep-ph].

[23] S. Inoue, M. J. Ramsey-Musolf, and Y. Zhang, *CP-violating phenomenology of flavor conserving two Higgs doublet models*, Phys. Rev. D **89** (2014) 11, 115023, arXiv:1403.4257 [hep-ph].

[24] D. Fontes, M. Mühlleitner, J. C. Romão, R. Santos, J. a. P. Silva, and J. Wittbrodt, *The C2HDM revisited*, JHEP **02** (2018) 073, arXiv:1711.09419 [hep-ph].

[25] W. Altmannshofer, S. Gori, N. Hamer, and H. H. Patel, *Electron EDM in the complex two-Higgs doublet model*, Phys. Rev. D **102** (2020) 11, 115042, arXiv:2009.01258 [hep-ph].

[26] M. J. Ramsey-Musolf and J. C. Vasquez, *Left-right symmetry and electric dipole moments. a global analysis*, Physics Letters B **815** (2021) 136136.

[27] V. D. Barger, T. Falk, T. Han, J. Jiang, T. Li, and T. Plehn, *CP violating phases in SUSY, electric dipole moments, and linear colliders*, Phys. Rev. D **64** (2001) 056007, arXiv:hep-ph/0101106.

[28] J. R. Ellis, J. S. Lee, and A. Pilaftsis, *Electric Dipole Moments in the MSSM Reloaded*, JHEP **10** (2008) 049, arXiv:0808.1819 [hep-ph].

[29] Y. Li, S. Profumo, and M. Ramsey-Musolf, *A Comprehensive Analysis of Electric Dipole Moment Constraints on CP-violating Phases in the MSSM*, JHEP **08** (2010) 062, arXiv:1006.1440 [hep-ph].

[30] K. Cheung, T.-J. Hou, J. S. Lee, and E. Senaha, *Higgs Mediated EDMs in the Next-to-MSSM: An Application to Electroweak Baryogenesis*, Phys. Rev. D **84** (2011) 015002, arXiv:1102.5679 [hep-ph].

[31] S. F. King, M. Muhlleitner, R. Nevzorov, and K. Walz, *Exploring the CP-violating NMSSM: EDM Constraints and Phenomenology*, Nucl. Phys. B **901** (2015) 526, arXiv:1508.03255 [hep-ph].

[32] T. N. Dao, D. N. Le, and M. Mühlleitner, *Leptonic anomalous magnetic and electric dipole moments in the CP-violating NMSSM with and without inverse seesaw mechanism*, Eur. Phys. J. C **82** (2022) 10, 954, arXiv:2207.12618 [hep-ph].

[33] K. Gaul and R. Berger, *Global analysis of CP-violation in atoms, molecules and role of medium-heavy systems*, arXiv:2312.08858 [hep-ph].

[34] R. Lafaye, T. Plehn, and D. Zerwas, *SFITTER: SUSY parameter analysis at LHC and LC*, arXiv:hep-ph/0404282.

[35] R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Supersymmetry*, Eur. Phys. J. C **54** (2008) 617, arXiv:0709.3985 [hep-ph].

[36] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas, and M. Duhrssen, *Measuring the Higgs Sector*, JHEP **08** (2009) 009, arXiv:0904.3866 [hep-ph].

[37] I. Brivio, S. Bruggisser, N. Elmer, E. Geoffray, M. Luchmann, and T. Plehn, *To Profile or To Marginalize – A SMEFT Case Study*, SciPost Phys. **16** (2024) 035, arXiv:2208.08454 [hep-ph].

[38] N. Elmer, M. Madigan, T. Plehn, and N. Schmal, *Staying on Top of SMEFT-Likelihood Analyses*, arXiv:2312.12502 [hep-ph].

[39] A. Ghosh, B. Nachman, T. Plehn, L. Shire, T. M. P. Tait, and D. Whiteson, *Statistical patterns of theory uncertainties*, SciPost Phys. Core **6** (2023) 045, arXiv:2210.15167 [hep-ph].

[40] J. de Vries, E. Mereghetti, R. G. E. Timmermans, and U. van Kolck, *The Effective Chiral Lagrangian From Dimension-Six Parity and Time-Reversal Violation*, Annals Phys. **338** (2013) 50, arXiv:1212.0990 [hep-ph].

[41] C. M. Maekawa, E. Mereghetti, J. de Vries, and U. van Kolck, *The Time-Reversal- and Parity-Violating Nuclear Potential in Chiral Effective Theory*, Nucl. Phys. A **872** (2011) 117, arXiv:1106.6119 [nucl-th].

[42] A. Manohar and H. Georgi, *Chiral Quarks and the Nonrelativistic Quark Model*, Nucl. Phys. B **234** (1984) 189.

[43] J. de Vries, A. Gnech, and S. Shain, *Renormalization of CP-violating nuclear forces*, Phys. Rev. C **103** (2021) 1, L012501, arXiv:2007.04927 [hep-ph].

[44] S. Weinberg, *Nuclear forces from chiral Lagrangians*, Phys. Lett. B **251** (1990) 288.

[45] Y. V. Stadnik and V. V. Flambaum, *Nuclear spin-dependent interactions: Searches for WIMP, Axion and Topological Defect Dark Matter, and Tests of Fundamental Symmetries*, Eur. Phys. J. C **75** (2015) 3, 110, arXiv:1408.2184 [hep-ph].

[46] J. C. Berengut, V. V. Flambaum, and E. M. Kava, *Search for variation of fundamental constants and violations of fundamental symmetries using isotope comparisons*, Phys. Rev. A **84** (2011) 042510, arXiv:1109.1893 [physics.atom-ph].

[47] Flavour Lattice Averaging Group (FLAG), Y. Aoki *et al.*, *FLAG Review 2021*, Eur. Phys. J. C **82** (2022) 10, 869, arXiv:2111.09849 [hep-lat].

[48] HERMES, A. Airapetian *et al.*, *Precise determination of the spin structure function g(1) of the proton, deuteron and neutron*, Phys. Rev. D **75** (2007) 012007, arXiv:hep-ex/0609039.

[49] M. Shifman, A. Vainshtein, and V. Zakharov, *Remarks on higgs-boson interactions with nucleons*, Physics Letters B **78** (1978) 4, 443.

[50] A. Anselm, V. Bunakov, V. Gudkov, and N. Uraltsev, *On the neutron electric dipole moment in the weinberg cp-violation model*, Physics Letters B **152** (1985) 1, 116.

[51] S. M. Barr, *T and P odd electron - nucleon interactions and the electric dipole moments of large atoms*, Phys. Rev. D **45** (1992) 4148.

[52] M. Bauer and T. Plehn, *Yet Another Introduction to Dark Matter: The Particle Physics Approach*, vol. 959 of *Lecture Notes in Physics*. Springer, 2019. arXiv:1705.01987 [hep-ph].

[53] E. S. Ensberg, *Experimental upper limit for the permanent electric dipole moment of* rb^{85} *by optical pumping techniques*, Phys. Rev. **164** (Dec, 1967) 270.

[54] E. S. Ensberg, *Experimental upper limit for the permanent electric dipole moment of* rb^{85} *by optical-pumping techniques*, Phys. Rev. **153** (Jan, 1967) 36.

[55] M. A. Player and P. G. H. Sandars, *An experiment to search for an electric dipole moment in the 3p2 metastable state of xenon*, Journal of Physics B: Atomic and Molecular Physics **3** (dec, 1970) 1620.

[56] S. Eckel, P. Hamilton, E. Kirilov, H. W. Smith, and D. DeMille, *Search for the electron electric dipole moment using Ω-doublet levels in PbO*, Phys. Rev. A **87** (2013) 5, 052130, arXiv:1303.3075 [physics.atom-ph].

[57] S. Eckel, A. O. Sushkov, and S. K. Lamoreaux, *A limit on the electron electric dipole moment using paramagnetic ferroelectric* $Eu_{0.5}Ba_{0.5}TiO_3$, Phys. Rev. Lett. **109** (2012) 193003, arXiv:1208.4420 [physics.atom-ph].

[58] L. Pondrom, R. Handler, M. Sheaff, P. T. Cox, J. Dworkin, O. E. Overseth, T. Devlin, L. Schachinger, and K. J. Heller, *New Limit on the Electric Dipole Moment of the* Λ *Hyperon*, Phys. Rev. D **23** (1981) 814.

[59] C. Abel *et al.*, *Measurement of the Permanent Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **124** (2020) 8, 081803, arXiv:2001.11966 [hep-ex].

[60] B. C. Regan, E. D. Commins, C. J. Schmidt, and D. DeMille, *New limit on the electron electric dipole moment*, Phys. Rev. Lett. **88** (2002) 071805.

[61] S. A. Murthy, D. Krause, Z. L. Li, and L. R. Hunter, *New Limits on the Electron Electric Dipole Moment from Cesium*, Phys. Rev. Lett. **63** (1989) 965.

[62] T. S. Roussy *et al.*, *A new bound on the electron's electric dipole moment*, Science **381** (2023) 46, arXiv:2212.11841 [physics.atom-ph].

[63] ACME, V. Andreev *et al.*, *Improved limit on the electric dipole moment of the electron*, Nature **562** (2018) 7727, 355.

[64] J. J. Hudson, D. M. Kara, I. J. Smallman, B. E. Sauer, M. R. Tarbutt, and E. A. Hinds, *Improved measurement of the shape of the electron*, Nature **473** (2011) 493.

[65] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, *Reduced Limit on the Permanent Electric Dipole Moment of Hg199*, Phys. Rev. Lett. **116** (2016) 16, 161601, arXiv:1601.04339 [physics.atom-ph]. [Erratum: Phys.Rev.Lett. 119, 119901 (2017)].

[66] B. Graner, Y. Chen, E. G. Lindahl, and B. R. Heckel, *Erratum: Reduced limit on the permanent electric dipole moment of* ^{199}Hg *[phys. rev. lett. 116, 161601 (2016)]*, Phys. Rev. Lett. **119** (Sep, 2017) 119901.

[67] N. Sachdeva *et al.*, *New Limit on the Permanent Electric Dipole Moment of* ^{129}*Xe using* ^{3}*He Comagnetometry and SQUID Detection*, Phys. Rev. Lett. **123** (2019) 14, 143003, arXiv:1902.02864 [physics.atom-ph].

[68] F. Allmendinger, I. Engin, W. Heil, S. Karpuk, H.-J. Krause, B. Niederländer, A. Offenhäusser, M. Repetto, U. Schmidt, and S. Zimmer, *Measurement of the permanent electric dipole moment of the* ^{129}Xe *atom*, Phys. Rev. A **100** (Aug, 2019) 022505.

[69] T. A. Zheng, Y. A. Yang, S. Z. Wang, J. T. Singh, Z. X. Xiong, T. Xia, and Z. T. Lu, *Measurement of the Electric Dipole Moment of Yb171 Atoms in an Optical Dipole Trap*, Phys. Rev. Lett. **129** (2022) 8, 083001, arXiv:2207.08140 [physics.atom-ph].

[70] M. Bishof *et al.*, *Improved limit on the* ^{225}*Ra electric dipole moment*, Phys. Rev. C **94** (2016) 2, 025501, arXiv:1606.04931 [nucl-ex].

[71] D. Cho, K. Sangster, and E. A. Hinds, *Search for time reversal symmetry violation in thallium fluoride using a jet source*, Phys. Rev. A **44** (1991) 2783.

[72] P. V. Coveney and P. G. H. Sandars, *PARITY AND TIME VIOLATING INTERACTIONS IN THALLIUM FLUORIDE*, J. Phys. B **16** (1983) 3727.

[73] J. de Vries, P. Draper, K. Fuyuto, J. Kozaczuk, and B. Lillard, *Uncovering an axion mechanism with the EDM portfolio*, Phys. Rev. D **104** (2021) 5, 055039, arXiv:2107.04046 [hep-ph].

[74] Particle Data Group, R. L. Workman and Others, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[75] T. Fleig and L. V. Skripnikov, *P , t -violating and magnetic hyperfine interactions in atomic thallium*, Symmetry **12** (2020) 4, .

[76] T. Fleig and M. Jung, *Model-independent determinations of the electron EDM and the role of diamagnetic atoms*, JHEP **07** (2018) 012, arXiv:1802.02171 [hep-ph].

[77] A.-M. Mårtensson-Pendrill and P. Öster, *Calculations of atomic electric dipole moments*, Physica Scripta **36** (sep, 1987) 444.

[78] V. A. Dzuba, V. V. Flambaum, and S. G. Porsev, *Calculation of P,T-odd electric dipole moments for diamagnetic atoms Xe-129, Yb-171, Hg-199, Rn-211, and Ra-225*, Phys. Rev. A **80** (2009) 032120, arXiv:0906.5437 [physics.atom-ph].

[79] L. V. Skripnikov, *Communication: Theoretical study of HfF+ cation to search for the T,P-odd interactions*, The Journal of Chemical Physics **147** (07, 2017) 021101, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4993622/13976956/021101_1_online.pdf.

[80] T. Fleig, $\mathcal{P},\mathcal{T}$-*odd and magnetic hyperfine-interaction constants and excited-state lifetime for* HfF^+, Phys. Rev. A **96** (2017) 4, 040502, arXiv:1706.02893 [physics.atom-ph].

[81] A. N. Petrov, N. S. Mosyagin, T. A. Isaev, and A. V. Titov, *Theoretical study of Hf F+ in search of the electron electric dipole moment*, Phys. Rev. A **76** (2007) 030501, arXiv:physics/0611254.

[82] E. R. Meyer, J. L. Bohn, and M. P. Deskevich, *Candidate molecular ions for an electron electric dipole moment experiment*, Phys. Rev. A **73** (Jun, 2006) 062108.

[83] L. V. Skripnikov, *Combined 4-component and relativistic pseudopotential study of ThO for the electron electric dipole moment search*, The Journal of Chemical Physics **145** (12, 2016) 214301, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4968229/15523128/214301_1_online.pdf.

[84] E. R. Meyer and J. L. Bohn, *Prospects for an electron electric dipole moment search in metastable ThO and ThF*, Phys. Rev. A **78** (2008) 010502, arXiv:0805.0161 [physics.atom-ph].

[85] M. Denis and T. Fleig, *In search of discrete symmetry violations beyond the standard model: Thorium monoxide reloaded*, The Journal of Chemical Physics **145** (12, 2016) 214307, https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4968597/15523220/214307_1_online.pdf.

[86] V. A. Dzuba, V. V. Flambaum, and C. Harabati, *Relations between matrix elements of different weak interactions and interpretation of the parity-nonconserving and electron electric-dipole-moment measurements in atoms and molecules*, Phys. Rev. A **84** (Nov, 2011) 052108.

[87] V. A. Dzuba, V. V. Flambaum, and C. Harabati, *Erratum: Relations between matrix elements of different weak interactions and interpretation of the parity-nonconserving and electron electric-dipole-moment measurements in atoms and molecules [phys. rev. a 84, 052108 (2011)]*, Phys. Rev. A **85** (Feb, 2012) 029901.

[88] A. Sunaga, M. Abe, M. Hada, and B. P. Das, *Relativistic coupled-cluster calculation of the electron-nucleus scalar-pseudoscalar interaction constant* W_s *in ybf*, Phys. Rev. A **93** (Apr, 2016) 042507.

[89] M. Abe, G. Gopakumar, M. Hada, B. P. Das, H. Tatewaki, and D. Mukherjee, *Application of relativistic coupled-cluster theory to the effective electric field in ybf*, Phys. Rev. A **90** (Aug, 2014) 022501.

[90] J. Ginges and V. Flambaum, *Violations of fundamental symmetries in atoms and tests of unification theories of elementary particles*, Physics Reports **397** (2004) 2, 63.

[91] Y. Ema, T. Gao, and M. Pospelov, *Improved Indirect Limits on Muon Electric Dipole Moment*, Phys. Rev. Lett. **128** (2022) 13, 131803, arXiv:2108.05398 [hep-ph].

[92] ACME, J. Baron *et al.*, *Methods, Analysis, and the Treatment of Systematic Errors for the Electron Electric Dipole Moment Search in Thorium Monoxide*, New J. Phys. **19** (2017) 7, 073029, arXiv:1612.09318 [physics.atom-ph].

[93] K. Yanase and N. Shimizu, *Large-scale shell-model calculations of nuclear Schiff moments of ^{129}Xe and ^{199}Hg*, Phys. Rev. C **102** (2020) 6, 065502, arXiv:2006.15142 [nucl-th].

[94] A. Butter, S. Murgia, T. Plehn, and T. M. P. Tait, *Saving the MSSM from the Galactic Center Excess*, Phys. Rev. D **96** (2017) 3, 035036, arXiv:1612.07115 [hep-ph].

[95] M. Klute, R. Lafaye, T. Plehn, M. Rauch, and D. Zerwas, *Measuring Higgs Couplings from LHC Data*, Phys. Rev. Lett. **109** (2012) 101801, arXiv:1205.2699 [hep-ph].

[96] A. Biekötter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, SciPost Phys. **6** (2019) 6, 064, arXiv:1812.07587 [hep-ph].

[97] I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou, S. Westhoff, and C. Zhang, *O new physics, where art thou? A global search in the top sector*, JHEP **02** (2020) 131, arXiv:1910.03606 [hep-ph].

[98] D. López-Val, T. Plehn, and M. Rauch, *Measuring extended Higgs sectors as a consistent free couplings model*, JHEP **10** (2013) 134, arXiv:1308.1979 [hep-ph].

[99] I. Brivio, S. Bruggisser, E. Geoffray, W. Killian, M. Krämer, M. Luchmann, T. Plehn, and B. Summ, *From models to SMEFT and back?*, SciPost Phys. **12** (2022) 1, 036, arXiv:2108.01094 [hep-ph].

[100] A. Hocker, H. Lacker, S. Laplace, and F. Le Diberder, *A New approach to a global fit of the CKM matrix*, Eur. Phys. J. C **21** (2001) 225, arXiv:hep-ph/0104062.

Single charged Higgs pair production in exclusive processes at the LHC

Laura Duarte,[1,*] Victor P. Gonçalves,[1,†] Daniel E. Martins,[2,‡] and Téssio B. de Melo[3,4,§]

[1] *Institute of Physics and Mathematics, Federal University of Pelotas, Postal Code 354, 96010-900, Pelotas, RS, Brazil*

[2] *The Henryk Niewodniczanski Institute of Nuclear Physics (IFJ) Polish Academy of Sciences (PAN), 31-342, Krakow, Poland*

[3] *Universidad Andrés Bello, Facultad de Ciencias Exactas, Departamento de Ciencias Físicas-Center for Theoretical and Experimental Particle Physics, Fernández Concha 700, Santiago, Chile*

[4] *Millennium Institute for Subatomic Physics at the High-Energy Frontier, SAPHIR, Chile*

arXiv:2403.01953v1 [hep-ph] 4 Mar 2024

Abstract

The production of a single charged Higgs boson pair by photon - photon interactions in pp collisions at the LHC is investigated in this exploratory study. We focus on the exclusive production, which is characterized by intact protons and two - rapidity gaps in the final state, and assume the type - I two - Higgs - doublet model, which still allows a light charged Higgs. Assuming the leptonic $H^{\pm} \to [\tau\nu_{\tau}]$ decay mode, we derive predictions for the transverse momentum, rapidity and invariant mass distributions of the $\tau^{+}\tau^{-}$ pair for different values of the charged Higgs mass. The contribution of different background processes are also estimated. Our results indicate that the contribution of the exclusive $H^{+}H^{-}$ production for the $[\tau^{+}\nu_{\tau}][\tau^{-}\nu_{\tau}]$ final state is non - negligible and can, in principle, be used to searching for a light charged Higgs.

PACS numbers:

Keywords:

*Electronic address: l.duarte@unesp.br

†Electronic address: barros@ufpel.edu.br

‡Electronic address: daniel.ernani@ifj.edu.pl

§Electronic address: tessiomelo@institutosaphir.cl

I. INTRODUCTION

The discovery of the Higgs boson at the LHC in 2012 is one the great triumphs of the Particle Physics, and represents the completion of the Standard Model (SM) [1, 2]. However, there are still many unanswered questions that suggest that the SM is an effective low - energy realization of a more complete and fundamental theory. Several scenarios for the beyond the Standard Model (BSM) physics predict the presence of extra physical Higgs boson states, which has motivated the searching for these additional states in various production and decay channels over a wide range of kinematical regimes at LEP, Tevatron and now at the LHC [3–15]. As these studies have yield negative, constraints have been placed on the associated masses and branching ratios of different decay channels. In particular, the existence of a light charged Higgs boson, with mass below the top quark mass, is still allowed only in a restrict number of BSM models [16–21].

The two-Higgs-Doublet model (2HDM) is one the simplest BSM frameworks that predict charged Higgs bosons [22, 23]. In this model, an additional complex doublet is added and its Higgs sector involves five scalars: CP-even neutral h and H, CP - odd neutral A, and a pair of charged Higgs $H^{\pm}$. Four distinct interaction modes arise when a Z_2 symmetry is introduced to prevent the Flavor Changing Neutral Currents (FCNC) at the tree level. Current experimental measurements imply that a light $H^{\pm}$ can only be accommodated in type-I 2HDM and type-X 2HDM [18, 24, 25]. In this paper, we will concentrate our analysis in the type - I 2HDM, where one of doublets couples to all fermions, and a light charged Higgs with mass below 100 GeV is still allowed.

Over the last decades, extensive phenomenological studies on the $H^{\pm}$ production in e^+e^-, $\gamma\gamma$, ep and pp collisions have been performed assuming the type - I 2HDM. In particular, a comprehensive analysis of the single charged Higgs pair production at the LHC has been recently performed in Ref.[24]. Such a study considers the charged Higgs pair production in inelastic processes, where both the incident protons breakup, and the dominant subprocesses are initiated by the quarks and gluons present in the proton wave function. Moreover, they have explored the entire parameter of type - I 2HDM and obtained the phenomenologically viable parameters. The authors also have demonstrated that the significance for the $pp \to H^+H^- \to [\tau\nu][\tau\nu]$ channel is large and that a future experimental analysis of this final state is a promising way for searching and discover the single charged Higgs boson.

The main goal in this paper is to extend the analysis performed in Ref. [24] for elastic processes, where the two incident protons remain intact in the final state. In order to protons remain intact, the H^+H^- pair should be produced by the interaction of color singlet objects, which can be a photon γ, a Z boson or a Pomeron $\mathbb{P}$, which is a color singlet particle with partonic structure. As the dominant channel is the $\gamma\gamma \to H^+H^-$ subprocess [26], we will focus on the process represented in Fig. 1, which is usually denoted the exclusive H^+H^- production, since only the H^+H^- pair is present in the final

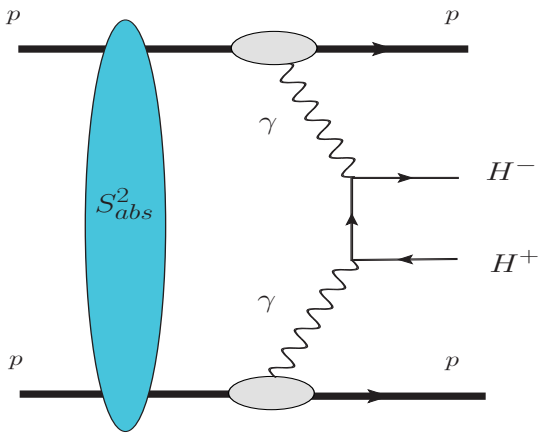


FIG. 1: Single charged Higgs pair production by $\gamma\gamma$ interactions in pp collisions at the LHC.

state. Our analysis is strongly motivated by the recent study performed in Ref. [27], where we have considered the exclusive production of a pair of doubly charged Higgs and demonstrated that such a process can be used to search for signatures of the type II seesaw mechanism and to obtain lower mass bounds on $H^{\pm\pm}$. We will focus on the $pp \to p \otimes H^+H^- \otimes p \to p \otimes [\tau\nu][\tau\nu] \otimes p$ channel, where $\otimes$ represents the presence of rapidity gaps in the final state, and will estimate the total cross-section and associated differential distributions on the transverse momentum, rapidity and invariant mass of the $\tau\tau$ pair, considering pp collisions at $\sqrt{s} = 14$ TeV and different values of the $H^{\pm}$ mass. A comparison with the predictions associated to the potential SM backgrounds will also be performed.

This paper is organized as follows. In the next section, we present a brief review of the type - I 2HDM and of the formalism used for the treatment of the H^+H^- pair production by photon-induced interactions in pp collisions. In Section III we discuss the backgrounds considered in our analysis and present our predictions for the invariant mass, transverse momentum and rapidity distributions, as well as for the total cross-section for the H^+H^- pair production in $\gamma\gamma$ interactions. Finally, in Section IV we summarize our main conclusions.

II. FORMALISM

A. The type-I 2HDM

The charged scalar bosons appear in several extensions of the SM [23, 24, 28–30]. A two-Higgs-doublet model (2HDM) is a simple extension of the SM by introducing an additional $SU(2)_L$ Higgs doublet, which predicts three neutral Higgs bosons and a pair of charged Higgs bosons $H^{\pm}$. To realize this model, the fermion content are the same as in the SM, while the scalar sector is extended with

two $SU(2)_L$ scalar doublets Φ_1 and Φ_2, which have weak hypercharge $Y = 1$:

$$\Phi_i = \begin{pmatrix} \phi_i^+ \\ \frac{v_i+\rho_i+i\eta_i}{\sqrt{2}} \end{pmatrix}, \qquad i = 1, 2 \ , \tag{1}$$

with v_1 and v_2 the vacuum expectation values (VEVs) of Φ_1 and Φ_2, respectively. v_1 and v_2 satisfy the relation $v = \sqrt{v_1^2 + v_2^2} = 246$ GeV, in order to successfully generate the electroweak symmetry breaking, and the ratio of v_2 and v_1 defines the mixing angle β as $\tan\beta = \frac{v_2}{v_1}$. In what follows, we use the simplified notation of $s_x = \sin x$, $c_x = \cos x$ and $t_x = \tan x$.

The physical mass eigenstates are given by

$$\begin{pmatrix} G^\pm \\ H^\pm \end{pmatrix} = R(\beta) \begin{pmatrix} \phi_1^\pm \\ \phi_2^\pm \end{pmatrix}, \qquad \begin{pmatrix} G \\ A \end{pmatrix} = R(\beta) \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \qquad \begin{pmatrix} H \\ h \end{pmatrix} = R(\beta) \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}. \tag{2}$$

Here $G^\pm$ and G are the Nambu-Goldstone bosons that are eaten as the longitudinal components of the massive gauge bosons. The rotation matrix is

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{3}$$

General 2HDMs are troubled by the presence of flavor changing neutral currents (FCNC) at tree level, as the gauge symmetries allow both doublets to couple to all fermions. In order to prevent large FCNC, additional symmetries may be employed, in order to forbid some of the offending couplings. A popular choice is to impose a discrete Z_2 symmetry[1], under which $\Phi_1 \to -\Phi_1$. The fermion Z_2 parities are such that each type of fermion couples to only one of the doublets. The different possibilities to fulfill this condition give rise to the well known four types of 2HDMs [23]: In the type-I model, only the doublet Φ_2 couples to all the fermions so all the quarks and charged leptons get their masses from the VEV of Φ_2 (ie. v_2); in the type-II, Φ_1 couples to charged leptons and down-type quarks and Φ_2 to up-type quarks; in the type-X, the charged leptons couple to Φ_1 and all the quarks couple to Φ_2, while in the type-Y model, Φ_1 couples to down-type quarks and Φ_2 to up-type quarks and charged leptons.

The most general scalar potential which is CP-conserving and invariant under the Z_2 symmetry (up to the soft-breaking term proportional to m_{12}^2) is given by:

$$\begin{aligned} V =& m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left(\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1 \right) + \frac{\lambda_1}{2} \left(\Phi_1^\dagger \Phi_1 \right)^2 + \frac{\lambda_2}{2} \left(\Phi_2^\dagger \Phi_2 \right)^2 \\ &+ \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} \left[\left(\Phi_1^\dagger \Phi_2 \right)^2 + \left(\Phi_2^\dagger \Phi_1 \right)^2 \right]. \end{aligned} \tag{4}$$

After spontaneous symmetry breaking, five physical scalars arise, three neutral h, H, A and a pair of charged scalars $H^\pm$. The remaining three scalars become the longitudinal components of the massive

[1] Continuous Abelian symmetries, either global or gauged, have also been considered in this context. See, e.g., [31–35].

$W^\pm$ and Z gauge bosons. As the CP symmetry is conserved, the neutral CP-even and CP-odd states do not mix, which means they can be diagonalized separately. The mixing angle for the CP-even sector is denoted by α, while β is the mixing angle for the CP-odd sector, as well as for the charged sector.

The eight parameters m_{ij}^2 and $\lambda_1 - \lambda_5$ in the potential (4) are replaced by the VEV v, the mixing angles α and β the scalar masses M_h, M_H, M_A and $M_H^\pm$, and the soft Z_2 breaking parameter $M^2 = m_{12}^2/s_\beta c_\beta$. In particular, the quartic coupling constants are given as [24]

$$\begin{aligned}
\lambda_1 &= \frac{1}{v^2}[(s_{\beta-\alpha} - c_{\beta-\alpha}t_\beta)^2 + M_H^2(s_{\beta-\alpha}t_\beta + c_{\beta-\alpha})^2 + M^2 t_\beta^2] \\
\lambda_2 &= \frac{1}{v^2}\left[M_h^2\left(s_{\beta-\alpha} + \frac{c_{\beta-\alpha}}{t_\beta}\right)^2 - \frac{M^2}{t_\beta^2} + M_H^2\left(\frac{s_{\beta-\alpha}}{t_\beta} - c_\beta\right)^2\right] \\
\lambda_3 &= \frac{1}{v^2}\left[(M_h^2 - M_H^2)\left\{s_{\beta-\alpha}^2 - s_{\beta-\alpha}c_{\beta-\alpha}\left(t_\beta - \frac{1}{t_\beta}\right) - c_{\beta-\alpha}^2\right\} + 2M_{H^\pm}^2 - M^2\right] \\
\lambda_4 &= \frac{1}{v^2}[M^2 + M_A^2 - 2M_{H^\pm}^2] \\
\lambda_5 &= \frac{1}{v^2}[M^2 - M_A^2].
\end{aligned} \tag{5}$$

The general Yukawa Lagrangian with two scalar doublets is

$$\begin{aligned}
-\mathcal{L}_Y = &\, y^{1d}\bar{Q}_L\Phi_1 d_R + y^{1u}\bar{Q}_L\tilde{\Phi}_1 u_R + y^{1e}\bar{L}_L\Phi_1 e_R \\
&+ y^{2d}\bar{Q}_L\Phi_2 d_R + y^{2u}\bar{Q}_L\tilde{\Phi}_2 u_R + y^{2e}\bar{L}_L\Phi_2 e_R,
\end{aligned} \tag{6}$$

where $Q_L^T = (u_L, d_L)$, $L_L^T = (\nu_l, l_l)$ and $y^{(1,2)(d,u,e)}$ are 3×3 matrices in family space. After imposing the Z_2 symmetry, some of these terms are forbidden. The production of a charged Higgs particle, depending on its mass with respect to the top quark, can be divided into light ($M_{H^\pm} \ll M_t$), intermediate ($M_{H^\pm} \sim M_t$) and heavy ($M_{H^\pm} \gg M_t$) scenarios [24, 36, 37]. Current searches impose stringent constraints on the mass of $H^\pm$ depending on the 2HDM type. For instance, the charged Higgs boson in type-II and type-Y is tightly constrained to be as heavy as $M_H^\pm \gtrsim 800$ GeV due to the measurements of the inclusive weak radiative B meson decay into $s\gamma$[38, 39]. Only type-I and type-X can accommodate a light $H^\pm$. As we are interested in studying relatively light charged scalars, with a mass ~ 100 GeV, we will focus here in the type-I 2HDM, in which such small masses are allowed provided that $\tan\beta$ is not too small. In this case, all the Yukawa couplings of $H^\pm$ are inversely proportional to $\tan\beta$ and the decay branching ratios into a fermion pair are proportional to the fermion mass. Restricting to the type-I model, the Yukawa Lagrangian (6) in the physical basis takes the form:

$$\begin{aligned}
-\mathcal{L}_Y^{\text{Type-I}} = &\frac{1}{t_\beta}\left\{\sum_f \frac{m_f}{v}\left(\frac{c_\alpha}{c_\beta}h\bar{f}f + \frac{s_\alpha}{c_\beta}H\bar{f}f\right) - iA(\bar{u}\gamma_5 u - \bar{d}\gamma_5 d - \bar{l}\gamma_5 l)\right. \\
&\left. + \left[\frac{\sqrt{2}V_{ud}}{v}H^+\bar{u}(m_u P_L - m_d P_R)d - \frac{\sqrt{2}m_l}{v}H^+\bar{\nu}_L l_R + h.c.\right]\right\}
\end{aligned} \tag{7}$$

with $f = u, d, l$. Notice that the coupling of the Higgs bosons to the fermions become suppressed when $t_\beta > 1$. The interaction of the scalars with the gauge bosons is given by,

$$\begin{aligned}\mathcal{L}_{\text{gauge}} = &\left(g m_W W_\mu^\dagger W^\mu + \frac{1}{2} g_Z m_Z Z_\mu Z^\mu\right)[s_{\beta-\alpha} h + c_{\beta-\alpha} H] \\ &+ \frac{g}{2} i[W_\mu^+[c_{\beta-\alpha} h - s_{\beta-\alpha} H] \overleftrightarrow{\partial^\mu} H^- - h.c.] - \frac{g}{2}[W_\mu^+ H^- \overleftrightarrow{\partial^\mu} A + h.c.] \\ &+ i\left[e A_\mu + \frac{g_Z}{2}(\sin^2\theta_W - \cos^2\theta_W) Z_\mu\right] H^+ \overleftrightarrow{\partial^\mu} H^- + \frac{g_Z}{2} Z_\mu [\cos(\beta-\alpha) A \overleftrightarrow{\partial^\mu} h - \sin(\beta-\alpha) A \overleftrightarrow{\partial^\mu} H],\end{aligned} \tag{8}$$

where $g_Z = g/\cos\theta_W$ and $f \overleftrightarrow{\partial^\mu} g \equiv (f\partial^\mu g - g\partial^\mu f)$.

We concentrate on the usual scenario where h is the SM-like 125 GeV Higgs boson, with being H the heavier neutral scalar. The pseudoscalar A can be either heavier or lighter than h. The parameter ranges are determined by the small $H^\pm$ mass scenario we are interested in, and also by the several experimental constraints: LEP experiments [3] have given limits on the mass of the charged Higgs boson in 2HDM from the charged Higgs searches in Drell-Yan events, $e^+e^- \to Z\gamma \to H^+H^-$, excluding $m_{H^\pm} \lesssim 80$ GeV (Type II) and $m_{H^\pm} \lesssim 72.5$ GeV (Type I) at 95% confidence level. Among the constraints from B meson decays (flavor physics constraints), the $B \to X_s\gamma$ decay [40] puts a very strong constraint on Type II and Type Y 2HDM, excluding $m_H^\pm \lesssim 580$ GeV and almost independently of t_β. For Type I and Type X, the $B \to X_s\gamma$ constraint is sensitive only for low t_β.

In Ref. [24], the authors have performed a comprehensive study and explored the entire parameter space for type - I 2DHM, deriving the current viable parameters for these models, taking into account the current theoretical and experimental constraints. In our analysis, we will make use of the results obtained in Ref. [24] and, in particular, we will estimate the exclusive cross-section for the parameters associated to the benchmark point 1 defined in Table III of that reference, in which $M_H = 138.6$ GeV, $M_A = 120.7$ GeV, $\tan\beta = 16.8$, $\sin(\beta-\alpha) = 0.975$, $m_{12}^2 = 1089.7$ GeV^2 and m_h being the mass of the observed Higgs boson. For $M_{H^\pm}$, three distinct values will be considered: $M_{H^\pm} = 80$, 100 and 140 GeV.

B. Single charged Higgs pair production by $\gamma\gamma$ interactions

Assuming the validity of the equivalent photon approximation (EPA) [41], the total cross section for the H^+H^- production by $\gamma\gamma$ interactions in pp collisions can be factorized in terms of the equivalent flux of photons into the proton projectiles and the $\gamma\gamma \to H^+H^-$ cross section, as follows

$$\sigma(pp \to p \otimes H^+H^- \otimes p) = S_{abs}^2 \int dx_1 \int dx_2\, \gamma_1^{el}(x_1) \cdot \gamma_2^{el}(x_2) \cdot \hat{\sigma}(\gamma\gamma \to H^+H^-)\ , \tag{9}$$

where x is the fraction of the proton energy carried by the photon and $\gamma^{el}(x)$ is the elastic equivalent photon distribution of the proton. The general expression for the elastic photon flux of the proton has

been derived in Ref. [42] and is given by

$$\gamma^{el}(x) = -\frac{\alpha}{2\pi}\int_{-\infty}^{-\frac{m^2x^2}{1-x}} \frac{dt}{t}\left\{\left[2\left(\frac{1}{x}-1\right)+\frac{2m^2x}{t}\right]H_1(t)+xG_M^2(t)\right\}\,, \tag{10}$$

where $t = q^2$ is the momentum transfer squared of the photon,

$$H_1(t) \equiv \frac{G_E^2(t)+\tau G_M^2(t)}{1+\tau} \tag{11}$$

with $\tau \equiv -t/m^2$, m being the proton mass, and where G_E and G_M are the Sachs elastic form factors. In our analysis, we will use the photon flux derived in Ref. [41], where an analytical expression is presented. Moreover, S^2_{abs} is the absorptive factor, which takes into account of additional soft interactions between incident protons which leads to an extra production of particles that destroy the rapidity gaps in the final state [43]. In our study, we will assume that $S^2_{abs} = 1$, which is a reasonable approximation since the contribution of the soft interactions is expected to be small in $\gamma\gamma$ interactions due to the long range of the electromagnetic interaction (For a more detailed discussion see, e.g., Ref. [26]). The cross section for the $\gamma\gamma \to H^+H^-$ subprocess, $\hat{\sigma}(\gamma\gamma \to H^+H^-)$, will be estimated using the type - I 2DHM and the events associated with the signal will be generated by MadGraph 5 [44].

A comment is in order here. The cross sections for the exclusive production of a given final state system, in general, two orders of magnitude smaller than for the production in inelastic proton - proton collisions, where both incident protons break up and a large number of particles is produced in addition to the final state. It turned out that the analysis of e.g. the H^+H^- production in inelastic collisions, generally involve serious backgrounds, thus making the search for new physics a hard task. In contrast, exclusive processes have smaller backgrounds and are characterized by a very clean final state, identified by the presence of two rapidity gaps, i.e. two regions devoid of hadronic activity separating the intact very forward protons from the central system. Such exclusive events can be clearly distinguished from the inelastic one by detecting the scattered protons in spectrometers placed in the very forward region close to the beam pipe, such as the ATLAS Forward Proton detector (AFP) [45, 46] and the CMS–Totem Precision Proton Spectrometer (CT–PPS) [47], and selecting events with two rapidity gaps in the central detector. For a more detailed discussion about the separation of exclusive processes we refer the interested reader to the recent studies performed in Refs. [48, 49].

III. RESULTS

In what follows we will present our results for the single charged Higgs pair production in exclusive processes considering pp collisions at $\sqrt{s} = 14$ TeV. In our analysis, we assume that the H^+H^- system decays leptonically, $H^+H^- \to [\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]$ and analyze two distinct experimental scenarios:

pp @ 14 TeV	**Signal**			**Backgrounds**	
	$pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p$			$pp \to pW^+W^-p \to p[\tau^+\nu_\tau][\tau^-\nu_\tau]p$	$pp \to p\tau^-\tau^+p$
$M_{H^\pm}$ [GeV]	80	110	140	-	-
σ(fb)	0.20	0.14	0.10	1.21	235500

TABLE I: Predictions at the generation level for the total cross-sections for the single charged Higgs pair production via photon-photon interactions in pp collisions at $\sqrt{s} = 14$ TeV, derived assuming the Type - I 2HDM and different values for the mass $M_{H^\pm}$. Results for the main backgrounds are also presented.

- **Scenario I:** the $\tau^+\tau^-$ pair is produced at central rapidities ($-2.0 \leq y(\tau\tau) \leq +2.0$) and both forward protons are tagged, which is the configuration that can be studied by the ATLAS and CMS Collaborations. We require both forward protons to be detected by Forward Proton Detectors (FPDs) and we will we assume an efficient reconstruction in the range $0.012 < \xi_{1,2} < 0.15$, where $\xi_{1,2} = 1 - p_{z1,2}/E_{\rm beam}$ is the fractional proton momentum loss on either side of the interaction point (side 1 or 2) and p_{z1} is the longitudinal momentum of the scattered proton on the side 1. This, in principle, allows one to measure masses of the central system by the missing mass method, $m_X = \sqrt{\xi_1\xi_2 s}$, starting from about 160 GeV.

- **Scenario II:** the $\tau^+\tau^-$ pair is produced at forward rapidities ($+2.0 \leq y(\tau\tau) \leq +4.5$), but the protons in the final state are not tagged, which is the case that can be analyzed by the LHCb Collaboration.

The signal is assumed to be the $pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p$ process, which will be generated using the MadGraph 5 [44, 50]. For the background we will consider the photon - induced processes $pp \to p \otimes W^+W^- \otimes p \to p \otimes [\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau] \otimes p$ and $pp \to p \otimes \tau^+\tau^- \otimes p$. All these backgrounds are also generated by MadGraph 5 [44].

In Table I we present our predictions for the total cross-sections associated with the signal and backgrounds considered in our analysis. The results for the signal have been derived assuming the type - I 2HDM for the benchmark point 1 [24], discussed in the previous Section, and different values for the mass $M_{H^\pm}$. The predictions for the signal at the generation level are similar to that derived in Ref. [26]. One has that the $pp \to p \otimes \tau^+\tau^- \otimes p$ process dominates the production of a $\tau^+\tau^-$ pair. However, such a process is characterized by a distinct topology in comparison to the other channels, where neutrinos (not seen by the detector) are also produced. In particular, this process is characterized by a pair with small acoplanarity [$\equiv 1 - (\Delta\phi/\pi)$], where $\Delta\phi$ is the angle between the τ particles, since the pair is dominantly produced back - to - back. Moreover, it is also characterized by an invariant mass $m_{\tau\tau}$ almost identical to the measured mass of the central system m_X, since other particle are not produced in addition to the $\tau^+\tau^-$ pair. Such expectations are confirmed by the results presented

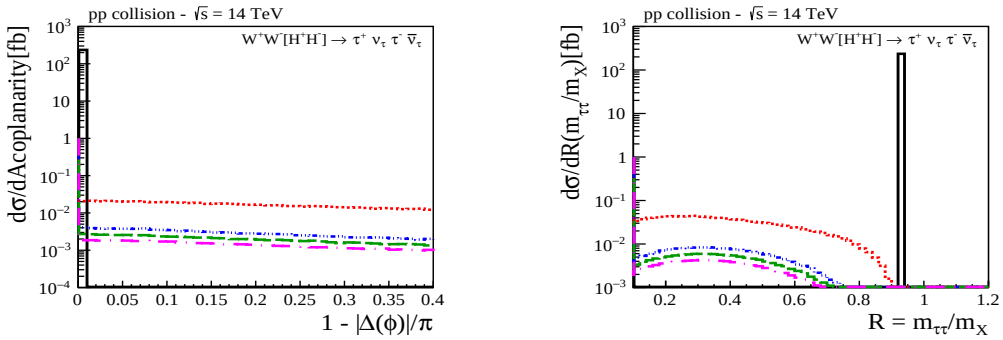


FIG. 2: Predictions for the acoplanarity distribution (left panel) and for the dependence on the ratio $R = m_{\tau\tau}/m_X$ (right panel) for the cross sections associated with the signal and backgrounds. The results for the signal have been derived assuming different values for the mass $M_{H^\pm}$.

in Fig. 2, where we show the acoplanarity distribution associated to the different channels (left panel) and the cross-sections as a function of the ratio $R = m_{\tau\tau}/m_X$ (right panel). These results suggest that the contribution of the $pp \to p \otimes \tau^+\tau^- \otimes p$ process can be strongly suppressed by assuming cuts on the acoplanarity and on the ratio R. In what follows, we will explore such expectations.

In Fig. 3 we present our predictions for the transverse momentum $[p_T(\tau\tau)]$, invariant mass $[m_{\tau\tau}]$ and rapidity $[y(\tau\tau)]$ distributions of the $\tau\tau$ pair associated with the $pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p$ and $pp \to p \otimes W^+W^- \otimes p \to p \otimes [\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau] \otimes p$ processes. The shape of the distributions for the signal assuming different values for the single charged Higgs mass are similar, decreasing in magnitude for larger values of $M_{H^\pm}$. Moreover, the position of the maximum in the $p_T(\tau\tau)$ and $m_{\tau\tau}$ distributions is dependent on the mass of the charged Higgs. Our results also indicate that the background predicts distributions that are similar to those from the signal, but with a larger normalization. In order to quantify the difference between the signal and background results for different values of $p_T(\tau\tau)$, $m_{\tau\tau}$ and $y(\tau\tau)$, we present in the lower panels of the plots shown in Fig. 3, the predictions for the ratio between the distributions associated with the W^+W^- and H^+H^- production, derived assuming distinct values of $M_{H^\pm}$. One has that the background dominates and is, in general, a factor ≥ 4 than the signal.

In what follows, we will investigate the impact of distinct kinematical cuts on the predictions for the total cross-sections considering the two scenarios discussed above. In particular, a cut on the ratio $R = m_{\tau\tau}/m_X$ will be considered in the case of the scenario I, since the protons in the final state are assumed to be tagged, and the central mass m_X can be estimated. In contrast, such a cut cannot be applied for the scenario II. Our results for the scenario I are presented in Table II. As expected from the results presented in Fig. 2, the cut on R is able to suppress the contribution associated with the $pp \to p\tau^-\tau^+p$ process, without impact on the other channels. The cut on the acoplanarity implies a small reduction of the W^+W^- background. On the other hand, if we impose that the $\tau^+\tau^-$

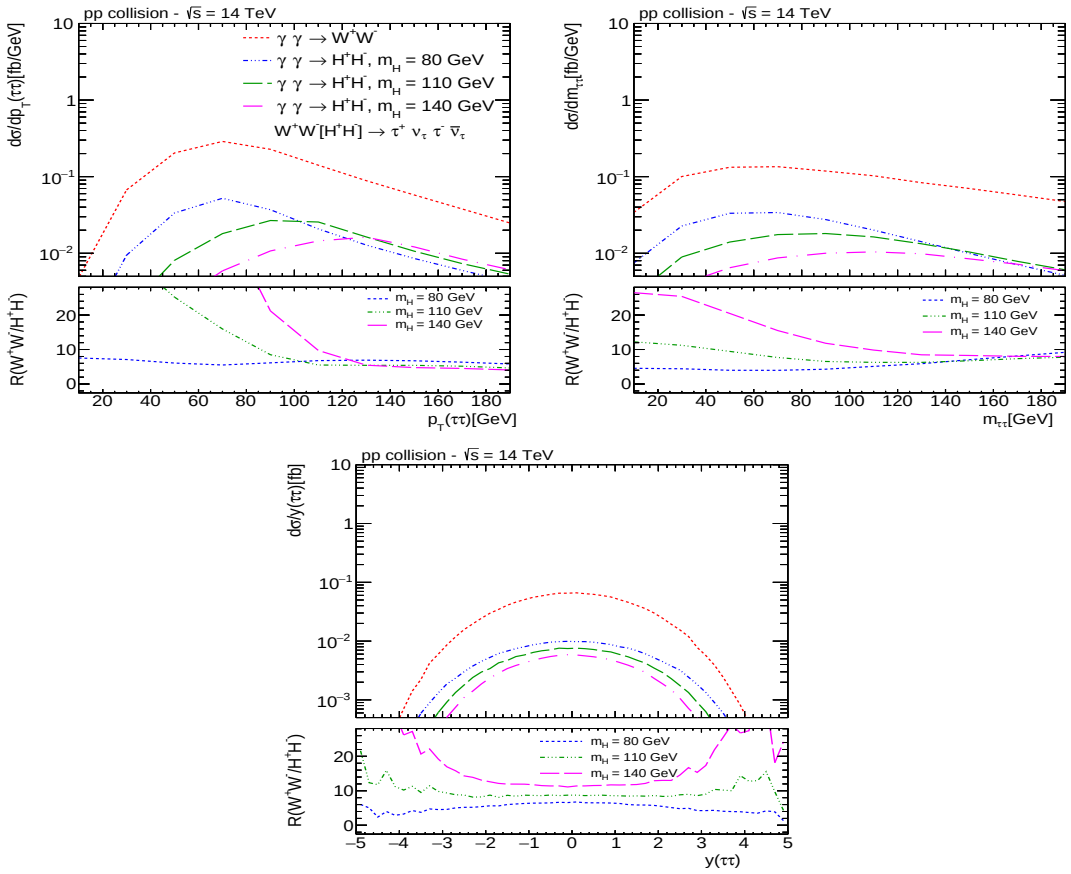


FIG. 3: Predictions for the transverse momentum $[p_T(\tau\tau)]$, invariant mass $[m_{\tau\tau}]$ and rapidity $[y(\tau\tau)]$ distributions of the $\tau\tau$ pair associated with the signal $[pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p]$, derived assuming distinct values of $M_{H^\pm}$ and considering pp collisions at $\sqrt{s} = 14$ TeV. The predictions for the background $pp \to p \otimes W^+W^- \otimes p \to p \otimes [\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau] \otimes p$ are also presented. The lower panels in the distinct plots present the ratio between the background and signal predictions.

pair must be produced at central rapidities, the predictions associated with the signal and background are reduced by $\approx 17\%$. One has that the signal predictions are of the order of ≈ 0.1 fb, while the background one is 1 fb. In addition, one has analyzed the impact of a cut on the invariant mass of the $\tau^+\tau^-$ pair. In particular, motivated by the results presented in Fig. 3 for the ratio between background and signal predictions for the invariant mass distributions, we have considered the selection of events in different ranges of $m_{\tau\tau}$, where this ratio assumes its smaller values. One has that such a cut implies the reduction of the total cross-sections, especially if events with smaller $m_{\tau\tau}$ are selected. However, in these cases we have the larger signal/background ratio, being $\approx 1/4$, in agreement with the results presented in Fig. 3. Such a result indicates that, for a single charged Higgs with a small mass, the exclusive H^+H^- production gives a non - negligible contribution for the $[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]$ final state.

In Table III we present the results associated with the scenario II, which can be investigated by the LHCb Collaboration. As explained before, for this scenario, the cut on R cannot be applied, since

pp @ 14 TeV	**Signal**			**Backgrounds**	
	$pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p$			$pp \to pW^+W^-p \to p[\tau^+\nu_\tau][\tau^-\nu_\tau]p$	$pp \to p\tau^-\tau^+p$
$M_{H^\pm}$ [GeV]	80	110	140	-	-
σ(fb) - w/o cuts	0.20	0.14	0.10	1.21	235500
$R \leq 0.9$	0.2	0.14	0.1	1.21	0.0
$[1 - (\Delta\phi/\pi)] \geq 0.01$	0.2	0.14	0.1	1.19	0.0
$-2.0 \leq y(\tau\tau) \leq +2.0$	0.16	0.12	0.086	1.00	0.0
$70.0 \leq m_{\tau\tau} \leq 90.0$ GeV	0.024	0.015	0.008	0.101	0.0
$100.0 \leq m_{\tau\tau} \leq 120.0$ GeV	0.097	0.049	0.025	0.410	0.0
$130.0 \leq m_{\tau\tau} \leq 150.0$ GeV	0.149	0.106	0.077	0.94	0.0

TABLE II: Predictions associated with the scenario I for the total cross-sections of single charged Higgs pair production via photon-photon interactions in pp collisions at $\sqrt{s} = 14$ TeV, derived assuming the Type - I 2HDM, different values for the mass $M_{H^\pm}$. The results were estimated considering a central detector and kinematical cuts on the acoplanarity, ratio R, rapidity and invariant mass of the $\tau\tau$ pair system. Results for the main backgrounds are also presented.

pp @ 14 TeV	**Signal**			**Backgrounds**	
	$pp \to pH^+H^-p \to p[\tau^+\nu_\tau][\tau^-\bar{\nu}_\tau]p$			$pp \to pW^+W^-p \to p[\tau^+\nu_\tau][\tau^-\nu_\tau]p$	$pp \to p\tau^-\tau^+p$
$M_{H^\pm}$ [GeV]	80	110	140	-	-
σ(fb) - w/o cuts	0.20	0.14	0.10	1.21	235500
$[1 - (\Delta\phi/\pi)] \geq 0.01$	0.19	0.14	0.098	1.19	0.0
$2.0 \leq y(\tau\tau) \leq 4.5$	0.019	0.011	0.006	0.093	0.0
$70.0 \leq m_{\tau\tau} \leq 90.0$ GeV	0.003	0.0015	0.0007	0.012	0.0
$100.0 \leq m_{\tau\tau} \leq 120.0$ GeV	0.002	0.0008	0.0006	0.009	0.0
$130.0 \leq m_{\tau\tau} \leq 150.0$ GeV	0.001	0.0009	0.0006	0.006	0.0

TABLE III: Predictions associated with the scenario II for the total cross-sections of single charged Higgs pair production via photon-photon interactions in pp collisions at $\sqrt{s} = 14$ TeV, derived assuming the Type - I 2HDM, different values for the mass $M_{H^\pm}$. The results were estimated considering a forward detector and kinematical cuts on the acoplanarity, rapidity and invariant mass of the $\tau\tau$ pair system. Results for the main backgrounds are also presented.

the central mass m_X cannot be reconstructed without the tagging of the protons in the final state. However, the results presented in Table III indicate that the cut on the acoplanarity is able to fully suppress the contribution associated with the $pp \to p\tau^-\tau^+p$ process. The selection of forward events implies the reduction of the cross - sections by almost one order of magnitude. Moreover, the cut on the invariant mass suppress the predictions by a factor ≥ 8, implying that the cross - sections become of the order of 10^{-3} fb, making a future experimental analysis of the exclusive H^+H^- production at forward rapidities a hard task.

IV. SUMMARY AND CONCLUDING REMARKS

Over the last decades, the study of photon - induced interactions in hadronic colliders became a reality, such that currently the Large Hadron Collider (LHC) is also considered a powerful photon - photon collider, which can be used to improve our understanding of the Standard Model as well as to searching for New Physics. The current data have already constrained several BSM scenarios, and more precise measurements are expected in the forthcoming years. Such an expectation has motivated the exploratory study performed in this paper, where we have considered the exclusive single charged Higgs pair production in pp collisions at $\sqrt{s} = 14$ TeV. One has assumed the type - I two - Higgs-Doublet model, which is one of the simplest BSM frameworks that predict charged Higgs bosons and that allows light charged Higgs with mass below 100 GeV. Our study complements the analysis performed in Ref. [24], where the single charged Higgs pair production in inelastic pp collisions was estimated and the current viable parameters for the type - I 2HDM were derived. The exclusive H^+H^- production cross - section was estimated considering different values for the charged Higgs and two distinct experimental configurations, whose are similar to those present in central (ATLAS/CMS) and forward (LHCb) detectors. We focused on the leptonic $H^{\pm} \to [\tau\nu_\tau]$ decay mode, which one has verified to be the final state with larger signal/background ratio. We have demonstrated that the background associated to the $pp \to p\tau^-\tau^+p$ can be fully removed by assuming a cut on the ratio $R = m_{\tau\tau}/m_X$ and/or in the acoplanarity. In contrast, the contribution of the $pp \to pW^+W^-p \to p[\tau^+\nu_\tau][\tau^-\nu_\tau]p$ background process dominates and generates distributions that are similar those predicted by the signal. However, the signal/background ratio is of the order of 1/4 for a light charged Higgs, which implies a non - negligible contribution for the $[\tau^+\nu_\tau][\tau^-\nu_\tau]$ final state. Such a promising result motivates the extension of this exploratory study by considering more sophisticated separation methods, as e.g. those used by the experimental collaborations in Refs. [11, 13], where kinematic variables that differentiate between the signal and backgrounds are identified and combined into a multivariate discriminant, with the output score of the boosted decision tree (BDT) used in order to separate the single charged signal from the SM background processes. We plan to perform such an extension in a forthcoming study.

Acknowledgments

D.E.M. thanks dearly, in the person of Janusz Chwastowski, the total support of the Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences (grant no. UMO2021/43/P/ST2/02279). T.B.M. acknowledges ANID-Chile grant FONDECYT No. 3220454 for financial support. This work was partially financed by the Brazilian funding agencies CNPq,

FAPERGS and INCT-FNA (processes number 464898/2014-5).

[1] G. Aad *et al.* [ATLAS], Phys. Lett. B **716**, 1-29 (2012)

[2] S. Chatrchyan *et al.* [CMS], Phys. Lett. B **716**, 30-61 (2012)

[3] [LEP Higgs Working Group for Higgs boson searches, ALEPH, DELPHI, L3 and OPAL], [arXiv:hep-ex/0107031 [hep-ex]].

[4] A. G. Akeroyd, S. Moretti and M. Song, Phys. Rev. D **101**, no.3, 035021 (2020)

[5] H. Bahl, T. Stefaniak and J. Wittbrodt, JHEP **06**, 183 (2021)

[6] S. Moretti, R. Santos and P. Sharma, Phys. Lett. B **760**, 697-705 (2016)

[7] M. Guchait and S. Moretti, JHEP **01**, 001 (2002)

[8] A. Arhrib, A. Jueid and S. Moretti, Phys. Rev. D **98**, no.11, 115006 (2018)

[9] A. M. Sirunyan *et al.* [CMS], JHEP **09** (2018), 007

[10] G. Aad *et al.* [ATLAS], Phys. Rev. Lett. **125** (2020) no.5, 051801

[11] M. Aaboud *et al.* [ATLAS], JHEP **09** (2018), 139

[12] G. Aad *et al.* [ATLAS], JHEP **06** (2021), 145

[13] A. M. Sirunyan *et al.* [CMS], JHEP **07** (2019), 142

[14] A. M. Sirunyan *et al.* [CMS], JHEP **01** (2020), 096

[15] A. Arhrib, R. Benbrik, H. Harouiz, S. Moretti and A. Rouchad, Front. in Phys. **8** (2020), 39

[16] A. G. Akeroyd, S. Moretti, K. Yagyu and E. Yildirim, Int. J. Mod. Phys. A **32**, no.23n24, 1750145 (2017)

[17] M. Drees, E. Ma, P. N. Pandita, D. P. Roy and S. K. Vempati, Phys. Lett. B **433**, 346-354 (1998)

[18] A. Arhrib, R. Benbrik, R. Enberg, W. Klemm, S. Moretti and S. Munir, Phys. Lett. B **774**, 591-598 (2017)

[19] M. Chakraborti, D. Das, M. Levy, S. Mukherjee and I. Saha, Phys. Rev. D **104** (2021) no.7, 075033

[20] M. Aoki, R. Guedes, S. Kanemura, S. Moretti, R. Santos and K. Yagyu, Phys. Rev. D **84** (2011), 055028

[21] H. Davoudiasl, W. J. Marciano, R. Ramos and M. Sher, Phys. Rev. D **89** (2014) no.11, 115008

[22] M. Aoki, S. Kanemura, K. Tsumura and K. Yagyu, Phys. Rev. D **80**, 015017 (2009)

[23] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher and J. P. Silva, Phys. Rept. **516**, 1-102 (2012)

[24] K. Cheung, A. Jueid, J. Kim, S. Lee, C. T. Lu and J. Song, Phys. Rev. D **105**, no.9, 095044 (2022)

[25] A. Arhrib, R. Benbrik, M. Krab, B. Manaut, S. Moretti, Y. Wang and Q. S. Yan, JHEP **10**, 073 (2021)

[26] P. Lebiedowicz and A. Szczurek, Phys. Rev. D **91**, 095008 (2015)

[27] L. Duarte, V. P. Goncalves, D. E. Martins, T. B. de Melo and F. S. Queiroz, Phys. Rev. D **107**, no.3, 035010 (2023)

[28] D. Eriklisson, F. Mahmoudi and O. Stal, JHEP **11**, 035 (2008)

[29] N. Craig and S. Thomas, JHEP **11**, 083 (2012)

[30] O. Eberhardt, U. Nierste and M. Wiebusch, JHEP **07**, 118 (2013)

[31] M. D. Campos, D. Cogollo, M. Lindner, T. Melo, F. S. Queiroz and W. Rodejohann, JHEP **08**, 092 (2017)

[32] S. M. Davidson and H. E. Logan, Phys. Rev. D **80**, 095008 (2009)

[33] D. A. Camargo, A. G. Dias, T. B. de Melo and F. S. Queiroz, JHEP **04** (2019), 129

[34] P. Ko, Y. Omura and C. Yu, Phys. Lett. B **717** (2012), 202-206

[35] P. Ko, Y. Omura and C. Yu, JHEP **01** (2014), 016

[36] M. Flechl, R. Klees, M. Kramer, M. Spira and M. Ubiali, Phys. Rev. D **91**, no.7, 075015 (2015)

[37] C. Degrande, R. Frederix, V. Hirschi, M. Ubiali, M. Wiesemann and M. Zaro, Phys. Lett. B **772**, 87-92 (2017)

[38] J. Haller, A. Hoecker, R. Kogler, K. Mönig, T. Peiffer and J. Stelzer, Eur. Phys. J. C **78**, no.8, 675 (2018)

[39] M. Misiak, H. M. Asatrian, R. Boughezal, M. Czakon, T. Ewerth, A. Ferroglia, P. Fiedler, P. Gambino, C. Greub and U. Haisch, *et al.* Phys. Rev. Lett. **114**, no.22, 221801 (2015)

[40] Y. Amhis *et al.* [HFLAV], [arXiv:1412.7515 [hep-ex]].

[41] V. M. Budnev, I. F. Ginzburg, G. V. Meledin and V. G. Serbo, Phys. Rept. **15**, 181 (1975).

[42] B. A. Kniehl, Phys. Lett. B **254**, 267 (1991).

[43] J. D. Bjorken, Phys. Rev. D **47**, 101 (1993).

[44] J. Alwall, M. Herquet, F. Maltoni *et al.*, JHEP **06**, 128 (2011).

[45] L. Adamczyk *et al.*, CERN-LHCC-2015-009, ATLAS-TDR-024.

[46] M. Tasevsky [ATLAS Collaboration], AIP Conf. Proc. **1654**, 090001 (2015).

[47] M. Albrow *et al.* [CMS and TOTEM Collaborations], CERN-LHCC-2014-021, TOTEM-TDR-003, CMS-TDR-13.

[48] V. P. Gonçalves, D. E. Martins, M. S. Rangel and M. Tasevsky, Phys. Rev. D **102**, no.7, 074014 (2020).

[49] D. E. Martins, M. Tasevsky and V. P. Goncalves, Phys. Rev. D **105**, no.11, 114002 (2022).

[50] N. D. Christensen and C. Duhr, Comput. Phys. Commun. **180**, 1614-1641 (2009)

Examining the critical phenomenon of pion parton distribution: Insights from the Moment Problem

Xiaobin Wang,[1,*] Zexin Wu,[1,†] Minghui Ding,[2,‡] and Lei Chang[1,§]

[1] *School of Physics, Nankai University, Tianjin 300071, China*
[2] *Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstraße 400, 01328 Dresden, Germany*

A recent study by Wang *et al.*(arXiv:2309.01417) proposed a novel connection between the nature of the parton distribution function (PDF) and the characteristics of its moments. In this study, we apply these findings to analyze the evolution of the pion valence quark PDF, garnering valuable qualitative insights. Firstly, we validate the non-negativity and continuity of the PDF across a wide range of scales, indicating the logical consistency of our chosen evolution scheme. Subsequently, we examine the unimodality of both the PDF and its transformed counterpart, the xPDF, i.e., the parton distribution function multiplied by the momentum fraction. We observe a smooth evolution of the peak position of the xPDF towards the small-x region with increasing scale, while intriguingly, the PDF undergoes a phase of bimodal competition as the energy scale evolves.

arXiv:2403.01937v1 [hep-ph] 4 Mar 2024

I. INTRODUCTION

Throughout history, there has been a great fascination with the intricate structure of particles. One pivotal advancement was Feynman's introduction of partons [1], which postulated that hadrons involved in high-energy collisions could be understood as combinations of numerous point-like partons. Subsequently, Bjorken applied this notion of partons to electron-nucleon deep inelastic scattering processes, elucidating the Bjorken scaling phenomenon observed in the structure function of the proton [2]. This scaling behavior is observed in the asymptotic limit, where the squared transferred momentum of the exchanged virtual photon Q^2 tends to infinity. In this limit, the structure function of the proton demonstrates scaling solely with respect to the light front momentum fraction x, regardless of Q^2. However, the experimentally and theoretically discovered Bjorken scaling violation catalyzed the formulation of the quantum field theory characterized by asymptotic freedom, quantum chromodynamics (QCD). QCD has now become the basis for describing high-energy collision processes, framing the connection between hadrons and partons through the concept of factorization [3]. This framework systematically separates short-range interactions between partons from long-range interactions associated with the color confinement, where one of the physical quantities describing the long-range interactions is the parton distribution function (PDF). It can be seen that the scale dependence phenomenon plays a crucial role in establishing the QCD factorization, and the PDF is scale-dependent.

At one extreme, in the ultra-high energy scale regime characterized by the asymptotic limit ($Q^2 \to \infty$), it has been found that for any given hadron, the first moments of the PDFs, denoting the light-front momentum fraction carried by various partons, exhibit different values. Valence quarks, sea quarks, and gluons exhibit $\langle x \rangle_q = 0$, $\langle x \rangle_s = 3/7$, and $\langle x \rangle_g = 4/7$, respectively [4]. This signifies that in the asymptotic limit, the valence quark PDF is effectively represented as a Dirac delta function centered at $x = 0$. Conversely, in the low energy scale regime, there is a widely discussed current assertion that there exists a hadron scale [5, 6] at which the valence quarks carry all of the light-front momentum of the hadron, with no presence of sea quarks or gluons at this energy scale. Assuming that this assertion holds and taking into account isospin symmetry, the valence quark PDF for the pion is a unimodal distribution, with a single peak positioned at $x = 1/2$ [7, 8]. This stark dichotomy raises a natural inquiry: How do these different manifestations of valence quark PDFs evolve from one configuration to the another?

A comprehensive understanding of the evolutionary path of these PDF profiles requires a full exploration of the underlying dynamics that govern the hadron structure at different energy scales. While the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi(DGLAP) evolution equation is widely recognized for describing PDFs at various scales, in recent years an all-orders evolution scheme has been developed [9, 10]. This innovative approach employs the process-independent (PI) effective charge of QCD to integrate the one-loop DGLAP equation [11]. In contrast to the conventional DGLAP scheme, where the running coupling exhibits infrared divergence, the process-independent effective charge employed in the all-orders evolution scheme is finite at all scales, from the infrared to the ultraviolet, and matches the behavior of the running coupling in the ultraviolet [12]. Notably, empirical validations have demonstrated the effectiveness of this evolution scheme in elucidating both meson and baryon systems [13, 14]. Given the compelling evidence for the all-orders evolution scheme, we opt to employ it herein.

In addition to the novel evolution scheme, an innovative perspective for studying PDFs has emerged: the application of the mathematical Moment Problem [15].

* wangxiaobin@mail.nankai.edu.cn
† wuzexin@mail.nankai.edu.cn
‡ m.ding@hzdr.de
§ leichang@nankai.edu.cn

Typically, theoretical methods [16] or Lattice QCD simulation [17–19] yield low-order moments, while the precise shape of the PDF remains elusive. Using the moment problem framework to analyze these moments provides a very effective way for extracting information regarding the PDF. Notably, this method furnishes insights into the location of the PDF's peaks, thereby facilitating the determination of the modal characteristics of the PDF, especially whether it exhibits a single peak, double peaks, or multiple peaks. This paper focuses on elucidating these modal features. Central to our study is to explore the evolutionary path of the pion valence quark PDF, i.e., the evolution from a single-peak distribution centered at $x = 1/2$ in the low-energy scale to a Dirac delta function centered at $x = 0$ in the high-energy scale, with the aim of investigating whether there are alterations in its modal characteristics during this evolution.

The paper is organized as follows: Sec. II compiles crucial inequalities from Ref. [15] that are relevant to the present study. Sec. III introduces the evolution scheme adopted in this study, namely, the all-orders evolution. Subsequently, Sec. IV presents the results of our analysis, which consists of two parts: examination for non-negativity and continuity, and examination of changes in peak position. Finally, the summary is presented in Sec. V.

II. THE MOMENT PROBLEM

In this section, we briefly introduce the moment problem and present essential inequalities extracted from Ref. [15] that are relevant to the present study. It is important to note that this section adopts a slightly more mathematical approach, and there may be variations in notation compared to other sections. For readers seeking a comprehensive understanding of the moment problem, we recommend consulting references such as Refs. [20, 21].

There exist various types of moment problems, and the specific type of moment problem under consideration can be formulated as follows: Given a sequence of real numbers $(s_k)_{k=0}^m$, it is required to find a non-negative function $f(x)$ that satisfies the equation:

$$s_k = \int_0^1 x^k f(x)\, dx \ (k = 0, 1, 2, \ldots, m)\,. \tag{1}$$

It is worth noting that this problem is not always solvable in general; its resolution requires the given sequence to satisfy certain special conditions. Ref. [15] used not only the non-negativity constraint imposed on $f(x)$ but also other pertinent properties to broaden the scope of the aforementioned problem. Consequently, they obtained the key inequalities essential for the analysis conducted in this study.

Prior to presenting the key inequalities, it is essential to introduce specific terminology and notation. It is termed unimodal if the function $f(x)$ satisfies the following conditions:

$$f(0) = f(1) = 0\,, \tag{2a}$$

$$\exists \lambda \in (0,1), \quad f'(x) \begin{cases} \geq 0, & 0 < x < \lambda \\ \leq 0, & \lambda < x < 1 \end{cases}, \tag{2b}$$

where λ corresponds to the position of the peak. We use $(s_{i+j})_{i,j=0}^n$ to represents the following Hankel matrix:

$$(s_{i+j})_{i,j=0}^n = \begin{pmatrix} s_0 & s_1 & \cdots & s_{n-1} & s_n \\ s_1 & s_2 & \cdots & s_n & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{n-1} & s_n & \cdots & s_{2n-2} & s_{2n-1} \\ s_n & s_{n+1} & \cdots & s_{2n-1} & s_{2n} \end{pmatrix}, \tag{3}$$

where i and j, represent the number of rows and columns of the matrix elements, respectively, with values ranging from 0 to n. In addition, we use $A \succ 0$ to indicate that the matrix A is positive definite.

We shall now present the key inequalities relevant to this study. Firstly, the necessary and sufficient condition for the existence of a non-negative continuous function $f(x)$ is that the sequence $(s_k)_{k=0}^m$ satisfies the following condition: for the even case $m = 2n$,

$$(s_{i+j})_{i,j=0}^n \succ 0 \quad \text{and} \quad (s_{i+j+1} - s_{i+j+2})_{i,j=0}^{n-1} \succ 0\,; \tag{4}$$

for the odd case $m = 2n + 1$,

$$(s_{i+j+1})_{i,j=0}^n \succ 0 \quad \text{and} \quad (s_{i+j} - s_{i+j+1})_{i,j=0}^n \succ 0\,. \tag{5}$$

Next, the necessary and sufficient condition for the existence of a non-negative, continuous, and unimodal function $f(x)$ is that the sequence $(s_k)_{k=0}^m$ satisfies the following conditions: for the even case $m = 2n$,

$$\begin{aligned} &\exists \lambda \in (0,1)\,, \quad (\lambda t_{i+j} - t_{i+j+1})_{i,j=0}^n \succ 0 \\ &\text{and} \quad [\lambda t_{i+j+1} - (\lambda+1) t_{i+j+2} + t_{i+j+3}]_{i,j=0}^{n-1} \succ 0\,; \end{aligned} \tag{6}$$

for the odd case $m = 2n + 1$,

$$\begin{aligned} &\exists \lambda \in (0,1)\,, \quad (\lambda t_{i+j+1} - t_{i+j+2})_{i,j=0}^n \succ 0 \\ &\text{and} \quad [\lambda t_{i+j} - (\lambda+1) t_{i+j+1} + t_{i+j+2}]_{i,j=0}^n \succ 0\,, \end{aligned} \tag{7}$$

where $t_k = -k s_{k-1}$.

There are two noteworthy observations regarding these inequalities: First, the conditions imposed on moments are all constituted by the positive definiteness of a specific matrix, whose dimensionality increases with the length of the moment sequence. Consequently, as the length of the moment sequence increases, meeting these conditions becomes progressively more challenging, and simultaneously, the judgement of these conditions can become progressively intricate. Second, these inequalities are only relevant to the existence of $f(x)$, indicating that the sequence $(s_k)_{k=0}^m$ satisfying the respective conditions may correspond to $f(x)$ not possessing the corresponding properties. However the reverse is perfectly clear, i.e., $(s_k)_{k=0}^m$ failing to meet the corresponding condition can not correspond to $f(x)$ exhibiting the specified property.

III. ALL-ORDERS EVOLUTION

In the subsequent section, we present the evolution scheme employed in this paper, namely, the all-orders evolution. The content presented herein is primarily drawn from Section VII of Ref. [9]. The all-orders evolution scheme is based on two foundational assumptions: Firstly, the existence of a hadron scale, denoted as ζ_H, at which the valence quarks carry all the momentum of the hadron, a concept originally introduced in Ref. [5]. Secondly, the existence of a process-independent effective charge in QCD, denoted as $\hat{\alpha}(k^2)$, which remains invariant under renormalization group transformations [9, 22]. For the pion valence quark PDF, the all-orders evolution scheme can be succinctly stated as follows:

$$\frac{\langle x^n\rangle^\zeta}{\langle x^n\rangle^{\zeta_H}} = \exp\left[-\frac{\gamma_0^n}{2\pi}\int_{\zeta_H}^{\zeta}\frac{dy}{y}\hat{\alpha}(y^2)\right], \tag{8}$$

where $\langle x^m\rangle^\zeta$ represents the m-th moment of the pion valence quark PDF at scale ζ ($\zeta_H = 0.331$ GeV), and

$$\gamma_0^n = -\frac{4}{3}\left(3 + \frac{2}{(n+1)(n+2)} - 4\sum_{j=1}^{n+1}\frac{1}{j}\right). \tag{9}$$

Moreover, we use an interpolation of the numerical results [12] as the specific form of $\hat{\alpha}(k^2)$, which is as follows:

$$\hat{\alpha}(k^2) = \frac{\gamma_m \pi}{\ln\left[\frac{\mathcal{K}^2(k^2)}{\Lambda_{\rm QCD}^2}\right]},\ \mathcal{K}^2(y) = \frac{a_0^2 + a_1 y + y^2}{b_0 + y}, \tag{10}$$

where $\gamma_m = 4/\beta_0$, $\beta_0 = 11 - (2/3)n_f$, $n_f = 4$, $\Lambda_{\rm QCD} = 0.234$GeV, with (in GeV^2)

a_0	a_1	b_0
0.104	0.0975	0.121

. (11)

For the sake of simplicity, we refrain from incorporating the corresponding relative errors of the parameters.

To elucidate the properties of $\hat{\alpha}(k^2)$, we compare it with QCD one-loop perturbative running coupling, as depicted in Fig. 1. It is noteworthy that $\hat{\alpha}(k^2)$ remains well defined at all scales without any singularities, and saturates in the infrared region. This property ensures the robust performance of the all-orders evolution scheme over a wide range of scales, which is essential for the investigation of critical phenomena in PDFs. Furthermore, in the ultraviolet region, $\hat{\alpha}(k^2)$ closely matches the QCD one-loop perturbative running coupling. However, a notable deviation between the two becomes apparent around the scale of 0.6 GeV.

IV. NUMERICAL RESULTS

At the outset of our investigation, we need a set of moments as input for the evolution. In this regard, we employ a conventional parametrization of the pion valence quark PDF at the hadron scale ζ_H, formulated as follows [23]:

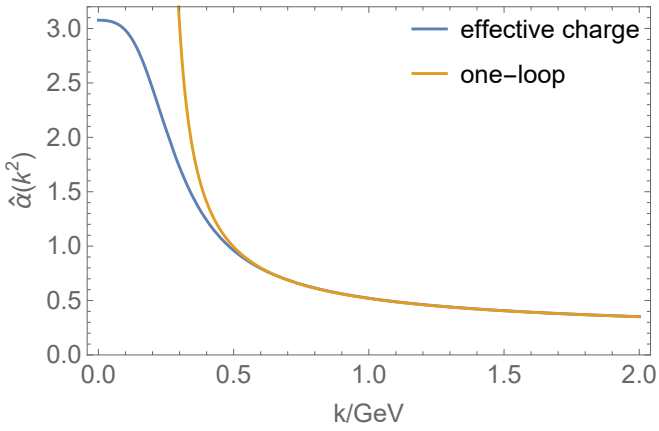


FIG. 1. Comparison between the process-independent effective charge and QCD one-loop perturbative running coupling.

$$u^\pi(x;\zeta_H) = n_\pi \ln[1 + x^2(1-x)^2/\rho^2], \tag{12}$$

where n_π is the normalization constant and $\rho = 0.0660$ is taken from the continuum Schwinger function method (CSM) prediction [24]. Next, we take as input to the evolution the zeroth to eleventh moments of the pion valence quark PDF at the hadron scale given in Table I.

n	0	1	2	3	4	5
$\langle x^n\rangle^{\zeta_H}$	1	0.5	0.29997	0.19995	0.14261	0.10660
n	6	7	8	9	10	11
$\langle x^n\rangle^{\zeta_H}$	0.08248	0.06554	0.05318	0.04390	0.03677	0.03116

TABLE I. The input to the evolution: the zeroth to eleventh moments of the pion valence quark PDF at the hadron scale.

A. Examination for non-negativity and continuity

Given the physical interpretation of the PDF, it is reasonable to anticipate its non-negativity and continuity at $\zeta < \infty$. To verify this conjecture, we conducted an examination of moments at various scales employing the 12 moments enumerated in Table I as input. Our investigation spans the range of scale ζ from ζ_H to 5.2 GeV, with a step size of 1 MeV. Remarkably, we consistently observed the preservation of non-negativity and continuity in the PDF moments over this scale range. This robust outcome is non-trivial and strongly supports to the viability of the all-orders evolution scheme. To further illustrate this point, we artificially float the value of the eleventh-order moment both upwards and downwards to delineate the extent of the 'safe zone' - the region where

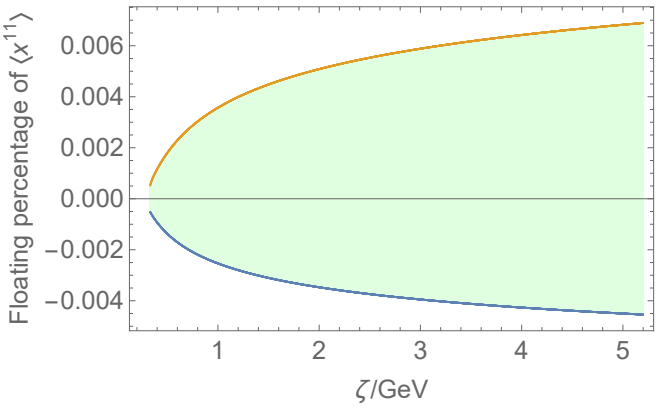


FIG. 2. The green region represents the safe zone of $\langle x^{11} \rangle$, where the non-negativity and continuity of PDF are maintained.

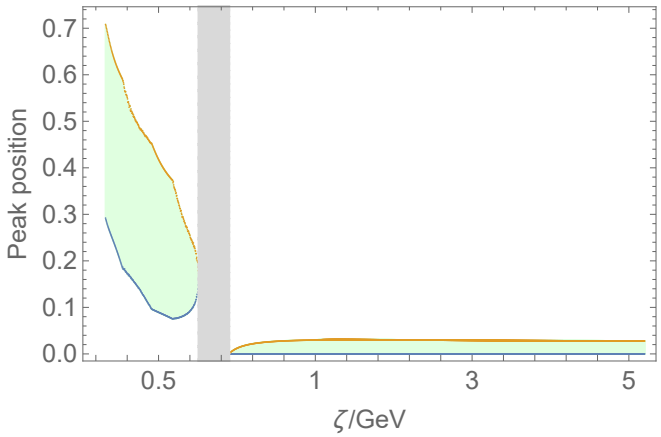


FIG. 3. The green region represents the possible ranges of the peak position of PDF, and the gray region indicates the breakdown of unimodality. Notably, regions that scale less than 1 GeV are enlarged to highlight them.

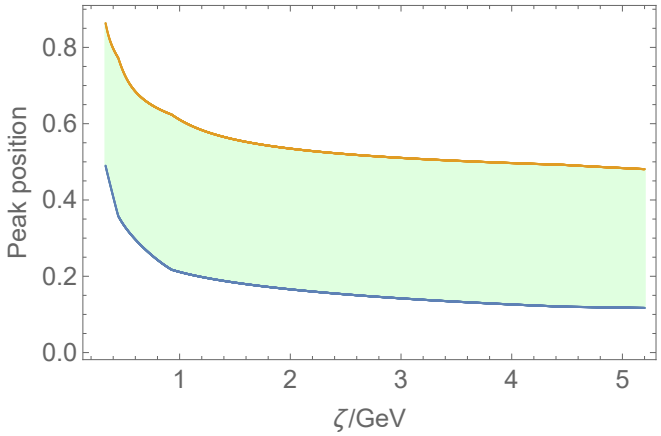


FIG. 4. The green region represents the possible range of the peak position of xPDF.

non-negativity and continuity are both preserved. The result of this analysis is depicted in Fig. 2.

In Fig. 2, it is evident that the safe zone exhibits remarkable narrowness, typically with a width of less than 0.01% of the floating percentage of the eleventh moments. Thus, it is truly remarkable that the all-orders evolution consistently operates within this exceedingly narrow region. Furthermore, the width of the safe region shows a tendency to expand with increasing scale, implying that the moment problem approach is more effective in the low-energy scale region.

B. Examination of changes in peak position

We proceed to analyze the unimodality of the PDF and the transformed one, xPDF. Unlike non-negativity and continuity, discussed previously, unimodality is not necessarily guaranteed. Nevertheless, based on the findings from various methods such as Lattice QCD [18, 19, 25, 26], experimental data [27, 28], global analysis [29, 30], DSE [31, 32], BLFQ-NJL [33], and χCQ [34], it is evident that at 5.2 GeV, the PDF exhibits an approximately unimodal behavior, diverging at $x = 0$, while the xPDF manifests as unimodal with its peak position approximately at $x = 0.3$. Notably, compared to their results at the hadron scale ζ_H, we observe that the peaks of both PDF and xPDF are shifted towards the small x region, a phenomenon that we will delve into further. Employing the aforementioned method, we determined the possible ranges of the positions of the PDF and xPDF peaks, as illustrated in Figs. 3 and 4, respectively.

These two figures provide valuable insights. Firstly, we observe that the results obtained at 5.2 GeV are consistent with the previous discussions, thereby affirming the robustness of our method. Additionally, we observe a notable distinction between the behavior of the PDF and the xPDF regarding the movement of the peak position towards the small x region. The xPDF exhibits a relatively straightforward pattern characterized by a continuous and monotonic variation of the peak range. Consequently, the evolution of the xPDF is naturally anticipated to exhibit a smooth variation. In contrast, the behavior of the PDF is notably more intricate, characterized by a distinct region where the unimodality of the PDF is disrupted. Consequently, the evolution of the PDF is expected to be considerably complex, and we propose that it can be qualitatively characteriszd in three stages. Initially, at low energy scale, before the disruption region, the peak of PDF, originally located at $x = 1/2$, gradually shifts towards the small x region while diminishing in peak magnitude, coinciding with the emergence of a new peak near $x = 0$. Subsequently, within the disrupted region, the height of the new peak gradually converges to that of the old peak, leading to the disruption of unimodality. Ultimately, at high energy scale, beyond the disrupted region, the prominence of the new peak surpasses that of the old peak, thereby reinstating

unimodality.

The bimodal competition process discussed above is highly innovative. Firstly, we note that the corresponding scale at which the unimodal disruption region appears aligns closely with the previously mentioned separation scale between the process-independent effective charge $\hat{\alpha}(k^2)$ and the QCD one-loop perturbative running coupling. In addition, the emergence of the two peaks occurs within distinct domains: one in the nonperturbative (low energy scale) region and the other in the perturbative (high energy scale) region. Therefore, we assert that the competition between the two peaks represents a unique critical phenomenon, wherein the new and old peaks correspond to perturbative and non-perturbative effects, respectively.

V. SUMMARY

In this study, we delve into the intricate domain of hadron physics by employing the moment problem as a pivotal tool to uncover critical phenomena within the foundational pion PDF.

We first introduce essential inequalities associated with the moment problem, which facilitate our understanding of the properties of PDF derived from its moments. Subsequently, we introduce the all-orders evolution scheme, and employ the moments in the form of conventional parameterizations at the hadron scale as inputs for the evolution process. Through this approach, we validate the non-negativity and continuity of the PDF, thereby providing evidence for the validity of the all-orders evolution scheme. Additionally, our finding indicates an enhanced effectiveness of the moment problem in the low energy scale region.

Moreover, our investigation extends to examine the unimodality of the PDF and its corresponding xPDF, thereby elucidating their qualitative evolution patterns. As the scale increases, while the peak position of the xPDF exhibits a gradual shift towards the small-x region, a notable critical phenomenon manifests in the PDF. At the interface of the nonperturbative and perturbative regimes, the PDF exhibits two distinct peaks, illustrating the competing interplay of the nonperturbative and perturbative aspects of QCD dynamics. This profound correspondence enhances our understanding of the intricate relationship between the PDF and the underlying strong interactions.

The method outlined in this paper holds significant potential for broader applicability, and we anticipate extending its scope to encompass other hadrons in future investigations. Furthermore, we believe that validating the correspondence between the two peaks of the PDF and the two effects of QCD - perturbation and non-perturbation - using alternative methods can further strengthen our findings.

ACKNOWLEDGMENTS

Work supported by National Natural Science Foundation of China (grant no. 12135007). MD is grateful for support by Helmholtz-Zentrum Dresden-Rossendorf High Potential Programme.

[1] R. P. Feynman, Phys. Rev. Lett. **23**, 1415 (1969).
[2] J. D. Bjorken, Phys. Rev. **179**, 1547 (1969).
[3] J. C. Collins, D. E. Soper, and G. F. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1 (1989), arXiv:hep-ph/0409313.
[4] G. Altarelli, Phys. Rept. **81**, 1 (1982).
[5] R. L. Jaffe and G. G. Ross, Phys. Lett. B **93**, 313 (1980).
[6] M. Diehl and P. Stienemeier, Eur. Phys. J. Plus **135**, 211 (2020), arXiv:1904.10722 [hep-ph].
[7] L. Chang and A. W. Thomas, Phys. Lett. B **749**, 547 (2015), arXiv:1410.8250 [nucl-th].
[8] M. Ding, K. Raya, D. Binosi, L. Chang, C. D. Roberts, and S. M. Schmidt, Chin. Phys. C **44**, 031002 (2020), arXiv:1912.07529 [hep-ph].
[9] K. Raya, Z.-F. Cui, L. Chang, J.-M. Morgado, C. D. Roberts, and J. Rodriguez-Quintero, Chin. Phys. C **46**, 013105 (2022), arXiv:2109.11686 [hep-ph].
[10] P.-L. Yin, Y.-Z. Xu, Z.-F. Cui, C. D. Roberts, and J. Rodríguez-Quintero, Chin. Phys. Lett. **40**, 091201 (2023), arXiv:2306.03274 [hep-ph].
[11] G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
[12] Z.-F. Cui, J.-L. Zhang, D. Binosi, F. de Soto, C. Mezrag, J. Papavassiliou, C. D. Roberts, J. Rodríguez-Quintero, J. Segovia, and S. Zafeiropoulos, Chin. Phys. C **44**, 083102 (2020), arXiv:1912.08232 [hep-ph].
[13] Y. Lu, L. Chang, K. Raya, C. D. Roberts, and J. Rodríguez-Quintero, Phys. Lett. B **830**, 137130 (2022), arXiv:2203.00753 [hep-ph].
[14] Y. Yu, P. Cheng, H.-Y. Xing, F. Gao, and C. D. Roberts, Contact interaction study of proton parton distributions (2024), arXiv:2402.06095 [hep-ph].
[15] X. Wang, M. Ding, and L. Chang, Sieving parton distribution function moments via the moment problem (2023), arXiv:2308.14871 [hep-ph].
[16] L. Chang, I. C. Cloet, J. J. Cobos-Martinez, C. D. Roberts, S. M. Schmidt, and P. C. Tandy, Phys. Rev. Lett. **110**, 132001 (2013), arXiv:1301.0324 [nucl-th].
[17] C. Alexandrou, S. Bacchio, I. Cloet, M. Constantinou, K. Hadjiyiannakou, G. Koutsou, and C. Lauer (ETM), Phys. Rev. D **103**, 014508 (2021), arXiv:2010.03495 [hep-lat].
[18] C. Alexandrou, S. Bacchio, I. Cloët, M. Constantinou, K. Hadjiyiannakou, G. Koutsou, and C. Lauer (ETM), Phys. Rev. D **104**, 054504 (2021), arXiv:2104.02247 [hep-lat].
[19] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, R. S. Sufian, and S. Zafeiropoulos, Phys. Rev. D **100**, 114512 (2019), arXiv:1909.08517 [hep-lat].
[20] K. Schmüdgen, Ten lectures on the moment problem (2020), arXiv:2008.12698 [math.FA].

[21] K. Schmüdgen, *The Moment Problem* (Springer, Leipzig, 2017).

[22] G. Grunberg, Phys. Rev. D **29**, 2315 (1984).

[23] Z. F. Cui, M. Ding, J. M. Morgado, K. Raya, D. Binosi, L. Chang, F. De Soto, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, Phys. Rev. D **105**, L091502 (2022), arXiv:2201.00884 [hep-ph].

[24] Y. Lu, Y.-Z. Xu, K. Raya, C. D. Roberts, and J. Rodríguez-Quintero, Pion distribution functions from low-order mellin moments (2023), arXiv:2311.08565 [hep-ph].

[25] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **99**, 074507 (2019), arXiv:1901.03921 [hep-lat].

[26] R. S. Sufian, C. Egerer, J. Karpie, R. G. Edwards, B. Joó, Y.-Q. Ma, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **102**, 054508 (2020), arXiv:2001.04960 [hep-lat].

[27] J. S. Conway *et al.*, Phys. Rev. D **39**, 92 (1989).

[28] M. Aicher, A. Schafer, and W. Vogelsang, Phys. Rev. Lett. **105**, 252003 (2010), arXiv:1009.2481 [hep-ph].

[29] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, Phys. Rev. Lett. **121**, 152001 (2018), arXiv:1804.01965 [hep-ph].

[30] N. Y. Cao, P. C. Barry, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum), Phys. Rev. D **103**, 114014 (2021), arXiv:2103.02159 [hep-ph].

[31] C. Chen, L. Chang, C. D. Roberts, S. Wan, and H.-S. Zong, Phys. Rev. D **93**, 074021 (2016), arXiv:1602.01502 [nucl-th].

[32] K. D. Bednar, I. C. Cloët, and P. C. Tandy, Phys. Rev. Lett. **124**, 042002 (2020), arXiv:1811.12310 [nucl-th].

[33] J. Lan, C. Mondal, S. Jia, X. Zhao, and J. P. Vary, Phys. Rev. D **101**, 034024 (2020), arXiv:1907.01509 [nucl-th].

[34] A. Watanabe, T. Sawada, and C. W. Kao, Phys. Rev. D **97**, 074015 (2018), arXiv:1710.09529 [hep-ph].

Prepared for submission to JHEP

Double-mixing CP violation in B decays

Wen-Jie Song, Yin-Fa Shen and Qin Qin

School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China

E-mail: jaysong@hust.edu.cn, syf70280@hust.edu.cn, qqin@hust.edu.cn

Abstract: We study a long-overlooked CP violation observable, termed double-mixing CP violation, which arises from the interference between two neutral meson oscillating paths involved in a decay chain. The double-mixing CP violation is beneficial for the precise test of the Standard Model CKM mechanism, as it offers the potential to extract weak phases without pollution from strong dynamics. To provide a comprehensive understanding of the double-mixing CP violation, we perform phenomenological analyses on the cascade decays of the B_d^0 and B_s^0 mesons. Our results show that the double-mixing CP violation can be very significant in certain decay channels, such as $B_s^0 \to \rho^0 K \to \rho^0(\pi^-\ell^+\nu_\ell)$. In addition, we employ the decay process $B_d^0 \to D^0 K \to (K^-\pi^+)(\pi\ell\nu)$ to demonstrate that the involved strong and weak phases can be directly determined from the experimental data without any theoretical inputs.

arXiv:2403.01904v1 [hep-ph] 4 Mar 2024

Contents

1 Introduction **2**

2 Formulae **3**

3 Numerical analysis **9**

3.1 B_d^0 decays 9

3.1.1 $B_d^0 \to J/\psi K \to J/\psi(\pi^\pm \ell^\mp \nu)$ 10

3.1.2 $B_d^0 \to J/\psi K \to J/\psi f_+$ 12

3.1.3 $B_d^0 \to DK \to (K^-\pi^+)(\pi^+\ell^-\bar{\nu}_\ell)$ 14

3.1.4 $B_d^0 \to DK_S^0 \to f_+ K_S^0$ 16

3.1.5 $B_d^0 \to DK_S^0 \to (K^-\pi^+)K_S^0$ 17

3.2 B_s^0 decays 19

3.2.1 $B_s^0 \to J/\psi K \to J/\psi f_+$ 20

3.2.2 $B_s^0 \to J/\psi K \to J/\psi(\pi^\pm \ell^\mp \nu_\ell)$ 20

3.2.3 $B_s^0 \to \rho^0 K \to \rho^0(\pi^\pm \ell^\mp \nu)$ 23

3.2.4 $B_s^0 \to DK \to (K^-\pi^+)(\pi^-\ell^+\nu_\ell)$ 26

3.2.5 $B_s^0 \to DK_S^0 \to (\pi^+\pi^-)K_S^0$ and $B_s^0 \to DK_S^0 \to (K^-\pi^+)K_S^0$ 29

4 Conclusions **30**

A Formulae for Category 3 **33**

1 Introduction

CP violation plays a crucial role in flavor physics. It is one of the fundamental criteria for explaining the matter-antimatter asymmetry of the universe [1]. However, the standard model (SM) does not account for a sufficiently large CP violation to explain the observed asymmetry [2, 3]. This suggests the presence of additional sources of CP violation beyond the standard Cabibbo-Kobayashi-Maskawa (CKM) mechanism [4, 5]. Consequently, the quest for non-standard CP violation is a vital undertaking in the quark flavor sector. One systematic approach involves testing the unitarity of the CKM matrix, which relies on measurements of CP violation in quark flavor processes to determine the values of the CKM phase angles. Therefore, it is imperative to conduct comprehensive investigations into CP violation in flavor physics, to precisely test and explore potential sources of CP violation beyond the framework of the CKM mechanism.

Since the unexpected discovery of indirect CP violation in the neutral K meson system in 1964 [6], significant theoretical and experimental advancements have been achieved [7–20], including the detection of CP violation in the B and D meson decays [10, 11, 20]. These studies have successfully observed three types of CP violation under the CKM mechanism in meson systems, along with measurements of CP-violating phases and testing of the unitarity of the CKM matrix. As no evidence of non-standard CP violation has been observed yet, collaborative efforts in both theoretical and experimental domains are essential to conduct more precise tests. In general, experimental precision of channels takes advantage of powerful flavor experiments such as BESIII [21], Belle II [22], and LHCb [23–25]. The primary theoretical challenge in improving precision lies in disentangling the effects of strong interactions within the SM and non-standard contributions. Despite numerous studies dedicated to strong interactions (see *e.g.* [26–35]), the theoretical predictions arising from strong interactions still struggle to align well with the current experimental precision. Therefore, there is an anticipation to provide CP violation observables that are less affected or unaffected by strong interactions.

Of the three types of observed CP violation—direct CP violation, CP violation in mixing, and CP violation in interference between a decay without mixing $M^0 \to f$ and a decay with mixing $M^0 \to \bar{M}^0 \to f$—the last one has the potential to avoid contamination from strong dynamics. The visualization of CP violation induced by the CKM mechanism typically involves interference between amplitudes with distinct CKM phases and CP-conserving phases, often including strong phases. However, the $M^0 \to f$ and $M^0 \to \bar{M}^0 \to f$ amplitudes have a relative CP-conserving phase induced by different time evolution, making strong phases not a necessity. In fact, the first significant CP violation effect observed by B factories in $B^0_d \to J/\psi K^0_S$ [10, 11] serves as a representative example of this type. The CP violation in the interference between $i \to M^0$ and $i \to \bar{M}^0 \to M^0$ [36] also capitalizes on the same advantage provided by neutral meson mixing. It motivates us to develop more CP violation observables that involve mixing.

In our recent study [37], we introduced a long-overlooked form of CP violation arising from the interference between two neutral meson oscillating paths involved in a decay chain. Such a decay, involving two neutral mesons $M^0_{1,2}$, typically occurs via two amplitudes:

$M_1^0 \to M_2^0 \to \bar{M}_2^0 \to f$ and $M_1^0 \to \bar{M}_1^0 \to \bar{M}_2^0 \to f$, as depicted in **Figure 1**, where the particles associated with M_2 are hidden for convenience. We refer to this type of CP violation as the double-mixing CP violation. This introduces a new experimental approach, allowing for measuring the dependence of CP violation on two time variables, specifically the evolution periods of $M_{1,2}$. Our previous research [37] revealed that the existence of double-mixing CP violation does not necessarily require a nonzero strong phase, offering the potential to extract weak phases without the influence of strong dynamics. In some cases where a strong phase is involved, it is also feasible to directly extract the strong phase from experimental data, thereby avoiding significant theoretical uncertainties. To provide a more comprehensive understanding, we will explore the double-mixing CP violation in typical B_d^0 and B_s^0 meson cascade decays in this study.

The rest of the paper is organized as follows. In section II, we provide the general formulas for the double-mixing CP violation in the process $M_1^0(t_1) \to M_2(t_2) \to f$, where M_2 is either M_2^0 or $\bar{M}_2^0$. In section III, we introduce a numerical analysis strategy to calculate the double-mixing CP violation in cascade decays of M_1^0, where M_1^0 can be either B_d^0 or B_s^0. Different initial and final states exhibit markedly distinct patterns. With frequently oscillating $B_{(s)}^0$ and K^0 involved, the double-mixing CP violation effects are generally more promising. In the case of semi-leptonic final states, such as $B_d^0 \to J/\psi K \to J/\psi(\pi\ell\nu)$, the total CP violation predominantly stems from contributions of the double-mixing CP violation. Conversely, the contribution of the double-mixing CP violation is exceedingly small when D^0 is involved, primarily due to the negligible mixing of D^0. We employ certain decay processes, such as $B_s^0 \to \rho^0 K \to \rho^0(\pi^\pm \ell^\mp \nu_\ell)$, to reveal that the double-mixing CP violation is possibly significantly large, exceeding 50%. These decay modes could serve as the initial experimental attempts to search for the double-mixing CP violation. Of greater significance, the double-mixing CP violation in certain decay channels under investigation can be utilized for CKM phase extraction, such as in the case of $B_d^0 \to D^0 K \to (K^-\pi^+)(\pi\ell\nu)$. Despite involving both strong and weak phases simultaneously, both phases can be extracted through experimental measurements of the double-mixing CP violation without relying on any theoretical inputs. In section IV, we summarize our results and present a conclusion.

2 Formulae

We adopt the conventional formulae that the mass eigenstates $M_{H,L}$ of the flavored neutral mesons are superpositions of their flavor eigenstates M^0 and $\bar{M}^0$ [38],

$$|M_{\rm H,L}\rangle = p\left|M^0\right\rangle \mp q\left|\bar{M}^0\right\rangle \ , \tag{2.1}$$

where q, p are complex coefficients. Then, the time evolution starting from a flavor eigenstate in the case of CPT conservation is given by,

$$\begin{aligned}\left|M^0(t)\right\rangle &= g_+(t)\left|M^0\right\rangle - \frac{q}{p}g_-(t)\left|\bar{M}^0\right\rangle, \\ \left|\bar{M}^0(t)\right\rangle &= g_+(t)\left|\bar{M}^0\right\rangle - \frac{p}{q}g_-(t)\left|M^0\right\rangle,\end{aligned} \tag{2.2}$$

where the functions $g_{\pm}(t)$ of time are given by

$$g_{\pm}(t) = \frac{1}{2}\left[\exp\left(-im_{\rm H}t - \frac{1}{2}\Gamma_{\rm H}t\right) \pm \exp\left(-im_{\rm L}t - \frac{1}{2}\Gamma_{\rm L}t\right)\right], \tag{2.3}$$

with $m_{\rm H(L)}$ and $\Gamma_{\rm H(L)}$ being the mass and decay width of the 'Heavy(Light)' mass eigenstate $M_{\rm H(L)}$, respectively.

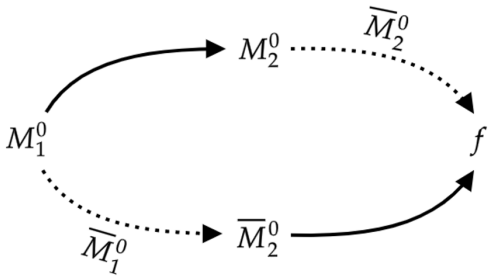


Figure 1. Two oscillation paths in the cascade decay $M_1^0 \to M_2 \to f$ with only $M_1^0 \to M_2^0$, $\bar{M}_2^0 \to f$ and their CP conjugate processes being allowed. The decay products associated with M_2 are not displayed.

The double-mixing CP violation will occur in the cascade decay containing at least two different neutral mesons. The process illustrated in **Figure 1** serves as a clear example. During the time interval t_1, a primary neutral meson M_1^0 either remains in its original state or oscillates into its antiparticle $\bar{M}_1^0$, which subsequently decays into a secondary neutral meson $\bar{M}_2^0(M_2^0)$. Other accompanying particles are not shown in the figure for simplicity. Similarly, the intermediate state $M_2(t_2)$, representing a time-evolved neutral meson, $M_2^0(t_2)$ or $\bar{M}_2^0(t_2)$, decays into a final state f. Here, t_2 denotes the time for the oscillation of $M_2^0(\bar{M}_2^0)$ in the rest frame of $M_2^0(\bar{M}_2^0)$. The time-dependent CP asymmetry in the chain $M_1^0(t_1) \to M_2(t_2) \to f$ can be defined as

$$\begin{aligned}
A_{\rm CP}(t_1,t_2) &\equiv \frac{|\mathcal{M}|^2(t_1,t_2) - |\bar{\mathcal{M}}|^2(t_1,t_2)}{|\mathcal{M}|^2(t_1,t_2) + |\bar{\mathcal{M}}|^2(t_1,t_2)} = \frac{N(t_1,t_2)}{D(t_1,t_2)}, \\
N(t_1,t_2) &= e^{-\Gamma_1 t_1}\Big[C_h(t_2)\cosh\frac{\Delta\Gamma_1 t_1}{2} + C_n(t_2)\cos\Delta m_1 t_1 \\
&\quad + S_n(t_2)\sin\Delta m_1 t_1 + S_h(t_2)\sinh\frac{\Delta\Gamma_1 t_1}{2}\Big], \\
D(t_1,t_2) &= e^{-\Gamma_1 t_1}\Big[C'_h(t_2)\cosh\frac{\Delta\Gamma_1 t_1}{2} + C'_n(t_2)\cos\Delta m_1 t_1 \\
&\quad + S'_n(t_2)\sin\Delta m_1 t_1 + S'_h(t_2)\sinh\frac{\Delta\Gamma_1 t_1}{2}\Big],
\end{aligned} \tag{2.4}$$

where the amplitude $\mathcal{M}(t_1,t_2)$ is the sum of amplitudes of all possible paths, while $\bar{\mathcal{M}}(t_1,t_2)$ represents the CP conjugate of $\mathcal{M}(t_1,t_2)$. The differences in widths and masses are defined as $\Delta\Gamma \equiv \Gamma_{\rm H} - \Gamma_{\rm L}$ and $\Delta m \equiv m_{\rm H} - m_{\rm L}$, respectively, and the average of widths is $\Gamma \equiv (\Gamma_{\rm H} + \Gamma_{\rm L})/2$. Here, the subscripts 1 and 2 denote the primary meson M_1 and the secondary

meson M_2, respectively. Evidently, according to the t_1 dependence, $N(t_1,t_2)$ and $D(t_1,t_2)$ can be written in four terms. Alternatively, a different t_1 function basis can be chosen, for example, rewriting the two terms in $N(t_1,t_2)$ as

$$e^{-\Gamma_1 t_1}\left[C_h(t_2)\cosh\frac{\Delta\Gamma_1 t_1}{2}+C_n(t_2)\cos\Delta m_1 t_1\right]\to C_+(t_2)|g_{+,1}(t_1)|^2+C_-(t_2)|g_{-,1}(t_1)|^2, \tag{2.5}$$

and the same substitution also applies to $D(t_1,t_2)$.

We will provide the explicit expressions of the terms in (2.4) for different cases. We categorize all cases into three groups based on the number of possible oscillation and decay paths. Besides the two-path scenario shown in **Figure 1**, there are also four- and eight-path ones, which will be discussed in more detail below.

Category 1

In the first category, we consider decay processes that occur via two possible paths, as exemplified in **Figure 1**. In this study, we are only interested in cases where direct CP violation is negligible, so we have $\langle\bar{M}_2^0|\bar{M}_1^0\rangle=\langle M_2^0|M_1^0\rangle e^{i\omega}$ with where ω being a pure weak phase, and $\langle f|\bar{M}_2^0\rangle=\langle\bar{f}|M_2^0\rangle$. The mixing parameters of $M_{1,2}$ are written as $(q/p)_{1,2}=|(q/p)|_{1,2}e^{-i\phi_{1,2}}$, with $\phi_{1,2}$ complex phases.

The amplitude $\mathcal{M}(t_1,t_2)$ of the process $M_1^0(t_1)\to M_2(t_2)\to f$ and its CP conjugate $\bar{\mathcal{M}}(t_1,t_2)$ are given by

$$\begin{aligned}\mathcal{M}(t_1,t_2)&=\langle f|M_2^0(t_2)\rangle\langle M_2^0|M_1^0(t_1)\rangle+\langle f|\bar{M}_2^0(t_2)\rangle\langle\bar{M}_2^0|M_1^0(t_1)\rangle,\\ \bar{\mathcal{M}}(t_1,t_2)&=\langle\bar{f}|M_2^0(t_2)\rangle\langle M_2^0|\bar{M}_1^0(t_1)\rangle+\langle\bar{f}|\bar{M}_2^0(t_2)\rangle\langle\bar{M}_2^0|\bar{M}_1^0(t_1)\rangle.\end{aligned} \tag{2.6}$$

Considering the time evolution of the neutral mesons (2.2) - (2.3), we can make the following substitutions

$$\begin{aligned}\langle M_2^0|M_1^0(t_1)\rangle&=g_{+,1}(t_1)\langle M_2^0|M_1^0\rangle, &\langle\bar{M}_2^0|M_1^0(t_1)\rangle&=-\frac{q_1}{p_1}g_{-,1}(t_1)\langle\bar{M}_2^0|\bar{M}_1^0\rangle,\\ \langle\bar{M}_2^0|\bar{M}_1^0(t_1)\rangle&=g_{+,1}(t_1)\langle\bar{M}_2^0|\bar{M}_1^0\rangle, &\langle M_2^0|\bar{M}_1^0(t_1)\rangle&=-\frac{p_1}{q_1}g_{-,1}(t_1)\langle M_2^0|M_1^0\rangle,\\ \langle f|M_2^0(t_2)\rangle&=-\frac{q_2}{p_2}g_{-,2}(t_2)\langle f|\bar{M}_2^0\rangle, &\langle\bar{f}|M_2^0(t_2)\rangle&=g_{+,2}(t_2)\langle\bar{f}|M_2^0\rangle,\\ \langle\bar{f}|\bar{M}_2^0(t_2)\rangle&=-\frac{p_2}{q_2}g_{-,2}(t_2)\langle\bar{f}|M_2^0\rangle, &\langle f|\bar{M}_2^0(t_2)\rangle&=g_{+,2}(t_2)\langle f|\bar{M}_2^0\rangle.\end{aligned} \tag{2.7}$$

Then, it arrives that the total time-dependent CP asymmetry (2.4).

The corresponding double-mixing CP asymmetry is reflected by the S_h and S_n terms, which are induced by the interference between the upper and lower paths in **Figure 1**. The results are given by

$$\begin{aligned}S_h(t_2)=\frac{e^{-\Gamma_2 t_2}}{2}\Big[&\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos(\omega-\phi_1+\phi_2)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|-\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\\ &+\sin\Delta m_2 t_2\sin(\omega-\phi_1+\phi_2)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|+\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\Big],\\ S_n(t_2)=\frac{e^{-\Gamma_2 t_2}}{2}\Big[&-\sinh\frac{\Delta\Gamma_2 t_2}{2}\sin(\omega-\phi_1+\phi_2)\left(\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|+\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|\right)\end{aligned} \tag{2.8}$$

$$+\sin\Delta m_2 t_2\cos\left(\omega-\phi_1+\phi_2\right)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|-\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\Big]. \tag{2.9}$$

The common factors $\left|\langle M_2^0|M_1^0\rangle\right|^2$ and $\left|\langle f|\bar{M}_2^0\rangle\right|^2$ are omitted in the above expression. The remaining terms in the numerator, which are not the focus of our current discussion, are given by

$$C_+(t_2)=\left(\left|\frac{q_2}{p_2}\right|^2-\left|\frac{p_2}{q_2}\right|^2\right)|g_{-,2}(t_2)|^2,\qquad C_-(t_2)=\left(\left|\frac{q_1}{p_1}\right|^2-\left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2\,. \tag{2.10}$$

The $C_+(t_2)$ term is contributed by purely the upper path in **Figure 1**, where $M_2^0\to\bar{M}_2^0$ oscillation is the only CP violation source, while the $C_-(t_2)$ term is contributed purely by the lower path, where CP violation is induced by $M_1^0\to\bar{M}_1^0$ oscillation. The terms contributing to $D(t_1,t_2)$ are given by

$$C'_+(t_2)=\left(\left|\frac{q_2}{p_2}\right|^2+\left|\frac{p_2}{q_2}\right|^2\right)|g_{-,2}(t_2)|^2,\qquad C'_-(t_2)=\left(\left|\frac{q_1}{p_1}\right|^2+\left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2, \tag{2.11}$$

$$\begin{aligned}S'_h(t_2)=&\frac{e^{-\Gamma_2 t_2}}{2}\Big[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\left(\omega-\phi_1+\phi_2\right)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|+\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\\&+\sin\Delta m_2 t_2\sin\left(\omega-\phi_1+\phi_2\right)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|-\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\Big],\end{aligned} \tag{2.12}$$

$$\begin{aligned}S'_n(t_2)=&\frac{e^{-\Gamma_2 t_2}}{2}\Big[-\sinh\frac{\Delta\Gamma_2 t_2}{2}\sin\left(\omega-\phi_1+\phi_2\right)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|-\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\\&+\sin\Delta m_2 t_2\cos\left(\omega-\phi_1+\phi_2\right)\left(\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|+\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|\right)\Big].\end{aligned} \tag{2.13}$$

For the convenience of further analysis, we define the two-dimensional time-dependent double-mixing CP asymmetry $A_{dm}(t_1,t_2)$, which is represented as

$$A_{dm}(t_1,t_2)=\frac{e^{-\Gamma_1 t_1}}{D(t_1,t_2)}\left(S_h(t_2)\sinh\frac{\Delta\Gamma_1 t_1}{2}+S_n(t_2)\sin\Delta m_1 t_1\right) \tag{2.14}$$

$$=A_{h,dm}(t_1,t_2)+A_{n,dm}(t_1,t_2). \tag{2.15}$$

It represents the double-mixing component within the $A_{\rm CP}(t_1,t_2)$ in (2.4). In fact, the remaining part is negligible.

For decay processes occurring via $M_1^0\to\bar{M}_2^0\to M_2^0\to f$ and $M_1^0\to\bar{M}_1^0\to M_2^0\to f$, the results can be directly obtained by applying the replacement $p_2/q_2\to q_2/p_2$ from (2.11) - (2.13), *i.e.*, $|p_2/q_2|\to|q_2/p_2|$ and $\phi_2\to-\phi_2$.

Category 2

In the second category, we consider that f is a CP eigenstate that can decay from both M_2^0 and $\bar{M}_2^0$. Unlike **Category 1**, there are four possible paths, as shown in **Figure 2**. Assuming negligible direct CP asymmetries, we can express $\langle\bar{M}_2^0|\bar{M}_1^0\rangle=\langle M_2^0|M_1^0\rangle e^{i\omega}$, and $\langle f|M_2^0\rangle=\langle f|\bar{M}_2^0\rangle$ under a suitable convention. It is important to note that in practical terms, M_2^0 here must be K^0. If it were D^0 decaying from B mesons, its antiparticle $\bar{D}^0$ could also decay from the same B meson, rendering this category inapplicable. This justifies

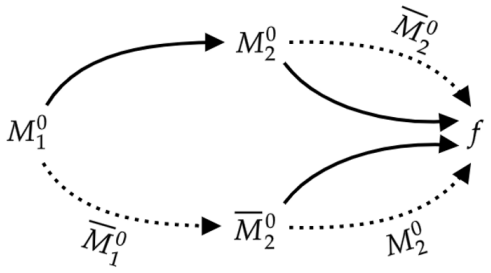


Figure 2. Four oscillation paths in the cascade decay $M_1^0 \to M_2 \to f$ where f is a CP eigenstate, with $M_1^0 \to M_2^0$, $\bar{M}_2^0 \to f$ and their CP conjugate processes being allowed.

the assumption $\langle f|M_2^0\rangle = \langle f|\bar{M}_2^0\rangle$ as CP violation is tiny in strange decays. The result for the CP asymmetry (2.4) is listed below.

To clarify the physical meaning, we separate the S_h and S_n terms in (2.4) into two parts, which are given by

$$S_{n,1}(t_2) = -\frac{e^{-\Gamma_2 t_2}}{2}\left[\left(\left|\frac{q_2}{p_2}\right| + \left|\frac{p_2}{q_2}\right|\right)\sin\Phi\sinh\frac{\Delta\Gamma_2 t_2}{2} - \left(\left|\frac{q_2}{p_2}\right| - \left|\frac{p_2}{q_2}\right|\right)\cos\Phi\sin\Delta m_2 t_2\right]\times \left(\left|\frac{q_1}{p_1}\right| + \left|\frac{p_1}{q_1}\right|\right), \tag{2.16}$$

$$S_{h,1}(t_2) = \frac{e^{-\Gamma_2 t_2}}{2}\left[\left(\left|\frac{q_2}{p_2}\right| + \left|\frac{p_2}{q_2}\right|\right)\cos\Phi\sinh\frac{\Delta\Gamma_2 t_2}{2} + \left(\left|\frac{q_2}{p_2}\right| - \left|\frac{p_2}{q_2}\right|\right)\sin\Phi\sin\Delta m_2 t_2\right]\times \left(\left|\frac{q_1}{p_1}\right| - \left|\frac{p_1}{q_1}\right|\right), \tag{2.17}$$

$$S_{n,2}(t_2) = -\left(\left|\frac{p_1}{q_1}\right| + \left|\frac{q_1}{p_1}\right|\right)\Big[\sin(\Phi-\phi_2)|g_{+,2}(t_2)|^2 + \sin(\Phi+\phi_2)|g_{-,2}(t_2)|^2\Big], \tag{2.18}$$

$$S_{h,2}(t_2) = \left(\left|\frac{q_1}{p_1}\right| - \left|\frac{p_1}{q_1}\right|\right)\Big[\cos(\Phi-\phi_2)|g_{+,2}(t_2)|^2 + \cos(\Phi+\phi_2)|g_{-,2}(t_2)|^2\Big], \tag{2.19}$$

where Φ is a shorthand for $\omega - \phi_1 + \phi_2$. The $S_{n,1}(t_2)$ and $S_{h,1}(t_2)$ terms are induced by the interference between two oscillation paths $M_1^0 \to M_2^0 \to \bar{M}_2^0$ and $M_1^0 \to \bar{M}_1^0 \to \bar{M}_2^0$, as well as the interference between $M_1^0 \to M_2^0$ and $M_1^0 \to \bar{M}_1^0 \to \bar{M}_2^0 \to M_2^0$. The $S_{n,2}(t_2)$ and $S_{h,2}(t_2)$ terms are induced by the interference between $M_1^0 \to M_2^0$ and $M_1^0 \to \bar{M}_1^0 \to \bar{M}_2^0$ and also the interference between $M_1^0 \to M_2^0 \to \bar{M}_2^0$ and $M_1^0 \to \bar{M}_1^0 \to \bar{M}_2^0 \to M_2^0$. The other terms involved in (2.4) are

$$\begin{aligned} C_n(t_2) = \frac{e^{-\Gamma_2 t_2}}{2}\Bigg\{&\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1+\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right] \\ &+ \sin\Delta m_2 t_2\sin\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1+\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right]\Bigg\} \\ &+ \frac{1}{2}\left[\left|\frac{q_2}{p_2}\right|^2\left(1+\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|^2\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right]|g_{-,2}(t_2)|^2 \end{aligned}$$

$$- \frac{1}{2}\left(\left|\frac{q_1}{p_1}\right|^2 - \left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2, \tag{2.20}$$

$$\begin{aligned} C_h(t_2) = & \frac{e^{-\Gamma_2 t_2}}{2}\left\{\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1-\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]\right. \\ & \left. + \sin\Delta m_2 t_2 \sin\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1-\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]\right\} \\ & + \frac{1}{2}\left[\left|\frac{q_2}{p_2}\right|^2\left(1-\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|^2\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]|g_{-,2}(t_2)|^2 \\ & + \frac{1}{2}\left(\left|\frac{q_1}{p_1}\right|^2 - \left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2, \end{aligned} \tag{2.21}$$

$$\begin{aligned} S'_n(t_2) = & -\frac{1}{2}\left(\left|\frac{q_1}{p_1}\right| - \left|\frac{p_1}{q_1}\right|\right)\left\{2\sin(\Phi-\phi_2)|g_{+,2}(t_2)|^2 + 2\sin(\Phi+\phi_2)|g_{-,2}(t_2)|^2\right. \\ & \left. + e^{-\Gamma_2 t_2}\left[\left(\left|\frac{q_2}{p_2}\right| + \left|\frac{p_2}{q_2}\right|\right)\sin\Phi\sinh\frac{\Delta\Gamma_2 t_2}{2} - \left(\left|\frac{q_2}{p_2}\right| - \left|\frac{p_2}{q_2}\right|\right)\cos\Phi\sin\Delta m_2 t_2\right]\right\}, \end{aligned} \tag{2.22}$$

$$\begin{aligned} S'_h(t_2) = & \frac{1}{2}\left(\left|\frac{q_1}{p_1}\right| + \left|\frac{p_1}{q_1}\right|\right)\left\{2\cos(\Phi-\phi_2)|g_{+,2}(t_2)|^2 + 2\cos(\Phi+\phi_2)|g_{-,2}(t_2)|^2\right. \\ & \left. + e^{-\Gamma_2 t_2}\left[\left(\left|\frac{q_2}{p_2}\right| + \left|\frac{p_2}{q_2}\right|\right)\cos\Phi\sinh\frac{\Delta\Gamma_2 t_2}{2} + \left(\left|\frac{q_2}{p_2}\right| - \left|\frac{p_2}{q_2}\right|\right)\sin\Phi\sin\Delta m_2 t_2\right]\right\}, \end{aligned} \tag{2.23}$$

$$\begin{aligned} C'_n(t_2) = & \frac{e^{-\Gamma_2 t_2}}{2}\left\{\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1-\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]\right. \\ & \left. + \sin\Delta m_2 t_2 \sin\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1-\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]\right\} \\ & + \frac{1}{2}\left[\left|\frac{q_2}{p_2}\right|^2\left(1-\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|^2\left(1-\left|\frac{q_1}{p_1}\right|^2\right)\right]|g_{-,2}(t_2)|^2 \\ & - \frac{1}{2}\left(\left|\frac{q_1}{p_1}\right|^2 + \left|\frac{p_1}{q_1}\right|^2 - 2\right)|g_{+,2}(t_2)|^2, \end{aligned} \tag{2.24}$$

$$\begin{aligned} C'_h(t_2) = & \frac{e^{-\Gamma_2 t_2}}{2}\left\{\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1+\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right]\right. \\ & \left. + \sin\Delta m_2 t_2 \sin\phi_2\left[\left|\frac{q_2}{p_2}\right|\left(1+\left|\frac{p_1}{q_1}\right|^2\right) - \left|\frac{p_2}{q_2}\right|\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right]\right\} \\ & + \frac{1}{2}\left[\left|\frac{q_2}{p_2}\right|^2\left(1+\left|\frac{p_1}{q_1}\right|^2\right) + \left|\frac{p_2}{q_2}\right|^2\left(1+\left|\frac{q_1}{p_1}\right|^2\right)\right]|g_{-,2}(t_2)|^2 \\ & + \frac{1}{2}\left(\left|\frac{q_1}{p_1}\right|^2 + \left|\frac{p_1}{q_1}\right|^2 + 2\right)|g_{+,2}(t_2)|^2. \end{aligned} \tag{2.25}$$

Again, if the decay occurs via $M_1^0 \to \bar{M}_2^0$ instead of $M_1^0 \to M_2^0$, the results can be directly obtained from (2.16) - (2.25) by replacing p_2/q_2 with q_2/p_2.

Category 3

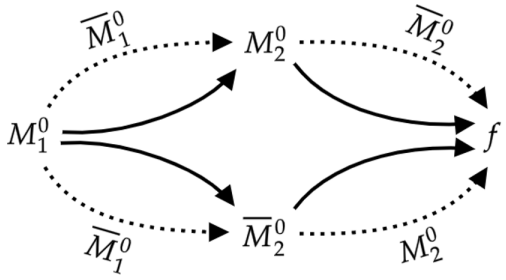


Figure 3. Eight oscillation paths in the cascade decay $M_1^0 \to M_2 \to f$, with $M_1^0 \to M_2^0, \bar{M}_2^0$, $\bar{M}_2^0 \to f, \bar{f}$ and their CP conjugate processes being allowed.

In the third category, M_1^0 can decay into both M_2^0 and $\bar{M}_2^0$, so the number of possible paths for $M_1^0 \to M_2 \to f$ is doubled compared with the second category, as shown in **Figure 3**. To evaluate the CP asymmetry of such a process, the following parameters need to be introduced,

$$\frac{\langle M_2^0|M_1^0\rangle}{\langle \bar{M}_2^0|M_1^0\rangle} = r_1 e^{i(\delta_1+\theta_1)}, \qquad \frac{\langle f|\bar{M}_2^0\rangle}{\langle f|M_2^0\rangle} = r_2 e^{i(\delta_2+\theta_2)}, \qquad \frac{\langle M_2^0|\bar{M}_1^0\rangle}{\langle \bar{M}_2^0|M_1^0\rangle} = e^{i\theta_3}, \tag{2.26}$$

where $\delta_{1,2}$ are the relative strong phases, $\theta_{1,2,3}$ are the relative weak phases, and $r_{1,2}$ are the magnitude ratios. All complete results can be found in Appendix 1, where (A.3) and (A.5) correspond to the double-mixing CP violation terms. This is the most complex and comprehensive scenario, and by making appropriate approximations for this category, we can derive the complete results of the preceding two categories.

3 Numerical analysis

In this section, we will conduct numerical analyses of the double-mixing CP violation for various channels. As B_d^0 and B_s^0 have very different oscillation rates, we will discuss their decays separately. The input parameters involved are detailed in Table 1.

3.1 B_d^0 decays

In this subsection, we discuss the double-mixing CP violation in cascade B_d^0 decays. For the B_d^0 meson, it is a reasonable approximation to assume that the decay width difference, denoted as $\Delta\Gamma_{B_d}$, is negligible [38, 39]. Consequently, all terms involving $\sinh \Delta\Gamma_1 t_1/2$ in (2.4) can be disregarded while $\cosh \Delta\Gamma_1 t_1/2$ effectively equals 1. Furthermore, the magnitude of ratio q_{B_d}/p_{B_d} is approximated to be 1 [38, 39]. Under these approximations, the double-mixing CP violation A_{dm} is simplified to $A_{n,dm}$. Here, $A_{n,dm}$ represents the component of CP violation induced by double-mixing interference and it is proportional to

Parameter	Value	Parameter	Value
$\lvert q_{B_d}/p_{B_d}\rvert$	1.0010 ± 0.0008 [39]	x_{B_d}	0.769 ± 0.004 [38]
$\lvert q_{B_s}/p_{B_s}\rvert$	1.0003 ± 0.0014 [39]	x_{B_s}	27.03 ± 0.09 [38]
$\lvert q_K/p_K\rvert$	0.996774 ± 0.000019 [38]	x_K	0.946 ± 0.002 [38]
$\lvert q_D/p_D\rvert$	0.995 ± 0.016 [39]	x_D	$(4.1 \pm 0.5) \times 10^{-3}$ [39]
ϕ_{B_d}	$(44.4 \pm 1.4)^\circ$ [38]	y_{B_d}	-0.0005 ± 0.0050 [38]
ϕ_{B_s}	$-(2.106 \pm 0.135)^\circ$ [38]	y_{B_s}	-0.064 ± 0.003 [38]
ϕ_K	$(0.176 \pm 0.001)^\circ$ [38]	y_K	-0.996506 ± 0.000016 [38]
ϕ_D	$-(177.5 \pm 1.2)^\circ$ [38]	y_D	$(6.2 \pm 0.6) \times 10^{-3}$ [39]
ω_{B_d}	$\left(0.00372^{+0.00025}_{-0.00023}\right)^\circ$ [38]	ω_{B_s}	$\left(-0.070^{+0.004}_{-0.005}\right)^\circ$ [38]
ω_K	$(43.52 \pm 0.05)^\circ$ [38]	ω_D	$\left(83.4446^{+32.2995}_{-36.5375}\right)^\circ$ [38]
ω_1	0° [38]	ω_2	$(65.54 \pm 1.55)^\circ$ [38]
ω_3	0° [38]	ω_4	$(-114.419 \pm 1.549)^\circ$ [38]
r_D	$-(5.857 \pm 0.017) \times 10^{-2}$ [39]	δ_D	$\left(7.2^{+7.9}_{-9.2}\right)^\circ$ [39]

Table 1. The input parameters and their values, with $x_M \equiv \Delta m_M/\Gamma_M$ and $y_M \equiv \Delta\Gamma_M/(2\Gamma_M)$, respectively.

$\sin \Delta m_1 t_1$. This simplification aids in a more straightforward analysis of the double-mixing CP violation effect in B_d^0 decays.

3.1.1 $B_d^0 \to J/\psi K \to J/\psi(\pi^\pm \ell^\mp \nu)$

We initially explore the decay channel $B_d^0 \to J/\psi K \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$. The decay mode $B_d^0 \to J/\psi K^0$, corresponding to the $b \to c\bar{c}s$ transition at the quark level, is frequently employed for the extraction of the CKM angle β because of the tiny pollution of strong interaction and the relatively substantial CP violation effect [38]. The cascading mixing effect in this decay channel where the secondary decay is semileptonic has been previously investigated in paper [40].

In this decay channel, only the paths $B_d^0 \to J/\psi K^0 \to J/\psi \bar{K}^0 \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$ and $B_d^0 \to \bar{B}_d^0 \to J/\psi \bar{K}^0 \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$ can contribute. This case aligns with the first category defined previously. As a reasonable approximation, we have neglected the direct CP violation in neutral K meson decay, assuming $\langle \pi^+\ell^-\bar{\nu}_\ell | \bar{K}^0 \rangle = \langle \pi^-\ell^+\nu_\ell | K^0 \rangle$. The primary decay amplitudes are related as $\langle J/\psi \bar{K}^0 | \bar{B}_d^0 \rangle = -\langle J/\psi K^0 | B_d^0 \rangle e^{i\omega_{B_d}}$, where ω_{B_d} is a pure weak phase, with the latest result given by $\arg(V_{cb}V_{cs}^\star/V_{cb}^\star V_{cs}) = \left(0.00372^{+0.00025}_{-0.00023}\right)^\circ$ [38]. It is important to note that, the minus sign originates from the convention $CP\left|J/\psi K^0\right\rangle = -\left|J/\psi \bar{K}^0\right\rangle$. Then, all the components in (2.4) are calculated as:

$$A_{n,dm}(t_1,t_2) = \frac{e^{-\left(\Gamma_{B_d}t_1+\Gamma_K t_2\right)}}{2D(t_1,t_2)}\Bigg\{\sinh\frac{\Delta\Gamma_K t_2}{2}\sin\Phi_1\left(\left|\frac{p_K}{q_K}\right| + \left|\frac{q_K}{p_K}\right|\right) - \sin\Delta m_K t_2 \cos\Phi_1\left(\left|\frac{q_K}{p_K}\right| - \left|\frac{p_K}{q_K}\right|\right)\Bigg\}\sin\Delta m_{B_d} t_1, \tag{3.1}$$

t_1/τ_{B_d}	t_2/τ_K	$A_{\rm CP}$
0∼3	0∼ 2	40.52%
0∼2	0∼ 2	39.76%
0∼3	0∼ 1	42.09%
0∼2	0∼ 1	48.04%
0∼3	0∼ 0.5	29.28%
0∼2	0∼ 0.5	38.76%

Table 2. The time integrated $A_{\rm CP}$ in $B_d^0 \to J/\psi K \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$.

$$A_{non-dm}(t_1,t_2) = \frac{1}{D(t_1,t_2)}\left(\left|\frac{q_K}{p_K}\right|^2 - \left|\frac{p_K}{q_K}\right|^2\right)|g_{-,K}(t_2)|^2|g_{+,B_d}(t_1)|^2, \tag{3.2}$$

where Φ_1 is defined as $\omega_{B_d} - \phi_{B_d} + \phi_K$, approximated as -2$\beta$. The denominator $D(t_1,t_2)$ is given by (2.4) with the t_2 functions substituted with (2.11) - (2.13). Hence, by measuring the double-mixing CP violation of this channel, the β angle can be extracted. The double-mixing CP violation $A_{n,dm}(t_1,t_2)$ is induced by the interference between the decay paths $B_d^0 \to J/\psi K^0 \to J/\psi\bar{K}^0 \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$ and $B_d^0 \to \bar{B}_d^0 \to J/\psi\bar{K}^0 \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$, corresponding to the $S_n(t_2)$ component in (2.4). Meanwhile, $A_{non-dm}(t_1,t_2)$ corresponds to the $C_+(t_2)$ term within $N(t_1,t_2)$ as defined in (2.4).

To investigate the double-mixing CP violation effect, we conducted a numerical analysis of both the total CP asymmetry $A_{\rm CP}$ and the double-mixing CP asymmetry A_{dm}. Notably, A_{dm} plays a crucial role in $A_{\rm CP}$, as evidenced by the mathematical formula in (3.1) and the corresponding numerical results in **Figure 4**. The left panel clearly shows that the maximum value of the two-dimensional time-dependent double-mixing CP asymmetry can even exceed 50%. When integrating t_2 from 0 to $2\tau_K$, as depicted in the middle panel, the double-mixing effect almost overlaps with the total CP asymmetry $A_{\rm CP}$, indicating its dominance. Its magnitude initially increases with t_1 and then decreases, a pattern attributed to its sine dependence on t_1. Within an interval of $4\tau_{B_d}$, it completes only half an oscillation period, consistent with the small value of x_{B_d} listed in Table 1. Similar trends are observed when considering the dependence on t_2, as shown in the right panel, where t_1 is integrated from 0 to $3\tau_{B_d}$. It rapidly increases within a period shorter than τ_K, and then decreases before stabilizing. This phenomenon can be explained by the fact that the numerator is mainly dominated by the first term in (3.1), which increases from 0 to τ_K and then stabilizes. and that the denominator is dominated by $C'_+(t_2)$ and $C'_-(t_2)$, which decreases from 0 to $0.5\tau_K$ and then steadily increases until around $t_2 \approx 2.5\tau_K$. A comprehensive two-dimensional time integration of $A_{\rm CP}(t_1,t_2)$ yields the results presented in Table 2. For experimental convenience, we perform the integration over different t_1 and t_2 intervals. The integrated $A_{\rm CP}$ fluctuates around 30-40% with different configurations.

The yield of $B_d^0 \to J/\psi K_S^0$ has been previously observed at LHCb [41] using data from Run 1 of the LHC at a center-of-mass energy $\sqrt{s} = 7\,\text{TeV}$ with an integrated luminosity 3 fb^{-1}. This observation involved the final states of $J/\psi \to \mu^+\mu^-$ and $K_S^0 \to \pi^+\pi^-$. Based

on this study, we estimate the yield of $B_d^0 \to J/\psi K \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$ with 300 fb^{-1} to be approximately 3.6×10^4.

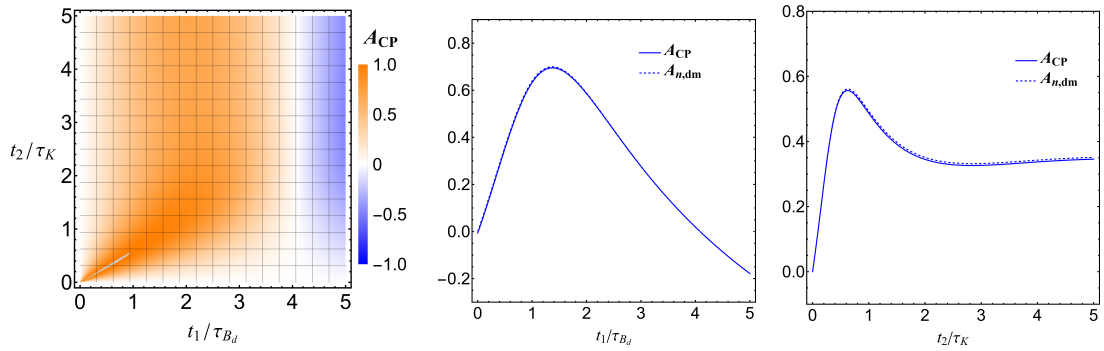


Figure 4. Time dependence of the double-mixing CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to J/\psi K(t_2) \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

Substituting the semileptonic final state with $\pi^-\ell^+\nu_\ell$ instead of $\pi^+\ell^-\bar{\nu}_\ell$ resulting in two viable decay paths, denoted as $B_d^0 \to J/\psi K^0 \to J/\psi(\pi^-\ell^+\nu_\ell)$ and $B_d^0 \to \bar{B}_d^0 \to J/\psi \bar{K}^0 \to J/\psi K^0 \to J/\psi(\pi^-\ell^+\nu_\ell)$. The corresponding CP asymmetries are then calculated to be

$$A_{n,dm}(t_1,t_2) = \frac{e^{-(\Gamma_{B_d}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\sin\Phi_1\left(\left|\frac{p_K}{q_K}\right|+\left|\frac{q_K}{p_K}\right|\right) - \sin\Delta m_K t_2\cos\Phi_1 \times\left(\left|\frac{q_K}{p_K}\right|-\left|\frac{p_K}{q_K}\right|\right)\Big\}\sin\Delta m_{B_d}t_1, \tag{3.3}$$

$$A_{non-dm}(t_1,t_2) = \frac{1}{D(t_1,t_2)}\left(-\left|\frac{q_K}{p_K}\right|^2+\left|\frac{p_K}{q_K}\right|^2\right)|g_{-,K}(t_2)|^2|g_{-,B_d}(t_1)|^2. \tag{3.4}$$

The denominator $D(t_1,t_2)$ here can be obtained by combining (2.4) with (2.11) - (2.13), while also performing the substitution $C'_+(t_2) \leftrightarrow C'_-(t_2)$, and adding an extra negative sign in front of $S'_n(t_2)$.

Compared to the final state $\pi^+\ell^-\bar{\nu}_\ell$, a key difference for the final state $\pi^-\ell^+\nu_\ell$ lies in an additional negative sign in the numerator of the non-double-mixing CP violation term $A_{non-dm}(t_1,t_2)$, as illustrated in **Figure 5**. The double-mixing CP violation remains significant, with the peak occurring later compared to the decay process with the final state $\pi^+\ell^-\bar{\nu}_\ell$. For instance, the peak in the middle panel of **Figure 5** occurs at around $t_1 = 3\tau_{B_d}$, in contrast to about $1.5\tau_{B_d}$ for the other final state. This difference is due to an additional negative sign in the term $S'_n(t_2)$ in (2.4). Moreover, the magnitude of CP violation increases significantly over time, reflecting the later occurrence of its peak in both t_1 and t_2.

3.1.2 $B_d^0 \to J/\psi K \to J/\psi f_+$

We now analyze the double-mixing CP violation in the decay chain $B_d^0 \to J/\psi K \to J/\psi f_+$. In this context, f_+ refers to CP-even final states, such as $\pi^+\pi^-$. Notably, the CP-even state under consideration is common to K^0 and $\bar{K}^0$. This will lead to four conceivable

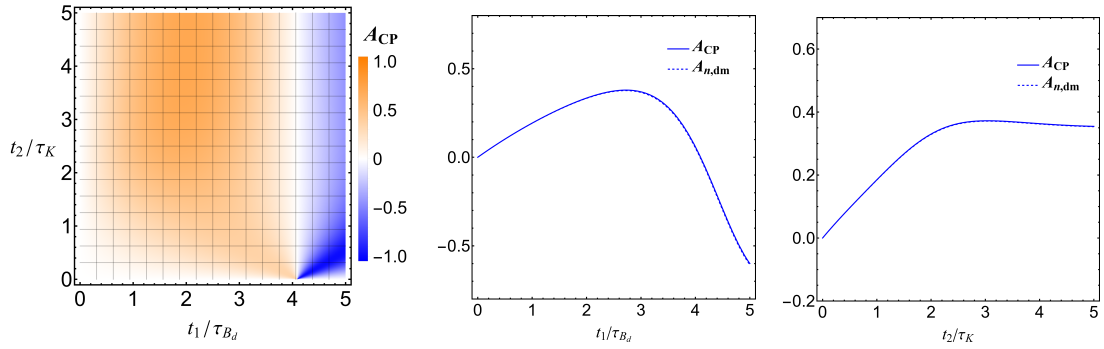


Figure 5. Time dependence of the double-mixing CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to J/\psi K(t_2) \to J/\psi(\pi^-\ell^+\nu_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

decay paths: $B_d^0 \to K^0 \to f_+$, $B_d^0 \to K^0 \to \bar{K}^0 \to f_+$, $B_d^0 \to \bar{B}_d^0 \to \bar{K}^0 \to f_+$ and $B_d^0 \to \bar{B}_d^0 \to \bar{K}^0 \to K^0 \to f_+$. Hence, it belongs to the second category in our classification. Considering the negligible direct CP violation in the neutral K meson($\langle f_+|K^0\rangle = \langle f_+|\bar{K}^0\rangle$) and the suppressed penguin amplitudes in $B_d^0 \to J/\psi K^0$ (both Cabbibo-suppressed and loop-suppressed as indicated in [38]), we obtain all components of the two-dimensional time-dependent CP asymmetry $A_{\rm CP}(t_1, t_2)$ in (2.4) as follows:

$$\begin{aligned} A_{n,dm}(t_1,t_2) = \frac{e^{-(\Gamma_{B_d}t_1+\Gamma_K t_2)}}{D(t_1,t_2)} \Bigg\{ & \sin\Delta m_K t_2 \left(\left|\frac{q_K}{p_K}\right| - \left|\frac{p_K}{q_K}\right| \right) \cos\Phi_1 \\ & - \sinh\frac{\Delta\Gamma_K t_2}{2} \left(\left|\frac{q_K}{p_K}\right| + \left|\frac{p_K}{q_K}\right| \right) \sin\Phi_1 \Bigg\} \sin\Delta m_{B_d} t_1, \end{aligned} \tag{3.5}$$

$$\begin{aligned} A_{non-dm}(t_1,t_2) = \frac{2e^{-\Gamma_{B_d}t_1}}{D(t_1,t_2)} \Big\{ & \sin(\phi_{B_d} - \omega_{B_d}) |g_{+,K}(t_2)|^2 \\ & + \sin(\phi_{B_d} - \omega_{B_d} + 2\phi_K) |g_{-,K}(t_2)|^2 \Big\} \sin\Delta m_{B_d} t_1. \end{aligned} \tag{3.6}$$

In this situation, $A_{n,dm}$ and A_{non-dm} constitute S_n in (2.4). Notably, $\Phi_1 \approx -2\beta$, $\phi_{B_d} - \omega_{B_d} \approx 2\beta$ and $\phi_K \approx 0$. Under the approximation that $|q_K/p_K| \approx 1$, $S_n(t_2)$ simplifies to

$$S_n(t_2) = 2\sin 2\beta e^{-\Gamma_S t_2}. \tag{3.7}$$

By taking $q_K/p_K \to 1$, we neglect the indirect CP violation of neutral kaon, so only $K_S^0 \to f_+$ contributes to the decay and it explains why only the $e^{-\Gamma_S t_2}$ term survives. As a result, the total CP asymmetry $A_{\rm CP}$ is expressed as

$$A_{\rm CP}(t_1,t_2) = \sin 2\beta \cdot \sin\Delta m_{B_d} t_1, \tag{3.8}$$

which is consistent with the formulas (13.74) and (13.82) in the " CP Violation in the Quark Sector " review of the PDG [38].

We conducted a numerical analysis of $A_{\rm CP}$ and A_{dm} for this process. The two-dimensional time-dependent CP violation result is shown in the left panel of **Figure 6**,

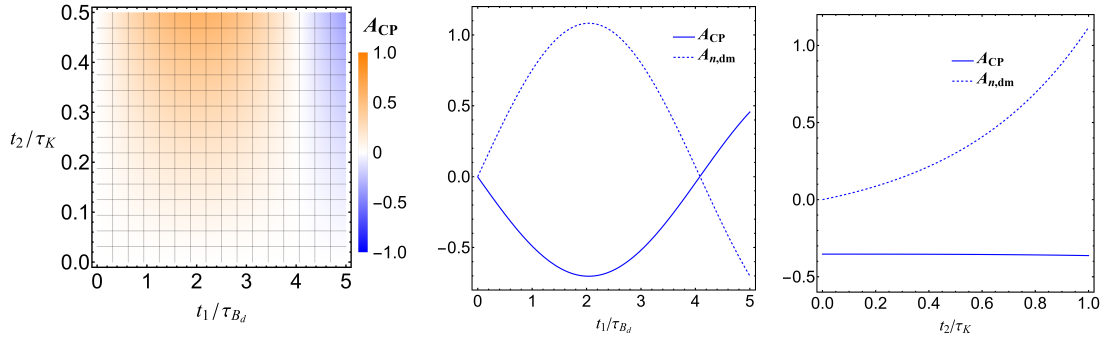


Figure 6. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to J/\psi K(t_2) \to J/\psi(\pi^+\pi^-)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

t_1/τ_{B_d}	t_2/τ_K	$A_{\rm CP}$
0∼3	0∼ 2	-35.36%
0∼2	0∼ 2	-32.35%
0∼3	0∼ 1	-35.30%
0∼2	0∼ 1	-32.28%
0∼3	0∼ 0.5	-35.31%
0∼2	0∼ 0.5	-32.27%

Table 3. Time integrated $A_{\rm CP}$ in $B_d^0 \to J/\psi K \to J/\psi(\pi^+\pi^-)$.

with its peak appearing around $t_1 = 2\tau_{B_d}$ and $t_2 > 0.5\tau_K$. By integrating t_2 from 0 to $2\tau_K$, the t_1-dependent result, displayed in the middle panel, reaches its peak at $t_1 = 2\tau_{B_d}$. Since both $A_{\rm CP}$ and $A_{n,dm}$ exhibit sine dependence on t_1, their oscillations are similar, but with different magnitudes and signs. It can be concluded that the double-mixing contribution and the non-double-mixing contribution have opposite signs, with the former having a smaller magnitude. The right panel of **Figure 6** shows that $A_{\rm CP}$ with t_1 integrated out is almost independent of t_2. On the other hand, the $A_{n,dm}$ part, with its hyperbolic sine dependence on t_2, increases with t_2 ranging from 0 to τ_K. The two-dimensional time integration of $A_{\rm CP}(t_1, t_2)$ for this channel is given in Table 3, demonstrating its insensitivity to changes in both t_1 and t_2, owing to the fact that most of the decays occur earlier.

3.1.3 $B_d^0 \to DK \to (K^-\pi^+)(\pi^+\ell^-\bar{\nu}_\ell)$

The decay $B_d^0 \to DK^\star$ is well-known for its precision in measuring the angle γ. First attempts to constraint γ using $B_d^0 \to DK^\star$ decays have adopted neutral D meson decays into $K_S^0\pi^+\pi^-$ [42] and suppressed final states such as $K^-\pi^+$, with relevant branching ratio and parameters measured in several studies [43–45]. Here, we consider a similar process $B_d^0 \to DK$, which was also briefly discussed in our previous work [37]. This channel, comprising two oscillating paths, $B_d^0 \to D^0K^0 \to D^0\bar{K}^0$ and $B_d^0 \to \bar{B}_d^0 \to D^0\bar{K}^0$, falls within the first

t_1/τ_{B_d}	t_2/τ_K	$A_{\rm CP}$
0∼3	0∼ 2	26.36%
0∼2	0∼ 2	31.28%
0∼3	0∼ 1	15.68%
0∼2	0∼ 1	19.79%
0∼3	0∼ 0.5	8.78%
0∼2	0∼ 0.5	11.52%

Table 4. Time integrated $A_{\rm CP}$ in $B_d^0 \to DK \to (K^-\pi^+)(\pi^+\ell^-\bar{\nu}_\ell)$.

category. The following relations are employed to evaluate the CP asymmetry,

$$\frac{A\big(\bar{B}_d^0 \to D^0\bar{K}^0\big)}{A\big(B_d^0 \to \bar{D}^0 K^0\big)} = e^{i\omega_1}, \quad \omega_1 = \arg\frac{V_{cb}V_{us}^\star}{V_{cb}^\star V_{us}}, \tag{3.9}$$

$$\frac{A\big(B_d^0 \to D^0 K^0\big)}{A\big(B_d^0 \to \bar{D}^0 K^0\big)} = r_B e^{i(\delta_B+\omega_2)}, \quad \omega_2 = \arg\frac{V_{ub}^\star V_{cs}}{V_{cb}^\star V_{us}}, \tag{3.10}$$

where r_B and δ_B are the magnitudes and the strong phase of the ratio, respectively. The parameters for the decay channel $B_d^0 \to DK^\star$ have been determined by the LHCb experiment [46, 47], and we take similar values for $B_d^0 \to DK$: $r_B = 0.366$ and $\delta_B = 164°$. We neglect the direct CP violation in neutral K meson decay, *i.e.*, $\langle f|\bar{K}^0\rangle = \langle \bar{f}|K^0\rangle$, as well as the amplitude of Cabibbo-suppressed process $\bar{D}^0 \to K^-\pi^+$ and the mixing of the neutral D meson. The double-mixing term is

$$A_{n,dm}(t_1,t_2) = \frac{e^{-\Gamma_{B_d}t_1}}{D(t_1,t_2)} S_n(t_2)\sin\Delta m_{B_d}t_1, \tag{3.11}$$

$$\begin{aligned} S_n(t_2) = e^{-\Gamma_K t_2} r_B \Big\{ &\left|\frac{q_K}{p_K}\right| [\sin(\delta_B+\delta_\omega)\sinh\frac{1}{2}\Delta\Gamma_K t_2 + \cos(\delta_B+\delta_\omega)\sin\Delta m_K t_2] \\ &- \left|\frac{p_K}{q_K}\right| [\sin(\delta_B-\delta_\omega)\sinh\frac{1}{2}\Delta\Gamma_K t_2 + \cos(\delta_B-\delta_\omega)\sin\Delta m_K t_2]\Big\}, \end{aligned} \tag{3.12}$$

where $\delta_\omega = \omega_2 - \omega_1 + \phi_{B_d} - \phi_K$, approximating $2\beta+\gamma$. Applying (2.11) - (2.13) yields the result of $D(t_1,t_2)$, with $C'_+(t_2)$ has an additional factor of r_B^2, and $S'_n(t_2)$ can be obtained by changing the negative sign in front of $|p_K/q_K|$ in (3.12) to a positive sign.

