arXiv:2402.17518v1 [hep-lat] 27 Feb 2024

The Strong CP Problem in the Quantum Rotor

D. Albandea,* G. Catumba,† and A. Ramos‡
IFIC (CSIC-UVEG), Edificio Institutos Investigación, Apt. 22085, E-46071 Valencia, Spain
(Dated: February 28, 2024)

Recent studies have claimed that the strong CP problem does not occur in QCD, proposing a new order of limits in volume and topological sectors when studying observables on the lattice. In order to shed light on this issue, we study the effect of the topological θ-term on a simple quantum mechanical rotor that allows a lattice description. The topological susceptibility and the θ-dependence of the energy spectrum are both computed using local lattice correlation functions. The sign problem is overcome by considering Taylor expansions in θ exploiting automatic differentiation methods for Monte Carlo processes. Our findings confirm the conventional wisdom on the strong CP problem.

I. INTRODUCTION

The strong CP problem remains as one of the puzzles of the Standard Model: why do the strong interactions conserve CP? In principle, the Lagrangian of Quantum Chromodynamics (QCD) admits a renormalizable gauge invariant θ-term,

$$\delta\mathcal{L}_\theta = \frac{\theta}{16\pi^2}\mathrm{tr}\left(F_{\mu\nu}\tilde{F}^{\mu\nu}\right) , \tag{1}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ is the field strength and A_μ the gauge connection. While the presence of such a term would break CP symmetry in QCD, experimental measurements of the electric dipole moment of the neutron constrain the coupling to be $|\theta| \lesssim 10^{-10}$ [1].

Many solutions have been proposed over the years. The existence of a massless quark would make the θ phase unphysical, but recent lattice simulations clearly disfavor such a solution [2]. Additionally, various alternatives beyond the Standard Model, such as a Peccei-Quinn symmetry [3] and Nelson-Barr type models [4, 5], have been explored.

It has been claimed in refs. [6, 7] that the effect of a such a θ-term, even if present, would not lead to observable consequences. These works argue that when computing correlation functions, the infinite volume limit should be taken *before* the summation over topological sectors. The consequence of such an order of limits (infinite volume at each fixed topological charge) would be the absence of θ-dependence from observables, and particularly from the energy spectrum of the theory.

We claim that, in fact, the order of limits is only important if one insists in computing *global* observables, i.e. those computed as an integral over the whole Euclidean space. Determining these observables requires extreme care with the role of boundary conditions and in taking the infinite volume limit. On the other hand they also represent quite unphysical setups, since we do not need to know the boundary conditions of the universe to measure the topological susceptibility, the proton mass, or the neutron electric dipole moment: all physical quantities of interest can be extracted from local correlators. Due to clustering, the dependence of local correlators on the boundary conditions or the finite volume is exponentially suppressed in theories with a mass gap, making the order of limits largely irrelevant.

In this paper, we aim to shed some light on this question by examining a simple toy model that, nevertheless, shares some key characteristics with QCD: the quantum rotor. We will compute the topological susceptibility from local correlators and show that, up to finite volume corrections, the result is independent of the choice of boundary conditions. The analytical computations will also be supported by numerical lattice simulations,[1] validating the approach used by the lattice community to answer these questions in QCD [8–10]. Additionally, we will determine the topological susceptibility using *master field* simulations [11], where a single gauge configuration in a very large Euclidean volume allows to determine expectation values as volume averages. Both local computations give the same non-zero value for the topological susceptibility, χ_t. Moreover, we will show that the spectrum of the theory has a dependence on θ.

Finally, the numerical study of this rather simple system faces several challenging problems present in lattice QCD. We will use several recent proposals to overcome these problems, making the study of this particular model a good test-bed for many state-of-the-art lattice techniques. Firstly, the issue of *topology freezing* [12–15] when approaching the continuum limit is solved by the use of winding transformations [16]. Secondly, simulating the theory at a non-zero value of θ leads to the so-called *sign problem* —see [17, 18] and references therein. We employ three different methods to overcome this, which

* david.albandea@uv.es
† gtelo@ific.uv.es
‡ alberto.ramos@ific.uv.es

[1] The code used for the simulations and analysis can be found in https://github.com/dalbandea/QuantumRotorExperiments.jl and https://igit.ific.uv.es/gtelo/qrotor.

rely on the use of imaginary values for θ along with analytical continuation [19–29]: on the one hand, we will extract the θ dependence by a fit to data from several simulations at different imaginary values of θ; on the other hand, we will explore two recent proposals that allow to extract series expansions in θ using automatic differentiation for Monte Carlo processes [30].

The paper is organized as follows: in section II we introduce the quantum rotor, along with the action and topological charge definitions in the continuum and on the lattice, including two different choices of discretization; in section III we study the order of limits argued in refs. [6, 7] and the conventional one using both global and local observables, where the latter are computed with lattice simulations; in section IV we obtain the continuum θ-dependence of the energy spectrum from lattice simulations using different boundary conditions —this computation requires overcoming the sign problem, whose proposed solutions are introduced in section IV A.

II. THE QUANTUM ROTOR

The quantum rotor describes a free particle of mass m on a ring of radius R. The system admits a θ-term and can be formulated at finite temperature $\beta = 1/T$ as a path integral with partition function [31, 32]

$$Z(\theta) = \int \mathcal{D}\phi \, e^{-S(\phi)+i\theta Q}, \tag{2}$$

where the action reads

$$S(\phi) = \frac{I}{2}\int_0^T dt\, \dot{\phi}(t)^2\,, \tag{3}$$

with $\phi(0) = \phi(T)$ and $-\pi < \phi(t) \leq \pi$, and where we have defined the moment of inertia $I = mR^2$. Note that using periodic boundary conditions leads to the quantization of the topological charge,

$$Q = \frac{1}{2\pi}\int_0^T dt\, \dot{\phi}(t) \ \in \mathbb{Z}. \tag{4}$$

To study the system on a lattice we divide the time extent $T \equiv \hat{T}a$ into a lattice of $\hat{T}$ points with spacing a, and express the moment of inertia measured in units of this lattice spacing as $\hat{I} = I/a$. There are several possibilities for the discretization of the lattice action, with the only condition being that the continuum action and topological charge in eqs. (3) and (4) are recovered in the limit $\hat{I} \to \infty$ with

$$\frac{T}{I} = \frac{\hat{T}}{\hat{I}} = \text{constant}. \tag{5}$$

Particularly, we will use the so-called *classical perfect* action [33],

$$S_{\text{cp}}(\phi) = \frac{\hat{I}}{2}\sum_{t=0}^{\hat{T}-1}((\phi_{t+1}-\phi_t) \bmod 2\pi)^2, \tag{6}$$

as well as the *standard* action,

$$S_{\text{st}}(\phi) = \frac{\hat{I}}{2}\sum_{t=0}^{\hat{T}-1}(1-\cos(\phi_{t+1}-\phi_t))\,. \tag{7}$$

One can show that the classical perfect action has smaller lattice artifacts and therefore a better behavior as one approaches the continuum limit [32]. However, the standard discretization is useful when using sampling algorithms that require the derivative of the action with respect to the field to be defined at all points of its domain. Particularly, we will use the Hybrid Monte Carlo (HMC) [34, 35] algorithm to overcome the sign problem when studying the θ dependence of the spectrum of the theory.

Analogously, we define the classical perfect topological charge,

$$Q_{\text{cp}} = \frac{1}{2\pi}\sum_{t=0}^{\hat{T}-1}((\phi_{t+1}-\phi_t) \bmod 2\pi), \tag{8}$$

and the standard topological charge,

$$Q_{\text{st}} = \frac{1}{2\pi}\sum_{t=0}^{\hat{T}-1}\sin(\phi_{t+1}-\phi_t). \tag{9}$$

Note that with periodic boundary conditions the classical perfect topological charge is exactly an integer, while the standard one is only an integer in the continuum limit.

III. ON THE ORDER OF LIMITS OF INFINITE VOLUME AND TOPOLOGICAL SECTORS

For theories with topology the computation of an observable can be decomposed as a sum of contributions from the different topological sectors,

$$\langle\mathcal{O}\rangle = \sum_Q \langle\mathcal{O}\rangle_Q \, p(Q), \tag{10}$$

where $p(Q)$ is the probability density of the topological sector with charge Q, and $\langle\ldots\rangle_Q$ denotes expectation value at fixed topological sector Q. While this equation is general, it requires the topological charge Q to be quantized. For gauge quantum field theories, the usual reasoning for the quantization of the topological charge arises from the requirement of finite saddle solutions for the action,[2] which constrains the gauge configurations at infinity to be pure gauge (i.e. a gauge transformation of zero).

[2] Not only the requirement of finite action solutions comes from the use of a semiclassical approximation, but also the topological quantization can be obtained for a finite volume with appropriate boundary conditions [36].

In practice, lattice calculations are restricted to finite Euclidean volumes. This, together with the above reasoning, led the authors of refs. [6, 7] to challenge the conventional order of limits to extract observables at infinite volume. Namely, they propose to compute observable quantities as

$$\langle \mathcal{O} \rangle = \lim_{N\to\infty} \lim_{T\to\infty} \sum_{|Q|<N} \langle \mathcal{O} \rangle_Q \, p(Q), \tag{11}$$

such that the volume is taken to infinity before the contributions from all the topological sectors are summed.

This claim can be verified in the quantum rotor, which can be trivially solved using the quantum mechanical formalism [32]. The energy levels of the system read

$$E_n = \frac{1}{2I}\left(n - \frac{\theta}{2\pi}\right)^2, \quad n \in \mathbb{Z}\,. \tag{12}$$

In Euclidean time with periodic boundary conditions we have the thermal partition function

$$Z(\theta) = \sum_{n\in\mathbb{Z}} e^{-TE_n} = \sum_{n\in\mathbb{Z}} e^{-\frac{T}{2I}\left(n-\frac{\theta}{2\pi}\right)^2}, \tag{13}$$

from which we can obtain the second moment of the topological charge at $\theta = 0$,

$$\begin{aligned}\langle Q^2 \rangle &= -\frac{\partial \log Z(\theta)}{\partial\theta}\Big|_{\theta=0} \\ &= \frac{T}{4I\pi^2}\left(1 - \frac{\sum_n n^2 e^{-\frac{T}{2I}n^2}}{\sum_n e^{-\frac{T}{2I}n^2}}\right).\end{aligned} \tag{14}$$

Note that for large T we have

$$\frac{\langle Q^2\rangle}{T} = \frac{1}{4I\pi^2}\left[1 - \frac{T}{I}\left(e^{-\frac{T}{2I}} - e^{-\frac{T}{I}}\right) + \dots\right], \tag{15}$$

i.e. a non-zero value with exponentially small corrections due to the finite size of the system. This leads to the usual result for the topological susceptibility,

$$\chi_t = \lim_{T\to\infty} \frac{\langle Q^2\rangle}{T} = \frac{1}{4I\pi^2}\,. \tag{16}$$

We can, instead, attempt to take the infinite T limit before summing over topological sectors. To do this, we can determine the probability distribution of each topological sector of charge Q. Considering $\theta = 0$ for simplicity we have

$$\begin{aligned} p(Q) =& \frac{1}{2\pi Z(0)} \int_{-\pi}^{\pi} d\theta\, Z(\theta) e^{-i\theta Q} \\ =& \frac{1}{Z(0)}\sqrt{\frac{2I\pi}{T}} \exp\left(-\frac{2I\pi^2}{T}Q^2\right). \end{aligned} \tag{17}$$

According to [6] the correct order of limits should now read

$$\begin{aligned}\chi_t &= \lim_{N\to\infty}\lim_{T\to\infty} \frac{1}{T}\sum_{|Q|<N} Q^2\, p(Q) \\ &= \lim_{N\to\infty}\lim_{T\to\infty} \frac{1}{T}\frac{\sum_{|Q|\le N} Q^2 \exp\left(-\frac{2\pi^2 I}{T}Q^2\right)}{\sum_{|Q|\le N}\exp\left(-\frac{2\pi^2 I}{T}Q^2\right)},\end{aligned} \tag{18}$$

which is trivially zero, and therefore contradicts the result in eq. (16).

To shed some light into this discrepancy, we propose another approach in which the order of limits is irrelevant. Due to locality, Euclidean two-point functions decay exponentially at large time separations, a fact that can be exploited to compute the topological susceptibility from local correlators. If one defines the topological charge density as $q(t) = \dot{\phi}(t)$ we have

$$\begin{aligned}\langle Q^2\rangle &= \sum_{t_1,t_2}\langle q(t_1)q(t_2)\rangle = T\sum_t \langle q(t)q(0)\rangle \\ &= T\sum_{t<R}\langle q(t)q(0)\rangle + \mathcal{O}(e^{-R/\xi})\,,\end{aligned} \tag{19}$$

where the last sum runs over a time extent R and ξ represents the correlation length of the system, which for the quantum rotor reads $\xi = (E_1 - E_0)^{-1} = 2I$. Therefore one can compute the topological susceptibility in a localized region of radius $R \ll T$,

$$\chi_t = \sum_{t<R}\langle q(t)q(0)\rangle + \mathcal{O}(e^{-R/\xi}), \tag{20}$$

where contributions from $t > R$ are exponentially suppressed. The following observations follow:

1. Since $R \ll T$ and finite volume effects in eq. (20) are $\mathcal{O}(e^{-T/\xi})$, they are even more exponentially suppressed. As we will see, any choice of boundary conditions must give the same result.

2. In eq. (20) there is no sum over topological sectors and, in fact, $\frac{1}{2\pi}\sum_{t<R} q(t)$ is not quantized. There is no order of limits to discuss.

It is easy to compute observables in the quantum rotor using open boundary conditions.[3] Concretely, the two-point correlation function of the topological density, which we derive in app. A, yields

$$\langle q(t_1)q(t_2)\rangle_{\text{cp, OBC}} = \delta_{t_1,t_2}\left[\frac{1}{4I\pi^2} + \mathcal{O}(I^{-2})\right]. \tag{21}$$

In this particular case, finite time-extent effects are completely absent and one can just set $R = 1$ in eq. (20). In the continuum limit we then have

$$\lim_{I\to\infty} I\chi_t = \frac{1}{4\pi^2}, \tag{22}$$

which coindices with the result in eq. (16).

[3] This amounts to setting $\phi_{\hat{T}} = \phi_{\hat{T}-1}$ in eqs. (6) to (9), which corresponds to Neumann boundary conditions in the continuum, i.e. $\dot{\phi}(T) = \dot{\phi}(0) = 0$.

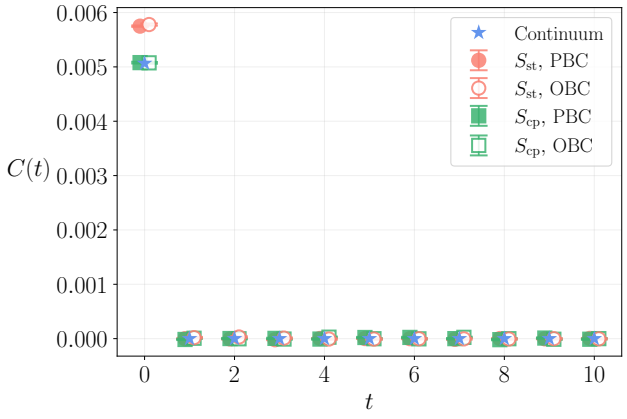


FIG. 1. The quantity $C(t) \equiv \langle q(t)q(0)\rangle$ as a function of t from simulations at $\hat{I} = 5$ and $\hat{T} = 100$ with the standard action (red circles) and classical perfect action (green squares) with open boundary conditions (open symbols) and periodic boundary conditions (filled symbols). The blue stars represent the continuum result.

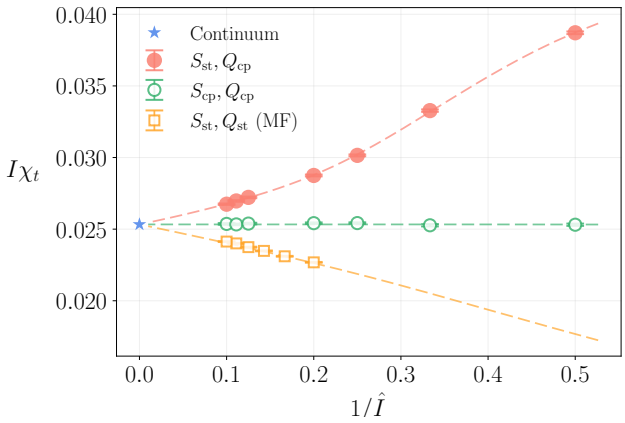


FIG. 2. Continuum extrapolation of χ_t computed from $\langle q(t)^2\rangle$ at various $\{\hat{T}, \hat{I}\}$ at constant $T/I = 20$ with the standard action and classical topological charge (red, filled circles), with the classical action and topological charge (green, open circles) and from master field simulations with the standard action and topological charge (orange, open squares), all with periodic boundary conditions. The topological charges of the master fields, from smaller to larger value of $\hat{I}$, are $Q = -83, -61, -233, 11, 24, -118$. The dashed lines represent the corresponding analytical results with open boundary conditions.

One can use lattice simulations to obtain the same result using periodic boundary conditions instead. Although in this model topology freezing is already apparent at rather small values of the coupling $\hat{I}$, we overcome it by using a combination of standard Metropolis updates [37] and winding transformations [16] (see app. C).

In fig. 1 we show the topological two-point function as a function of t for the standard action (red circles), and the classical perfect action (green squares), as well as for open (open symbols) and periodic (filled symbols) boundary conditions. The continuum result is also shown (blue stars). It is worth noting that the case with periodic boundary conditions should have a similar t-dependence as the correlator for open boundaries, $\langle q(t_1)q(t_2)\rangle \propto \delta_{t_1,t_2}$, up to exponentially small finite-volume effects.

On the other hand, the discrepancy between the standard and classical perfect actions is a discretization effect that disappears taking the continuum limit. This can be seen in fig. 2, where we show the continuum extrapolation of $I\chi_t$ with simulations at constant $T/I = 20$ for the classical perfect action (green, open circles) and for the standard action (red, filled circles) with periodic boundary conditions. The red and green dashed lines are the corresponding analytic curves with open boundary conditions, and one can see that both choices of discretization and boundary conditions give the same non-zero results for $\hat{I} \to \infty$.

Finally, the orange squares come from master field simulations using the standard action with periodic boundary conditions. Concretely, every point comes from the volume average of the local definition of the topological susceptibility in eq. (20) on a single master field configuration with $T = 10^6$, and with the standard definition of the topological charge of eq. (9). Although these configurations can have very large values of the topological charge, the result, once extrapolated to the continuum, matches eq. (16).

IV. ON THE θ DEPENDENCE OF THE SPECTRUM

The spectrum of the quantum rotor is changed by the presence of the θ-term, as seen in eq. (12). However, and similarly to the previous section, we are interested in obtaining this dependence explicitly from local correlators on the lattice. The obvious challenge in this computation is the sign problem arising from the imaginary term in the partition function of eq. (2) when $\theta \neq 0$.

A. Overcoming the sign problem on the lattice

A common approach around the sign problem is to perform simulations at an imaginary value of θ by defining $\theta_I = i\theta \in \mathbb{R}$. This makes the integrand in eq. (2) become real and allows to investigate the θ-dependence of observables using standard sampling algorithms. Analyticity around $\theta = 0$ of the observable O under study is also a required assumption, i.e.

$$O(\theta) = O^{(0)} + O^{(1)}\theta + O^{(2)}\theta^2 + \mathcal{O}(\theta^3), \qquad (23)$$

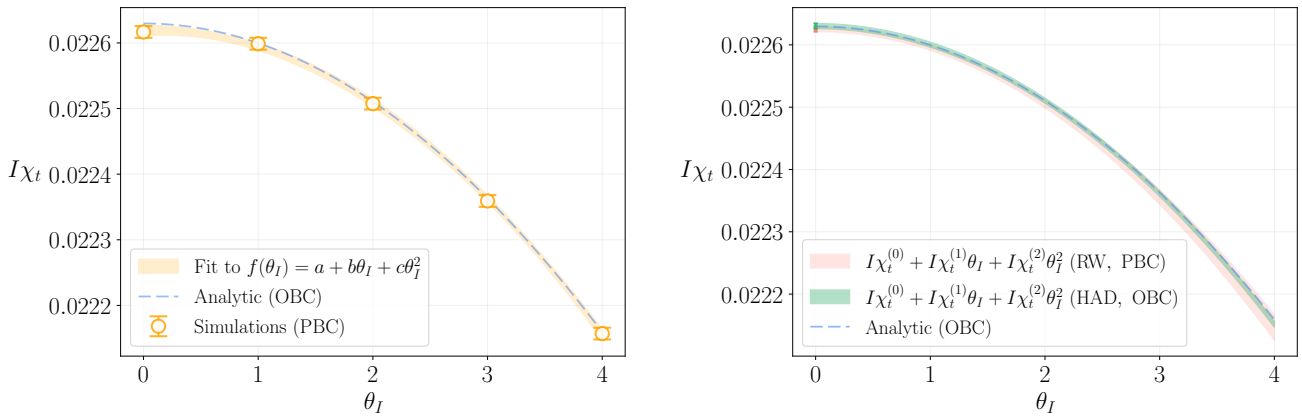


FIG. 3. θ_I dependence of $I\chi_t$ at $\hat{I} = 5$ and $\hat{T} = 100$ with periodic boundary conditions and with different methods. Left frame: results (orange circles) from 5 different simulations with 100k uncorrelated configurations each, along with their fit to the functional form $f(\theta_I) = a + b\theta_I + c\theta_I^2$. Right frame: results from single simulations with 500k uncorrelated configurations using reweighting and HAD (red and green points) at $\theta_I = 0$, along with the curve $I\chi_t^{(0)} + I\chi_t^{(1)}\theta_I + I\chi_t^{(2)}\theta_I^2$ (light-red and thick-green bands) obtained from the use of truncated polynomials. The reweighting was performed on a simulation with periodic boundary condition, while HAD was used with open boundary conditions. In both frames the analytic result from open boundary conditions is displayed (dashed line).

with the expansion coefficients

$$O^{(n)} = \frac{1}{n!}\frac{\partial^n O}{\partial\theta^n}\bigg|_{\theta=0}. \tag{24}$$

From this we can obtain the expansion coefficient $O^{(n)}$ from analytic continuation and fits to data. Particularly, in fig. 3 (left) we show results of $I\chi_t$ at different θ_I at $\hat{I} = 5$, along with a fit to the functional form $f(x) = a + bx + cx^2$ with parameters $a, b, c \in \mathbb{R}$.

This method has been used recently to study the θ-dependence of the energy spectrum of four-dimensional SU(N) Yang–Mills theories [29]. One of its drawbacks, however, is that it requires several simulations to extract the different expansion coefficients.

Alternatively, it is possible obtain arbitrarily high expansion coefficients by performing a single simulation where the Taylor expansion in θ is done automatically using the algebra of truncated polynomials [38]. From the point of view of the lattice implementation, the idea is to define a new field, $\tilde{\phi}$, as a power series expansion in θ up to some order K, i.e.

$$\tilde{\phi}(\theta) \equiv \tilde{\phi}^{(0)} + \tilde{\phi}^{(1)}\theta + \tilde{\phi}^{(2)}\theta^2 + \cdots + \tilde{\phi}^{(K)}\theta^K. \tag{25}$$

All basic operations and functions acting on these truncated polynomials are defined to be exact at each order, and therefore the computation of observables preserves the correct derivatives with respect to θ. For more details about the implementation of automatic differentiation[4] using truncated polynomials and applications to stochastic processes, see [30].

The simplest application of truncated polynomials to extract higher order derivatives is by using reweighting techniques. Starting from an already existent ensemble of configurations generated at $\theta = 0$, reweighting allows the computation of expectation values for $\theta \neq 0$ by using the identity

$$\langle O(\phi)\rangle_\theta = \frac{\langle e^{i\theta Q} O(\phi)\rangle_{\theta=0}}{\langle e^{i\theta Q}\rangle_{\theta=0}}, \tag{26}$$

where $\langle\cdots\rangle_\theta$ denotes expectation value with respect to the probability distribution $p(\phi) = \frac{1}{Z(\theta)} e^{-S(\phi)+i\theta Q}$.

To extract the Taylor expansion of $\langle O(\phi)\rangle_\theta$ with respect to θ up to order K we replace θ with a truncated polynomial $\tilde{\theta} = \sum_{k=0}^{K} \tilde{\theta}^{(k)}\theta^k$ where $\tilde{\theta}^{(1)} = 1$ and $\tilde{\theta}^{(k\neq 1)} = 0$. The factor $\exp(i\theta Q)$ and also the numerator and denominator of eq. (26) thus become truncated polynomials. Therefore, by performing reweighting on a single ensemble we obtain the full analytical dependence of the polynomial expansion. This is shown in fig. 3 (right), where by reweighting a single standard simulation at $\theta_I = 0$ with periodic boundary conditions we automatically obtain $I\chi_t^{(k)}$; particularly, we show $I\chi_t^{(0)} + I\chi_t^{(1)}\theta_I + I\chi_t^{(2)}\theta_I^2$ in the (light) red band, which agrees with the analytical result (dashed line) obtained with open boundary conditions.

An alternative application of truncated polynomials consists in a modification of sampling algorithms based on Hamiltonian dynamics inspired in Numerical Stochastic Perturbation theory [40, 41]. Particularly, we consider a modification of the HMC algorithm which we denote as Hamiltonian Automatic Differentiation (HAD), but

[4] Truncated polynomials form the basis of the forward implementation of automatic differentiation [39].

Method	$\chi_t^{(0)} \times 10^{-3}$	$\chi_t^{(2)} \times 10^{-6}$
Fit	4.5238(16)	-6.08(48)
Reweighting	4.52501(76)	-5.99(25)
HAD	4.52604(83)	-5.980(34)

TABLE I. Comparison of errors between different methods to obtain the θ_I-dependence, namely, the quadratic fit to the results from direct simulations at θ_I (fig. 3 left), reweighting, and HAD (fig. 3 right). These results correspond to simulations with $\hat{I} = 5$ and $\hat{T} = 100$, using the standard definitions of the action and topological charge.

this modification can be easily applied to other sampling methods. For the quantum rotor, since differentiability with respect to the field is required in the HMC equations of motion, it is necessary to use the standard discretization of the action and topological charge in eqs. (7) and (9). Concretely, with periodic boundary conditions the equations of motion read

$$\begin{aligned}\dot{\phi}_t =& \frac{\partial H(\pi,\phi)}{\partial \pi_t} = \pi_t, \\ \dot{\pi}_t =& -\frac{\partial H(\pi,\phi)}{\partial \phi_t} \\ =& -I\left[\sin(\phi_t - \phi_{t-1}) - \sin(\phi_{t+1} - \phi_t)\right] \\ &+ \theta_I \frac{1}{2\pi}\left[\cos(\phi_t - \phi_{t-1}) - \cos(\phi_{t+1} - \phi_t)\right],\end{aligned} \tag{27}$$

where $t \in [0, \hat{T} - 1]$, the dot denotes derivative with respect to Monte Carlo time, π is the canonical conjugate momenta of ϕ, and $H(\pi,\phi) = \sum_{t=0}^{\hat{T}-1} \pi_t^2/2 + S_{\rm cp}(\phi) - \theta_I Q_{\rm cp}$ is the Hamiltonian of the system.

To extract deritatives with respect to θ_I we replace it by a truncated polynomial, $\tilde{\theta}_I$. This way, ϕ and π become truncated polynomials as well, and eq. (27) becomes a set of differential equations where each order involves only terms of the same or lower order.[5]

The use of truncated polynomials within the solver of the equations of motion automatizes the sampling. Therefore, after a single HMC simulation with truncated polynomials[6] we obtain a Markov chain of N samples, $\{\tilde{\phi}_{(i)}\}_{i=1}^N$, that carry the derivatives with respect to θ_I. The Taylor expansion of observables is obtained by the computation of conventional expectation values using these samples.

In fig. 3 (right) we show the curve $I\chi_t^{(0)} + I\chi_t^{(1)}\theta_I + I\chi_t^{(2)}\theta_I^2$ in (thick) green obtained from a single HAD simulation, and one can appreciate that the predictions for high θ_I are more accurate than the ones obtained by reweighting. This can be seen more transparently in tab. I, where we show the results for $\chi_I^{(0)}$ and $\chi_I^{(2)}$ obtained with the three methods with an equivalent amount of statistics. Focusing on $\chi_t^{(2)}$ one can see that the reweighting result has an improved accuracy with respect to the value obtained by simulating directly at imaginary values of θ; on the other hand, the error obtained with HAD is an order of magnitude smaller, although one should also consider the additional computational overhead coming from the operations of truncated polynomials.[7]

B. Spectrum from lattice correlators

To extract the energy spectrum on the lattice we use the spectral decomposition of an operator O,

$$C(t) = \langle O(t)O(0)\rangle = \sum_k \langle 0|\hat{O}|k\rangle\langle k|\hat{O}|0\rangle e^{-t\Delta E_k}, \tag{28}$$

where $\Delta E_k = E_k - E_0$ are the energies of the system relative to the ground state. For $t \gg 1$ the only non-negligible contribution comes from the first energy difference,

$$C(t) \propto |\langle 1|\phi|0\rangle|^2 \left(e^{-\Delta E_1 t} + e^{-\Delta E_{-1} t}\right), \tag{29}$$

where we have used that $|\langle 1|\phi|0\rangle|^2 = |\langle -1|\phi|0\rangle|^2$. If we now do a series expansion in the energy difference,

$$\Delta E_n(\theta) = \Delta E_n^{(0)} + \Delta E_n^{(1)}\theta + \mathcal{O}(\theta^2), \tag{30}$$

we get

$$C(t) \propto e^{-Et}\left[1 + \frac{1}{2}\theta^2 t^2 (\Delta E_1^{(1)})^2 + \mathcal{O}(\theta^4)\right],$$

where we have used that $\Delta E_{-1}^{(1)} = \Delta E_1^{(1)}$. Therefore, by computing the correlator on the lattice with the methods

[5] The convergence of the equations of motion for each order is guaranteed for large values of $\hat{I}$. This is seen by noticing that the equation for each order can be rewritten as

$$\ddot{\phi}_t^{(n)} = -\frac{\partial^2 S}{\partial \phi_t^2}\phi_t^{(n)} + \text{lower order terms},$$

and that convergence requires the positivity of the linear factor

$$\frac{\partial^2 S}{\partial \phi_t^2} = \hat{I}\left[\cos(\phi_t - \phi_{t-1}) + \cos(\phi_{t+1} - \phi_t)\right].$$

Since $(\phi_t - \phi_{t-1}) \to 0$ in the continuum, $\hat{I} \to \infty$, the factor $\frac{\partial^2 S}{\partial \phi_t^2}$ remains positive. For all simulations performed, the absolute value of $\phi_{t+1} - \phi_t$ remained in the region where convergence is guaranteed.

[6] In the usual HMC formalism, the Metropolis accept–reject step is included to correct for the errors in the numerical integration of the equations of motion. Since this discrete step is not differentiable, it is not possible to perform together with the expansion of the Hamiltonian. However, it is enough to use a fine enough integration step-size such that the possible extrapolation to vanishing step-size is below statistical uncertainties.

[7] For a more detailed cost comparison between the reweighting and HAD methods, see [30].

outlined in sec. IV A, one can obtain $\Delta E^{(1)}$ from its large-t behavior. Concretely, with open boundary conditions one can perform a quadratic fit to the expression

$$\frac{C^{(2)}(t)}{C^{(0)}(t)} = \frac{1}{2}\Delta E_1^{(1)} t^2, \tag{31}$$

while for periodic boundary conditions the expression to fit reads (see app. B for a derivation)

$$\frac{C^{(2)}(t)}{C^{(0)}(t)} = \frac{1}{2}(\Delta E_1^{(1)})^2 \left[t^2 + \frac{(T^2 - 2tT)}{1 + e^{-\Delta E_1^{(0)}(T-2t)}} \right]. \tag{32}$$

To obtain the energy spectrum we use two different interpolating operators, $O_1(t) = \phi_t$ and $O_2(t) = \sin(\phi_t)$. Analogously to the study of the topological susceptibility in sec. III, we extract $\Delta E_1^{(1)}$ for increasing values of $\hat{I}$ following the line of constant physics $T/I = 20$ in order to take the continuum limit. In particular, in fig. 4 we plot the results for the dimensionless quantity $I\Delta E_1^{(1)}$, where all datapoints come from simulations with the standard action in eq. (7).

Irrespective of the choice of boundary conditions, the results obtained from local correlators match the analytical results that can be obtained with open boundary conditions (see app. B), and they have the correct continuum limit obtained in eq. (12) using the quantum mechanical formalism. It is also interesting to note that, while for this model the topological quantization depends only on the choice of boundary conditions, the analytic θ-dependence is nevertheless obtained even for choices that do not generate topological quantization.

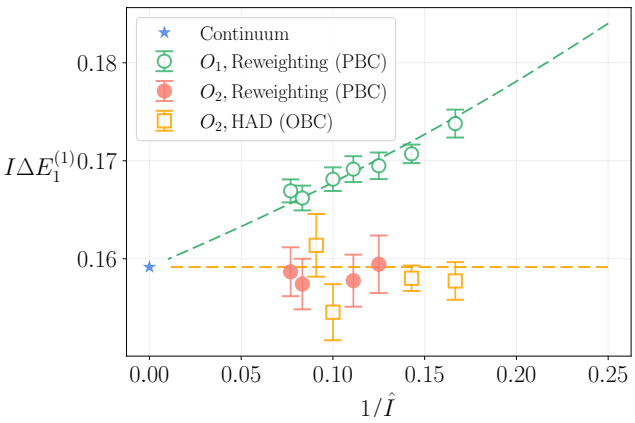


FIG. 4. Linear θ-correction to the first energy level, ΔE_1, for a periodic lattice with the standard discretization of the action at different values of $\hat{I}$ with constant $T/I = 20$. The results were obtained using the interpolating operator $O_1(t) = \phi_t$ with reweighting on a simulation with periodic boundary conditions (green, open circles), and using $O_2(t) = \sin(\phi_t)$ with HAD (yellow, open squares) and reweighting (red, filled circles) with open and periodic boundary conditions, respectively. The dashed lines represent the corresponding the analytic results obtained with open boundary conditions (see eq. (B3)).

V. CONCLUSIONS

We have studied a well-known toy model of QCD to explore a recent claim to solve the strong CP problem, where it is argued that physical observables do not depend on the θ angle of QCD because the infinite volume limit should be taken before the summation over topological sectors, an order of limits which makes θ disappear from the energy spectrum of the theory. Concretely, the authors of refs. [6, 7] consider global observables (i.e. integrals over all spacetime), where the order of limits is indeed relevant and subtleties about the choice of boundary conditions have to be addressed with care. For the case of local observables, their claim is based on the requirement of topological quantization, which is usually seen in the limit of infinite spacetime.

In the present work we argue that quantities of physical interest that can be compared with experiment are intrinsically local and therefore their dependence on the choice of boundary conditions or global questions such as the quantization of the global topological charge are completely irrelevant. We have studied observables in the quantum rotor model from local correlators, and we have shown that the requirement of topological quantization at any finite value of the lattice spacing and volume is unimportant. In fact, we have seen that even in cases for which the topology is not quantized, the local correlators deliver the correct θ-dependence of the observables.

Particularly we have determined the topological susceptibility and the spectrum of the theory from local correlators with different choices of boundary conditions (open and periodic), and also at fixed topological charge using volume averages on very large volumes (using master field simulations). We find perfect agreement in the θ-dependence between all the methods and the analytical results.

For the computation of the θ-dependence of the observables on the lattice we have studied two recent proposals that might yield a computational advantage with respect to the standard ways of alleviating the sign problem. These new proposals involve using truncated polynomials to automatically Taylor expand observables in θ. These methods have the advantage that one can obtain the full dependence on the first orders of θ from a single simulation at $\theta = 0$. Concretely, we have studied truncated polynomials using reweighting techniques and the HMC algorithm.

We conclude that at least for the quantities of physical interest that can be derived from local correlators, such as the susceptibility or the electric dipole moment of the neutron in the case of QCD, the order of limits is irrelevant, and a dependence on θ in the observables is still present. We still fail to understand why QCD decides to respect CP with such a high accuracy.

AKNOWLEDGEMENTS

We specially want to thank C. Tamarit for his patience explaining to us the arguments of [6, 7], P. Hernandez for the extremely useful discussions, and S. Cruz-Alzaga for his contribution at the early stages of this work. We acknowledge support from the Generalitat Valenciana grant PROMETEO/2019/083, the European projects H2020-MSCA-ITN-2019//860881-HIDDeN and 101086085-ASYMMETRY, and the national project PID2020-113644GB-I00 as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV). DA acknowledges support from the Generalitat Valenciana grants ACIF/2020/011 and PROMETEO/2021/083. GC and AR acknowledge financial support from the Generalitat Valenciana grant CIDEGENT/2019/040. The computations were performed on the local SOM clusters, funded by the MCIU with funding from the European Union NextGenerationEU (PRTR-C17.I01) and Generalitat Valenciana, ASFAE/2022/020. We also acknowledge the computational resources provided by Finis Terrae II (CESGA), Lluis Vives (UV), Tirant III (UV). The authors also gratefully acknowledge the computer resources at Artemisa, funded by the European Union ERDF and Comunitat Valenciana, as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV).

Appendix A: Analytical lattice two-point function with open boundary conditions

Consider the generating functional for correlation functions of the topological charge density on the lattice,

$$Z[J] = \int \left(\prod_{t=1}^{N} d\phi_t\right) \exp\left[- S(\phi) + \sum_{t=1}^{T/a} J_t q_t\right], \quad \text{(A1)}$$

with the source J_t and $q_t = \frac{1}{2\pi}(\phi_{t+1} - \phi_t) \bmod 2\pi$. The explicit form of the discretized actions in eqs. (6) and (7) and topological charge in eqs. (8) and (9) depend only on the combination $\phi_{t+1} - \phi_t$. This suggests the change of integration variables $\phi_t \to q_t$ defined by

$$\phi_t = \phi_{\hat{T}-1} - \sum_{i=t}^{\hat{T}-1} q_i, \quad \text{(A2)}$$

with $t \in [0, \hat{T}-1]$ and $q_{\hat{T}-1} = 0$. With open boundary conditions every time layer decouples leaving $\hat{T}-1$ identical independent integrals. For the case of the classical perfect action in eq. (6) we get

$$Z[J] = 2\pi \left(-\sqrt{\frac{\pi}{2\pi}}\right)^{N} \prod_{i=1}^{N-1} e^{-\frac{\left(J_i + i\frac{\theta}{2\pi}\right)^2}{2\hat{I}}} \quad \text{(A3)}$$
$$\times \left[\operatorname{erf}\left(\frac{-J_i - i\frac{\theta}{2\pi} - \pi\hat{I}}{\sqrt{2\hat{I}}}\right) + \operatorname{erf}\left(\frac{J_i + i\frac{\theta}{2\pi} - \pi\hat{I}}{\sqrt{2\hat{I}}}\right)\right].$$

Differentiating the functional integral with respect to J_{t_1} and J_{t_2} and setting these to zero gives us the connected two-point correlation function of the topological charge density,

$$\langle q_{t_1} q_{t_1} \rangle_c = \frac{1}{Z[0]} \left.\frac{\delta^2 Z[J]}{\delta J_1 \delta J_2}\right|_{J=0}. \quad \text{(A4)}$$

For a finite distance $t_1 - t_2 \neq 0$ the two-point function vanishes, while for $t_1 = t_2$ we get

$$\langle q_t q_t \rangle_c = \frac{1}{4\pi^2 \hat{I}} \Bigg\{ 1 + \frac{\sqrt{2\hat{I}}}{\sqrt{\pi} C^2(\theta, \hat{I})} \quad \text{(A5)}$$
$$\times \left[\frac{e^{-\frac{A^2(\hat{I}, -\theta)}{2\hat{I}}}}{\hat{I}} \left(C(\theta, \hat{I}) A(\hat{I}, -\theta) - \hat{I}\sqrt{2/\pi} \right) \right.$$
$$\left. + \frac{e^{-\frac{A^2(\hat{I}, \theta)}{2\hat{I}}}}{\hat{I}} \left(C(\theta, \hat{I}) A(\hat{I}, \theta) + \hat{I}\sqrt{2/\pi} \right) \right] \Bigg\},$$

where $A(\hat{I}, \theta) \equiv \pi\hat{I} + \frac{i\theta}{2\pi}$ and

$$C(\theta, \hat{I}) = \operatorname{erf}\left(\frac{-A(\hat{I}, \theta)}{\sqrt{2\hat{I}}}\right) + \operatorname{erf}\left(\frac{-A(\hat{I}, -\theta)}{\sqrt{2\hat{I}}}\right). \quad \text{(A6)}$$

Eq. (A5) gives the exact result of the correlation function for any value of $\hat{I}$ and θ. Additionally, the continuum value is recovered for $\hat{I} \to \infty$. The complete lack of finite-T effects is due to the choice of boundary conditions that effectively decouple the time layers.

The global definition of the topological susceptibility can also be computed from the above generating functional,

$$\chi_t(\theta)^{\text{global}} = -\frac{1}{T} \frac{\partial^2 \log Z[0]}{\partial \theta^2} = \frac{T-1}{T} \langle q_t q_t \rangle_c, \quad \text{(A7)}$$

recovering a similar expression up to finite-T effects.

Appendix B: Spectrum and θ-dependence from the lattice

1. Analytical results

The energy spectrum of the lattice theory can be computed analytically for different choices of the action discretization. We consider a Fourier transform of the transfer matrix [32] with respect to $\psi_t = \phi_{t+a} - \phi_t$,

$$A_n(\theta) \equiv e^{-aE_n\theta} = \int_{-\pi}^{\pi} d\psi \, \langle \phi_{t+a} | \, \mathcal{T} \, | \phi_t \rangle \, e^{-in\psi}. \quad \text{(B1)}$$

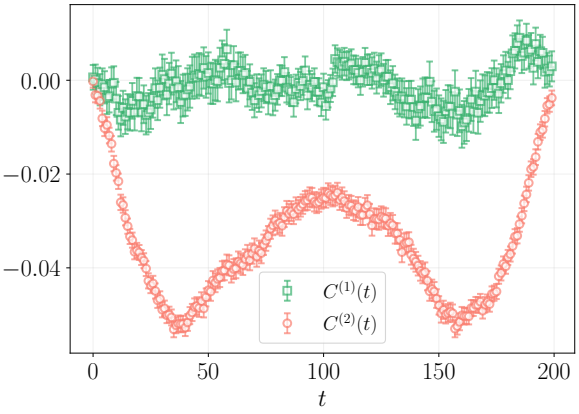


FIG. 5. Linear (green squares) and quadratic (red circles) contributions to the two-point function $C(t) = \langle \phi_t \phi_0 \rangle_c$ in the θ-expansion from a simulation at $\hat{I} = 10$ and $\hat{T} = 200$ with periodic boundary conditions and the standard action.

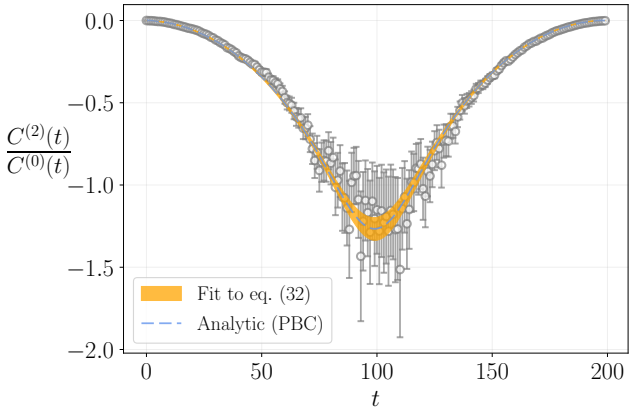


FIG. 6. Ratio $C^{(2)}(t)/C^{(0)}(t)$ from a simulation at $\hat{I} = 10$ and $\hat{T} = 200$ with periodic boundary conditions using the standard action. The resulting fit with the functional form in eq. (32) is shown (yellow band) together with the exact analytical result (blue dashed line).

The transfer matrix for the standard discretization of the action and topological charge of the action and topological charge reads

$$\langle \phi_{t+a} | \mathcal{T} | \phi_t \rangle_{\text{st}} = \sqrt{\frac{\hat{I}}{2\pi}} \exp\left\{ -\hat{I}(1-\cos\psi) + i\frac{\theta}{2\pi}\sin\psi \right\}, \tag{B2}$$

and the corresponding energies are

$$e^{-E_n} = \sqrt{\frac{\hat{I}}{2\pi}} \int_{-\pi}^{\pi} d\psi \, \exp\left\{ -\frac{\hat{I}}{a}(1-\cos\psi) \right\} \times \exp\left\{ in\psi - i\frac{\theta}{2\pi}\sin\psi \right\}. \tag{B3}$$

For the classical perfect discretization the transfer matrix is

$$\langle \phi_{t+a} | \mathcal{T} | \phi_t \rangle_{\text{cp}} = \sqrt{\frac{\hat{I}}{2\pi}} \exp\left\{ -\frac{\hat{I}}{2}((\phi_{t+1}-\phi_t) \bmod 2\pi)^2 + i\frac{\theta}{2\pi}(\phi_{t+1}-\phi_t) \bmod 2\pi \right\}, \tag{B4}$$

and the energies read

$$e^{-E_n} = \sqrt{\frac{\hat{I}}{2\pi a}} \int_{-\pi}^{\pi} d\psi \, \exp\left\{ -\frac{\hat{I}}{2}\psi^2 + i\psi\left(n - \frac{\theta}{2\pi}\right) \right\}. \tag{B5}$$

2. Lattice correlators

The correlation functions on a finite lattice are modified due to the choice of boundary conditions. For a periodic lattice, each term k in eq. (28) has an additional contribution $e^{-(T-t)\Delta E_k}$. Its behavior at large t is modified to

$$C(t) = 2Ae^{-\Delta E_1^{(1)} t}\left[1 + \left(\Delta E_1^{(0)} t\theta\right)^2\right] + 2Ae^{-\Delta E_1^{(1)}(T-t)}\left[1 + \left(\Delta E_1^{(0)}(T-t)\theta\right)^2\right], \tag{B6}$$

where we used $\Delta E_1^{(1)} = -\Delta E_{-1}^{(1)}$. To extract the linear correction to the energy we consider the ratio

$$\frac{C^{(2)}(t)}{C^{(0)}(t)} = \frac{1}{2}(\Delta E_1^{(1)})^2\left[t^2 + \frac{(T^2 - 2tT)}{1 + e^{-\Delta E_1^{(0)}(T-2t)}}\right]. \tag{B7}$$

In fig. 5 the two first contributions to the two-point correlator $\langle \phi_t \phi_0 \rangle_c$ in the θ expansion are shown for a periodic lattice for a simulation with the standard dicretization of the action using periodic boundary conditions. The linear term is compatible with zero, as expected from eq. (B6). The ratio in eq. (B7) is plotted in fig. 6. The fit is shown to match the exact value for this discretization obtained with the functional form in eq. (B7) and the energy values from eq. (B3).

Appendix C: Winding transformation

We define the winding transformation on a quantum rotor configuration as

$$\mathcal{W}^{\pm}:\ \phi_t \to \phi_t^{\mathcal{W}^{\pm}} = \phi_t \pm 2\pi t/\hat{T}, \tag{C1}$$

for $t \in [0, \hat{T}-1]$. Assuming a smooth enough configuration ϕ such that $(\phi_{t+1} - \phi_t \bmod 2\pi) \pm 2\pi/\hat{T} \in [-\pi, \pi)$ —which is the case close enough to the continuum—, one can show that the winding transformation changes its classical topological charge in one unit, i.e.

$$Q_{\rm cp}(\phi^{\mathcal{W}^{\pm}}) = Q_{\rm cp}(\phi) \pm 1. \tag{C2}$$

The change in the classical perfect action after the winding transformation is

$$\begin{aligned}\Delta S_{\rm cp}^{\mathcal{W}^{\pm}} &\equiv S_{\rm cp}(\phi^{\mathcal{W}^{\pm}}) - S_{\rm cp}(\phi) \\ &= 2\pi^2 \left(1 \pm 2Q_{\rm cp}(\phi)\right) \frac{\hat{I}}{\hat{T}},\end{aligned} \tag{C3}$$

and can be made small by increasing the time extent $\hat{T}$ of the lattice. One can use this transformation to propose new samples in a Metropolis algorithm, to be accepted with probability

$$p_{\rm acc}(\phi^{\mathcal{W}^{\pm}} \mid \phi) = \min\left\{1, \exp\left(-\Delta S_{\rm cp}^{\mathcal{W}^{\pm}}\right)\right\}. \tag{C4}$$

Note that one must use this update step in combination with other sampling algorithm to ensure ergodicity [16].

[1] T. E. Chupp, P. Fierlinger, M. J. Ramsey-Musolf, and J. T. Singh, Electric dipole moments of atoms, molecules, nuclei, and particles, Rev. Mod. Phys. **91**, 015001 (2019).

[2] Y. Aoki *et al.* (Flavour Lattice Averaging Group (FLAG)), FLAG Review 2021, Eur. Phys. J. C **82**, 869 (2022), arXiv:2111.09849 [hep-lat].

[3] R. D. Peccei and H. R. Quinn, Constraints imposed by CP conservation in the presence of pseudoparticles, Phys. Rev. D **16**, 1791 (1977).

[4] A. E. Nelson, Naturally Weak CP Violation, Phys. Lett. B **136**, 387 (1984).

[5] S. M. Barr, Solving the strong CP problem without the peccei-quinn symmetry, Phys. Rev. Lett. **53**, 329 (1984).

[6] W.-Y. Ai, J. S. Cruz, B. Garbrecht, and C. Tamarit, Absence of CP violation in the strong interactions, Physics Letters B **822**, 136616 (2021), arXiv:2001.07152 [hep-ph, physics:hep-th].

[7] W.-Y. Ai, J. S. Cruz, B. Garbrecht, and C. Tamarit, Consequences of the order of the limit of infinite spacetime volume and the sum over topological sectors for CP violation in the strong interactions, Physics Letters B **822**, 136616 (2021).

[8] A. Shindler, T. Luu, and J. de Vries, Nucleon electric dipole moment with the gradient flow: The θ-term contribution, Phys. Rev. D **92**, 094518 (2015), arXiv:1507.02343 [hep-lat].

[9] J. Dragos, T. Luu, A. Shindler, J. de Vries, and A. Yousif, Confirming the Existence of the strong CP Problem in Lattice QCD with the Gradient Flow, Phys. Rev. C **103**, 015202 (2021), arXiv:1902.03254 [hep-lat].

[10] L. Giusti and M. Lüscher, Topological susceptibility at $T > T_c$ from master-field simulations of the SU(3) gauge theory, Eur. Phys. J. C **79**, 207 (2019), arXiv:1812.02062 [hep-lat].

[11] M. Lüscher, Stochastic locality and master-field simulations of very large lattices, EPJ Web Conf. **175**, 01002 (2018), arXiv:1707.09758 [hep-lat].

[12] B. Alles, G. Boyd, M. D'Elia, A. Di Giacomo, and E. Vicari, Hybrid Monte Carlo and topological modes of full QCD, Phys. Lett. B **389**, 107 (1996), arXiv:hep-lat/9607049.

[13] L. Del Debbio, H. Panagopoulos, and E. Vicari, theta dependence of SU(N) gauge theories, JHEP **08**, 044, arXiv:hep-th/0204125.

[14] L. Del Debbio, G. M. Manca, and E. Vicari, Critical slowing down of topological modes, Phys. Lett. B **594**, 315 (2004), arXiv:hep-lat/0403001.

[15] S. Schaefer, R. Sommer, and F. Virotta (ALPHA), Critical slowing down and error analysis in lattice QCD simulations, Nucl. Phys. B **845**, 93 (2011), arXiv:1009.5228 [hep-lat].

[16] D. Albandea, P. Hernández, A. Ramos, and F. Romero-López, Topological sampling through windings, The European Physical Journal C **81**, 10.1140/epjc/s10052-021-09677-6 (2021).

[17] P. de Forcrand, Simulating QCD at finite density, PoS **LAT2009**, 010 (2009), arXiv:1005.0539 [hep-lat].

[18] C. Gattringer and K. Langfeld, Approaches to the sign problem in lattice field theory, Int. J. Mod. Phys. A **31**, 1643007 (2016), arXiv:1603.09517 [hep-lat].

[19] G. Bhanot and F. David, The Phases of the O(3) σ Model for Imaginary θ, Nucl. Phys. B **251**, 127 (1985).

[20] V. Azcoiti, G. Di Carlo, A. Galante, and V. Laliena, New proposal for numerical simulations of theta vacuum - like systems, Phys. Rev. Lett. **89**, 141601 (2002), arXiv:hep-lat/0203017.

[21] B. Alles and A. Papa, Mass gap in the 2D O(3) non-linear sigma model with a theta=pi term, Phys. Rev. D **77**, 056008 (2008), arXiv:0711.1496 [cond-mat.stat-mech].

[22] B. Alles, M. Giordano, and A. Papa, Behavior near $\theta = \pi$ of the mass gap in the two-dimensional O(3) non-linear sigma model, Phys. Rev. B **90**, 184421 (2014), arXiv:1409.1704 [hep-lat].

[23] H. Panagopoulos and E. Vicari, The 4D SU(3) gauge theory with an imaginary θ term, JHEP **11**, 119, arXiv:1109.6815 [hep-lat].

[24] M. D'Elia and F. Negro, θ dependence of the deconfinement temperature in Yang-Mills theories, Phys. Rev. Lett. **109**, 072001 (2012), arXiv:1205.0538 [hep-lat].