Our numerical analysis addresses both the total CP asymmetry $A_{\rm CP}$ and the double-mixing CP asymmetry A_{dm}. As illustrated in the left panel of **Figure 7**, the maximum value of two-dimensional time-dependent double-mixing results can surpass 50% when t_1 is small. From the t_1 and t_2 dependence displayed in the middle and right panels, respectively, it can be seen that the double-mixing effect dominates. This trend resembles that observed in $B_d^0 \to J/\psi K \to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$, albeit with differences attributed to the suppression factor r_B and a negative sign in δ_ω. With integrating t_1 from 0 to $3\tau_{B_d}$, a steady growth of $A_{\rm CP}$ with respect to t_2 is observed. This is supported by the time-integrated $A_{\rm CP}$ presented in **Table 4**, which shows sensitivity to changes in the upper limit of t_2.

Given that the weak phase involved in this decay $\delta_\omega \approx 2\beta + \gamma$, its extraction holds significant importance. It should be emphasized that the weak phase can be directly extracted from experimental data, as discussed in our previous study [37].

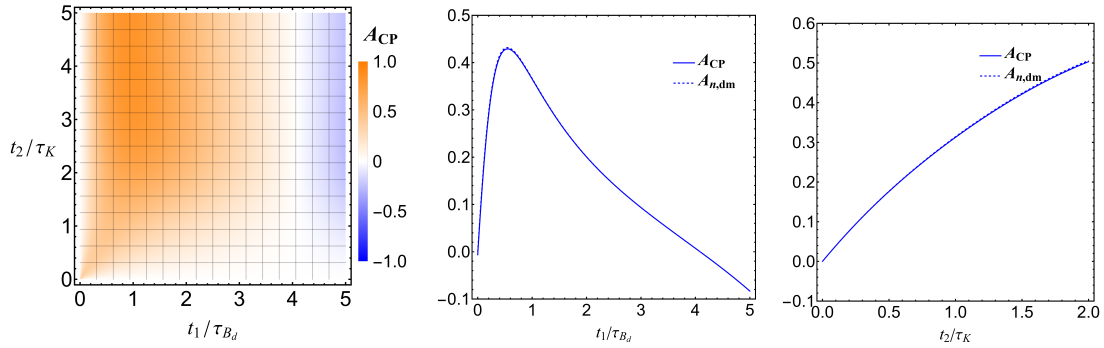


Figure 7. Time dependence of the double-mixing CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to DK(t_2) \to (K^-\pi^+)(\pi^+\ell^-\bar{\nu}_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

3.1.4 $B_d^0 \to DK_S^0 \to f_+ K_S^0$

Distinct from $B_d^0 \to DK \to (K^-\pi^+)(\pi^+\ell^-\bar{\nu}_\ell)$ discussed previously in section 3.1.3, when considering the CP-even final state like $f_+ = \pi^+\pi^-$ of D in $B_d^0 \to DK_S^0 \to f_+ K_S^0$, both intermediate D^0 and $\bar{D}^0$ contributions are comparable. With primary $b \to u$ and $b \to c$ transitions, both B_d^0 and $\bar{B}_d^0$ can decay into states involving either D^0 or $\bar{D}^0$. These make this channel a particularly intricate case, categorizing it within the third classification.

The terms related to the double-mixing CP violation are expressed as below:

$$\begin{aligned}
A_{n,dm,1}(t_1,t_2) = f(t_1,t_2)e^{-\Gamma_D t_2}\sinh\frac{1}{2}\Delta\Gamma_D t_2\Big\{ &-\left(\left|\frac{p_D}{q_D}\right| + \left|\frac{q_D}{p_D}\right|\right)\sin(\omega_1 - \phi_{B_d} - \phi_D) \\
&- r_B\left[\left|\frac{p_D}{q_D}\right|\sin(\delta_B + \Phi_2) - \left|\frac{q_D}{p_D}\right|\sin(\delta_B - \Phi_2)\right. \\
&\left.+\left|\frac{p_D}{q_D}\right|\sin(\delta_B + \Phi_3) - \left|\frac{q_D}{p_D}\right|\sin(\delta_B - \Phi_3)\right] \\
&- r_B^2\left(\left|\frac{p_D}{q_D}\right| + \left|\frac{q_D}{p_D}\right|\right)\sin(\Phi_2 - \omega_2)\Big\},
\end{aligned} \tag{3.13}$$

$$\begin{aligned}
A_{n,dm,2}(t_1,t_2) = f(t_1,t_2)e^{-\Gamma_D t_2}\sin\Delta m_D t_2\Big\{ &\left(\left|\frac{p_D}{q_D}\right| - \left|\frac{q_D}{p_D}\right|\right)\cos(\omega_1 - \phi_{B_d} - \phi_D) \\
&- r_B\left[\left|\frac{p_D}{q_D}\right|\cos(\delta_B + \Phi_2) - \left|\frac{q_D}{p_D}\right|\cos(\delta_B - \Phi_2)\right. \\
&\left.-\left|\frac{p_D}{q_D}\right|\cos(\delta_B + \Phi_3) + \left|\frac{q_D}{p_D}\right|\cos(\delta_B - \Phi_3)\right] \\
&- r_B^2\left(\left|\frac{p_D}{q_D}\right| - \left|\frac{q_D}{p_D}\right|\right)\cos(\Phi_2 - \omega_2)\Big\},
\end{aligned} \tag{3.14}$$

where $\Phi_2 = \omega_1 - \omega_2 - \phi_{B_d} + \phi_D$, $\Phi_3 = \omega_1 - \omega_2 - \phi_{B_d} - \phi_D$ and $f(t_1, t_2) = e^{-\Gamma_{B_d} t_1} \sin \Delta m_{B_d} t_1 / D(t_1, t_2)$, and the double-mixing CP violation is the sum of these two contributions. Combining (2.4) and (2.22) - (2.25) yields the denominator $D(t_1, t_2)$. The parameters r_B and δ_B have been defined in (3.9).

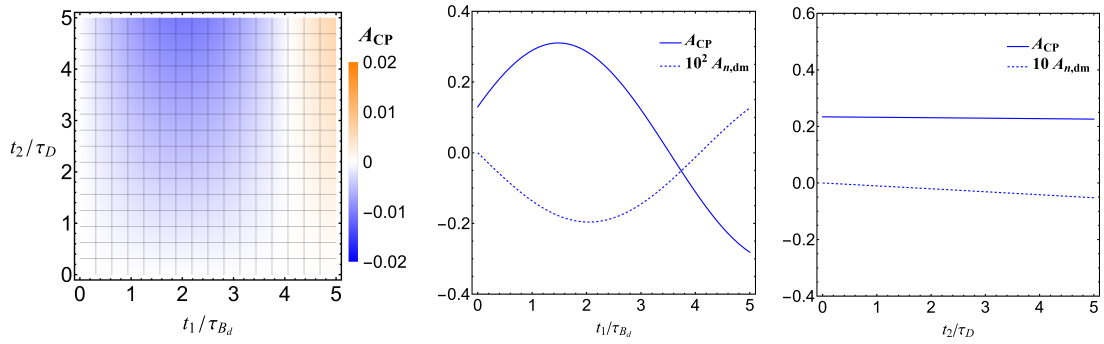


Figure 8. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to D(t_2) K_S^0 \to (\pi^+\pi^-) K_S^0$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $5\tau_D$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

We perform the numerical analysis assuming the parameter values $r_B = 0.366$ and $\delta_B = 164^\circ$ as in section 3.1.3. The results are depicted in **Figure 8**. On the left panel, the two-dimensional time-dependent CP violation exhibits a distribution akin to the decay channel $B_d^0 \to J/\psi K \to J/\psi(\pi^+\pi^-)$, but with peak values significantly lower. Integrating t_2 from 0 to $5\tau_D$, we present the t_1-dependent results in the middle panel, revealing tiny double-mixing CP violation. This is primarily attributed to the small magnitudes of x_D and y_D, both approximately 10^{-3}. These small values reduces both $\sinh \frac{1}{2}\Delta\Gamma_D t_2$ and $\sin \Delta m_D t_2$. The right panel of **Figure 8** shows the t_2 dependence of $A_{\rm CP}$ when integrating t_1 from 0 to $3\tau_{B_d}$. The near-linear dependency of A_{dm} on t_2 is primarily due to its denominator being dominated by the $\cosh y_D t_2$ term and its numerator by the $\sinh x_D t_2$ term. It is concluded that the CP asymmetry in this channel is dominated by the part induced by the interference between the $b \to c$ and $b \to u$ amplitudes, and the part induced by the interference between decays with and without B mixing, while the double-mixing CP violation is negligible owing to tiny D-mixing effect.

3.1.5 $B_d^0 \to D K_S^0 \to (K^-\pi^+) K_S^0$

The last B_d^0 decay channel selected is $B_d^0 \to D K_S^0 \to (K^-\pi^+) K_S^0$. The observation of a similar channel $\bar{B}_d^0 \to D^0 \bar{K}^{\star 0}$ was reported by [43, 44], but the branching ratio and relevant parameters were measured using self-tagging methods and thus the mixing effect was not considered. In this channel, the intermediate D meson can be either D^0 or $\bar{D}^0$, although the latter has a suppressed contribution. Therefore, this channel is categorized as the third type.

To present the result of the CP asymmetry, we introduce the amplitude ratio of the

t_1/τ_{B_d}	t_2/τ_D	$A_{\rm CP}$
0∼3	0∼ 5	-60.71%
0∼2	0∼ 5	-69.20%
0∼3	0∼ 4	-60.71%
0∼2	0∼ 4	-69.20%

Table 5. Two-dimensional time integration of $A_{\rm CP}(t_1, t_2)$ in $B_d^0(t_1) \to D(t_2)K_S^0 \to (K^-\pi^+)K_S^0$.

primary decay as in (3.9) and the ratio of the secondary decay as

$$\frac{A(\bar{D}^0 \to K^-\pi^+)}{A(D^0 \to K^-\pi^+)} = -r_D e^{-i\delta_D}, \tag{3.15}$$

where the numerical values for magnitude ratio r_D and the strong phase δ_D are listed in Table 1. The direct CP violation in the D meson decays is neglected due to the minimal weak phases between the first two generations of quarks. The double-mixing term in (2.4) is expressed as

$$\begin{aligned} A_{n,dm}(t_1,t_2) = &- \frac{e^{-(\Gamma_{B_d}t_1+\Gamma_D t_2)}}{2D(t_1,t_2)} \left[\left(\left|\frac{q_D}{p_D}\right| + \left|\frac{p_D}{q_D}\right| \right) \sin(\omega_1 - \phi_{B_d} - \phi_D) \sinh\frac{1}{2}\Delta\Gamma_D t_2 \right. \\ &\left. + \left(\left|\frac{q_D}{p_D}\right| - \left|\frac{p_D}{q_D}\right| \right) \cos(\omega_1 - \phi_{B_d} - \phi_D) \sin\Delta m_D t_2 \right] \sin\Delta m_{B_d} t_1 \ , \end{aligned} \tag{3.16}$$

where the contributions below $\mathcal{O}(10^{-3})$ have been neglected. It turns out that the corrections from the doubly Cabibbo-suppressed $D^0 \to K^+\pi^-$ decay are negligible, so practically only the interference term between $B_d^0 \to \bar{D}^0 \to D^0$ and $B_d^0 \to \bar{B}_d^0 \to D^0$ is retained. With also the contributions below $\mathcal{O}(10^{-3})$ neglected, the non-double-mixing CP asymmetry reads

$$\begin{aligned} A_{non-dm,1}(t_1,t_2) = r_B \Big\{ &\left(\left|\frac{p_D}{q_D}\right| \cos(\delta_B + \omega_2 - \phi_D) - \left|\frac{q_D}{p_D}\right| \cos(\delta_B - \omega_2 + \phi_D) \right) \sinh\frac{1}{2}\Delta\Gamma_D t_2 \\ &+ \left(\left|\frac{p_D}{q_D}\right| \sin(\delta_B + \omega_2 - \phi_D) - \left|\frac{q_D}{p_D}\right| \sin(\delta_B - \omega_2 + \phi_D) \right) \sin\Delta m_D t_2 \\ &+ 2r_D \sin(\delta_B - \delta_D) \sin\omega_2 \left(\cosh\frac{1}{2}\Delta\Gamma_D t_2 + \cos\Delta m_D t_2 \right) \Big\} f(t_1,t_2), \\ A_{non-dm,2}(t_1,t_2) = &- 2\frac{e^{-\Gamma_{B_d}t_1}}{D(t_1,t_2)} \Big\{ r_B \cos\delta_B \sin(\omega_1 - \omega_2 - \phi_{B_d}) + r_D \cos\delta_D \sin(\omega_1 - \phi_{B_d}) \Big\} \\ &\times |g_{+,D}(t_2)|^2 \sin\Delta m_{B_d} t_1, \end{aligned} \tag{3.17}$$

where $f(t_1,t_2) \equiv e^{-\Gamma_D t_2}|g_{+,B_d}(t_1)|^2/4D(t_1,t_2)$. The denominator $D(t_1,t_2)$ can be derived by combining (2.4) with (A.7) - (A.12), while maintaining consistency with all the approximations used in the numerator. The parameters r_B and δ_B have been defined in (3.9). The non-double-mixing CP asymmetry contains r_D in the expression (3.17), indicating that the doubly Cabibbo-suppressed $D^0 \to K^+\pi^-$ contribution is considerable here.

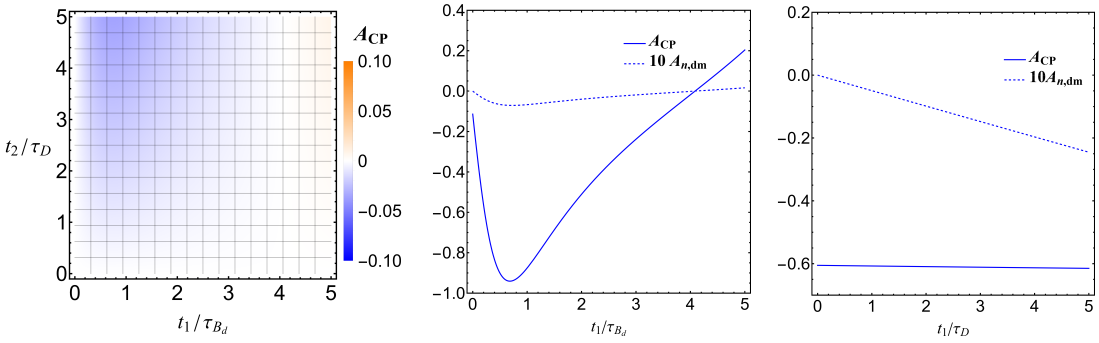


Figure 9. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_d^0(t_1) \to D(t_2)K_S^0 \to (K^-\pi^+)K_S^0$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $5\tau_D$) and t_2 (with t_1 integrated from 0 to $3\tau_{B_d}$), respectively.

We perform numerical analysis on this channel. As depicted in the left panel of **Figure 9**, the peak of the two-dimensional time-dependent double-mixing results is on the order of 10^{-2}, occurring at later t_2 and earlier t_1 times. When integrating t_2 from 0 to $5\tau_D$, the peak value of the double-mixing effect contributes only 0.8% to the total. The dependencies of $A_{\rm CP}$ and A_{dm} on t_1 are similar, differing only in amplitude magnitude. Upon integrating t_1 from 0 to $3\tau_{B_d}$, the t_2 dependence of $A_{\rm CP}$ is shown in the right panel. The dependence of A_{dm} on t_2 is nearly linear, attributed to its nearly hyperbolic sine dependence on t_2. However, due to the small value of y_D, the overall variation of A_{dm} with t_2 remains nearly linear. The double-mixing CP violation in the decay channel of this meson is relatively more significant than in the $\pi^+\pi^-$ final state, due to the presence of the r_D factor which suppresses the dominant non-double-mixing CP violation terms. The dominant contribution to the non-double-mixing CP violation arises from the interference between the $B_d^0 \to \bar{D}^0$ and $B_d^0 \to D^0$ path, *i.e.*, the third line of $A_{non-dm,1}(t_1, t_2)$. A slightly smaller contribution comes from the interference between the $B_d^0 \to D^0$ and $B_d^0 \to \bar{B}_d^0 \to D^0$ paths, *i.e.*, the first term of $A_{non-dm,2}(t_1, t_2)$ which involves the factor r_B. The two-dimensional time integration of $A_{\rm CP}(t_1, t_2)$ for this channel, presented in Table 5, suggests that the CP asymmetry $A_{\rm CP}$ is relatively insensitive to the decay of the D meson.

3.2 B_s^0 decays

In this subsection, we extend our investigation to the B_s^0 meson, by considering decays with the same final states as the B_d^0 decays. For simplicity, we retain the approximation:

$$q_{B_s}/p_{B_s} = e^{-i\phi_{B_s}}. \tag{3.18}$$

Unlike the B_d^0 meson, the B_s^0 meson exhibits a non-negligible difference in decay widths, as indicated in Table 1. Consequently, the double-mixing CP violation for the B_s^0 meson comprises both a component proportional to $\sin \Delta m_{B_s} t_1$, represented by $A_{n,dm}$, and a component proportional to $\sinh \frac{1}{2}\Delta\Gamma_{B_s} t_1$, represented by $A_{h,dm}$. Although the underlying calculation process is similar to that for the B_d^0 meson, the results for the B_s^0 meson are inherently more complicated, implying more fruitful phenomenology.

3.2.1 $B_s^0 \to J/\psi K \to J/\psi f_+$

We first consider the process $B_s^0 \to J/\psi K \to J/\psi f_+$, which originates from the $\bar{b} \to \bar{c}c\bar{d}$ quark-level transition. The first observation of such a mode $B_s^0 \to J/\psi K_S^0$ was reported by CDF along with the measured branching ratio [48]. In this channel, both the transitions $B_s^0 \to J/\psi \bar{K}^0$ and $B_s^0 \to \bar{B}_s^0 \to J/\psi K^0$ can occur, with final state f_+ selected as $\pi^+\pi^-$ being common to both K^0 and $\bar{K}^0$, resulting in four possible paths. This makes these channels fall into the second category of our previously defined classifications.

In the calculation of the CP violation of this channel, direct CP violation in the primary and secondary decays are neglected. The involved phase difference between the decay amplitude $\langle J/\psi K^0 | \bar{B}_s^0 \rangle$ and its charge conjugation $\langle J/\psi \bar{K}^0 | B_s^0 \rangle$ is given by $-e^{i\omega_{B_s}} = -(V_{cb}V_{cd}^\star)/(V_{cb}^\star V_{cd})$, with its numerical value detailed in Table 1. The negative sign arises from the convention $CP\,|J/\psi K^0\rangle = -\,|J/\psi \bar{K}^0\rangle$. The double-mixing CP asymmetry is given by

$$\begin{aligned} A_{n,dm}(t_1,t_2) = 2\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{D(t_1,t_2)}\Big\{ &\sinh\frac{1}{2}\Delta\Gamma_K t_2 \sin\Phi_4\left(\left|\frac{q_K}{p_K}\right| + \left|\frac{p_K}{q_K}\right|\right) \\ &+ \sin\Delta m_K t_2 \cos\Phi_4\left(\left|\frac{q_K}{p_K}\right| - \left|\frac{p_K}{q_K}\right|\right)\Big\} \sin\Delta m_{B_s} t_1, \end{aligned} \tag{3.19}$$

where $\Phi_4 \equiv \omega_{B_s} - \phi_{B_s} - \phi_K$. It is induced by the interference between $B_s^0 \to \bar{K}^0 \to \pi^+\pi^-$ and $B_s^0 \to \bar{B}_s^0 \to K^0 \to \bar{K}^0 \to \pi^+\pi^-$, as well as between the $B_s^0 \to \bar{K}^0 \to K^0 \to \pi^+\pi^-$ and $B_s^0 \to \bar{B}_s^0 \to K^0 \to \pi^+\pi^-$ pathways. The denominator $D(t_1,t_2)$ is obtained from (2.22) - (2.25), where the substitutions $|p_K/q_K| \leftrightarrow |q_K/p_K|$ and $\phi_K \to -\phi_K$ need to be performed.

We have conducted numerical analyses of this decay channel, obtaining the CP asymmetry and its time dependence, as illustrated in **Figure 10**. The left panel shows that the maximum value of the two-dimensional time-dependent CP asymmetry can exceed 50%, a substantial value. Integrating the time t_2 from 0 to $2\tau_K$, the middle panel shows the t_1 dependence of $A_{\rm CP}$. The rapid oscillation observed is due to the large x_s value of the B_s^0 meson. The total CP asymmetry is dominated by the S_n term, which can be further separated into the double-mixing part $A_{n,dm}$ and the non-double-mixing part, both of which have the sine dependence on t_1, as defined in (2.4). The double-mixing part has a magnitude approximately twice the total CP asymmetry but with an opposite sign. Integrating t_1 from 0 to τ_{B_s} reveals the t_2 dependence of $A_{\rm CP}$, as shown in the right panel. In this case, $A_{\rm CP}$ is almost zero because the cancellation between the double-mixing and non-double-mixing terms in S_n, while $A_{n,dm}$ varies approximately as $-\exp(y_K t_2)\sinh y_K t_2$ with respect to t_2.

3.2.2 $B_s^0 \to J/\psi K \to J/\psi(\pi^\pm \ell^\mp \nu_\ell)$

The decay process $B_s^0 \to J/\psi K \to J/\psi(\pi^- \ell^+ \nu_\ell)$ can occur via the two paths $B_s^0 \to J/\psi \bar{K}^0 \to J/\psi K^0 \to J/\psi(\pi^- \ell^+ \nu_\ell)$ and $B_s^0 \to \bar{B}_s^0 \to J/\psi K^0 \to J/\psi(\pi^- \ell^+ \nu_\ell)$; while; the process $B_s^0 \to J/\psi K \to J/\psi(\pi^+ \ell^- \bar{\nu}_\ell)$ can occur via $B_s^0 \to J/\psi \bar{K}^0 \to J/\psi(\pi^+ \ell^- \bar{\nu}_\ell)$ and $B_s^0 \to \bar{B}_s^0 \to J/\psi K^0 \to J/\psi \bar{K}^0 \to J/\psi(\pi^+ \ell^- \bar{\nu}_\ell)$. Therefore, they both align with the first category in our prior discussion.

In the calculation of the CP asymmetry of these channels, we still neglect the direct CP violation in both the primary and secondary decays. Then, we have $\langle f | K^0 \rangle = \langle \bar{f} | \bar{K}^0 \rangle$ and

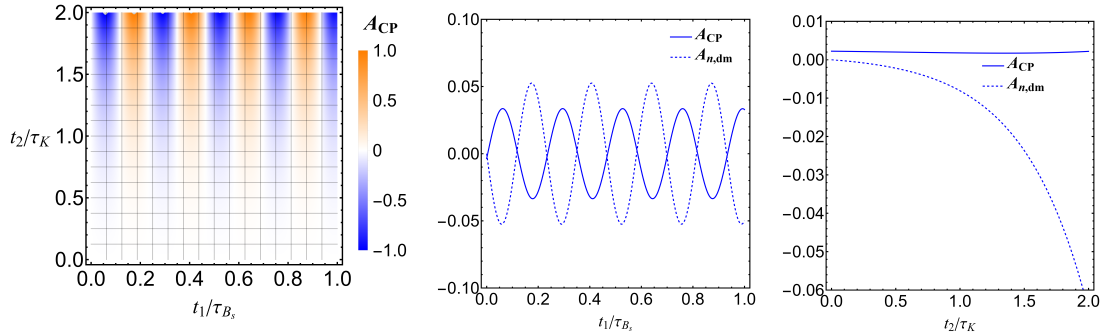


Figure 10. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to J/\psi K \to J/\psi f_+$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to τ_{B_s}), respectively.

the weak phase difference between the involved primary decay amplitudes is ω_{B_s} as introduced in previously. For the decay channel with the final state $\pi^-\ell^+\nu_\ell$, the corresponding double-mixing terms are given by

$$A_{h,dm}(t_1,t_2) = -\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\cos\Phi_4\left(\left|\frac{p_K}{q_K}\right|-\left|\frac{q_K}{p_K}\right|\right) + \sin\Delta m_K t_2\sin\Phi_4\left(\left|\frac{p_K}{q_K}\right|+\left|\frac{q_K}{p_K}\right|\right)\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \tag{3.20}$$

$$A_{n,dm}(t_1,t_2) = -\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{-\sinh\frac{1}{2}\Delta\Gamma_K t_2\left(\left|\frac{p_K}{q_K}\right|+\left|\frac{q_K}{p_K}\right|\right)\sin\Phi_4 + \sin\Delta m_K t_2\cos\Phi_4\left(\left|\frac{p_K}{q_K}\right|-\left|\frac{q_K}{p_K}\right|\right)\Big\}\sin\Delta m_{B_s}t_1. \tag{3.21}$$

Combining (2.4) with (2.11) - (2.13) and making the substitutions $|p_K/q_K| \leftrightarrow |q_K/p_K|$ and $\phi_K \to -\phi_K$ yields the denominator $D(t_1,t_2)$.

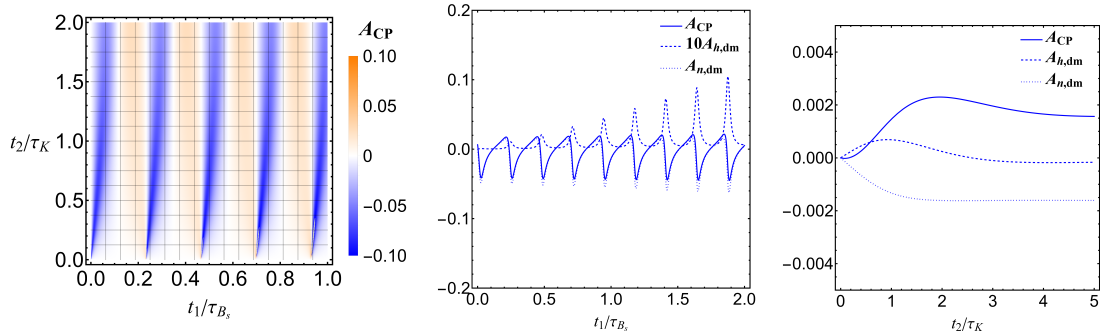


Figure 11. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to J/\psi K \to J/\psi(\pi^-\ell^+\nu_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to τ_K) and t_2 (with t_1 integrated from 0 to $2\tau_{B_s}$), respectively.

The numerical results for the $\pi^-\ell^+\nu_\ell$ channel are illustrated in **Figure 11**. The left panel shows the two-dimensional time dependence of the CP asymmetry, whose maximal values are of $\mathcal{O}(10\%)$. Integrating time t_2 from 0 to τ_K, it is observed from the middle panel that the total asymmetry is dominated by the $A_{n,dm}$ term. This is attributable to the sine dependence on t_1 and hyperbolic sine dependence on t_2 of $A_{n,dm}$, particularly for the large x_s. Conversely, the numerator of $A_{h,dm}$ exhibits a hyperbolic sine dependence on t_1 and a sine dependence on t_2. In the range of t_1 from 0 to $2\tau_{B_s}$, the numerator of $A_{h,dm}$ increases almost linearly while the denominator behaves like a sine function, so $A_{h,dm}$ experiences modulation over time, as reflected in its increasing peak value. Integrating over t_1 from 0 to $2\tau_{B_s}$, we observe the t_2 dependence of $A_{\rm CP}$ in the right panel of **Figure 11**. Now, the contributions of the C term and $A_{n,dm}$ term are comparable, slightly surpassing the $A_{h,dm}$ term.

When the final state is considered as $\pi^+\ell^-\bar{\nu}_\ell$, the results for the double-mixing terms are similar, but with an extra negative sign in the term $A_{h,dm}$. The corresponding expressions are

$$\begin{aligned} A_{h,dm}(t_1,t_2) = &-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\left[\cos\Phi_4\left(\left|\frac{q_K}{p_K}\right|-\left|\frac{p_K}{q_K}\right|\right)\right] \\ &-\sin\Delta m_K t_2\left[\sin\Phi_4\left(\left|\frac{q_K}{p_K}\right|+\left|\frac{p_K}{q_K}\right|\right)\right]\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \end{aligned} \tag{3.22}$$

$$\begin{aligned} A_{n,dm}(t_1,t_2) = &-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{-\sinh\frac{1}{2}\Delta\Gamma_K t_2\left[\sin\Phi_4\left(\left|\frac{q_K}{p_K}\right|+\left|\frac{p_K}{q_K}\right|\right)\right] \\ &-\sin\Delta m_K t_2\left[\cos\Phi_4\left(\left|\frac{q_K}{p_K}\right|-\left|\frac{p_K}{q_K}\right|\right)\right]\Big\}\sin\Delta m_{B_s}t_1. \end{aligned} \tag{3.23}$$

The denominator $D(t_1,t_2)$ can be obtained by combining (2.4) with (2.11) - (2.13), while also performing the substitutions $C'_+(t_2)\leftrightarrow C'_-(t_2)$, $|p_K/q_K|\leftrightarrow|q_K/p_K|$ and $\phi_K\to-\phi_K$, and adding an extra negative sign in front of $S'_n(t_2)$.

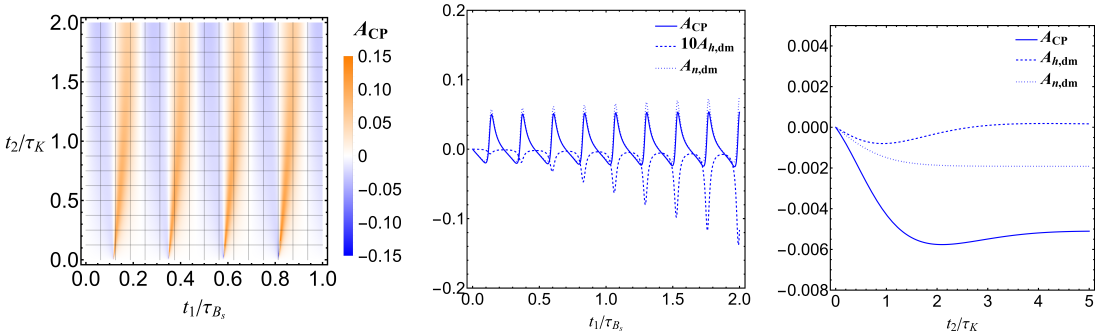


Figure 12. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0\to J/\psi K\to J/\psi(\pi^+\ell^-\bar{\nu}_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to τ_K) and t_2 (with t_1 integrated from 0 to $2\tau_{B_s}$), respectively.

Our numerical analysis for this channel is depicted in **Figure 12**. The left panel presents the two-dimensional time dependence. The difference between the results for this

channel and the $\pi^-\ell^+\nu_\ell$ channel lies in the modulation of the denominator. When t_2 is fixed, the dominant term in the denominator has a hyperbolic cosine dependence on t_1, which is the same in the results for the two different channels, while the sine term is a secondary contribution, and it contributes to the two channels with a sign difference. The middle panel shows the CP asymmetry with t_2 integrated from 0 to τ_K. While the $A_{n,dm}$ term still dominates, the growth magnitude of the $A_{h,dm}$ term is larger compared to the $\pi^-\ell^+\nu_\ell$ channel. Integrating over t_1 and analyzing t_2 dependence in the right panel, the $A_{h,dm}$ term emerges as dominant. Additionally, both the $A_{\rm CP}$ and $A_{n,dm}$ terms in the $\pi^+\ell^-\bar{\nu}_\ell$ channel display a sign difference from the $\pi^-\ell^+\nu_\ell$ channel.

3.2.3 $B_s^0 \to \rho^0 K \to \rho^0(\pi^\pm\ell^\mp\nu)$

The decay processes $B_s^0 \to \rho^0 K \to \rho^0(\pi^\pm\ell^\mp\nu)$ resemble to $B_s^0 \to J/\psi K^0_{L,S}(\to \pi^-\ell^+\nu_\ell)$, differing primarily in the phase difference ω_5. This difference arises from the phase between the decay amplitude $\langle\rho^0\bar{K}^0|B_s^0\rangle$ and its charge conjugate $\langle\rho^0 K^0|\bar{B}_s^0\rangle$, contrasting with the phase difference ω_{B_s} in the former. Considering $CP\,|\rho^0K^0\rangle = -\,|\rho^0\bar{K}^0\rangle$, we define

$$\frac{\langle\rho^0K^0|\bar{B}_s^0\rangle}{\langle\rho^0\bar{K}^0|B_s^0\rangle} = -e^{i\omega_5},\ \omega_5 = \arg\frac{V_{ub}V_{ud}^\star}{V_{ub}^\star V_{ud}}, \tag{3.24}$$

where again we have neglected the direct CP asymmetry in these processes.

For the final state $\pi^-\ell^+\nu_\ell$, two possible paths exist: $B_s^0 \to \rho^0\bar{K}^0 \to \rho^0K^0 \to \rho^0(\pi^-\ell^+\nu_\ell)$ and $B_s^0 \to \bar{B}_s^0 \to \rho^0K^0 \to \rho^0(\pi^-\ell^+\nu_\ell)$. Therefore, it is categorized as the first type in our classification. The results for the double-mixing terms are

$$\begin{aligned} A_{h,dm}(t_1,t_2) = &-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\left(\left|\frac{p_K}{q_K}\right| - \left|\frac{q_K}{p_K}\right|\right)\cos\Phi_5\sinh\frac{1}{2}\Delta\Gamma_K t_2 \\ &-\left(\left|\frac{p_K}{q_K}\right| + \left|\frac{q_K}{p_K}\right|\right)\sin\Phi_5\sin\Delta m_K t_2\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \end{aligned} \tag{3.25}$$

$$\begin{aligned} A_{n,dm}(t_1,t_2) = &-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\left(\left|\frac{p_K}{q_K}\right| + \left|\frac{q_K}{p_K}\right|\right)\sin\Phi_5\sinh\frac{1}{2}\Delta\Gamma_K t_2 \\ &+\left(\left|\frac{p_K}{q_K}\right| - \left|\frac{q_K}{p_K}\right|\right)\cos\Phi_5\sin\Delta m_K t_2\Big\}\sin\Delta m_{B_s}t_1, \end{aligned} \tag{3.26}$$

where $\Phi_5 \equiv \phi_{B_s} + \phi_K - \omega_5$. $D(t_1,t_2)$ can be obtained by replacing ω_{B_s} in the denominator of (3.20) with ω_5.

We have conducted the numerical analysis of both the total CP asymmetry $A_{\rm CP}$ and the double-mixing CP asymmetry A_{dm}, which are depicted in **Figure 13**. The two-dimensional time-dependent CP violation is depicted in the left panel, showing a peak value exceeding 50%. By integrating t_2 from 0 to τ_K, we obtain the t_1-dependent results displayed in the middle panel, where the double-mixing CP violation is notably significant. Here, $A_{\rm CP}$ predominantly arises from the double-mixing CP violation, particularly from the $A_{n,dm}$ term, which sums the contributions of the decays of K_S^0 and K_L^0. Integrating t_1 from τ_{B_s} to $5\tau_{B_s}$ yields the t_2-dependent results, showcased in the right panel. In this scenario, $A_{\rm CP}$ still mainly originates from the double-mixing CP violation, with the term $A_{h,dm}$ being dominant, particularly governed by the $K_S^0 - K_L^0$ interference.

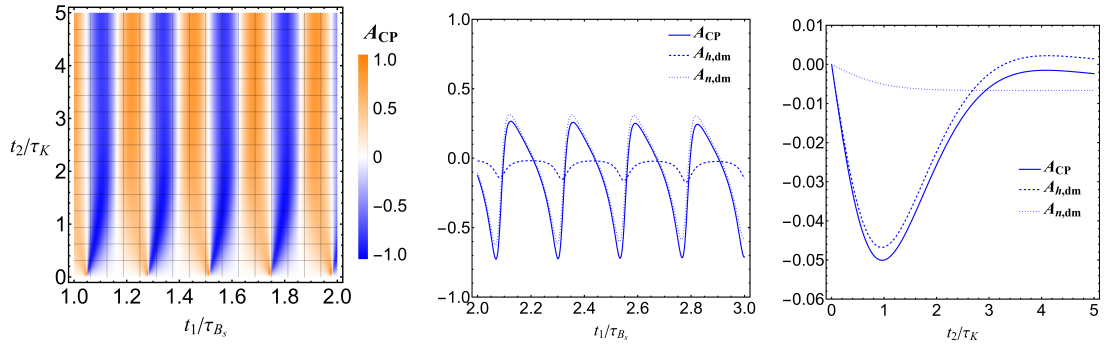


Figure 13. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to \rho^0 K \to \rho^0(\pi^-\ell^+\nu_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to τ_K) and t_2 (with t_1 integrated from τ_{B_s} to $5\tau_{B_s}$), respectively.

The results for the double-mixing terms of the $\pi^+\ell^-\bar{\nu}_\ell$ channel are similar to those of the $\pi^-\ell^+\nu_\ell$ channel with an extra overall negative sign in the term $A_{h,dm}$. They are calculated as

$$A_{h,dm}(t_1,t_2) = -\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\left[\cos\Phi_5\left(\left|\frac{q_K}{p_K}\right|-\left|\frac{p_K}{q_K}\right|\right)\right] + \sin\Delta m_K t_2\left[\sin\Phi_5\left(\left|\frac{q_K}{p_K}\right|+\left|\frac{p_K}{q_K}\right|\right)\right]\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \tag{3.27}$$

$$A_{n,dm}(t_1,t_2) = -\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\left[\sin\Phi_5\left(\left|\frac{q_K}{p_K}\right|+\left|\frac{p_K}{q_K}\right|\right)\right] - \sin\Delta m_K t_2\left[\cos\Phi_5\left(\left|\frac{q_K}{p_K}\right|-\left|\frac{p_K}{q_K}\right|\right)\right]\Big\}\sin\Delta m_{B_s}t_1, \tag{3.28}$$

where the denominator $D(t_1,t_2)$ can be obtained from the denominator of (3.22) by replacing ω_{B_s} with ω_5.

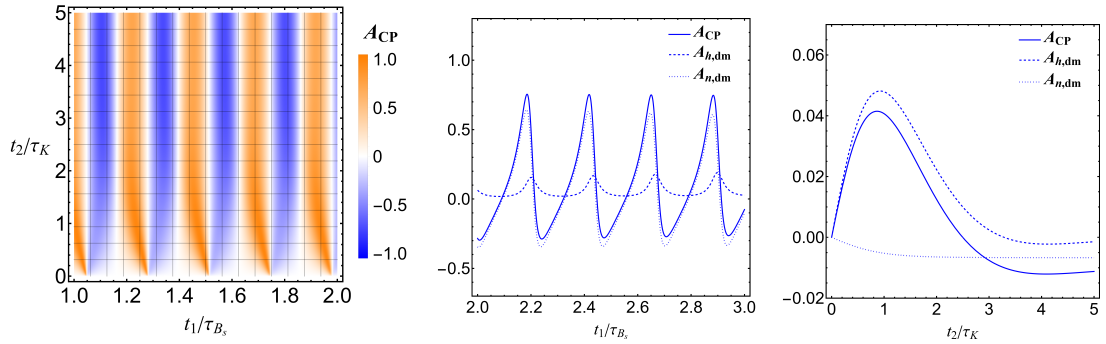


Figure 14. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to \rho^0 K \to \rho^0(\pi^+\ell^-\bar{\nu}_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to τ_K) and t_2 (with t_1 integrated from τ_{B_s} to $5\tau_{B_s}$), respectively.

Our analysis for this channel reveals a smaller CP violation compared to the $\pi^-\ell^+\nu_\ell$ final state, as displayed in **Figure 14**. The left panel displays the two-dimensional time dependence of the CP asymmetry. Integrating t_2 from 0 to τ_K, the t_1-dependent results, depicted in the middle panel, highlight the dominant role of separate contributions from K_S^0 and K_L^0 in the double-mixing CP violation. When integrating t_1 from τ_{B_s} to $5\tau_{B_s}$, the t_2-dependent results, as shown in the right panel, are predominantly influenced by the $K_S^0 - K_L^0$ interference. Notably, the value of the t_1-integrated CP violation is almost entirely positive, contrasting with the result for the $\pi^-\ell^+\nu_\ell$ final state. This distinction enables the experimental differentiation of these two processes.

We emphasize that the best approach for the first experimental attempt to measure the double-mixing CP violation would be to combine the two decay channels mentioned above. Considering the CP violation defined by the differences of the summed branching ratios

$$\frac{\mathcal{B}[B_s^0/\bar{B}_s^0 \to \rho^0 K \to \rho^0(\pi^-\ell^+\nu_\ell)] - \mathcal{B}[B_s^0/\bar{B}_s^0 \to \rho^0 K \to \rho^0(\pi^+\ell^-\bar{\nu}_\ell)]}{\mathcal{B}[B_s^0/\bar{B}_s^0 \to \rho^0 K \to \rho^0(\pi^-\ell^+\nu_\ell)] + \mathcal{B}[B_s^0/\bar{B}_s^0 \to \rho^0 K \to \rho^0(\pi^+\ell^-\bar{\nu}_\ell)]}\,, \tag{3.29}$$

initial tagging for B mesons in experiments is not necessary and the corresponding efficiency loss is prevented. It contains the double-mixing and non-double-mixing terms, which are calculated to be

$$\begin{aligned} A_{h,dm}(t_1,t_2) = &-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{D(t_1,t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\cos\Phi_5\left(\left|\frac{p_K}{q_K}\right| - \left|\frac{q_K}{p_K}\right|\right) \\ &- \sin\Delta m_K t_2\sin\Phi_5\left(\left|\frac{p_K}{q_K}\right| + \left|\frac{q_K}{p_K}\right|\right)\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \end{aligned} \tag{3.30}$$

$$A_{non-dm}(t_1,t_2) = \frac{e^{-\Gamma_{B_s}t_1}}{D(t_1,t_2)}\left(\left|\frac{p_K}{q_K}\right|^2 - \left|\frac{q_K}{p_K}\right|^2\right)|g_{-,K}(t_2)|^2\cosh\frac{1}{2}\Delta\Gamma_{B_s}t_1, \tag{3.31}$$

$$\begin{aligned} D(t_1,t_2) = &\, e^{-\Gamma_{B_s}t_1}\Big\{\left(\left|\frac{p_K}{q_K}\right|^2 + \left|\frac{q_K}{p_K}\right|^2\right)|g_{-,K}(t_2)|^2 + 2|g_{+,K}(t_2)|^2\Big\}\cosh\frac{1}{2}\Delta\Gamma_{B_s}t_1 \\ &- e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}\Big\{\sinh\frac{1}{2}\Delta\Gamma_K t_2\cos\Phi_5\left(\left|\frac{p_K}{q_K}\right| + \left|\frac{q_K}{p_K}\right|\right) \\ &- \sin\Delta m_K t_2\sin\Phi_5\left(\left|\frac{p_K}{q_K}\right| - \left|\frac{q_K}{p_K}\right|\right)\Big\}\sinh\frac{1}{2}\Delta\Gamma_{B_s}t_1. \end{aligned} \tag{3.32}$$

Apparently, the non-double-mixing term is negligible and the total CP violation is dominated by the double-mixing contribution.

We conducted a numerical analysis of this process, and the results are depicted in **Figure 15**. The left panel exhibits that the total CP asymmetry has a hyperbolic tangent dependence on t_1, consequently behaving linearly when t_1 is not too large. When integrating t_2 from 0 to τ_K, the total CP violation is predominantly contributed from the double-mixing CP violation, as displayed in the middle panel. It exhibits a linear dependence on t_1, which is consistent with the two-dimensional time dependence. The right panel shows the variation of CP violation with respect to t_2 after integrating t_1 from τ_{B_s} to $5\tau_{B_s}$. As t_2 increases, it first reaches a maximum and then rapidly decreases to nearly zero. This is because the primary contribution of the double-mixing CP violation in the numerator comes from the interference term between K_S^0 and K_L^0, *i.e.*, the second term in (3.32), which has a sine

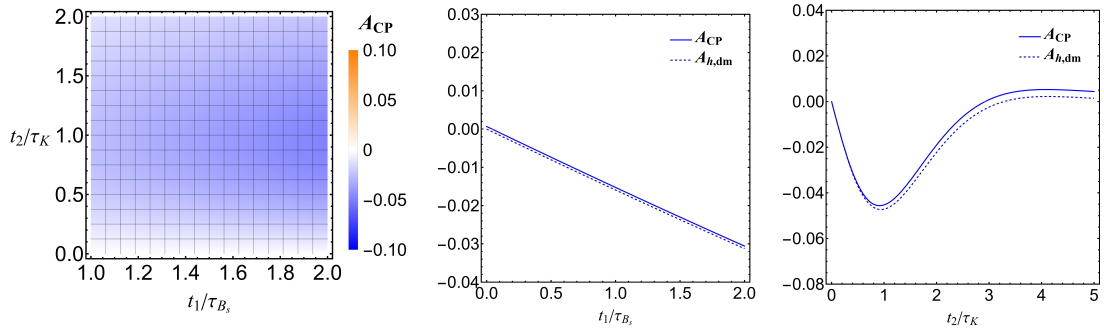


Figure 15. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0(t_1) \to \rho^0 K(t_2) \to \rho^0(\pi\ell\nu)$ defined by (3.29). The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to τ_K) and t_2 (with t_1 integrated from τ_{B_s} to $5\tau_{B_s}$), respectively.

t_1/τ_{B_s}	t_2/τ_K	$A_{\rm CP}$
0∼3	0∼ 2	-1.33%
0∼2	0∼ 2	-1.06%
0∼3	0∼ 1	-1.28%
0∼2	0∼ 1	-1.03%
0∼3	0∼ 0.5	-0.83%
0∼2	0∼ 0.5	-0.67%

Table 6. The time integrated $A_{\rm CP}$ in $B_s^0 \to \rho^0 K \to \rho^0(\pi\ell\nu_\ell)$.

dependence on t_2. Meanwhile, the denominator is mainly contributed by the term that depends on the hyperbolic cosine function of t_2, which continues to increase with time. The time-integrated results for the CP asymmetry (3.29) for this case are presented in Table 6, for experimental convenience. With different integrated time ranges, the values are basically around -1%.

3.2.4 $B_s^0 \to DK \to (K^-\pi^+)(\pi^-\ell^+\nu_\ell)$

We now focus on the decay channel $B_s^0 \to DK \to (K^-\pi^+)(\pi^-\ell^+\nu_\ell)$. This process involves four paths, even when we neglect the small mixing of the neutral D meson: $B_s^0 \to D^0\bar{K}^0 \to D^0K^0 \to f$, $B_s^0 \to \bar{D}^0\bar{K}^0 \to \bar{D}^0K^0 \to f$, $B_s^0 \to \bar{B}_s^0 \to D^0K^0 \to f$ and $B_s^0 \to \bar{B}_s^0 \to \bar{D}^0K^0 \to f$, where $f \equiv (K^-\pi^+)(\pi\ell\nu)$. This situation is very unique and does not belong to any of the categories we previously defined, but the expression for the CP violation can be derived analogously. We define the ratios between the decay amplitudes as

$$\begin{aligned}
\frac{A\left(\bar{B}_s^0 \to D^0K^0\right)}{A\left(B_s^0 \to \bar{D}^0\bar{K}^0\right)} &= e^{i\omega_3}, \quad \omega_3 = \arg\frac{V_{cb}V_{ud}^\star}{V_{cb}^\star V_{ud}}, \\
\frac{A\left(B_s^0 \to D^0\bar{K}^0\right)}{A\left(B_s^0 \to \bar{D}^0\bar{K}^0\right)} &= re^{i(\delta_4+\omega_4)}, \quad \omega_4 = \arg\frac{V_{ub}^\star V_{cd}}{V_{cb}^\star V_{ud}},
\end{aligned} \tag{3.33}$$

where $\omega_{3,4}$ are the weak phases and δ_4 is the strong phase. Then, the double-mixing and non-double-mixing CP violation terms defined in (2.4) are calculated as

$$
\begin{aligned}
A_{h,dm}(t_1,t_2) =& \frac{N_1}{D(t_1,t_2)} \\
=& \frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{p_K}{q_K}\right|\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\left[r\cos\left(\delta_4+\Phi_6\right)-r_D\cos\left(\Phi_7-\delta_D\right)\right] \\
&+\sin\Delta m_K t_2\left[-r\sin\left(\delta_4+\Phi_6\right)+r_D\sin\left(\Phi_7-\delta_D\right)\right]\Big\}\sinh\frac{\Delta\Gamma_{B_s}t_1}{2} \\
&-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{q_K}{p_K}\right|\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\left[r\cos\left(\delta_4-\Phi_6\right)-r_D\cos\left(\Phi_7+\delta_D\right)\right] \\
&+\sin\Delta m_K t_2\left[-r\sin\left(\delta_4-\Phi_6\right)-r_D\sin\left(\Phi_7+\delta_D\right)\right]\Big\}\sinh\frac{\Delta\Gamma_{B_s}t_1}{2} \\
&-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{p_K}{q_K}\right|rr_D\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}(r\cos\left(\Phi_8-\delta_D\right)-r_D\cos\left(\delta_4-\Phi_6\right)) \\
&+\sin\Delta m_K t_2(r\sin\left(\Phi_8-\delta_D\right)-r_D\sin\left(\delta_4-\Phi_6\right))\Big\}\sinh\frac{\Delta\Gamma_{B_s}t_1}{2} \\
&+\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}rr_D\left|\frac{q_K}{p_K}\right|\sinh\frac{\Delta\Gamma_{B_s}t_1}{2}\times \\
&\Big\{r\left(\sinh\frac{\Delta\Gamma_K t_2}{2}\cos\left(\Phi_8+\delta_D\right)-\sin\Delta m_K t_2\sin\left(\Phi_8+\delta_D\right)\right) \\
&-r_D\left(\sinh\frac{\Delta\Gamma_K t_2}{2}\cos\left(\delta_4+\Phi_6\right)+\sin\Delta m_K t_2\sin\left(\delta_4+\Phi_6\right)\right)\Big\},
\end{aligned}
\tag{3.34}
$$

$$
\begin{aligned}
A_{n,dm}(t_1,t_2) =& \frac{N_2}{D(t_1,t_2)} \\
=& \frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{p_K}{q_K}\right|\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\left[r\sin\left(\delta_4+\Phi_6\right)-r_D\sin\left(\Phi_7-\delta_D\right)\right] \\
&+\sin\Delta m_K t_2\left[r\cos\left(\delta_4+\Phi_6\right)-r_D\cos\left(\Phi_7-\delta_D\right)\right]\Big\}\sin\Delta m_{B_s}t_1 \\
&-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{q_K}{p_K}\right|\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\left[r\sin\left(\delta_4-\Phi_6\right)+r_D\sin\left(\Phi_7+\delta_D\right)\right] \\
&+\sin\Delta m_K t_2\left[r\cos\left(\delta_4-\Phi_6\right)-r_D\cos\left(\Phi_7+\delta_D\right)\right]\Big\}\sin\Delta m_{B_s}t_1 \\
&-\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}\left|\frac{p_K}{q_K}\right|rr_D\Big\{\sinh\frac{\Delta\Gamma_K t_2}{2}\left[-r\sin\left(\Phi_8-\delta_D\right)+r_D\sin\left(\delta_4-\Phi_6\right)\right] \\
&+\sin\Delta m_K t_2\left[r\cos\left(\Phi_8-\delta_D\right)-r_D\cos\left(\delta_4-\Phi_6\right)\right]\Big\}\sin\Delta m_{B_s}t_1 \\
&+\frac{e^{-(\Gamma_{B_s}t_1+\Gamma_K t_2)}}{2D(t_1,t_2)}rr_D\left|\frac{q_K}{p_K}\right|\sin\Delta m_{B_s}t_1\times \\
&\Big\{r\left(\sinh\frac{\Delta\Gamma_K t_2}{2}\sin\left(\Phi_8+\delta_D\right)+\sin\Delta m_K t_2\cos\left(\Phi_8+\delta_D\right)\right) \\
&+r_D\left(\sinh\frac{\Delta\Gamma_K t_2}{2}\sin\left(\delta_4+\Phi_6\right)-\sin\Delta m_K t_2\cos\left(\delta_4+\Phi_6\right)\right)\Big\},
\end{aligned}
\tag{3.35}
$$

$$
\begin{aligned}
A_{non-dm}(t_1,t_2) =& \frac{N_3}{D(t_1,t_2)} \\
=& \Big\{ |g_{+,B_s}(t_1)|^2 |g_{-,K}(t_2)|^2 \Big\{ \left|\frac{p_K}{q_K}\right|^2 \left[r^2 + r_D^2 - 2r_D r\cos\left(\delta_4+\delta_D+\omega_4\right)\right] \\
& - \left|\frac{q_K}{p_K}\right|^2 \left[r^2 + r_D^2 - 2r_D r\cos\left(\delta_4+\delta_D-\omega_4\right)\right] \Big\} - 2rr_D|g_{+,K}(t_2)|^2 \times \\
& \left[\cos\left(\delta_4-\omega_4-\delta_D\right) - \cos\left(\delta_4+\omega_4-\delta_D\right)\right] |g_{-,B_s}(t_1)|^2 \Big\} \frac{1}{D(t_1,t_2)}, \quad (3.36)
\end{aligned}
$$

where $\Phi_6 \equiv \phi_{B_s} + \phi_K - \omega_3 + \omega_4$, $\Phi_7 \equiv \phi_{B_s} + \phi_K - \omega_3$ and $\Phi_8 \equiv -\phi_{B_s} - \phi_K + \omega_3 - 2\omega_4$. The denominator $D(t_1,t_2)$ can be divided into three parts $N_1' + N_2' + N_3'$, which can be obtained by make some replacements to N_1, N_2, and N_3 in (3.34) - (3.36). N_1' and N_2' can be obtained by changing the sign in front of the exponential in the fourth line and the third-to-last line of (3.34) and (3.35) to an opposite sign. As for N_3', in addition to changing the minus sign in front of $|q_K/p_K|^2$ in the third line and the minus sign between the two cosines in the fourth line of (3.36) to a positive sign in N_3, we still need to incorporate an additional term $2|g_{-,B_s}(t_1)|^2|g_{+,K}(t_2)|^2\left(r_D^2 r^2+1\right)$ to the modified N_3.

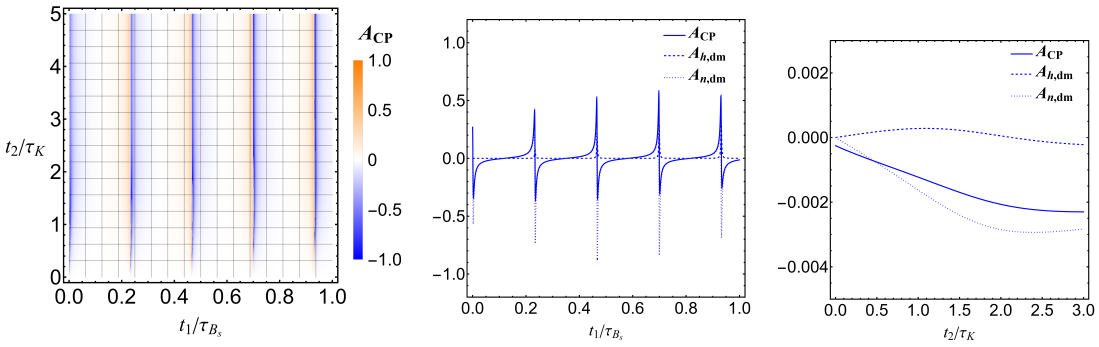


Figure 16. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to DK \to (K^-\pi^+)(\pi^-\ell^+\nu_\ell)$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $2\tau_K$) and t_2 (with t_1 integrated from 0 to τ_{B_s}), respectively.

The numerical result of the CP asymmetry in this process is shown in **Figure 16**. The two-dimensional time-dependent CP violation is shown in the left panel. While its peaks surpass 50%, it is only present in the proximity of specific moments. The result of integrating t_2 from 0 to $2\tau_K$ is shown in the middle panel. The contribution of $A_{n,dm}$ dominates except at the positions of positive peaks, as the values of $A_{h,dm}$ and A_{non-dm} are both positive and form upward pulses at each peak position, while $A_{n,dm}$ rapidly rises at each peak position and then directly becomes a negative pulse. Additionally, as the sum of $A_{h,dm}$ and A_{non-dm} is positive when $A_{n,dm}$ is negative, and their contribution is smaller than $A_{n,dm}$, the overall CP violation A_{CP} has larger positive peaks and smaller negative peaks compared to $A_{n,dm}$. The result of integrating t_1 from 0 to τ_{B_s} is shown in the right panel. $A_{h,dm}$ is one order of magnitude smaller than $A_{n,dm}$. The interference between $B_s^0 \to D^0\bar{K}^0 \to D^0K^0$ and $B_s^0 \to \bar{B}_s^0 \to D^0K^0$ is suppressed by the factor r.

Considering the contribution from $\bar{D}^0$, the interference between $B_s^0 \to \bar{D}^0\bar{K}^0 \to \bar{D}^0 K^0$ and $B_s^0 \to \bar{B}_s^0 \to D^0 K^0$ is suppressed by r_D, while the remaining interference contributions are all suppressed by factors of $r r_D$ or even more suppressed.

3.2.5 $B_s^0 \to DK_S^0 \to (\pi^+\pi^-)K_S^0$ and $B_s^0 \to DK_S^0 \to (K^-\pi^+)K_S^0$

We now analyze the decay modes $B_s^0 \to DK_S^0 \to f_D K_S^0$, with $f_D = \pi^+\pi^-, K^-\pi^+$. Measurements of similar channels have been performed by experiments. The branching ratio of $B_s^0 \to \bar{D}^0\bar{K}^{\star 0}$ has been reported by LHCb [49, 50], but the final state is self-tagged by the flavor of the $K^{\star 0}$, which excludes time-dependent analysis. The CP violation for $A_{\rm CP}(B_s^0 \to f_D K^{\star 0})$ has been measured [51]. Considering the possible decays $B_s^0 \to D^0 K_S^0$, $B_s^0 \to \bar{D}^0 K_S^0$, $D^0 \to f_D$, and $\bar{D}^0 \to f_D$, there are eight pathways for these two process, fitting them into what is defined as the third type.

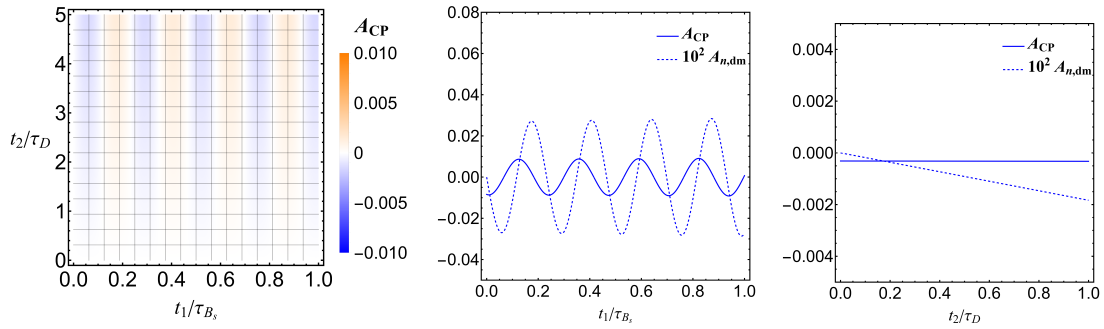


Figure 17. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to DK_S^0 \to (\pi^+\pi^-)K_S^0$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $5\tau_D$) and t_2 (with t_1 integrated from 0 to τ_{B_s}), respectively.

For the channel with $f_D = \pi^+\pi^-$, we derive the results for all the components of CP violation by setting $r_1 = r$, $\delta_1 = \delta_4$, $\theta_1 = \omega_4$, $\theta_2 = 0$, $r_2 = 1$, $\delta_2 = 0$ and $\theta_3 = \omega_3$ in (2.26), and then we obtain all the terms, which are shown in (A.1) - (A.12). For r and δ_4 defined in (3.33), we numerically use $r = 0.02$ and $\delta_4 = 194°$ as a benchmark. The values for ω_3 and ω_4 are given in Table 1. In this case, the dominant term in the double-mixing CP violation arises from the interference between the $B_s^0 \to \bar{D}^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ and $B_s^0 \to \bar{B}_s^0 \to D^0 K_S^0 \to \bar{D}^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ amplitudes as well as that between the $B_s^0 \to \bar{D}^0 K_S^0 \to D^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ and $B_s^0 \to \bar{B}_s^0 \to D^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ amplitudes. Our numerical results for this channel are presented in **Figure 17**, with the two-dimensional time dependence in the left panel. Its peak value is $\mathcal{O}(10^{-3})$. Integrating t_2 from 0 to $5\tau_D$ yields the t_1-dependent result, shown in the middle panel, revealing the double-mixing CP violation effect is of the order $\mathcal{O}(10^{-4})$. This is due to the small values x_D and y_D of the D meson as listed in Table 1. The t_2-integrated CP asymmetry is dominated by terms arising from the interference between the $B_s^0 \to \bar{D}^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ and $B_s^0 \to \bar{B}_s^0 \to D^0 K_S^0 \to (\pi^+\pi^-)K_S^0$ amplitudes, which have a cosine or hyperbolic cosine dependence on t_2 and have a sine dependence on t_1, namely $\sin \Delta m_{B_s} t_1$. Integrating t_1 from 0 to τ_{B_s}, the magnitude of the double-mixing term $A_{n,dm}$ climbs slowly as t_2 increases attributed to small mixing

parameters of D meson, and the total asymmetry $A_{\rm CP}$ does not show visible changes before t_2 becomes too large, as depicted in the right panel.

For the channel with $f_D = K^-\pi^+$, the double-mixing and non-double-mixing contributions to the CP asymmetry can be obtained by taking (3.16) and (3.17) with the substitutions $\phi_{B_d} \to \phi_{B_s}$, $r_B \to r$ and $\delta_B \to \delta_4$, and the replacement of the denominator $D(t_1, t_2)$ accordingly. For r and δ_4, we still numerically use $r = 0.02$ and $\delta_4 = 194°$. The presence of the suppression factor r_D and the strong phase δ_D results in a pattern distinct from that of the $\pi^+\pi^-$ channel. The numerical analysis for the double-mixing CP asymmetry A_{dm} is depicted in **Figure 18**. The left panel shows the two-dimensional time dependence, indicating the insignificance of the CP violation with $t_2 < 3\tau_D$, below the threshold of current experimental detection capacities. This double-mixing effect is also negligible in the integrated results over t_2 or t_1, presented on the middle and the right panel, respectively.

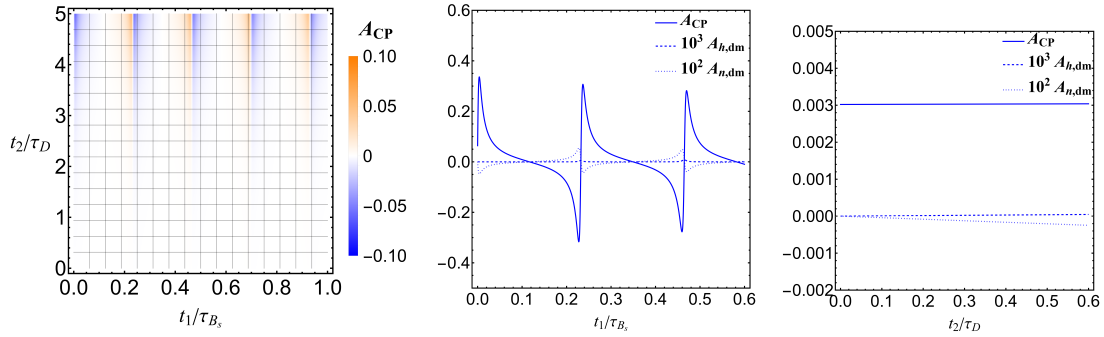


Figure 18. Time dependence of the CP asymmetry $A_{\rm CP}$ in $B_s^0 \to DK_S^0 \to (K^-\pi^+)K_S^0$. The left panel displays the two-dimensional time dependence. The middle panel and the right panel display the dependence on t_1 (with t_2 integrated from 0 to $5\tau_D$) and t_2 (with t_1 integrated from 0 to τ_{B_s}), respectively.

4 Conclusions

It was proposed recently by [37] that a long-overlooked CP violation effect exists in a decay chain with two neutral mesons involved, namely the double-mixing CP violation. This violation is induced by the interference among different oscillation paths of two neutral flavored mesons. In this work, we have analyzed the double-mixing CP violation in a series of cascade decay processes $M_1^0(t_1) \to M_2(t_2) \to f$, where the primary neutral meson M_1^0 is either $B_{d,s}^0$ mesons and the secondary neutral meson is D^0 or K^0 mesons. The involvement of two time parameters, t_1 and t_2, makes the two-dimensional time analysis of double-mixing CP violation allowable. We performed numerical analyses for each decay channel, providing the two-dimensional time-dependent double-mixing CP violation and the CP violation integrated over t_1 or t_2.

Our results show that the double-mixing CP violation can be very significant, with instances exceeding 50% – notably in the decay channel $B_s^0 \to \rho^0 K \to \rho^0(\pi\ell\nu)$. In similar cases with kaons reconstructed by semi-leptonic states, the total CP violation almost

entirely comes from the double-mixing CP violation. Moreover, our findings indicate that the double-mixing CP violation's dependency varies distinctly with t_1 and t_2: when integrating over t_2, the direct decays of K_S^0 and K_L^0 predominantly contribute to the CP violation, whereas integrating over t_1 highlights the dominance of $K_S^0 - K_L^0$ interference. We emphasize that one of the first experimental attempts to measure the double-mixing CP violation would be to combine the two decay channels $B_s^0 \to \rho^0 K \to \rho^0(\pi^-\ell^+\nu_\ell)$ and $B_s^0 \to \rho^0 K \to \rho^0(\pi^+\ell^-\bar{\nu}_\ell)$. This method does not require initial tagging of neutral B mesons, thus avoiding efficiency losses.

In comparison, the significance of double-mixing CP violation is less in decay channels involving D^0 as M_2^0, mainly due to the tiny mixing observed in neutral D mesons, such as in the decay channel $B_s^0 \to DK_S^0 \to (\pi^+\pi^-)K_S^0$. Nonetheless, double-mixing CP violation offers a major advantage by facilitating the extraction of CKM phase angles without theoretical inputs. For example, the decay channel $B_d^0 \to DK \to (K^-\pi^+)(\pi\ell\nu)$ involves both strong and weak phases, with the overall weak phase being approximated by $2\beta + \gamma$. This allows for the simultaneous extraction of strong and weak phases from experimental data on double-mixing CP violation, thereby directly testing the CKM mechanism.

Acknowledgement

The authors are grateful to Cai-Dian Lü, Wen-Bin Qian, Liang Sun, and Yue-Hong Xie for useful discussions. This work is supported by the Natural Science Foundation of China under grant No. 12375086.

References

[1] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5** (1967), 32-35

[2] W. Bernreuther, Lect. Notes Phys. **591** (2002), 237-293 [arXiv:hep-ph/0205279 [hep-ph]].

[3] L. Canetti, M. Drewes and M. Shaposhnikov, New J. Phys. **14** (2012), 095012 [arXiv:1204.4186 [hep-ph]].

[4] N. Cabibbo, Phys. Rev. Lett. **10** (1963), 531-533

[5] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973), 652-657

[6] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, Phys. Rev. Lett. **13** (1964), 138-140

[7] H. Burkhardt *et al.* [NA31], Phys. Lett. B **206** (1988), 169-176

[8] V. Fanti *et al.* [NA48], Phys. Lett. B **465** (1999), 335-348 [arXiv:hep-ex/9909022 [hep-ex]].

[9] A. Alavi-Harati *et al.* [KTeV], Phys. Rev. Lett. **83** (1999), 22-27 [arXiv:hep-ex/9905060 [hep-ex]].

[10] B. Aubert *et al.* [BaBar], Phys. Rev. Lett. **87** (2001), 091801 [arXiv:hep-ex/0107013 [hep-ex]].

[11] K. Abe *et al.* [Belle], Phys. Rev. Lett. **87** (2001), 091802 [arXiv:hep-ex/0107061 [hep-ex]].

[12] Y. Chao *et al.* [Belle], Phys. Rev. Lett. **93** (2004), 191802 [arXiv:hep-ex/0408100 [hep-ex]].

[13] B. Aubert *et al.* [BaBar], Phys. Rev. Lett. **93** (2004), 131801 [arXiv:hep-ex/0407057 [hep-ex]].

[14] P. del Amo Sanchez *et al.* [BaBar], Phys. Rev. D **82** (2010), 072004 [arXiv:1007.0504 [hep-ex]].

[15] A. Poluektov *et al.* [Belle], Phys. Rev. D **81** (2010), 112002 [arXiv:1003.3360 [hep-ex]].

[16] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **108** (2012), 101803 [arXiv:1112.3183 [hep-ex]].

[17] R. Aaij *et al.* [LHCb], Phys. Lett. B **712** (2012), 203-212 [erratum: Phys. Lett. B **713** (2012), 351] [arXiv:1203.3662 [hep-ex]].

[18] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **110** (2013) no.22, 221601 [arXiv:1304.6173 [hep-ex]].

[19] R. Aaij *et al.* [LHCb], JHEP **07** (2014), 041 [arXiv:1405.2797 [hep-ex]].

[20] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **122** (2019) no.21, 211803 [arXiv:1903.08726 [hep-ex]].

[21] M. Ablikim *et al.* [BESIII], Chin. Phys. C **44** (2020) no.4, 040001 [arXiv:1912.05983 [hep-ex]].

[22] E. Kou *et al.* [Belle-II], PTEP **2019** (2019) no.12, 123C01 [erratum: PTEP **2020** (2020) no.2, 029201] [arXiv:1808.10567 [hep-ex]].

[23] R. Aaij *et al.* [LHCb], Eur. Phys. J. C **73** (2013) no.4, 2373 [arXiv:1208.3355 [hep-ex]].

[24] R. Aaij *et al.* [LHCb], [arXiv:1808.08865 [hep-ex]].

[25] R. Aaij *et al.* [LHCb], [arXiv:2305.10515 [hep-ex]].

[26] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, Phys. Rev. Lett. **83** (1999), 1914-1917 [arXiv:hep-ph/9905312 [hep-ph]].

[27] M. Beneke, G. Buchalla, M. Neubert and C. T. Sachrajda, Nucl. Phys. B **591** (2000), 313-418 [arXiv:hep-ph/0006124 [hep-ph]].

[28] Y. Y. Keum, H. N. Li and A. I. Sanda, Phys. Rev. D **63** (2001), 054008 [arXiv:hep-ph/0004173 [hep-ph]].

[29] C. D. Lu, K. Ukai and M. Z. Yang, Phys. Rev. D **63** (2001), 074009 [arXiv:hep-ph/0004213 [hep-ph]].

[30] C. W. Bauer, S. Fleming, D. Pirjol and I. W. Stewart, Phys. Rev. D **63** (2001), 114020 [arXiv:hep-ph/0011336 [hep-ph]].

[31] C. W. Bauer, D. Pirjol and I. W. Stewart, Phys. Rev. D **65** (2002), 054022 [arXiv:hep-ph/0109045 [hep-ph]].

[32] H. n. Li, C. D. Lu and F. S. Yu, Phys. Rev. D **86** (2012), 036012 [arXiv:1203.3120 [hep-ph]].

[33] Q. Qin, H. n. Li, C. D. Lü and F. S. Yu, Phys. Rev. D **89** (2014) no.5, 054006 [arXiv:1305.7021 [hep-ph]].

[34] Q. Qin, Z. T. Zou, X. Yu, H. n. Li and C. D. Lü, Phys. Lett. B **732** (2014), 36-40 [arXiv:1401.1028 [hep-ph]].

[35] Q. Qin, C. Wang, D. Wang and S. H. Zhou, Front. Phys. (Beijing) **18** (2023) no.6, 64602 [arXiv:2111.14472 [hep-ph]].

[36] F. S. Yu, D. Wang and H. n. Li, Phys. Rev. Lett. **119** (2017) no.18, 181802 [arXiv:1707.09297 [hep-ph]].

[37] Y. F. Shen, W. J. Song and Q. Qin, [arXiv:2301.05848 [hep-ph]].

[38] R. L. Workman *et al.* [Particle Data Group], PTEP **2022** (2022), 083C01

[39] Y. S. Amhis *et al.* [HFLAV], Phys. Rev. D **107** (2023) no.5, 052008 [arXiv:2206.07501 [hep-ex]].

[40] Y. I. Azimov, [arXiv:hep-ph/9712288 [hep-ph]].

[41] R. Aaij *et al.* [LHCb], JHEP **06** (2015), 131 [arXiv:1503.07055 [hep-ex]].

[42] B. Aubert *et al.* [BaBar], Phys. Rev. D **79** (2009), 072003 [arXiv:0805.2001 [hep-ex]].

[43] P. Krokovny *et al.* [Belle], Phys. Rev. Lett. **90** (2003), 141802 [arXiv:hep-ex/0212066 [hep-ex]].

[44] B. Aubert *et al.* [BaBar], Phys. Rev. Lett. **96** (2006), 011803 [arXiv:hep-ex/0509036 [hep-ex]].

[45] B. Aubert *et al.* [BaBar], Phys. Rev. D **74** (2006), 031101 [arXiv:hep-ex/0604016 [hep-ex]].

[46] R. Aaij *et al.* [LHCb], JHEP **08** (2016), 137 [arXiv:1605.01082 [hep-ex]].

[47] R. Aaij *et al.* [LHCb], JHEP **06** (2016), 131 [arXiv:1604.01525 [hep-ex]].

[48] T. Aaltonen *et al.* [CDF], Phys. Rev. D **83** (2011), 052012 [arXiv:1102.1961 [hep-ex]].

[49] R. Aaij *et al.* [LHCb], Phys. Lett. B **706** (2011), 32-39 [arXiv:1110.3676 [hep-ex]].

[50] R. Aaij *et al.* [LHCb], Phys. Rev. D **90** (2014) no.7, 072003 [arXiv:1407.7712 [hep-ex]].

[51] R. Aaij *et al.* [LHCb], Phys. Rev. D **90** (2014) no.11, 112002 [arXiv:1407.8136 [hep-ex]].

A Formulae for Category 3

Here we present relevant analytical expressions for the terms of the double-mixing CP violation of **Category 3** in Section 2. Based on the definitions of the parameters provided in (2.26), we can derive the individual terms in (2.4) as

$$
\begin{aligned}
C_+(t_2) = e^{-\Gamma_2 t_2}\Bigg\{ & r_2 \sinh\frac{\Delta\Gamma_2 t_2}{2}\left[r_1^2\left(\left|\frac{q_2}{p_2}\right|\cos\Omega_3 - \left|\frac{p_2}{q_2}\right|\cos\Omega_4\right) + \left|\frac{p_2}{q_2}\right|\cos\Omega_3 - \left|\frac{q_2}{p_2}\right|\cos\Omega_4\right] \\
& - r_2 \sin\Delta m_2 t_2\left[r_1^2\left(\left|\frac{q_2}{p_2}\right|\sin\Omega_3 - \left|\frac{p_2}{q_2}\right|\sin\Omega_4\right) - \left|\frac{p_2}{q_2}\right|\sin\Omega_3 + \left|\frac{q_2}{p_2}\right|\sin\Omega_4\right] \\
& + r_1 \sinh\frac{\Delta\Gamma_2 t_2}{2}\left[r_2^2\left(\left|\frac{q_2}{p_2}\right|\cos\Omega_1 - \left|\frac{p_2}{q_2}\right|\cos\Omega_2\right) + \left|\frac{p_2}{q_2}\right|\cos\Omega_1 - \left|\frac{q_2}{p_2}\right|\cos\Omega_2\right] \\
& - r_1 \sin\Delta m_2 t_2\left[r_2^2\left(\left|\frac{q_2}{p_2}\right|\sin\Omega_1 - \left|\frac{p_2}{q_2}\right|\sin\Omega_2\right) - \left|\frac{p_2}{q_2}\right|\sin\Omega_1 + \left|\frac{q_2}{p_2}\right|\sin\Omega_2\right]\Bigg\} \\
& + 2r_1 r_2 |g_{+,2}(t_2)|^2\Big[\cos\left(\Omega_1 - \Omega_3\right) - \cos\left(\Omega_2 - \Omega_4\right)\Big] \\
& + 2r_1 r_2 |g_{-,2}(t_2)|^2\Big[\cos\left(\Omega_1 + \Omega_3\right) - \cos\left(\Omega_2 + \Omega_4\right)\Big] \\
& + \left(r_1^2 r_2^2 - 1\right)|g_{-,2}(t_2)|^2\left(\left|\frac{q_2}{p_2}\right|^2 - \left|\frac{p_2}{q_2}\right|^2\right),
\end{aligned}
\tag{A.1}
$$

where $\Omega_1 \equiv \delta_1 + \theta_1 - \phi_2$, $\Omega_2 \equiv \delta_1 - \theta_1 + \phi_2$, $\Omega_3 \equiv \delta_2 - \phi_2$ and $\Omega_4 \equiv \delta_2 + \phi_2$,

$$
C_-(t_2) = |g_{-,2}(t_2)|^2\left\{r_1^2\left(\left|\frac{q_1}{p_1}\right|^2\left|\frac{p_2}{q_2}\right|^2 - \left|\frac{p_1}{q_1}\right|^2\left|\frac{q_2}{p_2}\right|^2\right) + r_2^2\left(\left|\frac{q_1}{p_1}\right|^2\left|\frac{q_2}{p_2}\right|^2 - \left|\frac{p_1}{q_1}\right|^2\left|\frac{p_2}{q_2}\right|^2\right)\right\}
$$

$$
\begin{aligned}
&+ 2r_1r_2\left|\frac{q_1}{p_1}\right|^2\Big\{|g_{+,2}(t_2)|^2\cos(\Omega_2+\Omega_3)+|g_{-,2}(t_2)|^2\cos(\Omega_2-\Omega_3)\Big\}\\
&- 2r_1r_2\left|\frac{p_1}{q_1}\right|^2\Big\{|g_{+,2}(t_2)|^2\cos(\Omega_1+\Omega_4)+|g_{-,2}(t_2)|^2\cos(\Omega_1-\Omega_4)\Big\}\\
&+ e^{-\Gamma_2 t_2}\Bigg\{r_2\left|\frac{q_1}{p_1}\right|^2\left|\frac{q_2}{p_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3-\sin\Delta m_2 t_2\sin\Omega_3\right)\\
&+r_1^2r_2\left|\frac{q_1}{p_1}\right|^2\left|\frac{p_2}{q_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3+\sin\Delta m_2 t_2\sin\Omega_3\right)\\
&-r_2\left|\frac{p_1}{q_1}\right|^2\left|\frac{p_2}{q_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4-\sin\Delta m_2 t_2\sin\Omega_4\right)\\
&-r_1^2r_2\left|\frac{p_1}{q_1}\right|^2\left|\frac{q_2}{p_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4+\sin\Delta m_2 t_2\sin\Omega_4\right)\\
&+r_1\sinh\frac{\Delta\Gamma_2 t_2}{2}\left[\left|\frac{q_1}{p_1}\right|^2\cos\Omega_2\left(r_2^2\left|\frac{q_2}{p_2}\right|+\left|\frac{p_2}{q_2}\right|\right)-\left|\frac{p_1}{q_1}\right|^2\cos\Omega_1\left(r_2^2\left|\frac{p_2}{q_2}\right|+\left|\frac{q_2}{p_2}\right|\right)\right]\\
&+r_1\sin\Delta m_2 t_2\left[\left|\frac{q_1}{p_1}\right|^2\sin\Omega_2\left(r_2^2\left|\frac{q_2}{p_2}\right|-\left|\frac{p_2}{q_2}\right|\right)-\left|\frac{p_1}{q_1}\right|^2\sin\Omega_1\left(r_2^2\left|\frac{p_2}{q_2}\right|-\left|\frac{q_2}{p_2}\right|\right)\right]\Bigg\}\\
&+\left(r_1^2r_2^2+1\right)\left(\left|\frac{q_1}{p_1}\right|^2-\left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2,
\end{aligned}
\tag{A.2}
$$

$$
\begin{aligned}
S_{n,1}(t_2)=&\frac{e^{-\Gamma_2 t_2}}{2}\Big\{g(-,-,-,-)\times\sinh\frac{\Delta\Gamma_2 t_2}{2}\Big[r_1^2\sin(\Omega_5-2\Omega_6)+\sin\Omega_5\Big]\\
&+g(-,+,-,+)\times\sin\Delta m_2 t_2\Big[r_1^2\cos(\Omega_5-2\Omega_6)-\cos\Omega_5\Big]\Big\}\\
&+r_1r_2\left|\frac{q_2}{p_2}\right|\sin(\delta_1-\Omega_5+\Omega_6)e^{-\Gamma_2 t_2}\left\{\left|\frac{q_1}{p_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3-\sin\Delta m_2 t_2\sin\Omega_3\right]\right.\\
&\left.+\left|\frac{p_1}{q_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4+\sin\Delta m_2 t_2\sin\Omega_4\right]\right\}\\
&-r_1r_2\left|\frac{p_2}{q_2}\right|\sin(\delta_1+\Omega_5-\Omega_6)e^{-\Gamma_2 t_2}\left\{\left|\frac{q_1}{p_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3+\sin\Delta m_2 t_2\sin\Omega_3\right]\right.\\
&\left.+\left|\frac{p_1}{q_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4-\sin\Delta m_2 t_2\sin\Omega_4\right]\right\},
\end{aligned}
\tag{A.3}
$$

where $\Omega_5\equiv\theta_3-\phi_1-\phi_2$, $\Omega_6\equiv\theta_1-\phi_2$, and the $g(-,+,-,+)$ is defined as

$$
g(-,+,-,+)=-\left|\frac{q_1}{p_1}\right|\left|\frac{p_2}{q_2}\right|+r_2^2\left|\frac{q_1}{p_1}\right|\left|\frac{q_2}{p_2}\right|-r_2^2\left|\frac{p_1}{q_1}\right|\left|\frac{p_2}{q_2}\right|+\left|\frac{p_1}{q_1}\right|\left|\frac{q_2}{p_2}\right|,
$$

and others g functions with different sign arguments can be obtained by flip signs in the above expression accordingly,

$$
\begin{aligned}
S_{n,2}(t_2)=&-r_1^2r_2\left|\frac{q_1}{p_1}\right|\Big\{|g_{+,2}(t_2)|^2\sin(\Omega_3+\Omega_5-2\Omega_6)-|g_{-,2}(t_2)|^2\sin(\Omega_3-\Omega_5+2\Omega_6)\Big\}\\
&+r_2\left|\frac{q_1}{p_1}\right|\Big\{|g_{+,2}(t_2)|^2\sin(\Omega_3-\Omega_5)-|g_{-,2}(t_2)|^2\sin(\Omega_3+\Omega_5)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&+ r_1^2 r_2 \left|\frac{p_1}{q_1}\right| \Big\{ |g_{+,2}(t_2)|^2 \sin(\Omega_4 - \Omega_5 + 2\Omega_6) - |g_{-,2}(t_2)|^2 \sin(\Omega_4 + \Omega_5 - 2\Omega_6) \Big\} \\
&- r_2 \left|\frac{p_1}{q_1}\right| \Big\{ |g_{+,2}(t_2)|^2 \sin(\Omega_4 + \Omega_5) - |g_{-,2}(t_2)|^2 \sin(\Omega_4 - \Omega_5) \Big\} \\
&+ r_1 \sin(\delta_1 - \Omega_5 + \Omega_6) f_1(q,p) - r_1 \sin(\delta_1 + \Omega_5 - \Omega_6) f_1(p,q), \qquad \text{(A.4)}
\end{aligned}
$$

where

$$
f_1(q,p) = |g_{+,2}(t_2)|^2 \left(\left|\frac{q_1}{p_1}\right| + r_2^2 \left|\frac{p_1}{q_1}\right| \right) + \left|\frac{q_2}{p_2}\right|^2 |g_{-,2}(t_2)|^2 \left(r_2^2 \left|\frac{q_1}{p_1}\right| + \left|\frac{p_1}{q_1}\right| \right),
$$

and the function $f_1(p,q)$ is obtained by exchanging $p_{1,2} \leftrightarrow q_{1,2}$ in $f_1(q,p)$,

$$
\begin{aligned}
S_{h,1}(t_2) = \frac{e^{-\Gamma_2 t_2}}{2} \Big\{ & g(+,+,-,-) \sinh\frac{\Delta\Gamma_2 t_2}{2} \Big[r_1^2 \cos(\Omega_5 - 2\Omega_6) + \cos\Omega_5 \Big] \\
& - g(+,-,-,+) \sin\Delta m_2 t_2 \Big[r_1^2 \sin(\Omega_5 - 2\Omega_6) - \sin\Omega_5 \Big] \Big\} \\
& + r_1 r_2 \left|\frac{q_2}{p_2}\right| \cos(\delta_1 - \Omega_5 + \Omega_6) e^{-\Gamma_2 t_2} \left\{ \left|\frac{q_1}{p_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_3 - \sin\Delta m_2 t_2 \sin\Omega_3 \right] \right. \\
& \left. - \left|\frac{p_1}{q_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_4 + \sin\Delta m_2 t_2 \sin\Omega_4 \right] \right\} \\
& + r_1 r_2 \left|\frac{p_2}{q_2}\right| \cos(\delta_1 + \Omega_5 - \Omega_6) e^{-\Gamma_2 t_2} \left\{ \left|\frac{q_1}{p_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_3 + \sin\Delta m_2 t_2 \sin\Omega_3 \right] \right. \\
& \left. - \left|\frac{p_1}{q_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_4 - \sin\Delta m_2 t_2 \sin\Omega_4 \right] \right\}, \qquad \text{(A.5)}
\end{aligned}
$$

$$
\begin{aligned}
S_{h,2}(t_2) = & \, r_1^2 r_2 \left|\frac{q_1}{p_1}\right| \Big[|g_{+,2}(t_2)|^2 \cos(\Omega_3 + \Omega_5 - 2\Omega_6) + |g_{-,2}(t_2)|^2 \cos(\Omega_3 - \Omega_5 + 2\Omega_6) \Big] \\
& + r_2 \left|\frac{q_1}{p_1}\right| \Big[|g_{+,2}(t_2)|^2 \cos(\Omega_3 - \Omega_5) + |g_{-,2}(t_2)|^2 \cos(\Omega_3 + \Omega_5) \Big] \\
& - r_1^2 r_2 \left|\frac{p_1}{q_1}\right| \Big[|g_{+,2}(t_2)|^2 \cos(\Omega_4 - \Omega_5 + 2\Omega_6) + |g_{-,2}(t_2)|^2 \cos(\Omega_4 + \Omega_5 - 2\Omega_6) \Big] \\
& - r_2 \left|\frac{p_1}{q_1}\right| \Big[|g_{+,2}(t_2)|^2 \cos(\Omega_4 + \Omega_5) + |g_{-,2}(t_2)|^2 \cos(\Omega_4 - \Omega_5) \Big] \\
& + r_1 \cos(\delta_1 - \Omega_5 + \Omega_6) f_2(q,p) + r_1 \cos(\delta_1 + \Omega_5 - \Omega_6) f_2(p,q), \qquad \text{(A.6)}
\end{aligned}
$$

where

$$
f_2(q,p) = |g_{+,2}(t_2)|^2 \left(\left|\frac{q_1}{p_1}\right| - r_2^2 \left|\frac{p_1}{q_1}\right| \right) + \left|\frac{q_2}{p_2}\right|^2 |g_{-,2}(t_2)|^2 \left(r_2^2 \left|\frac{q_1}{p_1}\right| - \left|\frac{p_1}{q_1}\right| \right),
$$

and the function $f_2(p,q)$ is obtained by exchanging $p_{1,2} \leftrightarrow q_{1,2}$ in $f_2(q,p)$,

$$
\begin{aligned}
C'_+(t_2) = e^{-\Gamma_2 t_2} \Big\{ & r_2 \sinh\frac{\Delta\Gamma_2 t_2}{2} \left[r_1^2 \left(\left|\frac{q_2}{p_2}\right| \cos\Omega_3 + \left|\frac{p_2}{q_2}\right| \cos\Omega_4 \right) + \left|\frac{p_2}{q_2}\right| \cos\Omega_3 + \left|\frac{q_2}{p_2}\right| \cos\Omega_4 \right] \\
& - r_2 \sin\Delta m_2 t_2 \left[r_1^2 \left(\left|\frac{q_2}{p_2}\right| \sin\Omega_3 + \left|\frac{p_2}{q_2}\right| \sin\Omega_4 \right) - \left|\frac{p_2}{q_2}\right| \sin\Omega_3 - \left|\frac{q_2}{p_2}\right| \sin\Omega_4 \right]
\end{aligned}
$$

$$
\begin{aligned}
&+r_1 \sinh\frac{\Delta\Gamma_2 t_2}{2}\left[r_2^2\left(\left|\frac{q_2}{p_2}\right|\cos\Omega_1+\left|\frac{p_2}{q_2}\right|\cos\Omega_2\right)+\left|\frac{p_2}{q_2}\right|\cos\Omega_1+\left|\frac{q_2}{p_2}\right|\cos\Omega_2\right]\\
&-r_1\sin\Delta m_2 t_2\left[r_2^2\left(\left|\frac{q_2}{p_2}\right|\sin\Omega_1+\left|\frac{p_2}{q_2}\right|\sin\Omega_2\right)-\left|\frac{p_2}{q_2}\right|\sin\Omega_1-\left|\frac{q_2}{p_2}\right|\sin\Omega_2\right]\Big\}\\
&+2r_1r_2|g_{+,2}(t_2)|^2\Big[\cos(\Omega_1-\Omega_3)+\cos(\Omega_2-\Omega_4)\Big]\\
&+2r_1r_2|g_{-,2}(t_2)|^2\Big[\cos(\Omega_1+\Omega_3)+\cos(\Omega_2+\Omega_4)\Big]\\
&+\big(r_1^2r_2^2+1\big)|g_{-,2}(t_2)|^2\left(\left|\frac{q_2}{p_2}\right|^2+\left|\frac{p_2}{q_2}\right|^2\right)+2\big(r_1^2+r_2^2\big)|g_{+,2}(t_2)|^2,
\end{aligned}
\tag{A.7}
$$

$$
\begin{aligned}
C'_-(t_2)=|g_{-,2}(t_2)|^2&\left\{r_1^2\left(\left|\frac{q_1}{p_1}\right|^2\left|\frac{p_2}{q_2}\right|^2+\left|\frac{p_1}{q_1}\right|^2\left|\frac{q_2}{p_2}\right|^2\right)+r_2^2\left(\left|\frac{q_1}{p_1}\right|^2\left|\frac{q_2}{p_2}\right|^2+\left|\frac{p_1}{q_1}\right|^2\left|\frac{p_2}{q_2}\right|^2\right)\right\}\\
&+2r_1r_2\left|\frac{q_1}{p_1}\right|^2\Big[|g_{+,2}(t_2)|^2\cos(\Omega_2+\Omega_3)+|g_{-,2}(t_2)|^2\cos(\Omega_2-\Omega_3)\Big]\\
&+2r_1r_2\left|\frac{p_1}{q_1}\right|^2\Big[|g_{+,2}(t_2)|^2\cos(\Omega_1+\Omega_4)+|g_{-,2}(t_2)|^2\cos(\Omega_1-\Omega_4)\Big]\\
&+e^{-\Gamma_2 t_2}\left\{r_2\left|\frac{q_1}{p_1}\right|^2\left|\frac{q_2}{p_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3-\sin\Delta m_2 t_2\sin\Omega_3\right)\right.\\
&+r_1^2r_2\left|\frac{q_1}{p_1}\right|^2\left|\frac{p_2}{q_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3+\sin\Delta m_2 t_2\sin\Omega_3\right)\\
&+r_2\left|\frac{p_1}{q_1}\right|^2\left|\frac{p_2}{q_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4-\sin\Delta m_2 t_2\sin\Omega_4\right)\\
&+r_1^2r_2\left|\frac{p_1}{q_1}\right|^2\left|\frac{q_2}{p_2}\right|\left(\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4+\sin\Delta m_2 t_2\sin\Omega_4\right)\\
&+r_1\sinh\frac{\Delta\Gamma_2 t_2}{2}\left[\left|\frac{q_1}{p_1}\right|^2\cos\Omega_2\left(r_2^2\left|\frac{q_2}{p_2}\right|+\left|\frac{p_2}{q_2}\right|\right)+\left|\frac{p_1}{q_1}\right|^2\cos\Omega_1\left(r_2^2\left|\frac{p_2}{q_2}\right|+\left|\frac{q_2}{p_2}\right|\right)\right]\\
&+r_1\sin\Delta m_2 t_2\left[\left|\frac{q_1}{p_1}\right|^2\sin\Omega_2\left(r_2^2\left|\frac{q_2}{p_2}\right|-\left|\frac{p_2}{q_2}\right|\right)+\left|\frac{p_1}{q_1}\right|^2\sin\Omega_1\left(r_2^2\left|\frac{p_2}{q_2}\right|-\left|\frac{q_2}{p_2}\right|\right)\right]\Big\}\\
&+\big(r_1^2r_2^2+1\big)\left(\left|\frac{q_1}{p_1}\right|^2+\left|\frac{p_1}{q_1}\right|^2\right)|g_{+,2}(t_2)|^2,
\end{aligned}
\tag{A.8}
$$

$$
\begin{aligned}
S'_{n,1}(t_2)=\frac{e^{-\Gamma_2 t_2}}{2}&\Big\{g(-,-,+,+)\times\sinh\frac{\Delta\Gamma_2 t_2}{2}\Big[r_1^2\sin(\Omega_5-2\Omega_6)+\sin\Omega_5\Big]\\
&+g(-,+,+,-)\times\sin\Delta m_2 t_2\Big[r_1^2\cos(\Omega_5-2\Omega_6)-\cos\Omega_5\Big]\Big\}\\
&+r_1r_2\left|\frac{q_2}{p_2}\right|\sin(\delta_1-\Omega_5+\Omega_6)e^{-\Gamma_2 t_2}\left\{\left|\frac{q_1}{p_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_3-\sin\Delta m_2 t_2\sin\Omega_3\right]\right.\\
&-\left.\left|\frac{p_1}{q_1}\right|\left[\sinh\frac{\Delta\Gamma_2 t_2}{2}\cos\Omega_4+\sin\Delta m_2 t_2\sin\Omega_4\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
&- r_1 r_2 \left|\frac{p_2}{q_2}\right| \sin(\delta_1 + \Omega_5 - \Omega_6) e^{-\Gamma_2 t_2} \left\{ \left|\frac{q_1}{p_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_3 + \sin\Delta m_2 t_2 \sin\Omega_3 \right] \right. \\
&\left. - \left|\frac{p_1}{q_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_4 - \sin\Delta m_2 t_2 \sin\Omega_4 \right] \right\}, \qquad \text{(A.9)}
\end{aligned}
$$

$$
\begin{aligned}
S'_{n,2}(t_2) = &- r_1^2 r_2 \left|\frac{q_1}{p_1}\right| \left\{ |g_{+,2}(t_2)|^2 \sin(\Omega_3 + \Omega_5 - 2\Omega_6) - |g_{-,2}(t_2)|^2 \sin(\Omega_3 - \Omega_5 + 2\Omega_6) \right\} \\
&+ r_2 \left|\frac{q_1}{p_1}\right| \left\{ |g_{+,2}(t_2)|^2 \sin(\Omega_3 - \Omega_5) - |g_{-,2}(t_2)|^2 \sin(\Omega_3 + \Omega_5) \right\} \\
&- r_1^2 r_2 \left|\frac{p_1}{q_1}\right| \left\{ |g_{+,2}(t_2)|^2 \sin(\Omega_4 - \Omega_5 + 2\Omega_6) - |g_{-,2}(t_2)|^2 \sin(\Omega_4 + \Omega_5 - 2\Omega_6) \right\} \\
&+ r_2 \left|\frac{p_1}{q_1}\right| \left\{ |g_{+,2}(t_2)|^2 \sin(\Omega_4 + \Omega_5) + |g_{-,2}(t_2)|^2 \sin(\Omega_4 - \Omega_5) \right\} \\
&+ r_1 \sin(\delta_1 - \Omega_5 + \Omega_6) f_2(q,p) - r_1 \sin(\delta_1 + \Omega_5 - \Omega_6) f_2(p,q), \qquad \text{(A.10)}
\end{aligned}
$$

$$
\begin{aligned}
S'_{h,1}(t_2) = &\frac{e^{-\Gamma_2 t_2}}{2} \left\{ g(+,+,+,+) \sinh\frac{\Delta\Gamma_2 t_2}{2} \left[r_1^2 \cos(\Omega_5 - 2\Omega_6) + \cos\Omega_5 \right] \right. \\
&\left. - g(+,-,+,-) \sin\Delta m_2 t_2 \left[r_1^2 \sin(\Omega_5 - 2\Omega_6) - \sin\Omega_5 \right] \right\} \\
&+ r_1 r_2 \left|\frac{q_2}{p_2}\right| \cos(\delta_1 - \Omega_5 + \Omega_6) e^{-\Gamma_2 t_2} \left\{ \left|\frac{q_1}{p_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_3 - \sin\Delta m_2 t_2 \sin\Omega_3 \right] \right. \\
&\left. + \left|\frac{p_1}{q_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_4 + \sin\Delta m_2 t_2 \sin\Omega_4 \right] \right\} \\
&+ r_1 r_2 \left|\frac{p_2}{q_2}\right| \cos(\delta_1 + \Omega_5 - \Omega_6) e^{-\Gamma_2 t_2} \left\{ \left|\frac{q_1}{p_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_3 + \sin\Delta m_2 t_2 \sin\Omega_3 \right] \right. \\
&\left. + \left|\frac{p_1}{q_1}\right| \left[\sinh\frac{\Delta\Gamma_2 t_2}{2} \cos\Omega_4 - \sin\Delta m_2 t_2 \sin\Omega_4 \right] \right\}, \qquad \text{(A.11)}
\end{aligned}
$$

$$
\begin{aligned}
S'_{h,2}(t_2) = &\, r_1^2 r_2 \left|\frac{q_1}{p_1}\right| \left[|g_{+,2}(t_2)|^2 \cos(\Omega_3 + \Omega_5 - 2\Omega_6) + |g_{-,2}(t_2)|^2 \cos(\Omega_3 - \Omega_5 + 2\Omega_6) \right] \\
&+ r_2 \left|\frac{q_1}{p_1}\right| \left\{ |g_{+,2}(t_2)|^2 \cos(\Omega_3 - \Omega_5) + |g_{-,2}(t_2)|^2 \cos(\Omega_3 + \Omega_5) \right\} \\
&+ r_1^2 r_2 \left|\frac{p_1}{q_1}\right| \left[|g_{+,2}(t_2)|^2 \cos(\Omega_4 - \Omega_5 + 2\Omega_6) + |g_{-,2}(t_2)|^2 \cos(\Omega_4 + \Omega_5 - 2\Omega_6) \right] \\
&+ r_2 \left|\frac{p_1}{q_1}\right| \left\{ |g_{+,2}(t_2)|^2 \cos(\Omega_4 + \Omega_5) + |g_{-,2}(t_2)|^2 \cos(\Omega_4 - \Omega_5) \right\} \\
&+ r_1 \cos(\delta_1 - \Omega_5 + \Omega_6) f_1(q,p) + r_1 \cos(\delta_1 + \Omega_5 - \Omega_6) f_1(p,q). \qquad \text{(A.12)}
\end{aligned}
$$

arXiv:2403.01833v1 [hep-ph] 4 Mar 2024

Charged particle multiplicity fluctuation in $A-A$ collisions at RHIC and LHC energies using Angantyr model.

Pritindra Bhowmick,[1,*] Sadhana Dash,[2,†] Basanta Kumar Nandi,[2,‡] and Claude Pruneau[3,§]

[1] *Department of Physics, Indian Institute of Science Education and Research, Bhopal, India-462066*

[2] *Department of Physics. Indian Institute of Technology Bombay, Mumbai, India-400076*

[3] *Department of Physics and Astronomy, Wayne State University, Detroit, 48201, USA*

Event-by-event fluctuations of the charged particle multiplicity are studied for a wide range of centralities for Au−Au collisions at $\sqrt{s_{NN}}$ = 200 GeV, Pb−Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV and 5.02 TeV using the Pythia 8 Angantyr model. The centrality dependence of ω_{ch} observable, which quantifies the fluctuations in terms of scaled variance is studied for different pseudorapidity ranges and has been compared with those obtained from a simple participant superposition model. The ω_{ch} was found to be lower than the expectations from the participant model. The estimate would act like a baseline for current and future measurements of event-by-event fluctuations in the charged particle multiplicities in systems at LHC energies where no de-confined medium of quarks and gluons are formed.

I. INTRODUCTION

Comparisons of measurements of anisotropic flow and transverse momentum correlator carried out in heavy-ion collisions at RHIC and LHC with hydrodynamical models suggest the matter formed in these collisions behave as nearly perfect fluid. Hydro calculations indeed indicate that, in spite of the enormous energy and pressure fleetingly produced in these systems, they are characterized by small shear and bulk velocities per unit of entropy, commonly denoted η/s and ζ/s, respectively. With the emergence of new measurement techniques and substantial theoretical advances, much progress has been recently achieved in measuring η/s and ζ/s with increasing precision [1–8]. Shear and bulk viscosities are however only two of the many properties that characterize the bulk properties of systems and the matter they consist of. Other properties, such as the specific heat, the isothermal compressibility, and various chemical potentials are also interest to fully characterize the properties of QGP matter and the nature of the phase transition from QGP to a hadron gas. There is hope, however, that such properties can be measured based on event-by-event fluctuations of "global event observable" such the particle multiplicity detected in a specific kinematic range, the mean transverse momenta ($\langle p_{\rm T}\rangle$), the total transverse energies (E_T), as well as fluctuations of conserved quantities such as the net-charge and net-baryon number of particles produced in collisions. Nominally, event-by-event fluctuations of net-charge and net-baryon number are proportional to the QGP susceptibilities and measurements of these fluctuations shall provide considerable insight onto the properties of the QGP and the phase transition. However, technical difficulties arise in the evaluation of these fluctuations because net charge and net baryon fluctuations are constrained by charge and baryon conservation and are additionally strongly influenced by production and transport mechanisms. Similarly, matter properties such as the isothermal compressibility and the heat capacity, which are nominally related to fluctuations of the produced particle multiplicity and the average (event-wise) particle momentum, are also challenging to extract because energy-momentum conservation, charge conservation, etc, also constrain fluctuations of the produced particle multiplicity and the momentum av-

* pritindra18@iiserb.ac.in
† sadhana@phy.iitb.ac.in
‡ basanta@phy.iitb.ac.in
§ pruneau@physics.wayne.edu

erage of these particles. There is thus a need to learn how to account and remove known “trivial" effects in measurements of such fluctuations. To this end, it is useful to consider predictions of pQCD inspired elementary collision models such as Pythia8 and Herwig[9, 10].

The existence of a critical point at the QCD phase transition has been predicted to be associated with large event-by-event fluctuations in the aforementioned global observables [11, 12]. The relative fluctuations of the observables in A−A collisions have been found to be smaller compared to those in p−p collisions. The origin of fluctuations in both the systems is quite different and their measurement can provide information about the event-by-event initial state and final state fluctuations. The presence of thermal equilibration in A−A collisions makes it tricky to extract the information about the initial state from final state fluctuations. Therefore, the basic goal in heavy ion collisions have been to relate the event-by-event fluctuations of the final state to the thermodynamic properties of the medium.

The experimental measurement of event-by-event fluctuations in charged particle multiplicities in recent past searched for direct evidence of the QCD phase transition in different collision systems at different energies [13–16]. If one considers the medium formed in relativistic nuclear collisions as a Grand Canonical Ensemble, the isothermal compressibility (k_T) of the system can be related to the variance of the particle multiplicity near mid-rapidity as

$$\langle (N - \langle N \rangle)^2 \rangle = \frac{k_B T \langle N \rangle^2}{V} k_T \tag{1}$$

where $\langle N \rangle$ is the mean multiplicity, k_B is the Boltzmann constant , T is the temperature and V is the volume of the system . Thus, measurements of charged particle multiplicity fluctuations are expected to be a sensitive probe for critical behavior. A recent measurement by ALICE experiment on centrality dependence of multiplicity fluctuations yielded a value of $k_T = 27.9\ fm^3/GeV$ where the background fluctuation was estimated from the participant model [17, 19]. In this work, the multiplicity fluctuations of charged particles are obtained in Au−Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb−Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV and 5.02 TeV using the Angantyr model of Pythia 8. The fluctuations are estimated over a large range of centralities and the effect of acceptance is presented by estimating the observable in three different pseudo-rapidity intervals. The obtained results are also compared with the expectations of simple participant model and are expected to provide a baseline for current and future measurements and any significant deviation in experimental data might suggest the contribution of collectivity and final state effects.

II. MULTIPLICITY FLUCTUATION OBSERVABLE

The relative fluctuation, ω_X of an observable X is defined as

$$\omega_X = \frac{\langle X^2 \rangle - \langle X \rangle^2}{\langle X \rangle} \tag{2}$$

where $\langle . \rangle$ is the expectation value of the observable in the given event ensemble. Fluctuations in the number of charged particles, N_{ch}, can be quantified as the scaled variance of the charged particle multiplicity distribution

$$\omega_{ch} = \frac{\langle N_{ch}^2 \rangle - \langle N_{ch} \rangle^2}{\langle N_{ch} \rangle} \tag{3}$$

The measured value of ω_{ch} has contributions originating from both statistical as well as dynamical effects. In order to extract the information from dynamical fluctuations and associated physical mechanism, one needs to clearly understand the different sources of fluctuations such as impact parameter fluctuations, effect of limited acceptance of detector, fluctuations in the number of primary collisions, effect of resonance decays , effect of rescattering of secondaries and other sources of correlations.

In the present work, the estimation of fluctuations is also carried out using a simple participant superposition model. This model assumes that the nucleus-nucleus collision as the sum of contributions from many sources created in the early stage of the interaction. The multiplicity fluctuations has contributions largely due to different impact parameters. The density fluctuations within the nucleus will make this contribution non-zero even if we consider a narrow impact parameter window. The quantum fluctuations in the nucleon-nucleon cross sections can also lead to fluctuations in the number of particles produced by each source. These contributions are related to the initial volume of the interacting system and can lead to fluctuations in the number of participating nucleons, N_{part} and can be related to the initial size of the system. Therefore, the fluctuations in N_{ch} will have contributions due to fluctuations in N_{part} as well as the fluctuations in the number of particles produced per participant (n). Hence, one can express the multiplicity fluctuations in the absence of correlations between the multiplicities produced by different sources by the following expression

$$\omega_{ch,parti} = \omega_n + \langle n \rangle \omega_{N_{part}} \tag{4}$$

Although the particle production in central collisions is affected by nucleon-nucleon scattering, rescatterings between produced particles, and other effects, the estimations from such calculations and comparison with experimental data might divulge the validity of the superposition of nucleon-nucleon interactions in the case of heavy-ion collisions. One should also consider that ω_n will have a strong dependence on the acceptance of the detector.

Estimation of $\omega_{N_{part}}$:
The fluctuations in the impact parameter can reflect in the fluctuations in the number of participants. It has been estimated using the Pythia8 Angantyr model. The quantity $\langle n \rangle$ can be calculated by taking the ratio of the mean charged particle multiplicity for a given acceptance to the mean number of participants for the same centrality range.

Estimation of ω_n :
This term gives the fluctuation in the number of particles produced per participant. One needs to estimate the fraction f which is the ratio of the mean number of particles per participant accepted within the acceptance under study ($\langle n \rangle$) to the mean number of total particles produced per participant ($\langle m \rangle$). While each participant produces m charged particles, only a smaller fraction $f = \langle n \rangle / \langle m \rangle$ are accepted. Without carrying out a detailed analysis of the acceptance, one can make a simple statistical estimate assuming that the particles are accepted randomly, in which case n is binomially distributed with mean mf and $\sigma_n^2 = mf(1-f)$ for a fixed m. Therefore, the fluctuations in the n accepted particles is given as

$$\omega_n = 1 - f + f\omega_m \tag{5}$$

III. ANALYSIS METHOD

In the present study, the default heavy-ion model of Pythia 8, Angantyr model has been used to generate Pb−Pb events at $\sqrt{s_{NN}}$ = 2.76 TeV and 5.02 TeV and Au − Au collisions at $\sqrt{s_{NN}}$ = 200 GeV. The model is an extension of p−p collisional dynamics to nucleon-nucleus (p−A) and nucleus-nucleus (A−A) collisions in the Pythia 8 event generator [18]. For each A−A collision, the nucleons are randomly distributed as per Wood-Saxon distribution in the impact parameter space and the formulation of individual nuclear interactions is based on the Glauber model. The number of wounded nucleons are estimated from Glauber formalism with Gribov corrections to the diffractive excitation of the individual nucleon. The Glauber calculations include the fluctuations of nucleons in the projectile and target nuclei. The model includes two interaction methodologies for the projectile and target nucleons. In the first one, the interactions are treated as non-diffractive (ND) p−p collisions and are called primary ND interactions while in the second case, a wounded projec-

tile nucleon is allowed to have ND interactions with other target nucleons. These type of interactions are called secondary ND collisions. Depending on their interaction probability, various interactions between wounded nucleons in projectile and target are referred as elastic, ND, secondary ND, single-diffractive and double-diffractive interactions. The full Pythia8 machinery is used for parton-level event generation for these primary ND interactions (as well as diffractive interactions). Later, all the sub-events are stacked together to build an exclusive final state of the considered heavy-ion collision. The sub-events are treated independently where hadrons are produced using the Lund string fragmentation model. As the model is devoid of collective effects and does not assume the formation of any thermalised medium, it can act as a baseline model to understand the observables and the associated background which are affected by medium effects.

In the present study, the transverse momentum, p_T range of the charged particles is selected from $0.2 < pT < 2$ GeV/c. The event-by-event charged particle multiplicity has been determined for $|\eta| <$ 0.5, 0.8 and 1.5 to study the dependence of the multiplicity fluctuations on acceptance. The events are classified into different centrality classes using the two different centrality estimators, namely, the total charged particle multiplicity obtained in the η range $1.6 \leq |\eta| \leq 2.5$ (CENT_METH1) and the sum of transverse energy of all neutral particles produced (CENT_METH2). The study is important as the variance of the charged particle multiplicity distribution is influenced by the centrality estimating method and may play an important role in observed scaled variance of charged particle multiplicity distributions. The centrality determination methods are chosen such that the observables are independent of methods and thus there is no bias due to selection of particles in the same kinematic acceptance.

The distribution of the considered observables for centrality distribution and their division into different centralities is shown in Figure 1 for Pb−Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV. An additional check was performed to check the correlations between the obtained charged particle multiplicity in different η regions and the number of participants obtained by the two different methods by calculating the Pearson Coefficient between the variables. The correlation coefficient was also determined between the number of charged particles in the analysis range and the centrality estimators (charged particle multiplicity ($1.6 \leq |\eta| \leq 2.5$) and transverse energy of neutral particles) for different centrality classes. This is shown in lower panels of Figure 1 for Pb−Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV. The lower left panel shows a considerable difference between the two methods and it is more pronounced for peripheral collisions while there is a good agreement in the right panel. One can observe that the correlation coefficient between the shown observables is around 0.3 for central collisions and it decreases from central to peripheral collisions.

A given centrality class is a collection of events that comprises of a range of impact parameters or number of participating nucleons, N_{part}. This results in additional fluctuations in the number of produced particles within each centrality class. In order to infer about the genuine fluctuations arising from different dynamical processes, one needs to ensure that the fluctuations in N_{part} within a centrality class is minimal.The fluctuations arising due to wider centrality bin width is corrected or minimised by dividing the centrality bin into smaller bins and weighing the corresponding moments as

$$X = \frac{\sum_i n_i X_i}{\sum_i n_i} \tag{6}$$

where the index i runs over each multiplicity bin, X_i represents various measured observable for the i^{th} bin, and n_i is the number of events in the i^{th} centrality bin. This method is commonly known as centrality bin width correction method and the final results are presented for a centrality bin width of 5% after implementing the centrality bin-width corrections of 1% bin width [20, 21].

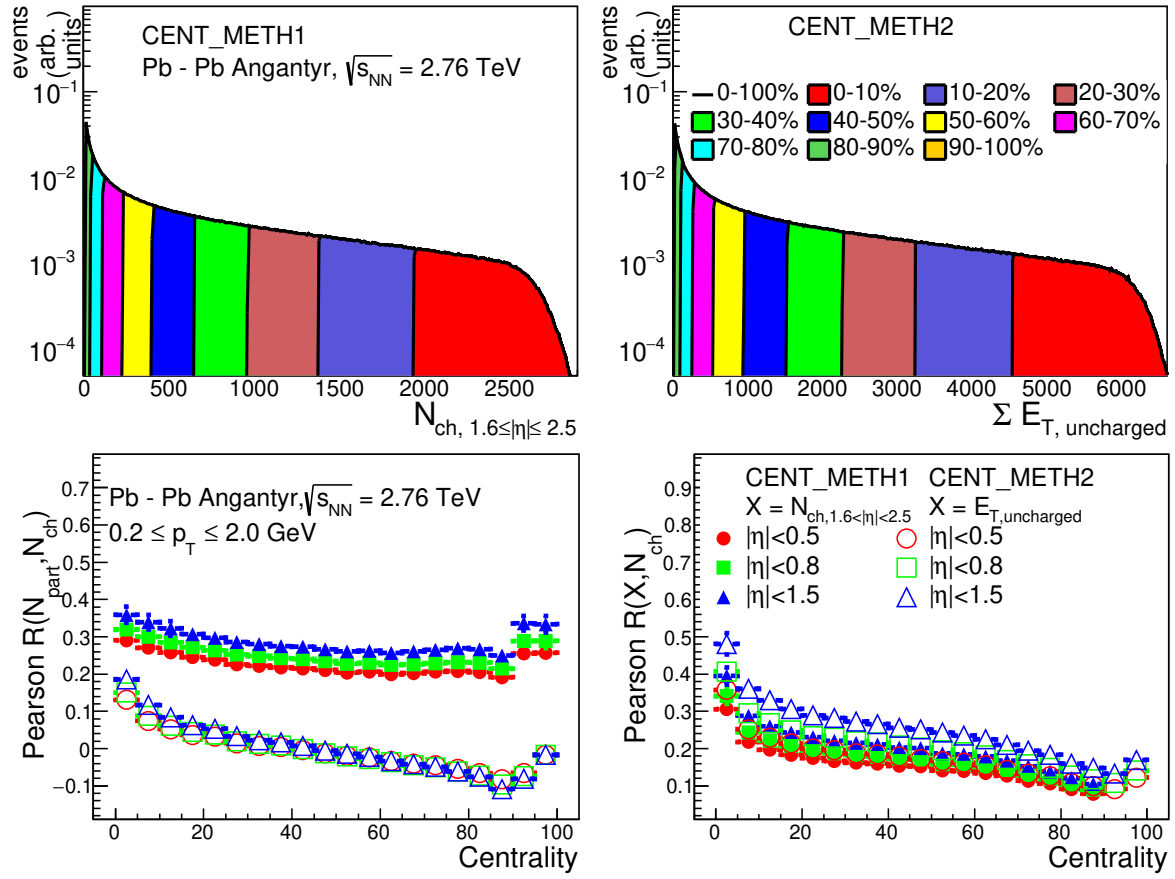


FIG. 1: (Color online) [Upper panels] : The distribution of charged particle multiplicity in the η range $1.6 \leq |\eta| \leq 2.5$ (left panel) and the distribution of transverse energy of neutral particles (right panel) in Pb−Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. The centrality division has also been shown with different coloured regions. [Bottom panels]: The Karl Pearson correlation coefficient for number of participant nucleons (N_part) and Charged particle multiplicity for different centrality classes (left panel). The Karl Pearson coefficient for charged particle multiplicity and centrality determination estimator (X) for three different acceptance regions

IV. RESULTS AND DISCUSSION

Figure 2 compares the mean, variance, skewness and kurtosis of the charged particle multiplicity distributions obtained for different η ranges as a function of centrality using two different methods of centrality determination. It can be observed that the variation of $\langle N_{ch} \rangle$ with $\langle N_{part} \rangle$ shows an expected increasing trend with $\langle N_{part} \rangle$ and there is a good agreement for the values obtained through the two centrality determination methods. However, the same is not true for the trend of variances of charged particle multiplicity distribution and the disagreement increases with an increase in η interval. The observed difference in principle can affect the values of ω_{ch}. The skewness and kurtosis decreases with $\langle N_{part} \rangle$ and is in agreement with the expectations of central limit theorem.

In order to estimate the ω_{ch} values from the wounded participant model, one needs to obtain the values of f, fraction of charged particles accepted to the total number of charged particles produced. The obtained value is used in Equation [5] to obtain the values of ω_n. The variation of this fraction with centrality for different pseudo-rapidity intervals is shown in Figure 3. It

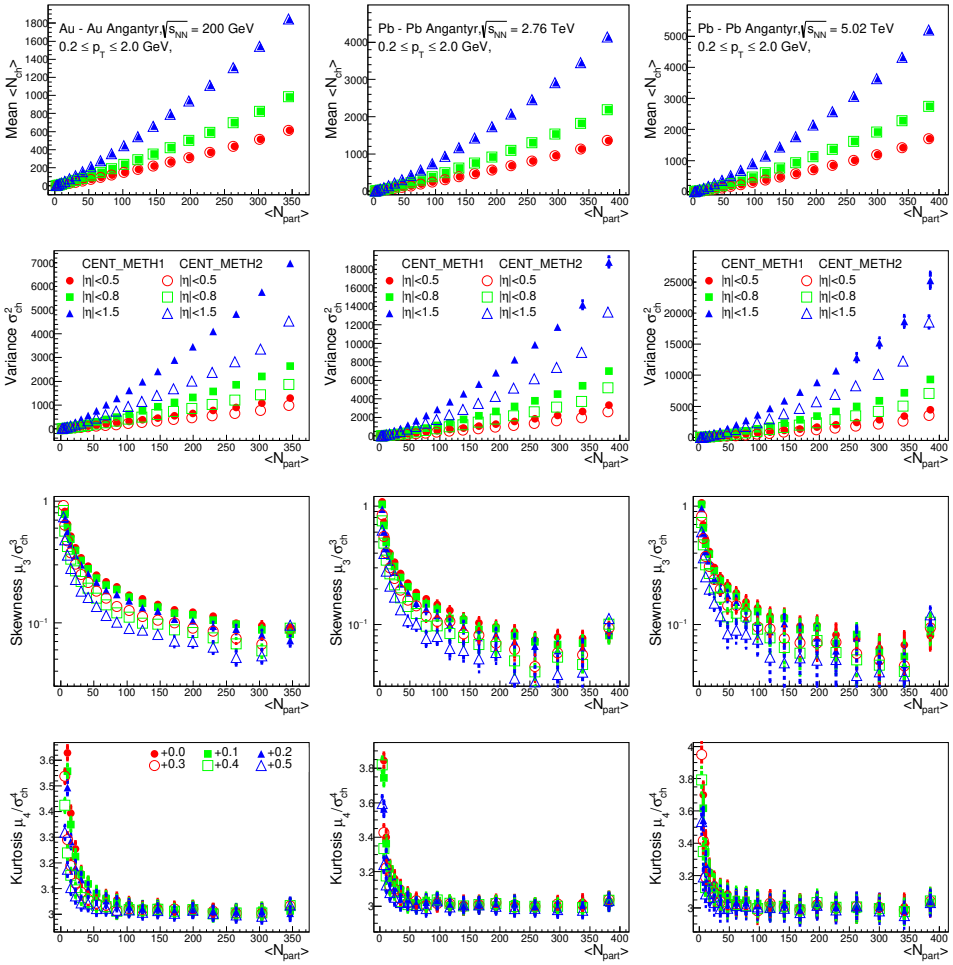


FIG. 2: (Color online) The variation of $\langle N_{ch} \rangle$, variance, skewness and kurtosis of charged particle multiplicity with $\langle N_{part} \rangle$(centrality) for different η ranges estimated for two centrality determination method. The solid and open markers are for two different methods of centrality estimation.

can be observed that f shows a slightly increasing trend with centrality. However, the values decrease with beam energy for similar pseudo-rapidity acceptance.

The variation of ω_{ch} obtained from the Angantyr model with centrality is compared with the expectations from the wounded participant model in Figure 4. The comparison is also shown for different η ranges. The left hand side panels show that the values obtained from the Angantyr model are consistently lower than the geometrical fluctuations. There is no significant centrality and beam energy dependence. However, one can observe that the central values show an increasing trend with centrality for ω_{ch} while it is opposite for the ones obtained from the simple participant model. The values in wider acceptance intervals are higher for all the energies and systems studied. The middle panel shows the same for a different centrality estimator and the trend remains the same. The right panel compares the estimations from two different centrality estimator methods and one can observe a noticeable difference between the obtained values. This observation can be attributed to different values of correlation coefficient between the charged particle multiplicity and the number of participants obtained from the two methods of centrality estimation as shown in Figure 1. This indicates towards the sensitivity

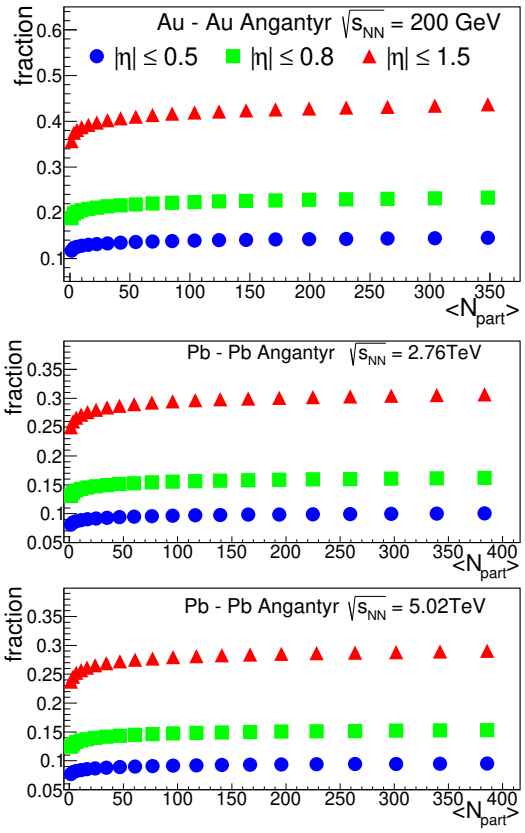


FIG. 3: (Color online) The variation of f with $\langle N_{part}\rangle$(centrality) for different η ranges estimated for the three collision systems.

of the multiplicity fluctuations studies on methods of centrality estimation. The values of the scaled variance obtained from the model remain significantly above the random (Poisson) expectation of 1.0.

A comparison of the ω_{ch} values obtained from Angantyr model and that measured by ALICE experiment [17] has been shown in Figure 5 for Pb−Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV. One can observe that the model under-predicts the data in lower centralities while the central values are in reasonable agreement for high values of $\langle N_{part}\rangle$. However, the trend shown for evolution with centrality is quite opposite. The model predicts a slightly increasing trend with $\langle N_{part}\rangle$ while the data shows a decreasing trend.

Figure 6 shows the fluctuations in terms of $R = \sigma^2/\mu^2$ as a function of the $\langle N_{part}\rangle$ for all the three systems. The dependence of the observable can be best described by the power law scaling as a function of $\langle N_{part}\rangle$. The observed scaling is independent of the system and beam energy and is in agreement with the results observed from different experiments. The details of the extracted fit parameters ($lnR = a_1 ln(\langle N_{part}\rangle) + a_0$) are shown in Table I.

The present study provides an estimation of the ω_{ch} obtained from Angantyr model and it will be a relevant baseline to the measured values from LHC data as the model does not consider the formation of a deconfined medium. An estimate from the participant model indicates that one needs to consider additional sources to account for background estimation of multiplicity fluctuations.

$\sqrt{s_{NN}}$	η range	a_0	a_1	b_0	b_1
Au+Au	$\|\eta\| < 0.5$	0.74	-1.084	0.69	-1.14
200 GeV	$\|\eta\| < 0.8$	0.48	-1.078	0.34	-1.14
	$\|\eta\| < 1.5$	0.16	-1.073	-0.12	-1.13
Pb+Pb	$\|\eta\| < 0.5$	0.56	-1.15	0.51	-1.20
2.76 TeV	$\|\eta\| < 0.8$	0.29	-1.14	0.16	-1.20
	$\|\eta\| < 1.5$	-0.05	-1.14	-0.30	-1.19
Pb+Pb	$\|\eta\| < 0.5$	0.56	-1.18	0.48	-1.23
5.02 TeV	$\|\eta\| < 0.8$	0.29	-1.17	0.13.	-1.22
	$\|\eta\| < 1.5$	-0.07	-1.16	-0.34	-1.20

TABLE I: A list of fit parameters for different collision systems. a_0 (b_0) and a_1 (b_1) are the intercept and the slope parameters extracted from the fit in CENT_METH1(CENT_METH2)

V. SUMMARY

A detailed study of event-by-event fluctuations in the multiplicities of charged particles has been carried out using the Pythia8 Angantyr model for Au−Au collisions at $\sqrt{s_{NN}}$ = 200 GeV and Pb−Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV and 5.02 TeV. The multiplicity fluctuation observable, ω_{ch} has been obtained for different centrality intervals as well as for different pseudo-rapidity classes. The centrality bin width correction has been implemented to take care of the contribution from impact parameter fluctuations. A small dependence on centrality method estimation is also observed which can be

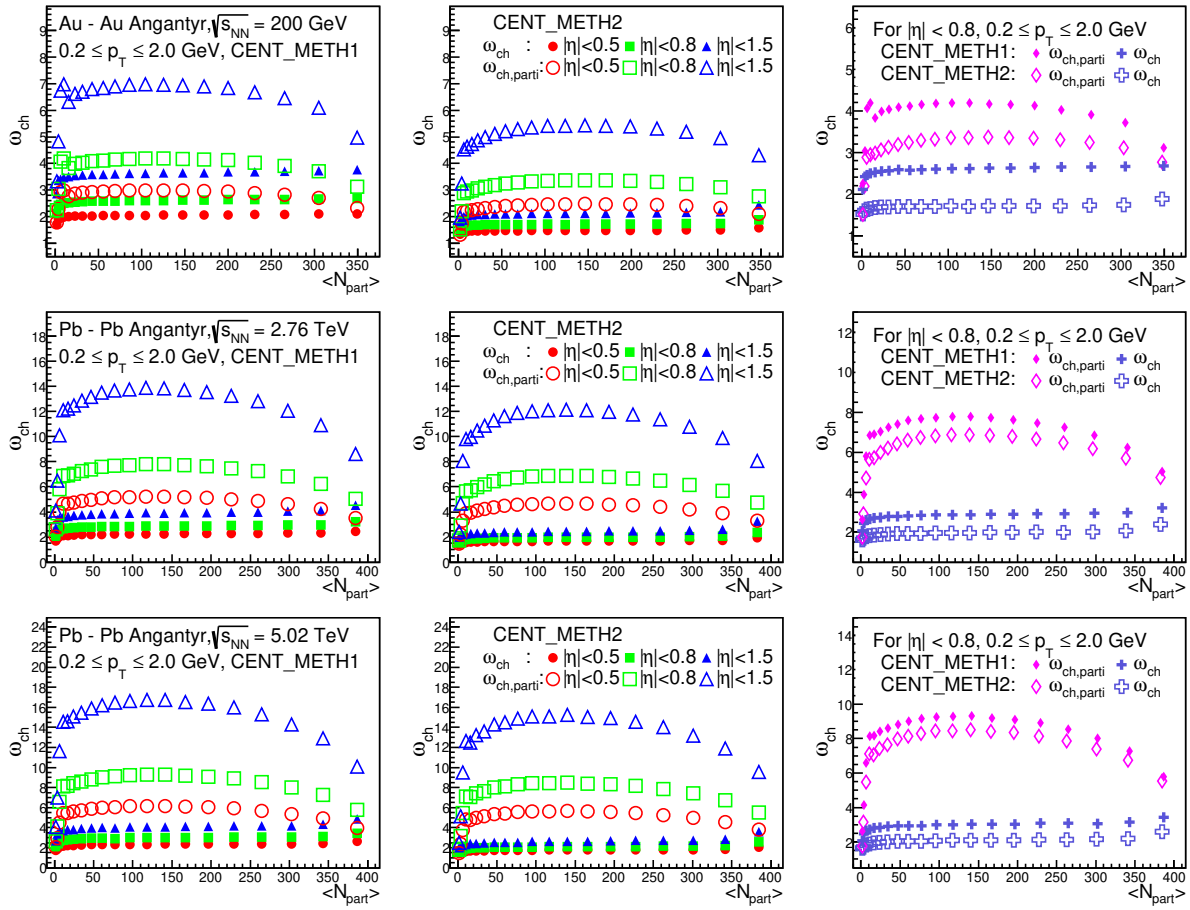


FIG. 4: (Color online) The variation of ω_{ch} with centrality for different collision systems at different centre of mass energies. The ω_{ch} values have been compared to the expectations of a simple model of participant overlap, $\omega_{ch,parti}$.

treated as a source of systematic error while estimating the ω_{ch} parameter. The observation was attributed to different values of correlation coefficient between the charged particle multiplicity and the number of participants obtained from the two methods of centrality estimation. The fluctuations in multiplicities are not observed to vary significantly as a function of centrality and beam energy. However, there is an observed centrality dependence of the multiplicity fluctuations of charged particles obtained from a simple participant model which shows a decrease with an increase in centrality. A decrease in η acceptance results in decreased multiplicity fluctuations which can be understood by a simple statistical analysis. The observed variance scaled with the square of mean multiplicity obeys a power law scaling with the number of participants for the three collision systems and the scaling is in agreement with the measured experimental data.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the Department of Science and Technology (DST), India for supporting the present work.

[1] D. Teany, Phys. Rev. **C 68**, 034913, (2003).

[2] R. A. Lacey *et al.*, Phys. Rev. Lett.**98**, 092301, (2007).

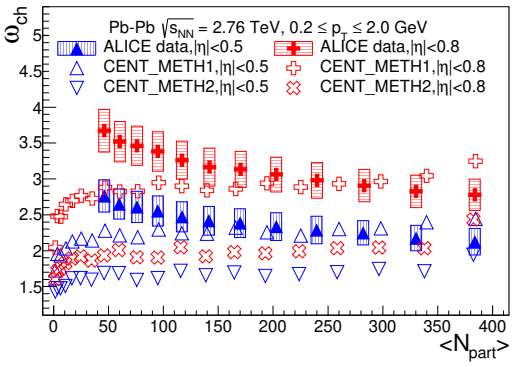


FIG. 5: (Color online) The variation of ω_{ch} as measured by ALICE experiment is compared with the predictions from Angantyr model for two different η regions.

[3] H.-J. Drescher, A. Dumitru, C. Gombeaud, and J.-Y. Ollitrault, Phys. Rev. **C 76**, 024905 (2007).

[4] K. Dusling and D. Teaney, Phys. Rev. **C 77**, 034905 (2008).

[5] Z. Xu, C. Greiner, and H. St¨ocker, Phys. Rev. Lett. **101**, 082302 (2008).

[6] D. Molnar and P. Huovinen, J. Phys. **G 35**, 104125 (2008).

[7] A. Adare *et al.*, (PHENIX Collaboration), Phys. Rev. Lett. **107**, 252301 (2011).

[8] K. Aamodt *et al.*, (ALICE Collaboration), Phys. Rev. Lett. **107**, 032301 (2011).

[9] T Sjostrand, S.'Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten,S. Mrenna, S. Prestel, C.O. Rasmussen , P.Z.Skands, Comput. Phys. Commun. **191**,159-177 (2015).

[10] J. Bellm, EPJ **C** 76(4), 196 (2016).

[11] M. Stephanov, K. Rajagopal, and E. Shuryak, Phys. Rev. **D 60**, 114028 (1999).

[12] F. Karsch, S. Ejiri, K. Redlich, Nucl. Phys. **A 774**, 619(2006)

[13] M.M. Aggarwal *et al.*, (WA98 Collaboration), Phys. Rev. **C 65**, 054912, (2002).

[14] C. Alt *et al.*, (NA49 Collaboration), Phys. Rev. **C 75**, 064904, (2007).

[15] S.S. Adler *et al.*, (PHENIX Collaboration), Phys. Rev. **C 76**, 034903 (2007).

[16] A. Adare *et al.*, (PHENIX Collaboration), Phys. Rev. **C 78**, 044902(2008).

[17] S. Acharya *et al.*, (ALICE Collaboration), Eur. Phys. J **C 81** 1032 (2021)

[18] C. Bierlich, G. Gustafson, L. Lnnblad and H. Shah, JHEP **1810**, 134 (2018).

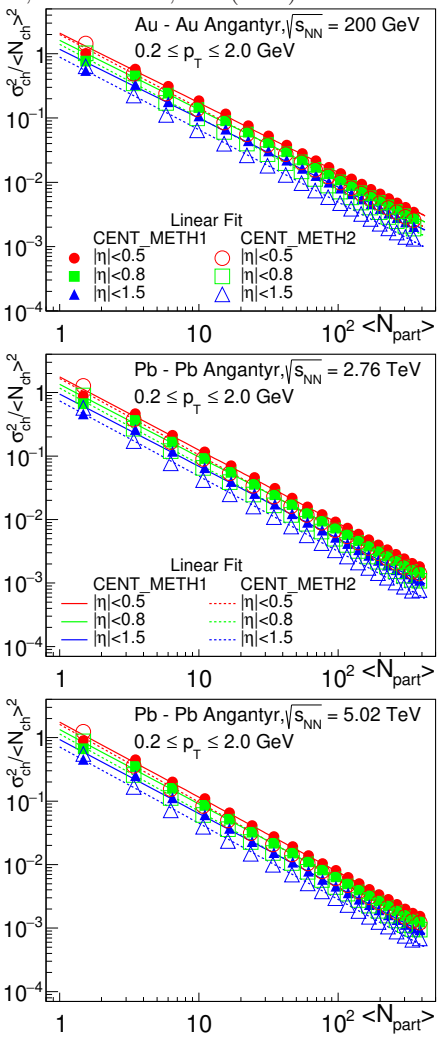


FIG. 6: (Color online) The σ^2/μ^2 for charged hadrons in the transverse momentum range $0.2 < p_T < 2.0$ GeV/c for different collision systems as a function of $< N_{part} >$. The solid line shows the fit function $lnR = a_1 ln(N_{part}) + a_0$ to the obtained values of R.

[19] Maitreyee Mukherjee, Sumit Basu, Arghya Chatterjee, Sandeep Chatterjee, Souvik Priyam Adhya, Sanchari Thakur, Tapan K. Nayak, Phys. Letts. **B 784**, 1-5 (2018).

[20] T. Sugiura, T. Nonaka, and S. Esumi, Phys. Rev. **C 100**, 044904 (2019).

[21] Maitreyee Mukherjee *et al.*, J. Phys. G: Nucl. Part. Phys. **43** , 085102 (2016).

arXiv:2403.01814v1 [hep-ph] 4 Mar 2024

The rotation effect on the deconfinement phase transition in holographic QCD

Jia-Hao Wang[1] and Sheng-Qin Feng[1,2,3,*]

[1] *College of Science, China Three Gorges University, Yichang 443002, China*

[2] *Center for Astronomy and Space Sciences and Institute of Modern Physics, China Three Gorges University, Yichang 443002, China*

[3] *Key Laboratory of Quark and Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China*

(Dated: March 5, 2024)

Abstract

Abstract: The impact of rotation on the deconfinement phase transition under the EM system of the soft and the hard wall models in holographic QCD is studied in this paper. The metric by cylindrical coordinates with rotation is introduced into the system to calculate the Hawking temperature. The first holographic study on the influence of the radius of a homogeneous rotating system on the phase diagram is proposed. It is found that the phase transition temperature hardly changes with the rotation angular velocity for a small rotation radius. Only with a larger rotation radius can the change in rotational angular velocity significantly alter the phase transition temperature. The phase transition temperature decreases rapidly with the increase of rotation angular velocity as the rotation radius increases.

* Corresponding author: fengsq@ctgu.edu.cn

I. INTRODUCTION

The phase and properties of matter become very extraordinary under rotation, which has recently attracted a lot of interest among people. This type of research is particularly relevant to the strong interacting matter in quantum chromodynamics (QCD). For example, astrophysical objects composed of dense QCD matter such as neutron stars can spin rapidly [1, 2]. Typical non-central nucleus-nucleus collisions can generate QCD matter that carries a non-zero angular momentum of the order of 10^4-$10^5\hbar$ with local angular velocities in the range of 0.01-0.1 GeV [3–5] in relativistic heavy ion collision experiments. An impressive progress [6] has also achieved to study the rotating QCD matter by using lattice simulations.

Exploring the effect of rotation on the phase transition of QCD matter is also very significant in relativistic heavy ion collisions. As is well known, non-central nucleus- nucleus collisions can generate strong external magnetic fields [7–12], which have interesting effects on the thermodynamics and phase diagrams of QCD matter [13–18]. The study of the magnetic catalytic and inverse magnetic catalytic properties of chiral condensates by using the NJL model [19–22], as well as the study of the effect of magnetic fields on phase diagrams by using holographic methods [23–31] are currently in progressing smoothly. Given the close analogy between magnetic field and rotation, it is tempting to ask how rotation could affect phase transitions. In this article, we will use holographic methods to study the effect of rotation on QCD phase transition.

It is well known that the confinement phase is at a lower temperature and density, while the deconfinement phase of QCD is at a higher temperature and density. The exploration of the phase structure of QCD is an important and challenging topic. How to probe the phase diagrams in the $T-\mu$ plane is a rather hard job because the QCD coupling constant becomes very large near the phase change region, and the traditional perturbation QCD method cannot be used. For a long time, the lattice QCD method is token as the only credible way to study the program. Although lattice QCD works well under zero baryon density, there is a sign problem when considering finite baryon density. However, the most interesting region in the QCD phase diagram is the region with finite baryon density. The QGP generated by heavy ion collisions, as well as compact stars in astrophysics, has a finite baryon density.

In holographic QCD, various approaches exist to study the deconfinement phase tran-

sition. Among these, using the Hawking-Page phase transition [32, 33] from gravitational theory to investigate the QCD deconfinement phase transition proves to be highly effective. The primary physical idea involves the phase transition between thermal AdS (tAdS) at low temperature and the Schwarzschild AdS black hole at high temperature in pure gauge theory [33], which corresponds to the confinement-deconfinement phase transition between hadron state and QGP state in QCD. Reference [34] delved into the Hawking-Page phase transition between the tAdS space and the Schwarzschild AdS black hole at zero chemical potential. Reference [35–37] examined the Hawking-Page phase transition between thermal charged (tc) AdS space and the Reissner-Nordström (RN) AdS black hole at nonzero chemical potential. Reference [38] conducted the first research work for the effect of magnetic fields on the confinement-deconfinement critical temperature in the hard wall model.

In heavy ion collisions, in addition to the presence of a strong magnetic field, the plasma also exhibits significant angular momentum. There have been some studies that introduce rotation into holography [39–41] and effective field theory [42, 43]. Braga [39] utilized the corresponding Hawking-Page phase transition between the thermal AdS space and the Schwarzschild AdS black hole, and selected a rotating cylindrical symmetric black hole model to study the confinement-deconfinement phase transition and thermodynamics of a rotating system with zero chemical potential. In this article, we will expand corresponding Hawking-Page phase transition into between thermal charged AdS space and RNAdS black hole, discussing the confinement-deconfinement phase transition and thermodynamic properties of non-zero chemical potentials under rotational conditions.

The purpose of this article is to investigate the effect of plasma rotation on the phase transition temperature T_c. This article will use holographic AdS/QCD models, especially hard wall [44–46] and soft wall [16, 47] models, to discuss the dependence of phase transition temperature T_c on rotational angular velocity, rotational radius and chemical potential. The structure of the paper is as follows: In Sec. II, we introduce the rotation into holographic QCD and discuss the consequent impacts on temperature and chemical potential. Sec. III delves into the effects of rotation on the deconfinement phase transition by utilizing both the hard and soft wall models. A summary and discussion are presented in Sec. IV.

II. THE HOLOGRAPHIC SYSTEM UNDER ROTATIONAL BACKGROUND

In this section, we will introduce some common theoretical settings in the context of rotation. To analyze the rotating charged black hole, it is necessary to assume that the spacetime has cylindrical symmetry [40, 41, 48–50]. Consequently, the metrics of Lorentzian signature in the cylindrical form of both AdS black hole (bh) and thermal charged (tc) AdS space can be uniformly reformulated as

$$ds^2_{bh,tc} = \frac{L^2}{z^2}(-f_{bh,tc}(z)dt^2 + d\vec{x}^2 + l^2 d\phi^2 + \frac{1}{f_{bh,tc}(z)}dz^2), \tag{1}$$

where L is AdS radius, ϕ is the angular coordinate describing the rotation, and l is the radius of the rotation axis. The metric function of Eq. (1) for the background geometry of AdS black hole is given as

$$f_{bh}(z) = 1 - mz^4 + q^2 z^6, \tag{2}$$

where m and q are the black hole mass and charge, respectively.

The background geometry of thermal charged AdS space is

$$f_{tc}(z) = 1 + q_1^2 z^6, \tag{3}$$

where q_1 is the charge of thermal charged AdS space. Because of boundary condition $f_{bh}(z_h) = 0$, one can obtain the mass as

$$m = \frac{1}{z_h^4} + q^2 z_h^2. \tag{4}$$

Next, the rotation will be introduced into the QCD medium system. According to Ref. [51], one obtains the rotating extension from the static configuration through a local Lorentz boost as

$$t \to \frac{1}{\sqrt{1 - l^2\omega^2}}(t + l^2\omega\phi), \tag{5}$$

and

$$\phi \to \frac{1}{\sqrt{1 - l^2\omega^2}}(\phi + \omega t), \tag{6}$$

where ω is the angular velocity. The corresponding transformation of the metric is given as

$$ds^2 = g_{tt}dt^2 + g_{t\phi}dtd\phi + g_{\phi t}d\phi dt + g_{\phi\phi}l^2 d\phi^2 + g_{zz}dz^2 + g_{xx}\sum_{i=1}^{2} dx_i^2, \tag{7}$$

with

$$g_{tt} = \frac{\gamma^2(\omega l)L^2}{z^2}(\omega^2 l^2 - f(z)), \tag{8}$$

$$g_{\phi\phi} = \frac{\gamma^2(\omega l)L^2}{z^2}(1 - \omega^2 l^2 f(z)), \tag{9}$$

$$g_{t\phi} = g_{\phi t} = \frac{\gamma^2(\omega l)L^2}{z^2}\omega l^2(1 - f(z)), \tag{10}$$

$$g_{zz} = \frac{L^2}{z^2 f(z)}, \tag{11}$$

$$g_{xx} = \frac{L^2}{z^2}, \tag{12}$$

where $\gamma(\omega l)$ is the Lorentz factor

$$\gamma(\omega l) = \frac{1}{\sqrt{1 - l^2\omega^2}}. \tag{13}$$

The Lorentz boost determines the relation between the Hawking temperature T of the black hole and the horizon position. Reestablishing the canonical form of rotation metric given by Eq. (7), as shown in Refs. [40, 48], one yields the following metric as

$$ds^2 = \frac{L^2}{z^2}\left[-N(z)^2 f(z)dt^2 + \frac{dz^2}{f(z)} + R(z)(d\phi + P(z)dt)^2 + \sum_{i=1}^{2} dx_i^2\right], \tag{14}$$

with

$$N(z)^2 = \frac{(1 - \omega^2 l^2)}{1 - f(z)\omega^2 l^2}, \tag{15}$$

$$R(z) = \gamma^2 l^2 - f(z)\gamma^2\omega^2 l^4, \tag{16}$$

$$P(z) = \frac{\omega(1 - f(z))}{1 - f(z)\omega^2 l^2}, \tag{17}$$

where $N(z)$ is the lapse function and $P(z)$ is the shift function.

Following Refs. [51, 52], one can obtain the following expressions for the Hawking temperature of rotating black holes:

$$T_H = -\frac{N(z_h)f'(z_h)}{4\pi}. \tag{18}$$

It should be noted that due to the Lorentz transformation of time and angular rotation components, the expression of the chemical potential differs from that of the static case. The quark chemical potential under rotation can be given as [53, 54]

$$\mu = A_\mu \chi^\mu|_{z=z_h} - A_\mu \chi^\mu|_{z=0}, \tag{19}$$

where $\chi = \partial_t + \Omega\partial_\phi$ is the killing vector, and the gauge potential A_μ under rotation is given as

$$A_\mu = A_t(\gamma\delta^t_\mu + \gamma\omega l^2\delta^\phi_\mu). \tag{20}$$

Substituting (20) into (19), one obtains the chemical potential as

$$\mu = \mu'\sqrt{1-\omega^2}. \tag{21}$$

The Ricci tensor R_{mn} and metric g_{mn} are second rank tensors, which transform under coordinate transformation of rotation to

$$T'_{\alpha\beta} = \frac{\partial x^m}{\partial x'^\alpha}\frac{\partial x^n}{\partial x'^\beta}T_{mn} \equiv M^{mn}_{\alpha\beta}T_{mn}. \tag{22}$$

However, for scalar coordinate transformations, the situation is different. For example: as one defines $\mathcal{R}$ as the curvature scalar, and $\mathcal{R}'$ as the curvature scalar with coordinate transformation of rotation, this transformation relation should be

$$\mathcal{R}' = g^{mn}R_{mn} = M^{mn}_{\alpha\beta}g^{\alpha\beta}M^{\alpha\beta}_{mn}R_{\alpha\beta} = g^{\alpha\beta}R_{\alpha\beta} = \mathcal{R}, \tag{23}$$

so, we can conclude that scalar quantity is not affected by the coordinate transformation of rotation.

III. THE PHASE DIAGRAM UNDER ROTATION BACKGROUND

In this section, we will discuss holographic QCD phase transition using the hard wall and soft wall model, respectively.

A. Hard wall model

Following Refs. [35, 37], the Euclidean action of the Einstein-Maxwell (EM) system describing holographic light quarks for the hard wall model is given by:

$$S = \int d^5x\sqrt{G}\left[\frac{1}{2\kappa^2}(-\mathcal{R}+2\Lambda) + \frac{1}{4g^2}F_{MN}F^{MN}\right], \tag{24}$$

where G is the determinant of metric, κ is the five-dimensional Newton constant, $\mathcal{R}$ is the Ricci scalar, Λ is the cosmological constant, g is the 5D gauge coupling constant, and F_{MN}

is the U(1) bulk gauge field strength tensor with $F_{MN} = \partial_M A_N - \partial_N A_M$. According to Refs. [35, 55, 56], the bulk gauge field A_M is

$$A_t = \frac{\mu}{\sqrt{1-\omega^2 l^2}} - \rho z^2 \ , \ A_i = A_z = 0 \ (i = 1, 2, 3), \tag{25}$$

where ρ is the quark number density, which is related to the black hole charge q

$$\rho = \sqrt{\frac{3g^2L^2}{2k^2}} q. \tag{26}$$

By imposing the Dirichlet boundary condition $A_t(z_h) = 0$ near the boundary at horizon, one can rewrite ρ as

$$\rho = \frac{\mu}{z_h^2\sqrt{1-\omega^2 l^2}}. \tag{27}$$

The five-dimensional Newton constant κ, and the 5D gauge coupling constant g are given as [56],

$$\frac{1}{\kappa^2} = \frac{N_c^2}{4\pi^2 L^3}, \tag{28}$$

and

$$\frac{1}{g^2} = \frac{N_c N_f}{4\pi^2 L}, \tag{29}$$

where N_c and N_f are the number of colors and flavors. From Eq. (26) to Eq. (29), one can obtain

$$q = \sqrt{\frac{2N_f}{3N_c}} \frac{\mu}{z_h^2\sqrt{1-\omega^2 l^2}}. \tag{30}$$

The boundary condition at the IR cut off $z = z_{IR}$ is

$$A(z_{IR}) = \alpha \frac{\mu}{\sqrt{1-\omega^2 l^2}}, \tag{31}$$

where α is a parameter, which can be determined as $\alpha = -\frac{1}{2}$ given by Ref. [35]. The quark number density ρ is given

$$\rho = \frac{3\mu}{2z_{IR}^2\sqrt{1-\omega^2 l^2}}. \tag{32}$$

By comparing Eq. (25) and Eq. (31), one can derive the charge of thermal charged AdS space q_1 as

$$q_1 = \sqrt{\frac{3N_f}{2N_c}} \frac{\mu}{z_{IR}^2\sqrt{1-\omega^2 l^2}}. \tag{33}$$

From Eq. (24), the corresponding on-shell gravity action of RNAdS BH and tcAdS space for the hard wall model are

$$S_{BH} = \frac{L^3 V_3}{\kappa^2} \int_0^\beta dt \int_\epsilon^{z_h} dz \left(\frac{4}{z^5} - 2q^2 z\right), \tag{34}$$

$$S_{tc} = \frac{L^3 V_3}{\kappa^2} \int_0^{\beta'} dt \int_\epsilon^{z_{IR}} dz (\frac{4}{z^5} - 2q_1^2 z), \tag{35}$$

where V_3 is the volume of the 3D space, $\epsilon(\epsilon \to 0)$ is the UV cut-off, β and β' are the time periodicity in RNAdS BH and tcAdS space, respectively. At $z = \epsilon$, β' is related to β as follows:

$$\beta' \sqrt{f_t(\epsilon)} = \beta \sqrt{f_b(\epsilon)}. \tag{36}$$

The on-shell action density difference between RNAdS BH and tcAdS space is given as:

$$\Delta\mathcal{E} = \frac{S_{BH} - S_{tc}}{V_3} = \frac{L^3}{\kappa^2} \left[\beta \int_\epsilon^{z_h} dz \left(\frac{4}{z^5} - 2q^2 z \right) - \beta' \int_\epsilon^{z_{IR}} dz \left(\frac{4}{z^5} - 2q_1^2 z \right) \right]. \tag{37}$$

The free energy density is given as $F = -(T \ln \mathcal{Z})/V_3$, where $\mathcal{Z}$ is the partition function. It should be noted that, in our pursuit of studying the thermodynamics of this hard wall model, we utilize the compactified imaginary time metric (Euclidean signature). Thus, the free energy is expressed as:

$$F = -\frac{T \ln \mathcal{Z}}{V_3} \approx \frac{T S_E}{V_3}. \tag{38}$$

Therefore, the difference of free energy density between RNAdS BH and tcAdS space is

$$\begin{aligned} \Delta F =& T\Delta\mathcal{E} = \frac{3N_c(-1+\omega^2 l^2) + N_f z_h^2 \mu^2}{18 N_c^2 \pi (1-\omega^2 l^2)^{\frac{3}{2}} z_{IR}^4 z_h^5} \\ & \times [-3N_c(-1+\omega^2 l^2)(z_{IR}^4 - 2z_h^4) + N_f z_{IR}^2 z_h^2 (2z_{IR}^2 - 9z_h^2)\mu^2], \end{aligned} \tag{39}$$

where $z_{IR} = 1/(0.323GeV)$ is obtained from the lightest ρ meson mass [34].

The dependence of ΔF on z_h with different μ and ω for the hard wall with $m = N_f/N_c = 1$ is manifested in Fig. 1.

When ΔF is positive, the tcAdS space is stable. On the contrary, when ΔF is negative, the RNAdS black hole is stable. According to Hawking-Page phase transition, the tcAdS space corresponds to confinement phase, the RNAdS black hole corresponds to deconfinement phase. Therefore, a positive (negative) ΔF suggests that the QCD system is in the confinement (deconfinement) phase. Thus, when the difference of free energy density equals zero, the corresponding horizon position marks the phase transition point for the Hawking-Page phase transition or confinement-deconfinement phase transition.

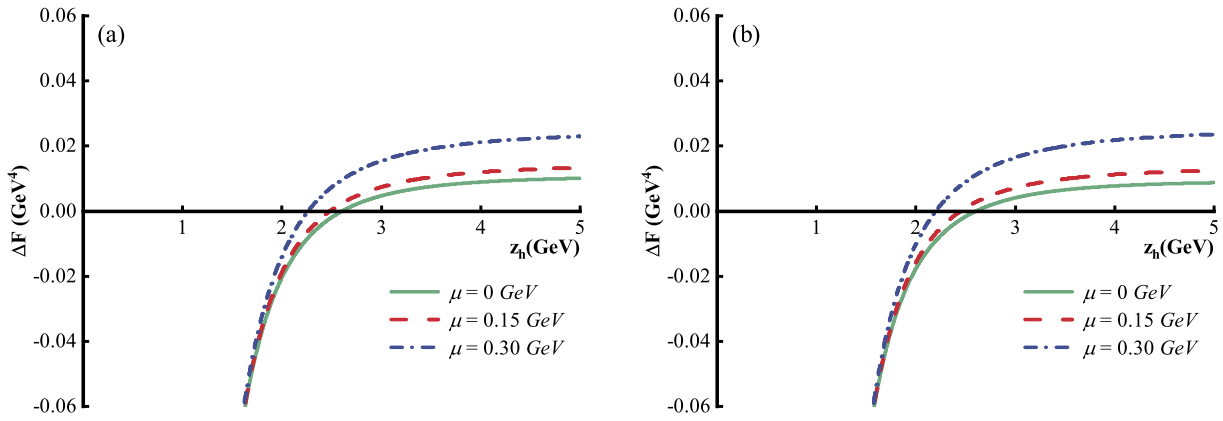


FIG. 1. The difference of free energy density of rotational system ΔF as a function of the horizon position z_h for different chemical potential for the hard wall model. (a) with $\omega = 0$ GeV. and (b) with $\omega = 0.5$ GeV

From Fig. 1(a) and 1(b), one finds that at a constant angular velocity ω, the horizon position of the phase transition point becomes small with the increase of the chemical potential μ. Comparing Fig.1 (a) and Fig. 1 (b), one also finds that at a fixed μ, the horizon position of the phase transition point becomes small with the increase of the angular velocity ω. It is found that chemical potential and angular velocity have the same effects on the horizon transition position., both of which cause the position of the phase transition point to decrease. Let's figure out what cause the reason in the following.

In the context of a rotating QCD system, we know that the introduction of rotation alters various physical quantities. Refs. [57, 58] point out that the influence of rotation on thermodynamic quantities should have a modification of the thermodynamic relation as

$$\epsilon = -P + Ts + \mu\rho + \omega J, \tag{40}$$

where ϵ, P, s, ρ and J are the energy density, pressure, entropy density, quark number density and angular momentum. In (40), it is evident that the energy density of the system also increases with ω and μ increase. Following the law of the black hole thermodynamics [59], the energy density of black hole correlates with its other properties as follows:

$$dE \propto dM \propto dA \propto dr \propto \frac{1}{z_h}, \tag{41}$$

where M, A and r denote the mass, area and radius of black hole, respectively. One can

find that the horizon position decreases with the increase of energy density. This explains why z_h decreases with the increase of ω and μ.

In order to describe the status of a two-dimensional ideal rotating system, two critical physical quantities must be considered: angular velocity and radii of rotation. Therefore, our study encompasses not only the variation of phase transitions with rotation and chemical potential but also its phase diagram on the radius of rotation.

In Fig. 2, the phase transition temperature is shown as a function of the angular velocity with different chemical potential for different rotational radiuses $l = 0.3\text{GeV}^{-1}$, $l = 0.6\text{GeV}^{-1}$ and $l = 0.9\text{GeV}^{-1}$, respectively. One finds that small angular velocity has less impact on transition temperature while large angular velocity induces a quick decrease of phase transition temperature. It is also found that the critical temperature T_c decreases with the increase of angular velocity but has weak dependence of chemical potential, which is analogous to some holographic work and PNJL model of deconfinement transition of light flavor [60–62]. There it has been found that the presence of anisotropy of rotation causes easier dissociation of the $Q\bar{Q}$ and that in phase transitions anisotropy takes a catalytic reduction of the critical temperature.

Since we are discussing the rotational system of QCD medium, the phase transition feature should rely on the finite size of the rotational system. Due to the cylindrical symmetry, these quantities are dependent on the rotation radius l. It would be interesting to probe how the phase transition feature in a strongly interacting rotating matter depends on the radius of the rotating system. The properties as a function of the radius of the rotating system may be connected to experimental observations in the future. And due to the limitation of the speed of light, it naturally causes the limitations $\omega l \leq 1$. Comparing Fig.2 (a) ($l = 0.3\text{GeV}^{-1}$), Fig.2 (b) ($l = 0.6\text{GeV}^{-1}$) and Fig.2 (c) ($l = 0.9\text{GeV}^{-1}$), one finds that the phase transition temperature hardly changes with the rotation angular velocity for small rotation radii ($\text{l} = 0.3\text{GeV}^{-1}$), but as the rotation radius increases, the phase transition temperature decreases rapidly with the increase of rotation angular velocity.

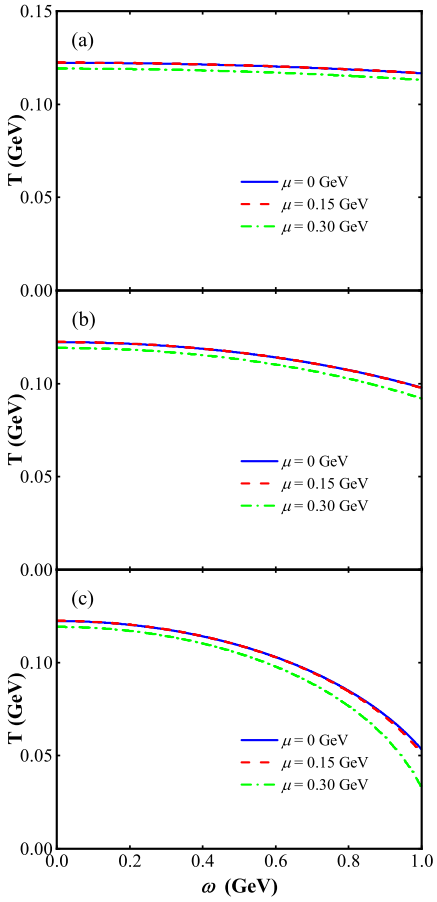


FIG. 2. Phase diagrams $T-\omega$ with different chemical potential of $\mu = 0$ GeV, $\mu = 0.15$ GeV and $\mu = 0.3$ GeV for the hard wall model. (a) for $l = 0.3$ GeV^{-1}, (b) for $l = 0.6$ GeV^{-1} and (c) for $l = 0.9$ GeV^{-1}.

B. Soft wall model

For soft wall model [36], the Euclidean action of the EM system is given by:

$$S = \int d^5x\sqrt{G}e^{-\Phi}\left[\frac{1}{2\kappa^2}(-\mathcal{R}+2\Lambda)+\frac{1}{4g^2}F_{MN}F^{MN}\right], \tag{42}$$

where Φ is is the non-dynamical scalar field with $\Phi = cz^2$, c is the IR energy parameter. The RNAdS BH charge q and tcAdS space charge q_1 under the rotation background in the soft wall model are given as

$$q = \sqrt{\frac{2N_f}{3N_c}}\frac{\mu}{z_h^2\sqrt{1-\omega^2}}, \tag{43}$$

and

$$q_1 = \sqrt{\frac{3N_f}{2N_c}}\frac{\mu c}{\sqrt{1-\omega^2}}. \tag{44}$$

The corresponding on-shell gravity action of RNAdS BH and tcAdS space are

$$S_{BH} = \frac{L^3 V_3}{\kappa^2}\int_0^{\beta} dt \int_{\epsilon}^{z_h} dz e^{-cz^2}(\frac{4}{z^5} - 2q^2 z), \tag{45}$$

$$S_{tc} = \frac{L^3 V_3}{\kappa^2}\int_0^{\beta'} dt \int_{\epsilon}^{\infty} dz e^{-cz^2}(\frac{4}{z^5} - 2q_1^2 z). \tag{46}$$

Contrasted with the hard wall model, the integral term is multiplied by a quadratic exponent of z, and the upper limit of the integral extends from z_{IR} to infinity for the soft wall model. The free energy density difference ΔF between RNAdS BH and tcAdS space is given as

$$\begin{aligned}\Delta F =& T\Delta\mathcal{E} = -\frac{1}{6cN_c\sqrt{1-\omega^2 l^2}z_h^4}e^{-cz_h^2} \\ &\times\{(6c^3 N_c z_h^4 e^{cz_h^2}(\omega^2 l^2 - 1)Ei(-cz_h^2) - 3c^2 z_h^2(3\mu^2 N_f z_h^2 e^{cz_h^2} + N_c(2 - 2\omega^2 l^2)) \\ &+ c(3N_c(e^{cz_h^2} - 2)(\omega^2 l^2 - 1) - 2\mu^2 N_f z_h^2 e^{cz_h^2}) + 4\mu^2 N_f(e^{cz_h^2} - 1)\},\end{aligned} \tag{47}$$

where the parameter c can be matched to the mass of the lightest ρ meson [34], denoted as $\sqrt{c} = 0.338\text{GeV}$, and $E_i(z)$ is the exponential integral function, which can be expressed as

$$Ei(z) = -\int_{-z}^{\infty} dt\frac{e^{-t}}{t}. \tag{48}$$

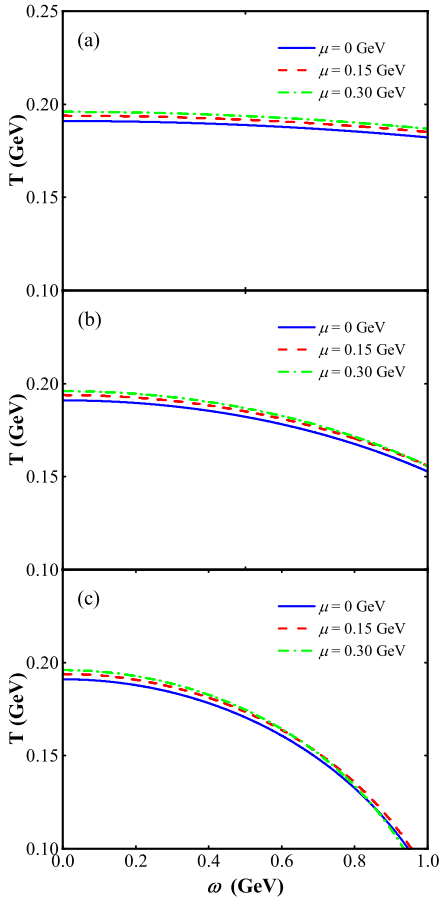


FIG. 3. Phase diagrams $T-\omega$ with different chemical potential of $\mu = 0$ GeV, $\mu = 0.15$ GeV and $\mu = 0.3$ GeV for the soft wall model. (a) for $l = 0.3$ GeV^{-1}, (b) for $l = 0.6$ GeV^{-1} and (c) for $l = 0.9$ GeV^{-1}.

Figure 3 shows the phase diagrams $T-\omega$ with different chemical potentials and different rotation radii for the soft wall model. It is found that the characteristics of the phase diagram of soft wall case are generally similar to that of hard wall case, but there are also some differences. For example, when analyzing the dependence of phase transition temperature on chemical potential, it is found that for the hard wall model, the larger the chemical potential is, the smaller the phase transition temperature is. However, for the soft wall model, the larger the chemical potential is, the larger the phase transition temperature is.

C. The comparison of phase diagram between the hard wall and the soft wall

The comparisons of phase diagram between the hard wall and the soft wall are published in Fig.4. Compared the soft wall model with hard wall model, the trend of phase transition temperature T_c decreasing with angular velocity ω is almost the same. However, the phase transition temperature given by soft wall model is obviously larger than that of hard wall model under the same conditions. As the rotation radius l increases, the decreasing speed of the phase transition with angular velocity for the soft wall model is faster than that of the hard wall model.

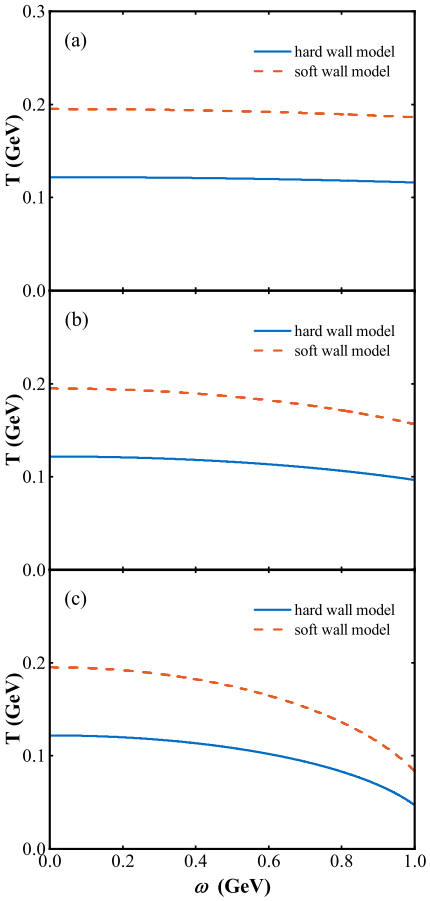


FIG. 4. The phase diagram $T-\omega$ comparison between the hard wall model and the soft wall model with chemical potential of $\mu = 0.2$ GeV. (a) for $l = 0.3$ GeV^{-1}, (b) for $l = 0.6$ GeV^{-1} and (c) for $l = 0.9$ GeV^{-1}.

IV. SUMMARY AND CONCLUSIONS

In this paper, we focus on the impact of rotation on the deconfinement phase transition under the EM system of the soft and the hard wall model in holographic QCD. The metric by cylindrical coordinates with rotation is introduced into the system to calculate the Hawking temperature. Whether it is a soft wall model or a hard wall model, the phase transition temperatures decrease with the increase of rotation angular velocity, but have weak dependence of chemical potential.

The first holographic study on the influence of the radius of a homogeneous rotating system on the phase diagram is established in the article. As we are discussing the rotating system of QCD medium, the phase transition characteristics should depend on the finite size of the rotating system. Due to the cylindrical symmetry of the rotating system, the rotation radius l has become an important characteristic quantity of the rotating system. Studying the dependence of phase transition characteristics of the strongly interacting rotating matter on the radius of the rotating system is an important research topic. It is found that the phase transition temperature hardly changes with the rotation angular velocity for a small rotation radius, but as the rotation radius increases, the phase transition temperature decreases rapidly with the increase of rotation angular velocity.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grants No. 11875178, No. 11475068, No. 11747115).

REFERENCES

[1] I. A. Grenier and A. K. Harding, C. R. Phys. **16**, 641 (2015).

[2] A. L. W. et al., Rev. Mod. Phys. **88**, 021001 (2016).

[3] Z. T. Liang and X. N. Wang, Phys. Rev. Lett. **94**, 102301 (2005).

[4] W.-T. Deng and X.-G. Huang, Phys. Rev. C **93**, 064907 (2016).

[5] L. G. Pang, H. Petersen, Q. Wang, and X. N. Wang, Phys. Rev. Lett. **117**, 192301 (2016).

[6] A. Yamamoto and Y. Hirono, Phys. Rev. Lett. **111**, 081601 (2013).

[7] V. Voronyuk, V. D. Toneev, W. Cassing, E. L. Bratkovskaya, V. P. Konchakovski, and S. A. Voloshin, Phys. Rev. C **83**, 054911 (2011).

[8] A. Bzdak and V. Skokov, Phys. Lett. B **710**, 171 (2012).

[9] W.-T. Deng and X.-G. Huang, Phys. Rev. C **85**, 044907 (2012).

[10] Y. Zhong, C.-B. Yang, X. Cai, and S.-Q. Feng, Adv. High Energy Phys. **2014**, 1 (2014).

[11] D. She, S.-Q. Feng, Y. Zhong, and Z.-B. Yin, Eur. Phys. J. A **54**, 48 (2018).

[12] Y. Guo, S. Shi, S. Feng, and J. Liao, Phys. Lett. B **798**, 134929 (2019).

[13] X. Cao, S. Qiu, H. Liu, and D. Li, J. High Energy Phys. 08 , 005 (2021).

[14] J. Grefa, M. Hippert, J. Noronha, J. Noronha-Hostler, I. Portillo, C. Ratti, and R. Rougemont, Phys. Rev. D **106**, 034024 (2022).

[15] J. Erlich, E. Katz, D. T. Son, and M. A. Stephanov, Phys. Rev. Lett. **95**, 261602 (2005).

[16] A. Karch, E. Katz, D. T. Son, and M. A. Stephanov, Phys. Rev. D **74**, 015005 (2006).

[17] T. M. Kelley, S. P. Bartz, and J. I. Kapusta, Phys. Rev. D **83**, 016002 (2011).

[18] Y.-Q. Sui, Y.-L. Wu, Z.-F. Xie, and Y.-B. Yang, Phys. Rev. D **81**, 014024 (2010).

[19] K. Xu, J. Chao, and M. Huang, Phys. Rev. D **103**, 076015 (2021).

[20] Y.-W. Qiu and S.-Q. Feng, Phys. Rev. D **107**, 076004 (2023).

[21] Y.-W. Qiu, S.-Q. Feng, and X.-Q. Zhu, Phys. Rev. D **108**, 116022 (2023).

[22] X. Zhu and S.-Q. Feng, Phys. Rev. D **107**, 016018 (2023).

[23] X. Chen, S.-Q. Feng, Y.-F. Shi, and Y. Zhong, Phys. Rev. D **97**, 066015 (2018).

[24] J. Deng and S.-Q. Feng, Phys. Rev. D **105** (2022), 10.1103/PhysRevD.105.026015.

[25] D. M. Rodrigues, E. F. Capossoli, and H. Boschi-Filho, Phys. Lett. B **780**, 37 (2018).

[26] D. M. Rodrigues, E. F. Capossoli, and H. Boschi-Filho, Phys. Rev. D **97**, 126001 (2018).

[27] D. M. Rodrigues, D. Li, E. F. Capossoli, and H. Boschi-Filho, Phys. Rev. D **98**, 106007 (2018).

[28] B. McInnes, Nucl. Phys. B **906**, 40 (2016).

[29] U. Gürsoy, M. Järvinen, and G. Nijs, Phys. Rev. Lett. **120**, 242002 (2018).

[30] U. Gürsoy, I. Iatrakis, M. Järvinen, and G. Nijs, J. High Energy Phys. 03 , 053 (2017).

[31] D. Dudal and T. G. Mertens, Phys. Rev. D **97**, 054035 (2018).

[32] S. W. Hawking and D. N. Page, Commun. Math. Phys. **87**, 577 (1983).

[33] E. Witten, Adv. Theor. Math. Phys. **2**, 505 (1998).

[34] C. P. Herzog, Phys. Rev. Lett. **98**, 091601 (2007).

[35] B.-H. Lee, C. Park, and S.-J. Sin, J. High Energy Phys. 07 , 087 (2009).

[36] C. Park, D.-Y. Gwak, B.-H. Lee, Y. Ko, and S. Shin, Phys. Rev. D **84**, 046007 (2011).

[37] C. Park, Phys. Rev. D **81**, 045009 (2010).

[38] K. A. Mamo, J. High Energy Phys. 05 , 121 (2015).

[39] N. R. F. Braga, L. F. Faulhaber, and O. C. Junqueira, Phys. Rev. D **105**, 106003 (2022).

[40] M. B. Gaete, L. Guajardo, and M. Hassaïne, J. High Energy Phys. 04 , 092 (2017).

[41] C. Erices and C. Martínez, Phys. Rev. D **97**, 024034 (2018).

[42] F. Sun and A. Huang, Phys. Rev. D **106**, 076007 (2022).

[43] Y. Jiang and J. Liao, Phys. Rev. Lett. **117**, 192302 (2016).

[44] J. Polchinski and M. J. Strassler, Phys. Rev. Lett. **88**, 031601 (2002).

[45] H. Boschi-Filho and N. R. F. Braga, Eur. Phys. J. C **32**, 529 (2004).

[46] H. Boschi-Filho and N. R. F. Braga, J. High Energy Phys. 05 , 009 (2003).

[47] C. A. B. Bayona, H. Boschi-Filho, N. R. F. Braga, and L. A. P. Zayas, Phys. Rev. D **77**, 046002 (2008).

[48] J. S. Lemos, Phys. Lett. B **353**, 46 (1995).

[49] J. P. S. Lemos and V. T. Zanchin, Phys. Rev. D **54**, 3840 (1996).

[50] H. Nadi, B. Mirza, Z. Sherkatghanad, and Z. Mirzaiyan, Nucl. Phys. B **949**, 114822 (2019).

[51] X. Chen, L. Zhang, D. Li, D. Hou, and M. Huang, J. High Energy Phys. 07 , 132 (2021).

[52] J. D. Brown, J. Creighton, and R. B. Mann, Phys. Rev. D **50**, 6394 (1994).

[53] M. Cvetic and S. S. Gubser, J. High Energy Phys. 04 , 024 (1999).

[54] M. M. Caldarelli, G. Cognola, and D. Klemm, Class. Quantum Grav. **17**, 399 (2000).

[55] S. W. Hawking and S. F. Ross, Phys. Rev. D **52**, 5865 (1995).

[56] S.-J. Sin, J. High Energy Phys. 10 , 078 (2007).

[57] Y.-Q. Zhao, S. He, D. Hou, L. Li, and Z. Li, J. High Energy Phys. 04 **2023**, 115 (2023).

[58] F. Sun, S. Li, R. Wen, A. Huang, and W. Xie, arXiv e-prints , arXiv:2310.18942 (2023).

[59] M. Natsuume, Lect. Notes Phys. (2015), 10.1007/978-4-431-55441-7.

[60] L. McLerran, K. Redlich, and C. Sasaki, Nucl. Phys. A **824**, 86 (2009).

[61] X. Chen, D. Li, D. Hou, and M. Huang, J. High Energy Phys. 03 **2020**, 073 (2020).

[62] H. Abuki, R. Anglani, R. Gatto, G. Nardulli, and M. Ruggieri, Phys. Rev. D **78**, 034034 (2008).

Dark matter semi-annihilation for inert scalar multiplets

Hugues Beauchesne[a] **and Cheng-Wei Chiang**[b,a]

[a] *Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan*

[b] *Department of Physics and Center for Theoretical Physics, National Taiwan University, Taipei 10617, Taiwan*

E-mail: beauchesneh@phys.ncts.ntu.edu.tw, chengwei@phys.ntu.edu.tw

Abstract: Dark matter semi-annihilation is a process through which two dark matter candidates annihilate to a single dark matter particle and a non-dark matter particle. Such processes are common when the symmetry stabilizing the dark matter differs from $\mathbb{Z}_2$ and can lead to qualitatively different phenomenology. In this work, we study the viability of semi-annihilation models including one or two inert multiplets. For one multiplet, we show that there does not exist any viable model in which semi-annihilation is efficient. For two multiplets, semi-annihilation can be efficient, but the number of viable and technically natural models is limited. We then perform a detailed study of the most promising model, showing that the correct relic abundance can be obtained for a wide range of masses.

arXiv:2403.01729v1 [hep-ph] 4 Mar 2024

Contents

1 Introduction **1**

2 One multiplet case **3**
2.1 Model 3
2.2 Comments 3

3 Two multiplets case **5**

4 Model definition **7**

5 Dark matter abundance computation **8**
5.1 Computation for unbroken electroweak symmetries 8
5.2 Computation for broken electroweak symmetries 11
5.3 Comparison of the methods 12

6 Constraints **12**
6.1 Direct detection experiments 13
6.2 Unitarity 13
6.3 Other constraints 14

7 Results **14**

8 Conclusion **18**

A $SU(2)$ tensors **19**
A.1 One multiplet case 19
A.2 Two multiplets case 20

1 Introduction

The preferred explanation for dark matter abundance has long been the standard freeze-out of Weakly Interacting Massive particles (WIMP). In this scenario, two dark matter particles of masses around the electroweak (EW) scale annihilate into Standard Model (SM) particles until the corresponding rate drops below the expansion rate of the Universe. Although this can naturally explain the observed dark matter abundance, standard WIMP thermal freeze-out is now under severe experimental constraints.

A potential explanation for dark matter that avoids many of the problems of standard thermal freeze-out is semi-annihilation [1]. In this mechanism, two dark matter candidates annihilate to a single dark matter particle and another particle that is not part of the dark

matter content. These processes are common in models in which the dark matter stability is explained by a symmetry other than the common $\mathbb{Z}_2$ symmetry. Semi-annihilation processes have the advantage that they, by their very nature, do not directly lead to direct detection signals. This generally results in a qualitatively different allowed parameter space. Previous works on semi-annihilation include Refs. [2–20].

Some of the most studied dark matter candidates are scalar multiplets.[1] Multiplets considered in previous studies include doublets [21–28], triplets [21, 23, 25, 26, 28–35], quadruplets [21, 23, 25, 28, 36–38], quintuplets [21, 23, 28, 34, 35, 39–43], sextuplets [21, 23, 28, 40, 44–46], septuplets [21, 23, 34, 35, 39–41, 43, 47, 48] and octuplets [28, 40, 44, 45]. Combinations of different multiplets were also considered in Refs. [36, 37, 49–55]. What almost all of these papers have in common is the assumption that dark matter is stabilized by a $\mathbb{Z}_2$ symmetry, which means that the dark matter abundance is set by the annihilation of dark matter particles to SM particles and not semi-annihilation.

In this paper, we study the possibility of semi-annihilation for inert multiplets. More precisely, we seek to determine which models of inert multiplets can lead to efficient semi-annihilation, while at the same time avoiding stable charged particles in a technically natural way and being compatible with known constraints. Models with one or two inert multiplets are considered. The most promising model is then examined in more detail by computing the dark matter density, studying various constraints, and mapping the region of parameter space compatible with the measured relic abundance.

We find the following results. Models with a single inert multiplet cannot lead to efficient semi-annihilation. For two multiplets, there exists a category of models that can lead to efficient semi-annihilation and no stable charged particles even at tree level. It also satisfies one of two possible conditions to suppress direct detection signal for inert scalar multiplets, i.e., the dark matter candidate being made up mostly of odd-dimensional multiplets of zero weak hypercharge or being almost pure CP-even or CP-odd scalars. Other models with two inert multiplets and efficient semi-annihilation exist, but cannot satisfy all of these properties simultaneously. For smaller multiplets, the existence of semi-annihilation processes allows multiplets to account for the correct relic abundance for masses varying by more than an order of magnitude. The allowed relative range shrinks as the size of the multiplets increases. Some combinations of gauge numbers can also lead to a strongly suppressed signal from indirect detection experiments.

The paper is organized as follows. Sec. 2 presents the case of a single multiplet and explains why it fails. Sec. 3 considers the case of two multiplets. The model that we will focus on is presented in Sec. 4. The details of the dark matter abundance computation are given in Sec. 5. The procedure through which constraints are applied is explained in Sec. 6. Plots of the allowed parameter space are shown in Sec. 7. Concluding remarks are given in Sec. 8, including a discussion of indirect detection signals. Some technical details about $SU(2)$ tensors are presented in Appendix A.

[1] Throughout this article, by "multiplet" we refer to the $SU(2)_L$ multiplets that are color-neutral under $SU(3)_C$.

2 One multiplet case

We begin by introducing the only unexcluded model that allows for semi-annihilation with only one multiplet. We then explain why it fails and what a viable model would require.

2.1 Model

Consider a complex scalar multiplet ϕ of dimension n and weak hypercharge 0. The relevant renormalizable potential is

$$V = V_0 + V_A + V_B + V_C, \tag{2.1}$$

where

$$\begin{aligned} V_0 &= -\mu^2|H|^2 + \lambda_0|H|^4 + m^2|\phi|^2, & V_A &= \lambda_1 A_{abc}\phi^a\phi^b\phi^c + \text{h.c.}, \\ V_B &= \sum_{r=1}^{\alpha} \lambda_2^r B^r_{abcd}(H^a)^\dagger H^b(\phi^c)^\dagger\phi^d, & V_C &= \sum_{r=1}^{\beta} \lambda_3^r C^r_{abcd}(\phi^a)^\dagger(\phi^b)^\dagger\phi^c\phi^d, \end{aligned} \tag{2.2}$$

where the mass parameters μ^2 and $m^2 > 0$, λ_1 has the dimension of mass, and $\lambda_{2,3}$ are dimensionless. The indices represented by Latin letters besides r correspond to $SU(2)$ indices and are summed over. The invariant tensors of the potential are presented in Appendix A.1. The A_{abc} term is non-zero for $n \in \{1, 5, 9, ...\}$. The upper limits on the sums are given by

$$\alpha = \min(2, n), \quad \beta = \left\lfloor \frac{n+1}{2} \right\rfloor, \tag{2.3}$$

where $\lfloor x \rfloor$ means x rounded down to the closest integer. The potential respects a $\mathbb{Z}_3$ symmetry under which only ϕ transforms in the following way:

$$\phi \to e^{2\pi i/3}\phi. \tag{2.4}$$

Unless $\mathbb{Z}_3$ is spontaneously broken, the lightest component of ϕ will be stable.

2.2 Comments

We now comment on the model and explain why it fails as a viable semi-annihilation model.

Besides the standard Higgs potential and multiplet masses, the potential consists of three types of terms. This separation will also apply to all other models defined in this work. Type A terms contain exactly three multiplets and in this potential take the form $\lambda_1\phi^3 + \text{h.c.}$ Without such terms, semi-annihilation would not be possible. Type B terms are those that involve two ϕ's and Higgs doublets. They are responsible for splitting the masses of the components of the multiplet and also contribute to certain semi-annihilation processes. Type C terms are dark matter self-interactions. Their exact details will not affect any computations in this paper.

This model has the nice property that it does not lead to any interactions between the neutral component of ϕ and the Z boson at tree level. This makes it far more likely to be consistent with the very stringent bounds from direct detection experiments. However, the model suffers from two problems, which will prove challenging for inert multiplet semi-annihilation in general.

First, the model suffers from a small naturalness problem. The potential includes a term that can be written as

$$\lambda_3^2 H^\dagger \tau^a H \phi^\dagger T^a \phi, \tag{2.5}$$

where τ^a and T^a are the appropriate $SU(2)$ generators. Besides small radiative corrections, this is the only source of mass splitting between the components of the multiplet. It leads to the lightest component of ϕ being either the one with the most positive or most negative value of T^3, depending on the sign of λ_3^2. Unless ϕ is a singlet, this results in the dark matter being electrically charged, which is not phenomenologically viable. Of course, the coefficient of this operator can be set to zero, but it is not technically natural as it does not correspond to a point of enhanced symmetry. In a more general case, this problem could be solved in three ways. First, this term is zero if the multiplet is real. This is often incompatible with the existence of symmetries allowing for semi-annihilation. Second, the multiplet can have a non-zero hypercharge. This can be done via $|Y| = (n-1)/2$, resulting in the neutral state being the one with the largest or smallest T^3, or via $Y = -1/2$, in which case a term of the form $\lambda H^2 \phi^2 + \text{h.c.}$ can be included. Third, the multiplet could mix with other fields. This will require the presence of another multiplet. See Refs. [40, 44] for enlightening discussions on the subject. In the end, this problem is more esthetical and is not sufficient to exclude the model. It is also not that difficult to have the lightest particle be neutral for scalars in the TeV or tens of TeV range when loop corrections are taken into account.

Second, the model suffers from a very serious practical problem that removes most of its appeal: It does not lead to efficient semi-annihilation. The reason is as follows. Two types of semi-annihilation processes are possible for this model:

1. $\phi^a \phi^b \to (\phi^c)^\dagger \ h$,
2. $\phi^a \phi^b \to (\phi^c)^\dagger \ B$,

where h is the Higgs boson and B denotes a gauge boson A, Z, W^+ or W^-. The conjugate processes are of course also possible.

For the first process, the cross section is vanishingly small. The reason is that, if electroweak symmetries were unbroken, this process would be forbidden by conservation of both isospin and weak hypercharge. The cross section must then be suppressed by some power of v/m, where v is the Higgs vacuum expectation value (VEV). Since electroweak multiplets must typically have a mass in the TeV range or higher to reproduce the correct dark matter abundance, this results in very inefficient processes of type 1 for this model.

For the second process, consider the limit in which v is zero. The corresponding cross section is

$$\sigma = \frac{1}{64\pi s p_A^2} \int_{t_1}^{t_0} |M|^2 dt, \tag{2.6}$$

with p_A being the center-of-mass three-momentum of one of the incoming particles and

$$\int_{t_1}^{t_0} |M|^2 dt \propto \frac{-3\sqrt{s(s-4m^2)} + 2(s+2m^2)\tanh^{-1}\left(\sqrt{\frac{s-4m^2}{s}}\right)}{s-m^2} \approx \frac{\Delta s^{5/2}}{90 m^5}, \tag{2.7}$$

where the approximation is at leading order in $\Delta s = s - 4m^2$. A short computation shows that this leads to $\sigma\beta \propto \beta^4$, where β is the Møller velocity, and therefore a thermally-averaged cross section $\langle\sigma\beta\rangle \propto T^2$, where T is the temperature. When the VEV of the Higgs is taken into account, some of these processes are not quite d-wave, but the leading terms in the temperature expansion of the thermally averaged cross section are suppressed by powers of v/m, resulting in semi-annihilation processes of type 2 being mostly irrelevant. If one considers two multiplets ϕ_1 and ϕ_2, one could include in the potential a term of the form $\lambda\phi_1^2\phi_2 + \text{h.c.}$ In the limit of zero Higgs VEV, the resulting semi-annihilation processes of type 2 would be p-wave if the incoming particles are distinct or d-wave otherwise, resulting in the same problem as for a single multiplet.

As a side note, this p-wave or d-wave suppression is unsurprising. Assume the collision $\phi_i\phi_j \to \phi_k^\dagger A$, with A being a massless gauge boson and i, j and k labelling the species. The gauge boson A has two possible helicities: ± 1. To conserve angular momentum, the amplitude must be at least p-wave. For $i \neq j$, the amplitude can be p-wave and this is indeed the result we observe. For $i = j$, the amplitude must be invariant under the interchange of the two incoming scalars and the amplitude therefore cannot be p-wave. This is why the amplitude is d-wave for identical scalars. The case of p-wave suppression was also noted for gluons and photons in Ref. [7], though the argument was different.

The only way to obtain efficient semi-annihilation for a renormalizable Lagrangian with only one multiplet would then be to include a term of the form $\lambda\phi^3 H + \text{h.c.}$ The resulting semi-annihilation of type 1 would be s-wave and would not be suppressed by any power of v/m. Unfortunately, this would only lead to stable charged particles, which is experimentally problematic and does not provide any dark matter candidate. This rules out the possibility of efficient semi-annihilation for only one inert multiplet.

From the discussion of this section, we are left with few options for efficient semi-annihilation of inert multiplets. First, one could add new fields beyond inert multiplets. This opens a few possibilities, but is beyond the scope of this paper. Second, the potential could include a term in $\lambda_1\phi_1^3 + \text{h.c.}$ and a term in $\lambda_2\phi_1\phi_2^\dagger H^\dagger + \text{h.c.}$ This would lead to semi-annihilation processes involving the Higgs boson. However, λ_2 is constrained by direct detection experiments to be small, which would limit the usefulness of the semi-annihilation processes. Finally, the only option left is to include in the potential a term of the form $\lambda\phi_i\phi_j\phi_k H + \text{h.c.}$ with i, j and k not all equal.

3 Two multiplets case

Armed with our experience from the one multiplet case, we now consider models of two inert multiplets semi-annihilation. We will try to determine which models lead to efficient semi-annihilation, are more likely to avoid the bounds from direct detection experiments, and can naturally avoid stable charged particles.

First, as explained in the last section, a model of two inert multiplets with efficient semi-annihilation must include in its potential a term of the form

$$V \supset \lambda\phi_1^2\phi_2 H + \text{h.c.} \tag{3.1}$$

The multiplet ϕ_2 has to be an even representation of $SU(2)$ and therefore a complex multiplet. For both ϕ_1 and ϕ_2 to be non-trivially charged under a stabilizing symmetry and thus have semi-annihilation, ϕ_1 must also be complex. We will refer to the weak hypercharge of ϕ_1 and ϕ_2 as Y_1 and Y_2 and their size as n_1 and n_2, respectively. By trying every combination of fields, one can verify that there is only a finite number of renormalizable type B terms compatible with Eq. (3.1). Of these, only some have both ϕ_1 and ϕ_2 non-trivially charged under a stabilizing symmetry and do not necessarily contain a stable charged particle. Besides $\lambda H^\dagger H \phi_i^\dagger \phi_i$ and $\lambda H^\dagger \tau^a H \phi_i^\dagger T^a \phi_i$ which are always allowed, the possibilities are:

- $V_{B_1} = \lambda \phi_2^2 (H^\dagger)^2 + \text{h.c.}$, with $Y_1 = -1/2$, $Y_2 = 1/2$ and $\mathbb{Z}_4$.
- $V_{B_2} = \lambda \phi_1 \phi_2^\dagger H^2 + \text{h.c.}$, with $Y_1 = -1/2$, $Y_2 = 1/2$ and $\mathbb{Z}_3$.
- $V_{B_3} = \lambda \phi_2^2 H^2 + \text{h.c.}$, with $Y_1 = 0$, $Y_2 = -1/2$ and $\mathbb{Z}_4$.
- $V_{B_4} = \lambda \phi_1 \phi_2^\dagger H^\dagger + \text{h.c.}$, with $Y_1 = 0$, $Y_2 = -1/2$ and $\mathbb{Z}_3$.
- No B term, with $U(1)$.

The symmetries above are respected by the A- and B-type terms and correspond to:

- $\mathbb{Z}_3$: $\phi_1 \to e^{2\pi i/3}\phi_1$ and $\phi_2 \to e^{2\pi i/3}\phi_2$.
- $\mathbb{Z}_4$: $\phi_1 \to i\phi_1$ and $\phi_2 \to -\phi_2$.
- $U(1)$: $\phi_1 \to e^{\pi\alpha i}\phi_1$ and $\phi_2 \to e^{-2\pi\alpha i}\phi_2$, with $\alpha \in \mathbb{R}$.

The potential can satisfy one and only one of these scenarios. These symmetries could be broken by type C terms, but doing so would not help satisfy any of our goals. A mass term of the form $m^2\phi_1^2 + \text{h.c.}$ would be incompatible with the existence of semi-annihilation processes and a term of the form $m^2\phi_2^2 + \text{h.c.}$ would actually be zero.

Second, without requiring precise accidental cancellations, the only two ways to suppress the direct detection signal of a stable scalar are:

- Split the masses of the CP-even and CP-odd parts of the neutral component,
- Have dark matter be mostly a component from an odd-dimensional multiplet of zero weak hypercharge.

We will require one of these conditions to be satisfied for all dark matter components. For V_{B_1}, the lightest component of ϕ_1 will be stable and cannot meet these requirements. For V_{B_2}, the lightest mass eigenstate will be stable and cannot meet these requirements.

Third, we would like the model to avoid the presence of stable charged particles in a technically natural way. For V_{B_3}, the lightest component of ϕ_1 will be stable and there is no technically natural way for it to be the neutral component at tree level unless it is a singlet. For $U(1)$, the lightest component of ϕ_1 will be stable and can only meet the direct detection requirements if $Y_1 = 0$. In this case, there is no natural way for the neutral component of ϕ_1 to be the lightest at tree level unless it is a singlet.

Ultimately, the case of V_{B_4} is the only scenario that meets all of the requirements without requiring ϕ_1 to be a singlet. It will be the focus of the rest of this work and will be presented in more detail in the next section. The cases of V_{B_1}, V_{B_3} and $U(1)$ are also not strictly excluded. They must, nevertheless, rely on loop corrections to avoid stable charged particles, satisfy special conditions to meet the direct detection constraints, or both. The scenario of V_{B_2} would, however, be very difficult to salvage.

4 Model definition

In this section, we present the model which will be the focus of the rest of this work.

Consider two complex scalar multiplets ϕ_1 and ϕ_2 of dimensions n_1 and n_2 and with weak hypercharge Y_1 and Y_2, respectively. We will assume throughout this work that $|n_1 - n_2| = 1$. The relevant potential is:

$$V = V_0 + V_A + V_B + V_C, \tag{4.1}$$

where

$$\begin{aligned}
V_0 =& -\mu^2|H|^2 + \lambda_0|H|^4 + m_1^2|\phi_1|^2 + m_2^2|\phi_2|^2,\\
V_A =& \lambda_1 A_{abcd}\phi_1^a\phi_1^b\phi_2^c H^d + \lambda_2 B_{abc}\phi_1^a\phi_1^b\phi_1^c + \text{h.c.},\\
V_B =& \sum_{r=1}^{\alpha}\lambda_3^r C^r_{abcd}(H^a)^\dagger H^b(\phi_1^c)^\dagger\phi_1^d + \sum_{r=1}^{\beta}\lambda_4^r D^r_{abcd}(H^a)^\dagger H^b(\phi_2^c)^\dagger\phi_2^d\\
&+ [\lambda_5 E_{abc}\phi_1^a(\phi_2^b)^\dagger(H^c)^\dagger + \text{h.c.}],\\
V_C =& \sum_{r=1}^{\gamma}\lambda_6^r F^r_{abcd}(\phi_1^a)^\dagger(\phi_1^b)^\dagger\phi_1^c\phi_1^d + \sum_{r=1}^{\delta}\lambda_7^r G^r_{abcd}(\phi_2^a)^\dagger(\phi_2^b)^\dagger\phi_2^c\phi_2^d\\
&+ \sum_{r=1}^{\epsilon}\lambda_8^r H^r_{abcd}(\phi_1^a)^\dagger(\phi_2^b)^\dagger\phi_1^c\phi_2^d,
\end{aligned} \tag{4.2}$$

with $\lambda_{2,5}$ having the dimension of mass and all the other λ's being dimensionless. The $SU(2)$ tensors are given in Appendix A.2. The conditions for the $SU(2)$ contractions to be non-zero are

$$\begin{aligned}
A:&\quad n_2 \in \{2n_1,\ 2n_1-2\ , 2n_1-4,\ ...\},\\
B:&\quad n_1 \in \{1, 5, 9,\ ...\},\\
E:&\quad n_1 - n_2 \in \{-1, 1\}.
\end{aligned} \tag{4.3}$$

All other contractions are always non-vanishing. The upper limits on the sums are given by

$$\alpha = \min(2, n_1)\,, \quad \beta = \min(2, n_2)\,, \quad \gamma = \left\lfloor \frac{n_1+1}{2} \right\rfloor, \quad \delta = \left\lfloor \frac{n_2+1}{2} \right\rfloor, \quad \epsilon = \min(n_1, n_2). \tag{4.4}$$

The potential respects a $\mathbb{Z}_3$ symmetry under which

$$\phi_1 \to e^{2\pi i/3}\phi_1, \quad \phi_2 \to e^{2\pi i/3}\phi_2. \tag{4.5}$$

Unless $\mathbb{Z}_3$ is spontaneously broken, the lightest mass eigenstate will be stable.

5 Dark matter abundance computation

In this section, we present our computation of the dark matter abundance. The computation is performed first under the assumption of unbroken electroweak symmetries and then under the assumption of broken electroweak symmetries. We will show that the two methods are consistent within the expected precision. Our implementation of Sommerfeld enhancement (SE) is also presented.

5.1 Computation for unbroken electroweak symmetries

To compute the dark matter abundance under the assumption of unbroken electroweak symmetries, we will consider scattering from one weak isospin eigenstate to another. This will simplify much of the formalism, especially concerning the Sommerfeld enhancement. We take inspiration from and expand on the results of Refs. [40, 44, 45, 56].

Consider a set of fields ϕ_1, ϕ_2 ... forming a state of isospin numbers J and M. We will refer to this state as $[\phi_1\phi_2\dots]^M_J$. Scattering will always be from a state J and M to a state of identical J' and M'. The states can easily be constructed by a combination of Clebsch-Gordan coefficients and charge conjugation matrices.

Cross sections

We begin by presenting the cross sections necessary for computing the dark matter abundance. All $2 \to 2$ cross sections can be decomposed as

$$\begin{aligned}\sigma([P_A P_B]^M_J \to [P_C P_D]^{M'}_J) &= \sigma_0([P_A P_B]_J \to [P_C P_D]_J)\delta^{MM'} \\ &= \frac{S_{CD}}{64\pi s p_A^2}\hat{M}([P_A P_B]_J \to [P_C P_D]_J)\delta^{MM'},\end{aligned} \tag{5.1}$$

where S_{CD} is 1/2 if P_C and P_D are identical and 1 otherwise and p_A is the center-of-mass three-momentum of P_A given by

$$p_A = \frac{\sqrt{(s-(m_A+m_B)^2)(s-(m_A-m_B)^2)}}{2\sqrt{s}}, \tag{5.2}$$

where s is the usual Mandelstam variable. For annihilation to Higgs pairs, the $\hat{M}$ coefficients are given by

$$\begin{aligned}
\hat{M}([\phi_1^\dagger\phi_1]_0 \to [H^\dagger H]_0) &= f_1\left(s, m_1, m_2, \sqrt{2n_1}\lambda_3^1, \sqrt{\frac{n_1}{2}}|\lambda_5|^2\right), \\
\hat{M}([\phi_2^\dagger\phi_2]_0 \to [H^\dagger H]_0) &= f_1\left(s, m_2, m_1, \sqrt{2n_2}\lambda_4^1, \frac{n_1}{\sqrt{2n_2}}|\lambda_5|^2\right), \\
\hat{M}([\phi_1^\dagger\phi_1]_1 \to [H^\dagger H]_1) &= f_1\left(s, m_1, m_2, \sqrt{\frac{(n_1^2-1)n_1}{24}}\lambda_3^2, \frac{(n_1-n_2)}{\sqrt{6}}\frac{\sqrt{n_1(n_1^2-1)}}{n_2}|\lambda_5|^2\right), \\
\hat{M}([\phi_2^\dagger\phi_2]_1 \to [H^\dagger H]_1) &= f_1\left(s, m_2, m_1, \sqrt{\frac{(n_2^2-1)n_2}{24}}\lambda_4^2, \frac{(n_1-n_2)}{\sqrt{6}}\sqrt{\frac{n_2^2-1}{n_2}}|\lambda_5|^2\right),
\end{aligned} \tag{5.3}$$

where

$$f_1(s, m_A, m_B, C_A, C_B) =$$
$$\sqrt{s(s-4m_A^2)}\left(C_A^2 + \frac{C_B^2}{(m_A-m_B)^2+m_B^2 s}\right) - 4C_A C_B \tanh^{-1}\left(\frac{\sqrt{s(s-4m_A^2)}}{s-2m_A^2+2m_B^2}\right). \tag{5.4}$$

The mass of the Higgs boson has been neglected, as will be the case for all computations in this subsection. For annihilation to pairs of gauge bosons, the $\hat{M}$ coefficients are given by

$$\begin{aligned}
\hat{M}([\phi_1^\dagger\phi_1]_0 \to [WW]_0) &= f_2\left(s, m_1, \frac{(n_1^2-1)\sqrt{n_1}}{4\sqrt{3}}g^2\right),\\
\hat{M}([\phi_2^\dagger\phi_2]_0 \to [WW]_0) &= f_2\left(s, m_2, \frac{(n_2^2-1)\sqrt{n_2}}{4\sqrt{3}}g^2\right),\\
\hat{M}([\phi_1^\dagger\phi_1]_2 \to [WW]_2) &= f_2\left(s, m_1, \frac{1}{2}\sqrt{\frac{n_1(n_1^2-1)(n_1^2-4)}{30}}g^2\right),\\
\hat{M}([\phi_2^\dagger\phi_2]_2 \to [WW]_2) &= f_2\left(s, m_2, \frac{1}{2}\sqrt{\frac{n_2(n_2^2-1)(n_2^2-4)}{30}}g^2\right),\\
\hat{M}([\phi_2^\dagger\phi_2]_0 \to [WB]_0) &= f_2\left(s, m_2, \sqrt{\frac{n_2(n_2^2-1)}{12}}Y_2 g g'\right),\\
\hat{M}([\phi_2^\dagger\phi_2]_0 \to [BB]_0) &= f_2\left(s, m_2, (Y_2 g')^2\right),
\end{aligned} \tag{5.5}$$

where g and g' are respectively the $SU(2)_L$ and $U(1)_Y$ gauge couplings and

$$f_2(s, m, C) = \frac{8C^2}{s}\left[(s+4m^2)\sqrt{s(s-4m^2)} - 8m^2(s-2m^2)\tanh^{-1}\left(\sqrt{\frac{s-4m^2}{s}}\right)\right]. \tag{5.6}$$

For semi-annihilation involving a Higgs boson, we will assume that the B tensor is zero, λ_2 is small or λ_5 is small. The latter condition must realistically be respected to meet direct detection constraints. If at least one of these conditions is satisfied, then the $\hat{M}$ coefficients are given by

$$\begin{aligned}
\hat{M}\left([\phi_1\phi_1]_{J_0} \to [\phi_2^\dagger H^\dagger]_{J_0}\right) &= \frac{4|\lambda_1|^2}{R}\frac{\sqrt{s-4m_1^2}(s-m_2^2)}{\sqrt{s}},\\
\hat{M}\left([\phi_1\phi_2]_{J_1} \to [\phi_1^\dagger H^\dagger]_{J_1}\right) &= 4|\lambda_1|^2\left|C_{n_1 n_2}^{R_1'}\right|^2\frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}(s-m_1^2)}{s},\\
\hat{M}\left([\phi_1\phi_2]_{J_2} \to [\phi_1^\dagger H^\dagger]_{J_2}\right) &= 4|\lambda_1|^2\left|C_{n_1 n_2}^{R_2'}\right|^2\frac{\sqrt{(s-(m_1+m_2)^2)(s-(m_1-m_2)^2)}(s-m_1^2)}{s},
\end{aligned} \tag{5.7}$$

where

$$\begin{aligned}
R &= 4\left\lfloor\frac{n_2}{4}\right\rfloor + 1, \qquad & R_1' &= n_1+1, \qquad & R_2' &= n_1-1,\\
J_0 &= \frac{R-1}{2}, & J_1 &= \frac{R_1'-1}{2}, & J_2 &= \frac{R_2'-1}{2},
\end{aligned} \tag{5.8}$$

and

$$\left|C^{R'}_{n_1 n_2}\right|^2 = \frac{1}{R'}\left(\frac{1}{2} - \frac{n_2}{4n_1}(R'-n_1)(n_1-n_2)(-1)^{\frac{n_1-1}{2}}\right). \tag{5.9}$$

For this model, semi-annihilation processes involving gauge bosons are either d-wave or forbidden by conservation of isospin or weak hypercharge. We therefore neglect them.

Boltzmann equation

We now proceed to derive the Boltzmann equation for the combined abundance of ϕ_1 and ϕ_2. We first define the effective cross sections

$$\begin{aligned}
\sigma(\phi_1\phi_1 \to \text{SM SM}) = &\ \frac{1}{2n_1^2}\sigma_0\left([\phi_1^\dagger\phi_1]_0 \to [H^\dagger H]_0\right) + \frac{3}{2n_1^2}\sigma_0\left([\phi_1^\dagger\phi_1]_1 \to [H^\dagger H]_1\right) \\
&+\frac{1}{2n_1^2}\sigma_0\left([\phi_1^\dagger\phi_1]_0 \to [WW]_0\right) + \frac{5}{2n_1^2}\sigma_0\left([\phi_1^\dagger\phi_1]_2 \to [WW]_2\right), \\
\sigma(\phi_2\phi_2 \to \text{SM SM}) = &\ \frac{1}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_0 \to [H^\dagger H]_0\right) + \frac{3}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_1 \to [H^\dagger H]_1\right) \\
&+\frac{1}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_0 \to [WW]_0\right) + \frac{5}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_2 \to [WW]_2\right) \\
&+\frac{1}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_0 \to [BB]_0\right) + \frac{3}{2n_2^2}\sigma_0\left([\phi_2^\dagger\phi_2]_1 \to [WB]_1\right), \\
\sigma(\phi_1\phi_1 \to \phi_2 \text{ SM}) = &\ \frac{R}{4n_1^2}\sigma_0\left([\phi_1\phi_1]_{J_0} \to [\phi_2^\dagger H^\dagger]_{J_0}\right), \\
\sigma(\phi_1\phi_2 \to \phi_1 \text{ SM}) = &\ \frac{R'_1}{2n_1 n_2}\sigma_0\left([\phi_1\phi_2]_{J_1} \to [\phi_1^\dagger H^\dagger]_{J_1}\right) + \frac{R'_2}{2n_1 n_2}\sigma_0\left([\phi_1\phi_2]_{J_2} \to [\phi_1^\dagger H^\dagger]_{J_2}\right).
\end{aligned} \tag{5.10}$$

We define the number density per entropy density s_E of ϕ_i as Y_i and $Y = Y_1 + Y_2$. Their equilibrium distributions at temperature T are given by

$$Y_i^{\text{eq}}(T) = \frac{g_i m_i^2 T}{2\pi^2 s_E} K_2(m_i/T), \tag{5.11}$$

where $g_i = 2n_i$ is the number of internal degrees of freedom of ϕ_i and $K_n(x)$ are the modified Bessel functions of the second kind.

As long as λ_5 is not vanishingly small, ϕ_1 and ϕ_2 will remain in relative chemical equilibrium during freeze-out via interactions with gauge or Higgs bosons. We can then further define effective thermally averaged cross sections as

$$\begin{aligned}
\langle\sigma\beta(\phi\phi \to \text{SM SM})\rangle &= \left(\frac{Y_1^{\text{eq}}}{Y^{\text{eq}}}\right)^2\langle\sigma\beta(\phi_1\phi_1 \to \text{SM SM})\rangle + \left(\frac{Y_2^{\text{eq}}}{Y^{\text{eq}}}\right)^2\langle\sigma\beta(\phi_2\phi_2 \to \text{SM SM})\rangle, \\
\langle\sigma\beta(\phi\phi \to \phi \text{ SM})\rangle &= \left(\frac{Y_1^{\text{eq}}}{Y^{\text{eq}}}\right)^2\langle\sigma\beta(\phi_1\phi_1 \to \phi_2 \text{ SM})\rangle + \frac{Y_1^{\text{eq}}Y_2^{\text{eq}}}{(Y^{\text{eq}})^2}\langle\sigma\beta(\phi_1\phi_2 \to \phi_1 \text{ SM})\rangle.
\end{aligned} \tag{5.12}$$

The current dark matter number density per entropy density can then be computed by solving a single Boltzmann equation given by

$$\frac{dY}{dx} = -s_E\frac{dt}{dx}\left[\langle\sigma\beta(\phi\phi \to \text{SM SM})\rangle(Y^2 - (Y^{\text{eq}})^2) + \langle\sigma\beta(\phi\phi \to \phi \text{ SM})\rangle Y(Y - Y^{\text{eq}})\right], \tag{5.13}$$

where we define $x = m_1/T$ and

$$\frac{dt}{dx} = \sqrt{\frac{45}{4\pi^3}} \frac{g_*^{1/2}}{h_{\rm eff}} \frac{M_{\rm Pl}}{m_1 T}, \quad g_*^{1/2} = \frac{h_{\rm eff}}{g_{\rm eff}^{1/2}} \left(1 + \frac{T}{3h_{\rm eff}} \frac{dh_{\rm eff}}{dT}\right), \tag{5.14}$$

with $g_{\rm eff}$ and $h_{\rm eff}$ respectively denoting the effective number of energy and entropy degrees of freedom and $M_{\rm Pl}$ being the Planck mass. Once the present-day value of Y has been obtained, the mass of the lightest eigenstate is computed, including one loop radiative corrections [21]. These two quantities are then easily combined to obtain the dark matter relic abundance.

Sommerfeld enhancement

Thermally averaged cross sections can be modified by the Sommerfeld enhancement because of the exchange of electroweak gauge bosons [57]. Being massless in the limit of unbroken electroweak symmetries, they mediate a Coulomb potential.

Consider two colliding particles P_A and P_B of weak isospins J_A and J_B and weak hypercharge Y_A and Y_B, respectively. Assume they form a state of weak isospin J. Define [58]

$$\alpha = \frac{g^2}{8\pi} \left[J(J+1) - J_A(J_A+1) - J_B(J_B+1)\right] + \frac{(g')^2}{4\pi} Y_A Y_B \eta, \tag{5.15}$$

where $\eta = -1$ for the collision of a particle and an antiparticle and $\eta = 1$ otherwise. The cross section is then enhanced by a multiplicative factor of [59]

$$S_{\rm SE} = -\frac{\pi\alpha}{\beta} \frac{1}{1 - e^{\frac{\pi\alpha}{\beta}}}. \tag{5.16}$$

This enhancement is taken into account in the computation of the thermally averaged cross sections. If λ_5 were to be sufficiently large, a significant Sommerfeld enhancement via Higgs exchange could be possible. However, direct detection constraints on λ_5 limit this effect to be small and we neglect it.

5.2 Computation for broken electroweak symmetries

We also perform the computation of the dark matter abundance when the VEV of the Higgs is set to its present-day value. Due to the presence of so many different masses, the amplitudes are cumbersome and computations are handled numerically.

Our procedure is as follows. Mass matrices are extracted from the Lagrangian and diagonalized. Interaction coefficients for mass eigenstates are then computed. The amplitudes for all annihilation and semi-annihilation processes involving gauge bosons or the Higgs boson are evaluated. The amplitudes are then used to compute the thermally averaged cross sections. The Boltzmann equation is solved numerically assuming all scalars maintain relative chemical equilibrium.

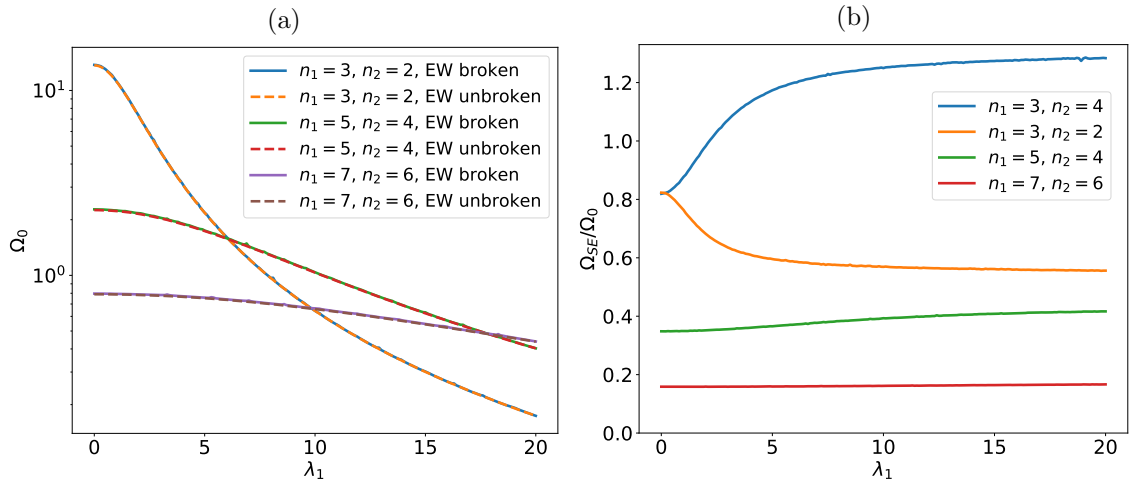


Figure 1: (a) Relic abundance Ω_0 for different choices of n_1 and n_2 and for the methods of broken and unbroken electroweak symmetries. The Sommerfeld enhancement is not included in this case. The cases of n_1 and $n_2 = n_1 \pm 1$ are nearly identical and only $n_2 = n_1 - 1$ is shown. (b) Ratio of the relic abundance with the Sommerfeld enhancement Ω_{SE} and without Ω_0. The case of $n_1 = 5$ and $n_2 = 6$ is nearly identical to $n_1 = 5$ and $n_2 = 4$ and is not included. For both plots, $m_1 = 10$ TeV, $m_2 = 11$ TeV and all other parameters are set to 0.

5.3 Comparison of the methods

We show in Fig. 1a examples of the dark matter abundance computed using both methods. For these examples, the two methods give results consistent within about 1%. Such deviations are of the expected size of $\mathcal{O}(v^2/m_i^2)$, though part of these deviations are unavoidably due to numerical errors. Obtaining higher precision would require tracking the VEV of the Higgs during the freeze-out process, which is beyond the scope of this paper, and higher numerical precision. As can be seen in Fig. 1b, the Sommerfeld enhancement can substantially modify the dark matter abundance, especially for larger multiplets. Note also that semi-annihilation can in fact be made less efficient by the Sommerfeld enhancement, as is noticeably the case for $n_1 = 3$ and $n_2 = 4$. Since both methods give identical results to good precision, we will perform all computations in the limit of unbroken electroweak symmetries for the rest of this paper, including the Sommerfeld enhancement.

6 Constraints

In this section, we discuss our implementation of direct detection and unitarity constraints. Other constraints that were considered but ultimately not implemented are also discussed. Though a full study is beyond the scope of this work, we will discuss constraints from indirect detection in the conclusion.

6.1 Direct detection experiments

The lightest mass eigenstate $\hat{\phi}_1$ can in principle interact at tree level with both the Z and Higgs bosons. We parametrize these interactions as

$$\mathcal{L} \supset \left[-i\hat{C}_{11} Z_\mu \hat{\phi}_1^\dagger \partial^\mu \hat{\phi}_1 + \text{h.c.}\right] - h\hat{\Omega}_{11}\hat{\phi}_1^\dagger \hat{\phi}_1. \tag{6.1}$$

Using the notation of Ref. [60], the spin-independent per-nucleon cross section is

$$\sigma_{\text{SI},N} = \frac{m_N^2 m_{\hat{\phi}_1}^2}{4\pi(m_N + m_{\hat{\phi}_1})^2}\left[\sum_q \frac{2g g_V^q}{c_W m_Z^2}\hat{C}_{11} F_1^{(N,q)}(0) - \frac{m_N}{m_h^2 m_{\hat{\phi}_1}}\frac{\hat{\Omega}_{11}}{v} f_{q,N}^{(0)}\right]^2, \tag{6.2}$$

where N is either a proton or a neutron, c_W (s_W) is the cosine (sine) of the Weinberg angle, $v \approx 246$ GeV is the Higgs VEV, and $g_V^q = T_q^3/2 - Q_q s_W^2$. For the Z exchange, we use the values $F_1^{(p,u)}(0) = F_1^{(n,d)}(0) = 2$, $F_1^{(p,d)}(0) = F_1^{(n,u)}(0) = 1$ and set all other $F_1^{(N,q)}(0)$ to 0. For Higgs exchange, we use $\sum_q f_{q,p}^{(0)} = 0.301$ and $\sum_q f_{q,n}^{(0)} = 0.307$ [61]. The values of $\sigma_{\text{SI},N}$ are very close for the proton and neutron and we take their average. Loop corrections to $\sigma_{\text{SI},N}$ were considered. However, a careful treatment reveals partial cancellations of the amplitudes such that they are safely below current detection limits for all multiplet sizes relevant to this work [62]. Constraints on $\sigma_{\text{SI},N}$ are applied using the results of the LZ experiment from Ref. [63].

6.2 Unitarity

The parameter λ_1 is constrained by unitarity. Consider the process $\phi_1\phi_1 \to \phi_2^\dagger H^\dagger$. In the limit of high energies, the amplitude is given in the basis of weak isospin eigenstates as

$$M\left([\phi_1\phi_1]_J^M \to [\phi_2^\dagger H^\dagger]_{J'}^{M'}\right) = \frac{2\lambda_1}{\sqrt{R}}\delta^{MM'}\delta_{JR}\delta_{J'R}, \tag{6.3}$$

where R was defined in Eq. (5.8). It can be decomposed in terms of the Legendre polynomials $P_\ell(\cos\theta)$ as

$$M\left([\phi_1\phi_1]_J^M \to [\phi_2^\dagger H^\dagger]_{J'}^{M'}\right) = 16\pi\sum_\ell (2\ell+1)(a_\ell)_{JJ'}^{MM'} P_\ell(\cos\theta). \tag{6.4}$$

This gives

$$(a_0)_{JJ'}^{MM'} = \frac{\lambda_1}{8\pi\sqrt{R}}\delta^{MM'}\delta_{JR}\delta_{J'R}. \tag{6.5}$$

In the basis of $[\phi_1\phi_1]_J^M \to [\phi_2^\dagger H^\dagger]_{J'}^{M'}$, $(a_0)_{JJ'}^{MM'}$ can be written as a diagonal matrix. Taking into account the presence of two identical particles in the incoming state, unitarity requires the largest eigenvalue of this matrix in absolute value to be smaller than $1/\sqrt{2}$. This imposes the simple condition

$$|\lambda_1| \leq 4\pi\sqrt{2R}. \tag{6.6}$$

As a side note, the fraction of $[\phi_1\phi_1]_J^M$ states that contribute to this scattering scales roughly as $1/R$. A higher upper limit on $|\lambda_1|^2$ proportional to R will partially compensate for this effect.

6.3 Other constraints

Other constraints were considered but were found to be irrelevant and as such were not used in the final results. We computed the oblique parameters [64] using the results of Refs. [65–67]. All relevant constraints were found to be subdominant to either direct detection constraints or the requirement of having the stable component be neutral. Unitarity constraints on parameters besides λ_1 were studied using the results of Ref. [67], but were found to be satisfied in all regions of parameter space considered in this work. The Higgs signal strengths were also studied using the method of Ref. [67], but the scalars proved to be simply too heavy to have any significant impact on the Higgs properties.

7 Results

We present in this section some scans of the parameter space.

Figs. 2, 3 and 4 show the allowed region in the space of m_1 and m_2/m_1 for different combinations of n_1 and n_2 when λ_3^1 is set respectively to 0, 1 and 5. Combinations of larger multiplets would be in conflict with the unitarity bounds of Ref. [56]. We also ignore the case of $n_1 = 1$ and $n_2 = 2$, as the presence of a singlet is not really aligned with the goals of this paper, is somewhat qualitatively different, and has previously been studied extensively (see, e.g., Ref. [6]). In each case, the parameter λ_1 is adjusted to reproduce the correct relic abundance. Contours of this parameter are also shown. All other parameters are set to zero, as their impact on the dark matter abundance would be small or as they are restricted to be small by other constraints. It is however assumed that λ_5 has a small non-zero value for ϕ_1 and ϕ_2 to maintain chemical equilibrium, but not sufficiently large to noticeably alter the relic abundance. The ratio m_2/m_1 is always taken as larger than 1, as this would otherwise be in conflict with direct detection experiments. The lower limit on m_1 comes from the impossibility of adjusting λ_1 to reproduce the relic abundance or possibly direct detection constraints when λ_3^1 is non-zero. The upper limit comes from the unitarity constraint of Eq. (6.6).

As can be seen, the presence of semi-annihilation processes allows m_1 to span a range of sometimes more than an order of magnitude. In scenarios with a $\mathbb{Z}_2$ symmetries, direct detection constraints often force the multiplet masses to be in a very narrow range, as the only freedom to change the abundance comes from heavily constrained interactions with the Higgs boson. The allowed range does tend to relatively shrink as the size of the multiplets increases. As was to be expected, increasing m_2/m_1 typically reduces the impact of semi-annihilation processes. The allowed range is especially narrow for values of m_2/m_1 larger than 2, as the process $\phi_1\phi_1 \to \phi_2^\dagger H^\dagger$ becomes kinematically forbidden. However, the plots show that semi-annihilation processes can still be very efficient even when m_1 and m_2 differ considerably. As such, semi-annihilation does not require careful fine-tuning between the masses to be efficient, as is sometimes the case with coannihilation [68]. Increasing the value of λ_3^1 results in some of the lower mass region being excluded by direct detection constraints for smaller multiplets, but shifts the upper limit on m_1 toward larger values. The behaviour around $m_2/m_1 = 1$ is non-trivial and comes from a combination of efficient coannhihilation and Sommerfeld enhancement with different signs of R, R_1' and R_2'.

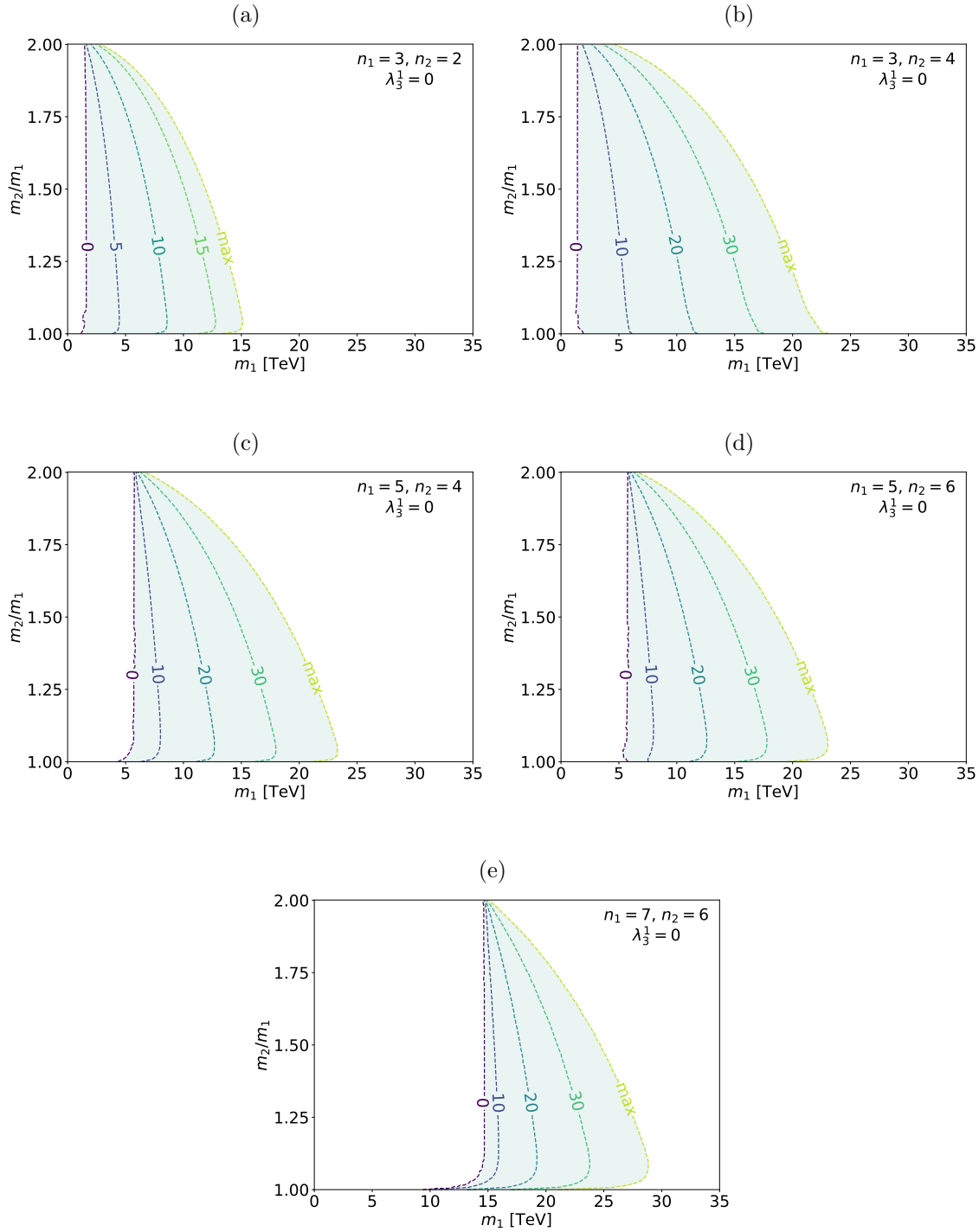


Figure 2: Parameter space as a function of m_1 and m_2/m_1 for different n_1 and n_2 when the parameter λ_3^1 is set to 0. All other parameters are set to 0. Contours of λ_1 are shown. The max value corresponds to the unitarity bound of Eq. (6.6). The green region is allowed.

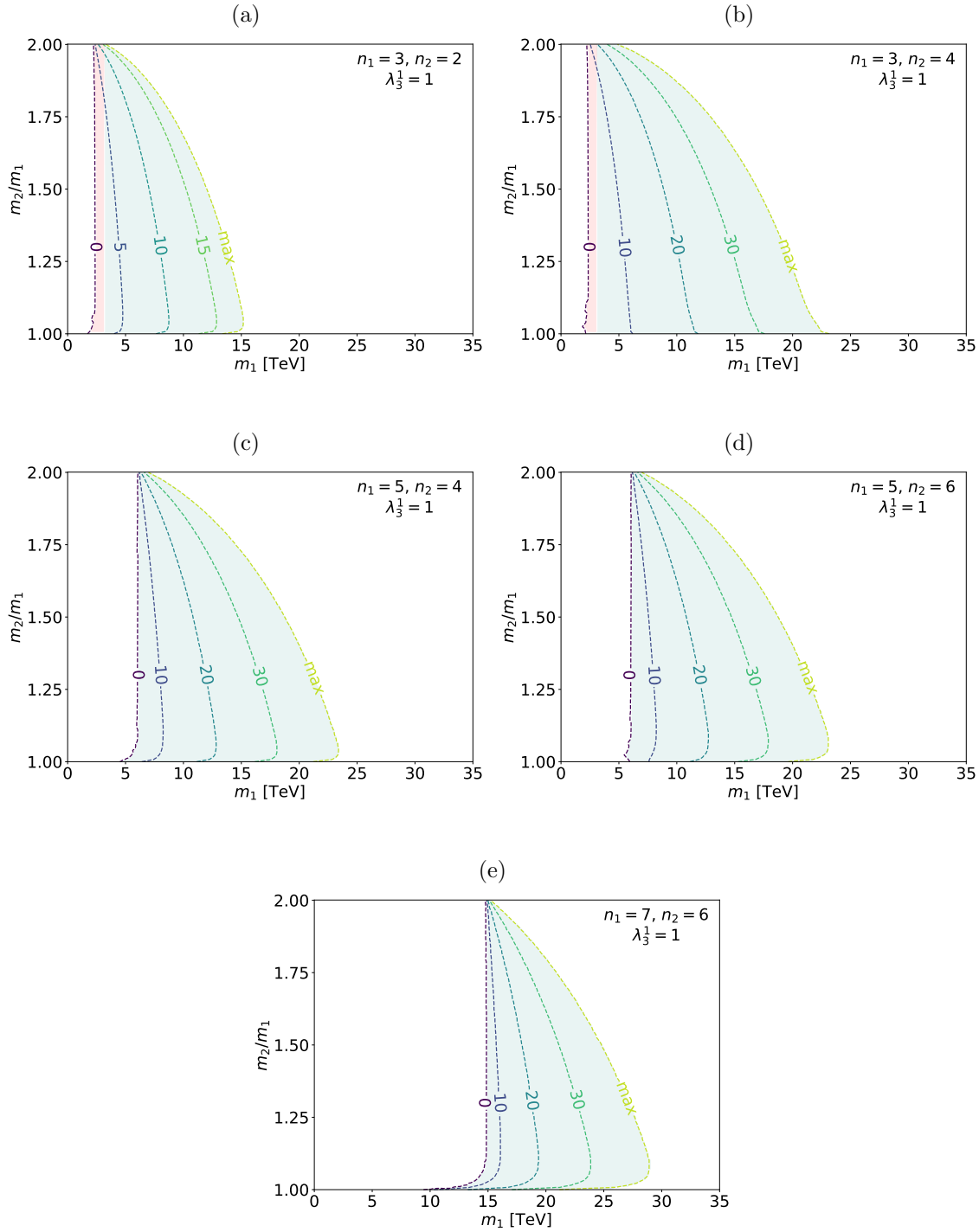


Figure 3: Same as Fig. 2 but for $\lambda_3^1 = 1$. The green region is allowed and the red region is forbidden by direct detection experiments.

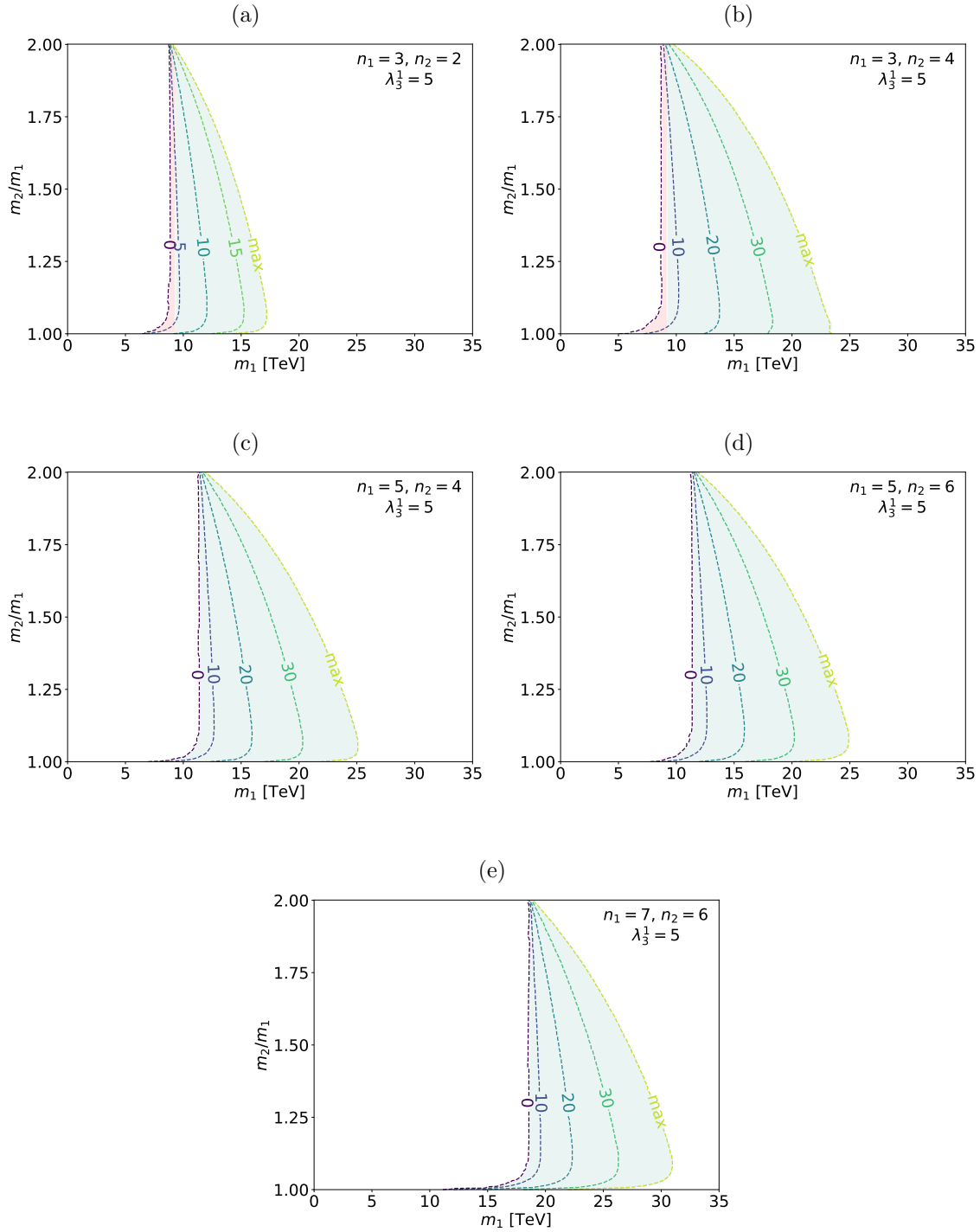


Figure 4: Same as Fig. 2 but for $\lambda_3^1 = 5$. The green region is allowed and the red region is forbidden by direct detection experiments.

8 Conclusion

The goal of this work is to determine which models of inert multiplets can lead to efficient semi-annihilation while at the same time being compatible with known constraints and naturally avoiding stable charged particles.

We found that semi-annihilation cannot be efficient in the case of a single multiplet. For two multiplets, there exists only one category of models that can lead to efficient semi-annihilation while avoiding strong constraints from direct detection experiments and stable charged particles at tree level. Other models exist, but must rely on loop corrections to avoid stable charged particles, meet certain conditions to avoid direct detection constraints, or both. For smaller multiplets, semi-annihilation allows the masses of the scalars to reproduce the correct relic abundance over a range of sometimes more than an order of magnitude. This relative range shrinks as the sizes of the multiplets increase.

As should be obvious, this paper only constitutes a first study of semi-annihilation of inert multiplets and many questions remain unanswered. We focused on what we considered to be the most attractive model, but some of the other options presented are far from forbidden. Furthermore, the model of this paper can indeed lead to the lightest particle being neutral at tree level, but meeting this condition is somewhat difficult numerically and it is not that hard for other models to avoid stable charged particles when loop corrections are taken into account. The other models would certainly warrant further studies. In addition, collider constraints were not considered in this work, as they are currently not very stringent and can easily be avoided. Potential signatures include charged particles stable on collider scales, disappearing tracks and missing energy. Prospects at future colliders already exist for inert multiplets [28, 34, 35, 38, 46, 69], but it would be interesting to determine how they are affected by the presence of semi-annihilation processes.

More importantly, we did not perform a dedicated analysis of indirect detection constraints. As is well known, indirect detection experiments now impose considerable constraints on inert multiplets, often to the point that standard $\mathbb{Z}_2$ scenarios struggle to accommodate the measured abundance [70, 71]. Performing a dedicated study of indirect detection constraints would entail two rather difficult tasks. First, one should compute a limit on the indirect detection cross section for semi-annihilation. Even though some work already exists on the subject [7, 10], we are unaware of any existing work general enough to cover all signals from inert multiplet semi-annihilation. Second, one would need to carefully compute the Sommerfeld-enhanced cross section for different astrophysical sources.

However, it is not necessary to perform such tasks to already tell that semi-annihilation can alleviate the tension with indirect detection experiments. The presence of semi-annihilation processes allows the dark matter to be considerably heavier. This by itself suppresses the constraints from annihilation to Standard Model particles. Semi-annihilation processes themselves can lead to an indirect detection signal, but the kinematics should lead to reduced limits. Most importantly, there exist combinations of gauge numbers for which semi-annihilation can actually be suppressed by the Sommerfeld enhancement. This is namely the case for $n_1 = 3$ and $n_2 = 4$. Such cases could therefore have negligible signals from indirect detection experiments. Many possibilities also exist when one is willing to

go beyond the simple model studied in this paper. As such, it could very well be that semi-annihilation of inert multiplets will become an increasingly attractive possibility as constraints from indirect detection experiments keep improving.

Acknowledgments

This work was supported by the National Science and Technology Council under Grant No. NSTC-111-2112-M-002-018-MY3, the Ministry of Education (Higher Education Sprout Project NTU-112L104022), and the National Center for Theoretical Sciences of Taiwan.

A $SU(2)$ tensors

We present in this section the different $SU(2)$ tensors used in the paper. Although not a very enlightening discussion, not explicitly stating them would lead to ambiguities on the meaning of certain coefficients as different bases exist in the literature.

A.1 One multiplet case

For the model with a single multiplet, the A_{abc} tensor is written as

$$A_{abc} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{00}_{JM j_3 m_3}, \tag{A.1}$$

where $C^{JM}_{j_1 m_1 j_2 m_2} = \langle j_1 j_2 m_1 m_2 | JM \rangle$ are the Clebsch-Gordan coefficients and

$$\begin{aligned} j_1 &= \frac{n-1}{2}, & j_2 &= \frac{n-1}{2}, & j_3 &= \frac{n-1}{2}, & J &= \frac{n-1}{2}, \\ m_1 &= \frac{n+1-2a}{2}, & m_2 &= \frac{n+1-2b}{2}, & m_3 &= \frac{n+1-2c}{2}. \end{aligned} \tag{A.2}$$

For the sake of convenience, the B^r_{abcd} tensors are written as

$$B^1_{abcd} = \delta_{ab}\delta_{cd}, \quad B^2_{abcd} = \tau^e_{ab} T^e_{cd}, \tag{A.3}$$

where τ^M_{ab} and T^M_{cd} are respectively the 2- and n-dimensional $SU(2)$ generators. The C^r_{abcd} tensors are given by

$$C^r_{abcd} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{JM}_{j_3 m_3 j_4 m_4}, \tag{A.4}$$

where

$$\begin{aligned} J &= n+1-2r, \\ j_1 &= \frac{n-1}{2}, & j_2 &= \frac{n-1}{2}, & j_3 &= \frac{n-1}{2}, & j_4 &= \frac{n-1}{2}, \\ m_1 &= \frac{n+1-2a}{2}, & m_2 &= \frac{n+1-2b}{2}, & m_3 &= \frac{n+1-2c}{2}, & m_4 &= \frac{n+1-2d}{2}, \end{aligned} \tag{A.5}$$

where $r \in \{1, 2, \ldots, \lfloor (n+1)/2 \rfloor\}$.

A.2 Two multiplets case

For the model with two multiplets, the A_{abcd} tensors are given by

$$A_{abcd} = \sum_{M,M'=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{JM'}_{j_3 m_3 j_4 m_4} C^{00}_{JMJM'}, \tag{A.6}$$

where

$$\begin{aligned} R &= 4\left\lfloor \frac{n_2}{4} \right\rfloor + 1, \qquad J = \frac{R-1}{2}, \\ j_1 &= \frac{n_1-1}{2}, \qquad j_2 = \frac{n_1-1}{2}, \qquad j_3 = \frac{n_2-1}{2}, \qquad j_4 = \frac{1}{2}, \\ m_1 &= \frac{n_1+1-2a}{2}, \quad m_2 = \frac{n_1+1-2b}{2}, \quad m_3 = \frac{n_2+1-2c}{2}, \quad m_4 = \frac{3-2d}{2}. \end{aligned} \tag{A.7}$$

The B_{abc} tensor is given by

$$B_{abc} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{00}_{JMj_3 m_3}, \tag{A.8}$$

where

$$\begin{aligned} j_1 &= \frac{n_1-1}{2}, \qquad j_2 = \frac{n_1-1}{2}, \qquad j_3 = \frac{n_1-1}{2}, \qquad J = \frac{n_1-1}{2}, \\ m_1 &= \frac{n_1+1-2a}{2}, \quad m_2 = \frac{n_1+1-2b}{2}, \quad m_3 = \frac{n_1+1-2c}{2}. \end{aligned} \tag{A.9}$$

The C^r_{abcd} and D^r_{abcd} tensors are written as

$$\begin{aligned} C^1_{abcd} &= \delta_{ab}\delta_{cd}, \quad C^2_{abcd} = \tau^e_{ab} T^e_{cd}, \\ D^1_{abcd} &= \delta_{ab}\delta_{cd}, \quad D^2_{abcd} = \tau^e_{ab} \hat{T}^e_{cd}, \end{aligned} \tag{A.10}$$

where T^M_{cd} and $\hat{T}^M_{cd}$ are respectively the $SU(2)$ generators of dimension n_1 and n_2. The E_{abc} tensor is given by

$$E_{abc} = C^{JM}_{j_1 m_1 j_2 m_2}, \tag{A.11}$$

where

$$\begin{aligned} j_1 &= \frac{n_2-1}{2}, \qquad j_2 = \frac{1}{2}, \qquad J = \frac{n_1-1}{2}, \\ m_1 &= \frac{n_2+1-2b}{2}, \quad m_2 = \frac{3-2c}{2}, \quad M = \frac{n_1+1-2a}{2}. \end{aligned} \tag{A.12}$$

The F^r_{abcd} tensor is written as

$$F^r_{abcd} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{JM}_{j_3 m_3 j_4 m_4}, \tag{A.13}$$

where

$$\begin{aligned} J &= n_1 + 1 - 2r, \\ j_1 &= \frac{n_1-1}{2}, \qquad j_2 = \frac{n_1-1}{2}, \qquad j_3 = \frac{n_1-1}{2}, \qquad j_4 = \frac{n_1-1}{2}, \\ m_1 &= \frac{n_1+1-2a}{2}, \quad m_2 = \frac{n_1+1-2b}{2}, \quad m_3 = \frac{n_1+1-2c}{2}, \quad m_4 = \frac{n_1+1-2d}{2}, \end{aligned} \tag{A.14}$$

where $r \in \{1, 2, \ldots, \lfloor (n_1+1)/2 \rfloor\}$. The G^r_{abcd} tensor is written as

$$G^r_{abcd} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{JM}_{j_3 m_3 j_4 m_4}, \tag{A.15}$$

where

$$\begin{aligned}
&J = n_2 + 1 - 2r, \\
&j_1 = \frac{n_2-1}{2}, \qquad j_2 = \frac{n_2-1}{2}, \qquad j_3 = \frac{n_2-1}{2}, \qquad j_4 = \frac{n_2-1}{2}, \\
&m_1 = \frac{n_2+1-2a}{2}, \quad m_2 = \frac{n_2+1-2b}{2}, \quad m_3 = \frac{n_2+1-2c}{2}, \quad m_4 = \frac{n_2+1-2d}{2},
\end{aligned} \tag{A.16}$$

where $r \in \{1, 2, \ldots, \lfloor (n_2+1)/2 \rfloor\}$. Finally, the H^r_{abcd} tensor is written as

$$H^r_{abcd} = \sum_{M=-J}^{J} C^{JM}_{j_1 m_1 j_2 m_2} C^{JM}_{j_3 m_3 j_4 m_4}, \tag{A.17}$$

where

$$\begin{aligned}
&J = \frac{n_1+n_2-2r}{2}, \\
&j_1 = \frac{n_1-1}{2}, \qquad j_2 = \frac{n_2-1}{2}, \qquad j_3 = \frac{n_1-1}{2}, \qquad j_4 = \frac{n_2-1}{2}, \\
&m_1 = \frac{n_1+1-2a}{2}, \quad m_2 = \frac{n_2+1-2b}{2}, \quad m_3 = \frac{n_1+1-2c}{2}, \quad m_4 = \frac{n_2+1-2d}{2},
\end{aligned} \tag{A.18}$$

where $r \in \{1, 2, \ldots, \min(n_1, n_2)\}$.

References

[1] F. D'Eramo and J. Thaler, "Semi-annihilation of Dark Matter," *JHEP* **06** (2010) 109, arXiv:1003.5912 [hep-ph].

[2] T. Hambye, "Hidden vector dark matter," *JHEP* **01** (2009) 028, arXiv:0811.0172 [hep-ph].

[3] B. Batell, "Dark Discrete Gauge Symmetries," *Phys. Rev. D* **83** (2011) 035006, arXiv:1007.0045 [hep-ph].

[4] G. Belanger, K. Kannike, A. Pukhov, and M. Raidal, "Impact of semi-annihilations on dark matter phenomenology - an example of Z_N symmetric scalar dark matter," *JCAP* **04** (2012) 010, arXiv:1202.2962 [hep-ph].

[5] I. P. Ivanov and V. Keus, "Z_p scalar dark matter from multi-Higgs-doublet models," *Phys. Rev. D* **86** (2012) 016004, arXiv:1203.3426 [hep-ph].

[6] G. Bélanger, K. Kannike, A. Pukhov, and M. Raidal, "Minimal semi-annihilating $\mathbb{Z}_N$ scalar dark matter," *JCAP* **06** (2014) 021, arXiv:1403.4960 [hep-ph].

[7] Y. Cai and A. Spray, "A Systematic Effective Operator Analysis of Semi-Annihilating Dark Matter," *JHEP* **02** (2017) 120, arXiv:1611.09360 [hep-ph].

[8] G. Arcadi, F. S. Queiroz, and C. Siqueira, "The Semi-Hooperon: Gamma-ray and anti-proton excesses in the Galactic Center," *Phys. Lett. B* **775** (2017) 196–205, arXiv:1706.02336 [hep-ph].

[9] A. Kamada, H. J. Kim, H. Kim, and T. Sekiguchi, "Self-Heating Dark Matter via Semiannihilation," *Phys. Rev. Lett.* **120** no. 13, (2018) 131802, arXiv:1707.09238 [hep-ph].

[10] F. S. Queiroz and C. Siqueira, "Search for Semi-Annihilating Dark Matter with Fermi-LAT, H.E.S.S., Planck, and the Cherenkov Telescope Array," *JCAP* **04** (2019) 048, arXiv:1901.10494 [hep-ph].

[11] C. E. Yaguna and O. Zapata, "Multi-component scalar dark matter from a Z_N symmetry: a systematic analysis," *JHEP* **03** (2020) 109, arXiv:1911.05515 [hep-ph].

[12] G. Bélanger, A. Pukhov, C. E. Yaguna, and O. Zapata, "The Z_5 model of two-component dark matter," *JHEP* **09** (2020) 030, arXiv:2006.14922 [hep-ph].

[13] G. Belanger, A. Mjallal, and A. Pukhov, "Two dark matter candidates: The case of inert doublet and singlet scalars," *Phys. Rev. D* **105** no. 3, (2022) 035018, arXiv:2108.08061 [hep-ph].

[14] C. E. Yaguna and O. Zapata, "Fermion and scalar two-component dark matter from a Z4 symmetry," *Phys. Rev. D* **105** no. 9, (2022) 095026, arXiv:2112.07020 [hep-ph].

[15] B. Díaz Sáez and K. Ghorbani, "Z $_3$ scalar dark matter with strong positron fluxes," *JCAP* **02** (2023) 002, arXiv:2203.09282 [hep-ph].

[16] G. Belanger, A. Mjallal, and A. Pukhov, "WIMP and FIMP dark matter in the inert doublet plus singlet model," *Phys. Rev. D* **106** no. 9, (2022) 095019, arXiv:2205.04101 [hep-ph].

[17] P. Bandyopadhyay, D. Choudhury, and D. Sachdeva, "Semiannihilation of fermionic dark matter," *Phys. Rev. D* **107** no. 1, (2023) 015020, arXiv:2206.05811 [hep-ph].

[18] W. Cheng, X. Liu, and R. Zhou, "Non-minimal coupling inflation and dark matter under the $\mathbb{Z}_3$ symmetry," *JCAP* **05** (2023) 049, arXiv:2206.12624 [hep-ph].

[19] A. Bas i Beneito, J. Herrero-García, and D. Vatsyayan, "Multi-component dark sectors: symmetries, asymmetries and conversions," *JHEP* **10** (2022) 075, arXiv:2207.02874 [hep-ph].

[20] N. Benincasa, A. Hryczuk, K. Kannike, and M. Laletin, "Phase Transitions and Gravitational Waves in a Model of $\mathbb{Z}_3$ Scalar Dark Matter," arXiv:2312.04627 [hep-ph].

[21] M. Cirelli, N. Fornengo, and A. Strumia, "Minimal dark matter," *Nucl. Phys. B* **753** (2006) 178–194, arXiv:hep-ph/0512090.

[22] M. Gustafsson, E. Lundstrom, L. Bergstrom, and J. Edsjo, "Significant Gamma Lines from Inert Higgs Dark Matter," *Phys. Rev. Lett.* **99** (2007) 041301, arXiv:astro-ph/0703512.

[23] T. Hambye, F. S. Ling, L. Lopez Honorez, and J. Rocher, "Scalar Multiplet Dark Matter," *JHEP* **07** (2009) 090, arXiv:0903.4010 [hep-ph]. [Erratum: JHEP 05, 066 (2010)].

[24] C. Garcia-Cely, M. Gustafsson, and A. Ibarra, "Probing the Inert Doublet Dark Matter Model with Cherenkov Telescopes," *JCAP* **02** (2016) 043, arXiv:1512.02801 [hep-ph].

[25] T. A. Chowdhury and S. Nasri, "The Sommerfeld Enhancement in the Scotogenic Model with Large Electroweak Scalar Multiplets," *JCAP* **01** (2017) 041, arXiv:1611.06590 [hep-ph].

[26] S. Jangid and P. Bandyopadhyay, "Distinguishing Inert Higgs Doublet and Inert Triplet Scenarios," *Eur. Phys. J. C* **80** no. 8, (2020) 715, arXiv:2003.11821 [hep-ph].

[27] A. Jueid, J. Kim, S. Lee, S. Y. Shim, and J. Song, "Phenomenology of the Inert Doublet Model with a global U(1) symmetry," *Phys. Rev. D* **102** no. 7, (2020) 075011, arXiv:2006.10263 [hep-ph].

[28] S. Bottaro, D. Buttazzo, M. Costa, R. Franceschini, P. Panci, D. Redigolo, and L. Vittorio, "The last complex WIMPs standing," *Eur. Phys. J. C* **82** no. 11, (2022) 992, arXiv:2205.04486 [hep-ph].

[29] T. Araki, C. Q. Geng, and K. I. Nagao, "Dark Matter in Inert Triplet Models," *Phys. Rev. D* **83** (2011) 075014, arXiv:1102.4906 [hep-ph].

[30] F.-X. Josse-Michaux and E. Molinaro, "Triplet scalar dark matter and leptogenesis in an inverse seesaw model of neutrino mass generation," *Phys. Rev. D* **87** no. 3, (2013) 036007, arXiv:1210.7202 [hep-ph].

[31] S. Yaser Ayazi and S. M. Firouzabadi, "Constraining Inert Triplet Dark Matter by the LHC and FermiLAT," *JCAP* **11** (2014) 005, arXiv:1408.0654 [hep-ph].

[32] N. Khan, "Exploring the hyperchargeless Higgs triplet model up to the Planck scale," *Eur. Phys. J. C* **78** no. 4, (2018) 341, arXiv:1610.03178 [hep-ph].

[33] W.-B. Lu and P.-H. Gu, "Mixed Inert Scalar Triplet Dark Matter, Radiative Neutrino Masses and Leptogenesis," *Nucl. Phys. B* **924** (2017) 279–311, arXiv:1611.02106 [hep-ph].

[34] K. Kadota and A. Spray, "Electroweak Multiplet Dark Matter at Future Lepton Colliders," *JHEP* **02** (2019) 017, arXiv:1811.00560 [hep-ph].

[35] S. Bottaro, D. Buttazzo, M. Costa, R. Franceschini, P. Panci, D. Redigolo, and L. Vittorio, "Closing the window on WIMP Dark Matter," *Eur. Phys. J. C* **82** no. 1, (2022) 31, arXiv:2107.09688 [hep-ph].

[36] S. S. AbdusSalam and T. A. Chowdhury, "Scalar Representations in the Light of Electroweak Phase Transition and Cold Dark Matter Phenomenology," *JCAP* **05** (2014) 026, arXiv:1310.8152 [hep-ph].

[37] C. Cai, Z.-H. Yu, and H.-H. Zhang, "CEPC Precision of Electroweak Oblique Parameters and Weakly Interacting Dark Matter: the Scalar Case," *Nucl. Phys. B* **924** (2017) 128–152, arXiv:1705.07921 [hep-ph].

[38] Y.-P. Zeng, C. Cai, D.-Y. Liu, Z.-H. Yu, and H.-H. Zhang, "Probing quadruplet scalar dark matter at current and future *pp* colliders," *Phys. Rev. D* **101** no. 11, (2020) 115033, arXiv:1910.09431 [hep-ph].

[39] M. Cirelli, A. Strumia, and M. Tamburini, "Cosmology and Astrophysics of Minimal Dark Matter," *Nucl. Phys. B* **787** (2007) 152–175, arXiv:0706.4071 [hep-ph].

[40] K. Earl, K. Hartling, H. E. Logan, and T. Pilkington, "Constraining models with a large scalar multiplet," *Phys. Rev. D* **88** (2013) 015002, arXiv:1303.1244 [hep-ph].

[41] C. Garcia-Cely, A. Ibarra, A. S. Lamperstorfer, and M. H. G. Tytgat, "Gamma-rays from Heavy Minimal Dark Matter," *JCAP* **10** (2015) 058, arXiv:1507.05536 [hep-ph].

[42] C. Cai, Z. Kang, Z. Luo, Z.-H. Yu, and H.-H. Zhang, "Scalar quintuplet minimal dark matter with Yukawa interactions: perturbative up to the Planck scale," *Chin. Phys. C* **43** no. 2, (2019) 023102, arXiv:1711.07396 [hep-ph].

[43] W. Chao, G.-J. Ding, X.-G. He, and M. Ramsey-Musolf, "Scalar Electroweak Multiplet Dark Matter," *JHEP* **08** (2019) 058, arXiv:1812.07829 [hep-ph].

[44] K. Earl, K. Hartling, H. E. Logan, and T. Pilkington, "Two viable large scalar multiplet models with a Z_2 symmetry," *Phys. Rev. D* **90** no. 5, (2014) 055029, arXiv:1311.3656 [hep-ph]. [Erratum: Phys.Rev.D 92, 039902 (2015)].

[45] H. E. Logan and T. Pilkington, "Large scalar multiplet dark matter in the high-mass region," *Phys. Rev. D* **96** no. 1, (2017) 015030, arXiv:1610.08835 [hep-ph].

[46] D.-Y. Liu, C. Cai, Z.-H. Yu, Y.-P. Zeng, and H.-H. Zhang, "Inert sextuplet scalar dark matter at the LHC and future colliders," *JHEP* **10** (2020) 212, arXiv:2008.06821 [hep-ph].

[47] Y. Cai, W. Chao, and S. Yang, "Scalar Septuplet Dark Matter and Enhanced $h \to \gamma\gamma$ Decay Rate," *JHEP* **12** (2012) 043, arXiv:1208.3949 [hep-ph].

[48] C. Cai, Z.-M. Huang, Z. Kang, Z.-H. Yu, and H.-H. Zhang, "Perturbativity Limits for Scalar Minimal Dark Matter with Yukawa Interactions: Septuplet," *Phys. Rev. D* **92** no. 11, (2015) 115004, arXiv:1510.01559 [hep-ph].

[49] O. Fischer and J. J. van der Bij, "Multi-singlet and singlet-triplet scalar dark matter," *Mod. Phys. Lett. A* **26** (2011) 2039–2049.

[50] O. Fischer and J. J. van der Bij, "The scalar Singlet-Triplet Dark Matter Model," *JCAP* **01** (2014) 032, arXiv:1311.1077 [hep-ph].

[51] C. Cheung and D. Sanford, "Simplified Models of Mixed Dark Matter," *JCAP* **02** (2014) 011, arXiv:1311.5896 [hep-ph].

[52] M. Kakizaki, A. Santa, and O. Seto, "Phenomenological signatures of mixed complex scalar WIMP dark matter," *Int. J. Mod. Phys. A* **32** no. 10, (2017) 1750038, arXiv:1609.06555 [hep-ph].

[53] X. Liu and L. Bian, "Dark matter and electroweak phase transition in the mixed scalar dark matter model," *Phys. Rev. D* **97** no. 5, (2018) 055028, arXiv:1706.06042 [hep-ph].

[54] C. Cai, Z. Kang, H.-H. Zhang, and Y.-P. Zeng, "Minimal dark matter in $SU(2)_L \times U(1)_Y \times U(1)_{B-L}$," *Phys. Lett. B* **784** (2018) 385–391, arXiv:1801.05594 [hep-ph].

[55] A. Filimonova and S. Westhoff, "Long live the Higgs portal!," *JHEP* **02** (2019) 140, arXiv:1812.04628 [hep-ph].

[56] K. Hally, H. E. Logan, and T. Pilkington, "Constraints on large scalar multiplets from perturbative unitarity," *Phys. Rev. D* **85** (2012) 095017, arXiv:1202.5073 [hep-ph].

[57] A. Sommerfeld, "Über die Beugung und Bremsung der Elektronen," *Annalen Phys.* **403** no. 3, (1931) 257–330.

[58] A. Strumia, "Sommerfeld corrections to type-II and III leptogenesis," *Nucl. Phys. B* **809** (2009) 308–317, arXiv:0806.1630 [hep-ph].

[59] I. Harris and L. M. Brown, "Radiative Corrections to Pair Annihilation," *Phys. Rev.* **105** (1957) 1656–1661.

[60] R. J. Hill and M. P. Solon, "Standard Model anatomy of WIMP dark matter direct detection II: QCD analysis and hadronic matrix elements," *Phys. Rev. D* **91** (2015) 043505, arXiv:1409.8290 [hep-ph].

[61] J. Ellis, N. Nagata, and K. A. Olive, "Uncertainties in WIMP Dark Matter Scattering Revisited," *Eur. Phys. J. C* **78** no. 7, (2018) 569, arXiv:1805.09795 [hep-ph].

[62] Q. Chen, G.-J. Ding, and R. J. Hill, "General heavy WIMP nucleon elastic scattering," *Phys. Rev. D* **108** no. 11, (2023) 116023, arXiv:2309.02715 [hep-ph].

[63] LZ Collaboration, J. Aalbers *et al.*, "First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment," *Phys. Rev. Lett.* **131** no. 4, (2023) 041002, arXiv:2207.03764 [hep-ex].

[64] M. E. Peskin and T. Takeuchi, "A New constraint on a strongly interacting Higgs sector," *Phys. Rev. Lett.* **65** (1990) 964–967.

[65] H. Beauchesne and C.-W. Chiang, "Is the Decay of the Higgs Boson to a Photon and a Dark Photon Currently Observable at the LHC?," *Phys. Rev. Lett.* **130** no. 14, (2023) 141801, arXiv:2205.10976 [hep-ph].

[66] H. Beauchesne and C.-W. Chiang, "Observability of the Higgs boson decay to a photon and a dark photon," *Phys. Rev. D* **108** no. 1, (2023) 015018, arXiv:2304.04165 [hep-ph].

[67] H. Beauchesne and C.-W. Chiang, "Contributions of inert electroweak multiplets to Higgs properties," *JHEP* **10** (2023) 170, arXiv:2308.06901 [hep-ph].

[68] K. Griest and D. Seckel, "Three exceptions in the calculation of relic abundances," *Phys. Rev. D* **43** (1991) 3191–3203.

[69] C.-W. Chiang, G. Cottin, Y. Du, K. Fuyuto, and M. J. Ramsey-Musolf, "Collider Probes of Real Triplet Scalar Dark Matter," *JHEP* **01** (2021) 198, arXiv:2003.07867 [hep-ph].

[70] J. Arakawa and T. M. P. Tait, "Is a Miracle-less WIMP Ruled out?," *SciPost Phys.* **11** no. 2, (2021) 019, arXiv:2101.11031 [hep-ph].

[71] T. Katayose, S. Matsumoto, S. Shirai, and Y. Watanabe, "Thermal real scalar triplet dark matter," *JHEP* **09** (2021) 044, arXiv:2105.07650 [hep-ph].

arXiv:2403.01727v1 [hep-ph] 4 Mar 2024

Identify the new state $Y(3872)$ as the P-wave $D\bar{D}^*/\bar{D}D^*$ resonance

Zi-Yang Lin [1, *] Jun-Zhang Wang [2, †] Jian-Bo Cheng [3, ‡] Lu Meng [4, §] and Shi-Lin Zhu [2, ¶]

[1] *School of Physics, Peking University, Beijing 100871, China*
[2] *School of Physics and Center of High Energy Physics, Peking University, Beijing 100871, China*
[3] *College of Science, China University of Petroleum, Qingdao, Shandong 266580, China*
[4] *Institut für Theoretische Physik II, Ruhr-Universität Bochum, D-44780 Bochum, Germany*

The BESIII Collaboration recently observed a new charmonium-like vector state $Y(3872)$ in $e^+e^- \to D\bar{D}$, which should be the first P-wave $D\bar{D}^*/\bar{D}D^*$ molecular resonance. The experimental and theoretical identification of the P-wave dimeson state holds paramount importance in enhancing our comprehension of the non-perturbative QCD and few-body physics. Its existence is firmly established in a unified meson-exchange model which simultaneously depicts the features of the $\chi_{c1}(3872)$, $Z_c(3900)$ and $T_{cc}(3875)$. This scenario can be directly examined in the $e^+e^- \to D\bar{D}^*/\bar{D}D^*$ cross section to see whether a resonance exists at the threshold. The credibility of the investigations is also ensured by the fact that the P-wave interaction dominantly arises from the well-known long-range pion exchange. Additionally, the existence of the P-wave resonance only depends on the interaction strength and is less sensitive to the potential shapes. We extensively calculate all systems up to P-wave with various quantum numbers and predict a dense population of the $D\bar{D}^*/\bar{D}D^*$ and DD^* states, where the S-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^-(1^{+-})$, P-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^+(0^{-+})$, and P-wave DD^* state with $I(J^P) = 0(0^-)$ are more likely to be observed in experiments.

Introduction.— Over the past two decades, a significant number of hadrons defying the spectra predicted by quark models have been observed in the heavy flavor sector, which are typically regarded as the exotica within the realm of Quantum Chromodynamics (QCD), see Refs. [1–7] for reviews. Delving into the structure and dynamics associated with these exotic states holds paramount importance in enhancing our comprehension of the non-perturbative features of low-energy QCD. These states also serve as promising examples for studying the general few-body physics.

Among these exotic states, the $\chi_{c1}(3872)$, $Z_c(3900)$, and $T_{cc}(3875)$ stand out as undeniable "star" examples, believed to be the first charmonium-like state [8], the first manifestly exotic charmonium-like state [9, 10], and the first doubly charmed tetraquark state [11, 12] observed in experiments, respectively. It is particularly intriguing that these three states are closely interconnected. The proximity of the former two states to the $D\bar{D}^*/\bar{D}D^*$ threshold and the latter one to the DD^* threshold positions them as strong candidates for corresponding hadronic molecules. Indeed, prior to the observation of $T_{cc}(3875)$, Li *et al.* had predicted a very loosely bound state of DD^* utilizing the one-boson-exchange model (OBE), with parameters established beforehand while investigating the $\chi_{c1}(3872)$ [13, 14].

In the realm of doubly heavy exotic states, such as the $\chi_{c1}(3872)$, $Z_c(3900)$, and $T_{cc}(3875)$, previous studies have predominantly focused on S-wave dimeson states, encompassing bound states, virtual states, or resonances. However, P-wave states near the threshold are of particular interest and arouse the attention in many fields of physics, see the halo nuclei as P-wave resonances in nuclear physics [15] and the P-wave Feshbach resonances in cold atomic physics [16]. Recently, the BESIII Collaboration discovered a new resonance in $e^+e^- \to D\bar{D}$[17]. Apart from the established 1^{--} states $\psi(3770)$, $\psi(4040)$, $\psi(4160)$, $\psi(4230)$, $\psi(4360)$, $\psi(4415)$, $\psi(4660)$, they observed a new resonance with a significance of over 20σ. Its mass and width are fitted to be $3872.5 \pm 14.2 \pm 3.0$ MeV and $179.7 \pm 14.1 \pm 7.0$ MeV, respectively. Hereafter we use $Y(3872)$ to denote this state. It is worth noting that the coupled-channel analysis of data from Belle and BESIII has the potential to generate a bump at this position without introducing new states. However, it appears to be very challenging in accurately depicting the nearby points [18, 19]. The newly observed state, locating exactly at the $D\bar{D}^*$ threshold, turns out a good candidate of the P-wave $D\bar{D}^*$ resonance.

In this work, we aim to identify it as the first P-wave dimeson state in the doubly heavy sector in the meson-exchange model. By relating the $Y(3872)$ to the S-wave states $\chi_{c1}(3872)$, $Z_c(3900)$ and $T_{cc}^+(3985)$, we make a unified description of these states in the one-boson-exchange (OBE) interaction. The resonance poles are obtained by solving the complex scaled Schrödinger equation in momentum space (The details can be found in Ref. [20]).

For the following three reasons, the predictions regarding the existence of the P-wave resonance are highly reliable. Just as the OBE model has provided a high-precision description of nuclear forces [21], meson-exchange models have also achieved notable success in elucidating heavy flavor hadronic molecules [22–30]. In the 1990s, Törnqvist predicted a deuteron-like $D\bar{D}^*$ bound state, which has been confirmed by the observa-

* lzy_15@pku.edu.cn
† wangjzh2022@pku.edu.cn
‡ jbcheng@upc.edu.cn
§ lu.meng@rub.de
¶ zhusl@pku.edu.cn

tion of the $\chi_{c1}(3872)$ [22, 23]. The interactions stemming from the exchange of π, η, ρ, ω, and σ particles naturally predict T_{cc} as a DD^* bound state once their parameters are determined in the $D\bar{D}^*/\bar{D}D^*$ systems, specifically the $\chi_{c1}(3872)$ [13, 14]. The interactions governing T_{cc} and $\chi_{c1}(3872)$ adhere to the G-parity rules. Given that both P-wave and S-wave states arise from the partial wave expansion of the same potential, the existence of these P-wave resonances could be firmly established once the S-wave interaction is fixed in depicting the $\chi_{c1}(3872)$, $Z_c(3900)$ and $T_{cc}^+(3985)$ states. Recently, similar ideas have been used to investigate the P-wave $D\bar{D}_1$, $D^*\bar{D}_1$ and $D^*\bar{D}_2^*$ states [31], inspired by the corresponding deeply bound S-wave states in Ref. [32]. The states of D^*N in the P-wave were also considered during the investigation of $\Lambda_c(2940)^+$ and its counterparts [33].

In principle, the generation of S-wave resonance poles typically hinges on the specifics of the potential, including its shape, coupled-channel effects, or the regularization method employed (see, for instance, Ref. [34]). Conversely, P-wave resonances can be generated by simply adjusting the strength of the interaction in a single-channel scenario. As the parameters in the potential are adjusted to make it less attractive (see Fig. 1), the bound state poles in the physical Riemann sheet tend to migrate into the unphysical sheet, manifesting as virtual states (S-wave) or resonances (higher partial waves)[35]. Consequently, the presence of S-wave virtual states and P- or higher-wave resonances relies primarily on the strength of the interaction rather than the shape of the potential. Thus the existence and the pole position of the P-wave resonance as well as the S-wave virtual state tend to be robust across different regularization methods and cutoff parameters.

Given the repulsion effect of the centrifugal barrier at the short range, the long-range interactions, specifically the one-pion-exchange (OPE) potential, will play a vital role in higher partial wave systems as the most peripheral part among the meson-exchange interactions. Indeed, taking the chiral effective field theory as an example, the leading-order chiral interaction of the P-wave system solely stems from the OPE interaction, while for the S-wave systems, the OPE interaction is accompanied with contact interactions to ensure the renormalization [7, 36–41]. Similar conclusions were also supported by analysing the lattice data [42]. Given that the coupling constants of OPE have been well determined by the partial decay width of D^*, P-wave resonances are typically predicted with high credibility.

Framework.— We adopt a framework established in Refs. [13, 14, 24, 25]. Under the heavy quark spin symmetry, the pseudoscalar D, the vector meson D^* and their antiparticles are combined into the superfield $\mathcal{H}$ and $\tilde{\mathcal{H}}$,

$$\mathcal{H} = \frac{1+\not{v}}{2}(P_\mu^*\gamma^\mu - P\gamma_5),\ \tilde{\mathcal{H}} = (\tilde{P}_\mu^*\gamma^\mu - \tilde{P}\gamma_5)\frac{1-\not{v}}{2}, \quad (1)$$

where $P = (D^0, D^+)$, $P_\mu^* = (D^{*0}, D^{*+})_\mu$, $\tilde{P} = (\bar{D}^0, D^-)^{\rm T}$, $\tilde{P}_\mu^* = (\bar{D}^{*0}, D^{*-})_\mu^{\rm T}$. $v = (1,0,0,0)$ is the velocity of the heavy meson. The conjugation of $\mathcal{H}$ and $\tilde{\mathcal{H}}$ is defined as $\bar{\mathcal{H}} = \gamma_0\mathcal{H}^\dagger\gamma_0$ and $\bar{\tilde{\mathcal{H}}} = \gamma_0\tilde{\mathcal{H}}^\dagger\gamma_0$. For the charge conjugation transformation, we adopt the convention $D \xrightarrow{\mathcal{C}} \bar{D}$ and $D^* \xrightarrow{\mathcal{C}} -\bar{D}^*$, namely $\mathcal{H} \xrightarrow{\mathcal{C}} C^{-1}\tilde{\mathcal{H}}^{\rm T}C$, where $C = i\gamma^2\gamma^0$. We include the π, η, σ, ρ, ω exchanges in the OBE model via the following Lagrangians,

$$\begin{aligned}\mathcal{L} &= g_s{\rm Tr}\left[\mathcal{H}\sigma\bar{\mathcal{H}}\right] + ig_a{\rm Tr}\left[\mathcal{H}\gamma_\mu\gamma_5\mathcal{A}^\mu\bar{\mathcal{H}}\right]\\ &+i\beta{\rm Tr}\left[\mathcal{H}v_\mu(\mathcal{V}^\mu-\rho^\mu)\bar{\mathcal{H}}\right] + i\lambda{\rm Tr}\left[\mathcal{H}\sigma_{\mu\nu}F^{\mu\nu}\bar{\mathcal{H}}\right]\\ &+g_s{\rm Tr}\left[\bar{\tilde{\mathcal{H}}}\sigma\tilde{\mathcal{H}}\right] + ig_a{\rm Tr}\left[\bar{\tilde{\mathcal{H}}}\gamma_\mu\gamma_5\mathcal{A}^\mu\tilde{\mathcal{H}}\right]\\ &-i\beta{\rm Tr}\left[\bar{\tilde{\mathcal{H}}}v_\mu(\mathcal{V}^\mu-\rho^\mu)\tilde{\mathcal{H}}\right] + i\lambda{\rm Tr}\left[\bar{\tilde{\mathcal{H}}}\sigma_{\mu\nu}F^{\mu\nu}\tilde{\mathcal{H}}\right] \quad (2)\end{aligned}$$

The vector meson fields ρ^μ and the pseudoscalar meson fields $\mathcal{M}$ are defined as

$$\rho^\mu = \frac{ig_V}{\sqrt{2}}\begin{pmatrix}\frac{\rho_0}{\sqrt{2}}+\frac{\omega}{\sqrt{2}} & \rho^+\\ \rho^- & -\frac{\rho_0}{\sqrt{2}}+\frac{\omega}{\sqrt{2}}\end{pmatrix}^\mu, \quad (3)$$

$$\mathbb{P} = \begin{pmatrix}\frac{\pi_0}{\sqrt{2}}+\frac{\eta}{\sqrt{6}} & \pi^+\\ \pi^- & -\frac{\pi_0}{\sqrt{2}}+\frac{\eta}{\sqrt{6}}\end{pmatrix}. \quad (4)$$

$F^{\mu\nu} = \partial^\mu\rho^\nu - \partial^\nu\rho^\mu - [\rho^\mu, \rho^\nu]$ represents the field strength tensor of vector mesons. $\mathcal{V}^\mu$ and $\mathcal{A}^\mu$ represent the vector and axial currents of pseudoscalar mesons, respectively

$$\begin{aligned}&\mathcal{V}^\mu = \frac{1}{2}[\xi^\dagger, \partial_\mu\xi], \quad \mathcal{A}^\mu = \frac{1}{2}\{\xi^\dagger, \partial_\mu\xi\},\\ &\xi = \exp(i\mathbb{P}/f_\pi). \quad (5)\end{aligned}$$

$f_\pi = 132$ MeV is the pion decay constant. The coupling constants are fixed to be consistent with Refs. [13, 14], resulting in a good depiction of $\chi_{c1}(3872)$ and a remarkable prediction of the T_{cc} state. The axial coupling constant $g_a = 0.59$ is extracted from the D^* width. The other coupling constants are $g_V = 5.8$, $\beta = 0.9$, $\lambda = 0.56$ GeV^{-1}, and $g_s = 0.76$. The isospin average masses of particles are taken from the Review of Particle Physics [43]: $m_\pi = 137$ MeV, $m_\eta = 548$ MeV, $m_\rho = 775$ MeV, $m_\omega = 783$ MeV, $m_D = 1867$ MeV, $m_{D^*} = 2009$ MeV. For the scalar meson exchange, we choose $m_\sigma = 600$ MeV.

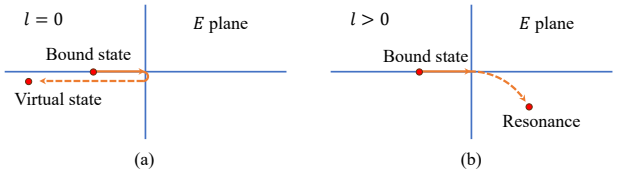


FIG. 1. Transition of the bound state pole to the virtual state (a) and resonance (b) for S-wave and higher partial waves, respectively by adjusting the strength of the potential to be less attractive [35]. The solid (dashed) lines represent the pole trajectories in the physical (unphysical) Riemann sheets.

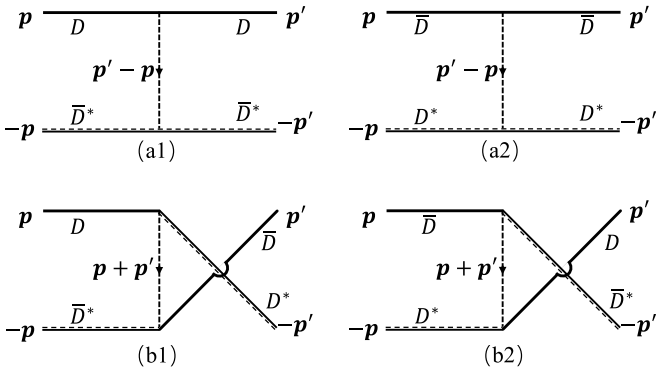


FIG. 2. The direct diagrams (upper row) and cross diagrams (lower row) in the OBE, where the transferred momenta are $\boldsymbol{p}' - \boldsymbol{p}$ and $\boldsymbol{p}' + \boldsymbol{p}$, respectively.

We construct the $D\bar{D}^*$ wave functions as the C-parity eigenstates for neutral channels,

$$|C = \pm\rangle = \frac{1}{\sqrt{2}}(|D(\boldsymbol{p})\bar{D}^*(-\boldsymbol{p})\rangle \mp |\bar{D}(\boldsymbol{p})D^*(-\boldsymbol{p})\rangle). \quad (6)$$

Here the momenta of the particles are shown explicitly. For the charged channels, we can similarly construct the wave functions as the eigenstates of G-parity. As depicted in Fig. 2, the transfer momentum in the cross diagrams corresponding to the u-channel turns out $\boldsymbol{k} = \boldsymbol{p} + \boldsymbol{p}'$, while it is $\boldsymbol{q} = \boldsymbol{p} - \boldsymbol{p}'$ in the direct diagrams or t-channel exchanges. It is explained in Supplemental Material [44] that the momentum labeling is crucial to get the correct P-wave interactions. To show the key mechanism of the P-wave resonance, we ignore the isospin breaking effect and adopt the time-component of $q^0, k^0 = 0$.

The effective potential for the $D\bar{D}^*/\bar{D}D^*$ system can be related to the DD^* potential up to a factor G_m $(-G_mG_{MM})$ for the direct (cross) diagrams, with G_m and G_{MM} as the G-parities of the exchanged meson and the $D\bar{D}^*/\bar{D}D^*$ system, respectively. It is noticeable that the G-parity rule for cross diagrams is different from that for direct diagrams. The specific effective potentials and the derivation of the G-parity rule are presented in Supplemental Material [44]. With the complex scaling method $p \to pe^{-i\theta}$, the resonance and bound state poles can be derived as the eigenenergy in the Schrödinger equation

$$E\phi(\boldsymbol{p}) = \frac{\boldsymbol{p}^2}{2\mu}\phi(\boldsymbol{p}) + \int V(\boldsymbol{p}, \boldsymbol{k})\phi(\boldsymbol{k})\frac{d^3\boldsymbol{k}}{(2\pi)^3}. \quad (7)$$

To search for virtual states, we adopt the method in Ref. [45].

To regularize the ultraviolet divergence in the integral, we introduce a monopole regulator to suppress the potential at the large momentum

$$V(\boldsymbol{p}', \boldsymbol{p}) \to V(\boldsymbol{p}', \boldsymbol{p})\frac{\Lambda^2}{p'^2 + \Lambda^2}\frac{\Lambda^2}{p^2 + \Lambda^2}. \quad (8)$$

Our final results are nearly irrelevant to the specific choice of regulator. The cutoff Λ is the only parameter to be determined. We adjust Λ to generate a pole at the threshold (a loosely bound state or a near-threshold virtual state) in the $^3S_0$ isosinglet $D\bar{D}^*/\bar{D}D^*$ system with the positive C-parity for neutral components, namely the 1^{++} channel corresponding to $\chi_{c1}(3872)$. Then we search for poles with different isospins, C-parities, obital angular momenta (S-wave and P-wave) in $D\bar{D}^*/\bar{D}D^*$ and DD^* systems.

Results and discussion.— The partial-wave potentials of $J = 1$ isosinglet $D\bar{D}^*/\bar{D}D^*$ systems are depicted in Fig. 3. In P-wave interactions, the significance of the pion exchange increases, whereas the S-wave interaction is predominantly governed by the ρ exchange. The $\chi_{c1}(3872)$ corresponds to the 1^{++} S-wave channel, exhibiting the most pronounced attraction. Its negative C-parity counterpart, the 1^{+-} S-wave channel, also displays an attractive potential. Consequently, these two S-wave channels may give rise to near-threshold bound states or virtual states. The 1^{-+} channel, serving as the P-wave counterpart of 1^{++}, demonstrates substantial repulsion, thus making it unlikely to produce poles near the threshold. However, the potential of the 1^{--} channel, which is the P-wave partner of the 1^{+-} channel, is attractive, suggesting a possible resonance pole corresponding to the $Y(3872)$.

In Fig. 4, we illustrate the pole trajectories of four particularly intriguing states: the $\chi_{c1}(3872)$, $T_{cc}(3875)$, $Z_c(3900)$, and the recently observed $Y(3872)$, as the cutoff parameter Λ varies from 0.4 GeV to 1.3 GeV. With a cutoff of around 0.5 GeV, the $\chi_{c1}(3872)$ manifests as a loosely bound state. The T_{cc} is also a near-threshold bound state, which agrees with the results in Ref. [14]. Simultaneously, the $Z_c(3900)$ emerges as a virtual state, aligning with the pole position deduced through a data-driven coupled-channel analysis in Ref. [19]. Remarkably, within this same cutoff range, a P-wave resonance materializes in the 1^{--} channel, corresponding to the $Y(3872)$ state. If the cutoff is increased to strengthen the attraction, the $Y(3872)$ resonance will move to the physical Riemann sheet and turns into a bound state, thereby confirming the $Y(3872)$ as indeed a P-wave resonance engendered by adjusting the interaction strength. The value of the cutoff differs from the results in Refs. [14, 46], since the regulator is different. However, our conclusion holds under different regulators. We test the results using the regulator and cutoff fixed in Ref. [46]. We validate these findings using the regulator and cutoff parameters established in Ref. [46]. The results indicate that as long as the cutoff is set to generate a loosely bound $\chi_{c1}(3872)$ state, a corresponding P-wave resonance emerges in the 1^{--} channel as the $Y(3872)$, while the poles of T_{cc} and $Z_c(3900)$ remain qualitatively unchanged, see the Supplemental Material [44].

We delve into the resonances with alternative quantum numbers, as summarized in Table I. We neglect the tiny imaginary part of virtual state pole arsing from the left-

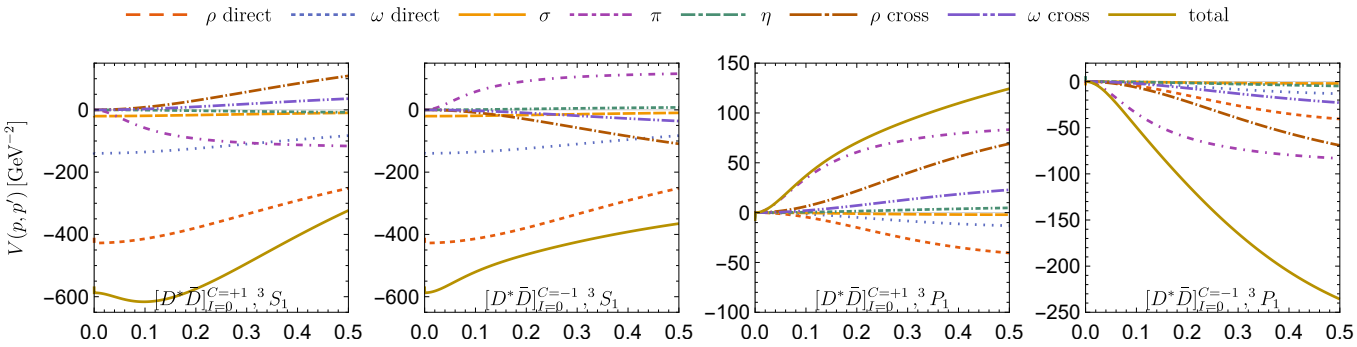


FIG. 3. The OBE potentials in the 1^{++}, 1^{+-}, 1^{-+} and 1^{--} isoscalar $D\bar{D}^*/\bar{D}D^*$ channels. The 1^{++} and 1^{-+} channel correspond to the $\chi_{c1}(3872)$ and its P-wave counterpart, respectively. The 1^{--} and 1^{+-} channel corresponds to the $Y(3872)$ and its S-wave counterpart, respectively. Only $p = p'$ cases are shown.

TABLE I. The poles in all channels of $D\bar{D}^*$ and DD^*, up to the orbital angular momentum $L = 1$. The B and V superscripts denote the bound state and the virtual state, respectively. Otherwise the pole refers to a resonance.

		$D\bar{D}^*, C=+$		$D\bar{D}^*, C=-$		DD^*	
		$I=0$	$I=1$	$I=0$	$I=1$	$I=0$	$I=1$
$\Lambda = 0.5$GeV	$1^+(^3S_1)$	-3.1^B, $\chi_{c1}(3872)$	-	-1.60^B	-35.6^V, $Z_c(3900)$	-0.41^B, $T_{cc}(3875)$	-
	$0^-(^3P_0)$	$-1.5-14.5i$	-	-	-	$-9.6-9.7i$	-
	$1^-(^3P_1)$	-	-	$-4.0-27.3i$, $Y(3872)$	-	$-31.7-70.6i$	-
	$2^-(^3P_2)$	$-42.6-39.4i$	-	$-21.3-50.7i$	-	$-37.8-40.9i$	-
$\Lambda = 0.6$GeV	$1^+(^3S_1)$	-6.5^B, $\chi_{c1}(3872)$	-	-5.8^B	-34.6^V, $Z_c(3900)$	-4.3^B, $T_{cc}(3875)$	-
	$0^-(^3P_0)$	$3.2-13.7i$	-	-	-	$-10.2-12.1i$	-
	$1^-(^3P_1)$	-	-	$2.0-27.3i$, $Y(3872)$	-	$-33.7-84.8i$	-
	$2^-(^3P_2)$	$-44.2-49.0i$	-	$-19.3-58.8i$	-	$-37.8-49.3i$	-

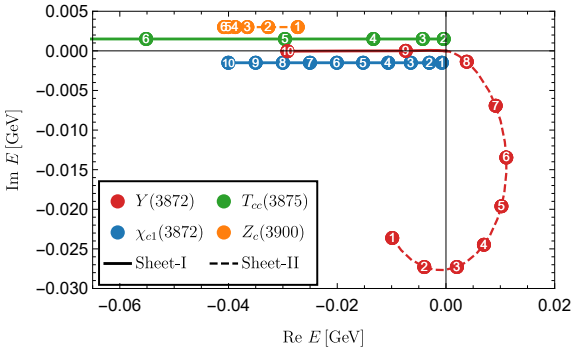


FIG. 4. The pole trajectories with the cutoff parameters correspond to $\chi_{c1}(3872)$, $T_{cc}(3875)$, $Z_c(3900)$ and the newly observed $Y(3872)$ states. The circled number 1-10 represent the increasing cutoff 0.4-1.3 GeV in order. The solid (dashed) lines represent the pole trajectories in the physical (unphysical) Riemann sheets. The poles on the negative real axis are slightly shifted for transparency.

hand cut of the pion exchange. For $D\bar{D}^*/\bar{D}D^*$ isospin singlets, aside from the $\chi_{c1}(3872)$ and $Y(3872)$, a loosely bound state exists in the S-wave partner channel of the $Y(3872)$. In channels with $J = 0$ and $J = 2$, resonances emerge in 0^{-+}, 2^{--}, and 2^{-+} channels. For the DD^* systems, in addition to the T_{cc} state as the S-wave isospin singlet, its P-wave partner with $J^P = 0^-$ also emerges as a near-threshold resonance. P-wave resonance poles are obtained in the 1^- and 2^- channels but are distant from the thresholds.

Concerning $I = 1$ channels, besides a virtual state pole in the S-wave 1^{+-} channel, corresponding to the $Z_c(3900)$, no additional states are obtained. This is because the isospin factor $\tau \cdot \tau$ is only $\frac{1}{3}$ of the $I = 0$ channels, rendering the potentials generally insufficient to generate bound states or P-wave resonances. For clarification, we do find resonance poles in 0^{-+}, 1^{--} isovector $D\bar{D}^*/\bar{D}D^*$ channels and 1^-, 2^- isovector DD^* channels, but they are very sensitive to the cutoff. They transform into virtual states rather than bound states when adjust-

ing the interaction strength, thereby are not the kind of P-wave resonances we refer to in Fig. 1 (b). Their existence depends on the particular regularization, making them less credible. So we omit them in the final results.

To sum up, aside from the $\chi_{c1}(3872)$, $Z_c(3900)$, $Y(3872)$, and T_{cc} states, the S-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^-(1^{+-})$, P-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^+(0^{-+})$, and P-wave DD^* state with $I(J^P) = 0(0^-)$ are more likely to be observed due to their proximity to the thresholds.

Conclusion and outlook.— The newly observed $Y(3872)$ is interpreted as the P-wave $D\bar{D}^*/\bar{D}D^*$ resonance in a novel scenario. The existence of the P-wave resonance is firmly established on a unified meson-exchange model which well depicts the features of $\chi_{c1}(3872)$, $Z_c(3900)$ and $T_{cc}(3875)$ simultaneously. Compared with the S-wave state, the P-wave interaction dominantly arises from the well-known long-range pion exchange, ensuring robust conclusions when shifting between different models. The appearance of the P-wave resonance is also quite natural, particularly when the P-wave channels lack sufficient attraction, thus rendering them less sensitive to the potential shape compared with the S-wave resonance. This mechanism contributes to the dense population of P-wave resonances in both the $D\bar{D}^*/\bar{D}D^*$ and DD^* systems, which is validated by our extensive calculations spanning all systems up to P-wave with various quantum numbers.

Furthermore, there is promise in identifying P-wave resonances in other systems. For instance, the odd-parity $X_1(2900)$ observed in LHCb alongside $X_0(2900)$ [47, 48] may be plausibly interpreted as the P-wave $\bar{D}^*K^*$ resonance. The $\psi(4220)$ state may potentially be interpreted as the P-wave $D_s^*\bar{D}_s^*$ resonance $Y(4220)$. Similarly, there may exist the P-wave $D_s^*\bar{D}_s/D_s\bar{D}_s^*$, $D_s\bar{D}_s$, $D^*\bar{D}^*$ and $D\bar{D}$ near-threshold resonances. One may also expect similar P-wave structures in the two bottom meson systems.

The scenario of identifying $Y(3872)$ as the P-wave resonance can be directly examined in the $e^+e^- \to D\bar{D}^*$ cross section to see whether a resonance exists at the threshold. In the case of the $Y(3872)$ being a resonance with the real part of the pole position below the threshold, its width comes from the decay to the $D\bar{D}^*$ final state, where the large width renders the decay allowable. Remarkably, for the pole below the threshold, the line shape deviates severely from the Breit-Wigner form and needs other parameterization methods like the Flatté form.

Among the abundant predictions, the S-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^-(1^{+-})$, P-wave $D\bar{D}^*/\bar{D}D^*$ state with $I^G(J^{PC}) = 0^+(0^{-+})$, and P-wave DD^* state with $I(J^P) = 0(0^-)$ are more likely to be observed due to their proximity to the thresholds. Unlike the 1^{--} $Y(3872)$ state, their decay mode to $D\bar{D}$ system is forbidden. Thus, these predictions could be searched in the hidden charmed channels, for example, the 1^{+-} state in $\eta_c\omega$, $J/\psi\eta$, $J/\psi\pi\pi$, the 0^{-+} state in $J/\psi\omega$, $\eta_c\pi\pi$ and $\chi_{c1}\pi\pi$, the 2^{-+} state in $J/\psi\omega$, $\chi_{c1}\pi\pi$, the 2^{--} state in $J/\psi\eta$, $\eta_c\omega$ etc.

ACKNOWLEDGMENTS

This project was supported by the National Natural Science Foundation of China (11975033, 12147168 and 12070131001). This project was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Project ID 196253076-TRR 110). J.Z.W. is also supported by the National Postdoctoral Program for Innovative Talent.

Appendix A: Sign problem of the u-channel potentials

In Fig. 2 of the main text, the u-channel diagrams of the OBE are involved. However, a prevailing misconception exists in much of the literature regarding these u-channel OBE diagrams. We aim to address this misconception and illustrate its impact, revealing that while it does not introduce errors for the systems with even orbital angular momentum, it does induce a sign alteration in the partial wave potential for odd orbital angular momentum.

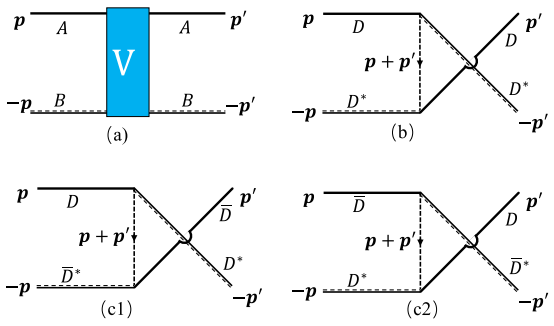


FIG. 5. The momentum labeling for the u-channel diagrams.

In the general elastic scattering depicted for two distinguishable particles, denoted as A and B, as illustrated in Fig. 5 (a), the nonlocal potential in the center-of-mass frame can be expressed as:

$$V(\boldsymbol{p}',\boldsymbol{p}) \equiv \langle \boldsymbol{p}'|\hat{V}|\boldsymbol{p}\rangle \equiv \langle A(\boldsymbol{p}')B(-\boldsymbol{p}')|\hat{V}|A(\boldsymbol{p})B(-\boldsymbol{p})\rangle, \tag{A1}$$

Here, we adopt the momentum of the particle A to label the two-particle states, denoted as $|\boldsymbol{p}\rangle$, which signifies $|A(\boldsymbol{p})B(-\boldsymbol{p})\rangle$. One can get the Lippmann-Schwinger equation in momentum space by sandwiching the operator equation $\hat{T} = \hat{V} + \hat{V}\hat{G}\hat{T}$ between initial and final two-body states and inserting complete basis between operators,

$$T(\boldsymbol{p}',\boldsymbol{p};E) = V(\boldsymbol{p}',\boldsymbol{p}) + \int \frac{d\boldsymbol{p}''}{(2\pi)^3} V(\boldsymbol{p}',\boldsymbol{p}'')G(E,\boldsymbol{p})T(\boldsymbol{p}'',\boldsymbol{p};E), \tag{A2}$$

with

$$G(E,\boldsymbol{p}) = \frac{1}{E - m_A - m_B - \frac{p^2}{2m_A} - \frac{p^2}{2m_B} + i\epsilon}. \quad \text{(A3)}$$

It is worth noting that in Eq. (A2), all the three-momenta refer specifically to those of particle A.

Now we can specify A and B as D and D^*, respectively taking OPE interaction as an example as shown in Fig 5 (b). Apparently, the momentum of the exchanged pion should be $\boldsymbol{k} = \boldsymbol{p} + \boldsymbol{p}'$ rather than the conventional $\boldsymbol{q} = \boldsymbol{p}' - \boldsymbol{q}$ used for the t-channel diagram.

For the $D^*\bar{D}/\bar{D}^*D$ system, one can construct the state with C-parity for the neutron channel,

$$|D\bar{D}^*/\bar{D}D^*, \{\beta, \boldsymbol{p}\}\rangle \equiv |D(\boldsymbol{p})\bar{D}^*(-\boldsymbol{p}) + \beta\bar{D}(\boldsymbol{p})D^*(-\boldsymbol{p})\rangle. \quad \text{(A4)}$$

To ensure the fixed C-parity, the $\bar{D}(D^*)$ in the second component should has the same momentum as the $D(\bar{D}^*)$ in the fist component. One can get

$$\hat{C}|D\bar{D}^*/\bar{D}D^*, \{\beta, \boldsymbol{p}\}\rangle = -\beta|D\bar{D}^*/\bar{D}D^*, \{\beta, \boldsymbol{p}\}\rangle. \quad \text{(A5)}$$

with the convention

$$\hat{C}|D(\boldsymbol{p})\rangle = |\bar{D}(\boldsymbol{p})\rangle;\ \hat{C}|D^*(\boldsymbol{p})\rangle = -|\bar{D}^*(\boldsymbol{p})\rangle \quad \text{(A6)}$$

Taking OPE as an example, the Feynman diagrams involved are shown in Fig. 5 (c1) and (c2). Once again, the momentum of the exchanged pion should be $\boldsymbol{k} = \boldsymbol{p} + \boldsymbol{p}'$.