[25] C. Bonati, M. D'Elia, and A. Scapellato, θ dependence in $SU(3)$ Yang-Mills theory from analytic continuation, Phys. Rev. D **93**, 025028 (2016), arXiv:1512.01544 [hep-

lat].
[26] S. Aoki, R. Horsley, T. Izubuchi, Y. Nakamura, D. Pleiter, P. E. L. Rakow, G. Schierholz, and J. Zanotti, The Electric dipole moment of the nucleon from simulations at imaginary vacuum angle theta, (2008), arXiv:0808.1428 [hep-lat].
[27] C. Bonati, M. D'Elia, P. Rossi, and E. Vicari, θ dependence of 4D $SU(N)$ gauge theories in the large-N limit, Phys. Rev. D **94**, 085017 (2016), arXiv:1607.06360 [hep-lat].
[28] C. Bonanno, M. D'Elia, and L. Verzichelli, The θ-dependence of the SU(N) critical temperature at large N, (2023), arXiv:2312.12202 [hep-lat].
[29] C. Bonanno, C. Bonati, M. Papace, and D. Vadacchino, The θ-dependence of the yang-mills spectrum from analytic continuation (2024), arXiv:2402.03096 [hep-lat].
[30] G. Catumba, A. Ramos, and B. Zaldivar, Stochastic automatic differentiation for Monte Carlo processes, (2023), arXiv:2307.15406 [hep-lat].
[31] N. Fjeldso, J. Midtdal, and F. Ravndal, Random walks of a quantum particle on a circle, (1987).
[32] W. Bietenholz, R. Brower, S. Chandrasekharan, and U.-J. Wiese, Perfect Lattice Topology: The Quantum Rotor as a Test Case, Physics Letters B **407**, 283 (1997), arXiv:hep-lat/9704015.
[33] N. S. Manton, An alternative action for lattice gauge theories, Phys. Lett. B **96**, 328 (1980).
[34] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid monte carlo, Physics Letters B **195**, 216 (1987).
[35] M. D. Brida and M. Lüscher, SMD-based numerical stochastic perturbation theory, Eur. Phys. J. C **77**, 308 (2017), arXiv:1703.04396 [hep-lat].
[36] S. Coleman, *Aspects of Symmetry: Selected Erice Lectures* (Cambridge University Press, 1985).
[37] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).
[38] A. Ramos, Formalseries.jl (2023), https://igit.ific.uv.es/alramos/formalseries.jl.
[39] A. Haro, Automatic differentiation tools in computational dynamical systems, (2011).
[40] F. Di Renzo, G. Marchesini, P. Marenzoni, and E. Onofri, Lattice perturbation theory on the computer, Nucl. Phys. B Proc. Suppl. **34**, 795 (1994).
[41] M. Dalla Brida and M. Lüscher, SMD-based numerical stochastic perturbation theory, Eur. Phys. J. C **77**, 308 (2017), arXiv:1703.04396 [hep-lat].

GW Backgrounds associated with PBHs

Guillem Domènech[a,b]*
[a] Institute for Theoretical Physics, Leibniz University Hannover,
Appelstraße 2, 30167 Hannover, Germany. and
[b] Max-Planck-Institut für Gravitationsphysik,
Albert-Einstein-Institut, 30167 Hannover, Germany

PBH formation requires high-density regions in the (random) density field filling the primordial universe. While only the largest (and so rarest) overdensities collapse to form PBHs, the rest cause large anisotropic stresses, which are the source of GWs. We provide an overview of the theoretical aspects of the GW backgrounds associated with PBHs from large primordial fluctuations. We consider GW backgrounds associated with PBH formation, PBH reheating and unresolved PBH binaries. We present several graphical summaries and illustrations for the busy reader.

arXiv:2402.17388v1 [gr-qc] 27 Feb 2024

I. INTRODUCTION

Whenever there are fluctuations on top of the isotropic and homogeneous primordial universe, there is a secondary generation of GWs. Kenji Tomita was the first to write about this effect in the '70s in Ref. [1], saying that "gravitational waves are induced by deformed density perturbations". Our interpretation of Ref. [1] is that by "deformed density perturbations" Tomita meant the anisotropic stress resulting from the presence of density perturbations. This will be clear later from Fig. 1. This effect, which we will call induced GWs, was later studied in the '90s by Matarrese, Pantano and Saez [2, 3] and by Matarrese, Mollerach and Bruni [4]. Ten years later, in 2006, we find a glimpse of the potential of induced GWs by Ananda, Clarkson and Wands [5], where they considered an "excess power in a single mode" from the power spectrum measured by CMB observations (and also using a notation very similar to the current one). A very nice work by Baumann, Steinhardt, Takahashi and Ichiki [6] hinted that induced GWs could be enhanced during a matter-dominated phase (see also Refs. [7, 8]). But, it was not until 2008 that Saito and Yokoyama [9, 10] made the connection between induced GWs and PBHs. Their idea was quickly followed by Bugaev and Klimai [11–13]. For more details on the early history of induced GWs and recent developments, we refer the reader to Ref. [14] for a recent review on the topic by the author of this chapter. Other helpful reviews can be found in Refs. [15–18].

Induced GWs are a crucial observable for any PBH scenario. As it will be evident by the end of this chapter, an induced GW background must be present if there is (or was) a significant fraction of PBHs in the universe. Though the opposite is not always true, the induced GW background might suggest, or strongly exclude, PBHs as a significant fraction of DM. This chapter will aim to introduce the main formulas and predictions for the GW backgrounds associated with PBHs, focusing on a qualitative and intuitive understanding of the physics behind such GWs and the

* guillem.domenech@itp.uni-hannover.de

interesting parameter space to be probed. The busy reader will find a graphical summary of the parameter space in Fig. 2 (for GWs associated with PBH formation) and in Fig. 6 (for GWs associated with PBH reheating). Their corresponding physical picture of the generation of induced GWs is respectively illustrated in Figs. 1 and 5.

The chapter is organized as follows. We first explore in § II general sources of GWs in cosmology. We then revisit in § III the definition of the energy density of GWs in cosmology, emphasizing potential ambiguities for induced GWs. In § IV, we present the general formulation of induced GWs and focus on the most studied. In § V, we review the GWs from the PBH-dominated early universe. The GW background from PBH binaries is explored in § VI. The chapter ends with final remarks in § VII.

II. BIRD'S EYE VIEW ON SECONDARY GWS

We start with a general discussion on GWs in cosmology, with the aim of understanding secondary GWs and how they differ from other sources. We use the notation "secondary GWs" to refer to any GW sourced at second order in cosmological perturbation theory using the linear solutions to cosmic fluctuations. Our basic framework, which we introduce in detail below, is a perfect fluid in a perturbed FLRW universe (for the basics of cosmology, we recommend Mukhanov's [19] and Baumann's [20] books). A perfect fluid could be an adiabatic perfect fluid with a given equation of state $w = P/\rho$, where P and ρ, respectively, are the pressure and energy density of the fluid. But, it could also be a scalar field, say φ. These two cases are described with the energy-momentum tensor given by

$$T_{\mu\nu} = (\rho + P)\, u_\mu u_\nu + P g_{\mu\nu}\,, \tag{2.1}$$

where $g_{\mu\nu}$ is our spacetime metric and u_μ the normalized 4-velocity of the fluid.[1] The adiabatic perfect fluid describes, e.g., a gas of ultra-relativistic particles when $w = 1/3$. This is the situation we encounter in standard Big Bang cosmology. A gas of non-relativistic particles has $w \ll 1$, which is the case of standard cold DM. On the other hand, the scalar field can be used to describe inflation, reheating scenarios, bubble nucleation and topological defects.

To get to the equations for GWs, we begin with Einstein Equations, namely $G_{\mu\nu} = T_{\mu\nu}$ with $G_{\mu\nu}$ the Einstein tensor, and perturb both the metric and the matter fields. For simplicity, we work in conformal time and in the shear-free slicing (also known as the Newton gauge) in which the line element reads

$$ds^2 = a^2 \left(-e^{2\Phi} d\tau^2 + e^{2\Psi} \left(e^h\right)_{ij} dx^i dx^j \right)\,, \tag{2.3}$$

where a is the scale factor, τ is conformal time, Φ and Ψ are the gravitational potentials, h_{ij} are the transverse-traceless degrees of freedom, and we neglected the transverse degrees of freedom, i.e.,

[1] The canonical scalar field case is recovered by identifying

$$u_\mu = \frac{\partial_\mu \varphi}{\sqrt{-\partial_\alpha \varphi \partial^\alpha \varphi}} \quad , \quad \rho = -\frac{1}{2}\partial_\alpha \varphi \partial^\alpha \varphi + V(\varphi)\,, \quad P = -\frac{1}{2}\partial_\alpha \varphi \partial^\alpha \varphi - V(\varphi)\,, \tag{2.2}$$

where $V(\varphi)$ is the potential of the scalar field.

vector fluctuations. Note that we took the exponential ansatz for later convenience when discussing subtleties of second-order perturbation theory. For the matter fields, we have that

$$u_0 = -a(1+\Psi) \quad , \quad u_i = a v_i \quad , \quad \rho = \rho(t) + \delta\rho \quad , \quad P = P(t) + \delta P \,, \tag{2.4}$$

where now $\rho(t)$ and $P(t)$ are the background components, i.e. homogeneous in space, and $\delta\rho$ and δP are the perturbations. Similarly, we take $\varphi \to \varphi(t) + \delta\varphi$ for the scalar field case.

With the perturbative expansion given by Eqs. (2.3) and (2.4), we may write the GW equation as [14, 21]

$$h_{ij}'' + 2\mathcal{H} h_{ij}' - \Delta h_{ij} = \left[4\partial_i\Phi\partial_j\Phi + 2a^2(\rho+P)v_i v_j\right]^{\rm TT} \,, \tag{2.5}$$

where $' = d/d\tau$, $\mathcal{H} = a'/a$, $\Delta = \delta^{ij}\partial_i\partial_j$ and we expanded Einstein equations up to second order in perturbation theory. Also, "TT" refers to the transverse-traceless component of a given tensor (see, e.g., Ref. [14] for the explicit form of the projector). There is no linear source to the GWs in Eq. (2.5) as a perfect fluid has no linear anisotropies. To derive the right-hand side of (2.5), we used that at linear order $\Psi = -\Phi$ as well as the properties of the transverse-traceless projector to write terms like $\Phi\partial_i\partial_j\Phi$ as $-\partial_i\Phi\partial_j\Phi$. It is interesting to note that in Eq. (2.5), the term with Φ comes from $G_{ij}^{(2)}$ and the term with v_i from $T_{ij}^{(2)}$, where the superscript (2) means the expansion at second order. Thus, the term with Φ can be understood as the "backreaction" of the metric, while the term with v_i is the standard matter source to GWs. With the above equation, we may qualitatively discuss cosmic sources of GWs.

Let us start with the case in which GWs are generated on scales much smaller than the Hubble radius (also called Hubble horizon) by a "spectator" field, i.e. the field responsible for the source of GWs is a subdominant component of the universe's energy density. This includes, e.g., nucleation of bubbles due to a first-order phase transition or the formation and evolution of a network of solitons, such as cosmic strings, domain walls, oscillons, etc. In this regime, also called the weak gravity limit, one may safely neglect the contribution from Φ in Eq. (2.5). This is because, on subhorizon scales, the term containing Φ in Eq. (2.5) is suppressed with respect to the v_i contribution by the ratio of energy densities between the spectator and the dominant field. Thus, GWs are sourced by the anisotropic stress resulting from the inhomogeneous velocity flows or scalar field profiles. For instance, in the case of bubble nucleation, one of the primary sources of GWs are the acoustic density waves following bubble collisions (though the collisions themselves and later turbulences also produce GWs; see Refs. [22, 23] for recent reviews). For cosmic strings, it may be better to think of Eq. (2.5) in terms of the quadrupole formula (as for a local source, the volume integral of T_{ij} is related to the second time derivative of the quadrupole). Then, a global cosmic string may be thought of as a "vibrating cylinder" with a length proportional to the Hubble radius. Its time-dependent quadrupole is then the source of GWs. There are, of course, other processes that lead to GW production, such as cosmic string loops and kinks. As the present discussion is very naive, and the above-mentioned process entirely non-linear, we refer the interested reader again to Refs. [22, 23] for recent reviews with more detailed discussions and references therein.

A somewhat similar situation applies to secondary GWs associated with PBHs. But, in contrast with previous cases, secondary GWs are generated by the dominant fluid in the universe. And, the

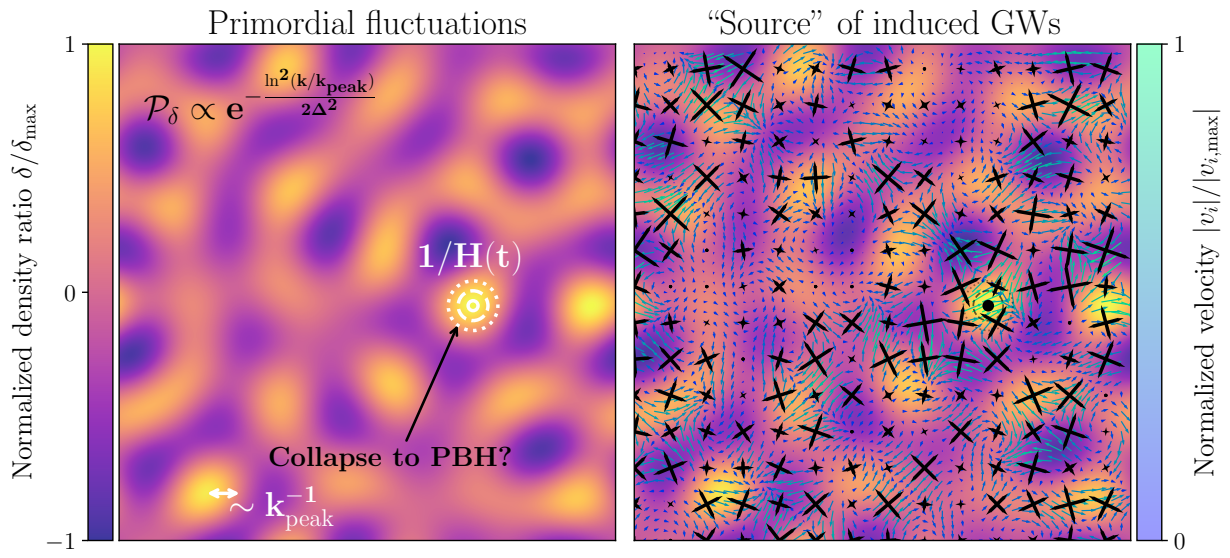


FIG. 1. Real space density field in 2D from Gaussian fluctuations with log-normal primordial spectrum with $\Delta = 0.01$ and $k_{\rm peak} = 0.005 L_{\rm box}^{-1}$ (i.e. the typical size of fluctuations is 20 times smaller than the box), all in arbitrary units. On the left panel, we show the density ratio and illustrate that only the rarest and largest fluctuations collapse to form PBHs. On the right panel, we show the velocity gradients and the eigenvectors times the eigenvalues of the traceless component of the energy-momentum tensor. This exemplifies that induced GWs are generated by the mean fluctuations where gradients are largest.

dominant fluid in the universe determines the *curvature fluctuations*. Thus, we cannot generally neglect the term $\partial_i\Phi\partial_j\Phi$ in Eq. (2.5). This is why one says that the GWs are induced by curvature fluctuations and are called induced GWs for short. We illustrate the physical picture of the generation of induced GWs in Fig. 1. There, we considered a universe filled with random Gaussian superhorizon curvature fluctuations with a sharp log-normal power spectrum, peaked at $k_{\rm peak}$ and with a large enough amplitude. The left plot of Fig. 1 shows the 2D real space distribution of the density field, say δ, or the curvature fluctuation. We see that only a few lumps (in this case, roughly one) have a large enough amplitude to eventually collapse and form a PBH once the Hubble horizon is approximately the size of the overdensity. The right plot of Fig. 1 illustrates with blue arrows the velocity flow (as being proportional to $\partial_i\delta$ or $\partial_i\Phi$) and with black crosses the eigenvectors of the traceless part of T_{ij} (proportional to v_iv_j). Larger velocities have larger arrows and a lighter blue colour. The size of the crosses is proportional to the eigenvalues of the traceless part of T_{ij}.[2] See how larger "sources" to the induced GWs are associated with larger velocity flows. We also see that while only one overdensity collapses to form a PBH, the rest source the induced GWs.

To put some units, we show in Fig. 2 the frequency of the induced GWs $f_{\rm iGWs}$ (see Eq. (4.14)), the typical mass of the PBHs $M_{\rm PBH,f}$ and the frequency related to the Innermost Stable Circular

[2] We only consider the traceless part of T_{ij} as in 2D there are no transverse-traceless degrees of freedom. Nevertheless, we expect that a similar situation applies to a more realistic 3D picture.

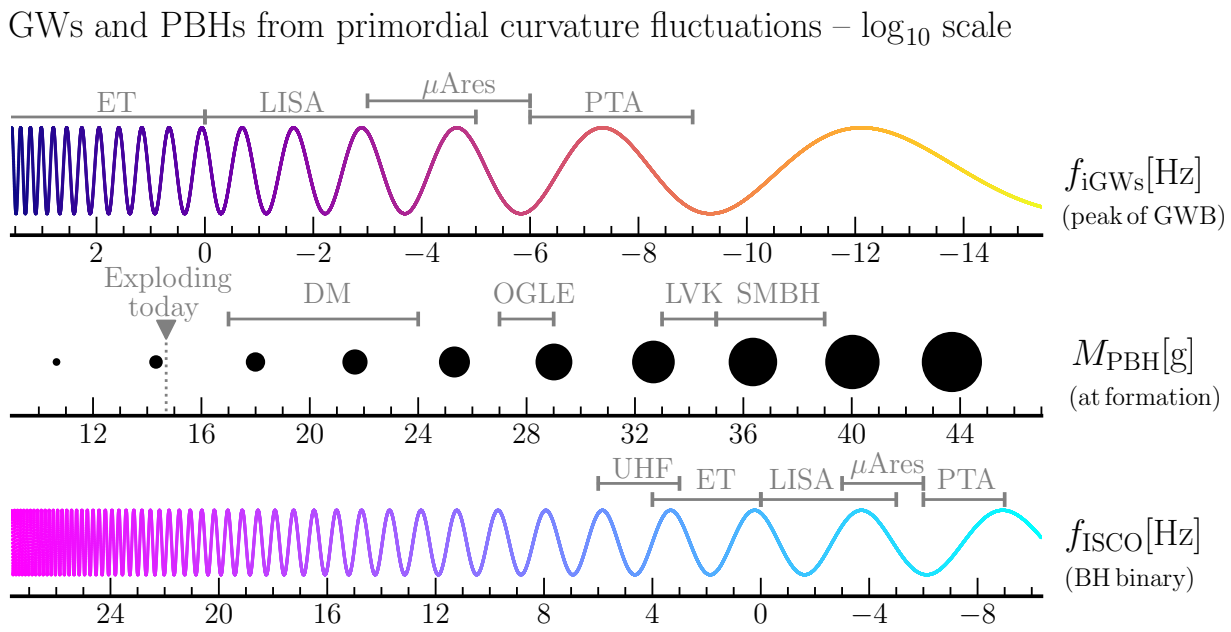


FIG. 2. Graphical summary of the relevant parameter space regarding GW frequency and PBH mass for the induced GWs (associated with PBH formation). In the top we show the frequency of induced GWs $\log_{10} f_{\rm iGWs}$ (4.14), in the middle we show PBH mass $\log_{10} M_{\rm PBH}$ and in the bottom we show the frequency of PBH mergers $\log_{10} f_{\rm ISCO}$ (6.7). They are all related to $M_{\rm PBH}$ by Eq. (4.13) through $k_{\rm peak} = a_{\rm f} H_{\rm f}$. We show the frequency range of current and future GW detectors such as Ultra-High-Frequency (UHF) GW detectors, ET, LISA, μ-Ares and PTAs. We also draw the interesting PBH mass windows: PBH as all DM [25, 32], planet mass PBH to explain OGLE candidate event [26, 27], the merger of solar mass PBHs at LVK [28–30] and PBH as the seeds of SMBH [31]. Note that the PTA window for induced GW overlaps with OGLE and LVK.

Orbit (ISCO), $f_{\rm ISCO}$, of the merger of an equal PBH binary (in the binary frame, see Eq. (6.7)). The redshift factor is the main difference between $f_{\rm iGW}$ and $f_{\rm ISCO}$. While induced GWs are generated at the same time PBHs form, the detected PBH binaries merge in the nearby universe. See how with GW detectors such as ET, LISA and PTAs, we can indirectly probe the existence of PBH with masses ranging from $\mathcal{O}(10^9\,{\rm g})$ to $\mathcal{O}(M_\odot)$ with induced GWs. In fact, there may already be hints of an induced GW background at nHz frequency from PTAs [24]. Interestingly, the same GW detectors can probe the mass range from $\mathcal{O}(M_\odot)$ to $\mathcal{O}(10^9 M_\odot)$ from the merger of PBHs binaries. In the same figure we illustrate the most interesting PBH mass windows to explain the DM [25], the planet mass candidates events in OGLE [26, 27], the BH mergers seen by the LVK collaboration [28–30] and the seeds of supermassive black holes (SMBHs) [31].

III. EFFECTIVE ENERGY DENSITY OF GWS IN COSMOLOGY

Before we dig into the details of the calculations of induced GWs, we briefly revisit the definition of the energy density of GWs in cosmology, as we will recursively use it to derive predictions. We will discuss important subtleties of the definition of energy density of the induced GWs at the end of this section. The derivation of the energy-momentum tensor can be found, e.g., in Maggiore's book [33]. A detailed discussion on the subtleties is given in Ref. [34], and references therein.

A. The pseudo Energy Momentum Tensor of GWs

To estimate the energy density carried by the GWs, we need to find the "backreaction" of *linear* tensor fluctuations onto the background metric. To do that, we expand Einstein equations to second order (as we did in Eq. (2.5)), and compute some kind of average. This average could be a volume and/or a time average, inside which several GW wavelengths or oscillations fit. In the case of stochastic fields, the volume/time average is equivalent to the ensemble average (by ergodicity). Expanding and taking the average, we formally find that

$$G^{(0)}_{\mu\nu}[a] \approx -\langle G^{(2)}_{\mu\nu}[a(\eta+h)]\rangle + T^{(0)}_{\mu\nu} \approx t^{\rm GW}_{\mu\nu} + T^{(0)}_{\mu\nu}\,, \tag{3.1}$$

where η refers to the Minkowski metric $\eta_{\mu\nu}$, h to h_{ij} of Eq. (2.3) and the brackets denote the chosen average. We moved the second order expansion $G^{(2)}_{\mu\nu}$ to the right-hand side so that we can interpret it as the energy-momentum tensor, which we call $t^{\rm GW}_{\mu\nu}$ in the last step. The contribution from linear fluctuations vanishes because of the linear equations of motion. After some algebra, one obtains that

$$t^{\rm GW}_{\mu\nu} = \frac{1}{4}\left\langle \partial_\mu h^{ij}\partial_\nu h_{ij} - \frac{1}{2} g_{\mu\nu}\partial_\sigma h^{ij}\partial^\sigma h_{ij}\right\rangle . \tag{3.2}$$

A fundamental assumption in deriving Eq. (3.2) is that we only consider subhorizon tensor fluctuations h_{ij}. This gives us an unambiguous notion of a slowly evolving, large-scale background and high-frequency, small-scale, GWs. It is not clear how to apply a similar procedure when both scales are similar, e.g., for superhorizon tensor modes. Also, note that $t^{\rm GW}_{\mu\nu}$ is called a pseudotensor as it only transforms as a tensor at the linear level in perturbation theory. It is a priory not well-defined at higher orders.

The energy density of the GWs is given by the 00 component of Eq. (3.2), namely $a^2\rho_{\rm GW} = t^{\rm GW}_{00}$ in conformal time. Since h_{ij} is a random field, we can compute the energy density in terms of its power spectrum. Working with Fourier modes,[3] the GW energy density fraction per $\ln k$, also referred to as the spectral density, is given by

$$\Omega_{\rm GW}(k,\tau) = \frac{1}{\rho_{\rm tot}}\frac{d\rho_{\rm GW}}{d\ln k} = \frac{k^2}{12\mathcal{H}^2}\sum_\lambda \overline{\mathcal{P}_{h,\lambda}(k,\tau)}\,, \tag{3.4}$$

[3] Our convention is that

$$h_{ij} = \frac{1}{(2\pi)^3}\sum_\lambda \int d^3k\, h_{\mathbf{k},\lambda}(\tau)\epsilon^\lambda_{ij}(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{3.3}$$

where polarization tensors are normalized, i.e. $(\epsilon^\lambda_{ij}(\mathbf{k}))^*\epsilon^{ij,\lambda'}(\mathbf{k}) = \delta^{\lambda\lambda'}$.

where $\mathcal{P}_{h,\lambda}(k,\tau)$ is the dimensionless strain power spectrum, defined by

$$\langle h_{\mathbf{k},\lambda}(\tau) h_{\mathbf{k}',\lambda'}(\tau)\rangle = (2\pi)^3\delta(\mathbf{k}+\mathbf{k}')\delta_{\lambda\lambda'}\tfrac{2\pi^2}{k^3}\mathcal{P}_{h,\lambda}(k,\tau)\,, \tag{3.5}$$

and we took an additional oscillation (time) average denoted with an overline, as the same average is done when looking for GW backgrounds. It should be noted that in deriving Eq. (3.4) we used that for subhorizon GWs $k \gg \mathcal{H}$ and that $h'_{\mathbf{k},\lambda} \sim k h_{\mathbf{k},\lambda}$. Eq. (3.4) is the quantity we will use to derive and compare our predictions with observations. The energy density of GWs is related to the strain noise variance.

B. Subtleties of induced GWs

Before ending this section, let us point out the fundamental ambiguity in the definition (3.2) of the energy density of GWs. We can convince ourselves that the definition is ambiguous by checking whether $\rho_{\rm GW}$ is invariant under a shift of the conformal time coordinate, e.g. $\bar\tau = \tau + T(\tau, x^i)$, where T is an infinitesimal gauge parameter. With some effort, one finds that the tensor perturbations defined as in Eq. (2.3) transform according to [21]

$$\bar h_{ij} = h_{ij} + \widehat{TT}^{kl}_{ij}\left[T\, h'_{kl} + \partial_k T \partial_l T\right] + \dots\,, \tag{3.6}$$

where "..." means higher order combinations in T and h_{ij} and we neglected vector perturbations. It turns out that what we called tensor perturbations $\bar h_{ij}$, in the barred time coordinate, are a mixture of tensor and scalar perturbations in the unbarred time coordinate. In other words, the definition of tensor perturbations beyond the linear order depends on the time slicing. The energy density of GWs as defined by Eq. (3.2) is certainly not invariant at second order under the time shift (3.6).

It is not so surprising that Eq. (3.2) becomes problematic at the second order. First, we considered only linear perturbations, and second, there is no local notion of energy in gravity, as we can always go to a local inertial (Minkowski) frame. Although this is a general problem, it is not so problematic for GW sources in the weak gravity regime, such as cosmic strings and phase transitions (where secondary GWs from Φ are already neglected), unless one deliberately chooses a strange time coordinate. But, it may be problematic for induced GWs since they are a second-order effect from Φ. For instance, considering that the unbarred time corresponds to the time in the Newton gauge, we have that $T = T(\Phi,\tau)$, with the exact functional form depending on the chosen barred time slicing. Since $h^{\rm iGW}_{ij} \propto \partial_i\Phi\partial_j\Phi$ all the terms in the right-hand side of Eq. (3.6) can be in principle comparable. This issue was first pointed out by Ref. [35] for a matter-dominated universe, in Ref. [36] in radiation domination and in general in Ref. [37]. See also Ref. [38]. It was then argued in Refs. [39, 40] that in radiation domination, the synchronous (which should describe the frame of a GW detector) and Newton gauges give the same results at late times. Then, in Ref. [34], it was shown that such approximate gauge invariance applies to other gauges and for any cosmological background (except matter domination; see below).

The crucial point to realize, which provides a partial resolution, is that we measure GWs far from the source [34]. The same is required, e.g., to define the GWs from binaries. In cosmology, "far"

from the source means that there is effectively no more production of GWs at some point in time. This is precisely the situation for induced GWs. The curvature perturbation Φ quickly decays once it enters the Hubble horizon, rendering the energy density of induced GWs practically independent of the time slicing.[4] For instance, the calculations in the Newton gauge, constant Hubble gauge, flat gauge and synchronous gauge (after proper gauge fixing) all yield the same results. The exceptions to this statement are (i) comoving slices and (ii) a dust-dominated universe. In exception (i), one can argue that such time slices are suitable for superhorizon but not for subhorizon physics. For example, on small scales, the velocity flow can become quite non-trivial and choosing a time slice that is comoving with the fluid everywhere is undoubtedly a convoluted choice. Thus, it is not surprising that the comoving slicing might contain gauge artefacts that look like GWs. In exception (ii), the curvature perturbation is constant on all scales, so there is a constant source to the induced GWs. In this case, one must follow the GW production until the end of the dust-dominated era to have an unambiguous definition of $\Omega_{\rm GW}$. The resolution to the ambiguity lies in a systematic study of the strain measured by a GW detector at second order in cosmological perturbation theory. More details and references on the gauge issue can be found in Ref. [14].

IV. GWS FROM PRIMORDIAL CURVATURE FLUCTUATIONS

After the general discussions in Secs. II and III, we focus on the mathematical formulation of induced GWs. We first take a general approach but quickly turn to the most studied case in Sec. IV B. Extensions of the formalism will be discussed in Sec. IV D. Details of the main calculations can be found in Refs. [14, 41, 42].

A. General formulation

Consider that, after inflation, there are some initial curvature fluctuations Φ in the universe. Since induced GWs are sourced by the dominant fluid in the universe, the scalar component of velocity fluctuations in Eqs. (2.5) is also related to the curvature perturbation via Einstein equations (the $0i$ component to be precise). For this reason, we can take a general approach to the source term in Eqs. (2.5) by saying that, in Fourier modes,[5] the source term can be split into a primordial value of the curvature perturbation $\Phi_{\mathbf{k}}^{\rm prim}$ times a time-dependent transfer function. A general equation for the induced tensor modes then reads

$$h''_{\mathbf{k},\lambda} + 2\mathcal{H}h'_{\mathbf{k},\lambda} + k^2 h_{\mathbf{k},\lambda} = 4\int \frac{d^3q}{(2\pi)^3} e^{ij}_{\lambda}(k) q_i q_j f(\tau, q, |\mathbf{k}-\mathbf{q}|)\Phi_{\mathbf{q}}^{\rm prim}\Phi_{|\mathbf{k}-\mathbf{q}|}^{\rm prim}\,, \tag{4.1}$$

where $f(\tau, q, |\mathbf{k}-\mathbf{q}|)$ is the transfer function of the source. We will discuss particular cases later. We can also write down a formal integral solution to Eq. (4.1) by means of the Green's function

[4] By almost independent, we mean that the prediction for the energy density of induced GWs is independent of the time slicing up to corrections suppressed by $(k/\mathcal{H})^2$ (a ratio which evaluated today is tiny) for a restricted class of time slicings.

[5] The Fourier expansion for Φ follows the same convention as Eq. (3.3) for h_{ij}.

method, which yields

$$h_{\mathbf{k},\lambda}(\tau) = \frac{4}{k^2}\int \frac{d^3q}{(2\pi)^3} e^{ij}_{\lambda}(\mathbf{k}) q_i q_j I(\tau, q, |\mathbf{k}-\mathbf{q}|) \Phi^{\rm prim}_{\mathbf{q}} \Phi^{\rm prim}_{|\mathbf{k}-\mathbf{q}|}\,, \tag{4.2}$$

where we defined for compactness

$$I(\tau, q, |\mathbf{k}-\mathbf{q}|) = k^2 \int_0^{\tau} d\tau'\, G_h(\tau,\tau') f(\tau', q, |\mathbf{k}-\mathbf{q}|)\,. \tag{4.3}$$

$I(\tau, q, |\mathbf{k}-\mathbf{q}|)$ is often called the kernel of induced GWs, and G_h in Eq. (4.3) is the Green's function for tensor modes which in radiation domination reads $kG_h(\tau,\tau') = \frac{a(\tau)}{a(\tau')}\sin(k(\tau-\tau'))$. The factors k^2 in Eqs. (4.2) and (4.3) are introduced for later convenience, as in this way, I is dimensionless.

With our solution, Eq. (4.2), we can compute the oscillation average of the dimensionless power spectrum that enters in the spectral density of GWs, Eq. (3.4). After some calculations, we obtain that

$$\begin{aligned}\overline{\mathcal{P}_{h,\lambda}(k,\tau)} = &\frac{4k}{\pi}\int \frac{d^3q}{(2\pi)^3}\int \frac{d^3q'}{(2\pi)^3} \overline{I(\tau, q, |\mathbf{k}-\mathbf{q}|) I(\tau, q', |\mathbf{k}-\mathbf{q}'|)} \\ &\times e^{ij}_{\lambda}(\mathbf{k}) q_i q_j e^{ij}_{\lambda}(-\mathbf{k}) q'_i q'_j \langle \Phi^{\rm prim}_{q} \Phi^{\rm prim}_{|\mathbf{k}-\mathbf{q}|} \Phi^{\rm prim}_{q'} \Phi^{\rm prim}_{|\mathbf{k}-\mathbf{q}'|}\rangle'\,,\end{aligned} \tag{4.4}$$

where we assumed that $\langle h_\lambda(k,\tau) h_{\lambda'}(k',\tau)\rangle$ follows the form of Eq. (3.5) so that we may subtract the factor $(2\pi)^3\delta_{\lambda\lambda'}\delta(\mathbf{k}+\mathbf{k}')$. We denote the remaining contribution to the 4-point function of Φ with a prime. Note that the oscillation average in Eq. (4.4) must be done once the tensor modes are deep inside the Hubble horizon, i.e. for $k\tau \gg 1$. We restrict now to the case of Gaussian primordial fluctuations (which is often a good approximation), where the 4-point function is determined solely by the 2-point function. In this case, we find that

$$\begin{aligned}\overline{\mathcal{P}_h(k,\tau)} &= \sum_{\lambda} \overline{\mathcal{P}_{h,\lambda}(k,\tau)} \\ &= 8\int_0^{\infty} dv \int_{|1-v|}^{1+v} du \left(\frac{4v^2-(1-u^2+v^2)^2}{4uv}\right)^2 \overline{I^2(\tau,k,v,u)} \mathcal{P}_{\Phi}(ku)\mathcal{P}_{\Phi}(kv)\,,\end{aligned} \tag{4.5}$$

where we introduced $v = q/k$ and $u = |\mathbf{k}-\mathbf{q}|/k$ for compactness.

It is worth noting that, in general, induced GWs are sourced by the typical curvature fluctuations, that is, $\langle\Phi^2\rangle^{1/2}$. In other words, the naive expectation is that $\langle\Phi^4\rangle \sim \mathcal{O}(\langle\Phi^2\rangle^2)$, unless the probability distribution function (PDF) of Φ is highly skewed (i.e. something very far from the Gaussian quantum fluctuations expected from inflation). In contrast, PBHs form from rare curvature fluctuations, which are determined by the tail of the PDF. Thus, while PBHs are exponentially sensitive to the root mean square of curvature fluctuations, induced GWs are proportional to the fourth power. From this, we conclude that the detection of induced GWs does not generally imply a significant fraction of PBHs in the universe. However, a substantial fraction of PBHs must have a detectable induced GW signal.

B. Semi-analytical formula in radiation domination

Let us now revisit the most considered case, primordial adiabatic (curvature) fluctuations in a radiation-dominated universe. The first semi-analytical formula was derived in Refs. [41, 43] and was generalized in Refs. [42, 44]. We will discuss more possibilities in Sec. IV D. In the case that one fluid dominates the universe, the $0i$ component of Einstein's equations yields

$$\Phi' + \mathcal{H}\Phi = -\frac{1}{2}a^2(\rho + P)v\,, \tag{4.6}$$

where v is defined through $v_i = \partial_i v$. Additionally, the curvature fluctuations evolve according to

$$\Phi(q\tau) = \Phi_{\mathbf{q}}^{\rm prim}\,\frac{3j_1(c_s q\tau)}{c_s q\tau}\,, \tag{4.7}$$

where j_1 is the spherical Bessel function of order 1.

Inserting the relation (4.6) between v and Φ in Eq. (2.5), and using the linear solution Eq. (4.7), we find the explicit form of $f(\tau, q, |\mathbf{k}-\mathbf{q}|)$ in Eq. (4.1). The transfer function $f(\tau, q, |\mathbf{k}-\mathbf{q}|)$ is a product of sines and cosines, and the integral $I(\tau, q, |\mathbf{k}-\mathbf{q}|)$ (4.3) can be carried out analytically. Then, combining Eq. (3.4) and Eq. (4.5), we find a compact form for the GWs induced during standard radiation domination (that is $c_s^2 = w = 1/3$), namely

$$\Omega_{\rm GW,*} = \int_0^\infty dv \int_{|1-v|}^{1+v} du\, \mathcal{T}_{RD}(u,v)\mathcal{P}_{\mathcal{R}}(ku)\mathcal{P}_{\mathcal{R}}(kv)\,, \tag{4.8}$$

where we used Eq. (4.11) and defined

$$\begin{aligned}\mathcal{T}_{RD}(u,v) = &\left(\frac{4v^2-(1-u^2+v^2)^2}{4u^2v^2}\right)^2 \\ &\times y^2\left\{\frac{\pi^2}{4}y^2\Theta[u+v-c_s^{-1}] + \left(1-\frac{1}{2}y\ln\left|\frac{1+y}{1-y}\right|\right)^2\right\}\,,\end{aligned} \tag{4.9}$$

and

$$y = \frac{u^2+v^2-c_s^{-2}}{2uv}\,. \tag{4.10}$$

The meaning of the subscript "*" will be explained in the next paragraph. Note that to obtain Eq. (4.9), we took the limit $k\tau \to \infty$ of the integral (4.3). This is a very good approximation since we are interested in the GWs generated in the very early universe. So, any correction scales at least as $\mathcal{O}(1/(k\tau_{\rm eq}))$ where $\tau_{\rm eq}$ is the time of matter-radiation equality. Recall that the comoving scale that entered the Hubble horizon at that time is $k_{\rm eq} \sim 0.01\,{\rm Mpc}^{-1}$, and we are interested in scales below a parsec. For the general form of the kernel for finite $k\tau$ see Ref. [41]. We also used that

$$\Phi = \frac{3(1+w)}{5+3w}\mathcal{R} = \frac{2}{3}\mathcal{R}\,, \tag{4.11}$$

since predictions from inflation are given in terms of the curvature perturbation on comoving slices $\mathcal{R}$. We note that the variable y, defined in Eq. (4.10), is far from arbitrary. It is actually related to the area of the triangle composed by 1, $c_s v$ and $c_s u$ (or k, $c_s q$ and $c_s|\mathbf{k}-\mathbf{q}|$ in the original variables) and it naturally appears when integrating the product of three Bessel functions. In the end, it is a very useful variable for the extension to general constant w and c_s. In the integration range of Eq. (4.5), we have that y is bounded by $-\infty < y < 1$.

Strictly speaking, instead of taking the limit $k\tau \to \infty$, one should integrate the kernel (4.3) until today, that is at $\tau = \tau_0$. However, for the GW frequencies of interest, it is unnecessary to do the integrals across all the radiation, matter and dark energy domination phases. We may use the fact that GWs, once inside enough the Hubble horizon, behave as radiation. From there on, we can track the energy density ratio of GWs by comparing it to the one of standard radiation. Thus, if we call τ_* the time at which a given GW frequency starts behaving as radiation, we can write [14, 45]

$$.\Omega_{\rm GW,0}h^2 = 1.62\times 10^{-5}\left(\frac{\Omega_{\rm rad,0}h^2}{4.18\times 10^{-5}}\right)\left(\frac{g_\rho(T_*)}{106.75}\right)\left(\frac{g_s(T_*)}{106.75}\right)^{-4/3}\Omega_{\rm GW,*}\,, \tag{4.12}$$

where the subscript "$*$" means evaluation at $\tau = \tau_*$, $\Omega_{\rm rad,0}$ is energy density ratio of radiation today [46] and $h = H_0/(100\,{\rm km/s/Mpc})$. $g_\rho(T_*)$ and $g_s(T_*)$ are the effective degrees of freedom in the energy density and entropy of the standard radiation fluid. Thus, our goal is to estimate $\Omega_{\rm GW,*}$ and use Eq. (4.12) to find the amplitude of the induced GW background today. In passing, using that PBHs form with a mass given by

$$M_{\rm PBH,f} = 4\pi\gamma M_{\rm pl}^2/H_{\rm f}\,, \tag{4.13}$$

where $\gamma \sim 0.2$ is an effective parameter (see, e.g., Ref. [47]), and that $k_{\rm peak} = a_{\rm f}H_{\rm f}$ where "f" refers to PBH formation, we can write $f_{\rm iGWs}$ of Fig. (2) in terms of $M_{\rm PBH,f}$ as

$$f_{\rm iGWs} \approx 1\,{\rm Hz}\left(\frac{M_{\rm PBH,f}}{10^{16}{\rm g}}\right)^{-1/2}\left(\frac{\gamma}{0.2}\right)^{1/2}\left(\frac{g_\rho(T_*)}{106.75}\right)^{1/4}\left(\frac{g_s(T_*)}{106.75}\right)^{-1/3}. \tag{4.14}$$

C. Typical features and general properties

Let us derive some useful and general properties of the induced GW spectrum from Eq. (4.8). This will provide us with intuition about the expected shape of the induced GW spectrum from a given primordial spectrum of curvature fluctuations, assuming that the primordial spectrum is peaked. We show a few examples of typical induced GW spectra in Fig. 7.

The first thing one may notice from the averaged square kernel Eq. (4.9) is that it contains a logarithm that diverges at $y = \pm 1$. The value $y = -1$ occurs when $u + v = c_s^{-1}$. The origin of such divergence is a resonant production of induced tensor modes. In fact, Eq. (2.5) is nothing but a damped harmonic oscillator with a periodic force proportional to Φ^2. So there is a resonance for $k = c_s(q + |\mathbf{k}-\mathbf{q}|)$, where k is the tensor and q the scalar wavenumbers.[6] Thus, for a peaked

[6] This intuition is only valid if the source in Fourier space (4.1) has a well-defined frequency. This only occurs when fluctuations of Φ have a typical scale, or in other words, the power spectrum of Φ is sufficiently peaked. From the point of view of the integral (4.5), the exact resonance is a zero-measure point. So if the primordial spectrum is broad, the resonance is smeared out.

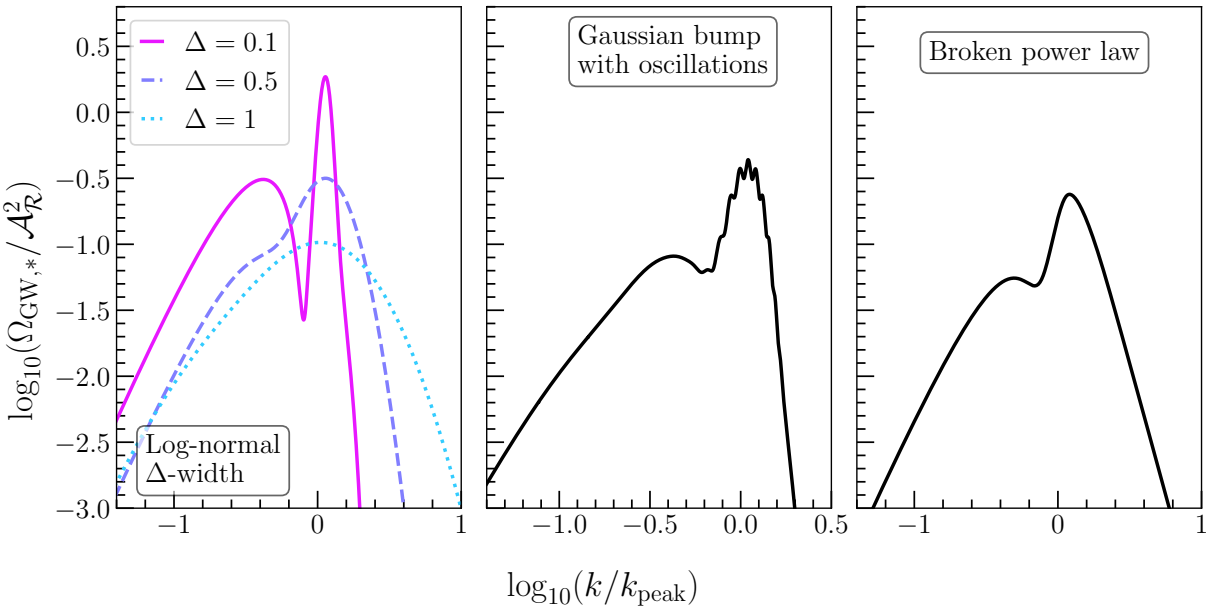


FIG. 3. Induced GW spectra in radiation domination computed numerically using SIGWfast [48]. On the left panel, we show the induced GWs from a log-normal peak for various values of the logarithmic width Δ. See how as the width increases the resonant peak is smoothed out. On the middle panel, we show the induced GW from a Gaussian bump with $\mathcal{O}(1)$ oscillations in the primordial spectrum from Ref. [49]. Note how the oscillations in the induced GW spectrum are only $\mathcal{O}(10\%)$ modulations at the peak. On the right panel, we show the induced GW spectrum from a broken power-law primordial spectrum with k^4 in the IR and k^{-5} in the UV. See how for all relatively sharp peaks, the induced GW spectrum has a similar shape. The IR tail of the induced GW spectrum in all cases goes as $k^3 \ln^2 k$ (except for the very sharp Δ case, in magenta, which initially is $k^2 \ln^2 k$).

enough spectra, the resonance lies at

$$k_{\rm res} = 2c_s k_{\rm peak}. \tag{4.15}$$

The resonant peak is clearly visible in Fig. 3, where we show the induced GW spectra of a log-normal primordial spectrum with logarithmic with Δ, a Gaussian bump with oscillations and a broken power-law spectrum, respectively from left to right. On the left plot, we also plot the induced GW spectrum for various values of the width. See how as one increases Δ the resonant peak disappears. For analytical approximations in the case of the log-normal primordial spectrum see Ref. [50]. For the Gaussian bump with oscillations see Refs. [49, 51–53]. And, for the broken power-law see Refs. [54–56].

The value $y = 1$ corresponds to the limit $u \sim v \gg 1$, which is related to long wavelength tensor modes $k \ll q$. Assuming that $\mathcal{P}_{\mathcal{R}}(k) \propto \mathcal{A}_{\mathcal{R}}(k/k_{\rm peak})^n$ (with appropriate cut-off) the very low

frequency, or far Infra-Red (IR), tail reads [54]

$$\Omega_{\rm GW,*}^{\rm IR}(k \ll k_{\rm peak}) \propto \begin{cases} \mathcal{P}_{\mathcal{R}}^2(k) & \text{for} \quad 0 < n < 3/2 \\ \left(\frac{k}{k_{\rm peak}}\right)^3 \ln^2\left(\frac{k}{k_{\rm peak}}\right) & \text{for} \quad n > 3/2 \end{cases} . \tag{4.16}$$

The first case of Eq. (4.16) is also valid for broad enough peaks.[7] This is for example the case of a broad log-normal for scales not too far from the peak. For scales very far from the peak, one has a clear exponential decay, and one recovers the universal k^3 scaling. The second case corresponds to sharp enough peaks. Note that for the case of a Dirac delta primordial spectrum (corresponding to the $\Delta \to 0$ limit of the log-normal), or a very sharp peak, the IR tail goes as k^2 instead of k^3. For more details on the universal IR scaling see Ref. [57]. The logarithmic running in the IR tail seems to be a characteristic of induced GWs in a radiation dominated universe [58, 59]. However, it is not present for general values of w [44].

Lastly, we consider the value $y \gg 1$, which corresponds to $k \gg k_{\rm peak}$. Assuming the same power-law ansatz but with $n < 0$, the high frequency, or Ultra-Violet (UV), the tail is given by [54]

$$\Omega_{\rm GW,*}^{\rm UV}(k) \propto \begin{cases} \mathcal{P}_{\mathcal{R}}^2(k) & \text{for} \quad 0 > n > -4 \\ \left(\frac{k}{k_{\rm peak}}\right)^4 \mathcal{P}_{\mathcal{R}}(k) & \text{for} \quad n < -4 \end{cases} . \tag{4.18}$$

From Eq. (4.18) we see that for sharply peaked primordial spectra the UV tail is quickly suppressed. In fact, for infinitely sharp primordial spectra (i.e. a Dirac delta), there is a sharp cut-off at $k = 2k_{\rm peak}$ due to momentum conservation. Namely, one cannot induce tensor modes with a frequency higher than the source term. Nevertheless, if the primordial spectrum is not too steep, we may obtain information from the UV tail. Examples of this are the inference of the magnitude of local non-Gaussianities [54] or the presence of a second phase of slow-roll inflation after the enhancement of fluctuations [60]. We also note that the k^4 scaling of the second case in Eq. (4.18) is sensitive to the value of c_s. When $c_s \to 1$, it gradually transitions to a k^2 scaling [60]. One can see the power-law scalings (or the sharp cut-off) in the induced GW spectra of Fig. 3. In the IR tail, all cases decay as $k^3 \ln^2 k$. Instead, the UV tail is more sensitive to the shape of the primordial spectrum.

We end with a helpful rough estimate of the peak amplitude of the GW spectrum today from Eqs. (4.12), (4.16) and (4.18), disregarding the presence of the resonant peak. This is given by

$$\Omega_{\rm GW,0}^{\rm peak} h^2 \approx 10^{-6}\, \mathcal{P}_{\mathcal{R}}^2(k_{\rm peak})\,, \tag{4.19}$$

which is good enough so that it also applies to a flat primordial spectrum (see Eq. (4.17)). From Eq. (4.19) we see that the amplitude of the primordial spectrum of fluctuations measured by CMB observations, around 10^{-9} [61], yields unobservable induced GWs. We need an enhancement of more than 4 orders of magnitude to have an observable GW signal.

[7] Looking at both Eqs. (4.16) and (4.18) we see that for broad enough spectra, a good estimate to the induced GW spectrum is given by $\Omega_{\rm GW,*}(k) \sim \mathcal{P}_{\mathcal{R}}^2(k)$. In the limit of a flat spectrum (that is $\mathcal{P}_{\mathcal{R}} \approx$ constant), one numerically finds that [41]

$$\Omega_{\rm GW,*}^{\rm flat} \approx 0.82\, \mathcal{P}_{\mathcal{R}}^2\,. \tag{4.17}$$

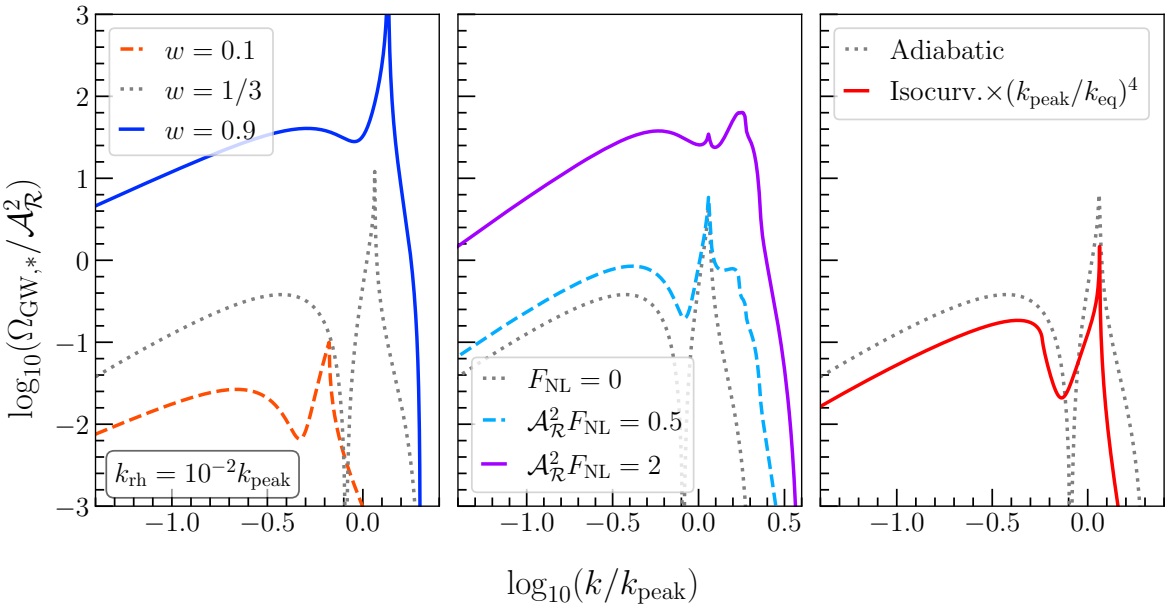


FIG. 4. Induced GW spectrum for different extensions assuming a Dirac delta peak in the primordial spectrum of fluctuations. The radiation domination result is shown in grey lines. On the left panel, we show the effects of softer (orange) and stiffer (blue) equations of state w from Ref. [42] (one may use SIGWfast [48] for the finite width case). See how the stiffer the w, the higher the GW spectrum and the sharper the resonant peak (which moves to higher frequencies). On the middle panel, we show the effects of local non-Gaussianity from Ref. [63] (data available at [64]). See how the effect of $F_{\rm NL}$ is to generate induced GW beyond the cut-off at $k = 2k_{\rm peak}$ from Gaussian fluctuations. On the right panel, we compare the induced GW spectrum from adiabatic and isocurvature initial conditions from Ref. [65]. Although the GW spectrum is qualitatively similar, they have a different shape of the resonant peak and the isocurvature-induced GWs have no destructive interference at $k = \sqrt{2} c_s k_{\rm peak}$.

D. Other extensions

It is important to consider physics beyond the simplest model of Gaussian primordial fluctuations and a radiation-dominated universe. First, we have not probed the early universe at such small scales, so one should be as agnostic as possible. Second, there are several attractive models of inflation and reheating (and much more; see, e.g., Ref. [62] for a review) which predict departures such as non-Gaussianities of primordial fluctuation and a matter or kinetic dominated phase of reheating, respectively with $w \ll 1$ (for an oscillating moduli field) and $w \approx 1$ (in quintessential inflation models). As we shall see, the induced GW spectrum depends on the initial conditions of primordial fluctuations (adiabatic/isocurvature and Gaussian/non-Gaussian) and the model of the primordial universe, e.g., the equation of state of the universe at the time of wave generation. Here we briefly mention below four different possibilities. We illustrate the effects of these possibilities on the induced GW spectrum in Fig. 4.