As far as we know, in much of the literature discussing u-channel OBE diagrams, there is a common mistake regarding the momentum of the exchanged meson, where it is erroneously taken as $\boldsymbol{q} = \boldsymbol{p}' - \boldsymbol{p}$. Using $\boldsymbol{q}$ as $\boldsymbol{k}$ is equivalent to substituting $V(\boldsymbol{p}', \boldsymbol{p})$ with $V(\boldsymbol{p}', -\boldsymbol{p})$. Consequently, the potential term in the Schrödinger equation becomes:

$$\begin{aligned} &\int d^3\boldsymbol{p} V(\boldsymbol{p}', -\boldsymbol{p})\phi_L(\boldsymbol{p}) \\ =\ &\int d^3\boldsymbol{p} V(\boldsymbol{p}', \boldsymbol{p})\phi_L(-\boldsymbol{p}) \\ =\ &(-1)^L \int d^3\boldsymbol{p} V(\boldsymbol{p}', \boldsymbol{p})\phi_L(\boldsymbol{p}) \end{aligned} \quad \text{(A7)}$$

where $\phi_L(\boldsymbol{p})$ represents wave function with orbital angular momentum L. One can see the mistake only affects states with odd L. Fortunately, the majority of past literature has focused on S-wave and D-wave systems, rendering this mistake inconsequential for them. Recent studies concerning P-wave systems have acknowledged this issue and adopted the correct notations [31, 42].

Appendix B: The OBE potentials and the G-parity rule

The effective potentials for the DD^* system in momentum space are listed as follows,

$$\begin{aligned} V_\sigma^D(\boldsymbol{p}', \boldsymbol{p}) &= -\frac{g_s^2}{\boldsymbol{q}^2 + m_\sigma^2}, \\ V_\pi^C(\boldsymbol{p}', \boldsymbol{p}) &= -\frac{g^2}{2f_\pi^2}\frac{(\boldsymbol{\epsilon}\cdot\boldsymbol{k})(\boldsymbol{\epsilon}'\cdot\boldsymbol{k})}{\boldsymbol{k}^2 - k_0^2 + m_\pi^2}\tau\cdot\tau, \\ V_\eta^C(\boldsymbol{p}', \boldsymbol{p}) &= -\frac{g^2}{6f_\pi^2}\frac{(\boldsymbol{\epsilon}\cdot\boldsymbol{k})(\boldsymbol{\epsilon}'\cdot\boldsymbol{k})}{\boldsymbol{k}^2 - k_0^2 + m_\eta^2}\mathbb{1}\cdot\mathbb{1}, \\ V_{\rho/\omega}^D(\boldsymbol{p}', \boldsymbol{p}) &= \frac{\frac{1}{4}\beta^2 g_V^2(\boldsymbol{\epsilon}\cdot\boldsymbol{\epsilon}')}{\boldsymbol{q}^2 + m_{\rho/\omega}^2} \times \begin{cases} \tau\cdot\tau, & \text{for } \rho, \\ \mathbb{1}\cdot\mathbb{1}, & \text{for } \omega, \end{cases} \\ V_{\rho/\omega}^C(\boldsymbol{p}', \boldsymbol{p}) &= \frac{\lambda^2 g_V^2}{\boldsymbol{k}^2 - k_0^2 + m_{\rho/\omega}^2}\{(\boldsymbol{k}\cdot\boldsymbol{\epsilon})(\boldsymbol{k}\cdot\boldsymbol{\epsilon}') \\ &\quad -\boldsymbol{k}^2(\boldsymbol{\epsilon}\cdot\boldsymbol{\epsilon}')\} \times \begin{cases} \tau\cdot\tau, & \text{for } \rho, \\ \mathbb{1}\cdot\mathbb{1}, & \text{for } \omega, \end{cases} \end{aligned} \quad \text{(B1)}$$

where D and C denotes the direct and cross diagrams, respectively. The isospin factors are

$$\tau\cdot\tau = \begin{cases} 1, & I=1,\ \mathrm{D}, \\ -3, & I=0,\ \mathrm{D}, \\ 1, & I=1,\ \mathrm{C}, \\ 3, & I=0,\ \mathrm{C}, \end{cases} \qquad \mathbb{1}\cdot\mathbb{1} = \begin{cases} 1, & I=1,\ \mathrm{D}, \\ 1, & I=0,\ \mathrm{D}, \\ 1, & I=1,\ \mathrm{C}, \\ -1, & I=0,\ \mathrm{C}. \end{cases} \quad \text{(B2)}$$

The results of the partial-wave expansion potential $V = (\boldsymbol{\epsilon}\cdot\boldsymbol{k})(\boldsymbol{\epsilon}'\cdot\boldsymbol{k})D(p', p, z)$ with $z = \boldsymbol{p}'\cdot\boldsymbol{p}/pp'$, are listed as follows,

$$\begin{aligned} V_S^{J=0} &= \frac{2\pi}{3}\int_{-1}^{1} D(p', p, z)(p^2 + p'^2 + 2pp'z)dz, \\ V_P^{J=0} &= 2\pi\int_{-1}^{1} D(p', p, z)\{(p^2 + p'^2)z + pp'(1+z^2)\}dz, \\ V_P^{J=1} &= 2\pi\int_{-1}^{1} D(p', p, z)\frac{1}{2}(z^2-1)pp'dz, \\ V_P^{J=2} &= \frac{2\pi}{5}\int_{-1}^{1} D(p', p, z)\{2(p^2 + p'^2)z \\ &\quad + \frac{1}{2}pp'(1+7z^2)\}dz. \end{aligned} \quad \text{(B3)}$$

In Fig. 6, we list the the OBE potentials of DD^* channels with various meson exchanges.

The effective potential for the $D\bar{D}^*/\bar{D}D^*$ system can be related to the DD^* potential up to a factor G_m ($-G_mG_{MM}$) for the direct (cross) diagrams, with G_m and G_{MM} as the G-parities of the exchanged meson and the $D\bar{D}^*/\bar{D}D^*$ system, respectively. In Fig. 7, the total potentials for all DD^* and $D\bar{D}^*/\bar{D}D^*$ up to P-wave with various quantum numbers are illustrated. One can read out the $D\bar{D}^*/\bar{D}D^*$ potential of the specific meson exchange from Fig. 6 via G-parity rule. It is noteworthy that the G-parity rule for cross diagrams is different from that in direct diagrams. We will present the derivation of the G-parity rule as follows.

With the C-parity convention in Eq. (A6) (the final result is irrelevant to the convention), the G-parity transformation reads

$$D = (-D^+, D^0) \xrightarrow{G} \bar{D} = (\bar{D}^0, D^-) \xrightarrow{G} -D,$$

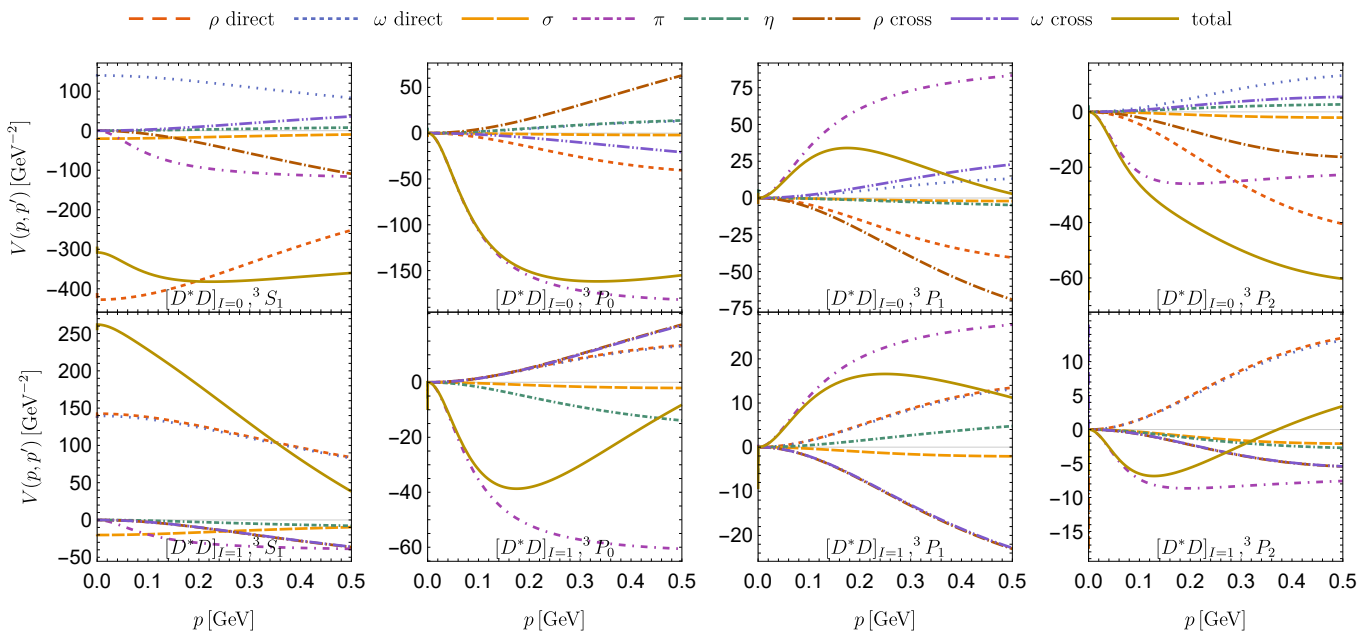


FIG. 6. The OBE potentials of DD^* channels are illustrated with various meson exchanges. Only $p = p'$ cases are depicted.

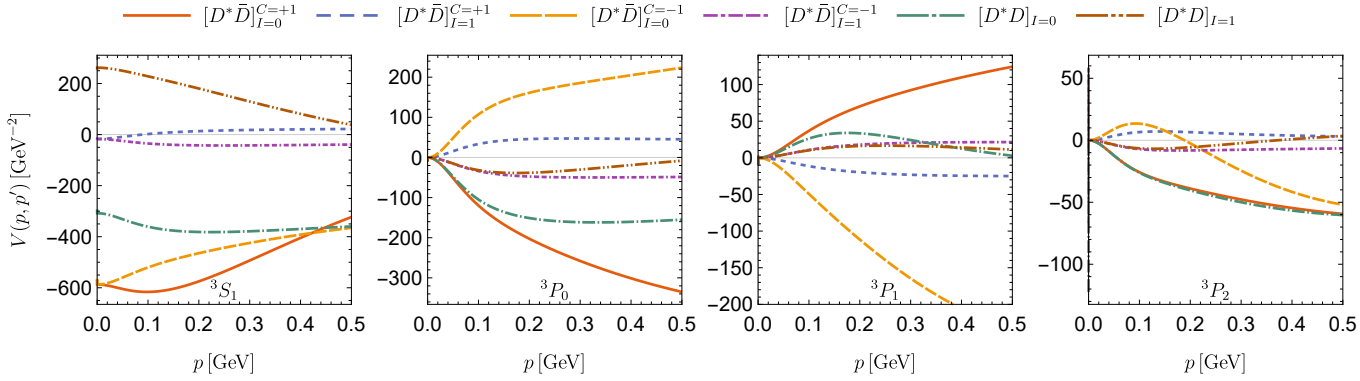


FIG. 7. The OBE potentials for all DD^* and $D\bar{D}^*/\bar{D}D^*$ systems up to P-wave with various quantum numbers. Only $p = p'$ cases are shown.

$$D^* = (-D^{*+}, D^{*0}) \xrightarrow{G} -\bar{D}^* = -(\bar{D}^{*0}, D^{*-}) \xrightarrow{G} -D^*, \tag{B4}$$

where the charmed mesons are written in the form of isospin doublets. The G-parity eigenstates can be constructed,

$$|D\bar{D}^*/\bar{D}D^*, G = \pm\rangle = \frac{1}{\sqrt{2}}(|D\bar{D}^*\rangle \pm |\bar{D}D^*\rangle). \tag{B5}$$

Since the exchanged mesons are eigenstates of the G-parity, we can apply the G-parity transformation to one of the vertex in $DD^* \to DD^*$, as shown in Fig. 8. The overall factor G_m arises from the G-parity of the exchanged meson. Then we derive

$$\begin{aligned} V^C_{\bar{D}D^*\to D\bar{D}^*} &= (-G_m)V^C_{DD^*\to DD^*}, \\ V^D_{\bar{D}D^*\to \bar{D}D^*} &= V^D_{DD^*\to DD^*}. \end{aligned} \tag{B6}$$

Combining Eq. (B5) and Eq. (B6), we derive

$$V_{D\bar{D}^*/\bar{D}D^*, G_{MM}} = G_m V^D_{DD^*\to DD^*} - G_m G_{MM} V^C_{DD^*\to DD^*}. \tag{B7}$$

Appendix C: Regulator dependence

We also estimate the uncertainty of our results arising from the different regulators. For example, we choose the following regulators for the direct and cross diagrams, respectively,

$$V^D(\boldsymbol{q}) \to V^D(\boldsymbol{q}) \left(\frac{\Lambda^2 - m^2}{\Lambda^2 + \boldsymbol{q}^2}\right)^2, \tag{C1}$$

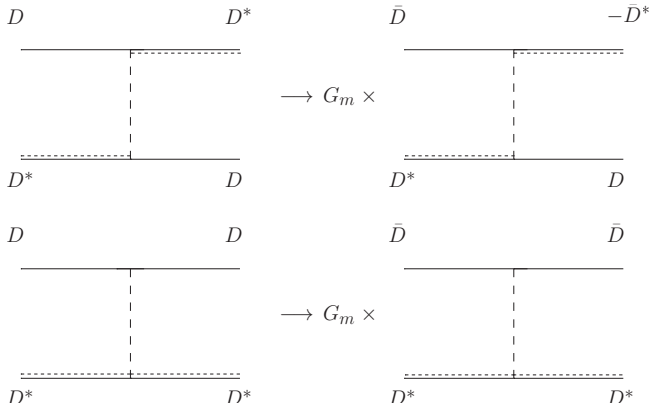


FIG. 8. The G-parity transformation for cross and direct diagrams, respectively. The signs are determined by the G-parity of the exchanged meson and Eq. (B4).

$$V^C(\boldsymbol{k}) \to V^C(\boldsymbol{k}) \left(\frac{\Lambda^2 - m^2}{\Lambda^2 + \boldsymbol{k}^2}\right)^2,$$

where the potentials from the direct and cross diagrams are the functions of $\boldsymbol{q}$ and $\boldsymbol{k}$, respectively. m is the mass of the transferred meson. The cutoff $\Lambda = 1.25$ and 1.35 GeV is adjusted to get the loosely bound state $\chi_{c1}(3872)$. The pole positions for all other channels are presented in Table II. One can see as long as the cutoff is set to generate a loosely bound $\chi_{c1}(3872)$ state, a corresponding P-wave resonance emerges in the 1^{--} channel as the $Y(3872)$, while the poles of T_{cc} and $Z_c(3900)$ remain qualitatively unchanged. Our predictions are robust under various regularization schemes.

[1] H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, The hidden-charm pentaquark and tetraquark states, Phys. Rept. **639**, 1 (2016), arXiv:1601.02092 [hep-ph].

[2] A. Esposito, A. Pilloni, and A. D. Polosa, Multiquark Resonances, Phys. Rept. **668**, 1 (2017), arXiv:1611.07920 [hep-ph].

[3] F.-K. Guo, C. Hanhart, U.-G. Meißner, Q. Wang, Q. Zhao, and B.-S. Zou, Hadronic molecules, Rev. Mod. Phys. **90**, 015004 (2018), arXiv:1705.00141 [hep-ph].

[4] Y.-R. Liu, H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, Pentaquark and Tetraquark states, Prog. Part. Nucl. Phys. **107**, 237 (2019), arXiv:1903.11976 [hep-ph].

[5] N. Brambilla, S. Eidelman, C. Hanhart, A. Nefediev, C.-P. Shen, C. E. Thomas, A. Vairo, and C.-Z. Yuan, The XYZ states: experimental and theoretical status and perspectives, Phys. Rept. **873**, 1 (2020), arXiv:1907.07583 [hep-ex].

[6] H.-X. Chen, W. Chen, X. Liu, Y.-R. Liu, and S.-L. Zhu, An updated review of the new hadron states, Rept. Prog. Phys. **86**, 026201 (2023), arXiv:2204.02649 [hep-ph].

[7] L. Meng, B. Wang, G.-J. Wang, and S.-L. Zhu, Chiral perturbation theory for heavy hadrons and chiral effective field theory for heavy hadronic molecules, Phys. Rept. **1019**, 1 (2023), arXiv:2204.08716 [hep-ph].

[8] S. K. Choi *et al.* (Belle), Observation of a narrow charmonium-like state in exclusive $B^{\pm} \to K^{\pm}\pi^+\pi^- J/\psi$ decays, Phys. Rev. Lett. **91**, 262001 (2003), arXiv:hep-ex/0309032.

[9] M. Ablikim *et al.* (BESIII), Observation of a Charged Charmoniumlike Structure in $e^+e^- \to \pi^+\pi^- J/\psi$ at $\sqrt{s}$ =4.26 GeV, Phys. Rev. Lett. **110**, 252001 (2013), arXiv:1303.5949 [hep-ex].

[10] Z. Q. Liu *et al.* (Belle), Study of $e^+e^- \to \pi^+\pi^- J/\psi$ and Observation of a Charged Charmoniumlike State at Belle, Phys. Rev. Lett. **110**, 252002 (2013), [Erratum: Phys.Rev.Lett. 111, 019901 (2013)], arXiv:1304.0121 [hep-ex].

[11] R. Aaij *et al.* (LHCb), Observation of an exotic narrow doubly charmed tetraquark, Nature Phys. **18**, 751 (2022), arXiv:2109.01038 [hep-ex].

[12] R. Aaij *et al.* (LHCb), Study of the doubly charmed tetraquark T_{cc}^+, Nature Commun. **13**, 3351 (2022), arXiv:2109.01056 [hep-ex].

[13] N. Li and S.-L. Zhu, Isospin breaking, Coupled-channel effects and Diagnosis of X(3872), Phys. Rev. D **86**, 074022 (2012), arXiv:1207.3954 [hep-ph].

[14] N. Li, Z.-F. Sun, X. Liu, and S.-L. Zhu, Coupled-channel analysis of the possible $D^{(*)}D^{(*)}, \overline{B}^{(*)}\overline{B}^{(*)}$ and $D^{(*)}\overline{B}^{(*)}$ molecular states, Phys. Rev. D **88**, 114008 (2013), arXiv:1211.5007 [hep-ph].

[15] C. A. Bertulani, H. W. Hammer, and U. Van Kolck, Effective field theory for halo nuclei, Nucl. Phys. A **712**, 37 (2002), arXiv:nucl-th/0205063.

[16] C. Chin, R. Grimm, P. Julienne, and E. Tiesinga, Feshbach resonances in ultracold gases, Reviews of Modern Physics **82**, 1225 (2010).

[17] M. Ablikim *et al.* (BESIII), Precise Measurement of Born Cross Sections for $e^+e^- \to D\bar{D}$ and Observation of One Structure between $\sqrt{s} = 3.80 - 4.95$ GeV, (2024), arXiv:2402.03829 [hep-ex].

[18] T. V. Uglov, Y. S. Kalashnikova, A. V. Nefediev, G. V. Pakhlova, and P. N. Pakhlov, Exclusive open-charm near-threshold cross sections in a coupled-channel approach, JETP Lett. **105**, 1 (2017), arXiv:1611.07582 [hep-ph].

[19] S. X. Nakamura, X. H. Li, H. P. Peng, Z. T. Sun, and X. R. Zhou, Global coupled-channel analysis of $e^+e^- \to c\bar{c}$ processes in $\sqrt{s} = 3.75 - 4.7$ GeV, (2023), arXiv:2312.17658 [hep-ph].

[20] Z.-Y. Lin, J.-B. Cheng, B.-L. Huang, and S.-L. Zhu, Partial widths from analytical extension of the wave function: Pc states, Phys. Rev. D **108**, 114014 (2023), arXiv:2305.19073 [hep-ph].

[21] R. Machleidt, K. Holinde, and C. Elster, The Bonn Meson Exchange Model for the Nucleon Nucleon Interaction, Phys. Rept. **149**, 1 (1987).

[22] N. A. Tornqvist, Possible large deuteron - like meson meson states bound by pions, Phys. Rev. Lett. **67**, 556 (1991).

[23] N. A. Tornqvist, From the deuteron to deusons, an analysis of deuteron - like meson meson bound states, Z. Phys. C **61**, 525 (1994), arXiv:hep-ph/9310247.

TABLE II. The poles in all channels of $D\bar{D}^*$ and DD^*, up to the orbital angular momentum $L = 1$ with the regularization in Eq. (C2). The B and V superscripts denote the bound state and the virtual state, respectively. Otherwise the pole refers to a resonance.

		$D\bar{D}^*, C = +$		$D\bar{D}^*, C = -$		DD^*	
		$I = 0$	$I = 1$	$I = 0$	$I = 1$	$I = 0$	$I = 1$
$\Lambda = 1.25$ GeV	$1^+(^3S_1)$	-0.40^B, $\chi_{c1}(3872)$	-	-25.0^V	-39.6^V, $Z_c(3900)$	-0.79^B, $T_{cc}(3875)$	-
	$0^-(^3P_0)$	$3.3 - 17.2i$	-	-	-	$-11.2 - 16.7i$	-
	$1^-(^3P_1)$	-	-	$4.4 - 39.9i$, $Y(3872)$	-	$-96.6 - 87.3i$	-
	$2^-(^3P_2)$	$-71.2 - 63.5i$	-	$-31.0 - 96.5i$	-	$-61.3 - 53.6i$	-
$\Lambda = 1.35$ GeV	$1^+(^3S_1)$	-2.8^B, $\chi_{c1}(3872)$	-	-2.2^V	-38.5^V, $Z_c(3900)$	-8.8^B, $T_{cc}(3875)$	-
	$0^-(^3P_0)$	$6.6 - 11.6i$	-	-	-	$-10.2 - 18.0i$	-
	$1^-(^3P_1)$	-	-	$10.2 - 33.7i$, $Y(3872)$	-	$-92.9 - 97.7i$	-
	$2^-(^3P_2)$	$-68.0 - 75.4i$	-	$-23.3 - 97.2i$	-	$-58.4 - 59.6$i	-

[24] X. Liu, Z.-G. Luo, Y.-R. Liu, and S.-L. Zhu, X(3872) and Other Possible Heavy Molecular States, Eur. Phys. J. C **61**, 411 (2009), arXiv:0808.0073 [hep-ph].

[25] Y.-R. Liu, X. Liu, W.-Z. Deng, and S.-L. Zhu, Is $X(3872)$ Really a Molecular State?, Eur. Phys. J. C **56**, 63 (2008), arXiv:0801.3540 [hep-ph].

[26] X. Liu, Y.-R. Liu, W.-Z. Deng, and S.-L. Zhu, Z+(4430) as a D(1)-prime D* (D(1) D*) molecular state, Phys. Rev. D **77**, 094015 (2008), arXiv:0803.1295 [hep-ph].

[27] G.-J. Ding, J.-F. Liu, and M.-L. Yan, Dynamics of Hadronic Molecule in One-Boson Exchange Approach and Possible Heavy Flavor Molecules, Phys. Rev. D **79**, 054005 (2009), arXiv:0901.0426 [hep-ph].

[28] Z.-F. Sun, J. He, X. Liu, Z.-G. Luo, and S.-L. Zhu, $Z_b(10610)^\pm$ and $Z_b(10650)^\pm$ as the $B^*\bar{B}$ and $B^*\bar{B}^*$ molecular states, Phys. Rev. D **84**, 054002 (2011), arXiv:1106.2968 [hep-ph].

[29] C. E. Thomas and F. E. Close, Is X(3872) a molecule?, Phys. Rev. D **78**, 034007 (2008), arXiv:0805.3653 [hep-ph].

[30] I. W. Lee, A. Faessler, T. Gutsche, and V. E. Lyubovitskij, X(3872) as a molecular DD* state in a potential model, Phys. Rev. D **80**, 094005 (2009), arXiv:0910.1009 [hep-ph].

[31] Z.-P. Wang, F.-L. Wang, G.-J. Wang, and X. Liu, Exploring Charmonium-like Molecular Resonances from Deeply Bound $D\bar{D}_1$, $D^*\bar{D}_1$, and $D^*\bar{D}_2^*$ Molecules, (2023), arXiv:2312.03512 [hep-ph].

[32] T. Ji, X.-K. Dong, F.-K. Guo, and B.-S. Zou, Prediction of a Narrow Exotic Hadronic State with Quantum Numbers JPC=0−, Phys. Rev. Lett. **129**, 102002 (2022), arXiv:2205.10994 [hep-ph].

[33] J. He, Y.-T. Ye, Z.-F. Sun, and X. Liu, The observed charmed hadron $\Lambda_c(2940)^+$ and the D^*N interaction, Phys. Rev. D **82**, 114029 (2010), arXiv:1008.1500 [hep-ph].

[34] J.-B. Cheng, B.-L. Huang, Z.-Y. Lin, and S.-L. Zhu, Z_{cs}, Z_c and Z_b states under the complex scaling method, Eur. Phys. J. C **83**, 1071 (2023), arXiv:2305.15787 [hep-ph].

[35] J. R. Taylor, *Scattering Theory: The Quantum Theory of Nonrelativistic Collisions* (John Wiley & Sons, Inc., New York, 1972).

[36] E. Epelbaum, H.-W. Hammer, and U.-G. Meissner, Modern Theory of Nuclear Forces, Rev. Mod. Phys. **81**, 1773 (2009), arXiv:0811.1338 [nucl-th].

[37] V. Baru, E. Epelbaum, A. A. Filin, F. K. Guo, H. W. Hammer, C. Hanhart, U. G. Meißner, and A. V. Nefediev, Remarks on study of X(3872) from effective field theory with pion-exchange interaction, Phys. Rev. D **91**, 034002 (2015), arXiv:1501.02924 [hep-ph].

[38] M.-L. Du, V. Baru, X.-K. Dong, A. Filin, F.-K. Guo, C. Hanhart, A. Nefediev, J. Nieves, and Q. Wang, Coupled-channel approach to Tcc+ including three-body effects, Phys. Rev. D **105**, 014024 (2022), arXiv:2110.13765 [hep-ph].

[39] B. Wang and L. Meng, Revisiting the DD* chiral interactions with the local momentum-space regularization up to the third order and the nature of Tcc+, Phys. Rev. D **107**, 094002 (2023), arXiv:2212.08447 [hep-ph].

[40] L. Meng, B. Wang, G.-J. Wang, and S.-L. Zhu, The hidden charm pentaquark states and $\Sigma_c\bar{D}^{(*)}$ interaction in chiral perturbation theory, Phys. Rev. D **100**, 014031 (2019), arXiv:1905.04113 [hep-ph].

[41] Z.-Y. Lin, J.-B. Cheng, and S.-L. Zhu, T_{cc}^+ and $X(3872)$ with the complex scaling method and $DD(\bar{D})\pi$ three-body effect, (2022), arXiv:2205.14628 [hep-ph].

[42] L. Meng, V. Baru, E. Epelbaum, A. A. Filin, and A. M. Gasparyan, Solving the left-hand cut problem in lattice QCD: $T_{cc}(3875)^+$ from finite volume energy levels, (2023), arXiv:2312.01930 [hep-lat].

[43] R. L. Workman *et al.* (Particle Data Group), Review of Particle Physics, PTEP **2022**, 083C01 (2022).

[44] Z.-Y. Lin, J.-Z. Wang, J.-B. Cheng, L. Meng, and S.-L. Zhu, See Supplemental Material for additional details on sign problem of the potential in cross diagrams, the OBE potentials, the G-parity rule and the uncertainties from various regulators, (2024).

[45] Y.-K. Chen, L. Meng, Z.-Y. Lin, and S.-L. Zhu, Vir-

tual states in the coupled-channel problems with an improved complex scaling method, Phys. Rev. D **109**, 034006 (2024), arXiv:2308.12424 [hep-ph].

[46] J.-B. Cheng, Z.-Y. Lin, and S.-L. Zhu, Double-charm tetraquark under the complex scaling method, Phys. Rev. D **106**, 016012 (2022), arXiv:2205.13354 [hep-ph].

[47] R. Aaij *et al.* (LHCb), A model-independent study of resonant structure in $B^+ \to D^+D^-K^+$ decays, Phys. Rev. Lett. **125**, 242001 (2020), arXiv:2009.00025 [hep-ex].

[48] R. Aaij *et al.* (LHCb), Amplitude analysis of the $B^+ \to D^+D^-K^+$ decay, Phys. Rev. D **102**, 112003 (2020), arXiv:2009.00026 [hep-ex].

March 5, 2024

Extranatural Warped Inflation

Toshiki Kawai[a,1] and Yoshiharu Kawamura[b,2]

[a] *Department of Physics, Hokkaido University, Sapporo 060-0810, Japan*
[b] *Department of Physics, Shinshu University, Matsumoto 390-8621, Japan*

arXiv:2403.01705v1 [hep-ph] 4 Mar 2024

Abstract

We investigate whether inflation in the early universe can be induced by an extra component of a five-dimensional U(1) gauge field in the Randall-Sundrum warped spacetime or not. We show that an effective potential obtained by quantum corrections can act as an inflaton potential by finding parameter regions consistent with Planck 2018 results. In our model, fields involved in inflation obey a different type of boundary condition from that of the visible particles, and a surviving Wilson line phase can play a specific role for physics beyond the standard model.

[1] E-mail: t-kawai@higgs3.sci.hokudai.ac.jp
[2] E-mail: haru@azusa.shinshu-u.ac.jp

1 Introduction

Inflation can solve both the flatness problem and the horizon problem at the same time and also explain temperature fluctuations of the cosmic microwave background (CMB) [1]. It is therefore almost certain that inflation occurred in the early universe. However, a specific mechanism or an origin of a quanta called "inflaton" is still a mystery.

An inflaton is assumed to be a scalar field in typical models, but its potential is, in general, not stable against quantum (gravity) corrections and should be fine-tuned to satisfy constraints of cosmological observations. This problem stems from the fact that radiative corrections on scalar fields cannot be controlled without any assistance of a powerful symmetry.

One way to solve this problem is to consider a higher-dimensional gauge theory, and then a Wilson line phase made of an extra component of a gauge field can become an inflaton whose potential can be determined by the gauge principle. In this scenario, a relevant part of potential is obtained as a function of the Wilson line phase radiatively, taking a finite value [2, 3]. Such an inflaton potential is robust owing to the gauge symmetry and its stability can be guaranteed. A famous five-dimensional (5d) model compactified on a circle S^1 has been proposed in the name of extranatural inflation [4]. In a similar setup, a model with both a radion and a Wilson line phase has been constructed, and it is shown that the radion is stabilized [5] and the Wilson line phase can be, in most cases, an inflaton [6].

Recently, constraints on cosmological parameters from various observations have become more stringent and inflation models can be selected more rigorously [7]. It is valuable to re-examine whether an extranatural inflation scenario is still valid today or not, in a wide variety of spacetime. It is known that the Randall-Sundrum (RS) warped spacetime can provide rich properties in phenomenological aspects of particle physics. For instance, RS1 model provides a solution of the hierarchy problem of the standard model [8], RS2 model explains the weakness of gravitational force on the TeV brane because the graviton localizes on the Planck brane [9], and gauge-Higgs unification (GHU) models in the warped spacetime can show promise as physics beyond the standard model (BSM) [10, 11, 12, 13]. Hence, it would be favorable if a riddle in cosmology such as inflation could be solved in the same background.

In this paper, we investigate whether inflation in the early universe can be induced by an extra component of a 5d U(1) gauge field in the warped background or not. In concrete, we examine whether an effective potential obtained by quantum corrections can act as an inflaton potential by finding parameter regions consistent with Planck 2018 results or not. The RS spacetime contains an orbifold $S^1/\mathbb{Z}_2$ and a different type of boundary condition called "conjugate boundary condition" can be imposed on fields to survive an extra component of a gauge boson in QED-like models and this enables fields involved in inflation to play a specific role for BSM.

This paper is organized as follows. In the next section, we present a 5d U(1) gauge theory on the warped spacetime. In section 3, we study an extranatural inflation scenario on the warped spacetime based on our model. The last section is devoted to conclusions and discussions.

2 U(1) gauge theory on a warped background

2.1 Randall-Sundrum metric and action integral

The spacetime is assumed to be 5d one with the RS metric given by [8, 9]

$$ds^2 = g_{MN}dx^M dx^N = e^{-2\sigma(y)}\eta_{\mu\nu}dx^\mu dx^\nu + dy^2, \tag{1}$$

where g_{MN} $(M, N = 0, 1, 2, 3, 5)$ is the 5d metric, $\eta_{\mu\nu}(= \mathrm{diag}\,(-1, 1, 1, 1))$ $(\mu, \nu = 0, 1, 2, 3)$ is the four-dimensional (4d) metric, x^M and x^μ are 5d and 4d coordinates, respectively, x^5 is also denoted as y, and $\sigma(y) = k|y|$ (for $-\pi R \le y \le \pi R$,$\sigma(y+2\pi R) = \sigma(y)$) with a constant k which is related to the cosmological constant Λ_5 in the bulk such as $\Lambda_5 = -6k^2$. The extra space $S^1/\mathbb{Z}_2$ is expanded in the range of $0 \le y \le \pi R$, and then $\sigma(y) = ky$ there.

We study a 5d U(1) gauge theory whose action integral is given by

$$S = \int d^4x \int_0^{\pi R} dy \left\{ \sqrt{-g} \left(-\frac{1}{4} g^{MP} g^{NS} C_{MN} C_{PS} + \mathscr{L}_{\mathrm{gf}} \right) + \det(e^A{}_M) \overline{\psi} \left(\Gamma^A e_A{}^M \mathscr{D}_M - M_\psi - c\sigma'(y) \right) \psi \right\}, \tag{2}$$

where $\sqrt{-g} = \sqrt{-\det g_{MN}}$, g^{MN} is the inverse of 5d metric, $C_{MN} = \partial_M C_N - \partial_N C_M$ is the field strength made of the 5d U(1) gauge boson C_M, $\mathscr{L}_{\mathrm{gf}}$ is a gauge-fixing term, $e^A{}_M$ and $e_A{}^M$ $(A = 0, 1, 2, 3, 5)$ are a fünfbein and its inverse, A is an index of 5d local Lorentz coordinates, Γ^A are the gamma matrices satisfying the relation $\{\Gamma^A, \Gamma^B\} = 2\eta^{AB}(= 2\,\mathrm{diag}\,(-1, 1, 1, 1, 1))$, ψ is a 5d fermion with a mass $M_\psi + c\sigma'(y)$ (M_ψ is a y-independent bulk mass, c is a y-dependent bulk mass coefficient, and $\sigma'(y) = d\sigma(y)/dy$)[3], and $\mathscr{D}_M$ is a covariant derivative defined by

$$\mathscr{D}_M = \partial_M + \frac{1}{8}\omega_{AB,M}[\Gamma^A, \Gamma^B] - ig_5 q_\psi C_M, \tag{3}$$

using the spin connection $\omega_{AB,M}$ and the gauge boson C_M. In (3), g_5 is a 5d gauge coupling, q_ψ is a U(1) charge of ψ, and each component of $\mathscr{D}_M$ is written as

$$\mathscr{D}_\mu = \partial_\mu - \frac{1}{2}\sigma'(y) e_{a\mu}\gamma^a\gamma^5 - ig_5 q_\psi C_\mu, \quad \mathscr{D}_y = \partial_y - ig_5 q_\psi C_y, \tag{4}$$

where $e_{a\mu} = \eta_{ab}e^b{}_\mu$ ($\eta_{ab}(= \mathrm{diag}\,(-1, 1, 1, 1))$ and $e^b{}_\mu$ are the 4d metric and a vierbein in the 4d local Lorentz coordinates, respectively), γ^a $(a = 0, 1, 2, 3)$ are the gamma matrices satisfying the relation $\{\gamma^a, \gamma^b\} = 2\eta^{ab}(= 2\,\mathrm{diag}\,(-1, 1, 1, 1))$, a and b are indexes of 4d local Lorentz coordinates, $\gamma^5 (\equiv -i\gamma^0\gamma^1\gamma^2\gamma^3)$ is 4d chirality operator, and $C_y = C_5(x, y)$. Here and hereafter x^μ on an argument of fields is denoted as x for short.

[3] We introduce both M_ψ and $c\sigma'(y)$ in a general standpoint, and we will see that $c\sigma'(y)$ is forbidden by imposing specific boundary conditions on fields in the next subsection.

2.2 Conjugate boundary conditions

We impose the so-called conjugate boundary conditions (CBCs) on 5d fields, in order to obtain a zero mode of C_y after $S^1/\mathbb{Z}_2$ compactification [14, 15, 16, 17].

The CBCs of $C_M(x,y)$ and $\psi(x,y)$ in the Dirac representation are given by

$$C_M(x,y+2\pi R) = C_M(x,y), \quad C_\mu(x,-y) = -C_\mu(x,y), \quad C_y(x,-y) = C_y(x,y), \tag{5}$$

$$\psi(x,y+2\pi R) = e^{i\beta}\psi(x,y), \quad \psi(x,-y) = \psi^{\rm C}(x,y) \equiv e^{i\theta_{\rm C}}\Gamma^2\psi^*(x,y), \tag{6}$$

where β is a constant called a twisted phase, the superscript C denotes a 4d charge conjugation, $\theta_{\rm C}$ is a real number, and the asterisk means the complex conjugation. Then, the covariant derivatives obey the relations:

$$\left(\partial_M + \frac{1}{8}\omega_{AB,M}(x,y+2\pi R)[\Gamma^A,\Gamma^B] - ig_5 q_\psi C_M(x,y+2\pi R)\right)\psi(x,y+2\pi R)$$
$$= \left(\partial_M + \frac{1}{8}\omega_{AB,M}(x,y)[\Gamma^A,\Gamma^B] - ig_5 q_\psi C_M(x,y)\right)\psi(x,y), \tag{7}$$

$$\left(\partial_\mu - \frac{1}{2}\sigma'(-y)e_{a\mu}\gamma^a\gamma^5 - ig_5 q_\psi C_\mu(x,-y)\right)\psi(x,-y)$$
$$= \left(\partial_\mu - \frac{1}{2}\sigma'(y)e_{a\mu}\gamma^a\gamma^5 - ig_5 q_\psi C_\mu(x,y)\right)^* \psi^{\rm C}(x,y), \tag{8}$$

$$\left(\partial_{-y} - ig_5 q_\psi C_y(x,-y)\right)\psi(x,-y) = -\left(\partial_y - ig_5 q_\psi C_y(x,y)\right)^* \psi^{\rm C}(x,y). \tag{9}$$

Based on the above CBCs and suitable ones for g_{MN} and $e^A{}_M$, we find that the requirement is fulfilled for $c = 0$ that the action integral S should be invariant under the translation $y \to y + 2\pi R$ and the $\mathbb{Z}_2$ transformation $y \to -y$. U(1) gauge theories with the CBCs own several interesting features. (1) The 4d-components of U(1) gauge boson C_μ have no zero mode, and the U(1) gauge symmetry is explicitly broken by the BCs. (2) The extra component of gauge boson C_y can have a zero mode and become a Wilson line phase. (3) 5d fermions can have constant bulk masses. (4) 5d fermions can produce Kaluza-Klein (KK) towers of 4d Majorana fermions.

For later convenience, we change a variable from y to $z = e^{ky}$. Then, the range of $S^1/\mathbb{Z}_2$ changes from $0 \le y \le \pi R$ to $1 \le z \le z_L (= e^{k\pi R})$, and (5) and (6) are rewritten by

$$C_M(x,z_L^2 z) = C_M(x,z), \quad C_\mu(x,1/z) = -C_\mu(x,z), \quad C_y(x,1/z) = C_y(x,z), \tag{10}$$

$$\psi(x,z_L^2 z) = e^{i\beta}\psi(x,z), \quad \psi(x,1/z) = \psi^{\rm C}(x,z) \equiv e^{i\theta_{\rm C}}\Gamma^2\psi^*(x,z), \tag{11}$$

respectively.

2.3 Mass spectrum

We examine a zero mode of C_y and a mass spectrum of fermion ψ. Setting $z = e^{ky}$, the metric (1) is rewritten as

$$ds^2 = \frac{1}{z^2}\left(\eta_{\mu\nu}dx^\mu dx^\nu + \frac{dz^2}{k^2}\right). \tag{12}$$

Then, the action integral is rewritten as

$$\begin{aligned} S = \int d^4x \int_1^{z_L} dz \left\{ \frac{1}{z}\left(-\frac{1}{4k}C_{\mu\nu}C^{\mu\nu} - \frac{k}{2}C_{\mu z}C^{\mu z} + \check{\mathscr{L}}_{\rm gf} \right) \right. \\ \left. + \frac{1}{k}\overline{\check{\psi}}\left(\gamma^\mu D_\mu + k\gamma^5 D_z - M_\psi\right)\check{\psi} \right\}, \end{aligned} \tag{13}$$

where $\check{\mathscr{L}}_{\rm gf} = \mathscr{L}_{\rm gf}/(kz^4)$, $\check{\psi} = \psi/z^2$, $C_{\mu z} = \partial_\mu C_z - \partial_z C_\mu$, $C_z = C_y/(ke^{ky})$, $D_\mu = \partial_\mu - ig_5 q_\psi C_\mu$, $D_z = \partial_z - ig_5 q_\psi C_z$, and $c = 0$ is taken by the CBCs.

The Lagrangian density in (13) contains a mixing term of C_μ and C_z such as

$$\frac{k}{z}(\partial_\mu C_z)(\partial^z C^\mu) = \left\{\partial_z\left(\frac{k}{z}C^z\right)\right\}(\partial_\mu C^\mu) + \text{total derivatives} \tag{14}$$

and the following gauge-fixing term is introduced, in order to cancel the above mixing term [18],

$$\check{\mathscr{L}}_{\rm gf} = \frac{1}{2\xi k}\left[\partial_\mu C^\mu - z\xi k\left\{\partial_z\left(\frac{k}{z}C^z\right)\right\}\right]^2, \tag{15}$$

where ξ is a gauge parameter. In this gauge-fixing, the equation of motion of C_z is given by

$$\Box_4 C_z - \xi k^2 \partial_z\left\{z\partial_z\left(\frac{1}{z}C_z\right)\right\} = 0, \tag{16}$$

where $\Box_4 = \eta^{\mu\nu}\partial_\mu\partial_\nu$. We now perform the KK mode expansion for C_z as

$$C_z(x,z) = \frac{1}{\sqrt{k}}\sum_{n=-\infty}^{\infty} h^{(n)}(z)C_z^{(n)}(x), \tag{17}$$

where $h^{(n)}(z)$ are a set of orthogonal functions satisfying $\int_1^{z_L}\frac{dz}{z}h^{(n)}(z)h^{(m)}(z) = \delta_{nm}$ and $C_z^{(n)}(x)$ are 4d fields satisfying the equation $\Box_4 C_z^{(n)}(x) = k^2\lambda_n^2 C_z^{(n)}(x)$ with constants λ_n. Inserting (17) into eq. (16) with $\xi = -1$, we obtain the eigenvalue equation on $h^{(n)}(z)$:

$$-\frac{d}{dz}\left\{z\frac{d}{dz}\left(\frac{1}{z}h^{(n)}(z)\right)\right\} = \lambda_n^2 h^{(n)}(z). \tag{18}$$

The function $h^{(0)}(z)$ corresponding to a zero mode obeys eq. (18) with $\lambda_0 = 0$, and it is given

by

$$h^{(0)}(z) = az + bz \ln z, \tag{19}$$

where a and b are some constants.

To survive $h^{(0)}(z)$, the Neumann boundary condition at both boundaries ($z = 1, z_L$) is imposed, i.e., $\partial(C_z/z)/\partial z = 0$ or $d(h^{(0)}/z)/dz = 0$, which comes from the requirement that surface terms appearing in the derivation of the action should vanish. The condition leads to $b = 0$, and the following normalized zero mode is obtained:

$$h^{(0)}(z) = \sqrt{\frac{2}{z_L^2 - 1}}\, z \quad \text{for } 1 \le z \le z_L. \tag{20}$$

Next, let us move on to the calculation of KK masses of the fermion $\check{\psi}(x, z)$. In the chiral representation, the equation of motion of $\check{\psi}(x, z) = (\xi_{\rm R}(x, z), \zeta_{\rm L}(x, z))^{\rm T}$ is given by

$$\begin{pmatrix} kD_z - M_\psi & \sigma^\mu D_\mu \\ \overline{\sigma}^\mu D_\mu & -kD_z - M_\psi \end{pmatrix} \begin{pmatrix} \xi_{\rm R}(x, z) \\ \zeta_{\rm L}(x, z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{21}$$

where $\sigma^\mu = (I, \boldsymbol{\sigma})$, $\overline{\sigma}^\mu = (-I, \boldsymbol{\sigma})$ (I is the 2×2 unit matrix and $\boldsymbol{\sigma}(= (\sigma^1, \sigma^2, \sigma^3))$ are Pauli matrices). We notice that the 4d U(1) gauge boson is absent in D_μ, because C_μ have no zero modes due to the CBCs (10). We now perform the KK mode expansion for $\xi_{\rm R}(x, z)$ and $\zeta_{\rm L}(x, z)$ as

$$\xi_{\rm R}(x, z) = \sum_{n=-\infty}^{\infty} f_{\rm R}^{(n)}(z) u_{\rm R}^{(n)}(x), \quad \zeta_{\rm L}(x, z) = \sum_{n=-\infty}^{\infty} f_{\rm L}^{(n)}(z) u_{\rm L}^{(n)}(x), \tag{22}$$

where $u_{\rm R}^{(n)}(x)$ and $u_{\rm L}^{(n)}(x)$ are 4d spinor fields satisfying the equations $\overline{\sigma}^\mu D_\mu u_{\rm R}^{(n)}(x) = k\check{\lambda}_n u_{\rm L}^{(n)}(x)$ and $\sigma^\mu D_\mu u_{\rm L}^{(n)}(x) = k\check{\lambda}_n u_{\rm R}^{(n)}(x)$ with constants $\check{\lambda}_n$. Inserting (22) into eq. (21), we obtain a same eigenvalue equation on $f_{\rm R}^{(n)}(z)$ and $f_{\rm L}^{(n)}(z)$:

$$-\left(D_z^2 - \frac{M_\psi^2}{k^2}\right) f^{(n)}(z) = \check{\lambda}_n^2 f^{(n)}(z), \tag{23}$$

where $f^{(n)}(z)$ represents both $f_{\rm R}^{(n)}(z)$ and $f_{\rm L}^{(n)}(z)$.

Now, we find that $f_{\rm R}^{(n)}(z)$ and $f_{\rm L}^{(n)}(z)$ are expressed by exponent functions, because eq. (23) has a same form as in the case of a flat spacetime. Using (11), $f^{(n)}(z)$ are determined as

$$f^{(n)}(z) = \frac{1}{\sqrt{z_L - 1}} \exp\left\{ i(2\pi n + \beta) \frac{\ln z}{\ln z_L^2} \right\} \tag{24}$$

with the relations $f^{(n)}(z_L^2 z) = e^{i\beta} f^{(n)}(z)$ and $f^{(n)}(1/z) = f^{(n)*}(z)$, and $u_{\rm R}^{(n)}(x)$ and $u_{\rm L}^{(n)}(x)$ turn out to form 4d Majorana spinors. Then, we see that $\psi(x, z)$ obeying CBCs has two degrees of

freedom. After integrating over the fifth coordinate in (13), we obtain the KK masses:

$$\left(m_n^{(\psi)}(\theta)\right)^2 = k^2 \left\{ \frac{2\pi n + \beta - q_\psi \theta(x)}{2(z_L - 1)} \right\}^2 + M_\psi^2, \tag{25}$$

where the Wilson line phase $\theta(x)$ is defined by

$$\theta(x) \equiv \int_{-\pi R}^{\pi R} dy\, g_5 C_y^{(0)}(x,y) = 2\int_1^{z_L} dz\, g_5 C_z^{(0)}(x,z) = 2g_5 \sqrt{\frac{z_L^2 - 1}{2k}}\, \phi(x). \tag{26}$$

In (26), we use the relation:

$$C_z^{(0)}(x,z) = \frac{1}{\sqrt{k}} h^{(0)}(z) C_z^{(0)}(x) = \frac{1}{\sqrt{k}} \sqrt{\frac{2}{z_L^2 - 1}}\, z\phi(x) \tag{27}$$

and denote $C_z^{(0)}(x)$ as $\phi(x)$ which will be treated as an inflaton later.

2.4 Effective potential

Let us derive the effective potential for the Wilson line phase $\theta(= \theta(x))$. Taking the standard procedure, a d-dimensional effective potential involving one degree of freedom at the one-loop level is given by

$$\begin{aligned} V_{\text{eff}}(\theta) &= \mp \frac{i}{2} \int \frac{d^d p}{(2\pi)^d} \ln\left\{ p^2 + m_n(\theta)^2 \right\} \\ &= \mp \frac{\Gamma(-d/2)}{2(4\pi)^{\frac{d}{2}}} \sum_n m_n(\theta)^d \quad \text{for} \begin{cases} + : \text{boson} \\ - : \text{fermion} \end{cases}, \end{aligned} \tag{28}$$

where $m_n(\theta)$ are KK masses and $\Gamma(-d/2)$ is the gamma function. Using (25) and the Poisson summation formula, we can get the effective potential involving a fermion $\psi(x,z)$ with two degrees of freedom:

$$\begin{aligned} V_{\text{eff}}(\theta) &= \frac{3k^4}{2\{2(z_L-1)\}^4 \pi^2} \sum_{n=1}^{\infty} \frac{\cos\{n(\beta - q_\psi\theta)\}}{n^5} B_{5/2}(\alpha n) + \cdots, \\ &= \frac{k^4}{2\{2(z_L-1)\}^4 \pi^2} \sum_{n=1}^{\infty} \frac{\cos\{n(\beta - q_\psi\theta)\}}{n^5} (\alpha^2 n^2 + 3\alpha n + 3) e^{-\alpha n} + \cdots, \end{aligned} \tag{29}$$

where the ellipsis stands for θ-independent terms including infinities, $\alpha = 2(z_L - 1)M_\psi/k$ and $B_{5/2}(\alpha n)$ is a function with $\nu = 5/2$ and $z = \alpha n$ defined by

$$B_\nu(z) \equiv \frac{2}{\Gamma(\nu)} \left(\frac{z}{2}\right)^\nu K_\nu(z), \quad B_\nu(0) = 1. \tag{30}$$

In (30), $K_\nu(z)$ is the modified Bessel function expressed by

$$K_\nu(z) = \frac{1}{2}\left(\frac{z}{2}\right)^{-\nu} \int_0^\infty dt\, t^{\nu-1} \exp\left(-t - \frac{z^2}{4t}\right) \quad \text{for } z \neq 0,\ |\arg z| < \frac{\pi}{4}. \tag{31}$$

We notice that the expression of $V_{\rm eff}(\theta)$ is similar to that of a flat spacetime such as $M^4 \times S^1$.

3 Gauge-Higgs inflation on a warped background

We will now see whether $V_{\rm eff}(\theta)$ given by (29) or its extension including contributions from several fermions can act as an inflaton potential or not.

3.1 Constraints on inflaton potential

The inflaton potential $V(\phi)$ is strictly constrained by cosmological observations. In particular, the allowed region of the tensor-to-scalar ratio r is getting smaller than that in the past. Constraints at the wavenumber $k_0 = 0.002$ Mpc^{-1} are listed as [7]

$$\text{slow-roll conditions :} \quad \epsilon \equiv \frac{M_{\rm Pl}^2}{2}\left(\frac{V'(\phi)}{V(\phi)}\right)^2 \ll 1, \quad |\eta| \equiv M_{\rm Pl}^2 \left|\frac{V''(\phi)}{V(\phi)}\right| \ll 1, \tag{32}$$

$$\text{e-folding number :} \quad N = \frac{1}{M_{\rm Pl}^2}\left|\int_{\phi_*}^{\phi_{\rm e}} \frac{V}{V'} d\phi\right| > 50, \tag{33}$$

$$\text{scalar power spectrum :} \quad \mathcal{P}_\mathcal{R} = \frac{1}{24\pi^2 M_{\rm Pl}^4}\frac{V(\phi_*)}{\epsilon_*} = (2.105 \pm 0.030) \times 10^{-9} \quad (68\,\%\,\text{CL}), \tag{34}$$

$$\text{spectral index :} \quad n_{\rm s} = 1 - 6\epsilon_* + 2\eta_* = 0.9661 \pm 0.0040 \quad (68\,\%\,\text{CL}), \tag{35}$$

$$\text{tensor-to-scalar ratio :} \quad r = 16\epsilon_* < 0.066 \quad (95\,\%\,\text{CL}), \tag{36}$$

$$\text{vacuum energy :} \quad V(\phi)|_{\rm min.} = 0, \tag{37}$$

where $M_{\rm Pl} = \sqrt{\hbar c^5/8\pi G} = 2.4353 \times 10^{18}$ GeV, the parameters with $*$ indicate the values at the horizon exit, and $\phi_{\rm e}$ denotes the value of ϕ at the end of inflation, which is limited by the conditions $\epsilon_{\rm e} \simeq 1$ and $|\eta_{\rm e}| \simeq 1$. Here, $\epsilon_{\rm e}$ and $\eta_{\rm e}$ are the values of ϵ and η at $\phi_{\rm e}$. The condition (37) corresponds to the fact that the current vacuum energy (dark energy) density is almost zero.

3.2 Model with a single fermion

From (26), let us use $\phi(= f\theta)$ in place of θ, where $f \equiv \sqrt{2k/(z_L^2 - 1)}/(2g_5)$. Then, the potential is given by

$$V(\phi) = \frac{k^4}{2\{2(z_L-1)\}^4\pi^2}\sum_{n=1}^{\infty}\frac{\cos\{n(\beta - q_\psi\phi/f)\}}{n^5}(\alpha^2 n^2 + 3\alpha n + 3)e^{-\alpha n} + c_{\rm 5d} + \cdots, \tag{38}$$

where a 5d cosmological constant term $c_{5\mathrm{d}}$ is introduced as a counter term to renormalize the vacuum energy density and to satisfy the condition (37).

A graph of $V(\phi)$ is depicted in Figure 1. Here, it is obtained by summing the function

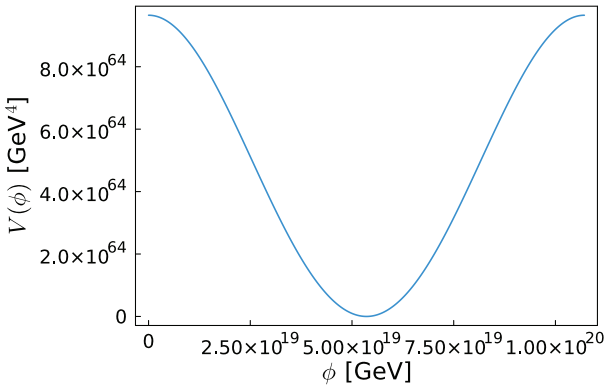


Figure 1: $V(\phi)$ $(0 < \phi/f < 2\pi)$ in the case with $f = 7M_{\mathrm{Pl}}$, $k = 2.465 \times 10^{18}$ GeV, $\alpha = 1$, $\beta = 0$, $q_\psi = 1$ and $2(z_L - 1) = 100$.

of n in (38) for $n = 1$ to 100 numerically, and taking $f = 7M_{\mathrm{Pl}}$, $k = 2.465 \times 10^{18}$ GeV, $\alpha = 2(z_L - 1)M_\psi/k = 1$, $\beta = 0$, $q_\psi = 1$ and $2(z_L - 1) = 100$. In this case, choosing $n_{\mathrm{s}} = 0.9641$ as an allowed value of n_{s}, the values of ϕ_*, ϵ_* and η_* are determined as $\phi_* = 1.81 \times 10^{19}$ GeV, $\epsilon_* = 0.0040$ and $\eta_* = -0.0060$, respectively, from the dependence of ϵ, η and n_{s} on ϕ, and the slow-roll conditions are fulfilled. Then, the tensor-to-scalar ratio becomes $r = 0.064$ and a suitable value of the scalar power spectrum is obtained as $\mathcal{P}_{\mathcal{R}} = 2.104 \times 10^{-9}$ consistent with (34). The n_{s}-r contours on the Planck 2018 results and our values are shown in Figure 2. When the inflation ended at $\phi_{\mathrm{e}} = 5.01 \times 10^{19}$ GeV with the e-folding number $N = 64$, we find that the conditions $\epsilon_{\mathrm{e}} \simeq 1$ and $|\eta_{\mathrm{e}}| \simeq 1$ are satisfied with $\epsilon_{\mathrm{e}} = 1.01$ and $\eta_{\mathrm{e}} = 1.00$.

Let us change the value of f keeping other parameters unchanged. The tensor-to-scalar ratio r can be reduced when the value of f is smaller. For instance, $f = 6M_{\mathrm{Pl}}$ with $n_{\mathrm{s}} = 0.9641$ can produce $r = 0.0273$ and this is within the 1σ region of BICEP2/Keck and BAO constraint, but, in this case, the e-folding number takes a larger value $N = 80$. The values of r and N for several values of f are listed in the Table 1. From them, we see that it is difficult to have a plot in the 1σ region for a model with a single fermion, keeping an appropriate value of N. This feature remains almost unchanged even if the values of α, β, q_ψ and n_{s} are altered.

Table 1: The tensor-to-scalar ratio r and the e-folding number N for several vaules of f.

	f	$7M_{\mathrm{Pl}}$	$6.5M_{\mathrm{Pl}}$	$6.3M_{\mathrm{Pl}}$	$6M_{\mathrm{Pl}}$
$n_{\mathrm{s}} = 0.9641$	r	0.064	0.048	0.0041	0.0273
	N	64	69	72	80
$n_{\mathrm{s}} = 0.9661$	r	0.055	0.039	0.031	0.017
	N	69	76	81	94

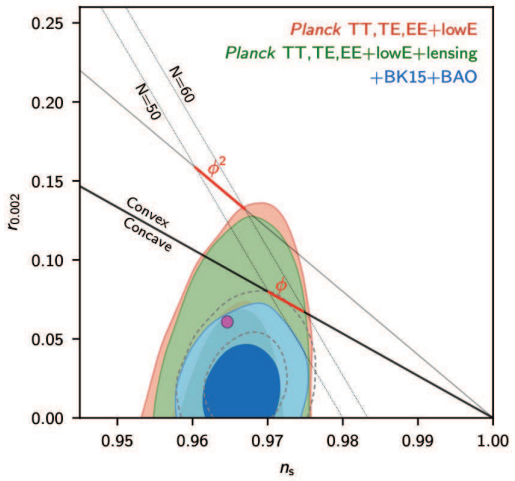


Figure 2: The $n_{\rm s}$-r contours on the Planck 2018 results [7] and the values obtained from our model with a single fermion in the case with $f = 7M_{\rm Pl}$. The pink dot represents $(n_{\rm s}, r) = (0.9641, 0.064)$ with $N = 64$. The vertical axis is denoted as $r_{0.002}$ in place of r to represent the observed value of r at the wavenumber $k_0 = 0.002$ Mpc^{-1}.

3.3 Model with multiple fermions

Next, let us consider a combination of these cos-type potentials. We have now two fermions ψ_1 and ψ_2. Each fermion has the values of α, β and q_ψ as shown in Table 2, and then the potential is written as

$$V(\phi) = \frac{3k^4}{2\{2(z_L-1)\}^4\pi^2}\left\{\sum_{n=1}^{\infty}\frac{\cos\{n(-\phi/f)\}}{n^5}B_{5/2}(\alpha_1 n) + \sum_{n=1}^{\infty}\frac{\cos\{n(\pi/2-2\phi/f)\}}{n^5}B_{5/2}(\alpha_2 n)\right\} + c_{\rm 5d} + \cdots . \tag{39}$$

Table 2: The values of α, β and q_ψ.

	α	β	q_ψ
ψ_1	$\alpha_1 = 0.01$	0	1
ψ_2	$\alpha_2 = 2.3$	$\frac{\pi}{2}$	2

A graph is described in Figure 3. Here, it is obtained by summing the function of n in (38) for $n = 1$ to 100 numerically, and taking $f = 11M_{\rm Pl}$, $k = 2.193\times10^{18}$ GeV and $2(z_L-1) = 100$. In this case, choosing $n_{\rm s} = 0.9661$ as an allowed value of $n_{\rm s}$, the values of ϕ_*, ϵ_* and η_* are determined as $\phi_* = 1.034\times10^{20}$ GeV, $\epsilon_* = 0.0021$ and $\eta_* = -0.011$, respectively, from the dependence of ϵ, η and $n_{\rm s}$ on ϕ, and the slow-roll conditions are fulfilled. Then, the tensor-to-scalar ratio becomes $r = 0.034$ and a suitable value of the scalar power spectrum is obtained as $\mathcal{P}_\mathcal{R} = 2.106\times10^{-9}$ consistent with (34). When the inflation ended at $\phi_{\rm e} = 6.63\times10^{19}$ GeV

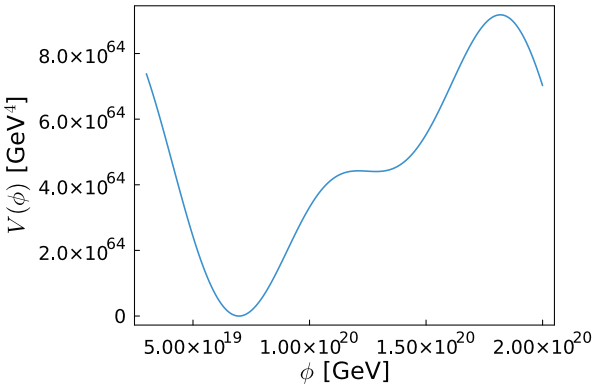


Figure 3: $V(\phi)$ in the case with $f = 11M_{\rm Pl}$, $k = 2.193 \times 10^{18}$ GeV and $2(z_L - 1) = 100$.

with the e-folding number $N = 70$, we find that the conditions $\epsilon_{\rm e} \simeq 1$ and $|\eta_{\rm e}| \simeq 1$ are satisfied with $\epsilon_{\rm e} = 1.00$ and $\eta_{\rm e} = 1.01$. In this way, this type of potential can give a smaller value of r.

Let us change the value of $n_{\rm s}$ keeping f unchanged[4]. Taking $n_{\rm s} = 0.9641$ as an allowed value of $n_{\rm s}$, the values of ϕ_*, ϵ_* and η_* are determined as $\phi_* = 1.025 \times 10^{20}$ GeV, $\epsilon_* = 0.00245$ and $\eta_* = -0.0106$, respectively. Then, we have $r = 0.039$ with $N = 65$ and hence even smaller N can be achieved. The $n_{\rm s}$-r contours on the Planck 2018 results and our values are shown in Figure 4.

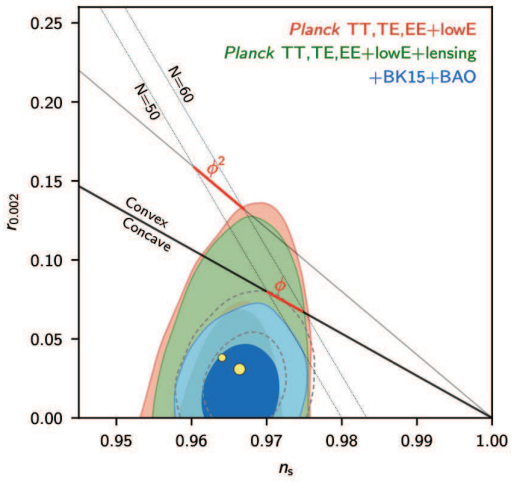


Figure 4: The $n_{\rm s}$-r contours on the Planck 2018 results [7] and the values obtained from our model with two fermions in the case with $f = 11M_{\rm Pl}$. The bigger and smaller dots represent $(n_{\rm s}, r) = (0.9661, 0.034)$ with $N = 70$ and $(0.9641, 0.039)$ with $N = 65$, respectively. The vertical axis is denoted as $r_{0.002}$ in place of r, as is the case with Figure 2.

In the same way, we can find a model with more fermions which can give an n_s-r plot inside the 1σ region. For instance, using three fermions ψ_1, ψ_2 and ψ_3 with α, β and q_ψ listed in Table 3, we can describe a graph of the potential $V(\phi)$ in Figure 5 and obtain $\phi_* =$

[4]To reproduce the appropriate value of $\mathcal{P}_\mathcal{R}$, the value of k or z_L should be changed. For instance, taking $k = 2.293 \times 10^{18}$ GeV with $2(z_L - 1) = 100$, we obtain $\mathcal{P}_\mathcal{R} = 2.103 \times 10^{-9}$.

9.30×10^{19} GeV, $\epsilon_* = 0.0025$, $\eta_* = -0.010$, $\mathcal{P}_\mathcal{R} = 2.103 \times 10^{-9}$ and $r = 0.040$ for $f = 10M_{\rm Pl}$, $k = 2.006 \times 10^{18}$ GeV, $2(z_L - 1) = 100$ and $n_{\rm s} = 0.9641$. In this case, the inflation ended at $\phi_{\rm e} = 6.28 \times 10^{19}$ GeV, $\epsilon_{\rm e} = 1.00$ and $\eta_{\rm e} = 0.98$ with the e-folding number $N = 66$.

Table 3: The values of α, β and q_ψ.

	α	β	q_ψ
ψ_1	0.01	0	1
ψ_2	2.1	$\frac{\pi}{2}$	2
ψ_3	2.1	$\frac{3\pi}{2}$	1

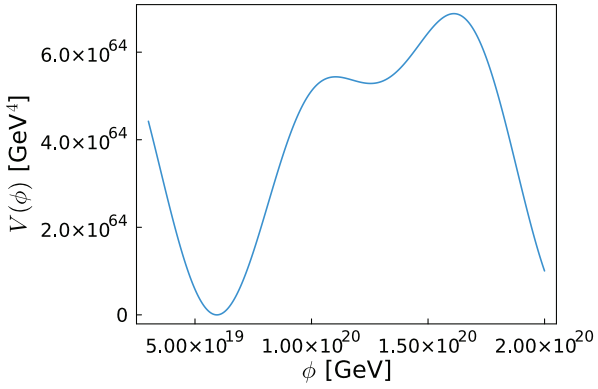


Figure 5: $V(\phi)$ in the case with $f = 10M_{\rm Pl}$, $k = 2.006 \times 10^{18}$ GeV and $2(z_L - 1) = 100$.

4 Conclusions and discussions

We have investigated whether inflation in the early universe can be induced by an extra component of a 5d U(1) gauge field coupled to fermions in the RS warped spacetime or not. Imposing CBCs on fields involved in inflation, we have shown that the one-loop effective potential for a Wilson line phase obtained by quantum corrections can behave as an inflaton potential by finding parameter regions consistent with Planck 2018 results. Our model shares basic properties with higher-dimensional models on a flat spacetime, e.g., a potential is generated radiatively, and is robust owing to the gauge symmetry, and then a model has a predictability for physics on inflation.

Both models with a single fermion and multiple fermions have allowed regions of parameters consistent with Planck 2018 results. The former model can give an $n_{\rm s}$-r plot inside the 2σ region of BICEP2/Keck and BAO constraint, but it is difficult to get them into the 1σ region while keeping an appropriate value of e-folding number. The latter cos-combined models can reduce the value of the tensor-to-scalar ratio r within the 1σ region with a suitable e-folding number.

The remaining issues and future prospects are as follows.

There is a fine-tuning problem that the (initial) vaules of ϕ and other parameters should be chosen in order to make ϕ as an inflaton work well. It would be better if they are determined from a fundamental theory. Or we would have to rely on the anthropic principle in the framework of multiverse.

While an inflaton can be treated independently of visible particles in our model, this feature can induce a problem how a reheating at the end of inflation occurs in the absence of couplings between the inflaton and the standard model particles at the tree level due to a mismatch of boundary conditions on $S^1/\mathbb{Z}_2$ [16].

It is necessary to consider contributions of a radion and/or modulus in order to identify an inflaton. In concrete, it is essential to study their stabilization and how inflation can be realized in a system with such several scalar fields including a Wilson line phase on the RS warped background.

It would be wonderful to construct a more realistic higher-dimensional gauge theory which solves riddles of particle physics and cosmology on a similar footing. More specifically, it would be fantastic that, in a system governed by the gauge principle, there are several Wilson line phases originated from extra components of gauge bosons, which behave as Higgs boson, an inflaton, a dark matter and so forth. A top-down approach would be useful to construct such a model in a more unified form. Superstring theory, for instance, provides many scalar and vector fields, and possesses many candidates of Higgs boson, an inflaton, a dark matter and so on. Hence, it is intriguing to investigate how a 5d effective theory on the warped geometry can be realized in such a top-down framework.

Acknowledgements

T.K. would like to thank Dr. O. Seto for the helpful discussions and advice. This work was supported in part by scientific grants from the Ministry of Education, Culture, Sports, Science and Technology under Grant No. 22K03632 (YK).

References

[1] D. H. Lyth and A. Riotto, "Particle physics models of inflation and the cosmological density perturbation", Phys. Rept. **314** , 1 (1999), arXiv:hep-ph/9807278, and references therein.

[2] Y. Hosotani, "Dynamical Mass Generation by Compact Extra Dimensions", Phys. Lett. B**126**, 309 (1983).

[3] Y. Hosotani, "Dynamics of Nonintegrable Phases and Gauge Symmetry Breaking", Annals. Phys. **190**, 233 (1989).

[4] N. Arkani-Hamed, H.-C. Cheng, P. Creminelli, and L. Randall, "Extranatural Inflation", Phys. Rev. Lett. **90**, 221302 (2003), arXiv:hep-th/0301218.

[5] Y. Abe, T. Inami, Y. Kawamura, and Y. Koyama, "Radion stabilization in the presence of a Wilson line phase", Prog. Theor. Exp. Phys. **2014**, 073B04 (2014), arXiv:1404.5125 [hep-th].

[6] Y. Abe, T. Inami, Y. Kawamura, and Y. Koyama, "Inflation from radion gauge-Higgs potential at Planck scale", Prog. Theor. Exp. Phys. **2015**, 093B03 (2015), arXiv:1504.06905 [hep-th].

[7] N. Aghanim et al. (Planck), "Planck 2018 results. VI. Cosmological parameters", Astron. Astrophys. **641**, [Erratum: Astron.Astrophys. 652, C4 (2021)], A6 (2020), arXiv:1807.06209 [astro-ph.CO].

[8] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", Phys. Rev. Lett. **83**, 3370-3373 (1999), arXiv:hep-ph/9905221.

[9] L. Randall and R. Sundrum, "An Alternative to compactification", Phys. Rev. Lett. **83**, 4690-4693 (1999), arXiv:hep-ph/9906064.

[10] A. D. Medina, N. R. Shah, and C. E. M. Wagner, "Gauge-Higgs Unification and Radiative Electroweak Symmetry Breaking in Warped Extra Dimensions", Phys. Rev. D **76**, 095010 (2007), arXiv:0706.1281 [hep-ph].

[11] Y. Hosotani, K. Oda, T. Ohnuma, and Y. Sakamura, "Dynamical Electroweak Symmetry Breaking in SO(5) × U(1) Gauge-Higgs Unification with Top and Bottom Quarks", Phys. Rev. D **78**, [Erratum: Phys. Rev.D **79**, 079902 (2009)], 096002 (2008), arXiv:0806.0480[hep-ph].

[12] Y. Hosotani, "New dimensions from gauge-Higgs unification", PoS CORFU2016, 026 (2017), arXiv:1702.08161 [hep-ph].

[13] S. Funatsu, H. Hatanaka, Y. Hosotani, Y. Orikasa, and N. Yamatsu, "CKM matrix and FCNC suppression in SO(5)×U(1)×SU(3) gauge-Higgs unification", Phys. Rev. D **101**, 055016 (2020), arXiv:1909.00190 [hep-ph].

[14] A. Hebecker and J. March-Russell, "The structure of GUT breaking by orbifolding", Nucl. Phys. B **625**, 128-150 (2002), arXiv:hep-ph/0107039.

[15] N. Haba, Y. Kawamura, and K.-y. Oda, "Dynamical Rearrangement of Theta Parameter in Presence of Mixed Chern-Simons Term", Phys. Rev. D **78**, 085021 (2008), arXiv:0803.4380 [hep-ph].

[16] Y. Abe, Y. Goto, Y. Kawamura, and Y. Nishikawa, "Conjugate boundary condition, hidden particles, and gauge-Higgs inflation", Mod. Phys. Lett. A **31**, 1650208 (2016), arXiv:1608.06393 [hep-ph].

[17] K. Kojima and Y. Okubo, "Early dark energy from a higher-dimensional gauge theory", Phys. Rev. D **106**, 063540 (2022), arXiv:2205.13777 [astro-ph.CO].

[18] C. Cśaki, S. Lombardo, and O. Telem, "TASI Lectures on Non-supersymmetric BSM Models", Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016 , edited by R. Essig and I. Low (WSP, 2018), 501-570, arXiv:1811.04279 [hep-ph].

Asymmetric Dark Matter and Sommerfeld Enhancement

Sujuan Qiu, Hoernisa Iminniyaz*, Wensheng Huo

School of Physics Science and Technology, Xinjiang University, Urumqi 830017, China

arXiv:2403.01682v1 [hep-ph] 4 Mar 2024

Abstract

We show that the relic density of asymmetric dark matter with long-range interactions is affected by the Sommerfeld enhancement. The annihilation cross section of asymmetric dark matter is enhanced due to the Sommerfeld effect. It results the relic density is decreased. Using the Planck date, the constraints on the asymmetry factor and coupling required to attain the observed dark matter density are obtained. We derive the upper bounds on the mass for asymmetric dark matter in $s-$wave and $p-$wave annihilation cases.

*Corresponding author, wrns@xju.edu.cn

1 Introduction

Recently a number of models of asymmetric Dark Matter (DM) are proposed as an alternative of symmetric DM for which the particle and antiparticle are identical [1, 2, 3]. Asymmetric DM models offer an effective explanation for the similarity of DM and baryon abundances, $\Omega_{DM} \sim \Omega_b$ [4]. In asymmetric DM scenarios, DM is not neutral, and self-conjugate, rather DM possesses particle-antiparticle asymmetry. In that case, the final abundance is determined not only by the annihilation cross section, but also by the asymmetry factor which is the difference between the particle and antiparticle abundances [5, 6]. The antiparticle abundance is largely depressed to the negligible levels when the thermal bath temperature is much lower than the DM mass. The DM abundances in the presence of a particle-antiparticle asymmetry is computed in detail in Refs.[5, 6].

In asymmetric DM models, asymmetric DM is assumed to couple to the light scalar or vector mediators [3, 7, 8, 9, 10]. The interaction between the asymmetric DM particles and antiparticles is appeared as long-range if the mediator is light enough. The wavefunction of DM particle and antiparticle is distorted by the long-range interaction. It is indeed the Sommerfeld effect [11], which leads the enhancement of DM annihilation rate at low velocities [12, 13]. The couplings required to obtain the observed DM density through the thermal freeze out mechanism of DM is suppressed due to the Sommerfeld effect [13, 14]. On the other hand, the late-time DM annihilation signals are enhanced for the certain set of couplings [15, 16, 17, 18, 19, 20, 21, 22]. Because the Sommerfeld enhancement relies on the coupling of DM to the light force mediator, for the same mass, asymmetric DM needs stronger couplings than symmetric DM, there maybe rather significant effect than symmetric DM in phenomenologically.

In Refs.[23, 24], the authors explored the effect of Sommerfeld enhancement on the relic density of asymmetric DM in detail when the light mediator is massless. The authors deduced constraints on the mass and couplings required to obtain the observed value of DM abundance in Ref.[23]. Ref.[25] discussed the impact of Sommerfeld enhancement on relic density in the case of s-wave annihilation when the mediator mass $m_\varphi \neq 0$ in model independent way. They concluded that the abundance of asymmetric DM is decreased due to the enhancement of the annihilation cross section. In the present work, we consider two minimal cases, in that scenarios, asymmetric DM is coupled to a light vector boson or a light scalar boson. We investigate the affect of Sommerfeld enhancement on the relic density of asymmetric DM for $s-$wave and $p-$wave annihilations in detail when the mediator mass $m_\varphi \neq 0$, especially the effect of $p-$wave Sommerfeld enhancement on the freeze out of asymmetric DM. We extend our previous analysics in several ways. First, we plot the ratio of asymmetric DM antiparticle to particle abundances with Sommerfeld enhancement as a function of the inverse-scaled tem-

perature. We find the antiparticle abundance is depleted faster while the annihilation cross section is enhanced by the Sommerfeld effect. The decrease of the ratio is significant for the larger couplings. We calculate the couplings required to attain the observed DM relic density as a function of the asymmetry factor. The coupling is larger for asymmetric case than the symmetric one with the Sommerfeld enhancement. The final abundance is largely determined by the asymmetry. When the asymmetry is large, the required mass bound is smaller. We also noticed that the maximum value of mass needed to satisfy the observed value of DM abundance is smaller than the symmetric DM. In our work, we ignored the effect of bound state formation on the relic density of asymmetric DM which affects the relic density of DM only around the unitarity bound [10, 14].

The paper is arranged as following. In section 2, we discuss the impact of Sommerfeld effect on the relic abundance of asymmetric DM for $s-$wave and $p-$wave annihilation when the mediator mass $m_\varphi \neq 0$. In section 3, we find contraints on the parameter spaces as couplings, mass and asymmetry when the annihilation cross section of asymmetric DM is modified by the Sommerfeld enhancement. The conclusion and summaries are in the last section.

2 Relic abundance of asymmetric DM including Sommerfeld enhancement

We discuss the effect of Sommerfeld enhancement on the relic density of asymmetric DM in this section. The number density of asymmetric DM particle χ (antiparticle $\bar{\chi}$) is evolved according to the following Boltzmann equation,

$$\frac{\mathrm{d}n_{\chi(\bar{\chi})}}{\mathrm{d}t} + 3Hn_{\chi(\bar{\chi})} \;=\; -\langle\sigma v\rangle(n_\chi n_{\bar{\chi}} - n_{\chi,\mathrm{eq}} n_{\bar{\chi},\mathrm{eq}})\,, \tag{1}$$

where $H = \pi T^2/M_{\mathrm{Pl}}\sqrt{g_*/90}$ is the expansion rate in the radiation dominated era, here $M_{\mathrm{Pl}} = 2.4\times10^{18}$ GeV is the reduced Planck mass, and g_* being the effective number of relativistic degrees of freedom. The equilibrium number densities are $n_{\chi(\bar{\chi}),\mathrm{eq}} = g_\chi\big[mT/(2\pi)\big]^{3/2}\mathrm{e}^{(-m\pm\mu_\chi)/T}$, where g_χ is the number of intrinsic degrees of freedom of the particle. Here the chemical potentials $\mu_{\bar{\chi}} = -\mu_\chi$ when the asymmetric DM is in equilibrium state. Thermally averaged Sommerfeld enhanced annihilation cross section is

$$\langle\sigma v\rangle = a\,\langle S_s\rangle + b\,\langle v^2\,S_p\rangle\,. \tag{2}$$

Here a is the $s-$wave contribution when $v\to 0$ and b is the $p-$wave contribution when $s-$wave is suppressed. The Sommerfeld enhancement factors for the massive mediator in the case of

$s-$wave and $p-$wave annihilations are given in [26, 27, 28, 29]. For $s-$wave annihilation,

$$S_s = \frac{2\pi\alpha}{v} \frac{\sinh\left(\frac{6mv}{\pi m_\phi}\right)}{\cosh\left(\frac{6mv}{\pi m_\phi}\right) - \cos\left(2\pi\sqrt{\frac{6\alpha m}{\pi^2 m_\phi} - \frac{9m^2v^2}{\pi^4 m_\phi^2}}\right)}\,, \tag{3}$$

where v is the relative velocity of the annihilating asymmetric DM particle χ and antiparticle $\bar{\chi}$, and α is the coupling strength. For $p-$wave annihilation, the Sommerfeld enhancement factor is

$$S_p = \frac{\left(6\alpha m/(\pi^2 m_\phi) - 1\right)^2 + 36m^2v^2/(\pi^4 m_\phi^2)}{1 + 36m^2v^2/(\pi^4 m_\phi^2)}\, S_s\,. \tag{4}$$

Then

$$\langle S_s\rangle = \frac{x^{3/2}}{2\sqrt{\pi}} \int_0^\infty v^2\; e^{-\frac{x}{4}v^2}\; S_s\, \mathrm{d}v\,. \tag{5}$$

$$\langle v^2\, S_p\rangle = \frac{x^{3/2}}{2\sqrt{\pi}} \int_0^\infty v^4\; e^{-\frac{x}{4}v^2}\; S_p\, \mathrm{d}v\,. \tag{6}$$

For convenience, Eq.(1) can be expressed in terms of $Y_{\chi(\bar{\chi})} = n_{\chi(\bar{\chi})}/s$, and $x = m/T$, here $s = 2\pi^2 g_{*s}/45\, T^3$ is the entropy density, where g_{*s} is the effective number of entropic degrees of freedom. Then

$$\frac{\mathrm{d}Y_{\chi(\bar{\chi})}}{\mathrm{d}x} = -\frac{\lambda\langle\sigma v\rangle}{x^2}\left(Y_\chi\; Y_{\bar{\chi}} - Y_{\chi,\mathrm{eq}}\; Y_{\bar{\chi},\mathrm{eq}}\right), \tag{7}$$

here the entropy conservation is used to derive Eq.(7) and $\lambda = 1.32\, m M_{\mathrm{Pl}}\sqrt{g_*}$, $g_* \simeq g_{*s}$ and $dg_{*s}/dx \simeq 0$.

The Boltzmann equation for particle (antiparticle) is rewritten as

$$\frac{\mathrm{d}Y_{\chi(\bar{\chi})}}{\mathrm{d}x} = -\frac{\lambda\langle\sigma v\rangle}{x^2}\left(Y^2_{\chi(\bar{\chi})} \mp \eta Y_{\chi(\bar{\chi})} - Y^2_{\mathrm{eq}}\right), \tag{8}$$

here $Y_\chi - Y_{\bar{\chi}} = \eta$ is used which is obtained by subtracting the Boltzmann equations for χ and $\bar{\chi}$ in Eq.(7), where η is a constant and

$$Y^2_{\mathrm{eq}} = Y_{\chi,\mathrm{eq}} Y_{\bar{\chi},\mathrm{eq}} = (0.145 g_\chi/g_*)^2\, x^3 e^{-2x}. \tag{9}$$

The Boltmann equation (8) is solved following the standard picture of particle evolution scenarios. It is supposed the asymmetric DM particles and antiparticles were in thermal equilibrium with the standard model particles in the early universe. When the temperature drops as $T < m$ for $m > |\mu_\chi|$, the interaction rate Γ falls below the expansion rate H. At this freeze out point, asymmetric DM is decoupled from equilibrium state and the number densities of asymmetric DM in a co–moving space almost become constant [5, 6, 30]. Following the method which is used in [6], the final abundance for antiparticle is determined as

$$Y_{\bar{\chi}}(x_\infty) = \frac{\eta}{\exp\left[\,1.32\,\eta\, m M_{\mathrm{Pl}}\sqrt{g_*}\int_{\bar{x}_F}^\infty \langle\sigma v\rangle/x^2\,\mathrm{d}x\,\right] - 1}\,, \tag{10}$$

here

$$Y_{\bar{\chi}}(x_\infty) \equiv \lim_{x\to\infty} Y_{\bar{\chi}}(x)\,. \tag{11}$$

The freeze out temperature $\bar{x}_F$ is fixed by using the standard procedure that the freeze out occurred at the point when the deviation of the relic abundance $Y_{\bar{\chi}}(\bar{x}_F)$ and $Y_{\bar{\chi},\mathrm{eq}}(\bar{x}_F)$ is the same order as the equilibrium value such as $Y_{\bar{\chi}}(\bar{x}_F) = (1+\xi)Y_{\bar{\chi},\mathrm{eq}}(\bar{x}_F)$. Here ξ is a constant and usually we take $\xi = \sqrt{2}-1$ [30]. The relic abundance for particle is obtained by using $Y_\chi = Y_{\bar{\chi}} + \eta$,

$$Y_\chi(x_\infty) = \frac{\eta}{1-\exp\left[-1.32\,\eta\, m M_{\mathrm{Pl}}\sqrt{g_*}\int_{x_F}^{\infty}\langle\sigma v\rangle/x^2\,\mathrm{d}x\right]}\,, \tag{12}$$

where x_F is the freeze out temperature for χ. Eqs.(10) and (12) are consistent with the constraint $Y_\chi = Y_{\bar{\chi}} + \eta$ only if $x_F = \bar{x}_F$. The total final DM relic density is

$$\Omega_{\mathrm{DM}}h^2 \;=\; 2.76\times10^8\,\left[Y_\chi(x_\infty) + Y_{\bar{\chi}}(x_\infty)\right]\,m, \tag{13}$$

where $\Omega_\chi = \rho_\chi/\rho_c$ with $\rho_\chi = n_\chi m = s_0 Y_\chi$ and $\rho_c = 3H_0^2 M_{\mathrm{Pl}}^2$, here $s_0 \simeq 2900$ cm^{-3} is the present entropy density, and H_0 is the Hubble constant.

Fig. 1 shows the ratio of the antiparticle abundance $Y_{\bar{\chi}}$ to particle abundance Y_χ as a function of the inverse-scaled temperature x for the Sommerfeld enhanced annihilation cross section. It is plotted using the numerical solutions of Eq.(8). Here panels (a) and (c) are for the $s-$wave annihilations, panels (b) and (d) are for $p-$wave annihilations. The particle and antiparticle abundances keep in the same amount before the decoupling from equilibrium state. After freeze out, the DM abundance becomes almost constant as stated previously which is shown in the figure. We find the affect of Sommerfeld enhancement is significant for the stronger coupling α and for the larger asymmetry factor η. In other words, the depletion of antiparticle abundance is faster when α is large. For $p-$wave annihilation, the decrease of the ratio is slower than $s-$wave annihilation.

3 Constraints

In this section we discuss two minimal cases in which asymmetric DM couples to a light vector or scalar boson. In the case of vector mediator, we consider the process $\chi\bar{\chi} \to 2\gamma$ which discussed in [23]. The Sommerfeld enhanced annihilation cross section times relative velocity is

$$(\sigma v)_{\mathrm{s-wave}} = \sigma_0\, S_s\,, \tag{14}$$

where the annihilation into two vector bosons is an $s-$wave process and the leading order cross section is $\sigma_0 = \pi\alpha^2/m^2$ [23].

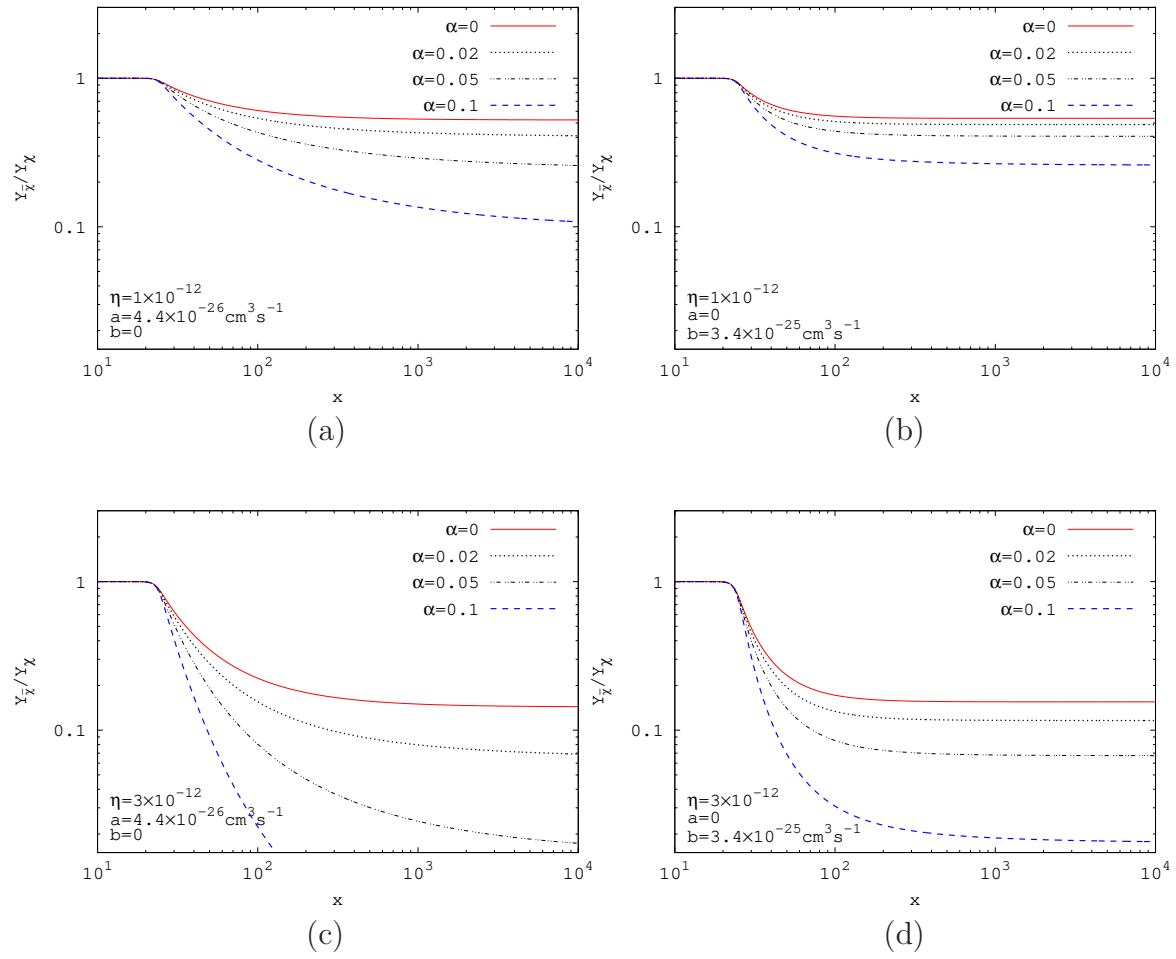


Figure 1: Ratio of asymmetric DM abundances $Y_{\bar{\chi}}$ and Y_χ as a function of x for the Sommerfeld enhanced annihilation cross section. Here $m = 130$ GeV, $m_\varphi = 0.25$ GeV, $g_\chi = 2$, $g_* = 90$.

For the scalar mediator, the process $\chi\bar{\chi} \to 2\varphi$ is considered. The annihilation cross section times relative velocity is given by

$$(\sigma v)_{\rm p-wave} = \sigma_1 \, v^2 S_p \,, \tag{15}$$

where the annihilation of two fermions into two scalar bosons is a $p-$wave process and the tree level cross section is $\sigma_1 = 3\pi\alpha^2/(8m^2)$ [23].

Following we insert Eq.(14) and Eq.(15) into Eq.(8) separately and using Planck data to find constraints on the coupling strength, asymmetry factor and mass. In Fig. 2, the relation between asymmetry factor η and the coupling strength α for $s-$wave($p-$wave) annihilation is shown for the value of total DM density $\Omega_{\rm DM} = 0.120$ [4]. Here σ_0 and σ_1 are evaluated using the correlated α. The abundance is not sensitive for the smaller value of η. When η is small, the symmetric case is recovered. Therefore, the contour line climbs vertically (countours are independent of η). On the other hand, the smaller η gives the minimal allowed value of

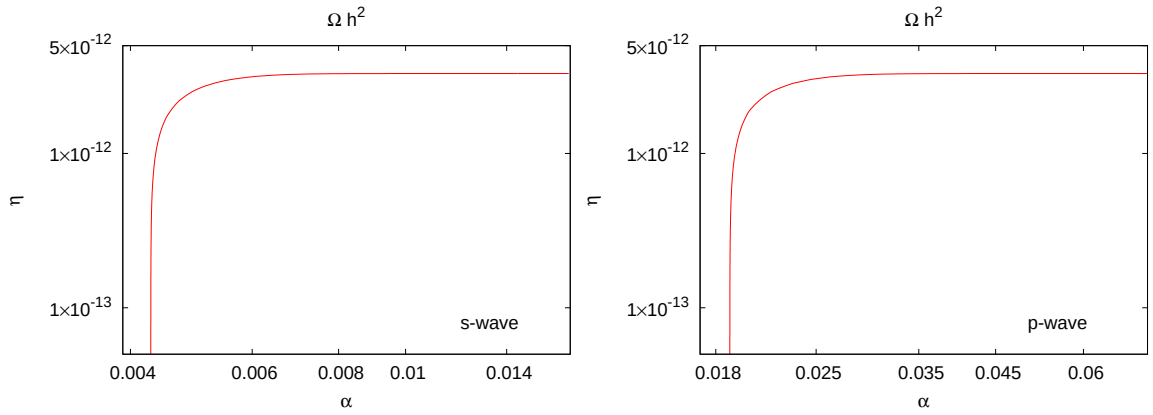


Figure 2: Contour plots of asymmetry factor η with coupling strength α for $s-$wave ($p-$wave) annihilation cross section when $\Omega_{\rm DM}h^2 = 0.120$ [4]. Here $m = 130$ GeV, $m_\varphi = 0.25$ GeV, $g_\chi = 2$, $g_* = 90$.

the coupling strength $\alpha = 0.0043$ for $s-$wave and $\alpha = 0.0188$ for $p-$wave annihilation. The lower bound of α is larger for $p-$wave annihilation. The curves flatten out rapidly while the asymmetry η is increased. The relic abundance is determined by the asymmetry. In that case, the abundance is not affected when the coupling strength is increased (independent of α). A larger α, corresponds to the larger enhancement of the cross section. It leads to a smaller relic density. This should be compensated by increasing η.

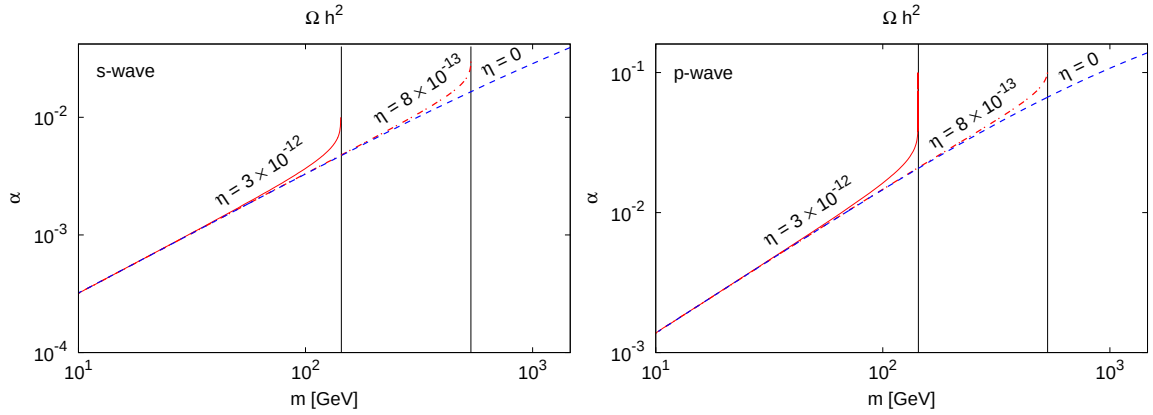


Figure 3: Contour plots of the coupling strength α and mass m for $s-$wave ($p-$wave) annihilation when $\Omega_{\rm DM}h^2 = 0.120$ [4]. Here $m_\varphi = 0.25$ GeV, $g_\chi = 2$, $g_* = 90$, where σ_0 and σ_1 are evaluated using the correlated α.

The coupling strength required to obtain the observed value of DM relic density with mass is plotted in Fig. 3. The mass bounds for $\eta = 3 \times 10^{-12}\,,\ 8 \times 10^{-13}$ are $m = 143\,,\ 535$ GeV respectively in panel (a); $m = 143\,,\ 532$ GeV in panel (b). Highly asymmetric DM has lower

mass bound. For the same mass bound, asymmetric DM needs larger coupling to attain the observed value of DM relic density compared to the symmetric case. For example, $m = 143$, the coupling $\alpha = 0.01$ for $\eta = 3 \times 10^{-12}$ and $\alpha = 0.005$ for $\eta = 0$. The coupling for $s-$wave annihilation is smaller than the case of $p-$wave annihilation.

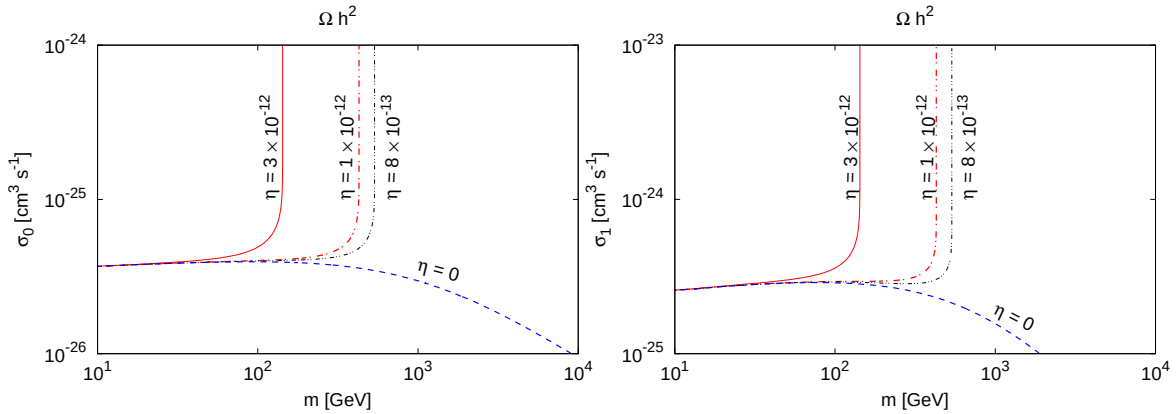


Figure 4: Contour plots of $s-$wave ($p-$wave) annihilation cross section $\sigma_0(\sigma_1)$ and the mass when $\Omega_{\rm DM}h^2 = 0.120$. Here $m_\phi = 0.25$ GeV, $g_\chi = 2$, $g_* = 90$, the coupling α is computed using the expression $\sigma_0 = \pi\alpha^2/m^2$ for $s-$wave and $\sigma_1 = 3\pi\alpha^2/(8m^2)$ for $p-$wave annihilations.

The contour plots of the cross section with mass are shown in Fig. 4. The allowed region is bounded by the cross section from below and by the maximum value of $m_{\rm max}$ from the right for different asymmetry. For $m \ll m_{\rm max}$, coupling constant α and $\sigma_0(\sigma_1)$ trace the symmetric DM ($\eta = 0$) case closely. When the cross section is small, the abundance is insensitive for the increased mass. On the other hand, while α is small, the effect of Sommerfeld enhancement is not significant for both the symmetric and asymmetric DM. When the mass reached the maximum value $m \simeq m_{\rm max}$, the Sommerfeld enhanced annihilation cross section is increased to satisfy the observed value of the DM relic density for asymmetric DM. In the symmetric case, for larger m, the coupling is stronger (see Fig. 3), then the Sommerfeld enhanced cross section reduces the relic density notably, the cross section should be small in order to obtain the required value of the relic density. Therefore, the cross section falls quickly when the mass is increased in symmetric case.

4 Summary and conclusions

When the asymmetric DM coupled to the light force carrier, the annihilation cross section is enhanced by the Sommerfeld effect. We discuss the effect of Sommerfeld enhancement on the relic density of asymmetric DM for $s-$wave and $p-$wave annihilations for the case of light

mediator $m_\varphi \neq 0$ in detail. We found the antiparticle abundance is depleted faster than the standard case due to the Sommerfeld enhanced annihilation cross section. The decrease of the ratio of antiparticle abundance to particle abundance is larger for the stronger coupling α.

We apply our method to two kinds of scenarios, asymmetric DM coupled to either light vector mediator or the scalar mediator. We use Planck data to find constraints on the coupling constant α, asymmetry factor η and the mass value. We found when η is small, the abundance is independent of η, it is indeed the symmetric case. The minimal value of the coupling strength are $\alpha = 0.0043$ for $s-$wave and $\alpha = 0.0188$ for $p-$wave annihilation. When η takes larger value, the relic abundance is determined by the asymmetry and it is independent of α. For the same mass bound $m_{\max}$, asymmetric DM needs larger coupling to attain the observed value of DM relic density compared to the symmetric case. When the cross section is small, the abundance is insensitive for the increased mass. When the mass reached the maximum value, the Sommerfeld enhanced annihilation cross section is increased to satisfy the observed value of the DM relic density for different asymmetry.

Our results are important when the affect of Sommerfeld enhancement is significant at low velocity limit. The Sommerfeld effect hints notable indirect detection signals from asymmetric DM antiparticle. This allows us to examine the asymmetric DM by the Cosmic Microwave Background (CMB) observation, the Milky ways and Dwarf galaxies.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (U2031204, 11765021) and Natural Science Foundation of Xinjiang Uygur Autonomous Region (2022D01C52).

References

[1] S. Nussinov, Phys. Lett. B **165**, (1985) 55; K. Griest and D. Seckel. Nucl. Phys. B **283**, (1987) 681; R. S. Chivukula and T. P. Walker, Nucl. Phys. B **329**, (1990) 445; D. B. Kaplan, Phys. Rev. Lett. **68**, (1992) 742; D. Hooper, J. March-Russell and S. M. West, Phys. Lett. B **605**, (2005) 228 [arXiv:hep-ph/0410114]; JCAP **0901** (2009) 043 [arXiv:0811.4153v1 [hep-ph]]; H. An, S. L. Chen, R. N. Mohapatra and Y. Zhang, JHEP **1003**, (2010) 124 [arXiv:0911.4463 [hep-ph]]; T. Cohen and K. M. Zurek, Phys. Rev. Lett. **104**, (2010) 101301 [arXiv:0909.2035 [hep-ph]]. D. E. Kaplan, M. A. Luty and K. M. Zurek, Phys. Rev. D **79**, (2009) 115016 [arXiv:0901.4117 [hep-ph]]; T. Cohen, D. J. Phalen, A. Pierce and K. M. Zurek, Phys. Rev. D **82**, (2010) 056001 [arXiv:1005.1655

[hep-ph]]; J. Shelton and K. M. Zurek, Phys. Rev. D **82**, (2010) 123512 [arXiv:1008.1997 [hep-ph]];

[2] A. Belyaev, M. T. Frandsen, S. Sarkar and F. Sannino, Phys. Rev. D **83** (2011), 015007 doi:10.1103/PhysRevD.83.015007 [arXiv:1007.4839 [hep-ph]].

[3] K. Petraki, L. Pearce and A. Kusenko, JCAP **07** (2014), 039 doi:10.1088/1475-7516/2014/07/039 [arXiv:1403.1077 [hep-ph]].

[4] N. Aghanim *et al.* [Planck], Astron. Astrophys. **641** (2020), A6 [erratum: Astron. Astrophys. **652** (2021), C4] doi:10.1051/0004-6361/201833910 [arXiv:1807.06209 [astro-ph.CO]].

[5] M. L. Graesser, I. M. Shoemaker and L. Vecchi, JHEP **1110**, (2011) 110 [arXiv:1103.2771 [hep-ph]].

[6] H. Iminniyaz, M. Drees and X. Chen, JCAP **1107**, (2011) 003 [arXiv:1104.5548 [hep-ph]].

[7] J. L. Feng, M. Kaplinghat, H. Tu and H. B. Yu, JCAP **07** (2009), 004 doi:10.1088/1475-7516/2009/07/004 [arXiv:0905.3039 [hep-ph]].

[8] P. Agrawal, F. Y. Cyr-Racine, L. Randall and J. Scholtz, JCAP **1708** (2017) 021 doi:10.1088/1475-7516/2017/08/021 [arXiv:1702.05482 [astro-ph.CO]].

[9] P. Agrawal, F. Y. Cyr-Racine, L. Randall and J. Scholtz, JCAP **05** (2017), 022 doi:10.1088/1475-7516/2017/05/022 [arXiv:1610.04611 [hep-ph]].

[10] M. Cirelli, P. Panci, K. Petraki, F. Sala and M. Taoso, JCAP **05** (2017), 036 doi:10.1088/1475-7516/2017/05/036 [arXiv:1612.07295 [hep-ph]].

[11] A. Sommerfeld, Ann. Phys. **403** (1931) 257-330

[12] J. Hisano, S. Matsumoto and M. M. Nojiri, Phys. Rev. D **67** (2003), 075014 doi:10.1103/PhysRevD.67.075014 [arXiv:hep-ph/0212022 [hep-ph]].

[13] J. Hisano, S. Matsumoto and M. M. Nojiri, Phys. Rev. Lett. **92** (2004), 031303 doi:10.1103/PhysRevLett.92.031303 [arXiv:hep-ph/0307216 [hep-ph]].

[14] B. von Harling and K. Petraki, JCAP **12** (2014), 033 doi:10.1088/1475-7516/2014/12/033 [arXiv:1407.7874 [hep-ph]].

[15] M. Pospelov and A. Ritz, Phys. Lett. B **671** (2009), 391-397 doi:10.1016/j.physletb.2008.12.012 [arXiv:0810.1502 [hep-ph]].

[16] J. D. March-Russell and S. M. West, Phys. Lett. B **676** (2009), 133-139 doi:10.1016/j.physletb.2009.04.010 [arXiv:0812.0559 [astro-ph]].

[17] K. Petraki, M. Postma and J. de Vries, JHEP **04** (2017), 077 doi:10.1007/JHEP04(2017)077 [arXiv:1611.01394 [hep-ph]].

[18] J. Hisano, S. Matsumoto, M. M. Nojiri and O. Saito, Phys. Rev. D **71** (2005), 063528 doi:10.1103/PhysRevD.71.063528 [arXiv:hep-ph/0412403 [hep-ph]].

[19] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer and N. Weiner, Phys. Rev. D **79** (2009), 015014 doi:10.1103/PhysRevD.79.015014 [arXiv:0810.0713 [hep-ph]].

[20] H. An, M. B. Wise and Y. Zhang, Phys. Rev. D **93** (2016) no.11, 115020 doi:10.1103/PhysRevD.93.115020 [arXiv:1604.01776 [hep-ph]].

[21] H. An, M. B. Wise and Y. Zhang, Phys. Lett. B **773** (2017), 121-124 doi:10.1016/j.physletb.2017.08.010 [arXiv:1606.02305 [hep-ph]].

[22] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg and P. Walia, Phys. Rev. Lett. **118** (2017) no.14, 141802 doi:10.1103/PhysRevLett.118.141802 [arXiv:1612.00845 [hep-ph]].

[23] I. Baldes and K. Petraki, JCAP **09** (2017), 028 doi:10.1088/1475-7516/2017/09/028 [arXiv:1703.00478 [hep-ph]].

[24] A. Abudurusuli and H. Iminniyaz, [arXiv:2001.08404 [hep-ph]].

[25] A. Sulitan, H. Iminniyaz and M. Baoxia, Chin. J. Phys. **70** (2021), 117-124 doi:10.1016/j.cjph.2020.12.015 [arXiv:2009.13410 [hep-ph]].

[26] A. Kamada, H. J. Kim and T. Kuwahara, JHEP **12** (2020), 202 doi:10.1007/JHEP12(2020)202 [arXiv:2007.15522 [hep-ph]].

[27] M. Duerr, K. Schmidt-Hoberg and S. Wild, JCAP **09** (2018), 033 doi:10.1088/1475-7516/2018/09/033 [arXiv:1804.10385 [hep-ph]].

[28] J. L. Feng, M. Kaplinghat and H. B. Yu, Phys. Rev. D **82** (2010) 083525 doi:10.1103/PhysRevD.82.083525 [arXiv:1005.4678 [hep-ph]].

[29] S. Cassel, J. Phys. G **37** (2010), 105009 doi:10.1088/0954-3899/37/10/105009 [arXiv:0903.5307 [hep-ph]].

[30] R. J. Scherrer and M. S. Turner, Phys. Rev. D **33**, (1986) 1585, Erratum-ibid. D **34**, (1986) 3263.

arXiv:2403.01677v1 [hep-ph] 4 Mar 2024

Elliptic anisotropy of open-charm hadron in p–Pb collisions at LHC energies from parton scatterings

Siyu Tang,[1,2] Yuan Lu,[3] Chao Zhang,[4,*] and RenZhuo Wan[2,3,†]

[1] *School of Mathematical and Physical Sciences, Wuhan Textile University, Wuhan 430200, China*
[2] *Hubei Key Laboratory of Digital Textile Equipment, Wuhan Textile University, Wuhan 430200, China*
[3] *School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan 430200, China*
[4] *School of Science, Wuhan University of Technology, Wuhan, 430200, China*

(Dated: March 5, 2024)

The elliptic azimuthal anisotropy coefficient (v_2) of open-charm hadron at midrapidity ($|\eta < 1|$) was studied in p–Pb collisions at $\sqrt{s_{\rm NN}} = 8.16$ TeV using a multi-phase transport model (AMPT). By implementing an additional heavy quark-antiquark pair production trigger in the AMPT, we obtained a simultaneously description of the $p_{\rm T}$ spectrum and v_2 of D^0 meson. Then the predictions for the v_2 of charm hadrons including D^+, D_s^+ and Λ_c^+ in p–Pb collisions are provided for the first time. We found that the v_2 of open-charm hadron follows the number-of-constituent-quark (NCQ) scaling in high-multiplicity p–Pb collisions, and is significantly affected by the parton scattering process. These findings further demonstrate the importance of partonic degrees of freedom in small collision systems for heavy flavors, and provide referential value for future measurements of azimuthal anisotropy at the LHC energies.

Keywords: quark-gluon plasma, heavy flavor, azimuthal anisotropy, small collision systems, LHC

I. INTRODUCTION

The heavy-ion collisions at ultra-relativistic energies aim to investigate the properties of nuclear matter under extremely high temperatures and energy densities, known as the quark-gluon plasma (QGP) [1, 2]. The QGP exhibits the behavior of a nearly perfect fluid, with a low shear viscosity to entropy density ratio [3, 4], η/s. The study of the azimuthal anisotropy of final-state particles produced in heavy-ion collisions is a significant approach to constrain the transport properties of the QGP [5, 6]. This anisotropy is characterised in terms of the Fourier coefficients $v_n =< \cos[n(\varphi - \Psi_n)] >$, where the φ is the azimuthal angle of the final-state particle angle and Ψ_n is the symmetry-plane angle in the collision for the n-th harmonic [7, 8]. The second order coefficient, v_2, referred to the elliptic flow, is connected to the almond-shaped overlap area formed by colliding nuclei and, as a result, constitutes the primary source of anisotropy in non-central collisions.

Heavy quarks (charm and beauty) predominantly originate from initial hard-scattering processes characterised by timescales shorter than the QGP formation time, typically around 0.1∼1 fm/c, therefore they experience the whole evolution of the QGP, and interact with the constituents of QGP medium [9–11]. Such interactions accompanied by the energy loss lead to a strong modification of the open heavy-flavor hadron (i.e. mesons and baryons that carry one charm or bottom quark/anti-quark) yield in heavy-ion collisions with respect to pp collisions, which are widely observed in experiments from the Relativistic Heavy Ion Collider (RHIC) and the Large Hadron Collider (LHC) [12–15]. Further understanding to the interactions of heavy quarks with QGP medium can be gained by analyzing the elliptic flow, v_2 of open heavy-flavour hadrons in heavy-ion collisions. Recent measurements at RHIC [16, 17] and the LHC [18, 19] show that the non-strange D-meson v_2 at low $p_{\rm T}$ is lower than that of pions and protons, following the the hypothesis of a mass hierarchy. It indicates that the charm quarks participate in the collective expansion of the medium, as well as undergo recombination with flowing light quarks. Additionally, studying the v_2 of D mesons with strange-quark content (D_s) is also very interesting as it allows further investigation into the effect of charm quark hadronization on D-meson flow [19].

In recent years, the flow-like phenomena of heavy flavors are also observed in small collision systems. The first measurement of elliptic azimuthal anisotropies for prompt D^0 mesons performed by the CMS collaboration, indicates that the D^0 v_2 has a sizable value and is lower than light-flavor hadron results [20]. The ALICE collaboration has measured a significant v_2 of electrons and muons from heavy-flavor hadron decays in p–Pb collisions at mid and forward rapidities [21–23]. However the origin of such collectivity of heavy flavors in small collision systems is still debated. In

* Correspondence email address: chaoz@whut.edu.cn
† Correspondence email address: wanrz@wtu.edu.cn

hydrodynamics-based models, a sizable v_2 is expected to result from significant interactions between charm quarks and the QGP medium, but it also generates the suppression of particle yield at high $p_{\rm T}$, in contradistinction to the observed unity of $R_{\rm pPb}$ for charm hadrons [24–26]. The Color-Glass Condensate (CGC) calculations within the dilute-dense formalism considering the initial-state effect can reproduce the v_2 of non-prompt D mesons well in p–Pb collisions at mid rapidity [27, 28], but it overestimates the data at forward rapidity [23]. Recent studies about resolving the $R_{\rm pA}$ and v_2 puzzle of D^0 mesons with the transport models demonstrates the importance of both parton interactions and the Cronin effect in the high-multiplicity p–Pb collisions [29].

In addition, an approximate number-of-constituent-quark (NCQ) scaling of v_2 for light flavor hadrons was observed in high-multiplicity p–Pb collisions by the ALICE and CMS collaborations [20, 30]. It triggered the discussions about the existence of partonic degree of freedom in small systems. The viscous hydrodynamics combined with the linearized Boltzmann transport (LBT) model, including various hadronization mechanisms can well describe the identified particle v_2 at intermediate $p_{\rm T}$ [31]. Our previous studies [32] based on a multi-phase transport (AMPT) model demonstrated that the parton scatterings plus quark coalescence can also reproduce the NCQ scaling behavior for light flavor hadrons. However similar studies were still missing for heavy flavors in p–Pb collisions. Probing the partonic collectivity for heavy quarks is crucial in searching possible formation of the QGP in the small systems at the LHC energies.

In this work, we incorporate an additional charm quark–antiquark($c\bar{c}$) pair trigger in the AMPT model to simultaneously describe the v_2 and $p_{\rm T}$ spectrum of D^0 mesons in high-multiplicity p–Pb collisions. The first study of the NCQ scaling of v_2 for open-charm hadrons, including D^0, D^+, D_s^+ and Λ_c^+, is performed over the transverse momentum region from 0 to 10 GeV/c. We also investigate how the parton cascade mechanism implemented in AMPT affect elliptic anisotropy of charm hadrons in small collision systems.

II. THE HEAVY-FLAVOR TRIGGERED AMPT MODEL

In this analysis, we employed the string melting version of the AMPT model (v2.26t9b, available online) [33], which has been demonstrated to successfully describe numerous observables at both RHIC and LHC energies. The AMPT is a hybrid framework that include four main processes: initial conditions, parton scatterings, hadronization and hadronic rescatterings. The initial conditions are handled by the heavy-ion jet interaction generator (HIJING) two-component model [34], which explicates particle production in the context of both a soft and a hard component. The soft component is modeled by the formation of excited strings in non-perturbative processes, while the hard component involves the production of minijets from hard processes. In these hard processes, hard partons are produced with a momentum transfer larger than the cutoff momentum p_0, to regulates the total minijet production cross section, which can be expressed as:

$$\frac{{\rm d}\sigma^{cd}}{{\rm d}p_{\rm T}^2{\rm d}y_1{\rm d}y_2} = K\sum_{a,b} x_1 f_a(x_1,p_{\rm T}^2) x_2 f_b(x_2,p_{\rm T}^2)\frac{d\sigma^{ab\to Q\bar{Q}}}{d\hat{t}} \tag{1}$$

where y_1 and y_2 represent the rapidities of produced partons, f_a and f_b are the parton distribution function of parton type a and b in a nucleon, and $\sigma^{ab\to cd}$ is the cross section for parton a and b to produce the minijets c and d. Then the total minijet cross section can be written as

$$\sigma_{\rm jet} = \sum_{c,d}\frac{1}{1+\delta_{cd}}\int_{p_0^2}^{s/4} dp_{\rm T}^2 dy_1 dy_2 \frac{d\sigma^{cd}}{dp_T^2 dy_1 dy_2} \tag{2}$$

The differential cross section of heavy-quark pair in HIJING is evaluated by the perturbative quantum chromodynamics (pQCD) at leading order, which has the same form as Eq. 1:

$$\frac{{\rm d}\sigma^{Q\bar{Q}}}{{\rm d}p_{\rm T}^2{\rm d}y_1{\rm d}y_2} = K\sum_{a,b} x_1 f_a(x_1,p_{\rm T}^2) x_2 f_b(x_2,p_{\rm T}^2)\frac{d\sigma^{ab\to Q\bar{Q}}}{d\hat{t}} \tag{3}$$

where the same minimum transverse momentum cut p_0 is used in calculating $\sigma^{Q\bar{Q}}$ as done in Eq.2 for $\sigma_{\rm jet}$. However, as described in [35], the heavy quark has large mass, which can naturally regulates the heavy quark total cross section. Therefore, the p_0 cutoff may result in an additional suppression of heavy quark production. For this issue, we employed an approach involving the introduction of $c\bar{c}$ trigger [36–38] to significantly enhance the production rate of heavy quarks. This trigger algorithm was initially developed to produce dijet event in HIJING model [34], then was extended to trigger $c\bar{c}$ pair productions ($q+\bar{q}\to Q+\bar{Q}, g+g\to Q+\bar{Q}$). This approach has been demonstrated to be analogous to a recently proposed extended AMPT version [35, 38], where the p_0 cutoff was removed for heavy quark production.

In the string melting mechanism, the produced light and heavy quarks are converted into primordial hadrons based on the Lund fragmentation [33]. Then these primordial hadrons are converted into partons according to their flavor and spin structures, thus form a dense partonic matter. The evolution of the partonic matter was simulated using Zhang's Parton Cascade (ZPC)

model [39]. This model incorporates two-body elastic scatterings with a cross-section defined by the simplified equation below:

$$\sigma_{gg} \simeq \frac{9\pi\alpha_s^2}{2\mu^2}, \tag{4}$$

where the α_s is the strong coupling constant, and μ is the Debye screening mass. In this study, the α_s was set to 0.33, and we adjusted the parton scattering cross section σ ($\sigma = 0, 0.2, 0.5$mb) by varying the parameter μ. After the partons stop scattering, the nearest two (or three) quarks are combined into mesons (or baryons) with a quark coalescence model. The subsequent hadronic rescattering processes are described by an extended relativistic transport (ART) model [40] including both elastic and inelastic scatterings for baryon-baryon, baryon-meson, and meson-meson interactions.

III. RESULTS AND DISCUSSIONS

Figure 1 shows the $p_{\rm T}$ spectrum of open charm mesons (D^0, D^+, D_s) and baryons (Λ_c^+) measured by the ALICE [41, 42] in p–Pb collisions at 5.02 TeV and the comparisons with AMPT model calculations. Following the previous studies [29], the parton cascade cross section in the AMPT is set to 0.5 mb. One can see that the calculations without $c\bar{c}$ trigger (labeled as "AMPT normal") underestimate all charm hadron yields. The violet full triangles represent the $p_{\rm T}$ spectrum obtained in the AMPT model with an additional $c\bar{c}$ trigger (labeled as "AMPT w/ c-trig."), which significantly enhanced the $p_{\rm T}$ spectrum compared to the "AMPT normal". The results from "AMPT w/ c-trig." provide a reasonable description of D^0 spectrum, but overestimate the D^0 and D_s^+ spectrum. Since the $p_{\rm T}$ slops of these two AMPT calculations are same, we build two sets of new event samples to describe the D^+ and D_s spectrum with the fractions determined by the data. The green lines shown in Fig 1 represent the results from such new event samples, with 60% of the "AMPT normal" plus 40% of the "AMPT w/ c-trig." for D^+ and 40% of the "AMPT normal" plus 60% of the "AMPT w/ c-trig." for D_s^+. In addition, the data of Λ_c^+ $p_{\rm T}$ spectrum is still underestimated by the model calculations even the $c\bar{c}$ trigger is implemented. It may be ascribed to the coalescence model implemented in current AMPT version, where the baryons are formed only after the formation of meson by simply combining three nearest partons regardless of the relative momentum among the coalescing partons [43]. In the following, the two new mixed samples are used to investigate the v_2 of D^+ and D_s^+, and the event sample from "AMPT w/ c-trig." is used for D^0 and Λ_c^+.

In order to directly compare the calculations of v_2 with the data, we exactly follow the two-particle correlation method employed by the CMS experiments [20]. The identified particles within the rapidity range -1.46 $< y_{\rm cm} <$ 0.54 are regarded as the particle of interest (denoted "poi" particles), then they are correlated with the reference charged particles with $0.3 < p_{\rm T} < 3$ GeV/c in -2.4 $< \eta <$ 2.4 (denoted "ref" particles). The azimuthal

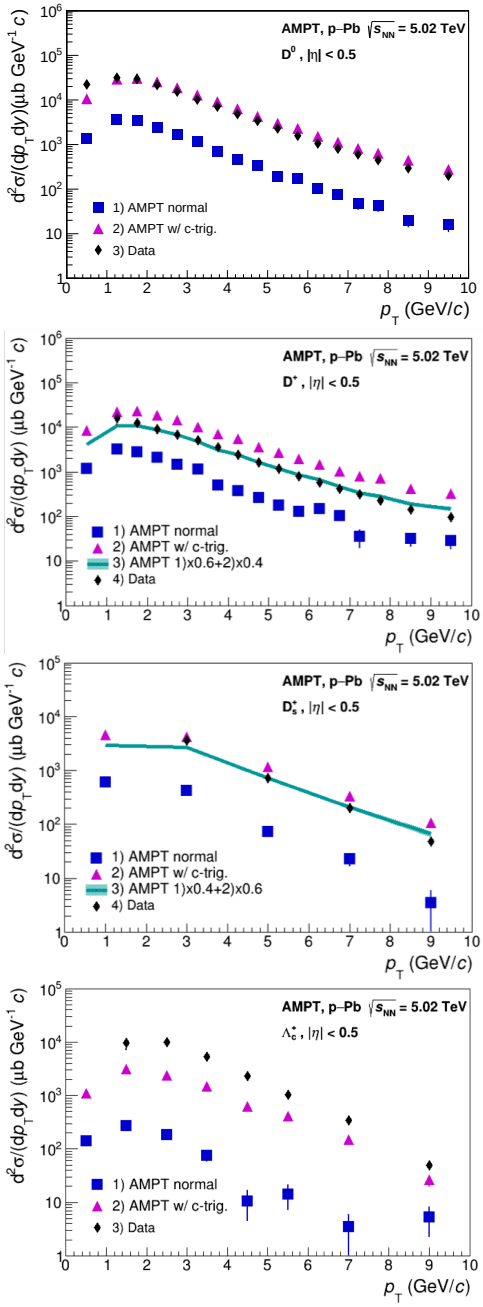


Figure 1. (Color online) The cross section of D^0, D^+, D_s^+ and Λ_c^+ as a function of $p_{\rm T}$ in p–Pb collisions at 5.02 TeV obtained from AMPT model calculations for different configurations, is compared to the ALICE measurement [41].

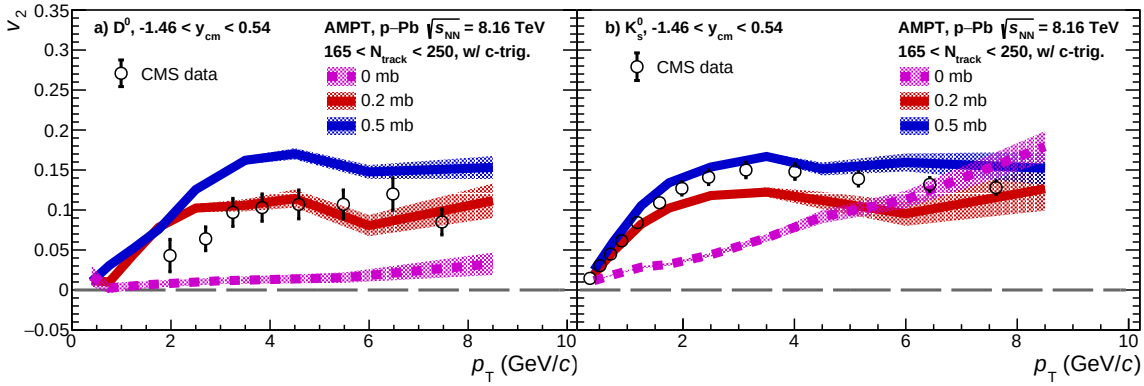


Figure 2. (Color online) The v_2 of K_s^0 and D^0 as a function of $p_{\rm T}$ obtained from the AMPT model calculations with charm quark-anitquark trigger. The results with three set of parton cross section values are shown.

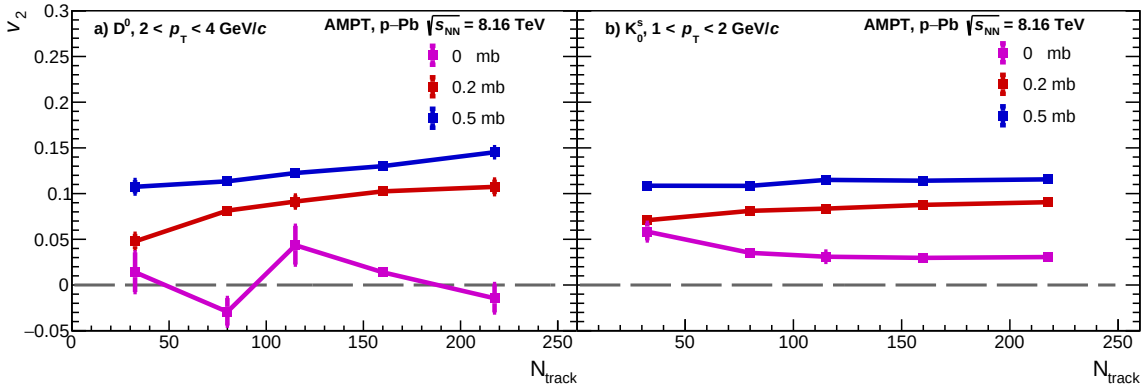


Figure 3. (Color online) The v_2 of K_s^0 and D^0 as a function of $N_{\rm track}$ obtained from the AMPT model calculations with charm quark-anitquark trigger. The results with three set of parton cross section values are shown.

correlation distribution of these two emission particles can be expanded in the Fourier series as follows:

$$\frac{{\rm d}N^{\rm pair}}{d\Delta\varphi}(\Delta\eta > 1) \propto 1 + 2\sum_{n=1}^{\infty} V_{n\Delta}({\rm poi}, {\rm ref}){\rm cos}[n(\Delta\varphi)], \tag{5}$$

where $\Delta\eta$ and $\Delta\varphi$ are the differences in η and φ of each particle pair, and the $|\Delta\eta| > 1$ is applied to suppress the nonflow contribution. Assuming factorization of the Fourier coefficients, the v_2 of the particle of interest can be obtained by

$$v_2({\rm poi}) = \frac{V_{2\Delta}({\rm poi}, {\rm ref})}{\sqrt{V_{2\Delta}({\rm ref}, {\rm ref})}}, \tag{6}$$

Figure 2(a) and Fig 2(b) show the $p_{\rm T}$-differential v_2 of D^0 and K_s^0 in p–Pb collisions at 8.16 TeV obtained from the AMPT calculations with the charm-quark trigger, and the comparisons with the CMS data [20]. The high multiplicity events within $165 < N_{\rm track} < 250$ are selected[1], where $N_{\rm track}$ is the number of charged particles with $p_{\rm T} > 0.4$ GeV/c within $|\eta| < 2.4$. Three sets of parton cross section values are tested in the model calculations. One can see that the v_2 of K_s^0 and D^0 has similar $p_{\rm T}$ trend and magnitude with the same and nonzero cross section value (σ= 0.2, 0.5 mb). The D^0 v_2 obtained with a parton cross section of 0.2 mb provide a good description of data for $p_{\rm T} > 3$ GeV/c while the results from 0.5 mb is systematically higher than the data. For K_s^0, the v_2 from 0.5 mb is closer to data compared to that from $\sigma = 0.2$ mb. It suggests that the scattering cross section among light quarks is different from that between charm quark and light quarks, and they may need to be determined from data separately,

[1] Note that it is a slightly looser selection compared to the cut $185 < N_{\rm track} < 250$ used in data [20], however its effect on the v_2 is negligible, as discussed in Fig. 3

as mentioned in an improved AMPT version proposed in [29]. In addition, we calculate the v_2 of D^0 and K^0_s when the parton scattering process is turn off (σ = 0 mb), as the dash lines shown in Fig. 2. A very small and finite value is obtained for the D^0 v_2, while the K^0_s v_2 is significant and increase with the increasing $p_{\rm T}$. Similar phenomenon was observed in previous v_2 analysis in quark level[29]. It indicates that the charm hadron v_2 is mostly generated from the parton scatterings, while for the light flavor hadron, the contribution from the initial state correlation before parton scattering process (or nonflow) is not negligible especially at high $p_{\rm T}$ region.

Figure 3(a) and Figure 3(b) present the $p_{\rm T}$-integrated v_2 of K^0_s and D^0 as a function of $N_{\rm track}$. When the implemented cross section value is nonzero (σ=0.2, 0.5 mb), the v_2 values slightly increase from low-multiplicity to high-multiplicity events, indicating the larger elliptic anisotropy in more central collisions for both K^0_s and D^0. Same conclusions for $K^{\pm}$, $\pi^{\pm}$ and proton were obtained in our previous studies [32], reflecting the changing dynamic conditions and particle production mechanisms in different collision zones. On the other hand, when we exclude the parton scattering process (σ = 0 mb), the v_2 of K^0_s decreases with the increasing $N_{\rm track}$, while the D^0 v_2 fluctuates around 0. It is consistent with our findings about the $p_{\rm T}$-differential v_2 results in Fig. 2, which hints a larger nonflow contribution to the v_2 of light quarks at low-multiplicity events compared to charm quarks.

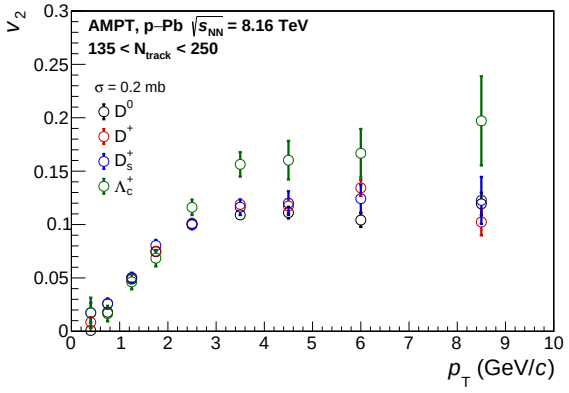


Figure 4. (Color online) The v_2 of D^0, D^+, D^+_s and Λ^+_c as a function of $p_{\rm T}$ obtained from the AMPT model calculations with charm quark-anitquark trigger.

To further investigate the elliptic anisotropy of open-charm hadrons in small collision systems, we extend the calculations of v_2 to D^+_s, D^+ and Λ^+_c, as shown in the Fig. 4. In order to reproduce the data as possible, the AMPT model with charm quark-antiquark trigger is used for the D^0 and Λ^+_c and mixed event samples introduced in Fig. 1 are used for D^+ and D^+_s, as discussed above. The σ = 0.2 mb is used for all charm-hadron species. One can see that the v_2 of D^0, D^+ and D^+_s are consistent within uncertainties, which is similar to the conclusions in heavy-ion collisions [18]. The v_2 of Λ^+_c is larger than that of charm meson for $p_{\rm T} > 2$ GeV/c, indicating that meson-baryon grouping behaviors also present in heavy flavor sector. Future measurements about the elliptic flow of charm baryons can provide more constrain on our calculations.

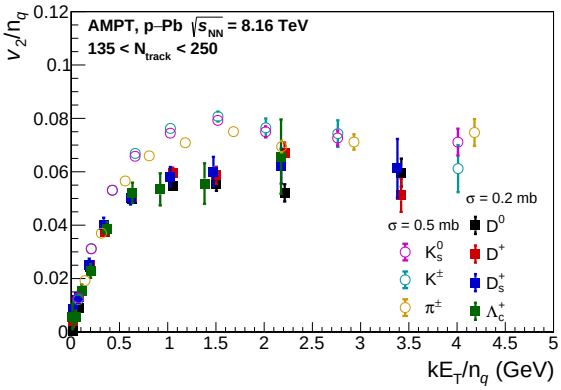


Figure 5. (Color online) The $kE_{\rm T}$-differential v_2 of D^0, D^+, D^+_s and Λ^+_c scaled by the the number of constituent quark. The comparison to the results of K^0_s, $K^{\pm}$ and $\pi^{\pm}$ is also shown.

As discussed in Ref. [44], the observed meson-baryon particle type grouping for flow measurements in heavy-ion collisions indicates the collective behavior at the partonic level, which can be further studied by means of the NCQ scaling technique proposed in Ref. [45]. In this work, we investigate the NCQ scaling behavior of open-charm hadrons in p–Pb collisions for the first time. Figure 5 (left) shows the $v_2/n_{\rm q}$ as the function of $kE_{\rm T}/n_{\rm q}$ for D^0, D^+, D^+_s and Λ^+_c, and the comparison with the results of light-flavor hadron including K^0_s, $\pi^{\pm}$, $K^{\pm}$ is also shown. The n_q is the number of constituent quark in mesons (n_q = 2) and baryons (n_q = 3), and the $p_{\rm T}$ is replaced by the transverse kinetic energy $kE_{\rm T}$ ($kE_{\rm T} = m_{\rm T} - m_0 = \sqrt{p_{\rm T}^2 + m_0^2} - m_0$) in consideration of the different mass of hadrons. We found that all charm hadrons show a set of similar v_2 values after the NCQ scaling, confirming that the quark degree of freedom in flowing matter can also be probed for heavy quark in the transport model. On the other hand, the v_2 after the NCQ scaling for K^0_s obtained with the cross section σ = 0.5 mb is compatible to results of $\pi^{\pm}$ and $K^{\pm}$, which shows a larger value than charm hadrons. It suggests that the weaker collective behavior of charm quark compared to light quarks is mainly attributed to their different parton scatterings.

IV. SUMMARY

In summary, we systematically studied the elliptic anisotropy of open charm hadrons in p–Pb collisions at $\sqrt{s_{\rm NN}}$ 8.16 TeV by means of introducing an additional charm quark-antiquark pair production trigger in the AMPT model. The implementation of such trigger provides an efficient way to simultaneously describe the $p_{\rm T}$ spectrum and v_2 of D^0. In addition, we provided new predictions for the v_2 of other charm hadrons including D^+, D_s^+ and Λ_c^+ in small collision systems that that can be tested in future experimental measurements. We argue that, the v_2 of open-charm hadron follows the NCQ scaling in high-multiplicity p–Pb collisions at LHC energies with a proper parton cross section value. These findings demonstrate the importance of parton scatterings for generating the collectivity of heavy quark in p–Pb collisions. Future studies about more types of heavy-flavor hadron, including charmonium and bottom hadron, can provide further understandings on the transport properties of heavy quark, then help to search for the possible formation of the hot medium in small systems.

V. ACKNOWLEDGEMENT

This work was supported by Natural Science Foundation of Hubei Provincial Education Department under Grant (Q20131603) and Key Laboratory of Quark and Lepton Physics (MOE) in Central China Normal University under Grant (QLPL2022P01, QLPL202106).

[1] E. V. Shuryak, Phys. Lett. B **78**, 150 (1978).
[2] E. V. Shuryak, Phys. Rept. **61**, 71 (1980).
[3] G.-Y. Qin, H. Petersen, S. A. Bass, and B. Muller, Phys. Rev. C **82**, 064903 (2010).
[4] D. Teaney and L. Yan, Phys. Rev. C **83**, 064904 (2011).
[5] J.-Y. Ollitrault, Phys. Rev. D **46**, 229 (1992).
[6] S. A. Voloshin, Nucl. Phys. A **827**, 377C (2009).
[7] S. Voloshin and Y. Zhang, Z. Phys. C **70**, 665 (1996).
[8] A. M. Poskanzer and S. A. Voloshin, Phys. Rev. C **58**, 1671 (1998).
[9] H. van Hees, V. Greco, and R. Rapp, Phys. Rev. C **73**, 034913 (2006).
[10] N. Brambilla *et al.*, Eur. Phys. J. C **71**, 1534 (2011).
[11] A. Andronic *et al.*, Eur. Phys. J. C **76**, 107 (2016).
[12] M. Gyulassy and L. McLerran, Nucl. Phys. A **750**, 30 (2005).
[13] J. Adams *et al.* (STAR), Nucl. Phys. A **757**, 102 (2005).
[14] K. Adcox *et al.* (PHENIX), Nucl. Phys. A **757**, 184 (2005).
[15] B. Muller, J. Schukraft, and B. Wyslouch, Ann. Rev. Nucl. Part. Sci. **62**, 361 (2012).
[16] L. Adamczyk *et al.* (STAR), Phys. Rev. Lett. **118**, 212301 (2017).
[17] (2023), 10.1016/j.physletb.2023.138071.
[18] S. Acharya *et al.* (ALICE), Phys. Lett. B **813**, 136054 (2021).
[19] F. Grosa (ALICE), EPJ Web Conf. **171**, 18007 (2018).
[20] A. M. Sirunyan *et al.* (CMS), Phys. Rev. Lett. **121**, 082301 (2018).
[21] S. Acharya *et al.* (ALICE), Phys. Rev. Lett. **122**, 072301 (2019).
[22] J. Adam *et al.* (ALICE), Phys. Lett. B **753**, 126 (2016).
[23] S. Acharya *et al.* (ALICE), Phys. Lett. B **846**, 137782 (2023).
[24] X. Du and R. Rapp, JHEP **03**, 015 (2019).
[25] S. Cao and X.-N. Wang, Rept. Prog. Phys. **84**, 024301 (2021).
[26] Y. Xu, S. Cao, G.-Y. Qin, W. Ke, M. Nahrgang, J. Auvinen, and S. A. Bass, Nucl. Part. Phys. Proc. **276-278**, 225 (2016).
[27] C. Zhang, C. Marquet, G.-Y. Qin, S.-Y. Wei, and B.-W. Xiao, Phys. Rev. Lett. **122**, 172302 (2019).
[28] C. Zhang, C. Marquet, G.-Y. Qin, Y. Shi, L. Wang, S.-Y. Wei, and B.-W. Xiao, Phys. Rev. D **102**, 034010 (2020).
[29] C. Zhang, L. Zheng, S. Shi, and Z.-W. Lin, Phys. Lett. B **846**, 138219 (2023).
[30] B. B. Abelev *et al.* (ALICE), Phys. Lett. B **726**, 164 (2013).
[31] W. Zhao, C. M. Ko, Y.-X. Liu, G.-Y. Qin, and H. Song, Phys. Rev. Lett. **125**, 072301 (2020).
[32] S. Tang, L. Zheng, X. Zhang, and R. Wan, (2023), 10.48550/arXiv.2303.06577.
[33] Z.-W. Lin, C. M. Ko, B.-A. Li, B. Zhang, and S. Pal, Phys. Rev. C **72**, 064901 (2005).
[34] M. Gyulassy and X.-N. Wang, Comput. Phys. Commun. **83**, 307 (1994).
[35] L. Zheng, C. Zhang, S. S. Shi, and Z. W. Lin, Phys. Rev. C **101**, 034905 (2020).
[36] H. Wang, J.-H. Chen, Y.-G. Ma, and S. Zhang, Nucl. Sci. Tech. **30**, 185 (2019).
[37] H. Wang and J.-H. Chen, Nucl. Sci. Tech. **32**, 2 (2021).
[38] H. Wang and J.-H. Chen, Nucl. Sci. Tech. **33**, 15 (2022).
[39] B. Zhang, Comput. Phys. Commun. **109**, 193 (1998).
[40] B.-A. Li and C. M. Ko, Phys. Rev. C **52**, 2037 (1995).
[41] S. Acharya *et al.* (ALICE), JHEP **12**, 092 (2019).
[42] S. Acharya *et al.* (ALICE), Phys. Rev. C **104**, 054905 (2021).
[43] Y. He and Z.-W. Lin, Phys. Rev. C **96**, 014910 (2017).
[44] S. Afanasiev *et al.* (PHENIX), Phys. Rev. Lett. **99**, 052301 (2007).
[45] D. Molnar and S. A. Voloshin, Phys. Rev. Lett. **91**, 092301 (2003).

arXiv:2403.01676v1 [hep-ph] 4 Mar 2024

Production of X_b via radiative transition of $\Upsilon(10753)$

Shi-Dong Liu,[1] Hao-Dong Cai,[1] Zu-Xin Cai,[1,2] Hong-Shuo Gao,[1] Gang Li,[1,*] Fan Wang,[1] and Ju-Jun Xie[2,3,4]

[1] *College of Physics and Engineering, Qufu Normal University, Qufu 273165, China*
[2] *Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China*
[3] *School of Nuclear Science and Technology, University of Chinese Academy of Sciences, Beijing 101408, China*
[4] *Southern Center for Nuclear-Science Theory (SCNT), Institute of Modern Physics, Chinese Academy of Sciences, Huizhou 516000, Guangdong, China*

(Dated: March 5, 2024)

We studied the radiative transitions between the $\Upsilon(10753)$, the S-D mixed state of the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$, and the X_b, the heavy quark flavor symmetry counterpart of the $X(3782)$ in the bottomonium sector. The radiative transition was assumed to occur through the intermediate bottom mesons, including P-wave $B_1^{(\prime)}$ mesons as well as the S-wave $B^{(*)}$ ones. The consideration of the $B_1^{(\prime)}$ mesons leads to the couplings to be in S-wave, and hence enhances the contributions of the intermediate meson loops. The radiative decay width for the $\Upsilon(10753) \to \gamma X_b$ is predicted to be order of 10 keV, corresponding to a branching fraction of 10^{-4}. Based on the theoretical results, we strongly suggest to search for the X_b in the $e^+e^- \to \gamma X_b$ with $X_b \to \pi\pi\chi_{b1}$ near $\sqrt{s} = 10.754$ GeV, and it is hoped that the calculations here could be tested by the future Belle II experiments.

I. INTRODUCTION

Studies of exotic states have received considerable attention since 2003 when Belle collaboration observed the $X(3872)$ state in the $\pi^+\pi^- J/\psi$ invariant mass spectrum [1]. The candidates for exotic states are usually referred to collectively as XYZ states, representing a group of particles that cannot be easily explained by the traditional quark model. Understanding the behavior and characteristics of these XYZ states is crucial for advancing our knowledge of the strong force and its role in the structure of matter. With the development of the experimental techniques, a fair amount of XYZ states have been directly observed, especially in the charmonium sector, for instance, the well-studied $X(3872)$ [1–6] and $Z_c(3900)$ [7–10]. However, in the bottomonium sector, there are only two exotic states, named as $Z_b(10610)$ and $Z_b(10650)$ [11–14]. More details on theoretical and experimental studies of the XYZ states can be found in the reviews [15–21].

The $X(3872)$ is the most well-studied exotic state either from the experimental or theoretical points of view, so that its properties are widely used as inputs to predict new hadronic states in the heavy quark sector. It has been well known that the $X(3872)$ mass is (3871.65 ± 0.06) MeV, extraordinarily close to the $D^{*0}\bar{D}^0$ threshold (3871.69 MeV), and its quantum numbers $J^{PC} = 1^{++}$ [22]. Employing the heavy quark flavor symmetry of the c and b quarks, it's natural to expect that an analogous state with a mass near the $B^{*0}\bar{B}^0$ threshold (10.604 GeV) exists in the bottomonium sector, with the same quantum numbers as the $X(3872)$ or other commonality. Such state is usually named for short as X_b [23]. The possible value of the X_b mass has been calculated in the framework of tetraquark model [24–26] and using the mesonic molecule interpretation [27, 28], which lies between 10.5 and 10.7 GeV.

By following the successful way to understanding the $X(3872)$, the productions and decays of the X_b have been investigated extensively. The partial decay width for the process $X_b \to \omega\Upsilon(1S)$, for instance, was theoretically predicted to be tens of keVs [29], under the interpretation that the X_b is a $B^*\bar{B}$ bound state. However, searching for the X_b in various experiments appears to be fruitless [15] (and references therein). Clear signal of the X_b was not observed in the $\pi^+\pi^-\Upsilon(1S)$ [30, 31] or $\pi^+\pi^-\pi^0\Upsilon(1S)$ [32, 33] invariant mass distribution, while the $X(3872)$ was clearly observed in the similar distributions of pions plus the J/ψ. Therefore, to searching for the X_b, we need to find other possible channels, for instance the $X_b \to \pi\pi\chi_{bJ}$, of which the partial width has been predicted to reach tens of keVs [34].

The production of the $X(3872)$ in the radiative decay of higher charmonium(-like) was firstly observed by the BESIII Collaboration in 2014 [35]. This observation is consistent with the early theoretical prediction for the radiative transition process between the 1^{--} charmonium states and the $X(3872)$, where the 1^{--} should be a D-wave charmonium or a $D_1\bar{D}$ molecule (the $Y(4260)$, for instance). Thanks to the high mass of the X_b, its production via the radiative decays of higher bottomonia, especially the 1^{--} states, is greatly expected. However, the theoretical production ratio of the X_b as a $B^*\bar{B}$ molecule in the processes $\Upsilon(5S, 6S) \to \gamma X_b$ is only of the order 10^{-6} [36], and thus the experimental observation seems difficult. Imitating the case of the $X(3872)$, the newly observed $\Upsilon(10753)$ [33, 37] is favored for the X_b production since the $\Upsilon(10753)$ is likely to be S-D mixed state of the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$ [38–42].

In this work, we calculated the production of the X_b in the radiative process $\Upsilon(10753) \to \gamma X_b$ using a nonrelativistic effective field theory. We regarded the X_b as the $B^*\bar{B}$ molecule and the $\Upsilon(10753)$ as the $4S$-$3D$ mixed state. Moreover, the radiative transition was assumed

* gli@qfnu.edu.cn

to occur through the intermediate bottom meson loops, including the P-wave $B_1^{(\prime)}$ mesons as well as the S-wave $B^{(*)}$ ones. The loops including the $B_1^{(\prime)}$ mesons are enhanced due to the S-wave couplings.

The rest of the paper is organized as follows. In Sec. II, we present the theoretical framework used in this work. Then in Sec. III the numerical results are presented, and a brief summary is given in Sec. IV.

II. THEORETICAL CONSIDERATION

Similar to the previous works as done in Refs. [38–41], we interpret the $\Upsilon(10753)$ as a $4S$-$3D$ mixed state, then the wave function of the $\Upsilon(10753)$ is written as

$$\tilde{\Upsilon}(10753) = \tilde{\Upsilon}(4S)\sin\theta + \tilde{\Upsilon}_1(3\,^3D_1)\cos\theta\,, \qquad (1)$$

where θ is a mixing angle to describe the proportion of the partial waves. $\tilde{\Upsilon}(4S)$ and $\tilde{\Upsilon}_1(3\,^3D_1)$ describe the wave functions of the pure $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$ states, respectively. The $\Upsilon(10580)$ was usually regraded as the $4S$ state [22]. When taking into account the S-D mixing, the $\Upsilon(10580)$ could be interpreted as another $4S$-$3D$ mixture, and accordingly has the wave function

$$\tilde{\Upsilon}(10580) = \tilde{\Upsilon}(4S)\cos\theta - \tilde{\Upsilon}(3\,^3D_1)\sin\theta\,. \qquad (2)$$

The mixing angle θ can be obtained by fitting the well-measured dielectron decay width of the $\Upsilon(10580)$ [38, 41, 43], which depends on its wave function and mass. The estimation yields the mixing angle $\theta = 23.4^\circ \sim 36.1^\circ$. Such mixing angles indicate that the $\Upsilon_1(3\,^3D_1)$ is the dominant component to form the $\Upsilon(10753)$ with proportion of about $65\% \sim 84\%$. In the following, in order to predict the decay width of the $\Upsilon(10753) \to \gamma X_b$, we shall adopt an angle of $\theta = 33^\circ$ for the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$ mixing with masses of 10.612 GeV and 10.675 GeV, respectively, which was predicted using the modified Godfrey-Isgur model [44].

A. Intermediate Bottom Meson Loops

The bottomonia $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$ are both above the open-bottom threshold so that they are expected to dominantly decay into the bottom-antibottom meson pair, and then the pair could couple further to the final states by exchanging a proper bottom meson. This process is widely described by the triangle meson loop mechanism, which proves to be important in the decays and productions of many heavy quarkonia and exotic states. In the case of the radiative transition of the $\Upsilon(10753)$ to X_b, the loops made of the S-wave bottom mesons with the quantum numbers $s_l^P = 1/2^+$ of the light degrees of freedom are shown in Fig. 1.

In view of the quantum numbers $J^{PC} = 1^{--}$ for the initial bottomonia and final photon, they both couple to

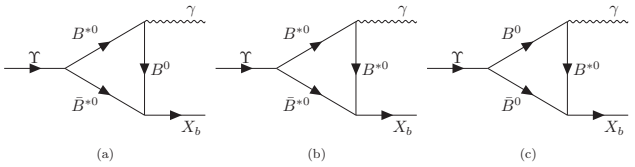


FIG. 1. Triangle Feynman diagrams for the radiative process $\Upsilon(10753) \to \gamma X_b$ via the S-wave bottom meson loops. Each diagram has its corresponding charged one. The symbol Υ stands for the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$.

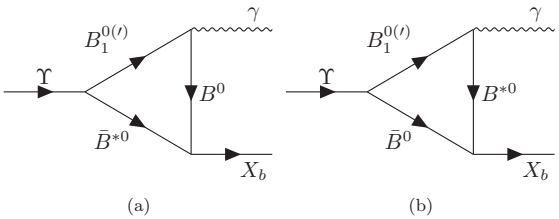


FIG. 2. Feynman diagrams for the $\Upsilon(10753) \to \gamma X_b$ via the loops associated with the P-wave bottom mesons $B_1^{(\prime)}$.

the S-wave $B^{(*)}$ mesons in a P wave. The other coupling for the X_b of $J^{PC} = 1^{++}$ to the S-wave $B^{(*)}$ mesons occurs in an S wave. When we replace the intermediate meson that connects the initial state $\Upsilon(4S)$ [$\Upsilon_1(3\,^3D_1)$] and the photon by a P-wave bottom mesons B_1' with $s_l^P = 1/2^+$ [B_1 with $s_l^P = 3/2^+$], as shown in Fig. 2, all the couplings are then allowed to be in an S wave. Near threshold, the S-wave contribution is usually more important than that from the P-wave. Thus, in comparison with the loops in Fig. 1, the contributions of the loops in Fig. 2 to the radiative transition $\Upsilon(10753) \to \gamma X_b$ are likely to be more important. This importance will be qualitatively analyzed in terms of the power counting and verified by the numerical calculations.

B. Effective Lagrangians

Similar to the case of the $X(3872)$ [45], we also consider the X_b as a pure molecule of the $B\bar{B}^* + \text{c.c}$, of which the neutral and charged components are assumed to be of equal proportion,

$$|X_b\rangle = \frac{1}{2}\left(|B^0\bar{B}^{*0}\rangle + |\bar{B}^0 B^{*0}\rangle + |B^+B^{*-}\rangle + |B^-B^{*+}\rangle\right). \qquad (3)$$

The effective coupling of the X_b to a pair of bottom and antibottom mesons is then given by

$$\begin{aligned}\mathcal{L}_X =& \frac{1}{2}g_X^0 X_b^{\dagger i}(B^0\bar{B}^{*0i} + \bar{B}^0 B^{*0i}) \\ &+ \frac{1}{2}g_X^c X_b^{\dagger i}(B^+B^{*-i} + B^-B^{*+i}) + \text{H.c.}\,, \end{aligned} \qquad (4)$$

where the constants g_X's describe the coupling strength of the X_b to the bottom meson pairs. Here and later

the symbols with (without) the dagger index represent the outgoing (incoming) fields relative to the coupling vertices.

The coupling, g_X, could be extracted from the binding energy (ϵ_X) of the X_b with respect to the mass threshold of its two components [46]:

$$g_X = \left(\frac{16\pi}{\mu_B}\sqrt{\frac{2\epsilon_X}{\mu_B}}\right)^{1/2}, \tag{5}$$

where $\mu_B = m_B m_{B^*}/(m_B+m_{B^*})$ and $\epsilon_B = m_B+m_{B^*}-m_{X_b}$ with the m's being the masses of the particles indicated by the subscripts. Given the quite small mass difference of the neutral and charged bottom mesons, the g_X^0 and g_X^c are taken to be equal. [1]. In addition, we adopt the binding energy ϵ_X from 2 to 100 MeV, similar to the values used in the recent work [34, 36].

Since we assume that the $\Upsilon(10753)$ is a mixture of the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$, to calculate the decay width of the $\Upsilon(10753)\to\gamma X_b$ we should also know the interactions of the S- and D-wave bottomonia with the bottom mesons. Within the framework of the nonrelativistic effective field theory, the interactions of the S-wave bottomonia and the bottom-antibottom meson pair read [46–49]

$$\mathcal{L}_S = \mathrm{i}\frac{g_S}{2}\langle \bar{H}_a^\dagger \vec{\sigma}\cdot\overleftrightarrow{\partial} H_a^\dagger J\rangle + \frac{g_S'}{2}\langle(\bar{H}_a^\dagger S_a^\dagger + \bar{S}_a^\dagger H_a^\dagger)J\rangle + \mathrm{H.c.} \tag{6}$$

Here $H_a = \vec{V}_a\cdot\vec{\sigma}+P_a$ and $S_a = \vec{V}'_{1a}\cdot\vec{\sigma}+P_{0a}$ are the spin doublets formed by the $1/2^-$ and $1/2^+$ bottom mesons. In this work, $\vec{V}_a = (B^{*0},\, B^{*+})$ and $P_a = (B^0,\, B^+)$ denoting the vector and pseudoscalar bottom mesons with $s_l^P = 1/2^-$, respectively, while $V'_{1a} = (B_1'^0,\, B_1'^+)$ and $P_{0a} = (B_0,\, B_0^+)$ for the $s_l^P = 1/2^+$ bottom mesons. Using the convention in Ref. [46], the charge conjugated fields for the heavy bottom mesons are $\bar{H}_a = -\vec{\bar{V}}\cdot\vec{\sigma}+\bar{P}_a$ and $\bar{S}_a = -\vec{\bar{V}}'_{1a}\cdot\vec{\sigma}+\bar{P}_{0a}$. The $J = \vec{\Upsilon}\cdot\vec{\sigma}+\eta_b$ represents the spin doublet of the S-wave bottomonia $\Upsilon(nS)$ and $\eta_b(nS)$. Conventionally, $A\overleftrightarrow{\partial}B \equiv A(\partial B)-(\partial A)B$, the $\vec{\sigma}$ stands for the Pauli matrices, and the subscript a is the light flavor index. After tracing operation in spinor space (indicated by $\langle\cdots\rangle$), the Lagrangian in Eq. (6) is explicitly written as

$$\begin{aligned}
\mathcal{L}_S =& \,\mathrm{i}g_S\Upsilon^i(\bar{V}_a^{\dagger j}\overleftrightarrow{\partial}^i V_a^{\dagger j} - \bar{V}_a^{\dagger j}\overleftrightarrow{\partial}^j V_a^{\dagger i} - \bar{V}_a^{\dagger i}\overleftrightarrow{\partial}^j V_a^{\dagger j})\\
&- g_S\epsilon^{ijk}\Upsilon^i(\bar{V}_a^{\dagger k}\overleftrightarrow{\partial}^j P_a^\dagger + \bar{P}_a^\dagger\overleftrightarrow{\partial}^j V_a^{\dagger k}) + \mathrm{i}g_S\Upsilon^i\bar{P}_a^\dagger\overleftrightarrow{\partial}^i P_a^\dagger\\
&+ \epsilon^{ijk}g_S\eta_b\bar{V}_a^{\dagger i}\overleftrightarrow{\partial}^j V_a^{\dagger k} + \mathrm{i}g_S\eta_b(\bar{P}_a^\dagger\overleftrightarrow{\partial}^i V_a^{\dagger i} - \bar{V}_a^{\dagger i}\overleftrightarrow{\partial}^i P_a^\dagger)\\
&+ g_S'\eta_b(\bar{P}_{0a}^\dagger P_a^\dagger + \bar{P}_a^\dagger P_{0a}^\dagger) - g_S'\eta_b(\bar{V}'^{\dagger i}_{1a}V_a^{\dagger i} + \bar{V}_a^{\dagger i}V'^{\dagger i}_{1a})\\
&+ g_S'\Upsilon^i(\bar{P}_a^\dagger V'^{\dagger i}_{1a} + \bar{P}_{0a}^\dagger V_a^{\dagger i} - \bar{V}_a^{\dagger i}P_{0a}^\dagger - \bar{V}'^{\dagger i}_{1a}P_a^\dagger)\\
&+ \mathrm{i}g_S'\epsilon^{ijk}\Upsilon^i(\bar{V}_a^{\dagger k}V'^{\dagger j}_{1a} + \bar{V}'^{\dagger k}_{1a}V_a^{\dagger j}) + \mathrm{H.c.},
\end{aligned} \tag{7}$$

[1] According to the world average masses of the B^0 and B^* [22] and the predicted mass around 10.562 GeV for the X_b [27], the relative difference of the g_X^0 and g_X^c do not exceed 0.5%.

where the coupling constants g_S and g_S' will be determined later.

The interactions of the D-wave bottomonium $\Upsilon_1(3\,^3D_1)$ with a pair of bottom and antibottom mesons are written as [46]

$$\begin{aligned}
\mathcal{L}_D =& \,\mathrm{i}\frac{g_D}{2}\langle\bar{H}_a^\dagger\sigma^i\overleftrightarrow{\partial}^j H_a^\dagger J^{ij}\rangle\\
&+ \frac{g_D'}{2}\langle(\bar{T}_a^{\dagger j}\sigma^i H_a^\dagger - \bar{H}_a^\dagger\sigma^i T_a^{\dagger j})J^{ij}\rangle + \mathrm{H.c.},
\end{aligned} \tag{8}$$

where J^{ij} represents the field for the D-wave bottomonium $\Upsilon_1(3\,^3D_1)$ in the two-component notation [46]

$$J^{ij} = \frac{\sqrt{15}}{10}(\Upsilon_1^i\sigma^j + \Upsilon_1^j\sigma^i) - \frac{1}{\sqrt{15}}\delta^{ij}\vec{\Upsilon}_1\cdot\vec{\sigma}, \tag{9}$$

and T_a^i is the field for the $s_l^P = 3/2^+$ bottom mesons [46],

$$T_a^i = \sqrt{\frac{2}{3}}V_{1a}^i + \mathrm{i}\frac{1}{\sqrt{6}}\epsilon^{ijk}V_{1a}^j\sigma^k. \tag{10}$$

It should be pointed out that for the $s_l^P = 3/2^+$ mesons with the total angular-momentum of 2 are not considered in this work. Hence, $\vec{V}_{1a} = (B_1^0,\, B_1^+)$. The antibottom meson field $\bar{T}_{1a}^i$ is described as

$$\bar{T}_a^i = \sqrt{\frac{2}{3}}\bar{V}_{1a}^i - \mathrm{i}\frac{1}{\sqrt{6}}\epsilon^{ijk}\bar{V}_{1a}^j\sigma^k. \tag{11}$$

The tracing evaluation yields the Lagrangian for the $\Upsilon_1(3\,^3D_1)$,

$$\begin{aligned}
\mathcal{L}_D =& \,\mathrm{i}g_D\frac{\sqrt{15}}{3}\Upsilon_1^i\bar{P}_a^\dagger\overleftrightarrow{\partial}^i P_a^\dagger\\
&+ g_D\frac{\sqrt{15}}{6}\epsilon^{ijk}\Upsilon_1^i(\bar{P}_a^\dagger\overleftrightarrow{\partial}^j V_a^{\dagger k} + \bar{V}_a^{\dagger k}\overleftrightarrow{\partial}^j P_a^\dagger)\\
&+ \mathrm{i}\frac{g_D}{2\sqrt{15}}\Upsilon_1^i(4\bar{V}_a^{\dagger j}\overleftrightarrow{\partial}^i V_a^{\dagger j} - \bar{V}_a^{\dagger j}\overleftrightarrow{\partial}^j V_a^{\dagger i} - \bar{V}_a^{\dagger i}\overleftrightarrow{\partial}^j V_a^{\dagger j})\\
&+ g_D'\frac{\sqrt{10}}{2}\Upsilon_1^i(\bar{V}_{1a}^{\dagger i}P_a^\dagger - \bar{P}_a^\dagger V_{1a}^{\dagger i})\\
&+ \mathrm{i}g_D'\frac{\sqrt{10}}{4}\epsilon^{ijk}\Upsilon_1^i(\bar{V}_{1a}^{\dagger k}V_a^{\dagger j} + \bar{V}_a^{\dagger k}V_{1a}^{\dagger j}) + \mathrm{H.c.}
\end{aligned} \tag{12}$$

The photonic coupling to the bottom mesons with $s_l^P = 1/2^-$ is written as [46, 50]

$$\mathcal{L}_{HH\gamma} = \frac{e\beta}{2}\langle H_a^\dagger H_b\vec{\sigma}\cdot\vec{B}Q_{ab}\rangle + \frac{eQ'}{2m_{Q'}}\langle H_a^\dagger\vec{\sigma}\cdot\vec{B}H_a\rangle, \tag{13}$$

where $B^k = \epsilon^{ijk}\partial^i A^j$ is the magnetic field, $Q_{ab} = \mathrm{diag}(-1/3,\, 2/3)$ denotes the charge matrix of the light d and u quarks, and $Q' = -1/3$ and $m_{Q'}$ stand for the b-quark charge and its mass, respectively.

In addition, the radiative transition of the $1/2^+$ and $3/2^+$ bottom mesons to the $1/2^-$ ones is described by the following Lagrangian [36, 46]

$$\mathcal{L}_{S/TH\gamma} = -\frac{\mathrm{i}e\tilde{\beta}}{2}\langle H_a^\dagger S_b\vec{\sigma}\cdot\vec{E}Q_{ba}\rangle + \langle T_a^i H_b^\dagger C_{ab}\rangle E^i \tag{14}$$

TABLE I. Coupling constants g_S and g_D (Units: GeV$^{-3/2}$) we employed in the calculations. Their estimations are based on the theoretical and experimental data in Refs. [22, 44].

g_S , g_D	$B\bar{B}$	$B\bar{B}^*$ + c.c.	$B^*\bar{B}^*$
$\Upsilon(4S)$	0.776	0.776	0.776
$\Upsilon_1(3\,^3D_1)$	0.157	0.376	1.879

where E^i is the electric field. The C_{ab} in the second term is a formalistic 2×2 diagonal matrix in the form of diag(C^d, C^u), of which the elements describe the coupling strength. Explicitly,

$$\begin{aligned}\mathcal{L}_{HH\gamma} &= \mathrm{i}e\beta Q_{ab}\partial^i A^j (V_a^{\dagger i}V_b^j - V_a^{\dagger j}V_b^i)\\ &+ e\beta Q_{ab}\epsilon^{ijk}\partial^i A^j (P_a^\dagger V_b^k + V_a^{\dagger k}P_b)\\ &+ \mathrm{i}\frac{eQ'}{m_{Q'}}\partial^i A^j (V_a^{\dagger j}V_a^i - V_a^{\dagger i}V_a^j)\\ &+ \frac{eQ'}{m_{Q'}}\epsilon^{ijk}\partial^i A^j (V_a^{\dagger k}P_a + P_a^\dagger V_a^k) \end{aligned} \tag{15}$$

and

$$\begin{aligned}\mathcal{L}_{S/TH\gamma} &= e\tilde{\beta}Q_{ab}\epsilon^{ijk}\partial^0 A^i V_a^{\dagger j}V_{1b}'^k\\ &- \mathrm{i}e\tilde{\beta}Q_{ab}\partial^0 A^i (V_a^{\dagger i}P_{0b} + P_a^\dagger V_{1b}'^i)\\ &+ \mathrm{i}\sqrt{\frac{2}{3}}C_{ab}\epsilon^{ijk}\partial^0 A^i V_{1a}^j V_b^{\dagger k}\\ &+ 2\sqrt{\frac{2}{3}}C_{ab}\partial^0 A^i P_b^\dagger V_{1a}^i \end{aligned} \tag{16}$$

III. NUMERICAL RESULTS

To proceed, we should evaluate the coupling constants g's in Eqs. (6) and (8), the parameter β in Eq. (13), and, $\tilde{\beta}$ and C_{ab} in Eq. (14). The way to evaluating the constants g_S and g_D is the same as that in our recent work [41] so that the details are not repeated. Notice that due to the factor 1/2 in Eqs. (6) and (8), the values of g_S and g_D here are twice of those in Ref. [41]. Their values are summarized in Table I.

However, the coupling constants g_S' and g_D' cannot be directly determined in terms of the way to the g_S and g_D, since the threshold of the $B^{(*)}B_1^{(\prime)}$ exceeds the masses of the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$. In order to give reasonable estimation of the g_S' and g_D', we then, considering the heavy quark symmetry, assume that the ratios g_S'/g_S and g_D'/g_D are heavy-flavor-independent, although the g's are all heavy-flavor-dependent separately. Given the predictions for the $\psi(4S)\to D_1'\bar{D}$ [51, 52], the value of g_S' for the $\psi(4S)$ varies from 0.11 to 0.42 GeV$^{-1/2}$, whereas the prediction for the $\psi(4S)\to D\bar{D}$ yields $g_S\approx 0.15$ GeV$^{-3/2}$ for the $\psi(4S)$. As a result, g_S'/g_S is in the range [0.73, 2.8] GeV. Similarly, in the case of the $\psi_1(3\,^3D_1)$, g_D is estimated to be between 0.1 and 0.3 GeV$^{-3/2}$, and g_D' ranges from 0.2 and 0.5 GeV$^{-1/2}$ according to the predictions in Refs. [51, 52]. Considering the intermediate values, it gives the ratio $g_D'/g_D\approx 1.7$ GeV.

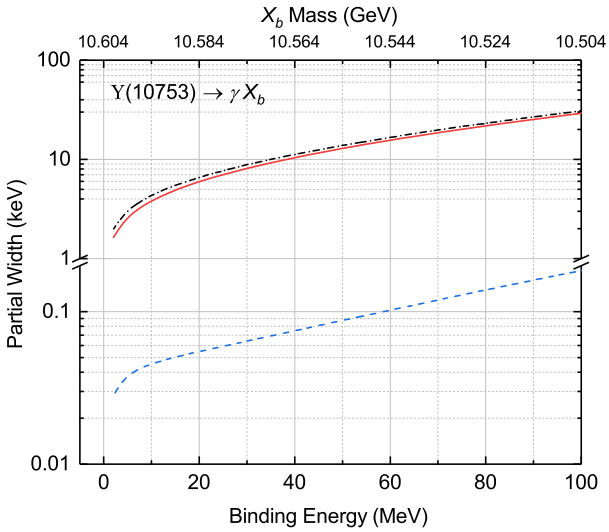


FIG. 3. Partial decay width of the radiative transition $\Upsilon(10753)\to\gamma X_b$ for different binding energies of the X_b relative to the $B^*\bar{B}$ threshold. In calculations, g_S' is taken to be $1.5g_S$ and g_D' is fixed to be $1.7g_D$. Moreover, $\tilde{\beta}=1.75$ GeV^{-1}, $C^d=-0.15$ GeV^{-1}, and $C^u=0.26$ GeV^{-1}. The blue dashed and black dash-dotted lines describe, respectively, the contributions of the loops in Figs. 1 and 2, while the red solid line denotes all the loop contributions.

In Refs. [53, 54], the radiative width for the B^{*+} is (0.40 ± 0.03) keV and for the B^{*0} it is (0.13 ± 0.03) keV. Using the Lagrangian in Eq. (13) together with the mass $m_b=4.18$ GeV [22], we get $\beta=2.1$ GeV^{-1}. Likewise, based on the interactions between the $B_1^{(\prime)}$ and B in Eq. (14) and the radiative widths of the $B_1^{(\prime)}\to\gamma B^{(*)}$ predicted in Ref. [55], $\tilde{\beta}$ is estimated to be in the range[1.5, 2.0] GeV^{-1}, and C^u is between 0.22 and 0.31 GeV^{-1} for the $B_1^+\to\gamma B^{(*)+}$, while C^d varies from -0.17 to -0.12 GeV^{-1} for the $B_1^0\to\gamma B^{(*)0}$. Notice that the minus sign is assigned for the later case, analogous to the radiative decay of B^* meson [53, 54].

Figure 3 exhibits the partial width for the radiative decay $\Upsilon(10753)\to\gamma X_b$ as a function of the X_b binding energy from 2 to 100 MeV, equivalently the mass between 10.504 and 10.602 GeV as indicated by the upper axis labels. It is seen that with increasing the binding energy, the radiative width increases. In particular, if the X_b as a molecule assumes a binding energy of 50 MeV, corresponding to a mass of about 10.554 GeV, we predict the radiative width

$$\Gamma[\Upsilon(10753)\to\gamma X_b]\sim 13\ \mathrm{keV}\,, \tag{17}$$

which yields a branching fraction of order 10^{-4}, two orders of magnitude larger than the processes $\Upsilon(5S,\,6S)\to$

γX_b [36].

When we look separately at the contributions from the loops in Figs. 1 and 2, indicated by the blue dashed line and the black dash-dotted line, respectively, it is clearly shown that the radiative process $\Upsilon(10753) \to \gamma X_b$ is predominantly governed by the loops in Fig. 2. The dominance of the loops involving the P-wave $B_1^{(\prime)}$ mesons, compared to the loops composed entirely of the S-wave $B^{(*)}$ mesons, is consistent with the power counting: For the loops in Fig. 1, the initial vertex is in a P wave and produces a momentum (see the Lagrangian in Eqs. (7) and (12)). This momentum has to be contracted with the external photon momentum q, and hence the initial vertex could be counted as q [46]. As a result, within the nonrelativistic framework, the loop integral scales as [36, 46, 49]

$$\frac{v^5}{(v^2)^3} q^2 = \frac{E_\gamma^2}{v} \,. \tag{18}$$

Here v can be understood as the average of the intermediate bottom meson velocities. The velocity can be estimated by $\sqrt{2|m_1 + m_2 - M_{i(f)}|/(m_1 + m_2)}$, where m_1 and m_2 are the masses of the bottom mesons related to the initial meson of mass M_i or the final meson of mass M_f [46].

For the loops in Fig. 2, the initial vertex is in an S wave thanks to the positive-parity bottom mesons $B_1^{(\prime)}$. In this case, the vertex is independent of the momentum (see Eqs. (7) and (12)). Therefore, such loop integral scales as [36, 46, 49]

$$\frac{v^5}{(v^2)^3} q m_B = m_B \frac{E_\gamma}{v} \,, \tag{19}$$

where m_B, the mass of the bottom meson, is introduced to balance the dimensions between Eqs. (18) and (19). According to the estimations at the beginning of this section, the coupling constants for the diagrams in Figs. 1 and 2 are nearly of the same order of magnitude. Thus the contributions from the loops in Fig. 2, when compared to those in Fig. 2, are enhanced by a factor of $m_B/E_\gamma \sim \mathcal{O}(30)$, agreeing with the numerical results shown in Fig. 3.

Theoretically, the $1/2^+$ bottom meson B_1' has a large width. The predictions in Ref. [55] show that $\Gamma_{B_1'}$ is around 130 MeV, which is about twice times smaller than the width (~ 240 MeV) predicted in Ref. [56]. For the $3/2^+$, the width was predicted to be about 20 MeV, agreeing with the measured data between 27.5 and 31 MeV [22]. In order to considering the width effect, especially for the B_1' mesons, we assume the mass spectrum to be described by the Breit-Wigner formula [36, 49, 57],

$$f(s, m, \Gamma) = \frac{1}{\pi} \frac{m\Gamma}{(s - m^2)^2 + m^2\Gamma^2} \,. \tag{20}$$

Here s is the mass squared of the meson in question, m is the central mass, and Γ is the meson width. Then the amplitude is given by

$$\mathcal{M} = \frac{1}{W} \int_{s_1}^{s_h} f(s, m_{B_1^{(\prime)}}, \Gamma_{B_1^{(\prime)}}) \mathcal{M}(m_{B_1^{(\prime)}} \to \sqrt{s})\, \mathrm{d}s \,, \tag{21}$$

where $\mathcal{M}(m_{B_1^{(\prime)}} \to \sqrt{s})$ represents the amplitude expression without considering the $B_1^{(\prime)}$ width, but the $B_1^{(\prime)}$ mass is replaced by the square root of the integration variable, $\sqrt{s}$. Additionally, $W = \int_{s_1}^{s_h} f(s, m_{B_1^{(\prime)}}, \Gamma)\, \mathrm{d}s$ with $s_l = m_{B_1^{(\prime)}}^2$ and $s_h = (m_{B_1^{(\prime)}} + \Gamma_{B_1'})^2$.

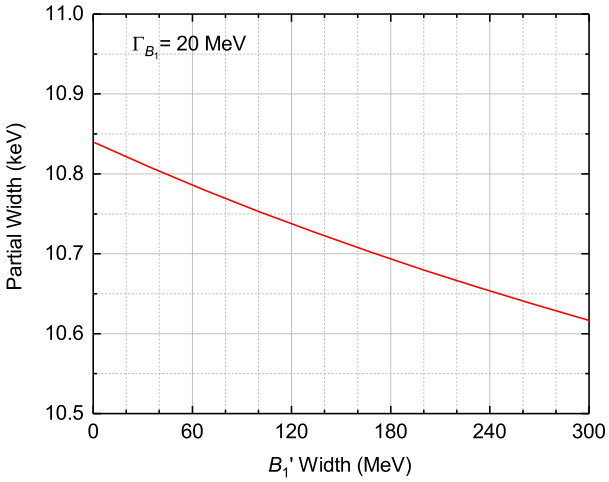


FIG. 4. Radiative decay width for the process $\Upsilon(10753) \to \gamma X_b$ as a function of the B_1' width. The B_1 width is fixed to be 20 MeV. The couplings are the same as those in Fig. 3.

In Fig. 4, the B_1' width dependence of the radiative decay width for the $\Upsilon(10753) \to \gamma X_b$ is shown. In the present calculations, the B_1 width is fixed to be 20 MeV because of its smallness in comparison with that of the B_1'. It is seen that the radiative decay width is slightly dependent on the B_1' width, decreasing less than 3% when the B_1' width is increased to 300 MeV. It should be noted that the possible variation of the couplings resulting from the change of the $B_1^{(\prime)}$ width is not considered in our calculations, which might give rise to extra effect on the radiative decay width.

IV. SUMMARY

In this work, we calculated the partial width for the radiative transition $\Upsilon(10753) \to \gamma X_b$, using a nonrelativistic effective field theory. In the calculations, we considered the X_b, the heavy quark flavor symmetry counterpart of the $X(3872)$ in the bottomonium sector, as a bound state of the $B^*\bar{B} + \text{c.c}$, and the $\Upsilon(10753)$ as an S-D mixed state of the $\Upsilon(4S)$ and $\Upsilon_1(3\,^3D_1)$. Moreover, the radiative transition was assume to occur through the intermediate bottom mesons, including the P-wave $B_1^{(\prime)}$ mesons as well as the S-wave $B^{(*)}$ ones.

It is found that the possible effect of the large width of the B_1' meson on the radiative decay width might be of minor importance, if the couplings do not change substantially with the $B_1^{(\prime)}$ width. Specially, our calculated results indicate that the radiative decay width is of order 10 keV when the X_b mass is around 10.56 GeV, corresponding a branching fraction of about 10^{-4}. This bigness of the radiative width implies that searching for the X_b via the process $\Upsilon(10753) \to \gamma X_b$ is promising. Recent experiments by Belle II Collaboration [33] did not find the X_b in $e^+e^- \to \gamma X_b$ with $X_b \to \omega\Upsilon(1S)$ at $\sqrt{s} = 10.745$ GeV. However, given our recent study [34], we suggest to hunt for the X_b in the channel $e^+e^- \to \gamma X_b$ with $X_b \to \pi\pi\chi_{b1}$ near $\sqrt{s} = 10.754$ GeV.

ACKNOWLEDGMENTS

This work is partly supported by the National Natural Science Foundation of China under Grants No. 12105153, No. 12075133, No. 12047503, and No. 12075288, and by the Natural Science Foundation of Shandong Province under Grants No. ZR2021MA082 and No. ZR2022ZD26. It is also supported by Taishan Scholar Project of Shandong Province (Grant No.tsqn202103062), the Higher Educational Youth Innovation Science and Technology Program Shandong Province (Grant No. 2020KJJ004).

[1] S. Choi, S. Olsen, K. Abe, *et al.* (Belle), Phys. Rev. Lett. **91**, 262001 (2003), arxiv:hep-ex/0309032.
[2] B. Aubert, R. Barate, D. Boutigny, *et al.* (BaBar), Phys. Rev. D **71**, 071103 (2005), arxiv:hep-ex/0406022.
[3] V. Abazov, B. Abbott, M. Abolins, *et al.* (D0), Phys. Rev. Lett. **93**, 162002 (2004), arxiv:hep-ex/0405004.
[4] T. Aaltonen, J. Adelman, T. Akimoto, *et al.* (CDF), Phys. Rev. Lett. **103**, 152001 (2009), arxiv:0906.5218 [hep-ex].
[5] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, *et al.* (CMS), JHEP **04** (4), 154, arxiv:1302.3968 [hep-ex].
[6] R. Aaij, C. Abellan Beteta, B. Adeva, *et al.* (LHCb), Phys. Rev. Lett. **110**, 222001 (2013), arxiv:1302.6269 [hep-ex].
[7] M. Ablikim *et al.* (BESIII), Phys. Rev. Lett. **110**, 252001 (2013), arxiv:1303.5949 [hep-ex].
[8] Z. Liu, C. Shen, C. Yuan, *et al.* (Belle), Phys. Rev. Lett. **110**, 252002 (2013), arxiv:1304.0121 [hep-ex].
[9] T. Xiao, S. Dobbs, A. Tomaradze, and K. K. Seth, Phys. Lett. B **727**, 366 (2013), arxiv:1304.3036 [hep-ex].
[10] V. M. Abazov, B. K. Abbott, B. S. Acharya, *et al.* (D0), Phys. Rev. D **98**, 052010 (2018), arxiv:1807.00183 [hep-ex].
[11] A. Bondar, A. Garmash, R. Mizuk, *et al.* (Belle), Phys. Rev. Lett. **108**, 122001 (2012), arxiv:1110.2251 [hep-ex].
[12] I. Adachi, K. Adamczyk, H. Aihara, *et al.* (Belle), arXiv:1207.4345 [hep-ex] 10.48550/arXiv.1207.4345 (2012), arxiv:1207.4345 [hep-ex].
[13] P. Krokovny, A. Bondar, I. Adachi, *et al.* (Belle), Phys. Rev. D **88**, 052016 (2013), arxiv:1308.2646 [hep-ex].
[14] A. Garmash, A. Bondar, A. Kuzmin, *et al.* (Belle), Phys. Rev. D **91**, 072003 (2015), arxiv:1403.0992 [hep-ex].
[15] N. Brambilla, S. Eidelman, C. Hanhart, A. Nefediev, C.-P. Shen, C. E. Thomas, A. Vairo, and C.-Z. Yuan, Phys. Rept. **873**, 1 (2020), arxiv:1907.07583 [hep-ex].
[16] F.-K. Guo, C. Hanhart, U.-G. Meißner, Q. Wang, Q. Zhao, and B.-S. Zou, Rev. Mod. Phys. **90**, 015004 (2018), arxiv:1705.00141 [hep-ph].
[17] R. F. Lebed, R. E. Mitchell, and E. S. Swanson, Prog. Part. Nucl. Phys. **93**, 143 (2017), arxiv:1610.04528 [hep-ph].
[18] Y. S. Kalashnikova and A. V. Nefediev, Phys. Usp. **62**, 568 (2019), arxiv:1811.01324 [hep-ph].
[19] H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, Phys. Rept. **639**, 1 (2016), arxiv:1601.02092 [hep-ph].
[20] L. Meng, B. Wang, G.-J. Wang, and S.-L. Zhu, Phys. Rept. **1019**, 2266 (2023), arxiv:2204.08716 [hep-ph].
[21] V. Baru, E. Epelbaum, A. Filin, C. Hanhart, and A. Nefediev, JHEP **06** (6), 158, arxiv:1704.07332 [hep-ph].
[22] R. Workman, V. Burkert, V. Crede, *et al.* (Particle Data Group), PTEP **2022**, 083C01 (2022).
[23] W.-S. Hou, Phys. Rev. D **74**, 017504 (2006), arxiv:hep-ph/0606016.
[24] D. Ebert, R. N. Faustov, and V. O. Galkin, Phys. Lett. B **634**, 214 (2006), arxiv:hep-ph/0512230.
[25] A. Ali, C. Hambrock, I. Ahmed, and M. J. Aslam, Phys. Lett. B **684**, 28 (2010), arxiv:0911.2787 [hep-ph].
[26] R. D. Matheus, S. Narison, M. Nielsen, and J.-M. Richard, Phys. Rev. D **75**, 014005 (2007), arxiv:hep-ph/0608297.
[27] N. A. Törnqvist, Z. Phys. C **61**, 525 (1994), arxiv:hep-ph/9310247.
[28] F.-K. Guo, C. Hidalgo-Duque, J. Nieves, and M. Pavón Valderrama, Phys. Rev. D **88**, 054007 (2013), arxiv:1303.6608 [hep-ph].
[29] G. Li and Z. Zhou, Phys. Rev. D **91**, 034020 (2015), arxiv:1502.02936 [hep-ph].
[30] G. Aad, B. Abbott, J. Abdallah, *et al.* (ATLAS), Phys. Lett. B **740**, 199 (2015), arxiv:1410.4409 [hep-ex].
[31] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, *et al.* (CMS), Phys. Lett. B **727**, 57 (2013), arxiv:1309.0250 [hep-ex].
[32] X. He, C. Shen, C. Yuan, *et al.* (Belle), Phys. Rev. Lett. **113**, 142001 (2014), arxiv:1408.0504 [hep-ex].
[33] I. Adachi, L. Aggarwal, H. Ahmed, *et al.* (Belle-II), Phys. Rev. Lett. **130**, 091902 (2023), arxiv:2208.13189 [hep-ex].
[34] Z.-S. Jia, Z.-H. Zhang, W.-H. Qin, and G. Li, arXiv:2311.15527 [hep-ph] (2023), arxiv:2311.15527 [hep-ph].
[35] M. Ablikim, M. Achasov, X. Ai, *et al.* (BESIII), Phys. Rev. Lett. **112**, 092001 (2014), arxiv:1310.4101 [hep-ex].
[36] X.-Y. Wang, Z.-X. Cai, G. Li, S.-D. Liu, C.-S. An, and J.-J. Xie, Eur. Phys. J. C **83**, 186 (2023), arxiv:2301.07365 [hep-ph].
[37] R. Mizuk, A. Bondar, I. Adachi, *et al.* (Belle), JHEP **10** (10), 220, arxiv:1905.05521 [hep-ex].
[38] Y.-S. Li, Z.-Y. Bai, Q. Huang, and X. Liu, Phys. Rev. D

104, 034036 (2021), arxiv:2106.14123 [hep-ph].
[39] Y.-S. Li, Z.-Y. Bai, and X. Liu, Phys. Rev. D **105**, 114041 (2022), arxiv:2205.04049 [hep-ph].
[40] Z.-Y. Bai, Y.-S. Li, Q. Huang, X. Liu, and T. Matsuki, Phys. Rev. D **105**, 074007 (2022), arxiv:2201.12715 [hep-ph].
[41] S.-D. Liu, Z.-X. Cai, Z.-S. Jia, G. Li, and J.-J. Xie, Phys. Rev. D **109**, 014039 (2024), arxiv:2312.02761 [hep-ph].
[42] I. Adachi, L. Aggarwal, H. Ahmed, *et al.* (Belle-II), arXiv:2312.13043 [hep-ex] (2023), arxiv:2312.13043 [hep-ex].
[43] A. M. Badalian, B. L. G. Bakker, and I. V. Danilkin, Phys. Atom. Nucl. **73**, 138 (2010), arxiv:0903.3643 [hep-ph].
[44] J.-Z. Wang, Z.-F. Sun, X. Liu, and T. Matsuki, Eur. Phys. J. C **78**, 915 (2018), arxiv:1802.04938 [hep-ph].
[45] F.-K. Guo, C. Hanhart, Yu. S. Kalashnikova, U.-G. Meißner, and A. V. Nefediev, Phys. Lett. B **742**, 394 (2015), arxiv:1410.6712 [hep-ph].
[46] F.-K. Guo, C. Hanhart, U.-G. Meißner, Q. Wang, and Q. Zhao, Phys. Lett. B **725**, 127 (2013), arxiv:1306.3096 [hep-ph].
[47] F.-K. Guo, C. Hanhart, G. Li, U.-G. Meißner, and Q. Zhao, Phys. Rev. D **83**, 034013 (2011), arxiv:1008.3632 [hep-ph].
[48] F.-K. Guo, C. Hanhart, G. Li, U.-G. Meißner, and Q. Zhao, Phys. Rev. D **82**, 034025 (2010), arxiv:1002.2712 [hep-ph].
[49] Q. Wu, D.-Y. Chen, and F.-K. Guo, Phys. Rev. D **99**, 034022 (2019), arxiv:1810.09696 [hep-ph].
[50] J. Hu and T. Mehen, Phys. Rev. D **73**, 054003 (2006), arxiv:hep-ph/0511321.
[51] L.-C. Gui, L.-S. Lu, Q.-F. Lü, X.-H. Zhong, and Q. Zhao, Phys. Rev. D **98**, 016010 (2018), arxiv:1801.08791 [hep-ph].
[52] J.-Z. Wang, D.-Y. Chen, X. Liu, and T. Matsuki, Phys. Rev. D **99**, 114003 (2019), arxiv:1903.07115 [hep-ph].
[53] H.-M. Choi, Phys. Rev. D **75**, 073016 (2007), arxiv:hep-ph/0701263.
[54] S.-l. Zhu, W.-Y. P. Hwang, and Z.-s. Yang, Mod. Phys. Lett. A **12**, 3027 (1997), arxiv:hep-ph/9610412.
[55] I. Asghar, B. Masud, E. S. Swanson, F. Akram, and M. Atif Sultan, Eur. Phys. J. A **54**, 127 (2018), arxiv:1804.08802 [hep-ph].
[56] M.-L. Du, M. Albaladejo, P. Fernandez-Soler, F.-K. Guo, C. Hanhart, U.-G. Meißner, J. Nieves, and D.-L. Yao, Phys. Rev. D **98**, 094018 (2018), arxiv:1712.07957 [hep-ph].
[57] Q. Wu, D.-Y. Chen, and T. Matsuki, Eur. Phys. J. C **81**, 193 (2021), arxiv:2102.08637 [hep-ph].

MeV gamma rays from Q-ball decay

Shinta Kasuya[a], Masahiro Kawasaki[b,c], Naomi Tsuji[a,d]
[a] *Physics Division, Faculty of Science, Kanagawa University, Kanagawa 221-8686, Japan*
[b] *Institute for Cosmic Ray Research, the University of Tokyo, Chiba 277-8582, Japan*
[c] *Kavli Institute for the Physics and Mathematics of the Universe (WPI), Todai Institutes for Advanced Study, the University of Tokyo, Chiba 277-8582, Japan*
[d] *Interdisciplinary Theoretical and Mathematical Science Program (iTHEMS), RIKEN, Saitama 351-0198, Japan*
(Dated: March 4, 2024)

We study the supersymmetric Q balls which decay at present and find that they create a distinctive spectrum of gamma rays at around O(10) MeV. The charge of the Q ball is lepton numbers in order for the lifetime to be as long as the present age of the universe, and the main decay products are light leptons. However, as the charge of the Q ball decreases, the decay channel into pions becomes kinematically allowed towards the end of the decay, and the pions are produced at rest. Immediately, π^0 decays into two photons with the energy of 67.5 MeV, half the pion mass, which exhibits a unique emission line. In addition, $\pi^\pm$ decay into $\mu^\pm$, which further decay with emitting internal bremsstrahlung, whose spectrum has a sharp cutoff at $\sim$50 MeV. If the observations would find these peculiar features of the gamma-ray spectrum in the future, it could be a smoking gun of the supersymmetric Q-ball decay at present.

arXiv:2403.01675v1 [hep-ph] 4 Mar 2024

I. INTRODUCTION

There exist non-topological solitons, Q balls [1], which consist of scalar fields such as squarks and sleptons in supersymmetric theories [2]. Q balls with large charge can form after inflation [3, 4] à la Affleck-Dine [5]. Large Q balls with Q being the baryon number are stable against the decay into nucleons, and could be dark matter of the universe [3]. Such Q ball dark matter may be detected in the large volume detectors [6–9]. If detected, it is not only the detection of the dark matter, but also might be the observational clue of the supersymmetry (SUSY).

On the other hand, large Q balls with lepton charge is not stable against the decay into light leptons, but they could long lived. Their decay leaves some observational consequences such that they may result in the source of the 511 keV line gamma rays from the galactic center [10], provide late-time entropy production [11], or enhance the secondary gravitational waves [12, 13].

Here we study yet another way to find the observational trace of SUSY Q balls with Q being the lepton number. We look for clues of the Q balls whose lifetime is just as long as the present age of the universe. Since the charge of the Q ball (lepton number in this case) should be very large for such lifetime, the mass per unit charge is a little bit larger than the electron (positron) mass. Therefore, Q balls decay into neutrinos, anti-neutrinos, electrons, and positrons in the first place, decreasing the Q-ball charge. As the charge gets smaller, the mass per unit charge becomes larger, and finally the decay channels to pions open up towards the end of the decay, and hence the pions are produced almost at rest. This is the unique feature of the Q-ball decay.

Produced pions decay immediately. Neutral pions decay into two photons with the energy of half the pion mass, which shows a 67.5 MeV emission line, while charged pions decay into muons, which further decay emitting photons through internal bremsstrahlung (IB). The resulting spectrum of the photons has such peculiar features in the 1–100 MeV range.

The structure of the paper is as follows. In the next section, we provide basic characters of the Q ball considered in this paper. In Sec.III, we investigate the process of the decreasing charge Q for the Q-ball decay at present, and consider the decay products and their fractions in Sec. IV. In Sec.V, we estimate the gamma-ray spectra produced by the Q-ball decay for both the galactic and extragalactic emissions. Finally we conclude in Sec. VI. We explain the internal bremsstrahlung in pion and muon decays in App.A, and the upper bound on Q-ball abundance in App.B.

II. Q BALLS IN THE GAUGE MEDIATION

The Q ball is the energy minimum configuration of the scalar fields for the fixed charge Q [1]. Those scalar fields Φ, called flat directions, are all classified in terms of gauge-invariant monomials in the minimal supersymmetric standard model [14, 15]. In the gauge-mediated SUSY breaking scenario, the scalar potential is written as

$$V(\Phi) = M_F^4 \left(\log \frac{|\Phi|^2}{M_S^2} \right)^2 + m_{3/2}^2 \left(1 + K \log \frac{|\Phi|^2}{M_{\rm P}^2} \right) |\Phi|^2, \tag{1}$$

where $m_{3/2}$ is the gravitino mass and $M_{\rm P} = 2.4 \times 10^{18}$ GeV is the Planck mass. The first term comes from the gauge-mediation effects above the messenger scale M_S [16]. M_F is related to the F component of a gauge-singlet chiral multiplet in the messenger sector, and its range is given by [9]

$$4 \times 10^4 \text{ GeV} \lesssim M_F \lesssim 0.1 \left(m_{3/2} M_{\rm P} \right)^{1/2}. \tag{2}$$

The second term is due to the gravity-mediation effects. Here, K is a coefficient of the one-loop corrections to

the mass, and negative for most of the flat directions with $|K|$ =0.01–0.1. Since the gravitino mass is relatively small in the gauge-mediation models, the second term dominates over the first one for field values larger than

$$\phi_{\rm eq} \simeq \frac{\sqrt{2}M_F^2}{m_{3/2}}, \tag{3}$$

where we define $\Phi = \frac{1}{\sqrt{2}}\phi e^{i\theta}$.

There are two types of the Q ball depending on which term of the potential is dominant when it forms. If the first term of Eq.(1) dominates the potential, the gauge-mediation type Q balls are created, while the so-called new-type Q balls are produced when the second term overwhelms the first [17]. The important property which concerns with following discussion is the mass per unit charge of the Q ball, ω_Q, equivalent to the effective mass of Φ inside the Q ball:

$$\omega_Q \simeq \begin{cases} \sqrt{2}\pi\zeta M_F Q^{-1/4} & \text{(gauge-med. type)}, \\ m_{3/2} & \text{(new type)}, \end{cases} \tag{4}$$

where $\zeta \simeq 2.5$ [18].

As shown later, we need very large charge Q so that the new-type Q balls form for the field amplitude well above ϕ_{eq}. The charge of the Q ball is thus estimated as [19]

$$Q = \beta_N \left(\frac{\phi_0}{m_{3/2}}\right)^2, \tag{5}$$

where $\beta_N \simeq 0.02$ [20], and ϕ_0 is the amplitude of Φ at the beginning of its oscillations. The mass and size of this type of the Q ball are given respectively by [17, 21]

$$M_Q \simeq m_{3/2}Q, \quad R_Q \simeq |K|^{-1/2} m_{3/2}^{-1}. \tag{6}$$

III. Q-BALL DECAY

Q-ball decay occurs if some decay particles carry the same kind of the charge of the Q ball and the mass of all the decay particles is less than the mass of the Q ball per unit charge ω_Q. Since the decay products carrying lepton numbers are light fermions, once the Fermi sea is filled, further decay proceeds only when produced fermions escape from the surface of the Q ball. The upper bound of the decay rate is thus determined by the maximum outgoing flow of the fermions [22]. This saturation occurs when the field value is large inside the Q ball, which is the case here. Then the decay rate is estimated as [22]

$$\Gamma_Q \simeq \frac{1}{Q}\frac{\omega_Q^3}{192\pi^2} 4\pi R_Q^2 \simeq \frac{m_{3/2}}{48\pi|K|Q}, \tag{7}$$

where we use Eq.(6) in the last equality. Since we consider the Q balls that decay at present, its lifetime is set to be $\tau_Q = \Gamma_Q^{-1} \simeq t_0 \simeq 13.8$ Gyr, which leads to the charge of the Q ball as

$$Q \simeq 2.2 \times 10^{38} \left(\frac{|K|}{0.02}\right)^{-1} \left(\frac{m_{3/2}}{\rm MeV}\right). \tag{8}$$

From Eqs.(5) and (8), we obtain the charge of the formed new-type Q ball and the gravitino mass respectively as

$$Q_{\rm D} \simeq 7.9 \times 10^{38} \left(\frac{\beta_N}{0.02}\right)^{1/3} \left(\frac{|K|}{0.02}\right)^{-2/3} \left(\frac{\phi_0}{0.3M_{\rm P}}\right)^{2/3}, \tag{9}$$

$$m_{3/2} \simeq 3.6 \text{ MeV} \left(\frac{\beta_N}{0.02}\right)^{1/3} \left(\frac{|K|}{0.02}\right)^{1/3} \left(\frac{\phi_0}{0.3M_{\rm P}}\right)^{2/3}. \tag{10}$$

Notice that the scenario works for broader parameter space, but here we take $\phi_0 = 0.3M_{\rm P}$ and $|K| = 0.02$ as a typical example.[1]

As the charge of the Q ball decreases, the new-type Q ball transforms into the gauge-mediation type [23] when the charge becomes

$$Q_{\rm tr} \simeq 4\left(\frac{M_F}{m_{3/2}}\right)^4 \simeq 1.8 \times 10^{38} \left(\frac{m_{3/2}}{3.6 \text{ MeV}}\right)^{-2}, \tag{11}$$

where we use the upper bound of M_F in Eq.(2) in the second equality. Thereafter, ω_Q increases as the charge diminishes as in the upper line in Eq.(4). Since pions consist of quark and anti-quark pairs, the decay channel to pions opens when $\omega_Q = m_\pi/2$, where the charge of the Q ball becomes

$$Q_\pi \simeq \left(\frac{2\sqrt{2}\pi\zeta}{m_\pi}\right)^4 M_F^4 \simeq 4.8 \times 10^{36} \left(\frac{m_{3/2}}{3.6 \text{ MeV}}\right)^2. \tag{12}$$

Here the upper bound of M_F in Eq.(2) is used in the last equality. As the decay into pions has just come to be kinematically allowed at that time, pions are produced at rest. This is the unique property of the Q-ball decay.

In Fig. 1, we show the evolution of the charge of the Q ball by the purple arrow. Here we plot the charge at the formation (5), that for the decay at present (8), that when the transformation from the new-type into the gauge-mediation type takes place (11), and that when the decay channel to pions opens (12) as the function of $m_{3/2}$, for $\phi_0 = 0.3M_{\rm P}$ and $|K| = 0.02$.

IV. DECAY PRODUCTS

We must identify the particles emitted by the Q-ball decay. To this end, we choose the $\Phi^5 = \mathcal{Q}u\mathcal{Q}ue$ direc-

[1] For the gravitino mass smaller than 100 keV, the gauge-mediation type Q balls can form and decay at present to produce same signatures of MeV gamma rays.

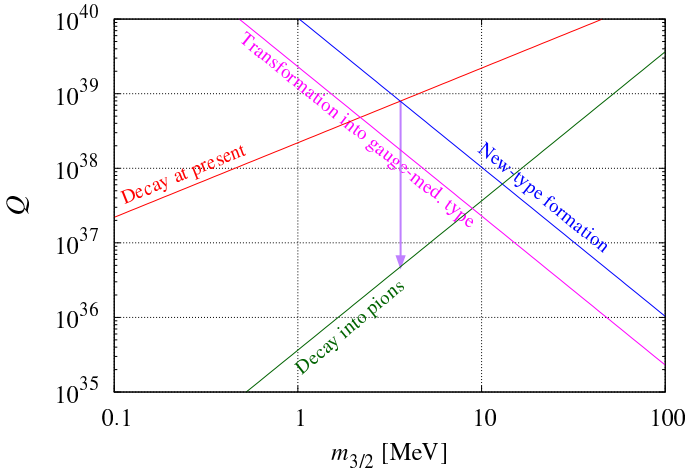


FIG. 1: Evolution of the Q-ball charge is shown in the purple arrow. We show the charge at the formation (5), that for the decay at present (8), that when the transformation from the new-type into the gauge-mediation type takes place (11), and that when the decay channel to pions opens (12) as the function of $m_{3/2}$, for $\phi_0 = 0.3M_{\rm P}$ and $|K| = 0.02$, in blue, red, magenta, and green lines, respectively.

tion as a concrete example below.[2] $\mathcal{Q}$ represents $\mathrm{SU}(2)_L$-doublet squarks, u and e are $\mathrm{SU}(2)_L$-singlet squarks and sleptons, respectively. We further assume that they consist only of the fields in the first generation, so that it can be written as $\tilde{u}_L\tilde{u}_R^*\tilde{d}_L\tilde{u}_R^*\tilde{e}_R^*$. Since this direction carries lepton number $L = -1$, the field Φ possesses $L = -1/5$.

In the process of the Q-ball decay, all the decay particles must have masses smaller than ω_Q, and some should carry anti-lepton numbers. In addition, the electric charge neutrality holds through the decay process, since the Q ball consists of the flat direction which is gauge invariant. Thus the composition changes before and after the decay channel into pions opens. Before that time, it is kinematically allowed to decay only into electrons, positrons, neutrinos, and anti-neutrinos.[3] Selectrons $\tilde{e}_R^*$ carrying anti-lepton number decay into anti-electron neutrinos $\bar{\nu}_e$. Since the Q-ball decay is saturated, there is no phase space left for $\bar{\nu}_e$ to be further produced. Therefore, the other decay particles are the following six patterns of combinations: $\nu_\mu\bar{\nu}_\mu\nu_\mu\bar{\nu}_\mu$, $\nu_\mu\bar{\nu}_\mu\nu_\tau\bar{\nu}_\tau$, $\nu_\tau\bar{\nu}_\tau\nu_\tau\bar{\nu}_\tau$, $\nu_\mu\bar{\nu}_\mu e^+e^-$, $\nu_\tau\bar{\nu}_\tau e^+e^-$, and $e^+e^-e^+e^-$.

On the other hand, after the decay into pions are kinematically allowed, anti-lepton numbers could be conveyed by e^+, in addition to $\bar{\nu}_e$, since π^- exists in the decay particles to compensate the electric charge of positrons. The phase spaces of produced $\bar{\nu}_e$ and e^+ are filled by the saturated decay so that they cannot be included in the remaining decay particles. In the case of $\bar{\nu}_e$, the combinations of other decay particles are $\nu_\mu\bar{\nu}_\mu\nu_\mu\bar{\nu}_\mu$, $\nu_\mu\bar{\nu}_\mu\nu_\tau\bar{\nu}_\tau$, $\nu_\tau\bar{\nu}_\tau\nu_\tau\bar{\nu}_\tau$, $\pi^+\nu_e e^-$, $\pi^+\nu_\mu\mu^-$, $\pi^0\nu_\mu\bar{\nu}_\mu$, $\pi^0\nu_\tau\bar{\nu}_\tau$, $\pi^+\pi^-$, or $\pi^0\pi^0$, while they are $\pi^-\pi^0$, $\pi^-\nu_\mu\bar{\nu}_\mu$, or $\pi^-\nu_\tau\bar{\nu}_\tau$ for e^+.

Now we are ready to estimate the fraction of the produced particle per unit anti-lepton number or per unit charge of the Q ball. After pions are kinematically allowed to produced from the Q-ball decay, the numbers of particles per unit Q-ball charge are obtained as $f_{\pi^0} = 1/12$ for π^0, $f_{\pi^+} = 1/20$ for π^+, $f_{\pi^-} = 1/15$ for π^-, $f_{\mu^-} = 1/60$ for μ^-, while $f_{e^+} = 2/15$ for e^+ when the decay channel to pions is not open.

V. PHOTON SPECTRUM

A. Galactic gamma-ray emission

Pions are produced at rest just after they are kinematically allowed, and immediately they emit photons which has peculiar spectrum as explained in the following. π^0 decays into two photons with the energy of half the π^0 mass, 67.5 MeV. We first consider the galactic gamma-ray emission. The spectrum of this line gamma rays can be obtained as

$$\varepsilon_\gamma^2\frac{d\Phi_{2\gamma}}{d\varepsilon_\gamma} = \frac{1}{4\pi}\frac{\Omega_Q\mathcal{D}}{\omega_Q\Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}}f_{\pi^0}2\left(\frac{\varepsilon_\gamma}{\Delta\varepsilon_\gamma}\right)\varepsilon_\gamma, \qquad (13)$$

where Ω_Q is the density parameter of the Q ball. $\Delta\tau_Q = \xi t_0$, where ξ represents the uncertainty of the lifetimes of the Q balls due to some variance in their sizes. $\Delta\varepsilon_\gamma/\varepsilon_\gamma$ is the energy resolution of the observation. $Q_\pi/Q_{\rm D}$ implies the fraction of the charge of the Q ball that can decay into pions, and f_{π^0} is the number of π^0 produced per unit Q-ball charge, as described in the previous section. $\mathcal{D}$ is a D-factor, given by

$$\mathcal{D} = \frac{1}{\Delta\Omega}\int_{\Delta\Omega}d\ell db\cos b\int_0^{s_{\rm max}}ds\rho_{\rm DM}(r(\ell,b,s)), \qquad (14)$$

where ℓ and b are galactic longitude and latitude, respectively. $\Delta\Omega$ denotes the solid angle of the observation. s is the distance along the line of sight. The distance from the galactic center is expressed as $r(\ell,b,s) = \left(R_{GC}^2 - 2sR_{\rm GC}\cos\psi + s^2\right)^{1/2}$, where ψ is the angle between the direction to the galactic center and the line of sight so that $\cos\psi = \cos\ell\cos b$. $s_{\rm max} = \left(R_{\rm MW}^2 - \sin^2\psi R_{\rm GC}^2\right)^{1/2} + R_{\rm GC}\cos\psi$ with $R_{MW} = 30$ kpc being assumed as the size of the dark halo. We adopt the Navaro-Frenk-White profile [25] for the dark matter halo density, written as

$$\rho_{\rm DM}(r) = \frac{\rho_0}{\left(\frac{r}{r_0}\right)\left(1+\frac{r}{r_0}\right)^2}. \qquad (15)$$

Here we take $\rho_0 = 0.53$ GeV/cm^3 and $r_0 = 15$ kpc so that we have the solar local density of $\rho_{\rm DM}(R_{\rm GC}) =$

[2] Similar argument holds in other leptonic directions.

[3] The branching ratio to photons can be extremely suppressed due to the very small left-right mixings for squarks and sleptons for the first and second generations [24].

0.4 GeV/cm^3, where $R_{\rm GC} = 8.23$ kpc is the distance to the galactic center from the sun [26].

On the other hand, π^+ and π^- decay into μ^+ and μ^-, respectively. Photons are emitted through the internal bremsstrahlung with the branching ratio [27–29]

$$\frac{dB_{\pi\rm IB}}{d\varepsilon_\gamma} = \frac{\alpha}{2\pi}\frac{1}{(1-p)^2}\frac{2}{m_{\pi^\pm}}\frac{1}{x} \times \left[\frac{\left\{x^2+4(1-p)(1-x)\right\}(x^2+p-1)}{1-x} + \left\{2(1-p)(1+p-x)+x^2\right\}\log\frac{1-x}{p}\right], \quad (16)$$

where $\alpha = 1/137$ is the fine structure constant, $p = (m_\mu/m_{\pi^\pm})^2$, and $x = 2\varepsilon_\gamma/m_{\pi^\pm}$. See App. A for the details. The spectrum of IB photons is thus given by

$$\varepsilon_\gamma^2\frac{d\Phi_{\pi\rm IB}}{d\varepsilon_\gamma} = \frac{1}{4\pi}\frac{\Omega_Q\mathcal{D}}{\omega_Q\Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}}f_{\pi^\pm}\frac{dB_{\pi\rm IB}}{d\varepsilon_\gamma}\varepsilon_\gamma^2, \quad (17)$$

where $f_{\pi^\pm} = f_{\pi^+} + f_{\pi^-} = 7/60$ since the spectrum of internal bremsstrahlung is the same for both π^+ and π^- decays.

In addition, μ^+ and μ^- decay into e^+ and e^-, respectively, similarly emitting internal bremsstrahlung with the branching ratio [30]

$$\frac{dB_{\mu\rm IB}}{d\varepsilon_\gamma} = \frac{\alpha}{2\pi}\frac{2}{m_\mu}\frac{2(1-y)}{3y} \times \left[-\frac{17}{2}-\frac{3}{4}(1-y)+\frac{16}{3}(1-y)^2-\frac{55}{12}(1-y)^3 + \left\{3-2(1-y)^2+2(1-y)^3\right\}\log\frac{1-y}{q}\right], \quad (18)$$

where $q = (m_e/m_\mu)^2$ and $y = 2\varepsilon_\gamma/m_\mu$. See App. A for the details. The spectrum of the IB photons is then written as

$$\varepsilon_\gamma^2\frac{d\Phi_{\mu\rm IB}}{d\varepsilon_\gamma} = \frac{1}{4\pi}\frac{\Omega_Q\mathcal{D}}{\omega_Q\Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}}f_{\mu^\pm}\frac{dB_{\mu\rm IB}}{d\varepsilon_\gamma}\varepsilon_\gamma^2, \quad (19)$$

where $f_{\mu^\pm} = f_{\pi^\pm} + f_{\mu^-} = 2/15$, since the charged pions decay into muons.

We illustrate the spectra (13), (17) and (19) in Fig. 2. Here we consider the region with $|\ell| < 30^\circ$ and $|b| < 5^\circ$, and set $\Omega_Q/\xi \simeq 8\times10^{-9}$, well below the rough upper bound explained in App. B.[4] Also shown are the observations of the gamma-ray diffuse emission from the inner galactic region measured by COMPTEL in 1–30 MeV [31] and EGRET in 30–100 MeV [32]. We adopt $\Delta\varepsilon_\gamma/\varepsilon_\gamma = 0.15$ in Eq.(13) to compare with the EGRET data [33]. It should be noted that the spectra obtained by COMPTEL and EGRET include the galactic diffuse emission, and there may be an excess component in addition to the galactic diffuse emission to reconcile with the observed spectra (See, e.g., Refs. [32, 34, 35]). The excess could be contributed from the gamma rays originated from the Q-ball decay, unresolved sources [36], dark matter [37–39] and/or cosmic neutrinos [40], or may be explained by the uncertainties of the model of the galactic diffuse emission (e.g., Ref. [41]).

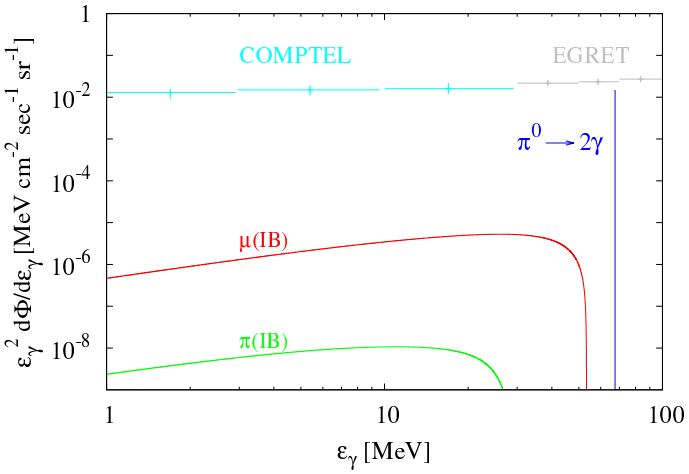


FIG. 2: Photon spectra from the Q-ball decay. We display the spectra (13), (17) and (19) in blue, green, and red lines, respectively. Observations of the galactic diffuse gamma-ray emission are also shown in cyan (COMPTEL) and gray points (EGRET).

We can see that the unique feature of the spectrum is the line-gamma rays at $\varepsilon_\gamma = 67.5$ MeV. Astrophysical sources could hardly produce such an emission line. It could be detected by future observations such as e-ASTROGAM [42], AMEGO-X [43], GRAMS [44], SMILE-3 [45], GECCO [46], and a CubeSat for MeV observations (MeVCube) [47].[5][6] For example, GRAMS has 67.5 MeV line sensitivity of $\sim 10^{-6}$ MeV/cm^2/s [44]. Detection of this line by some future observations would prove this Q-ball decay scenario. If not, it may result in lowering the upper bound on Ω_Q. In addition, the Q-ball decay produces almost the same amount of π^0 and $\pi^\pm$, so that photons from the radiative muon decay through the internal bremsstrahlung will appear in the energy range $\varepsilon_\gamma \lesssim 50$ MeV with a sharp cutoff at around 50 MeV. Con-

[4] Since Q-ball abundance would be by far much larger than this value for usual formation à la Affleck-Dine, it is necessary to reduce it. One mechanism is considered in Ref. [10] using resonant decay of the flat direction in most of the region of the universe, and another one may be late-time entropy production before nucleosynthesis.

[5] Although the gamma-ray data by Fermi-LAT below 100 MeV is available, there is no dedicated search for 67.5 MeV line emission because of the technical difficulty at the lower energy band.

[6] COSI-SMEX [48], to be launched in 2027, is not sensitive to the gamma rays above 5 MeV.

tinuum sensitivity of GRAMS may reach to observe this internal bremsstrahlung on the top of the diffuse background for a satellite mission [44]. Measurements of the 67.5 MeV line and/or this broad spectrum with a cutoff at around 50 MeV could be the evidence of the Q-ball decay model.

B. Extragalactic gamma-ray emission

Let us now study the extragalactic gamma-ray emission. For $\pi^0 \to 2\gamma$, taking into account the cosmic expansion, we can estimate the present flux as [49, 50]

$$\varepsilon_\gamma^2 \frac{d\Phi_{2\gamma}^{\rm ext}}{d\varepsilon_\gamma} = \frac{\varepsilon_\gamma^2}{4\pi}\int_{(1-\xi)t_0}^{t_0} dt' \frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}} f_{\pi^0} \times 2(1+z)\delta\left[\varepsilon_\gamma' - \frac{m_{\pi^0}}{2}\right], \quad (20)$$

where ρ_{c0} is the present critical density of the universe, z is the redshift, and $\varepsilon_\gamma' = (1+z)\varepsilon_\gamma$ is the photon energy produced at the decay time for the photon with energy ε_γ at present. The redshift z is related to cosmic time t as

$$\frac{dt}{dz} = -(1+z)^{-5/2} H_0^{-1}\left[\Omega_{\rm M} + \frac{\Omega_\Lambda}{(1+z)^3}\right]^{-1/2}, \quad (21)$$

where H_0 is the Hubble parameter at present, Ω_M and Ω_Λ are the present density parameters of matter and cosmological constant, respectively, and the flat universe is assumed. Therefore, we obtain the flux as

$$\varepsilon_\gamma^2 \frac{d\Phi_{2\gamma}^{\rm ext}}{d\varepsilon_\gamma} = \frac{1}{2\pi}\frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q H_0}\frac{Q_\pi}{Q_{\rm D}} f_{\pi^0}\varepsilon_\gamma \left(\frac{2\varepsilon_\gamma}{m_{\pi^0}}\right)^{3/2} \times \left[\Omega_{\rm M} + \Omega_\Lambda\left(\frac{2\varepsilon_\gamma}{m_{\pi^0}}\right)^3\right]^{-1/2}. \quad (22)$$

Notice that the spectrum has a rather narrow shape where the upper and lower edges are respectively located at the energy of 67.5 MeV and $67.5/(1+z_\xi)$ MeV, where z_ξ corresponds to the redshift at the cosmic time $(1-\xi)t_0$.

For the internal bremsstrahlung from pions, we have

$$\varepsilon_\gamma^2 \frac{d\Phi_{\pi{\rm IB}}^{\rm ext}}{d\varepsilon_\gamma} = \frac{\varepsilon_\gamma^2}{4\pi}\int_{(1-\xi)t_0}^{t_0} dt' \frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}} f_{\pi^\pm} \times (1+z)\frac{dB_{\pi{\rm IB}}}{d\varepsilon_\gamma'}(\varepsilon_\gamma'), \quad (23)$$

where Eq.(16) is used. Changing the variable of integration to z, we obtain

$$\varepsilon_\gamma^2 \frac{d\Phi_{\pi{\rm IB}}^{\rm ext}}{\varepsilon_\gamma} = \frac{\varepsilon_\gamma}{4\pi}\frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q H_0}\frac{Q_\pi}{Q_{\rm D}} f_{\pi^\pm}\frac{\alpha}{2\pi}\frac{1}{(1-p)^2} \times \int_0^{z_\xi} dz \frac{K_\pi\left((1+z)\frac{2\varepsilon_\gamma}{m_{\pi^\pm}}\right)}{\sqrt{\Omega_{\rm M}(1+z)^5 + \Omega_\Lambda(1+z)^2}}, \quad (24)$$

where

$$K_\pi(x) = \frac{\left\{x^2 + 4(1-p)(1-x)\right\}(x^2+p-1)}{1-x} + \left\{2(1-p)(1+p-x) + x^2\right\}\log\frac{1-x}{p}. \quad (25)$$

On the other hand, the spectrum of the internal bremsstrahlung from muons can be calculated as

$$\varepsilon_\gamma^2 \frac{d\Phi_{\mu{\rm IB}}^{\rm ext}}{d\varepsilon_\gamma} = \frac{\varepsilon_\gamma^2}{4\pi}\int_{(1-\xi)t_0}^{t_0} dt' \frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q}\frac{Q_\pi}{Q_{\rm D}} f_{\mu^\pm} \times (1+z)\frac{dB_{\mu{\rm IB}}}{d\varepsilon_\gamma'}(\varepsilon_\gamma'), \quad (26)$$

where Eq.(18) is exploited. Converting the variable of integration to z, we get

$$\varepsilon_\gamma^2 \frac{d\Phi_{\mu{\rm IB}}^{\rm ext}}{d\varepsilon_\gamma} = \frac{\varepsilon_\gamma}{4\pi}\frac{\Omega_Q \rho_{c0}}{\omega_Q \Delta\tau_Q H_0}\frac{Q_\pi}{Q_{\rm D}} f_{\mu^\pm} \times \int_0^{z_\xi} dz \frac{2\left[J_+\left((1+z)\frac{2\varepsilon_\gamma}{m_\mu}\right) + J_+\left((1+z)\frac{2\varepsilon_\gamma}{m_\mu}\right)\right]}{\sqrt{\Omega_{\rm M}(1+z)^5 + \Omega_\Lambda(1+z)^2}}, \quad (27)$$

where Eqs.(A5) and (A6) are used.

We plot the spectra of (22), (24) and (27) in Fig. 3. The integrals in Eqs.(24) and (27) are estimated numerically. Cosmological parameters are set to be $\Omega_M = 0.311$, $\Omega_\Lambda = 0.689$, and $H_0 = 67.7$ km/s/Mpc [51]. We illustrate the cases for $\xi = 0.1$ and 0.5, which correspond respectively to $z_\xi \simeq 0.1$ and 0.8. Also shown are the observations of the extragalactic diffuse gamma-ray emission by EGRET (above 30 MeV) [52], COMPTEL (below 30 MeV) [53], and the Solar Maximum Mission (SMM) Gamma-Ray Spectrometer in the energy range 0.3–8.0 MeV [54]. The origin of this extragalactic MeV gamma-ray background remains unknown, and contributions from seyfert galaxies or blazars are suggested (e.g., Refs. [55, 56]).

We can see a thick line-like emission at $\lesssim 67.5$ MeV, and broad spectra of internal bremsstrahlung with sharp cutoffs at 50 MeV and 30 MeV for muon and pion decays, respectively. They have similar features as in the galactic diffuse gamma-ray spectrum. Future MeV gamma-ray observations could prove or falsify these signals from the Q-ball decay.

VI. CONCLUSION

We have investigated the observational clues of the SUSY Q balls which decay at present. Such Q balls are realized by having very large lepton numbers as the charge of the Q ball in the gauge-mediated SUSY breaking models. The new-type Q balls form and transform

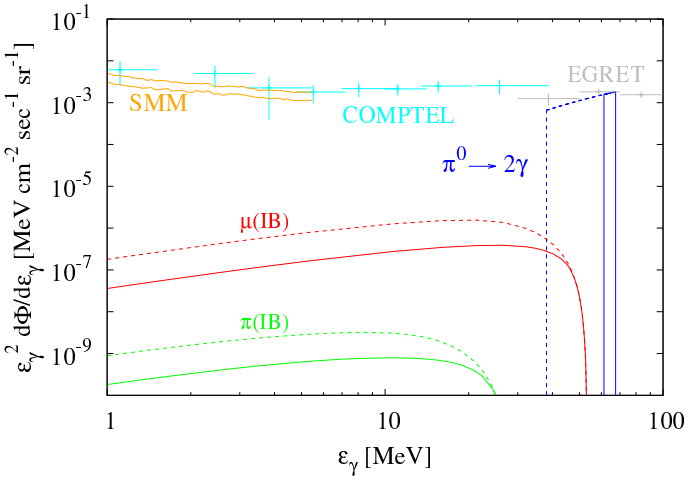


FIG. 3: Extragalactic photon spectra from the Q-ball decay. We display the spectra (22) in blue, (24) in green, and (27) in red lines. Solid and dashed lines denote the cases with $\xi = 0.1$ and 0.5, respectively. Observations of the extragalactic diffuse gamma-ray emission are also shown in cyan (COMPTEL) and gray (EGRET) points, and in orange lines (SMM).

to the gauge-mediation type towards the end of the decay. Then the mass per unit charge ω_Q increases as the charge Q decreases, and subsequently the decay channels into pions become kinematically allowed to create pions at rest. This is the distinctive mechanism of the Q-ball decay to produce pions at rest.

Neutral pions decay into two photons with the energy of half the pion mass, making the sharp emission line at 67.5 MeV. If detected, it would be a smoking gun of the SUSY Q ball since it is the unique feature of the Q-ball decay and no astrophysical sources would create such an emission line. In addition, charged pions decay into muons which further decay radiatively to emit internal bremsstrahlung whose spectrum has a sharp edge at around 50 MeV. We may expect that these photons could be detected by future MeV gamma-ray observations.

Acknowledgments

The authors are grateful to Tsuguo Arakami for useful comments. This work is supported by JSPS KAKENHI Grant Nos. 20H05851(M. K.), 21K03567(M. K.), and 22K14064 (N. T.).

Appendix A: Internal bremsstrahlung

When charged pions decay into muons, they partially emit internal bremsstrahlung. Its differential branching ratio is written as [27–29]

$$\frac{dB_{\pi\mathrm{IB}}}{dxdw} = \frac{\alpha}{2\pi}\frac{1}{(1-p)^2}\frac{1-w+p}{x^2(x+w-1-p)} \times\left[x^2+2(1-x)(1-p)-\frac{2xp(1-p)}{x+w-1-p}\right], \tag{A1}$$

where $\alpha = 1/137$ is the fine structure constant, $p = (m_\mu/m_{\pi^\pm})^2$, $x = 2\varepsilon_\gamma/m_{\pi^\pm}$, and $w = 2E_\mu/m_{\pi^\pm}$ is the normalized muon energy. w- and x-ranges are

$$\begin{aligned} 2\sqrt{p} \le\ & w \le 1+p, \\ 1-\tfrac{w}{2}-\tfrac{\sqrt{w^2-4p}}{2} \le\ & x \le 1-\tfrac{w}{2}+\tfrac{\sqrt{w^2-4p}}{2}, \end{aligned} \tag{A2}$$

respectively. Integration with respective to w over the w-range of (A2) leads to

$$\frac{dB_{\pi\mathrm{IB}}}{dx} = \frac{\alpha}{2\pi}\frac{1}{(1-p)^2}\frac{1}{x} \times\left[\frac{\left\{x^2+4(1-p)(1-x)\right\}(x+p-1)}{1-x} +\left\{2(1-p)(1+p-x)+x^2\right\}\log\frac{1-x}{p}\right]. \tag{A3}$$

On the other hand, polarized muons are produced by the decay of charged pions at rest. The photon spectrum by radiative muon decay is given by [30]

$$\frac{dB(\mu^\pm\to e^\pm\nu\bar\nu\gamma)}{dyd\cos\theta_\gamma} = \frac{1}{y}\left[J_+(y)\left(1\pm P_\mu\cos\theta_\gamma\right) +J_-(y)\left(1\mp P_\mu\cos\theta_\gamma\right)\right], \tag{A4}$$

where $y = 2\varepsilon_\gamma/m_\mu$, P_μ is a magnitude of the muon spin polarization vector, θ_γ is the angle between the muon spin polarization and the photon momentum. $J_+(y)$ and $J_-(y)$ are written respectively as

$$J_+(y) = \frac{\alpha}{6\pi}(1-y)\left[\left(3\log\frac{1-y}{q}-\frac{17}{2}\right) +\left(-3\log\frac{1-y}{q}+7\right)(1-y) +\left(2\log\frac{1-y}{q}-\frac{13}{3}\right)(1-y)^2\right], \tag{A5}$$

$$J_-(y) = \frac{\alpha}{6\pi}(1-y)^2\left[\left(3\log\frac{1-y}{q}-\frac{31}{4}\right) +\left(-4\log\frac{1-y}{q}+\frac{29}{3}\right)(1-y) +\left(2\log\frac{1-y}{q}-\frac{55}{12}\right)(1-y)^2\right], \tag{A6}$$

where $q = (m_e/m_\mu)^2$. Since the produced muons have no specific direction, we must integrate Eq.(A4) with respect to θ_γ over the whole angles, and hence obtain the spectrum as

$$\begin{aligned}\frac{dB_{\mu\mathrm{IB}}}{dy} &= \frac{2}{y}\left[J_+(y)+J_-(y)\right] \\ &= \frac{\alpha}{2\pi}\frac{2(1-y)}{3y} \\ &\quad\times\left[-\frac{17}{2}-\frac{3}{4}(1-y)+\frac{16}{3}(1-y)^2-\frac{55}{12}(1-y)^3\right. \\ &\quad\left.+\left\{3-2(1-y)^2+2(1-y)^3\right\}\log\frac{1-y}{q}\right]. \quad \text{(A7)}\end{aligned}$$

Appendix B: Rough constraint on Ω_Q

Q-ball decay produces positrons with energy $\lesssim \omega_Q$ which may annihilate with electrons at the galactic center to generate 511 keV gamma rays. Since the morphology of the dark matter halo in our galaxy is still not known, we make very rough estimate for the upper bound of the Q-ball abundance according to Ref. [57]. Assuming the half of the total 511 keV flux is emitted from an angular region of 9° circle, we have

$$\frac{M_{(<9^\circ)}\Omega_Q}{\omega_Q\Delta\tau_Q}f_{e^+}\left[\frac{f}{4}+(1-f)\right]=\frac{1}{2}\Phi_{511}4\pi R_{\mathrm{GC}}^2, \quad \text{(B1)}$$

where f is the fraction of positrons which annihilate via positronium, $\Phi_{511}(\simeq 10^{-3}\ \mathrm{cm^{-2}s^{-1}})$ is the observed total flux of 511 keV line, and $\Delta\tau_Q = \xi\tau_Q$. We assume $\tau_Q \simeq t_0$. The total mass within the 9° circle is given by

$$M_{(<9^\circ)} = \int_{(<9^\circ)}\rho_{\mathrm{DM}}(r)4\pi r^2 dr. \quad \text{(B2)}$$

Therefore, we obtain the constraint on the density parameter as

$$\begin{aligned}\frac{\Omega_Q}{\xi} &\lesssim 3.0\times10^{-7}\left(\frac{\Phi_{511}}{10^{-3}\ \mathrm{cm^{-2}s^{-1}}}\right) \\ &\quad\times\left(\frac{R_{\mathrm{GC}}}{8.23\ \mathrm{kpc}}\right)^2\left(\frac{\omega_Q}{3.6\mathrm{MeV}}\right)\left(\frac{\tau_Q}{t_0}\right). \quad \text{(B3)}\end{aligned}$$

[1] S. R. Coleman, Nucl. Phys. B **262**, no.2, 263 (1985).
[2] A. Kusenko, Phys. Lett. B **405**, 108 (1997).
[3] A. Kusenko and M. E. Shaposhnikov, Phys. Lett. B **418**, 46-54 (1998).
[4] S. Kasuya and M. Kawasaki, Phys. Rev. D **61**, 041301 (2000).
[5] I. Affleck and M. Dine, Nucl. Phys. B **249**, 361-380 (1985).
[6] A. Kusenko, V. Kuzmin, M. E. Shaposhnikov and P. G. Tinyakov, Phys. Rev. Lett. **80**, 3185-3188 (1998).
[7] J. Arafune, T. Yoshida, S. Nakamura and K. Ogure, Phys. Rev. D **62**, 105013 (2000).
[8] S. Kasuya and M. Kawasaki, Phys. Rev. D **64**, 123515 (2001).
[9] S. Kasuya, M. Kawasaki and T. T. Yanagida, PTEP **2015**, no.5, 053B02 (2015).
[10] S. Kasuya and F. Takahashi, Phys. Rev. D **72**, 085015 (2005).
[11] S. Kasuya and F. Takahashi, JCAP **11**, 019 (2007).
[12] S. Kasuya, M. Kawasaki and K. Murai, JCAP **05**, 053 (2023).
[13] M. Kawasaki and K. Murai, JCAP **01**, 050 (2024).
[14] T. Gherghetta, C. F. Kolda and S. P. Martin, Nucl. Phys. B **468**, 37-58 (1996).
[15] M. Dine, L. Randall and S. D. Thomas, Nucl. Phys. B **458**, 291-326 (1996).
[16] A. de Gouvea, T. Moroi and H. Murayama, Phys. Rev. D **56**, 1281-1299 (1997).
[17] S. Kasuya and M. Kawasaki, Phys. Rev. Lett. **85**, 2677-2680 (2000).
[18] J. Hisano, M. M. Nojiri and N. Okada, Phys. Rev. D **64**, 023511 (2001).
[19] S. Kasuya and M. Kawasaki, Phys. Rev. D **62**, 023512 (2000).
[20] T. Hiramatsu, M. Kawasaki and F. Takahashi, JCAP **06**, 008 (2010).
[21] K. Enqvist and J. McDonald, Nucl. Phys. B **538**, 321-350 (1999).
[22] A. G. Cohen, S. R. Coleman, H. Georgi and A. Manohar, Nucl. Phys. B **272**, 301-321 (1986).
[23] S. Kasuya and M. Kawasaki, Phys. Rev. D **80**, 123529 (2009).
[24] S. P. Martin, Adv. Ser. Direct. High Energy Phys. **18**, 1-98 (1998).
[25] J. F. Navarro, C. S. Frenk and S. D. M. White, Astrophys. J. **462**, 563-575 (1996).
[26] H. W. Leung, J. Bovy, J. T. Mackereth, J. A. S. Hunt, R. R. Lane and J. C. Wilson, Mon. Not. Roy. Astron. Soc. **519**, no.1, 948-960 (2022).
[27] G. Bressi, G. Carugno, E. Conti, A. T. Meneguzzo, S. Cerdonio and D. Zanello, Nucl. Phys. B **513**, 555-572 (1998).
[28] S. G. Brown and S. A. Bludman, Phys. Rev. **136**, B1160-B1163 (1964).
[29] D. E. Neville, Phys. Rev. **124**, 2037-2046 (1961).
[30] Y. Kuno and Y. Okada, Rev. Mod. Phys. **73**, 151-202 (2001).
[31] A. W. Strong, H. Bloemen, R. Diehl, W. Hermsen and V. Schoenfelder, Astrophys. Lett. Commun. **39**, 209 (1999).
[32] A. W. Strong, I. V. Moskalenko and O. Reimer, Astrophys. J. **613**, 962-976 (2004).
[33] D. J. Thompson, Rept. Prog. Phys. **71**, 116901 (2008).
[34] M. Ackermann *et al.* [Fermi-LAT], Astrophys. J. **750**, 3 (2012).
[35] E. Orlando, Mon. Not. Roy. Astron. Soc. **475**, no.2, 2724-2742 (2018).
[36] N. Tsuji, Y. Inoue, H. Yoneda, R. Mukherjee and H. Odaka, Astrophys. J. **943**, no.1, 48 (2023).
[37] K. K. Boddy and J. Kumar, Phys. Rev. D **92**, no.2,

023533 (2015).
[38] T. Binder, S. Chakraborti, S. Matsumoto and Y. Watanabe, JHEP **01**, 106 (2023).
[39] J. G. Christy, J. Kumar and A. Rajaraman, Phys. Rev. D **106**, no.12, 123021 (2022).
[40] K. Fang, J. S. Gallagher and F. Halzen, Astrophys. J. **933**, no.2, 190 (2022).
[41] L. Bouchet, A. W. Strong, T. A. Porter, I. V. Moskalenko, E. Jourdain and J. P. Roques, Astrophys. J. **739**, 29 (2011).
[42] A. De Angelis *et al.* [e-ASTROGAM], JHEAp **19**, 1-106 (2018).
[43] H. Fleischhack, PoS **ICRC2021**, 649 (2021).
[44] T. Aramaki, P. Hansson Adrian, G. Karagiorgi and H. Odaka, Astropart. Phys. **114**, 107-114 (2020).
[45] A. Takada, T. Takemura, K. Yoshikawa, Y. Mizumura, T. Ikeda, Y. Nakamura, K. Onozaka, M. Abe, K. Hamaguchi and H. Kubo, *et al.* Astrophys. J. **930**, no.1, 6 (2022).
[46] A. Moiseev, PoS **ICRC2021**, 648 (2021).
[47] G. Lucchetta, M. Ackermann, D. Berge and R. Bühler, JCAP **08**, no.08, 013 (2022).
[48] J. A. Tomsick, A. Zoglauer, C. Sleator, H. Lazar, J. Beechert, S. Boggs, J. Roberts, T. Siegert, A. Lowell and E. Wulf, *et al.*, BAAS **51**, 98 (2019).
[49] M. Kawasaki and T. Yanagida, Phys. Lett. B **399**, 45-48 (1997).
[50] T. Asaka, J. Hashiba, M. Kawasaki and T. Yanagida, Phys. Rev. D **58**, 083509 (1998).
[51] N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020) [erratum: Astron. Astrophys. **652**, C4 (2021)].
[52] A. W. Strong, I. V. Moskalenko and O. Reimer, Astrophys. J. **613**, 956-961 (2004).
[53] G. Weidenspointer *et al.*, AIP Conf. Proc. **510**, 467-475 (2000).
[54] K. Watanabe, M. D. Leising, G. H. Share, AIP Conf. Proc. **510**, 471-470 (2000).
[55] Y. Inoue, T. Totani and Y. Ueda, Astrophys. J. Lett. **672**, L5 (2008).
[56] M. Ajello, L. Costamante, R. M. Sambruna, N. Gehrels, J. Chiang, A. Rau, A. Escala, J. Greiner, J. Tueller and J. V. Wall, *et al.* Astrophys. J. **699**, 603-625 (2009).
[57] D. Hooper and L. T. Wang, Phys. Rev. D **70**, 063506 (2004).

Electroweak Phase Transition in Singlet Extensions of The Standard Model with Dimension-Six Operators

V.K. Oikonomou[1,2], Apostolos Giovanakis[1]
[1)] *Department of Physics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece and*
[2)] *L.N. Gumilyov Eurasian National University - Astana, 010008, Kazakhstan*

The significance of the electroweak phase transition is undeniable, and although initially it was believed that it was second-order, it is now believed that it is a first-order transition. However, it is not a strong first-order phase transition in the context of the Standard Model and the remedy to this issue is to use the Higgs portal and directly couple the Higgs to a hidden scalar sector. This can result in a strong electroweak phase transition, while the couplings to a hidden scalar are constrained by several phenomenological constraints, such as the sphaleron rate criterion and the branching ratio of the Higgs to invisible channels. In this work, we consider the standard singlet extensions of the Standard Model, including dimension-six non-renormalizable operators that couple a real singlet scalar field with the Higgs doublet. As a result, we examine the effects of those Higgs-singlet couplings on the electroweak phase transition. The effective theory, where the non-renormalizable couplings originate from, is considered to be active beyond 15 TeV. As we show, the Universe experiences a two-step electroweak phase transition, a primary phase transition in the singlet sector at a high temperature, and then a subsequent first-order phase transition from the singlet vacuum to the electroweak vacuum. The singlet's phase transition can either be second-order or first-order, depending on the singlet mass and its couplings to the Higgs. In particular, we show that the dimension-six operator assists in generating a strong electroweak phase transition in regions of the parameter space that were excluded in the previous singlet extensions of the Standard Model. This is further apparent for low singlet masses $m_S < m_H/2$ which are rarely taken into account in the literature due to the invisible branching ratio of the Higgs boson. In some limited cases in the parameter space, the electroweak phase transition is weakened by the presence of the higher order operator.

PACS numbers: 04.50.Kd, 95.36.+x, 98.80.-k, 98.80.Cq,11.25.-w

arXiv:2403.01591v1 [hep-ph] 3 Mar 2024

I. INTRODUCTION

The most mysterious eras of our Universe are the primordial eras, and specifically the inflationary era and the reheating era. These two eras are strongly related and to date no direct evidence of the occurrence of these two exists. Much light on these two mysterious eras is expected to be shed by the upcoming stage 4 Cosmic Microwave Background (CMB) experiments [1, 2] and of course by the future interferometric gravitational wave experiments, like the LISA, BBO, DECIGO, and the Einstein telescope [3–11]. The stage 4 CMB experiments will directly probe the B-modes of the CMB polarization, if they exist, and the interferometers will probe the stochastic primordial gravitational waves. The latter is already confirmed to exist at nanohertz frequencies, in 2023 by the NANOGrav collaboration [12] and by the EPTA [13], the PPTA [14] and the CPTA [15] collaborations. The future seems to be fruitful in this research line.

Moreover, due to the complete absence of particle detections at the LHC since the detection of the Higgs particle, and recall that the LHC functions at a center-of-mass energy 13.6 TeV, modern high energy physics relies on gravitational wave experiments and astrophysical observations in order to address fundamental problems related to baryogenesis, electroweak phase transition, and dark matter. However we must mention that the particles colliding at the LHC are basically hadrons, so have some non negligible form factors, so the actual theories and energies probed at the LHC might actually correspond to lower energies than 13.6 TeV center-of-mass. To date, the electroweak phase transition is believed to be a first-order phase transition [16], although earlier studies indicated that it was a second-order phase transition [17]. The order of the electroweak phase transition has triggered a great deal of heated debate in the last decades. In the past years, it was shown in numerous studies using perturbative calculations that the electroweak phase transition is a first-order phase transition [16]. On the other hand, it is understood that this topic is still controversial. In fact, recent simulations and studies, indicate that with the Higgs mass being ~ 125 GeV, the phase transition is even a crossover, and not a first order phase transition, in the customary sense, see for example [18, 19] and for an older lattice simulation, see [17] which points out to a second order phase transition. A first-order phase transition would provide non-equilibrium conditions and thus one of the Sakharov criteria [20] would hold true and the baryogenesis could occur during the electroweak phase transition. First-order phase transitions also produce a stochastic gravitational wave background, and are thus observationally important [11, 21–44]. The problem with the electroweak phase transition in the Standard Model (SM) is that it is not strong enough, and thus singlet extensions

of the SM active through the Higgs portal have been proposed to solve it [25, 28, 42–70]. In this work, we consider the standard real singlet extension, including higher order non-renormalizable interactions between the Higgs field and a real singlet scalar field and we concretely examine the effects of these Higgs-singlet couplings on the electroweak phase transition. The dimension-six non-renormalizable operators are considered to originate from a weakly coupled effective theory that remains active well beyond 15 TeV. As we demonstrate, the Universe experiences two phase transitions, a primary phase transition in the singlet scalar sector at high temperatures, and then the ordinary electroweak phase transition. The singlet scalar phase transition can either be a second-order or a first-order phase transition, which depends solely on the singlet scalar mass and its couplings to the Higgs boson. More importantly, we demonstrate that due to the presence of the dimension-six non-renormalizable operator, the electroweak phase transition is stronger in some regions of the parameter space, which were excluded in the standard singlet extensions of the SM without the higher order operators. However, we show that, in some limited cases for some parameter values, the electroweak phase transition is actually weakened by the presence of the higher order operator.

This paper is organized as follows: In section II, we present the effective potential for the SM enriched with the singlet extension, including the higher order non-renormalizable operators. In section III, we discuss the constraints from the electroweak baryogenesis which must be taken into account in our analysis, in sections IV and V, we showcase the allowed parameter space of our model and we also discuss various phenomenological constraints and constraints from invisible Higgs decay. Finally, in section VI, we study and present all the different possibilities for the electroweak phase transition in our singlet-extended SM, and the conclusions follow at the end of the article.

II. SM EFFECTIVE POTENTIAL WITH SINGLET EXTENSIONS AND HIGHER ORDER OPERATORS

We extend the SM by introducing a real singlet scalar field ϕ [25, 28, 42–70] equipped with a $\mathbb{Z}_2$ discrete symmetry, under which $\phi \to -\phi$ and all other SM fields remain unaffected. In addition, we shall assume that a higher dimensional operator of the singlet scalar field is weakly coupled to the Higgs sector. This dimension six higher order non-renormalizable operator originates from an effective theory active at a scale M, which will be assumed to be way higher than the electroweak scale, of the order $M = 15 - 100$ TeV, a fact that is further motivated by the lack of new particle observations in the LHC, beyond the electroweak symmetry breaking scale [25]. The dimensionless coupling λ is the Wilson coefficient of the higher order effective theory and we shall assume that the theory is strongly coupled. In general, the coefficient of the higher order term is considered small such that $\lambda/M^2 < 10^{-4}$ GeV^{-2} and the contribution of the higher dimensional operator to the invisible Higgs decays is omitted. The tree-level potential for the doublet Higgs field and the singlet scalar field is given by,

$$V_0(H,\phi) = -\mu_H^2|H|^2 + \lambda_H|H|^4 - \frac{\mu_S^2}{2}\phi^2 + \frac{\lambda_S}{4}\phi^4 + \lambda_{HS}|H|^2\phi^2 + \frac{\lambda}{M^2}|H|^2\phi^4 \tag{1}$$

where $m_H = \sqrt{2}\mu_H$ is the Higgs boson mass, $\lambda_H > 0$ is the Higgs self-coupling and λ_{HS} is the Higgs-singlet interaction coupling which is assumed to be positive. In this work, we consider one of the simplest higher-dimensional operators in the real singlet extensions to the Standard Model. The $D > 6$ operators are suppressed and it can be easily shown that their effect on the electroweak phase transition is not so strong for valid values of the Wilson coefficients. In principle, dimension eight and higher order operators can be added, due to some non-perturbative physics motivated Lagrangian terms of the form $\sim |H|^2\cos(\frac{\phi}{M})$, but the dominant term is always the dimension 6 operator. The tree-level potential of the singlet scalar field is determined by the four parameters μ_S, λ_S, λ_{HS} and λ and the energy scale M. The SM Higgs doublet is parameterized as follows,

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix}\chi_1 + i\chi_2 \\ h + i\chi_3\end{pmatrix} \tag{2}$$

where h is the Higgs boson, χ_1, χ_2, χ_3 are three Goldstone bosons, and $v = \mu_H/\sqrt{\lambda_H}$ is the electroweak symmetry breaking minimum of the Higgs scalar field h at zero temperature, without the singlet sector. The tree-level potential (1) in the unitary gauge takes the following form,

$$V_0(h,\phi) = -\frac{\mu_H^2}{2}h^2 + \frac{\lambda_H}{4}h^4 - \frac{\mu_S^2}{2}\phi^2 + \frac{\lambda_S}{4}\phi^4 + \frac{\lambda_{HS}}{2}h^2\phi^2 + \frac{\lambda}{2M^2}h^2\phi^4 \tag{3}$$

which is shown in Fig. 1. In the context of the SM, we consider only the dominant contributions in the one-loop finite temperature effective potential coming from the gauge bosons, the top quark (the heaviest fermion), the Higgs and

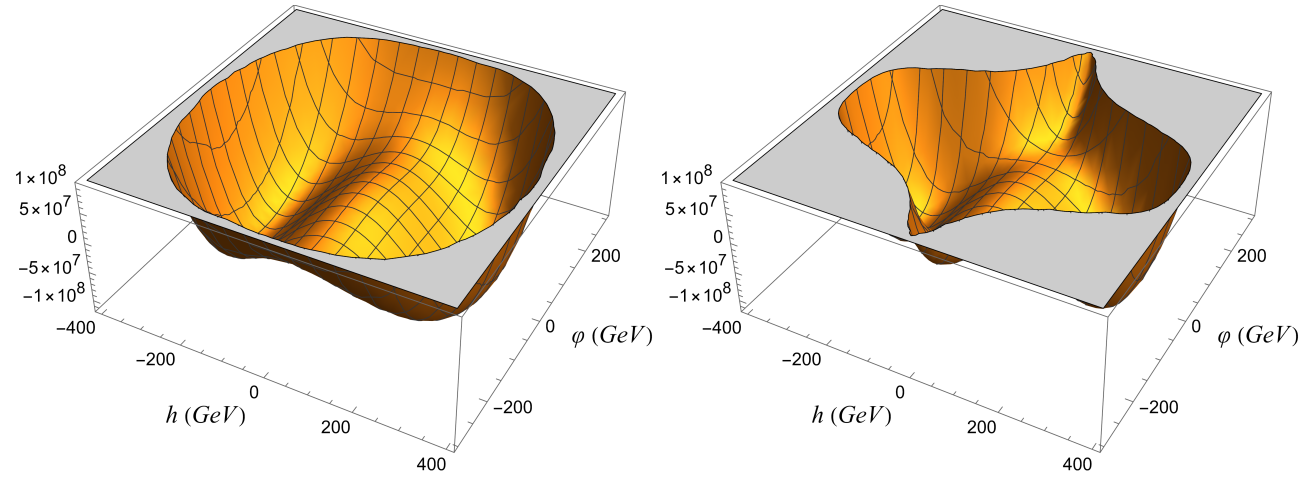


FIG. 1. The tree-level potential in the field space for $\lambda = 0$ (left) and $\lambda/M^2 = 2 \times 10^{-5}$ GeV^{-2} (right), where the global minimum is the electroweak vacuum. We assumed that $m_S = 62.5$ GeV, $\lambda_{HS} = 0.1$ and $a = 0.1$.

Goldstone bosons. The effective masses of the Higgs boson, the singlet scalar field and those fields which are coupled to the background fields h and ϕ are,

$$m_h^2(h,\phi) = -\mu_H^2 + 3\lambda_H h^2 + \lambda_{HS}\phi^2 + \frac{\lambda}{M^2}\phi^4 \tag{4}$$

$$m_\chi^2(h,\phi) = -\mu_H^2 + \lambda_H h^2 + \lambda_{HS}\phi^2 + \frac{\lambda}{M^2}\phi^4 \tag{5}$$

$$m_S^2(h,\phi) = -\mu_S^2 + 3\lambda_S \phi^2 + \lambda_{HS} h^2 + \frac{6\lambda}{M^2} h^2\phi^2 \tag{6}$$

$$m_W^2(h) = \frac{g^2}{4}h^2 \tag{7}$$

$$m_Z^2(h) = \frac{g^2 + g'^2}{4}h^2 \tag{8}$$

$$m_t^2(h) = \frac{y_t^2}{2}h^2 \tag{9}$$

where g,g' and y_t are the $SU(2)_L$, $U(1)_Y$ and top quark Yukawa couplings, respectively, and $m_h = 125$ GeV, $m_W = 80.4$ GeV, $m_Z = 91.2$ GeV and $m_t = 173$ GeV at the current vacuum state of the Universe ($h = v$).

The full effective potential is defined as the sum of the tree-level potential and the one-loop finite-temperature effective potential which splits into a zero-temperature and a temperature-dependent part, which will be presented explicitly in the next two sections.

A. Zero Temperature Corrections

The zero-temperature one-loop contribution to the effective potential is called the Coleman-Weinberg (CW) potential [71] and is computed as the sum of all one-particle irreducible Feynman diagrams with zero external momenta and a single loop. In the $\overline{\text{MS}}$ renormalization scheme, it is written as,

$$V_1^i(h,\phi) = (-1)^{F_i} n_i \frac{m_i^4(h,\phi)}{64\pi^2}\left[\ln\left(\frac{m_i^2(h,\phi)}{\mu_R^2}\right) - C_i\right] \tag{10}$$

where $i = \{h, \chi, \phi, W, Z, t\}$ counts the particles that mainly contribute to the CW potential and couple with the Higgs doublet and the singlet scalar field, $F_i = 1$ (0) for fermions (bosons), n_i is the number of degrees of freedom of each particle i, μ_R denotes the renormalization scale, and $C_i = 3/2$ (5/6) for scalars and fermions (gauge bosons). The degrees of freedom of each particle i are,

$$n_h = 1, \quad n_\chi = 3, \quad n_\phi = 1, \quad n_W = 6, \quad n_Z = 3, \quad n_t = 12 \tag{11}$$

The choice of the renormalization scale leads to an uncertainty in the critical temperature T_c and to other related quantities [66, 72–74]. Different schemes can be followed to compute the CW potential and handle this uncertainty, such as the on-shell (OS) and on-shell-like schemes (OS-like). In the literature, the OS-like scheme is commonly adopted, since it has a fixed prescription for the renormalization scales for each particle and the tree-level relations among the parameters are valid at higher loop corrections, like in the OS scheme. For instance, in this scheme, the CW potential takes the form,

$$V_1^{(OS)}(h,\phi) = \sum_i \frac{(-1)^{F_i} n_i}{64\pi^2} \left[m_i^4(h,\phi) \left(\ln \frac{m_i^2(h,\phi)}{m_i^2(v,0)} - \frac{3}{2} \right) + 2 m_i^2(h,\phi) m_i^2(v,0) \right], \tag{12}$$

However, both OS schemes suffer from an infrared divergence which originates from the Goldstone bosons as they acquire a zero mass at zero temperature. This subtlety can be rectified, by the proposed methods in Refs. [75, 76], but in our study, we extensively work in the $\overline{\text{MS}}$ scheme.

In Ref. [66], it was concluded that varying μ_R from $m_t/2$ to $2m_t$ the OS-like and $\overline{\text{MS}}$ schemes agreed with each other within the scale uncertainties which were $(3.8 - 6.2)\%$ in the critical temperature T_c and $(10 - 23)\%$ in the ratio v_c/T_c. In the case of $\mu_R = m_t/2$ the two results were approximately identical with the highest accuracy. In our study we have chosen $\mu_R = m_t/2$. A similar analysis was also developed in Ref. [72]. Lastly, the renormalization scale-dependence could be eliminated by the Renormalization Group Equations (RGE) improvement for the CW potential [77–81], but this approach is left to be implemented in future studies.

B. Finite Temperature Corrections to the Effective Potential

The temperature dependence of the finite-temperature one-loop potential is interpreted in terms of the free energy of an ideal gas. The effective potential at finite temperature contains the effective potential at zero temperature which was presented in the previous section. Thus, we focus on its temperature-dependent component[1] for each particle i and it is equal to,

$$V_T^i(h,\phi,T) = (-1)^{F_i} \frac{n_i T^4}{2\pi^2} \int_0^\infty dx x^2 \ln \left[1 - (-1)^{F_i} \exp\left(-\sqrt{x^2 + \frac{m_i^2(h,\phi)}{T^2}} \right) \right] \tag{13}$$

where the thermal functions integrals are defined as follows,

$$J_{B/F}\left(y^2\right) = \int_0^\infty dx x^2 \ln \left[1 \mp \exp\left(-\sqrt{x^2 + y^2} \right) \right] \tag{14}$$

where the subscript B (F) stands for bosons (fermions). The thermal functions can be computed numerically, but they also admit a high-temperature expansion (for $|y^2| \ll 1$) [81], which will be applied in this work, and is given by,

$$J_B(y^2) = -\frac{\pi^4}{45} + \frac{\pi^2}{12} y^2 - \frac{\pi^3}{6} y^3 - \frac{1}{32} y^4 \log\left(\frac{y^2}{a_b} \right) + O(y^6) \tag{15}$$

$$J_F(y^2) = -\frac{7\pi^4}{360} - \frac{\pi^2}{12} y^2 - \frac{1}{32} y^4 \log\left(\frac{y^2}{a_f} \right) + O(y^6) \tag{16}$$

where $a_b = \pi^2 \exp\left(3/2 - 2\gamma_E\right)$, $a_f = 16\pi^2 \exp\left(3/2 - 2\gamma_E\right)$, ζ denotes the Riemann ζ-function, and γ_E is the Euler-Mascheroni constant. Numerical analysis shows that the high-temperature expansion up to the logarithmic term is

[1] This component vanishes at zero temperature.

accurate to better than 5% for $y \leq 1.6$ for fermions and $y \leq 2.2$ for bosons [82, 83]. Moreover, this approximation for both the potential and its derivatives also agrees with the exact form to better than approximately 10% for the values $y \lesssim (1-3)$, depending on the function [84].

The symmetry restoration at high temperature implies that conventional perturbation theory breaks down near the critical temperature [81]. As a matter of fact, if perturbation theory were to remain valid, considering the temperature-independence of the tree-level potential, temperature-dependent radiative corrections should be incapable of restoring the symmetry. In particular, it can be proved that the one-loop approximation in terms of small coupling constants, breaks down at a high temperature, due to the appearance of infrared divergences for the zero Matsubara modes of bosonic degrees of freedom [40]. Therefore, higher loop corrections that are contained in the daisy diagrams have to be resummed in the perturbative expansion. In principle, the so-called daisy resummation is a method aiming to sum all powers of the coupling constant resulting in a theory where the effective mass is $m_i^2 \to M_i^2 = m_i^2 + \Pi_i(T)$, which is called the thermal mass, and $\Pi_i(T)$ is the temperature-dependent self-energy corresponding to the one-loop resummed diagrams to leading powers of the temperature. This replacement in the full temperature-dependent effective potential is usually described by the Parwani scheme [85] and leads to,

$$V_{\text{eff}}(h,\phi,T) = V_0(h,\phi) + \sum_i \left[V_1^i\left(m_i^2(h,\phi) + \Pi_i(T)\right) + V_T^i\left(m_i^2(h,\phi) + \Pi_i(T), T\right)\right] \tag{17}$$

A similar method, developed by Arnold and Espinosa [86], resums only the zero Matsubara modes that cause the infrared divergence. In the Arnold-Espinosa scheme, the resummed full effective potential is written as follows,

$$V_{\text{eff}}(h,\phi,T) = V_0(h,\phi) + \sum_i \left[V_1^i\left(m_i^2(h,\phi)\right) + V_T^i\left(m_i^2(h,\phi), T\right) + V_{ring}^i\left(m_i^2(h,\phi), T\right)\right] \tag{18}$$

The last term is added to incorporate the daisy resummation[2] and it is given by,

$$V_{ring}^i\left(m_i^2(h,\phi), T\right) = \frac{\overline{n}_i T}{12\pi}\left[m_i^3(h,\phi) - \left(m_i^2(h,\phi) + \Pi_i(T)\right)^{3/2}\right] \tag{19}$$

for the particles $i = \{h, \chi, \phi, W, Z, \gamma\}$ and $\overline{n}_i = \{1, 3, 1, 2, 1, 1\}$ is the modified number of degrees of freedom which takes into account that only the longitudinal polarizations of the gauge bosons contribute to the temperature-dependent self-energy [16, 86]. A comparison between these schemes is presented in Refs. [72, 73], where it is apparently shown that the numerical differences between the two schemes regarding the critical temperature and the ratio v_c/T_c are relatively small. Finally, the Arnold-Espinosa scheme is widely preferred in beyond the Standard Model physics and it will be adopted throughout this paper.

In this model, the temperature-dependent self-energy for the scalar fields are computed as,

$$\Pi_h(T) = \Pi_\chi(T) = \left(\frac{3g^2}{16} + \frac{g'^2}{16} + \frac{y_t^2}{4} + \frac{\lambda_H}{2} + \frac{\lambda_{HS}}{12}\right) T^2 \tag{20}$$

$$\Pi_S(T) = \left(\frac{\lambda_S}{4} + \frac{\lambda_{HS}}{3} + \frac{\lambda v^2}{2M^2}\right) T^2 \tag{21}$$

The thermal corrections of the gauge bosons require special treatment as the transverse gauge fields have zero thermal corrections [16]. According to Ref. [16], we compute the temperature-dependent self-energy of the longitudinal gauge bosons in the high-temperature limit,

$$\Pi_{W_L}(T) = \frac{11}{6} g^2 T^2 \tag{22}$$

$$\Pi_{Z_L}(h,T) = \frac{11}{6} g^2 T^2 - \frac{g'^2}{4} h^2 \tag{23}$$

$$\Pi_{\gamma_L}(h,T) = \frac{11}{6} g'^2 T^2 + \frac{g'^2}{4} h^2 \tag{24}$$

---

[2] This scheme is equivalent to the Parwani scheme in the high-temperature limit [84].

where the gauge boson mass matrix in the basis $(A^1_\mu, A^2_\mu, A^3_\mu, B_\mu)$ is non-diagonal

$$M^2_{GB}(h,T) = \begin{pmatrix} \frac{1}{4}g^2h^2 + \frac{11}{6}g^2T^2 & 0 & 0 & 0 \\ 0 & \frac{1}{4}g^2h^2 + \frac{11}{6}g^2T^2 & 0 & 0 \\ 0 & 0 & \frac{1}{4}g^2h^2 + \frac{11}{6}g^2T^2 & -\frac{1}{4}gg'h^2 \\ 0 & 0 & -\frac{1}{4}gg'h^2 & \frac{1}{4}g'^2h^2 + \frac{11}{6}g'T^2 \end{pmatrix} \tag{25}$$

Nevertheless, the photon and the Z boson are not mass eigenstates at high temperature, since there is an additional mixing term between the Z boson and the photon [73]. Therefore, the gauge boson mass matrix has the following eigenvalues for the longitudinal photon and Z boson:

$$M^2_{Z_L} = \frac{1}{2}\left[\frac{1}{4}\left(g^2+g'^2\right)h^2 + \frac{11}{6}\left(g^2+g'^2\right)T^2 + \sqrt{\left(g^2-g'^2\right)^2\left(\frac{1}{4}h^2+\frac{11}{6}T^2\right)^2 + \frac{g^2g'^2}{4}h^4}\right] \tag{26}$$

$$M^2_{\gamma_L} = \frac{1}{2}\left[\frac{1}{4}\left(g^2+g'^2\right)h^2 + \frac{11}{6}\left(g^2+g'^2\right)T^2 - \sqrt{\left(g^2-g'^2\right)^2\left(\frac{1}{4}h^2+\frac{11}{6}T^2\right)^2 + \frac{g^2g'^2}{4}h^4}\right] \tag{27}$$

In the limit $m^2_W(h)/T^2 \ll 1$, the full resummed mass eigenvalues (26) and (27) can be effectively approximated by those in Eqs. (23) and (24). The numerical difference between these expressions is small, which indicates that one could treat the photon and Z boson as mass eigenstates [73].

Finally, it is essential to mention that the full effective potential can have imaginary contributions in the case of negative squared effective masses. This holds true especially for the scalar fields due to the logarithmic and cubic terms in the thermal functions (15) and (16), since the gauge bosons and the top quark have always a positive squared effective mass. In general, the effective mass in the logarithmic term is cancelled by its counterpart in the one-loop zero-temperature correction. Moreover, the daisy resummation could potentially cure the imaginary part originating from the cubic term which signals the breakdown of the perturbative expansion[3]. Nevertheless, the $m^2_i(h,\phi) + \Pi_i(T)$ in the ring correction can be negative for certain temperatures and field values. As a result, the effective potential is still complex and we consider only the real part of the full effective potential and ensure the field's stability during the phase transition, provided its imaginary counterpart remains sufficiently insignificant. A relevant discussion can be found in Ref. [37].

To sum up, in the Arnold-Espinosa scheme, the full effective potential that describes the dynamics of the phase transition of our singlet-extended SM reads,

$$\begin{aligned} V_{\text{eff}}(h,\phi,T) = &-\frac{\mu_H^2}{2}h^2 + \frac{\lambda_H}{4}h^4 - \frac{\mu_S^2}{2}\phi^2 + \frac{\lambda_S}{4}\phi^4 + \frac{\lambda_{HS}}{2}h^2\phi^2 + \lambda\frac{h^2\phi^4}{2M^2} \\ &+ \sum_i (-1)^{F_i} n_i \frac{m_i^4(h,\phi)}{64\pi^2}\left[\ln\left(\frac{m_i^2(h,\phi)}{\mu_R^2}\right) - C_i\right] \\ &+ \sum_i \frac{n_i T^4}{2\pi^2} J_B\left(\frac{m_i^2(h,\phi)}{T^2}\right) - 12\frac{T^4}{2\pi^2}J_F\left(\frac{m_t^2(h)}{T^2}\right) \\ &+ \sum_i \frac{\overline{n}_i T}{12\pi}\left[m_i^3(h,\phi) - \left(M_i^2(h,\phi,T)\right)^{3/2}\right] \end{aligned} \tag{28}$$

where $i = \{h, \phi, \chi, W, Z, \gamma\}$ corresponds to the bosons in the extended SM.

Our main results are based on the high-temperature expansion (15) and (16) and its validity is primarily checked by the condition on the value of the ratio M_i/T, which was previously explained. If these conditions are violated, we numerically compute the full effective potential (28). Therefore, according to the high-temperature expansion, the

[3] Weinberg and Wu claimed that the imaginary part of the effective potential could also be interpreted as a decay rate of a state of the scalar fields [87]

full effective potential (28) becomes,

$$
\begin{aligned}
V_{\rm eff}^{\rm HT}(h,\phi,T) = & -\frac{\mu_H^2}{2}h^2 + \frac{\lambda_H}{4}h^4 - \frac{\mu_S^2}{2}\phi^2 + \frac{\lambda_S}{4}\phi^4 + \frac{\lambda_{HS}}{2}h^2\phi^2 + \frac{\lambda}{2M^2}h^2\phi^4 \\
& \frac{m_h^2(h,\phi)}{24}T^2 - \frac{T}{12\pi}\left[M_h^2(h,\phi,T)\right]^{3/2} + \frac{m_h^4(h,\phi)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{3}{2}\right] + \\
& \frac{3m_\chi^2(h,\phi)}{24}T^2 - \frac{3T}{12\pi}\left[M_\chi^2(h,\phi,T)\right]^{3/2} + \frac{3m_\chi^4(h,\phi)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{3}{2}\right] + \\
& \frac{m_\phi^2(h,\phi)}{24}T^2 - \frac{T}{12\pi}\left[M_\phi^2(h,\phi,T))\right]^{3/2} + \frac{m_\phi^4(h,\phi)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{3}{2}\right] + \\
& \frac{6m_W^2(h)}{24}T^2 - \frac{4T}{12\pi}m_W^3(h) - \frac{2T}{12\pi}\left[M_{W_L}^2(h,T)\right]^{3/2} + \frac{6m_W^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{5}{6}\right] + \\
& \frac{3m_Z^2(h)}{24}T^2 - \frac{2T}{12\pi}m_Z^3(h) - \frac{T}{12\pi}\left[M_{Z_L}^2(h,T)\right]^{3/2} + \frac{3m_Z^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{5}{6}\right] + \\
& \frac{12m_t^2(h)}{48}T^2 - \frac{12m_t^4(h)}{64\pi^2}\left[\ln\left(\frac{a_f T^2}{\mu_R^2}\right) - \frac{3}{2}\right] - \frac{T}{12\pi}\left[M_{\gamma_L}^2(h,T)\right]^{3/2}
\end{aligned}
\tag{29}
$$

where the effective and thermal masses are given above. It is interesting to notice that only the cubic terms of the gauge bosons do not cancel each other in the high-temperature expansion of the effective potential and the ring corrections, in contrast with the rest of the fields which have the same degrees of freedom in the ring correction. Our main aim is to analyze in detail the phase transitions that occur at finite temperature for the effective potential (29). This is a cumbersome procedure though, since the parameter space is highly constrained from various aspects and several limitations and constraints apply that reduce significantly the range of the values of the free parameters. Thus before we actually study the phase transition, in the next few sections we shall narrow down the values of the free parameters that are phenomenologically acceptable and comply with all the experimental and theoretical constraints.

III. ELECTROWEAK BARYOGENESIS CONSTRAINTS

Initially, the vacuum of the Universe was stabilized at the origin $(h,\phi) = (0,0)$. At a high temperature T_s a phase transition starts in the ϕ direction from the origin to a non-zero vacuum expectation value (VEV). This phase transition can be either first or second-order, depending on the singlet extension model's parameters. As the Universe cools down, a second local minimum appears in the full effective potential, where a barrier is gradually formed between the two local minima. At the temperature T_c the electroweak phase transition occurs as a first-order phase transition when the two local minima $(0, v_s')$ and $(v_c, 0)$ are degenerate as it is illustrated in Figs. 4 and 5. Therefore, the transition from the false vacuum to the true vacuum proceeds via thermal tunneling, when the bubbles of the broken phase nucleate within the surrounding plasma of the phase with the false vacuum.

The observed baryon asymmetry of the Universe can be attributed to the electroweak baryogenesis (EWBG). This is a physical mechanism in the early Universe that generates an asymmetry between baryons and anti-baryons in the electroweak phase transition. This asymmetry is established if the so-called Sakharov criteria [20] are fulfilled, which namely are: (i) baryon number violation, (ii) C-CP violation, and (iii) departure from thermal equilibrium. The EWBG satisfies these criteria requiring a first-order phase transition [81, 88–90]. Baryon creation in EWBG occurs at the vicinity of the expanding bubble walls. More specifically, ahead of the bubble wall, CP and C asymmetries in particle number densities can be produced by CP-violating interactions of the plasma with the expanding bubble wall of the true vacuum [91]. Then, these asymmetries diffuse into the symmetric phase in front of the bubble wall, where they are converted to baryons by electroweak sphalerons [92]. In the broken phase, the rate of sphaleron transitions can be strongly suppressed to avoid washing out the generated baryons. Hence, it is necessary for a successful EWBG scenario, that the baryon asymmetry generated at the expanding bubble wall is not washed out by sphalerons within the broken phase. In particular, EWBG requires a strong first-order electroweak phase transition, characterized by the following condition which provides an approximation to a factor in the rate of sphaleron transitions in the broken phase,

$$
\frac{v_c}{T_c} > 0.6 - 1.4 \tag{30}
$$

The lower bound of this so-called sphaleron rate criterion[4] varies between 0.6 and 1.4 due to some uncertainties in the precise calculation, but it is conventionally taken to be one [93, 94]. In our study, we also consider the lower limit in this range for the sake of completeness. The critical temperature T_c and the Higgs VEV υ_c are computed numerically in this paper.

In fact, the three Sakharov criteria are satisfied in the SM [81]: the baryon number is violated by non-perturbative effects, CP violation can be induced by the CKM phase in the fermion mass matrix, and the Universe is out of equilibrium due to the electroweak phase transition. However, CP violation is insufficient to produce large enough chiral asymmetries to explain the observed baryon-to-entropy ratio. Furthermore, the electroweak phase transition should be strong enough first-order, a requirement that is not met by the SM as demonstrated in Appendix A, thus the motivation for singlet extensions is clear. These drawbacks in SM motivate beyond the Standard Model physics which should suggest an extension of the Higgs potential to realize a strong first-order electroweak phase transition and an extra source of CP violation. The former is successfully achieved by singlet scalar extensions of SM, two-Higgs doublet models and higher order operators to SM [44], and this occurs in our case too.

CP-violating sources in EWBG can be obtained by a plethora of models such as real and complex singlet scalar extensions [49, 95–98], two-Higgs doublet models [99] and composite Higgs models [100]. A promising CP-violating source could be a dimension-six operator which is introduced as an effective field theory [49]. This extension couples the singlet with the top quark's mass considering the singlet particle as a dark matter candidate and modifying top quark's mass for non-zero values of the singlet scalar field so that it is given by

$$y_t \bar{Q}_L H \left(1 + \frac{\eta}{\Lambda^2}\phi^2\right) t_R + h.c. \tag{31}$$

where η is the complex phase and Λ, is the energy scale of the effective field theory. Consequently, the top quark acquires a complex phase that varies in space along the profile of the bubble wall, offering the crucial source of CP violation needed to explain baryon asymmetry.

This higher dimension operator does not affect the critical temperature and VEVs as it vanishes at both the Higgs and singlet direction. This operator mostly rises the height of the barrier at temperature T_c which leads to a thinner bubble wall. It is additionally assumed that ϕ/Λ remains very small resulting in a negligible contribution to the effective potential which does not significantly affect the dynamics of the phase transition and it is not included in our analysis of the effective potential, see also [49, 95]. However, a reasonable claim could be that, the fact that this higher dimension operator does not affect the critical temperature and VEVs since it vanishes at both the Higgs and singlet direction could lead to a deeper minimum away from these points? This could be possible, but it is clearly mentioned that the ratio ϕ/Λ remains very small resulting in a negligible contribution to the effective potential and a deeper minimum cannot be generated in the context of a valid perturbative analysis. The aforementioned higher dimension operator is further discussed in [49]. The choice of the phase η is not restricted by any additional experimental constraints due to the forbidden Higgs-singlet mixing and the behavior of the singlet at high temperatures. As a result, this phase can be chosen to be maximally CP-violating and it easily generates sufficient baryon asymmetry for EWBG [49]. However, potential loop contributions to electric dipole moments might exist. This is an important aspect that we did not take into account in this work to simplify the discussion. The two-loop Barr-Zee contributions to the electric dipole moments are discussed in [100]. However, this would require Higgs-singlet mixing, which is not the case in our model. In that scenario CP would be broken spontaneously at high temperature.

IV. ALLOWED PARAMETER SPACE FOR THE MODEL AND PHENOMENOLOGICAL CONSTRAINTS

The parameter space of the model is generally described by the three parameters of the real singlet extension μ_S, λ_S, and λ_{HS} and the Wilson coefficient[5] λ. In addition, a strong EWPT can be realized by a parameter space that is mainly restricted by the sphaleron rate criterion, the vacuum structure, the values of couplings and on the fact whether these values break perturbation theory, and the invisible Higgs decay width.

Firstly, we proceed on the premise that the current vacuum state of the Universe is described by the Higgs VEV at zero temperature with broken $SU(2) \times U(1)_Y$ symmetry, while the singlet has a zero VEV with unbroken $\mathbb{Z}_2$ symmetry. As a result, we demand that the Higgs minimum is the global minimum of the effective potential at zero temperature by requiring,

$$V_0(\upsilon, 0) < V_0(0, \upsilon_s) \Rightarrow \lambda_S > \lambda_H \frac{\mu_S^4}{\mu_H^4} \tag{32}$$

[4] This ratio does not respect gauge invariance and special treatment is required to derive the gauge invariant form. However, this approach affects the numerical results much less than the two-loop approximation [93].

[5] The effective field theory is considered active at the scale $M = 15$ TeV in the following sections.

where v_s is the zero-temperature VEV of the singlet scalar field in the ϕ direction and it is assumed that $\mu_S^2 > 0$ for a two-step electroweak phase transition[6]. We additionally demand that the tree-level potential is bounded from below, which leads to $\lambda_S > 0$ and $\lambda > 0$. Also the condition (32) is essential to ensure that the electroweak vacuum is the global minimum of the effective potential at zero temperature and the parameter values we shall use ensure that the electroweak vacuum is always the lowest minimum of the effective potential. This vacuum structure further implies that the singlet scalar field acquires the following mass squared at zero temperature, which is required to be positive,

$$m_S^2 = -\mu_S^2 + \lambda_{HS} v^2 > 0 \tag{33}$$

An expression for the minimum value of λ_S can then be derived by equations (32) and (33), and it is,

$$\lambda_S^{min} = \frac{\lambda_H}{\mu_H^4} \left(m_S^2 - \lambda_{HS} v^2\right)^2 \tag{34}$$

Accordingly, the quartic coupling can be cast into the form,

$$\lambda_S = \lambda_S^{min} + a \tag{35}$$

where a is a positive parameter[7]. Consequently, it is more convenient and suitable to study the parameter space in terms of m_S, λ_{HS}, and λ for a given parameter a. However, it is worth mentioning that Eq. (32) could be rewritten in the equivalent form in terms of the other parameters,

$$\mu_S^2 < v^2 \sqrt{\lambda_H \lambda_S} \Rightarrow \lambda_{HS} < \frac{m_S^2}{v^2} + \sqrt{\lambda_H \lambda_S} \tag{36}$$

Secondly, the validity of the one-loop approximation can be violated by large coupling constants. Thus, we impose the perturbativity of the couplings in high energy scales, while the Renormalization Group Equations (RGE) can be solved at one loop for the couplings λ_{HS}, λ_H and λ_S (see Appendix B). The RGE evolution of the gauge couplings and the top quark Yukawa coupling remains the same as in the case of the SM. Moreover, the contribution of the higher order operator to the RGEs is omitted since the effective field theory is a weakly coupled theory with $\lambda/M^2 < 10^{-4}$ GeV^{-2}. Following Ref. [65], the RGE evolution shows that the model remains perturbative up to scales $10 - 100$ TeV depending on the Higgs-singlet coupling, as well as $\lambda_S^{min} < 8$ is imposed for a reliable perturbative analysis which is equivalent to $\lambda_{HS} < 5$ with the singlet mass ranging from $0 - 550$ GeV.

Moving on, the tree-level potential has an inherent $\mathbb{Z}_2$ symmetry, allowing the singlet scalar to act as a dark matter candidate [49, 55, 59, 64, 68, 69, 102–104]. Hence, the unbroken $\mathbb{Z}_2$ symmetry at zero temperature ensures the stability of the dark matter particle and prohibits the mixing between the Higgs field and the singlet scalar field. As a consequence, the couplings of the Higgs boson to fermions and gauge bosons remain identical to the SM, as well as further constraints from electroweak precision tests and Higgs coupling modifications are ruled out due to the forbidden Higgs-singlet mixing.

Lastly, the EWBG can be probed by gravitational wave and collider measurements [25, 43, 44, 64, 65]. In the case of $m_S > m_H/2$, the number of potential collider signatures are highly restricted due to the forbidden Higgs-singlet mixing. However, future colliders, such as a 100 TeV hadron collider, could test this scenario, probing the direct production of the singlet states, and the modification in the triple Higgs couplings and Zh cross section [65]. Regarding the gravitational waves, their signals from an electroweak first-order phase transition can be detected in the future observatories [43, 44, 64], such as LISA.

V. CONSTRAINTS FROM INVISIBLE HIGGS DECAY

If $m_S < m_H/2$, the decay $h \to \phi\phi$ is kinematically allowed and it contributes to the invisible decay width of the Higgs boson. The ATLAS and CMS experiments have reported numerous results of searches for invisible Higgs decays over the last years. The branching ratio of the Higgs to the invisible singlet sector is set to $BR_{inv} < 0.11 - 0.19$ at 95% CL, where the upper and lower limits of the range correspond to the results from the ATLAS [105] and the CMS collaboration [106], respectively. In Ref. [107], those results are analyzed further. However, the most recent result from combination of searches, which is not included in Ref. [107], indicates $BR_{inv} < 0.107$ at 95% CL [108].

---

[6] See section VI.

[7] It is commonly set to $a = 0.1$ for $m_S \geq m_H/2$ [40, 65, 66, 70, 101].

The Higgs decay width to visible channels is $\Gamma_{vis} = 4.07$ MeV for $m_H = 125$ GeV. Thus, if the branching ratio of the Higgs to the invisible particles is fixed at $BR_{inv} < 0.19$, the upper bound on the invisible decay width of the Higgs boson is,

$$\Gamma(h \to \phi\phi) < 0.955 \text{ MeV} \tag{37}$$

where this decay width is given by,

$$\Gamma(h \to \phi\phi) = \frac{\lambda_{HS}^2 v^2}{32\pi m_H}\sqrt{1 - \frac{4m_S^2}{m_H^2}} \tag{38}$$

Hence, this bound leads to an additional constraint for the interaction coupling,

$$\lambda_{HS} < \sqrt{\frac{32\pi m_H}{v^2}\left(1 - \frac{4m_S^2}{m_H^2}\right)^{-1/2}\Gamma_m(h \to \phi\phi)} \tag{39}$$

where $\Gamma_m(h \to \phi\phi)$ is the upper bound on the invisible decay width of the Higgs boson. Requiring $\mu_S^2 \geq 0$ and imposing the previous constraint result in,

$$\frac{m_S^2}{v^2} < \lambda_{HS} < \sqrt{\frac{32\pi m_H}{v^2}\left(1 - \frac{4m_S^2}{m_H^2}\right)^{-1/2}\Gamma_m(h \to \phi\phi)} \tag{40}$$

which is completely independent of the coupling λ_S (or equivalently a) and the Wilson coefficient λ.

Consequently, the allowed singlet masses are divided into two regions,

$$m_S \leq 30.19 \text{ GeV} \quad \text{and} \quad m_S \geq 62.43 \text{ GeV}$$

where the two separated regions are merged only for $BR_{inv} > 0.59$ which does not agree with the experimental data [107]. We gathered the allowed singlet masses in Fig. 2. The lower mass region generally has the upper bound of $\lambda_{HS} = 0.014$, whereas the higher mass region is characterized by a wider range of values for λ_{HS}, but it has almost a fixed singlet mass.

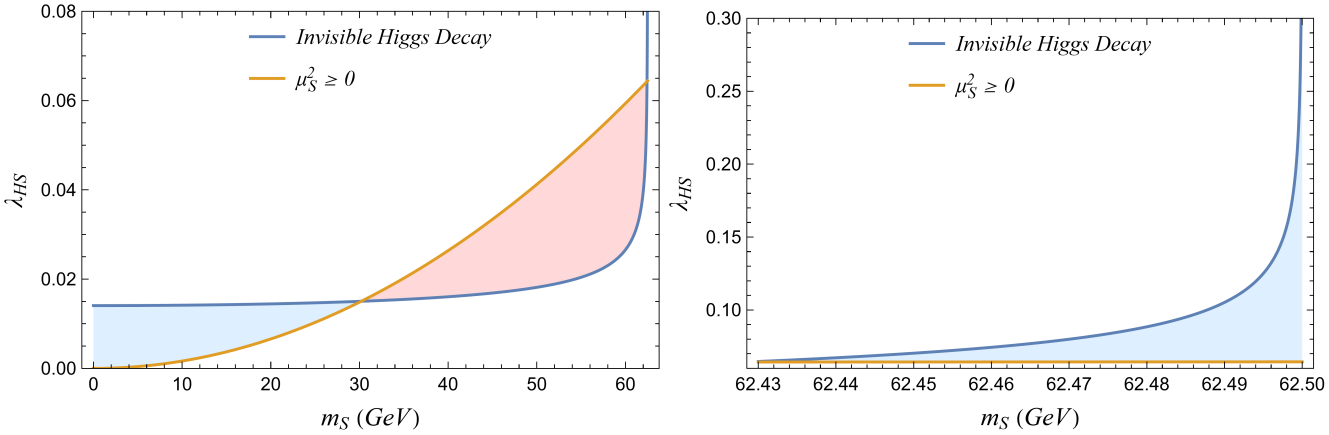


FIG. 2. **Left**: The parameter space (blue shaded) for $m_S < 62.5$ GeV which satisfies the invisible Higgs Decay constraint (42) and $\mu_S^2 \geq 0$ setting $BR_{inv} < 0.19$. **Right**: The higher mass region, $m_S \geq 62.43$ GeV, which satisfies the invisible Higgs Decay constraint (42) and $\mu_S^2 \geq 0$ setting $BR_{inv} < 0.19$.

The invisible Higgs boson branching ratio can be also set to the $BR_{inv} < 0.11$ at 95% CL in order to yield the invisible decay width,

$$\Gamma(h \to \phi\phi) < 0.503 \text{ MeV} \tag{41}$$

with the following allowed mass regions which satisfy the condition (40),

$$m_S \leq 25.45 \text{ GeV} \quad \text{and} \quad m_S \geq 62.48 \text{ GeV}$$

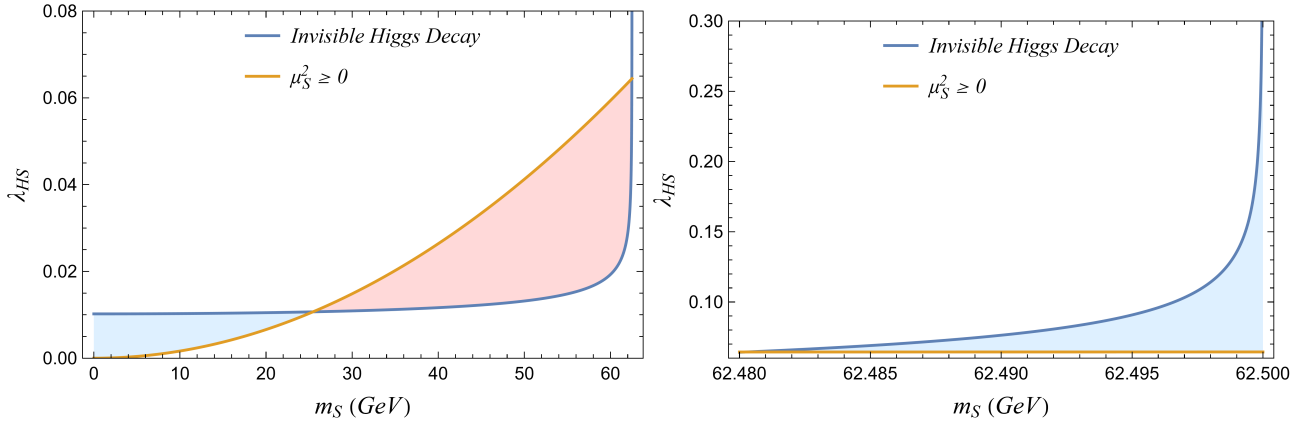


FIG. 3. **Left**: The parameter space (blue shaded) for $m_S < 62.5$ GeV which satisfies the invisible Higgs Decay constraint (42) and $\mu_S^2 \geq 0$ setting $BR_{inv} < 0.11$. **Right**: The higher mass region, $m_S \geq 62.48$ GeV, which satisfies the invisible Higgs Decay constraint (42) and $\mu_S^2 \geq 0$ setting $BR_{inv} < 0.11$.

It is immediately obvious that this upper bound further excludes the low-mass region, because of the lower maximum value of the coupling λ_{HS} compared to the previous case.

Taking into account the sphaleron rate criterion (30), the parameter space is eliminated further, since very low values of the coupling λ_{HS} generate a one-step or a two-step EWPT with a weak first-order phase transition. Therefore, the lower bound on the Higgs-singlet coupling can be higher than the one depicted in Fig. 2 or 3.

VI. THE ELECTROWEAK PHASE TRANSITION IN THE SINGLET-EXTENDED SM: A DETAILED DESCRIPTION

Now we shall analyze the behavior of the singlet-extended effective potential (29) in the full two dimensional configuration space spanned by the Higgs and the singlet scalar fields (h, ϕ). We shall describe in some detail the behavior of the full effective potential at finite temperature and in the next subsections we shall constrain the free parameters of the model, in order to be phenomenologically viable from various aspects. The results presented in Figs. 4 and 5 summarize some of the main results of this paper.

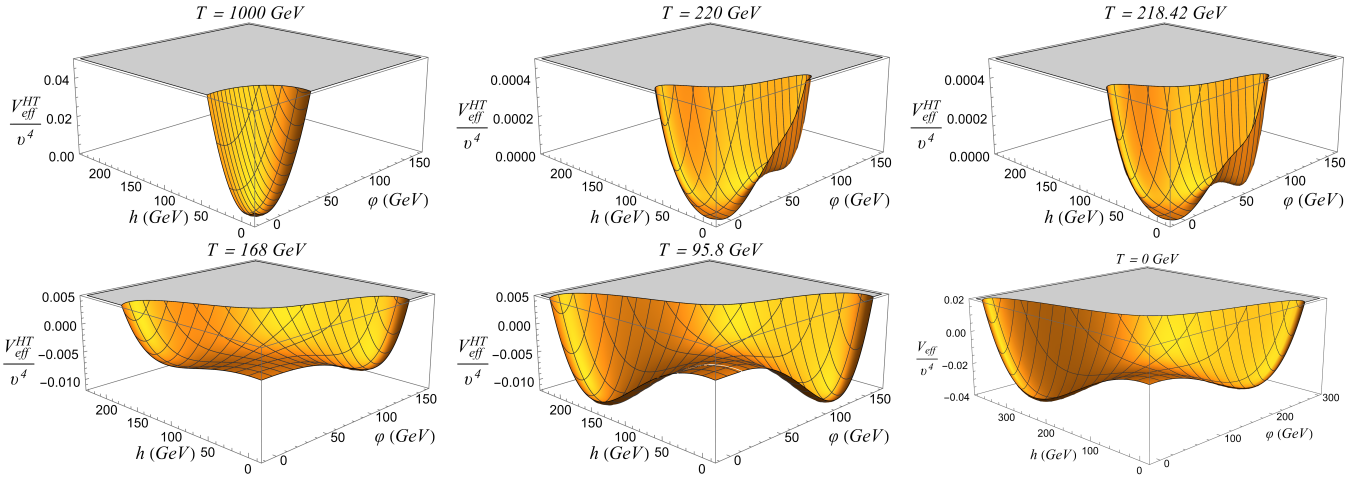


FIG. 4. The full effective potential during the strong electroweak phase transition as the temperature decreases. In this example, the singlet's phase transition is first-order using a point of the parameter space with $m_S = 500$ GeV, $\lambda_{HS} = 4.3$, $\lambda/M^2 \simeq 2 \times 10^{-5}$ GeV^{-2} and $a = 0.1$.

In this work, the electroweak phase transition is considered as a two-step phase transition which consists of a primary first-order or second-order phase transition in the ϕ direction from the origin $(h, \phi) = (0, 0)$ to a non-zero

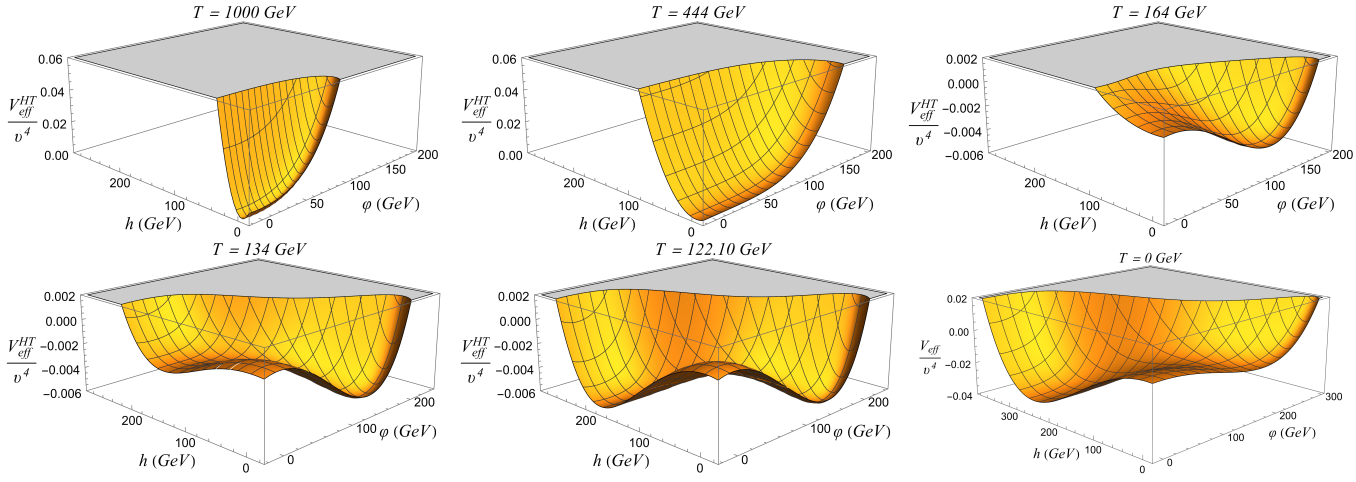


FIG. 5. The full effective potential during the strong electroweak phase transition as the temperature decreases. In this example, the singlet's phase transition is second-order using a point of the parameter space with $m_S = 62.5$ GeV, $\lambda_{HS} = 0.15$, $\lambda/M^2 \simeq 2 \times 10^{-5}$ GeV^{-2} and $a = 0.1$.

VEV and a subsequent first-order phase transition from $(0, \upsilon_s')$ to the Higgs vacuum $(\upsilon_c, 0)$. In the early Universe, as the temperature decreases, the Higgs minimum is formed, and a barrier is created between the Higgs and the singlet minimum as shown in Figs. 4 and 5. Then a strong first-order phase transition signals the tunneling to the electroweak vacuum which remains the global minimum at zero temperature. Instead of this cosmological history, the electroweak symmetry breaking could alternatively take place as a one-step strong first-order phase transition from the origin $(0, 0)$ to the electroweak vacuum $(\upsilon_c, 0)$ along the Higgs direction, while the singlet vacuum is stabilized on the origin with zero VEV. This one-step scenario can be driven by loop effects requiring $\mu_S^2 < 0$ [38, 65, 69]. A comprehensive study on the classification of numerous models to thermally driven, tree-level driven, and loop-driven is given in Ref. [38]. Nevertheless, a one-step phase transition can be also realized in the case of $\mu_S^2 > 0$ due to thermal effects.

Furthermore, an instability can be developed along the singlet's direction at extremely high temperatures for large values of the couplings λ, λ_{HS} and λ_S. This instability can also take place without the dimension six operator which was included in the Lagrangian, but in our model, it is mainly caused by large λ and a which importantly affect the singlet's effective potential. However, this behavior does not alter the above discussion because the strong first-order transition is still realized with $T_c \ll T_s$. Nevertheless we need to stress that such an instability could lead to a high temperature transition from the $\langle s \rangle = 0$ minimum to the $\langle s \rangle = \upsilon_{s'}$ vacuum, at a temperature way higher than the one corresponding to the electroweak phase transition. This phase transition would generate defects such as domain walls and these could remain as remnants of this phase transition. This feature is mentionable and it occurs for the aforementioned values of the parameter, however it relies on the fact that the temperature indeed reached such high values, which is not certain. We needed to report this, because it is a peculiar and mentionable feature, which must be treated with reluctancy however, due to the extreme temperature that it occurs and also because, if it occurs, it occurs well before the electroweak phase transition. More importantly, from a physical point of view, perturbation theory does not break in the singlet sector, however the ratio of the singlet mass over the corresponding temperature would be nearly zero, $m_S/T_s \sim 0$, thus normally, such high temperatures and the corresponding transitions in the effective potential should be disregarded, in the same way we disregarded the effect of small mass quarks in the Higgs effective potential. We mention this issue for completeness though and in order to provide a spherical and complete view of the parameter space, for the convenience of the reader.

In addition, a common issue in this thermal history arises when topological defects emerge due to the spontaneous symmetry breaking of the discrete $\mathbb{Z}_2$ symmetry, while the singlet scalar field acquires a non-zero VEV at a high temperature [109]. Namely, electroweak baryogenesis can be highly affected by the generated domain walls after the high-temperature spontaneous symmetry breaking. Recently, it was shown in Ref. [110] that the inclusion of a dimension-six singlet scalar field operator in the usual real singlet extensions can resolve this problem considering a scenario in which the vacuum state never respected the $\mathbb{Z}_2$ symmetry. This scenario is not examined further in this study, leaving such an analysis and the implications of topological defects for future work. However, it is noticeable that in our model, the $\mathbb{Z}_2$ symmetry could be never restored at high temperature due to the aforementioned instability developed along the ϕ direction in the effective potential. Namely, one observes that the singlet VEV remains non-zero for large λ and a at extremely high temperatures $T > T_c$. As a result, in these cases, the $\mathbb{Z}_2$ symmetry is not restored

and the issue of the domain walls is avoided.

Finally, the influence of the higher order operator on the electroweak phase transition can be comprehensively understood by its mathematical aspects. In the Higgs direction, the non-zero Wilson coefficient changes the thermal mass of the singlet (21), where λ is compared with the couplings λ_{HS} and λ_S, which is then expressed in terms of singlet mass m_S, the coupling λ_{HS}, and the parameter a. This means that the relation between their values can play an essential role for the effect of the higher-order operator. In the ϕ direction, on the other hand, the Wilson coefficient mainly contributes to the effective mass of the Higgs boson (4), the Goldstone bosons (5), and the singlet (6), while the thermal mass of the singlet remains the same as in the Higgs direction. As a consequence, in the direction with $h = 0$, the effect of the Wilson coefficient is primarily determined by the coupling λ_{HS} and the singlet mass m_S. Therefore, it is worth mentioning that the non-zero Wilson coefficient significantly changes the effective potential (29) in the ϕ direction.

In the next subsections, the two-step strong electroweak phase transition is studied by dividing the parameter space of the singlet extension with the dimension six operator in three regions: the low-mass region ($m_S < m_H/2$), the Higgs resonance region ($m_S = m_H/2$), and the high-mass region ($m_S > m_H/2$).

A. High-mass Region: $m_S > m_H/2$

First of all, the two-step electroweak phase transition can be described by the parameter space for $m_H < 2m_S$ in Fig. 6 with zero Wilson coefficient, adopting the common parametrization with $a = 0.1$. As it has been previously mentioned, this parameter space includes a number of small regions with a one-step electroweak phase transition. In particular, it was found that the parameter space with $\mu_S \lesssim 90$ GeV corresponds to a region with both scenarios. On the other hand, imposing the criterion for a strong first-order phase transition, reduces the generic two-step parameter space with $\lambda = 0$. More specifically, it is mainly changed for $m_S > 200$ GeV and low Higgs-singlet couplings as the criterion (30) requires higher coupling[8] λ_{HS} in comparison with the generic case of a two-step phase transition. Those higher values of the lower bound for the coupling λ_{HS} vary with different singlet masses and they are depicted by the blue dotted line in Fig. 6. These results generally agree with similar studies [64, 65, 70, 101]. The renormalization scale

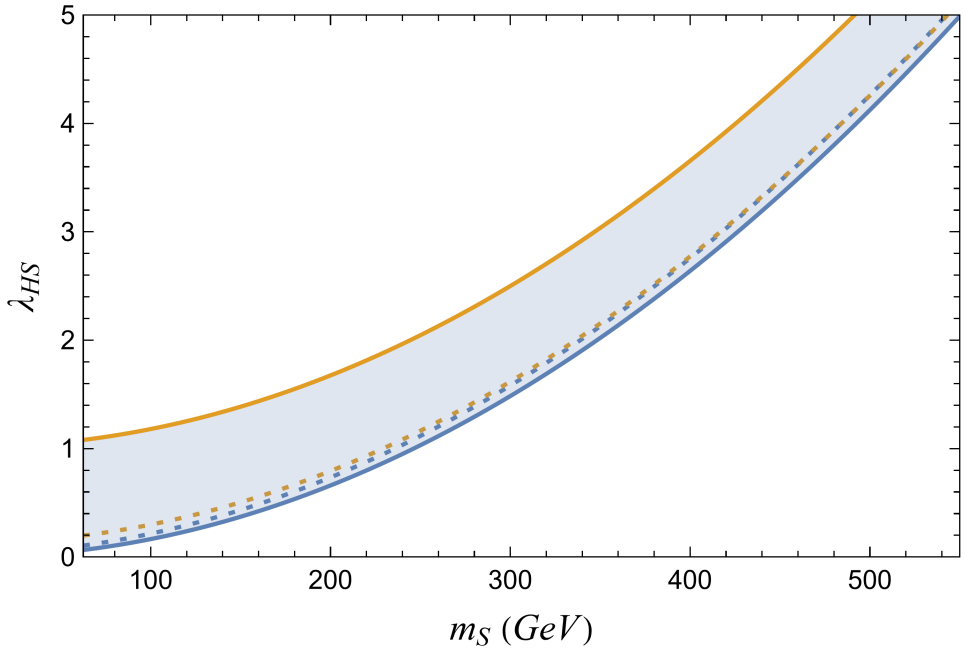


FIG. 6. The blue region is the parameter space of the singlet model for a two-step electroweak phase transition with $m_S > m_H/2$, $a = 0.1$ and $\lambda = 0$. The orange dotted line describes the constant $\mu_S = 90$ GeV. The parameter space of the strong two-step electroweak phase transition with $v_c/T_c > 1$ corresponds to the blue region which is bounded from below by the blue dotted line. The critical temperature in this parameter space varies from $T_c \simeq 30 - 140$ GeV.