1. General equation of state of the very early universe

We know the universe had to be dominated by relativistic particles of the standard model around the time of BBN. More precisely, the radiation in the universe should have thermalised at $T \sim 4\,\mathrm{MeV}$ [66–69] for a successful BBN. We also have strong evidence for a period of cosmic inflation from CMB observations [61]. However, we do not know how or when the universe was re-heated. So, it is entirely plausible that there was a re-heating phase for $T > 4\,\mathrm{MeV}$ with a different equation of state w than radiation, e.g. a stiff fluid or a scalar field. If induced GWs (and PBHs) are generated during this w-dominated period, the shape of the induced GW spectral density will be modified. This is studied in detail in Refs. [42, 44]. The main changes are: (i) an enhancement/suppression due to the different background expansion, (ii) a different position of the resonant peak at $k_{\rm res} = 2c_s k_{\rm peak} = 2\sqrt{w}k_{\rm peak}$ and (iii) a different power-law in the low-frequency tail going as $f^{3-2|b|}$ where $b = (1-3w)/(1+3w)$. We illustrate these effects on the left of Fig. 4. In this way, induced GWs also probe the content of the primordial universe.

2. Primordial non-Gaussianity and Trispectrum

As we discussed around Eq. (4.4), the induced GW spectrum is determined by the 4-point function of curvature fluctuations. So, strictly speaking, induced GWs are also sensitive to any departure from Gaussian fluctuations. In general, this means that induced GWs probe the trispectrum of primordial fluctuations. However, if such fluctuations are generated during inflation, one expects perturbative departures from Gaussian fluctuations. And, therefore, the induced spectrum would only be slightly modified compared to the Gaussian case (see, e.g., Ref. [70] for a general treatment later applied to scale-invariant fluctuations). But, one could also assume a general ansatz for primordial non-Gaussianity and study its effects, regardless of the dynamics during inflation. The most studied case is local non-Gaussianity, which is parametrized as [71]

$$\Phi(x) = \Phi_{\rm Gauss}(x) + F_{\rm NL}(\Phi^2_{\rm Gauss}(x) - \langle\Phi^2_{\rm Gauss}(x)\rangle)\,, \tag{4.20}$$

where $\Phi_{\rm Gauss}$ is the Gaussian variable. See Ref. [54, 58, 72] for early works and Ref. [63] for a precise systematic approach. As shown in the middle panel of Fig. 4, the local non-Gaussian contribution adds a resonant peak at $k = 3c_s k_{\rm peak}$ and dominates over the Gaussian contribution for $\mathcal{A}^2_{\mathcal{R}} F_{\rm NL} > 1$. Non-Gaussianity does not alter the predictions for the IR tail of the induced GW spectrum, though. Note that there is no reason to stop the ansatz (4.20) at Φ^2, especially if non-Gaussianities are large, as one could include higher powers of Φ. See Ref. [73] for a nice general treatment. As a curious possibility, we also mention that a parity-violating trispectrum could lead to chiral-induced GWs, though only as a small correction [74].

3. Isocurvature initial conditions

Observations of the CMB tell that primordial fluctuations on the largest scales are adiabatic [61]. This means that there is a time slice of uniform total energy density with only curvature

perturbations but no relative number density fluctuations. But, at much smaller scales, this might not be the case. These initial conditions are called isocurvature [75, 76]. Isocurvature fluctuations do not directly source induced GWs (see Ref. [65] for the original work and Ref. [17] for a recent review). Instead, as the universe evolves, isocurvature fluctuations induce curvature fluctuations that later induce GWs. The price to pay is that the amplitude of induced GWs is suppressed by the energy density ratio of the subdominant field (responsible for isocurvature) with respect to the dominant field (responsible for curvature). For DM isocurvature fluctuations with a peaked isocurvature spectrum at $k_{\rm peak}$, the induced GW spectrum is suppressed by a factor $(k_{\rm eq}/k_{\rm peak})^4$, where $k_{\rm eq} \approx 0.01\,{\rm Mpc}^{-1}$ is the comoving scale at the standard matter-radiation equality. Interestingly, the amplitude of the isocurvature spectrum can be very large, say $\mathcal{A}_{\mathcal{S}} \gg 1$, and still be compatible with cosmological perturbation theory (up to the time fluctuations become non-linear and might form PBHs [77]). In this case, one has a detectable isocurvature-induced GW signal. Another interesting possibility is the GWs induced by the isocurvature from Poisson statistical fluctuations. This is a general signal when discrete objects or solitons exist, such as PBH, cosmic strings, oscillons, etcetera. This generic prediction has been dubbed "Universal Gravitational Waves of Solitons" in Refs. [78, 79]. In Sec. V, we study in detail the GWs from PBH number density fluctuations.

4. Induced GWs during inflation

We mainly discussed the GWs induced by primordial fluctuations and the different extensions after inflation. However, induced GW can also be generated during inflation as a secondary effect from vacuum fluctuations of spectator fields, such as scalar, vector, or tensor fields. The GW spectrum, in this case, is more model-dependent; therefore, we only cite the relevant literature. For scalar-induced GWs, e.g., in multifield models of inflation, we refer the reader to Refs. [80–83]. One typical case for vector field is axion inflation with Abelian or non-Abelian gauge fields, see Refs. [84–86]. For more references, we refer the reader to Komatsu's review on the topic [87]. Two form fields can also source GWs at second order [88], though not at the linear order. Additional tensor fields can linearly source GWs [89]. For a review on linear GWs from metric vacuum fluctuations during inflation, see Ref. [90].

V. GWS FROM THE PBH DOMINATED EARLY UNIVERSE

There is one scenario beyond the standard adiabatic primordial fluctuations which is interesting regarding the associated GW signals. This is the case when ultra-light PBHs dominate the very early universe and evaporate before BBN, reheating the universe via Hawking radiation [91–93]. Most relevant to secondary GWs is the fact that, as shown in Refs. [94–96] (see also Ref. [97]), an epoch of dust domination (such as a gas of PBHs) may enhance the production of induced GWs. Adding that, due to the discrete nature of PBHs, there are number density fluctuations from the Poisson noise in the PBH gas [98], one finds a unique induced GW signature of the PBH-dominated universe [98–100]. Below, we focus on the induced GW signatures following Ref. [99]. The physics

of BH evaporation is discussed in detail in Chapter 22.

A. The PBH dominated universe

For simplicity, let us assume a monochromatic PBH mass function. In that case, there are two free parameters in the PBH reheating scenario: the PBH mass at formation $M_{\rm PBH,f}$ and the initial fraction β of PBHs. The mass is proportional to the Hubble parameter at formation via Eq. (4.13). The initial fraction is related to the *mean* number density of PBHs per Hubble volume, that is

$$\beta = \frac{\rho_{\rm PBH,f}}{\rho_{\rm rad,f}} = \frac{4\pi\gamma}{3H_{\rm f}^3} n_{\rm PBH,f}\,, \tag{5.1}$$

where we used that $\rho_{\rm PBH,f} = M_{\rm PBH,f} \times n_{\rm PBH,f}$ and $\rho_{\rm rad,f} \approx 3H_{\rm f}^2 M_{\rm pl}^2$. We can also compute the mean distance between PBHs from the mean number density. The comoving wavenumber associated with this distance is given by [98]

$$k_{\rm UV} = \left(\frac{4\pi n_{{\rm PBH},f} a_{\rm f}}{3}\right)^{1/3} = \beta^{1/3}\gamma^{-1/3}\, k_{\rm f}\,, \tag{5.2}$$

where $k_{\rm f} = \mathcal{H}_{\rm f}$ is the comoving wavenumber that entered the horizon at formation.[8] This means that PBHs are initially displaced over super-Hubble distances.

The most interesting point is to note that, in fact, we have a gas of PBHs spread all over the universe. And, while it has on average a homogeneous distribution, there are statistical fluctuations. Namely, by chance, PBHs will be somewhere closer and somewhere sparser leading to number density fluctuations. Since PBHs form very rarely and, therefore, are mostly distributed uniformly in space, the distribution of PBHs follows Poisson statistics [98]. The typical scale of the fluctuations is set by $k_{\rm UV}$ (5.2). The dimensionless power spectrum of PBH number density fluctuations at formation is given by [98]

$$\mathcal{P}_{\delta_{\rm PBH,f}}(k) = \frac{2}{3\pi}\left(\frac{k}{k_{\rm UV}}\right)^3 \Theta(k_{\rm UV} - k)\,, \tag{5.3}$$

where $\delta_{\rm PBH,f} = \delta n_{\rm PBH,f}/n_{\rm PBH,f}$ and $\Theta(x)$ is the Heaviside theta function. Note that the picture of a PBH gas is only valid if we have at least one PBH in a given volume. Thus, in the PBH gas approximation, we have a cut-off at the mean inter-PBH distance, namely at $k = k_{\rm UV}$. Hence, the notation "UV" to refer to the high wavenumber cut-off of the spectrum. We illustrate the PBH number density fluctuations in Fig. 5.

After formation, PBHs evaporate via Hawking radiation. This leads to a cosmic time-dependent PBH mass given by [96]

$$M_{\rm PBH}(t) \approx M_{\rm PBH,f}\left(1 - \frac{t}{t_{\rm eva}}\right)^{1/3} \quad \text{where} \quad t_{\rm eva} = \frac{M_{\rm PBH,f}^3}{M_{\rm pl}^4}\frac{160}{3.8\pi g_H(T_{\rm PBH})}\,. \tag{5.4}$$

[8] Note that if such PBHs formed from the collapse of primordial fluctuations then $k_{\rm f}$ is equivalent to what we called $k_{\rm peak}$ in previous sections.

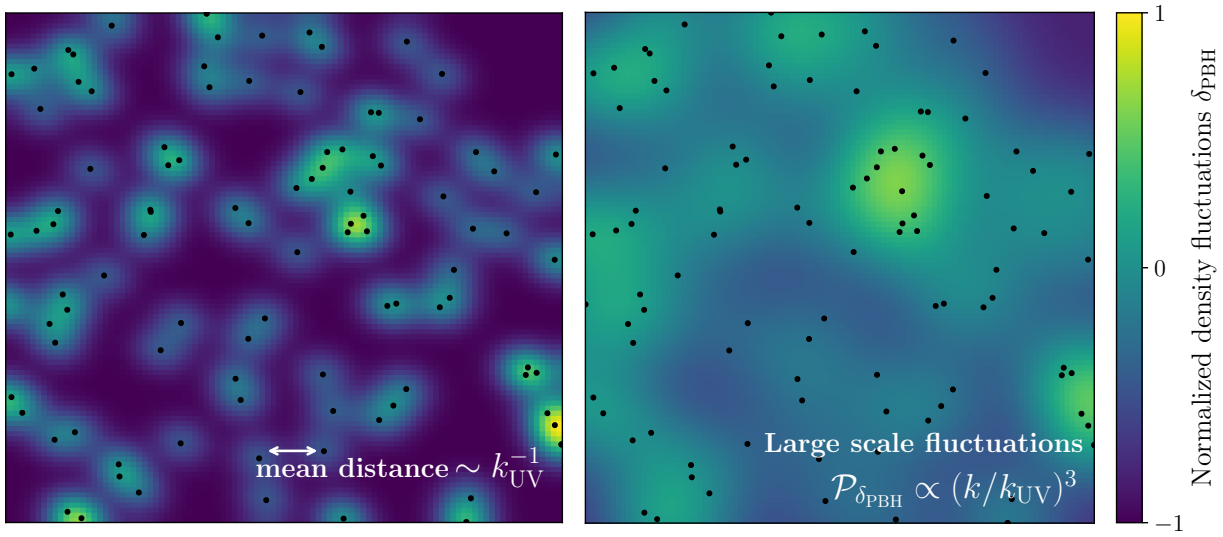


FIG. 5. PBH gas in the early universe. After formation, PBH are randomly distributed in space following a uniform distribution. This randomness leads to number density fluctuations. On the left panel, we show the kernel density estimate for the PBH number density in a fluid picture, smoothing on scales similar to the mean distance between PBHs. On the right panel, we show the kernel density estimate using a larger smoothing scale. See how number density fluctuations lead to large-scale density fluctuations of the PBH gas.

The subscript "eva" refers to evaluation at evaporation. Because of PBH number density conservation, we have that $n_{\rm PBH} \propto a^{-3}$. This result together with Eq. (5.4) completely determines the time evolution of $\rho_{\rm PBH}$. Note that, during the time that Hawking evaporation is not effective, we have that $\rho_{\rm PBH}/\rho_{\rm r} \propto a$ and, eventually, PBHs might dominate the very early universe. The requirement that PBHs dominate the very early universe, which is equivalent to the requirement that PBHs evaporate after the Early PBH-radiation Equality (let us call this time "eeq"), yields

$$\beta > 6 \times 10^{-10} \left(\frac{g_H(T_{\rm PBH,f})}{108}\right)^{1/2} \left(\frac{\gamma}{0.2}\right)^{-1/2} \left(\frac{M_{\rm PBH,f}}{10^4 {\rm g}}\right)^{-1} , \tag{5.5}$$

where $g_H(T_{\rm PBH})$ are the spin weighted degrees of freedom and $T_{\rm PBH,f}$ is the Hawking temperature of the PBH. For the standard model of particle physics, one has $g_H(T_{\rm PBH}) \approx 108$. The temperature of the radiation filling the universe after PBH evaporation (which follows from saying that $H_{\rm eva} = 2/(3t_{\rm eva})$, where the numerical factor is that of a matter-dominated universe) is given by

$$T_{\rm eva} \approx 2.8 \times 10^4 \, {\rm GeV} \left(\frac{M_{\rm PBH,f}}{10^4 {\rm g}}\right)^{-3/2} \left(\frac{g_H(T_{\rm PBH,f})}{108}\right)^{1/2} \left(\frac{g_\rho(T_{\rm eva})}{106.75}\right)^{-1/4} . \tag{5.6}$$

Imposing BBN bounds, namely that $T_{\rm eva} > 4\,{\rm MeV}$ [66–69], yields an upper bound on the PBH

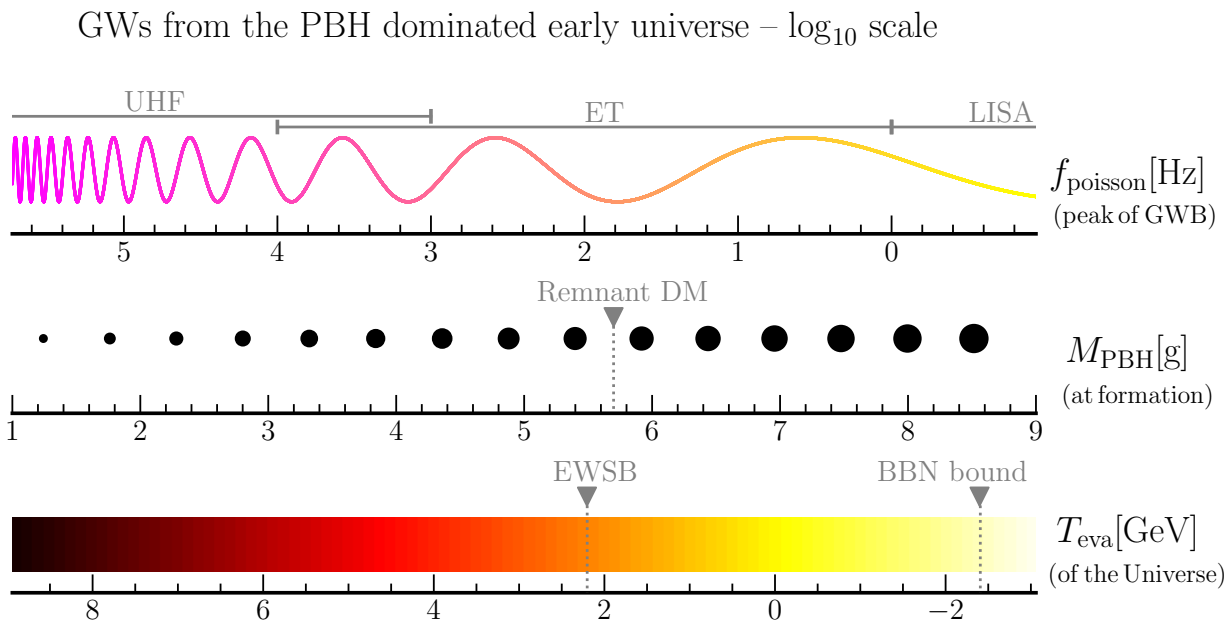


FIG. 6. Graphical summary of the GW signatures of the PBH-dominated universe. In the top panel, we show the frequency of the induced GWs associated with PBH reheating, $\log_{10} f_{\rm possion}$ (5.9). GW detectors in the relevant frequency range are UHF GW detectors, ET and LISA. In the middle panel, we show the corresponding PBH mass at formation, $\log_{10} M_{\rm PBH,f}$. We also show the PBH remnant scenario as DM [101], which requires $M_{\rm PBH,f} \approx 5\times 10^5$g [102]. In the bottom panel, we give the temperature of the radiation filling the universe right after PBH evaporation $\log_{10} T_{\rm eva}$ (5.6). We show the BBN constraint $T \sim 4\,$MeV [66–69] as well as the temperature corresponding to ElectroWeak Symmetry Breaking (EWSB), which is relevant for leptogenesis/baryogenesis, see e.g. Ref. [103]. Interestingly, leptogenesis via Right-handed neutrinos requires $M_{\rm PBH,f} \lesssim 10\,$g [104, 105].

mass as

$$M_{{\rm PBH},f} < 5\times 10^8 {\rm g}\,. \tag{5.7}$$

The dynamics of the PBH reheating scenario are illustrated in Fig. 7.

It is also useful to compute the wavenumber corresponding to the Horizon size at evaporation (e.g. relative to the Hubble radius today). This is given by

$$k_{\rm eva} \approx 4.7\times 10^{11}{\rm Mpc}^{-1}\left(\frac{M_{\rm PBH,f}}{10^4{\rm g}}\right)^{-3/2}\left(\frac{g_H(T_{\rm PBH})}{108}\right)^{1/2} \times\left(\frac{g_\rho(T_{\rm eva})}{106.75}\right)^{1/4}\left(\frac{g_s(T_{\rm eva})}{106.75}\right)^{-1/3}, \tag{5.8}$$

Since $T_{\rm eva}$ (5.6) only depends on $M_{\rm PBH,f}$ and not β, the same applies to $k_{\rm eva}$. The frequency associated with the numerical factor in Eq. (5.8), namely $4.7\times 10^{11}{\rm Mpc}^{-1}$, is about $7.3\times 10^{-4}\,$Hz. It turns out that, because of number density conservation (which implies $a\propto n_{\rm PBH}^{-1/3}$), $k_{\rm eva}$ computed

in terms of $k_{\rm f}$ has the same β dependence as $k_{\rm UV}$ (5.2). This implies that the ratio $k_{\rm UV}/k_{\rm eva}$ is independent of β. And, since $k_{\rm eva}$ from (5.8) depends on $M_{\rm PBH,f}$, the frequency associated with the cut-off of Poisson fluctuations (say $f_{\rm poisson} = k_{\rm UV}/2\pi$) reads

$$f_{\rm possion} \approx 1.7\times 10^3{\rm Hz}\left(\frac{M_{{\rm PBH},f}}{10^4{\rm g}}\right)^{-5/6}\left(\frac{g_H(T_{\rm PBH})}{108}\right)^{1/6}\left(\frac{g_\rho(T_{\rm eva})}{106.75}\right)^{1/4}\left(\frac{g_s(T_{\rm eva})}{106.75}\right)^{-1/3}. \quad (5.9)$$

This frequency is where the induced GW spectrum from PBH Poisson fluctuations peaks. We show the parameter space for $f_{\rm poisson}$, $M_{\rm PBH,f}$ and $T_{\rm eva}$ in Fig. (6). See how GW detectors such as the LVK collaboration, Einstein Telescope and Cosmic Explorer can probe a wide range of masses of ultra-light PBHs if such PBHs dominate the universe.

B. GWs from PBH number density fluctuations

As mentioned at the beginning of the section, PBHs are initially inhomogeneously distributed in space, according to Poisson statistics. The initial conditions of such number density fluctuations are isocurvature in nature [98, 99]. This is because as each PBH forms, it leaves a "hole" in the original radiation fluid. Thus, by energy conservation, one must have in the fluid picture that $\delta\rho_{\rm PBH}+\delta\rho_{\rm rad}=0$ initially. This is nothing but isocurvature initial fluctuations. For instance, in the Newton gauge and on superhorizon scales, one has that $2\Phi\sim\delta\rho_{\rm tot}=\delta\rho_{\rm PBH}+\delta\rho_{\rm rad}=0$, i.e. there are no initial curvature fluctuations. But, as the universe evolves, and the mean PBH energy density starts to overtake that of the radiation fluid (see Fig. (6)), there is a conversion from isocurvature to curvature fluctuations. The conversion is completed once PBHs dominate the universe. Then, if we call S_i the initial isocurvature, one finds that the curvature fluctuations at PBH domination are given by $\Phi(k\ll k_{\rm eeq})=S_i/5$ and $\Phi(k\gg k_{\rm eeq})\approx\frac{3}{4}S_i(k/k_{\rm eeq})^{-2}$ (see e.g. Ref. [75]), where $k_{\rm eeq}=\sqrt{2}\,\beta^{2/3}\gamma^{1/3}k_{\rm UV}$ is the scale that enters the horizon at PBH-radiation equality. Scales with $k\gg k_{\rm eeq}$ are suppressed because they entered the horizon during the radiation-dominated epoch when the subhorizon curvature fluctuations decay.

During the PBH-dominated phase, curvature fluctuations remain constant on all scales due to the zero pressure of the PBH gas. This also implies that PBH density fluctuations grow. During that period, there is a constant source of the induced GWs. Because of this, the induced GW spectrum for wavenumbers such that $k_{\rm eeq}<k<k_{\rm UV}$ turns out to be flat with an amplitude approximately given by [98]

$$\Omega^{\rm flat}_{\rm GW,PBH-dom}\approx 2\times 10^{13}\left(\frac{M_{\rm PBH,f}}{10^4{\rm g}}\right)^{4/3}\beta^{16/3}. \quad (5.10)$$

However, the biggest production of induced GWs happens right after PBH evaporate completely. The source of the enhancement is that by the time PBHs evaporate, the number of density fluctuations has grown very much. Then, the almost sudden transition from PBH to radiation domination (see in Fig. (7) that the transition takes less than a quarter of efold) leads to large radiation density fluctuations (this is called the "poltergeist" mechanism in Ref. [96]). Velocity fluctuations of the radiation fluid after evaporation are very large. From Eq. (4.6), we find that after evaporation

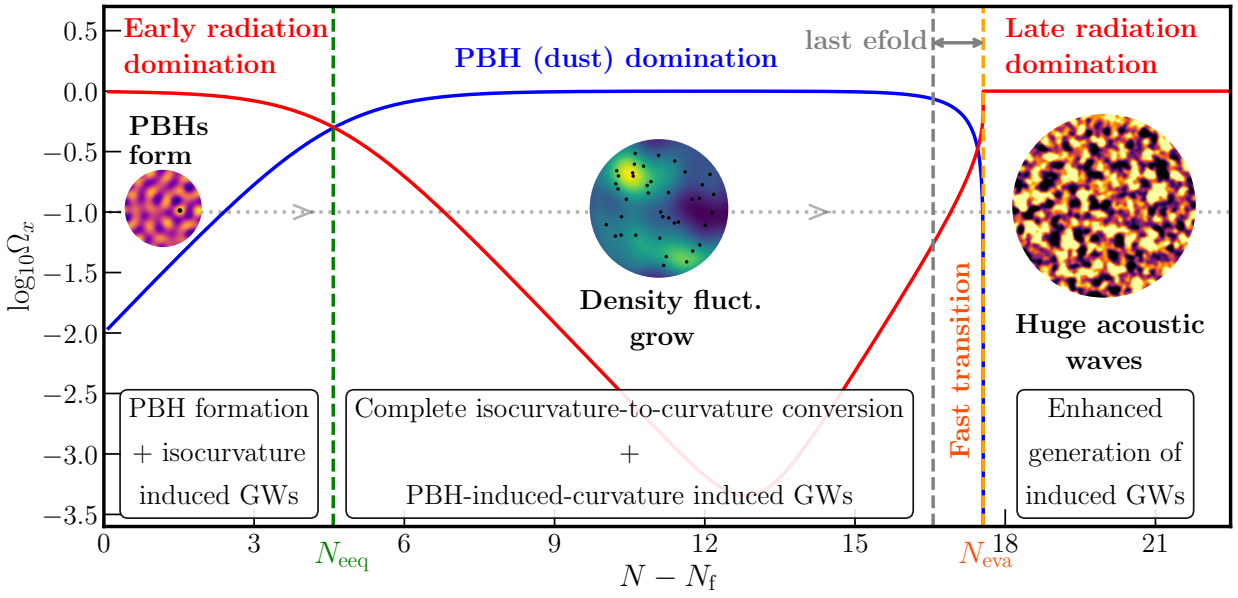


FIG. 7. Illustration of the universe's evolution in the PBH reheating scenario. We show $\Omega_{\rm PBH} = \rho_{\rm PBH}/\rho_{\rm tot}$ (with a blue line) and $\Omega_{\rm rad} = \rho_{\rm rad}/\rho_{\rm tot}$ (with a red line) in terms of efolds N from PBH formation $N_{\rm f}$. PBHs start to dominate the universe at $N_{\rm eeq}$ and evaporate completely at $N_{\rm eva}$. During the early radiation domination, PBHs form, but the induced GW signal associated with the formation and PBH isocurvature are respectively of too high frequency or too low amplitude. After $N_{\rm eeq}$, PBH isocurvature has been converted into curvature fluctuations and induced GW are more efficiently produced (derived in Ref. [98]). Because the complete evaporation of PBHs is very fast, there is an enhanced production of induced GWs right after PBH evaporation [96] (estimated in Ref. [99]).

$v \sim \Phi'/\mathcal{H} \propto (k/k_{\rm eva})^2$. Taking into account that the transition is not instantaneous leads to a suppression factor $\sim (k/k_{\rm eva})^{-1/3}$ [96] (that comes from the fact that PBHs lose energy according to $M_{\rm PBH} \propto (1-t/t_{\rm eva})^{1/3}$). An approximation to the induced GW spectrum after PBH evaporation is given by [99, 100]

$$\Omega_{\rm GW,eva}(k) \approx \Omega^{\rm peak}_{\rm GW,eva} \left(\frac{k}{k_{\rm UV}}\right) \Theta(k_{\rm UV} - k)\,, \tag{5.11}$$

where

$$\Omega_{\rm GW,eva} \approx 10^{30} \beta^{16/3} \left(\frac{\gamma}{0.2}\right)^{8/3} \left(\frac{g_H(T_{\rm PBH})}{108}\right)^{-17/9} \left(\frac{M_{\rm PBH,f}}{10^4 {\rm g}}\right)^{34/9}\,. \tag{5.12}$$

Note that Eq. (5.12) is larger than Eq. (5.10) by a factor $M^{22/9}_{\rm PBH,f}$. We show on the left of Fig. (8) the shape of the induced GW spectrum after PBH evaporation for different values of the PBH mass. On the right of Fig. (8), we show the bounds on β as a function of PBH mass and the corresponding amplitude of the induced GW spectrum peak (5.12).

It is important to note, though, that while the amplitude of Eq. (5.12) is very sensitive to the monochromatic mass function assumption (as a broad mass function leads to a more gradual

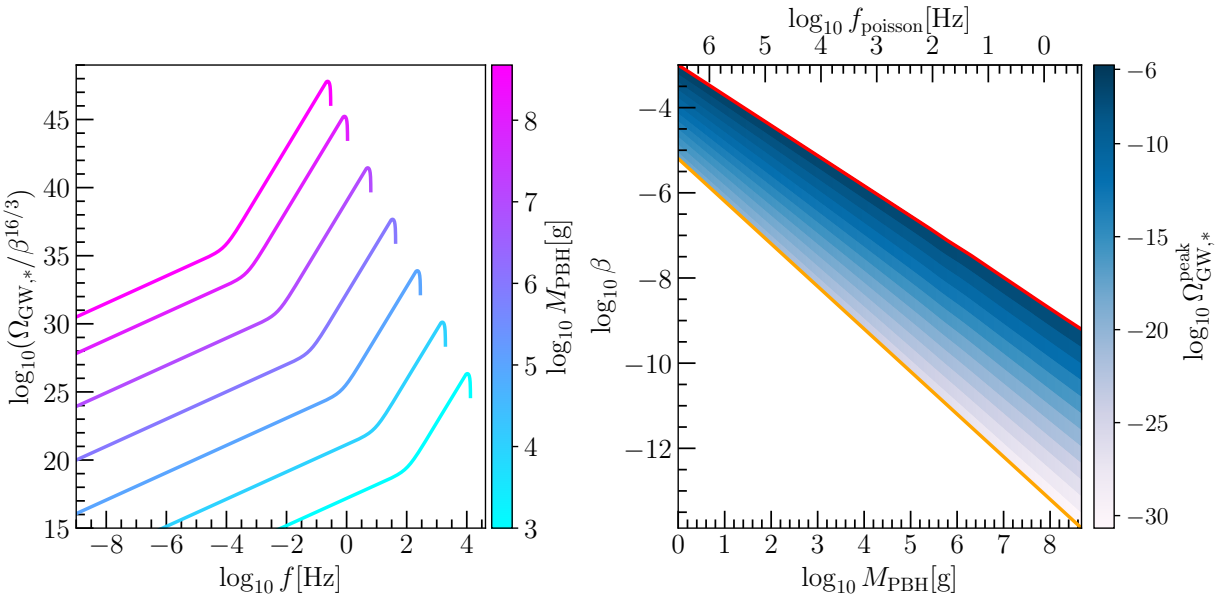


FIG. 8. On the left panel, we show the GW spectrum of GWs induced by PBH density fluctuations after PBH evaporation in terms of frequency for various values of $M_{\rm PBH,f}$ (shown in colour code) and normalized to $\beta^{16/3}$. The larger the mass, the higher the amplitude and the lower the peak frequency. On the right panel, we show the allowed parameter space of initial PBH fraction β requiring PBH domination (lower bound (5.5), orange line) and using BBN constraints on the effective number of relativistic species (upper bound (5.13), red line). In colour code, we show the corresponding amplitude of the peak of the GW spectrum (5.12).

transition to radiation domination), the amplitude of Eq. (5.10) is not as it only depends on the length PBH dominated phase. We can now use Eq. (5.12) to derive the current upper bounds on the initial fraction of PBHs using BBN bounds [106, 107] on the effective number of relativistic species (yielding $\Omega_{\rm GW,0}h^2 \lesssim 10^{-6}$). This leads us to

$$\beta < 1.1\times 10^{-6}\left(\frac{\gamma}{0.2}\right)^{-1/2}\left(\frac{g_H(T_{\rm PBH})}{108}\right)^{17/48}\left(\frac{g_*(T_{\rm eva})}{106.75}\right)^{1/16}\left(\frac{M_{\rm PBH,f}}{10^4{\rm g}}\right)^{-17/24}. \tag{5.13}$$

We end this section by mentioning shortcomings in the GW calculations of the PBH-dominated universe. First, although curvature fluctuations remain well inside the regime of cosmological perturbation theory, density fluctuations eventually become non-linear during PBH domination. For adiabatic primordial fluctuations, one may impose a cut-off to the primordial spectrum as in Ref. [96] to ensure that all fluctuations are linear at all times. However, PBH density fluctuations are intrinsic to the model, so an arbitrary cut-off may not be as justified. More accurate predictions require more sophisticated numerical simulations. Second, the GW enhancement is highly sensitive to the monochromatic mass assumption. A broad mass function will lead to a slower transition to the late radiation-dominated regime. According to Ref. [96], the resulting induced GW spectrum is exponentially sensitive to the width of the mass function. It would be interesting to do a more

detailed study.

VI. GW BACKGROUND FROM PBH BINARIES

For completeness, we dedicate the last section of this chapter to the GW background from unresolved PBH binaries. We want to mention that this short section does not do justice to such an interesting topic. The interested reader is encouraged to read Ref. [108] for a recent review (see also Refs. [109, 110] for earlier reviews) and Refs. [111, 112] for applications to PBHs.

The GW background from PBH binaries has a different nature than the induced GW background. The former results from the PBH themselves as a collection of unresolved PBH mergers, while the latter is generated by the large-scale structure of fluctuations (very few of which collapse to PBHs). The first is the local, continuous, source of GWs (active until PBHs evaporate), while the second is only generated in the very early universe. The GW spectral density from unresolved PBH binaries is computed as the integrated energy flux in GWs per logarithmic frequency bin normalized to the critical energy density of the universe ρ_c [108, 112, 113]. Schematically one has that [108]

$$\Omega_{\rm GW}(f) = \frac{1}{\rho_c} f F_{\rm GW}(f)\,, \tag{6.1}$$

where $F_{\rm GW}$ is the integrated flux per frequency bin.[9] The integrated flux is nothing but an integral over time of the flux emitted per binary times the rate at which PBHs merge. The flux per binary is computed as the emitted energy in GWs per binary and per frequency bin, say $dE_{\rm GW}/df$, times the inverse area that the GWs expanded, namely $\frac{1}{4\pi r^2}\frac{dE_{\rm GW}}{df}$ where r is the distance to the source. A nice analytical fit to $dE_{\rm GW}/df$, including the inspiral and merger phases, for non-precessing binaries is provided by Refs. [114, 115] (they provide the template for the strain in frequency domain but the energy is proportional to f^2 times the strain squared). The rate at which the PBH binaries merge as a function of time (i.e. the number of mergers per unit time and unit volume) for PBHs is computed in Chapter 17. Here, we call it $\mathcal{R}_{\rm merge}(t)$. Using the cosmic redshift as a time variable, i.e. $1+z = a(t)^{-1}$, the spectral density of the GW background from PBH binaries is given by [108]

$$\Omega_{\rm GW}^{\rm binaries}(f) = \frac{f}{3H_0^2 M_{\rm pl}^2}\int_0^{z_{\rm max}} dz \frac{\mathcal{R}_{\rm merge}(z)}{(1+z)H(z)}\frac{dE_{\rm GW}}{df_s}\,, \tag{6.2}$$

where we used the notation f_s to denote the frequency in the source frame, which is related to observed frequency by $f_s = f(1+z)$. If the GW background is composed by the merger of long-lived PBHs (i.e. $M_{\rm PBH,f} > 10^{15}\,{\rm g}$), then only the nearby binaries (in the cosmological sense) contributes to the integral. In this case, we can take the Hubble parameter of the ΛCDM universe, i.e. $H(z) = H_0\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}$. The explicit expression for $dE_{\rm GW}/df_s$ for non-spinning

[9] To be precise, $F_{\rm GW}$ also contains an integral over the direction of observation [108]. However, we assume an isotropic background of binaries and consider the average GW energy density.

PBHs (which is the expectation for PBH binaries at late times) reads [108, 112]

$$\frac{dE_{\rm GW}}{df_s} = \frac{(G\pi)^{2/3}\mathcal{M}_c^{5/3}}{3} f_s^{-1/3} \times \begin{cases} (1-(1.44-2.68\eta)\nu^{2/3})^2 & \text{for } f_s < f_1\,, \\ \omega_m f_s (1-1.89\nu+1.66\nu^2) & \text{for } f_1 \le f_s < f_2\,, \\ \omega_r f_s^{7/3} \frac{\nu_4^4}{4(\nu-\nu_2)^2+\nu_4^2} & \text{for } f_2 \le f_s < f_3\,, \end{cases} \tag{6.3}$$

where we defined $\nu_n \equiv (\pi G \mathcal{M} f_n)^{1/3}$, $\mathcal{M} = m_1 + m_2$ is the total mass, $\mathcal{M}_c = (m_1 m_2)^{3/5}/\mathcal{M}^{1/5}$ is the chirp mass and $\eta = m_1 m_2/\mathcal{M}^2$. The numerical parameters in Eq. (6.3) are given by

$$\begin{aligned} \nu_1^3 &= 0.07 + 0.64\eta - 0.06\eta^2 - 7.09\eta^3 \;\;,\;\; \nu_2^3 = 0.19 + 0.15\eta - 0.02\eta^2 + 2.33\eta^3\,, \\ \nu_3^3 &= 0.32 - 0.13\eta - 0.27\eta^2 + 4.92\eta^3 \;\;,\;\; \nu_4^3 = 0.09 - 0.41\eta + 1.83\eta^2 - 2.87\eta^3\,. \end{aligned} \tag{6.4}$$

Lastly, upper integration limit in Eq. (6.2) is given by $z_{\rm max} = f_3/f - 1$.

We end this section with some rough estimates for equal mass PBH binaries from Ref. [116] to develop a bit of intuition. The merger rate of equal PBH binaries evaluated today in its simplest form is given by [28, 29, 117, 118]

$$\mathcal{R}_{\rm merge} \approx 4\times 10^6\,{\rm Gpc}^{-3}{\rm yr}^{-1} \frac{{\rm f}_{\rm PBH}^2}{\left({\rm f}_{\rm PBH}^2+\sigma_{\rm eq}^2\right)^{21/74}} \left(\frac{M_{\rm PBH}}{M_\odot}\right)^{-32/37}\,, \tag{6.5}$$

where ${\rm f}_{\rm PBH}$ is the fraction of PBHs as dark matter and $\sigma_{\rm eq}^2 \approx 2.5\times 10^{-5}$. Putting some numbers and using that only the closest binaries, which are the loudest, contribute to the peak of the GW spectrum, we find that

$$\Omega_{\rm GW,binary}^{\rm peak} h^2 \approx 1.6\times 10^{-8} \left(\frac{M_{\rm PBH,f}}{M_\odot}\right)^{5/37} \left(\frac{{\rm f}_{\rm PBH}}{0.01}\right)^{53/37} \left(1+0.25\left(\frac{{\rm f}_{\rm PBH}}{0.01}\right)^{-2}\right)^{-21/74}\,. \tag{6.6}$$

The peak of the GW spectrum is then roughly at the ISCO frequency, namely

$$f_{\rm ISCO} \approx 2.2\,{\rm kHz}\left(\frac{M_{\rm PBH,f}}{M_\odot}\right)^{-1}\,. \tag{6.7}$$

This concludes our estimates for the GW background from unresolved PBH binaries.

VII. FINAL REMARKS

The induced GW background is a decisive observable for the PBH scenario from large primordial fluctuations. In fact, it is generally expected that any mechanism that forms PBHs is always accompanied by a loud GW signal (as discussed in Sec. II). The absence of an induced GW background would leave any PBH scenario in doubt. But, most important to us, any hint (or even discovery) of PBHs in a given mass range provides us with strong motivation to look for the induced GW background (see Sec. IV). From the GW background, we may also learn about the type of primordial density fluctuations (e.g. adiabatic/isocurvature and Gaussian/non-Gaussian) as well as the

content of the very early universe (e.g. the equation of state w); see Sec. IV D. We may also be able to test whether PBHs once dominated the very early universe, reheating it via Hawking radiation. And, it might give us hints of PBH remnants as DM, as we saw in Sec. V. We may also be able to observe the GW background from the unresolved PBH binaries (see Sec. VI), providing further evidence to the PBH scenario. Besides the interesting observational consequences, induced GWs also motivate us to improve our theoretical understanding of GW energy density when dealing with second-order cosmological perturbation theory (as discussed in Sec. III).

Interesting research directions that we have not discussed in this chapter are the study of statistical properties of the GW background (see, e.g. Ref. [119]) and GW background anisotropies (see, e.g., Ref. [120–122] for the general formulation and Ref. [123–129] for applications to PBHs) as additional probes/evidence of the PBH scenario. We have also not discussed high-frequency GWs from Hawking evaporation of PBHs, which might be testable by its effects on the effective number of relativistic species at CMB and BBN [107, 130–134]. There is also the possibility to induce GWs from scalar-tensor and tensor-tensor mixings, see e.g. Refs. [135–138].

We end this chapter by pointing out some directions worth exploring. First, one must examine the extensions beyond the simplest assumptions (radiation domination and Gaussian fluctuations) in more detail. This includes exploring motivated models that modify the expansion history of the early universe and derive more accurate kernels, studying how departures from General Relativity affect the induced GW spectrum, and, perhaps, exploring different types of non-Gaussianity of primordial fluctuations. Second, it would be very interesting to derive accurate predictions using numerical simulations for the GWs from the PBH-dominated early universe, as this is a unique way to probe the existence of PBHs that evaporated long ago. Lastly, it would be remarkable from the theoretical point of view to find a rigorous solution to the gauge ambiguity of induced GWs.

ACKNOWLEDGMENTS

I would like to thank all of my collaborators from whom I learned a lot and without whom I would have probably never explored induced GWs to the extent I did. These are V. Atal, S. Balaji, P. Bari, N. Bartolo, E. Dimastrogiovanni, M. Fasiello, G. Franciolini, A. Malhotra, S. Matarrese, A. Naruko, C. Lin, S. Passaglia, S. Patil, S. Pi, S. Renaux-Petel, M. Sasaki, V. Takhistov, G. Tasinato, T. Vargas, A. Wang, J. Wang and L. Witkowski. I also thank J. Fedrow, K. Inomata, E. Komatsu, T. Papanikolaou, S. Profumo, R. Samanta, J. Turner, C .Unal, Z. Weiner and S. Young for many stimulating discussions

[1] K. Tomita, Non-Linear Theory of Gravitational Instability in the Expanding Universe, Progress of Theoretical Physics **37**, 831 (1967), https://academic.oup.com/ptp/article-pdf/37/5/831/5234391/37-5-831.pdf.

[2] S. Matarrese, O. Pantano, and D. Saez, A General relativistic approach to the nonlinear evolution of collisionless matter, Phys. Rev. D **47**, 1311 (1993).

[3] S. Matarrese, O. Pantano, and D. Saez, General relativistic dynamics of irrotational dust: Cosmological implications, Phys. Rev. Lett. **72**, 320 (1994), arXiv:astro-ph/9310036.

[4] S. Matarrese, S. Mollerach, and M. Bruni, Second order perturbations of the Einstein-de Sitter universe, Phys. Rev. D **58**, 043504 (1998), arXiv:astro-ph/9707278.

[5] K. N. Ananda, C. Clarkson, and D. Wands, The Cosmological gravitational wave background from primordial density perturbations, Phys. Rev. D **75**, 123518 (2007), arXiv:gr-qc/0612013.

[6] D. Baumann, P. J. Steinhardt, K. Takahashi, and K. Ichiki, Gravitational Wave Spectrum Induced by Primordial Scalar Perturbations, Phys. Rev. D **76**, 084019 (2007), arXiv:hep-th/0703290.

[7] H. Assadullahi and D. Wands, Gravitational waves from an early matter era, Phys. Rev. D **79**, 083511 (2009), arXiv:0901.0989 [astro-ph.CO].

[8] L. Alabidi, K. Kohri, M. Sasaki, and Y. Sendouda, Observable induced gravitational waves from an early matter phase, JCAP **05**, 033, arXiv:1303.4519 [astro-ph.CO].

[9] R. Saito and J. Yokoyama, Gravitational wave background as a probe of the primordial black hole abundance, Phys. Rev. Lett. **102**, 161101 (2009), [Erratum: Phys.Rev.Lett. 107, 069901 (2011)], arXiv:0812.4339 [astro-ph].

[10] R. Saito and J. Yokoyama, Gravitational-Wave Constraints on the Abundance of Primordial Black Holes, Prog. Theor. Phys. **123**, 867 (2010), [Erratum: Prog.Theor.Phys. 126, 351–352 (2011)], arXiv:0912.5317 [astro-ph.CO].

[11] E. Bugaev and P. Klimai, Induced gravitational wave background and primordial black holes, Phys. Rev. D **81**, 023517 (2010), arXiv:0908.0664 [astro-ph.CO].

[12] E. V. Bugaev and P. A. Klimai, Bound on induced gravitational wave background from primordial black holes, JETP Lett. **91**, 1 (2010), arXiv:0911.0611 [astro-ph.CO].

[13] E. Bugaev and P. Klimai, Constraints on the induced gravitational wave background from primordial black holes, Phys. Rev. D **83**, 083521 (2011), arXiv:1012.4697 [astro-ph.CO].

[14] G. Domènech, Scalar Induced Gravitational Waves Review, Universe **7**, 398 (2021), arXiv:2109.01398 [gr-qc].

[15] C. Yuan and Q.-G. Huang, A topic review on probing primordial black hole dark matter with scalar induced gravitational waves, iScience **24**, 102860 (2021), arXiv:2103.04739 [astro-ph.GA].

[16] E. Bagui et al. (LISA Cosmology Working Group), Primordial black holes and their gravitational-wave signatures, (2023), arXiv:2310.19857 [astro-ph.CO].

[17] G. Domènech, Cosmological gravitational waves from isocurvature fluctuations, AAPPS Bull. **34**, 4 (2024), arXiv:2311.02065 [gr-qc].

[18] G. Domènech and M. Sasaki, Probing Primordial Black Hole Scenarios with Terrestrial Gravitational Wave Detectors, (2024), arXiv:2401.07615 [gr-qc].

[19] V. Mukhanov, Physical foundations of cosmology (Cambridge University Press, Cambridge, UK New York, 2005).

[20] D. Baumann, Cosmology (Cambridge University Press, 2022).

[21] G. Domènech and M. Sasaki, Hamiltonian approach to second order gauge invariant cosmological perturbations, Phys. Rev. D **97**, 023521 (2018), arXiv:1709.09804 [gr-qc].

[22] C. Caprini and D. G. Figueroa, Cosmological Backgrounds of Gravitational Waves, Class. Quant. Grav. **35**, 163001 (2018), arXiv:1801.04268 [astro-ph.CO].

[23] R. Roshan and G. White, Using gravitational waves to see the first second of the Universe, (2024), arXiv:2401.04388 [hep-ph].

[24] A. Afzal et al. (NANOGrav), The NANOGrav 15 yr Data Set: Search for Signals from New Physics, Astrophys. J. Lett. **951**, L11 (2023), arXiv:2306.16219 [astro-ph.HE].

[25] B. Carr and F. Kuhnel, Primordial Black Holes as Dark Matter: Recent Developments, Ann. Rev. Nucl. Part. Sci. **70**, 355 (2020), arXiv:2006.02838 [astro-ph.CO].

[26] P. Mroz et al., No large population of unbound or wide-orbit Jupiter-mass planets 10.1038/nature23276 (2017), arXiv:1707.07634 [astro-ph.EP].

[27] H. Niikura, M. Takada, S. Yokoyama, T. Sumi, and S. Masaki, Constraints on Earth-mass primordial black holes from OGLE 5-year microlensing events, Phys. Rev. D **99**, 083503 (2019), arXiv:1901.07120 [astro-ph.CO].

[28] 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**, 201301 (2016), arXiv:1603.00464 [astro-ph.CO].

[29] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914, Phys. Rev. Lett. **117**, 061101 (2016), [Erratum: Phys.Rev.Lett. 121, 059901 (2018)], arXiv:1603.08338 [astro-ph.CO].

[30] G. Franciolini, V. Baibhav, V. De Luca, K. K. Y. Ng, K. W. K. Wong, E. Berti, P. Pani, A. Riotto, and S. Vitale, Searching for a subpopulation of primordial black holes in LIGO-Virgo gravitational-wave data, Phys. Rev. D **105**, 083526 (2022), arXiv:2105.03349 [gr-qc].

[31] B. Carr and J. Silk, Primordial Black Holes as Generators of Cosmic Structures, Mon. Not. Roy. Astron. Soc. **478**, 3756 (2018), arXiv:1801.00672 [astro-ph.CO].

[32] A. M. Green and B. J. Kavanagh, Primordial Black Holes as a dark matter candidate, J. Phys. G **48**, 4 (2021), arXiv:2007.10722 [astro-ph.CO].

[33] M. Maggiore, Gravitational Waves. Vol. 1: Theory and Experiments, Oxford Master Series in Physics (Oxford University Press, 2007).

[34] G. Domènech and M. Sasaki, Approximate gauge independence of the induced gravitational wave spectrum, Phys. Rev. D **103**, 063531 (2021), arXiv:2012.14016 [gr-qc].

[35] J.-C. Hwang, D. Jeong, and H. Noh, Gauge dependence of gravitational waves generated from scalar perturbations, Astrophys. J. **842**, 46 (2017), arXiv:1704.03500 [astro-ph.CO].

[36] J.-O. Gong, Analytic Integral Solutions for Induced Gravitational Waves, Astrophys. J. **925**, 102 (2022), arXiv:1909.12708 [gr-qc].

[37] K. Tomikawa and T. Kobayashi, Gauge dependence of gravitational waves generated at second order from scalar perturbations, Phys. Rev. D **101**, 083529 (2020), arXiv:1910.01880 [gr-qc].

[38] M. Sipp and B. M. Schaefer, Scalar-induced gravitational waves in a ΛCDM cosmology, Phys. Rev. D **107**, 063538 (2023), arXiv:2212.01190 [astro-ph.CO].

[39] V. De Luca, G. Franciolini, A. Kehagias, and A. Riotto, On the Gauge Invariance of Cosmological Gravitational Waves, JCAP **03**, 014, arXiv:1911.09689 [gr-qc].

[40] K. Inomata and T. Terada, Gauge Independence of Induced Gravitational Waves, Phys. Rev. D **101**, 023523 (2020), arXiv:1912.00785 [gr-qc].

[41] K. Kohri and T. Terada, Semianalytic calculation of gravitational wave spectrum nonlinearly induced from primordial curvature perturbations, Phys. Rev. D **97**, 123532 (2018), arXiv:1804.08577 [gr-qc].

[42] G. Domènech, Induced gravitational waves in a general cosmological background, Int. J. Mod. Phys. D **29**, 2050028 (2020), arXiv:1912.05583 [gr-qc].

[43] J. R. Espinosa, D. Racco, and A. Riotto, A Cosmological Signature of the SM Higgs Instability: Gravitational Waves, JCAP **09**, 012, arXiv:1804.07732 [hep-ph].

[44] G. Domènech, S. Pi, and M. Sasaki, Induced gravitational waves as a probe of thermal history of the universe, JCAP **08**, 017, arXiv:2005.12314 [gr-qc].

[45] K. Inomata, M. Kawasaki, K. Mukaida, Y. Tada, and T. T. Yanagida, Inflationary primordial black holes for the LIGO gravitational wave events and pulsar timing array experiments, Phys. Rev. D **95**,

123510 (2017), arXiv:1611.06130 [astro-ph.CO].
[46] N. Aghanim et al. (Planck), Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020), [Erratum: Astron.Astrophys. 652, C4 (2021)], arXiv:1807.06209 [astro-ph.CO].
[47] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Primordial black holes—perspectives in gravitational wave astronomy, Class. Quant. Grav. **35**, 063001 (2018), arXiv:1801.05235 [astro-ph.CO].
[48] L. T. Witkowski, SIGWfast: a python package for the computation of scalar-induced gravitational wave spectra, (2022), arXiv:2209.05296 [astro-ph.CO].
[49] J. Fumagalli, S. Renaux-Petel, and L. T. Witkowski, Oscillations in the stochastic gravitational wave background from sharp features and particle production during inflation, JCAP **08**, 030, arXiv:2012.02761 [astro-ph.CO].
[50] S. Pi and M. Sasaki, Gravitational Waves Induced by Scalar Perturbations with a Lognormal Peak, JCAP **09**, 037, arXiv:2005.12306 [gr-qc].
[51] M. Braglia, X. Chen, and D. K. Hazra, Probing Primordial Features with the Stochastic Gravitational Wave Background, JCAP **03**, 005, arXiv:2012.05821 [astro-ph.CO].
[52] J. Fumagalli, S. e. Renaux-Petel, and L. T. Witkowski, Resonant features in the stochastic gravitational wave background, JCAP **08**, 059, arXiv:2105.06481 [astro-ph.CO].
[53] L. T. Witkowski, G. Domènech, J. Fumagalli, and S. Renaux-Petel, Expansion history-dependent oscillations in the scalar-induced gravitational wave background, JCAP **05** (05), 028, arXiv:2110.09480 [astro-ph.CO].
[54] V. Atal and G. Domènech, Probing non-Gaussianities with the high frequency tail of induced gravitational waves, JCAP **06**, 001, [Erratum: JCAP 10, E01 (2023)], arXiv:2103.01056 [astro-ph.CO].
[55] W.-T. Xu, J. Liu, T.-J. Gao, and Z.-K. Guo, Gravitational waves from double-inflection-point inflation, Phys. Rev. D **101**, 023505 (2020), arXiv:1907.05213 [astro-ph.CO].
[56] J. Liu, Z.-K. Guo, and R.-G. Cai, Analytical approximation of the scalar spectrum in the ultraslow-roll inflationary models, Phys. Rev. D **101**, 083535 (2020), arXiv:2003.02075 [astro-ph.CO].
[57] R.-G. Cai, S. Pi, and M. Sasaki, Universal infrared scaling of gravitational wave background spectra, Phys. Rev. D **102**, 083528 (2020), arXiv:1909.13728 [astro-ph.CO].
[58] R.-g. Cai, S. Pi, and M. Sasaki, Gravitational Waves Induced by non-Gaussian Scalar Perturbations, Phys. Rev. Lett. **122**, 201101 (2019), arXiv:1810.11000 [astro-ph.CO].
[59] C. Yuan, Z.-C. Chen, and Q.-G. Huang, Log-dependent slope of scalar induced gravitational waves in the infrared regions, Phys. Rev. D **101**, 043019 (2020), arXiv:1910.09099 [astro-ph.CO].
[60] S. Balaji, G. Domenech, and J. Silk, Induced gravitational waves from slow-roll inflation after an enhancing phase, JCAP **09**, 016, arXiv:2205.01696 [astro-ph.CO].
[61] Y. Akrami et al. (Planck), Planck 2018 results. X. Constraints on inflation, Astron. Astrophys. **641**, A10 (2020), arXiv:1807.06211 [astro-ph.CO].
[62] R. Allahverdi et al., The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe, Open J. Astrophys. **4**, 10.21105/astro.2006.16182 (2021), arXiv:2006.16182 [astro-ph.CO].
[63] P. Adshead, K. D. Lozanov, and Z. J. Weiner, Non-Gaussianity and the induced gravitational wave background, (2021), arXiv:2105.01659 [astro-ph.CO].
[64] P. Adshead, K. Lozanov, and Z. Weiner, ngsigw-results (2023).
[65] G. Domènech, S. Passaglia, and S. Renaux-Petel, Gravitational waves from dark matter isocurvature, JCAP **03** (03), 023, arXiv:2112.10163 [astro-ph.CO].
[66] M. Kawasaki, K. Kohri, and N. Sugiyama, Cosmological constraints on late time entropy production, Phys. Rev. Lett. **82**, 4168 (1999), arXiv:astro-ph/9811437.
[67] M. Kawasaki, K. Kohri, and N. Sugiyama, MeV scale reheating temperature and thermalization of neutrino background, Phys. Rev. D **62**, 023506 (2000), arXiv:astro-ph/0002127.