[8] The difference between the lowest λ_{HS} for satisfying the sphaleron rate criterion with lower bound 0.6 and 1 is negligible.

dependence and other theoretical uncertainties in the perturbative analysis (which are discussed in Refs. [66, 72–74]) may introduce small differences between our results and the literature. Moreover, the authors of Ref. [70] claimed that larger parts of the two-step electroweak phase transition parameter space are ruled out since relativistic speeds during the phase transition must not be reached by the expanding bubble walls and it is required that bubbles do nucleate at finite temperature. Nonetheless, we do not consider these effects in the present study, leaving such improvement for future work.

The presence of the higher order operator with $\lambda < 10^2$ does not noticeably affect the aforementioned parameter space of the strong electroweak phase transition. The lowest coupling λ_{HS} for a strong phase transition, slightly increases with $\lambda \lesssim 10^3$ and the critical temperature is significantly increased, whereas the ratio υ_c/T_c shows a downward trend. This effect is strengthened by very high and very low singlet masses. On the other hand, the maximum value of the coupling λ_{HS}, which is shown in Fig. 6, is not changed by the higher order operator. In this case, the effect of the non-zero Wilson coefficient is highly weakened by lower singlet masses. The comparison between the parameter space of the singlet extension with $\lambda = 0$ and $\lambda = 10^4$ is presented in Fig. 7. It is clearly seen that the influence of the dimension-6 operator is not crucial for $a = 0.1$. On the contrary, this behavior dramatically alters for higher values of the parameter a and it is extensively studied for lower singlet masses.

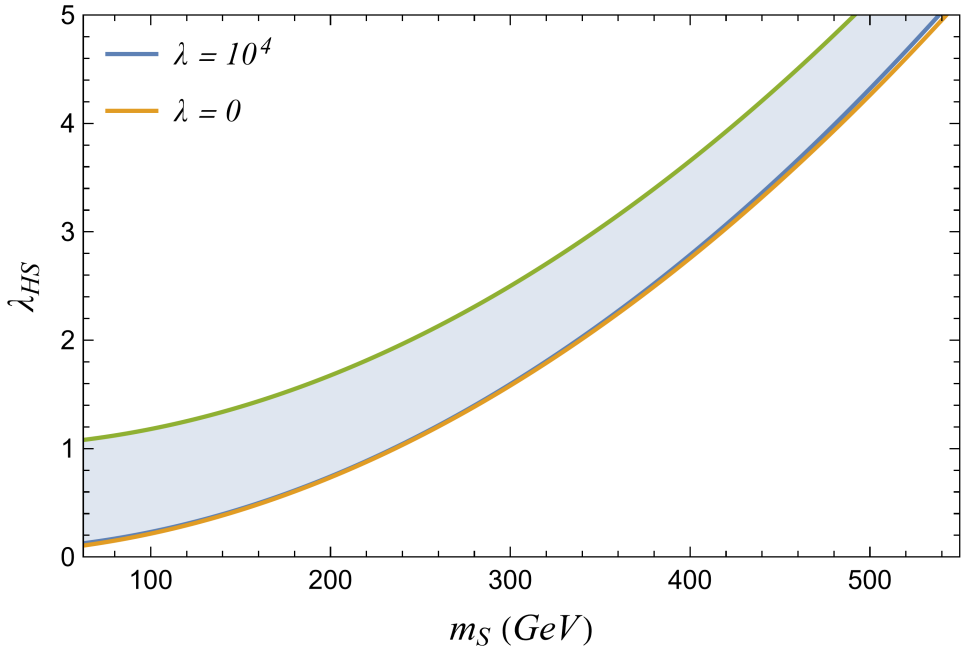


FIG. 7. The parameter space (blue region) of the singlet extension with a dimension-six operator ($\lambda = 10^4$) in order to realize a strong two-step electroweak phase transition ($\upsilon_c/T_c > 1$). The orange line shows the lower bound of the parameter space of the singlet extension which is represented by a blue dotted line in Fig. 6. In both cases, we take $a = 0.1$.

The behavior of the full effective potential at finite temperature is depicted in Fig. 4 for the case at hand, by taking into account all the phenomenological constraints obtained in this subsection, and for various temperatures. In this example, the singlet's phase transition is first-order using a point of the parameter space with $m_S = 500$ GeV, $\lambda_{HS} = 4.3$, $\lambda/M^2 \simeq 2 \times 10^{-5}$ GeV^{-2} and $a = 0.1$, but it can also be a second-order phase transition for lower singlet masses and Higgs-singlet couplings. As it can be seen, at high temperatures, the effective potential is symmetric, and as the temperature decreases, the singlet vacuum with a barrier between the second vacuum and the origin is developed. As the temperature decreases, the Higgs vacuum is generated with another barrier between the latter and the origin also existing. Accordingly, the Higgs vacuum is deeper than the singlet vacuum and the electroweak symmetry breaking phase transition occurs.

B. Higgs Resonance Region: $m_S = m_H/2$

The Higgs resonance region is highly motivated by dark matter physics since the singlet should have a mass near $m_S = 62.5$ GeV as a viable dark matter candidate [102–104] and it is strongly restricted by the direct dark matter searches by LUX, XENON1T and XENONnT experiments [49, 59, 64, 102–104]. While in the majority of the

parameter space without the dimension-six operator, the assumption of the real singlet as a dark matter candidate is inconsistent with a strong first-order electroweak phase transition, it has been shown that in the Higgs resonance region, the dark matter relic density can constitute an important proportion of the total observed dark matter density [102–104].

Additionally, the full effective potential at finite temperature is illustrated in Fig. 5 using a point of the parameter space with $\lambda_{HS} = 0.15$, $\lambda/M^2 \simeq 2\times10^{-5}$ GeV^{-2}, and $a = 0.1$ and a similar behavior is followed in the total parameter space with $m_S = m_H/2$. The singlet's phase transition is generally second-order in this region in which the two-step electroweak phase transition satisfies the sphaleron rate criterion.

Firstly, the common value of $a = 0.1$ is selected in order to compute the critical temperature T_c, the critical temperature T_s of the singlet's phase transition and the sphaleron rate criterion for various values of the Wilson coefficient and the Higgs-singlet coupling which are presented in Fig. 8, 9, 10 and 11. The electroweak phase transition is considerably affected by the higher order operator for $\lambda > 10^3$ and the parameter space is restricted, as the ratio v_c/T_c drops for a non-zero Wilson coefficient, and a higher λ_{HS} is necessary to generate the strong electroweak phase transition. Moreover, the phase transition in the ϕ direction occurs at a very high temperature for $\lambda/M^2 \simeq 4\times10^{-5}$ GeV^{-2}, which shows the strong influence of the higher order operator on the singlet's dynamics.

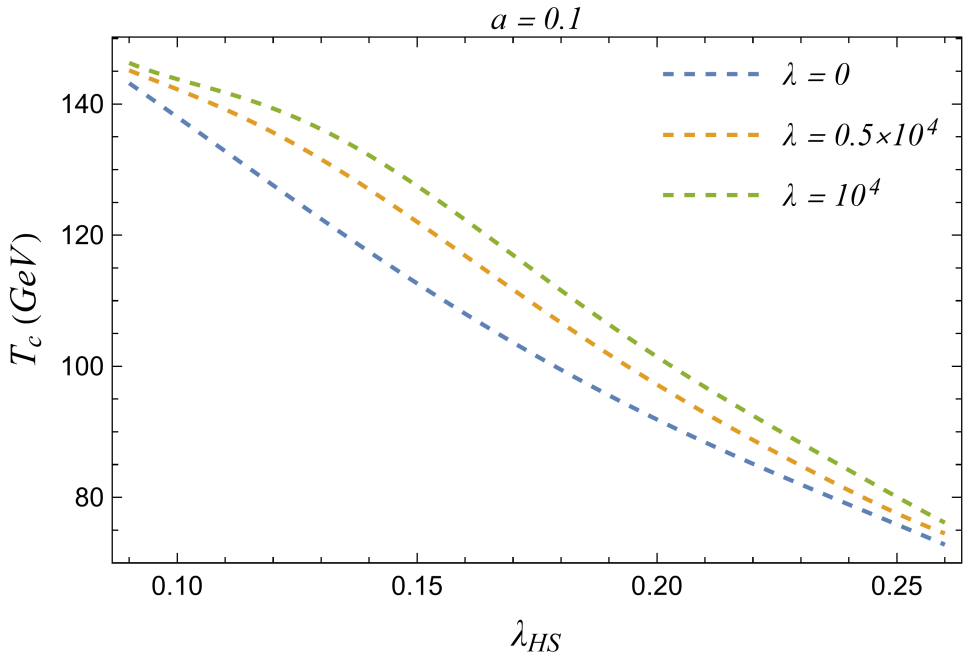


FIG. 8. The critical temperature (T_c) as a function of the Higgs-singlet coupling (λ_{HS}) for $m_S = 62.5$ GeV and $a = 0.1$.

On the other hand, the parameter space of the strong electroweak phase transition is expanded by taking $a > 0.4$. This is an important feature of our model which we need to further highlight at this point. This is clearly illustrated in the case of $a = 1$ in Fig. 10, where the criterion $v_c/T_c > 1$ can be satisfied by much lower λ_{HS} compared to the model with zero Wilson coefficient. As a result, if $\lambda/M^2 \gtrsim 10^{-4}$ GeV^{-2}, a strong electroweak phase transition can be realized for every low Higgs-singlet coupling with $\mu_S^2 > 0$. Hence, the higher order operator could assist a strong electroweak phase transition in regions of the parameter space which were excluded in the previous singlet extensions of the SM.

Regarding the singlet's phase transition, the trend of the critical temperature T_s is similar to the case of $a < 0.4$ and T_s sharply increases as the Wilson coefficient increases. In addition, the previously mentioned instability in the ϕ direction takes place for $\lambda > 3650$ and $\lambda_{HS} > 0.52$ with $a = 1$ which is also showcased by the fact that $\lambda_{HS} = 0.52$ leads to $T_s \simeq 8768$ GeV. It is noticeable that lower values of a lead to this instability for higher values of the Wilson coefficient and the Higgs-singlet coupling.

In Figs. 10 and 8, the critical temperature with $\lambda \neq 0$ approaches the critical temperature with $\lambda = 0$ for larger values of the Higgs-singlet coupling. This behavior shows that the effect of the higher order operator is stronger for low values of λ_{HS}. This is also apparent for the ratio for the sphaleron rate criterion in Fig. 9 and 10. However, in the case of $a = 0.1$, the influence of the higher order operator is maximized for an intermediate low value of the Higgs-singlet coupling and not for the lower values.

Finally, it is essential to mention that if the value of $a < 0.1$, at zero temperature the Higgs VEV v is not much

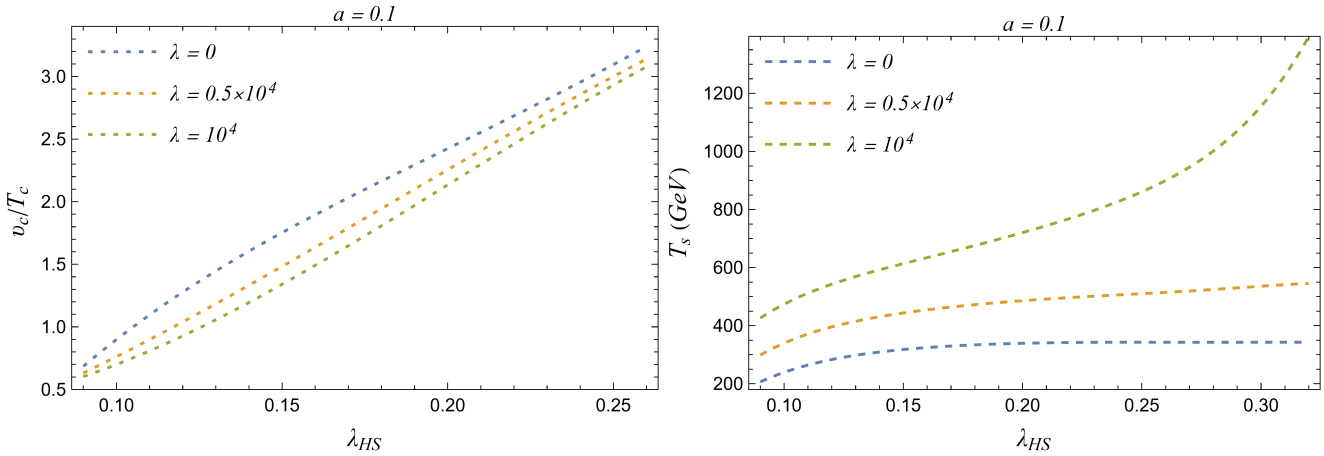


FIG. 9. **Left**: The sphaleron rate criterion as a function of the coupling λ_{HS} for $m_S = 62.5$ GeV and $a = 0.1$ in the case of zero and non-zero Wilson coefficient. **Right**: The critical temperature of the singlet's second-order phase transition T_s as a function of the coupling λ_{HS} for $m_S = 62.5$ GeV and $a = 0.1$.

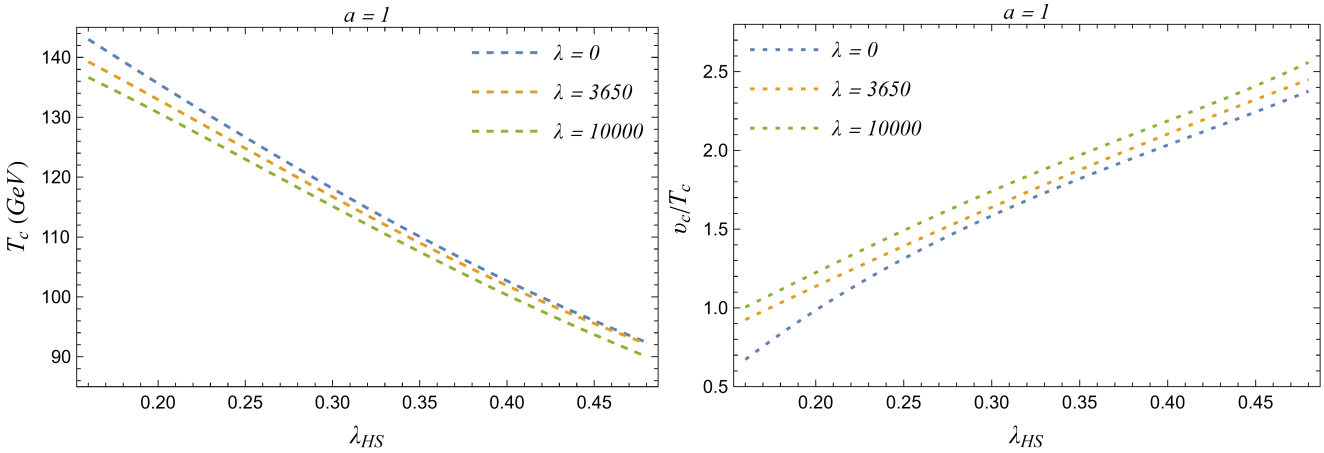


FIG. 10. The electroweak phase transition with $m_S = 62.5$ GeV and $a = 1$: **Left**: The critical temperature (T_c) as a function of the Higgs-singlet coupling (λ_{HS}). **Right**: The sphaleron rate criterion as a function of the coupling λ_{HS} in the case of zero and non-zero Wilson coefficient.

deeper than the singlet VEV υ_s. Namely, the condition (32) is satisfied for every a in Eq. (35), but low values of a do not satisfy the condition,

$$V_0(0, \upsilon_s) \gg V_0(\upsilon, 0) \tag{42}$$

which ensures that the Higgs vacuum is by far more energetically favorable compared to the singlet vacuum.

C. Low-mass Region: $m_S < m_H/2$

In the low mass region, the behavior of the full effective potential remains the same as in the Higgs resonance region, while the high-temperature phase transition in the ϕ direction is primarily second-order in the allowed parameter space.

It has been already shown that the parameter space for $m_S < m_H/2$ is severely eliminated by the invisible decay width of the Higgs boson and the condition $\mu_S^2 > 0$. It is initially assumed that the branching ratio of the Higgs boson to invisible particles is set to $BR_{inv} < 0.19$. However, the lower mass region $m_S \lesssim 30$ GeV, which is allowed by these constraints, is completely excluded by the sphaleron rate criterion considering $\lambda = 0$ and $a \gtrsim 0.05$ due to the low values of the Higgs-singlet coupling. On the contrary, large values of $\lambda > 10^3$ and $a \gtrsim 0.05$ can easily assist the strong phase transition for $m_S \leq 1$ GeV. This is clearly illustrated if we take $\lambda = 2 \times 10^4$ and $a = 0.75$ as a strong electroweak phase transition ($\upsilon_c/T_c > 0.6$) occurs for every coupling λ_{HS} and $m_S < 10$ GeV. This is shown in Fig.

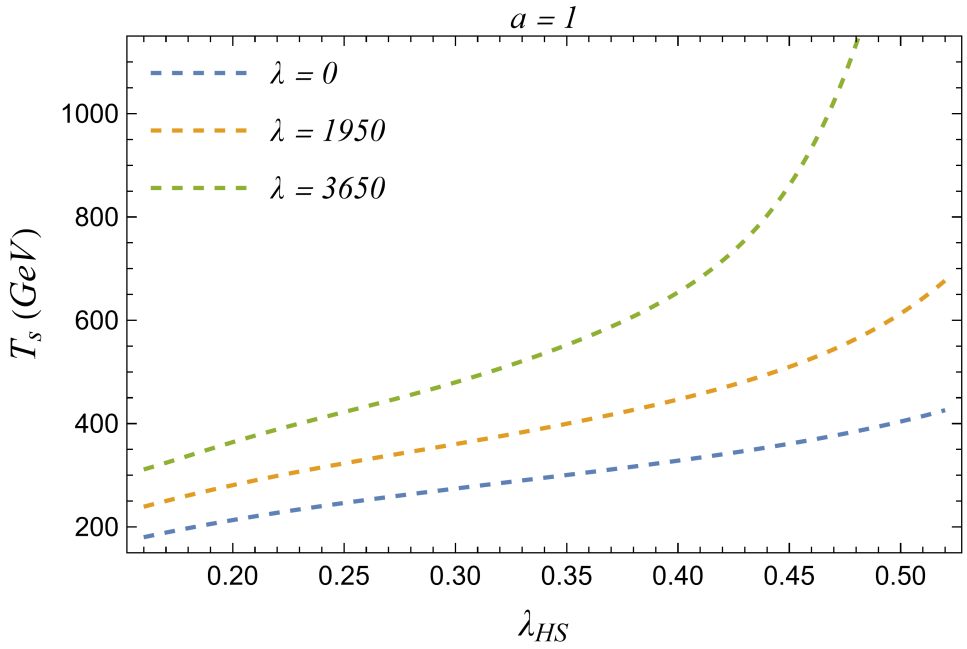


FIG. 11. The critical temperature of the singlet's second-order phase transition T_s as a function of the coupling λ_{HS} for $m_S = 62.5$ GeV and $a = 1$ in the case of zero and non-zero Wilson coefficient.

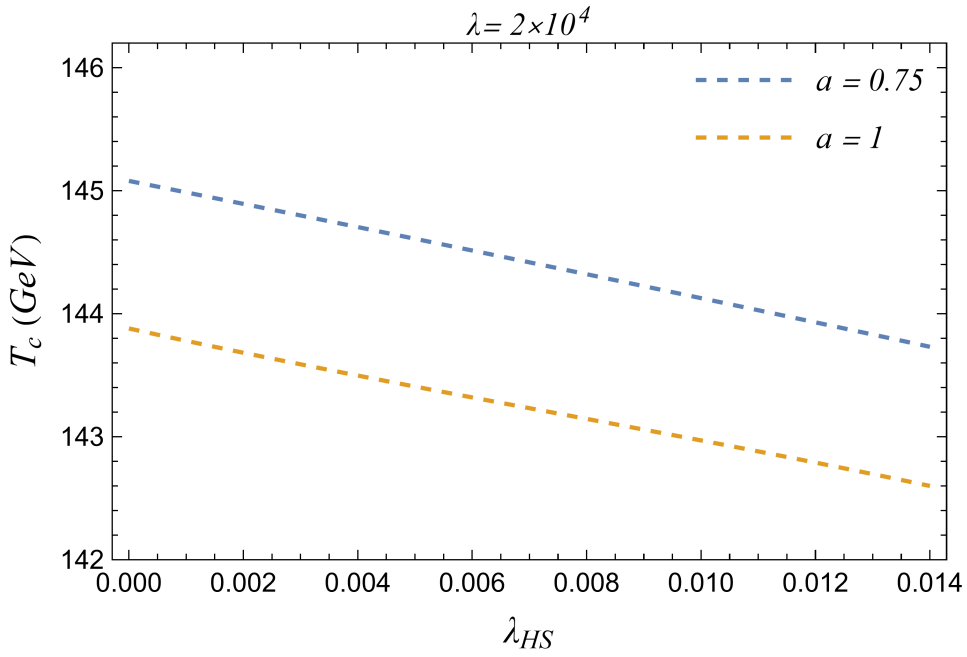


FIG. 12. The critical temperature as a function of the Higgs-singlet coupling for $a = 0.75$ and 1 and $m_S \leq 0.1$ GeV.

12. More specifically, the lower bound of λ_{HS} with a singlet mass $m_S = 0.1$ GeV and $\lambda = 2 \times 10^4$ can generate a strong electroweak phase transition and it is at least 10^6 times smaller than the corresponding value in the singlet extension without the higher order operator. Therefore, the presence of the higher order operator allows very low interaction couplings which easily avoid the constraint imposed by the branching ratio $BR_{inv} < 0.10$.

In contrast, the previous behavior is dramatically changed for values $a \lesssim 0.05$, as the zero Wilson coefficient can lead to a strong first-order phase transition, whereas the presence of the higher order operator weakens the electroweak phase transition. The electroweak phase transition with zero Wilson coefficient mainly occurs for $m_S < 10$ GeV in order to have a two-step phase transition, but this upper limit on the mass increases for very low values of the

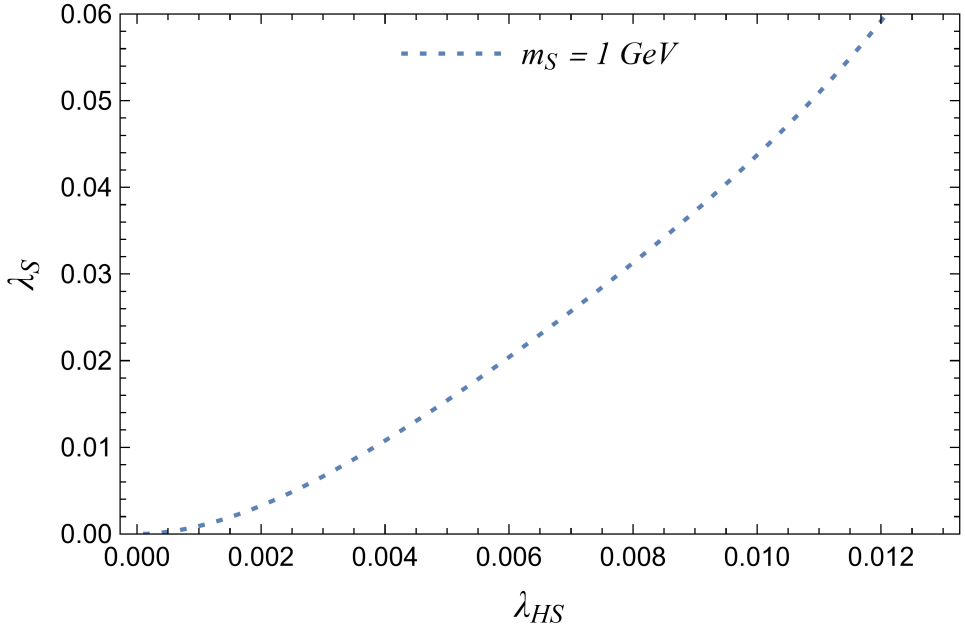


FIG. 13. The dependence of the lowest λ_{HS}, which is allowed by all the previous constraints imposing the criterion $v_c/T_c > 0.6$, on the parameter$a$ for $m_S = 1$GeV and $\lambda = 0$. This dependence was numerically computed to explain qualitatively that the parameter space expands its lower bound for low values of the quartic coupling λ_S.

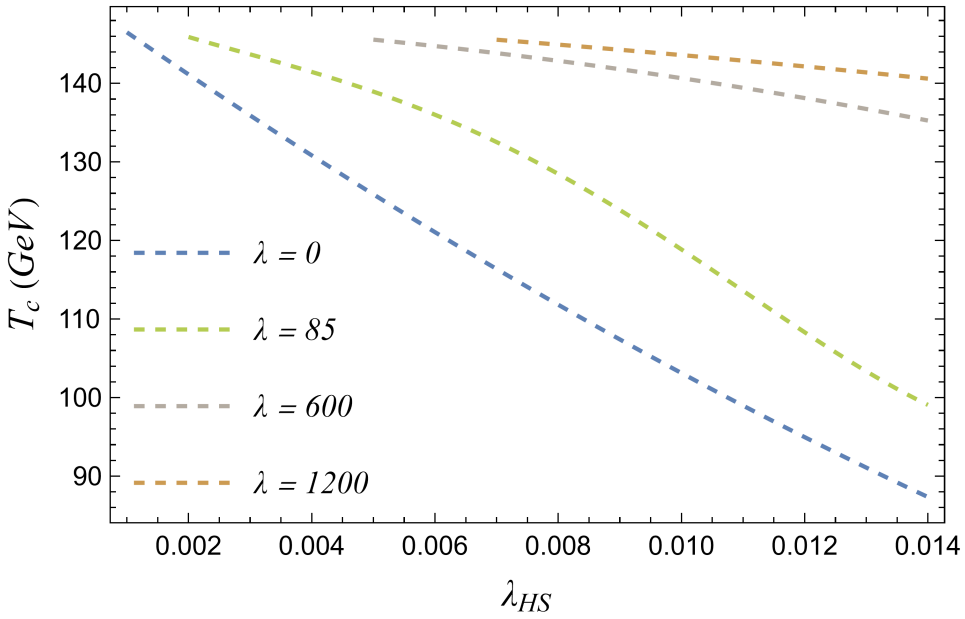


FIG. 14. The critical temperature as a function of the coupling λ_{HS} for $m_S \leq 1$GeV and $a = 0.001$ in the case of zero and non-zero Wilson coefficient $\lambda = 0$ (blue), 85 (green), 600 (gray), 1200 (brown).

parameter a and λ_{HS}. This behavior is also illustrated in Fig. 13 since it shows that the lowest allowed value of λ_{HS} decreases as the parameter a drops for a given singlet mass.

Some typical results are presented for $m_S \leq 1$ GeV and $a = 0.001$ in Fig. 14 and 15. Firstly, it can be clearly seen that the lower bound of λ_{HS} is strongly constrained by a non-zero total Wilson coefficient $\lambda/M^2 \simeq 10^{-7} - 10^{-5}$ GeV^{-2}. In addition, the full effective potential around the critical temperature T_c remains the same for masses between $0-1$ GeV with maximum deviations in the critical temperature of the order of 10^{-2}. In Fig. 14, the trend of critical temperature T_c is similar to one in the previous mass region, whereas the temperature T_s in Fig. 15 stabilizes for large values of λ_{HS}, in contrast to the behavior described in the previous sections.

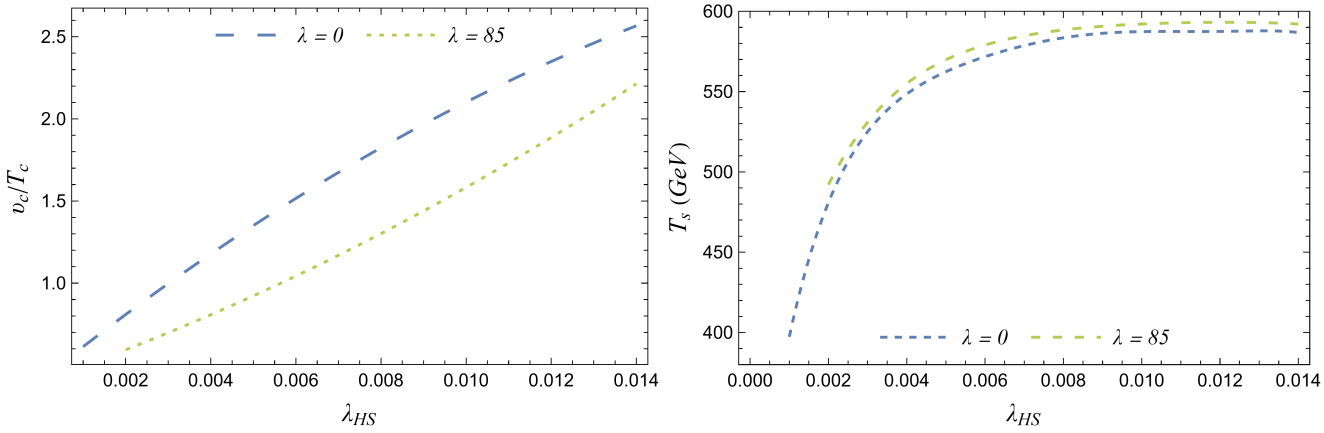


FIG. 15. **Left**: The sphaleron rate criterion as a function of the coupling λ_{HS} for $m_S \leq 1$GeV and $a = 0.001$ in the case of zero and non-zero Wilson coefficient. **Right**: The critical temperature of the singlet's second-order phase transition T_s as a function of the coupling λ_{HS} for $m_S \leq 0.1$GeV and $a = 0.001$ in the case of zero and non-zero Wilson coefficient.

In the case of $BR_{inv} < 0.11$ in the lower mass region the previous conclusions do not change as this constraint only decreases the maximum allowed value of the Higgs-singlet coupling to $\lambda_{HS}^{max} = 0.10$, which subsequently excludes some scenarios, especially for $\lambda = 0$.

In the second region of the parameter space for $m_S < m_H/2$, the singlet mass is strictly restricted to almost the half mass of the Higgs boson due to constraints in Eq. (40) (Fig. 2 and 3). As a result, there is no noticeable difference between this case and the Higgs resonance regarding the second step in the two-step electroweak phase transition, while the phase transition in the ϕ direction is slightly affected. Namely, the results on the critical temperature T_c and the Higgs VEV v_c remain the same as the results in the previous section.

Assuming $BR_{inv} < 0.19$, the constraint (40) leads to singlet mass $m_S = 62.48$ GeV and coupling $\lambda_{HS} = 0.88$ as the only strong electroweak phase transition scenario[9] for $a = 0.1$ in the standard singlet extension. Thus, the constraint (40) imposes that $m_S \geq 62.48$ GeV and leads to a narrow parameter space[10]. On the other hand, the higher order operator can decrease the lower bound on the Higgs-singlet coupling due to the sphaleron rate criterion as it was discussed in the previous section. It is remarkable that in the higher mass region the coupling $\lambda_{HS} > 0.065$ (allowed by Eq. (40)) generates a viable electroweak phase transition for $\lambda/M^2 \simeq 9 \times 10^{-5}$ GeV^{-2} and $a = 0.75$. Consequently, the inclusion of the dimension-6 operator can contribute to the occurrence of a strong electroweak first-order phase transition in the higher mass region of the parameter space for $m_S < m_H/2$. This discussion is analogous in the case of $BR_{inv} < 0.11$ because the lower bound of the Higgs-singlet coupling remains unaffected, whereas the size of the parameter space differs, with the parameter space being highly restricted for $m_S < 62.499$ GeV.

VII. CONCLUSIONS AND DISCUSSION

Our current perception for the Universe during its primordial classical and quantum evolution stages is rather vague. Too many questions have to be answered, and the current and future CMB and gravitational wave experiments are expected to shed light on these stages of our Universe's evolution. The mechanism that gave masses to the matter particles is the electroweak symmetry breaking and this is believed to have occurred via a first-order phase transition. However, the electroweak phase transition in the SM is not a sufficiently strong first-order phase transition. To this end, in this article we studied the electroweak phase transition in the context of the real singlet extensions of the SM, including dimension-six non-renormalizable operators which couple the singlet scalar field with the Higgs doublet. We showed that the electroweak phase transition occurs as a two-step phase transition, which consists of the singlet's phase transition at high temperature and a subsequent strong first-order phase transition in the Higgs and SM sector. In addition, considering a viable CP-violation source, such as a dimension-6 operator which couples the singlet to the top quark mass, the electroweak baryogenesis can be realized to describe the baryon asymmetry in the current Universe. This scenario is also compatible with assuming the singlet particle as dark matter candidate with mass

[9] The sphaleron rate criterion is satisfied: $\frac{v_c}{T_c} \simeq 0.61$.

[10] The maximum value of λ_{HS} in Eq. (39) sharply increases as it approaches half of the Higgs mass. This maximum value for $m_S = 62.49998$ GeV reaches nearly half of the upper bound of λ_{HS} in the parameter space shown in Fig. 7.

nearly half of the Higgs mass, which is highly restricted by the direct dark matter searches. As we previously discussed, the parameter space is majorly reduced by the sphaleron rate criterion, the vacuum structure, the perturbativity of couplings, and the invisible Higgs decay width. Consequently, we studied the two-step strong electroweak phase transition which respects those constraints for $m_S = 0 - 550$ GeV. In this model, the critical temperature of the first-order phase transition varies from $T_c \simeq 30 - 200$ GeV depending on the parameters of the model.

The impact of the dimension-six operator is summarized as follows:

- The presence of the dimension-six operator mainly modifies the thermal mass of the singlet in the Higgs direction which alters the first-order phase transition, while the singlet's phase transition is strongly affected by the non-zero Wilson coefficient, and its critical temperature increases for higher Wilson coefficients. In addition, the effect of the higher order operator is washed out by large values of the Higgs-singlet coupling and is weakened for coefficients $\lambda < 10^3$ in most cases.

- The dimension-six operator can generally assist in generating the strong electroweak phase transition for low Higgs-singlet couplings which were not allowed in the standard singlet extensions of the SM.

- The parameter space of our model for $m_S > m_H/2$ and $a = 0.1$ remains approximately the same as in the standard singlet extension. In the Higgs resonance region, the singlet particle could be a dark matter candidate, and the parameter space for $a > 0.4$ is expanded for large values of the Wilson coefficient, whereas this trend is reversed for $a < 0.4$.

- The low-mass region is significantly excluded by the invisible Higgs decay width and the sphaleron rate criterion in the context of the standard singlet extensions. Nevertheless, a strong electroweak phase transition can be generated for low singlet masses $m_S < m_H/2$ by including the dimension-six operator with $\lambda/M^2 \gtrsim 5 \times 10^{-6}$ GeV^{-2} and $a \gtrsim 0.05$.

A study we did not include in our analysis, is the study of the stochastic gravitational wave background that corresponds to the first-order phase transitions occurring in our model. It is expected that the collisions between the bubbles of vacua during the first-order phase transition can generate a stochastic gravitational wave background in the frequencies probed by current and future gravitational wave experiments. We aim to analyze the stochastic gravitational wave background generated by our model in a near future work.

ACKNOWLEDGMENTS

This research has been is funded by the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan (V.K.O) (Grant No. AP19674478).

APPENDIX A: ELECTROWEAK PHASE TRANSITION IN THE SM

The electroweak phase transition can be realized in the context of the SM as a first-order phase transition [16, 40, 81, 83, 86, 91, 111, 112], but it is insufficiently strong to generate the baryon asymmetry. The dynamics of the phase transition are described by the temperature-dependent effective potential of the Higgs field including the dominant contributions of the gauge bosons, the top quark and the Goldstone bosons. This effective potential is explicitly presented in section II and in the Arnold-Espinosa scheme it is written as

$$\begin{aligned} V_{eff}^{SM}(h,T) = &-\frac{\mu_H^2}{2}h^2 + \frac{\lambda_H}{4}h^4 + \sum_i (-1)^{F_i} n_i \frac{m_i^4(h)}{64\pi^2}\left[\ln\left(\frac{m_i^2(h)}{\mu_R^2}\right) - C_i\right] \\ &+ \sum_i \frac{n_i T^4}{2\pi^2} J_B\left(\frac{m_i^2(h)}{T^2}\right) - 12\frac{T^4}{2\pi^2} J_F\left(\frac{m_t^2(h)}{T^2}\right) \\ &+ \sum_i \frac{\overline{n}_i T}{12\pi}\left[m_i^3(h) - \left(M_i^2(h,T)\right)^{3/2}\right] \end{aligned} \tag{43}$$

where $i = \{h, \chi, W, Z, \gamma\}$ corresponds to the bosons in the SM.

The high-temperature expansion of the thermal functions (15) and (16) for the temperature-dependent one-loop effective potential is valid with very high accuracy in the case of the phase transition. Hence, the high-temperature

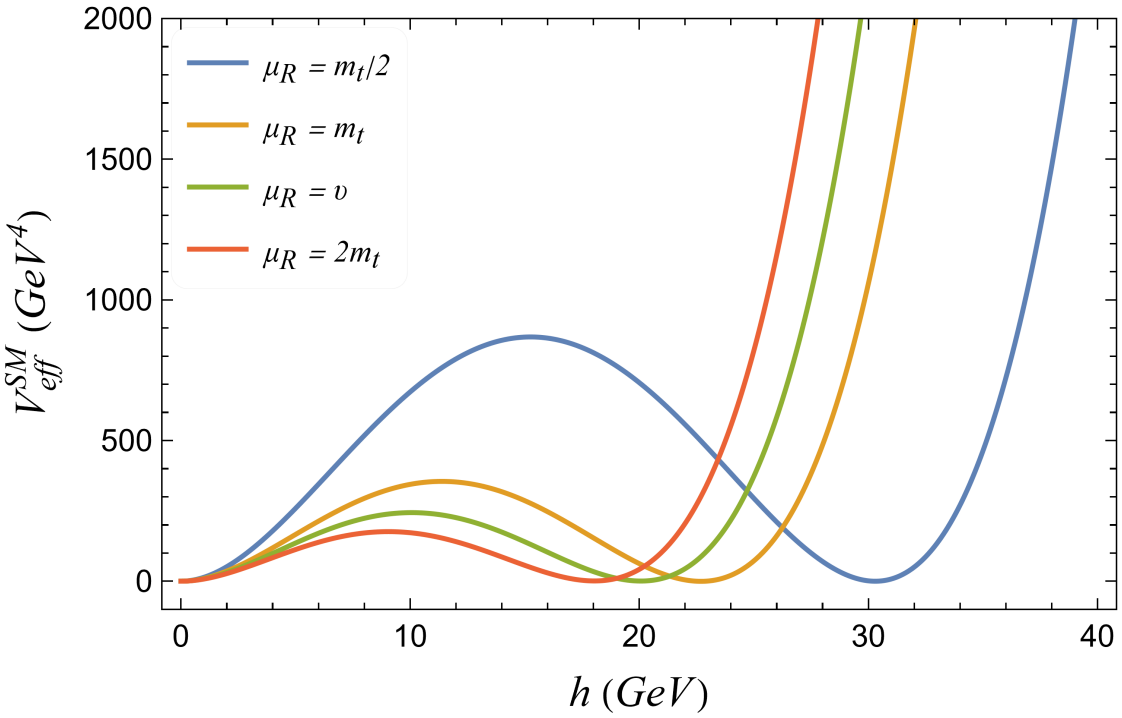


FIG. 16. The full effective potential of the SM at the critical temperature for numerous renormalization scales $\mu_R = m_t/2$ (blue), m_t (orange), v (green) and $2m_t$ (red).

expansion is implemented in the full effective potential which reads

$$\begin{aligned} V_{eff}^{SM}(h,T) = &-\frac{\mu_H^2}{2}h^2 + \frac{\lambda_H}{4}h^4 + \frac{m_h^2(h)}{24}T^2 - \frac{T}{12\pi}\left[m_h^2(h)+\Pi_h(T)\right]^{3/2} + \frac{m_h^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{3}{2}\right] + \\ &\frac{3m_\chi^2(h)}{24}T^2 - \frac{3T}{12\pi}\left[m_\chi^2(h)+\Pi_\chi(T)\right]^{3/2} + \frac{3m_\chi^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{3}{2}\right] + \\ &\frac{6m_W^2(h)}{24}T^2 - \frac{4T}{12\pi}m_W^3(h) - \frac{2T}{12\pi}\left[m_W^2(h)+\Pi_{W_L}(T)\right]^{3/2} + \frac{6m_W^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{5}{6}\right] + \\ &\frac{3m_Z^2(h)}{24}T^2 - \frac{2T}{12\pi}m_Z^3(h) - \frac{T}{12\pi}\left[M_{Z_L}^2(h,T)\right]^{3/2} + \frac{3m_Z^4(h)}{64\pi^2}\left[\ln\left(\frac{a_b T^2}{\mu_R^2}\right) - \frac{5}{6}\right] + \\ &\frac{12m_t^2(h)}{48}T^2 - \frac{12m_t^4(h)}{64\pi^2}\left[\ln\left(\frac{a_f T^2}{\mu_R^2}\right) - \frac{3}{2}\right] - \frac{T}{12\pi}\left[M_{\gamma_L}^2(h,T)\right]^{3/2} \end{aligned} \tag{44}$$

where the effective masses are given by

$$m_h^2(h) = -\mu_H^2 + 3\lambda_H h^2 \tag{45}$$

$$m_\chi^2(h,\phi) = -\mu_H^2 + \lambda_H h^2 \tag{46}$$

and Eqs. (7), (8), (9), while the temperature-dependent self-energy of the Higgs and the Goldstone bosons reads

$$\Pi_h(T) = \Pi_\chi(T) = \left(\frac{3g^2}{16} + \frac{g'^2}{16} + \frac{y_t^2}{4} + \frac{\lambda_H}{2}\right)T^2 \tag{47}$$

and the thermal masses of the gauge bosons are given by Eqs. (26), (27) and (22).

The electroweak phase transition is studied by varying the renormalization scale for different values. The full effective potential (44) is illustrated in Fig. for $\mu_R = m_t/2$ (blue), m_t (orange), v (green) and $2m_t$ (red).

It is immediately apparent that the sphaleron rate criterion is not satisfied as well as the maximum ratio is achieved for $\mu_R = m_t/2$ with $v_c/T_c \simeq 0.2 < 0.6$; namely, the electroweak phase transition is not a strong enough first-order

μ_R (GeV)	T_c (GeV)	υ_c (GeV)
86.5	148.987	30
173.0	148.326	23
246.2	148.002	20
346.0	147.693	18

TABLE I. The critical temperature of the electroweak phase transition (T_c) and the Higgs VEV (υ_c) for different values of the renormalization scale.

phase transition[11]. Similar calculations can be found in Refs. [16, 40, 81, 83, 91]. In the context of the perturbative analysis, the sphaleron rate criterion can be translated into an upper bound on the mass of the Higgs boson, such as $m_h \lesssim 42$ GeV [81] which showcases the lack of a strong phase transition in SM, whereas this bound is $m_h \lesssim 72 - 80$ GeV in lattice calculations [113–116]. While the height of the barrier in the effective potential and the Higgs VEV during the phase transition clearly change for different renormalization scales, the critical temperature is slightly dropped by less than 1% for higher values of the renormalization scale as it is presented in Table I.

APPENDIX B: ONE-LOOP BETA FUNCTIONS

The RGEs for the parameters of the real singlet extension to the SM are presented here:

$$16\pi^2\beta_{\lambda_H} = 24\lambda_H^2 - 3\left(3g^2 + g'^2 - 4y_t^2\right)\lambda_H + \frac{1}{2}\lambda_{HS}^2 + \frac{3}{8}\left(3g^4 + 2g^2g'^2 + g'^2\right) - 6y_t^4 \tag{48}$$

$$16\pi^2\beta_{\lambda_{HS}} = 4\lambda_{HS}^2 + \left(12\lambda_H + 6\lambda_S + 6y_t^2 - \frac{9}{2}g^2 - \frac{3}{2}g'^2\right)\lambda_{HS} \tag{49}$$

$$16\pi^2\beta_{\lambda_S} = 18\lambda_S^2 + 2\lambda_{HS}^2 \tag{50}$$

$$16\pi^2\beta_g = -\frac{19}{6}g^3 \tag{51}$$

$$16\pi^2\beta_{g'} = \frac{41}{6}g'^3 \tag{52}$$

$$16\pi^2\beta_{g_s} = -7g_s^3 \tag{53}$$

$$16\pi^2\beta_{y_t} = \frac{9}{2}y_t^3 - \left(\frac{9}{4}g^2 + \frac{17}{12}g'^2 + 8g_s^2\right)y_t \tag{54}$$

where the one-loop beta function is defined as

$$\beta_g = \mu\frac{dg}{d\mu} \tag{55}$$

[1] K. N. Abazajian *et al.* [CMB-S4], [arXiv:1610.02743 [astro-ph.CO]].
[2] M. H. Abitbol *et al.* [Simons Observatory], Bull. Am. Astron. Soc. **51** (2019), 147 [arXiv:1907.08284 [astro-ph.IM]].

[11] Likewise, the authors of Ref. [83] argued that the sphaleron rate criterion is not satisfied for $m_h = 125$ GeV in SM. In particular, they calculated that $\upsilon_c/T_c \simeq 0.2$ for $m_h = 120$ GeV and $m_t = 170$ GeV, considering the high-temperature expansion without the daisy resummation

[3] S. Hild, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia and M. Beker, *et al.* Class. Quant. Grav. **28** (2011), 094013 doi:10.1088/0264-9381/28/9/094013 [arXiv:1012.0908 [gr-qc]].

[4] J. Baker, J. Bellovary, P. L. Bender, E. Berti, R. Caldwell, J. Camp, J. W. Conklin, N. Cornish, C. Cutler and R. DeRosa, *et al.* [arXiv:1907.06482 [astro-ph.IM]].

[5] T. L. Smith and R. Caldwell, Phys. Rev. D **100** (2019) no.10, 104055 doi:10.1103/PhysRevD.100.104055 [arXiv:1908.00546 [astro-ph.CO]].

[6] J. Crowder and N. J. Cornish, Phys. Rev. D **72** (2005), 083005 doi:10.1103/PhysRevD.72.083005 [arXiv:gr-qc/0506015 [gr-qc]].

[7] T. L. Smith and R. Caldwell, Phys. Rev. D **95** (2017) no.4, 044036 doi:10.1103/PhysRevD.95.044036 [arXiv:1609.05901 [gr-qc]].

[8] N. Seto, S. Kawamura and T. Nakamura, Phys. Rev. Lett. **87** (2001), 221103 doi:10.1103/PhysRevLett.87.221103 [arXiv:astro-ph/0108011 [astro-ph]].

[9] S. Kawamura, M. Ando, N. Seto, S. Sato, M. Musha, I. Kawano, J. Yokoyama, T. Tanaka, K. Ioka and T. Akutsu, *et al.* [arXiv:2006.13545 [gr-qc]].

[10] A. Weltman, P. Bull, S. Camera, K. Kelley, H. Padmanabhan, J. Pritchard, A. Raccanelli, S. Riemer-Sørensen, L. Shao and S. Andrianomena, *et al.* Publ. Astron. Soc. Austral. **37** (2020), e002 doi:10.1017/pasa.2019.42 [arXiv:1810.02680 [astro-ph.CO]].

[11] P. Auclair *et al.* [LISA Cosmology Working Group], [arXiv:2204.05434 [astro-ph.CO]].

[12] G. Agazie *et al.* [NANOGrav], doi:10.3847/2041-8213/acdac6 [arXiv:2306.16213 [astro-ph.HE]].

[13] J. Antoniadis, P. Arumugam, S. Arumugam, S. Babak, M. Bagchi, A. S. B. Nielsen, C. G. Bassa, A. Bathula, A. Berthereau and M. Bonetti, *et al.* [arXiv:2306.16214 [astro-ph.HE]].

[14] D. J. Reardon, A. Zic, R. M. Shannon, G. B. Hobbs, M. Bailes, V. Di Marco, A. Kapur, A. F. Rogers, E. Thrane and J. Askew, *et al.* doi:10.3847/2041-8213/acdd02 [arXiv:2306.16215 [astro-ph.HE]].

[15] H. Xu, S. Chen, Y. Guo, J. Jiang, B. Wang, J. Xu, Z. Xue, R. N. Caballero, J. Yuan and Y. Xu, *et al.* doi:10.1088/1674-4527/acdfa5 [arXiv:2306.16216 [astro-ph.HE]].

[16] M. E. Carrington, Phys. Rev. D **45** (1992), 2933-2944 doi:10.1103/PhysRevD.45.2933

[17] K. Kajantie, M. Laine, K. Rummukainen and M. E. Shaposhnikov, Nucl. Phys. B **466** (1996), 189-258 doi:10.1016/0550-3213(96)00052-1 [arXiv:hep-lat/9510020 [hep-lat]].

[18] C. E. M. Wagner, LHEP **2023** (2023), 466 doi:10.31526/lhep.2023.466 [arXiv:2311.06949 [hep-ph]].

[19] M. Laine, G. Nardini and K. Rummukainen, JCAP **01** (2013), 011 doi:10.1088/1475-7516/2013/01/011 [arXiv:1211.7344 [hep-ph]].

[20] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5** (1967), 32-35 doi:10.1070/PU1991v034n05ABEH002497

[21] R. Apreda, M. Maggiore, A. Nicolis and A. Riotto, Nucl. Phys. B **631** (2002), 342-368 doi:10.1016/S0550-3213(02)00264-X [arXiv:gr-qc/0107033 [gr-qc]].

[22] R. M. Schabinger and J. D. Wells, Phys. Rev. D **72** (2005), 093007 doi:10.1103/PhysRevD.72.093007 [arXiv:hep-ph/0509209 [hep-ph]].

[23] A. Kusenko, Phys. Rev. Lett. **97** (2006), 241301 doi:10.1103/PhysRevLett.97.241301 [arXiv:hep-ph/0609081 [hep-ph]].

[24] J. McDonald, Phys. Rev. D **50** (1994), 3637-3649 doi:10.1103/PhysRevD.50.3637 [arXiv:hep-ph/0702143 [hep-ph]].

[25] M. Chala, C. Krause and G. Nardini, JHEP **07** (2018), 062 doi:10.1007/JHEP07(2018)062 [arXiv:1802.02168 [hep-ph]].

[26] H. Davoudiasl, R. Kitano, T. Li and H. Murayama, Phys. Lett. B **609** (2005), 117-123 doi:10.1016/j.physletb.2005.01.026 [arXiv:hep-ph/0405097 [hep-ph]].

[27] I. Baldes, T. Konstandin and G. Servant, Phys. Lett. B **786** (2018), 373-377 doi:10.1016/j.physletb.2018.10.015 [arXiv:1604.04526 [hep-ph]].

[28] A. Noble and M. Perelstein, Phys. Rev. D **78** (2008), 063518 doi:10.1103/PhysRevD.78.063518 [arXiv:0711.3018 [hep-ph]].

[29] R. Zhou, J. Yang and L. Bian, JHEP **04** (2020), 071 doi:10.1007/JHEP04(2020)071 [arXiv:2001.04741 [hep-ph]].

[30] D. J. Weir, Phil. Trans. Roy. Soc. Lond. A **376** (2018) no.2114, 20170126 doi:10.1098/rsta.2017.0126 [arXiv:1705.01783 [hep-ph]].

[31] M. B. Hindmarsh, M. Lüben, J. Lumma and M. Pauly, SciPost Phys. Lect. Notes **24** (2021), 1 doi:10.21468/SciPostPhysLectNotes.24 [arXiv:2008.09136 [astro-ph.CO]].

[32] X. F. Han, L. Wang and Y. Zhang, Phys. Rev. D **103** (2021) no.3, 035012 doi:10.1103/PhysRevD.103.035012 [arXiv:2010.03730 [hep-ph]].

[33] H. L. Child and J. T. Giblin, Jr., JCAP **10** (2012), 001 doi:10.1088/1475-7516/2012/10/001 [arXiv:1207.6408 [astro-ph.CO]].

[34] M. Fairbairn and R. Hogan, JHEP **09** (2013), 022 doi:10.1007/JHEP09(2013)022 [arXiv:1305.3452 [hep-ph]].

[35] C. Caprini, M. Hindmarsh, S. Huber, T. Konstandin, J. Kozaczuk, G. Nardini, J. M. No, A. Petiteau, P. Schwaller and G. Servant, *et al.* JCAP **04** (2016), 001 doi:10.1088/1475-7516/2016/04/001 [arXiv:1512.06239 [astro-ph.CO]].

[36] S. J. Huber, T. Konstandin, G. Nardini and I. Rues, JCAP **03** (2016), 036 doi:10.1088/1475-7516/2016/03/036 [arXiv:1512.06357 [hep-ph]].

[37] C. Delaunay, C. Grojean and J. D. Wells, JHEP **04** (2008), 029 doi:10.1088/1126-6708/2008/04/029 [arXiv:0711.2511 [hep-ph]].

[38] D. J. H. Chung, A. J. Long and L. T. Wang, Phys. Rev. D **87** (2013) no.2, 023509 doi:10.1103/PhysRevD.87.023509 [arXiv:1209.1819 [hep-ph]].

[39] G. Barenboim and J. Rasero, JHEP **07** (2012), 028 doi:10.1007/JHEP07(2012)028 [arXiv:1202.6070 [hep-ph]].

[40] E. Senaha, Symmetry **12** (2020) no.5, 733 doi:10.3390/sym12050733

[41] C. Grojean and G. Servant, Phys. Rev. D **75** (2007), 043507 doi:10.1103/PhysRevD.75.043507 [arXiv:hep-ph/0607107 [hep-ph]].
[42] A. Katz and M. Perelstein, JHEP **07** (2014), 108 doi:10.1007/JHEP07(2014)108 [arXiv:1401.1827 [hep-ph]].
[43] A. Alves, T. Ghosh, H. K. Guo, K. Sinha and D. Vagie, JHEP **04** (2019), 052 doi:10.1007/JHEP04(2019)052 [arXiv:1812.09333 [hep-ph]].
[44] P. Athron, C. Balázs, A. Fowlie, L. Morris and L. Wu, [arXiv:2305.02357 [hep-ph]].
[45] S. Profumo, M. J. Ramsey-Musolf and G. Shaughnessy, JHEP **08** (2007), 010 doi:10.1088/1126-6708/2007/08/010 [arXiv:0705.2425 [hep-ph]].
[46] P. H. Damgaard, D. O'Connell, T. C. Petersen and A. Tranberg, Phys. Rev. Lett. **111** (2013) no.22, 221804 doi:10.1103/PhysRevLett.111.221804 [arXiv:1305.4362 [hep-ph]].
[47] A. Ashoorioon and T. Konstandin, JHEP **07** (2009), 086 doi:10.1088/1126-6708/2009/07/086 [arXiv:0904.0353 [hep-ph]].
[48] D. O'Connell, M. J. Ramsey-Musolf and M. B. Wise, Phys. Rev. D **75** (2007), 037701 doi:10.1103/PhysRevD.75.037701 [arXiv:hep-ph/0611014 [hep-ph]].
[49] J. M. Cline and K. Kainulainen, JCAP **01** (2013), 012 doi:10.1088/1475-7516/2013/01/012 [arXiv:1210.4196 [hep-ph]].
[50] M. Gonderinger, H. Lim and M. J. Ramsey-Musolf, Phys. Rev. D **86** (2012), 043511 doi:10.1103/PhysRevD.86.043511 [arXiv:1202.1316 [hep-ph]].
[51] S. Profumo, L. Ubaldi and C. Wainwright, Phys. Rev. D **82** (2010), 123514 doi:10.1103/PhysRevD.82.123514 [arXiv:1009.5377 [hep-ph]].
[52] M. Gonderinger, Y. Li, H. Patel and M. J. Ramsey-Musolf, JHEP **01** (2010), 053 doi:10.1007/JHEP01(2010)053 [arXiv:0910.3167 [hep-ph]].
[53] V. Barger, P. Langacker, M. McCaskey, M. Ramsey-Musolf and G. Shaughnessy, Phys. Rev. D **79** (2009), 015018 doi:10.1103/PhysRevD.79.015018 [arXiv:0811.0393 [hep-ph]].
[54] C. Cheung, M. Papucci and K. M. Zurek, JHEP **07** (2012), 105 doi:10.1007/JHEP07(2012)105 [arXiv:1203.5106 [hep-ph]].
[55] T. Alanne, K. Tuominen and V. Vaskonen, Nucl. Phys. B **889** (2014), 692-711 doi:10.1016/j.nuclphysb.2014.11.001 [arXiv:1407.0688 [hep-ph]].
[56] J. R. Espinosa, T. Konstandin and F. Riva, Nucl. Phys. B **854** (2012), 592-630 doi:10.1016/j.nuclphysb.2011.09.010 [arXiv:1107.5441 [hep-ph]].
[57] J. R. Espinosa and M. Quiros, Phys. Rev. D **76** (2007), 076004 doi:10.1103/PhysRevD.76.076004 [arXiv:hep-ph/0701145 [hep-ph]].
[58] V. Barger, P. Langacker, M. McCaskey, M. J. Ramsey-Musolf and G. Shaughnessy, Phys. Rev. D **77** (2008), 035005 doi:10.1103/PhysRevD.77.035005 [arXiv:0706.4311 [hep-ph]].
[59] J. M. Cline, K. Kainulainen, P. Scott and C. Weniger, Phys. Rev. D **88** (2013), 055025 [erratum: Phys. Rev. D **92** (2015) no.3, 039906] doi:10.1103/PhysRevD.88.055025 [arXiv:1306.4710 [hep-ph]].
[60] C. P. Burgess, M. Pospelov and T. ter Veldhuis, Nucl. Phys. B **619** (2001), 709-728 doi:10.1016/S0550-3213(01)00513-2 [arXiv:hep-ph/0011335 [hep-ph]].
[61] M. Kakizaki, S. Kanemura and T. Matsui, Phys. Rev. D **92** (2015) no.11, 115007 doi:10.1103/PhysRevD.92.115007 [arXiv:1509.08394 [hep-ph]].
[62] K. Enqvist, S. Nurmi, T. Tenkanen and K. Tuominen, JCAP **08** (2014), 035 doi:10.1088/1475-7516/2014/08/035 [arXiv:1407.0659 [astro-ph.CO]].
[63] J. R. Espinosa and M. Quiros, Phys. Lett. B **305** (1993), 98-105 doi:10.1016/0370-2693(93)91111-Y [arXiv:hep-ph/9301285 [hep-ph]].
[64] A. Beniwal, M. Lewicki, J. D. Wells, M. White and A. G. Williams, JHEP **08** (2017), 108 doi:10.1007/JHEP08(2017)108 [arXiv:1702.06124 [hep-ph]].
[65] D. Curtin, P. Meade and C. T. Yu, JHEP **11** (2014), 127 doi:10.1007/JHEP11(2014)127 [arXiv:1409.0005 [hep-ph]].
[66] C. W. Chiang, Y. T. Li and E. Senaha, Phys. Lett. B **789** (2019), 154-159 doi:10.1016/j.physletb.2018.12.017 [arXiv:1808.01098 [hep-ph]].
[67] P. S. B. Dev, F. Ferrer, Y. Zhang and Y. Zhang, JCAP **11** (2019), 006 doi:10.1088/1475-7516/2019/11/006 [arXiv:1905.00891 [hep-ph]].
[68] K. Ghorbani and P. H. Ghorbani, J. Phys. G **47** (2020) no.1, 015201 doi:10.1088/1361-6471/ab4823 [arXiv:1804.05798 [hep-ph]].
[69] P. Ghorbani, Phys. Dark Univ. **33** (2021), 100861 doi:10.1016/j.dark.2021.100861 [arXiv:2010.15708 [hep-ph]].
[70] G. Kurup and M. Perelstein, Phys. Rev. D **96** (2017) no.1, 015036 doi:10.1103/PhysRevD.96.015036 [arXiv:1704.03381 [hep-ph]].
[71] S. R. Coleman and E. J. Weinberg, Phys. Rev. D **7** (1973), 1888-1910 doi:10.1103/PhysRevD.7.1888
[72] P. Athron, C. Balazs, A. Fowlie, L. Morris, G. White and Y. Zhang, JHEP **01** (2023), 050 doi:10.1007/JHEP01(2023)050 [arXiv:2208.01319 [hep-ph]].
[73] D. Croon, O. Gould, P. Schicho, T. V. I. Tenkanen and G. White, JHEP **04** (2021), 055 doi:10.1007/JHEP04(2021)055 [arXiv:2009.10080 [hep-ph]].
[74] O. Gould and T. V. I. Tenkanen, JHEP **06** (2021), 069 doi:10.1007/JHEP06(2021)069 [arXiv:2104.04399 [hep-ph]].
[75] S. P. Martin, Phys. Rev. D **90** (2014) no.1, 016013 doi:10.1103/PhysRevD.90.016013 [arXiv:1406.2355 [hep-ph]].
[76] J. Elias-Miro, J. R. Espinosa and T. Konstandin, JHEP **08** (2014), 034 doi:10.1007/JHEP08(2014)034 [arXiv:1406.2652 [hep-ph]].
[77] A. Andreassen, W. Frost and M. D. Schwartz, Phys. Rev. D **91** (2015) no.1, 016009 doi:10.1103/PhysRevD.91.016009 [arXiv:1408.0287 [hep-ph]].

[78] A. Andreassen, W. Frost and M. D. Schwartz, Phys. Rev. Lett. **113** (2014) no.24, 241801 doi:10.1103/PhysRevLett.113.241801 [arXiv:1408.0292 [hep-ph]].
[79] M. Bando, T. Kugo, N. Maekawa and H. Nakano, Phys. Lett. B **301** (1993), 83-89 doi:10.1016/0370-2693(93)90725-W [arXiv:hep-ph/9210228 [hep-ph]].
[80] C. Tamarit, Phys. Rev. D **90** (2014) no.5, 055024 doi:10.1103/PhysRevD.90.055024 [arXiv:1404.7673 [hep-ph]].
[81] M. Quiros, [arXiv:hep-ph/9901312 [hep-ph]].
[82] M. Laine and A. Vuorinen, Lect. Notes Phys. **925** (2016), pp.1-281 Springer, 2016, doi:10.1007/978-3-319-31933-9 [arXiv:1701.01554 [hep-ph]].
[83] G. W. Anderson and L. J. Hall, Phys. Rev. D **45** (1992), 2685-2698 doi:10.1103/PhysRevD.45.2685
[84] D. Curtin, P. Meade and H. Ramani, Eur. Phys. J. C **78** (2018) no.9, 787 doi:10.1140/epjc/s10052-018-6268-0 [arXiv:1612.00466 [hep-ph]].
[85] R. R. Parwani, Phys. Rev. D **45** (1992), 4695 [erratum: Phys. Rev. D **48** (1993), 5965] doi:10.1103/PhysRevD.45.4695 [arXiv:hep-ph/9204216 [hep-ph]].
[86] P. B. Arnold and O. Espinosa, Phys. Rev. D **47** (1993), 3546 [erratum: Phys. Rev. D **50** (1994), 6662] doi:10.1103/PhysRevD.47.3546 [arXiv:hep-ph/9212235 [hep-ph]].
[87] E. J. Weinberg and A. q. Wu, Phys. Rev. D **36** (1987), 2474 doi:10.1103/PhysRevD.36.2474
[88] M. Trodden, Rev. Mod. Phys. **71** (1999), 1463-1500 doi:10.1103/RevModPhys.71.1463 [arXiv:hep-ph/9803479 [hep-ph]].
[89] A. Riotto, [arXiv:hep-ph/9807454 [hep-ph]].
[90] A. Riotto and M. Trodden, Ann. Rev. Nucl. Part. Sci. **49** (1999), 35-75 doi:10.1146/annurev.nucl.49.1.35 [arXiv:hep-ph/9901362 [hep-ph]].
[91] D. E. Morrissey and M. J. Ramsey-Musolf, New J. Phys. **14** (2012), 125003 doi:10.1088/1367-2630/14/12/125003 [arXiv:1206.2942 [hep-ph]].
[92] A. G. Cohen, D. B. Kaplan and A. E. Nelson, Ann. Rev. Nucl. Part. Sci. **43** (1993), 27-70 doi:10.1146/annurev.ns.43.120193.000331 [arXiv:hep-ph/9302210 [hep-ph]].
[93] H. H. Patel and M. J. Ramsey-Musolf, JHEP **07** (2011), 029 doi:10.1007/JHEP07(2011)029 [arXiv:1101.4665 [hep-ph]].
[94] K. Fuyuto and E. Senaha, Phys. Rev. D **90** (2014) no.1, 015015 doi:10.1103/PhysRevD.90.015015 [arXiv:1406.0433 [hep-ph]].
[95] V. Vaskonen, Phys. Rev. D **95** (2017) no.12, 123515 doi:10.1103/PhysRevD.95.123515 [arXiv:1611.02073 [hep-ph]].
[96] F. P. Huang, Z. Qian and M. Zhang, Phys. Rev. D **98** (2018) no.1, 015014 doi:10.1103/PhysRevD.98.015014 [arXiv:1804.06813 [hep-ph]].
[97] M. Jiang, L. Bian, W. Huang and J. Shu, Phys. Rev. D **93** (2016) no.6, 065032 doi:10.1103/PhysRevD.93.065032 [arXiv:1502.07574 [hep-ph]].
[98] B. Grzadkowski and D. Huang, JHEP **08** (2018), 135 doi:10.1007/JHEP08(2018)135 [arXiv:1807.06987 [hep-ph]].
[99] J. M. Cline, K. Kainulainen and A. P. Vischer, Phys. Rev. D **54** (1996), 2451-2472 doi:10.1103/PhysRevD.54.2451 [arXiv:hep-ph/9506284 [hep-ph]].
[100] J. R. Espinosa, B. Gripaios, T. Konstandin and F. Riva, JCAP **01** (2012), 012 doi:10.1088/1475-7516/2012/01/012 [arXiv:1110.2876 [hep-ph]].
[101] B. Jain, S. J. Lee and M. Son, Phys. Rev. D **98** (2018) no.7, 075002 doi:10.1103/PhysRevD.98.075002 [arXiv:1709.03232 [hep-ph]].
[102] P. Athron *et al.* [GAMBIT], Eur. Phys. J. C **77** (2017) no.8, 568 doi:10.1140/epjc/s10052-017-5113-1 [arXiv:1705.07931 [hep-ph]].
[103] P. Athron, J. M. Cornell, F. Kahlhoefer, J. Mckay, P. Scott and S. Wild, Eur. Phys. J. C **78** (2018) no.10, 830 doi:10.1140/epjc/s10052-018-6314-y [arXiv:1806.11281 [hep-ph]].
[104] L. Feng, S. Profumo and L. Ubaldi, JHEP **03** (2015), 045 doi:10.1007/JHEP03(2015)045 [arXiv:1412.1105 [hep-ph]].
[105] [ATLAS], ATLAS-CONF-2020-052.
[106] A. M. Sirunyan *et al.* [CMS], Phys. Lett. B **793** (2019), 520-551 doi:10.1016/j.physletb.2019.04.025 [arXiv:1809.05937 [hep-ex]].
[107] R. L. Workman *et al.* [Particle Data Group], PTEP **2022** (2022), 083C01 doi:10.1093/ptep/ptac097
[108] [ATLAS], Phys. Lett. B **842** (2023), 137963 doi:10.1016/j.physletb.2023.137963 [arXiv:2301.10731 [hep-ex]].
[109] Y. B. Zeldovich, I. Y. Kobzarev and L. B. Okun, Zh. Eksp. Teor. Fiz. **67** (1974), 3-11 SLAC-TRANS-0165.
[110] A. Angelescu, F. Goertz and A. Tada, JHEP **10** (2022), 019 doi:10.1007/JHEP10(2022)019 [arXiv:2112.12087 [hep-ph]].
[111] M. Dine, R. G. Leigh, P. Y. Huet, A. D. Linde and D. A. Linde, Phys. Rev. D **46** (1992), 550-571 doi:10.1103/PhysRevD.46.550 [arXiv:hep-ph/9203203 [hep-ph]].
[112] L. Dolan and R. Jackiw, Phys. Rev. D **9** (1974), 3320-3341 doi:10.1103/PhysRevD.9.3320
[113] M. Gurtler, E. M. Ilgenfritz and A. Schiller, Phys. Rev. D **56** (1997), 3888-3895 doi:10.1103/PhysRevD.56.3888 [arXiv:hep-lat/9704013 [hep-lat]].
[114] M. Laine and K. Rummukainen, Nucl. Phys. B Proc. Suppl. **73** (1999), 180-185 doi:10.1016/S0920-5632(99)85017-8 [arXiv:hep-lat/9809045 [hep-lat]].
[115] F. Csikor, Z. Fodor and J. Heitger, Phys. Rev. Lett. **82** (1999), 21-24 doi:10.1103/PhysRevLett.82.21 [arXiv:hep-ph/9809291 [hep-ph]].
[116] Y. Aoki, F. Csikor, Z. Fodor and A. Ukawa, Phys. Rev. D **60** (1999), 013001 doi:10.1103/PhysRevD.60.013001 [arXiv:hep-lat/9901021 [hep-lat]].

arXiv:2403.01240v1 [hep-ph] 2 Mar 2024

Analyzing the transport coefficients and observables of a rotating QGP medium in kinetic theory framework with a novel approach to the collision integral

Shubhalaxmi Rath* and Sadhana Dash†

Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India

Abstract

In the present work, we have studied how the rotation of the QGP medium affects the transport coefficients and observables in heavy ion collisions. For the noncentral collisions, although most of the angular momentum gets carried away by the spectators, there still remains a finite angular momentum with a finite range of angular velocity, which thus incites rotation in the produced matter. As a result, various properties of the QGP medium are likely to be modulated by the rotation. We have calculated the transport coefficients and observables, such as electrical conductivity, thermal conductivity, Knudsen number, elliptic flow, specific heat at constant pressure, specific heat at constant volume, trace anomaly, thermal diffusion constant and isothermal compressibility using the kinetic theory to see the effect of rotation on them. In particular, we have used the novel relaxation time approximation for the collision integral in the relativistic Boltzmann transport equation to derive the transport coefficients and compared them with their values in the relaxation time approximation within the kinetic theory approach in conjunction with the finite angular velocity. We have found that the angular velocity plays an important role and enhances the flow of charge and heat in the medium. Further, as compared to the relaxation time approximation, the electrical and thermal conductivities have smaller values in the novel relaxation time approximation and these differences between the conductivities in the said approximations are more pronounced at high temperature than at low temperature. Furthermore, all the aforesaid observables are found to be sensitive to the rotation of the QGP medium.

*shubhalaxmirath@gmail.com

†sadhana@phy.iitb.ac.in

Contents

1 Introduction **2**

2 Rotating QGP medium **5**

2.1 Charge conduction in rotating QGP medium 7

2.2 Heat conduction in rotating QGP medium 10

2.3 Observables in rotating QGP medium 12

3 Results and discussions **17**

4 Conclusions **24**

1 Introduction

The possibility of creating the extreme state of matter that existed in the beginning of the universe, *i.e.* quark-gluon plasma (QGP) in ultrarelativistic heavy ion collisions at Relativistic Heavy Ion Collider (RHIC) and Large Hadron Collider (LHC) has elevated the exploration and understanding of the properties of such matter. In addition, the study of the evolution of the primordial matter, its equation of state and the phase structure within the quantum chromodynamics (QCD) at finite temperature, density and other extreme conditions are some of the important goals of the heavy ion collisions. There exist evidences that, in the noncentral heavy ion collisions, the nuclei colliding at ultrarelativistic energies could possess a large initial angular momentum of the order $J \propto b\sqrt{s_{NN}}$, where b and $\sqrt{s_{NN}}$ are the impact parameter and the center-of-mass energy, respectively. Although a large part of the angular momentum is carried away by the spectators, there still a significant fraction of the angular momentum varying between $10^3\hbar$ and $10^5\hbar$ with angular velocity in the range 0.01 GeV - 0.1 GeV or even more is retained in the interaction zone due to the inhomogeneity of the colliding nuclei in the transverse plane, which further gets transformed to the produced fireball [1–4]. Thus, the larger overlap region between the colliding nuclei or impact parameter and large collision energy may result into higher angular momentum of the QGP. The fireball or QGP may sustain the large angular momentum for a longer time due to the conservation of total angular momentum. Thus, QGP may be perceived as a rotating system and accordingly, its properties could be significantly altered. Consequently, the initial angular momentum of the QGP may give rise to significant observable effects in heavy ion collisions, for

example, global polarization of the thermal photons, dileptons and final hadrons with spin, enhancement of the elliptic flow and broadening of the transverse momentum spectra [5, 6]. Like the rotating QGP, other systems where rotating matter exists are the spinning neutron stars [7] and the nonrelativistic bosonic cold atoms [8]. Rotating QGP medium with finite magnetic field also induces some interesting phenomenological effects, such as chiral vortical effect [9, 10], chiral vortical wave [11], chiral magnetic effect [12, 13], chiral magnetic wave [14, 15], spin alignment of vector mesons [5], emission of circularly polarized photons [16] and spin polarization of observed particles in heavy ion collisions [17–22].

In a system, angular momentum exhibits the vorticity, and the coupling between the rotational motion and the quantum spin can result into polarization of hadrons, or in other words, the polarization of emitted hadrons in heavy ion collisions can be treated as a signature of vorticity in the produced matter. The initial conditions of the collisions, *i.e.* the energy and the impact parameter mainly decide the amount of polarization of observable particles. A review on the progress of theoretical and experimental activities on the vortical behavior of the QGP is found in ref. [23]. The relativistic rotational effect could also influence the thermodynamics and the critical temperature of QGP to hadronic phase transition in QCD. Most of the observations in different approaches, such as the Nambu-Jona-Lasinio model [3, 24], the local density approximation with the boundary condition [25], the hadron resonance gas model [26] and the holographic QCD approach [27] had found that the rotation decreases the critical temperature, whereas an increase of the critical temperature was reported in lattice simulations [28, 29]. According to the study using the Nambu-Jona-Lasinio model, it was estimated that the rotation also decreases the critical temperature of chiral symmetry breaking due to the suppression of chiral condensate in the similar regime [30]. Thus, like the high temperature, the strong magnetic field and large chemical potential, the rapid rotation of QGP matter can be treated as a kind of extreme condition. Out of different external conditions, magnetic field directly affects the quark degrees of freedom and indirectly affects the gluon degrees of freedom only through the quark loops, unlike the temperature that affects all degrees of freedom in QCD. The rotation could also affect quark as well as gluon degrees of freedom.

Recently, a lot of effort has been put into the investigation of various features of magnetized and fast rotating QCD matter due to the emergence of extremely strong magnetic field and large angular momentum in the noncentral events of heavy ion collisions. Both magnetic field and rotation would significantly alter the properties of hot and dense matter. Studies on the rotating free fermion states were carried out in references [18, 31–34], whereas the system of interacting fermions was studied in references [25, 30, 35] using effective field theoretical models, in references [36, 37] using the holographic model and in ref. [24] using the Nambu-Jona-Lasinio model. Further, the impact of vorticity on the

dilepton production rate was investigated in ref. [38], where it was reported that the dilepton yield gets suppressed due to the presence of vorticity and the degree of suppression depends on the initial value of vorticity. The effect of magnetic field on different properties of the thermal QCD medium has been extensively studied using different methods [39–63], while the effect of rotation attracts growing interests recently. However, studies on the thermodynamic and transport properties of fast rotating QGP medium are still relatively few in the literature.

In this work, we intend to study the transport properties related to charge and heat, and some associated observables for the QGP medium subjected to the extreme conditions of rapid rotation and high temperature. It would be interesting to see how the rapid rotation modulates the flow of charge and heat in QGP medium. In addition, the observables, such as the Knudsen number, elliptic flow, specific heat at constant pressure, specific heat at constant volume, trace anomaly, thermal diffusion constant and isothermal compressibility are also most likely to be affected by the rotation of the QGP medium. Consequently, it can be expected that this study on the transport phenomena and observables of QGP medium with rotation will provide deeper insight into the properties of the strongly interacting system. We determine the transport coefficients, such as the electrical conductivity and the thermal conductivity using the relativistic Boltzmann transport equation in the novel relaxation time approximation within the kinetic theory approach. The relaxation time approximation may defy the local particle number conservation in the medium, because the charge is not conserved instantaneously but only on the average over a cycle. However, the validity of the conservation laws can be satisfied by using the novel relaxation time approximation. In this method, the collision integral takes a modified form to ensure the validity of the energy-momentum conservation laws [64, 65]. Thus the novel relaxation time approximation preserves the fundamental properties of the collision associated with the microscopic conservation laws and is imperative when considering the momentum-dependent relaxation time. In addition, the rotational effect is incorporated through the parton distribution functions. Further, the interactions among partons are included through their thermal masses in the quasiparticle model. In this model, the quark-gluon plasma is treated as a medium consisting of thermally massive quasiparticles.

The present paper is organized as follows. In section 2, the electrical and thermal conductivities of a rotating QGP medium are studied by calculating them using the relativistic Boltzmann transport equation in the novel relaxation time approximation within the kinetic theory approach. Then the effects of rotation on the observables, such as the Knudsen number, elliptic flow, specific heat at constant pressure, specific heat at constant volume, trace anomaly, thermal diffusion constant and isothermal compressibility are studied. The results on the aforesaid conductivities and observables are discussed in

section 3. In section 4, the results of this work are concluded.

2 Rotating QGP medium

In a rotating QGP medium, the phase-space distribution functions of partons get significantly affected. In such rotating system with angular velocity ω, the density operator can be written as

$$\hat{\rho} = \frac{1}{Z} e^{\beta\left(-\hat{H}+\mu\hat{N}+\omega\hat{J}\right)} P_V, \tag{1}$$

where $\hat{H}$, $\hat{N}$ and $\hat{J}$ represent the Hamiltonian operator, the number operator and the angular momentum operator, respectively. Here, $T = \beta^{-1}$ and P_V denotes the projector onto the localized states $|h_V\rangle$, *i.e.* $P_V = \sum_{h_V} |h_V\rangle\langle h_V|$, which forms a complete set of quantum states for the system in a finite region V. The partition function, Z of the rotating system is expressed as

$$Z = Tr\left[e^{\beta\left(-\hat{H}+\mu\hat{N}+\omega\hat{J}\right)} P_V\right]. \tag{2}$$

For a relativistic gas, such as QGP, the partition function can be written as a product of single particle partition functions. Thus, eq. (2) can be calculated by computing the matrix elements of operators $\hat{h}$, $\hat{n}$ and $\hat{j}$, suitable for single particles as

$$\langle p,\tau|e^{\beta\left(-\hat{h}+\mu\hat{n}+\omega\hat{j}\right)} P_V|p,\sigma\rangle = \left(e^{\beta(u_\mu p^\mu + a)} + b\right)^{-1} \langle p,\tau|e^{\beta\omega\hat{j}} P_V|p,\sigma\rangle, \tag{3}$$

where $a = -\mu$, $+\mu$ and 0 for quarks, antiquarks and gluons, respectively, and $b = +1$ for quarks as well as for antiquarks and $b = -1$ for gluons. In the above equation, τ and σ represent the polarization states. The matrix element appearing on the right-hand side of eq. (3) can be determined by the analytical continuation to the imaginary ω values. In this process, $\beta\omega$ is replaced by $-i\phi$ through the rotation $R_{\hat{\omega}}(\phi)$ around the axis ω by an angle ϕ. Now, the matrix element on the right-hand side of eq. (3) can be expanded as

$$\begin{aligned}\langle p,\tau|e^{\beta\omega\hat{j}} P_V|p,\sigma\rangle &= \langle p,\tau|R_{\hat{\omega}}(\phi) P_V|p,\sigma\rangle \\ &= \sum_{\sigma'} \int d^3\mathrm{p}' \langle p,\tau|R_{\hat{\omega}}(\phi)|p',\sigma'\rangle\langle p',\sigma'|P_V|p,\sigma\rangle.\end{aligned} \tag{4}$$

The matrix element of the representation of a rotation involves a Dirac delta function and a Wigner rotation matrix, *i.e.*,

$$\langle p,\tau|R_{\hat{\omega}}(\phi)|p',\sigma'\rangle = \delta^3\left(p - R_{\hat{\omega}}(\phi)(p')\right) D^S\left(\left[R_{\hat{\omega}}(\phi)(p')\right]^{-1} R_{\hat{\omega}}(\phi)\,[p']\right)_{\tau\sigma'}. \tag{5}$$

The matrix element of the projector P_V can be obtained by using the quantum field theoretical framework [66] as

$$\langle p',\sigma'|P_V|p,\sigma\rangle = \frac{1}{2}\sqrt{\frac{\varepsilon}{\varepsilon'}} \int d^3\mathrm{x}\; e^{i\mathbf{x}\cdot(\mathbf{p}-\mathbf{p}')} \left(D^S\left([p']^{-1}\,[p]\right) + D^S\left([p']^{\dagger}\,[p]^{\dagger-1}\right)\right)_{\sigma'\sigma} \langle 0|P_V|0\rangle, \tag{6}$$

where $\langle 0|P_V|0\rangle$ is the vacuum expectation value of the projector P_V. For large volume V, $\langle 0|P_V|0\rangle \to 1$, because P_V tends to unity in this case. Substituting eq. (5) and eq. (6) in eq. (4), we have

$$\begin{aligned}\langle p,\tau|R_{\hat{\omega}}(\phi)P_V|p,\sigma\rangle &= \int d^3\mathrm{x}\; e^{i\mathbf{x}\cdot\left(\mathbf{p}-R_{\hat{\omega}}(\phi)^{-1}(\mathbf{p})\right)} \\ &\times\frac{1}{2}\left(D^S\left([p]^{-1}\,R_{\hat{\omega}}(\phi)\,[p]\right)+D^S\left([p]^{\dagger}\,R_{\hat{\omega}}(\phi)\,[p]^{\dagger-1}\right)\right)_{\tau\sigma}. \quad (7)\end{aligned}$$

Using the unitarity of the Wigner rotation, *i.e.*

$$D^S\left([R_{\hat{\omega}}(\phi)(p')]^{-1}\,R_{\hat{\omega}}(\phi)\,[p']\right)=D^S\left([R_{\hat{\omega}}(\phi)(p')]^{\dagger}\,R_{\hat{\omega}}(\phi)\,[p']^{\dagger-1}\right),$$

eq. (7) can be analytically extended to the imaginary angles. Thus the matrix element in eq. (3) takes the following form,

$$\begin{aligned}\langle p,\tau|e^{\beta\omega\hat{j}}P_V|p,\sigma\rangle &= \int d^3\mathrm{x}\; e^{i\mathbf{x}\cdot\left(\mathbf{p}-R_{\hat{\omega}}(i\beta\omega)^{-1}(\mathbf{p})\right)} \\ &\times\frac{1}{2}\left(D^S\left([p]^{-1}\,R_{\hat{\omega}}(i\beta\omega)\,[p]\right)+D^S\left([p]^{\dagger}\,R_{\hat{\omega}}(i\beta\omega)\,[p]^{\dagger-1}\right)\right)_{\tau\sigma}. \quad (8)\end{aligned}$$

It can be observed that, $R_{\hat{\omega}}(i\beta\omega)=e^{\beta\omega J}$. The matrix element in eq. (3) has a spatial integral form, which allows us to write the phase-space distribution function as

$$\begin{aligned}f(x,p)_{\tau\sigma} &= \left(e^{\beta(u_\mu p^\mu+a)}+b\right)^{-1} e^{i\mathbf{x}\cdot\left(\mathbf{p}-R_{\hat{\omega}}(i\beta\omega)^{-1}(\mathbf{p})\right)} \\ &\times\frac{1}{2}\left(D^S\left([p]^{-1}\,R_{\hat{\omega}}(i\beta\omega)\,[p]\right)+D^S\left([p]^{\dagger}\,R_{\hat{\omega}}(i\beta\omega)\,[p]^{\dagger-1}\right)\right)_{\tau\sigma}, \quad (9)\end{aligned}$$

where u_μ represents the four-velocity of fluid. For high temperature QCD medium, $\beta\omega$ is small, *i.e.* $\beta\omega<1$. Thus the difference between momenta in the exponent of eq. (9) can be approximated by the lowest order term in $\beta\omega$,

$$\begin{aligned}\mathbf{p}-R_{\hat{\omega}}(i\beta\omega)^{-1}(\mathbf{p}) &= \mathbf{p}-[\cosh\beta\omega\;\mathbf{p}-i\sinh\beta\omega\;\hat{\omega}\times\mathbf{p}+(1-\cosh\beta\omega)\,\mathbf{p}\;\hat{\omega}\hat{\omega}] \\ &\simeq i\beta\;\boldsymbol{\omega}\times\mathbf{p}. \quad (10)\end{aligned}$$

Now, the phase-space distribution function becomes

$$\begin{aligned}f(x,p)_{\tau\sigma} &= \left(e^{\beta(u_\mu p^\mu+a)}+b\right)^{-1} e^{-\beta\mathbf{x}\cdot(\boldsymbol{\omega}\times\mathbf{p})} \\ &\times\frac{1}{2}\left(D^S\left([p]^{-1}\,R_{\hat{\omega}}(i\beta\omega)\,[p]\right)+D^S\left([p]^{\dagger}\,R_{\hat{\omega}}(i\beta\omega)\,[p]^{\dagger-1}\right)\right)_{\tau\sigma} \\ &= \left(e^{\beta(u_\mu p^\mu+a)}+b\right)^{-1} e^{\beta\mathbf{p}\cdot\mathbf{v}} \\ &\times\frac{1}{2}\left(D^S\left([p]^{-1}\,R_{\hat{\omega}}(i\beta\omega)\,[p]\right)+D^S\left([p]^{\dagger}\,R_{\hat{\omega}}(i\beta\omega)\,[p]^{\dagger-1}\right)\right)_{\tau\sigma}, \quad (11)\end{aligned}$$

where $\mathbf{v}=\boldsymbol{\omega}\times\mathbf{x}$ has been used. The above equation is the unnormalized single-particle phase-space distribution function. By taking the trace of the matrix (11), one can obtain

the phase-space density in (x,p) as

$$\begin{aligned} f(x,p) &= \sum_{\sigma} f(x,p)_{\sigma\sigma} \\ &= \left(e^{\beta(u_\mu p^\mu + a)} + b\right)^{-1} e^{\beta \mathbf{p}\cdot\mathbf{v}} \chi(\beta\omega). \end{aligned} \tag{12}$$

In the local rest frame, $\mathbf{v} = 0$. So, we have

$$f(x,p) = \frac{1}{e^{\beta(u_\mu p^\mu + a)} + b} \chi(\beta\omega), \tag{13}$$

where $\chi(\beta\omega)$ represents the angular velocity-dependent part and is given by

$$\begin{aligned} \chi(\beta\omega) &= Tr\left[D^S\left(R_{\hat{\omega}}(i\beta\omega)\right)\right] \\ &= \frac{\sinh\left(\left(S+\frac{1}{2}\right)\beta\omega\right)}{\sinh\left(\frac{\beta\omega}{2}\right)}. \end{aligned} \tag{14}$$

For quarks and antiquarks, spin $S = \frac{1}{2}$ and for gluons, $S = 1$. Accordingly, they have different expressions for $\chi(\beta\omega)$. Thus, quark, antiquark and gluon distribution functions at finite rotation can be written as

$$f_q = f_0^q \chi_q(\beta\omega), \tag{15}$$

$$\bar{f}_q = \bar{f}_0^q \chi_{\bar{q}}(\beta\omega), \tag{16}$$

$$f_g = f_0^g \chi_g(\beta\omega), \tag{17}$$

respectively. The quantities in the above equations are defined as follows,

$$f_0^q = \frac{1}{e^{\beta(u_\mu p^\mu - \mu)} + 1} = \frac{1}{e^{\beta\left(\omega_f - \mu\right)} + 1}, \quad \bar{f}_0^q = \frac{1}{e^{\beta(u_\mu p^\mu + \mu)} + 1} = \frac{1}{e^{\beta\left(\omega_f + \mu\right)} + 1},$$

$$f_0^g = \frac{1}{e^{\beta u_\mu p^\mu} - 1} = \frac{1}{e^{\beta\omega_g} - 1}, \quad \chi_q(\beta\omega) = \frac{\sinh(\beta\omega)}{\sinh(\beta\omega/2)},$$

$$\chi_{\bar{q}}(\beta\omega) = \frac{\sinh(\beta\omega)}{\sinh(\beta\omega/2)}, \quad \chi_g(\beta\omega) = \frac{\sinh(3\beta\omega/2)}{\sinh(\beta\omega/2)},$$

where ω_f is the energy of the fth flavor of quark (antiquark) and ω_g is the energy of the gluon in the rotating QGP medium. We note that the flavor degrees of freedom are taken into account explicitly. Further, the chemical potentials for flavors, u, d and s are kept the same ($\mu_f = \mu$).

2.1 Charge conduction in rotating QGP medium

When a rotating QGP medium is exposed to an external electric field, an electric current density gets induced and its spatial component is written as

$$J^i = \sum_f g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3 \omega_f} p^i \left[q_f \delta f_q + \bar{q}_f \delta \bar{f}_q\right], \tag{18}$$

where g_f, q_f ($\bar{q}_f$) and δf_q ($\delta \bar{f}_q$) represent the degeneracy factor, electric charge and infinitesimal change in the quark (antiquark) distribution function of fth flavor, respectively. The Ohm's law relates the spatial current density with the electric field and the electrical conductivity as

$$J^i = \sigma_{\rm el} E^i \ . \tag{19}$$

In order to calculate the infinitesimal change δf_q, we use the relativistic Boltzmann transport equation (RBTE) in the novel relaxation time approximation (RTA). In this approximation, the collision integral takes a modified form for the validity of the conservation laws [64, 65]. However, in the frequently used relaxation time approximation, the conservation laws are satisfied, if the relaxation time is momentum-independent. The RBTE in the novel RTA is written as

$$p^\mu \frac{\partial f_q'}{\partial x^\mu} + q_f F^{\rho\sigma} p_\sigma \frac{\partial f_q'}{\partial p^\rho} = -\frac{p_\nu u^\nu}{\tau_f} \Bigg[\delta f_q - \frac{\langle (\omega_f/\tau_f)\, \delta f_q \rangle_0}{\langle \omega_f/\tau_f \rangle_0} + \frac{P_1^{(0)} \left\langle (\omega_f/\tau_f)\, P_1^{(0)} \delta f_q \right\rangle_0}{\left\langle (\omega_f/\tau_f)\, P_1^{(0)} P_1^{(0)} \right\rangle_0} + \frac{p^{\langle\mu\rangle} \left\langle (\omega_f/\tau_f)\, p_{\langle\mu\rangle} \delta f_q \right\rangle_0}{(1/3) \left\langle (\omega_f/\tau_f)\, p_{\langle\mu\rangle} p^{\langle\mu\rangle} \right\rangle_0} \Bigg] , \tag{20}$$

where τ_f represents the novel relaxation time, $f_q' = \delta f_q + f_q$ with δf_q denoting the infinitesimal change of the quark distribution function due to the external field and $F^{\rho\sigma}$ is the electromagnetic field strength tensor, whose components are related to the electric and magnetic fields. The novel relaxation time for quarks (antiquarks), τ_f ($\tau_{\bar{f}}$) is momentum-dependent with the following form,

$$\tau_{f(\bar{f})} = (\beta \omega_f)^\xi \, t_{f(\bar{f})}, \tag{21}$$

where ξ is an arbitrary constant that controls the energy dependence of the relaxation time and $t_{f(\bar{f})}$ is momentum-independent, whose form is given [67] by

$$t_{f(\bar{f})} = \frac{1}{5.1 T \alpha_s^2 \log(1/\alpha_s) \left[1 + 0.12(2N_f + 1)\right]} \ . \tag{22}$$

For the calculations, the value of ξ can be taken up to 1, for example, in QCD kinetic theories, $\xi = \frac{1}{2}$ [68], while in scalar field theories $\xi = 1$ [69].

In eq. (20), $P_1^{(0)}$ and $p^{\langle\mu\rangle}$ are respectively defined as

$$P_1^{(0)} = 1 - \frac{\langle \omega_f/\tau_f \rangle_0 \, \omega_f}{\left\langle \omega_f^2/\tau_f \right\rangle_0}, \tag{23}$$

$$p^{\langle\mu\rangle} = \Delta^{\mu\nu} p_\nu, \tag{24}$$

where $\Delta^{\mu\nu} = g^{\mu\nu} - u^\mu u^\nu$. The momentum integral of A (A can be different terms appearing in eq. (20)) relative to the local equilibrium distribution function is defined as

$$\langle A \rangle = \int \frac{d^3{\rm p}}{(2\pi)^3 \omega_f} A f_q \ . \tag{25}$$

Using equations (23), (24), (25) and integration by parts, the second (say L^2), third (say L^3) and fourth (say L^4) terms appearing inside the square bracket of eq. (20) can be calculated to take the following forms,

$$L^2 = \delta f_q, \tag{26}$$

$$L^3 = \frac{\delta f_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right) \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right)}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right)^2}, \tag{27}$$

$$L^4 = \frac{3\ \delta f_q\ \mathrm{p} \int d\mathrm{p}\ \frac{\mathrm{p}^3}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\tau_f} f_q}. \tag{28}$$

For the calculation of electrical conductivity, we use the components of $F^{\rho\sigma}$ associated with only electric field. Moreover, in case of a spatially homogeneous distribution function with the steady-state condition, we use $\frac{\partial f_q'}{\partial \mathbf{r}} = 0$ and $\frac{\partial f_q'}{\partial t} = 0$. Thus, RBTE (20) becomes

$$q_f \mathbf{E} \cdot \mathbf{p} \frac{\partial f_q'}{\partial p_0} + q_f p_0 \mathbf{E} \cdot \frac{\partial f_q'}{\partial \mathbf{p}} = -\frac{p_0}{\tau_f} \delta f_q J, \tag{29}$$

where the expression of J is written as

$$J = \frac{\left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right) \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right)}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_f}\omega_f f_q}\right)^2} + \frac{3\mathrm{p} \int d\mathrm{p}\ \frac{\mathrm{p}^3}{\tau_f} f_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\tau_f} f_q}. \tag{30}$$

In the antiquark case, $\bar{J}$ is given by

$$\bar{J} = \frac{\left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}} \bar{f}_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}}\omega_f \bar{f}_q}\right) \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}} \bar{f}_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}} \bar{f}_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}}\omega_f \bar{f}_q}\right)}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}} \bar{f}_q \left(1 - \frac{\omega_f \int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}} \bar{f}_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\tau_{\bar{f}}}\omega_f \bar{f}_q}\right)^2} + \frac{3\mathrm{p} \int d\mathrm{p}\ \frac{\mathrm{p}^3}{\tau_{\bar{f}}} \bar{f}_q}{\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\tau_{\bar{f}}} \bar{f}_q}. \tag{31}$$

Using equations (15), (16) and (21) in equations (30) and (31), both J and $\bar{J}$ can be

simplified, and they are respectively written as

$$J = \frac{\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}f_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}f_0^q}\right)\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}f_0^q\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}f_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}f_0^q}\right)}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}f_0^q\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}f_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}f_0^q}\right)^2} + \frac{3\mathrm{p}\int d\mathrm{p}\ \frac{\mathrm{p}^3}{\omega_f^\xi}f_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_f^\xi}f_0^q}, \tag{32}$$

$$\bar{J} = \frac{\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}\bar{f}_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}\bar{f}_0^q}\right)\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}\bar{f}_0^q\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}\bar{f}_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}\bar{f}_0^q}\right)}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}\bar{f}_0^q\left(1-\frac{\omega_f\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^\xi}\bar{f}_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^2}{\omega_f^{\xi-1}}\bar{f}_0^q}\right)^2} + \frac{3\mathrm{p}\int d\mathrm{p}\ \frac{\mathrm{p}^3}{\omega_f^\xi}\bar{f}_0^q}{\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_f^\xi}\bar{f}_0^q}. \tag{33}$$

From eq. (29), δf_q is obtained as

$$\delta f_q = \frac{2\tau_f q_f \beta}{\omega_f J}\mathbf{E}\cdot\mathbf{p}f_0^q\left(1-f_0^q\right)\chi_q(\beta\omega)\ . \tag{34}$$

Similarly, $\delta\bar{f}_q$ for antiquark is determined as

$$\delta\bar{f}_q = \frac{2\tau_{\bar{f}} \bar{q}_f \beta}{\omega_f \bar{J}}\mathbf{E}\cdot\mathbf{p}\bar{f}_0^q\left(1-\bar{f}_0^q\right)\chi_{\bar{q}}(\beta\omega)\ . \tag{35}$$

Substituting the values of δf_q and $\delta\bar{f}_q$ in eq. (18), using eq. (21) and then comparing with eq. (19), we get the electrical conductivity for a rotating QGP medium as

$$\sigma_{\mathrm{el}} \ = \ \frac{\beta^{1+\xi}}{3\pi^2}\sum_f g_f q_f^2\int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_f^{2-\xi}}\left[\frac{t_f}{J}f_0^q\left(1-f_0^q\right)\chi_q(\beta\omega)+\frac{t_{\bar{f}}}{\bar{J}}\bar{f}_0^q\left(1-\bar{f}_0^q\right)\chi_{\bar{q}}(\beta\omega)\right]. \tag{36}$$

2.2 Heat conduction in rotating QGP medium

In a thermal medium, the heat flow four-vector is related to the energy diffusion and the enthalpy diffusion as

$$Q_\mu = \Delta_{\mu\alpha}T^{\alpha\beta}u_\beta - h\Delta_{\mu\alpha}N^\alpha, \tag{37}$$

where $T^{\alpha\beta}$ is the energy-momentum tensor, N^α represents the particle flow four-vector, $\Delta_{\mu\alpha} = g_{\mu\alpha} - u_\mu u_\alpha$, the enthalpy per particle $h = (\varepsilon + P)/n$ with ε, P and n denoting the energy density, the pressure and the particle number density, respectively. N^α and $T^{\alpha\beta}$ are also known as the first and second moments of the distribution function, respectively

with the following expressions,

$$N^{\alpha} = \sum_{f} g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3 \omega_f} p^{\alpha} \left[f_q + \bar{f}_q \right], \tag{38}$$

$$T^{\alpha\beta} = \sum_{f} g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3 \omega_f} p^{\alpha} p^{\beta} \left[f_q + \bar{f}_q \right]. \tag{39}$$

Using N^{α} and $T^{\alpha\beta}$, it is possible to get $n = N^{\alpha} u_{\alpha}$, $\varepsilon = u_{\alpha} T^{\alpha\beta} u_{\beta}$ and $P = -\Delta_{\alpha\beta} T^{\alpha\beta}/3$. In the local rest frame, the heat flow is purely spatial, because $Q_{\mu} u^{\mu} = 0$. Hence, the spatial component of the heat flow remains finite and is expressed as

$$Q^{i} = \sum_{f} g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3} \frac{p^{i}}{\omega_f} \left[(\omega_f - h_f) \delta f_q + (\omega_f - \bar{h}_f) \delta \bar{f}_q \right]. \tag{40}$$

The Navier-Stokes equation relates the heat flow with the gradients of temperature and pressure as

$$Q^{i} = -\kappa \delta^{ij} \left[\partial_j T - \frac{T}{\varepsilon + P} \partial_j P \right], \tag{41}$$

where κ is the thermal conductivity. For the calculation of thermal conductivity, the electromagnetic field strength part is not required, so it can be dropped from the relativistic Boltzmann transport equation (20). Now, expanding the gradient of the particle distribution function in terms of the gradients of the flow velocity and the temperature in the novel relaxation time approximation, we have

$$-p^{\mu} f_0^q \left(1 - f_0^q\right) \chi_q(\beta\omega) \left[\left(u_{\alpha} p^{\alpha}\right) \partial_{\mu} \beta + \beta \partial_{\mu} \left(u_{\alpha} p^{\alpha}\right) - \partial_{\mu} \left(\beta \mu\right) \right] + p^{\mu} f_0^q \partial_{\mu} \left(\chi_q(\beta\omega)\right) = -\frac{p_{\nu} u^{\nu}}{\tau_f} \left[\delta f_q - \frac{\langle (\omega_f/\tau_f)\, \delta f_q \rangle_0}{\langle \omega_f/\tau_f \rangle_0} + \frac{P_1^{(0)} \left\langle (\omega_f/\tau_f)\, P_1^{(0)} \delta f_q \right\rangle_0}{\left\langle (\omega_f/\tau_f)\, P_1^{(0)} P_1^{(0)} \right\rangle_0} + \frac{p^{\langle\mu\rangle} \left\langle (\omega_f/\tau_f)\, p_{\langle\mu\rangle} \delta f_q \right\rangle_0}{(1/3) \left\langle (\omega_f/\tau_f)\, p_{\langle\mu\rangle} p^{\langle\mu\rangle} \right\rangle_0} \right]. \tag{42}$$

After solving eq. (42), δf_q can be obtained as

$$\begin{aligned} \delta f_q \;=\; & -\frac{\beta \tau_f f_0^q \left(1 - f_0^q\right)}{J} \chi_q(\beta\omega) \left[\frac{\left(\omega_f - h_f\right) p^{j}}{T \omega_f} \left(\partial_j T - \frac{T}{n h_f} \partial_j P \right) + p_0 \frac{DT}{T} - \frac{p^{\mu} p^{\alpha}}{p_0} \nabla_{\mu} u_{\alpha} \right. \\ & \left. + TD\left(\frac{\mu}{T}\right) \right] - \frac{\beta \tau_f f_0^q}{J} \sinh\left(\frac{\beta\omega}{2}\right) \left[TD\left(\frac{\omega}{T}\right) + \frac{T p^{\alpha}}{p_0} \nabla_{\alpha} \left(\frac{\omega}{T}\right) \right]. \end{aligned} \tag{43}$$

In the similar way, $\delta \bar{f}_q$ for antiquark is determined as

$$\begin{aligned} \delta \bar{f}_q \;=\; & -\frac{\beta \tau_{\bar{f}} \bar{f}_0^q \left(1 - \bar{f}_0^q\right)}{\bar{J}} \chi_{\bar{q}}(\beta\omega) \left[\frac{\left(\omega_f - \bar{h}_f\right) p^{j}}{T \omega_f} \left(\partial_j T - \frac{T}{n \bar{h}_f} \partial_j P \right) + p_0 \frac{DT}{T} - \frac{p^{\mu} p^{\alpha}}{p_0} \nabla_{\mu} u_{\alpha} \right. \\ & \left. - TD\left(\frac{\mu}{T}\right) \right] - \frac{\beta \tau_{\bar{f}} \bar{f}_0^q}{\bar{J}} \sinh\left(\frac{\beta\omega}{2}\right) \left[TD\left(\frac{\omega}{T}\right) + \frac{T p^{\alpha}}{p_0} \nabla_{\alpha} \left(\frac{\omega}{T}\right) \right]. \end{aligned} \tag{44}$$

Substituting the values of δf_q and $\delta \bar{f}_q$ in eq. (40), using eq. (21) and then comparing with eq. (41), we get the thermal conductivity for a rotating QGP medium as

$$\kappa = \frac{\beta^{2+\xi}}{6\pi^2}\sum_f g_f \int d\mathrm{p}\, \frac{\mathrm{p}^4}{\omega_f^{2-\xi}}\left[\frac{t_f}{J}(\omega_f - h_f)^2 f_0^q\,(1-f_0^q)\,\chi_q(\beta\omega)\right.$$
$$\left. + \frac{t_{\bar{f}}}{\bar{J}}(\omega_f - \bar{h}_f)^2 \bar{f}_0^q\left(1-\bar{f}_0^q\right)\chi_{\bar{q}}(\beta\omega)\right]. \tag{45}$$

2.3 Observables in rotating QGP medium

The observables such as the Knudsen number, elliptic flow, specific heat at constant pressure, specific heat at constant volume, trace anomaly, thermal diffusion constant and isothermal compressibility are also likely to be affected by the rotation of the QGP medium.

The Knudsen number is the ratio of the mean free path (λ) to the characteristic length scale (l) of the medium, *i.e.* $\Omega = \lambda/l$. For the medium to be in the equilibrium state, l should be larger than λ. Using $\lambda = 3\kappa/(vC_V)$, the Knudsen number becomes

$$\Omega = \frac{3\kappa}{lvC_V}\ , \tag{46}$$

where v and C_V are the relative speed and the specific heat at constant volume, respectively. The Knudsen number explains the competition between the microscopic and macroscopic length scales of the concerned system. If Ω is less than unity, then the equilibrium hydrodynamics can be applicable and in the hydrodynamic limit, the system assumes local thermodynamical equilibrium. On the other hand, the Knudsen number is large for rarefied gases and the regime of extremely large Ω is known as the free streaming particle regime. In heavy ion collisions, the degree of thermalization in the produced matter can be characterized by the Knudsen number. The experimental extraction of the Knudsen number relies on various quantities, such as the overlap area between two colliding nuclei in the transverse plane of the collision zone, eccentricity, multiplicity density, centrality etc. It is reported in ref. [70] that the inverse of the Knudsen number can be obtained experimentally as it is directly proportional to the multiplicity density and inversely proportional to the transverse area of the collision zone. In this work, we have

used $v \simeq 1$ and $l = 4$ fm, and calculated C_V from the energy-momentum tensor as follows,

$$\begin{aligned} C_V &= \frac{\partial(u_\mu T^{\mu\nu} u_\nu)}{\partial T} \\ &= \sum_f g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3}\, \omega_f \left[\frac{\partial f_q}{\partial T} + \frac{\partial \bar{f}_q}{\partial T}\right] + g_g \int \frac{d^3\mathrm{p}}{(2\pi)^3}\, \omega_g \frac{\partial f_g}{\partial T} \\ &= \frac{1}{2\pi^2}\sum_f g_f \int d\mathrm{p}\ \mathrm{p}^2 \omega_f \left[\chi_q(\beta\omega)\frac{\partial f_0^q}{\partial T} + f_0^q \frac{\partial \chi_q(\beta\omega)}{\partial T} + \chi_{\bar{q}}(\beta\omega)\frac{\partial \bar{f}_0^q}{\partial T} + \bar{f}_0^q \frac{\partial \chi_{\bar{q}}(\beta\omega)}{\partial T}\right] \\ &\quad + \frac{1}{2\pi^2} g_g \int d\mathrm{p}\ \mathrm{p}^2 \omega_g \left[\chi_g(\beta\omega)\frac{\partial f_0^g}{\partial T} + f_0^g \frac{\partial \chi_g(\beta\omega)}{\partial T}\right]. \end{aligned} \tag{47}$$

In the above equation, the values of different partial derivatives are given by

$$\frac{\partial f_0^q}{\partial T} = \beta^2(\omega_f - \mu) f_0^q \left(1 - f_0^q\right), \tag{48}$$

$$\frac{\partial \bar{f}_0^q}{\partial T} = \beta^2(\omega_f + \mu) \bar{f}_0^q \left(1 - \bar{f}_0^q\right), \tag{49}$$

$$\frac{\partial f_0^g}{\partial T} = \beta^2 \omega_g f_0^g \left(1 + f_0^g\right), \tag{50}$$

$$\frac{\partial \chi_q(\beta\omega)}{\partial T} = \frac{\partial \chi_{\bar{q}}(\beta\omega)}{\partial T} = -\beta^2 \omega \sinh\left(\frac{\beta\omega}{2}\right), \tag{51}$$

$$\frac{\partial \chi_g(\beta\omega)}{\partial T} = -2\beta^2 \omega \sinh(\beta\omega). \tag{52}$$

Now, eq. (47) becomes

$$\begin{aligned} C_V &= \frac{\beta^2}{2\pi^2}\sum_f g_f \int d\mathrm{p}\ \mathrm{p}^2 \omega_f \left[\left\{(\omega_f - \mu)\, f_0^q \left(1 - f_0^q\right) + (\omega_f + \mu)\, \bar{f}_0^q \left(1 - \bar{f}_0^q\right)\right\} \frac{\sinh(\beta\omega)}{\sinh(\beta\omega/2)}\right. \\ &\quad \left. - \left(f_0^q + \bar{f}_0^q\right) \omega \sinh(\beta\omega/2)\right] + \frac{\beta^2}{2\pi^2} g_g \int d\mathrm{p}\ \mathrm{p}^2 \omega_g \left[\omega_g f_0^g \left(1 + f_0^g\right) \frac{\sinh(3\beta\omega/2)}{\sinh(\beta\omega/2)}\right. \\ &\quad \left. - 2 f_0^g \omega \sinh(\beta\omega)\right]. \end{aligned} \tag{53}$$

Using the values of κ (45) and C_V (53) in eq. (46), the Knudsen number for the QGP medium at finite rotation can be explored.

The elliptic flow (v_2) is one of the key observables in ultrarelativistic heavy ion collisions at terrestrial laboratories. In such collisions, the overlapping region of bombarding nuclei is the origin of generation of the elliptic flow, where the pressure gradient converts the initial spatial anisotropy (quantified by the eccentricity ϵ) into the momentum anisotropy. The elliptic flow develops gradually with the evolution of the system. The anisotropy of the momentum distributions can be completely attained only once all parts of the system get involved in the process. It also gives the information about the degree of interactions/reinteractions among the particles produced in heavy ion collisions. Thus, its magnitude is associated with the equilibrium property of the thermal medium and

any deviation from equilibrium can reduce the magnitude of such flow. Experimentally, elliptic flow can be analyzed by selecting events in a centrality class. The quantities which influence the elliptic flow, such as the density, the transverse size and the eccentricity, fluctuate from one event to the other, thus causing dynamical fluctuations of the elliptic flow. The elliptic flow is closely related to the Knudsen number and is defined [70–72] as

$$v_2 = \frac{v_2^h}{1+\frac{\Omega}{\Omega_h}}. \tag{54}$$

Here, v_2^h represents the elliptic flow in the hydrodynamic limit ($\Omega \to 0$ limit) and Ω_h is the value of the Knudsen number calculated by observing the transition between the hydrodynamic regime and the free streaming particle regime. The transport calculation [72] has estimated $\Omega_h \approx 0.7$ and $v_2^h \approx 0.1$. The presence of magnetic field, temperature and chemical potential were observed to influence the elliptic flow [45, 73, 74]. Similarly, the finite rotation could also affect the elliptic flow. Substituting the value of Ω (46) in eq. (54), the effect of finite rotation on the elliptic flow can be investigated.

Like the specific heat at constant volume, it is possible to determine the specific heat at constant pressure from the energy-momentum tensor as

$$\begin{aligned}
C_P &= \frac{\partial(u_\mu T^{\mu\nu}u_\nu - \Delta_{\mu\nu}T^{\mu\nu}/3)}{\partial T} \\
&= \sum_f g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3}\,\omega_f \left[\frac{\partial f_q}{\partial T}+\frac{\partial \bar{f}_q}{\partial T}\right] + \frac{1}{3}\sum_f g_f \int \frac{d^3\mathrm{p}}{(2\pi)^3}\,\frac{\mathrm{p}^2}{\omega_f}\left[\frac{\partial f_q}{\partial T}+\frac{\partial \bar{f}_q}{\partial T}\right] \\
&\quad + g_g \int \frac{d^3\mathrm{p}}{(2\pi)^3}\,\omega_g \frac{\partial f_g}{\partial T} + \frac{1}{3} g_g \int \frac{d^3\mathrm{p}}{(2\pi)^3}\,\frac{\mathrm{p}^2}{\omega_g}\frac{\partial f_g}{\partial T} \\
&= \frac{1}{2\pi^2}\sum_f g_f \int d\mathrm{p}\ \mathrm{p}^2\omega_f \left[\chi_q(\beta\omega)\frac{\partial f_0^q}{\partial T} + f_0^q \frac{\partial \chi_q(\beta\omega)}{\partial T} + \chi_{\bar{q}}(\beta\omega)\frac{\partial \bar{f}_0^q}{\partial T} + \bar{f}_0^q \frac{\partial \chi_{\bar{q}}(\beta\omega)}{\partial T}\right] \\
&\quad + \frac{1}{6\pi^2}\sum_f g_f \int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_f}\left[\chi_q(\beta\omega)\frac{\partial f_0^q}{\partial T} + f_0^q \frac{\partial \chi_q(\beta\omega)}{\partial T} + \chi_{\bar{q}}(\beta\omega)\frac{\partial \bar{f}_0^q}{\partial T} + \bar{f}_0^q \frac{\partial \chi_{\bar{q}}(\beta\omega)}{\partial T}\right] \\
&\quad + \frac{1}{2\pi^2} g_g \int d\mathrm{p}\ \mathrm{p}^2\omega_g \left[\chi_g(\beta\omega)\frac{\partial f_0^g}{\partial T} + f_0^g \frac{\partial \chi_g(\beta\omega)}{\partial T}\right] \\
&\quad + \frac{1}{6\pi^2} g_g \int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_g}\left[\chi_g(\beta\omega)\frac{\partial f_0^g}{\partial T} + f_0^g \frac{\partial \chi_g(\beta\omega)}{\partial T}\right].
\end{aligned} \tag{55}$$

Substituting the values of partial derivatives in the above equation and simplifying, we

get

$$C_P = \frac{\beta^2}{6\pi^2}\sum_f g_f \int d\mathrm{p}\left(3\mathrm{p}^2\omega_f + \frac{\mathrm{p}^4}{\omega_f}\right)\Big[\left\{(\omega_f-\mu)\,f_0^q\,(1-f_0^q)+(\omega_f+\mu)\,\bar{f}_0^q\,\left(1-\bar{f}_0^q\right)\right\} \times\frac{\sinh(\beta\omega)}{\sinh(\beta\omega/2)} - \left(f_0^q+\bar{f}_0^q\right)\omega\sinh(\beta\omega/2)\Big] + \frac{\beta^2}{6\pi^2}g_g\int d\mathrm{p}\left(3\mathrm{p}^2\omega_g+\frac{\mathrm{p}^4}{\omega_g}\right)\left[\omega_g f_0^g\,(1+f_0^g)\,\frac{\sinh(3\beta\omega/2)}{\sinh(\beta\omega/2)} - 2f_0^g\omega\sinh(\beta\omega)\right] \quad (56)$$

From C_P (56) and C_V (53), the adiabatic index can be determined and studied to understand the effect of rotation on it, which is defined as

$$\gamma = \frac{C_P}{C_V}. \quad (57)$$

This ratio measures how much internal energy is getting converted into work in any compression/expansion process. Different systems possess different γ values, *e.g.* for a nonrelativistic ideal gas, $\gamma = \frac{5}{3}$ and for a system with massless particles, $\gamma = \frac{4}{3}$. The adiabatic index for an ideal gas is also associated with the degrees of freedom, so, for monatomic, diatomic and triatomic gases, the values of γ are different. As C_P and C_V for the system of quarks, antiquarks and gluons are influenced by the extreme conditions, their ratio is supposed to change in such conditions. Thus the adiabatic index depends on the features of the medium and its physical conditions. Since the rotation significantly affects the properties of the QGP medium, the value of the adiabatic index must be affected.

The trace anomaly is also one of the important thermodynamic observables, that characterizes the deviation of the system from the conformal limit, where the energy density (ε) equals 3 times the pressure ($3P$). In the ideal case or in the noninteracting case, *i.e.* when the quarks and gluons are idealized to be noninteracting, massless and there is no net baryon number in the system, the trace anomaly vanishes and the conformal symmetry is approached. On the other hand, if the trace anomaly is finite, it describes the existence of interaction between partons and the generation of mass. It has been observed that the trace anomaly could approach zero for large values of temperature in QCD due to the asymptotic freedom [75]. The study on the variation of the trace anomaly with the temperature is also very important in determining the phase transition temperature. In this work, we have explored how the trace anomaly gets altered by the rotation of the

QGP medium. The trace anomaly or interaction measure is defined as

$$\begin{aligned}\Gamma &= \varepsilon - 3P \\ &= u_\mu T^{\mu\nu} u_\nu + \Delta_{\mu\nu} T^{\mu\nu} \\ &= \frac{1}{2\pi^2} \sum_f g_f \int d\mathrm{p} \left(\mathrm{p}^2 \omega_f - \frac{\mathrm{p}^4}{\omega_f} \right) \left[f_0^q \chi_q(\beta\omega) + \bar{f}_0^q \chi_{\bar{q}}(\beta\omega) \right] \\ &\quad + \frac{1}{2\pi^2} g_g \int d\mathrm{p} \left(\mathrm{p}^2 \omega_g - \frac{\mathrm{p}^4}{\omega_g} \right) f_0^g \chi_g(\beta\omega). \end{aligned} \tag{58}$$

The thermal diffusion constant (D^T) is associated with the rate of heat transfer in a system. It appears in the diffusion equation for heat as

$$\frac{\partial T}{\partial t} = D^T \nabla^2 T, \tag{59}$$

which is carried out at constant pressure. The thermal diffusion constant in the first order relativistic viscous hydrodynamics can be defined in terms of the thermal conductivity and the specific heat at constant pressure as

$$D^T = \frac{\kappa}{C_P}. \tag{60}$$

Larger value of the thermal diffusion constant reflects the fact that the heat transfer is faster in the medium. Through κ (45) and C_P (56), it is possible to discern the influence of rotation on the thermal diffusion constant in QGP medium.

Another important observable is the isothermal compressibility (k^T), which can also be modulated by the rotation of the QGP medium. It acts as one of the thermodynamic response functions related to the equation of state, which governs the evolution of the system. It can also be useful in discerning the nature of the phase transition of the matter produced in heavy ion collisions. Depending on the center-of-mass energy, k^T may also change. Thermodynamically, the isothermal compressibility is defined as

$$k^T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T, \tag{61}$$

where V denotes the volume. Thus the isothermal compressibility measures how much the volume of the system decreases with the increase of pressure at a constant temperature. Generally, an increase in the applied pressure leads to a decrease in volume, so the negative sign appearing in eq. (61) makes the value of k^T positive. Smaller value of the isothermal compressibility indicates that the matter is stiffer. This observable can also provide the information about the deviation of a system from perfect fluid behavior. In heavy ion collisions, the isothermal compressibility can be calculated from the event-by-event fluctuation of charged particle multiplicity distributions [76]. For precise idea about the

equation of state of the system, k^T has utmost significance. For calculation, the isothermal compressibility can be transformed into a tractable form as

$$\begin{aligned} k^T &= \frac{1}{n}\left(\frac{\partial n}{\partial P}\right)_T \\ &= \frac{1}{n}\left(\frac{\partial n/\partial\mu}{\partial P/\partial\mu}\right)_T. \end{aligned} \tag{62}$$

Here, the expressions of n, $\partial n/\partial\mu$ and $\partial P/\partial\mu$ are respectively written as

$$n = \frac{1}{2\pi^2}\sum_f g_f \int d\mathrm{p}\ \mathrm{p}^2\left[f_0^q\chi_q(\beta\omega) + \bar{f}_0^q\chi_{\bar{q}}(\beta\omega)\right] + \frac{1}{2\pi^2}g_g\int d\mathrm{p}\ \mathrm{p}^2 f_0^g\chi_g(\beta\omega), \tag{63}$$

$$\frac{\partial n}{\partial\mu} = \frac{\beta}{2\pi^2}\sum_f g_f \int d\mathrm{p}\ \mathrm{p}^2\left[f_0^q\left(1-f_0^q\right)\chi_q(\beta\omega) - \bar{f}_0^q\left(1-\bar{f}_0^q\right)\chi_{\bar{q}}(\beta\omega)\right], \tag{64}$$

$$\frac{\partial P}{\partial\mu} = \frac{\beta}{6\pi^2}\sum_f g_f \int d\mathrm{p}\ \frac{\mathrm{p}^4}{\omega_f}\left[f_0^q\left(1-f_0^q\right)\chi_q(\beta\omega) - \bar{f}_0^q\left(1-\bar{f}_0^q\right)\chi_{\bar{q}}(\beta\omega)\right]. \tag{65}$$

The isothermal compressibility at finite rotation can be obtained by using the values of n, $\partial n/\partial\mu$ and $\partial P/\partial\mu$ in eq. (62) in the similar regime.

3 Results and discussions

In the present work, the quasiparticle masses or the thermal masses of particles within the quasiparticle model are considered. These masses are generated due to the interactions of partons with the surrounding thermal medium. At finite temperature and finite chemical potential, the quasiparticle masses (squared) of quark and gluon up to one-loop are given [77–79] by

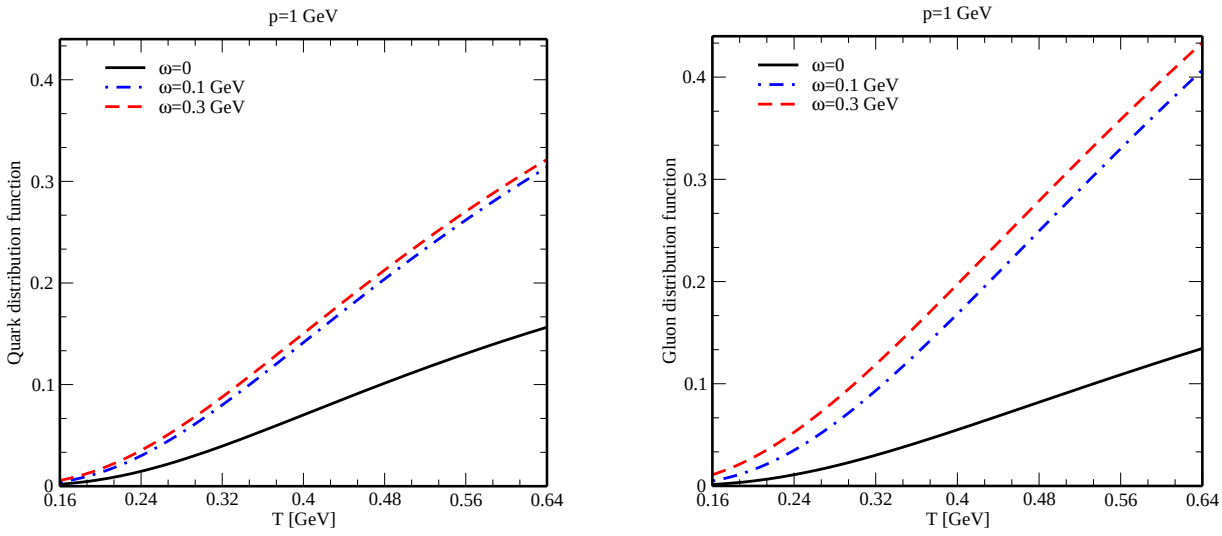
$$m_{fT}^2 = \frac{g^2T^2}{6}\left(1+\frac{\mu_f^2}{\pi^2T^2}\right), \tag{66}$$

$$m_{gT}^2 = \frac{g^2T^2}{6}\left(N_c+\frac{N_f}{2}+\frac{3}{2\pi^2T^2}\sum_f\mu_f^2\right), \tag{67}$$

respectively. Here, g is the one-loop running coupling with the following [80] form,

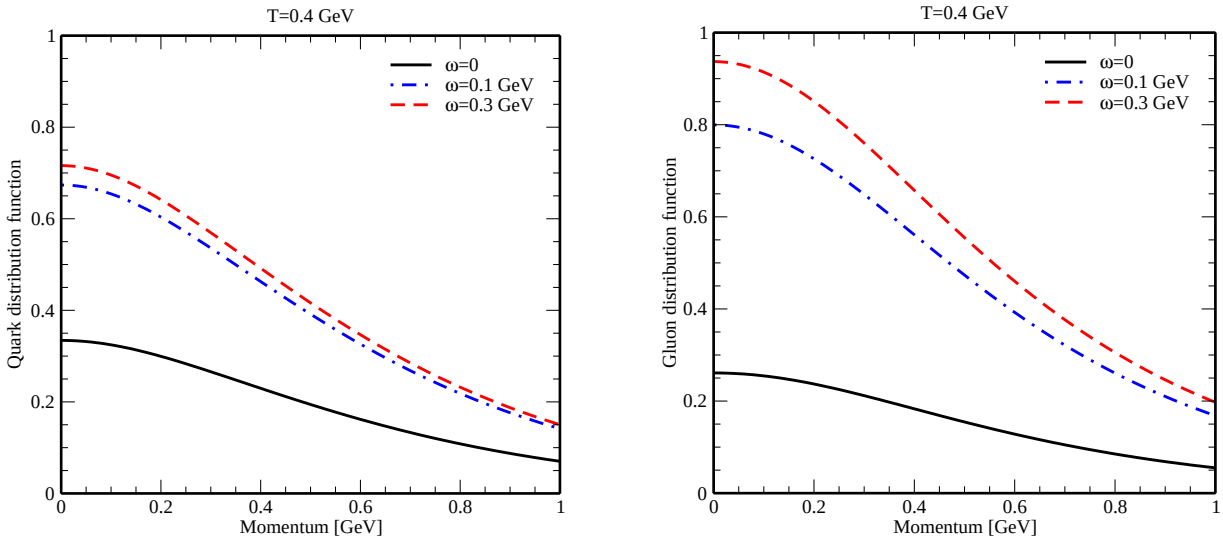
$$g^2 = 4\pi\alpha_s = \frac{48\pi^2}{(11N_c-2N_f)\ln\left(\Lambda^2/\Lambda^2_{\overline{\mathrm{MS}}}\right)}, \tag{68}$$

where $\Lambda_{\overline{\mathrm{MS}}} = 0.176$ GeV, $\Lambda = 2\pi\sqrt{T^2+\mu_f^2/\pi^2}$ for electrically charged particles (quarks and antiquarks) and $\Lambda = 2\pi T$ for gluons. Further, we note that, in the calculation, the chemical potential is set at 0.06 GeV.



a b

Figure 1: Variations of (a) the quark distribution function and (b) the gluon distribution function with temperature for different values of angular velocity.



a b

Figure 2: Variations of (a) the quark distribution function and (b) the gluon distribution function with momentum for different values of angular velocity.

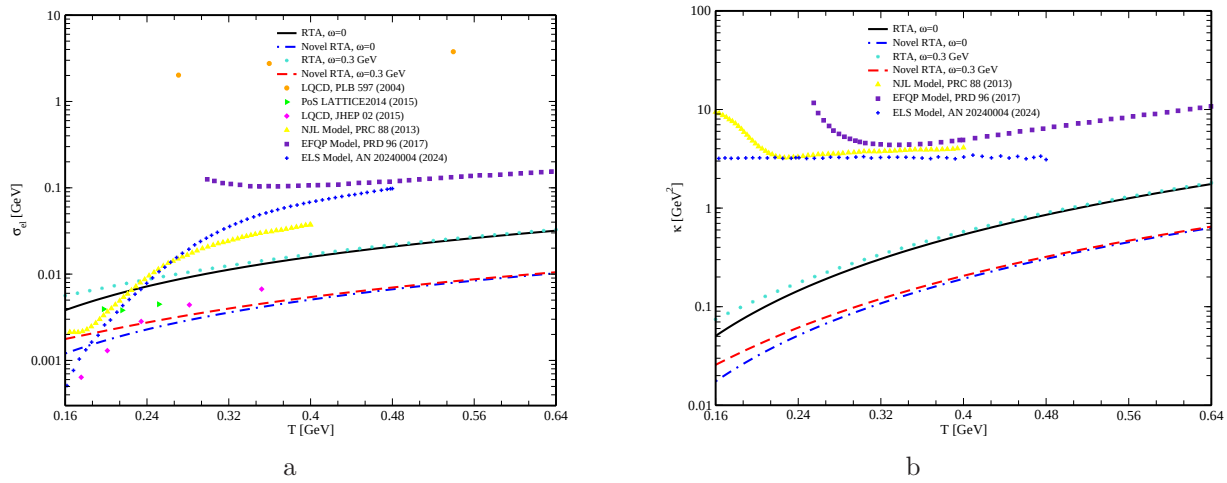


Figure 3: Variations of (a) the electrical conductivity and (b) the thermal conductivity with temperature for different values of angular velocity.

To understand the effect of rotational QGP medium on the transport coefficients and observables, let us first observe how the particle distribution functions get altered in the similar environment. Figure 1 shows the u quark and gluon distribution functions as functions of temperature at a fixed momentum for different values of the angular velocity, whereas figure 2 displays the same as functions of momentum at a fixed temperature. It is observed that the distribution functions increase with the increase of angular velocity. The effect of rotation on parton distribution functions is more evident at high temperatures and low momenta. It can be further noticed from figures 1 and 2 that, as compared to the quark distribution function, the gluon distribution function is more influenced by the rotation of the QGP medium. This can be perceived from the fact that the number of gluons in adjoint representation is larger than the number of quarks in fundamental representation at high temperatures, so the overall response of gluons to rotation exceeds that of quarks. These observations on the parton distribution functions would help greatly in exploring the effect of rotation on various transport coefficients and observables in kinetic theory approach.

Figure 3 depicts the variations of electrical ($\sigma_{\rm el}$) and thermal (κ) conductivities of the QGP medium with temperature for different values of angular velocity in the relaxation time approximation as well as in the novel relaxation time approximation. Observation shows that the emergence of angular velocity gives an increasing effect to both $\sigma_{\rm el}$ and κ, and their deviations from $\omega = 0$ case are less prominent at high temperatures. In the novel relaxation time approximation, a decrease in $\sigma_{\rm el}$ and κ is observed as compared to the frequently used relaxation time approximation. Difference between the results in these two approximations is mainly attributed to the different relaxation times and collision integrals in both cases. From figures 3a and 3b, it is seen that the distinction

between the RTA and the novel RTA is more conspicuous at higher temperatures both for the nonrotating and rotating QGP mediums. Thus the rotation facilitates the transport of charge and heat in the medium. Further, we have compared our results with the results obtained from other models, such as the lattice QCD [81–83], the Nambu-Jona-Lasinio (NJL) model [84], the effective fugacity quasiparticle (EFQP) model [85] and the extended linear sigma (ELS) model [86]. We note that the aforesaid models including the lattice simulation with the rotation are at an early stage of development, and the results on electrical and thermal conductivities at finite rotation are not yet available in these models, so we have used the results in the absence of rotation from these models. The electrical conductivity in the hot phase of the QCD plasma is extracted from a quenched lattice measurement of the Euclidean time vector correlator for $1.5 \leq T/T_c \leq 3$ in ref. [81]. This study estimates the value of the electrical conductivity that lies above our results at zero angular velocity as well as at finite angular velocity. In ref. [82], the electrical conductivity has been calculated using the vector correlation function within the quenched lattice QCD and this result is found to be in close proximity to our results throughout the considered temperature range. Lattice result in ref. [83] lies slightly below our result calculated using RTA method, whereas it is closer to our result determined using the novel RTA method. Upon comparing our results with the NJL model [84], EFQP model [85] and ELS model [86] calculations, it is found that the results of the aforesaid models lie above our results on electrical conductivity in the absence as well as in the presence of finite angular velocity. However, as compared to the EFQP model, the NJL model and the ELS model estimations on electrical conductivity are in close agreement with our results at low temperatures. We note that the lattice QCD calculations on the thermal conductivity are not available in literature, so we have only compared our results on the thermal conductivity with that in the NJL model [84], EFQP model [85] and ELS model [86]. We have found that the results from these models lie above our result over the entire range of temperature for both nonrotating and rotating medium cases. However at high temperatures, our result on the thermal conductivity is closer to the estimations of these models.

In figure 4, the effects of rotation on the Knudsen number (Ω) and the elliptic flow (v_2) have been displayed. Observation shows a decrease in the Knudsen number (figure 4a) and an increase in the elliptic flow (figure 4b) when rotation is introduced in the medium. This behavior of the Knudsen number at finite angular velocity is corroborated by the observations on the thermal conductivity and the specific heat at constant volume in the similar regime. The increase of v_2 at finite rotation also agrees with the observation in ref. [6]. Thus the medium approaches towards the equilibrium with the increase of rotation. Further, the novel relaxation time approximation estimates smaller value of Ω and larger value of v_2 as compared to their corresponding values in the relaxation time

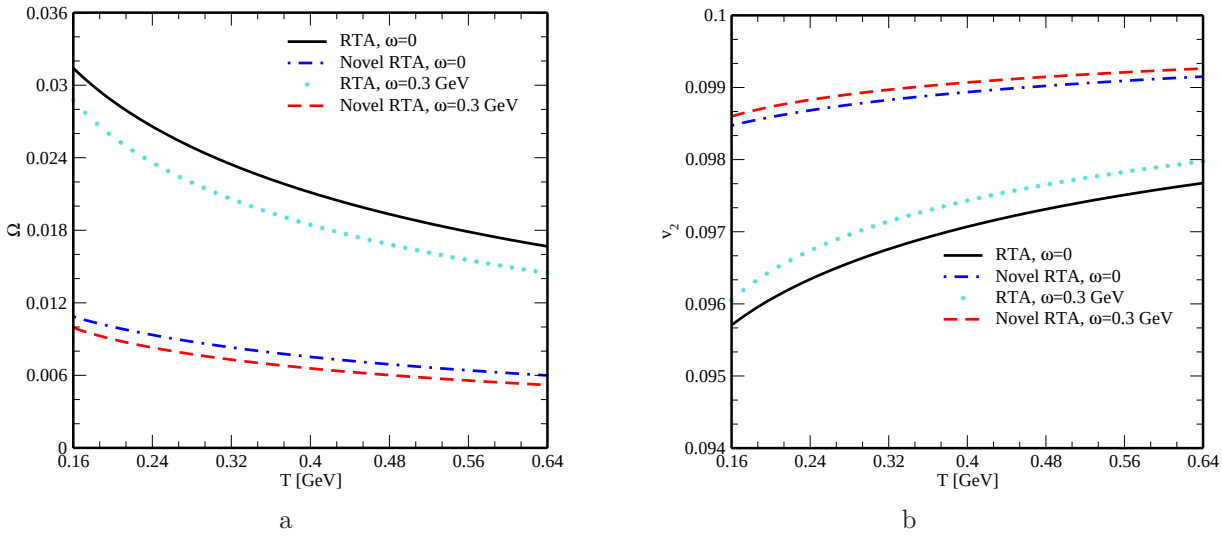
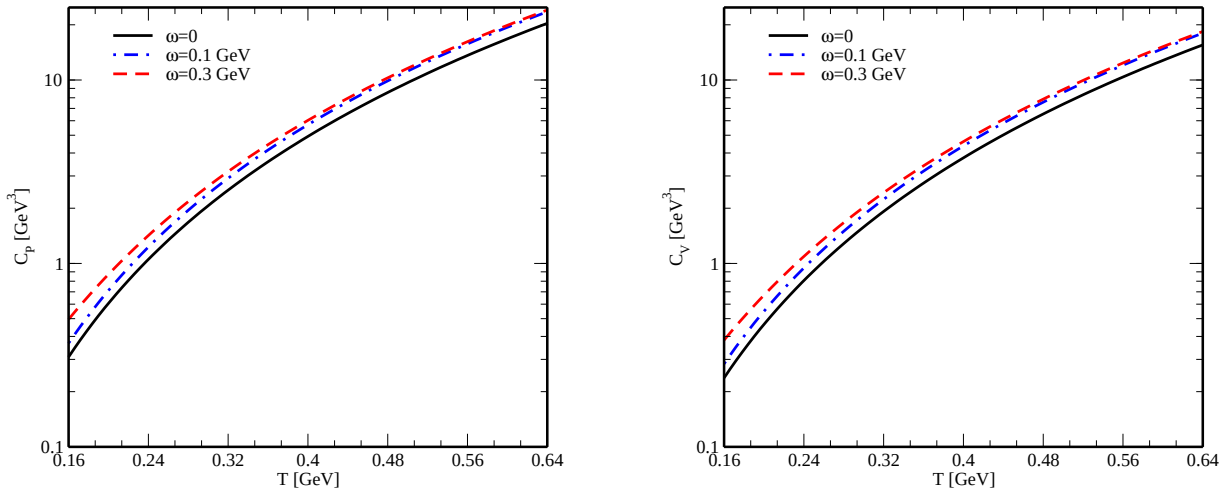


Figure 4: Variations of (a) the Knudsen number and (b) the elliptic flow with temperature for different values of angular velocity.

approximation. This indicates an enhancement of the interactions among the produced particles in heavy ion collisions and faster thermalization of the medium.

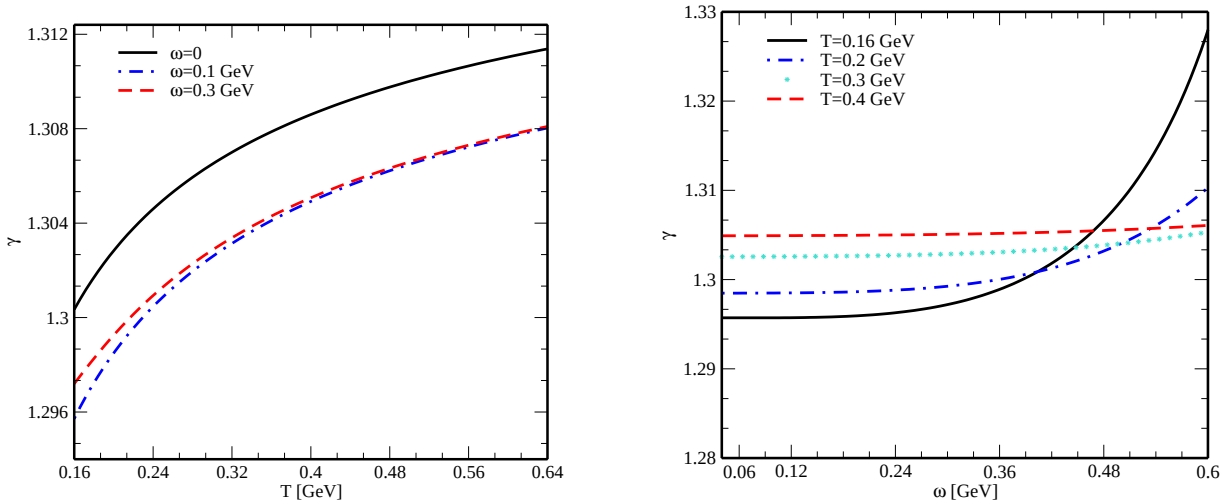
Figure 5 shows the effects of rotation on the specific heat at constant pressure (C_P) and the specific heat at constant volume (C_V) over a given range of the temperature. It is observed that the emergence of rotation enhances both C_P and C_V, with the magnitude of the former one being larger than that of the latter one for all temperatures. Thus, for a rotating medium, the changes in enthalpy and energy density with temperature are larger as compared to the nonrotating case. The relative behavior between C_P and C_V can be discerned through their ratio γ or the adiabatic index. Figure 6 displays how the adiabatic index varies with temperature and with angular velocity. It is observed that the adiabatic index becomes larger as the medium gets hotter, whereas, if the rotation is introduced in the medium, γ gets decreased as compared to the $\omega = 0$ case (figure 6a). On the other hand, within the rotating medium, γ is found to increase with the increase of angular velocity. Further, the trend of variation of γ with the angular velocity is not same for all temperatures, rather the increase of γ with angular velocity becomes faster for lower temperatures (figure 6b). In low angular velocity regime, the variation of γ is meagre, however, a noticeable increasing behavior of γ is observed in high angular velocity regime. This indicates that the change in enthalpy is larger than the change in energy density for an increase of one unit in temperature and this difference gets enhanced as the rotation of the QGP medium becomes more rapid.

In figure 7, the variations of the trace anomaly in units of its value in the absence of rotation with temperature and with angular velocity have been depicted. It is observed



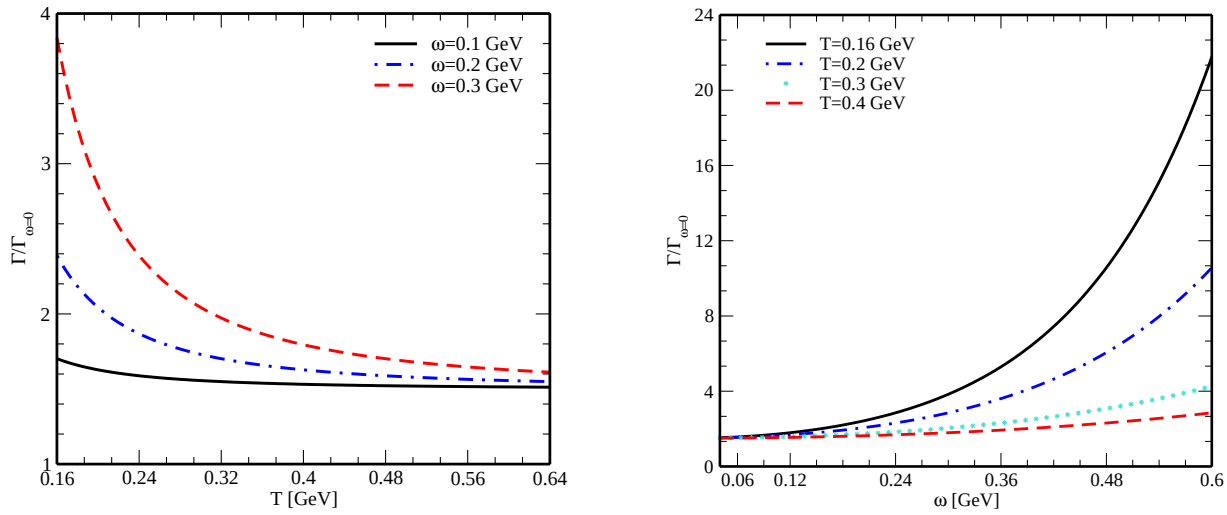
a b

Figure 5: Variations of (a) the specific heat at constant pressure and (b) the specific heat at constant volume with temperature for different values of angular velocity.



a b

Figure 6: Variations of the ratio γ (C_P/C_V) (a) with temperature for different values of angular velocity and (b) with angular velocity for different values of temperature.



a b

Figure 7: Variations of the trace anomaly scaled with its value in the absence of rotation (a) as a function of temperature for different values of angular velocity and (b) as a function of angular velocity for different values of temperature.

that, with the increase of temperature, the deviation of the trace anomaly at finite angular velocity from that at zero angular velocity decreases (figure 7a). Thus, at high temperatures, rotation of the medium has a meagre impact on the trace anomaly. Our observation has further found an increasing trend of the trace anomaly as the angular velocity increases and the slope decreases if the medium becomes hotter (figure 7b). The effect of angular velocity on the trace anomaly seems to be more pronounced at lower temperatures. Thus the rotation makes the constituents of the medium more interactive and as a result, it takes the medium a bit away from the conformal symmetry.

Figure 8 displays how the thermal diffusion constant (D^T) and the isothermal compressibility (κ^T) get modulated by the rotation with varying temperature. It is observed that the rotation of the medium reduces the thermal diffusion constant over the entire range of temperature (figure 8a). This reduction in D^T is mainly attributed to the variations of κ and C_P in the similar regime. With the angular velocity, the increase in the specific heat at constant pressure dominates over the increase in the thermal conductivity, therefore, a decrease in the thermal diffusion constant at finite angular velocity is observed. Smaller value of the thermal diffusion constant indicates slower heat transfer in a rotating medium as compared to that in a nonrotating medium. With the increase of temperature, isothermal compressibility is found to decrease (figure 8b). Further, the introduction of angular velocity also reduces the isothermal compressibility and the matter may show nearly perfect fluid behavior. This suggests that with the increasing angular velocity, it is difficult to compress the system. Thus, it is inferred that the matter becomes

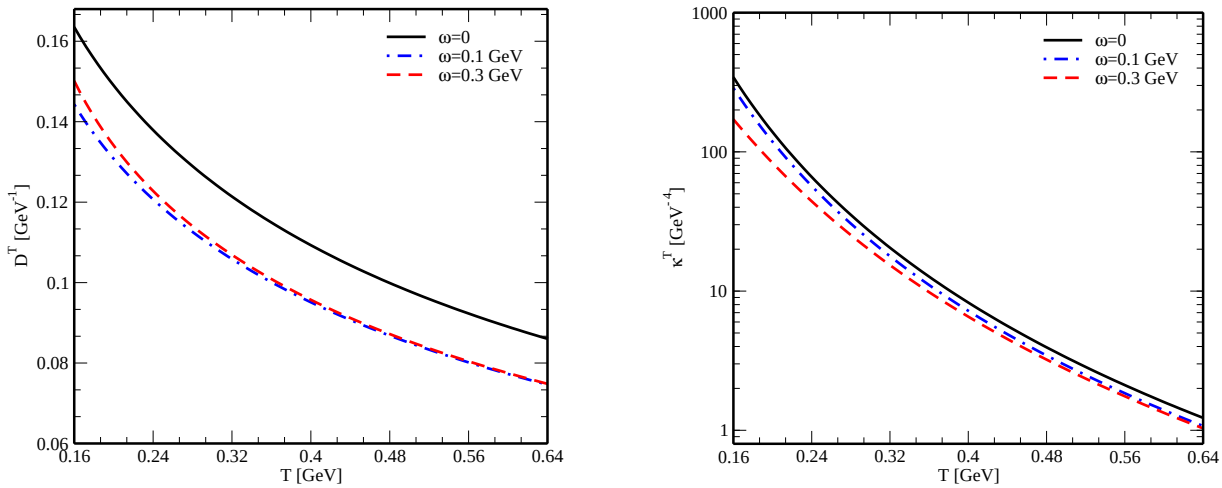


Figure 8: Variations of (a) the thermal diffusion constant and (b) the isothermal compressibility with temperature for different values of angular velocity.

stiffer as the rotation becomes more rapid.

4 Conclusions

In this work, we explored the transport coefficients and observables of a rotating QGP medium. We calculated the electrical and thermal conductivities using the relativistic Boltzmann transport equation in the novel relaxation time approximation within the kinetic theory approach. We also compared our results with the frequently used relaxation time approximation. We observed that the introduction of rotation enhances the conduction of charge and heat in the medium. In addition, the novel relaxation time approximation estimates smaller values of the electrical and thermal conductivities as compared to their corresponding values in the relaxation time approximation. We further explored the effects of rotation on various observables, such as the Knudsen number, elliptic flow, specific heat at constant pressure, specific heat at constant volume, trace anomaly, thermal diffusion constant and isothermal compressibility. From the behaviors of these observables, we found that the rotation leaves a significant influence on the equilibrium property of the medium, the interactions among particles produced in heavy ion collisions, the adiabatic index, the conformal symmetry, the rate of heat transfer in the medium and the stiffness of the medium.

Acknowledgments

One of us (S. R.) would like to thank the Indian Institute of Technology Bombay for the Institute postdoctoral fellowship and S. D. acknowledges the SERB Power Fellowship, SPF/2022/000014 for the support on this work.

References

[1] W. T. Deng and X. G. Huang, *Vorticity in heavy-ion collisions*, Phys. Rev. C **93**, 064907 (2016).

[2] Y. Jiang, Z. W. Lin and J. Liao, *Rotating quark-gluon plasma in relativistic heavy-ion collisions*, Phys. Rev. C **94**, 044910 (2016).

[3] X. Wang, M. Wei, Z. Li and M. Huang, *Quark matter under rotation in the NJL model with vector interaction*, Phys. Rev. D **99**, 016018 (2019).

[4] M. Wei, C. A. Islam and M. Huang, *Production rate and ellipticity of lepton pairs from a rotating hot and dense QCD medium*, Phys. Rev. D **105**, 054014 (2022).

[5] Z. T. Liang and X. N. Wang, *Spin alignment of vector mesons in non-central $A + A$ collisions*, Phys. Lett. B **629**, 20 (2005).

[6] F. Becattini, F. Piccinini and J. Rizzo, *Angular momentum conservation in heavy ion collisions at very high energy*, Phys. Rev. C **77**, 024906 (2008).

[7] E. Berti, F. White, A. Maniopoulou and M. Bruni, *Rotating neutron stars: an invariant comparison of approximate and numerical space-time models*, Mon. Not. R. Astron. Soc. **358**, 923 (2005).

[8] A. L. Fetter, *Rotating trapped Bose-Einstein condensates*, Rev. Mod. Phys. **81**, 647 (2009).

[9] D. E. Kharzeev and D. T. Son, *Testing the Chiral Magnetic and Chiral Vortical Effects in Heavy Ion Collisions*, Phys. Rev. Lett. **106**, 062301 (2011).

[10] D. T. Son and P. Surowka, *Hydrodynamics with Triangle Anomalies*, Phys. Rev. Lett. **103**, 191601 (2009).

[11] Y. Jiang, X. G. Huang and J. Liao, *Chiral vortical wave and induced flavor charge transport in a rotating quark-gluon plasma*, Phys. Rev. D **92**, 071501 (2015).

[12] K. Fukushima, D. E. Kharzeev and H. J. Warringa, *Chiral magnetic effect*, Phys. Rev. D **78**, 074033 (2008).

[13] D. E. Kharzeev, L. D. McLerran and H. J. Warringa, *The effects of topological charge change in heavy ion collisions: "Event by event P and CP violation"*, Nucl. Phys. A **803**, 227 (2008).

[14] D. E. Kharzeev and H. U. Yee, *Chiral magnetic wave*, Phys. Rev. D **83**, 085007 (2011).

[15] Y. Burnier, D. E. Kharzeev, J. Liao and H. U. Yee, *Chiral Magnetic Wave at Finite Baryon Density and the Electric Quadrupole Moment of the Quark-Gluon Plasma*, Phys. Rev. Lett. **107**, 052303 (2011).

[16] A. Ipp, A. Di Piazza, J. Evers and C. H. Keitel, *Photon polarization as a probe for quark-gluon plasma dynamics*, Phys. Lett. B **666**, 315 (2008).

[17] B. Betz, M. Gyulassy and G. Torrieri, *Polarization probes of vorticity in heavy ion collisions*, Phys. Rev. C **76**, 044901 (2007).

[18] F. Becattini and F. Piccinini, *The ideal relativistic spinning gas: Polarization and spectra*, Ann. Phys. (Amsterdam) **323**, 2452 (2008).

[19] F. Becattini, V. Chandra, L. D. Zanna and E. Grossi, *Relativistic distribution function for particles with spin at local thermodynamical equilibrium*, Ann. Phys. **338**, 32 (2013).

[20] W. Florkowski, A. Kumar and R. Ryblewski, *Relativistic hydrodynamics for spin-polarized fluids*, Prog. Part. Nucl. Phys. **108**, 103709 (2019).

[21] S. Bhadury, W. Florkowski, A. Jaiswal, A. Kumar and R. Ryblewski, *Relativistic Spin Magnetohydrodynamics*, Phys. Rev. Lett. **129**, 192301 (2022).

[22] V. E. Ambruş, R. Ryblewski and R. Singh, *Spin waves in spin hydrodynamics*, Phys. Rev. D **106**, 014018 (2022).

[23] F. Becattini and M. A. Lisa, *Polarization and Vorticity in the Quark-Gluon Plasma*, Ann. Rev. Nucl. Part. Sci. **70**, 395 (2020).

[24] M. N. Chernodub and S. Gongyo, *Interacting fermions in rotation: chiral symmetry restoration, moment of inertia and thermodynamics*, J. High Energy Phys. **1701**, 136 (2017).

[25] S. Ebihara, K. Fukushima and K. Mameda, *Boundary effects and gapped dispersion in rotating fermionic matter*, Phys. Lett. B **764**, 94 (2017).

[26] Y. Fujimoto, K. Fukushima and Y. Hidaka, *Deconfining phase boundary of rapidly rotating hot and dense matter and analysis of moment of inertia*, Phys. Lett. B **816**, 136184 (2021).

[27] X. Chen, L. Zhang, D. Li, D. Hou and M. Huang, *Gluodynamics and deconfinement phase transition under rotation from holography*, J. High Energy Phys. **2107**, 132 (2021).

[28] V. V. Braguta, A. Y. Kotov, D. D. Kuznedelev and A. A. Roenko, *Study of the Confinement/Deconfinement Phase Transition in Rotating Lattice SU(3) Gluodynamics*, JETP Lett. **112**, 9 (2020).

[29] V. V. Braguta, A. Y. Kotov, D. D. Kuznedelev and A. A. Roenko, *Influence of relativistic rotation on the confinement-deconfinement transition in gluodynamics*, Phys. Rev. D **103**, 094515 (2021).

[30] Y. Jiang and J. Liao, *Pairing Phase Transitions of Matter under Rotation*, Phys. Rev. Lett. **117**, 192302 (2016).

[31] B. R. Iyer, *Dirac field theory in rotating coordinates*, Phys. Rev. D **26**, 1900 (1982).

[32] F. Becattini and L. Tinti, *Thermodynamical inequivalence of quantum stress-energy and spin tensors*, Phys. Rev. D **84**, 025013 (2011).

[33] V. E. Ambruş and E. Winstanley, *Rotating quantum states*, Phys. Lett. B **734**, 296 (2014).

[34] V. E. Ambruş and E. Winstanley, *Rotating fermions inside a cylindrical boundary*, Phys. Rev. D **93**, 104014 (2016).

[35] H. -L. Chen, K. Fukushima, X. -G. Huang and K. Mameda, *Analogy between rotation and density for Dirac fermions in a magnetic field*, Phys. Rev. D **93**, 104052 (2016).

[36] B. McInnes, *Angular momentum in QGP holography*, Nucl. Phys. B **887**, 246 (2014).

[37] B. McInnes, *A rotation/magnetism analogy for the quark-gluon plasma*, Nucl. Phys. B **911**, 173 (2016).

[38] B. Singh, J. R. Bhatt and H. Mishra, *Probing vorticity in heavy ion collision with dilepton production*, Phys. Rev. D **100**, 014016 (2019).

[39] S. Rath and B. K. Patra, *Revisit to electrical and thermal conductivities, Lorenz and Knudsen numbers in thermal QCD in a strong magnetic field*, Phys. Rev. D **100**, 016009 (2019).

[40] S. Rath and B. K. Patra, *One-loop QCD thermodynamics in a strong homogeneous and static magnetic field*, J. High Energy Phys. **1712**, 098 (2017).

[41] A. Bandyopadhyay, B. Karmakar, N. Haque and M. G. Mustafa, *Pressure of a weakly magnetized hot and dense deconfined QCD matter in one-loop hard-thermal-loop perturbation theory*, Phys. Rev. D **100**, 034031 (2019).

[42] S. Rath and B. K. Patra, *Thermomagnetic properties and Bjorken expansion of hot QCD matter in a strong magnetic field*, Eur. Phys. J. A **55**, 220 (2019).

[43] B. Karmakar, R. Ghosh, A. Bandyopadhyay, N. Haque and M. G. Mustafa, *Anisotropic pressure of deconfined QCD matter in presence of strong magnetic field within one-loop approximation*, Phys. Rev. D **99**, 094002 (2019).

[44] B. Feng, *Electric conductivity and Hall conductivity of the QGP in a magnetic field*, Phys. Rev. D **96**, 036009 (2017).

[45] S. Rath and S. Dash, *Effects of weak magnetic field and finite chemical potential on the transport of charge and heat in hot QCD matter*, Eur. Phys. J. A **59**, 25 (2023).

[46] K. Hattori and D. Satow, *Electrical conductivity of quark-gluon plasma in strong magnetic fields*, Phys. Rev. D **94**, 114032 (2016).

[47] L. Thakur and P. K. Srivastava, *Electrical conductivity of a hot and dense QGP medium in a magnetic field*, Phys. Rev. D **100**, 076016 (2019).

[48] K. Fukushima and Y. Hidaka, *Electric Conductivity of Hot and Dense Quark Matter in a Magnetic Field with Landau Level Resummation via Kinetic Equations*, Phys. Rev. Lett. **120**, 162301 (2018).

[49] S. Rath and B. K. Patra, *Effect of magnetic field on the charge and thermal transport properties of hot and dense QCD matter*, Eur. Phys. J. C **80**, 747 (2020).

[50] M. Kurian, S. Mitra, S. Ghosh and V. Chandra, *Transport coefficients of hot magnetized QCD matter beyond the lowest Landau level approximation*, Eur. Phys. J. C **79**, 134 (2019).

[51] G. S. Denicol *et al.*, *Nonresistive dissipative magnetohydrodynamics from the Boltzmann equation in the 14-moment approximation*, Phys. Rev. D **98**, 076009 (2018).

[52] A. Das, H. Mishra and R. K. Mohapatra, *Transport coefficients of hot and dense hadron gas in a magnetic field: A relaxation time approach*, Phys. Rev. D **100**, 114004 (2019).

[53] Seung-i. Nam and Chung-W. Kao, *Shear viscosity of quark matter at finite temperature under an external magnetic field*, Phys. Rev. D **87**, 114003 (2013).

[54] S. Rath and S. Dash, *Momentum transport properties of a hot and dense QCD matter in a weak magnetic field*, Eur. Phys. J. C **82**, 797 (2022).

[55] S. Rath and B. K. Patra, *Momentum and its affiliated transport coefficients for hot QCD matter in a strong magnetic field*, Phys. Rev. D **102**, 036011 (2020).

[56] S. Rath and B. K. Patra, *Viscous properties of hot and dense QCD matter in the presence of a magnetic field*, Eur. Phys. J. C **81**, 139 (2021).

[57] K. Hattori, X.-G. Huang, D. H. Rischke and D. Satow, *Bulk viscosity of quark-gluon plasma in strong magnetic fields*, Phys. Rev. D **96**, 094009 (2017).

[58] S. Li and Ho-U. Yee, *Shear viscosity of the quark-gluon plasma in a weak magnetic field in perturbative QCD: Leading log*, Phys. Rev. D **97**, 056024 (2018).

[59] H. van Hees, C. Gale and R. Rapp, *Thermal photons and collective flow at energies available at the BNL Relativistic Heavy-Ion Collider*, Phys. Rev. C **84**, 054906 (2011).

[60] C. Shen, U. W. Heinz, J.-F. Paquet and C. Gale, *Thermal photons as a quark-gluon plasma thermometer reexamined*, Phys. Rev. C **89**, 044910 (2014).

[61] K. Tuchin, *Magnetic contribution to dilepton production in heavy-ion collisions*, Phys. Rev. C **88**, 024910 (2013).

[62] K. A. Mamo, *Enhanced thermal photon and dilepton production in strongly coupled $N = 4$ SYM plasma in strong magnetic field*, J. High Energy Phys. **1308**, 083 (2013).

[63] K. Fukushima, K. Hattori, H.-U. Yee and Y. Yin, *Heavy quark diffusion in strong magnetic fields at weak coupling and implications for elliptic flow*, Phys. Rev. D **93**, 074028 (2016).

[64] C. Cercignani and G. M. Kremer. "The Relativistic Boltzmann Equation: Theory and Applications". Birkhäuser Basel, 2002.

[65] G. S. Rocha, G. S. Denicol and J. Noronha, *Novel Relaxation Time Approximation to the Relativistic Boltzmann Equation*, Phys. Rev. Lett. **127**, 042301 (2021).

[66] F. Becattini and L. Ferroni, *The microcanonical ensemble of the ideal relativistic quantum gas with angular momentum conservation*, Eur. Phys. J. C **52**, 597 (2007).

[67] A. Hosoya and K. Kajantie, *Transport coefficients of QCD matter*, Nucl. Phys. B **250**, 666 (1985).

[68] K. Dusling, G. D. Moore and D. Teaney, *Radiative energy loss and v_2 spectra for viscous hydrodynamics*, Phys. Rev. C **81**, 034907 (2010).

[69] E. Calzetta and B. L. Hu, *Nonequilibrium quantum fields: Closed-time-path effective action, Wigner function, and Boltzmann equation*, Phys. Rev. D **37**, 2878 (1988).

[70] R. S. Bhalerao, J.-P. Blaizot, N. Borghini and J.-Y. Ollitrault, *Elliptic flow and incomplete equilibration at RHIC*, Phys. Lett. B **627**, 49 (2005).

[71] H.-J. Drescher, A. Dumitru, C. Gombeaud and J.-Y. Ollitrault, *Centrality dependence of elliptic flow, the hydrodynamic limit, and the viscosity of hot QCD*, Phys. Rev. C **76**, 024905 (2007).

[72] C. Gombeaud and J.-Y. Ollitrault, *Covariant transport theory approach to elliptic flow in relativistic heavy ion collision*, Phys. Rev. C **77**, 054904 (2008).

[73] S. Rath and S. Dash, *Impact of nonextensivity on the transport coefficients of a magnetized hot and dense QCD matter*, Eur. Phys. J. C **83**, 867 (2023).

[74] R. K. Mohapatra, P. S. Saumia and A. M. Srivastava, *Enhancement of flow anisotropies due to magnetic field in relativistic heavy-ion collisions*, Mod. Phys. Lett. A **26**, 2477 (2011).

[75] J. O. Andersen, N. Haque, M. G. Mustafa and M. Strickland, *Three-loop hard-thermal-loop perturbation theory thermodynamics at finite temperature and finite baryonic and isospin chemical potential*, Phys. Rev. D **93**, 054045 (2016).

[76] S. Mrówczyński, *Hadronic matter compressibility from event-by-event analysis of heavy-ion collisions*, Phys. Lett. B **430**, 9 (1998).

[77] E. Braaten and R. D. Pisarski, *Simple effective Lagrangian for hard thermal loops*, Phys. Rev. D **45**, R1827 (1992).

[78] A. Peshier, B. Kämpfer and G. Soff, *From QCD lattice calculations to the equation of state of quark matter*, Phys. Rev. D **66**, 094003 (2002).

[79] J.-P. Blaizot, A. Ipp, A. Rebhan and U. Reinosa, *Asymptotic thermal quark masses and the entropy of QCD in the large - N_f limit*, Phys. Rev. D **72**, 125005 (2005).

[80] J. I. Kapusta and C. Gale. Finite Temperature Field Theory Principles and Applications (Cambridge University Press, Cambridge, 2006).

[81] S. Gupta, *The electrical conductivity and soft photon emissivity of the QCD plasma*, Phys. Lett. B **597**, 57 (2004).

[82] H.-T. Ding, O. Kaczmarek and F. Meyer, *Vector spectral functions and transport properties in quenched QCD*, PoS **LATTICE2014**, 216 (2015).

[83] G. Aarts, C. Allton, A. Amato, P. Giudice, S. Hands and J.-I. Skullerud, *Electrical conductivity and charge diffusion in thermal QCD from the lattice*, J. High Energy Phys. **02**, 186 (2015).

[84] R. Marty, E. Bratkovskaya, W. Cassing, J. Aichelin and H. Berrehrah, *Transport coefficients from the Nambu-Jona-Lasinio model for $SU(3)_f$*, Phys. Rev. C **88**, 045204 (2013).

[85] S. Mitra and V. Chandra, *Transport coefficients of a hot QCD medium and their relative significance in heavy-ion collisions*, Phys. Rev. D **96**, 094003 (2017).

[86] A. N. Tawfik and A. Elshehri, *Extended linear-sigma model: transport properties of QCD matter at finite temperature*, Astron. Nachr. **20240004** (2024).

arXiv:2403.01224v1 [hep-ph] 2 Mar 2024

Study of identified particle production as a function of transverse event activity classifier, S_T in p−p collisions

Rahul Verma,* Vishu Saini,† Basanta Nandi,‡ and Sadhana Dash§
[1] *Department of Physics*
Indian Institute of Technology Bombay,
Mumbai 400076, India

A new observable, S_T, is introduced in terms of the sum of the transverse momentum of charged particles ($\sum_i p_{T_i}$) produced in proton proton (p−p) collisions at LHC energies to probe the underlying events (UE). The UE are defined as those aspects of proton-proton collisions that are not attributed to the primary hard scattering process, but rather to the accompanying interactions of the rest of the proton. The conventional approach of studying underlying events is usually carried out by defining topological regions with respect to the leading particle in an event. The transverse region is generally sensitive to UE and various classifiers have been used to discriminate the extent of UE activity regions. The production of identified particles like $\pi^{\pm}$, $K^{\pm}$, p , K^0_S, and Λ^0 are studied in different ranges of transverse activity classifier in p−p collisions at $\sqrt{s} = 13$ TeV using pQCD inspired PYTHIA 8 event generator. A comparative analysis of the identified particle spectra, mean multiplicity and mean transverse momentum has been carried out with respect to S_T and the performance of this new observable is gauged by comparing the results with previously defined R_T observable.

I. INTRODUCTION

Recently, several effects typical of heavy-ion phenomenology have been observed in high-multiplicity proton-proton (p−p) collisions and p−Pb collisions. The strangeness enhancement [1], collectivity [2, 3] and anisotropic flows [4] etc. which were well accepted signatures of creation of a hot and dense state composed of deconfined partons, called the Quark Gluon Plasma, have been reported in heavy-ion collisions. However, these signatures have also been reported in small systems like p−p and p−Pb [5–8]. This has intrigued the heavy-ion community to understand the underlying dynamics in small collision systems. In order to understand the origin of such effects in small collision systems where the effects of hot and dense medium is not expected to manifest, it becomes imperative to study and comprehend the production of identified particles as a function of the underlying event activity [9, 10] observables. The "underlying event" can be defined as those aspects of proton-proton collision that are not attributed to the hard scattering process, but rather to the accompanying interactions of the rest of the proton [9, 11]. Underlying event is an essential element of the hadron-hadron interaction environment. Its contributions to the final state of a given hadron-hadron collision are vast and thus to perform precise standard model measurements (and also search for physics beyond the standard model at hadron colliders like LHC), it is necessary to have a good understanding of these events. However, such events have contributions from both hard (perturbative QCD) and soft (non-perturbative QCD) processes. Therefore, one resorts to various phenomenological models (usually implemented in Monte Carlo event generators), to constrain parameters tuned to relevant experimental data. Various UE-sensitive variables have been proposed [9, 10, 12] to disentangle hard process-dominated events from soft ones. Moreover, UE activity of an event is also found to be related to geometry of energy momentum flow in the event [9, 11]. The conventional way to disentangle particle production associated with soft and hard processes is usually carried out by defining three kine-

* rahul.verma@iitb.ac.in
† vishusaini220301@gmail.com
‡ basanta@iitb.ac.in
§ sadhana@phy.iitb.ac.in

matic regions depending on the angular distance with respect to the leading charged particle. These regions have different sensitivities to UE activity. Various event shape variables like transverse sphericity , thrust, sphericity etc. also provide an alternative way to classify events as jetty (dominated by hard processes) and isotropic (dominated by soft processes) events. A transverse activity classifier, R_T [10] was also proposed to quantify UE activity of an event.

In this work, a new transverse activity classifier, S_T is introduced and the production of identified particles like $\pi^{\pm}$, $K^{\pm}$, p, K^0_S, and Λ^0 as a function of the proposed UE activity classifier has been investigated in p−p collisions at $\sqrt{s} = 13$ TeV. One of the primary aims of the present study is to explore the p_T spectra, multiplicity distribution and strangeness and baryon production in regions of varying UE activity quantified by S_T. The study has been performed by dividing the kinematic region into **toward**, **away**, and **transverse** regions where the **transverse** region is most sensitive to UE activity. The particle production has been studied as a function of transverse activity classifiers like R_T and S_T in order to compare their performances.

II. UNDERLYING EVENT OBSERVABLE

It is nearly impossible to uniquely separate the underlying event activity from the hard scattering process on an event-by-event basis in experimental studies. However, UE activity is intimately related to the topology of an event. For example, a typical hard scattering event is characterized by a burst, of hadrons travelling approximately in the same direction as that of initial beam particles or as a collection of hadrons (called *jets*) with large momentum in transverse direction, that are approximately back to back in azimuthal angle, ϕ. As two high p_T jets are back to back in azimuthal angle, ϕ, one can use the topology of collision event to study the UE. This is carried out on an event-by-event basis, by selecting a leading object in the event which is generally the charged particle having the highest transverse momentum in the event, the p_T^{lead} . The axis of direction of the p_T^{lead} in the event is used to define regions in the $\eta - \phi$ plane which have different sensitivities to the UE. The axis of direction of p_T^{lead} particle is well-defined for all events. It is also correlated with the axis of hard scattering in high p_T events [9]. The azimuthal angle difference between the p_T^{lead} and other associated particles, $|\Delta\phi| = |\phi - \phi_{lead}|$, is used to define the following three kinematic regions as shown in Table 1 and Figure 1)

Azimuthal angle difference	**Region**
$\|\Delta\phi\| < 60°$	Toward region
$60° < \|\Delta\phi\| < 120°$	Transverse region
$\|\Delta\phi\| > 120°$	Away region

TABLE I. The three topological regions.

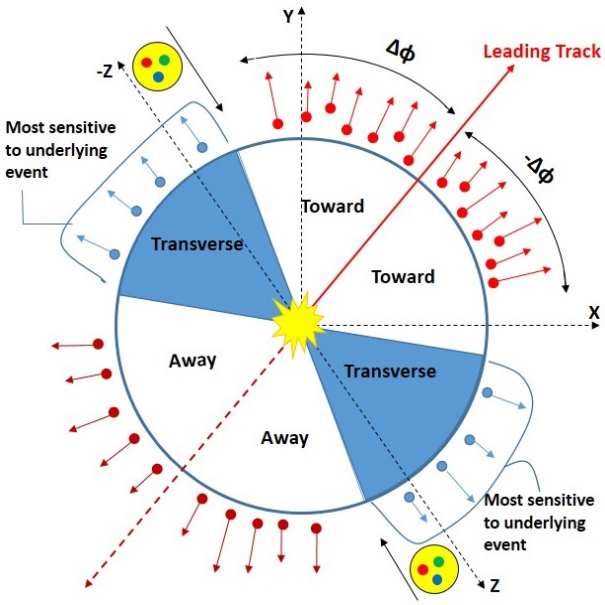


FIG. 1. The three topological regions

Out of these three regions, the **transverse** region is most sensitive to UE activity as it is perpendicular to the axis of hardest scattering and thus has the lowest level of activity from hard scattering. The particle production associated with hard and soft processes is explored by investigating the particles produced in **toward**, **away** and **transverse** regions. The $\Delta\phi$ distribution for the three regions is shown in Figure 2.

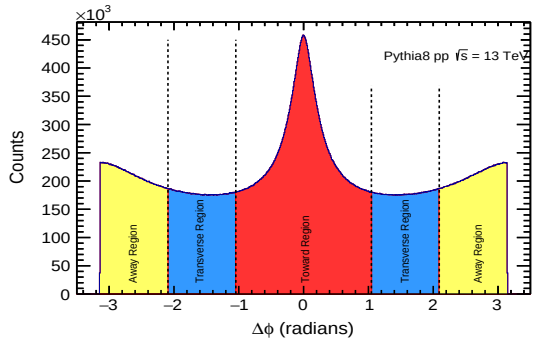


FIG. 2. The $\Delta\phi$ distribution of charged particles with respect to the leading particle in the three topological regions.

II.1. Transverse activity classifiers

The UE activity of an event is directly proportional to the activity in the transverse region of that event as this region is most sensitive of the three regions to UE activities. One can use UE-sensitive variables like the charged-particle multiplicity, N^{ch}, and the scalar sum of p_T of considered particles, $\sum_i p_{T_i}$ to construct transverse activity (or UE activity) classifier. One such classifier, called the relative transverse activity classifier, R_T was proposed [10, 13], to quantify UE activity. R_T of an event is defined as:

$$R_T = \frac{N_T^{ch}}{< N_T^{ch} >} \tag{1}$$

where N_T^{ch} is the number of charged particles in the transverse region of an event and $< N_T^{ch} >$ refers to the mean charged particle multiplicity. If UE activity in a particular event is high leading to the production of large number of particles in the transverse region, the corresponding R_T of that event will be higher. Similarly, if the UE activity of an event is low, a lesser number of particles are produced in the transverse region and thus R_T of the considered event will be low. One can observe that the choice of construction of such a variable classifies the events around $R_T = 1$ with "higher-than-average" UE from the "lower-than-average" UE. In similar lines, it is interesting and worth trying to explore the scalar sum of p_T, $\sum_i p_{T_i}$, of charged particles in transverse region to quantify UE activity of an event. The proposed classifier is denoted as S_T and is defined as :

$$S_T = \frac{\sum_i p_{T_i}^T}{< \sum_i p_{T_i} >} \tag{2}$$

where $\sum p_T^T$ is the $\sum_i p_{T_i}$ of charged particles in the transverse region of an event and $< \sum_i p_{T_i} >$ is the mean of $\sum p_T$. As observed for R_T, $S_T = 1$ also divides the events with "higher-than-average" UE activity from the "lower-than-average" UE activity. One of the primary goals of this study is to investigate the performance of this novel classifier S_T to probe UE activity of events as compared to the R_T classifier.

It was observed previously that above a certain lower threshold of p_T^{lead} ($p_T^{lead} > 5$ GeV/c), a plateau region in the transverse region was reached [9]. This plateau region signifies that the mean values of UE quantities (like mean charged-particle multiplicity density, $d\langle N_{ch}\rangle/d\eta$, and the $\sum_i p_{T_i}$) calculated over a fully inclusive set of particles show little dependence on the p_T^{lead} i.e the soft processes that contribute to UE are independent of p_T^{lead}. The slow rise of UE plateau in the transverse regions is possibly due to the additional contributions from wide-angle radiation associated with the hard scattering. In the present study, the particles that lie in the plateau region (5 GeV/c $< p_T^{lead} <$ 40 GeV/c) are considered to ensure that the UE observables in the transverse region remain nearly independent of the p_T of the leading particle.

III. ANALYSIS METHOD

Pythia 8.308 [14, 15], a general-purpose Monte Carlo (MC) event generator that can simulate initial and final-state parton showers, multipartonic interactions, hadronization, and particle decays has been used to generate p−p collisions at $\sqrt{s}$ =13 TeV. In Pythia, multi-partonic interactions (MPI) allow the generation of a large number of final state string objects over limited transverse space. It has been observed that inclusion of MPI is necessary for a correct description of UE in event genera-

tors like PYTHIA or Herwig [16]. From an experimental point of view, as UE consists of all activity that accompanies the primary hard scattering, it essentially consists of MPI and interactions between constituents of beam remnants. This is implemented in the Lund string model [17, 18]. However, this traditional model is expected to be modified in high parton density environments due to the inter-string interactions. Rope Hadronization (RH) is a model that extends the Lund string model to describe environments with many overlapping strings, such as in heavy ion collisions or high multiplicity p−p collisions. The idea is to let the strings overlap and act coherently forming a rope which hadronizes with a larger, effective string tension [19, 20]. This model is tested and observed to improve strange to non-strange ratios and the near-side long range correlations observed in high multiplicity p−p collisions [21, 22]. One essential ingredient used in the hadronization models is Color-Reconnection (CR). It is the mechanism that describes the formation of hadronizing strings between the outgoing partons. The implementation of this mechanism is based on the minimization of the total string potential energy by connecting different final state partons such that the total string length remains as short as possible [23, 24]. In Pythia 8, the QCD-based color reconnection scheme and rope hadronization has been shown to explain the enhancement of baryons [25]. It also explained the similarities in the multiplicity evolution of the p_T spectra among different colliding systems (p−p, p−Pb, and Pb−Pb). The present study includes the effect of color reconnection as well as rope hadronization to explore the strange particle and baryon production in varying topological regions of UE activity. The study also aims to compare the sensitivity of R_T and S_T . The kinematic cuts $p_T \geq 0.15$ GeV/c and $|\eta| \leq 0.8$ are applied to the particles to approximate experimental sensitivity.

IV. RESULTS AND DISCUSSIONS

IV.1. R_T and S_T distributions

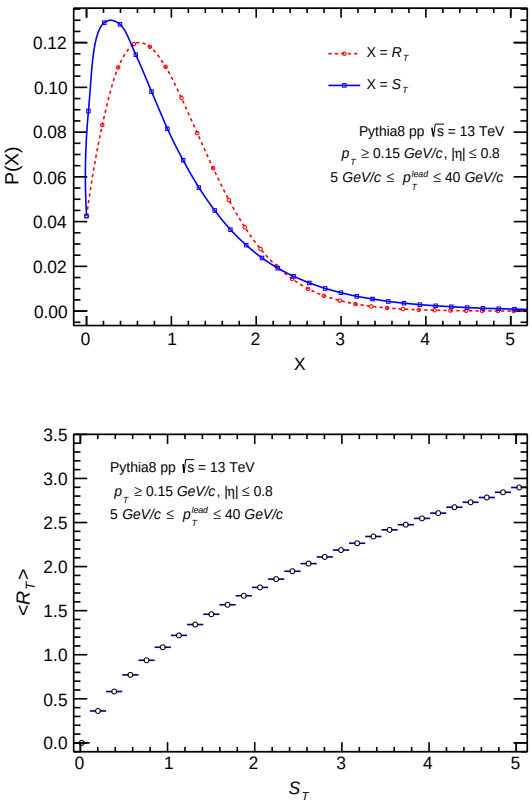


FIG. 3. (Upper Panel)The S_T and R_T distribution for events with $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV. (Lower panel)The variation of $< R_T >$ with S_T.

Figures 3 shows the S_T and R_T distributions obtained using charged particles in p−p collisions at $\sqrt{s} = 13$ TeV. One can observe that higher S_T (or R_T) events (i.e. high UE activity) are rarer than low UE activity events. In the bottom panel, where the variation of $< R_T >$ is shown as a function of S_T, one notes that it deviates from linearity at values of S_T greater than 1.5. Nevertheless, large values of S_T corresponds to large values of $< R_T >$ showing a positive correlation between the two observables.

Figure 4 shows the mean transverse momentum ($< p_T >$) of the charged particles as a function of

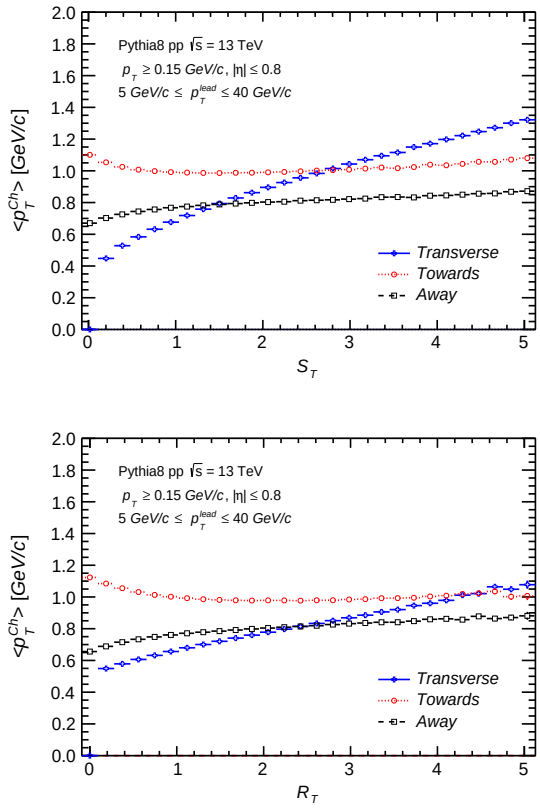


FIG. 4. The variation of $< p_T >$ of charged particles with respect to S_T (upper panel) and R_T (lower panel) for events with $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV for the three topological regions.

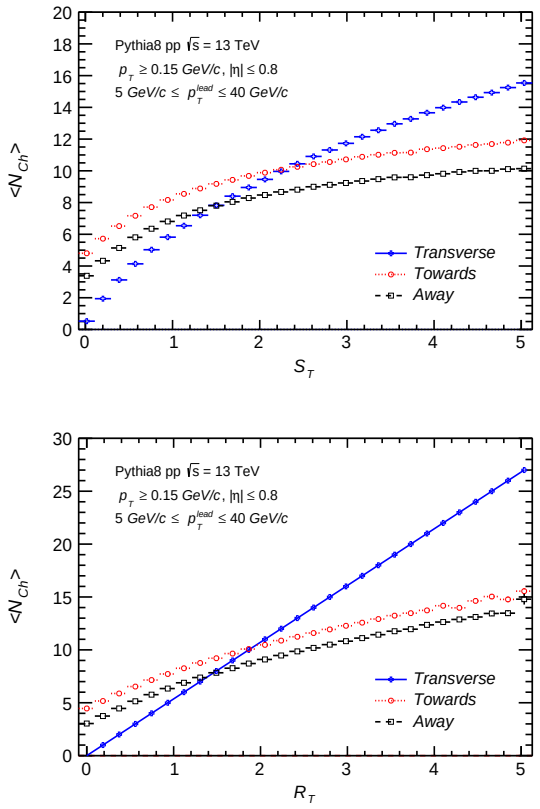


FIG. 5. The variation of $< N_{ch} >$ with S_T (upper panel) and R_T (lower panel) for events with $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV for the three topological regions.

S_T(and R_T) in the three topological regions. In the toward region, $< p_T >$ of the charged particles is largest for low R_T and S_T regime due to dominance of jets fragmenting into numerous particles in this region. It slowly decreases and saturates for S_T (and R_T) values > 1.5. In the away side region which is dominated by away-side jet, the $< p_T >$ slightly increases with R_T and S_T. However, in the transverse region, $< p_T >$ values are lower than toward and away regions for lower ranges of S_T and R_T. It increases with R_T and S_T and the increase can be attributed to dominance of UE activity. The increase is steeper in case of S_T and can be seen as a consequence of the auto-correlation effect arising due the structure of the observable. Therefore, it is worth observing the variation of mean charged particle multiplicity ($< N_{ch} >$) as a function of S_T. Figure 5 shows the evolution of $< N_{ch} >$ as a function of S_T(and R_T) in the three topological regions. Here, one clearly observes the $< N_{ch} >$ in toward region is consistently higher than away region for the considered range of S_T (and R_T). However, in the transverse region, for lower values of S_T (and R_T) there is a crossing over with the towards and away region. There is a strong rise in $< N_{ch} >$ in the transverse region for values of $S_T > 1.5$ which signals towards an increase in transverse activity emanating primarily from underlying events. A similar trend is also seen for R_T where the increase is stronger than S_T and can be attributed to the autocorrelation effect. The study was also performed for different identified particles to see the evolution of their multiplicity with S_T (and R_T). Figures 6 and 7 show the variation of mean multiplicity of pions, kaons and protons as a function of S_T

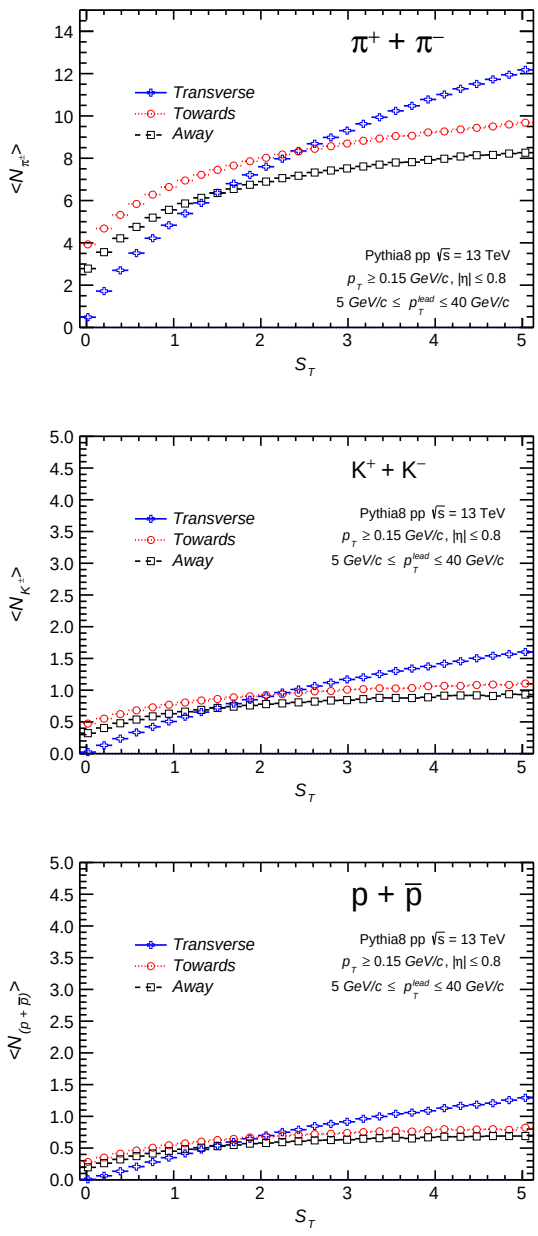


FIG. 6. The variation of mean multiplicity of pions(upper panel), kaons(middle panel) and protons(bottom panel) with S_T in p−p collisions at $\sqrt{s} =$ 13 TeV for the three topological regions.

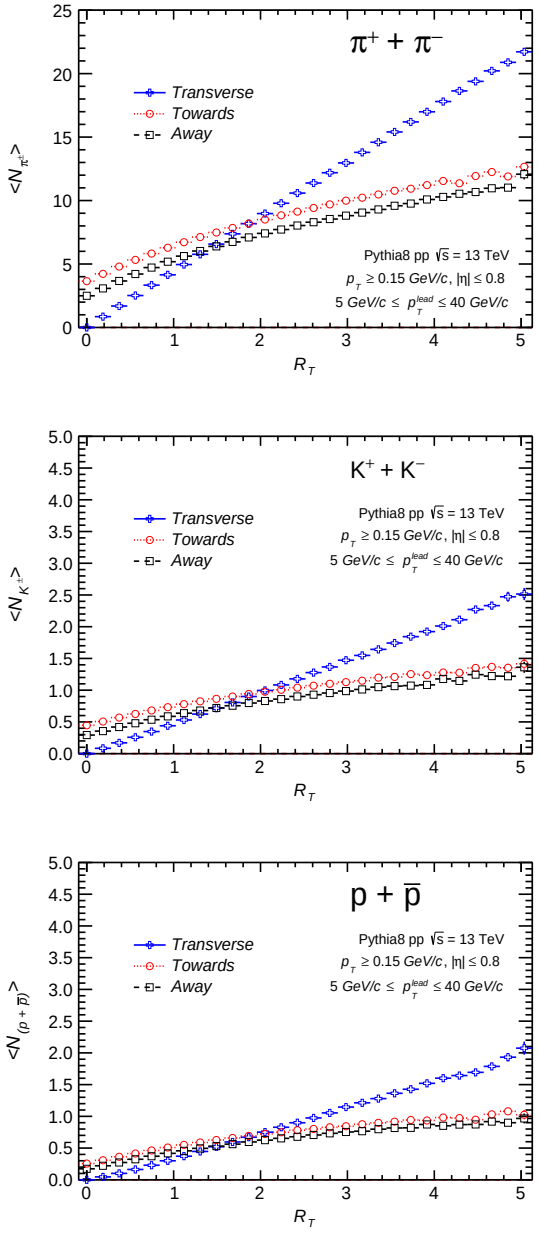


FIG. 7. The variation of mean multiplicity of pions(upper panel), kaons(middle panel) and protons(bottom panel) with R_T in p−p collisions at $\sqrt{s} =$ 13 TeV for the three topological regions.

and R_T, respectively. The sensitivities for different topological regions is also shown. The trend is similar to that observed for charged multiplicity case. However, neutral V_0 particles like K^0_S and Λ^0, do not exhibit autocorrelation effects as both S_T and R_T are defined in terms of charged particles. One can observe that values for toward region remain consistently higher than the other regions up to S_T and $R_T \sim 2.0$ indicating the dominance of V_0 production by fragmenting jets. The gradual increase for transverse region with increase in S_T values indicates that hadronisation mechanism which is sensitive to the non-perturbative effects affects the strangeness and baryon content of the

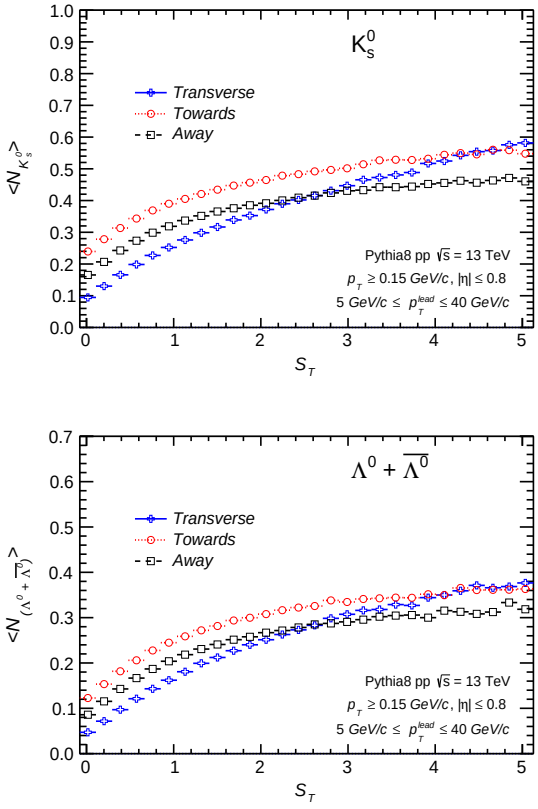


FIG. 8. The variation of mean number of K^0_S (upper panel) and Λ^0 (lower panel) with S_T in p−p collisions at $\sqrt{s} = 13$ TeV for the three topological regions.

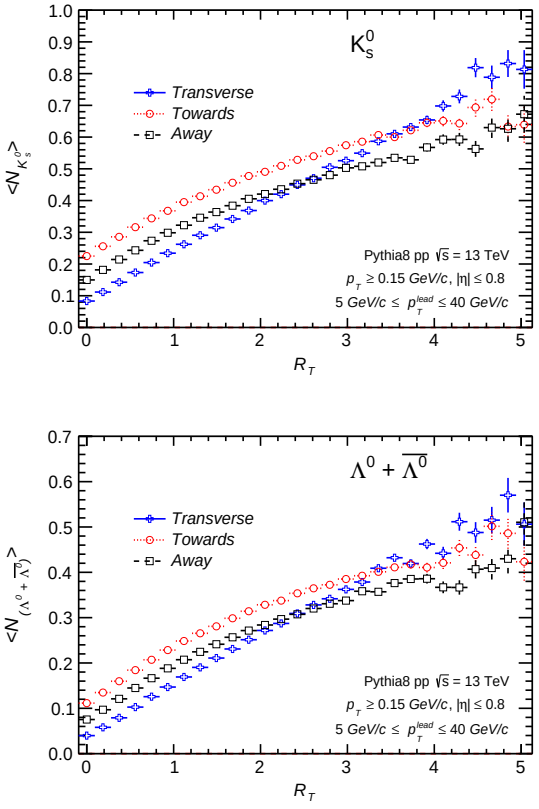


FIG. 9. The variation of mean number of K^0_S (upper panel) and Λ^0 (lower panel) with R_T in p−p collisions at $\sqrt{s} = 13$ TeV for the three topological regions.

final state.

IV.2. Transverse momentum distributions

The evolution of the p_T distributions of identified particles in different regions of transverse activity quantified by R_T and S_T can provide additional information about the particle production mechanism. Figures 10, 11, 12, 13, 14 and 15 show the p_T spectra of charged pions, kaons and protons in different classes of S_T and R_T. The same is shown for K^0_S and Λ^0 in Figure 16, 17, 18 and 19. The p_T spectra obtained in **transverse**, **towards** and **away** regions are shown in the top, middle and bottom panels, respectively. The corresponding ratios to the S_T (and R_T) integrated spectra is shown in the bottom section of each figure. It can be observed that for all particle species, spectral shapes harden with increasing UE activity i.e. with increasing values of S_T and R_T in the transverse region and is reminiscent of radial flow effects. However, in the toward and away region, the differentiation of these classes is not pronounced and the spectra is harder for low values of S_T and R_T for toward as. well as away region . This is observed for all the identified particle species studied and can also be seen from ratio to $S_T \geq 0$ (and $R_T \geq 0$) integrated spectrum in lower panels of the figures. Additionally, it can be observed that the p_T spectra in the transverse region is more strongly differentiated by S_T classes than R_T classes for charged particles. This effect seems to be a manifestation of the construction of the observable as this differentiation vanished for neutral particles.

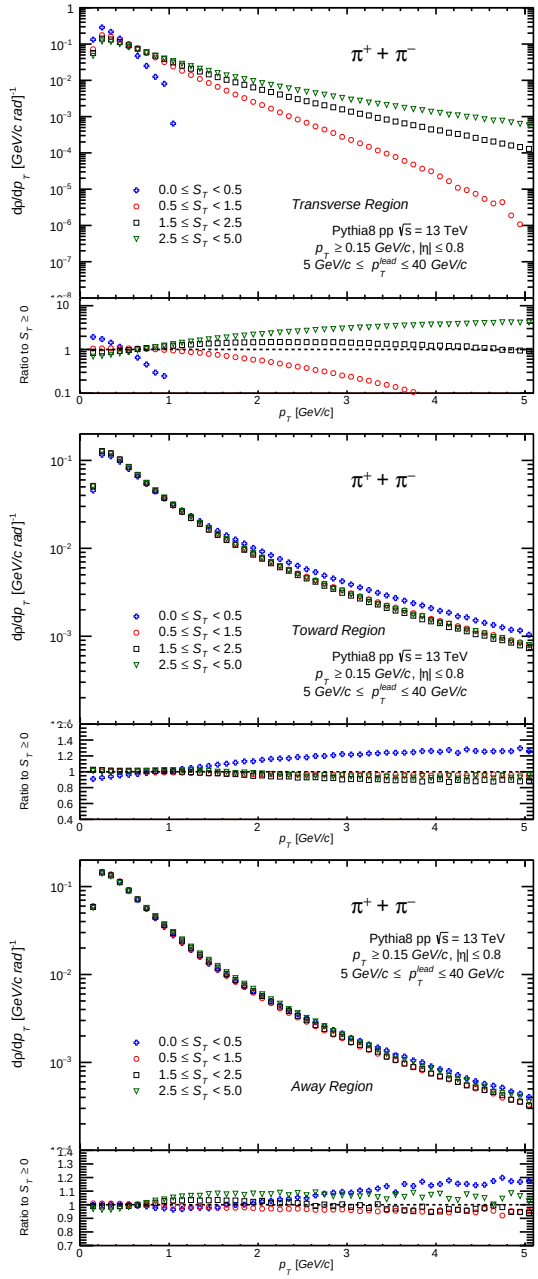


FIG. 10. The p_T spectra of charged pions in transverse, toward and away regions for different S_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

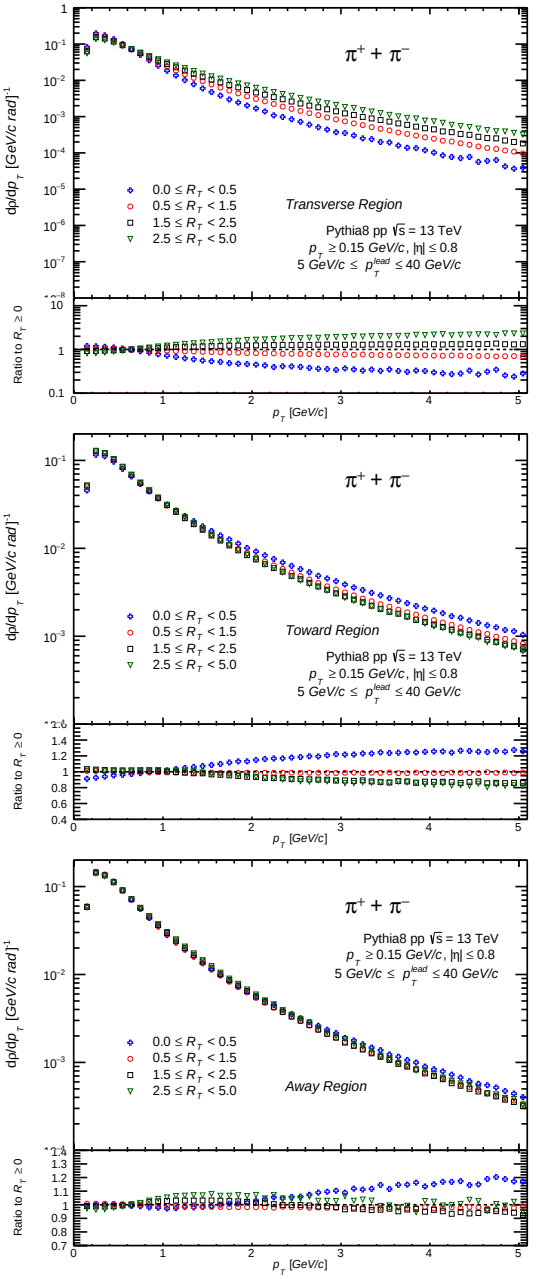


FIG. 11. The p_T spectra of charged pions in transverse, toward and away regions for different R_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

IV.3. Mean p_T

The evolution of the $< p_T >$ of pions, kaons ,protons, K^0_S and Λ^0 were studied as a function of S_T (and R_T) for the three topological regions. Figures 20, 21, 22, 23, and 24 show the variation of $< p_T >$ with respect to S_T and R_T. One can observe that $< p_T >$ of the toward region is consistently higher for all the identified particles for lower ranges of S_T and R_T. The values for away region is also higher than transverse region for lower ranges. This is expected as these regions are dominated by particle production via jet fragmentation and hence carry the highest p_T. How-

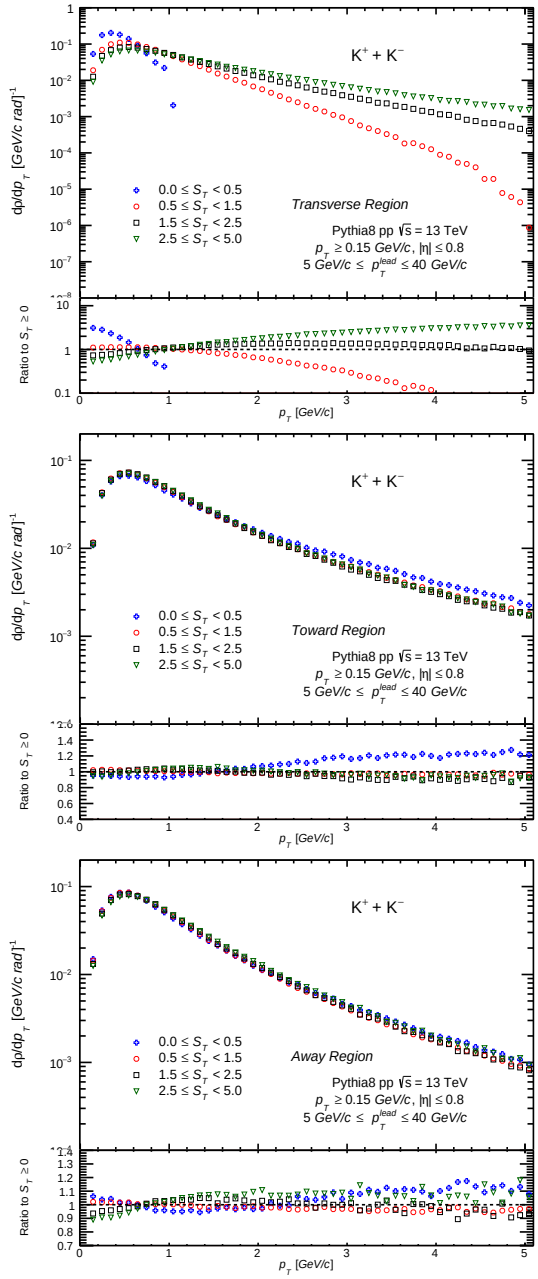


FIG. 12. The p_T spectra of charged kaons in transverse, toward and away regions for different S_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

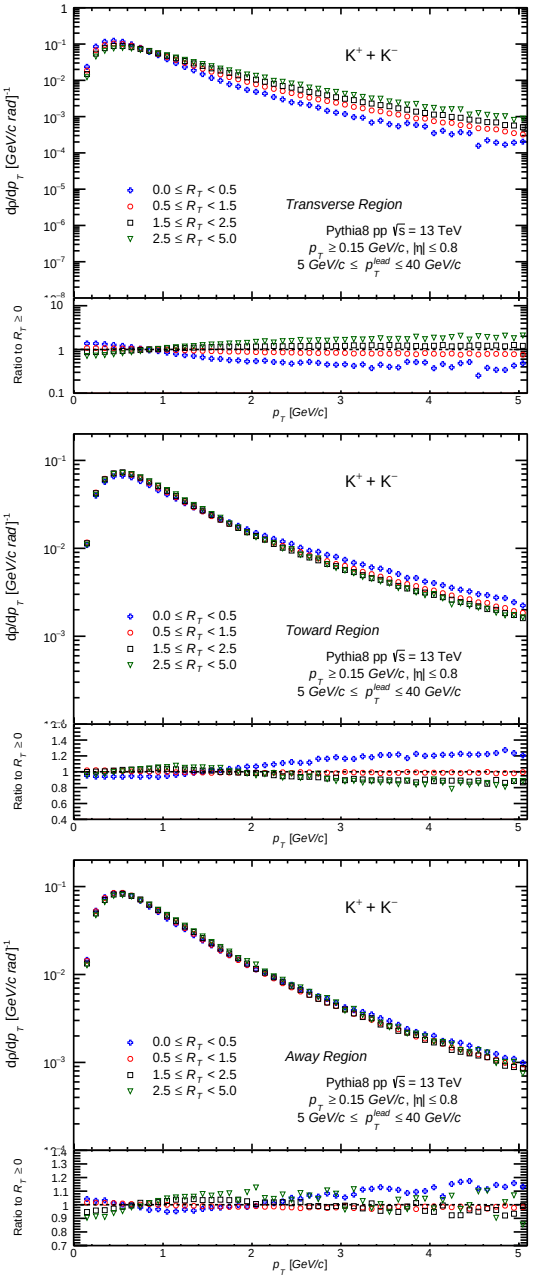


FIG. 13. The p_T spectra of charged kaons in transverse, toward and away regions for different R_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

ever, the trend observed for pions is different than other identified particles. The $< p_T >$ of pions are considerably higher for lower ranges showing a slight decrease with S_T (and R_T) and remains more or less uniform throughout. The values for other particles show a consistent increasing trend with S_T (and R_T). The values of $< p_T >$ of identified charged particles are highest in the transverse region for higher ranges of S_T and R_T where the transverse activity due to underlying events dominates. For charged particles, there is a smooth crossing over of the $< p_T >$ in transverse region with the one of away and toward region for $S_T \sim 1.5$ and $R_T \sim 1.5$. The same is not observed for the

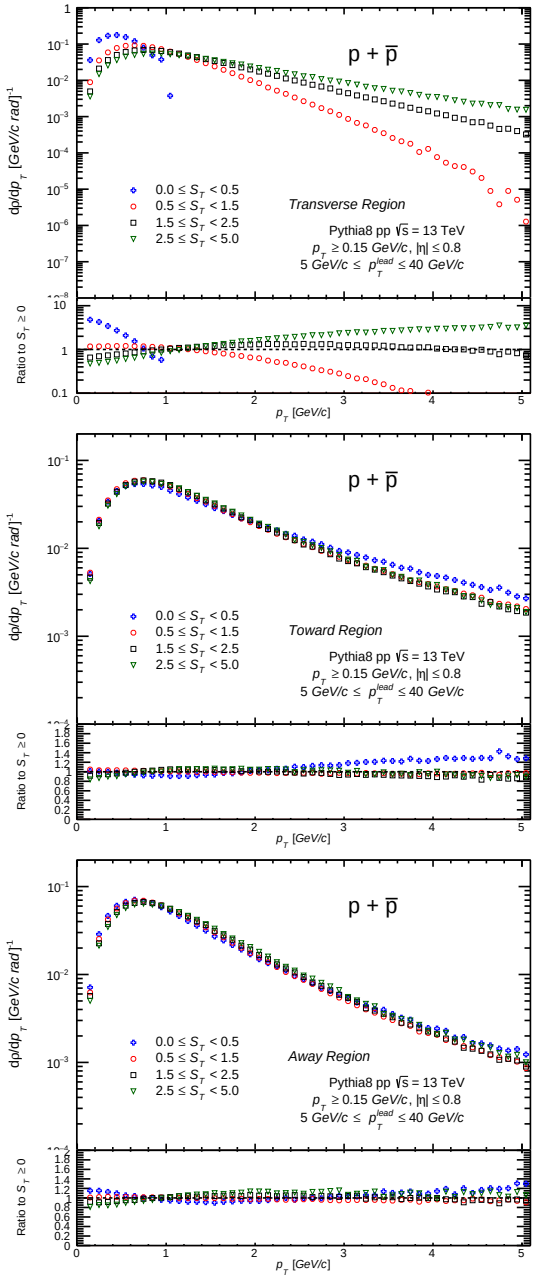


FIG. 14. The p_T spectra of protons in transverse, toward and away regions for different S_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

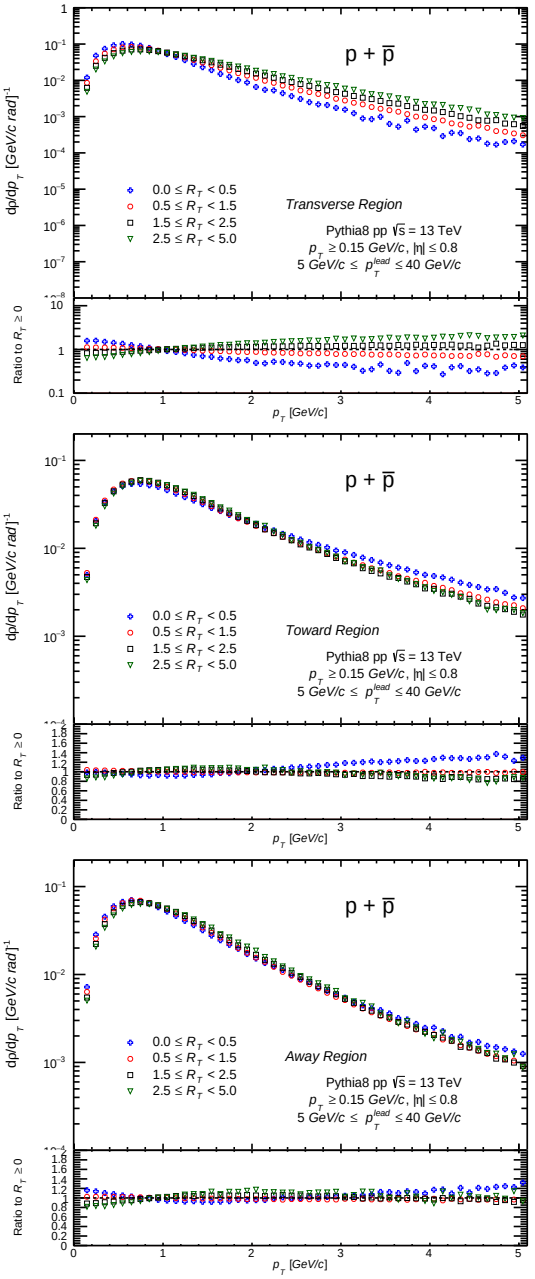


FIG. 15. The p_T spectra of protons in transverse, toward and away regions for different R_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

neutral particles for which the contribution from toward region dominates throughout. The crossing point for the V_0 particles with away region is seen at S_T , $R_T \sim 2.5$ but with towards region happen at larger values. This indicates that the contribution to $< p_T >$ is essentially driven by jet fragmentation for neutral particles while the charged particles are affected by underlying events.

V. SUMMARY

The production of various species of particles $\pi^{\pm}$, $K^{\pm}$, $p + \bar{p}$, K^0_S , and Λ^0) in p−p collisions at

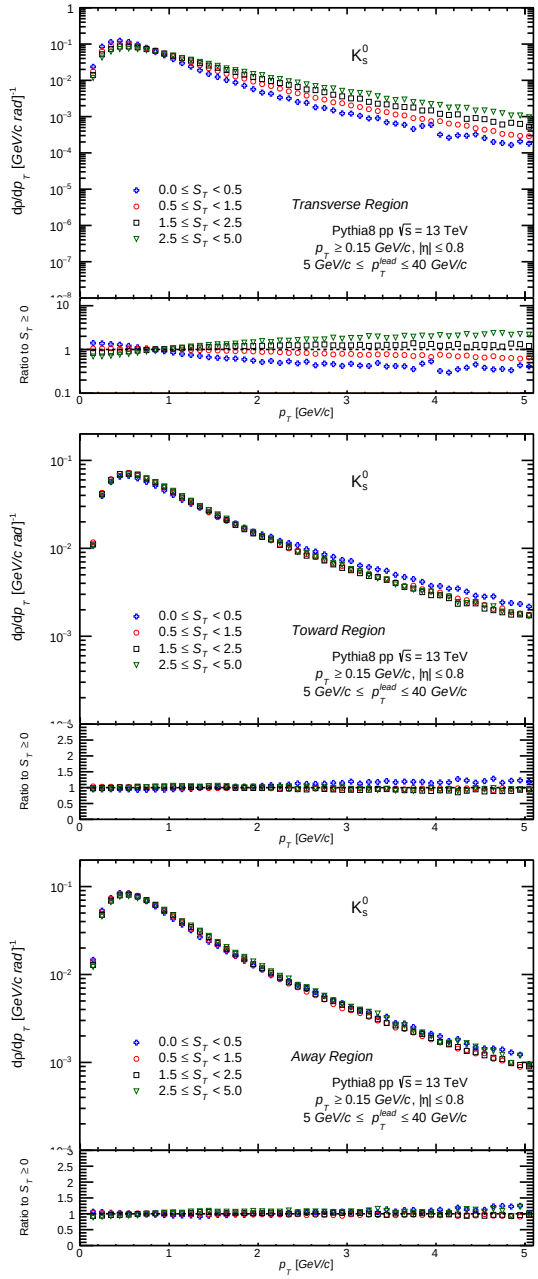


FIG. 16. The transverse momentum spectra of K^0_S particles in transverse, toward and away regions for different S_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

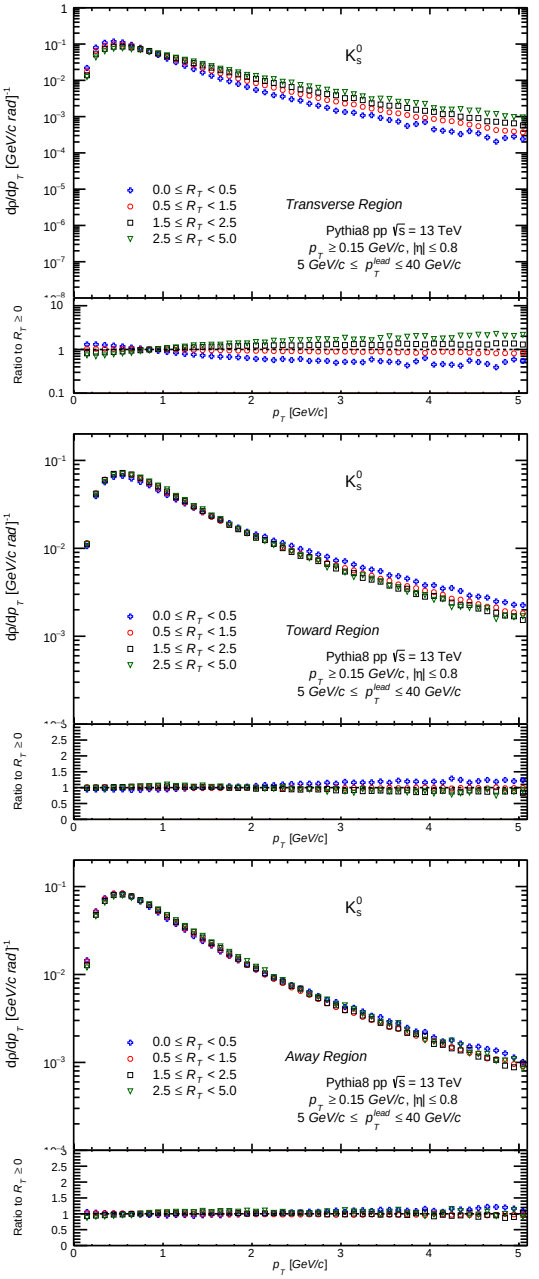


FIG. 17. The transverse momentum spectra of K^0_S particles in transverse, toward and away regions for different R_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

$\sqrt{s} = 13$ TeV is studied as a function of transverse activity classifier, S_T, in the three topological regions using pQCD inspired PYTHIA 8 event generator. The classifier S_T was introduced for the first time to investigate its performance with respect to the previously defined R_T. The transverse activity classifiers R_T and S_T are used to investigate differentially in the three topological regions where the particle production mechanisms are expected to have contributions emanating from soft processes (**Transverse** region) and hard processes (**Toward** and **Away** regions). The evolution of mean multiplicity and mean transverse momentum for identified particles were studied as a function of

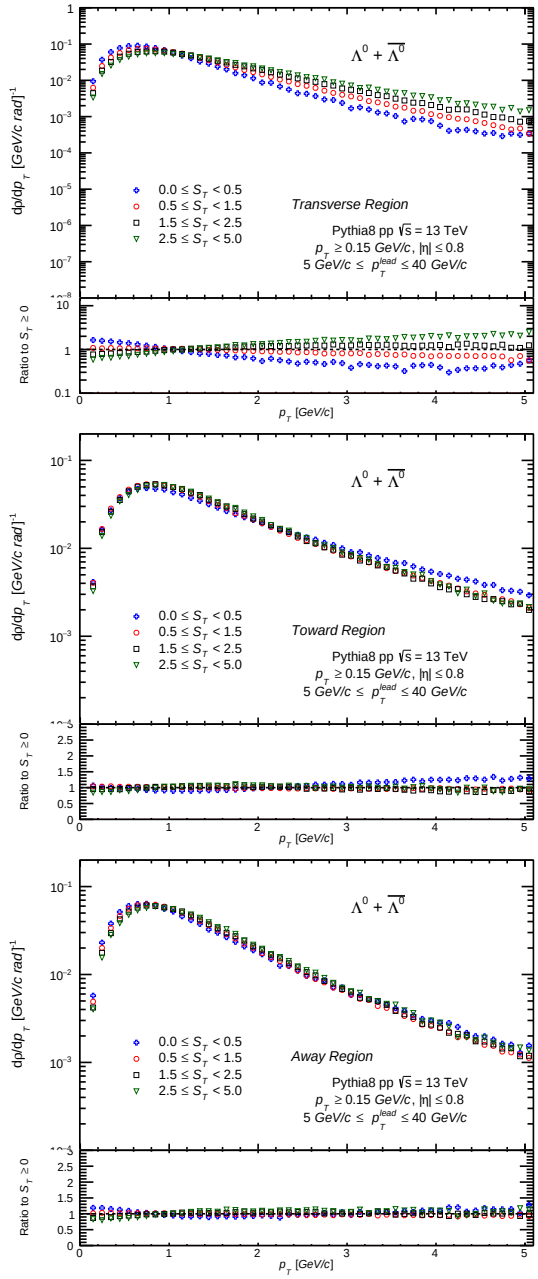


FIG. 18. The transverse momentum spectra of Λ^0 in transverse, toward and away regions for different S_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

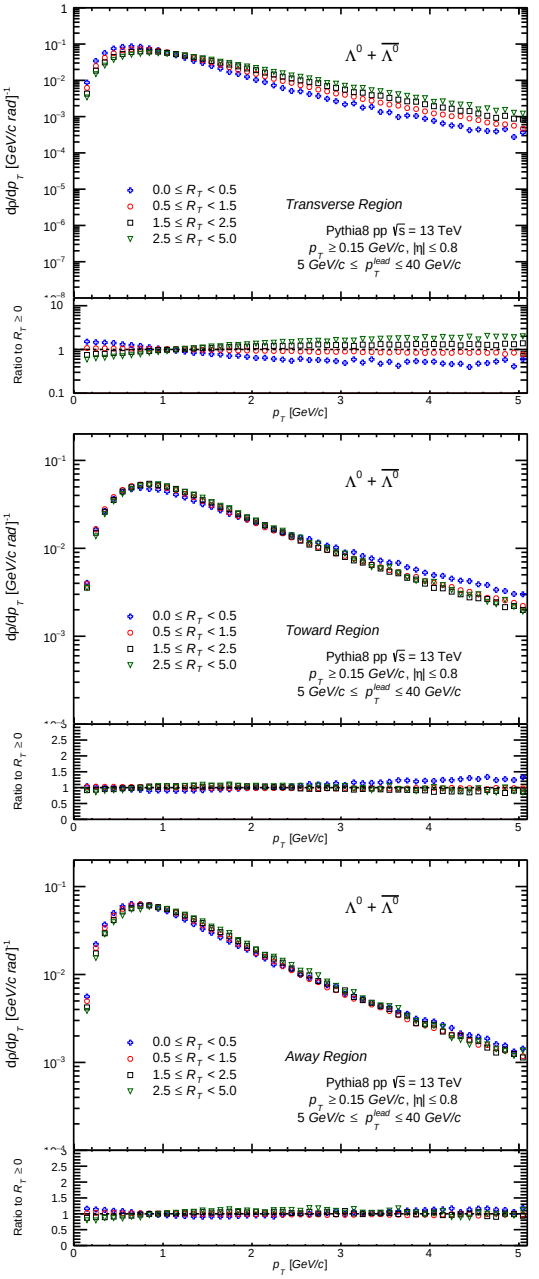
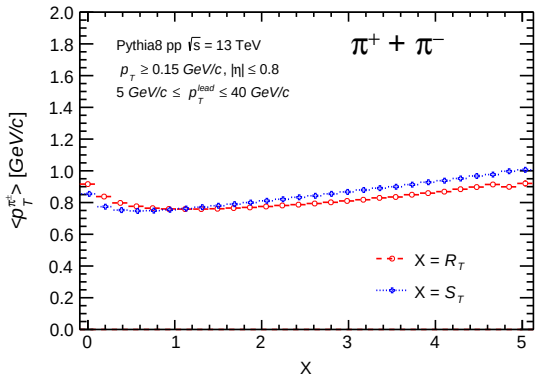
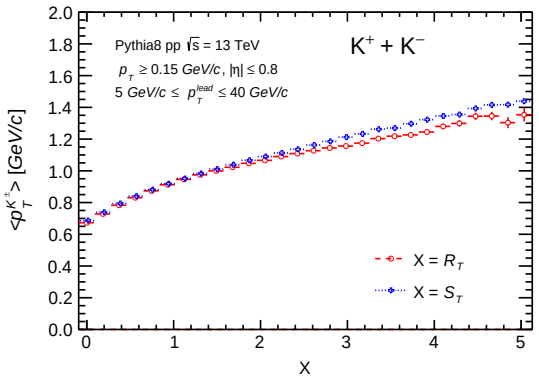
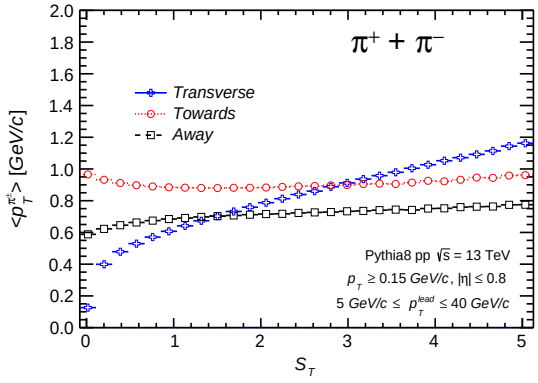


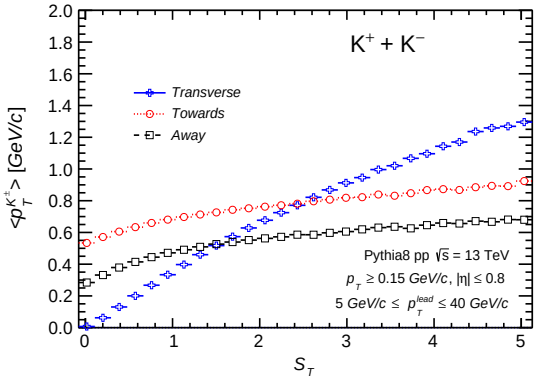
FIG. 19. The transverse momentum spectra of Λ^0 in transverse, toward and away regions for different R_T classes in p−p collisions at $\sqrt{s} = 13$ TeV.

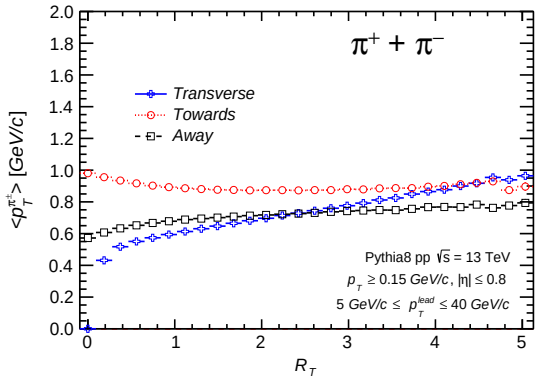
S_T. It was observed that the particle production in the transverse region is dominated by underlying events for higher ranges of S_T. However, for the production of V_0 particles, there was an observed dominance from hard processes. The p_T spectra of the considered particle species were observed to be sensitive to UE activity. Moreover, it was observed that p_T spectra of identified charged particles were strongly differentiated by S_T classes than R_T classes. Therefore, it can be observed that by measuring the production of particles as a function of these transverse activity classifiers, one can study various aspects of underlying events. The results presented in this work are expected to











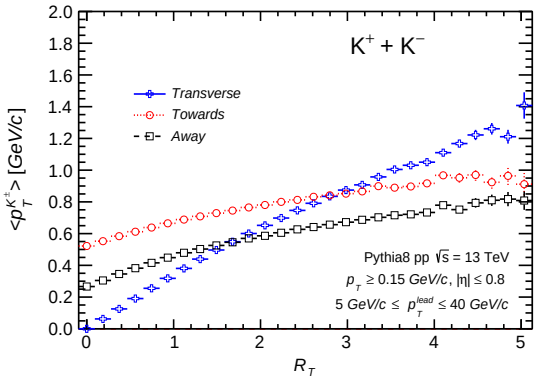


FIG. 20. (Upper panel)The variation of $< p_T >$ of pions with respect to S_T (and R_T) with $\mathrm{l}p_T^{lead} \geq 5$ GeV/c for p−p collisions at $\sqrt{s} = 13$ TeV. The middle and bottom panels show the evolution of $< p_T >$ of pions as a function of S_T and R_T, respectively. The comparison is shown for the three topological regions.

FIG. 21. (Upper panel)The variation of $< p_T >$ of kaons with respect to S_T (and R_T) with leading $p_T^{lead} \geq 5$ GeV/c for p−p collisions at $\sqrt{s} = 13$ TeV. The middle and bottom panels show the evolution of $< p_T >$ of kaons as a function of S_T and R_T, respectively. The comparison is shown for the three topological regions.

shed more light on the understanding of underlying event activity at LHC energies. The obtained results can act as a baseline for the upcoming experimental measurements at LHC energies and can help constrain the Monte-Carlo models.

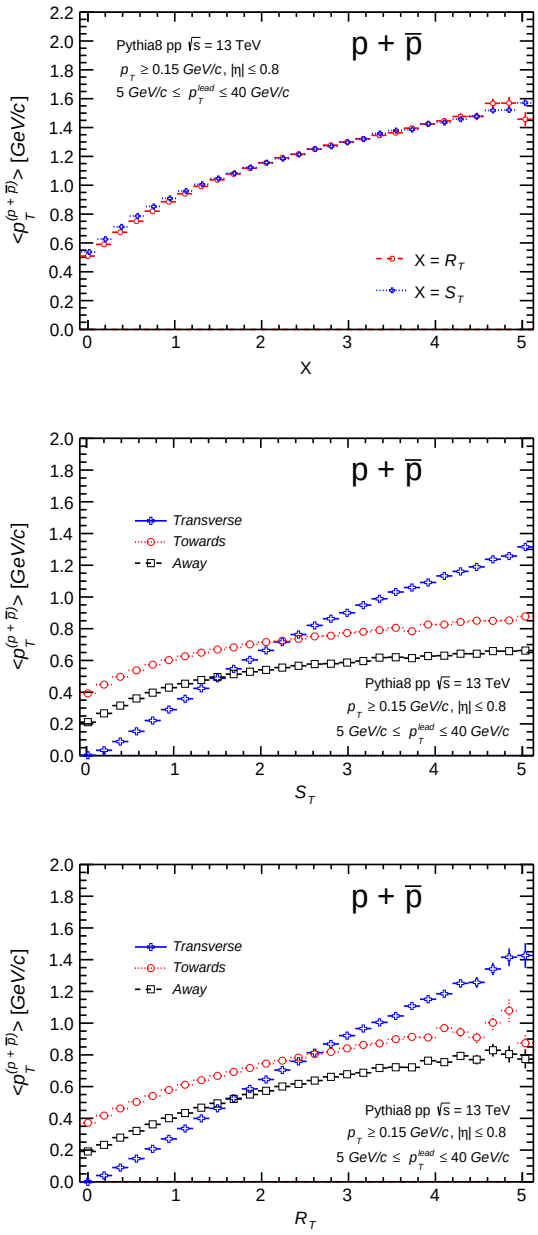


FIG. 22. (Upper panel)The variation of $< p_T >$ of protons with respect to S_T (and R_T) for $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV. The middle and bottom panels show the evolution of $< p_T >$ of protons as a function of S_T and R_T, respectively. The comparison is shown for the three topological regions.

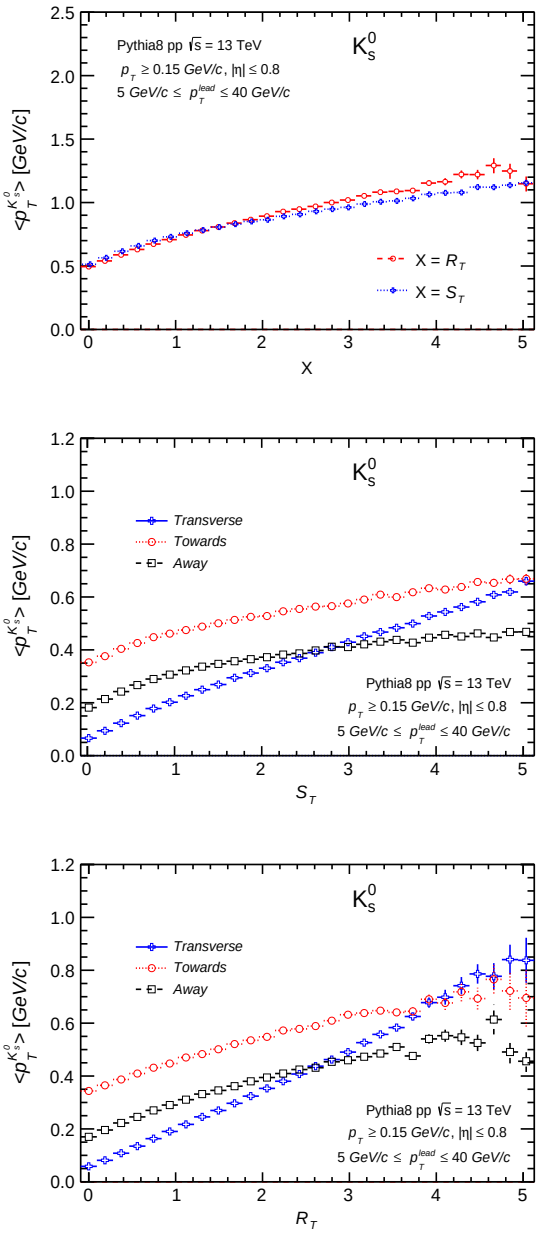


FIG. 23. (Upper panel)The variation of $< p_T >$ of K_S^0 with respect to S_T (and R_T) for $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV. The middle and bottom panels show the evolution of $< p_T >$ of K_S^0 as a function of S_T and R_T, respectively. The comparison is shown for he three topological regions.

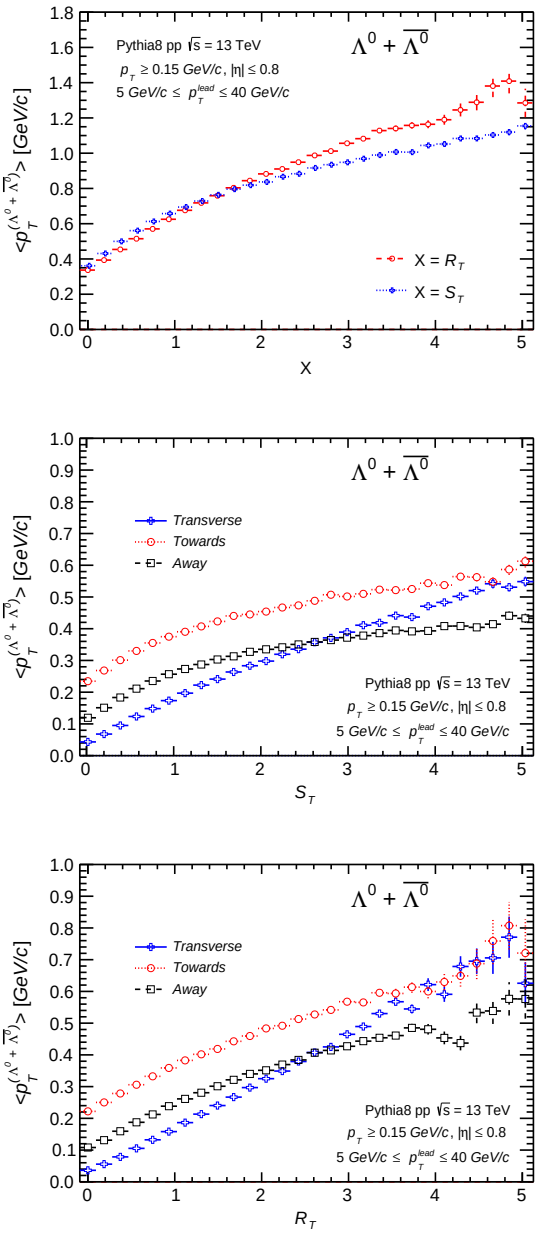


FIG. 24. (Upper panel)The variation of $< p_T >$ of Λ^0 with respect to S_T (and R_T) for $p_T^{lead} \geq 5$ GeV/c in p−p collisions at $\sqrt{s} = 13$ TeV. The middle and bottom panels show the evolution of $< p_T >$ of Λ^0 as function of S_T and R_T, respectively. The comparison is shown for the three topological regions.

VI. ACKNOWLEDGEMENTS

Sadhana Dash would like to acknowledge and thank SERB Power fellowship, **SPF/2022/000014** for supporting the present work.

[1] B. Abelev et al. [ALICE Collaboration] Phys. Lett. B **728**, 216-227, (2014)
[2] J. Adams et al. [STAR Collaboration] Phys. Rev. Lett. **95** 152301, (2005)
[3] S Chatrchyan *et al.*, (CMS Collaboration), JHEP **07**, 076, (2011).
[4] J. Adams et al. [ALICE Collaboration] Phys. Rev. Lett. **116** 132302, (2016)
[5] V Khachatryan *et al.*, (CMS Collaboration), JHEP **09**, 091 (2010).
[6] V Khachatryan *et al.*, (CMS Collaboration), Phys. Rev. Lett. **116(17)**,172302 (2016).
[7] J. Adams *et al.*, (ALICE Collaboration),Nature Physics**13**,535-539 (2017).
[8] G Aad *et al.*, (ATLAS Collaboration), Phys. Rev. Lett. **116(17)**,172301 (2016).
[9] Georges Aad et al.(ATLAS Collaboration), Phys. Rev. **D 83**, 112001, (2011).
[10] T. Martin, P. Skands, and S. Farrington, Eur. Phys. J. C **76**, 299 (2016)
[11] Rick Field, Annu. Rev. Nuclear and Particle Science**62**, 453-483 (2012).
[12] Krishna Kumar and Sadhana Dash, Eur. Phys. J. A **58**, 148(2022).
[13] S. Acharya et al, (ALICE Collaboration), JHEP 06 (2023) 027.
[14] Torbjörn Sjöstrand, Stefan Ask, Jesper R Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O Rasmussen, and Peter Z Skands, Comput. phys. commun. **191**, 159–177 (2015).
[15] P. Skands, S. Carrazza, J. Rojo, Eur. Phys. J. C **74(8)**, 3024 (2014).
[16] J. Bellm, EPJ **C** 76(4), 196 (2016).
[17] B Anderson, G. Gustafson and B. Soderberg, Z. Phys. **C 20**, 317 (1983).
[18] B Anderson and G. Gustafson, Z. Phys. **C 3**,223 (1980).
[19] T. S. Biro, H. B. Nielson and J. Knoll, Nucl. Phys. **B 245**, 449, (1984).
[20] C.Bierlich, G. Gustafson, L. Lonnblad and A. Tarasov, J. High Energ. Phys. **2015: 148**, (2015).
[21] Ranjit Nayak, Subhadip Pal, and Sadhana Dash, Phys. Rev. **D 100**, 074023 (2019).
[22] Pritam Chakraborty and Sadhana Dash, Phys. Rev. **C 102**, 055202, (2020).
[23] Christian Bierlich and Jesper Roy Christiansen, Phys. Rev **D 92**,094010 (2015).
[24] Jesper Roy Christiansen and P. Z.Skands , JHEP, **08**, 003 (2015).
[25] Ankita Goswami, Ranjit Nayak, Basanta Nandi, and Sadhana Dash, Eur. Phys. J. C **81(11)**, 1-9 (2021).

Strangeness plus-one ($S = +1$) resonance-state P_0^{+*} via $K^+n \to K^{*0}p$

Dayoung Lee [1, 2, *] and Seung-il Nam* [1, 2, 3, †]

[1] *Department of Physics, Pukyong National University (PKNU), Busan 48513, Korea*
[2] *Center for Extreme Nuclear Matters (CENuM), Korea University, Seoul 02841, Korea*
[3] *Asia Pacific Center for Theoretical Physics (APCTP), Pohang 37673, Korea*

(Dated: March 5, 2024)

In our current study, we delve into the peak-like structure observed during the reaction process of $K^+n \to K^0p$ at approximately $\sqrt{s} \sim 2.5$ GeV. Our focus centers on exploring the potential $S = +1$ resonance $P_0^{+*} \equiv P_0^*$ as an excited state within the extended vector-meson and baryon (VB) antidecuplet. To achieve this aim, we employ the effective Lagrangian method in conjunction with the (u,t)-channel Regge approach, operating within the tree-level Born approximation. We thoroughly examine various spin-parity quantum numbers for the resonance, resulting in a compelling description of the data, where $M_{P_0^*} \approx 2.5$ GeV and $\Gamma_{P_0^*} \approx 100$ MeV. Furthermore, we propose an experimental technique to amplify the signal-to-noise ratio (S/N) for accurately measuring the resonance. Notably, our findings reveal that background interference diminishes significantly within the K^* forward-scattering region in the center-of-mass frame when the K^* is perpendicularly polarized to the reaction plane. Additionally, we explore the recoil-proton spin asymmetry to definitively determine the spin and parity of the resonance. This study stands to serve as a valuable reference for designing experimental setups aimed at investigating and comprehending exotic phenomena in QCD. Specifically, our insights will inform future J-PARC experiments, particularly those employing higher kaon beam energies.

Keywords: Pentaquark, vector-meson and baryon interaction, $S = +1$ production process, effective Lagrangian method, Regge approach, polarization, resonance, forward-scattering cross-section.

arXiv:2403.01191v1 [hep-ph] 2 Mar 2024

I. INTRODUCTION

Quantum Chromodynamics (QCD) stands as the foundational principle governing the strong interactions among standard-model particles, portraying various hadrons through the non-perturbative interactions of quarks and gluons. It showcases confinement and asymptotic freedom within the framework of non-Abelian color-SU(3) gauge theory. While the minimal composition for baryons is represented by qqq and for mesons by $q\bar{q}$, QCD does not forbid more intricate compositions such as $qqqq\bar{q}$ (pentaquark) and $q\bar{q}q\bar{q}$ (tetraquark), termed *Exotics*. Experimental evidence supporting these exotics has accumulated over decades. The initial observation of the tetraquark meson $X(3872)$, which defies the simple quark model, dates back to 2003, reported by the Belle experiment [1]. Recently, its existence was reaffirmed by the LHCb [2] and CMS [3] experiments, specifying $J^{PS} = 1^{++}$. Additionally, the Belle experiments reported evidence for $Y(4660)$ and $Z(4430)$ in 2007 [4]. Further tetraquark states have been identified via experiments conducted by LHCb, Belle, BES III, Fermilab, and others [5]. Regarding baryons, LHCb reported the $P_c^+(4312, 4440, 4457)$, observed in decay into J/ψ and p [6]. Recently, the same facility detected a novel pentaquark state with the strange-quark s ($udsc\bar{c}$) in the decay of $B^- \to J/\psi\Lambda\bar{p}$ [7]. While not yet definitively confirmed, these pentaquarks can potentially be understood as bound states of vector-meson and baryon (VB), denoted as $P_c^+ \sim D^*\Sigma_c$.

Remarkably, all experimentally observed and confirmed exotics exhibit heavy (charm) flavor. A theoretical rationale for this peculiarity concerning the stability of heavy pentaquarks was elucidated through color-spin interactions among quarks, contrasting with colorless hadronic interactions [8]. While the $S = +1$ light-pentaquark $\Theta^+ \sim uudd\bar{s}$ was reported by the LEPS collaboration [9], supported by the non-topological chiral-quark soliton model [10], its existence remains disputed and unresolved due to varied outcomes across experimental facilities [11]. Furthermore, generating the $S = +1$ resonance dynamically proves challenging within the framework of Weinberg-Tomozawa (WT) type chiral interactions, primarily due to their repulsive nature. Nevertheless, the suggestion of an $S = 0$ pentaquark-like bound-state $P_s^+ \sim uuds\bar{s} \sim K^*\Sigma$, akin to P_c^+ in heavy flavor, was put forth by one of the present authors and collaborators [12, 13], aiming to elucidate the bump structure observed near $\sqrt{s} \approx 2.1$ GeV in $\gamma p \to \phi p$ [14]. In Ref. [14], the existence of P_s^+ was found to be crucial for reproducing the nontrivial structure observed in the spin-density matrix elements (SDMEs), in addition to successfully describing the angular-dependent cross-sections.

*E-mail: ldyoung0421@pukyong.ac.kr
†E-mail: sinam@pknu.ac.kr (corresponding author)

Typeset by REVTEX

Given the current circumstances outlined above, it becomes of paramount importance to explore the potential for observing evidence of light-pentaquark states, either as bound states or resonances within the VB framework, to further enhance our understanding of QCD regarding exotic particles. Additionally, even in the $S = +1$ channel, there may exist a resonance or bound-state pentaquark, despite the lack of support from chiral dynamics [15]. Hence, in the present study, we undertake a theoretical examination of relatively aged data pertaining to the $S = +1$ channel reaction process, such as $K^+n \to K^{*0}p$ [16]. Intriguingly, this reaction process exhibits a peak-like structure around $\sqrt{s} = 2.5$ GeV, potentially indicative of an $S = +1$ resonance state, which we provisionally designate as $P_0^{+*} \equiv P_0^*$. We aim to elucidate this hypothetical pentaquark resonance P_0^* in our current investigation. Drawing upon our previous work [13], where we assigned the light-flavor VB pentaquark as $P_s \sim qqqs\bar{s}$, assuming it as a member of the extended VB antidecuplet flavor structure akin to Θ^+, we can extend this assignment to other possible states: $P_0 \sim qqqq\bar{s}$, $P_1 \equiv P_s \sim qqqs\bar{s}$, $P_2 \equiv P_{ss} \sim qqss\bar{s}$, and $P_3 \equiv P_{sss} \sim qsss\bar{s}$, where q represents the flavor-SU(2) light quarks (u and d), being categorized by the numbers of the strange quarks they include.

In our previous study [13], we set the mass of P_s to be approximately 2.071 GeV, guided by the chiral-unitary approach [12]. Assuming $P_s(2071)$ as the $S = 0$ ground-state pentaquark of the extended VB antidecuplet and considering the mass difference between antidecuplets due to the inertia moment of the non-topological chiral soliton, approximately ~ 180 MeV [10], the ground-state P_0 mass is estimated to be around 1.890 GeV. Consequently, we regard the hypothetical resonance in the vicinity of the bump structure at 2.5 GeV as a resonance state of P_0, denoted as $P_0^*(2500)$. Furthermore, we highlight several advantageous features of the chosen process: Firstly, since all hadrons in the final state of $K^+n \to K^{*0}p \to \pi^\mp K^\pm p$ can be measured experimentally in principle, the Fermi-motion effects of the deuteron-target experiment can be minimized. Secondly, in the s channel, it is possible to measure the resonance simultaneously coupling to PB and VB channels, allowing for the study of its production mechanism in two distinct dynamics. Lastly, this process involves the vector meson with polarizations, enabling the selection of specific quantum numbers for the resonance and control over its production mechanism.

To achieve this objective, we employ a straightforward tree-level Born approximation calculation utilizing the effective Lagrangian method. We account for the Regge trajectories in the t and u channels, incorporating phenomenological form factors to characterize the spatial extension of the hadrons. To replicate the observed peak-like structure in the total cross-section, potentially indicative of the $S = +1$ VB resonance ($\equiv P_0^{+*}$), we introduce a relativistic Breit-Wigner type contribution with $M_R = 2.5$ GeV and $\Gamma_R = 100$ MeV. Our numerical calculations reveal that, alongside the resonance contribution in the s channel, Reggeized π-exchange in the t channel and $\Lambda(1116)$ one in the u channel dominate the reaction process. We explore various spin-parity quantum numbers theoretically ($J^P = 1/2^\pm, 3/2^\pm$) since experimental data to determine them precisely are lacking. Notably, all tested spin-parity quantum numbers qualitatively reproduce the data well. To mitigate the signal-to-noise (S/N) ratio, we propose a specific combination of scattering angle and polarization ($\theta_{K^*} \sim 0$ and $k_K \cdot \epsilon_{K^*} \sim 0$) for future experiments, surpassing current programs like the K1.8 beam-line at J-PARC, which require a kaon beam energy of approximately 3.0 GeV to access the P_0^* region. Additionally, we recommend utilizing the recoil-proton spin asymmetry to ascertain the spin parity of the resonance.

The current study is structured as follows: Section II presents the theoretical framework. In Section III, we present the numerical results along with detailed explanations and discussions. Finally, the concluding section provides a summary and outlines future directions.

II. THEORETICAL FRAMEWORK

In this section, we will provide a concise overview of the theoretical framework governing the present reaction process. Relevant Feynman diagrams for the process are illustrated in Fig. 1, where the four momenta of the particles involved are also defined. The Yukawa interaction vertices are described as follows:

$$\begin{aligned}
\mathcal{L}_{PBB'} &= \frac{g_{PBB'}}{M_K}\bar{B}'^{\mu}\Theta_{\mu\nu}(X)(\partial^\nu K)\Gamma_5 B + \text{h.c.},\\
\mathcal{L}_{VBB'} &= -\frac{ig_{VBB'}}{M_{K^*}}\bar{B}'^{\mu}\gamma^\nu(\partial_\mu K^*_\nu - \partial_\nu K^*_\mu)\Gamma_5\gamma_5 B + \text{h.c.},\\
\mathcal{L}_{PBB} &= ig_{PBB}\bar{B}\Gamma_5\gamma_5 K^\dagger B + \text{h.c.},\\
\mathcal{L}_{VBB} &= g_{VBB}\bar{B}\gamma^\mu\Gamma_5 V^\dagger_\mu B + \text{h.c.},\\
\mathcal{L}_{PPV} &= ig_{PPV}V^\mu[P\partial_\mu P^\dagger - P^\dagger\partial_\mu P] + \text{h.c.},\\
\mathcal{L}_{PBB} &= \Big(\frac{f_{PBB}}{M_P}\Big)\bar{B}\gamma_5(\not{\partial}P)B.
\end{aligned} \tag{1}$$

In this context, B', B, P, and V denote the fields corresponding to spin-3/2, spin-1/2 baryons, pseudoscalar, and vector mesons, respectively. We utilize the notation $\Gamma_5 = (I_{4\times4}, \gamma_5)$ to represent the parity $(+,-)$ of B'. The term

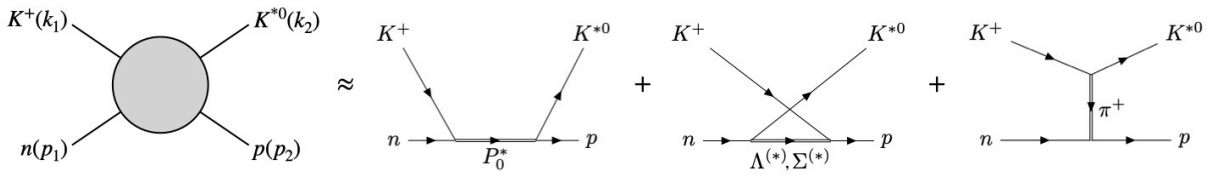


FIG. 1: Feynman diagrams pertinent to the process $K^+n \to K^{*0}p$ within the tree-level Born approximation are presented.

$\Theta_{\mu\nu}(X)$ signifies the off-shell parameter X-dependent term within the Rarita-Schwinger formalism [17].

$$\Theta_{\mu\nu}(X) = g_{\mu\nu} + X\gamma_\mu\gamma_\nu. \tag{2}$$

As illustrated in Fig. 1, our analysis encompasses the P_0^* diagram in the s channel, hyperon ($\Lambda^{(*)}$, $\Sigma^{(*)}$) exchange in the u channel, and π exchange in the t channel. Given the absence of experimentally confirmed $S = +1$ baryon states to date, we exclusively focus on the hypothetical VB pentaquark P_0^*. Concerning the t channel, while there could be additional meson exchanges, such as the axial-vector meson $b_1(1235)$, we exclude them due to the scant information available regarding their decays [11].

Following straightforward calculations, the invariant amplitudes for the Feynman diagrams depicted in Fig. 1 are presented as follows:

$$\begin{aligned}
i\mathcal{M}^s_{P_0^*(s=1/2)} &= -g_{K^*NP_0^*}g_{KNP_0^*}\bar{u}(p_2)\Gamma_5\not{\epsilon}\frac{\not{q}_s + M_{P_0^*}}{s - M^2_{P_0^*} - iM_{P_0^*}\Gamma_{P_0^*}}\gamma_5\Gamma_5 u(p_1),\\
i\mathcal{M}^s_{P_0^*(s=3/2)} &= -\frac{ig_{K^*NP_0^*}g_{KNP_0^*}}{M_{K^*}M_K}\bar{u}(p_2)\gamma_5\Gamma_5(k_{2\mu}\epsilon_\nu - k_{2\nu}\epsilon_\mu)\gamma^\nu\frac{(\not{q}_s + M_{P_0^*})\,\bar{g}^{\mu\alpha}(q_s)}{s - M^2_{P_0^*} + iM_{P_0^*}\Gamma_{P_0^*}}\big(g_{\alpha\beta} + X\gamma_\alpha\gamma_\beta\big)k_1^\beta\Gamma_5 u(p_1),\\
i\mathcal{M}^u_Y &= -\frac{i}{2}g_{KNY}g_{K^*NY}\bar{u}(p_2)\Gamma_5\frac{(\not{q}_u + M_Y)\not{\epsilon}}{u - M_Y^2 - iM_Y\Gamma_Y}\Gamma_5\gamma_5 u(p_1),\\
i\mathcal{M}^u_{\Sigma^*(1385)} &= \frac{i}{2}\frac{g_{K^*N\Sigma^*}}{M_{K^*}}\frac{g_{KN\Sigma^*}}{M_K}\bar{u}(p_2)k_1^\nu\,(g_{\mu\nu} + X\gamma_\mu\gamma_\nu)\,\frac{\bar{g}^{\mu\alpha}(q_u)\,(\not{q}_u + M_{\Sigma^*})}{u - M^2_{\Sigma^*} + iM_{\Sigma^*}\Gamma_{\Sigma^*}}\gamma^\beta\,(k_{2\alpha}\epsilon_\beta - k_{2\beta}\epsilon_\alpha)\,\gamma_5 u(p_1),\\
i\mathcal{M}^t_\pi &= -\frac{2if_{\pi NN}}{M_\pi}\frac{g_{PPV}}{M_\pi}\overline{u}(p_2)\gamma_5\frac{\not{q}_t(\epsilon\cdot k_1)}{t - M_\pi^2}u(p_1).
\end{aligned} \tag{3}$$

Here, the notation $(s, u, t) = (q_s^2, q_u^2, q_t^2)$ denotes the Mandelstam variables. In our calculations, we utilize the spin-1/2 hyperons $Y = \Lambda(1116, 1/2^+)$, $\Sigma(1192, 1/2^+)$, $\Lambda(1405, 1/2^-)$, and $\Lambda(1670, 1/2^-)$, along with the hyperon resonance $\Sigma^*(1385, 3/2^+)$. The polarization vector of K^* is represented by ϵ_μ. The spin sum of the Rarita-Schwinger field yields the following expression:

$$\bar{g}^{\mu\nu}(q) = g^{\mu\nu} - \frac{1}{3}\gamma^\mu\gamma^\nu - \frac{2}{3M^2}q^\mu q^\nu + \frac{q^\mu\gamma^\nu + q^\nu\gamma^\mu}{3M}. \tag{4}$$

Here, M and Γ represent the mass and full decay width of the particle corresponding to the field. All relevant couplings are established based on experimental data and theoretical frameworks [11, 12, 18], as summarized in Table I, except for those for P_0^*. To streamline our analysis, we denote the combined coupling as $g_{P^*0} \equiv g_{K^*NP_0^*}g_{KNP_0^*}$ hereafter.

To account for the spatial extension of the hadrons, we incorporate a phenomenological form factor, defined as follows:

$$F(x) = \frac{\Lambda^4}{\Lambda^4 + (x - M_x^2)}. \tag{5}$$

Hyperon (Y)	g_{K^*NY}	g_{KNY}	Combined
$\Lambda(1116, 1/2^+)$	-4.26	-13.4	57.08
$\Lambda^*(1405, 1/2^-)$	$0.2 - 2.7i$	$2.5 + 0.9i$	$2.93 - 6.57i$
$\Lambda^*(1670, 1/2^-)$	$-0.2 + 0.8i$	$0.2 - 0.6i$	$-0.52 - 0.04i$
$\Sigma(1192, 1/2^+)$	-2.46	4.09	-10.06
$\Sigma^*(1385, 3/2^+)$	-5.48	-6.94	38.03
π	$f_{\pi NN} = 0.989$	$g_{\pi KK^*} = 3.76$	3.72

TABLE I: Strong couplings for the hyperon and pion vertices and their combinations derived from theoretical models and experimental data [11, 12, 18].

In this expression, x represents the Mandelstam variable, and M_x denotes the off-shell mass of the corresponding particle. The cutoff parameter Λ will be adjusted to accurately reproduce the experimental data.

Given that the available data from Ref. [16] extends up to $\sqrt{s} \sim 5$ GeV, which approaches the asymptotic limit $s \to \infty$, it becomes imperative to account for the higher-spin states of the exchange particles. To address this, we incorporate the t- and u-channel Regge trajectories. Following the methodology outlined in Ref. [19], the u-channel amplitudes for the $\Lambda(1116)$, $\Sigma(1192)$, and $\Sigma^*(1385)$ are adjusted using the Regge approach, wherein each trajectory $\alpha(x)$ is characterized by the intercept α'.

$$\begin{aligned}
\tilde{\mathcal{M}}^u_{\Lambda,\Sigma}(s,u) &= C_{\Lambda,\Sigma}(u)\mathcal{M}'^u_{\Lambda,\Sigma}\left(\frac{s}{s_{\Lambda,\Sigma}}\right)^{\alpha_{\Lambda,\Sigma}(u)-\frac{1}{2}}\Gamma\left[\frac{1}{2}-\alpha_{\Lambda,\Sigma}(u)\right]\alpha'_{\Lambda},\\
\tilde{\mathcal{M}}^u_{\Sigma^*}(s,u) &= C_{\Sigma^*}(u)\mathcal{M}'^u_{\Sigma^*}\left(\frac{s}{s_{\Sigma^*}}\right)^{\alpha_{\Sigma^*}(u)-\frac{3}{2}}\Gamma\left[\frac{3}{2}-\alpha_{\Sigma^*}(u)\right]\alpha'_{\Sigma^*}.
\end{aligned} \tag{6}$$

Note that the amplitude $\mathcal{M}'$ represents the expression in Eq. (3) after removing the denominator part of the propagator. The trajectories are specified as $\alpha_\Lambda(u) = -0.65 + 0.94u$, $\alpha_\Sigma(u) = -0.79 + 0.87u$, and $\alpha_{\Sigma^*}(u) = -0.27 + 0.9u$. Additionally, the phenomenological momentum-dependent hadronic coefficient is defined as follows:

$$C_Y(u) = \left[\frac{\epsilon_Y\Lambda_Y^2}{\Lambda_Y^2 - u}\right]^2. \tag{7}$$

Finally, the t-channel Reggeized amplitude for the π reads

$$\tilde{\mathcal{M}}^t_{\pi}(s,u) = C_\pi(u)\mathcal{M}'^t_{\pi}\left(\frac{s}{s_\pi}\right)^{\alpha_\pi(t)}\frac{\pi\alpha'_\pi}{\Gamma[\alpha_X(t)+1]\sin[\pi\alpha_\pi(t)]}. \tag{8}$$

Here, $C_\pi(t) = a_\pi/(1 - t/\Lambda_\pi^2)^2$. The Regge parameters and cutoffs for the form factors are provided in Table II. It is important to mention that we adopt $X = 0$ for simplicity in our analysis.

III. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present the numerical results for the reaction process along with detailed discussions. The cutoffs for the form factors in Eq. (5) are chosen to be 1.0 GeV and 0.45 GeV for the (s,t) and u channels, respectively, by fitting the data as shown in Table II. Firstly, in panel (a) of Fig. 2, we depict the total cross-sections as functions of the center-of-mass energy (cm)($E_{\rm cm}$), showcasing the non-resonant u- and t-channel contributions separately. The lines represent contributions from the total (solid), π (dotted), $\Lambda(1116)$ (dashed), $\Lambda(1670)$ (long-dashed), $\Lambda(1670)$ (dot-dashed), $\Sigma(1192)$ (long-short-dashed), and $\Sigma^*(1385)$ (long-dot-dashed). Experimental data are obtained from both charged (circle) and neutral (square) channels [16, 20–25]. We observe that the $\Lambda(1116)$ exchange in the u-channel is crucial for reproducing the strength of the cross-section near the threshold, while the contributions from other hyperons are almost negligible, as inferred from their coupling strengths with the Regge coefficients, as depicted in Tables I and II. Conversely, the π-exchange contribution in the t channel is significant for describing the relatively higher-energy region. However, these non-resonant contributions fail to explain the peak-like structure at $E_{\rm cm} \approx 2.5$ GeV. In the figure, data points beyond the description provided by the background (BKG) are denoted with a hollow circle. Several explanations can potentially account for the observed peak-like structure in the cross-sections: 1) Single or combinations of resonances, which are the focus of our current investigation; 2) Opening of meson-baryon scattering channels [26]; 3) Interferences of scattering amplitudes [27]; and 4) Non-trivial kinematic singularities [28].

Considering the scattering of ground-state vector-meson and baryon multiplets $(8, 10, \bar{10})$ in the $S = +1$ channel, the possibility of channel opening appears less likely due to their masses, which are approximately $M_{\rm max} = M_{K^*} + M_\Delta \approx$

Hyperon (Y)	η_Y	s_Y [GeV^2]	Λ_Y [GeV]
$\Lambda(1116, 1/2^+)$	2.60	1.10	0.45
$\Sigma(1192, 1/2^+)$	0.66	1.10	0.45
$\Sigma^*(1385, 3/2^+)$	0.66	1.10	0.45
π	$\alpha_\pi = 0.53$	$s_\pi = 1.00$	$\Lambda_\pi = 1.00$

TABLE II: Parameters for the Regge approach and cutoffs in the present work [19].

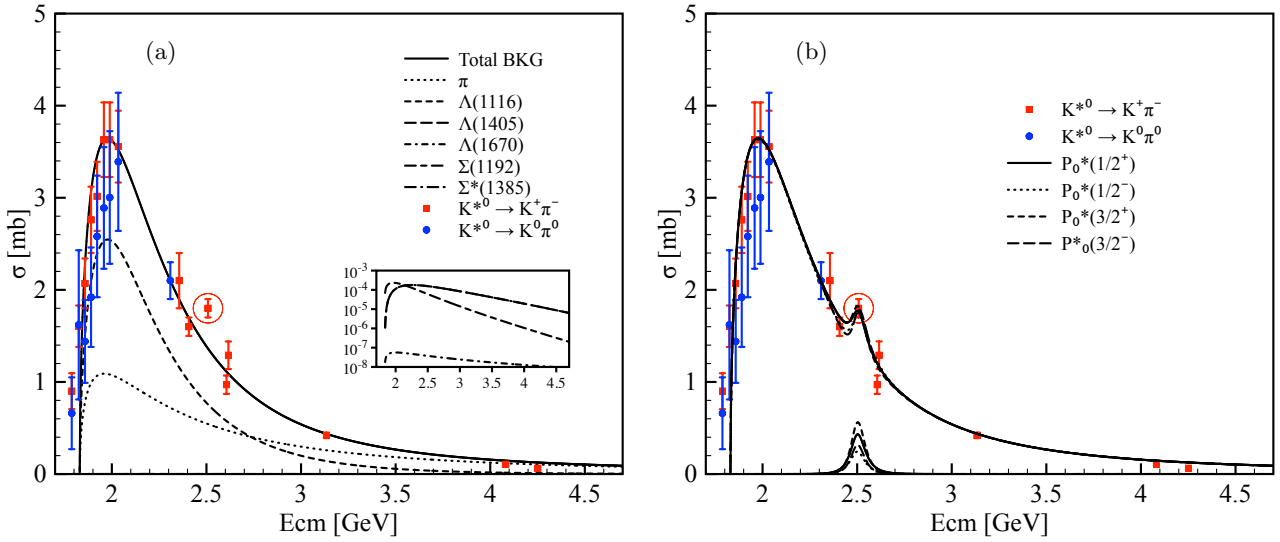


FIG. 2: (a) Total cross-sections for the non-resonant contributions as functions of $E_{\rm cm}$, serving as the background (BKG), including the total (solid), π (dotted), $\Lambda(1116)$ (dashed), $\Lambda(1670)$ (long-dashed), $\Lambda(1670)$ (dot-dashed), $\Sigma(1192)$ (long-short-dashed), and $\Sigma^*(1385)$ (long-dot-dashed) components. (b) Total cross-sections with four different spin-parity quantum numbers of the P_0^*: $J^P = 1/2^+$ (solid), $J^P = 1/2^-$ (dotted), $J^P = 3/2^+$ (dashed), and $J^P = 3/2^-$ (long-dashed). Experimental data are obtained from various sources, including Refs. [16, 20–25]. Data points that deviate from the background description are marked with hollow circles.

2.12 GeV. Although scatterings involving higher multiplets may contribute, this possibility is beyond the scope of our current investigation. We also explore the interferences between the u- and t-channel amplitudes with various phase factors, yielding no significant structures. Another potential source of interference could arise from crossing the invariant-mass bands on the Dalitz plot for $K^+n \to \pi^+K^-p$, as a function of $M_{\pi^+K^-}$ and M_{K^-p}, for instance. It's worth noting that similar considerations regarding interference possibilities have been explored theoretically and experimentally for $\gamma p \to \pi^+K^-p$, with a focus on ϕ production [27]. It was found that interference effects were negligible. Nontrivial effects, such as the triangle singularity [28], could also be introduced. For example, one might consider a triangle diagram consisting of the P-V-B internal lines in the present work, where higher-mass meson-baryon cuts can lead to singularities. However, achieving a peak at $E_{\rm cm} \approx 2.5$ GeV would require very high-mass hadrons. Nonetheless, this remains an intriguing subject for future exploration, and we leave it for subsequent studies.

Therefore, our focus lies in identifying the presence of a resonance within the observed structure. It's important to note that we explore four different spin-parity scenarios for P_0^*, namely $J^P = 1/2^\pm$, $3/2^\pm$, as there is currently no experimental data available for reference. It has been verified that all these scenarios can qualitatively reproduce the peak-like structure effectively by adjusting the combined couplings as follows:

$$g_{P_0^*}^{1/2^+} = 0.65, \quad g_{P_0^*}^{1/2^-} = 0.15, \quad g_{P_0^*}^{3/2^+} = 0.70, \quad g_{P_0^*}^{3/2^-} = 0.23. \tag{9}$$

It's worth noting that the full decay width of P_0^* remains consistent across all spin-parity quantum numbers, set at $\Gamma_{P_0^*} = 100$ MeV. In panel (b) of Fig. 2, we present the total cross-sections for each case separately: $J^P = 1/2^+$ (solid line), $J^P = 1/2^-$ (dotted line), $J^P = 3/2^+$ (dashed line), and $J^P = 3/2^-$ (long-dashed line).

In Fig. 3, the numerical results for the differential cross-sections for the angular dependence $d\sigma/d\cos\theta$ are presented for $E_{\rm cm} = (2.0, 2.5, 3.0)$ GeV in panels (a), (b), and (c), respectively, illustrating each contribution. Here, θ denotes the scattering angle of the final state K^* in the cm frame. Notably, strong forward and backward scatterings are observed from the hyperons in the u channel and from π in the t channel, respectively. Conversely, the s-channel contribution yields almost flat curves, as anticipated. Around $E_{\rm cm} = 2.5$ GeV, corresponding to the resonance region (where we use $3/2^-$), a slight yet discernible angular dependence emerges. As the energy surpasses the resonance region, the angular dependence markedly separates into forward- and backward-scattering regions. Panel (d) presents the angular dependence as a function of $E_{\rm cm}$ and $\cos\theta$. Notably, the resonance signal becomes more pronounced around $\cos\theta = 0$, where the contributions from the t and u channels are diminished or negligible. These observations suggest a potential method for experimentally isolating the resonance P_0^* contribution, which we will discuss below.

Now, we are in a position to discuss how to enhance the S/N for the P_0^* resonance. As previously illustrated, the dominant contributions in the present reaction process originate from the u- and t-channel interactions. Therefore, by focusing solely on the forward-scattering region, where the hyperon-induced backward scattering is negligible, we

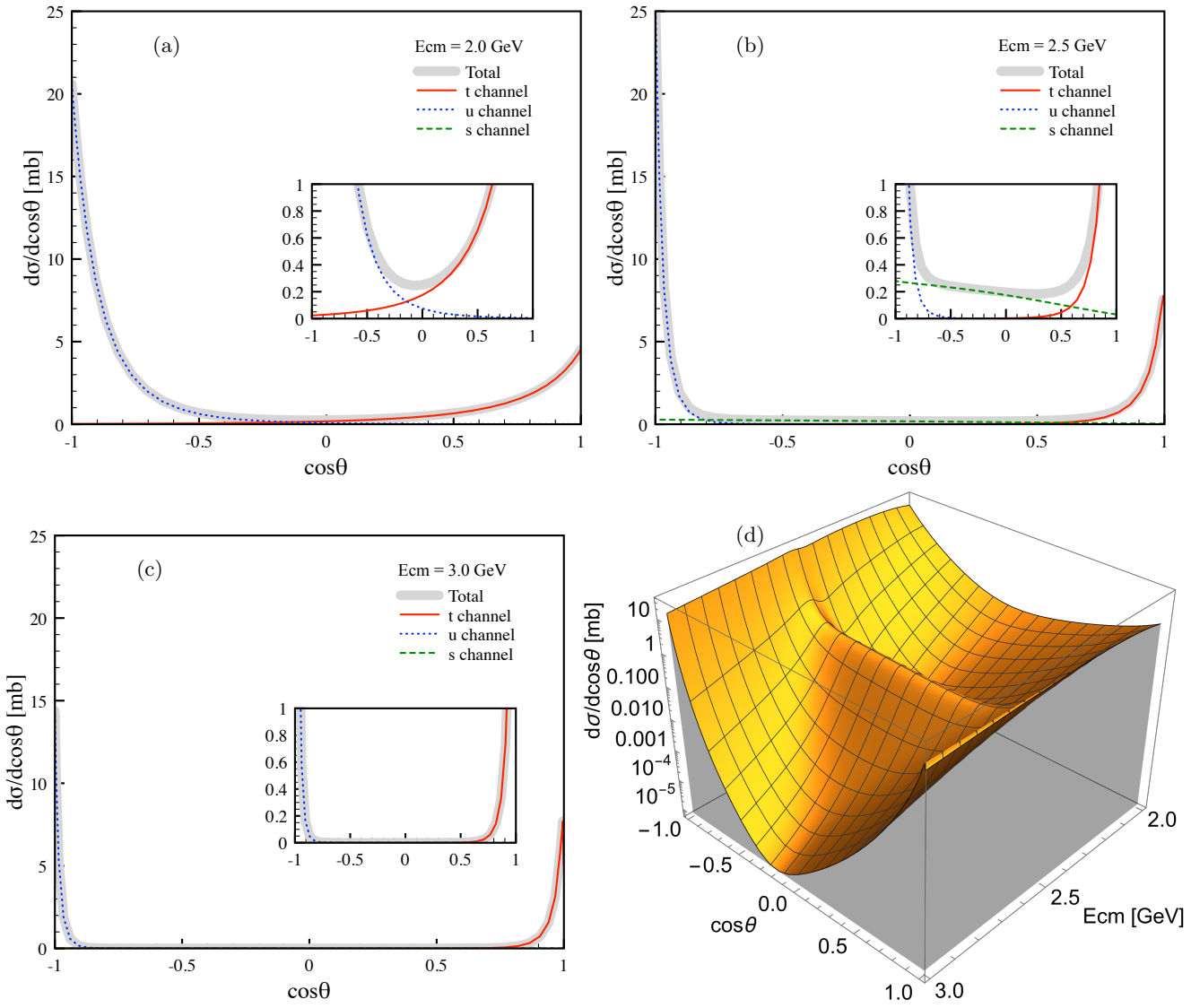


FIG. 3: The differential cross-sections for the angular dependence $d\sigma/d\cos\theta$ [mb] at $E_{\rm cm} = (2.0, 2.5, 3.0)$ GeV are depicted in panels (a), (b), and (c) as functions of $\cos\theta$, illustrating each contribution. Here, θ represents the scattering angle of the final state K^* in the cm frame. Panel (d) displays $d\sigma/d\cos\theta$ as a function of $E_{\rm cm}$ and $\cos\theta$. In this panel, we exclusively showcase the results for $P_0^*(3/2^-)$, as other quantum-number states exhibit negligible differences.

can effectively eliminate this background. The subsequent step involves reducing the contribution from the t-channel π-exchange. Remarkably, due to its vector-meson nature, the invariant amplitude for the π exchange in Eq. (3) is directly proportional to $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*}$, where $\boldsymbol{k}_K$ represents the three-momentum of the initial K^+ meson. Consequently, when the polarization of K^* is fixed perpendicular to the reaction plane, where all particle momenta are defined, the π-exchange contribution diminishes. In panel (a) of Fig. 4, we display the differential cross-sections $d\sigma/d\cos\theta$ as a function of $E_{\rm cm}$ for $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*} = 0$ at $\theta = 0$ (solid), $\pi/4$ (dotted), and $\pi/2$ (dashed) in the cm frame. Here, we specifically focus on the result for the $P_0^*(3/2^-)$, as other quantum-number states exhibit minimal variation. At $\theta = 0$, where K^* is scattered forwardly, the P_0^* signal is isolated with minimal background contamination. As the angle increases, the S/N value worsens, particularly at $\theta = \pi/4$. Further increasing it to $\theta = \pi/2$ enhances the peak due to interference between the resonance and the u-channel contribution, although the background contamination also increases. In panel (b) of Fig. 4, we plot the differential cross-section as a function of $\cos\theta$ and $E_{\rm cm}$. Across a wide range of angles ($-0.5 \lesssim \cos\theta \leq 1$), clear resonance signals are observed, facilitating their measurement in experiments. Additionally, we confirm that the *unpolarized* cross-sections exhibit minimal dependence on the spin-parity quantum numbers. Therefore, the discussions provided for Figs. 3 and 4 remain applicable to other quantum-number cases without loss of generality.

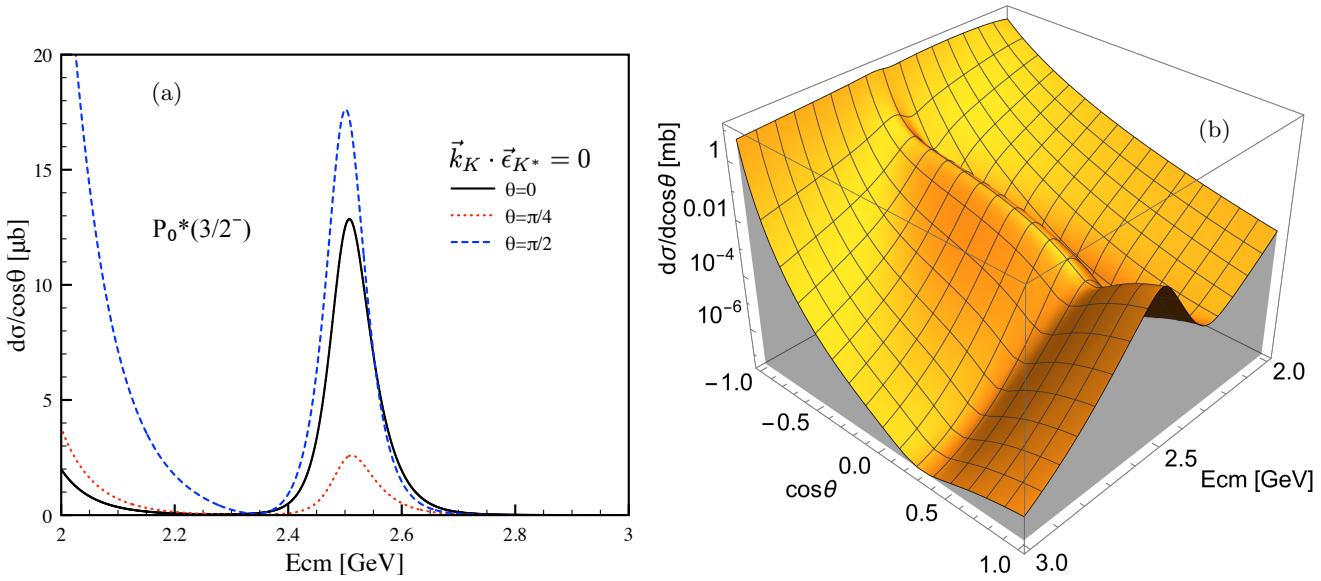


FIG. 4: (a) Differential cross-sections $d\sigma/d\cos\theta$ for $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*} = 0$ for $\theta = 0$ (solid), $\pi/4$ (dotted), and $\pi/2$ (dashed) in the cm frame. Here, we only present the results for the $P_0^*(3/2^-)$, as other quantum-number states do not significantly differ. (b) The same as (a), plotted as a function of $\cos\theta$ and $E_{\rm cm}$ for $\boldsymbol{k}K \cdot \boldsymbol{\epsilon}K^* = 0$ at $\theta = 0$.

Finally, we would like to propose a physical observable for determining the spin-parity quantum number of the P_0^* resonance. The polarizations of the involved particles, such as the target and recoil nucleons, and K^*, are invaluable for this determination. However, achieving a definite neutron polarization inside the deuteron target poses challenges in experimental setups. As discussed earlier, to maximize the S/N, it's advantageous to set the K^* polarization as $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*} = 0$ (perpendicular configuration). Consequently, we rely on controlling the recoil-proton polarization for quantum-number determination. Hence, we introduce the recoil-proton spin asymmetry (RSA) defined as follows, with $\boldsymbol{\epsilon}_{K^*}$ perpendicular to the reaction plane:

$$\Sigma_\perp \equiv \frac{d\sigma_\uparrow/d\cos\theta - d\sigma_\downarrow/d\cos\theta}{d\sigma_\uparrow/d\cos\theta + d\sigma_\downarrow/d\cos\theta}\Big|_{\boldsymbol{k}_K\cdot\boldsymbol{\epsilon}_{K^*}=0}, \tag{10}$$

where the notation $(\uparrow,\downarrow)$ denotes the (up, down)-spin polarization of the recoil proton, respectively. In panel (a) of Fig. 5, we illustrate the RSAs as functions of $\cos\theta$ at $E_{\rm cm} = 2.5$ GeV. Note that the curves mainly comprise contributions from the s-channel and u-channel, as the t-channel contribution is eliminated by fixing the polarization as $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*} = 0$. The spin-1/2 cases exhibit much smaller asymmetry compared to the spin-3/2 ones. The $P_0^*(1/2^\pm)$ amplitudes show weak angular dependence and are relatively small due to their Lorentz structures in the vicinity of $E_{\rm cm} = 2.5$ GeV as follows:

$$u^\dagger(p_2)\gamma_0\gamma_2\left(E_{\rm cm}\gamma_0 \pm M_{P_0^*}\right)\gamma_5 u(p_1) \sim A(E_{\rm cm} \pm M_{P_0^*}) + B(E_{\rm cm} \mp M_{P_0^*})(\cos\theta, \sin\theta), \quad (A,B) \in R. \tag{11}$$

On the contrary, the $P_0^*(3/2^\pm)$ amplitudes exhibit highly nontrivial angular dependencies arising from the spin-sum of the Rarita-Schwinger field and the kinematic factor $\boldsymbol{k}_K \cdot \boldsymbol{k}_{K^*} \propto t$, as shown in Eq. (3). Consequently, the spin-parity quantum numbers can be determined from $\Sigma_\perp(\cos\theta)$ for the spin-3/2 cases, whereas fixing the parity for the spin-1/2 cases proves to be more challenging. In panel (b) of Fig. 5, the numerical results for the RSA as functions of $E_{\rm cm}$ are presented. Each spin-parity case exhibits a distinctive curve shape above $E_{\rm cm} \approx 2.2$ GeV. Once more, the spin-3/2 cases demonstrate different energy dependence, whereas the spin-1/2 ones remain small and flat due to the same reasons as discussed above.

IV. SUMMARY AND FUTURE PERSPECTIVES

In our current study, we delve into the peak-like structure observed around $\sqrt{s} \sim 2.5$ GeV in the reaction $K^+n \to K^{*0}p$. We consider the potential $S = +1$ resonance $P_0^{+*} \equiv P_0^*$ as an excited state within the extended vector-meson and baryon (VB) antidecuplet, particularly in the s channel. To explore this, we employ the effective Lagrangian

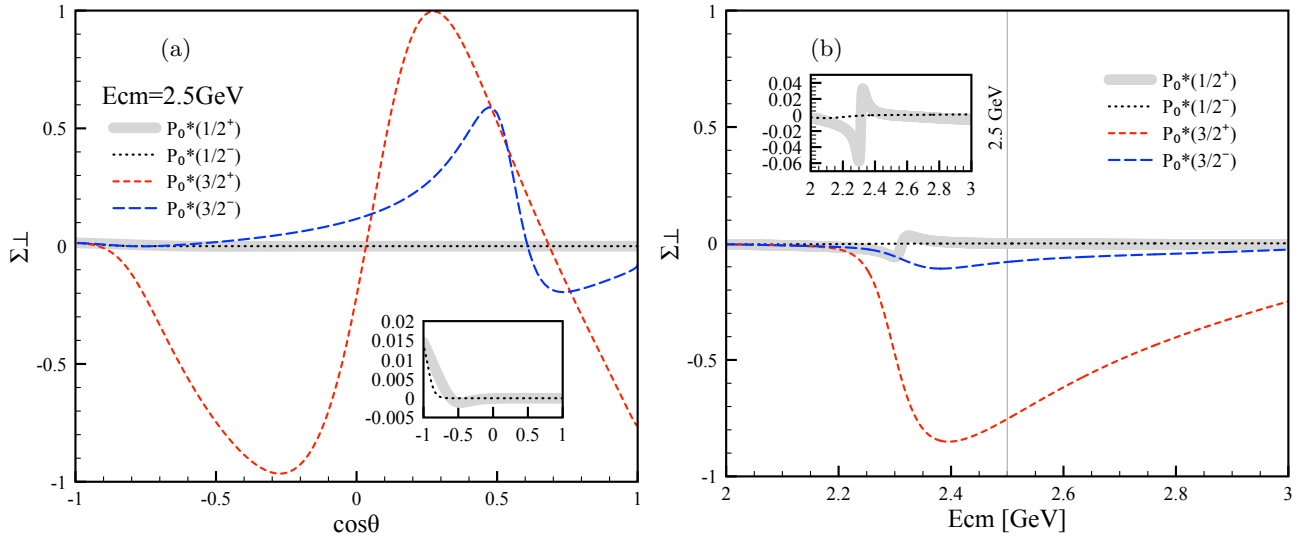


FIG. 5: (a) Recoil-proton spin asymmetry (RSA) in Eq. (10) as functions of $E_{\rm cm}$ for different spin-parity quantum numbers of P^*0, with $\boldsymbol{\epsilon}K^*$ perpendicular to the reaction plane. (b) $\Sigma_\perp$ as functions of $\cos\theta$ at $E_{\rm cm} = 2.5$ GeV following the same approach.

method in conjunction with the Regge approach at the tree-level Born approximation. We examine various hyperon (Y) exchanges in the u channel and pion (π) ones in the t channel. Below, we list important findings of the present studies:

- We find that the currently available experimental data around $E_{\rm cm} = 2.5$ GeV cannot be adequately explained solely by the tree-level Born diagrams incorporating Regge contributions as the background (BKG). However, they are well accounted for by a resonance state characterized by $M_{P_0^*} = 2.5$ GeV and $\Gamma_{P_0^*} = 100$ MeV, alongside its spin-parity quantum numbers $J^P = 1/2^\pm$ and $3/2^\pm$. To further elucidate this, we estimate the combined coupling constants for the resonance, denoted as $g_{P_0^*} \equiv g_{KNP_0^*} g_{K^*NP_0^*}$, by fitting the data, paving the way for future theoretical studies.

- The angular dependence of the reaction process, represented by $d\sigma/d\cos\theta$, is primarily influenced by the contributions from the t-channel π and u-channel $\Lambda(1116)$. These contributions result in significant enhancements of both forward and backward scattering in the center-of-mass (cm) frame, which gradually diminish as $E_{\rm cm}$ increases. However, in contrast, for all spin-parity states of P_0^*, it becomes evident that the resonance contribution is comparatively smaller than the background (BKG) and only becomes notable around Ecm $= 2.5$ GeV, particularly in the vicinity of $\cos\theta = 0$.

- To accurately study the resonance contribution, we have devised an experimental approach aimed at enhancing the signal-to-noise ratio (S/N) by taking into account the structure of the predominant background signals (BKGs). By concentrating on the forward-scattering region and orienting the polarization of the K^* perpendicular to the reaction plane, it becomes feasible to effectively circumvent the effects originating from the u channel and deactivate those from the t channel, respectively. This strategy yields a S/N value close to unity. As the scattering angle increases towards $\cos\theta \approx 0$, the contamination stemming from the u-channel becomes more pronounced, yet the S/N value remains reasonably high.

- Considering the ongoing reaction process, we delve into exploring a polarized observable known as the recoil-proton spin asymmetry (RSA) with $\boldsymbol{k}_K \cdot \boldsymbol{\epsilon}_{K^*} = 0$ to discern the spin-parity quantum numbers of the P_0^*. We discern distinctive angular dependencies of the RSAs for the $J^P = 3/2^\pm$ states. However, determining the parity for the spin-1/2 states proves to be challenging due to their weak and minimal angular dependencies at Ecm $= 2.5$ GeV. Furthermore, we analyze the energy dependencies of the RSA for prospective experiments.

We wish to underscore that the potential $S = +1$ resonance state P_0^*, proposed to account for the currently available data on $K^+n \to K^{*0}p$, holds promise for illuminating our understanding of exotic phenomena in QCD, surpassing traditional chiral interaction theories. The conjecture that P_0^* pertains to the resonance of the extended VB antidecuplet, akin to the PB antidecuplet for Θ^+, remains inadequately elucidated within existing theories. Nevertheless, ongoing discourse contemplates plausible explanations for these antidecuplet resonances [29]. Furthermore, more

realistic theoretical studies will be conducive to future experimental analyses, i.e., $K^+n \to K^{*0}p \to \pi^{0,\mp}K^{0,\pm}p$ for instance. Such endeavors will yield valuable insights into P_0^*, particularly in the K^0p invariant-mass domain. Relevant investigations are underway and will be detailed elsewhere.

Acknowledgment

The authors extend their gratitude for the fruitful discussions with Jung Keun Ahn (Korea University), Shin Hyung Kim (Kyungpook National University), Ghil-Seok Yang (Hoseo University), Kanchan Pradeepkumar Khemchandani (Federal University of São Paulo), and Alberto Martinez Torres (University of São Paulo). This work received support from the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (2018R1A5A1025563, 2022R1A2C1003964, and 2022K2A9A1A06091761).

[1] S. K. Choi *et al.* [Belle], Phys. Rev. Lett. **91**, 262001 (2003).
[2] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **110**, 222001 (2013).
[3] A. M. Sirunyan *et al.* [CMS], Phys. Rev. Lett. **128**, no.3, 032001 (2022).
[4] G. Cotugno, R. Faccini, A. D. Polosa and C. Sabelli, Phys. Rev. Lett. **104**, 132005 (2010).
[5] H. X. Chen, W. Chen, X. Liu, Y. R. Liu and S. L. Zhu, Rept. Prog. Phys. **86**, no.2, 026201 (2023).
[6] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **115**, 072001 (2015).
[7] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **131**, no.3, 031901 (2023).
[8] W. Park, S. Cho and S. H. Lee, Phys. Rev. D **99**, no.9, 094023 (2019).
[9] T. Nakano *et al.* [LEPS], Phys. Rev. C **79**, 025210 (2009).
[10] D. Diakonov, V. Petrov and M. V. Polyakov, Z. Phys. A **359**, 305-314 (1997).
[11] P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020).
[12] K. P. Khemchandani, H. Kaneko, H. Nagahiro and A. Hosaka, Phys. Rev. D **83**, 114041 (2011).
[13] S. i. Nam, Phys. Rev. D **103**, no.5, 054040 (2021).
[14] S. I. Shim, Y. Kim and S. i. Nam, [arXiv:2402.07392 [hep-ph]].
[15] Private communications with Kanchan P. Khemchandani (Federal University of São Paulo).
[16] G. Giacomelli *et al.* [Bologna-Glasgow-Rome-Trieste], Nucl. Phys. B **111**, 365-396 (1976).
[17] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[18] V. G. J. Stoks and T. A. Rijken, AIP Conf. Proc. **334**, 379-382 (1995).
[19] S. H. Kim, J. K. Ahn, S. H. Kim, S. i. Nam and M. K. Cheoun, Phys. Rev. C **107**, no.6, 065202 (2023).
[20] D. Cords *et al.*, Phys. Rev. D **4**, 1974 (1971).
[21] S. L. Baker *et al.*, Nucl. Phys. B **99**, 211-231 (1975).
[22] S. Goldhaber *et al.*, Phys. Rev. L **15**, 737 (1965).
[23] K. Buchner *et al.*, Nucl. Phys. B **45**, 333-378 (1972).
[24] G. Bassompierre *et al.*, Nucl. Phys.B **16**, 125-154 (1970).
[25] K. Hendrickx *et al.*, Nucl. Phys. B **112**, 189-212 (1976).
[26] J. K. Ahn, S. Yang and S. i. Nam, Phys. Rev. D **100**, no.3, 034027 (2019).
[27] S. Y. Ryu *et al.* [LEPS], Phys. Rev. Lett. **116**, no.23, 232001 (2016).
[28] A. P. Szczepaniak, Phys. Lett. B **747**, 410-416 (2015).
[29] G. S. Yang, S. i. Nam, in progress.

arXiv:2403.00983v1 [hep-ph] 1 Mar 2024

Heavy neutral lepton search and $\mu \to e\gamma$ constraints in case of type-I seesaw

Stefano Morisi[1,2]

[1] *Dipartimento di Fisica "Ettore Pancini", Università degli studi di Napoli "Federico II", Complesso Univ. Monte S. Angelo, I-80126 Napoli, Italy*
[2] *INFN - Sezione di Napoli, Complesso Univ. Monte S. Angelo, I-80126 Napoli, Italy*

Abstract

Within type-I seesaw mechanism it is possible to have large (order one) light-heavy neutrino mixing even in case of low right-handed neutrino mass scale (of the order of GeV). This implies large lepton flavor violation. As example we consider the process $\mu \to e\gamma$ that can have a branching up to 10^{-8} within type-I seesaw (in contrast with the tiny value 10^{-54} expected). Such an enhancing of lepton flavor violation can be used to constraint the parameter space of long lived particle experiments.

Observation of neutrino oscillation is an evidence that neutrino are massive and that flavor neutrino states do not coincide with the massive one. The unitary lepton mixing matrix U_ν connecting the two basis introduced by Pontecorvo-Maki-Nakagawa-Sakata (PMNS) has been observed esperimentally to be very different from the identity. Indeed one of the three angles parametrizing U is close to be maximal $\sin^2\theta_{23} \sim 0.5$ and one is large $\sin^2\theta_{12} \sim 0.3$ while the third is small but not zero $\sin^2\theta_{13} \sim 0.02$. Lepton mixing suggests that in the Standard Model can be present Lepton Flavor Violation (LFV) phenomena like $\mu \to e\gamma$. Early computation of this process mediated by the three active light neutrinos gives [1, 2] (see [3, 4] for a recent overview)

$$Br(\mu \to e\gamma) \approx \frac{3\alpha_e}{32\pi}\left|U^*_{\nu_{13}}U_{\nu_{23}}\frac{\Delta m^2_{31}}{M^2_W}\right|^2 \sim 10^{-54}\,, \tag{1}$$

that is very far from actual experimental sensitivity that is $7.5\cdot 10^{-13}$ [5].

If the standard model is extended by means of n right-handed neutrino N_k (where $k = 1, .., n$) having Majorana mass given by $n \times n$ mass matrix M_N, the neutrino mass matrix is a $(3+n) \times (3+n)$ matrix

$$M_\nu = \begin{pmatrix} 0 & m_D \\ m_D^T & M_N \end{pmatrix}, \tag{2}$$

where $m_D = Y_D v$ is the $3 \times n$ Dirac mass matrix (v is the standard model vev) and Y_D is the corresponding Yukawa coupling and we assume $m_D \ll M_N$. Without loss of generality we can go in the basis where M_N is diagonal. The neutrino mass matrix M_ν is diagonalized by a $(3+n) \times (3+n)$ unitary matrix V given in block form by

$$V = \begin{pmatrix} U_\nu & U_{\nu N} \\ U_{\nu N}^\dagger U_\nu & \mathbb{I} \end{pmatrix} + \mathcal{O}(\theta^2), \tag{3}$$

where $U_{\nu N} = m_D \cdot M_N^{-1}$ is a $3 \times n$ matrix that mixes light and heavy neutrino. It follows that the 3×3 lepton mixing matrix U_ν is a sub-block of the unitary matrix V and therefore there is violation of unitarity in PMNS that is typically parametrized by $\theta^2 \equiv U_{\nu N}\, U_{\nu N}^\dagger$, see for instance [6]. Block diagonalizing M_ν one obtain the well know (type-I) seesaw relation for the three light active neutrinos

$$m_\nu = -m_D \frac{1}{M_N} m_D^T. \tag{4}$$

Using this expression naively, namely assuming only one active neutrino with mass m_ν and one right-handed neutrino with mass m_N, it follows that

$$\theta^2 \sim m_\nu / m_N, \tag{5}$$

that is suppressed even for light m_N, indeed $\theta^2 \sim [10^{-10} - 10^{-25}]$ for $m_N \sim [10^{-1} - 10^{14}]$ GeV. This estimation does not really change in the $3+n$ realistic case.

In the case of type-I seesaw the branching ratio of the process $\mu \to e\gamma$ is given by

$$Br(\mu \to e\gamma) \approx \frac{3\alpha_e}{32\pi} \left| \sum_{k=4}^{n} (U_{\nu N}^*)_{1k} (U_{\nu N})_{2k} F(x_k) \right|^2, \tag{6}$$

where $x_k = m_{N_k}^2 / m_W^2$ and $F(x) = (10 - 43x + 78x^2 - 49x^3 + 18x^3 \log x + 4x^4)/(6(1-x)^4)$ [7]. Because of eq. (1), the contribution from light active neutrino is negligible, so here we

consider only heavy right-handed neutrino in the sum of (6). Therefore in case of type-I seesaw naively it is expected that $Br(\mu \to e\gamma) \sim \theta^2$ is suppressed.

Even if this suppression is true in some limit, this is not the most general result, in fact θ^2 can be (theoretically) up to $10^{-1} - 10^{-2}$ (as long as $m_D \ll M_N$ is guaranteed). For large θ^2 the branching (6) is enhanced and can be up to 10^{-8} [8] (see also [9] for an effective approach). The aim of the present paper is to update the main idea of [8] in a different language that can be useful for experiments searching for long lived particles like heavy neutral leptons, see for instance ANUBIS [10, 11], MATHUSLA [12], SHADOWS [13], NA62 [14, 15], FASER [16], CODEX-b [17].

Heavy neutral leptons (here right-handed neutrino Ns) can be produced from D, B meson decay, gauge boson W, Z, standard model Higgs H and top quark. Indeed in the minimal type-I scenario with n right handed neutrino, N_k enter in the charged and neutral current that leads to a coupling of N_k with Z and W bosons,

$$\mathcal{L} \supset -\frac{g}{\sqrt{2}} Z_\mu \, \overline{\nu_L}_{\alpha k} \gamma^\mu N_k \, (U_{\nu N})_{\alpha k} - \frac{g}{\sqrt{2}} Z_\mu \, \bar{\ell}_{\alpha k} \gamma^\mu N_k \, (U_{\nu N})_{\alpha k} \, . \tag{7}$$

Such a couplings are at the origin of both N production and decay. Then heavy neutral leptons decay quite far from the production point depending on the $U_{\nu N} \sim \theta$ mixing. Being such a mixing quite small in case of heavy neutral leptons, the lifetime can be up to $\tau_N < 0.1 s$ (this upper limit come from Big-Bang Nucleosynthesis constraints). As a consequence the decay length can be much bigger then $100\, m$ and so any detector can catch a small fraction of long-lived particle decay. For this reason all this experiment try to maximize the distance from the interaction point and the detector. Just to give an idea the distance is about $20\, m$ for ANUBIS and CODEX-b, $200\, m$ for MATHUSLA and $480\, m$ for FASER. In [18] has been shown that the dominant branching of heavy neutral lepton N is into hadrons, but decays into leptons are also possible.

The rate for production and decay of N are both proportional to $U_{\nu N}$. The mixing parameters that are typically considered in long lived experiments are

$$U_\alpha^2 = \sum_{i=1}^{3} |(U_{\nu N})_{\alpha\, i}|^2 \, , \qquad U^2 = \sum_\alpha |U_\alpha|^2 \, , \tag{8}$$

where $\alpha = e$, μ, τ. The sensitivity of heavy neutral lepton experiments is typically reported in the $(U_\alpha^2 - m_N)$ or $(U^2 - m_N)$ plane.

To understand the origin of the enhancing of θ^2 we need to go deeply into the detail of type-I seesaw mechanism. The Dirac Yukawa coupling Y_D can be parametrized in terms of the physical observable, namely the masses of the light active neutrino and the

parameters of the PMNS mixing matrix and the right handed masses by means of the Casas-Ibarra parametrization [8]

$$Y_D = v^{-1} U_{PMNS} \sqrt{m_\nu^{diag}}\, R \sqrt{M_N^{diag}}\,, \tag{9}$$

where R is an arbitrary complex $3 \times n$ orthogonal matrix. From relation (4) it is possible to fit the two square mass differences if $n \geq 2$. The minimal case with $n = 2$ predicts one massless light active neutrino. In the following for simplicity we will consider the case $n = 2$ and degenerate heavy right-handed neutrino

$$m_N \equiv (M_N)_{11} = (M_N)_{22}.$$

When $n = 2$, the matrix R is given by (for normal neutrino mass ordering considered here)

$$R = \begin{pmatrix} 0 & 0 \\ \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}, \tag{10}$$

where $\beta = x + i\,y$ is an arbitrary complex number. The value of θ^2 strongly depends on the parameter y while only mildly on the parameter x that for simplicity we assume to be $x = 0$. The parameter y can be in principle very large as soon as the seesaw regime is preserved, namely $m_D \ll M_N$. In the present analysis we take $0 < y < 30$. The fact that θ^2 is not suppressed by the neutrino mass m_ν as in eq. (5) is possible only for large values of y. If y is large enough the magnitude of the neutrino Yukawa couplings could be of order one even for $m_N \sim$ GeV. This seems to be in contradiction with common sense (5) but is a possibility. Using large value of y is therefore possible to obtain an enhancing of $Br(\mu \to e\gamma)$. Barring large y is possible in case of low-scale seesaw mechanism, for a review see [19]. A study of large lepton flavor violation coming from unitarity violation in case of low-scale seesaw is given for instance in [20].

Here we assume for simplicity as benchmark case the following choice of the parameters appearing in (9): $\sin^2\theta_{23} \sim 0.5$, $\sin^2\theta_{12} \sim 0.3$, $\sin^2\theta_{13} \sim 0.02$, $m_{\nu 1} = 0$, $m_{\nu 2} = \sqrt{\Delta m^2_{12}}$, $m_{\nu 3} = \sqrt{\Delta m^2_{13}}$ where $\Delta m^2_{12} \simeq 7 \cdot 10^{-5}\, eV$, $\Delta m^2_{13} \simeq 2 \cdot 10^{-3}\, eV$. Moreover the Dirac and Majorana phases as well as the parameter x are taken to be zero. With all these assumptions it follows that $Br(\mu \to e\gamma) \sim \theta^2$ depends only by the two free parameters y and m_N. For each set of y and m_N value chosen, the neutrino mass matrix M_ν is fixed and we obtain numerically the $\mu \to e\gamma$ branching from (6), m_D from (9) and the mixing matrix $U_{\nu N}$ and therefore also the corresponding parameter U^2_α and U^2 from (8).

We graph U^2_e as function of m_N marginalizing with respect to $Br(\mu \to e\gamma) < 7.5 \cdot 10^{-13}$ (similar graphs can be obtained for U^2_μ, U^2_τ, U^2). The result is shown in figure (1) where

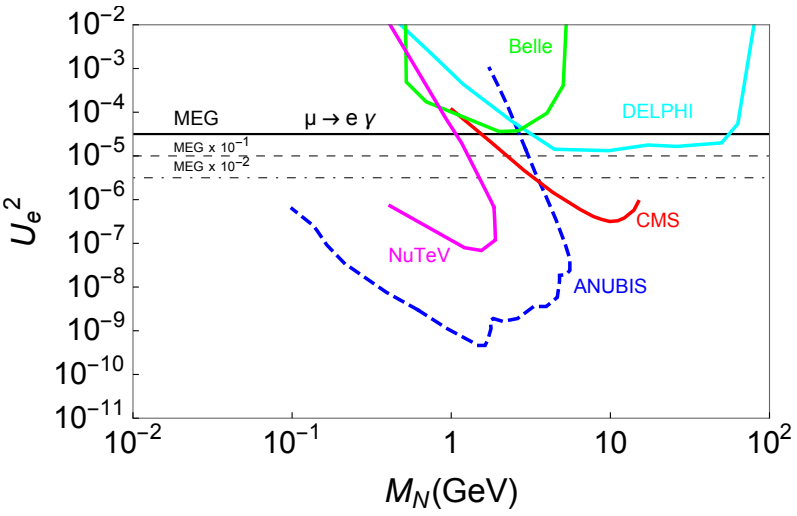


Figure 1: Light-heavy neutrino mixing parameter U_e^2 as a function of the right-handed neutrino mass m_N, the horizontal line represents MEG constraint, for details see the text.

we report with continues lines existing experimental limits (see for instance [15]). In figure we report with the dashed line in the graph the expected sensitivity of ANUBIS [21] taken as representative of long lived particle experiments. The horizontal continuous line is the limit coming from MEG. In order to better understand the role of $Br(\mu \to e\gamma)$ in the $(U_e^2,\, m_N)$ plane, we show with dashed horizontal lines the constraints coming assuming a sensitivity of MEG improved by a factor 10 and 100.