[68] S. Hannestad, What is the lowest possible reheating temperature?, Phys. Rev. D **70**, 043506 (2004), arXiv:astro-ph/0403291.

[69] T. Hasegawa, N. Hiroshima, K. Kohri, R. S. L. Hansen, T. Tram, and S. Hannestad, MeV-scale reheating temperature and thermalization of oscillating neutrinos by radiative and hadronic decays of massive particles, JCAP **12**, 012, arXiv:1908.10189 [hep-ph].

[70] S. Garcia-Saenz, L. Pinol, S. Renaux-Petel, and D. Werth, No-go theorem for scalar-trispectrum-induced gravitational waves, JCAP **03**, 057, arXiv:2207.14267 [astro-ph.CO].

[71] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Measuring primordial non-Gaussianity in the cosmic microwave background, Astrophys. J. **634**, 14 (2005), arXiv:astro-ph/0305189.

[72] C. Unal, Imprints of Primordial Non-Gaussianity on Gravitational Wave Spectrum, Phys. Rev. D **99**, 041301 (2019), arXiv:1811.09151 [astro-ph.CO].

[73] K. T. Abe, R. Inui, Y. Tada, and S. Yokoyama, Primordial black holes and gravitational waves induced by exponential-tailed perturbations, JCAP **05**, 044, arXiv:2209.13891 [astro-ph.CO].

[74] S. Garcia-Saenz, Y. Lu, and Z. Shuai, Scalar-induced gravitational waves from ghost inflation and parity violation, Phys. Rev. D **108**, 123507 (2023), arXiv:2306.09052 [gr-qc].

[75] H. Kodama and M. Sasaki, Evolution of Isocurvature Perturbations. 2. Radiation Dust Universe, Int. J. Mod. Phys. A **2**, 491 (1987).

[76] K. A. Malik and D. Wands, Adiabatic and entropy perturbations with interacting fluids and fields, JCAP **02**, 007, arXiv:astro-ph/0411703.

[77] S. Passaglia and M. Sasaki, Primordial black holes from CDM isocurvature perturbations, Phys. Rev. D **105**, 103530 (2022), arXiv:2109.12824 [astro-ph.CO].

[78] K. D. Lozanov, M. Sasaki, and V. Takhistov, Universal Gravitational Wave Signatures of Cosmological Solitons, (2023), arXiv:2304.06709 [astro-ph.CO].

[79] K. D. Lozanov, M. Sasaki, and V. Takhistov, Universal gravitational waves from interacting and clustered solitons, Phys. Lett. B **848**, 138392 (2024), arXiv:2309.14193 [astro-ph.CO].

[80] M. Biagetti, M. Fasiello, and A. Riotto, Enhancing Inflationary Tensor Modes through Spectator Fields, Phys. Rev. D **88**, 103518 (2013), arXiv:1305.7241 [astro-ph.CO].

[81] Z. Zhou, J. Jiang, Y.-F. Cai, M. Sasaki, and S. Pi, Primordial black holes and gravitational waves from resonant amplification during inflation, Phys. Rev. D **102**, 103527 (2020), arXiv:2010.03537 [astro-ph.CO].

[82] J. Fumagalli, G. A. Palma, S. Renaux-Petel, S. Sypsas, L. T. Witkowski, and C. Zenteno, Primordial gravitational waves from excited states, JHEP **03**, 196, arXiv:2111.14664 [astro-ph.CO].

[83] K. Inomata, Bound on induced gravitational waves during inflation era, Phys. Rev. D **104**, 123525 (2021), arXiv:2109.06192 [astro-ph.CO].

[84] N. Barnaby, E. Pajer, and M. Peloso, Gauge Field Production in Axion Inflation: Consequences for Monodromy, non-Gaussianity in the CMB, and Gravitational Waves at Interferometers, Phys. Rev. D **85**, 023525 (2012), arXiv:1110.3327 [astro-ph.CO].

[85] J. Garcia-Bellido, M. Peloso, and C. Unal, Gravitational waves at interferometer scales and primordial black holes in axion inflation, JCAP **12**, 031, arXiv:1610.03763 [astro-ph.CO].

[86] A. Maleknejad, Axion Inflation with an SU(2) Gauge Field: Detectable Chiral Gravity Waves, JHEP **07**, 104, arXiv:1604.03327 [hep-ph].

[87] E. Komatsu, New physics from the polarized light of the cosmic microwave background, Nature Rev. Phys. **4**, 452 (2022), arXiv:2202.13919 [astro-ph.CO].

[88] T. Fujita, H. Nakatsuka, I. Obata, and S. Young, Inflation with two-form field: the production of primordial black holes and gravitational waves, JCAP **09**, 017, arXiv:2202.02401 [astro-ph.CO].

[89] M. A. Gorji and M. Sasaki, Primordial-tensor-induced stochastic gravitational waves, Phys. Lett. B **846**, 138236 (2023), arXiv:2302.14080 [gr-qc].

[90] M. C. Guzzetti, N. Bartolo, M. Liguori, and S. Matarrese, Gravitational waves from inflation, Riv. Nuovo Cim. **39**, 399 (2016), arXiv:1605.01615 [astro-ph.CO].

[91] B. J. Carr, Some cosmological consequences of primordial black-hole evaporations, Astrophys. J. **206**, 8 (1976).

[92] J. E. Lidsey, T. Matos, and L. A. Urena-Lopez, The Inflaton field as selfinteracting dark matter in the brane world scenario, Phys. Rev. D **66**, 023514 (2002), arXiv:astro-ph/0111292.

[93] R. Anantua, R. Easther, and J. T. Giblin, GUT-Scale Primordial Black Holes: Consequences and Constraints, Phys. Rev. Lett. **103**, 111303 (2009), arXiv:0812.0825 [astro-ph].

[94] K. Inomata, K. Kohri, T. Nakama, and T. Terada, Enhancement of Gravitational Waves Induced by Scalar Perturbations due to a Sudden Transition from an Early Matter Era to the Radiation Era, Phys. Rev. D **100**, 043532 (2019), arXiv:1904.12879 [astro-ph.CO].

[95] K. Inomata, K. Kohri, T. Nakama, and T. Terada, Gravitational Waves Induced by Scalar Perturbations during a Gradual Transition from an Early Matter Era to the Radiation Era, JCAP **10**, 071, arXiv:1904.12878 [astro-ph.CO].

[96] K. Inomata, M. Kawasaki, K. Mukaida, T. Terada, and T. T. Yanagida, Gravitational Wave Production right after a Primordial Black Hole Evaporation, Phys. Rev. D **101**, 123533 (2020), arXiv:2003.10455 [astro-ph.CO].

[97] M. Pearce, L. Pearce, G. White, and C. Balázs, Gravitational Wave Signals From Early Matter Domination: Interpolating Between Fast and Slow Transitions, (2023), arXiv:2311.12340 [astro-ph.CO].

[98] T. Papanikolaou, V. Vennin, and D. Langlois, Gravitational waves from a universe filled with primordial black holes, JCAP **03**, 053, arXiv:2010.11573 [astro-ph.CO].

[99] G. Domènech, C. Lin, and M. Sasaki, Gravitational wave constraints on the primordial black hole dominated early universe, JCAP **04**, 062, [Erratum: JCAP 11, E01 (2021)], arXiv:2012.08151 [gr-qc].

[100] G. Domènech, V. Takhistov, and M. Sasaki, Exploring evaporating primordial black holes with gravitational waves, Phys. Lett. B **823**, 136722 (2021), arXiv:2105.06816 [astro-ph.CO].

[101] J. H. MacGibbon, Can Planck-mass relics of evaporating black holes close the universe?, Nature **329**, 308 (1987).

[102] G. Domènech and M. Sasaki, Gravitational wave hints black hole remnants as dark matter, Class. Quant. Grav. **40**, 177001 (2023), arXiv:2303.07661 [gr-qc].

[103] D. Hooper and G. Krnjaic, GUT Baryogenesis With Primordial Black Holes, Phys. Rev. D **103**, 043504 (2021), arXiv:2010.01134 [hep-ph].

[104] S. Datta, A. Ghosal, and R. Samanta, Baryogenesis from ultralight primordial black holes and strong gravitational waves from cosmic strings, JCAP **08**, 021, arXiv:2012.14981 [hep-ph].

[105] N. Bernal, C. S. Fong, Y. F. Perez-Gonzalez, and J. Turner, Rescuing high-scale leptogenesis using primordial black holes, Phys. Rev. D **106**, 035019 (2022), arXiv:2203.08823 [hep-ph].

[106] R. H. Cyburt, B. D. Fields, K. A. Olive, and E. Skillman, New BBN limits on physics beyond the standard model from ^4He, Astropart. Phys. **23**, 313 (2005), arXiv:astro-ph/0408033.

[107] A. Arbey, J. Auffinger, P. Sandick, B. Shams Es Haghi, and K. Sinha, Precision calculation of dark radiation from spinning primordial black holes and early matter-dominated eras, Phys. Rev. D **103**, 123549 (2021), arXiv:2104.04051 [astro-ph.CO].

[108] T. Regimbau, The Quest for the Astrophysical Gravitational-Wave Background with Terrestrial Detectors, Symmetry **14**, 270 (2022).

[109] T. Regimbau, The astrophysical gravitational wave stochastic background, Res. Astron. Astrophys. **11**, 369 (2011), arXiv:1101.2762 [astro-ph.CO].

[110] P. A. Rosado, Gravitational wave background from binary systems, Phys. Rev. D **84**, 084004 (2011), arXiv:1106.5795 [gr-qc].

[111] S. Wang, T. Terada, and K. Kohri, Prospective constraints on the primordial black hole abundance from the stochastic gravitational-wave backgrounds produced by coalescing events and curvature perturbations, Phys. Rev. D **99**, 103531 (2019), [Erratum: Phys.Rev.D 101, 069901 (2020)], arXiv:1903.05924 [astro-ph.CO].

[112] M. Braglia, J. Garcia-Bellido, and S. Kuroyanagi, Testing Primordial Black Holes with multi-band observations of the stochastic gravitational wave background, JCAP **12** (12), 012, arXiv:2110.07488 [astro-ph.CO].

[113] X.-J. Zhu, E. Howell, T. Regimbau, D. Blair, and Z.-H. Zhu, Stochastic Gravitational Wave Background from Coalescing Binary Black Holes, Astrophys. J. **739**, 86 (2011), arXiv:1104.3565 [gr-qc].

[114] P. Ajith et al., A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries, Phys. Rev. D **77**, 104017 (2008), [Erratum: Phys.Rev.D 79, 129901 (2009)], arXiv:0710.2335 [gr-qc].

[115] P. Ajith et al., Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins, Phys. Rev. Lett. **106**, 241101 (2011), arXiv:0909.2867 [gr-qc].

[116] G. Domènech, Were recently reported MHz events planet mass primordial black hole mergers?, Eur. Phys. J. C **81**, 1042 (2021), arXiv:2110.00550 [astro-ph.CO].

[117] T. Nakamura, M. Sasaki, T. Tanaka, and K. S. Thorne, Gravitational waves from coalescing black hole MACHO binaries, Astrophys. J. Lett. **487**, L139 (1997), arXiv:astro-ph/9708060.

[118] Y. Ali-Haïmoud, E. D. Kovetz, and M. Kamionkowski, Merger rate of primordial black-hole binaries, Phys. Rev. D **96**, 123523 (2017), arXiv:1709.06576 [astro-ph.CO].

[119] M. Braglia, J. Garcia-Bellido, and S. Kuroyanagi, Tracking the origin of black holes with the stochastic gravitational wave background popcorn signal, Mon. Not. Roy. Astron. Soc. **519**, 6008 (2023), arXiv:2201.13414 [astro-ph.CO].

[120] C. R. Contaldi, Anisotropies of Gravitational Wave Backgrounds: A Line Of Sight Approach, Phys. Lett. B **771**, 9 (2017), arXiv:1609.08168 [astro-ph.CO].

[121] N. Bartolo, D. Bertacca, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, Anisotropies and non-Gaussianity of the Cosmological Gravitational Wave Background, Phys. Rev. D **100**, 121501 (2019), arXiv:1908.00527 [astro-ph.CO].

[122] N. Bartolo et al. (LISA Cosmology Working Group), Probing anisotropies of the Stochastic Gravitational Wave Background with LISA, JCAP **11**, 009, arXiv:2201.08782 [astro-ph.CO].

[123] G. Scelfo, N. Bellomo, A. Raccanelli, S. Matarrese, and L. Verde, GW×LSS: chasing the progenitors of merging binary black holes, JCAP **09**, 039, arXiv:1809.03528 [astro-ph.CO].

[124] S. Wang, V. Vardanyan, and K. Kohri, Probing primordial black holes with anisotropies in stochastic gravitational-wave background, Phys. Rev. D **106**, 123511 (2022), arXiv:2107.01935 [gr-qc].

[125] E. Dimastrogiovanni, M. Fasiello, A. Malhotra, and G. Tasinato, Enhancing gravitational wave anisotropies with peaked scalar sources, JCAP **01**, 018, arXiv:2205.05644 [astro-ph.CO].

[126] S. Profumo and F. Yang, Anisotropy of the stochastic gravitational wave background from sub-horizon-collapsed primordial black hole mergers, Phys. Rev. D **108**, 103528 (2023), arXiv:2306.07454 [astro-ph.CO].

[127] N. Bartolo, D. Bertacca, V. De Luca, G. Franciolini, S. Matarrese, M. Peloso, A. Ricciardone, A. Riotto, and G. Tasinato, Gravitational wave anisotropies from primordial black holes, JCAP **02**, 028, arXiv:1909.12619 [astro-ph.CO].

[128] J.-P. Li, S. Wang, Z.-C. Zhao, and K. Kohri, Primordial non-Gaussianity f $_{NL}$ and anisotropies in scalar-induced gravitational waves, JCAP **10**, 056, arXiv:2305.19950 [astro-ph.CO].

[129] Y.-H. Yu and S. Wang, Anisotropies in Scalar-Induced Gravitational-Wave Background from Inflaton-Curvaton Mixed Scenario with Sound Speed Resonance, (2023), arXiv:2310.14606 [astro-ph.CO].
[130] D. Hooper, G. Krnjaic, and S. D. McDermott, Dark Radiation and Superheavy Dark Matter from Black Hole Domination, JHEP **08**, 001, arXiv:1905.01301 [hep-ph].
[131] I. Masina, Dark matter and dark radiation from evaporating primordial black holes, Eur. Phys. J. Plus **135**, 552 (2020), arXiv:2004.04740 [hep-ph].
[132] I. Masina, Dark Matter and Dark Radiation from Evaporating Kerr Primordial Black Holes, Grav. Cosmol. **27**, 315 (2021), arXiv:2103.13825 [gr-qc].
[133] T. C. Gehrman, B. Shams Es Haghi, K. Sinha, and T. Xu, The primordial black holes that disappeared: connections to dark matter and MHz-GHz gravitational Waves, JCAP **10**, 001, arXiv:2304.09194 [hep-ph].
[134] A. Ireland, S. Profumo, and J. Scharnhorst, Primordial gravitational waves from black hole evaporation in standard and nonstandard cosmologies, Phys. Rev. D **107**, 104021 (2023), arXiv:2302.10188 [gr-qc].
[135] Z. Chang, X. Zhang, and J.-Z. Zhou, Gravitational waves from primordial scalar and tensor perturbations, Phys. Rev. D **107**, 063510 (2023), arXiv:2209.07693 [astro-ph.CO].
[136] Y.-H. Yu and S. Wang, Primordial Gravitational Waves Assisted by Cosmological Scalar Perturbations, (2023), arXiv:2303.03897 [astro-ph.CO].
[137] P. Bari, N. Bartolo, G. Domènech, and S. Matarrese, Gravitational waves induced by scalar-tensor mixing, Phys. Rev. D **109**, 023509 (2024), arXiv:2307.05404 [astro-ph.CO].
[138] R. Picard and K. A. Malik, Induced gravitational waves: the effect of first order tensor perturbations, (2023), arXiv:2311.14513 [astro-ph.CO].

UT-WI-09-2024

Positivity in Amplitudes from Quantum Entanglement

Rafael Aoude,[1] Gilly Elor,[2] Grant N. Remmen,[3] and Olcyr Sumensari[4]

[1] *Higgs Centre for Theoretical Physics, School of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3JZ, Scotland, United Kingdom*
[2] *Weinberg Institute, Department of Physics, University of Texas at Austin, Austin, TX 78712, United States*
[3] *Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, United States*
[4] *IJCLab, Pôle Théorie (Bat. 210), CNRS/IN2P3 et Université, Paris-Saclay, 91405 Orsay, France*

We show that positivity of the imaginary part of forward elastic amplitudes for perturbative scattering is equivalent to consistency of the entanglement generated by the S-matrix, for states with arbitrary internal quantum numbers such as flavor. We also analyze "disentanglers," certain highly entangled initial states for which the action of the S-matrix is to decrease subsystem entanglement.

arXiv:2402.16956v1 [hep-th] 26 Feb 2024

Introduction.—Scattering amplitudes are among the most foundational and compelling observables in physics. As the fundamental objects of study in quantum field theory, their structure underpins much of modern high-energy physics, from phenomenology to string theory. The mathematical structure of amplitudes has been studied for decades, and important theorems arising from unitarity and locality are well known, including constraints on their asymptotic growth and analytic structure, e.g., the bounds of Froissart, Martin, Lehmann, et al. [1–3]. Moreover, in the modern amplitudes program, it has been shown that unitarity and locality themselves can emerge from more primordial mathematical structures [4, 5].

Chief among the many useful and famous facts about scattering amplitudes is the optical theorem, which relates the imaginary part of an amplitude—at forward kinematics, where the outgoing and incoming states are identical and particles pass straight through each other—to its cross section. Systematic use of this fact, dubbed "positivity" of amplitudes, has especially in the last two decades allowed the laws of physics themselves to be constrained using analytic dispersion relations, implying numerous different bounds on the space of coefficients of higher-dimension operators in effective field theories (EFTs) [6–10], ranging from the standard model EFT [11–16] to quantum corrections to gravity [17–21].

Meanwhile, entanglement is the quintessentially novel property of quantum mechanics [22, 23]. Historically, it was experimentally verified in classic laboratory demonstrations of violations of Bell's inequality [23, 24]. Excitingly, the highest-energy probes of entanglement can now be implemented in colliders, such as measurements of spin entanglement in top quark systems recently observed by the ATLAS Collaboration at the LHC [25] as proposed in Ref. [26]; see also Refs. [27–30]. These measurements open novel directions in constraining possible new physics using quantum information [31–33].

Quantum information theory also provides many positivity bounds on physics of an a priori different sort than dispersion relations. These bounds take the form of inequalities on quantum entanglement, that is, the von Neumann entropy $S(\hat{\rho}) = -\mathrm{Tr}[\hat{\rho}\log\hat{\rho}]$ for a density matrix $\hat{\rho}$. Various such bounds can be derived from quantum mechanics alone as consequences of linear algebra, including monogamy of entanglement, subadditivity, strong subadditivity, the Araki-Lieb inequality, etc. [34]. By considering other quantum information theoretic measures, such as relative entropy, yet more relations can be found [35]. In addition, entanglement entropy for states of conformal field theories with holographic spacetime duals are subject to additional bounds, such as monogamy of mutual information [36] or reflected entropy inequalities [37, 38], which can be proved geometrically via the AdS/CFT correspondence. Underpinning all of these results is the basic requirement of positivity of entanglement entropy, $S(\hat{\rho}) \geq 0$.

An interesting question is whether these two notions of positivity—that is, the amplitudes' positivity and entanglement entropy—can be related. Do quantum information bounds place additional constraints on EFTs? Or is the optical theorem somehow itself a statement about the entanglement structure of the S-matrix? While there are many realizations of the unitarity in quantum field theory, in this paper we aim to connect those associated with these two manifestations of positivity. The intersection of quantum entanglement and the S-matrix has not been fully investigated, though there have been interesting observations made about the relation between entanglement extremization and low-energy emergent symmetries [39–43], as well as between entanglement and the Yang-Mills/gravity double copy [44]. In notable work, Ref. [45] considered the quantum Tsallis entropy and showed that, subject to certain conditions on an unentangled initial state, it grows under perturbative scattering. The deeper question of whether positivity of the quantum entanglement itself can *imply* amplitudes' positivity, in a broader context of states with general quantum numbers, has not been explored. This question—focusing on discrete Hilbert spaces and asking the converse of whether positivity can emerge from entanglement—is especially interesting from both a quantum computing and phenomenological standpoint, as many particles in the standard model possess flavor or polarization.

In this Letter, we address this problem and show that

amplitudes' positivity is *precisely equivalent* to positivity of entropy, for a general perturbative quantum theory. Specifically, we will show that the linearized entropy,[1] defined as

$$\begin{aligned} \mathcal{E}[\Omega] &= 1 - \mathrm{Tr}_A\left[\hat{\rho}_A^2\right] , \\ \hat{\rho}_A &= \mathrm{Tr}_B \hat{\rho}_{AB} , \;\; \hat{\rho}_{AB} = \hat{S}|\Omega\rangle\langle\Omega|\hat{S}^\dagger , \end{aligned} \tag{1}$$

for a two-to-two scattering process through S-matrix $\hat{S}$, is nonnegative *if and only if* the corresponding amplitude has positive imaginary part in the forward limit, i.e.,

$$\mathcal{E}\left[|\Omega^{\mathrm{prod}}\rangle\right] \geq 0 \quad \Longleftrightarrow \quad \mathrm{Im}\,\mathcal{A} > 0 \,. \tag{2}$$

Furthermore, we will show that the elastic optical theorem implies $\mathcal{E}\left[|\Omega^{\mathrm{prod}}\rangle\right] \geq 0$. We will write a general initial two-particle state in Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ as $|\Omega\rangle = \sum_{ab} \Omega_{ab} |k_1, a\rangle_A \otimes |k_2, b\rangle_B$, with definite momenta k_i and summed over internal quantum numbers, and Ω^{prod} refers to a product (unentangled) initial state. For the basis vectors we use the notation $|k_1, a\rangle_A \otimes |k_2, b\rangle_B \equiv |k_1, a; k_2, b\rangle$, and write $\hat{\rho}_A$ for the reduced density matrix in $\mathcal{H}_A$ after tracing out degrees of freedom in $\mathcal{H}_B$. The results of this work represent an exciting connection between entanglement and the S-matrix, opening the door to future insights that might be gained in the amplitudes program via the use of powerful results from quantum information theory. From a phenomenological perspective, the study of entanglement in EFTs could yield insights to model building principles in a similar spirit to Refs. [39–43].

Entanglement from Two-to-Two Scattering.—The entanglement entropy can be computed perturbatively, order by order, in a quantum field theory [46]. Furthermore, such calculations have been shown to be valid in a Wilsonian EFT [47, 48] where one deals with effective operators in the infrared. In practice, this amounts to tracing over continuous degrees of freedom. For example, Refs. [49–52] built upon the results of Ref. [47] to compute momentum-space entanglement generated in scattering (with and without spin entanglement). In this paper, we are interested in the entanglement generated between scalar fields with an internal quantum number, e.g., flavor; we will address the generalization to fermions or vectors in future work. We thus compute the traces of operators $\hat{\mathcal{O}}$ as $\mathrm{Tr}_B[\hat{\mathcal{O}}] = \sum_c \int_u \langle u, c|_B \hat{\mathcal{O}} |u, c\rangle_B$. We use the shorthand $\int_p \equiv \int d^3p/[(2\pi)^3 2E_p]$ throughout. The reduced density matrix in Eq. (1) of a scalar particle in $\mathcal{H}_A$ is computed by tracing out degrees of freedom in $\mathcal{H}_B$,

$$\hat{\rho}_A = \frac{1}{N} \sum_c \int_u \langle u, c|_B \hat{S}|\Omega\rangle\langle\Omega|\hat{S}^\dagger|u, c\rangle_B \,. \tag{3}$$

Here N is a factor that ensures that the incoming state (for which $\hat{S}$ is replaced by $\hat{\mathbb{1}}$) is properly normalized.

We now compute the linearized entanglement generated by two-to-two scalar scattering $\phi_i(k_1)\phi_j(k_2) \to \phi_m(p_1)\phi_n(p_2)$, induced by a perturbative S-matrix $\hat{S} = \hat{\mathbb{1}} + i\hat{T}$ but *without assuming unitarity*. The indices i, j, m, n denote internal quantum numbers and p/k are outgoing/incoming momentum. This computation holds both for the case where scattering proceeds through a four-scalar contact term or where the scattering in generated by a mediating particle.

As stated above we will *not* assume unitarity. Our assumption is simply that there exists an S-matrix, which we will take to be a *state-independent general linear operator* acting on the Hilbert space. We will *not* assume a priori that $\hat{S}$ preserves normalizations, i.e., we will not use $\hat{S}^\dagger\hat{S} = \hat{\mathbb{1}}$.[2] In this way, failure of $\hat{S}$ to square to unity can be thought of as modeling an open quantum system, in which the initial state evolves into some inaccessible degrees of freedom. However, we need not invoke this picture and can instead consider a closed quantum system of fixed Hilbert space with a (not necessarily unitary) S-matrix. We will assume that $\hat{S}$ can, at worst, decrease the normalization of a state, so that the linearized entropy (1) is by definition positive. Perturbative scattering amplitudes are defined by acting on state vectors with the transition matrix $\hat{T}$, i.e., $\langle p_1, m; p_2, n|\hat{T}|k_1, a; k_2, b\rangle = (2\pi)^4\delta^4(k_i - p_f)\,\mathcal{A}^{mn}_{ab}(k_1k_2 \to p_1p_2)$. Likewise, we define the Hermitian conjugate as $\langle p_1, m; p_2, n|\hat{T}^\dagger|k_1, a; k_2, b\rangle = (2\pi)^4\delta^4(k_i - p_f)\,\mathcal{A}^{\dagger mn}_{ab}(k_1k_2 \to p_1p_2)$. While momentum conservation is enforced through the delta functions in the definitions above, *we do not demand conservation of the internal symmetry*. Note that in defining $\mathcal{A}^\dagger$ we have simply taken the Hermitian conjugate of the S-matrix element. We need not assume, e.g., that $\mathcal{A}^\dagger$ is the CPT conjugate of $\mathcal{A}$, nor any other assumption that would require unitarity. Furthermore, since we have not assumed unitarity, we also do not a priori have the optical theorem, and so the sign of $\mathrm{Im}\,\mathcal{A}$ is as yet unfixed. The two-particle incoming pure state $|\Omega\rangle$ is defined below Eq. (1). The internal quantum numbers of the incoming states live in $\mathcal{H}_A \otimes \mathcal{H}_B$, which we take as discrete, e.g., $\mathbb{C}^n \otimes \mathbb{C}^n$, so that we can use the above expressions to compute the entanglement generated in scattering.

[1] The linearized entropy equals the von Neumann entropy at leading order, linearized around a maximally mixed state. The quantity $\gamma \equiv \mathrm{Tr}_A\left[\hat{\rho}_A^2\right]$ is known as the purity and ranges over $1/d \leq \gamma \leq 1$ where d is the dimension of $\mathcal{H}_A$. The case $\gamma = 1$ corresponds to a pure state and $\gamma = 1/d$ to a maximally mixed state.

[2] Here, $\hat{S}^\dagger$ is defined as simply the Hermitian conjugate of $\hat{S}$, i.e., where $\hat{S}$ acts with left multiplication on vectors $|\psi\rangle$ in Hilbert space, $\hat{S}^\dagger$ acts with right multiplication on covectors $\langle\psi|$.

We compute the entanglement for a general S-matrix in Eq. (1) by inserting into Eq. (3) the identity $\hat{\mathbb{1}} = \sum_n \hat{\mathbb{1}}_n$, where $\hat{\mathbb{1}}_n = (\prod_{i=1}^n \int_{q_i}) |\{q_i\}\rangle\langle\{q_i\}|$ is the n-particle identity operator, and the sum is over all n-particle intermediate states, $\{q_i\} \equiv \{q_1, q_2, \ldots, q_n\}$. One-particle states are normalized as $\langle k, a|l, b\rangle = (2\pi)^3 2E_k \delta^3(k-l)\delta_{ab}$ following the conventions in Ref. [52]. The linearized entanglement in Eq. (1) may then be compactly written as

$$\mathcal{E}[\Omega] = 1 - \frac{1}{N^2} \sum_{ab} \int_{q_1} \int_{q_2} \mathcal{F}(q_1, q_2)_{ab} \mathcal{F}(q_2, q_1)_{ba} \quad (4)$$

in terms of a scalar function,

$$\mathcal{F}(q_1, q_2)_{ab} \equiv \sum_m \int_l \langle q_1, a; l, m|\hat{S}|\Omega\rangle\langle\Omega|\hat{S}^\dagger|q_2, b; l, m\rangle \,. \quad (5)$$

Again, we emphasize that the $\hat{S}^\dagger$ appearing above does not arise from any assumption of unitarity, but instead simply from the definition of bras, kets, and the inner product.

After various manipulations, we find the leading-order change in entanglement of the two scalars generated by scattering:

$$\begin{aligned} \Delta\mathcal{E}[\Omega] = 4 \left(\frac{1}{2E_{k_1} 2E_{k_2}} \frac{T}{V} \right) \times \\ \times \operatorname{Im}\Big[\sum_{abij} \Omega_{ij} \mathcal{A}^{ij}_{ab}(k_1 k_2 \to k_1 k_2) (\Omega^\dagger \cdot \Omega \cdot \Omega^\dagger)_{ab} \Big] \,, \end{aligned} \quad (6)$$

through $O(g^2)$, where g is a small coupling in the perturbative expansion $\mathcal{A} \sim g^2$. Here $\Delta\mathcal{E} = \mathcal{E}_f - \mathcal{E}_i$ and $\mathcal{E}_i[\Omega] = 1 - \operatorname{Tr}\left[(\Omega^\dagger\Omega)^2\right]$, which vanishes for a pure initial state. Divergences in spacetime volume are defined as $V \equiv (2\pi)^3 \delta^3_V(0)$ and $T \equiv 2\pi\delta_T(0)$. The seemingly divergent prefactor in Eq. (6), which we henceforth denote as $\mathcal{N} \equiv \left(\frac{1}{2E_{k_1} 2E_{k_2}} \frac{T}{V}\right)$, is an artifact of using planes waves in our derivation and is shown to be equal to unity in the Supplemental Material with the aid of a wave-packet formulation. For properly normalized initial states Ω, with $\operatorname{Tr}\left[\Omega^\dagger\Omega\right] = 1$, we have $N = 2E_{k_1} 2E_{k_2} V^2$. Interestingly, the entanglement computation has automatically selected amplitudes with forward kinematics from the integral in Eq. (4). See the Supplemental Material for the full derivation, including details on how forward kinematics arise due to momentum conservation and the tracing out states in the $O(g^2)$ term.[3]

Amplitudes' Positivity from Entanglement.—Using our calculations from the previous section, we are now poised to demonstrate the primary result of this paper summarized in Eq. (2), i.e., that positivity of entanglement, computed without assuming unitarity, implies the positivity of elastic amplitudes.

We are interested in entanglement that is generated by the S-matrix, so as such we will consider a state where the two initial particles are unentangled, so that $\mathcal{E}_i = 0$. By Schmidt decomposition, this means that there exist unique vectors $|\alpha\rangle_A \in \mathcal{H}_A$ and $|\beta\rangle_B \in \mathcal{H}_B$—given by some sums over the basis vectors $|a\rangle_A$ and $|b\rangle_B$, respectively—such that $|\Omega^{\text{prod}}\rangle = |k_1, \alpha\rangle_A \otimes |k_2, \beta\rangle_B$ is a direct product state. In that case, $(\Omega \cdot \Omega^\dagger \cdot \Omega)_{ab}$ is simply Ω_{ab}, so from Eq. (6) the leading-order entanglement generated by scattering is

$$\Delta\mathcal{E}[\Omega^{\text{prod}}] = 4\mathcal{N} \operatorname{Im} \mathcal{A}^{\alpha\beta}_{\alpha\beta}(k_1 k_2 \to k_1 k_2) \,. \quad (7)$$

The amplitude on the right-hand side is kinematically forward, as well as elastic in the internal quantum numbers. Critically, $\Delta\mathcal{E}[\Omega^{\text{prod}}] \geq 0$ since $\mathcal{E}_i = 0$ and $\mathcal{E}_f \geq 0$, i.e., the entanglement is necessarily nonnegative. Thus, positivity of entanglement also implies positivity of the imaginary part of the forward amplitude (which need not have been positive a priori without unitarity, $\hat{S}^\dagger\hat{S} = \hat{\mathbb{1}}$). If we were to invoke the optical theorem at this point, we would have $\operatorname{Im} \mathcal{A}^{\alpha\beta}_{\alpha\beta}(k_1 k_2 \to k_1 k_2) = 2E_{\text{cm}} p_{\text{cm}} \sigma_{\alpha\beta}$, where $\sigma_{\alpha\beta}$ is the cross section for $|\alpha\rangle_A \otimes |\beta\rangle_B \to$ anything and which as a physical area is by definition positive, E_{cm} is the energy of the two-particle initial state in the center-of-mass frame, and $p_{\text{cm}} = v_1 E_{k_1} = v_2 E_{k_2}$ is the momentum of either of the initial particles. Thus, the optical theorem implies positivity of the entanglement for elastic scattering,

$$\Delta\mathcal{E}[\Omega^{\text{prod}}] = 2(T/V)(v_1 + v_2)\sigma_{\alpha\beta} \geq 0 \,. \quad (8)$$

Note that for this product initial states, this leading-order tree-level entanglement is only nonvanishing when it has support on the resonance (as shown in the Supplemental Material). However, this conclusion generalizes to all weakly coupled theories, as one can view any loop-level process as an infinite sum over trees, as shown in App. B of Ref. [10]. We leave a detailed study of higher-order contributions to future work.

These conclusions generalize to the quantum Tsallis entropy $\mathcal{E}_n = \frac{1}{n-1}(1 - \operatorname{Tr}_A[\hat{\rho}_A^n])$, for which the linearized entanglement is the $n=2$ case. As above, $\mathcal{E}_n$ is nonnegative and vanishes for a pure state.[4] For an initial product state, as shown in the Supplemental Material, we have $\Delta\mathcal{E}_n[\Omega^{\text{prod}}] = \frac{2n}{n-1}\mathcal{N} \operatorname{Im} \mathcal{A}^{\alpha\beta}_{\alpha\beta}(k_1 k_2 \to k_1 k_2)$ for all integers $n \geq 2$.

[3] A related computation carried out in Ref. [45] did not fix forward kinematics, precisely because their derivation involved evoking the optical theorem throughout, setting $i(\hat{T} - \hat{T}^\dagger)$ to $-\hat{T}\hat{T}^\dagger$.

[4] The Tsallis entropy is bijectively related to the nth Rényi entropy, $S_n(\hat{\rho}) = \frac{1}{1-n} \log \operatorname{Tr}[\hat{\rho}^n]$, and both reduce to the von Neumann entropy in the limit $n \to 1$.

Thus, we have shown that positivity of the linear entropy and positivity of the amplitudes are *equivalent*, as advertised in Eq. (2). Both are manifestations of unitarity, but in a priori very different contexts, namely dispersion relations and consistency of quantum information. From this perspective, all positivity bounds on EFTs—including their many consequences including applications to quantum gravity, the EFT-hedron [10], etc.—are all quantum information theoretic statements, demanding consistency of the entanglement produced among the decay products.

Disentanglers.—For entangled initial states, $\Delta\mathcal{E}$ in Eq. (6) need not always be positive. Such "disentangler states" occasionally arise when the S-matrix itself acts as a "Maxwell's demon" or "disentangler," decreasing the entanglement of the final state relative to the initial one. Similar states were explored in Ref. [44], and their existence can be anticipated simply from the fact that in any quantum theory the action of the S-matrix on a state may be reversed through acting with $\hat{S}^\dagger$; such objects have found extensive use in tensor network contexts [53, 54].

We now demonstrate the existence of such disentangler states and highlight their role in hindering us from deriving the *generalized* optical theorem from entanglement.[5] Recall that the generalized optical theorem is the statement that $\mathcal{A}^{ij}_{ab}$ is a positive definite matrix, a property that is preserved under unitary transformations.

Let us Schmidt-decompose the initial state,

$$|\Omega\rangle = \sum_k \lambda_k |\alpha_k\rangle \otimes |\beta_k\rangle, \;\; \text{i.e.,} \;\; \Omega_{ij} = \sum_k \lambda_k \alpha_i^{(k)} \beta_j^{(k)} \,, \quad (9)$$

for some real $\lambda_k \geq 0$ and orthonormal vectors $|\alpha_k\rangle$ and $|\beta_k\rangle$, so Eq. (6) now becomes

$$\Delta\mathcal{E}[\Omega] = 4\mathcal{N}\,\mathrm{Im}\left[\sum_{kk'} \lambda_k \alpha_j^{*(k)} \beta_i^{*(k)} \mathcal{A}^{ij}_{ab} \lambda^3_{k'} \alpha_a^{(k')} \beta_b^{(k')}\right]. \quad (10)$$

Note that without internal flavor indices, we would have $\Delta\mathcal{E} \propto \mathrm{Im}\mathcal{A}(k_1 k_2 \to k_1 k_2) \propto \sigma \geq 0$, in which case disentangler states would not arise. However, in the presence of the internal degrees of freedom considered here, this is no longer the case as we will now demonstrate.

Let us write the basis of $|i\rangle \otimes |j\rangle$ states as $|I\rangle$, each denoted by some vector v_I

$$\Delta\mathcal{E}[\Omega] = 4\mathcal{N}\,\mathrm{Im}\left[\sum_{k,k'} \lambda_k \lambda^3_{k'} v_I^{(k)*} \mathcal{A}^I_{I'} v_{I'}^{(k')}\right]. \quad (11)$$

Without loss of generality, we take each v_I to be real (a complex v_I can made real via a unitary transformation, which we absorb into $\mathcal{A}$). Thus, we can pull everything but the amplitude out of the imaginary part of Eq. (10), yielding

$$\Delta\mathcal{E}[\Omega] = 4\mathcal{N} \sum_{k,k'} \lambda_k \lambda^3_{k'} x^{(k)} \cdot x^{(k')} \,, \quad (12)$$

where we have used the generalized optical theorem to decompose $\mathrm{Im}\,\mathcal{A}^I_{I'} = M^{IJ} M_{JI'}$ for some matrices M and defined the vectors $x^{(k)} = M \cdot v^{(k)}$. The M matrices—which can be taken to be real [15, 21]—describe the cuts of the four-particle amplitudes, that is, the (real and imaginary parts of) amplitudes for a given two-particle state described by an element of the Schmidt basis to go to some exchanged state. By simple algebra and relabeling of the indices k, k', we have:

$$\Delta\mathcal{E}[\Omega] = \frac{\mathcal{N}}{4} \sum_{k,k'} \left[(\lambda_k + \lambda_{k'})^4 - (\lambda_k - \lambda_{k'})^4\right] x^{(k)} \cdot x^{(k')}. \quad (13)$$

Since the generalized optical theorem is both necessary and sufficient for a healthy perturbative ultraviolet completion, it is possible to design a completion with arbitrary $x^{(k)} \cdot x^{(k')}$. One can think of $x^{(k)} \cdot x^{(k')}$ as defining a Riemannian metric $g_{kk'}$, i.e., $g_{kk'}$ is positive definite. For an *un*-entangled initial state, the Schmidt decomposition has only one nonzero value of λ, so that $\Delta\mathcal{E}$ is manifestly nonnegative since all the $\lambda_k - \lambda_{k'}$ differences vanish. However, for general $|\Omega\rangle$, we now demonstrate that disentangler states for which $\Delta\mathcal{E}[\Omega] < 0$ can arise.

As an illustration that $\Delta\mathcal{E}$ can be negative, consider for simplicity a scenario with only two nonzero λ_k values

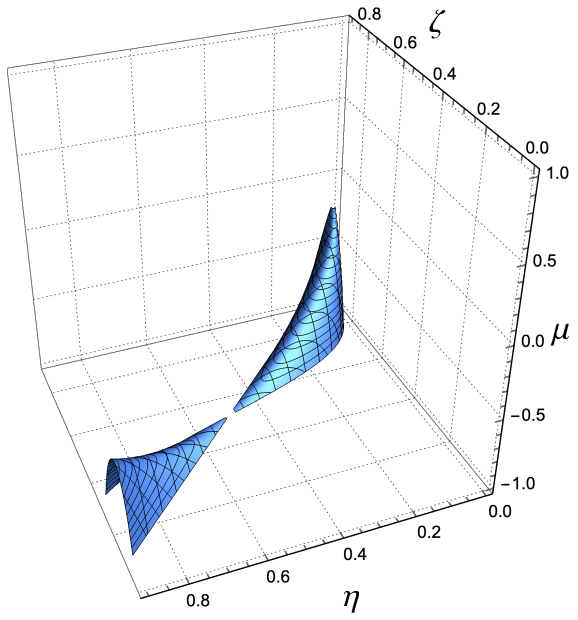


FIG. 1. The blue surface and volume below it corresponds to the parameter space described in text for a disentangler state, for which $\Delta\mathcal{E} < 0$.

[5] Understanding the connection between entanglement and the generalized optical theorem is particularly interesting in light of recent positivity bounds that make use of it [15, 16, 21, 55].

in the Schmidt basis. For example, each particle could be a qubit, leaving us with a four-dimensional joint Hilbert space, and we can choose to scatter an incoming Bell state, e.g., $(|1\rangle|1\rangle+|2\rangle|2\rangle)/\sqrt{2}$ or $(|1\rangle|2\rangle-|2\rangle|1\rangle)/\sqrt{2}$. Let us define $\xi = \lambda_2/\lambda_1 > 0$ and write $x^{(1)} = (\cos\theta_1, \sin\theta_1)$ as a general unit vector and $x^{(2)} = \nu(\cos\theta_2, \sin\theta_2)$ for some $\nu > 0$. (Roughly speaking, ν describes the ratio in the overall size of the couplings of the two elements of the Schmidt basis to the ultraviolet states.) Writing $\cos(\theta_1 - \theta_2) = \mu$, $\xi = \zeta/(1-\zeta)$, and $\nu = \eta/(1-\eta)$, the parameter space is given by $\mu \in [-1, 1]$, $\zeta \in [0, 1]$, and $\eta \in [0, 1]$. Note that $\zeta = 1$ or 0 corresponds to a pure state and $\zeta = 1/2$ to a maximally mixed state. We indeed find that a fraction of the parameter space leads to disentanglers, as shown by the region on and under the blue surface in Fig. 1. Note that disentanglers only appear when $\mu < 0$, corresponding to the physical picture where vectors x^1 and x^2 are pointing away from each other such that the element $g_{12} < 0$ is possible.

We observe in passing that if $x^{(k)} \cdot x^{(k')}$ is diagonal, equal to $\gamma_k \delta_{kk'}$ for $\gamma_k > 0$, then from Eq. (13) $\Delta\mathcal{E}[\Omega] = 4\mathcal{N}\sum_k \gamma_k \lambda_k^4 \geq 0$ for any λ_k. This situation corresponds to the case where there is no mixing among the exchanged states to which each two-particle state in the Schmidt basis couples, that is, the scattering describes a sum over disjoint superselection sectors.

Moreover, for projector states,[6] where $\lambda_k = c$ for $k \leq n$ for some n and $\lambda_k = 0$ for $k > n$, $\Delta\mathcal{E}$ is positive for all S-matrices, since $\Delta\mathcal{E} = 2\mathcal{N}c^4|X^{(n)}|^2 \geq 0$, defining the vector $X^{(n)} \equiv \sum_{k=1}^n x^{(k)}$. The converse is also true: if λ_k does not take the projector form, it is a disentangler state for some S-matrix. Without loss of generality, for λ_k not a projector, there exist $c_1 > c_2$ for which we can write $\lambda_k = c_1$ for $k \leq n$ for some n and $\lambda_{k+1} = c_2$. We pick the S-matrix such that $x^{(k)} = 0$ for $k > n+1$ and $X^{(n)} \cdot x^{(n+1)} = -\sqrt{|X^{(n)}|^2 |x^{(n+1)}|^2}$, as permitted by Cauchy's inequality. Defining $q = \sqrt{|x^{(n+1)}|^2/|X^{(n)}|^2}$, we have $\Delta\mathcal{E}/|X^{(n)}|^2 = 2(c_1 - qc_2)(c_1^3 - qc_2^3)$. Since $c_1 > c_2 > 0$, it is always possible to choose q such that $c_1/c_2 < q < c_1^3/c_2^3$, in which case $\Delta\mathcal{E} < 0$. This construction completes the proof that the projector form for the Schmidt coefficients is both necessary and sufficient to prevent disentangler states—for which scattering acts as a Maxwell's demon—for arbitrary S-matrices.

Discussion and Future Directions.—In this work we have shown that amplitudes' positivity—a consequence of unitarity in the form of the optical theorem—implies positivity of entanglement, and vice versa. Could there be other consequences of unitary that are related to entanglement? Could entanglement be used as a "guiding principle" of sorts for effective field theories? This Letter paves the way to such explorations.

As discussed above, constraints on Wilson coefficients of EFT operators are traditionally derived from unitarity. Positivity of entanglement can be used to replicate (and perhaps even go beyond) such constraints. Consider as a simple example collinear two-to-two photon scattering arising from the EFT operator $\frac{C}{M_\Phi^2}(F_{\mu\nu}F^{\mu\nu})^2$, which can be generated via a heavy scalar Φ coupled to $F_{\mu\nu}F^{\mu\nu}$. In Ref. [6], the bound on the Wilson coefficient was shown to be simply $C > 0$. In the forward limit the photon has only transverse modes and behaves as a complex scalar. As shown in the Supplemental Material, positivity of entanglement

$$\Delta\mathcal{E}_{\gamma\gamma}[\Omega^{\text{prod}}] \propto Cs^2\delta(s - M_\Phi^2) \tag{14}$$

implies the positivity bound. We leave it to future work to extend such a study beyond this simplest of cases. For instance, one could consider a theory with two scalars, allowing for a CP phase, as well scattering through other operators considered in Ref. [11]. Such studies could also lead to new insights for the emergence of symmetries, in a similar spirit as Refs. [39, 40]. It would also be interesting to explore more generally the entanglement structure of theories with decays (where the leading-order entanglement will go as $\mathcal{A}_{1\to2}\mathcal{A}_{1\to2}^\dagger$), mass mixing, and particle-antiparticle oscillations. Another avenue for future work is to generalize the results of this letter to theories with fermions or vectors where the interplay between spin, momentum, and internal degrees of freedom can be studied. Finally, extending these results beyond leading order in perturbation theory could perhaps allow us to derive the generalized optical theorem.

[6] The result about projector states and Maxwell's demon was also found in Ref. [45] in the context of a product initial two-particle state that is not necessarily pure.

ACKNOWLEDGMENTS

We thank Cliff Cheung, Ian Low, and Zhewei Yin for comments. This work was supported in part by the European Union's Horizon research and innovation programme under the Marie Sklodowska-Curie grant agreements No. 860881-HIDDeN and No. 101086085-ASYMMETRY. R.A. was funded by the STFC grant "Particle Physics at the Higgs Centre." The research of G.E. is supported by the National Science Foundation (NSF) Grant Number PHY-2210562, by a grant from University of Texas at Austin, and by a grant from the Simons Foundation. G.N.R. is supported by the James Arthur Postdoctoral Fellowship at New York University. G.E. thanks the Aspen Center for Physics (supported by NSF grant PHY-2210452) for their hospitality while this work was in progress.

[1] M. Froissart, "Asymptotic behavior and subtractions in the Mandelstam representation," *Physical Review* **123** (1961) 1053.

[2] A. Martin, "Unitarity and high-energy behavior of scattering amplitudes," *Phys. Rev.* **129** (1963) 1432.

[3] L. Lukaszuk and A. Martin, "Absolute upper bounds for pi pi scattering," *Nuovo Cim. A* **52** (1967) 122.

[4] N. Arkani-Hamed, L. Rodina, and J. Trnka, "Locality and Unitarity of Scattering Amplitudes from Singularities and Gauge Invariance," *Phys. Rev. Lett.* **120** (2018) 231602, arXiv:1612.02797 [hep-th].

[5] N. Arkani-Hamed, H. Frost, G. Salvatori, P.-G. Plamondon, and H. Thomas, "All Loop Scattering as a Counting Problem," arXiv:2309.15913 [hep-th].

[6] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion," *JHEP* **10** (2006) 014, arXiv:hep-th/0602178.

[7] A. Jenkins and D. O'Connell, "The Story of $\mathcal{O}$: Positivity constraints in effective field theories," arXiv:hep-th/0609159.

[8] A. Nicolis, R. Rattazzi, and E. Trincherini, "Energy's and amplitudes' positivity," *JHEP* **05** (2010) 095, arXiv:0912.4258 [hep-th]. [Erratum: *JHEP* **11** (2011) 128].

[9] G. Dvali, A. Franca, and C. Gomez, "Road Signs for UV-Completion," arXiv:1204.6388 [hep-th].

[10] N. Arkani-Hamed, T.-C. Huang, and Y.-t. Huang, "The EFT-Hedron," *JHEP* **05** (2021) 259, arXiv:2012.15849 [hep-th].

[11] G. N. Remmen and N. L. Rodd, "Consistency of the Standard Model Effective Field Theory," *JHEP* **12** (2019) 032, arXiv:1908.09845 [hep-ph].

[12] G. N. Remmen and N. L. Rodd, "Flavor Constraints from Unitarity and Analyticity," *Phys. Rev. Lett.* **125** (2020) 081601, arXiv:2004.02885 [hep-ph]. [Erratum: *Phys. Rev. Lett.* **127** (2021) 149901].

[13] G. N. Remmen and N. L. Rodd, "Signs, spin, SMEFT: Sum rules at dimension six," *Phys. Rev. D* **105** (2022) 036006, arXiv:2010.04723 [hep-ph].

[14] G. N. Remmen and N. L. Rodd, "Spinning sum rules for the dimension-six SMEFT," *JHEP* **09** (2022) 030, arXiv:2206.13524 [hep-ph].

[15] C. Zhang and S.-Y. Zhou, "Convex Geometry Perspective on the (Standard Model) Effective Field Theory Space," *Phys. Rev. Lett.* **125** (2020) 201601, arXiv:2005.03047 [hep-ph].

[16] T. Trott, "Causality, unitarity and symmetry in effective field theory," *JHEP* **07** (2021) 143, arXiv:2011.10058 [hep-ph].

[17] B. Bellazzini, C. Cheung, and G. N. Remmen, "Quantum Gravity Constraints from Unitarity and Analyticity," *Phys. Rev. D* **93** (2016) 064076, arXiv:1509.00851 [hep-th].

[18] C. Cheung and G. N. Remmen, "Positivity of Curvature-Squared Corrections in Gravity," *Phys. Rev. Lett.* **118** (2017) 051601, arXiv:1608.02942 [hep-th].

[19] C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," *JHEP* **12** (2014) 087, arXiv:1407.7865 [hep-th].

[20] Y. Hamada, T. Noumi, and G. Shiu, "Weak Gravity Conjecture from Unitarity and Causality," *Phys. Rev. Lett.* **123** (2019) 051601, arXiv:1810.03637 [hep-th].

[21] N. Arkani-Hamed, Y.-t. Huang, J.-Y. Liu, and G. N. Remmen, "Causality, unitarity, and the weak gravity conjecture," *JHEP* **03** (2022) 083, arXiv:2109.13937 [hep-th].

[22] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum mechanical description of physical reality be considered complete?," *Phys. Rev.* **47** (1935) 777.

[23] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics Physique Fizika* **1** (1964) 195.

[24] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.* **23** (1969) 880.

[25] **ATLAS Collaboration**, G. Aad et al., "Observation of quantum entanglement in top-quark pairs using the ATLAS detector," arXiv:2311.07288 [hep-ex].

[26] Y. Afik and J. R. M. n. de Nova, "Entanglement and quantum tomography with top quarks at the LHC," *Eur. Phys. J. Plus* **136** (2021) 907, arXiv:2003.02280 [quant-ph].

[27] A. J. Barr, "Testing Bell inequalities in Higgs boson decays," *Phys. Lett. B* **825** (2022) 136866, arXiv:2106.01377 [hep-ph].

[28] C. Severi, C. D. E. Boschi, F. Maltoni, and M. Sioli, "Quantum tops at the LHC: from entanglement to Bell inequalities," *Eur. Phys. J. C* **82** (2022) 285, arXiv:2110.10112 [hep-ph].

[29] J. A. Aguilar-Saavedra and J. A. Casas, "Improved tests of entanglement and Bell inequalities with LHC tops," *Eur. Phys. J. C* **82** (2022) 666, arXiv:2205.00542 [hep-ph].

[30] R. Ashby-Pickering, A. J. Barr, and A. Wierzchucka, "Quantum state tomography, entanglement detection and Bell violation prospects in weak decays of massive particles," *JHEP* **05** (2023) 020, arXiv:2209.13990 [quant-ph].

[31] R. Aoude, E. Madge, F. Maltoni, and L. Mantani, "Quantum SMEFT tomography: Top quark pair production at the LHC," *Phys. Rev. D* **106** (2022) 055007, arXiv:2203.05619 [hep-ph].

[32] M. Fabbrichesi, R. Floreanini, and E. Gabrielli, "Constraining new physics in entangled two-qubit systems: top-quark, tau-lepton and photon pairs," *Eur. Phys. J. C* **83** (2023) 162, arXiv:2208.11723 [hep-ph].

[33] C. Severi and E. Vryonidou, "Quantum entanglement and top spin correlations in SMEFT at higher orders," *JHEP* **01** (2023) 148, arXiv:2210.09330 [hep-ph].

[34] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[35] G. Lindblad, "Expectations and entropy inequalities for finite quantum systems," *Commun.Math. Phys.* **39** (1974) 111.

[36] P. Hayden, M. Headrick, and A. Maloney, "Holographic Mutual Information is Monogamous," *Phys. Rev. D* **87** (2013) 046003, arXiv:1107.2940 [hep-th].

[37] S. Dutta and T. Faulkner, "A canonical purification for the entanglement wedge cross-section," *JHEP* **03** (2021) 178, arXiv:1905.00577 [hep-th].

[38] N. Bao, A. Chatwin-Davies, and G. N. Remmen, "Entanglement wedge cross section inequalities from replicated geometries," *JHEP* **07** (2021) 113,

arXiv:2106.02640 [hep-th].

[39] S. R. Beane, D. B. Kaplan, N. Klco, and M. J. Savage, "Entanglement Suppression and Emergent Symmetries of Strong Interactions," *Phys. Rev. Lett.* **122** (2019) 102001, arXiv:1812.03138 [nucl-th].

[40] I. Low and T. Mehen, "Symmetry from entanglement suppression," *Phys. Rev. D* **104** (2021) 074014, arXiv:2104.10835 [hep-th].

[41] R. Aoude, M.-Z. Chung, Y.-t. Huang, C. S. Machado, and M.-K. Tam, "Silence of Binary Kerr Black Holes," *Phys. Rev. Lett.* **125** (2020) 181602, arXiv:2007.09486 [hep-th].

[42] B.-T. Chen, M.-Z. Chung, Y.-t. Huang, and M. K. Tam, "Minimal spin deflection of Kerr-Newman and supersymmetric black hole," *JHEP* **10** (2021) 011, arXiv:2106.12518 [hep-th].

[43] M. Carena, I. Low, C. E. M. Wagner, and M.-L. Xiao, "Entanglement Suppression, Enhanced Symmetry and a Standard-Model-like Higgs Boson," arXiv:2307.08112 [hep-ph].

[44] C. Cheung and G. N. Remmen, "Entanglement and the double copy," *JHEP* **05** (2020) 100, arXiv:2002.10470 [hep-th].

[45] C. Cheung, T. He, and A. Sivaramakrishnan, "Entropy growth in perturbative scattering," *Phys. Rev. D* **108** (2023) 045013, arXiv:2304.13052 [hep-th].

[46] V. Rosenhaus and M. Smolkin, "Entanglement Entropy: A Perturbative Calculation," *JHEP* **12** (2014) 179, arXiv:1403.3733 [hep-th].

[47] V. Balasubramanian, M. B. McDermott, and M. Van Raamsdonk, "Momentum-space entanglement and renormalization in quantum field theory," *Phys. Rev. D* **86** (2012) 045014, arXiv:1108.3568 [hep-th].

[48] M. H. M. Costa, J. v. d. Brink, F. S. Nogueira, and G. a. I. Krein, "Momentum space entanglement from the Wilsonian effective action," *Phys. Rev. D* **106** (2022) 065024, arXiv:2207.12103 [hep-th].

[49] S. Seki, I. Y. Park, and S.-J. Sin, "Variation of Entanglement Entropy in Scattering Process," *Phys. Lett. B* **743** (2015) 147, arXiv:1412.7894 [hep-th].

[50] R. Peschanski and S. Seki, "Entanglement Entropy of Scattering Particles," *Phys. Lett. B* **758** (2016) 89, arXiv:1602.00720 [hep-th].

[51] R. Peschanski and S. Seki, "Evaluation of Entanglement Entropy in High Energy Elastic Scattering," *Phys. Rev. D* **100** (2019) 076012, arXiv:1906.09696 [hep-th].

[52] J. Fan, Y. Deng, and Y.-C. Huang, "Variation of entanglement entropy and mutual information in fermion-fermion scattering," *Phys. Rev. D* **95** (2017) 065017, arXiv:1703.07911 [hep-th].

[53] G. Evenbly and G. Vidal, "Algorithms for entanglement renormalization," *Phys. Rev. B* **79** (2009) 144108, arXiv:0707.1454 [cond-mat.str-el].

[54] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency conditions for an AdS multiscale entanglement renormalization ansatz correspondence," *Phys. Rev. D* **91** (2015) 125036, arXiv:1504.06632 [hep-th].

[55] M. Freytsis, S. Kumar, G. N. Remmen, and N. L. Rodd, "Multifield positivity bounds for inflation," *JHEP* **09** (2023) 041, arXiv:2210.10791 [hep-th].

[56] M. Blasone, F. Dell'Anno, S. De Siena, and F. Illuminati, "Entanglement in neutrino oscillations," *EPL* **85** (2009) 50002, arXiv:0707.4476 [hep-ph].

[57] E. K. Akhmedov and A. Y. Smirnov, "Neutrino oscillations: Entanglement, energy-momentum conservation and QFT," *Found. Phys.* **41** (2011) 1279, arXiv:1008.2077 [hep-ph].

[58] B. Bellazzini, "Softness and amplitudes' positivity for spinning particles," *JHEP* **02** (2017) 034, arXiv:1605.06111 [hep-th].

[59] D. A. Kosower, B. Maybee, and D. O'Connell, "Amplitudes, Observables, and Classical Scattering," *JHEP* **02** (2019) 137, arXiv:1811.10950 [hep-th].

Supplemental Material for Positivity in Amplitudes from Quantum Entanglement

Rafael Aoude, Gilly Elor, Grant N. Remmen, and Olcyr Sumensari

In this Supplemental Material, we first provide a detailed derivation of $\Delta\mathcal{E}[\Omega]$ in Eq. (6) of the main text of this Letter. We begin by deriving an expression for the initial state entanglement, followed by a derivation of Eq. (6) generated by both a quartic contact term and by exchange of a heavy mediator. Critically, this exercise demonstrates the validly of our result to leading order in perturbation theory. We will also see the details of how forward kinematics are automatically selected. We include a computation of the next-highest order term ($O(g^4)$) in the perturbation expansion. We conclude with a section providing an alternative derivation of $\Delta\mathcal{E}[\Omega]$ using a wave function formulation. This approach will also provide justification for setting the seemingly divergent prefactor $\mathcal{N}$ in Eq. (6) to one. Finally we include some details on extending these results to a more general case considering the Tsallis entropy.

$\Delta\mathcal{E}[\Omega]$ Derivation

We now present details of the computation of $\Delta\mathcal{E}[\Omega]$, the linearized entanglement in Eq. (6) generated in $\phi_i(k_1)\phi_j(k_2) \to \phi_k(p_1)\phi_l(p_2)$ scattering. First, we will compute the entanglement for a general initial state $|\Omega\rangle$ using the formalism discussed in this paper. We then compute the leading-order entanglement generated by the final state through the action of a perturbative S-matrix. We do this first in the case where the scattering proceeds through a four-point scalar contact term. We then repeat this computation in the case where the scattering arises through the exchange of a heavy particle. In both cases, we find that Eq. (6) describes the leading-order linearized entanglement generated through the scattering, as must have been the case. We present our derivation in terms of a toy model, but the result holds for, e.g., photon scattering $\gamma\gamma \to \Phi \to \gamma\gamma$ mediated by a heavy scalar Φ, where transverse photon helicities can be treated as an internal symmetry in the forward limit.

Throughout this derivation, our starting point will be a general two-particle incoming state,

$$|\Omega\rangle = \sum_{ij} \Omega_{ij} |k_1, i\rangle_A \otimes |k_2, j\rangle_B \equiv \sum_{ij} \Omega_{ij} |k_1, i; k_2, j\rangle \,, \tag{S1}$$

where $\phi_i \in \mathcal{H}_A$ and $\phi_j \in \mathcal{H}_B$. Initial states are normalized such that $\mathrm{Tr}[\Omega^\dagger \Omega] = 1$, as discussed in the main text. In the absence of a scalar mass mixing term, we may normalize the one-particle inner product as follows:

$$\langle p_1, i | p_2, j\rangle = (2\pi)^3 2E_{p_1} \delta_{ij} \delta^3(p_1 - p_2) \,. \tag{S2}$$

Throughout, as in the main text, we use the shorthand $\int_u \equiv \int d^3u/(2\pi)^3(2E_u)$, and define spacelike and timelike divergences as $V \equiv (2\pi)^3 \delta_V^3(0)$ and $T \equiv (2\pi)\delta_T(0)$.

Entanglement of the initial state

We begin with the straightforward derivation of the initial state entanglement in Eq. (S1). The full system density matrix of the two-particle initial state is constructed as follows:

$$\hat{\rho}^i_{AB} = |\Omega\rangle\langle\Omega| = \sum_{ijab} \Omega_{ij} \Omega^\dagger_{ab} |k_1, i\rangle_A \otimes |k_2, j\rangle_B \langle k_1, a|_A \otimes \langle k_2, b|_B \,. \tag{S3}$$

The reduced density matrix is computed by tracing over particle B, and using Eq. (S2) this yields

$$\begin{aligned} \mathrm{Tr}_B[\hat{\rho}_{AB}] &= \sum_{ijabc} \Omega_{ij} \Omega^\dagger_{ab} \int_u \Big(\langle u, c|_B |k_2, j\rangle_B \langle k_2, b|_B |u, c\rangle_B \Big) |k_1, i\rangle_A \langle k_1, a|_A \\ &= 2E_{k_2} V \sum_{ija} \Omega_{ij} \Omega^\dagger_{aj} |k_1, i\rangle_A \langle k_1, a|_A \,. \end{aligned} \tag{S4}$$

The density matrix must be properly normalized so that $\mathrm{Tr}_A \hat{\rho}_{nA} = 1$:

$$\begin{aligned} N = \mathrm{Tr}_A[\mathrm{Tr}_B \hat{\rho}^i_{AB}] &= 2E_{k_2} V \sum_{ijac} \int_c \Omega_{ij} \Omega^\dagger_{aj} \langle u, c|k_1, i\rangle \langle k_1, a|u, c\rangle \\ &= 2E_{k_2} 2E_{k_1} V^2 \mathrm{Tr}[\Omega^\dagger \Omega] = 2E_{k_2} 2E_{k_1} V^2 \,. \end{aligned} \tag{S5}$$

The linear entanglement is then

$$\begin{aligned}\mathcal{E}_i[\Omega] &= 1 - \frac{1}{N^2}\mathrm{Tr}_A[\hat{\rho}_A^2] \\ &= 1 - \frac{1}{(2E_{k_1})^2 V}\sum_c \sum_{ija}\sum_{i'j'a'} \Omega_{ij}\Omega^\dagger_{aj}\Omega_{i'j'}\Omega^\dagger_{a'j'}\int_u \langle u,c|k_1,i\rangle\langle k_1,a|k_1,i'\rangle\langle k_1,a'|u,c\rangle \\ &= 1 - \sum_{ija}\sum_{i'j'a'}\Omega_{ij}\Omega^\dagger_{aj}\Omega_{i'j'}\Omega^\dagger_{a'j'}\delta_{ai'}\delta_{a'i} = 1 - \sum_{ijj'a}\Omega_{ij}\Omega^\dagger_{aj}\Omega_{aj'}\Omega^\dagger_{ij'} = 1 - \mathrm{Tr}[(\Omega^\dagger\Omega)^2]\,.\end{aligned} \tag{S6}$$

Note that for a pure state $\mathrm{Tr}[(\Omega^\dagger\Omega)^2] = 1$ and the linear entanglement vanishes, $\mathcal{E}_i[\Omega^{\rm pure}] = 0$, as must be the case.

Entanglement derivation for a contact term

Consider four-scalar scattering through the following quartic contact term,

$$\mathcal{L}_{\rm contact} = \frac{g^2}{M_X^2} c_{ij}^{kl}\phi_i\phi_j\phi_k\phi_l\,. \tag{S7}$$

This can be generated in the UV through, for instance, the exchange of a heavy scalar X. The initial state (S1) and initial state entanglement (S5) are given above. We are now interested in the final state density matrix and entanglement, the definitions of which we repeat here for convenience:

$$\hat{\rho}^f_{AB} = \hat{S}|\Omega\rangle\langle\Omega|\hat{S}^\dagger\,, \quad \mathcal{E}^f[\Omega] = 1 - \frac{1}{N^2}\mathrm{Tr}_A[\mathrm{Tr}_B\hat{\rho}^f_{AB}]^2\,, \quad \mathrm{Tr}_B\hat{\rho}^f_{AB} = \frac{1}{N}\sum_c\int_u \langle u,c|_B\hat{S}|\Omega\rangle\langle\Omega|\hat{S}^\dagger|u,c\rangle_B\,. \tag{S8}$$

Once again, the normalization is fixed by the initial state. Vector products are given by Eq. (S2).

To evaluate Eq. (S8), a useful trick is to insert the identity operator on Fock space, which we expand out to work at lowest order in perturbation theory: $\hat{S}|\Omega\rangle = \hat{\mathbb{1}}_1\hat{S}|\Omega\rangle + \hat{\mathbb{1}}_2\hat{S}|\Omega\rangle + \cdots$, where $\hat{\mathbb{1}}_1$ denotes the identity on the subspace of one-particle states, $\hat{\mathbb{1}}_2$ on two-particle states, etc. (see Eq. (S18) below). At present, the $\hat{\mathbb{1}}_1$ term does not contribute and the leading-order sum over states becomes (where the q denote final momenta and i,j are internal flavor indices):

$$\begin{aligned}\mathbb{1}_2\hat{S}|\Omega\rangle &= \sum_{ijab}\int_{q_1}\int_{q_2}|q_1,i;q_2,j\rangle\langle q_1,i;q_2,j|(1+i\hat{T})\Omega_{ab}|k_1,a;k_2,b\rangle \\ &= |\Omega\rangle + i2\pi\sum_{ijab}\int_{q_2}\frac{1}{2E_{q_1}}\Omega_{ab}\delta(E_{k_1}+E_{k_2}-E_{q_1}-E_{q_2})\mathcal{A}^{ab}_{ij}(k_1k_2\to q_1q_2)|q_1,i;q_2,j\rangle\Big|_{q_1=q_2-k_1-k_2}\,,\end{aligned} \tag{S9}$$

and similarly for the conjugate $\langle\Omega|\hat{S}^\dagger\mathbb{1}_2$. Plugging into Eq. (S8), we compute the reduced density matrix:

$$\begin{aligned}N\hat{\rho}_A &= \sum_{jaa'}\int\frac{d^3u}{2E_u}(2E_{k_2})^2\Omega_{ac}\Omega^\dagger_{a'j}(2\pi)^3(\delta^3(k_2-u))^2|k_1,a\rangle_A\langle k_1,a'|_A \\ &- \left[i\pi\sum_{iaa'bj}\int\frac{d^3u}{2E_uE_{q_1}}2E_{k_2}\delta^3(k_2-u)\Omega_{a'j}\Omega^\dagger_{ab}\delta(E_{k_1}+E_{k_2}-E_{q_1}-E_u)\mathcal{A}^{ab\dagger}_{ij}(q_1u\to k_1k_2)|k_1,a'\rangle_A\langle q_1,i|_A + \text{h.c.}\right]_{\substack{q_1=\\k_1+k_2-u}} \\ &+ \sum_{iaa'bb'j}\int\frac{d^3u\,\pi^2\Omega_{ab}\Omega^\dagger_{a'b'}}{(2\pi)^3 2E_uE_{q_1}^2}(\delta(E_{k_1}+E_{k_2}-E_{q_1}-E_u))^2\mathcal{A}^{ab}_{ij}(k_1k_2\to q_1u)\mathcal{A}^{a'b'\,\dagger}_{ij}(q_1u\to k_1k_2)|q_1,i\rangle_A\langle q_1,i|_A\Big|_{\substack{q_1=\\k_1+k_2-u}}\,.\end{aligned} \tag{S10}$$

Critically, note that, at tree level, the fourth term is $O(g^4)$ as it arises from the product of two *four-point* amplitudes, while the second and third terms are $O(g^2)$. Computing the momentum integrals, we have to leading order,

$$\begin{aligned}N\,\hat{\rho}_A &= 2E_{k_2}V\sum_{jaa'}\Omega_{aj}\Omega^\dagger_{a'j}|k_1,a\rangle_A\langle k_1,a'|_A \\ &\quad - 2\left[i\frac{T}{2E_{k_1}}\sum_{iaa'bj}\Omega_{a'j}\Omega^\dagger_{ab}\mathcal{A}^{ab\dagger}_{ij}(k_1k_2\to k_1k_2)|k_1,a'\rangle_A\langle k_1,i|_A + \text{h.c.}\right] + O(g^4),\end{aligned} \tag{S11}$$

where we have used $(2\pi)^3\delta^3(0) \equiv V$ in the first line. We can now understand, pragmatically, how the kinematic limit arises; integrating over u momentum in Eq. (S10) selects $u = k_2$, so that $q_1 = k_1 + k_2 - u = k_1$. The energy delta function thus becomes $\delta(0)$ and we define $2\pi\delta(0) \equiv T$. The result in Eq. (S11) may now be squared and plugged into Eq. (S8) to compute the linearized entanglement to leading order $O(g^2)$, which after evaluating the integrals and summing of internal symmetry indices, becomes

$$\begin{aligned} N^2 \operatorname{Tr}_A\left[\hat{\rho}_A^2\right] &= (2E_{k_1}2E_{k_2}V^2)^2 \operatorname{Tr}[(\Omega^\dagger\cdot\Omega)^2] \\ &\quad - 4\,(2E_{k_2}2E_{k_1}V^3T)\operatorname{Im}\Big[\sum_{abij}\Omega_{ab}\mathcal{A}^{ab}_{ij}(k_1k_2\to k_1k_2)(\Omega^\dagger\cdot\Omega\cdot\Omega^\dagger)_{ij}\Big] + O(g^4)\,, \end{aligned} \tag{S12}$$

using $i(X - X^\dagger) = -2\,\mathrm{Im}X$. The resulting change in linearized entanglement generated by scattering to leading order becomes Eq. (6), namely,

$$\Delta\mathcal{E}_{\text{contact}}[\Omega] = 4\left(\frac{1}{2E_{k_1}2E_{k_2}}\frac{T}{V}\right)\operatorname{Im}\Big[\sum_{abij}\Omega_{ab}\mathcal{A}^{ab}_{ij}(k_1k_2\to k_1k_2)(\Omega^\dagger\cdot\Omega\cdot\Omega^\dagger)_{ij}\Big] + O(g^4)\,, \tag{S13}$$

where we have subtracted off the the initial state entanglement fond in Eq. (S6), i.e., $\mathcal{E}_i = 1 - \operatorname{Tr}\left[(\Omega^\dagger\Omega)^2\right]$. For a pure state $\Delta\mathcal{E}_{\text{contact}}[\Omega^{\text{pure}}] \propto \mathrm{Im}\mathcal{A} \propto \mathrm{Im}\left(g^2c/M_X^2\right) = 0$ since the coupling is real, and the leading-order contribution arises at $O(g^4)$, i.e., at one-loop order. However $\mathcal{E}_{\text{contact}}[\Omega]$ will generally be nonvanishing for an impure initial state:

$$\Delta\mathcal{E}_{\text{contact}}[\Omega] = 4\left(\frac{1}{2E_{k_1}2E_{k_2}}\frac{T}{V}\right)\frac{g^2}{M_X^2}\operatorname{Im}\Big[\sum_{abij}\Omega_{ab}c^{ab}_{ij}(\Omega^\dagger\cdot\Omega\cdot\Omega^\dagger)_{ij}\Big] + O(g^4)\,, \tag{S14}$$

where we have plugged in the amplitude from Eq. (S7). We could of course dress the contact term with derivatives, in which case we can view this result in an EFT context: at linear order in the Wilson coefficients, the entanglement generated by an EFT vanishes for a pure state, for scattering below the mass scale of the ultraviolet completion.

The $O(g^4)$ contribution to the entropy includes both the one-loop part of the amplitude $\mathcal{A}^{\text{1-loop}}$, as well as terms going like the square of the tree-level amplitude,

$$\begin{aligned} \Delta\mathcal{E}[\Omega]_{O(g^4)} &= 4\lambda\sum_{ijmn}\operatorname{Im}\big[\Omega_{ij}\,\mathcal{A}^{ij\,\text{1-loop}}_{mn}\,(\Omega^\dagger\cdot\Omega\cdot\Omega^\dagger)_{nm}\big] \\ &\quad - 2\lambda^2\sum_{ijmn}\sum_{i'j'm'n'}\Big\{\delta_{mm'}\,\Omega_{ij}(\Omega^\dagger)_{j'i'}\,\mathcal{A}^{ij\,\text{tree}}_{mn}\,(\Omega^\dagger\cdot\Omega)_{nn'}\mathcal{A}^{\dagger\,i'j'\,\text{tree}}_{m'n'} \\ &\qquad + \delta_{nn'}\,\Omega_{ij}(\Omega^\dagger)_{j'i'}(\Omega\cdot\Omega^\dagger)_{m'm}\,\mathcal{A}^{ij\,\text{tree}}_{mn}\mathcal{A}^{\dagger\,i'j'\,\text{tree}}_{m'n'} \\ &\qquad - \operatorname{Re}\Big[\Omega_{ij}\Omega_{i'j'}(\Omega^\dagger)_{n'm}\,\mathcal{A}^{ij\,\text{tree}}_{mn}\,(\Omega^\dagger)_{nm'}\mathcal{A}^{i'j'\,\text{tree}}_{m'n'}\Big]\Big\}\,, \end{aligned} \tag{S15}$$

where $\lambda = (2\pi)^4\delta^{(4)}(0)/\mathcal{N}$ and all amplitudes are in the forward limit.

Entanglement derivation for massive exchange

In the previous subsection, we computed the entanglement in two-to-two scalar scattering $\Delta\mathcal{E}[\Omega]$ that arises from a contact operator, namely Eq. (S7), and have shown that this is equivalent to Eq. (6). To verify the robustness of Eq. (6) we now compute $\Delta\mathcal{E}[\Omega]$ in the context of a theory with massive exchange, which reduces to Eq. (S7) in the low-energy limit:

$$\mathcal{L} = g f_{ijk} X_i\phi_j\phi_k\,. \tag{S16}$$

Here X is a heavy scalar, and the whole Hilbert space is assumed to be decomposed as $\mathcal{H} = \mathcal{H}_X \otimes \mathcal{H}_A \otimes \mathcal{H}_B$. Once again the initial state is given by

$$|\Omega\rangle = \sum_{ij}\Omega_{ij}|\phi_i(k_1)\rangle_A \otimes |\phi_j(k_2)\rangle_B \equiv \sum_{ij}\Omega_{ij}|\phi_i(k_1)\phi_j(k_2)\rangle\,, \tag{S17}$$

This is equivalent to Eq. (S3), but with a slight change of notation. We distinguish a state of the heavy scalar as $|X_m(Q)\rangle_X$. Since in the ultraviolet one can form on-shell X particles as intermediate states, we must sum over X when inserting the identity operator,

$$\hat{\mathbb{1}} = \sum_m \int_Q |X_m(Q)\rangle_X \langle X_m(Q)|_X + \sum_{ij} \int_{q_1} \int_{q_2} |\phi_i(q_1)\phi_j(q_2)\rangle\langle\phi_i(q_1)\phi_j(q_2)| + \cdots , \tag{S18}$$

which we write as $\hat{\mathbb{1}} = \hat{\mathbb{1}}_1 + \hat{\mathbb{1}}_2$. Since the $\hat{T}$ matrix will now generate $2 \leftrightarrow 1$ processes involving X, we define amplitudes generically now as

$$\langle p_f, a_f|\hat{T}|p_i, a_i\rangle = (2\pi)^4 \delta^4 (p_f - p_i)\, \mathcal{A}^{a_i}_{a_f} (p_i \to p_f) , \tag{S19}$$

where $p_{f/i}$ and $a_{f/i}$ stand for any number of final and initial state momenta (X or ϕ) and internal symmetry indices. Inner products between states are $\langle\phi_i(p_1)|_A|\phi_j(p_2)\rangle_A = (2\pi)^3\delta^3(p_1 - p_2)\delta_{ij}$ and $\langle X|\phi\rangle = 0$, since there is no X-ϕ mass mixing.[7]

We now compute the reduced density matrix by tracing out the ϕ in $\mathcal{H}_B$, $\hat{\rho}_A = \mathrm{Tr}_B \hat{\rho}_{AB}$, as follows:

$$\hat{\rho}_A = \frac{1}{N} \sum_c \int_u \langle\phi_c(u)|_B (\hat{\mathbb{1}}_1 \hat{S} + \hat{\mathbb{1}}_2 \hat{S})|\Omega\rangle\langle\Omega|(\hat{S}\hat{\mathbb{1}}_1^\dagger + \hat{S}^\dagger \mathbb{1}_2)|\phi_c(u)\rangle_B . \tag{S20}$$

The matrix elements $\langle u, c|_B \hat{\mathbb{1}}_2 \hat{S}|\Omega\rangle$ and its conjugate are essentially equivalent to those of the previous contact-term computation (of course with the amplitude itself now as given by the Lagrangian in this subsection). We note that the contribution from the internal X particle given by $\hat{\mathbb{1}}_1$ vanishes,

$$\langle\phi_c(u)|_B \hat{S}\hat{\mathbb{1}}_1|\Omega\rangle = 0 = \langle\phi_c(u)|_B \hat{\mathbb{1}}_1 \hat{S}|\Omega\rangle , \tag{S21}$$

since in this theory there is no mass mixing or other way to convert $\phi \leftrightarrow X$. This makes intuitive sense since because, *to leading order*, neither the final nor the initial state has overlap with $\mathcal{H}_X$. Thus we conclude that the density matrix is as we previously computed,

$$\Delta\mathcal{E}_{\mathrm{UV}}[\Omega] = 4 \left(\frac{1}{2E_{k_1} 2E_{k_2}} \frac{T}{V}\right) \mathrm{Im}\Big[\sum_{abij} \Omega_{ij} \mathcal{A}^{ij}_{ab}(k_1 k_2 \to k_1 k_2)(\Omega^\dagger \cdot \Omega \cdot \Omega^\dagger)_{ab}\Big] + O(g^4) . \tag{S22}$$

For a narrow X resonance, the imaginary part of the amplitude has support on the delta function $\delta(s - M_X^2)$.

The toy model considered above has a realization as collinear two-to-two photon scattering. A Lagrangian with a dilaton coupling $\mathcal{L} \supset \frac{g}{f}\Phi F_{\mu\nu}F^{\mu\nu}$, permits Φ-mediated light-by-light scattering, which at low energies generates the effective operator $\frac{C}{M_\Phi^2}(F_{\mu\nu}F^{\mu\nu})^2$, where $C = g^2/2f^2 > 0$; see, e.g., the model in Ref. [11]. In the forward limit, all polarizations are transverse to all momenta and thus behave like a complex internal quantum number [58], so the computations above for a scalar theory can be applied. For product initial states with polarizations ϵ_1 and ϵ_2, the leading-order entanglement becomes

$$\Delta\mathcal{E}_{\gamma\gamma}[\Omega^{prod}] \propto C\, s^2\, \mathrm{Im}\Big[\frac{(\epsilon_1 \cdot \epsilon_2)(\epsilon_1 \cdot \epsilon_2)^\star + (\epsilon_1 \cdot \epsilon_2^\star)(\epsilon_2 \cdot \epsilon_1^\star)}{s - M_\Phi^2 + i\epsilon}\Big] \propto C s^2 \delta(s - M_\Phi^2) . \tag{S23}$$

Wave Function Derivation

In the main text, we have encountered the factor of $\mathcal{N} = \left(\frac{1}{2E_{k_1} 2E_{k_2}} \frac{T}{V}\right)$, which we expect to be one. This factor is due to scattering plane wave states, which have infinite norm. Here, we try justify why we can ignore this factor in front with the cost of adding wave functions. We do so by scattering properly normalized states, which are obtained with the aid of wave packets. Hence, we do not scatter purely plane waves but states $|\psi\rangle$ with a given flavor index a,

$$|\psi, a\rangle \equiv \int_p \psi(p)|p, a\rangle, \qquad \int_p \equiv \int \hat{d}^4 p\; \hat{\delta}(p^2 - m^2)\Theta(p^0), \qquad \langle\psi, b|\psi, a\rangle = \delta^b_a \int_p ||\psi(p)||^2 = \delta^b_a , \tag{S24}$$

---

[7] Note that if we had $\mathcal{L} = m_{ij}\phi_i\phi_j + g f_{ijk}\phi_i\phi_j\phi_k$, such vector products would not vanish and one could study the entanglement generated in mass mixing and more generically in particle-antiparticle oscillations. Studying such systems using our framework could, e.g., shed new insights on entanglement generated in neutrino oscillations [56, 57].

where the hatted notation means $\hat{dx} = dx/2\pi$ and $\hat{\delta}(x) = 2\pi\delta(x)$, so that we do not need to carry around factors of 2π [59]. It is also clear that from the definitions that the momentum-space representation of the wave function is $\langle k, b|\psi, a\rangle = \psi(k)\delta^b_a$. The most general pure initial state is given by

$$|\text{in}\rangle = \sum_{ab}\int_{k_1,k_2} \psi_{12}(k_1,k_2)\Omega_{ab}|k_1,a;k_2,b\rangle\,, \tag{S25}$$

which in principle could be entangled in momenta and flavor space. We want a product state in momentum space $\psi_{12}(k_1,k_2) = \psi_1(k_1)\psi_2(k_2)$ while still permitting entanglement in flavor space. Our initial state becomes

$$|\text{in}\rangle \Rightarrow |\Omega\rangle \equiv \sum_{ab}\int_{k_1,k_2} \psi_1(k_1)\psi_2(k_2)\Omega_{ab}|k_1,a;k_2,b\rangle = \sum_{ab}\Omega_{ab}|\psi_1,a;\psi_2,b\rangle\,. \tag{S26}$$

The normalization of these states $\langle\Omega|\Omega\rangle = 1$ implies that $\sum_{ab}\Omega^\dagger_{ab}\Omega_{ab} = 1$. We then evolve this initial state via the S-matrix $|\text{out}\rangle = \hat{S}|\text{in}\rangle = (1+i\hat{T})|\text{in}\rangle$ and keep terms up to linear order in $\hat{T}$. Performing the partial trace over $\mathcal{H}_B$, we obtain the following reduced density matrix

$$\begin{aligned}\rho_A = &\sum_{crs}\Omega_{rc}\Omega^\dagger_{sc}|\psi_1,r\rangle\langle\psi_1,s| \\ &+\left[i\sum_{abi}\sum_{nm}\Omega_{ab}\Omega^\dagger_{nm}\int_{q_1,q_2,k_1,k_2}\hat{\delta}^{(4)}(k_1{+}k_2{-}q_1{-}q_2)\psi_1(k_1)\psi_2(k_2)\psi^\dagger_2(q_2)\mathcal{A}^{ab}_{im}(k_1{+}k_2{\to}q_1{+}q_2)|q_1,i\rangle\langle\psi_1,n| + \text{h.c.}\right].\end{aligned} \tag{S27}$$

For the trace of ρ_A^2, it is also useful to define the smeared amplitude, which will be an object acting purely on the flavor space,

$$\boldsymbol{\mathcal{A}}^{ab}_{cd} = \int_{p_1,p_2,p_3,p_4}\hat{\delta}^{(4)}(p_1{+}p_2{-}p_3{-}p_4)\psi_1(p_1)\psi_2(p_2)\psi^\dagger_1(p_3)\psi^\dagger_2(p_4)\mathcal{A}^{ab}_{cd}(p_1{+}p_2{\to}p_3{+}p_4)\,. \tag{S28}$$

Squaring the reduced density matrix and finally tracing over $\mathcal{H}_A$ leads to

$$\text{tr}[\rho_A^2] = \sum_{crs}\Omega_{rc}\Omega^\dagger_{sc}\left[\sum_{c'r's'}\Omega_{sc'}\Omega^\dagger_{rc'} + 2i\sum_{abi}\sum_{nm}\Omega_{ab}\Omega^\dagger_{nm}\left(\boldsymbol{\mathcal{A}}^{ab}_{im}\delta^s_i\delta^n_r - \boldsymbol{\mathcal{A}}^{ib}_{nm}\delta^s_a\delta^i_r\right)\right], \tag{S29}$$

which can be put in simplified notation, $\Omega_{ab}\equiv|\boldsymbol{\Omega}\rangle$ and $\Omega^\dagger_{nm}\equiv\langle\boldsymbol{\Omega}|$, to obtain the entanglement

$$\mathcal{E}[\boldsymbol{\Omega}] = 1 - \langle\boldsymbol{\Omega}|\boldsymbol{\Omega}\boldsymbol{\Omega}^\dagger|\boldsymbol{\Omega}\rangle + 4\,\text{Im}[\langle\boldsymbol{\Omega}|\boldsymbol{\Omega}\boldsymbol{\Omega}^\dagger\boldsymbol{\mathcal{A}}|\boldsymbol{\Omega}\rangle]\,. \tag{S30}$$

In the case where the initial state is a product state, $(\Omega\cdot\Omega^\dagger\cdot\Omega)_{rc'} = \Omega_{rc'}$, i.e., $\boldsymbol{\Omega}\boldsymbol{\Omega}^\dagger|\boldsymbol{\Omega}\rangle = |\boldsymbol{\Omega}\rangle$. Then

$$\mathcal{E}[\boldsymbol{\Omega}] = 4\,\text{Im}[\langle\boldsymbol{\Omega}|\boldsymbol{\mathcal{A}}|\boldsymbol{\Omega}\rangle]\,. \tag{S31}$$

Coming back to $\text{tr}[\rho_A^2]$, if we were to calculate the quantum Tsallis entropy, we would need the nth power instead,

$$\text{tr}[\rho_A^n] = \langle\boldsymbol{\Omega}|(\boldsymbol{\Omega}\boldsymbol{\Omega}^\dagger)^{n-1}|\boldsymbol{\Omega}\rangle + 2n\,\text{Im}[\langle\boldsymbol{\Omega}|(\boldsymbol{\Omega}\boldsymbol{\Omega}^\dagger)^{n-1}\boldsymbol{\mathcal{A}}|\boldsymbol{\Omega}\rangle]\,, \tag{S32}$$

which in the case of a product initial state reduces to $\mathcal{E}_n[\boldsymbol{\Omega}] = \frac{2n}{n-1}\,\text{Im}[\langle\boldsymbol{\Omega}|\boldsymbol{\mathcal{A}}|\boldsymbol{\Omega}\rangle]$.

The details of the wave function in the smeared amplitude should not be relevant to our discussion here. However, it is enlightening to show how one recovers the usual story when scattering infinite-norm states by choosing a Dirac delta for the squared wave function, cf. the final expression in Eq. (S24). For example, choosing

$$||\psi_1(p)||^2 \propto \delta^{(3)}(p - P_\text{a}) \qquad ||\psi_2(p)||^2 \propto \delta^{(3)}(p - P_\text{b})\,, \tag{S33}$$

since the smeared amplitude is given by a combination of wave functions $\psi_1(p_1)\psi_2(p_2)\psi^\dagger_1(p_3)\psi^\dagger_2(p_4)$, this will only have support when $p_1 = p_3 = P_\text{a}$ and $p_2 = p_4 = P_\text{b}$, automatically selecting forward kinematics.

Effective Potential and Vacuum Stability in the Litim-Sannino Model

Tom Steudtner [a,b]

[a] *TU Dortmund University, Department of Physics, Otto-Hahn-Str.4, D-44227 Dortmund, Germany*
[b] *Department of Physics, University of Cincinnati, Cincinnati, OH 45221, USA*

E-mail: tom2.steudtner@tu-dortmund.de

Abstract: We revisit the scalar potential in the Litim-Sannino model. We compute for the first time the full quantum corrections to the classical potential and show that they significantly ameliorate the stability analysis at the UV fixed point. The quantum effective potential is computed at two-loop order and the numerical precision is further improved using resummations and parameter optimisations. As a result, we find a consistent widening of the UV conformal window across various approximations.

arXiv:2402.16950v1 [hep-th] 26 Feb 2024

Contents

1 Introduction **1**

2 Litim-Sannino Model **2**

3 Effective Potential **4**
- 3.1 Fixed-Order Potential 4
- 3.2 Renormalisation Group Improvement 7
- 3.3 Fixed-Point Potential 9
- 3.4 Away from the UV Fixed Point 12

4 Impact on Conformal Window **13**
- 4.1 Uncertainties 14
- 4.2 Conformal Expansion 15
- 4.3 Padé Resummations 17
- 4.4 Resummation via β Functions 18

5 Discussion **21**

1 Introduction

Historically the paradigm of asymptotic freedom [1, 2] has been a centre piece for constructing UV-complete four-dimensional QFTs. More recently, a complementary path was discovered with the Litim-Sannino model [3] and its equivalent formulations [4]. The crucial difference is that these QFTs approach conformality via an interacting UV fixed point in the high energy limit. And yet, the Litim-Sannino model retains all benefits of asymptotic freedom: both are consistently defined in the deep UV, cutoff-free, and their predictivity is enhanced by the applicability of perturbation theory. This has enabled a wide range of studies and model building applications, see e.g. [3–23]. In particular, the UV conformal window has been investigated with increasingly high precision [24–27], zoning in on the mechanism causing the eventual demise of the UV fixed point. These studies have systematically uncovered that the conformal window is entirely within a perturbative regime, and that an instability of the scalar potential is one of the most plausible explanation for its upper end. Therefore, a dedicated investigation of vacuum stability is vital to explore the conformal window of the Litim-Sannino model.

While the fixed-point analysis was conducted up to high loop orders, the vacuum stability analysis employed in previous works [24–27] is actually a classical one, as first discussed in [28]. Quantum corrections have been investigated in [6] by employing a renormalisation

group (RG) resummation of the classical potential. This was sufficient to match the leading-order precision of the UV fixed point at the time. Some aspects of symmetry breaking were later studied in [14]. In this work, we demonstrate that the (semi-) classical approximation of the potential is not on par with the precision of the latest fixed-point results [26]. We then extend the scalar stability analysis by computing the effective potential, which fully accounts for quantum corrections.

The quantum effective potential was originally put forward in [29] and further developed in [30]. A comprehensive review of early works is found in [31]. As of today, generalised formulæ to compute the effective potential for any renormalisable QFT are available up to three loops [32, 33] in Landau gauge, as well as at two loop in more general choices of gauge fixing [34]. Since the potential is an observable and overall RG invariant, these fixed-order results may be enhanced by RG resummations, see e.g. [35–40]. On the other hand, the potential depends on classical background-field variables, which may introduce an explicit dependence on gauge-fixing parameters [41–43].

This work is structured as follows: Sec. 2 gives a brief introduction to the Litim-Sannino model, our notation, and the stability of its scalar potential at classical level. The computation and discussion of the effective potential follows in Sec. 3. Sec. 4 investigates how the enhanced precision influences the conformal window. We close with a final discussion in Sec. 5.

2 Litim-Sannino Model

The Litim-Sannino model [3] features a non-abelian $SU(N_c)$ gauge group and an unbroken chiral flavour symmetry $U(N_f)_L \times U(N_f)_R$. It contains N_f Dirac fermions ψ in the fundamental representation of the gauge group, as well as N_f^2 complex but neutral scalars ϕ, which are bifundamentals under the global symmetry. The renormalisable Lagrangian compatible with the symmetry group reads

$$\begin{aligned}\mathscr{L} = &-\tfrac{1}{4} F^{A\mu\nu} F^{A}{}_{\mu\nu} + \mathscr{L}_{\text{gf}} + \mathscr{L}_{\text{gh}} \\ &+ \operatorname{tr}\Big[\overline{\psi} i \not{D} \psi\Big] - y \operatorname{tr}\Big[\overline{\psi}\left(\phi\, \mathscr{P}_R + \phi^\dagger \mathscr{P}_L\right)\psi\Big] \\ &+ \operatorname{tr}\Big[\partial^\mu \phi^\dagger \partial_\mu \phi\Big] - m^2 \operatorname{tr}\Big[\phi^\dagger \phi\Big] - u \operatorname{tr}\Big[\phi^\dagger \phi \phi^\dagger \phi\Big] - v \operatorname{tr}\Big[\phi^\dagger \phi\Big] \operatorname{tr}\Big[\phi^\dagger \phi\Big],\end{aligned} \tag{2.1}$$

where traces are taken over flavour and gauge indices while $\mathscr{L}_{\text{gf}}$ and $\mathscr{L}_{\text{gh}}$ refer to gauge-fixing and ghost terms, respectively. This family of theories contains gauge interactions, g, coupled to the fermions, one real Yukawa coupling, y, mediating between the fermionic and scalar sector, as well as two scalar self-interactions u and v.

The intriguing aspect of this theory is the occurrence of a weakly interacting UV fixed point. In particular, this fixed point can be brought under strict perturbative control in the planar Veneziano limit [44], where $N_{f,c} \to \infty$ while the ratio N_f/N_c is a finite, continuous parameter. To that end, we introduce the rescaled couplings [45]

$$\alpha_g = \frac{N_c\, g^2}{(4\pi)^2}, \qquad \alpha_y = \frac{N_c\, y^2}{(4\pi)^2}, \qquad \alpha_u = \frac{N_f\, u}{(4\pi)^2}, \qquad \alpha_v = \frac{N_f^2\, v}{(4\pi)^2}, \tag{2.2}$$

which are finite in the Veneziano limit and absorb leading powers of $N_{f,c}$. Thus, the explicit dependence on $N_{f,c}$ is traded for a single expansion parameter

$$\epsilon = \frac{N_f}{N_c} - \frac{11}{2}\,, \tag{2.3}$$

which is continuously tunable within the range $[-\frac{11}{2}, \infty)$. The couplings at the UV fixed point can be expanded as power series in ϵ (conformal expansion)

$$\alpha_i^* = \sum_{\ell=1}^{\infty} \alpha_i^{(\ell)} \epsilon^\ell\,, \tag{2.4}$$

Thus, if ϵ is small but positive, the UV fixed point is perturbative and its existence guaranteed. The coefficients $\alpha_i^{(\ell)}$ in (2.4) only receive contributions up to $(\ell+1)$ loops from the gauge- as well as ℓ loops from the Yukawa and quartic β functions. Currently, the coefficients are determined up to $\ell = 3$ [26].

The UV fixed point prevails within a parameter range $(0, \epsilon_{\max}]$, known as the UV conformal window. Its extent and the dynamics constraining it have been investigated in [24–27]. The primary reason for the demise of the UV fixed point appears to be vacuum instability, i.e. scalar potential is not bounded from below at the UV fixed point.

The classical stability of the potential has been considered in [28] and will be reviewed below. The potential consists only a classical field $\bar{\phi}$, which can be reduced by the applying the $U(N_f)_L \times U(N_f)_R$ global flavour symmetry without loss of generality. In particular, the complex matrix field $\bar{\phi}$ can be diagonalised with real components h_i:

$$\bar{\phi} = \mathrm{diag}(h_1, \ldots, h_{N_f})\,. \tag{2.5}$$

In order to check if the potential is bounded from below, the asymptotic limit of large field values $|h_i| \to \infty$ is of interest. Thus, only the quartic part of the classical potential is relevant, which simplifies to

$$V = u\,\mathrm{tr}\big[\bar{\phi}^\dagger\bar{\phi}\,\bar{\phi}^\dagger\bar{\phi}\big] + v\,\mathrm{tr}\big[\bar{\phi}^\dagger\bar{\phi}\big]\,\mathrm{tr}\big[\bar{\phi}^\dagger\bar{\phi}\big] = u\sum_i h_i^4 + v\sum_{i,j} h_i^2 h_j^2\,. \tag{2.6}$$

For the potential is bounded from below, the quartic part has to be positive for all possible directions $(h_1, \ldots, h_{N_f})^\intercal$ in the space of real moduli. The initial normalisation of the moduli vector is irrelevant we can fix it by $\sum_i h_i^2 = \mathscr{N}$. Using this condition, the minimum of the quartic potential is found by solving

$$\frac{\partial}{\partial h_k}\left(V - \Lambda\left[\mathscr{N} - \sum_i h_i^2\right]\right) = 4h_k\left(u\,h_k^2 + v\,\mathscr{N} + \tfrac{1}{2}\Lambda\right) = 0\,. \tag{2.7}$$

where Λ is a Lagrange parameter. The solution of (2.7) implies that each modulus h_k is either 0 or has a non-vanishing value $h_k = h$, which is universal for all moduli. Assuming that n out of the N_f moduli are non-zero, the quartic potential becomes

$$V_{\mathrm{cl}} = h^4\big[n\,u + n^2\,v\big]\,. \tag{2.8}$$

As the potential is manifestly stable (unstable) for $u, v \gtrless 0$, sufficient stability criteria are found for both couplings have opposite signs. The value of n minimising a positive potential (2.8) depends on the sign of each coupling, leading to the stability conditions

$$\begin{aligned} u > 0 \geq v : &\qquad N_f\, u + N_f^2\, v > 0\,, \\ u \leq 0 \leq v : &\qquad\qquad u + v > 0\,. \end{aligned} \tag{2.9}$$

However, only the first condition is compatible with the Veneziano limit

$$\alpha_u > 0\,, \quad \alpha_u + \alpha_v > 0\,, \tag{2.10}$$

while the second one is voided. Thus, we arrive at the expression

$$\frac{V_{\text{cl}}}{(4\pi)^2} = [\alpha_u + \alpha_v] h^4 \tag{2.11}$$

for the classical potential. In the next section, we review and compute quantum corrections to this potential, extending the results of [6].

3 Effective Potential

3.1 Fixed-Order Potential

Now, we extend the stability consideration to quantum level using the effective potential. For a more detailed introduction see e.g. [29–31]. Starting from the generating functional of connected Green's functions $W[J]$

$$e^{iW[J]} = \int \mathscr{D}\phi \; e^{i\int \mathrm{d}^4x \; [\mathscr{L}(\phi) + J\phi]}\,, \tag{3.1}$$

the effective action is given by a Legendre transformation

$$\Gamma[\bar{\phi}] = W[J] - \int \mathrm{d}^4x \; J\bar{\phi} \tag{3.2}$$

with a classical field variable $\bar{\phi} = \langle\phi\rangle = \frac{\delta W[J]}{\delta J}$. Note that the procedure (3.2) is equivalent to expanding each field ϕ around a classical background $\phi \mapsto \phi + \bar{\phi}$ and integrating out the quantum fluctuations ϕ

$$e^{i\Gamma[\bar{\phi}]} = \int \mathscr{D}\phi \; e^{i\int \mathrm{d}^4x \; [\mathscr{L}(\bar{\phi}+\phi) + J\phi]}\,. \tag{3.3}$$

The parameter J is now a functional of the background field and defined via $J(\bar{\phi}) = -\frac{\delta\Gamma[\bar{\phi}]}{\delta\bar{\phi}}$, defined such that $\Gamma[\bar{\phi}]$ has a minimum at $\bar{\phi} = 0$. The nth functional derivative of $\Gamma[\bar{\phi}]$ with respect to $\bar{\phi}$ yields the sum of all connected, 1-particle irreducible n-point graphs when evaluated at $\bar{\phi} = 0$. The effective potential $V_{\text{eff}}(\bar{\phi})$ is part of $\Gamma[\bar{\phi}]$ with vanishing momenta

$$\Gamma[\bar{\phi}] = \int \mathrm{d}^4x \left[-V_{\text{eff}}(\bar{\phi}) + \mathscr{O}(\partial\bar{\phi})\right], \tag{3.4}$$

and includes the classical potential as well as quantum corrections. For determining V_{eff}, it is sufficient to choose the background field to be constant. Thus, for computational purposes it is convenient to evaluate $\Gamma[\bar{\phi}]$ at $\bar{\phi}(x) = \varphi$, which amounts to the shift of the classical field by a constant value φ in the action. Subsequently, φ can be absorbed into the masses and couplings of the theory and hence resummed in loop computations.

There are two approaches to obtain the potential:

1. The effective action, $\Gamma[\varphi]$ may be computed directly from (3.3) [30, 46], by integrating out the field $\phi(x)$, which yields

$$V_{\text{eff}}(\varphi) = V_{\text{cl}}(\varphi) - \frac{i}{2}\int \frac{\mathrm{d}^4k}{(2\pi)^4}\,\ln\left[\det i\mathscr{P}_\phi(k,\varphi)\right] + \left\langle \exp\left(i\int d^4x\,\mathscr{L}_{\text{int}}\right)\right\rangle \tag{3.5}$$

   for the effective potential. Here V_{cl} denotes the classical (tree-level) potential, $\mathscr{P}_\phi$ the propagator of the quantum field ϕ, and $\mathscr{L}_{\text{int}}$ is Lagrangian of cubic and higher interaction terms in ϕ, all of which containing the background-field variable φ. Note that the contribution $J\phi$ in (3.3) subtracts all tadpole diagrams. The last term of (3.5) consists of all vacuum diagrams at two-loops and higher. See [30] for more details.

2. Alternatively [47], one may instead use the relation

$$\left.\frac{\delta\Gamma[\bar{\phi}]}{\delta\bar{\phi}}\right|_{\bar{\phi}=\varphi} = -\frac{\partial V_{\text{eff}}}{\partial\varphi}\,. \tag{3.6}$$

   Here the left hand side is by definition the 1-particle irreducible tadpole of the quantum field ϕ in the original theory where the constant shift $\phi \to \phi + \varphi$ was introduced. Thus, φ remains as a parameter which has to be integrated in order to obtain the effective potential.

Note that since φ is classical, it can be defined in a minimal form, reduced by the application of all anomaly-free global symmetries without loss of generality. In our specific case, this allows us to choose $\varphi = \text{diag}(h_1, \ldots, h_{N_f})$ [28] in accordance with the discussion of Sec. 2. Since we are only interested in the Veneziano limit, we can further simplify

$$\varphi_{ij} = h\,\delta_{ij}\,, \tag{3.7}$$

which corresponds to the ground state corresponding to (2.10). By absorbing the classical field h into the couplings of quantum fields, we can resum its effect in loop computations. This softly breaks the global symmetry group as it introduces additional superrenormalisable couplings, including mass terms. In our particular case, the $U(N_f)_L \times U(N_f)_R$ symmetry reduces to the vectorial subgroup $U(N_f)_V$, while the axial part is broken by the classical field. Overall, the complex scalar decomposes into the classical background field, a real singlet scalar S and pseudo-scalar $\tilde{S}$, as well as the adjoint fields R and I

$$\phi_{jk} = h\,\delta_{jk} + \frac{\delta_{jk}}{\sqrt{2N_f}}\left[S + i\,\tilde{S}\right] + R_{jk} + iI_{jk} \tag{3.8}$$

with the respective mass terms

$$\begin{aligned} m_R^2 &= m^2 + 2h^2\left(3\,u + N_f\,v\right), \\ m_I^2 &= m^2 + 2h^2\left(u + N_f\,v\right), \\ m_S^2 &= m^2 + 6h^2\left(u + N_f\,v\right), \\ m_{\tilde{S}}^2 &= m^2 + 2h^2\left(u + N_f\,v\right), \\ m_\psi^2 &= y^2\,h^2\,. \end{aligned} \tag{3.9}$$

Moreover, eq. (3.8) generates a tadpole interaction

$$-\mathscr{L}_{\text{Tad}} = \sqrt{2N_f}h\left[m^2 + 2h^2\left(u + v\,N_f\right)\right]S\,. \tag{3.10}$$

The decomposition of the scalar field leads to a plethora of Yukawa

$$-\mathscr{L}_{\text{Yuk}} = \frac{y}{\sqrt{2N_f}}\overline{\psi}_{ia}\left(S + i\gamma_5\,\tilde{S}\right)\psi_{ia} + y\,\overline{\psi}_{ia}\left(R_{ij} + i\gamma_5\,I_{ij}\right)\psi_{ja}\,, \tag{3.11}$$

scalar cubic

$$\begin{aligned} h^{-1}V_{\text{Cubic}} = {} & 4u\,\text{tr}[RRR + RII] \\ & + 2\sqrt{\frac{2}{N_f}}(3\,u + N_f\,v)\,S\,\text{tr}RR + 2\sqrt{\frac{2}{N_f}}(u + N_f\,v)\,S\,\text{tr}II \\ & + 4\sqrt{\frac{2}{N_f}}\,u\,\tilde{S}\,\text{tr}RI + \sqrt{\frac{2}{N_f}}(u + N_f\,v)\,S\left(S^2 + \tilde{S}^2\right) \end{aligned} \tag{3.12}$$

and quartic interactions

$$\begin{aligned} V_{\text{Quartic}} = {} & v\,\text{tr}^2\left[RR + II\right] + u\,\text{tr}[RRRR + IIII + 4RRII - 2RIRI] \\ & + 2\sqrt{\frac{2}{N_f}}\,u\left[S\,\text{tr}\left(RRR + RII\right) + \tilde{S}\,\text{tr}\left(IRR + III\right)\right] \\ & + \left(v + \frac{3u}{N_f}\right)\left[S^2\,\text{tr}RR + \tilde{S}^2\,\text{tr}II\right] + \left(v + \frac{u}{N_f}\right)\left[S^2\,\text{tr}II + \tilde{S}^2\,\text{tr}RR\right] \\ & + \frac{4u}{N_f}\,S\tilde{S}\,\text{tr}RI + \frac{1}{4}\left(v + \frac{u}{N_f}\right)\left(S^2 + \tilde{S}^2\right)^2\,. \end{aligned} \tag{3.13}$$

In the Veneziano limit, it is expected that merely the scalars R and I are contributing to loop interactions. The same picture emerges for sister theories of the Litim-Sannino model, featuring orthogonal and symplectic gauge groups, Majorana fermions, and a single $SU(N_f)$ flavour symmetry [4]. In the orthogonal case, the scalar ϕ is in the symmetric representation of $SU(N_f)$. Breaking the diagonal component breaks the global symmetry down to $SO(N_f)$, with R and I being in the symmetric representation. In the symplectic case, scalars ϕ are in the antisymmetric representation of $SU(N_f)$ while R, I are in the antisymmetric representation of $Sp(N_f)$ after the breaking. An overview is found in Tab. 1.