The main result of this analysis is that constraints coming from $\mu \to e\gamma$ lepton flavor violation process is in agreement with the actual constraints coming from other experiments. In particular such a limit are of the same order for masses $1\,GeV \lesssim m_N \lesssim 80\,GeV$. However above $80\,GeV$ MEG provide new limits. In principle MEG limits can be extended up to grand unified scale, but above 100 GeV the future heavy neutral leptons experiments are not sensitive. If MEG sensitivity will be improved by a factor 100, then $\mu \to e\gamma$ constraints could dominate for $m_N \gtrsim 10\,GeV$.

In summary in this analysis we provide a prove of the potentiality of lepton flavor violation in discriminating standard type-I seesaw with the interplay of long lived particle experiments.

References

[1] S. T. Petcov, Sov. J. Nucl. Phys. **25** (1977), 340 [erratum: Sov. J. Nucl. Phys. **25** (1977), 698; erratum: Yad. Fiz. **25** (1977), 1336] JINR-E2-10176.

[2] S. M. Bilenky, S. T. Petcov and B. Pontecorvo, Phys. Lett. B **67** (1977), 309 doi:10.1016/0370-2693(77)90379-3

[3] M. Lindner, M. Platscher and F. S. Queiroz, Phys. Rept. **731** (2018), 1-82 doi:10.1016/j.physrep.2017.12.001 [arXiv:1610.06587 [hep-ph]].

[4] M. Ardu and G. Pezzullo, Universe **8** (2022) no.6, 299 doi:10.3390/universe8060299 [arXiv:2204.08220 [hep-ph]].

[5] K. Afanaciev *et al.* [MEG II], [arXiv:2310.12614 [hep-ex]].

[6] S. Antusch, E. Cazzato, M. Drewes, O. Fischer, B. Garbrecht, D. Gueter and J. Klaric, JHEP **09** (2018), 124 doi:10.1007/JHEP09(2018)124 [arXiv:1710.03744 [hep-ph]].

[7] B. He, T. P. Cheng and L. F. Li, Phys. Lett. B **553** (2003), 277-283 doi:10.1016/S0370-2693(02)03258-6 [arXiv:hep-ph/0209175 [hep-ph]].

[8] J. A. Casas and A. Ibarra, Nucl. Phys. B **618** (2001), 171-204 doi:10.1016/S0550-3213(01)00475-8 [arXiv:hep-ph/0103065 [hep-ph]].

[9] S. Antusch, C. Biggio, E. Fernandez-Martinez, M. B. Gavela and J. Lopez-Pavon, JHEP **10** (2006), 084 doi:10.1088/1126-6708/2006/10/084 [arXiv:hep-ph/0607020 [hep-ph]].

[10] M. Bauer, O. Brandt, L. Lee and C. Ohm, [arXiv:1909.13022 [physics.ins-det]].

[11] A. Shah *et al.* [ANUBIS], PoS **EPS-HEP2023** (2024), 051 doi:10.22323/1.449.0051 [arXiv:2401.11604 [hep-ex]].

[12] D. Curtin, M. Drewes, M. McCullough, P. Meade, R. N. Mohapatra, J. Shelton, B. Shuve, E. Accomando, C. Alpigiani and S. Antusch, *et al.* Rept. Prog. Phys. **82** (2019) no.11, 116201 doi:10.1088/1361-6633/ab28d6 [arXiv:1806.07396 [hep-ph]].

[13] W. Baldini, A. Balla, J. Bernhard, A. Calcaterra, V. Cafaro, A. Ceccucci, V. Cicero, P. Ciambrone, H. Danielsson and G. D'Alessandro, *et al.* [arXiv:2110.08025 [hep-ex]].

[14] E. Cortina Gil *et al.* [NA62], JINST **12** (2017) no.05, P05025 doi:10.1088/1748-0221/12/05/P05025 [arXiv:1703.08501 [physics.ins-det]].

[15] M. Drewes, J. Hajer, J. Klaric and G. Lanfranchi, JHEP **07** (2018), 105 doi:10.1007/JHEP07(2018)105 [arXiv:1801.04207 [hep-ph]].

[16] A. Ariga *et al.* [FASER], Phys. Rev. D **99** (2019) no.9, 095011 doi:10.1103/PhysRevD.99.095011 [arXiv:1811.12522 [hep-ph]].

[17] V. V. Gligorov, S. Knapen, M. Papucci and D. J. Robinson, Phys. Rev. D **97** (2018) no.1, 015023 doi:10.1103/PhysRevD.97.015023 [arXiv:1708.09395 [hep-ph]].

[18] F. Kling and S. Trojanowski, Phys. Rev. D **97** (2018) no.9, 095016 doi:10.1103/PhysRevD.97.095016 [arXiv:1801.08947 [hep-ph]].

[19] S. M. Boucenna, S. Morisi and J. W. F. Valle, Adv. High Energy Phys. **2014** (2014), 831598 doi:10.1155/2014/831598 [arXiv:1404.3751 [hep-ph]].

[20] D. V. Forero, S. Morisi, M. Tortola and J. W. F. Valle, JHEP **09** (2011), 142 doi:10.1007/JHEP09(2011)142 [arXiv:1107.6009 [hep-ph]].

[21] M. Hirsch and Z. S. Wang, Phys. Rev. D **101** (2020) no.5, 055034 doi:10.1103/PhysRevD.101.055034 [arXiv:2001.04750 [hep-ph]].

An exciting hint towards the solution of the neutron lifetime puzzle?

Benjamin Koch[1, 2, *] and Felix Hummel[1]

[1] *Institut für Theoretische Physik, Technische Universität Wien, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria*

[2] *Instituto de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago, Chile*

(Dated: March 5, 2024)

We revisit the neutron lifetime puzzle, a discrepancy between beam and bottle measurements of the weak neutron decay. Since both types of measurements are realized at different times after the nuclear production of free neutrons, we argue that the existence of an excited state could be responsible for the different lifetimes. We elaborate on the required properties of such a state and construct a concrete toy model with several of the required characteristics.

arXiv:2403.00914v1 [hep-ph] 1 Mar 2024

Contents

I. Introduction 1
A. The neutron in the quark model 1
B. The lifetime puzzle 2
C. Structure of the paper 3

II. Excited states hypothesis 3

III. A model with three quarks 5
A. Spin wave functions 6
B. Spatial wave function 6
C. Weak decay 7
D. Electromagnetic transition 10

IV. Discussion 10
A. Other measurements 10
B. Parameter space 11
C. Analogies 12
D. How to test the hypothesis? 12

V. Conclusion 13

Acknowledgements 13

References 13

I. INTRODUCTION

A. The neutron in the quark model

The neutron is one of the main constituents of nuclear matter. It is a composite state, whose properties are ruled by the strong and the electroweak interactions between the lightest quarks of the Standard Model of particle physics. It is of pivotal importance in many phenomena ranging from Big Bang Nucleosynthesis to experimental particle physics [1]

Even though, a detailed understanding of the low energy properties of this particle in terms of fundamental degrees of freedom is an open field of research, it is possible to understand several properties in terms of much simpler models. In our discussion we will make use of the language and notation of the quark model. In this model, protons and neutrons are composite particles made up of quarks. Protons consist of a particular combination two "up" (u) quarks

*Electronic address: benjamin.koch@tuwien.ac.at

and one "down" (d) quark, while neutrons consist of combinations of one up quark and two down quarks. Quarks carry a fractional electric charge, and the combination of quarks in protons and neutrons results in particles with integer electric charges.

Isospin describes the similarity between protons and neutrons. It was introduced by Werner Heisenberg and later developed further by Eugene Wigner [2, 3]. Algebraically, the isospin operator $\vec{I}$, can be represented analogously to the spin operator of spin one half particles $\vec{S}$ in terms of the Pauli matrices $\sim \vec{\sigma}$, which are a representation of the group $SU(2)$. In the quark model it is imposed that both $\vec{I}$ and $\vec{S}$ are (approximate) symmetries. Thus actions of the group memebers are to be understood as symmetry transformations. Based on this, one imposes that the neutron wave function is an eigenfunction of ($\vec{I}^2$, I_z) with the eigenvalues $(3/4, -1/2)$ while the proton has the eigenvalues $(3/4, +1/2)$. Further imposing that these particles carry spin $\pm 1/2$ like their three constituents singles out a unique state for the neutron (e.g. with spin up $\uparrow$ [4])

$$|n\uparrow\rangle = \frac{1}{3\sqrt{2}} \left(|udd\rangle|dud\rangle|ddu\rangle\right) \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} |\downarrow\uparrow\uparrow\rangle \\ |\uparrow\downarrow\uparrow\rangle \\ |\uparrow\uparrow\downarrow\rangle \end{pmatrix}. \tag{1}$$

The corresponding unique state for the proton is obtained analogously by replacing $u \leftrightarrow d$. If the isospin symmetry were an exact symmetry protons and neutrons would have the same energy and thus the same mass. The observed difference between the neutron mass m_n and the proton mass m_p and the aforementioned neutron decay show that I_z must be broken. In the above description, this breaking is modeled by associating a larger mass to up quarks than to down quarks $m_u < m_d$. This generates a mass splitting between $|n\uparrow\rangle$ and $p\uparrow\rangle$. Also the conservation of spin is only an approximate concept, due to the presence of gluons, virtual quark antiquark pairs, and the respective angular momenta. Interestingly, the presence of these virtual particles can be effectively described with constituent quarks, which are gauge invariant "dressed" color neutral states [5]. This, and other, more sophisticated models of the neutron, in combination with experimental efforts allowed to learn more and more details about the properties such as mass, composition, and lifetime [1]. Thus, it is fair to say that quark models have proven to be useful for our understanding of the internal structure of hadrons [6].

Since we will exemplify our ideas with a model which has also three quarks, a word of caution is in place here. There are several phenomenological aspects which can hardly be captured even by our most sophisticated theoretical models, not to speak by simple three-quark models.

- Mass:
  The simple quark model fails when one attempts to predict the precise mass of hadrons [4].

- Spin:
  The structure of the spin distribution inside of a neutron wave-function is largely unknown (for a review see [7]). However, there is certain agreement that the valence quarks only carry a fraction of the neutron spin, which was at times labeled as "spin crisis" [8–10].

- Radius:
  The radius of the proton has been deduced from complementary measurements and the resulting 4% disagreement became widely known as the "proton radius puzzle" (for a review see [11]).

- Neighbors:
  The inner structure of protons and neutrons seems to be strongly sensitive to the neighboring hadrons within a nucleus. This puzzling behaviour is known as EMC effect [12].

- Lifetime:
  Last but not least, there is a tension between different measurements of the neutron's lifetime (see e.g. [13, 14]).

Thus, any three-quark model should be understood as useful tool for qualitative understanding rather than for precise quantitative modelling.

The last item on the above list will be the subject of interest of this paper. It will therefore be summarized in the following subsection.

B. The lifetime puzzle

In this study, we will focus on a discrepancy in the lifetime of the neutron, known as the Neutron Lifetime Puzzle (NλP):

Beam neutrons have an about 10 sec longer lifetime than Ultra Cold Neutrons (UCN) in bottle traps [13, 15–19]. The corresponding normalized difference for UCNs in gravitational bottle traps and average beam results is

$$\Delta\tau = \tau_{\rm beam} - \tau_{\rm bottle} = (8.6 \pm 2.2)\,{\rm s}. \tag{2}$$

One expects this quantity to be compatible with zero and the fact that it is significantly different from zero is known as the NλP. This puzzle has persisted over the years, and recently, a possible explanation due to exotic decay channels was excluded [31, 49]. An alternative theoretical conjecture is based on neutron oscillations [20–23]. Also this possibility has, been largely constrained [24]. Further, conjectures include Kaluza–Klein states [25], modifications of the CKM matrix induced at the TeV scale [26], and dark photons [27]. Also concerns about the systematic error in beam experiments have been raised [28], but a contradicting response was given in [29]. Thus, the status of how to interpret, or understand the NλP (2) is still inconclusive. What makes this problem even more interesting is the fact that it is not only a pressing problem for theoretical model building, it is a disagreement between two complementary types of experiments. Clearly, it would be helpful to have more results with beam experiments, but meanwhile we opt to trust the results of our experimental colleagues. Thus, a solution of this discrepancy is urgently needed.

C. Structure of the paper

In the following section II we present our working hypothesis and formulate necessary conditions for a solution of the NλP in terms of an excited state. In section III we provide a toy model that is able to fulfill most of these necessary conditions. In section IV we explore the parameter space of the toy model and discuss observational issues and known analogies in physics. Conclusions are given in section V.

II. EXCITED STATES HYPOTHESIS

Beam neutrons and UCNs are very similar with respect to most important characteristics. There are, however, some interesting differences. First of all there is the difference in velocity. While neutron beams operate typically with velocity at the order of $\mathcal{O}(10^3\,{\rm m/s})$, neutrons in bottles have velocities of $\mathcal{O}(10^0\,{\rm m/s})$. Another distinction is time. Beam neutrons are measured very shortly after their production in the reactor. The time scale for this is of the order of $t_{\rm beam} \sim 25ms$. Bottle neutrons, instead, have to undergo a process of slowing, orienting, and cleaning, which means that their weak decay is measured a long time after their production $t_{\rm bottle} > 300\,{\rm s}$ [17].

This difference is the motivation for our proposal. Let us assume that the neutron wave functions can have excited states, which are not present for neutrons within a strongly bound hadronic ensemble. For example, let us further assume that there is a ground state Ψ_g and an excited state Ψ_e with corresponding the lifetimes under beta decay of

$$\begin{aligned} \tau_g &= 878\,{\rm s}, \\ \tau_e \equiv \tau_g + |\Delta\tau_e| &\geq \tau_g + |\Delta\tau| \end{aligned} \tag{3}$$

The states neutron states shall be connected by an electromagnetic channel with half-life time τ_γ

$$\Psi_e \;\rightarrow\; \Psi_g + \gamma. \tag{4}$$

The characteristic times for the transitions shall be

$$t_{\rm beam}(\approx 2\times 10^{-2}\,{\rm s}) \ll \tau_\gamma < t_{\rm bottle} < \tau_g(=878\,{\rm s}) < \tau_e(=888\,{\rm s}). \tag{5}$$

Then the change rate of the corresponding particle numbers in these states is

$$\begin{aligned} \dot{n}_e &= -n_e\left(\frac{1}{\tau_\gamma}+\frac{1}{\tau_e}\right), \\ \dot{n}_g &= n_e\left(\frac{1}{\tau_\gamma}\right) - n_g\left(\frac{1}{\tau_g}\right). \end{aligned} \tag{6}$$

Since the two states Ψ_e, Ψ_g are, from the experimental perspective, indistinguishable, both would be counted together

$$n_n = n_g + n_e, \tag{7}$$

and the lifetime would be extracted from

$$\tau_n = -\frac{n_n}{\dot{n}_n}. \tag{8}$$

This observable is plotted in Figure 1, where we solved the equations (6) numerically for $\tau_\gamma = 4\,\mathrm{s}$ and $n_e(0) = 1$, $n_g(0) = 0$. One sees clearly that such a constellation would offer a potential solution for the lifetime puzzle. Apart

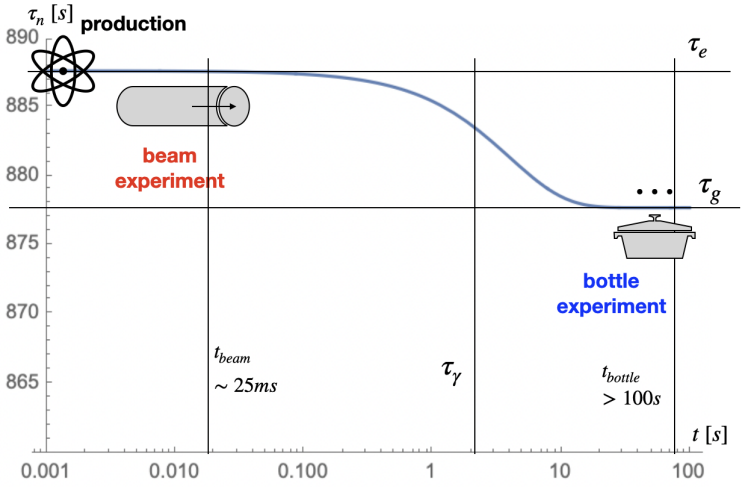


FIG. 1: Effective lifetime (8) as a function of the time for an exemplary value of $\tau_\gamma = 4\,\mathrm{s}$.

from the numerical solution, the time hierarchy (5) also allows to get an analytical understanding of how the two different lifetimes for beam and bottle experiments come about. For this we define

$$\Delta\tau_e = \tau_e - \tau_g. \tag{9}$$

The lifetime puzzle arises from count rates at two different sites:

beam) Since $t_{\mathrm{beam}} \ll$ all other times, one can identify the beam abundances as the initial abundances $n_i(0)$. Imposing that $\tau_n(0) \equiv \tau_g + |\Delta\tau|$ and using relation (8) one can solve for the initial relative abundances that mimic this value

$$\frac{n_e(0)}{n_g(0)} \approx \frac{\Delta\tau}{\Delta\tau_e - \Delta\tau} + \mathcal{O}\left(\frac{\Delta\tau}{\tau_g}\right)^2. \tag{10}$$

From this relation one realizes that $\Delta\tau_e \geq \Delta\tau$ and that for the extremal case $\Delta\tau_e \to \Delta\tau$ all initial states need to be excited $n_g(0) \to 0$.

bottle) Since, $t_{\mathrm{bottle}} \gg$ "all other time-scales", all remaining neutrons are in the ground state. Thus,

$$\tau_n(t_{\mathrm{bottle}}) = \tau_g. \tag{11}$$

Let us briefly summarize the necessary ingredients that we identified to make the proposal work:

a) There is at least one excited state of the free neutron Ψ_e above the ground state Ψ_g.

b) The lifetime of the excited state under beta decay is larger than, or equal to, the lifetime of the ground state plus the absolute value of the NλP.

c) The lifetime for the electromagnetic transition $\Psi_e \to \Psi_g$ lies between the time-scales of beam and bottle experiments ($t_{\mathrm{beam}} < \tau_\gamma < t_{\mathrm{bottle}}$). This hierarchy (5) of lifetimes is sketched in Figure (2).

d) The energy difference between Ψ_e and Ψ_g is smaller than $\sim 10\,\mathrm{MeV}$, such that they can be excited with the initial kinetic energy available from nuclear fission.

e) When a free neutron is created from a nuclear reaction, the initial abundancy of the excited state is (10).

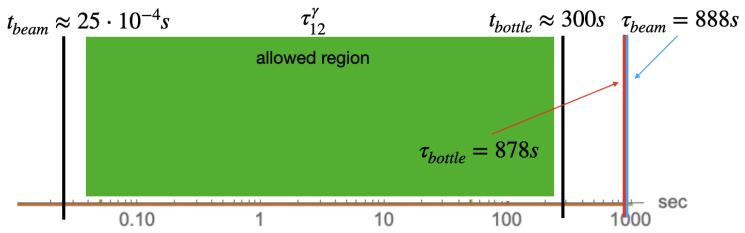


FIG. 2: a

f) The state Ψ_e is not populated if the neutron is strongly bound within a nucleus.

As add on to condition c) one should also consider other observables from different experiments. This topic will be touched in section IV. If these conditions are realized, they imply a preferred decay cascade for the initially produced state Ψ_e. This cascade is indicated by Figure 3.

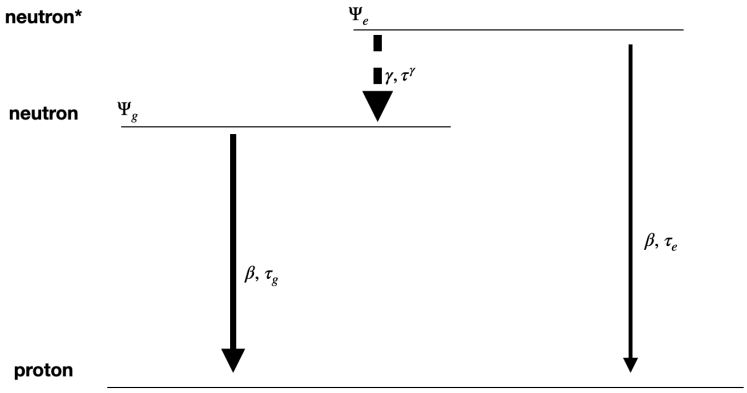


FIG. 3: Schematic energy levels of the cascade. The thickness of the arrows indicates the increased probability of certain channels.

The items (a–f) are necessary conditions for an alternative solution of the NλP. How could this be realized? Differences in lifetimes under beta decay of states with identical couplings at the quark level can be conceived in two different ways. The first possibility is an enhanced phase space. The decay of an excited state with larger energy has a larger phase space, than the decay of a ground state, which can result in a reduced lifetime. The second possibility are selection rules. If the quantum numbers of the excited state are such that a beta decay is (partially) forbidden or disfavored. While the former option would worsen the NλP, the latter option is worth to be explored further. For this, we will recur to a simple toy model, where a neutron is built of three quarks udd, which are assumed to carry it's total spin, while neglecting the contributions of internal angular momentuma and virtual particles.

III. A MODEL WITH THREE QUARKS

The purpose of this section is to give an idea of how the above hypothesis can be implemented in terms of a simple but concrete model. Let us assume that a neutron consists of three constituent quarks udd. The two down quarks are assumed to be distinguishable from the up quark but indistinguishable among themselves. In the quark model, presented in the introduction, isospin symmetry is broken since $m_u > m_d$. This breaking is, however, only responsible for the explanation of the neutron as an excited state of the proton. It is not sufficient to implement the hypothesis of an excited neutron state far below the Roper resonance (concerning this resonance in the quark model e.g. [6]). Thus, we need to consider an effect, which splits the energy of the free neutron into non-equivalent states. It can not be related to isospin, or strong interaction, because these energies would be too large. Further, the gauge dependent degrees of freedom such as color do not appear in the constituent quark picture [5]. Instead of strong color dynamics, we assume a low spin-orbit coupling, such that not only total angular momentum, but also spin symmetry is conserved. Thus, the wave functions of the neutron in either state $\Psi = \Phi\Xi$ is a product of a spatial part $\Phi(\vec{r}_{d,1}, \vec{r}_{d,2}, \vec{r}_u)$ and

a spin part $\Xi_{\sigma_{d,1},\sigma_{d,2},\sigma_u}$, where $\vec{r}_{d,1}, \vec{r}_{d,2}, \vec{r}_u$ and $\sigma_{d,1}, \sigma_{d,2}, \sigma_u$ denote the spatial and the spin coordinates of the ddu quarks, respectively, in the restframe of the center of gravity. As mentioned, color and isospin wave functions are not considered, because they operate at different energy scales. The total wave function must be anti-symmetric under exchange of the two d-quark coordinates

$$\Phi(\vec{r}_{d,1}, \vec{r}_{d,2}, \vec{r}_u)\Xi_{\sigma_{d,1},\sigma_{d,2},\sigma_u} = -\Phi(\vec{r}_{d,2}, \vec{r}_{d,1}, \vec{r}_u)\Xi_{\sigma_{d,2},\sigma_{d,1},\sigma_u}. \tag{12}$$

We suppose that the total spin of the neutron is 1/2 in accordance with the measured spin of neutrons at all times. Excited states of the neutron with a total spin 3/2 are known as delta baryons, but these are far heavier than the neutron itself [1].

A. Spin wave functions

We start considering the possible spin parts Ξ of the model wave functions of the neutron and later turn to the respective spatial parts. First, the d-spins can be combined to a total spin of the d-quarks S_d and the corresponding z-projection $S_{d,z}$. We label the spin states with a quantum number for total spin and a quantum number for the $\hat{z}$ projection $|S_d, S_{d,z}\rangle_d$. There are three symmetric d-triplet states

$$|1,+1\rangle_d = |\uparrow\uparrow\rangle_d, \qquad |1,0\rangle_d = \frac{1}{\sqrt{2}}\left(|\uparrow\downarrow\rangle_d + |\downarrow\uparrow\rangle_d\right), \qquad |1,-1\rangle_d = |\downarrow\downarrow\rangle_d, \tag{13}$$

and one anti-symmetric d-singlet state

$$|0,0\rangle_d = \frac{1}{\sqrt{2}}\left(|\uparrow\downarrow\rangle_d - |\downarrow\uparrow\rangle_d\right). \tag{14}$$

There are four linearly independent combinations of these states with the $|\downarrow\rangle_u$, $|\uparrow\rangle_u$ spin states of the u-quark that are eigenstate of $\hat{S}^2$ with a total spin S of 1/2. Using the notation $|S, S_z, S_d\rangle$, the two possibilities having $S_z = +1/2$ are

$$\Xi_1 = \left|\frac{1}{2}, \frac{1}{2}, 1\right\rangle = \sqrt{\frac{2}{3}}|1,1\rangle_d|\downarrow\rangle_u - \frac{1}{\sqrt{3}}|1,0\rangle_d|\uparrow\rangle_u, \tag{15}$$

$$\Xi_2 = \left|\frac{1}{2}, \frac{1}{2}, 0\right\rangle = |0,0\rangle_d|\uparrow\rangle_u. \tag{16}$$

Whether Ξ_1 or Ξ_2 belongs to the proposed excited state of the neutron, should it exist, remains to be determined from the corresponding spatial part. Note that these states can also be represented in the product space of the three spinor indices as follows

$$\Xi_1 = \frac{1}{\sqrt{6}}\big(2|\uparrow\uparrow\rangle_d|\downarrow\rangle_u - |\uparrow\downarrow\rangle_d|\uparrow\rangle_u - |\downarrow\uparrow\rangle_d|\uparrow\rangle_u\big), \tag{17}$$

$$\Xi_2 = \frac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle_d|\uparrow\rangle_u - |\downarrow\uparrow\rangle_d|\uparrow\rangle_u\big). \tag{18}$$

This is the spin-part of the wave-functions, which is completed by a multiplicative spatial part.

B. Spatial wave function

We do not attempt to fully model the spatial part of the wave function. Quantitative models are only known for heavy mesons, such as charmonium or bottomonium. Here, we only want to assess the necessary conditions for the hypothesized excited state of the neutron. In the simplest approximation, the wave functions are products of one-body wave functions denoted by $|ij\rangle_d|k\rangle_u = \phi_{d,i}(r_{d,1})\phi_{d,j}(r_{d,2})\phi_{u,k}(r_u)$ where $\phi_{d,i}$ and $\phi_{u,k}$ are the spatial wave functions of the down and the up quarks, respectively. The functions $\phi_{d,i}$ are mutually orthogonal to each other and ordered in the one-body energy. From the antisymmetry condition condition (12) we see that Φ_1 has to be antisymmetric under the exchange of two d-quark positions. Analogously, the spatial part of Φ_2 has to be symmetric under this exchange. The permissible product wave functions with lowest one-body energies are

$$|\Phi_1\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle_d|0\rangle_u - |10\rangle_d|0\rangle_u\big), \qquad |\Phi_2\rangle = |00\rangle_d|0\rangle_u. \tag{19}$$

In an harmonic-oscillator-like constraining potential $\phi_{d,0}$ has zero nodes, while $\phi_{d,1}$ has a nodal plane. The kinetic energy $|01\rangle_d$ would be about double the kinetic energy of the spatially symmetric wave function $|00\rangle_d$ if the constraining potential were to remain fixed. If, on the other hand, the constraining potential is allowed to become anisotropic in that situation, the kinetic energy of the wave function Φ_1 can be lowered while still being greater than Φ_2 by softening the potential in the direction perpendicular to the nodal plane while hardening it in other two directions. For a neutron inside a nucleus this is unlikely due to the constrained space. However, for a free neutron this may be conceivable.

Let us now consider the electrodynamic interaction $\hat{V}_e$ between the quarks. Assuming that $\hat{V}_e$ does not act on the spin degrees of freedom, the first-order contributions for the two states are

$$V_1^{(1)} = \langle \Phi_1 \Xi_1 | \hat{V}_e | \Phi_1 \Xi_1 \rangle = \langle 01|_d \langle 0|_u \hat{V}_e |01\rangle_d |0\rangle_u - \langle 10|_d \langle 0|_u \hat{V}_e |01\rangle_d |0\rangle_u \tag{20}$$

$$V_2^{(1)} = \langle \Phi_2 \Xi_2 | \hat{V}_e | \Phi_2 \Xi_2 \rangle = \langle 00|_d \langle 0|_u \hat{V}_e |00\rangle_d |0\rangle_u \tag{21}$$

We expect the term $\langle 01|_d \hat{V}_e |01\rangle_d$ to be positive for the two d-quarks since they have the same charge. By the same argument we expect $-\langle 10|_d \hat{V}_e |01\rangle_d$ to be negative, albeit smaller in magnitude. Furthermore, two d-quarks in the $|01\rangle_d$ state are spatially more separated than in the $|00\rangle_d$ state, such that we expect the following ordering of the interaction energy contributions

$$-\langle 10|_d \hat{V}_e |01\rangle_d < 0 < \langle 01|_d \hat{V}_e |01\rangle_d < \langle 00|_d \hat{V}_e |00\rangle_d. \tag{22}$$

Assuming the electrodynamic interaction between the up quark and one of down quarks is of similar magnitude, the electrodynamic interaction orders $V_1^{(1)} < V_2^{(1)}$. This provides a further means of counteracting the larger kinetic energy of Φ_1, such that the energy of Φ_1 and Φ_2 may split in the desired energy range. Note that an analogous phenomenon causes ferromagnetism in solid state physics.

This is, of course, a very crude model. However, it serves the purpose of exemplifying how the necessary ingredient d) (a low energy splitting between the states Ψ_1 and Ψ_2) can be produced by weak ferromagnetic effects. With this we write

$$\Delta E = \xi_E^2 \mu_d^2 \frac{\mu_0}{\pi r_0^3}\,. \tag{23}$$

Here, ξ_E is a dimensionless phenomenological parameter for the energy splitting. Just to compare this energy with the typical magnetic dipole-dipole interaction, we chose the dimensionful parameters in terms of the magnetic moment μ_d of a down quark and in terms of the neutron radius r_0. Thus, even though (23) is expressed as ξ_E-times a typical dipole-dipole energy, its origin lies in ferromagnetic behaviour.

C. Weak decay

The weak decay of the neutron is mediated by a W^- boson with spin one. This boson couples to the down quarks, which have their respective spin. Using form factors for the neutron and the proton wave function (ψ_n, ψ_p), the main part of this coupling is effectively described in terms of the Lagrangian [34]

$$\mathcal{L} = -\frac{G_F}{\sqrt{2}} W_\mu \bar{\psi}_p \gamma^\mu (g_V + g_A \gamma^5) \psi_n + \mathcal{O}\left(\dots\right)_{WM,S}\,, \tag{24}$$

where G_F is the Fermi coupling, W_μ is a weak current vector, g_V is the vector coupling, g_A axial coupling with the ratio $\lambda = g_A/g_V = -1.275$. In this estimate we further omitted the subleading higher order corrections, induced weak magnetism (WM) between the outgoing states, and the typically vanishing induced scalar terms (S) (for a review see e.g. [35, 36]). Since our states Ψ_g and Ψ_e differ in the spin structure of their down quarks, we want to explore how likely it is that the weak current couples to an $|\uparrow\rangle_d$ or a $|\downarrow\rangle_d$, given the condition that the neutron is initially in an $|\uparrow\rangle_n$ state. The discussion is done in the rest frame of the neutron and the orbital angular momentum is negligible. This means that conservation of total angular momentum implies the conservation of total spin. Thus, if the proton in the final state has an opposite spin direction $|\downarrow\rangle_p$ as the initial neutron $|\uparrow\rangle_n$, the weak current was a state with total spin one and also its spin projected to the $\hat{z}$ direction with the value one $W(1,1)$. Similarly, if the proton has the same spin direction $|\uparrow\rangle_p$ as the neutron, then the weak current was a state with spin one and with a $\hat{z}$ projection with the value zero $W(1,0)$. The final states are different. This means that there are no interference terms between the amplitudes of $W(1,0)$ and $W(1,1)$ processes. A $W(1,-1)$ configuration of the weak current is not possible in this

case, because it would leave the proton with a spin larger than 1/2. The ratio of the $W(1,0)$ and $W(1,1)$ channels can be obtained from the ratio of the corresponding partial decay widths, which reads for the ground state Ψ_g

$$r \equiv \frac{\Gamma_1^{W(1,0)}}{\Gamma_1^{W(1,1)}}. \tag{25}$$

The sum of these partial widths gives the total width under weak decays of the neutron state Ψ_g and analogously the state Ψ_2

$$\frac{1}{\tau_{n,1}} = \Gamma_1 = \Gamma_1^{W(1,0)} + \Gamma_1^{W(1,1)}, \tag{26}$$

$$\frac{1}{\tau_{n,2}} = \Gamma_2 = \Gamma_2^{W(1,0)} + \Gamma_2^{W(1,1)}. \tag{27}$$

We want to estimate how the width Γ_1 behaves in comparison with Γ_2. For this, some intermediate steps are necessary. First we need to specify the spinor wave function. The spinor of the initial state neutron is taken without loss of generality as

$$\psi_n^{\uparrow} = \psi_0(1,0,0,0)^T. \tag{28}$$

After the decay, the proton will carry a momentum $\vec{p}_p = (p_x, p_y, p_z)$. The weak current will be proportional to the 4-momentum $W_\mu \sim P_\mu^W$ and the momentum of the proton $\vec{p}_p$ will be in opposite direction of this weak current vector

$$P_\mu^W = (p_0^W, -\vec{p}_p). \tag{29}$$

The Dirac spinor of the final proton can be either in an $|\uparrow\rangle_p$ state

$$\psi_p^{\uparrow} = \psi_0 \begin{pmatrix} 1 \\ 0 \\ p_z/(E_p + m_p) \\ (p_x + ip_y)/(e_p + m_p) \end{pmatrix}, \tag{30}$$

or in a $|\downarrow\rangle_p$ state

$$\psi_p^{\downarrow} = \psi_0 \begin{pmatrix} 0 \\ 1 \\ (p_x - ip_y)/(e_p + m_p) \\ -p_z/(E_p + m_p) \end{pmatrix}. \tag{31}$$

These wave functions can be used in the definitions of the widths

$$\Gamma_1^{W(1,0)} = C_1 \int d\Omega_p d|p_p| \left| \bar{\psi}_p^{\uparrow} P_\mu^W \gamma^\mu (g_V + g_A\gamma^5) \psi_n^{\uparrow} \right|^2, \tag{32}$$

$$\Gamma_1^{W(1,1)} = C_1 \int d\Omega_p d|p_p| \left| \bar{\psi}_p^{\downarrow} P_\mu^W \gamma^\mu (g_V + g_A\gamma^5) \psi_n^{\uparrow} \right|^2, \tag{33}$$

where C_1 is a global constant, which is the same for both channels. Thus, the ratio (25) gives

$$r = \frac{\int d\Omega_p d|p_p| \left| \bar{\psi}_p^{\uparrow} P_\mu \gamma^\mu (g_V + g_A\gamma^5) \psi_n^{\uparrow} \right|^2}{\int d\Omega_p d|p_p| \left| \bar{\psi}_p^{\downarrow} P_\mu \gamma^\mu (g_V + g_A\gamma^5) \psi_n^{\uparrow} \right|^2} \approx \frac{6}{\lambda^2} - \frac{3(m_n - m_p)^2}{2m_p^2 \lambda} = 3.7. \tag{34}$$

Here, we used an expansion $|\vec{p}|/c \ll m_p$. To understand what this ratio means for the down-quarks in our model we split the width into partial widths of the two spin configurations in (15)

$$\Gamma_1 = \frac{2}{3}\Gamma_{|1,1\rangle_d|\downarrow\rangle_u} + \frac{1}{3}\Gamma_{|1,0\rangle_d|\uparrow\rangle_u}. \tag{35}$$

subsequently we further split these two partial widths into their weak decay channels (either $W(1,0)$ or $W(1,1)$)

$$\Gamma_1 = \frac{2}{3}\left(\Gamma^{W(1,0)}_{|1,1\rangle_d|\downarrow\rangle_u} + \Gamma^{W(1,1)}_{|1,1\rangle_d|\downarrow\rangle_u}\right) + \frac{1}{3}\left(\Gamma^{W(1,0)}_{|1,0\rangle_d|\uparrow\rangle_u} + \Gamma^{W(1,1)}_{|1,0\rangle_d|\uparrow\rangle_u}\right). \tag{36}$$

Recombining, we can also define $\Gamma_1 = \Gamma_1^{W(1,0)} + \Gamma_1^{W(1,1)}$ where

$$\Gamma_1^{W(1,0)} \equiv \Gamma^{W(1,0)}_{|1,1\rangle_d|\downarrow\rangle_u} + \Gamma^{W(1,0)}_{|1,0\rangle_d|\uparrow\rangle_u}, \tag{37}$$

$$\Gamma_1^{W(1,1)} \equiv \Gamma^{W(1,1)}_{|1,1\rangle_d|\downarrow\rangle_u} + \Gamma^{W(1,1)}_{|1,0\rangle_d|\uparrow\rangle_u}. \tag{38}$$

From the quark content of these states we realize that the weak current in a $W(1,0)$ configuration can couple to the down quarks independent of their spin orientation

$$\Gamma^{W(1,0)}_{|1,0\rangle_d|\downarrow\rangle_u} = \Gamma^{W(1,0)}_{|1,1\rangle_d|\downarrow\rangle_u} = \Gamma_1^{W(1,0)}. \tag{39}$$

Instead, the weak current in a $W(1,1)$ configuration can only couple to a down quark if it is in an $|\downarrow\rangle_d$ state. This poses no restriction to the coupling to $|1,1\rangle_d|\downarrow\rangle_u$ states, but it reduces the coupling to the $|1,0\rangle_d|\downarrow\rangle_u$ by a factor of two

$$\Gamma^{W(1,1)}_{|1,0\rangle_d|\uparrow\rangle_u} = \frac{1}{2}\Gamma^{W(1,1)}_{|1,1\rangle_d|\downarrow\rangle_u} = \frac{3}{5}\Gamma_1^{W(1,1)}. \tag{40}$$

Now, following the same logic as in (36), we split the Γ_2 width into weak $W(1,0)$ and $W(1,1)$ contributions

$$\Gamma_2 = \left(\Gamma^{W(1,0)}_{|0,0\rangle_d|\uparrow\rangle_u} + \Gamma^{W(1,1)}_{|0,0\rangle_d|\uparrow\rangle_u}\right). \tag{41}$$

The down quark content of $|0,0\rangle_d|\uparrow\rangle_u$ and of $|1,0\rangle_d|\uparrow\rangle_u$ are both formed of sums of $|\downarrow\uparrow\rangle_d$ and $|\uparrow\downarrow\rangle_d$, with the only difference that one is symmetric and the other is antisymmetric. This sign is, however, irrelevant for the weak width, since the weak current couples identically at the quark level for both terms. Thus, we can deduce

$$\Gamma^{W(1,0)}_{|0,0\rangle_d|\uparrow\rangle_u} = \Gamma^{W(1,0)}_{|1,0\rangle_d|\uparrow\rangle_u} = \Gamma_1^{W(1,0)}, \tag{42}$$

$$\Gamma^{W(1,1)}_{|0,0\rangle_d|\uparrow\rangle_u} = \Gamma^{W(1,1)}_{|1,0\rangle_d|\uparrow\rangle_u} = \frac{3}{5}\Gamma_1^{W(1,1)},$$

where (40) was used. Inserting this into (41) and using the ratio (34) we get

$$\Gamma_2 = \Gamma_1 \frac{r+\frac{3}{5}}{r+1} \approx 0.9\,\Gamma_1. \tag{43}$$

In terms of lifetimes under weak decay this means

$$\frac{\tau_2 - \tau_1}{\tau_1} = \frac{2}{5r+3} \approx 0.09. \tag{44}$$

The lifetime of the state Ψ_2 is 9% longer than that of the state Ψ_1. In our hypothesis we needed the lifetime of the excited state to be longer than that of the the ground state. This means that we have to identify Ψ_1 with the ground state and Ψ_2 with the excited state

$$\Psi_1 \equiv \Psi_g, \qquad\qquad \Psi_2 \equiv \Psi_e. \tag{45}$$

Therefore, the exchange energy that gives rise to (23) needs to be such, that it favours the spin-parallel configuration of the down quarks over an anti-parallel configuration. Interestingly, the possibility of such ferromagnetic properties of quarks has also been discussed for nuclear matter at low density [37–40].

Interpreting (44) in terms of initial conditions using (10) we find that the initial percentage of Ψ_e states reaching the beam experiment should be

$$\frac{n_e(0)}{n_n(0)} = \frac{|\Delta\tau|}{2\tau_g}(3+5r) \approx 12\%. \tag{46}$$

Note that this discussion uses simplifying approximations which do not strongly affect the result or the conclusion. The first approximation was on the phasespace. We assumed that Γ_1 and Γ_2 have approximately the same phase-space. This is not exactly right since the phase-space of the excited state is slightly larger than the phase-space of the ground state. This correction is, however for practical purposes negligible. For example, if $\Delta E \approx 1.4 \times 10^5\,\mathrm{eV}$, the phase-space correction is of order

$$\frac{\Delta E}{(m_n - m_p)c^2} \approx 10^{-4}. \tag{47}$$

The second approximation was on the angular momentum. The result (46) was obtained by imposing conservation of spin $\vec{S}$ in the emission of W^μ. The conserved quantity is, however, the total angular momentum which consists of spin plus orbital angular momentum $\vec{L}$

$$\vec{J} = \vec{S} + \vec{L}. \tag{48}$$

The maximal change in orbital angular momentum is achieved when the W^μ is radiated with maximal momentum ($|\vec{p}^W| \leq (m_n - m_p)c$) and from a maximal radius r and orthogonal to the symmetry axis $\hat{x}$

$$|\Delta L|_{max} = r(m_n - m_p)c. \tag{49}$$

If we use $r \approx 2r_n$ and compare the result to the spin $1/2\hbar$, we see that

$$\frac{|\Delta L|_{max}}{\hbar/2} \approx 10^{-2}, \tag{50}$$

thus, relative corrections of (44) due to orbital angular momentum effects are only of the order of 1%.

D. Electromagnetic transition

The electromagnetic transition between Ψ_e and Ψ_g can be due to electromagnetic multi-pole radiation. Given the wide range of allowed lifetimes and the absence of a reliable model for the spatial wave-functions, we will be satisfied with a classical ballpark estimate. As example, let us consider magnetic quadrupole radiation. The classical rate for magnetic quadrupole radiation is

$$\frac{dP}{d\tau} = \frac{\mu_0 c k^6}{1440\pi}|Q|^2. \tag{51}$$

The square of a quadrupole moment $|Q|^2$ formed by two dipoles scales with the square of the radius r_0 and the square dipole moment of the individual dipoles. Thus, in absence of concrete spatial distributions, we parameterize this unknown property with a quadrupole distribution parameter ξ_Q by writing

$$|Q|^2 = 80\xi_Q^2 r_0^2 \mu_B^2. \tag{52}$$

Here, the multiplicative factor 80 was taken as reference from a "billard-ball" toy model.

Now, we use the energy difference between the excited state (23) and the ground state $k = \Delta E/(\hbar c)$. Identifying the integrated rate with this energy difference gives an estimate for the life-time of the electromagnetic channel

$$\tau_\gamma = \frac{18 c^5 \hbar^6 \pi}{r_0^2 \Delta E^5 \mu_0 \mu_B^2 \xi_Q^2}. \tag{53}$$

If we now use the estimate of ΔE, that was found before in (23), we get

$$\tau_\gamma = 3.6 \times 10^4 \, \frac{c^5 \hbar^6 r_0^{13}}{\mu_0^6 \mu_B^{12} \xi_E^{10} \xi_Q^2} \approx 10\,\mu\text{s}\, \frac{1}{\xi_E^{10}\xi_Q^2}. \tag{54}$$

The relevance of these relations will be explored in subsection IV B.

IV. DISCUSSION

A. Other measurements

An important fact that should be mentioned in this context is that the neutron mass is measured to extreme precision ($\pm 2\,\text{eV}$). However, this measurement involves deuterium as baryonic bound state. For this reason we formulated the hypothesis f), that in the strongly interacting environment of the baryon, the relatively weak excitation of the free neutron can not be formed. Thus, the excited state of the neutron should also not show up in mass measurements that involve baryonic bound states.

There are also experiments that measure the mass of the free neutron without recurring to baryonic bound states. These experiments have a precision of $\pm\Delta m \approx 10^4\,\mathrm{eV}/c^2$ [41, 42]. There are two possible causes that could explain why these experiments did not see the excited neutron state in their measurements. First, the experiments are performed at times t larger than τ_γ. Second, the excitation energy lies below the mass sensitivity of these experiments

$$\Delta E < \Delta E_m = \Delta m c^2 \approx 2 \times 10^4\,\mathrm{eV}. \tag{55}$$

Apart from mass measurements, there is a plethora of experiments performed with free neutrons. They have to be revisited as part of the process of testing our hypothesis. This will, be our task for upcoming projects and collaborations.

B. Parameter space

Let us now discuss the viability of the parameter space of our model. The only adjustable parameters of our model are the factors of the energy splitting ξ_E (23) and the quadrupole moment ξ_Q (52). These parameters have to be confronted with the minimal conditions, needed to make the proposal work. First and foremost, the lifetime has to be in the window shown in Figure 2. Further, the excitation energies must lie below the known precision of the free (no bound state involved) neutron mass (55). Solving the first two conditions one obtains

$$\xi_E^5 \;\geq\; \frac{190 c^{5/2}\hbar^3 r_0^{13/2}}{\mu_0^3\mu_B^6\xi_Q\sqrt{t_{bottle}}} \approx 2\cdot 10^{-4}\frac{1}{\xi_Q}, \tag{56}$$

$$\xi_E^5 \;\leq\; \frac{190 c^{5/2}\hbar^3 r_0^{13/2}}{\mu_0^3\mu_B^6\xi_Q\sqrt{t_{beam}}} \approx 0.06\frac{1}{\xi_Q}. \tag{57}$$

Further, inserting the condition (55) into relation (23) and solving for ξ_E one obtains

$$\xi_E \leq 25\left(\frac{r_0^3\Delta E_{max}}{\mu_0\mu_B^2}\right)^{5/2} \approx 2\cdot 10^{-4}. \tag{58}$$

These conditions are plotted in Figure 4. One realizes from this figure that the parameter space is quite constrained,

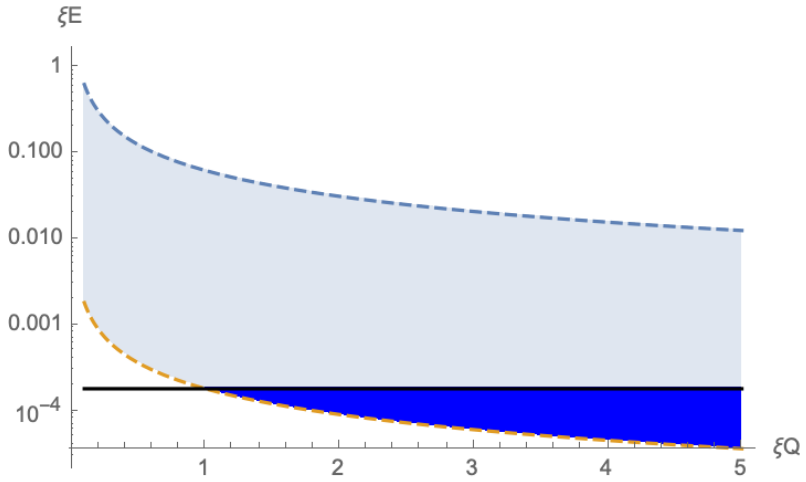


FIG. 4: Sketch of the parameter space ξ_Q and ξ_E. The bright blue region satisfies the two conditions (56 and 57), which are represented by the dashed lines (blue for 56 and orange for 57). The condition (58) is saturated at the black line, such that the dark blue region satisfies all three conditions ((56, 57, and 58). Here, $\Delta E_m = 20\,\mathrm{keV}$. The gray dotted line is for $t_{bottle} = 15\,\mathrm{s}$, which corresponds to the limiting case, when the intersection of (58) and condition (57) becomes to large to be realized by simple dipoles within a neutron radius.

giving minimal and maximal values for both ξ_E and ξ_Q. This also figure illustrates two more interesting aspects: First, the minimal value of ξ_Q of the deep blue region is of the order of one, which is surprisingly close to a naive

estimate of the quadrupole moment. Second, for reasonable values of $\xi_Q < 4$, the gray dotted curve suggests that the electromagnetic transition time is more on the bottle end of the window between the beam time and the bottle time

$$15\,\mathrm{s} < \tau_\gamma < 300\,\mathrm{s}. \tag{59}$$

Keeping in mind that we were working with a very bold model, the obtained regions for ξ_E and f_r are quite reasonable.

C. Analogies

The characteristics of an excited state that has a longer lifetime than the ground state, as proposed here for the free neutron, is quite unusual. Nevertheless there are some cases with this characteristics known in nuclear physics. Interestingly, nuclear isomers become metastable, if the spin structure between excited and ground state is largely different. For example, certain nuclear isomers, such as ^{180m}Ta are stable in contrast to their ground state [43]. Also only recently, other isomers with only tiny excitation energy have been discovered [44].

Another analogy in nuclear physics can be found in the magnetic dipole interactions of currents within atomic nuclei. In these nuclei it is also magnetic dipoles and exchange currents that are responsible for the existence of excited states of otherwise degenerate states with the same nuclear mass [45]. In this context we also want to mention the resonance that was found in polarized neutron - proton scattering [46].

D. How to test the hypothesis?

The hypothesis of an excited state in the context of the NλP can be tested directly. Such tests would involve designing experiments or observations aimed at detecting the predicted characteristics of this state. Here are several potential approaches to test the hypothesis directly:

- Perform a beam experiment at later time, which means constructing longer beam pipes with length L. Check whether the deduced decay time depends on L. This narrows the window between $t_{\mathrm{beam}} = L/v_n$ and t_{bottle} from below.
- Repeat a bottle experiment at earlier times t_{bottle}. Check whether the deduced decay time depends on t_{bottle}. This narrows the window from above. For our toy model we expect relatively large values for τ_γ (59), which speaks for this method.
- Search for electromagnetic signatures of the transition between Ψ_e and Ψ_g along beam pipes, UCN cooling facilities, or before filling UCN bottle containers. For example, correlations between γ backgrounds from neutron beams have been measured and compared to simulations to a few percent level [47]. Such backgrounds will be of crucial importance. If for instance $\tau_\gamma \approx t_{\mathrm{beam}}$ then the signature will have to compete with the background of reactor photons. Instead, if $\tau_\gamma \approx t_{\mathrm{bottle}}$ like in (59), the signal will have to compete with the background generated by secondary reactions of the actual weak decays.
- Try to re-populate the excited state with fine-tuned external radiation.
- Recalculate the angular distributions of the neutron decay products in analogy to (36 and 41) and compare to the nucleon decay parameters measured e.g. in [48, 49].

Apart from these tests, one can also look for indirect signatures in all sorts of precision experiments with free neutrons, as long as these experiments are performed at $t < t_{\mathrm{bottle}}$. Other neutron rich environments such as the early universe during Big Bang Nucleosynthesis, or neutron stars, are probably insensitive to the seeing the excited state due to the immense red-shift of the former and the huge gravitational binding energy of the latter.

One interesting test opportunity we want to mention also is the systematic 3.9σ energy shift observed in the qBounce experiment [50]. Since the energies in qBounce depend on the neutron mass, this relative shift of $\approx 10^{-3}$ could be associated to a mass shift due to an excited state.

V. CONCLUSION

In this paper we revisited the neutron lifetime puzzle, assuming that both mutual contradicting experimental results are correct. Our exploration has led us to propose a novel perspective on the NλP. The discrepancy could be explained if: The conditions (a–f) from section II are fulfilled. We then presented a three-quark toy model utilizing the spin configurations (15) and (16). Through this model, we demonstrated that the conditions a)-d) can be satisfied quite naturally, while e) and f) are posed as additional assumption.

It is crucial to note that our presented toy model is intended to serve as a starting point for further exploration rather than a definitive explanation. Our primary goal is to emphasize the possibility that the NλP may suggest the presence of an excited state with the outlined characteristics. While a substantial portion of our effort went into constructing and discussing a specific model, we encourage further investigations and alternative approaches to validate and refine our proposed hypothesis of an excited state.

In essence, this study opens a pathway for future research to delve deeper into the nature of neutron decay and the potential existence of an excited states, providing valuable insights that could contribute to resolving the Neutron Lifetime Puzzle.

Acknowledgements

B.K. was supported by the grants P 31702-N27 and P 33279-N. We thank H. Abele and M. Loewe for detailed comments. Further thanks to A. Santoni, A. Serebrov, and F. Wietfeldt for discussion.

[1] R. L. Workman *et al.* [Particle Data Group], PTEP **2022** (2022), 083C01 doi:10.1093/ptep/ptac097
[2] W. Heisenberg, Z. Phys. **77** (1932), 1-11 doi:10.1007/BF01342433
[3] E. Wigner, Phys. Rev. **51** (1937), 106-119 doi:10.1103/PhysRev.51.106
[4] W. Greiner, B. Mueller, Springer, doi:10.1007/978-3-540-87843-8.
[5] M. Lavelle and D. McMullan, Phys. Rept. **279** (1997), 1-65 doi:10.1016/S0370-1573(96)00019-1 [arXiv:hep-ph/9509344 [hep-ph]].
[6] G. Eichmann, H. Sanchis-Alepuz, R. Williams, R. Alkofer and C. S. Fischer, Prog. Part. Nucl. Phys. **91** (2016), 1-100 doi:10.1016/j.ppnp.2016.07.001 [arXiv:1606.09602 [hep-ph]].
[7] A. Deur, S. J. Brodsky and G. F. De Téramond, doi:10.1088/1361-6633/ab0b8f [arXiv:1807.05250 [hep-ph]].
[8] G. Veneziano, Mod. Phys. Lett. A **4** (1989), 1605 doi:10.1142/S0217732389001830
[9] M. Anselmino, B. L. Ioffe and E. Leader, Sov. J. Nucl. Phys. **49** (1989), 136 NSF-ITP-88-94.
[10] G. C. Nayak, JHEP **03** (2018), 101 doi:10.1007/JHEP03(2018)101 [arXiv:1802.02864 [hep-ph]].
[11] C. E. Carlson, Prog. Part. Nucl. Phys. **82** (2015), 59-77 doi:10.1016/j.ppnp.2015.01.002 [arXiv:1502.05314 [hep-ph]].
[12] D. F. Geesaman, K. Saito and A. W. Thomas, Ann. Rev. Nucl. Part. Sci. **45** (1995), 337-390 doi:10.1146/annurev.ns.45.120195.002005
[13] S. Paul, Nucl. Instrum. Meth. A **611**, 157 (2009) doi:10.1016/j.nima.2009.07.095 [arXiv:0902.0169 [hep-ex]].
[14] A. P. Serebrov and A. K. Fomin, Phys. Procedia **17** (2011), 199-205 doi:10.1016/j.phpro.2011.06.037 [arXiv:1104.4238 [nucl-ex]].
[15] F. E. Wietfeldt and G. L. Greene, Rev. Mod. Phys. **83**, no. 4, 1173 (2011). doi:10.1103/RevModPhys.83.1173
[16] A. P. Serebrov *et al.*, KnE Energ. Phys. **3**, 121 (2018). doi:10.18502/ken.v3i1.1733
[17] F. M. Gonzalez *et al.* [UCNτ], Phys. Rev. Lett. **127** (2021) no.16, 162501 doi:10.1103/PhysRevLett.127.162501 [arXiv:2106.10375 [nucl-ex]].
[18] R. W. Pattie, Jr. *et al.*, Science **360**, no. 6389, 627 (2018) doi:10.1126/science.aan8895 [arXiv:1707.01817 [nucl-ex]].
[19] V. F. Ezhov *et al.*, JETP Lett. **107**, no. 11, 671 (2018) [Pisma Zh. Eksp. Teor. Fiz. **107**, no. 11, 707 (2018)] doi:10.1134/S0021364018110024, 10.7868/S0370274X18110036 [arXiv:1412.7434 [nucl-ex]].
[20] I. Altarev, C. A. Baker, G. Ban, K. Bodek, M. Daum, P. Fierlinger, P. Geltenbort, K. Green, M. G. D. van der Grinten and E. Gutsmiedl, *et al.* Phys. Rev. D **80** (2009), 032003 doi:10.1103/PhysRevD.80.032003 [arXiv:0905.4208 [nucl-ex]].
[21] Z. Berezhiani, Eur. Phys. J. C **79**, no. 6, 484 (2019) doi:10.1140/epjc/s10052-019-6995-x [arXiv:1807.07906 [hep-ph]].
[22] Z. Berezhiani and A. Vainshtein, Phys. Lett. B **788**, 58 (2019) doi:10.1016/j.physletb.2018.11.014 [arXiv:1809.00997 [hep-ph]].
[23] W. Tan, Universe **9** (2023) no.4, 180 doi:10.3390/universe9040180 [arXiv:2302.07805 [nucl-ex]].
[24] L. J. Broussard, J. L. Barrow, L. DeBeer-Schmitt, T. Dennis, M. R. Fitzsimmons, M. J. Frost, C. E. Gilbert, F. M. Gonzalez, L. Heilbronn and E. B. Iverson, *et al.* Phys. Rev. Lett. **128** (2022) no.21, 212503 doi:10.1103/PhysRevLett.128.212503 [arXiv:2111.05543 [nucl-ex]].
[25] G. Dvali, M. Ettengruber and A. Stuhlfauth, [arXiv:2312.13278 [hep-ph]].

[26] B. Belfatto, R. Beradze and Z. Berezhiani, Eur. Phys. J. C **80** (2020) no.2, 149 doi:10.1140/epjc/s10052-020-7691-6 [arXiv:1906.02714 [hep-ph]].

[27] D. Barducci, M. Fabbrichesi and E. Gabrielli, Phys. Rev. D **98** (2018) no.3, 035049 doi:10.1103/PhysRevD.98.035049 [arXiv:1806.05678 [hep-ph]].

[28] A. P. Serebrov, M. E. Chaikovskii, G. N. Klyushnikov, O. M. Zherebtsov and A. V. Chechkin, Phys. Rev. D **103** (2021) no.7, 074010 doi:10.1103/PhysRevD.103.074010 [arXiv:2011.13272 [nucl-ex]].

[29] F. E. Wietfeldt, R. Biswas, J. Caylor, B. Crawford, M. S. Dewey, N. Fomin, G. L. Greene, C. C. Haddock, S. F. Hoogerheide and H. P. Mumm, *et al.* Phys. Rev. D **107** (2023) no.11, 118501 doi:10.1103/PhysRevD.107.118501

[30] M. Klopf, E. Jericha, B. Märkisch, H. Saul, T. Soldner and H. Abele, Phys. Rev. Lett. **122**, no. 22, 222503 (2019) doi:10.1103/PhysRevLett.122.222503 [arXiv:1905.01912 [hep-ex]].

[31] D. Dubbers, H. Saul, B. Märkisch, T. Soldner and H. Abele, Phys. Lett. B **791**, 6 (2019) doi:10.1016/j.physletb.2019.02.013 [arXiv:1812.00626 [nucl-ex]].

[32] J. I. Collar and C. M. Lewis, [arXiv:2203.00750 [physics.ins-det]].

[33] A. N. Khan, Nucl. Phys. B **986** (2023), 116064 doi:10.1016/j.nuclphysb.2022.116064 [arXiv:2201.10578 [hep-ph]].

[34] S. Weinberg, Phys. Rev. **112** (1958), 1375-1379 doi:10.1103/PhysRev.112.1375

[35] B. Fornal, Universe **9** (2023) no.10, 449 doi:10.3390/universe9100449 [arXiv:2306.11349 [hep-ph]].

[36] J. Byrne, "An overview of neutron decay," International Workshop on Quark Mixing, CKM Unitarity, 15-25, www.physi.uni-heidelberg.de/Publications/ckm_byrne.pdf

[37] T. Tatsumi, Phys. Lett. B **489** (2000), 280-286 doi:10.1016/S0370-2693(00)00927-8 [arXiv:hep-ph/9910470 [hep-ph]].

[38] T. Maruyama and T. Tatsumi, Nucl. Phys. A **693** (2001), 710-730 doi:10.1016/S0375-9474(01)00811-9 [arXiv:nucl-th/0010018 [nucl-th]].

[39] T. Tatsumi, T. Maruyama and E. Nakano, Prog. Theor. Phys. Suppl. **153** (2004), 190-197 doi:10.1143/PTPS.153.190 [arXiv:hep-ph/0312347 [hep-ph]].

[40] A. Niegawa, Prog. Theor. Phys. **113** (2005), 581-601 doi:10.1143/PTP.113.581 [arXiv:hep-ph/0404252 [hep-ph]].

[41] V. V. Nesvizhevsky, H. G. Borner, A. M. Gagarski, A. K. Petoukhov, G. A. Petrov, H. Abele, S. Baessler, G. Divkovic, F. J. Ruess and T. Stoferle, *et al.* Phys. Rev. D **67** (2003), 102002 doi:10.1103/PhysRevD.67.102002 [arXiv:hep-ph/0306198 [hep-ph]].

[42] T. Jenke, P. Geltenbort, H. Lemmel and H. Abele, Nature Phys. **7** (2011), 468-472 doi:10.1038/nphys1970

[43] B. Lehnert, M. Hult, G. Lutter and K. Zuber, Phys. Rev. C **95** (2017) no.4, 044306 doi:10.1103/PhysRevC.95.044306 [arXiv:1609.03725 [nucl-ex]].

[44] T. Sikorsky, J. Geist, D. Hengstler, S. Kempf, L. Gastaldo, C. Enss, C. Mokry, J. Runke, C. E. Düllmann and P. Wobrauschek, *et al.* Phys. Rev. Lett. **125** (2020) no.14, 142503 doi:10.1103/PhysRevLett.125.142503 [arXiv:2005.13340 [nucl-ex]].

[45] K. Heyde, P. von Neumann-Cosel and A. Richter, Rev. Mod. Phys. **82** (2010), 2365-2419 doi:10.1103/RevModPhys.82.2365 [arXiv:1004.3429 [nucl-ex]].

[46] P. Adlarson *et al.* [WASA-at-COSY], Phys. Rev. Lett. **112** (2014) no.20, 202301 doi:10.1103/PhysRevLett.112.202301 [arXiv:1402.6844 [nucl-ex]].

[47] Qiang Li et. al Nuclear Instruments and Methods in Physics Research Section A: **980** (2020) 164506 doi:10.1016/j.nima.2020.164506 .

[48] J. Reich, H. Abele, M. A. Hoffmann, S. Baessler, P. von Bulow, D. Dubbers, U. Peschke, V. Nesvizhevsky and O. Zimmer, Nucl. Instrum. Meth. A **440** (2000), 535-538 doi:10.1016/S0168-9002(99)01032-3

[49] M. Klopf, E. Jericha, B. Märkisch, H. Saul, T. Soldner and H. Abele, Phys. Rev. Lett. **122** (2019) no.22, 222503 doi:10.1103/PhysRevLett.122.222503 [arXiv:1905.01912 [hep-ex]].

[50] J. Micko, F. Di Pumpo, J. Bosina, S. S. Cranganore, T. Jenke, M. Pitschmann, S. Roccia, R. I. P. Sedmik and H. Abele, [arXiv:2203.16375 [gr-qc]].

MITP-24-033
March 5, 2024

Two-loop contributions of axion-like particles to electromagnetic and chromomagnetic form factors

Matthias Neubert[a,b] and Marvin Schnubel[c]

[a] *PRISMA$^+$ Cluster of Excellence & Mainz Institute for Theoretical Physics Johannes Gutenberg University, 55099 Mainz, Germany*

[b] *Department of Physics & LEPP, Cornell University, Ithaca, NY 14853, U.S.A.*
[c] *Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, U.S.A.*

arXiv:2403.00913v1 [hep-ph] 1 Mar 2024

Axions and axion-like particles emerge in many models for physics beyond the Standard Model. Thus, they have gained increasing research interest in both experimental and theoretical physics apart from their original proposition as a solution to the strong CP-problem. Among other aspects it has recently been shown that ALPs can potentially provide a solution to the long-lasting discrepancy between theory and experiment of the anomalous magnetic moment of the muon. Provided that the ALP has flavor-violating couplings to leptons, they can also mediate flavor-violating decays like $\mu \to e\gamma$. Both processes are mediated through related form factors that we compute to two-loop order. We further show numerical implications of our calculations and how they might affect constraints on ALP couplings derived from experiments.

Contents

1 Introduction **1**

2 ALPs coupling to the SM **3**
2.1 The effective ALP Lagrangian 3
2.2 Form factors 5

3 Off-shell ALP-photon vertex **6**

4 General expressions for the form factors at two-loops **7**
4.1 Two-loop contributions proportional to $c_{\gamma\gamma}^2$ 11

5 Application to $(g-2)_\mu$ and $\mu \to e\gamma$ **11**
5.1 Application to a_μ 12
5.2 Application to $\mu \to e\gamma$ 14

6 Generalization to gluons and application to the chromomagnetic moment of the top quark **17**

7 Conclusions **18**

References **20**

1 Introduction

Axions and axion-like particles (to which we will commonly refer to as ALPs in this work) are well-motivated extensions of the Standard Model (SM). As an emergent pseudo Nambu-Goldstone boson from the spontaneous breaking of a global $U(1)$ symmetry, they are part of many new physics models. Initially, they were introduced as a solution to the strong CP-problem by Peccei and Quinn and others [1–4], though it was soon noticed that ALPs can have serious connections to other open SM questions, too. One example is the question of the large hierarchy in fermion masses with the proposed solution of the Froggatt–Nielsen mechanism that introduces a new scalar field together with a high energy $U(1)$ symmetry under which the fermions are charged appropriately and which is spontaneously broken at lower energies. Identifying this underlying $U(1)$ symmetry with the Peccei–Quinn symmetry, or equivalently, identifying the phase of the new scalar particle with an ALP, results in an ALP with couplings to all SM fermions with a possibly rich flavor structure [5–7]. Furthermore, especially light ALPs have been shown to be viable cold dark matter candidates [8] and be able to provide a stochastic gravitational wave background in the early universe [9–11]. Consequently, ALPs have gained strong research interest by theory and experimental initiatives likewise in recent years.

The QCD axion naturally couples to gluons when it is supposed to provide a solution to the strong CP problem and in fact most concrete ALP models feature strong couplings of the ALP to gauge bosons. Whereas in KSVZ models the SM fermions are uncharged

under the PQ symmetry [12, 13], additional ALP-fermion couplings are present in DSFZ models [14, 15]. Additionally, the mass of the ALP and the photon coupling are inversely related to each other, resulting in a narrow band in the parameter space where the ALP is a solution to the strong CP problem [16]. In this work we take a model-independent approach and make no further assumptions on the ALP coupling structure. We also keep the ALP mass and its photon coupling as free parameters.

In [17–19] it was shown that an ALP coupling to any SM particle in the UV region will eventually generate couplings to all SM particles in the low-energy effective theory by subsequently evolving the operators from higher to lower scales through their renormalization group (RG) equations and, at the low scale, matching the effective theory to the Standard Model. Whenever a mass threshold is crossed in the evolution process, the respective particle is integrated out and and a new matching is performed. This procedure is repeated until at energies below $\mu_0 \approx 2\,\text{GeV}$ the theory is matched onto a chiral perturbation theory (χPT), where the ALP interacts directly with the hadrons instead. Most importantly, these running and matching effects generate effective quark flavor-changing couplings in the down sector, even if the underlying UV theory is flavor-blind or -conserving. This opens a huge variety of possibilities to probe UV ALP couplings with numerous flavor experiments.

Since the SM features the individual lepton numbers as accidental symmetries, evolution effects alone cannot generate ALP couplings that are lepton flavor-changing. In principle, loops containing flavor-oscillating neutrinos could give rise to flavor-changing processes also in the charged sector. However, they are suppressed by factors of $\Delta m_\nu^2/m_W^2 \approx 10^{-26}$, and hence are typically neglected. If such lepton flavor-violating are already present at tree-level in the UV region, in [17] it was shown that they do not receive corrections from the evolution procedure.

QCD axions are typically thought to be light, $m_\text{a} < 1\,\text{eV}$. Hence, QCD axions and ALPs in that mass region are best probed by light shining through wall (LSW) experiments, stellar and astrophysical probes. Bounds coming from flavor experiments are most stringent in the mass range of $\sim 0.1\,\text{MeV}$ to $10\,\text{GeV}$. For this work, we will therefore assume the ALP mass to lie in this mass region. If the ALP is heavier than $\sim 10\,\text{GeV}$, best limits arise from searches for exotic decays of the Z and Higgs boson, as well as searches for direct production at colliders. We refer the interested reader to [17, 18, 20–34] for detailed overview over studies of ALPs with both flavor-conserving and flavor-violating couplings.

In a common normalization of the ALP Lagrangian, the ALP couplings to gauge bosons are written as $c_{VV}\frac{\alpha_V}{4\pi}\frac{a}{f}F_{\mu\nu}\tilde{F}^{\mu\nu}$ where $F^{\mu\nu}$ is the gauge boson field strength tensor and $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$ its dual[1]. Pulling out a factor of $\alpha_V/(4\pi)$ of the Wilson coefficients ensures the scale independence of these couplings up to two-loop order. Consequently, a $\psi_1 \to \psi_2\gamma$ one-loop Feynman diagram featuring an ALP-photon coupling is, naïvely estimated, contributing at the same order as the two-loop diagram where the direct ALP-photon coupling is replaced by a fermion loop. The work at hand is therefore dedicated to study the impact of such two-loop effects thoroughly. Note that this is not only an academic exercise, but an important contribution to consistently study ALP effects in processes like

[1] We choose the convention $\epsilon^{0123} = 1$.

the decay $\mu \to e\gamma$ and the anomalous magnetic moment of the muon and the electron, $(g-2)_\mu$ and $(g-2)_e$.

Two-loop ALP corrections to dipole moments have been calculated previously in the context of the anomalous magnetic moment of the muon in [35]. In this work, we first re-evaluate the calculation in a different basis, eliminating the need for an additional subtraction to render the expressions finite. We then generalize the computation to flavor-changing currents. Additionally we transfer our findings to dipoles involving non-abelian fields and study the related contribution to the chomomagnetic moment of the top quark.

2 ALPs coupling to the SM

2.1 The effective ALP Lagrangian

The Lagrangian containing the ALP and its interactions with Standard Model particles is given by

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{D\leq 5} =& \frac{1}{2}\left(\partial^\mu a\right)\left(\partial_\mu a\right) - \frac{m_{a,0}^2}{2}a^2 + \frac{\partial_\mu a}{f}\sum_f \left(\bar{\psi}_L \boldsymbol{k}_F \gamma^\mu \psi_L + \bar{\psi}_R \boldsymbol{k}_f \gamma^\mu \psi_R\right) + c_{GG}\frac{\alpha_s}{4\pi}\frac{a}{f}G^a_{\nu\mu}\tilde{G}^{\mu\nu,a} \\ &+ c_{\gamma\gamma}\frac{\alpha}{4\pi}\frac{a}{f}F_{\mu\nu}\tilde{F}^{\mu\nu} + c_{\gamma Z}\frac{\alpha}{2\pi s_w c_w}\frac{a}{f}F_{\mu\nu}\tilde{Z}^{\mu\nu} + c_{ZZ}\frac{\alpha}{4\pi s_w^2 c_w^2}\frac{a}{f}Z_{\mu\nu}\tilde{Z}^{\mu\nu}\,,\end{aligned} \tag{2.1}$$

where $G^a_{\mu\nu}$ and $F_{\mu\nu}$ are the gluon and photon field-strength tensors and $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$ and equivalently for $\tilde{G}$ are the dual tensors. The sine and cosine of the weak mixing angle are abbreviated to $\sin\theta_w \equiv s_w$ and $\cos\theta_w \equiv c_w$, respectively. The couplings are defined below the electroweak scale, however, we keep the top quark and Z bosons as propagating degrees of freedom. Because of their high mass, the contributions of Z bosons to the processes studied here are negligible. At the classical level the Lagrangian the ALP couplings to the SM fields are protected by an approximate shift symmetry $a \to a +$ constant. While the derivative couplings to fermions naturally feature this symmetry, the additional terms in the couplings to gauge bosons can be removed by field redefinitions. Due to instanton effects the coupling to gluons only respects a discrete version of the shift symmetry, realized as $a \to a + n\pi f/c_{GG}$, where n is a natural number [3, 4]. The effective ALP mass m_a is given as the sum of an explicitly shift-symmetry breaking ALP mass $m_{a,0}$ and the contribution of non-perturbative QCD dynamics and reads at lowest order in chiral perturbation theory [13, 36, 37]

$$m_a^2 = m_{a,0}^2\left[1 + \mathcal{O}\left(\frac{f_\pi^2}{f^2}\right)\right] + c_{GG}^2\frac{f_\pi^2 m_\pi^2}{f^2}\frac{2m_u m_d}{(m_u+m_d)^2}\,, \tag{2.2}$$

with $f_\pi \approxeq 130\,\text{MeV}$ the pion decay constant.

We find that for the computation of $\psi_1 \to \psi_2 V$ $(V = g, \gamma)$ form factors it is helpful to work with an alternative formulation of the Lagrangian in a different basis by applying the equations of motion to the SM fermions

$$\begin{aligned}\mathcal{L}_{\text{eff}}^{D\leq 5} =& \frac{1}{2}\left(\partial^\mu a\right)\left(\partial_\mu a\right) - \frac{m_a^2}{2} + \mathcal{L}_{\text{ferm}} \\ &+ \tilde{c}_{GG}\frac{\alpha_s}{4\pi}\frac{a}{f}G^a_{\nu\mu}\tilde{G}^{\mu\nu,a} + \tilde{c}_{\gamma\gamma}\frac{\alpha}{4\pi}\frac{a}{f}F_{\mu\nu}\tilde{F}^{\mu\nu}\,.\end{aligned} \tag{2.3}$$

Here,

$$\tilde{c}_{\gamma\gamma} = c_{\gamma\gamma} + \sum_f N_c^f Q_f^2 c_{ff}, \qquad \text{and} \qquad \tilde{c}_{GG} = c_{GG} + \frac{1}{2}\sum_q c_{qq}\,, \tag{2.4}$$

where the sum of the first term runs over all SM fermions f and in the second term it runs over all quark states q. The couplings c_{ff} and c_{qq} are related to the couplings in the Lagrangian (2.1) via

$$c_{f_i f_i} = [k_f]_{ii} - [k_F]_{ii}\,. \tag{2.5}$$

It is only in these combinations that the coupling parameters of the original Lagrangian can appear in physical observables [17, 19]. The ALP-fermion Lagrangian is given by

$$\begin{aligned}\mathcal{L}_{\text{ferm}} = -\frac{ia}{2f}\sum_f \big[&(m_{f_i} - m_{f_j})[k_f + k_F]_{ij}\bar{f}_i f_j \\ &+ (m_{f_i} + m_{f_j})[k_f - k_F]_{ij}\bar{f}_i\gamma_5 f_j\big]\,.\end{aligned} \tag{2.6}$$

The suppression scale f of the dimension-5 operators is related to the scale of global symmetry breaking by $\Lambda = 4\pi f$. Often one decides to eliminate f in favor of the axion decay constant f_a under the relation $f_a \equiv -f/(2c_{GG})$. In the literature, this is often done when dealing with QCD axions. Note that the ALP-fermion couplings are suppressed with the fermion masses in this alternative formulation, allowing us to neglect couplings to neutrinos.

In the Lagrangians (2.1) and (2.3) we pulled out a normalization factor $\alpha_i/(4\pi)$ for the gauge boson couplings, as can be found in many explicit ALP models in the literature [18, 19, 32, 38]. This ensures that the ALP-gauge boson couplings c_{VV} are scale independent at least up to two-loop order, and the scale dependence of the $\tilde{c}_{VV}$ couplings is fully governed by the evolution of the diagonal fermion couplings [17, 19, 39]. As a consequence, we need to take certain two-loop graphs into account when studying $\psi_1 \to \psi_2 V$ observables for the initial assumption that $c_{ff} \sim c_{VV} \sim \mathcal{O}(1)$, because they contribute with the same power of the QED/QCD coupling constant as other one-loop diagrams. This circumstance is illustrated in figure 1. Here we present an estimate for the size of the contribution of representative classes of one and two-loop diagrams for the flavor conserving process $\psi_1 \to \psi_1\gamma$. We limit ourselves to one exchange of an ALP, as further exchanges are suppressed by additional factors of $1/f$. The diagrams (a) and (b) represent the two possible classes of one-loop diagrams. Diagrams (e) and (f) are loop corrections to the aforementioned cases and are therefore of subleading power in perturbation theory. This is not true for diagrams (c) and (d). From the tentative estimate one expects that inserting the fermion loop into the ALP-photon vertex in (c) contributes at the same order of perturbation theory as the direct contribution (b). For a consistent treatment of ALP induced $\psi_1 \to \psi_2\gamma$ form factors this contribution must therefore be taken into account. It is similar to the well-known Barr-Zee diagrams [40] when the ALP is exchanged with a Higgs boson. Due to our choice of normalization the two-loop diagram with two ALP-photon vertices (d) is severely suppressed in perturbation theory. In models where the ALP-photon coupling is enhanced instead, they can give major contributions to the amplitude.

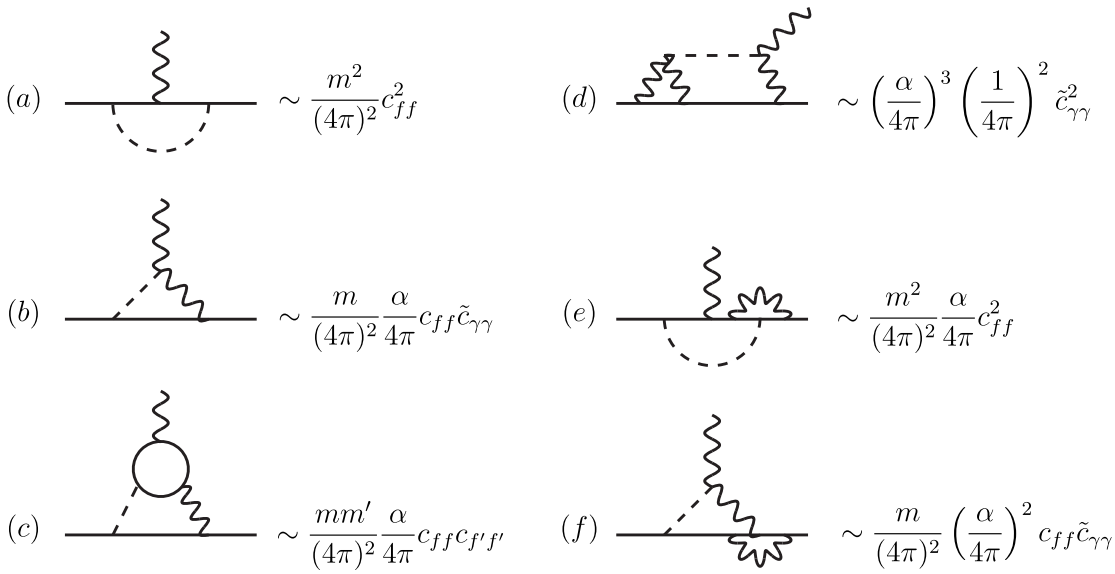


Figure 1: Representative diagrams contributing to ALP-induced $\psi \to \psi\gamma$ form factors up to two-loop order, excluding multiple ALP exchanges. Diagrams (a) and (b) are the one-loop diagrams. Diagrams (e) and (f) present loop-corrections to (a) and (b), respectively, and are therefore suppressed. Diagram (c) contributes at the same order in perturbation theory as diagram (b), even though it is a two-loop correction, and therefore its contribution must be taken into account for a consistent treatment if $c_{ff} \sim c_{\gamma\gamma}$ is assumed. In our choice of normalization of the ALP-gauge boson coupling diagram (d) is highly suppressed in perturbation theory.

2.2 Form factors

In this section we will show how to derive the electromagnetic form factors relevant for $\psi_1 \to \psi_2\gamma$ processes. This can easily be generalized to the non-abelian case with the obvious replacements and insertions of color factors.

The matrix element of a $\psi_1 \to \psi_2\gamma$ process can generally be written as

$$\mathcal{M}^\mu = \bar{\ell}_1(p_1)\Gamma^\mu \ell_2(p_2)\,, \tag{2.7}$$

where p_1 (p_2) denotes the momentum of the initial- (final-) state particle. We furthermore define $p \equiv p_1 + p_2$, and $q \equiv p_1 - p_2$ is the momentum of the outgoing photon. If both initial- and final-state fermions are identical the matrix element can be parametrized in terms of four linearly independent form factors as follows

$$\begin{aligned}\bar{u}_i(p_2)\Gamma^\mu(p_1,p_2)\,u_i(p_1) = \bar{u}_i(p_2)\Big[&F_2^{i\to i}(q^2)\big(p^\mu - 2m_i\gamma^\mu\big) + 2m_i F_3^{i\to i}(q^2)\gamma^\mu \\ &+ F_2^{5,i\to i}(q^2)p^\mu\gamma_5 + F_3^{5,i\to i}(q^2)\Big(q^\mu + \frac{q^2}{2m_i}\gamma^\mu\Big)\gamma_5\Big]u_i(p_1)\,.\end{aligned} \tag{2.8}$$

Note that two additional form factors $F_1^{(5)}$ that are proportional to γ^μ and $\gamma^\mu\gamma_5$, respectively, can be eliminated through application of the Ward identity $q_\mu\Gamma^\mu = 0$. At tree-level

in the SM equation (2.8) takes the form $\Gamma^{\mu}_{\mathrm{SM},0} = Qe\gamma^{\mu}$, where Q is the electric charge of the fermion. If the fermions are leptons of flavor i, one can read off the anomalous magnetic moment defined by $a_i = \frac{(g-2)_i}{2}$ as

$$a_i = \frac{2m_i}{e} F_2^{i\to i}(q^2 = 0). \tag{2.9}$$

Furthermore, the electric dipole moment d_i of a fermion with flavor i is given by

$$|d_i| = \frac{1}{2}\left|F_2^{5,i\to i}(q^2 = 0)\right|. \tag{2.10}$$

In the case that initial and final states are of different fermion flavors instead, the parametrization reads

$$\begin{aligned}
&\bar{u}_j(p_2)\Gamma^{\mu}(p_1, p_2)\, u_i(p_1) = \\
&\quad \bar{u}_j(p_2)\Big[F_2^{i\to j}(q^2)\big(p^{\mu} - (m_i + m_j)\gamma^{\mu}\big) + F_3^{i\to j}(q^2)\Big(q^{\mu} - \frac{q^2}{m_i - m_j}\gamma^{\mu}\Big) \\
&\qquad + F_2^{5,i\to j}(q^2)\big(p^{\mu} + (m_i - m_j)\gamma^{\mu}\big)\gamma_5 + F_3^{5,i\to j}(q^2)\Big(q^{\mu} + \frac{q^2}{m_i + m_j}\gamma^{\mu}\Big)\gamma_5\Big] u_i(p_1)\,.
\end{aligned} \tag{2.11}$$

Note that in equations (2.8) and (2.11) we keep the momentum q of the photon general, even though it would be sufficient to take $q^2 \equiv 0$, i.e. the on-shell limit, for our purposes. The reason why we do so is that if the photon is off-shell it can give rise to secondary lepton pair production that could induce experimental signatures like $\mu^- \to e^- e^+ e^-$.

Flavor-changing neutral currents (FCNCs) are loop-suppressed in the Standard Model. Since the lepton numbers are individually conserved, lepton flavor-changing processes are even strictly forbidden. Hence, we assume for this work that the necessary flavor change stems from an ALP interaction instead. The branching ratio of such an ALP-induced flavor-changing decay is given by

$$\mathrm{Br}(\psi_1 \to \psi_2\gamma) = \frac{m_1^3}{8\pi\Gamma_1}\left(|F_2^{1\to 2}(0)|^2 + |F_2^{5,1\to 2}(0)|^2\right), \tag{2.12}$$

with Γ_1 the decay width of the initial state fermion. This formula is valid up to leading order in the expansion in m_2^2/m_1^2. The mass hierarchies between the different fermion generations ensure that this is indeed a sufficient approximation.

3 Off-shell ALP-photon vertex

As an intermediate step, we compute the fermion loop correction to the ALP-photon vertex, see figure 2. Here, we keep one of the photons as well as the ALP off-shell. The momentum of the ALP is denoted by k, and p and q represent the momenta of the off- and on-shell photons, respectively. The loop integral is finite and can be evaluated in $d = 4$ spacetime

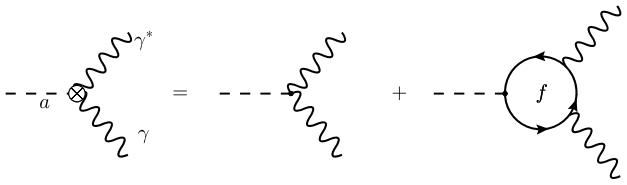


Figure 2: Feynman diagrams contributing to the effective ALP-photon vertex, including the one-loop correction from fermion loops.

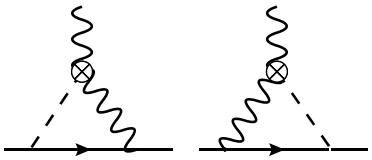


Figure 3: Feynman diagrams for $\psi_1 \to \psi_2\gamma$ form factors with an effective ALP-photon coupling.

dimensions. Including both tree-level and loop-induced diagrams, we obtain for the ALP-photon vertex

$$\begin{aligned}\Gamma^{\mu\alpha}_{a\gamma\gamma^\star}(p^2,k^2) =& \frac{i\alpha}{\pi f}\epsilon^{\mu\alpha\beta\rho}q_\beta p_\rho \\ &\times\left\{c_{\gamma\gamma}+\sum_f N_c^f Q_f^2 c_{ff}\left[1-\int_0^1 \mathrm{d}x\int_0^1 \mathrm{d}y\frac{m_f^2}{m_f^2-x\bar{x}(yp^2+\bar{y}k^2)-i\epsilon}\right]\right\},\end{aligned} \tag{3.1}$$

with $\bar{x}=1-x$ and the polarization indices of off- and on-shell photons denoted as α and μ. Since $k=p+q$ we may replace $yp^2+\bar{y}k^2 = k^2-2y\,k\cdot q = (k-y\,q)^2$ in the denominator. We find that the term involving $k\cdot q$ can be neglected when initial and final state lepton coincide, as in the diagrams contributing to $(g-2)_\mu$. This renders the y-integration trivial. Note that this is not true for diagrams contributing to flavor-changing processes. It can be observed that heavy fermions decouple from the vertex, whereas light fermions contribute 1 inside the rectangular bracket. For the vertex with two on-shell photons a fermion is considered light when it fulfills $m_f^2 \ll m_a^2$, in the case of one off-shell photon the relevant condition is $m_f^2 \ll |k^2|$ instead.

4 General expressions for the form factors at two-loops

The one-loop diagrams contributing to the form factors have already been calculated in the literature [18, 33, 38], which is why we focus on the two-loop graphs here. Moreover, we only consider diagrams with one ALP exchange, as graphs with further ALP exchanges are suppressed by additional factors of the large new physics scale $1/f$.

We insert the vertex function (3.1) into the diagrams, giving rise to the graphs shown in figure 3. Since the off-shell vertex falls off like $1/k^2$ for large loop momenta (see (3.1)), the

graphs with internal fermion loops are UV finite and can be evaluated in $d = 4$ spacetime dimensions. For the processes we are interested in, only the form factors $F_2^{(5)}(q^2)$ yield a contribution. In the relevant limit of $q^2 = 0$, i.e. an on-shell photon in the final state, we obtain

$$F_2^{(5)}(q^2=0) = -\frac{e\alpha q}{16\pi^3}(m_1 \pm m_2)\frac{|k_f|_{12} \mp |k_F|_{12}}{f}\left(\frac{\tilde{c}_{\gamma\gamma}}{f}\mathcal{I}_1 + \sum_f N_c^f Q_f^2 \frac{c_{ff}}{f}\mathcal{I}_2\right), \tag{4.1}$$

where q is the charge of the initial and final state fermion, and Q_f and N_c^f are the charge and number of flavors of the fermions in the one-loop correction in the effective ALP-photon vertex. $\mathcal{I}_1$ and $\mathcal{I}_2$ are the one and two-loop parameter integrals, respectively, and read

$$\begin{aligned}
\mathcal{I}_1 =& \int_0^1 \mathrm{d}x \int_0^x \mathrm{d}y \Bigg(2\delta_2 + \ln\left(\frac{\mu^2}{\Delta_1}\right) + \ln\left(\frac{\mu^2}{\Delta_2}\right) \\
&+ \frac{m_1+m_2}{2}(x-y)\left[\frac{m_1(-1+x)+m_2(1-y)}{\Delta_1} + (m_1 \leftrightarrow m_2)\right]\Bigg), \\
\mathcal{I}_2 =& \frac{1}{2}\int_0^1 \mathrm{d}x \frac{m_f^2}{x\bar{x}} \int_0^1 \mathrm{d}y \left(4(A+A') - (m_1+m_2)(m_1 B + m_2 B') + (m_1^2 - m_2^2)(a + a' - C - C')\right)
\end{aligned} \tag{4.2}$$

with

$$\begin{aligned}
\Delta_1 =& m_a^2(1-x) + m_1^2(-1+x)(x-y) + m_2^2(1-y)(x-y), \\
\Delta_2 =& m_a^2(1-x) + m_1^2(1-y)(x-y) + m_2^2(-1+x)(x-y).
\end{aligned} \tag{4.3}$$

In (4.2) the functions a, A, B and C and their primed counterparts are Passarino-Veltman coefficient functions. They are obtained through the relations (with $\bar{m}^2 = m_f^2/(x\bar{x})$)

$$\begin{aligned}
\int \frac{\mathrm{d}^4k}{(2\pi)^4} \frac{k^\mu}{(k^2-m_a^2)(k-q)^2((k-yq)^2-\bar{m}^2)((k-p_1)^2-m_2^2)} &= \frac{1}{(4\pi)^2}\left(a p_1^\mu + b q^\mu\right), \\
\int \frac{\mathrm{d}^4k}{(2\pi)^4} \frac{k^\mu k^\nu}{(k^2-m_a^2)(k-q)^2((k-yq)^2-\bar{m}^2)((k-p_1)^2-m_2^2)} & \\
= \frac{1}{(4\pi)^2}\left(A g^{\mu\nu} + B p_1^\mu p_1^\nu + C(p_1^\mu q^\nu + q^\mu p_1^\nu) + D q^\mu q^\nu\right).&
\end{aligned} \tag{4.4}$$

The primed coefficients follow equivalently with the replacement $(p_1, m_2) \to (p_2, m_1)$.

While it is possible to obtain analytic expressions for $\mathcal{I}_1$ and reduce $\mathcal{I}_2$ to one unevaluated parameter integral, the equations are too long and unhandy to present them in this article in a convenient form. In the phenomenologically relevant cases, $m_1 = m_2$ and $m_2 \ll m_1$, they can be cast into more compact representations, however

$$\begin{aligned}
\mathcal{I}_1^{m_1=m_2} =& \delta_2 + \ln\left(\frac{\mu^2}{m_1^2}\right) + 3 - h_2\left(\frac{m_a^2}{m_1^2}\right), \\
\mathcal{I}_1^{m_2 \ll m_1} =& \delta_2 + \ln\left(\frac{\mu^2}{m_1^2}\right) + 2 + g_2\left(\frac{m_a^2}{m_1^2}\right)
\end{aligned} \tag{4.5}$$

with

$$\begin{aligned} h_2(x) =& 1 - \frac{x}{3} + \frac{x^2}{6}\ln x + \frac{x+2}{3}\sqrt{(4-x)x}\arccos\frac{\sqrt{x}}{2}\,, \\ g_2(x) =& -\frac{x^2\ln x}{2(x-1)} + \frac{1}{2}(x-1)\ln(x-1)\,. \end{aligned} \tag{4.6}$$

Here, $\delta_2 = -3$ is a scheme-dependent constant derived from treating the Levi-Civita-tensor as a d-dimensional object[2]. If it is seen as a $d = 4$ dimensional object instead, $\delta_2 = 0$ [32]. The two-loop terms simplify to

$$\begin{aligned} \mathcal{I}_2^{m_1=m_2} =& \int_0^1 \mathrm{d}x F\left(y_x, r_1\right)\,, \\ \mathcal{I}_2^{m_2\ll m_1} =& \int_0^1 \frac{\mathrm{d}x}{x\bar{x}} \int_0^1 \mathrm{d}u \Bigg\{ (1-u)u r_1 \left[-(r_a - r_1 u)x\bar{x}\ln\left(\frac{(r_a - r_1 u)x\bar{x}}{1 - r_1 u x\bar{x}}\right) - \ln(1 - r_1 u x\bar{x})\right] \\ &+ 3(1-u)(1-(r_a - r_1 u)x\bar{x})\left[\mathrm{Li}_2\left(\frac{-1+(r_a - r_1 u)x\bar{x}}{(r_a - r_1 u)x\bar{x}}\right) - \mathrm{Li}_2\left(\frac{-1 + r_a x\bar{x}}{(r_a - r_1 u)x\bar{x}}\right)\right] \\ &+ \int_0^1 \mathrm{d}y \int_0^{1-u} \mathrm{d}v \int_0^{1-u-v} \mathrm{d}w \frac{r_a(1-u-v-w) + \frac{w}{x\bar{x}} + \frac{3}{2}r_1 u(1-v-wy)}{\left(r_a(1-u-v-w) + \frac{w}{x\bar{x}} + \frac{3}{2}r_1 u(1-v-wy)\right)^2} \Bigg\}\,. \end{aligned} \tag{4.7}$$

Here, we introduced the abbreviations $r_a = m_a^2/m_f^2$ and $r_1 = m_1^2/m_f^2$. The parameter integrals in the last row can all be computed analytically. However, their expressions are unwieldy long, which is why we decided to omit them here. In the first line

$$F(y_x, r_1) = \frac{1}{1-y_x}\left[h_2\left(\frac{r_a}{r_1}\right) - h_2\left(\frac{1}{x\bar{x}r_1}\right)\right]\,, \quad \text{and} \quad y_x = x(1-x)r_a\,. \tag{4.8}$$

The function $\mathcal{I}_2^{m_1=m_2}$ in the equal mass limit has first been computed in [35]. In this calculation the loop function in the heavy mass limit needed to be subtracted to obtain the correct result.The reason why a similar procedure is not necessary in our calculation is that we changed the basis by applying the equations of motions, and as a result the ALP features no derivative couplings in this alternative form of the Lagrangian (2.3). The advantage of this alternative formulation becomes clear after a closer inspection of the effective ALP-photon vertex (3.1). In the infinite mass limit $m_f^2 \gg k^2$ the rectangular bracket including the two-loop contribution vanishes. In the opposite limit $m^2 \ll k^2$ the additional propagator renders the two-loop contribution finite, allowing us to perform the calculation in $d = 4$ spacetime dimensions regardless of the mass of the inner fermion. However, we want to emphasize here that both basis choices are of course equivalent, and consequently our results fully agree with those of Ref. [35].

The remaining parameter integrals can be readily integrated numerically. In certain limits we find it interesting to report some useful explicit results. The limits $m_a^2 \gg m_1^2$ and $m_f^2 \gg m_{a,1}^2$ yield the same functional behavior for both the flavor-conserving and flavor-changing case.

[2]The one-loop diagrams are divergent and hence need to be regularized with dimensional regularization.

Limit $m_a^2 \gg m_1^2$ In the limit that the ALP is much heavier than the initial state fermion, one can take the limit $r = \frac{r_a}{r_1} \to \infty$ to obtain

$$\begin{aligned} f(r_a) &= \lim_{r\to\infty} \int_0^1 \mathrm{d}x\, F(y_x, r_1) \\ &= \frac{-4}{\sqrt{r_a(r_a-4)}} \left\{ \frac{\pi^2}{12} + \ln^2\left(\frac{1}{2}\left(\sqrt{r_a} - \sqrt{r_a-4}\right)\right) + \mathrm{Li}_2\left[-\frac{1}{4}\left(\sqrt{r_a} - \sqrt{r_a-4}\right)^2\right]\right\}, \end{aligned} \tag{4.9}$$

Its asymptotic behavior is such that

$$\begin{aligned} f(r_a) &= \ln\frac{m_f^2}{m_a^2} - 2 + \mathcal{O}\left(\frac{m_a^2}{m_f^2}\right); && m_f^2 \gg m_a^2\,, \\ f(r_a) &= -\frac{m_f^2}{m_a^2}\left[\ln^2\left(\frac{m_a^2}{m_f^2}\right) + \frac{\pi^2}{3}\right] + \mathcal{O}\left(\frac{m_f^4}{m_a^4}\right); && m_f^2 \ll m_a^2\,. \end{aligned} \tag{4.10}$$

The fact that a light internal fermion decouples in the Barr-Zee graphs is more general and holds also if $m_a \sim m_1$. It follows that for an light internal fermion ($m_f^2 \ll m_a^2$) the sum of all contributions proportional to c_{ff} is

$$N_c^f Q_f^2 c_{ff}\left[\ln\frac{\mu^2}{m_1^2} + \delta_2 + 3 - h_2\left(\frac{m_a^2}{m_1^2}\right) + \mathcal{O}\left(\frac{m_f^2}{m_a^2}\right)\right]. \tag{4.11}$$

In essence this means that virtual loop momenta are cut off for values below the scale $\max(m_a, m_1)$.

Limit $m_f^2 \gg m_{a,1}^2$ In the limit that the inner fermion is much heavier than both the initial state fermion and the ALP, one finds that

$$\lim_{r_1, r_a\to 0} \int_0^1 \mathrm{d}x\, F(y_x, r_1) = -\ln\frac{m_f^2}{m_1^2} + h_2\left(\frac{m_a^2}{m_1^2}\right) - \frac{7}{2} + \mathcal{O}\left(\frac{m_{a,1}^2}{m_f^2}\right). \tag{4.12}$$

It follows that for an internal heavy fermion ($m_f^2 \gg m_a^2$) the sum of all contributions proportional to c_{ff} is

$$N_c^f Q_f^2 c_{ff}\left[\ln\frac{\mu^2}{m_f^2} + \delta_2 - \frac{1}{2} + \mathcal{O}\left(\frac{m_{a,1}^2}{m_f^2}\right)\right]. \tag{4.13}$$

In essence, the virtual loop momenta are cut off for values below the scale m_f. For scales $\mu \ll m_f$ the heavy fermion decouples and can be integrated out.

Limit $m_1^2 = m_2^2 = m_f^2 \gg m_a^2$ The limit where the initial and final state fermion coincide with the inner loop-fermion and the ALP is very light light is mainly relevant in the case of the chromomagnetic moment of the top quark. The integral in the first line of (4.7) then reads

$$\lim_{r_1=1, r_a\to 0} \int_0^1 \mathrm{d}x\, F\left(y_x, r_1\right) = \frac{2}{3} + \frac{2\pi^2}{9} - \frac{2\pi}{3}\frac{m_a}{m_f} + \left(\frac{1}{3} + \frac{\pi^2}{9}\right)\frac{m_a^2}{m_f^2} + \mathcal{O}\left(\frac{m_a^3}{m_f^3}\right). \tag{4.14}$$

Note that for this special case the expansion is linear in m_a/m_f, giving that the ALP mass effects are not as strongly suppressed as in the other limits that we have discussed here. The opposite case that $m_1^2 = m_2^2 = m_f^2 \ll m_a^2$ is already covered with the second equation in (4.10).

4.1 Two-loop contributions proportional to $c_{\gamma\gamma}^2$

Two-loop diagrams involving two ALP-photon vertices yield amplitudes proportional to $c_{\gamma\gamma}^2$ and are highly suppressed in our normalization of the Lagrangian, see for example the scaling of diagram (d) in figure 1. In models where the ALP-photon coupling is enhanced instead, they can give substantial contributions. The relevant diagrams consist of light-by-light (LbL) scattering diagrams and ALP insertions into the internal photon propagator that are similar to the corresponding contribution of pions to the SM vacuum polarization. The diagrams are shown in figure 4. We find it instructive to to give an estimate of their size and compare their effects with the two-loop diagrams described in the previous section, since they arise at the same loop order. In [41] the contributing diagrams have been calculated to leading logarithmic order. It was shown that while the LbL contribution gives rise to double logarithmic corrections, the vacuum polarization contribution only yields single logarithmic behavior and can therefore often be neglected. In [42] the contributions where derived more thoroughly through the means of solving the renormalization group equations. Furthermore it was noticed that in the previous work the single logarithmic contribution was underestimated, yet the conclusion that the double logarithmic part gives the dominant contribution was correct. Translating their findings to the quantities used in this work, the contributions proportional to $c_{\gamma\gamma}^2$ read

$$F_2^{c_{\gamma\gamma}^2}(q^2=0) = \frac{e m_\mu^3}{(4\pi)^2}\left(\frac{\alpha}{\pi}\right)^2\left(\frac{c_{\gamma\gamma}}{f}\right)^2\frac{3}{8}\left[\ln^2\frac{\mu^2}{m_\mu^2} + \left(2\delta_2 + \frac{56}{9} - 2h_2\left(x_\mu\right) + \frac{x_\mu}{3}\right)\ln\frac{\mu^2}{m_\mu^2}\right], \tag{4.15}$$

with $x_\mu = m_a^2/m_\mu^2$. The scale μ should be chosen as $\mu \sim \max(m_\mu, m_a)$. However, if the ALP is too heavy it should be integrated out from the low-energy effective theory instead. Above the electroweak (EW) scale contributions from Z bosons need to be taken into account as well. In our model where we all Z bosons as propagating degrees of freedom also below the EW scale, their contributions are negligible.

The diagrams in figure 4 are very similar to SM diagrams with LbL scattering with neutral pions and pion insertions in the photon propagator. They have been studied intensely in the literature as they are an important ingredient of the SM content supplying to the anomalous magnetic moment of the muon a_μ. Note that diagrams of this type cannot give contributions to flavor-changing observables since the flavor-change must be provided directly by the ALP.

5 Application to $(g-2)_\mu$ and $\mu \to e\gamma$

In this section we apply our findings to the phenomenological studies of the anomalous magnetic moment of the muon a_μ and the lepton flavor-violating decay $\mu \to e\gamma$. The

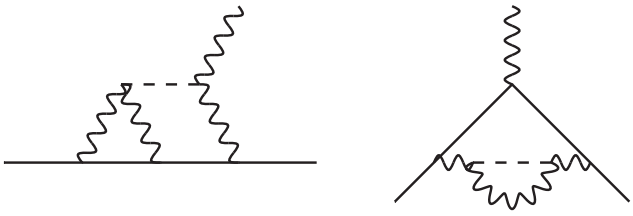


Figure 4: Feynman diagrams for the 2-loop contribution to the flavor-conserving process proportional to the ALP-photon coupling only. Left: Light-by-light scattering. Right: Photon vacuum polarization.

anomalous magnetic moment defined via $a_\mu = (g-2)_\mu/2$ is one of the most precisely measured quantities in particle physics, measured in the Brookhaven [43] and Fermilab [44] experiments. However, it deviates from its predicted value in the SM. How large this deviation is cannot be stated without ambiguity. The reason for this is that there are currently two methods used for determining the hadronic vacuum polarization (HVP), the part which is currently giving the dominating contribution to the theory uncertainty. In the data-driven approach the HVP is extracted from e^+e^--scattering data, resulting in a discrepancy between the measurement of a_μ and the prediction of over 5σ [45][3]. Alternatively, it can be obtained using only input from lattice calculations by the Budapest, Marseille and Wuppertal (BMW) collaboration (BMW) [67], yielding a better agreement with the experimental result. Since the lepton numbers are individually conserved in the SM, lepton flavor-violating (LFV) decays like $\mu \to e\gamma$ are forbidden. Currently, the best constraint on the respective muon branching ratio was obtained by the MEG collaboration and reads $\mathrm{Br}(\mu \to e\gamma) < 4.2 \times 10^{-13}$ [68]. Taking neutrino oscillations into account the SM expectation is $\mathrm{Br}(\mu \to e\gamma)|_{\mathrm{SM}} \approx 10^{-52}$. Therefore LFV observables present an excellent probe of new physics.

5.1 Application to a_μ

An ALP coupling to SM particles gives rise to several contributions to a_μ. At the one-loop level, they have been calculated in [32, 33]. Here, we extend this computation by including the two-loop graph with an inserted fermion-loop. The final result is given by

$$\delta a_\mu = \frac{m_\mu^2}{f^2}\left(K_{a_\mu}(f,\mu) - \frac{c_{\mu\mu}^2}{16\pi^2} h_1\left(\frac{m_a^2}{m_\mu^2}\right) - \frac{\alpha}{8\pi^3} c_{\mu\mu} \underbrace{\left[\tilde{c}_{\gamma\gamma} \mathcal{I}_1^{m_1=m_2} + \sum_f N_c^f Q_f^2 c_{ff} \mathcal{I}_2^{m_1=m_2} \right]}_{=c_{\gamma\gamma}^{\mathrm{eff}}} \right), \tag{5.1}$$

and we have ignored the numerically sub-dominant contributions from the Z-boson. The parameter function h_1 was first calculated in [32] and reads

$$h_1(x) = 1 + 2x + (1-x)x\ln x - 2x(3-x)\sqrt{\frac{x}{4-x}}\arccos\frac{\sqrt{x}}{2}\,. \tag{5.2}$$

[3]The theory initiative paper derives their result based on various contributions to the SM prediction [46–66].

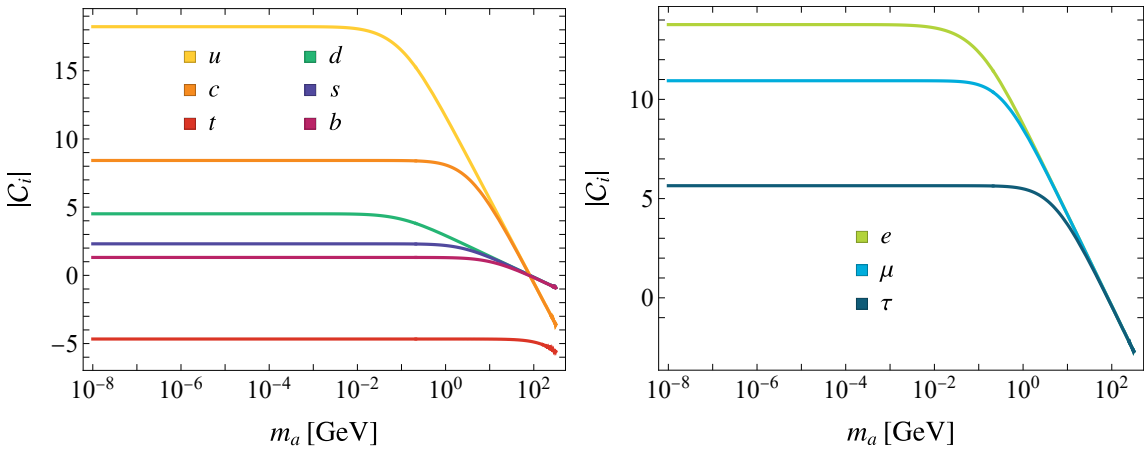


Figure 5: Coefficients $\mathcal{C}_i$ in (5.3) for quarks (left) and leptons (right). When the ALP is light when compared with the fermion in the loop, the effective coupling tends to constant plateau, and the value is entirely governed by the mass of the loop fermion. In the case that the ALP is the heaviest particle the effective coupling scales as $\sim N_c^f Q_f^2 \ln \mu^2/m_a^2$. Hence the behavior is dictated by the ALP mass and the slope depends on the fermion type, i.e. up or down-type quark or lepton.

To disentangle the effects of the different fermion species in the loop, we decompose the effective ALP-photon coupling $c_{\gamma\gamma}^{\text{eff}}$ defined in (5.1) into

$$c_{\gamma\gamma}^{\text{eff}} = c_{\gamma\gamma}\mathcal{I}_1^{m_1=m_2} + \sum_i \mathcal{C}_i c_{ii}\,, \tag{5.3}$$

where i runs over all SM fermion states. In figure 5 we show the coefficients $\mathcal{C}_i$ for the contributions of the individual fermions for quarks (left) and leptons (right). The shape of the functions can be explained as follows: If the ALP is very heavy, the light fermion decouples and the contribution to the effective photon coupling is only dependent on the ALP mass, as dictated by equation (4.11). Essentially the coupling scales as $\lim_{m_a \gg m_f} c_{\gamma\gamma}^{\text{eff}} \sim N_c^f Q_f^2 \ln \mu^2/m_a^2$, explaining the linear pattern in the logarithmically scaled plot as well as the different slopes for up and down-type quarks, and leptons, respectively, based on the number of colors. In the opposite regime, $m_a^2 \ll m_f^2$, the behavior is instead governed by (4.13). The effective coupling reaches a plateau whose value only depends on the mass of the loop fermion. Note further that, in the up and down quark sector respectively, light quarks give a bigger contribution than heavy quarks do, as long as the quark can be considered light when compared with the ALP. This is expected, since in this regime essentially $\mathcal{C}_i \sim Q_{f_i}^2 \ln \mu^2/m_{f_i}^2$. To get a feeling how strong these contributions are, we compare the value of the plateau for $m_a^2 \ll m_f^2$ with the value of the coefficient of $c_{\gamma\gamma}$ in that limit. We find $\lim_{m_a \to 0} \mathcal{I}_1^{m_1=m_2} = 13.8$, where we have chosen the renormalization scale as the top quark mass $\mu = m_t$. Hence, especially light fermions give rise to sizable contributions to the effective ALP-photon coupling if one takes all ALP couplings to be of the same order $c_{\gamma\gamma} \sim c_{ff}$.

We find it instructive to compare the effects from the different terms contributing to

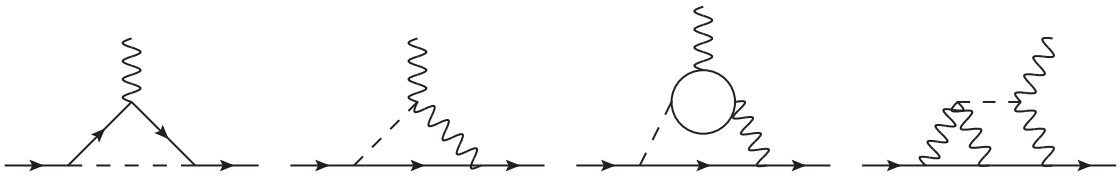


Figure 6: Leading order diagrams involving an ALP that contribute to the anomalous magnetic moment of the muon up to two-loop level. The contributions are proportional to $c_{\mu\mu}^2, c_{\mu\mu}\tilde{c}_{\gamma\gamma}, c_{\mu\mu}c_{ff}, c_{\gamma\gamma}^2$, reading from left to right. Note that for second and third diagram there exists a mirror diagram, where the ALP leg is interchanged with the photon leg. In the third diagram the inner loop fermion can run both clockwise and counter-clockwise. The fourth diagram has a mirror diagram where the ALP-photon vertex with one on-shell photon is exchanged with the other, additionally a diagram where the two rightmost legs are interchanged and the mirror of that one, too.

δa_μ (see Fig. 6 for comparison). For this reason we decompose δa_μ into

$$\delta a_\mu = \frac{m_\mu^2}{f^2} K_{a\mu}(f,\mu) + \frac{1}{f^2}\left(c_{\mu\mu}^2 \delta a_\mu^{\mu\mu} + c_{\mu\mu}\tilde{c}_{\gamma\gamma}\delta a_\mu^{\mu\gamma} + c_{\mu\mu}c_{ff}\delta a_\mu^{\mu f} + c_{\gamma\gamma}^2 \delta a_\mu^{\gamma\gamma}\right) . \tag{5.4}$$

Combinations of ALP-SM particle couplings missing here such as terms proportional to c_{ff}^2 are of higher loop order and thus further suppressed. We will neglect these terms here. The magnitude of the individual terms is compared in figure 7. Here we only keep leptons in the loop in $\delta a_\mu^{\mu f}$. The far dominant contribution over all ALP masses is $\delta a_\mu^{\mu\mu}$, exceeding the others by multiple orders of magnitude. For this reason we exclude this contribution in the overview for better visibility. It can be seen from this figure that the two-loop contribution is of roughly the same order as the one-loop contribution with a tree-level ALP-photon coupling and therefore should not be neglected when including the latter one. Note that the LbL and photon polarization diagrams combined (proportional to $c_{\gamma\gamma}^2$) are several orders of magnitude weaker than the $c_{\mu\mu}c_{ff}$ and $c_{\mu\mu}c_{\gamma\gamma}$ contribution. The results worked out in this section can easily be transferred to the similar cases of the anomalous magnetic moment of the electron a_e and, in principle, to that of the tau-lepton a_τ. The latter is however experimentally too weakly constrained to give meaningful constraints on new physics models. Since it is not entirely clear what the Standard Model predicts for the value of a_μ, we refrain from deriving bounds on the ALP couplings here.

5.2 Application to $\mu \to e\gamma$

An ALP with tree-level flavor-changing couplings can give rise to LFV decays like $\mu \to e\gamma$. Unlike quark flavor-changing couplings, lepton flavor-changing couplings cannot be generated through evolution and matching corrections, thus they need to be imprinted already at the high energy scale, where the PQ symmetry is still unbroken [17, 19]. Explicit models featuring such a flavor structure include for example axiflavon models, where the PQ symmetry is identified with the $U(1)$ symmetry of the Froggatt-Nielsen mechanism that could potentially explain the mass hierarchy of the SM fermions.

Diagrams contributing to the decay rate of $\mu \to e\gamma$ are given in figure 3 for the case that an effective ALP-photon coupling is present. Additionally penguin-like diagrams where the

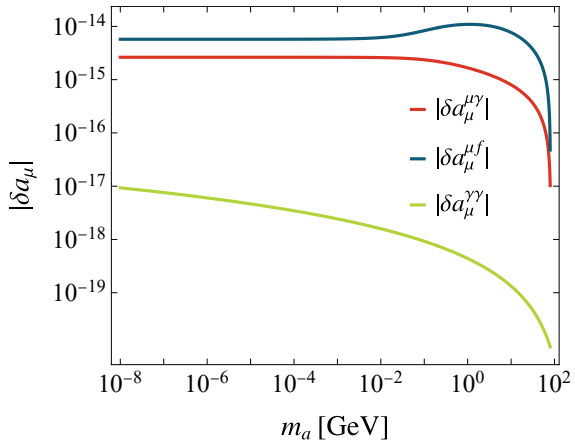


Figure 7: Comparison between the absolutes of different contributions in δa_μ. Note that the LbL and photon polarization terms (proportional to $c_{\gamma\gamma}^2$) are several order of magnitude weaker than the other terms. We assumed $c/f = 1\,\text{TeV}^{-1}$, where c is any ALP-SM coupling.

ALP-loop closes on the fermion line also contribute. The branching ratio is given by

$$\text{Br}(\mu \to e\gamma) = \frac{m_\mu^3}{8\pi\Gamma_\mu}\left(|F_2(0)|^2 + |F_2^5(0)|^2\right)\,, \tag{5.5}$$

and the form factors read

$$F_2^{(5)}(0) = \frac{-em_\mu}{64\pi}\frac{|k_E|_{e\mu} \mp |k_e|_{e\mu}}{f}\left(\frac{c_{\mu\mu}}{f}g_1\left(\frac{m_a^2}{m_\mu^2}\right) + \underbrace{\frac{\alpha}{\pi}\left[\frac{\tilde{c}_{\gamma\gamma}}{f}\mathcal{I}_1^{m_e\ll m_\mu} + \sum_f N_c^f Q_f^2 \frac{c_{ff}}{f}\mathcal{I}_2^{m_e\ll m_\mu}\right]}_{=c_{\gamma\gamma}^{\text{eff}}}\right), \tag{5.6}$$

where $\mathcal{I}_1^{m_e\ll m_\mu}$ and $\mathcal{I}_2^{m_e\ll m_\mu}$ have been given in eqs. (4.6) and (4.7), respectively, and

$$g_1(x) = 2x^{3/2}\sqrt{4-x}\arccos\frac{\sqrt{x}}{2} + 1 - 2x + \frac{x^2(3-x)}{1-x}\ln x \tag{5.7}$$

Note that by taking the limit $m_e \ll m_\mu$ we neglect terms that scale as $m_e^2/m_\mu^2 = \mathcal{O}(10^{-5})$ in the branching ratio, making this a reasonable approximation.

As before, we like to show the numerical implications of including the fermion-loop induced ALP-photon coupling to the effective ALP-photon coupling. We decompose the coupling into

$$c_{\gamma\gamma}^{\text{eff}} = c_{\gamma\gamma}\mathcal{I}_1^{m_2\ll m_1} + \sum_i \mathcal{C}_i c_{ii}\,, \tag{5.8}$$

in accordance with (5.3). The results for the real parts of the coefficients are shown in figure 8. Overall, the picture looks very similar to the flavor-conserving case in figure 5. While the effective coupling tends to a constant for ALPs that are very light, it yields a logarithmic function dependent only on the ALP mass with the slope in the logarithmic plot dictated by the number of colors of the inner loop fermion in the case that the ALP is heavy. Again,

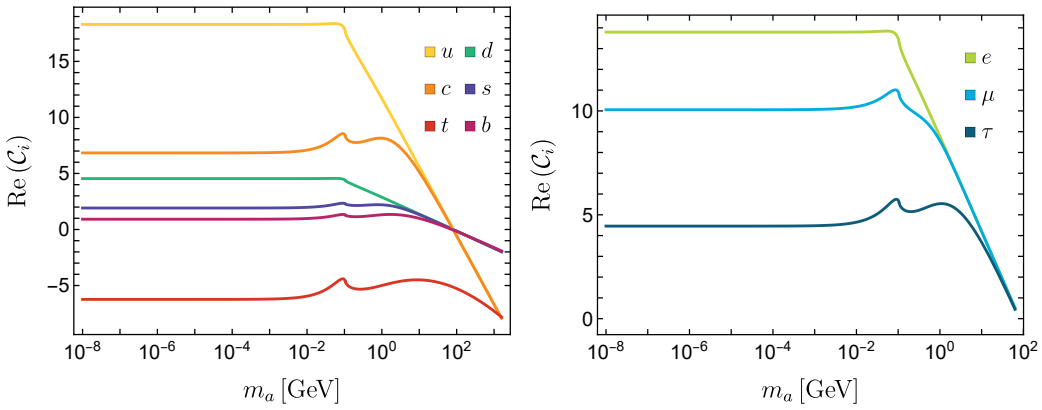


Figure 8: Effective ALP-photon coupling $c_{\gamma\gamma}^{\text{eff}}$ in $\mu \to e\gamma$ for (left) a quark and (right) a lepton in the loop. For this plot we assumed $c_{ff}/f = 1\,\text{TeV}^{-1}$. Note that for large ALP masses the functions become degenerate and the slope depends on the only on the number of colors and the electric charge of the fermions in the loop.

it is notable that light fermions in the loop give a larger contribution than heavy fermions. The cusps and wiggles of the curves at values $m_a \sim m_\mu$ are due to the fact that $c_{\gamma\gamma}^{\text{eff}}$ has imaginary values for ALPs lighter than the muon mass. We want to give a feeling of the size of the two-loop effects also in the flavor-changing case. If all c_{ii} are set to zero and only $c_{\gamma\gamma}$ is kept, the real part of the effective ALP-photon coupling in the vanishing ALP mass limit reads $\lim_{m_a \to 0} \text{Re}\, \mathcal{I}^{m_2 \ll m_1} = 13.8$. Comparing this value to the value of the plateaus for small ALP masses shows that also in ALP mediated flavor-changing form factors a loop-induced ALP-photon coupling is of the same order of magnitude than a tree-level one.

As an illustration how different coupling structures of the ALP imprint themselves on constraints on the ALP parameter space with one lepton flavor-changing coupling, we show the area of $c_{e\mu}$– m_a parameter space derived from the latest limit on the muon branching ratio $\text{Br}(\mu \to e\gamma) < 4.2 \times 10^{-13}$ in figure 9. The bound was obtained from the MEG collaboration [68]. Here, the LFV ALP coupling is defined as $c_{e\mu} = \sqrt{|k_E|^2_{e\mu} + |k_e|^2_{e\mu}}$. We assume that additionally to the LFV coupling a diagonal coupling to the other leptons is present, too. In the plot we distinguish the following cases: In blue (yellow) we show the scenario where the ALP only couples to leptons, neglecting (taking into account) the additional two-loop piece from inserting a lepton loop in the effective ALP-photon coupling. The areas shaded in orange (red) and light green (dark green) represent the corresponding scenarios when the ALP features, additionally to the lepton coupling, a coupling to up-type quarks and down-type quarks, respectively. We keep the ALP mass and its LFV coupling as free parameters and set all lepton couplings universally to $c_{\ell\ell}/f = 1\,\text{TeV}^{-1}$. In a similar way, $c_{uu}/f = c_{cc}/f = c_{tt}/f = 1\,\text{TeV}^{-1}$ and $c_{dd} = c_{ss} = c_{bb} = 1\,\text{TeV}^{-1}$ are fixed in the second and third scenario, respectively. Note that taking the two-loop contributions into account generally leads to a weakening of the derived bounds. Even though the absolute

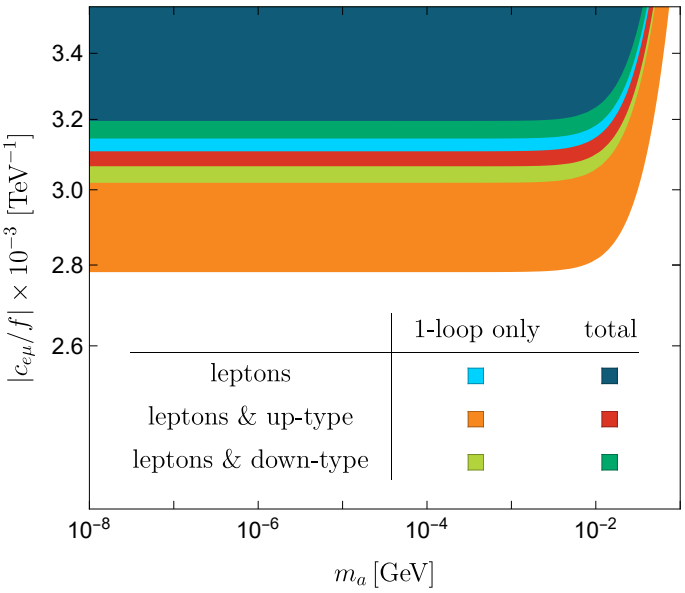


Figure 9: Parameter space excluded by the non-observation of $\mu \to e\gamma$ for different coupling structures of the ALP, i.e. a coupling only to leptons (light blue and dark blue), leptons and up-type quarks (orange and red) and leptons and down-type quarks (light green and dark green). To visualize the effect of our findings, we show the limits for the given coupling structures on the flavor-violating ALP coupling $c_{e\mu} = \sqrt{|k_E|^2_{e\mu} + |k_e|^2_{e\mu}}$ when the two-loop contribution is neglected (first color) and taken into account (second color). The diagonal ALP couplings are assumed to be $c_{ff}/f = 1\,\text{TeV}^{-1}$.

bound on the LFV coupling does not change too much when the two-loop diagrams are considered because the dominant contribution is still given by the penguin-like diagrams proportional to $c_{e\mu}c_{\mu\mu}$, relative corrections of $\approx 10\,\%$ are encountered.

6 Generalization to gluons and application to the chromomagnetic moment of the top quark

The computations presented for the anomalous magnetic moment of the muon in section 5.1 are easily generalizable to non-abelian gauge bosons, giving rise to chromomagnetic moments of quarks. Especially the chromomagnetic dipole moment of the top quark $\hat{\mu}_t$ is an excellent probe for new physics above the electroweak scale [69–71]. It is defined as the coefficient of the operator

$$\mathcal{L} \supset -\hat{\mu}_t \frac{g_s}{2m_t} \bar{t}\sigma^{\mu\nu} T^a \, t \, G^a_{\mu\nu} \,. \tag{6.1}$$

The latest result on its value was derived by the CMS collaboration in two independent measurements [72, 73] and yields the exclusion limit

$$-0.014 < \text{Re}(\hat{\mu}_t) < 0.004 \tag{6.2}$$

at 95 % confidence level. The ALP could possibly induce a non-vanishing chromomagnetic moment via the diagrams given in figure 6, where one has to exchange the photons by

gluons. Its contribution is given by

$$\hat{\mu}_t = \frac{m_t^2}{32\pi^2 f^2}\left\{c_{tt}^2 h_1\left(\frac{m_a^2}{m_t^2}\right) + \frac{2\alpha_s}{\pi} c_{tt}\left[\tilde{c}_{GG}\mathcal{I}_1^{m_1=m_2} + \frac{1}{2}\sum_q c_{qq}\mathcal{I}_2^{m_1=m_2}\right]\right\}. \tag{6.3}$$

The loop functions are given in equations (5.2), (4.5) and (4.7), respectively. For ALPs that are significantly lighter than the top quark the two-loop contribution is given by (4.14). Terms proportional to $\tilde{c}_{GG}^2$ are suppressed by an additional factor of α_s^2. Using RG methods, the leading logarithmic contributions have been first derived in [74]. When we want to derive limits on the ALP parameters entering (6.3) from the experimental bound (6.2) we must however mention an important caveat. The measurement puts constraints on the value of $\hat{\mu}_t$ through the operator (6.1) solely. Any bound derived here therefore implicitly implies that the ALP's influence is mostly covered by aforementioned operator. This means especially that the ALP does not alter the underlying $pp \to t\bar{t}$ production process significantly. We consider this assumption justified for the heaviest ALPs $m_a \gtrsim m_\Upsilon \approx 10.58\,\text{GeV}$.

We show our derived bounds in figure 10 for varying ALP masses for the three cases that the ALP couples only to the top quark (red), the top quark and gluons (green), and the top quark and down-type quarks (blue). In light red we show the derived bounds for top couplings only in the case that the two-loop contribution $\mathcal{I}_2^{m_1=m_2}$ is neglected. Note that similar to the $\mu \to e\gamma$ discussion including two-loop terms leads to a weakening of the derived constraint on the ALP parameter. We assume that all non-vanishing couplings are of the same order, i.e. $c_{tt} \sim c_{GG} \sim c_{qq} \sim c$. The case in which the ALP couples uniformly to all up-type quarks is not portrayed, because in the plot it would be indistinguishable from the second case (top and gluon coupling). This observation is easy to explain: The one-loop contribution is proportional to the coupling structure $\tilde{c}_{GG} = c_{GG} + 1/2\sum c_{qq}$. The coupling c_{tt} must be present to generate a non-vanishing $\hat{\mu}_t$ in the first place. Hence, taking the additional couplings to up and charm quarks into account has the same effect as keeping a tree-level ALP-gluon coupling, for the assumption that all ALP couplings are of the same order. For all quarks other than the top quark the two-loop contribution $\mathcal{I}_2^{m_1=m_2}$ is negligible. Comparing the two red regions in the plot, we emphasize that the two-loop corrections entering the effective ALP-gluon coupling should not be neglected. For an ALP with a mass of $m_a = 100\,\text{GeV}$, the corrections to the derived bound are of the order of $\sim 10\,\%$.

7 Conclusions

Axion-like particles are among the best motivated models for physics beyond the Standard Model. Since they arise as pseudo-Nambu-Goldstone bosons of a high-energy spontaneously broken $U(1)$ symmetry, they are part of many models. Furthermore they have the potential to tackle multiple issues of the SM at once, such as the strong CP problem and the absence of dark matter in the SM. In an often used formulation of the most general ALP-Lagrangian the Wilson coefficient associated with the ALP-photon coupling (or a general ALP to gauge boson coupling) is suppressed by a factor of $\alpha/(4\pi)$. The advantage of this normalization

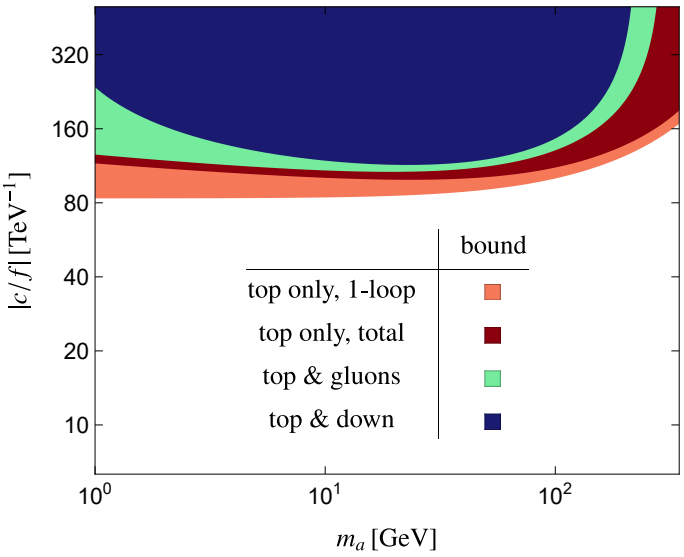


Figure 10: Parameter space excluded by searches for the chromomagnetic dipole moment of the top quark. Excluded regions are shown for the case when the ALP couples only to top quarks (light and dark red), additionally to gluons (green), and top quarks and down-type quarks (blue). All analysis assume that the couplings are of the same order $c_{tt} \sim c_{GG} \sim c_{qq} \sim c$. The scenario where the ALP couples uniformly to all up-type quarks coincides with the one where a coupling to top quarks and gluons is present, assuming all couplings are of the same order of magnitude. In the light red region we neglected the contribution from the two-loop graph, giving rise to an overestimation of the excluded parameter space of $\sim 10\,\%$ for $m_a = 100\,\text{GeV}$.

is that up to two-loop order the gauge-boson couplings of the ALP are scale independent. However, this comes at the price that when the ALP-photon coupling is replaced by an ALP coupling to a fermion loop that radiates the two photons, this additional loop contributes with the same power of the QED coupling constant. For a consistent treatment of ALP-photon couplings it is therefore necessary to include this contribution into the computations of Feynman diagrams.

In this work we have studied the impact of the added fermion loop contribution to $\psi_1 \to \psi_2\gamma$ form factors. Examples where such form factors appear in phenomenological studies are the lepton flavor-violating decay $\mu \to e\gamma$, as well as the flavor-conserving observable of the anomalous magnetic moment of the muon a_μ. For the physically relevant cases that either the two fermions ψ_1 and ψ_2 are the same, or that the final state fermion is much lighter than the initial state one, we were able to derive explicit expressions for both the one-loop as well as the fermion-loop induced two-loop contribution to the form factor. The mass hierarchy between the different fermion families ensures that the assumption $m_2 \ll m_1$ in the flavor-violating case is indeed justified. We find that generally for both flavor-conserving and -violating processes the two-loop piece shows the same behavior in certain mass limits: If the ALP is much heavier than the loop fermion, the latter decouples in the Barr-Zee graphs and in essence the virtual loop momenta are cut off for values below the highest mass scale in the process. If on the other hand the inner loop fermion is much

heavier than both the ALP and the initial state fermion, the loop fermion can be integrated out.

Eventually, we applied our findings to the two lepton observables of $(g-2)_\mu$ and $\mu \to e\gamma$. We find for both scenarios that the effective ALP-photon coupling including the one and two-loop contributions shows a similar behavior. When the inner fermion is heavy, the ALP-photon coupling tends to a constant dependent only on the fermion's mass. When the ALP is the heaviest particle instead, the effective coupling is a linear function of the ALP mass in logarithmic plots and the slope is dictated by the number of flavors in the loop. We showed that neglecting the two-loop contribution results in an overestimation of the effects of up to $\approx 10\,\%$. For the LFV observable $\mu \to e\gamma$ we presented the impact of this finding on the excluded parameter space in the three scenarios where the ALP features additionally to the LFV coupling a coupling to leptons, leptons and up-type quarks, and leptons and down-type quarks, respectively. A similar analysis is currently not possible for the anomalous magnetic moment of the muon, because it is unclear what the SM predicted value is due to the discrepancy of the two generally used methods to extract the hadronic vacuum polarization, i.e. the term that is currently driving the SM error budget.

In the last section we eventually generalized our results to the study of the chromomagnetic moment of the top quark. We found that the insertion of the fermion loop has an impact only when the fermion in the loop itself is a top quark, too. Neglecting this effect however gives rise to an overestimation of the excluded parameter space of about $10\,\%$.

Acknowledgments

M.S. would like to thank Claudia Cornella for useful discussions. The research of M.N. and M.S. are supported by the Cluster of Excellence *Precision Physics, Fundamental Interactions and Structure of Matter* (PRISMA$^+$ – EXC 2118/1) within the German Excellence Strategy (project ID 390831469). M.S. gratefully acknowledges support from the Alexander von Humboldt Foundation as a Feodor Lynen Fellow. M.S. is supported by the United States Department of Energy under Grant Contract No. DESC0012704.

References

[1] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[2] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev.* **D16** (1977) 1791–1797.

[3] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223–226.

[4] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[5] Y. Ema, K. Hamaguchi, T. Moroi and K. Nakayama, *Flaxion: a minimal extension to solve puzzles in the standard model*, *JHEP* **01** (2017) 096, [1612.05492].

[6] L. Calibbi, F. Goertz, D. Redigolo, R. Ziegler and J. Zupan, *Minimal axion model from flavor*, *Phys. Rev.* **D95** (2017) 095009, [1612.08040].

[7] T. Alanne, S. Blasi and F. Goertz, *Common source for scalars: Flavored axion-Higgs unification*, *Phys. Rev. D* **99** (2019) 015028, [1807.10156].

[8] A. Ringwald, *Alternative dark matter candidates: Axions*, *PoS* **NOW2016** (2016) 081, [1612.08933].

[9] C. S. Machado, W. Ratzinger, P. Schwaller and B. A. Stefanek, *Audible Axions*, *JHEP* **01** (2019) 053, [1811.01950].

[10] C. S. Machado, W. Ratzinger, P. Schwaller and B. A. Stefanek, *Gravitational wave probes of axionlike particles*, *Phys. Rev. D* **102** (2020) 075033, [1912.01007].

[11] E. Madge, W. Ratzinger, D. Schmitt and P. Schwaller, *Audible axions with a booster: Stochastic gravitational waves from rotating ALPs*, *SciPost Phys.* **12** (2022) 171, [2111.12730].

[12] J. E. Kim, *Weak Interaction Singlet and Strong CP Invariance*, *Phys. Rev. Lett.* **43** (1979) 103.

[13] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys. B* **166** (1980) 493–506.

[14] M. Dine, W. Fischler and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett. B* **104** (1981) 199–202.

[15] A. R. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*, *Sov. J. Nucl. Phys.* **31** (1980) 260.

[16] A. Ringwald, *Exploring the Role of Axions and Other WISPs in the Dark Universe*, *Phys. Dark Univ.* **1** (2012) 116–135, [1210.5081].

[17] M. Bauer, M. Neubert, S. Renner, M. Schnubel and A. Thamm, *The Low-Energy Effective Theory of Axions and ALPs*, *JHEP* **04** (2021) 063, [2012.12272].

[18] M. Bauer, M. Neubert, S. Renner, M. Schnubel and A. Thamm, *Flavor probes of axion-like particles*, 2110.10698.

[19] M. Chala, G. Guedes, M. Ramos and J. Santiago, *Running in the ALPs*, *Eur. Phys. J. C* **81** (2021) 181, [2012.09017].

[20] A. Davidson and M. A. H. Vozmediano, *The Horizontal Axion Alternative: The Interplay of Vacuum Structure and Flavor Interactions*, *Nucl. Phys. B* **248** (1984) 647–670.

[21] R. D. Peccei, T. T. Wu and T. Yanagida, *A VIABLE AXION MODEL*, *Phys. Lett. B* **172** (1986) 435–440.

[22] L. M. Krauss and F. Wilczek, *A SHORTLIVED AXION VARIANT*, *Phys. Lett. B* **173** (1986) 189–192.

[23] C. Q. Geng and J. N. Ng, *Flavor Connections and Neutrino Mass Hierarchy Invariant Invisible Axion Models Without Domain Wall Problem*, *Phys. Rev. D* **39** (1989) 1449.

[24] A. Celis, J. Fuentes-Martin and H. Serodio, *An invisible axion model with controlled FCNCs at tree level*, *Phys. Lett. B* **741** (2015) 117–123, [1410.6217].

[25] D. S. M. Alves and N. Weiner, *A viable QCD axion in the MeV mass range*, *JHEP* **07** (2018) 092, [1710.03764].

[26] L. Di Luzio, F. Mescia, E. Nardi, P. Panci and R. Ziegler, *Astrophobic Axions*, *Phys. Rev. Lett.* **120** (2018) 261803, [1712.04940].

[27] K. Choi, S. H. Im, C. B. Park and S. Yun, *Minimal Flavor Violation with Axion-like Particles*, *JHEP* **11** (2017) 070, [1708.00021].

[28] J. Martin Camalich, M. Pospelov, P. N. H. Vuong, R. Ziegler and J. Zupan, *Quark Flavor Phenomenology of the QCD Axion*, *Phys. Rev. D* **102** (2020) 015023, [2002.04623].

[29] G. B. Gelmini, S. Nussinov and T. Yanagida, *Does Nature Like Nambu-Goldstone Bosons?*, *Nucl. Phys. B* **219** (1983) 31–40.

[30] A. A. Anselm, N. G. Uraltsev and M. Y. Khlopov, *mu —> e FAMILON DECAY*, *Sov. J. Nucl. Phys.* **41** (1985) 1060.

[31] M. Bauer, M. Neubert and A. Thamm, *LHC as an Axion Factory: Probing an Axion Explanation for $(g-2)_\mu$ with Exotic Higgs Decays*, *Phys. Rev. Lett.* **119** (2017) 031802, [1704.08207].

[32] M. Bauer, M. Neubert and A. Thamm, *Collider Probes of Axion-Like Particles*, *JHEP* **12** (2017) 044, [1708.00443].

[33] M. Bauer, M. Neubert, S. Renner, M. Schnubel and A. Thamm, *Axionlike Particles, Lepton-Flavor Violation, and a New Explanation of a_μ and a_e*, *Phys. Rev. Lett.* **124** (2020) 211803, [1908.00008].

[34] M. Bauer, M. Neubert, S. Renner, M. Schnubel and A. Thamm, *Consistent Treatment of Axions in the Weak Chiral Lagrangian*, *Phys. Rev. Lett.* **127** (2021) 081803, [2102.13112].

[35] M. A. Buen-Abad, J. Fan, M. Reece and C. Sun, *Challenges for an axion explanation of the muon $g-2$ measurement*, *JHEP* **09** (2021) 101, [2104.03267].

[36] W. A. Bardeen, S. H. H. Tye and J. A. M. Vermaseren, *Phenomenology of the New Light Higgs Boson Search*, *Phys. Lett. B* **76** (1978) 580–584.

[37] P. Di Vecchia and G. Veneziano, *Chiral Dynamics in the Large n Limit*, *Nucl. Phys. B* **171** (1980) 253–272.

[38] C. Cornella, P. Paradisi and O. Sumensari, *Hunting for ALPs with Lepton Flavor Violation*, *JHEP* **01** (2020) 158, [1911.06279].

[39] K. G. Chetyrkin, B. A. Kniehl, M. Steinhauser and W. A. Bardeen, *Effective QCD interactions of CP odd Higgs bosons at three loops*, *Nucl. Phys. B* **535** (1998) 3–18, [hep-ph/9807241].

[40] S. M. Barr and A. Zee, *Electric Dipole Moment of the Electron and of the Neutron*, *Phys. Rev. Lett.* **65** (1990) 21–24.

[41] W. J. Marciano, A. Masiero, P. Paradisi and M. Passera, *Contributions of axionlike particles to lepton dipole moments*, *Phys. Rev. D* **94** (2016) 115033, [1607.01022].

[42] A. M. Galda and M. Neubert, *ALP-LEFT Interference and the Muon $(g-2)$*, 2308.01338.

[43] Muon g-2 collaboration, G. W. Bennett et al., *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys. Rev. D* **73** (2006) 072003, [hep-ex/0602035].

[44] Muon g-2 collaboration, B. Abi et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (2021) 141801, [2104.03281].

[45] T. Aoyama et al., *The anomalous magnetic moment of the muon in the Standard Model*, *Phys. Rept.* **887** (2020) 1–166, [2006.04822].

[46] M. Davier, A. Hoecker, B. Malaescu and Z. Zhang, *Reevaluation of the Hadronic Contributions to the Muon g-2 and to alpha(MZ)*, *Eur. Phys. J. C* **71** (2011) 1515, [1010.4180].

[47] T. Aoyama, M. Hayakawa, T. Kinoshita and M. Nio, *Complete Tenth-Order QED Contribution to the Muon g-2*, *Phys. Rev. Lett.* **109** (2012) 111808, [1205.5370].

[48] T. Aoyama, T. Kinoshita and M. Nio, *Theory of the Anomalous Magnetic Moment of the Electron*, *Atoms* **7** (2019) 28.

[49] A. Czarnecki, W. J. Marciano and A. Vainshtein, *Refinements in electroweak contributions to the muon anomalous magnetic moment*, *Phys. Rev. D* **67** (2003) 073006, [hep-ph/0212229].

[50] C. Gnendiger, D. Stöckinger and H. Stöckinger-Kim, *The electroweak contributions to $(g-2)_\mu$ after the Higgs boson mass measurement*, *Phys. Rev. D* **88** (2013) 053005, [1306.5546].

[51] M. Davier, A. Hoecker, B. Malaescu and Z. Zhang, *Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon $g-2$ and $\alpha(m_Z^2)$ using newest hadronic cross-section data*, *Eur. Phys. J. C* **77** (2017) 827, [1706.09436].

[52] A. Keshavarzi, D. Nomura and T. Teubner, *Muon $g-2$ and $\alpha(M_Z^2)$: a new data-based analysis*, *Phys. Rev. D* **97** (2018) 114025, [1802.02995].

[53] G. Colangelo, M. Hoferichter and P. Stoffer, *Two-pion contribution to hadronic vacuum polarization*, *JHEP* **02** (2019) 006, [1810.00007].

[54] M. Hoferichter, B.-L. Hoid and B. Kubis, *Three-pion contribution to hadronic vacuum polarization*, *JHEP* **08** (2019) 137, [1907.01556].

[55] M. Davier, A. Hoecker, B. Malaescu and Z. Zhang, *A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to $\boldsymbol{\alpha(m_Z^2)}$*, *Eur. Phys. J. C* **80** (2020) 241, [1908.00921].

[56] A. Keshavarzi, D. Nomura and T. Teubner, *$g-2$ of charged leptons, $\alpha(M_Z^2)$, and the hyperfine splitting of muonium*, *Phys. Rev. D* **101** (2020) 014029, [1911.00367].

[57] A. Kurz, T. Liu, P. Marquard and M. Steinhauser, *Hadronic contribution to the muon anomalous magnetic moment to next-to-next-to-leading order*, *Phys. Lett. B* **734** (2014) 144–147, [1403.6400].

[58] K. Melnikov and A. Vainshtein, *Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment revisited*, *Phys. Rev. D* **70** (2004) 113006, [hep-ph/0312226].

[59] P. Masjuan and P. Sanchez-Puertas, *Pseudoscalar-pole contribution to the $(g_\mu-2)$: a rational approach*, *Phys. Rev. D* **95** (2017) 054026, [1701.05829].

[60] G. Colangelo, M. Hoferichter, M. Procura and P. Stoffer, *Dispersion relation for hadronic light-by-light scattering: two-pion contributions*, *JHEP* **04** (2017) 161, [1702.07347].

[61] M. Hoferichter, B.-L. Hoid, B. Kubis, S. Leupold and S. P. Schneider, *Dispersion relation for hadronic light-by-light scattering: pion pole*, *JHEP* **10** (2018) 141, [1808.04823].

[62] A. Gérardin, H. B. Meyer and A. Nyffeler, *Lattice calculation of the pion transition form factor with $N_f = 2+1$ Wilson quarks*, *Phys. Rev. D* **100** (2019) 034520, [1903.09471].

[63] J. Bijnens, N. Hermansson-Truedsson and A. Rodríguez-Sánchez, *Short-distance constraints for the HLbL contribution to the muon anomalous magnetic moment*, *Phys. Lett. B* **798** (2019) 134994, [1908.03331].

[64] G. Colangelo, F. Hagelstein, M. Hoferichter, L. Laub and P. Stoffer, *Longitudinal short-distance constraints for the hadronic light-by-light contribution to $(g-2)_\mu$ with large-N_c Regge models*, *JHEP* **03** (2020) 101, [1910.13432].

[65] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung et al., *Hadronic Light-by-Light Scattering Contribution to the Muon Anomalous Magnetic Moment from Lattice QCD*, *Phys. Rev. Lett.* **124** (2020) 132002, [1911.08123].

[66] G. Colangelo, M. Hoferichter, A. Nyffeler, M. Passera and P. Stoffer, *Remarks on higher-order hadronic corrections to the muon g−2*, *Phys. Lett. B* **735** (2014) 90–91, [1403.7512].

[67] S. Borsanyi et al., *Leading hadronic contribution to the muon magnetic moment from lattice QCD*, *Nature* **593** (2021) 51–55, [2002.12347].

[68] MEG collaboration, A. M. Baldini et al., *Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, *Eur. Phys. J. C* **76** (2016) 434, [1605.05081].

[69] D. Atwood, A. Kagan and T. G. Rizzo, *Constraining anomalous top quark couplings at the Tevatron*, *Phys. Rev. D* **52** (1995) 6264–6270, [hep-ph/9407408].

[70] P. Haberl, O. Nachtmann and A. Wilch, *Top production in hadron hadron collisions and anomalous top - gluon couplings*, *Phys. Rev. D* **53** (1996) 4875–4885, [hep-ph/9505409].

[71] K.-m. Cheung, *Probing the chromoelectric and chromomagnetic dipole moments of the top quark at hadronic colliders*, *Phys. Rev. D* **53** (1996) 3604–3615, [hep-ph/9511260].

[72] CMS collaboration, A. M. Sirunyan et al., *Measurement of the top quark forward-backward production asymmetry and the anomalous chromoelectric and chromomagnetic moments in pp collisions at $\sqrt{s}$ = 13 TeV*, *JHEP* **06** (2020) 146, [1912.09540].

[73] CMS collaboration, A. M. Sirunyan et al., *Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Rev. D* **100** (2019) 072002, [1907.03729].

[74] A. M. Galda, M. Neubert and S. Renner, *ALP — SMEFT interference*, *JHEP* **06** (2021) 135, [2105.01078].