	Dirac $SU(N_c)$	Majorana $SO(N_c)$	Majorana $Sp(N_c)$
Unbroken symmetry	$SU(N_f)_L \times SU(N_f)_R$	$SU(N_f)$	$SU(N_f)$
Weyl fermions	$2N_cN_f$	N_cN_f	N_cN_f
Scalar DOF ϕ	$2N_f^2$	$N_f(N_f+1)$	$N_f(N_f-1)$
Remnant Symmetry	$SU(N_f)_V$	$SO(N_f)$	$Sp(N_f)$
Weyl fermions	$2N_cN_f$	N_cN_f	N_cN_f
Scalar DOF R, I	N_f^2-1	$\frac{1}{2}N_f(N_f+1)-1$	$\frac{1}{2}N_f(N_f-1)-1$
Scalar DOF S, P	1	1	1

Table 1. Overview of fermionic and scalar degrees of freedom with and without symmetry breaking for the Litim-Sannino model and its sister theories [4].

3.2 Renormalisation Group Improvement

The effective potential depends on renormalisation group scale μ explicitly and implicitly through the running of the background field h, couplings α_i, mass m^2 and the R_ξ gauge fixing parameter ξ. However, as the potential is overall renormalisation group invariant and follows a Callan-Symanzik relation

$$\frac{\mathrm{d}}{\mathrm{d}\ln\mu}V_{\text{eff}} = \left(\frac{\partial}{\partial\ln\mu} + \gamma_h h\frac{\partial}{\partial h} + \beta_j\frac{\partial}{\partial\alpha_j} + \beta_\xi\frac{\partial}{\partial\xi} + \gamma_{m^2}\frac{\partial}{\partial\ln m^2}\right)V_{\text{eff}} = 0\,. \tag{3.14}$$

Moreover, we can make the ansatz

$$\frac{V_{\text{eff}}}{(4\pi)^2} = \alpha_h(z,\kappa,\alpha_i,\xi)\,h^4, \quad \text{where} \quad z = \frac{(4\pi)^2}{N_f}\frac{h^2}{\mu^2} \quad \text{and} \quad \kappa = \frac{m^2}{\mu^2}\,, \tag{3.15}$$

which retains higher orders of the classical field in the variable z. In the same manner, the scalar mass parameter is swapped for the dimensionless variable κ without loss of generality. The potential is bounded from below if $\alpha_h > 0$ for all values of z. The RG invariance condition then reads

$$\left(-\frac{2\gamma_h}{1-\gamma_h} + \frac{\partial}{\partial\ln z} + \frac{1}{2}\frac{\beta_j}{1-\gamma_h}\frac{\partial}{\partial\alpha_j} + \frac{1}{2}\frac{\beta_\xi}{1-\gamma_h}\frac{\partial}{\partial\xi} + \frac{\kappa}{2}\frac{\gamma_{m^2}-2}{1-\gamma_h}\frac{\partial}{\partial\kappa}\right)\alpha_h = 0\,. \tag{3.16}$$

Note that the term encoding the running of the gauge fixing parameter can always be switched off by choosing the Landau gauge $\xi = 0$, as $\beta_\xi \propto \xi$. Due to the RG invariance, it is sufficient obtain a solution of the effective potential close to the fixed point in order to argue about vacuum stability.

At the fixed point, β functions as well as κ vanish, while z remains a tunable parameter, containing the background field dependence. We obtain

$$0 = \left(-\frac{2\gamma_h^*}{1-\gamma_h^*} + \frac{\partial}{\partial\ln z}\right)\alpha_h(z,0,\alpha_i^*,\xi^*)\,. \tag{3.17}$$

Solving this equation and expressing α_i^*, ξ^* through their ϵ expansion yields a resummed h-dependence

$$\alpha_h(z,\,0,\,\alpha_i^*,\,\xi^*) = \alpha_w^*(z_0,\,\epsilon)\left(\frac{z}{z_0}\right)^{2\gamma_h^*/(1-\gamma_h^*)}\,, \tag{3.18}$$

where z_0 is chosen as an arbitrary positive parameter. Obviously, a parameter change in z_0 amounts to

$$\alpha_w^*(z_0,\epsilon)=\left(\frac{z_0}{z_0'}\right)^{2\gamma_h^*/(1-\gamma_h^*)}\alpha_w^*(z_0',\epsilon)\,. \tag{3.19}$$

Thus, terms in the effective potential with higher orders in h, which are contained in z, are resummed in the last factor of eq. (3.18). Most notably, this term is always positive. Hence the boundedness from below of the quantum effective potential at the fixed point is determined by the sign of α_w^* and independent of the field values z. In particular, α_w^* can be computed in perturbation theory by calculating the effective potential in a loop expansion at $z=z_0$:

$$\alpha_w^*=\sum_{\ell=0}^{\infty}\alpha_w^{(\ell)}>0\,. \tag{3.20}$$

Importantly, the stability of the effective potential does not depend on the choice of the parameter z_0, which can be fixed in a way that optimises the perturbative expansion. In particular, logarithmic terms $\propto\epsilon^2\ln\epsilon/z_0$ are already present at one-loop in α_w^*. For e.g. the trivial choice $z_0=1$, such terms are under perturbative control as they vanish for $\epsilon\to 0$. However, these logarithms become much more sizeable than non-logarthmic contributions in the limit $\epsilon\to 0$. On the other hand, the choice $z_0=\epsilon$ eliminates all terms $\propto\ln\epsilon/z_0$, which improves the perturbative convergence for every order in $\alpha_w^{(\ell)}$. We will return to this opportunity in Sec. 4.

The leading term in (3.20) represents the tree-level of the effective potential

$$\alpha_w^{(0)}=\alpha_u^*+\alpha_v^*\,, \tag{3.21}$$

which coincides with the classical potential and is independent of z_0. In [6, 24–27], the condition $\alpha_w^{(0)}>0$ has been adopted as a criterion for vacuum stability. This rests on the argument in [6] that the classical potential $\alpha_w^{(0)}$ controls the sign of leading-log resummations of the background field in the effective potential. Ideed, using the loop expansion $\gamma_h^*=\sum_{\ell=1}^{\infty}\gamma_h^{(\ell)}$, the resummation (3.18) at the fixed point reads

$$\alpha_h^*=\sum_{n=0}^{\infty}\frac{2^n}{n!}\left[\ln\frac{z}{z_0}\right]^n\left\{\alpha_w^{(0)}\left(\gamma_h^{(1)}\right)^n+\left[\alpha_w^{(1)}\gamma_h^{(1)}+n\,\alpha_w^{(0)}\left(\gamma_h^{(2)}+\gamma_h^{(1)2}\right)\right]\left(\gamma_h^{(1)}\right)^{n-1}+\ldots\right\}. \tag{3.22}$$

The first term in the braces of (3.22) is the leading-log (LL) contribution, which is $\propto\alpha_w^{(0)}$. The second term is next-to-leading-log (NLL) and hence suppressed with an additional loop order. In general, powers $\ln^n z$ involve contributions up N^{n}LL which contain n-loop quantum corrections to the classical potential $\alpha_w^{(n)}$. Boundedness from below is determined in the limit $z\to\infty$ where the term $n\to\infty$ in (3.22) dominates, which means that all loop corrections to α_w^* contribute to its coefficient and thus vacuum stability. The classical stability condition $\alpha_w^{(0)}>0$ ensures the positivity of LL term which is naïvely the largest contribution. Although pertubatively suppressed, the subleading-log corrections may still overpower the LL numerically, in particular at the edge of classical stability where $\alpha_w^{(0)}\approx 0$. This scenario is accounted for when considering the full α_w^*. Thus, the focus of this work is the computation of quantum corrections $\alpha_w^{(\ell>0)}$, in order to improve the prediction on vacuum stability.

3.3 Fixed-Point Potential

In this section, the stability of the scalar potential in the high-energy limit is determined by computing the quantity α_w, defined in (3.18) at the UV fixed point. Since all fixed-point values $\alpha^*_{g,y,u,v}$ of classically marginal couplings are known up to $\propto \epsilon^3$ in the conformal expansion, the effective potential needs to be calculated up to two-loop order to reach the same precision for α_w. The effective potential has been determined by a direct loop computation using the computational setup of `MaRTIn` [48]. Both methods outlined in Sec. 3.1 have been pursued and yield identical results. Finally, literature expressions applicable to generic renormalisable QFTs [32, 34] have been utilised to cross check the results. At this loop order, there is no dependence on the gauge fixing. Concretely, we obtain

$$\alpha_w^{(0)} = \alpha_u + \alpha_v\,, \tag{3.23}$$

$$\alpha_w^{(1)} = (\alpha_u+\alpha_v)^2\left[L_I-\tfrac{3}{2}\right]+(3\,\alpha_u+\alpha_v)^2\left[L_R-\tfrac{3}{2}\right]-\left(\tfrac{11}{2}+\epsilon\right)\alpha_y^2\left[L_F-\tfrac{3}{2}\right], \tag{3.24}$$

$$\begin{aligned}
\alpha_w^{(2)} = {}& \left(\tfrac{11}{2}+\epsilon\right)^2\left[4\,\Phi_{RF}-(3-L_F)^2\right]\alpha_y^3+\left(\tfrac{11}{2}+\epsilon\right)\left[9-8\,L_F+3\,L_F^2\right]\alpha_y^2\,\alpha_g\\
&-(11+2\epsilon)\left[9+4\,L_F-L_I-15L_R+L_F\,(L_I+9\,L_R-5\,L_F)\right]\alpha_y^2\,\alpha_u\\
&-(11+2\epsilon)\left[2+2\,L_F-L_I-5L_R+L_F\,(L_I+3\,L_R-2\,L_F)\right]\alpha_y^2\,\alpha_v\\
&-(11+2\epsilon)\left[(\alpha_u+\alpha_v)\,\Phi_{IF}+2(3\alpha_u+\alpha_v)\,\Phi_{RF}\right]\alpha_y^2\\
&+\left[50-10\,L_F^2+2(L_F-2)(9\,L_R+L_I)+\Phi_{IF}+9\,\Phi_{RF}\right]\alpha_y\,\alpha_u^2\\
&+\left[10-2\,L_F^2+2(L_F-2)(L_R+L_I)+\Phi_{IF}+\Phi_{RF}\right]\alpha_y\,\alpha_v^2\\
&+\left[40-8\,L_F^2+4(L_F-2)(3\,L_R+L_I)+2\,\Phi_{IF}+6\,\Phi_{RF}\right]\alpha_y\,\alpha_u\,\alpha_v\\
&+\left[192-243\,S_2-\Phi_{RI}-24\,L_I-168\,L_R+L_I^2+18\,L_I L_R+45\,L_R^2\right]\alpha_u^3\\
&+\left[108-81\,S_2-3\,\Phi_{RI}-40\,L_I-104\,L_R+6\,L_I^2+24\,L_I L_R+30\,L_R^2\right]\alpha_u^2\,\alpha_v\\
&+4(2-L_I-L_R)(3-L_I-2\,L_R)\alpha_u\,\alpha_v^2+(2-L_I-L_R)^2\alpha_v^3\,.
\end{aligned} \tag{3.25}$$

For convenience we define the couplings

$$\alpha_F = \left(\tfrac{11}{2}+\epsilon\right)\alpha_y\,,\quad \alpha_I = 2(\alpha_u+\alpha_v)\,,\quad \alpha_R = 2(3\,\alpha_u+\alpha_v)\,, \tag{3.26}$$

which stem from the squared masses (3.9) of fermions and adjoint scalars R and I as the unbroken scalar mass parameter m^2 is a relevant operator and vanishes at the fixed point $m^2\to 0$. We have employed these couplings as arguments for the (poly-) logarithmic terms

$$L_X = \ln[\alpha_X/z_0]\,,\qquad \Phi_{XY} = \Phi\left(\frac{\alpha_X}{4\,\alpha_Y}\right), \tag{3.27}$$

where $X, Y = F, R, I$. Note that these contributions are the only dependencies of the parameter z_0. The Φ_{XY} are loop functions defined in [49] in terms of hypergeometric functions

$$\Phi(z) = 4z\left[(2-\ln 4z)\;{}_2F_1\left(\left.\begin{matrix}1,1\\3/2\end{matrix}\right|z\right)-\frac{\partial}{\partial a}\;{}_2F_1\left(\left.\begin{matrix}a,1\\3/2\end{matrix}\right|z\right)\Big|_{a=1}-\frac{\partial}{\partial c}\;{}_2F_1\left(\left.\begin{matrix}1,1\\c\end{matrix}\right|z\right)\Big|_{c=3/2}\right] \tag{3.28}$$

or equivalently

$$\Phi(z) = \begin{cases} \frac{4}{\sqrt{1/z-1}}\mathrm{Cl}_2\left(2\arcsin\sqrt{z}\right) & 0 \le z < 1, \\ 8\ln 2 & z = 1, \\ \frac{1}{\sqrt{1-1/z}}\left[\frac{\pi^2}{3} - \ln^2 4z + 2\ln^2\frac{1-\sqrt{1-1/z}}{2} - 4\mathrm{Li}_2\left(\frac{1-\sqrt{1-1/z}}{2}\right)\right] & z > 1 \end{cases} \tag{3.29}$$

in the non-negative parameter range, where Li_2 is a dilogarithm and Cl_2 denotes the Clausen function. For large arguments, the function diverges as

$$\lim_{z\to\infty}\Phi(z) = \frac{\pi^2}{3} + \ln^2 4z\,. \tag{3.30}$$

Finally, S_2 is a constant with the value

$$S_2 = \tfrac{1}{9}\Phi\left(\tfrac{1}{4}\right) = \tfrac{4}{9\sqrt{3}}\mathrm{Cl}_2\left(\tfrac{\pi}{3}\right) \approx 0.260434\ldots\,. \tag{3.31}$$

In the region where the potential is classically stable $\alpha_u > 0$, $\alpha_u + \alpha_v > 0$, the squared masses (3.9) of fermions and all scalars are positive. This changes at the point of classical flatness $\alpha_u + \alpha_v = 0$ where L_I becomes singular. However, the effective potential remains well-defined

$$\alpha_w^{(0)}\big|_{\alpha_u+\alpha_v=0} = 0\,, \tag{3.32}$$

$$\alpha_w^{(1)}\big|_{\alpha_u+\alpha_v=0} = 4\,\alpha_u^2\left[L_R - \tfrac{3}{2}\right] - \left(\tfrac{11}{2}+\epsilon\right)\alpha_y^2\left[L_F - \tfrac{3}{2}\right], \tag{3.33}$$

$$\begin{aligned}\alpha_w^{(2)}\big|_{\alpha_u+\alpha_v=0} = {}& \left(\tfrac{11}{2}+\epsilon\right)^2\left[4\,\Phi_{RF} - (3-L_F)^2\right]\alpha_y^3 + \left(\tfrac{11}{2}+\epsilon\right)\left[9 - 8\,L_F + 3\,L_F^2\right]\alpha_y^2\,\alpha_g \\ & - (11+2\epsilon)\left[7 - 10 L_R + L_F\,(2 + 6\,L_R - 3\,L_F) + 4\Phi_{RF}\right]\alpha_y^2\,\alpha_u \\ & + 4\left[5 + \Phi_{RF} - L_F^2 + 2\,L_R(L_F - 2)\right]\alpha_y\,\alpha_u^2 \\ & + \left[104 + \tfrac{2}{3}\pi^2 - 162 S_2 - 88\,L_R + 24\,L_R^2\right]\alpha_u^3\,.\end{aligned} \tag{3.34}$$

In general, there are two symptoms of the effective potential coupling α_w that indicate the breakdown of the UV fixed point:

(i) $\mathrm{Re}\,\alpha_w < 0$ signals that the potential is not bounded from below and therefore no ground state of finite energy exists.

(ii) $\mathrm{Im}\,\alpha_w \neq 0$ is an indication that the effective potential ceases to be convex [50], which suggests an additional minimum with $h \neq 0$. The imaginary part can be interpreted as the decay width between degenerate vacua [31, 51].

In the classical approximation $\alpha_w = \alpha_w^{(0)}$, both phenomena coincide at $\alpha_u^* + \alpha_v^* < 0$. For one, this suggests that the classical potential is not bounded from below as $h \to \pm\infty$, and there is no ground state of finite energy. Furthermore, the condition also turns the tree-level mass of the I field tachyonic, $m_I^2 < 0$, implying symmetry breaking into two non-trivial vacua and therefore a non-convex potential. At loop level, $m_I^2 < 0$ generates an imaginary part through L_I, Φ_{IF} and Φ_{RI}.

However, both mechanisms (i) and (ii) receive corrections at loop level, such that it is not clear *a priori* which one is the dominant effect to herald scalar instability. In particular, even the imaginary part implied by $m_I^2 < 0$ may potentially be cancelled by resumming higher-order corrections. The question whether $\mathrm{Im}\,\alpha_w \neq 0$ is of physical consequence or merely signals a deficiency of the perturbative expansion is indeed non-trivial. The parameter region close to vacuum instability is numerically problematic as $|m_I^2|$ is very tiny or vanishing, implying large logarithmic contributions $\propto \ln m_I^2$ to the effective potential at each loop order and thus weakening the perturbative convergence. This is known as the Goldstone catastrophe [39, 52–55]. The impact of the Goldstone catastrophe can be mitigated by resumming diagrams that correspond to soft contributions to the light or vanishing scalar mass. This yields corrections to their masses $\tilde{m}_I^2 = m_I^2 + \delta m_I^2$. The corrected mass $\tilde{m}_I^2$ is then used in place of m_I^2, especially in logarithms where the numerical impact of small $|m_I^2|$ is large. This corresponds to resumming higher-order terms in the effective potential. In our case the strategy is implemented by means of the softened coupling

$$\tilde{\alpha}_I = \alpha_I + \Delta_I \tag{3.35}$$

where the shift admits a loop expansion $\Delta_I = \sum_{\ell=1}^{\infty} \Delta_I^{(\ell)}$. In order to retain the non-logarithmic terms of the effective potential up to $\propto \epsilon^3$ while also addressing the Goldstone catastrophe with the same precision, we make the replacement $\alpha_I \mapsto \tilde{\alpha}_I$ in the logarithmic terms, yielding the alternative definitions

$$L_I = \ln\left[\frac{\alpha_I + \Delta_I}{z_0}\right], \quad \Phi_{IF} = \Phi\left[\frac{\alpha_I + \Delta_I}{4\alpha_F}\right], \quad \Phi_{RI} = \Phi\left[\frac{\alpha_R}{4(\alpha_I + \Delta_I)}\right] \tag{3.36}$$

in (3.23)–(3.25) as well as the shift

$$\alpha_w^{(2)} \mapsto \alpha_w^{(2)} - \tfrac{1}{2}(\alpha_u + \alpha_v)\Delta_I^{(1)}. \tag{3.37}$$

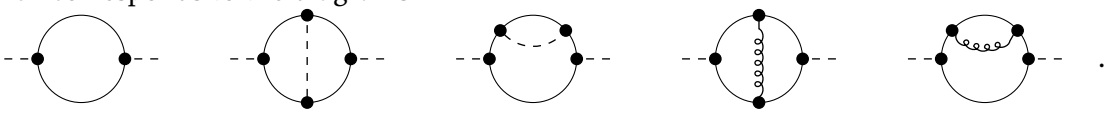
to avoid double counting of resummed contributions. This mechanism allows to trade the imaginary part of α_w for a real contribution while improving the numerical reliability. The choice of Δ_I is not unique but the impact of different prescriptions diminishes with higher loop orders of the effective potential. The literature [39, 52–55], advocates choices related to the self-energy of the Goldstone boson at zero momentum $\Pi_I(0)$, which can be interpreted as a resummation of soft leg corrections to its mass. Self-interactions of the Goldstone bosons are often neglected. In our case, we choose to resum Yukawa and gauge interactions to implement a sufficiently effective resummation

$$\Delta_I = \sum_{\ell=1}^{\infty} \Delta_I^{(\ell)} = \frac{N_f}{h^2}\Pi_I(0)\Big|_{\alpha_{I,R}=0}. \tag{3.38}$$

We have computed the self-energy to two loop order to extract $\tilde{\alpha}_I$ up to order $\propto \epsilon^3$, and obtain in the loop expansion

$$\begin{aligned}
\Delta_I^{(1)} &= (11+2\epsilon)\alpha_y^2\left[1 - L_F\right], \\
\Delta_I^{(2)} &= (11+2\epsilon)\alpha_g\alpha_y^2\left[5 - 5\,L_F + 3\,L_F^2\right] - \tfrac{1}{2}(11+2\epsilon)^2\alpha_y^3\left[6 - 5\,L_F + L_F^2 - 4\,\Phi_{RF}\right],
\end{aligned} \tag{3.39}$$

which corresponds to the diagrams



3.4 Away from the UV Fixed Point

As the effective potential is RG invariant, vacuum stability at the UV fixed point also guarantees the potential to be bounded from below in the IR. Thus, scalar stability reigns along the weakly coupled trajectory connecting the interacting UV fixed point with IR freedom. More explicitly, we follow the relevant trajectory

$$\alpha_g = \alpha_g^*(\epsilon) + \delta, \qquad \alpha_{y,u,v} = \alpha_{y,u,v}^*(\epsilon) + \sum_{n=1}^{\infty} c_{y,u,v}^{(n)}(\epsilon)\,\delta^n \tag{3.40}$$

which emanates from the fixed point at $\delta = 0$ and leads into the weakly ($\delta < 0$) or strongly ($\delta > 0$) coupled regime. The RG flow of the effective coupling δ at leading order is related to the relevant critical exponent $\beta_\delta = \mathrm{d}\delta/\mathrm{d}\ln\mu = \vartheta_1\,\delta + \mathcal{O}(\delta^2)$. The RG invariance of the effective potential away from the fixed point is encoded in the condition

$$0 = \left(-\frac{2\gamma_h}{1-\gamma_h} + \frac{\partial}{\partial \ln z} + \frac{1}{2}\frac{\beta_\delta}{1-\gamma_h}\frac{\partial}{\partial\delta}\right)\alpha_h(z,\delta,\epsilon)\,. \tag{3.41}$$

In accordance with (3.18) we apply the ansatz

$$\alpha_h(z,\,\delta,\,\epsilon) = \alpha_w(\delta,\,z_0,\,\epsilon)\left(\frac{z}{z_0}\right)^{2\gamma_h^*/(1-\gamma_h^*)}, \tag{3.42}$$

yields a condition

$$0 = \left(\frac{\gamma_h - \gamma_h^*}{1-\gamma_h^*} - \frac{\beta_\delta}{4}\frac{\partial}{\partial\delta}\right)\alpha_w(\delta,\,z_0,\,\epsilon) \tag{3.43}$$

that is valid away from the UV fixed point and solved by

$$\alpha_w = \alpha_w^*\,\exp\left\{\frac{4}{1-\gamma_h^*}\int_0^\delta \mathrm{d}\hat\delta\,\frac{\gamma_h(\hat\delta)-\gamma_h^*}{\beta_{\hat\delta}}\right\}. \tag{3.44}$$

Eqs. (3.42) and (3.44) imply that if the effective potential is stable at the UV fixed point through $\alpha_w^* > 0$, it also remains stable away from it.

The scalar mass operator m^2 as in (2.1) is relevant and vanishes at the UV fixed point. If it is switched on at some IR scale $m^2(\mu_{\rm IR}) > 0$, all scalars eventually decouple and the theory consists of free fermions and vector bosons in the low-energy limit, with an intact $U(N_f)_L \times U(N_f)_R$ global symmetry. If $m^2(\mu_{\rm IR}) = 0$, all fields including scalars remain massless until IR freedom is reached. For $m^2(\mu_{\rm IR}) < 0$, spontaneous symmetry breaking occurs and the scalar sector develops a vacuum expectation value h as in (3.8). Other mechanisms to violate the global symmetry in the IR without spoiling the UV fixed point require an explicit breaking by relevant operators [14]. A dynamical symmetry breaking by turning relevant couplings to be irrelevant via the RG cannot occur in the weak coupling regime.

The other trajectory emanating from the UV fixed point leads away from the weak regime towards strong couplings. Again, spontaneous symmetry breaking occurs if $m^2(\mu_{\rm IR}) < 0$. For $m^2(\mu_{\rm IR}) = 0$, the theory may reach a putative interacting IR fixed point, a scenario compatible with the observation of fixed-point mergers [24–26]. Alternatively, the IR trajectory may also lead into a strongly coupled regime, where an additional scale parameter is generated via dimensional transmutation, and non-perturbative phenomena arise such as the breaking of global symmetries and confinement. For larger value of ϵ, an indication for non-perturbative effects causing a symmetry breaking in the effective potential is $\gamma_h^* = 1$, which causes a breakdown of the solution (3.18).

4 Impact on Conformal Window

In this section we discuss the UV conformal window of the Litim-Sannino model, namely the value $\epsilon_{\rm max}$ at which the fixed point ceases to exist. Several mechanisms are known that may cause such a disappearance [56]. As established by previous works [24, 26, 27], the conformal window appears to be within a weakly-interacting regime $|\alpha^*_{g,y,u,v}(\epsilon)| \ll 1$ and not limited due to strong-coupling effects. Rather, the UV fixed point may disappear into the complex plane due to a fixed-point merger. This occurs when an IR fixed-point solution exists at $\epsilon < \epsilon_{\rm mer}$ that approaches the UV one with increasing ϵ, causing both solutions to become complex at $\epsilon > \epsilon_{\rm mer}$. We distinguish two subcategories of this phenomenon: single- and double-trace mergers. A single-trace merger is characterised by the relevant critical exponent turning marginal

$$\vartheta_1(\epsilon_{\rm mer,1}) = 0\,, \tag{4.1}$$

signalling the disappearance of the UV fixed point. Typically, the single-trace couplings $\alpha^*_{g,y,u}(\epsilon < \epsilon_{\rm mer,1})$ of the colliding UV and IR fixed point solutions are different before this merger. The second type of merger is due to the algebraic decoupling in the planar Veneziano limit. The single-trace RG evolution via $\beta_{g,y,u}$ is independent of the double-trace quartic α_v, while its own β function is only quadratic in it

$$\beta_v = f_0 + f_1\,\alpha_v + f_2\,\alpha_v^2 \tag{4.2}$$

up to all loop orders [57]. Thus, each fixed-point solution $\alpha^*_{g,y,u}$ implies up to two real solutions $\alpha_v^{*\pm}$ for the double-trace quartic with the critical exponents

$$\vartheta_3^{\pm} = \pm\sqrt{f_1^{*2} - 4 f_0^* f_2^*}\,. \tag{4.3}$$

For the UV fixed point, only α_v^{*+} is physical while α_v^{*-} exhibits an unstable potential [26]. A double-trace merger occurs when both solutions collide such that

$$\alpha_v^{*+}(\epsilon_{\rm mer,2}) = \alpha_v^{*-}(\epsilon_{\rm mer,2}) \qquad \text{and} \qquad \vartheta_3^{\pm}(\epsilon_{\rm mer,2}) = 0. \tag{4.4}$$

While there is evidence of a single-trace merger, a double-trace merger seems to be absent in the ϵ-expansion [26]. However, constraints to the conformal window due to instability of the fixed-point potential, $\epsilon_{\rm vac}$, appear to be slightly tighter. To settle which mechanism is more

dominant for the upper end of the conformal window, more precision in their determination is required. Thus, in the following analysis we improve upon previous works [26, 27] by bringing the stability analysis on par with available loop orders of the β functions. To do so, we first take stock of all uncertainties for the conformal-window estimate.

4.1 Uncertainties

Computing the effective potential at the UV fixed point within a finite order in perturbation theory, we are confronted with the following sources of uncertainty.

1. Uncertainty of the fixed-point couplings $\alpha^*_{g,y,u,v}(\epsilon)$ and critical exponents $\vartheta_{1,2,3,4}(\epsilon)$. The state-of-the.art is to extract these quantities from four-loop gauge, three-loop Yukawa and three-loop quartic β functions, 433 in short. Thus, $\alpha^*_{g,y,u,v}$ are determined up to the third coefficient $\propto \epsilon^3$ in the conformal expansion and do not receive contributions from higher loops. We will also use various techniques to resum higher powers in ϵ, though the reliability of such approaches will remain an open question and will be used to estimate the residual uncertainty.

2. Uncertainty from missing loop orders of the effective potential. In order to match the $\propto \epsilon^3$ expansion of the fixed-point couplings, the effective potential is required to two-loop order. Given that the fixed point remains in a perturbative regime, higher-order contributions to the potential should be increasingly irrelevant. To verify this, we compare the impact of various loop orders in this section.

3. Ambiguity of the Goldstone resummation Δ_I. As discussed earlier, a softened parameter $\tilde{\alpha}_I$ (3.35) is introduced to improve the numerical stability through resummation. This procedure moves the hints of instability from the imaginary to the real part of the effective potential. However, Δ_I is not uniquely defined. Here we keep track on how the choice (3.39) alters the real part of the unresummed potential and ensure that $\tilde{\alpha}_I(\epsilon_{\rm vac}) > 0$ as $\alpha_w(\epsilon_{\rm vac}) = 0$.

4. Suitable choice of z_0. The overall stability estimate is independent of the choice of z_0. This parameter is introduced in (3.18) as a mediator between the exact field resummation and the perturbatively determined coefficient α_w in the effective potential. The convergence of the loop expansion of α_w can be optimised by a smart choice of z_0. Moreover, an uncertainty estimate may be obtained by varying z_0.

Let us now develop a strategy regarding the last point. The parameter z_0 only appears within the logarithms (3.27). Thus, a natural choice of z_0 would be one of the couplings α_F, α_R or $\tilde{\alpha}_I$ to cancel corresponding terms logarithmic terms, or an intermediate value. Values of z_0 outside the range spanned by α_F, α_R and $\tilde{\alpha}_I$ cause unnecessarily large logarithmic contributions. At the fixed point the hierarchy

$$0 < \tilde{\alpha}_I \ll \alpha_R \approx \epsilon < \alpha_F \ll 1 \tag{4.5}$$

usually holds. This is transparently seen in the leading-order expansion

$$0 < 0.13\,\epsilon \ll 0.92\,\epsilon \approx \epsilon < 1.16\,\epsilon \ll 1\,. \tag{4.6}$$

This suggests that z_0 should be chosen within $\alpha_F \geq z_0 \geq \tilde{\alpha}_I$ to optimize α_w. In this range the largest impact of varying z_0 is due to the rift $\tilde{\alpha}_I \ll \alpha_{R,F}$. Fortunately, logarithmic terms in the effective potential are softened by the couplings in front.

This can be seen for instance in (3.24), which contains both the terms $\propto \alpha_R^2 \ln \alpha_R/z_0$ and $\propto \alpha_I^2 \ln \tilde{\alpha}_I/z_0$. The reader is reminded that by definition (3.35), $\tilde{\alpha}_I$ and α_I differ only by a higher loop expression. The choice $z_0 \approx \tilde{\alpha}_I$ cancels the logarithm $\propto \alpha_I^2 \ln \tilde{\alpha}_I/z_0$ minimizes. However, as $\alpha_I^2 \ll \alpha_R^2$, the term was already smaller than the other $\alpha_R^2 \ln \alpha_R/z_0$. In fact the latter term $\propto \alpha_R^2 \ln \alpha_R/\tilde{\alpha}_I$ becomes even more sizeable due to the logarithm. Instead, the converse choice $z_0 \approx \alpha_R$ is more meaningful: the logarithm $\propto \alpha_R^2 \ln \alpha_R/z_0$ is traded in for a contribution $\propto \alpha_I^2 \ln \tilde{\alpha}_I/\alpha_R$. While the logarithm is equally large in both cases, the perturbative suppression by the coupling in front is much stronger in the second approach. Conclusively, z_0 should be chosen away from $\tilde{\alpha}_I$ to optimise the perturbativity of loop corrections to the effective potential. More precisely, the parameter should be varied in the range of remaining logarithmic scales

$$\alpha_F \geq z_0 \geq \alpha_R \,, \tag{4.7}$$

which also yields an uncertainty estimate. For the perturbative expansion of the effective potential to be reliable, we expect that for each loop order the uncertainty range due to (4.7) is decreasing and higher-loop ranges are roughly nested within lower-order ones.

4.2 Conformal Expansion

In this section, we investigate the implications of scalar stability from the ϵ expansion of the effective potential. The expansion is reliable up to $\propto \epsilon^3$, after which higher loop orders in the β functions and the effective potential are required. We start by recalling the classical stability condition $\alpha_w^{(0)} > 0$ with

$$\alpha_w^{(0)} = \alpha_u^* + \alpha_v^* \approx 0.0625\,\epsilon - 0.1915\,\epsilon^2 - 1.6200\,\epsilon^3 + \mathcal{O}(\epsilon^4) \,. \tag{4.8}$$

This result is improved by taking quantum corrections into account

$$\begin{aligned} \alpha_w \approx\ & 0.0625\,\epsilon \\ & - \left(0.2127 + 0.0263 \ln \frac{\epsilon}{z_0}\right)\epsilon^2 \\ & - \left(1.0536 - 0.0372 \ln \frac{\epsilon}{z_0} - 0.0055 \ln^2 \frac{\epsilon}{z_0}\right)\epsilon^3 + \mathcal{O}(\epsilon^4) \,. \end{aligned} \tag{4.9}$$

At order $\propto \epsilon$, both conditions are equivalent and the contribution is manifestly positive, pointing towards stability. At higher orders the quantum effective potential depends on z_0. For $z_0 = \epsilon$, the $\propto \epsilon^2$ correction to the effective potential is negative, indicating an instability at a certain value for ϵ_{vac}. Moreover, the coefficient is more negative than the classical one, implying that the conformal window may be smaller once quantum corrections are included. The $\propto \epsilon^3$ coefficients are again negative in both cases, reinforcing the evidence for the loss of vacuum stability. However, the quantum effective coefficient is larger than in the classical potential, thus relaxing the constraint on the conformal window.

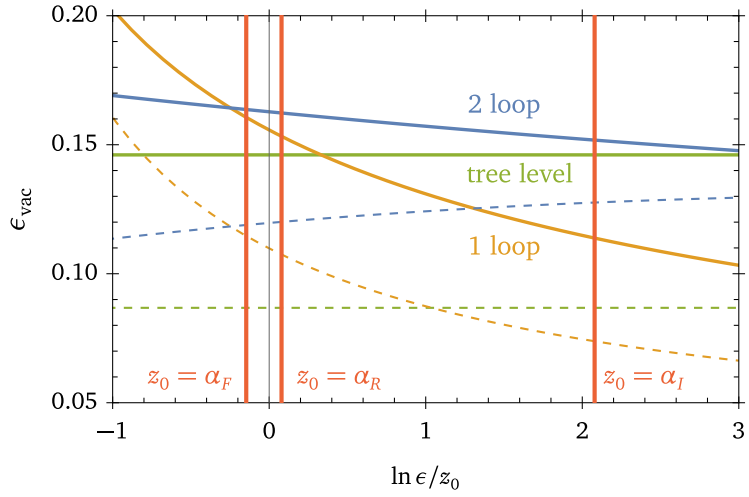


Figure 1. Upper bound of the conformal window $\epsilon_{\rm vac}$ due to vacuum instability depending on the parameter z_0. The bound is extracted from the conformal expansion up to $\sim \epsilon^3$ (solid) and its [2/1] Padé resummation (dashed) of the effective potential at tree (green), one-loop (yellow) and two-loop (blue) level. Red lines mark special choices of z_0.

The dependence of the upper values of the conformal windows on the parameter z_0 for tree-level, one and two loops in the effective potential is depicted in Fig. 1 (solid lines). We find both one- and two-loop corrections from the effective potential to predict a widening of the conformal window over the classical bound. However, the two-loop result shows better convergence as it is less dependent on z_0, especially in the range (4.7) (vertical red lines). We expect this trend to continue to higher loops where the z_0 dependence should diminish.

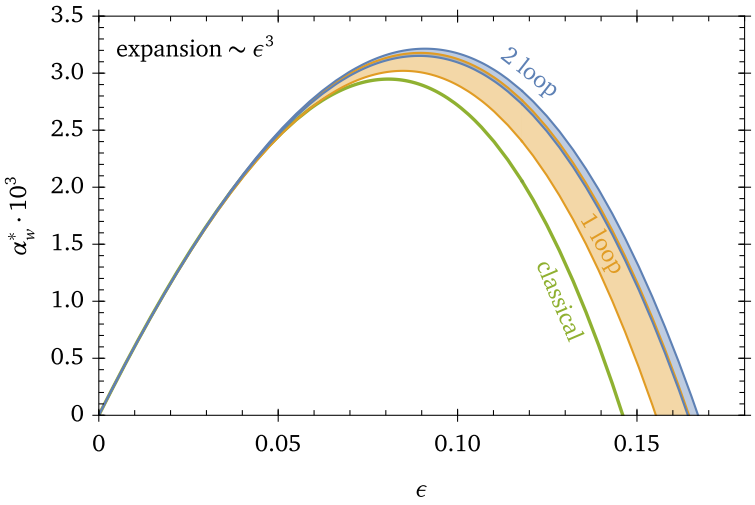


Figure 2. Effective fixed-point coupling α_w^* expanded up to $\propto \epsilon^3$, using only the classical (green), one- and two-loop approximation in the uncertainty range (4.7) (yellow and blue bands).

The improvement of the fixed-point potential by quantum corrections in the ϵ expansion

is depicted in Fig. 2. Classical and improved potential differ notably starting around $\epsilon \approx 0.05$. Consequentially, the classical stability estimate $\epsilon^{\text{exp}}_{\text{vac,cl}} \approx 0.146$ is lifted decisively to

$$\epsilon^{\text{exp}}_{\text{vac}} \approx 0.166(1) \tag{4.10}$$

at two loops. Both results are more stringent than the estimate of the single-trace merger $\epsilon_{\text{mer},1} \approx 0.249$ [26]. The width of the uncertainty band of the effective potential is decreases from one to two loops. However, their overlap diminishes for larger ϵ, slightly hinting at an inadequacy of the ϵ expansion at such values.

Finally, we note that an imaginary part of α_w is not manifest in the ϵ expansion. It would be heralded by $\tilde{\alpha}_I < 0$ (3.35). However, we find that our resummation prescription

$$\begin{aligned}\tfrac{1}{2}\tilde{\alpha}_I \approx{}& 0.0625\,\epsilon \\ &+\left(0.0166 - 0.2438\,\ln\frac{\epsilon}{z_0}\right)\epsilon^2 \\ &+\left(0.0519 - 0.3509\,\ln\frac{\epsilon}{z_0} - 0.0513\,\ln^2\frac{\epsilon}{z_0}\right)\epsilon^3 + \mathcal{O}(\epsilon^4),\end{aligned} \tag{4.11}$$

is manifestly positive in the conformal expansion within the z_0-range (4.7).

4.3 Padé Resummations

While the strict ϵ expansion is reliable for small ϵ and consistent with respect to higher-loop corrections, its convergence is rather poor. To remedy this, we employ Padé resummations in ϵ, serving as an estimate for the impact of higher powers in the ϵ expansion. There are two approaches to this strategy. For one, the resummation may be applied to the fixed-point couplings $\alpha^*_{i=g,y,u,v} = \sum_{n=1}^{3} c_n^{(i)}\,\epsilon^n$. The relevant Padé approximations read

$$[2/1]: \quad \alpha^* \approx c_1\,\epsilon + \frac{c_2^2\,\epsilon^2}{c_2 - c_3\,\epsilon}, \qquad [1/2]: \quad \alpha^* \approx \frac{c_1^3\,\epsilon}{c_1^2 + c_2^2\,\epsilon^2 - c_1\epsilon(c_2 + c_3\,\epsilon)}. \tag{4.12}$$

The second ansatz is to apply the $[2/1]$ Padé resummation directly to $\alpha^*_w(\epsilon)$. The $[1/2]$ approximant is not helpful as it remains stable by construction. Note that both approaches are identical if only the classical potential is considered.

The results are summarised in Fig. 3. As a reference, the $[2/1]$ of the classical potential with $\epsilon^{[2/1]}_{\text{vac,cl}} \approx 0.0868$ is shown in green. The corresponding resummation of the effective potential α^*_w is shown in red. Unfortunately, the prediction for the loss of vacuum stability appears to be under poor perturbative control since the two-loop prediction 0.1233(5) (darker shades) does not overlap with the one-loop result 0.115(5) (lighter shades) and should be disregarded. Note that the $[2/1]$ approximation is depicted by dashed lines in Fig. 1, highlighting that all resummations appear to predict a tighter conformal window than the ϵ expansion.

Returning to Fig. 3, perturbative reliablility is slightly better when the Padé resummation is applied to the couplings $\alpha^*_{g,y,u,v}$, where two-loop effective potentials are approximately contained within the one-loop uncertainties for most values of ϵ. Using a $[2/1]$ resummation (yellow), the two-loop effective potential falls out of perturbative control around $\epsilon \approx 0.1$ and

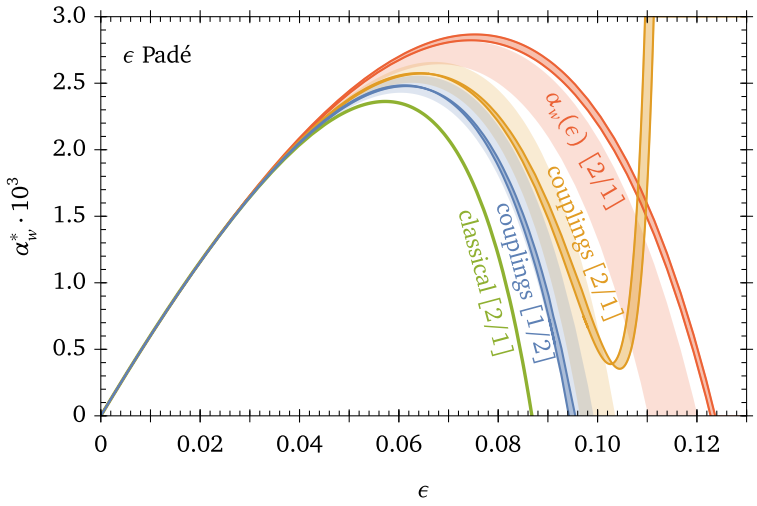


Figure 3. Padé resummed fixed-point potential α_w^* in various approximations. This includes the [2/1] approximants of the conformal expansion for the classical (green) and quantum effective potential (red), as well as the [2/1] (yellow) and [1/2] (blue) resummations of couplings $\alpha^*_{g,y,u,v}(\epsilon)$ as they contribute to α_w^*. The z_0 uncertainty range (4.7) for one- and two-loop effective potentials are displayed as shaded regions.

never enters an unstable regime. On the other hand, the [1/2] approximant (blue) yields a two-loop prediction

$$\epsilon_{\text{vac}}^{\text{Padé}} \approx 0.0949(6) \tag{4.13}$$

that is completely within the one-loop uncertainty. Note that this is larger than the single-trace merger $\epsilon_{\text{mer},1}^{[2/1]} \approx 0.091$ obtained by the [3/1] resummation of ϑ_1 [26].

4.4 Resummation via β Functions

Another method to estimate higher powers in the ϵ expansion is to utilise the fixed-point conditions $\beta_{g,y,u,v} = 0$ to fix the couplings $\alpha^*_{g,y,u,v}$ exactly. As this approach is independent of the ϵ expansion, it is more sensitive to phenomena beyond strict perturbative control. In this section, we will investigate these conditions at orders `lmn`, meaning `l` loops in the gauge, `m` in the Yukawa as well as `n` loops in the quartic β functions. In order to expand fixed-point couplings up to $\propto \epsilon^{\texttt{n-1}}$ requires at least the loop combinations `(n+1)nn` and needs to be paired with the `(n-1)`-loop effective potential.

Before discussing scalar stability, we recollect the findings of [24, 27] regarding fixed-point mergers. Fig. 4 displays all fixed-point couplings $\alpha^*_{g,y,u,v}(\epsilon)$ both for 433 (dashed lines) and the previous order in the conformal expansion 322 (solid lines). The UV fixed point is drawn in blue and its double-trace counterpart in green; both show good agreement between 433 and 322 for small ϵ. At 322, however, the UV fixed point collides with an IR fixed point outside of strict perturbative control (red) via a single-trace merger at $\epsilon_{\text{mer},1} \approx 0.0972$. On the other hand, the conformal window at 433 persists up to a double-trace merger at $\epsilon_{\text{mer},2} \approx 0.152$.

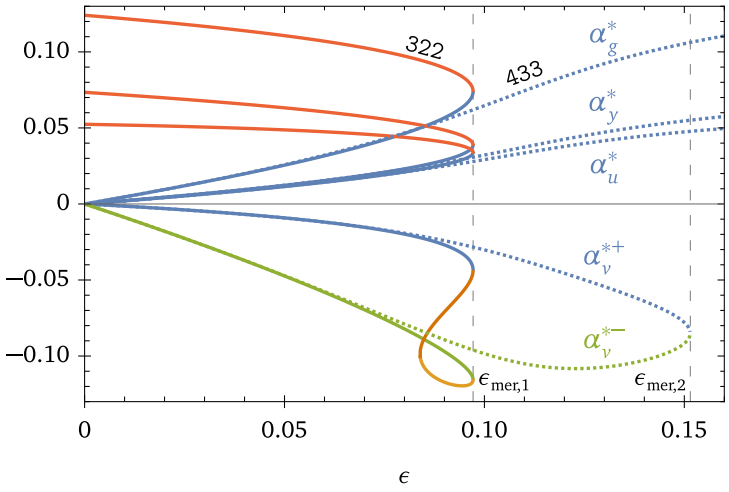


Figure 4. Fixed-point couplings $\alpha^*_{g,y,u,v}(\epsilon)$ in the 322 (solid) and 433 (dashed) approximation, showing the UV fixed point (blue), its second double-trace quartic (green), and other solutions (red, yellow, brown). The single-trace ($\epsilon_{\mathrm{mer},1}$) and double-trace merger ($\epsilon_{\mathrm{mer},2}$) of 322 and 433, respectively, are marked by dashed vertical lines.

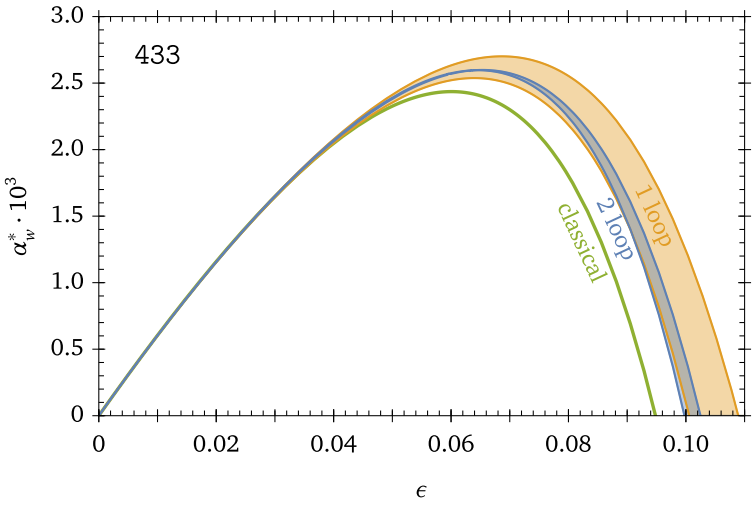


Figure 5. Fixed-point potential in the 433 approximation at tree- (green), one-loop (yellow) and two-loops (blue) level, with the z_0 uncertainty range (4.7) (shaded).

We now turn towards the stability analysis, where we will find tighter constraints than from the mergers. The tree-level and effective potential in the 433 approximation are displayed in Fig. 5. The two-loop uncertainty is much smaller and approximately contained within the one-loop one. It reads

$$\epsilon_{\mathrm{vac}}^{433} \approx 0.1011(13)\,, \tag{4.14}$$

which is a significant improvement over the tree-level result quoted in [27]. In fact, the tree-level result is not even contained within the uncertainty range at one-loop.

Approx.	Tree level	1 loop	2 loop	Merger
433	0.0948	0.1048(42)	0.1011(13)	$\epsilon_{\text{mer},2} \approx 0.1515$
432	0.0883	0.0948(25)	0.0929(5)	$\epsilon_{\text{mer},2} \approx 0.1364$
423	0.1284	0.169(23)	0.1451(58)	$\epsilon_{\text{mer},1} \approx 0.4223$
322	0.0867	0.0913(16)	0.090(4)	$\epsilon_{\text{mer},1} \approx 0.0972$

Table 2. Bound on the UV conformal window at loop order `lmn`, from vacuum instability (ϵ_{vac}), using the tree-level, one- and two loop effective potential as well as the merger. The uncertainty is estimated using (4.7).

Furthermore, a range of relevant loop combinations for $\beta_{g,y,u,v}$ are collected in Tab. 2. In each case, the merger is only subdominant for the end of the conformal window. Overall, the same picture as at 433 emerges: each approximation finds that the one- and two-loop potentials are within a perturbative agreement and admit roughly a 10% widening of the conformal window with respect to the classical potential. In this sense, the resummation using β functions appears to be more consistent than Padé approximants.

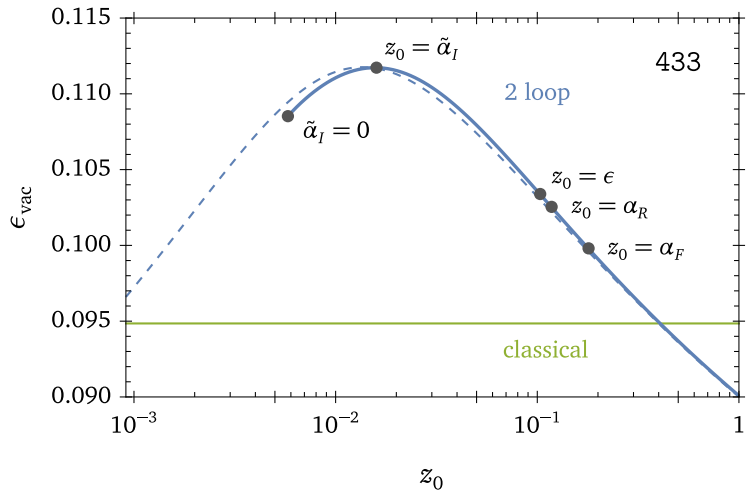


Figure 6. Upper bound for the UV conformal window ϵ_{vac} at 433 and its z_0 dependence. The green line utilises only the classical stability condition. The solid blue curve takes into account the two-loop effective potential, and two-loop softening of $\tilde{\alpha}_I$ (3.35). The curve is only drawn for values of z_0 where $\text{Im}\,\alpha_w(\epsilon_{\text{vac}}) \neq 0$ as $\text{Re}\,\alpha_w(\epsilon_{\text{vac}}) = 0$. The dashed blue line does not employ the soft resummation (3.35). It shows the bound from $\text{Re}\,\alpha_w(\epsilon_{\text{vac}})$ ignoring the imaginary part of α_w entirely. Several points highlight unique choices of z_0, the left and rightmost represent the ones taken into account for the uncertainty estimate (4.7). Within this range, the prediction lies well above the classical one.

Next we scrutinise the precision and reliability of our 433 prediction. To that end, we study the z_0-dependence of the stability prediction in more detail in Fig. 6 (blue curve) and contrast with the classical bound (green). The impact of the soft resummation in $\tilde{\alpha}_I$ is displayed by comparing the resummed potential (solid) with the real part of the unresummed one (dashed), ignoring its imaginary part. We observe that both agree well within the range

(4.7). We therefore expect that a different resummation prescription in the definition of $\tilde{\alpha}_I$ would have a similarly small impact. Conversely, if z_0 inches closer towards $\tilde{\alpha}_I$, the difference between the resummed and unresummed potential increases. Also, the prediction for ϵ_{vac} increases significantly, showcasing the numerical inaccuracies for z_0 outside the range (4.7) as argued before. Naturally, at $z_0 \approx \tilde{\alpha}_I$ the prediction ϵ_{vac} is very sensitive to the definition of $\tilde{\alpha}_I$. As z_0 is decreased even further, we eventually see the resummation prescription fail as the effective potential gains an imaginary part through $\tilde{\alpha}_I < 0$ while the real part stays positive.

5 Discussion

In this work, the prediction for the scalar stability of the UV fixed point in the Litim-Sannino model has been significantly enhanced using loop corrections, renormalisation-group improvement, resummation of soft corrections and parameter optimisations. In particular, the conformal window has been estimated using the conformal expansion (4.10) as well as resummations based on Padé approximants (4.13) and β functions (4.14). In each of these approximations, we find a significant widening of the conformal window with respect to previous works [26, 27]. As a result, smaller integer values of (N_c, N_f) may be sufficient for a UV fixed point.

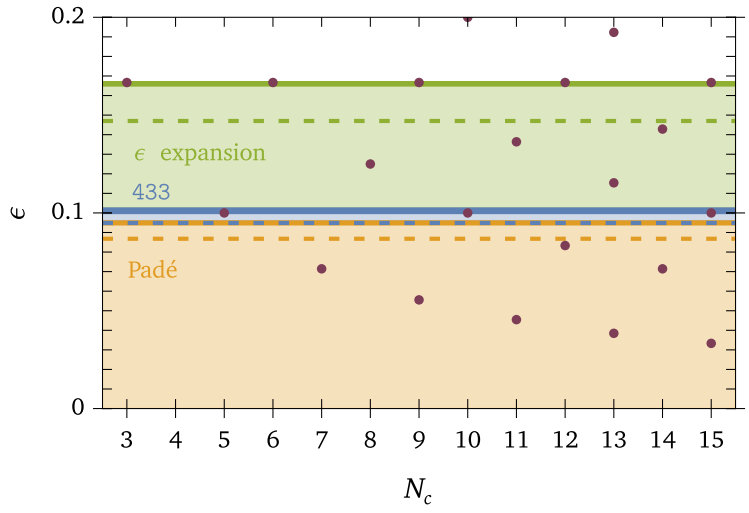


Figure 7. Conformal-window estimates from vacuum-instability estimates using the classical (dashed) or two-loop quantum improved potential (solid) in the conformal expansion (green), Padé resummation (yellow) and 433 approximation. Dots mark integer choices of $N_{f,c}$ closest to the conformal window.

This is displayed in Fig. 7. Each dot represents an integer value of $N_{f,c}$, neglecting finite-N corrections. The extend of the conformal window for each approximation is marked in different colours. The difference between dashed and solid lines mark the improvement due to quantum corrections of the effective potential. It is only due to this enhancement that the conformal expansion points towards $(N_c, N_f) = (3, 17)$ as the smallest multiplicity within

the window. Similarly, the 433 approximation admits (5, 28) into the window. Neither of these theories were considered viable in [26, 27].

In our analysis, we have verified that uncertainties in the stability predictions overall decrease with higher loops in the effective potential. Moreover, we have used the consecutive nestedness of these regions as sanity criterion for our approximations. In light of this, we find soft hints that the convergence of the conformal expansion is too weak for a reliable prediction of vacuum instability at the current order. In general, the conformal-window predictions from ϵ expansions and resummation techniques do not agree at the orders available to us. This is because of unknown higher-loop corrections to the β functions, which also represent the largest source of uncertainty. More so, this uncertainty prohibits any conclusion about the upper bound of the conformal window. In particular, a fixed-point merger is still in contest with vacuum stability as the mechanism behind the breakdown of conformality. Higher loops or non-perturbative arguments are required to settle this question.

Acknowledgement

The author is indebted to Emmanuel Stamou for valuable discussions and comments on the manuscript.

References

[1] D.J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343.

[2] H.D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346.

[3] D.F. Litim and F. Sannino, *Asymptotic safety guaranteed*, *JHEP* **12** (2014) 178 [1406.2337].

[4] A.D. Bond, D.F. Litim and T. Steudtner, *Asymptotic safety with Majorana fermions and new large N equivalences*, *Phys. Rev. D* **101** (2020) 045006 [1911.11168].

[5] F. Sannino and I.M. Shoemaker, *Asymptotically Safe Dark Matter*, *Phys. Rev. D* **92** (2015) 043518 [1412.8034].

[6] D.F. Litim, M. Mojaza and F. Sannino, *Vacuum stability of asymptotically safe gauge-Yukawa theories*, *JHEP* **01** (2016) 081 [1501.03061].

[7] N.G. Nielsen, F. Sannino and O. Svendsen, *Inflation from Asymptotically Safe Theories*, *Phys. Rev. D* **91** (2015) 103521 [1503.00702].

[8] D.H. Rischke and F. Sannino, *Thermodynamics of asymptotically safe theories*, *Phys. Rev. D* **92** (2015) 065014 [1505.07828].

[9] A. Codello, K. Langæble, D.F. Litim and F. Sannino, *Conformal Gauge-Yukawa Theories away From Four Dimensions*, *JHEP* **07** (2016) 118 [1603.03462].

[10] A.D. Bond, G. Hiller, K. Kowalska and D.F. Litim, *Directions for model building from asymptotic safety*, *JHEP* **08** (2017) 004 [1702.01727].

[11] N.A. Dondi, V. Prochazka and F. Sannino, *Conformal Data of Fundamental Gauge-Yukawa Theories*, *Phys. Rev. D* **98** (2018) 045002 [1712.05388].

[12] A.D. Bond and D.F. Litim, *More asymptotic safety guaranteed*, *Phys. Rev. D* **97** (2018) 085008 [1707.04217].

[13] A.D. Bond and D.F. Litim, *Asymptotic safety guaranteed in supersymmetry*, *Phys. Rev. Lett.* **119** (2017) 211601 [1709.06953].

[14] S. Abel and F. Sannino, *Radiative symmetry breaking from interacting UV fixed points*, *Phys. Rev. D* **96** (2017) 056028 [1704.00700].

[15] D. Barducci, M. Fabbrichesi, C.M. Nieto, R. Percacci and V. Skrinjar, *In search of a UV completion of the standard model — 378,000 models that don't work*, *JHEP* **11** (2018) 057 [1807.05584].

[16] G. Hiller, C. Hormigos-Feliu, D.F. Litim and T. Steudtner, *Anomalous magnetic moments from asymptotic safety*, 1910.14062.

[17] G. Hiller, C. Hormigos-Feliu, D.F. Litim and T. Steudtner, *Asymptotically safe extensions of the Standard Model with flavour phenomenology*, in *54th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 415–418, 2019 [1905.11020].

[18] G. Hiller, C. Hormigos-Feliu, D.F. Litim and T. Steudtner, *Model Building from Asymptotic Safety with Higgs and Flavor Portals*, *Phys. Rev. D* **102** (2020) 095023 [2008.08606].

[19] S. Bißmann, G. Hiller, C. Hormigos-Feliu and D.F. Litim, *Multi-lepton signatures of vector-like leptons with flavor*, *Eur. Phys. J. C* **81** (2021) 101 [2011.12964].

[20] R. Bause, G. Hiller, T. Höhne, D.F. Litim and T. Steudtner, *B-anomalies from flavorful U(1)′ extensions, safely*, *Eur. Phys. J. C* **82** (2022) 42 [2109.06201].

[21] A.D. Bond and D.F. Litim, *Asymptotic safety guaranteed for strongly coupled gauge theories*, *Phys. Rev. D* **105** (2022) 105005 [2202.08223].

[22] G. Hiller, D.F. Litim and K. Moch, *Fixed Points in Supersymmetric Extensions of the Standard Model*, 2202.01264.

[23] G. Hiller, T. Höhne, D.F. Litim and T. Steudtner, *Portals into Higgs vacuum stability*, *Phys. Rev. D* **106** (2022) 115004 [2207.07737].

[24] A.D. Bond, D.F. Litim, G. Medina Vazquez and T. Steudtner, *UV conformal window for asymptotic safety*, *Phys. Rev. D* **97** (2018) 036019 [1710.07615].

[25] A.D. Bond, D.F. Litim and G.M. Vazquez, *Conformal windows beyond asymptotic freedom*, *Phys. Rev. D* **104** (2021) 105002 [2107.13020].

[26] D.F. Litim, N. Riyaz, E. Stamou and T. Steudtner, *Asymptotic Safety Guaranteed at Four Loop*, 2307.08747.

[27] A.V. Bednyakov and A.I. Mukhaeva, *On the asymptotic safety in the Litim-Sannino model at four loops*, 2312.12128.

[28] A. Paterson, *Coleman-Weinberg symmetry breaking in the chiral $SU(N) \times SU(N)$ linear σ model*, *Nucl. Phys. B* **190** (1981) 188.

[29] S.R. Coleman and E.J. Weinberg, *Radiative Corrections as the Origin of Spontaneous Symmetry Breaking*, *Phys. Rev. D* **7** (1973) 1888.

[30] R. Jackiw, *Functional evaluation of the effective potential*, *Phys. Rev. D* **9** (1974) 1686.

[31] M. Sher, *Electroweak Higgs Potentials and Vacuum Stability*, *Phys. Rept.* **179** (1989) 273.

[32] S.P. Martin, *Two Loop Effective Potential for a General Renormalizable Theory and Softly Broken Supersymmetry*, *Phys. Rev. D* **65** (2002) 116003 [hep-ph/0111209].

[33] S.P. Martin, *Effective potential at three loops*, *Phys. Rev. D* **96** (2017) 096005 [1709.02397].

[34] S.P. Martin and H.H. Patel, *Two-loop effective potential for generalized gauge fixing*, *Phys. Rev. D* **98** (2018) 076008 [1808.07615].

[35] B.M. Kastening, *Renormalization group improvement of the effective potential in massive phi**4 theory*, *Phys. Lett. B* **283** (1992) 287.

[36] C. Ford, D.R.T. Jones, P.W. Stephenson and M.B. Einhorn, *The Effective potential and the renormalization group*, *Nucl. Phys. B* **395** (1993) 17 [hep-lat/9210033].

[37] M. Bando, T. Kugo, N. Maekawa and H. Nakano, *Improving the effective potential*, *Phys. Lett. B* **301** (1993) 83 [hep-ph/9210228].

[38] M. Bando, T. Kugo, N. Maekawa and H. Nakano, *Improving the effective potential: Multimass scale case*, *Prog. Theor. Phys.* **90** (1993) 405 [hep-ph/9210229].

[39] A. Andreassen, W. Frost and M.D. Schwartz, *Consistent Use of Effective Potentials*, *Phys. Rev. D* **91** (2015) 016009 [1408.0287].

[40] A. Andreassen, W. Frost and M.D. Schwartz, *Consistent Use of the Standard Model Effective Potential*, *Phys. Rev. Lett.* **113** (2014) 241801 [1408.0292].

[41] N.K. Nielsen, *On the Gauge Dependence of Spontaneous Symmetry Breaking in Gauge Theories*, *Nucl. Phys. B* **101** (1975) 173.

[42] R. Fukuda and T. Kugo, *Gauge Invariance in the Effective Action and Potential*, *Phys. Rev. D* **13** (1976) 3469.

[43] I.J.R. Aitchison and C.M. Fraser, *Gauge Invariance and the Effective Potential*, *Annals Phys.* **156** (1984) 1.

[44] G. Veneziano, *Some Aspects of a Unified Approach to Gauge, Dual and Gribov Theories*, *Nucl. Phys. B* **117** (1976) 519.

[45] G. 't Hooft, *A Planar Diagram Theory for Strong Interactions*, *Nucl. Phys. B* **72** (1974) 461.

[46] R. Fukuda and E. Kyriakopoulos, *Derivation of the Effective Potential*, *Nucl. Phys. B* **85** (1975) 354.

[47] S.Y. Lee and A.M. Sciaccaluga, *Evaluation of Higher Order Effective Potentials with Dimensional Regularization*, *Nucl. Phys. B* **96** (1975) 435.

[48] J. Brod, L. Hüdepohl, E. Stamou and T. Steudtner, *MaRTIn – Manual for the "Massive Recursive Tensor Integration"*, 2401.04033.

[49] A.I. Davydychev and J.B. Tausk, *Two loop selfenergy diagrams with different masses and the momentum expansion*, *Nucl. Phys. B* **397** (1993) 123.

[50] Y. Fujimoto, L. O'Raifeartaigh and G. Parravicini, *Effective Potential for Nonconvex Potentials*, *Nucl. Phys. B* **212** (1983) 268.

[51] E.J. Weinberg and A.-q. Wu, *Understanding Complex Perturbative Effective Potentials*, *Phys. Rev. D* **36** (1987) 2474.

[52] S.P. Martin, *Taming the Goldstone contributions to the effective potential*, *Phys. Rev. D* **90** (2014) 016013 [1406.2355].

[53] J. Elias-Miro, J.R. Espinosa and T. Konstandin, *Taming Infrared Divergences in the Effective Potential*, *JHEP* **08** (2014) 034 [1406.2652].

[54] N. Kumar and S.P. Martin, *Resummation of Goldstone boson contributions to the MSSM effective potential*, *Phys. Rev. D* **94** (2016) 014013 [1605.02059].

[55] J. Braathen and M.D. Goodsell, *Avoiding the Goldstone Boson Catastrophe in general renormalisable field theories at two loops*, *JHEP* **12** (2016) 056 [1609.06977].

[56] D.B. Kaplan, J.-W. Lee, D.T. Son and M.A. Stephanov, *Conformality lost*, *Phys. Rev. D* **80** (2009) 125005 [0905.4752].

[57] E. Pomoni and L. Rastelli, *Large N Field Theory and AdS Tachyons*, *JHEP* **04** (2009) 020 [0805.2261].

arXiv:2402.16859v1 [cond-mat.mes-hall] 19 Jan 2024

Negative magnetoresistance induced by longitudinal photons in Dirac/Weyl semimetals.

J. L. Acosta Avalo* and H. Pérez Rojas** ,
*Instituto Superior de Tecnologías y Ciencias Aplicadas (INSTEC). Universidad de la Habana, Ave Salvador Allende, No. 1110, Vedado, La Habana, 10400 Cuba. and
**Instituto de Cibernética, Matemática y Física (ICIMAF), Calle E esq 15, No. 309, Vedado, La Habana, 10400 Cuba.

A low-energy model is built to study systems such as Dirac/Weyl semimetals, according to statistical quantum electrodynamics formalism. We report that the introduction of a pseudoscalar, associated to longitudinal photons propagating along a magnetic field **B**, could transforms a Dirac semimetal into a Weyl semimetal with a pair of Weyl nodes for each point of Dirac. The nodes are separated by a pseudovector electric field induced dynamically along **B** associated to a chiral effect on the Fermi surface. A topological quantum transition is produced between a chiral-and non chiral symmetry phase. A general expression to the longitudinal magnetoconductivity is found. It provides the possibility of generalizing the usual expressions of the magnetoconductivity reported in the literature. This has a quadratic dependence on **B**, which is associated with a positive contribution to the magnetoconductivity. This is a prominent signature of the chiral magnetic effect in Dirac/Weyl systems in parallel electric and magnetic fields. We report a chiral effect induced by longitudinal photons associated to a negative longitudinal magnetoresistance in Dirac systems via an axial anomaly relation. We show some numerical results, and reproduced with a high level of accuracy some of the experimental results, in the low temperature region, obtained to the magnetoresistance of $ZrTe_5$ and Na_3Bi. We believe that a wide variety of these semimetals can be studied by using our general expression to the negative longitudinal magnetoresistance.

Keywords: Dirac/Weyl semimetals−negative longitudinal magnetoresistance−chiral symmetry breaking−statistical quantum electrodynamics

I. INTRODUCTION

Nowadays the influence of the magnetic fields in relativistic quantum systems, like the three-dimensional $(3+1)$ Dirac semimetals: cadmium arsenide (Cd_3As_2), trisodium bismuthide (Na_3Bi)[1–4] and zirconium pentatelluride ($ZrTe_5$)[5] has enabled experimental studies of the quantum dynamics of relativistic field theory in condensed matter systems. The relativistic theory of charged chiral fermions in three spatial dimensions possesses the so-called chiral anomaly (discussed by Adler-Bell-Jackiw[6,7]), that is, non-conservation of chiral charge induced by gauge fields with non-trivial topology, for example, by parallel electric and magnetic fields. The existence of chiral quasiparticles in Dirac and Weyl semimetals opens the possibility to observe the effects of the chiral anomaly in magnetized low-energy systems.

Weyl and Dirac semimetals are three dimensional phases of matter with gapless electronic excitations that are protected by topology and symmetry. As the three dimensional analogs of graphene the semimetals have low-energy quasiparticles near the Fermi surface, which are described by the Dirac and Weyl equations respectively[8–10]. The Bismuth is an example of a semimetal whose effective theory at low-energies includes Dirac fermions in $3+1$ dimensions[11]. Their electronic states in the vicinity of the Weyl nodes have a definite chirality, which is found associated with unique topological and electromagnetic properties.

In the absence of any symmetry one could obtain accidental two-fold degeneracies of bands in a three dimensional solid[12]. The dispersion in the vicinity of these band touching points is generically linear and is given by the Weyl equation. Perhaps, the best known example is the graphene, where a linear dispersion relation is obtained by the two dimensional massless Dirac equation. These two-dimensional $(2+1)$ carbon sheets provide a condensed matter analogue of a $2+1$-dimensional quantum electrodynamics[13]. Remarkably several of the defining physical properties of Weyl fermions, such as the so-called chiral anomaly, continue to hold in this non relativistic condensed matter context. The Adler-Bell-Jackiw anomaly can have nontrivial effects as pointed out in[14], which established the link between band touchings in three dimensional crystals, named Weyl nodes[8], and chiral fermions.

On the other hand the discovery of topological insulators in two and three dimensions[15–25], has led to an explosion of activity in the study of topological aspects of band structures[26–28]. These have interesting connections to gapless states. The surfaces of topological insulators in $3+1$ feature a gapless Dirac dispersion, analogous to the two dimensional semimetal graphene, but with important differences in the number of nodal points. The transition between topological and trivial phases proceeds through a gapless state. For example a $3+1$ topological to trivial insulator transition proceeds through the $3+1$ Dirac dispersion, in the presence of both time reversal and inversion symmetry[29].

In this paper we will mainly be concerned in the three dimensional crystals with linear dispersion fermionic excitations, which are describe by the massless $3+1$ Weyl and Dirac equations. These condensed matter phenomena offer a scenario where predictions made by relativistic theories may be found, showing also new properties to appear only in the condensed matter context.

One of the most interesting properties of the Dirac and Weyl semimetals discussed in the literature is related to transport phenomena, such as the chiral magnetic effect via axial anomalies. In[30] it was shown that a magnetic field in presence of imbalanced chirality induces a current along the magnetic field. This is so-called Chiral Magnetic Effect. This effect have already been observed in $ZrTe_5$[5], Na_3Bi[31], Cd_3As_2[32] and $Bi_{1-x}Sb_x$ at $x \approx 0.03$[33]. In several papers related to this effect the chirality is introduced through a nonvanishing chiral chemical potential factor[34,35]. In this sense, our paper is closer to the approach of[36–38], in which only the electromagnetic chemical potential associated to the (conserved) electromagnetic charge is used. However, to obtain a full understanding of the chiral magnetic effect, it is desirable to include the dynamics leading to a net chirality. In our paper, the pseudovector longitudinal mode, propagating along an external magnetic field $\mathbf{B}$, is associated to the chiral charge density in a magnetized medium[37,38].

The most prominent signature of the chiral magnetic effect in Dirac systems in parallel electric and magnetic fields is a positive contribution to the conductivity that has a quadratic dependence on magnetic field[5,31]. As a result, we report a general expression for the negative longitudinal magnetoresistance induced by longitudinal photons in semimetals. We remark that a general expression associated to the longitudinal negative magnetoresistance was calculated in[38]. It could be applied in the relativistic context and in astrophysical models.

In this paper, we describe the Fermi surface modification in the presence of $\mathbf{B}$ and a pseudovector electric field associated to longitudinal photons (non-trivial topology) in 3+1 condensed matter. We report a topological quantum transition (chiral-to non chiral symmetry system transition) between a Dirac-and Weyl semimetal. The Fermi surface modification implies that a Dirac semimetal is transformed into a Weyl semimetal. The resulting Weyl semimetal has one pair of Weyl nodes for every Dirac point. I will show that each pair of nodes is separated by a dynamically induced pseudovector (axial vector) E_3 associated with longitudinal photons propagating in the direction of $\mathbf{B}$. The results will be independent of the electrical charge symmetry-C. The magnitude of the pseudovector E_3 is determined by the parameters of a small radiation field previously considered.

A low-energy model will be built to study systems such as Dirac/Weyl semimetals. We start of general results related to the propagation of electromagnetic waves, the electromagnetic current vector, and the axial current in a magnetized relativistic fermionic system[37,38], where the low-energy limit is taking into account. Our results are based on the quantum field theory formalism at finite temperature[39] $T \neq 0$ (T is given in energy units) and density $\mu \neq 0$ (the medium is C-non invariant)[40,41]. Our approach could has a great relevance in the study and understanding of the properties of new emerging materials related to Dirac and Weyl semimetals. This study could be applied to a broad range of Weyl/Dirac semimetals what often emerge at quantum transitions between conventional and topological insulators.

In our low-energy model the axial character will refer to the average density in momentum space of those electrons and holes having a common helicity, that is, right-R or left-L-handed helicity. If a perturbative pseudovector electric field $\mathbf{E}$ (associated to a longitudinal pure electric mode) is applied to a $3+1$ Dirac semimetal, in thermodynamic equilibrium in presence of $\mathbf{B}$, such that $\mathbf{E} \parallel \mathbf{B}$, this is transformed (Fermi surface transformation) into a Weyl semimetal (has one pair of Weyl nodes for every Dirac point), and it produces an axial current leading to the breaking of the previously existing statistical chiral balance of the densities of charged particles. The chirality-changing transitions between Weyl nodes is produced. The charges will move according to the relative directions of $\mathbf{E}$ and $\mathbf{B}$. In the case they are parallel, electrons decrease their negative momentum and holes increase their positive momentum, producing a chiral R-current. For the antiparallel case we have the inverse effect, leading to an L-current, where the sign of the pseudoscalar $\mathfrak{G} = \frac{1}{4}\mathbf{E}\cdot\mathbf{B}$ plays a fundamental role. This gives a phenomenological basis to our chirality-changing transitions (chiral current generation) between the Weyl nodes. A quantitative approach is given below.

A. About the chiral current induced by longitudinal photons in a magnetized relativistic medium

In a relativistic medium under the action of an external constant field $\mathbf{B}$ (which we assume as parallel to the z axis), electrons and positrons move in bound states characterized by energy levels given by (we shall use $\hbar = c = 1$ units, except for specific calculations)[37,38]:

$$\varepsilon_{n_l,p_3} = \sqrt{p_3^2 + m^2 + |e|B(2n_l + 1 - sgn(e)s_3)}, \quad (1)$$

where p_3 is the momentum along B, m is the electron mass, $s_3 = \pm 1$ are spin eigenvalues along x_3 and $n_l = 0, 1, ...$ are the Landau quantum numbers. These are two-fold spin degenerate, except the ground state ε_0 in which $n_l = 0$, and for electrons it is $s_3 = -1$ and for positrons $s_3 = 1$. Quantum states are also degenerate with regard to the orbit's center coordinates[42]. Although higher Landau quantum numbers contribute both to paramagnetic and diamagnetic terms[43], we will make here the fundamental assumption that the system is close

to be degenerate and $2eB > \mu^2 - m^2$, where $m^2, eB > T^2$, so that the contribution of the first Landau levels becomes dominant whereas that from higher Landau levels is negligibly small.

In a medium, as different from vacuum, there is a non-vanishing four-velocity vector $u_\mu \neq 0$. We must recall that also in absence of external fields in the charged medium there are three electromagnetic modes, two transverse and one longitudinal[39]. The two transverse modes correspond to photon spin projections ± 1 along its momentum $\mathbf{k}$, their dispersion equations being not on the light cone. The third mode is a pure electrical longitudinal one (zero spin) and its dispersion equation in the infrared limit has a solution for $\omega = 0, \mathbf{k} \neq 0$ which accounts for the Debye mass screening[39], having close formal analogy to the Yukawa force[38].

In a quantum relativistic electron-positron plasma under the action of an external field $\mathbf{B}$ generated by a four-potential A^{ext}_μ, the photon self-energy tensor structure determine the modes of propagation in both the C-invariant and non-invariant cases[40,41,44,45] (see the Appendix A 1), as well as the electric currents. In our magnetized plasma is assumed a ionic positive background for the case $\mu \neq 0$ to guarantee total charge neutrality, although other effects can be neglected. The presence of $\mathbf{B}$ breaks the spatial symmetry, leaving invariances for rotations around $\mathbf{B}$, and for space translations along it. This determines different eigenmodes according to the direction of propagation and the linear in the electric field expression of the currents. The dispersion equations for photons propagating in the medium can be solved in any direction. For instance, for the case of propagation parallel to $\mathbf{B}$ (in which we are interested)[41], it was described the chiral magnetic effect associated to longitudinal photons[38] while for transverse modes it was obtained the relativistic Hall conductivity as well as the Faraday Effect[46,47].

The Schwinger-Dyson equation for the photon in Fourier space, with $\nu = 1, 2, 3, 4$, is[39]:

$$[k^2 g_{\mu\nu} - \Pi_{\mu\nu}(k|A^{ext}_\mu)]a^\nu(k) = 0, \tag{2}$$

where $g_{\mu\nu}$ is the Minkowski metric, in the form $g_{\mu\nu} = (1, 1, 1, -1)$ (it corresponds to the analytic continuation $k_4 \to i\omega$ from Euclidean metric), and $k^2 = k_3^2 + k_\perp^2 - \omega^2$. Here k_3 and $k_\perp$ are respectively the components of the photon four-momentum in directions parallel and perpendicular to $\mathbf{B}$, and ω its energy. The total external electromagnetic field is $A^{ext}_\mu + a_\mu$, where a_μ is a small perturbative radiation field (its electric field $E \ll B$).

The quantum corrections are given by the photon self-energy tensor $\Pi_{\mu\nu}(k|A^{ext}_\mu)$, which was calculated in magnetized vacuum and in the one-loop approximation in[45], and in a magnetized medium in[40,41,44]. According to[40,45], the diagonalization of the photon self-energy tensor in vacuum and in a medium leads to the equation: $\Pi_{\mu\nu} b^{\nu(i)} = \eta_i b^{(i)}_\mu$, having three non-vanishing eigenvalues η_i and three eigenvectors $b^{(i)}_\mu$ for $i = 1, 2, 3$, (the photon four-vector k_ν has zero eigenvalue) corresponding to three photon propagation modes (see (A8) in Appendix A 1). For each mode it is obtained a dispersion law $k^2 = \eta_i(k_3, k_\perp, \omega, B)$[40,41,44,45]. In addition to the two transverse modes, there is a longitudinally polarized mode along $\mathbf{B}$ given by the pseudovector: $b^{(2)}_\mu(k) = a c^{(2)}_\mu$[37,38]. Here $c^{(2)}_\mu = R_2(F^* k)_\mu$ is a normalized pseudovector (the normalization parameter is $R_2 = 1/B z_1^{1/2}$, see (A4)), and $F^*_{\mu\nu}$ is the dual of the electromagnetic field tensor $F_{\mu\nu}$. The parameter a (which has dimension of vector potential) is determined by the applied perturbative electric field. Its electric polarization vector (see (A9) in Appendix A 1) being in the direction along $\mathbf{B}$[45]

$$\mathbf{E}_B = E^{(2)}\mathbf{e}_B = a(k_3^2 - \omega^2)^{\frac{1}{2}}\mathbf{e}_B, \tag{3}$$

where $\mathbf{e}_B = \mathbf{B}/B$ is a unit pseudovector. As pointed out earlier, the longitudinal mode is not on the light cone, that is $k_3^2 - \omega^2 \neq 0$[45]. From now on we will call $z_1 = k_3^2 - \omega^2$.

The electromagnetic current as a function of $A^{ext}_\mu + a_\mu$ depends on the two relativistic invariants: $\mathfrak{F} =\simeq \frac{1}{2}B^2$ and $\mathfrak{G} = \mathbf{B} \cdot \mathbf{E}$. Notice that for the case of propagation along $\mathbf{B}$, the pseudoscalar $\mathfrak{G} \neq 0$ only for the longitudinal mode $b^{(2)}_\mu$, independently of the C-symmetry of the system. An expansion of the electromagnetic current density in functional series of a_ν gives:

$$j_\mu(A^{ext}_\mu + a_\mu) = j_\mu(A^{ext}_\mu) + \frac{\delta j_\mu}{\delta A^{ext}_\nu} a_\nu + ... \ , \tag{4}$$

its linear term in a_ν is[46,47]

$$j_i = \Pi_{i\nu} a^\nu = Y_{ij} E_j, \tag{5}$$

where $E_j = i(\omega a_j - k_j a_0)$ is the electric field, with $a_4 = i a_0$ and $k_4 = i\omega$, also $j_\mu(A^{ext}_\mu) = N_0 \delta_{\mu 4}$, with N_0 the net density of charged particles in the ground state. The term $Y_{ij} = \Pi_{ij}/i\omega$ is the complex conductivity. The third term in (5) comes from the second one by using the four-dimensional transversality of $\Pi_{\mu\nu}$ due to gauge invariance, $\Pi_{\mu\nu} k^\nu = 0$[40,41,44,45]. In (4), a_μ is in general a linear function of the eigenmodes $b^{(i)}_\mu$. Below we particularize to the case in which the eigenvector $a_\mu = b^{(2)}_\mu$, for which the electric field vector is parallel to $\mathbf{B}$ (notice that only terms contaning odd number of $b^{(2)}_\mu$ legs in (4) lead to pseudovector terms).

Now we must observe that in the linear-in-E_i approximation of j_i (5), and from the eigenvalue equation of $\Pi_{\mu\nu}$, one gets also[37,38]:

$$j_i = \Pi_{i\nu} a^\nu = s b^{(2)}_i, \tag{6}$$

where we can write the scalar $s = c^{(2)\mu}\Pi^{\nu}_{\mu}c^{(2)}_{\nu}$, which is the eigenvalue of the photon self-energy tensor corresponding to the longitudinal mode (see (A5) in Appendix A 1). The scalar s in the one-loop approximation is[38,40,41,44,45]:

$$s = -\frac{e^3B}{\pi^2}\sum_{n=0}^{\infty}\int_{-\infty}^{\infty}\frac{dp_3}{\varepsilon_q}[\alpha_n\varepsilon^2_{n,0}(2p_3k_3+z_1)] \times \frac{[n^p(\varepsilon_q)+n^e(\varepsilon_q)-1]}{4z_1p_3(p_3+k_3)+z_1^2-4\omega^2\varepsilon^2_{n,0}}, \quad (7)$$

where $n^{e,p} = [1+e^{(\varepsilon_q\mp\mu)/T}]^{-1}$ are the electron and positron densities in momentum space, $\varepsilon_{n,0} = \sqrt{m^2+|e|Bn}$, with $n = 2n_l+1+sgn(e)s_3$, $\alpha_n = 2-\delta_{n,0}$ and $q = (n, p_3)$. Here the -1 inside the square brackets accounts for the quantum vacuum limit ($\mu = T = 0$). From the expression for the imaginary part of the scalar s (calculated in one-loop approximation) obtained in[41] (details of these calculations are given in Appendix A 2), we have

$$Im[s] = -\frac{e^3B}{2\pi}\sum_{n=0}^{\infty}\alpha_n\varepsilon^2_{n,0}S_n, \quad (8)$$

here

$$S_n = \frac{\Delta N+\theta\Delta H}{\Lambda}, \quad (9)$$

where $\theta = \theta(-4\varepsilon^2_{n,0}-z_1)$ is the Heaviside step function, and $\Delta N = [N(\varepsilon_r)-N(\varepsilon_r+\omega)]$, $\Delta H = [H(-\varepsilon_s)+H(\omega+\varepsilon_s)-2]$, where the term $N = n^e(\varepsilon_r)+n^p(\varepsilon_r)$ while the term $H = n^e(\varepsilon_s)+n^p(\omega-\varepsilon_s)$. Here $\Lambda = \sqrt{z_1(z_1+4\varepsilon^2_{n,0})}$, and $\varepsilon_s = (\omega z_1+|k_3|\Lambda)/2z_1$, $\varepsilon_r = (-\omega z_1+|k_3|\Lambda)/2z_1$ with $r, s = (n, \omega, k_3)$ are the fermion energies in a magnetic field in terms of the longitudinal mode energy ω and momentum k_3. The term ΔN accounts for the excitation of particles $[\varepsilon(p_3, n) \longrightarrow \varepsilon(p_3+k_3, n)]$ by increasing their momentum along $\mathbf{B}$, while ΔH accounts for the pair creation (only in the region $z_1 < -4\varepsilon^2_{n,0}$), both due to the interaction with the longitudinal mode, the Landau quantum numbers being unchanged[47]. Pauli's principle demands that vacant states must exist both for the occurrence of excitation and pair creation processes, for a fixed n.

The current j_ν is also a pseudovector ($b^{(2)}_\nu$ is a pseudovector), which is a necessary condition for the breaking of chiral symmetry[38]. An anomaly relation in a medium of massive fermions in presence of $\mathbf{B}$ (analogous to the Adler-Bell-Jackiw relation[6]) was calculated in[38]. This anomaly relation is associated to a chiral current induced by longitudinal photons.

As different from[36], we established a difference between the chiral symmetry breaking due to nonzero mass (which we may name as scalar or dynamical), compatible with thermodynamical equilibrium (since in it at each instant an equal number of left and right particles are expected to be on the average) and the pseudoscalar chiral symmetry breaking, arising from the nonvanishing term $\mathfrak{G} = \frac{1}{4}\mathbf{E}\cdot\mathbf{B}$, which leads to non-equilibrium processes of electric current and transport of charge[38].

It is easy to find a gauge transformation (in which it is obtained $b^{(2)}_3 = (k_4/z_1)E_3$) leading to

$$j_3 = \frac{sk_4}{z_1}E_3 = \frac{\Pi_{33}}{k_4}E_3, \quad (10)$$

where $E_3 = E^{(2)}(\mathbf{e}_B\cdot\mathbf{e}_3)$[38]. We only considered the real part of j_3, as we will restrict ourselves to the imaginary part of $\Pi_{\mu\nu}$. From now on we will restrict only to the real frequency and momentum ($k_3^2 > z_1$).

On the other hand the conductivity associated to the pseudovector mode is proportional to the corresponding eigenvalue of the photon self-energy tensor in the medium. The structure of this current is given in terms of the scattering and pair creation of electrons and positrons resulting from the decay of the longitudinal photons. Thus, we start of general expression to the real conductivity, which can be expressed explicitly in terms of the imaginary part of $\Pi_{\mu\nu}$ as[40,41,44,45]:

$$\sigma_{ij} = \frac{Im[\Pi_{ij}]}{\omega}. \quad (11)$$

The contribution to the current density j_i in (5) due to conductivity can then be written in the general form as: $j_i = \sigma^0_{ij}E_j + (E\times S)_i$, where $\sigma^0_{ij} = Im[\Pi^S_{ij}]/\omega$ and $S_i = \frac{1}{2}\epsilon^{ijk}\sigma^H_{jk}$ is a pseudovector associated with $\sigma^H_{jk} = Im[\Pi^A_{ij}]/\omega$, ϵ^{ijk} is the third rank antisymmetric unit tensor. Π^S_{ij}, Π^A_{ij} are the symmetric and antisymmetric parts of $\Pi_{\mu\nu}$. The first term corresponds to the Ohm current and the second is the Hall current[46]. The Hall and Faraday effects occur for the C-non-symmetric case; as different from the chiral magnetic effect, which occurs in both the C-symmetric and C-non-symmetric cases[38]. Here we will only work with the Ohm current term associated to the longitudinally polarized mode.

II. NEGATIVE MAGNETORESISTANCE INDUCED BY LONGITUDINAL PHOTONS IN DIRAC/WEYL SEMIMETALS

In this section, a low-energy model will be built to study systems such as Dirac/Weyl semimetals, by using the results of the statistical quantum electrodynamics formalism described above for a magnetized relativistic medium.

From (1) and considering the substitution $c \to v_F^2/c$[46], we get the following expression:

$$\varepsilon_{p_3,n} = \pm\sqrt{v_F^2 p_3^2 + m^{*2}c^4 + \frac{|e|\hbar v_F^2 B}{c}(2n_l + 1 - sgn(e)s_3)}, \tag{12}$$

which represents the dispersion relation for a low-energy fermionic system in $3+1$ dimensions in presence of a magnetic field, where $m^* \simeq 0.03\, m$ (m is the electron mass) corresponds to the effective mass, which is comparable to typical Dirac semimetal Cd_3As_2 values in $3+1$ dimensions[48–50]. The presence of a massive term in (12) implies that the band structure has a energy gap between the valence band (negative energy levels) and the conduction band (positive energy levels). As it is known in the massless limit, the basic energy level corresponds to the values $\varepsilon_{p_3,0} = \pm v_F|p_3|$. The valence band and the conduction bands touch for $p_3 = 0$. This contact point is called Dirac point. In the case of graphene (where the charge carriers behave as non-massive Dirac fermions in a two-dimensional hexagonal lattice) there are 6 points of Dirac, which lie at the corners of the first Brillouin zone (BZ).

If we now consider fermions with a very small effective mass, and the limit $\omega \to 0$, where we have $|k_3| \gg \omega$, that is, conditions very close to low-energy limit[38]. From (3), and taking into account the kinematics of absorption processes (see Appendix A 2) for excitation of particles in presence of $\mathbf{B}$ and longitudinal photons, it is possible introduce in (12), where $p_3 = p_3' - k_3$ (see (A16)), a pseudovectorial term (axial vector). This axial term is associated with the longitudinal photons propagation along $\mathbf{B}$. These are responsible of chiral symmetry breaking in the system[38], so we have

$$\varepsilon_{p_3,n,\sigma} = \pm v_F\sqrt{(\sigma p_3 + \frac{\eta\hbar|E_3|l_c}{|e|})^2 + |e|\hbar n B/c}, \quad n \geq 1,$$

$$\varepsilon_{p_3,0,\sigma} = v_F(\sigma p_3 + \frac{\eta\hbar|E_3|l_c}{|e|}), \quad n = 0, \tag{13}$$

where $\eta = sgn(\mathbf{e}_B \cdot \mathbf{e}_3) = \pm 1$ are associated with the pseudovector character of the electric field E_3, and $\sigma = \pm 1$ correspond to the different Weyl nodes. We have introduced a characteristic length l_c equal to the distance between Weyl nodes. The curves associated to the dispersion relation (13) are shown in Fig. 1. Notice the (13) corresponds to a dispersion relation of particles in presence of both the field $\mathbf{B}$ and the pseudovector field E_3. In Fig. 1 we can observe a qualitatively different character (compared to higher Landau levels) of the dispersion relations in the LLL given by the straight lines whose signs of slope correlate with the Weyl nodes. This correlation is due to the complete polarization of quasiparticle pseudospins in the LLL.

In Fig. 1, we can observe that the presence of longitudinal photons associated to the parameter E_3 has the effect not only of changing the relative position of the Weyl nodes (a magnitude $k_3 \sim E_3$ in the momentum space, with $\eta = +1$) but of inducing a chiral asymmetry on the Fermi surface. The two dispersion relations are the mirror images of each other. That is, particles with helicity-R and antiparticles (hollow-vacancy states) with helicity-L are produced. The R-particles will move towards the conduction band, while the L-antiparticles will move towards the valence band. In this way, the presence of a pseudoscalar, associated with longitudinal photons propagating along $\mathbf{B}$, transforms a Dirac semimetal into a Weyl semimetal with a pair of Weyl nodes for each point of Dirac. The nodes are separated by a pseudovector electric field induced dynamically in the direction of $\mathbf{B}$. We report a topological quantum transition (Fermi surface transformation) between a chiral symmetry-and non chiral symmetry phase.

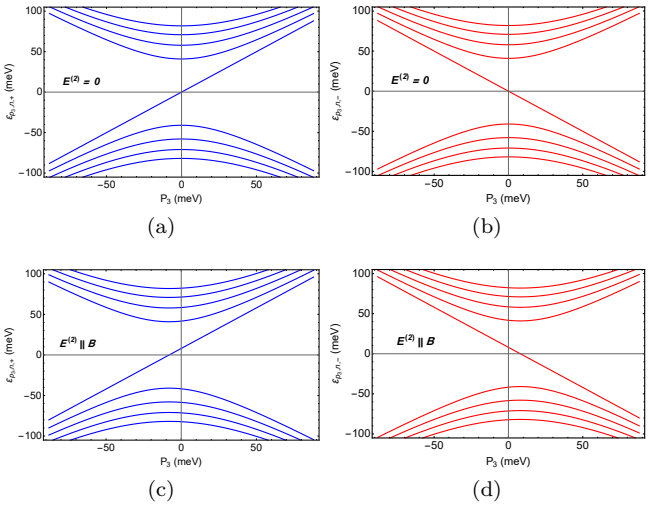


FIG. 1: Dispersion relations of the Weyl fermions to different Weyl nodes and their Fermi surfaces. a, b: system in equilibrium (with chiral symmetry) in presence only of $\mathbf{B}$. c, d: chiral symmetry broken in presence of both the field $\mathbf{B}$ and the electric field $\mathbf{E}^{(2)}$, the lattter related to longitudinal photons ($\mathbf{E}^{(2)} \parallel \mathbf{B}$). Typical values to these semimetals were used: $B = 6$ T, $v_F = 5 \times 10^7\ cm/s$, and $E_3 = 1\ MV/cm$, associated to $k_3 = 2.44 \times 10^5\ cm^{-1}$ and $l_c = 3.33 \times 10^{-8}\ cm$, this last corresponding to typical values of Na_3Bi[51].

A. Negative longitudinal magnetoresistance

Let us consider the specific case of current along $\mathbf{B}$, which is chiral non-symmetric. At the low-energy limit $|k_3| \gg \omega$, the expression (5) is valid for the excitation of particles (at conditions near the quantum degeneracy of the Fermion gas, as ω decreases significantly below $2m$, the pair creation contribution to $Im[s]$, see (8), is zero), which is guaranteed by the gauge invariance of a_μ[38]. Taking into account (6), and doing an appropriate gauge transformation (in which it is obtained the pseudovector potential $b_3^{(2)} = E_3/k_4$, with $E_3 = a|k_3|(\mathbf{e}_B \cdot \mathbf{e}_3)$,

see (3) in the low frequency limit), one gets:

$$j_3 = \frac{Im[s]}{\omega} E_3, \tag{14}$$

where $\sigma = \sigma^0_{33} = Im[s]/\omega$ is the magnetoconductivity associated to the longitudinal mode. Thus, as a result of (8), and (14), we obtain:

$$j_3 = \frac{e^3|k_3|}{4\pi z_1}\frac{\sinh(\frac{\omega}{2T})}{\omega}\sum_{n=0}^{n_{\mu,T}}\left[\frac{\alpha_n\varepsilon^2_{n,0}}{\lambda}\frac{1}{\cosh(\frac{\omega}{2T})+\cosh(\frac{\lambda-\mu}{T})} + \frac{1}{\cosh(\frac{\omega}{2T})+\cosh(\frac{\lambda+\mu}{T})}\right](\mathbf{E}^{(2)}\cdot\mathbf{B}), \tag{15}$$

where $\mathbf{E}^{(2)} = E^{(2)}\mathbf{e}_3$, $\lambda = (|k_3|\sqrt{1+\frac{4\varepsilon^2_{n,0}}{z_1}})/2$ and $n_{\mu,T}$ represents the contribution of the first Landau levels. We must recall the fundamental assumption that the system is close to be degenerate and $eB > \mu^2, T^2$, so that the contribution of the higher Landau levels is negligible. The second term inside the sum in (15) is associated to the hole-particles (hollow-vacancy states) with charge positive ($\varepsilon_q + \mu$ term in the Fermi-Dirac distribution). Taking now the limit $\omega \to 0$ in (15), we obtain[38]:

$$j_3 = \frac{e^3}{8\pi|k_3|T}\sum_{n=0}^{n_{\mu,T}}\left[\frac{\alpha_n\varepsilon^2_{n,0}}{\lambda}\frac{1}{1+\cosh(\frac{\lambda-\mu}{T})} + \frac{1}{1+\cosh(\frac{\lambda+\mu}{T})}\right](\mathbf{E}^{(2)}\cdot\mathbf{B}), \tag{16}$$

where $\mathbf{E}^{(2)} = a|k_3|\mathbf{e}_3$. We see easily that (16) tends smoothly to a nonzero value as $\mu \to 0$. This means that the chiral current is nonzero in both C-symmetric and C-non-symmetric cases[38]. In this respect, it differs from the Faraday rotation and from the Hall current, reported in the literature[46], which vanish in the C-symmetric case. From (14) and the expression (16) (returning to units $c, \hbar$) the longitudinal magnetoconductivity at finite temperature T and density μ can be written as:

$$\sigma = \frac{e^3Bc^3m^{*2}}{4\pi\hbar^2k_3T}\left[\frac{(2\varepsilon_0)^{-1}}{1+\cosh[\frac{(\varepsilon_0-\mu)}{T}]} + \frac{\varepsilon_1^{-1}}{1+\cosh[\frac{(\varepsilon_1-\mu)}{T}]}\right] + \frac{e^4B^2v_F^3\tau_c}{2\pi\hbar c^2l_ck_3T}\left[\frac{\varepsilon_1^{-1}}{1+\cosh[\frac{(\varepsilon_1-\mu)}{T}]}\right], \tag{17}$$

where

$$\varepsilon_0 = \frac{1}{2}\sqrt{\hbar^2v_F^2k_3^2 + 4m^{*2}c^4}, \tag{18}$$

$$\varepsilon_1 = \frac{1}{2}\sqrt{\hbar^2v_F^2k_3^2 + 4m^{*2}c^4 + 8eB\hbar v_F^2/c}, \tag{19}$$

and τ_c^{-1} is the rate of chirality-changing transitions. In (17) we have only considered the electronic transport contribution (first term inside bracket in (15), the hole current term could be include easily) and $n_{\mu,T} = 0, 1$, where T, μ are fixed by condition $eB > \mu^2, T^2$. The relation (17) corresponds to a negative longitudinal magnetoresistance related to a chiral current generation along $\mathbf{B}$. This electric current between two Weyl nodes is induced by longitudinal photons. A general expression to the magnetoconductivity at finite temperature and density associated to the scattering and pair creation was calculated in[37,38].

If we now consider fermions with a very small effective mass (massless fermions) we get that:

$$\sigma = \frac{e^4B^2v_F^3\tau_c}{2\pi\hbar c^2l_c}\cdot f(B, k_3, \mu, T), \tag{20}$$

where we have defined the function f, which depends on field $\mathbf{B}$, the k_3-momentum associated with the pseudovector E_3, chemical potential-μ, and temperature T. The function f can be written as:

$$f(B, k_3, \mu, T) = \frac{1}{T}\frac{1}{k_3\varepsilon_1}\frac{1}{1+\cosh[\frac{(\varepsilon_1-\mu)}{T}]}. \tag{21}$$

We can observe in (20) the quadratic dependence on field $\mathbf{B}$. It is associated with a positive contribution to the magnetoconductivity, which is a prominent signature of the chiral magnetic effect in Dirac/Weyl systems in parallel electric and magnetic fields. Our relation (20) could describe the negative longitudinal magnetoresistance of a wide variety of Dirac/Weyl semimetals such as Na_3Bi and $ZrTe_5$[5,31], where $f(B, k_3, \mu, T)$ provides the possibility of generalizing the usual expressions of the magnetoconductivity reported in the literature.

If $\varepsilon_1 \simeq \mu$ is considered (only the first term in the expansion of $\cosh[\frac{(\varepsilon_1-\mu)}{T}]$ is used). From (20) and (21), it obtains the following expression for the magnetoconductivity:

$$\sigma = \frac{e^4B^2v_F^3\tau_c}{4\pi\hbar c^2l_cTk_3\varepsilon_1}. \tag{22}$$

If we now consider $\varepsilon_1 \geq \mu$ (the second term in the expansion of $\cosh[\frac{(\varepsilon_1-\mu)}{T}]$ is added). From (20) and (21), we obtain:

$$\sigma = \frac{e^4B^2v_F^3\tau_c}{4\pi\hbar c^2l_cTk_3\varepsilon_1}\frac{1}{1+\frac{1}{4T^2}(\varepsilon_1^2 - 2\varepsilon_1\mu + \mu^2)}. \tag{23}$$

In Fig. 2 we show the magnetoconductivity (a, c) and the negative longitudinal magnetoresistance ρ (b, d) corresponding to (22)-top and (23)-down respectively. The magnetoconductivity oscillations in (c, d) at low temperatures is associated with the Shubnikov-de Haas effect[49].

The curves (c, d) reproduced with a high level of accuracy some of the experimental results, in the low temperature region, reported for $ZrTe_5$ and Na_3Bi (see[5,31]). We can observed that the upward curvature of the magnetoresistance decrease when the temperature increase. At hight temperatures: $T^2 > eB$ could have positive contribution to the longitudinal magnetoresistance, even for certain values of the magnetic field. It depends on the semimetal considered. As can be seen in[5,31] the temperature range for which the longitudinal magnetoresistance is negative is different depending on whether it is the Na_3Bi or the $ZrTe5$[5,31]).

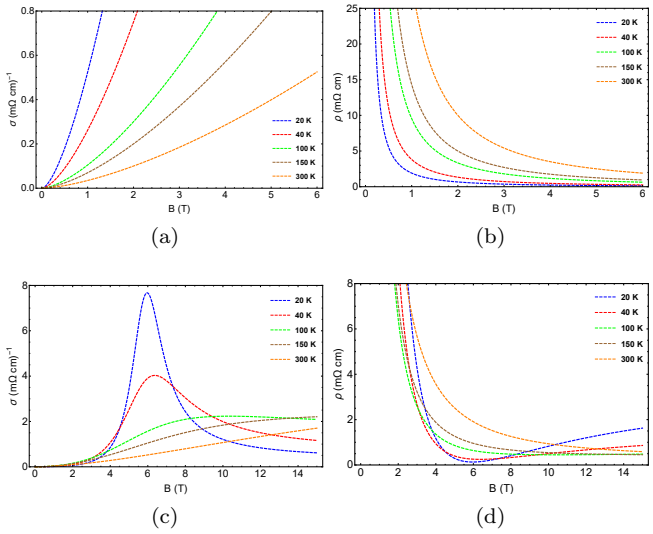


FIG. 2: **a,b**, Magnetoconductivity and the negative longitudinal magnetoresistence as a function of $\mathbf{B}$ to $\varepsilon_1 \simeq \mu$. **c,d**, Magnetoconductivity and the negative longitudinal magnetoresistence as a function of $\mathbf{B}$ to $\varepsilon_1 \geq \mu$. Typical values to these semimetals were used: $v_F = 5 \cdot 10^7\ cm/s$, $E_3 = 1\ MV/cm$, associated to $k_3 = 2.44 \cdot 10^5\ cm^{-1}$, $\tau_c = 6.66 \cdot 10^{-14}\ s$, $\mu = 40\ meV$ and $l_c = 3.33 \cdot 10^{-8}\ cm$, this last corresponding to typical values of Na_3Bi[51].

III. CONCLUSIONS

A low-energy model was built to study systems such as Dirac/Weyl semimetals. We started of general results related to the propagation of electromagnetic waves, the electromagnetic current vector, and the axial current in a magnetized relativistic fermionic system, according to statistical quantum electrodynamics formalism.

We conclude, that as a consequence of a pseudoscalar, associated to longitudinal photons propagating along $\mathbf{B}$, a Dirac semimetal could be transformed into a Weyl semimetal with a pair of Weyl nodes for each point of Dirac. The nodes are separated by a pseudovector electric field induced dynamically along $\mathbf{B}$. A topological quantum transition is produced between a chiral-and non chiral symmetry phase in presence of a pseudovector electric current parallel to $\mathbf{B}$. The pseudovector field E_3 is not just changing the relative position of the Weyl nodes in momentum space a magnitude k_3, but also to induce a chiral effect on the Fermi surface. That is, particles with helicity-R and antiparticles-(hollow-vacancy states)-L are produced. The R-particles will move towards the conduction band, while the L-holes will move towards the valence band.

A general expression to the longitudinal magnetoconductivity at finite temperature and density was found. This has a quadratic dependence on magnetic field $\mathbf{B}$, which is associated with a positive contribution to the magnetoconductivity. This is a prominent signature of the chiral magnetic effect in Dirac/Weyl systems in parallel electric and magnetic fields. In our case, the pseudovector electric field supplies a momentum to the system, and thermodynamic equilibrium as well as chiral balance are broken via an axial anomaly relation. We reported a chiral effect induced by longitudinal photons associated to a negative longitudinal magnetoresistance in Dirac systems. This negative magnetoresistance is valid for relativistic quantum systems associated to the scattering and pair creation. A general expression to longitudinal magnetoconductivity was calculated in[38].

We show some numerical results to the negative longitudinal magnetoresistance, and reproduced with a high level of accuracy some of the experimental results (in the low temperature region) reported to the magnetoresistance of $ZrTe_5$ and Na_3Bi[5,31]. The function $f(T, k_3, B, \mu)$ in (20) provides the possibility of generalizing the usual expressions of the magnetoconductivity reported in the literature, and to study the influence of the magnetic field $\mathbf{B}$ and the pseudovector electric field in these semimetals. We believe that a wide variety of these semimetals can be studied by using our general expression to the negative longitudinal magnetoresistance.

ACKNOWLEDGMENTS

The authors thanks the Abdus Salam ICTP OEA Office for support under Net-35.

Appendix A

1. Electromagnetic modes for propagation along B

From the structure of $\Pi_{\mu\nu}$ in a magnetized plasma, in the case of non-vanishing temperature T and chemical potential μ[40,41], we can find the polarization properties of three electromagnetic eigenmodes propagating in the system[40,45]. Under those conditions the polarization tensor may be expanded in terms of six independent transverse tensors[41]

$$\Pi_{\mu\nu} = \sum_{n=1}^{6} \pi^{(i)} \Psi^{(i)}_{\mu\nu}. \quad \text{(A1)}$$

As is shown in[40], symmetric properties in quantum statistics, reduce the number of the basic tensors from an initial set of 9 to a final set of 6. The basic tensors are:

$$\begin{aligned} \Psi^{(1)}_{\mu\nu} &= k^2 T_{\mu\nu}, \quad \Psi^{(2)}_{\mu\nu} = (Fk)_\mu (Fk)_\nu \\ \Psi^{(3)}_{\mu\nu} &= -k^2 T^\lambda_\mu F^2_{\lambda\eta} T^\eta_\nu, \quad \Psi^{(4)}_{\mu\nu} = R_\mu R_\nu, \end{aligned} \quad \text{(A2)}$$

with $T_{\mu\nu} = (g_{\mu\nu} - k_\mu k_\nu / k^2)$, and $R_\mu = (u_\mu - k_\mu (uk)/k^2)$; where u_μ is the four-velocity of the medium, and $F_{\mu\nu}$ is the electromagnetic tensor. We express tensor quantities in Minkowski metric, in the form $g_{\mu\nu} = (1, 1, 1, -1)$, understanding the order $\mu, \nu = 1, 2, 3, 0$ as it was agreed after (2). The tensors (A2) are symmetric in the indexes μ, ν while the following ones are antisymmetric

$$\begin{aligned} \Psi^{(5)}_{\mu\nu} &= (uk)[k_\mu (Fk)_\nu - k_\nu (Fk)_\mu + k^2 F_{\mu\nu}] \\ \Psi^{(6)}_{\mu\nu} &= u_\mu (Fk)_\nu - u_\nu (Fk)_\mu + (uk) F_{\mu\nu}. \end{aligned} \quad \text{(A3)}$$

We introduce a set of orthonormal vectors which are the eigenvectors of $\Pi_{\mu\nu}$ in the limit $\mu = 0$ and $T = 0$

$$\begin{aligned} c^{(1)}_\mu &= R_1 (F^2 k)_\mu k^2 - k_\mu (k F^2 k), \quad c^{(2)}_\mu = R_2 (F^* k)_\mu \\ c^{(3)}_\mu &= R_3 (Fk)_\mu, \quad c^{(4)}_\mu = R_4 k_\mu, \end{aligned} \quad \text{(A4)}$$

where $R_i, (i = 1, 2, 3, 4)$ are normalization parameters, and $F^*_{\mu\nu}$ is the dual of the electromagnetic tensor $F_{\mu\nu}$. Using these vectors we can obtain the scalars:

$$p = c^{(1)\mu} \Pi^\nu_\mu c^{(1)}_\nu, \quad s = c^{(2)\mu} \Pi^\nu_\mu c^{(2)}_\nu \quad \text{(A5)}$$

$$t = c^{(3)\mu} \Pi^\nu_\mu c^{(3)}_\nu, \quad r = c^{(3)\mu} \Pi^\nu_\mu c^{(1)}_\nu \quad \text{(A6)}$$

and the pseudoscalars:

$$q = c^{(2)\mu} \Pi^\nu_\mu c^{(1)}_\nu, \quad v = c^{(2)\mu} \Pi^\nu_\mu c^{(3)}_\nu \quad \text{(A7a)}$$

From (A1) we can find the polarization properties of three electromagnetic eigenmodes of $\Pi_{\mu\nu}$[40,45]. In the case of propagation along the magnetic field $\mathbf{B}$ in a magnetized medium, the eigenmodes of $\Pi_{\mu\nu}$, are:

$$\begin{aligned} b^{(2)}_\mu &= a c^{(2)}_\mu \, e^{i(k_3 x_3 - \omega t)} \\ b^{(1,3)}_\mu &= b(c^{(1)}_\mu \pm i c^{(3)}_\mu) \, e^{i(\mathbf{k}_\perp \cdot \mathbf{r}_\perp - \omega t)}, \end{aligned} \quad \text{(A8)}$$

with eigenvalues $\eta^{(2)} = s$ and $\eta^{(1,3)} = t \pm \sqrt{-r^2}$ respectively. Here $\mathbf{r}_\perp$ is the coordinate vector in the plane-(x, y), and a, b are parameters, which have dimensions of vector potential. The electric and magnetic fields associated by these modes are obtained from the equations:

$$\mathbf{E}^{(i)} = -\frac{\partial \mathbf{b}^{(i)}}{\partial x_0} - \frac{\partial \, b^{(i)}_0}{\partial \mathbf{x}}, \quad \mathbf{H}^{(i)} = \nabla \times \mathbf{b}^{(i)}, \quad \text{(A9)}$$

with $(i = 1, 2, 3)$. For the case of C-symmetric ($\mu = 0$), the mode $b^{(3)}_\mu$ is a transverse plane polarized wave, whose electric unit vector is $\mathbf{E}^{(3)}_u = \mathbf{e}_\perp \times \mathbf{e}_3$, orthogonal to the plane $(\mathbf{B}, \mathbf{k})$. Where we are defining $\mathbf{e}_\perp = \frac{\mathbf{k}_\perp}{k_\perp}$ and $\mathbf{e}_3 = \frac{\mathbf{k}_3}{k_3}$ as the transverse and parallel unit vectors respectively. The mode $b^{(2)}_\mu$ is pure electric and longitudinal with $\mathbf{E}^{(2)}_u = \mathbf{e}_B$ (we recall that $\mathbf{e}_B = \mathbf{B}/B$ is a pseudovector), whereas $b^{(1)}_\mu$ is transverse $E^{(1)}_u = \mathbf{e}_\perp$. In this C-symmetric case $\eta^{(1)} = \eta^{(3)}$, and the circular polarization unit vectors $(\mathbf{E}^{(1)}_u \pm i\mathbf{E}^{(3)}_u)/\sqrt{2}$ are common eigenvectors of Π_{ij} and of the rotation generator matrix A^{3ij} .

In the case of C-non-symmetric ($\mu \neq 0$). The second mode $b^{(2)}_\mu$ is the same pure longitudinal wave that in the C-symmetric case. The transverse modes $b^{(1,3)}_\mu$ describe circularly polarized waves in the plane orthogonal to $\mathbf{B}$ having different eigenvalues, typical of Faraday effect.

2. Calculation of $Im[s]$

The denominator D of the integral s (see (7), which have singularities due to D) given by:

$$D = 4 z_1 p_3 (p_3 + k_3) + z_1^2 - 4\omega^2 \varepsilon^2_{n,0}, \quad \text{(A10)}$$

where $z_1 = k_3^2 - \omega^2$ and $\varepsilon^2_{n,0} = m^2 + 2enB$, it can be written in the form symmetric under the exchange $\varepsilon_q \leftrightarrow \varepsilon_{q'}$, $\omega \leftrightarrow -\omega$[41]

$$\begin{aligned} D^{-1} = \frac{1}{8 \varepsilon_{q'} \varepsilon_q \omega} \Big(& \frac{1}{\varepsilon_{q'} - \varepsilon_q - \omega + i\epsilon} - \frac{1}{\varepsilon_{q'} - \varepsilon_q + \omega + i\epsilon} \\ & - \frac{1}{\varepsilon_{q'} + \varepsilon_q - \omega + i\epsilon} + \frac{1}{\varepsilon_{q'} + \varepsilon_q + \omega + i\epsilon} \Big), \end{aligned} \quad \text{(A11)}$$

where $\varepsilon_{q'} = \sqrt{(p_3 + k_3)^2 + m^2 + 2enB}$ and $\varepsilon_q = \sqrt{p_3^2 + m^2 + 2enB}$, with $q = (n, p_3)$. The first pair of singularities are related to excitation of particles to higher energies and the second two are connected to the pair creation. We have added an infinitesimal positive imaginary part $i\epsilon$ to ω, and by using the relation:

$$\frac{1}{s - \omega - i\epsilon} = P \frac{1}{s - \omega} + i\pi \delta(s - \omega), \quad \text{(A12)}$$

where P corresponds to the principal value in the expression, we get for the imaginary part of D^{-1}[41]

$$ImD^{-1}=\pm\frac{\pi}{8\varepsilon_q\varepsilon_{q'}\omega}[\delta(\varepsilon_{q'}-\varepsilon_q\mp\omega)+\delta(\varepsilon_{q'}-\varepsilon_q\pm\omega) - \delta(\varepsilon_{q'}+\varepsilon_q\mp\omega)], \quad \text{(A13)}$$

where the $\pm$ signs applies respectively to $\omega \gtrless 0$. We can use now (A13) to obtain the imaginary part the escalar s (see (7)) according to the relation:

$$\int_{-\infty}^{\infty} dp_3 f(p_3)\delta(g(p_3)) = \sum_m \frac{f(p_3^m)}{\mid g'(p_3^m)\mid}, \quad \text{(A14)}$$

where p_3^m, with $m=(1,2)$ are the roots of $g(p_3)=0$. It may be easily shown that while p_3 runs within $(-\infty < p_3 < \infty)$, the denominator of the expression (7) may vanish only for real z_1[41]. Thus, the integral in (7) represents an analytic function in the z_1 plane except possible singularities located somewhere on the real axis, which corresponds with the absorption region ($Im[\Pi_{33}]$ is responsible of absorption process for the longitudinal mode), where

$$p_3^{(1,2)} = \frac{-k_3 z_1 \pm \omega\Lambda}{2z_1}, \quad \text{(A15)}$$

are the roots of denominator in (7)[41] and $\Lambda = \sqrt{z_1(z_1+4\varepsilon_{n,0}^2)}$. In our case $g(p_3) = \omega \pm (\varepsilon_{q'} \pm \varepsilon_q)$, thus:

$$\omega = \varepsilon_{q'} \pm \varepsilon_q, \;\; k_3 = p_3' \pm p_3, \quad \text{(A16)}$$

and the corresponding values of the energies are given by:

$$\varepsilon_r = \frac{-\omega z_1 + |k_3|\Lambda}{2z_1} \quad \text{(A17)}$$

$$\varepsilon_s = \frac{\omega z_1 + |k_3|\Lambda}{2z_1} \quad \text{(A18)}$$

where $r,s=(n,\omega,k_3)$. The $\pm$ signs in (A16) correspond to the pair creation (ε_s) and excitation cases (ε_r) respectively. By substituting these expressions it is easy to obtain:

$$\mid \frac{d}{dp_3}(g(p_3)) \mid = \frac{\Lambda}{2\varepsilon_q^m\varepsilon_{q'}^m}. \quad \text{(A19)}$$

In the evaluation of the integral (7) containing the second delta (A13), the following exchange is made $p_3+k_3 \leftrightarrow -p_3$, $n' \leftrightarrow n$.

[1] Z. Wang, et al, Phys. Rev. B, 85, 195320 (2012).
[2] Z. Wang, H. Weng, Q. Wu, X. Dai and Z. Fang, Phys. Rev. B, 88, 125427 (2013).
[3] S. Borisenko, et al, Phys. Rev. Lett, 113, 027603 (2014).
[4] Z. K. Liu, et al, Science, 343, 864-867 (2014).
[5] Q. Li, D. E. Kharzeev, Ch. Zhang, Y. Huang, I. Pletikosic, A. V. Fedorov, R. D. Zhong, J. A. Schneeloch, G. D. Gu and T. Valla, Nature Physics, 12:550-554, (2016).
[6] S. L. Adler, Phys. Rev, 177, 2426-2438 (1969).
[7] J. S. Bell and R. Jackiw, Nuovo Cim. A, 60, 47-61 (1967).
[8] X. Wan, A. M. Turner, A. Vishwanath and S.Y. Savrasov, Phys. Rev. B, 83(20), 205101 (2011).
[9] A. A. Burkov and L. Balents, Phys. Rev. Lett, 107, 127205 (2011).
[10] Y. Chen, L. Bergman, and A. A. Burkov, Phys. Rev. B, 88, 125105 (2013).
[11] L. A. Fal'kovskii, Sov. Phys. Usp, 11, 1 [Usp. Fiz. Nauk 94, 1] (1969).
[12] C. Herring, Phys. Rev, 52(4), 365 (1937).
[13] G. W. Semenoff, Phys. Rev. Lett, 53(26), 2449 (1984).
[14] H. B. Nielsen and M. Ninomiya, Phys. Lett. B, 130(6), 389 (1983).
[15] B. A. Bernevig, T. L. Hughes and S. C. Zhang, Science, 314(5806), 1757 (2006).
[16] Y. Chen, J. Analytis, J. H. Chu, Z. Liu, S. K. Mo, X. L. Qi, H. Zhang, D. Lu, X. Dai, Z. Fang, et al, Science, 325(5937), 178 (2009).
[17] L. Fu and C. L. Kane, Phys. Rev. B, 76(4), 045302 (2007).
[18] L. Fu, C. L. Kane and E. J. Mele, Phys. Rev. Lett, 98(10), 106803 (2007).
[19] F. D. M. Haldane, Phys. Rev. Lett. 61(18), (1988).
[20] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. Cava and M. Z. Hasan, Nature, 452(7190), 970 (2008).
[21] C. L Kane and E. J. Mele, Phys. Rev. Lett, 95(22), 226801 (2005).
[22] J. E. Moore and L. Balents, Phys. Rev. B, 75(12), 121306 (2007).
[23] R. Roy, Phys. Rev. B, 79(19), 195322 (2009).
[24] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. Hor, R. Cava, et al, Nature Physics, 5(6), 398 (2009).
[25] H. Zhang, C. X. Liu, X. L. Qi, X. Dai, Z. Fang, and S. C. Zhang, Nature Physics, 5(6), 438 (2009).
[26] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys, 82, 3045 (2010).
[27] X. L. Qi and S. C. Zhang, Rev. Mod. Phys, 83, 1057 (2011).
[28] D. X. Qu, Y. Hor, J. Xiong, R. Cava and N. Ong, Science, 329(5993), 821 (2010).
[29] S. Murakami, New Journal of Physics, 9(9), 356 (2007).
[30] D. E. Kharzeev, L. D. McLerran and H. J. Warringa, Nucl Phys. A, 803, 227-253 (2008).
[31] J. Xiong, S. K. Kushwaha, T. Liang, J. W. Krizan, M. Hirschberger, W. Wang, R. J. Cava and N. P. Ong, Science, 350(6259), 413-416 (2015).
[32] C. Z. Li, et al. Nature Commun. 6, 10137 (2015).
[33] H. J. Kim, et al. Phys. Rev. Lett, 111, 246603 (2013).
[34] A. Y. Alekseev, V. V. Cheianov and J. Fröhlich, Phys. Rev. Lett, 81, 3503-3506 (1998).

[35] K. Fukushima, D. E. Kharzeev and H. J. Warringa, Phys. Rev. D, 78, 074033 (2008).
[36] A. Vilenkin, Phys. Rev. D, 22, 3080-3084 (1980).
[37] J. L. Acosta Avalo and H. Pérez Rojas, Astronomische Nachrichten, 336(8-9), 890–894 (2015).
[38] J. L. Acosta Avalo and H. Pérez Rojas, Nucl. Phys. B, 909, 230-242 (2016).
[39] E. S. Fradkin, Proc of the P. N. Lebedev Phys Inst, No. (29), consultants Bureau (1967).
[40] H. Pérez Rojas and A. E. Shabad, Ann. Phys, 121, 432-455 (1979).
[41] H. Pérez Rojas and A. E. Shabad, Ann. Phys, 138, 1-35 (1982).
[42] M. H. Johnson and B. A. Lippmann, Phys. Rev, 76, 828-832 (1949).
[43] M. Chaichian, S. S. Masood, C. Montonen, A. Pérez Martínez, and H. Pérez Rojas, Quantum Magnetic Collapse, Phys. Rev. Lett, 84, 5261-5264 (2000).
[44] I. A. Batalin and A. E. Shabad, Zh. Eksp. Teor. Fiz, 60, 894-900 (1971).
[45] A. E. Shabad, Ann. Phys, 90, 166-195 (1975).
[46] L. Cruz Rodríguez, A. Pérez Martínez, H. Pérez Rojas and E. Rodríguez Querts, Phys. Rev. A, 88, 052126 (2013).
[47] R. G. Felipe, A. Pérez Martines, H. Pérez Rojas, Modern Phys Lett. B, 4, 1103-1109 (1990).
[48] L. He, et a, Phys. Rev. B, 113, 246402 (1985).
[49] G. Kamm, D. Gillespie, A. Ehrlich, T. Wieting and F. Levy, Phys. Rev. B, 31, 7617-7623 (1985).
[50] L. P. He, et al, Phys. Rev. Lett, 113, 246402 (2014).
[51] N. P. Armitage, E. J. Mele and A. Vishwanath, Rev. Mod. Phy, 90(1), 015001 (2018).

Prepared for submission to JHEP

Cluster Scanning: a novel approach to resonance searches

I. Oleksiyuk,[a,b] **J. A. Raine,**[a] **M. Krämer,**[c] **S. Voloshynovskiy,**[b] **T. Golling**[a]

[a] *Département de Physique Nucléaire et Corpusculaire, University of Geneva, 1211 Geneva, Switzerland*

[b] *Centre Universitaire d'Informatique, University of Geneva, 1211 Geneva, Switzerland*

[c] *Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, 52074 Aachen, Germany*

E-mail: ivan.oleksiyuk@unige.ch, john.raine@unige.ch, mkraemer@physik.rwth-aachen.de, svyatoslav.voloshynovskyy@unige.ch, tobias.golling@unige.ch

Abstract: We propose a new model-independent method for new physics searches called Cluster Scanning. It uses the k-means algorithm to perform clustering in the space of low-level event or jet observables, and separates potentially anomalous clusters to construct a signal-enriched region. The invariant mass spectra in these two regions are then used to determine whether a resonant signal is present. A pseudo-analysis on the LHC Olympics dataset with a Z' resonance shows that Cluster Scanning outperforms the widely used 4-parameter functional background fitting procedures, reducing the number of signal events needed to reach a 3σ significant access by a factor of 0.61. Emphasis is placed on the speed of the method, which allows the test statistic to be calibrated on synthetic data.

arXiv:2402.17714v1 [hep-ph] 27 Feb 2024

Contents

1 Introduction **1**

2 Dataset **3**
2.1 Jet images 3

3 Method **4**
3.1 Bump hunt 4
3.1.1 Signal enrichment 4
3.1.2 Background estimation 5
3.1.3 Test Statistic definition and calibration 5
3.1.4 Significance evaluation 6
3.2 Cluster scanning 6
3.3 Discussion 10
3.4 Idealised CS 11

4 Results **12**

5 Conclusions and outlook **14**

A Idealised fit and n-parameter fit pseudo-analysis **15**

B Sparsity of the jet images **17**

C Hyperparameter selection and motivation **17**

D Gaussianity of cluster scanning bin entries **20**

E Outlier robust estimators **20**

F Calibration distributions **21**

1 Introduction

The Standard Model (SM) is the current apex of theoretical physics, describing the electromagnetic, weak and strong interactions with unparalleled precision. Unfortunately, it is still far from complete, as several phenomena remain unexplained. In order to create a "theory of everything", one would not only need to combine the SM with general relativity, but also provide an explanation for many other issues, including the existence of neutrino masses, the origin of the matter-antimatter asymmetry, and most importantly, the origin of dark matter. To solve these problems, researchers are collaborating to formalise new

theories, design, build and carry out new experiments, as well as simulate and analyze research data.

One of the most renowned experimental facilities, the Large Hadron Collider (LHC) was constructed with the purpose of testing the SM in the high energy regime. The last elementary particle predicted by the SM, the Higgs boson, was discovered in 2012 [1, 2]. Since then, LHC research has shifted towards precision measurements and searches for beyond the Standard Model (BSM) effects.

Many extensions of the SM imply the existence of as yet undiscovered massive particles, often associated with proposed new symmetry groups. If a new particle has a narrow decay width, the straightforward method is to search for a resonant peak in the spectrum of a mass-like observable, such as the invariant mass of a dijet event. However, such a bump hunt is not completely free of assumptions. Often complex analytical functions need to be chosen to model the background distribution, with the possibility to introduce spurious signals and varying sensitivity under the assumption of different functional forms. Furthermore, additional observables or fiducial cuts need to be chosen and optimised to enhance sensitivity in the case where potential signal yields are low, causing searches to become more model-specific.

Over the past decade, machine learning-based algorithms have become increasingly popular for solving a multitude of problems. Deep learning, in particular, has gained popularity for various tasks, with large neural networks being utilised. For example, many methods were implemented to perform anomaly detection (AD) tasks in various industries. Some of these AD methods have been repurposed and extended to support BSM searches [3–79] (see Refs. [80–83] for a comparison of various ML assisted BSM methods and Refs. [84, 85] for a comparison of weakly supervised and unsupervised approaches). The ATLAS collaboration produced the first experimental results for such searches applied to experimental data using weakly supervised methods [86] and unsupervised ML anomaly detection methods [74, 87]. However, these efforts have not observed any significant deviations from the SM expectation.

Many AD approaches rely on the assumption that any new signal would form a set of outliers. However, in a bump hunt the assumption is instead that any new signal would be localised in some feature space, in particular in an invariant mass spectrum. Weakly supervised approaches, on the other hand, aim to enhance the sensitivity by applying a cut on a classifier trained directly on the data. However, in both instances the same bump hunt restrictions apply with either functional forms or input observables impacting the sensitivity to a model.

In this work we introduce a new data-driven method, cluster scanning (CS), which builds on the foundations of the bump hunt but addresses several limitations. By leveraging more information from the event CS is able to enhance sensitivity to potential signals without enforcing any model specific assumptions, and can also provide a direct estimate of the background distribution. The proposed approach complements existing techniques and is designed to be computationally efficient.

The paper is structured as follows. In Section 2, we briefly describe the LHCO R&D dataset [88], commonly used to benchmark the performance of anomaly detection tech-

niques, and introduce our data preprocessing steps. Section 3 touches on the general topic of bump-hunting strategies in the literature, introduces the novel CS method, and discusses similarities and differences between them. In Section 4 we provide the results of applying CS in an anomaly search. Finally, we draw conclusions in Section 5.

2 Dataset

The LHCO R&D dataset consists of one million background Standard Model dijet events (also subsequently referred to as QCD) and 100 000 signal BSM $Z' \to XY$ events, where massive particles with $m_X = 500\,\text{GeV}$ and $m_Y = 100\,\text{GeV}$ decay into quark-antiquark pairs. The resonance itself has a mass of $m_{Z'} = 3.5\,\text{GeV}$. This anomaly model is discussed in detail in Ref. [89].

All the events were produced using Pythia 8.219 [90] and Delphes 3.4.1 [91–93] using default settings. The jets were clustered using an anti-k_T algorithm [94] with $R = 1$ using FastJet [95] with a `python` interface provided through the `pyjet` library in Scikit-HEP [96]. Jets are required to have $p_T > 1.2\,\text{TeV}$ and fall within $|\eta| < 2.5$.

2.1 Jet images

In addition to the di-jet invariant mass (m_{jj}) of the event, used in a bump hunt, we extract additional information from the image representations of the two jets. This allows for a more model agnostic approach than selecting specific jet substructure observables. The jet images are processed following a prescription similar to that used in Ref. [6, 97–99] from the η, ϕ and p_T of the jet constituents. Individual jet images are centred, rotated, and flipped in order to provide a consistent input to a convolutional neural network, reducing the number of symmetries the ML method would need to learn.

The jet images are cropped to $[-0.8, 0.8] \times [-0.8, 0.8]$ in $\eta - \phi$ space relative to the jet centre, binned with a 40×40 pixel grid, and normalised such that the sum of all pixels is equal to one. Fig. 1 shows the average jet images for QCD background, and the separate averages of all lighter (mostly Y) and heavier (mostly X) jets in each Z' event.

Despite being used in many applications, the jet image representation has two main drawbacks, namely the sparsity of non-zero pixels (see app. B) and the imbalance in the magnitudes of their intensities. This is particularly problematic for approaches that depend on the L_2 (Euclidean) distance. We address both of these problems with the solutions introduced in Refs. [57, 100].

To take the soft constituents into account, which have intensities orders of magnitudes lower than hard constituents, we apply a non-linear scaling to all pixels of $I_{ij} \to I_{ij}^{\gamma}$. To address sparsity we convolve (smear) the whole image with a two-dimensional Gaussian kernel with an isotropic standard deviation σ_k. We find that using a value of $\gamma = 0.5$ for the pixel scaling alongside $\sigma_k = 1$ for the Gaussian kernel provides a adequate solution to both issues without excessive impact on the structure of the jets.

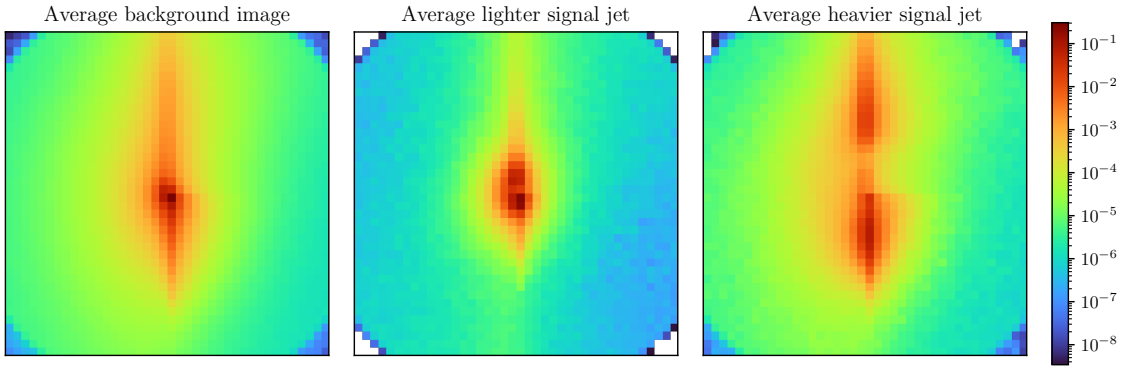


Figure 1: From left to right: average of all 2M available QCD jet images, average image of all 100K lighter jets in a Z' event and average image of all 100K heavier jets in a Z' event before smearing and pixel scaling.

3 Method

3.1 Bump hunt

The bump hunt approach is a standard method used to search for excesses over a non-resonant background in HEP (high-energy physics) data. This method usually follows four main steps that we briefly discuss below. Each of these steps is a complex topic in itself with several different approaches in the literature, thus for our study, we choose only simplified and basic approaches.

3.1.1 Signal enrichment

Signal enrichment, in general, refers to the selection of a subset of experimental data in such a manner that the fraction of signal events in it is increased compared to the initial sample. Most often, this is done by cutting out a region of the observable space where the signal is expected to be abundant compared to the rest of the space, typically using a theoretical model of the signal of interest.

These approaches, despite being sensitive to specific signal processes, make the search less model-agnostic and are ill-suited for general anomaly detection searches. Alternatively, one can hope to define a signal-rich region of the experimental data using a plethora of unsupervised ML (machine learning) techniques, which are expected to provide enhanced sensitivity over a wider range of potential signal processes.

In our particular example of LHCO data, we choose to explore a wide, smoothly falling region of the spectrum of dijet events with invariant dijet mass m_{jj} from 3000 GeV to 4600 GeV. We choose this lower bound to avoid the turn on curve of the mass distribution, resulting from the jet trigger, and the upper bound is selected to remain in a region with relatively high statistics, so that we work in the region where the fit functions from Subsection 3.1.2 are applicable. This interval contains, in total, around 380,000 QCD events and nearly all Z' events. We divide this region into 16 non-intersecting bins with 100 GeV width each, as in Ref. [101, 102].

3.1.2 Background estimation

To perform a hypothesis test, one must first postulate a null hypothesis, which in counting experiments takes form of the expected background coming from the Standard Model processes. Often the background prediction relies on a theoretical basis to calculate the cross sections of the hard process and a simulation to account for detector response and measurement uncertainties. Still there are a number of searches where theory and simulation cannot provide a reliable background estimate. In these cases the background has to be estimated from the data itself in an empiric manner, using some general assumptions.

In dijet-like searches a background is often estimated by fitting a function of the form

$$f(x) = p_1(1-x)^{p_2}x^{p_3+p_4\ln(x)+p_5\ln(x^2)} \tag{3.1}$$

to a smoothly falling part of the dijet mass distribution [103–119], where $x = m_{jj}/\sqrt{s}$. This function is referred to as the "n-parameter dijet fit function", where n is the number of nonzero free parameters p_i used in the function. Despite being a good fit to the simulated data, this functional form is still an empirical assumption and thus is subjects to a systematic error. Furthermore, after applying some selection criteria on the events which could be correlated with m_{jj}, this function may no longer well describe the resulting distribution.

More advanced methods of fitting, such as the Sliding Window Fit (SWIFT) [120] and the ABCD method used in [121, 122] are other methods that reduce the assumption of a functional form but introduce their own assumptions instead. However, due to the simplicity and wide use of the n-parameter fit function, we choose to use global 3-parameter and 4-parameter function fits as the benchmark analysis strategy. Further details of the (pseudo-)analysis on the LHCO R&D data performed using these background estimates are given in Appendix A. To access the upper bound on the performance of all background estimation methods, we use the underlying background distribution as an idealised fit, i.e. a fit with no systematic error. The (pseudo-)analysis using this is also described in Appendix A.

3.1.3 Test Statistic definition and calibration

There are several ways to calculate a global test statistic for two spectra. In HEP one of the more popular tests in model agnostic searches, called BumpHunter [123], relies on the maximal local significance (MLS) as the test statistic, where it is computed using a range of different windows over the spectrum. One of the benefits of the MLS test statistic is its simplicity and that it is well suited for signals that give rise to narrow, localised resonances. Here the MLS is applied to the binned m_{jj} distributions of the data. Given a set $\mathcal{B} = \{b_1, ..., b_{n_{\text{bins}}}\}$ of non-intersecting bins with $N_{\text{sig},b}$ events or jets from the signal-rich (experimental) distribution and $N_{\text{bkg},b}$ events or jets from the background estimation, the MLS can be written as

$$\text{MLS} = \max_{b\in\mathcal{B}} Z_b = \max_{b\in\mathcal{B}} (\text{CDF}^{-1}_{\mathcal{N}(0,1)}(\text{CDF}_{Poisson(N_{\text{bkg},b})}(N_{\text{sig},b}))\,, \tag{3.2}$$

where CDF is the cumulative density function of the respective distribution. In equation 3.2 only overdensities are taken into account, i.e. $Z_b > 0$ only for $N_{\text{sig},b} > N_{\text{bkg},b}$ as we are searching for a resonance.

For bins with $N_{\text{sig/bkg},b} \gg 1$ one can approximate the Poisson distribution with a normal distribution $\mathcal{N}(N_{\text{sig/bkg},b}, \sqrt{N_{\text{sig/bkg},b}})$. Equation 3.2 then reduces to a much simpler form

$$\text{MLS} = \max_{b\in\mathcal{B}} Z_b = \max_{b\in\mathcal{B}} \frac{N_{\text{bkg},b} - N_{\text{sig},b}}{\sqrt{N_{\text{bkg},b}}} \,. \tag{3.3}$$

Although some test statistics, like χ^2, have well-known distributions, other more unusual test statistics, like the BumpHunter test statistic, require calibration. This is commonly done by modelling its distribution using Monte Carlo simulation.

Moreover, as systematic uncertainties arise from the definition of a signal region selection and the background estimate, this calibration should be performed even in the case where the distribution is known a priori. The calibration for the BumpHunter test statistic is performed in Ref. [123] by running pseudo-experiments in which the counts in each bin are varied according to Poisson's law. This can be extended to higher dimensions by resampling the background events with bootstrapping. By calculating the test statistic for each of our bootstrapped background-only pseudo-experiments, we obtain the distribution of the test statistic in the background-only hypothesis. To ensure good modelling of the tail of the test statistic distribution, which corresponds to large significance values in the presence of signal, a large number of pseudo-experiments is required.

3.1.4 Significance evaluation

To obtain a calibrated p-value for a given value of the test statistic t, one counts the number of background only pseudo-experiments exceeding this value $N_{>t}$ and divides it by the total number of pseudo-experiments done, N_{tot}.

The (one-sided) significance is computed using the inverse cumulative density function of the normal distribution $Z = \text{CDF}^{-1}_{\mathcal{N}(0,1)}(1 - p\text{-value})$.

In the case of $N_{>t} = 0$ arising from the limited number of pseudo-experiments, we instead set a lower bound:

$$p < \frac{1}{N_{tot}}, \; Z > \text{CDF}^{-1}_{\mathcal{N}(0,1)}(1 - \frac{1}{N_{tot}}) \,. \tag{3.4}$$

For every experiment with added signal events, we still bootstrap the background (for consistency) and combine it with a given number of signal events chosen at random from 100,000 signal events (around 5% of events fall outside of the evaluation region). Due to statistical fluctuations we also perform several pseudo-experiments in the signal enriched case in order to obtain a robust estimate of the significance for each level of signal doping.

3.2 Cluster scanning

In this section we present a novel approach called Cluster Scanning, which follows the same bump hunting scheme, but relies on a distinct set of assumptions than the commonly

employed methods and thus has several favorable characteristics. Our approach can be divided into several key steps given below, with the hyperparameters chosen in order to search for narrow resonances in the m_{jj} spectrum of the LHCO R&D data. The motivation for these hyperparameters in each step and the argumentation on how to choose them for a different application case is given in App. C.

Training region selection: We select a narrow m_{jj} window [3000, 3100] GeV for training of the k-means algorithm. This window contains 56,486 original background events. In this publication, we focus on relatively small signal injections that include only 5% or less of the total number of Z' signals available. Therefore the training region is expected to contain 89 signal events or less, which can be regarded as negligible. Despite observing a qualitative improvement in performance in case the training region matches the resonant peak and thus has a larger portion of signals events involved in clustering, in an actual analysis the position of the peak will be unknown, thus we choose to discuss a more representative case given here, when the training region happens to be in the tail of the signal peak and thus has a negligible number of signal events.

K-means Clustering: We apply a mini-batch k-means clustering algorithm with $k = 50$ implemented in the SCIKIT-LEARN [124] library, with a batch size of 2048 on the set containing jet images of the leading two jets from each event in this m_{jj} window. The mini-batch implementation is chosen due to its computational speed. The seeding of the cluster centroids is performed using the K-MEANS++ prescription described and motivated in Ref. [125].

Cluster Spectra: After performing the fit of k-centroids to the data in the training region, we fix the centroid positions and evaluate how many jet images from each of the 16 m_{jj} bins of the evaluation region, defined in Subsection 3.1.1, fall into each of the k clusters $N_{i,b}$ where $i \in \{1...k\}$, $b \in \{1...n_{bins}\}$. Fig. 2 shows the resulting 50 normalised cluster spectra $N_{i,b}/\sum_b(N_{i,b})$ for one pseudo-experiment with signal injection.

Per bin standardisation: We note that in each bin the normalised cluster spectra follow an approximately normal distribution with several outliers from the anomalous clusters (see discussion in App. D). Therefore we standardise the normalised cluster spectra in each bin using outlier robust estimators (described in App. E) for mean and standard deviation with an outlier factor of 0.2. Here we make the assumption that the majority of the signal is located in a small number of clusters, and the rest of the clusters are signal depleted. Figure 3 shows the cluster spectra from Fig. 2 after normalising with the outlier robust estimator.

Selecting anomalous clusters: Utilising the assumption that the signal is localised in m_{jj}, we select potentially signal-rich cluster spectra as those with a deviation of more than a threshold value of $\theta = 3$ standard deviations from the robust mean in the positive direction as we are only interested in a resonance leading to excess of events. The rest of the clusters are labelled as signal-depleted. The threshold and the selected signal-rich clusters are shown in Fig. 3 in red.

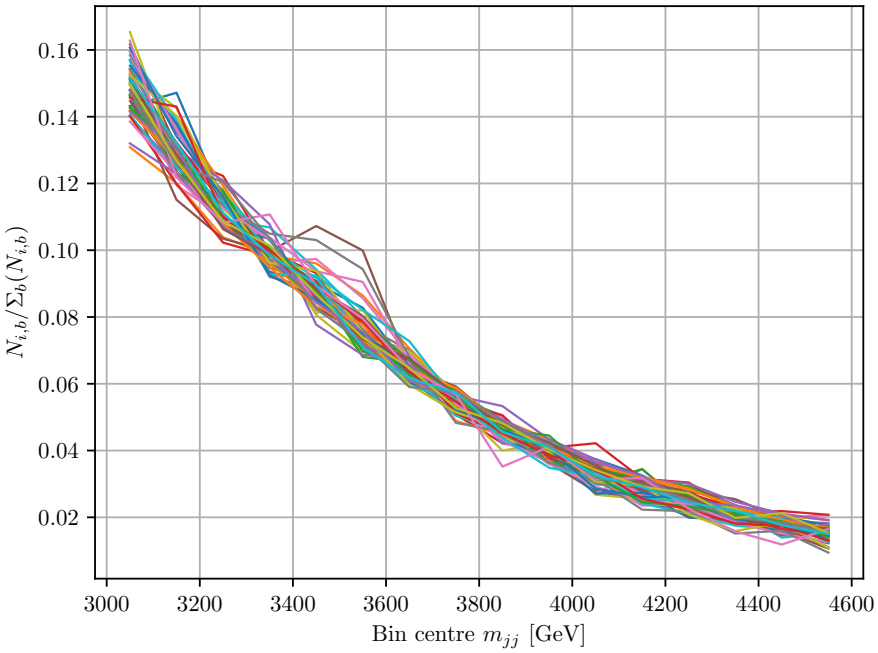


Figure 2: The m_{jj} distributions for the jets in each of the 50 clusters, each normalised to unity. Here, 5,000 signal events have been injected into the evaluation dataset, which corresponds to 5% of the total available signal events.

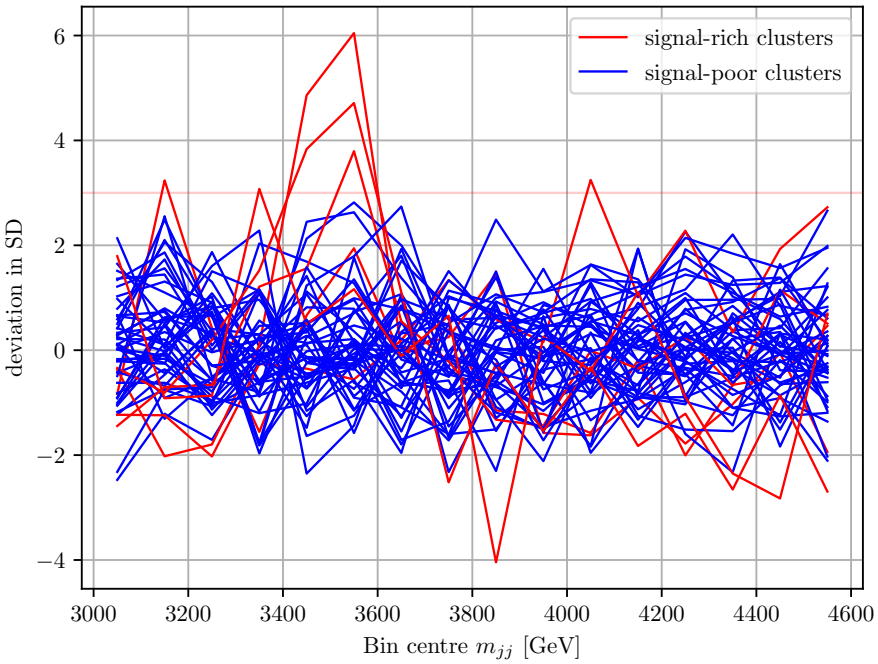


Figure 3: Spectra in Fig. 2 standardised over clusters in each bin. Potentially signal-rich cluster spectra are shown in red.

Signal-rich and signal-depleted regions: After the selection, we combine the non-normalised distributions corresponding to our selected signal-rich clusters. This results in a signal-rich spectrum $N_{\text{sig},b}$ with an example shown in red in Fig. 4.

The remaining cluster spectra are combined to form a signal-depleted spectrum $N_{\text{poor},b}$. The estimate of the background is then constructed by normalising it to the same total entries as in signal-rich spectrum, namely $N_{\text{bkg},b} = N_{\text{poor},b} \frac{\sum_b N_{\text{sig},b}}{\sum_b N_{\text{poor},b}}$. It is shown in blue in Fig. 4.

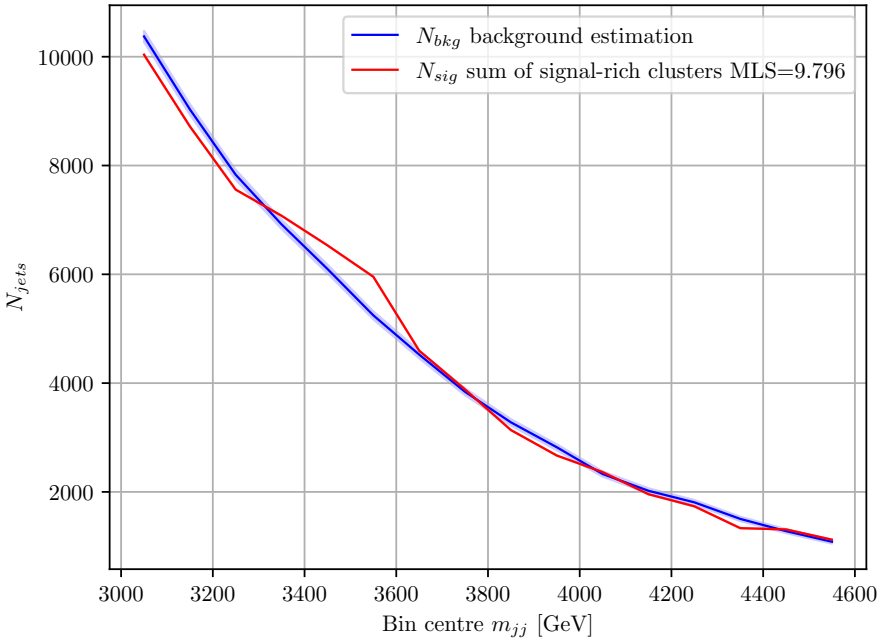


Figure 4: Curves corresponding to the sum of signal-rich and signal-poor spectra in Fig. 3. The blue signal-poor curve is rescaled to have the same total jet number as the signal-rich curve. The coloured region around the blue curve is a $\sigma_{\text{bkg},b} = \sqrt{N_{\text{bkg},b}}$ Poisson deviation after recaling used to compute MLS.

Test statistic: As previously discussed, to test the significance of an observed excess we use the simple maximum local significance, as defined in Equation 3.2. It may occur that no cluster is selected as anomalous. In this case we assign a default value of 0, in order to show good agreement with the null-hypothesis expectation. This is similar in motivation to setting the value for an observed deficit in events to zero. Following the discussion in Subsection 3.1 for the calibration process, we construct 3,900 pseudo-experiments using bootstrap resampling on 1 million background events. The distribution of the test statistic is discussed in App. F.

Ensembling: Different initialisations lead to a broader distribution over the final test statistic obtained with cluster scanning. In order to obtain a final value for the test statistic, the cluster scanning method is performed 15 times with independent initialisations. The

mean of the test statistic from all the runs forms the final ensembled test statistic. The distribution of this statistic is presented in app. F.

3.3 Discussion

As we can see, CS follows the general bump hunt strategy, but introduces novel approaches for the first two steps of this strategy. First of all, CS selects the most anomalous looking clusters to define the signal-enriched region, and constructs a background estimate from the rest of the clusters. Notably though, this selection is completely data-driven and does not target a specific family of signal models. However, CS relies on a set of assumptions that fundamentally differ from those commonly used in other anomaly detection approaches.

Search for overdensity instead of outliers: Most anomaly search methods like Autoencoders [100] and SVDDs [82] rely on outlier detection, namely, identifying the data instances that lie in a region of very low probability density or outside the support of the "normal" distribution. Notably, while all normal events share similar characteristics and exhibit easily recognisable trends, anomalous data, such as defects or fraud, can differ in numerous ways and are thus given a wide prior. Although model-agnostic searches should accommodate a wide range of possible anomaly models, it is usually assumed that a signal is produced by only one or a few unknown BSM process. Thus, all anomalous events have many features in common and exhibit some similarity to SM events, as any new particle must radiate and decay into SM particles to be detectable.

Therefore we use the localisation of anomalies in both low-level (e.g. jet images) and high-level variable (e.g. m_{jj}) space as the first main assumption of the CS method. Localisation of anomalies in low-level variable space means that only a few out of all clusters contain a fraction of anomalies much higher than the rest of the clusters. This way clustering plays a role of data-driven binning in low-level variable space. Localisation in m_{jj} gives us a possibility to distinguish these anomaly rich clusters from the rest, namely, by searching for an overdensity in m_{jj} in one cluster spectrum compared to all others. Thus, CS is able to select a signal-rich region of events by leveraging the assumption of signal being localised rather than consisting of outliers.

Although semi-supervised methods based on CWoLa (see Refs. [101, 102, 126–130]) and density estimation methods are also sensitive to overdensities, they usually require construction of a background template, which until recent developments [129, 130] was preferably constructed for a smooth distribution of low dimensionality, typically using a few high-level observables. In this publication, we show that CS is able to draw significant improvement from a high-dimensional distribution of low-level jet observables. In this way, it can be considered less signal-model dependent than the methods that rely on hand-crafted high-level observables.

Assume cluster mass independence instead of smoothness: CS proposes a solution to the second step of the analysis, namely, it estimates the form of the background by combining the signal-depleted clusters. In this way, we do not rely on any assumptions on smoothness or on a particular functional form of the background-only spectrum in

m_{jj}, which are heavily relied upon by most other bump hunt methods, such as the global functional fit mentioned in Subsection 3.1, SWIFT [120] and even Gaussian processes [131].

Instead, the second main assumption of CS is that in the background-only case, the assigned cluster centroid is approximately independent from m_{jj}. Ideally, we would want the distribution of background events over m_{jj} in each cluster to be identical within statistical uncertainty, such that the probability of a jet belonging to a cluster and having a specific mass factorises, $p(i, m_{jj}) = p(i)p(m_{jj})$, or at least that the correlation is weak. This would minimise the rate of incorrectly identified signal-enriched clusters. In practice, although Fig. 2 shows that the distributions all follow a similar trend, there are still some systematic deviations. These are a result of the finite width in m_{jj} of the training window and slight correlations between the distribution of the jet constituents and m_{jj} arising from the transverse momenta of the two non-resonant jets depending on m_{jj}. Therefore, for the selection of the clusters, we estimate the full uncertainty (including Poisson fluctuations and uncertainty from mass dependence) separately for each experiment and for bin based on the sample of our k cluster spectra values.

Unlike in methods with sliding window approach [102, 120], in CS the fit only needs to be performed once. Moreover k-means clustering is a simple classical algorithm that typically requires less training than deep learning approaches, making CS a relatively fast analysis method. This is important in the context of the ensembling and calibration, which both require a large number of analysis iterations, and are thus a notable obstacle to incorporating deep learning in HEP analysis under the constraint in computing resources. Fast analysis is also advantageous for testing its efficiency for simulated BSM events in order to produce the exclusion limits (see the RECAST [132] framework). Moreover, CS avoids other disadvantages inherent to sliding window approaches, such as limited search range due to the definition of the sidebands and the need to optimise sideband and signal window widths.

3.4 Idealised CS

Despite choosing a narrow m_{jj} window to reduce mass dependence systematics, the variables that we use for clustering are in general not independent of m_{jj}. Thus, we observe the background-only spectra of some clusters do not just statistically fluctuate around the expected shape of the background, but exhibit some degree of smooth mass sculpting. This affects the performance of the method by introducing false positives at the cluster selection stage and increasing the discrepancy between signal and background spectra when evaluating the test statistic. This may be partially remedied by a more sophisticated method of selecting anomalous clusters or a better background estimate, both of which rely on more assumptions. These studies are outside the scope of this publication. However, to give an upper bound on the performance one may achieve with such improvements we propose an idealised version of clusters scanning.

Idealised CS version requires us to modify the distribution of the jets between the clusters. First, we count the numbers of jets that fall into each cluster in the first m_{jj} bin. If no mass dependence were present, the fractions of QCD jets in each cluster $x_{\mathrm{QCD},i,b} = N_{\mathrm{QCD},i,b} / \sum_i N_{\mathrm{QCD},i,b}$ should be independent of bin number b within statis-

tical uncertainties. To simulate this case in all the consecutive bins except the first we distribute the QCD jets in these bins among clusters using a multinomial distribution with weights equal to the fractions obtained in the first bin $x_{\mathrm{QCD},i,b} = x_{\mathrm{QCD},i,1}$, thus generating cluster spectra that follow the original background spectrum with statistical fluctuations, i.e. the case with no mass dependence. The signal jets are distributed as before according to which cluster they belong to, such that the fractions of Z' jets may differ between different bins. This is done because we assume that only the background is distributed roughly proportionally between clusters, which is equivalent to assumption 2, but not the signal.

This distribution of jets creates idealised cluster spectra for each clustering, and the rest of the algorithm remains unchanged.

4 Results

As a proof of concept we perform an analysis applying CS and global fit based bump-hunting with the above mentioned hyperparameters to the LHCO R&D dataset with different amounts of signal injection, given in figures either as an absolute number of injected events ϵ or as a signal to background ratio S/B of events in the considered [3000, 4600] Gev m_{jj} region. For each pseudo-experiment with signal injection we calculate the significance Z as discussed in Subsection 3.1.4 using the calibration test statistic distribution. For each signal injection level we run 100 pseudo-experiments with bootstrapped background data and randomly sampled signal events. As a reference for the significance and its statistical variation for each contamination level, we report the median significance of these pseudo-experiments and 0.25 and 0.75 quantiles. We define the ratio between the median significance provided by CS to the significance of a baseline method as the significance improvement (SI). We also quote the relation between the number of events needed to obtain a 3σ evidence in each analysis strategy.

Figure 5a shows how the global significance of CS and the parametric fit-based methods depends on the signal contamination in the pseudo-experiment. It characterises the performance of these realistic analysis strategies, which do not use any truth information for the evaluation of the test statistic, thus including all the systematical uncertainties coming from partially fulfilled assumptions needed for the respective method.

We observe that although 3- and 4-parameter fits give approximately the same results, CS outperforms them by a significant margin in the region from 1500 to 4000 signal events. Beyond 3000 signal events, the significance yield from CS is limited by the number of bootstrap pseudo-experiments in the calibration set, but the lower bound on its significance still remains substantially higher than the significance of the parametric fits. This is the most interesting region as there the transition between non-significant signal (below 1σ) and new physics evidence (above 3σ) takes place. Looking at the lower subplot in Fig. 5a we see that CS gives us an SI of 2 and higher on the majority of regions of interest. We can also see that CS produces a 3σ evidence for only 61% of the events needed to obtain this evidence with the parametric fit. This shows that although both suffer from fit and assumption induced systematic uncertainties, CS has a clear advantage over parametric fitting procedures.

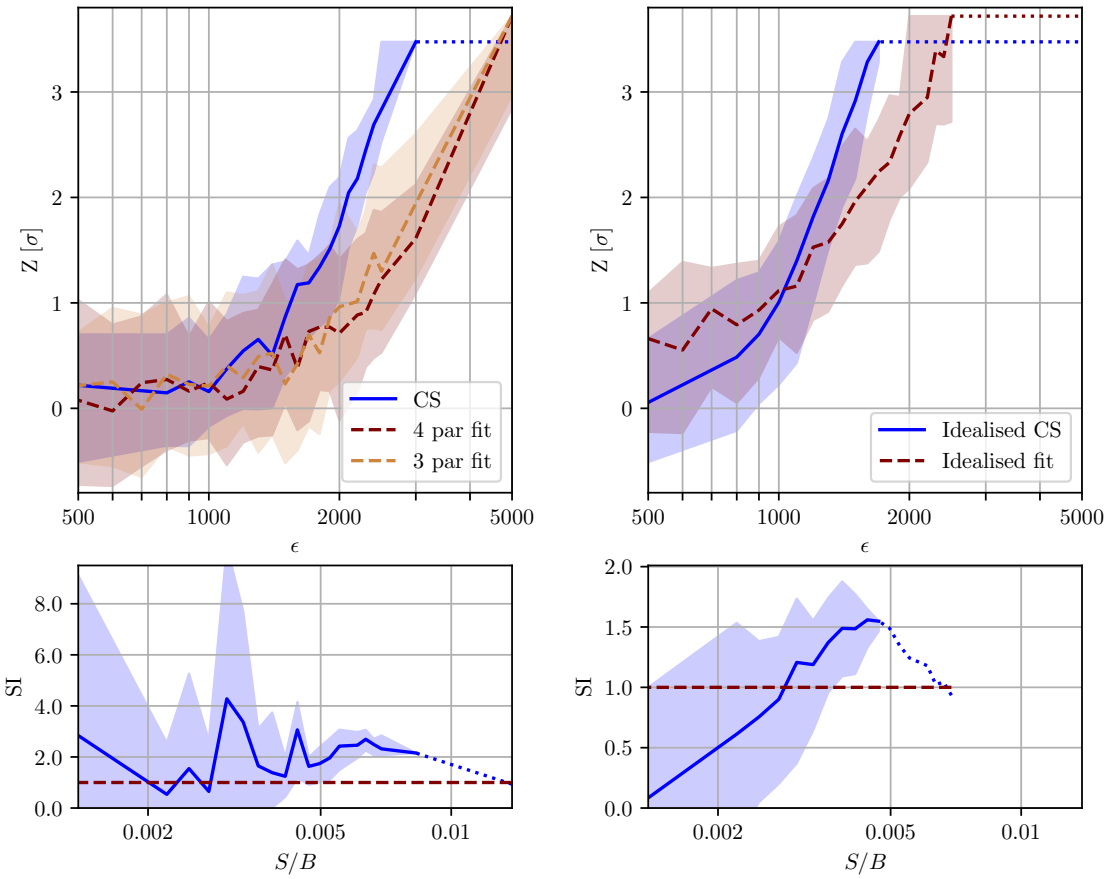


Figure 5: Upper figure: Median and the quartile bounds of global significance of the signal contaminated pseudo-experiments as a function of the number of signal events ϵ injected shown for the (a) realistic and (b) idealised analysis methods. The dotted lines mark lower bounds, as there was not enough statistics to access higher significance levels. Bottom figure: Significance improvement of the CS method compared to the 4-parameter fit for the (a) and the significance improvement of idealised CS compared to idealised fit for the (b) as a function of the signal-to-background ratio S/B.

Above we have also described the idealised version of the cluster scanning method and in Appendix A the analysis with an idealised background fit. Both methods rely on event labels to remove systematic uncertainties introduced by the limitations of the assumptions of our methods and to make the background estimate in both cases close to the true background, with only statistical fluctuation taken in consideration. This is done to separate the influence of additional information, namely the low-level observables used in the analysis, from the systematic uncertainties introduced by the fits, and to construct the upper bound on the performance of our methods.

Figure 5b shows how the global significance of both idealised methods depend on the signal contamination in the pseudo-experiment. It can be seen that one needs substantially

less signal for it to be significant in the idealised methods compared to the realistic methods, as it removes false-positives induced by systematic uncertainties in the fits. Still, we see that the idealised CS outperforms the idealised functional fit on the majority of the interval between 1000 and 2200 signal events. Looking at the lower subplot in Fig. 5b we see that by using CS in the region of interest we gain a significance improvement factor of up to 1.5. We can also see that CS produces a 3σ evidence with only 69% of the events needed to gain this evidence with the idealised fit. This shows that in the case of negligible systematic uncertainties, CS gives an improvement over any smooth fit as, in addition to just using information from m_{jj}, it also makes use of the low-level event information. From the difference between idealised and non-idealised CS we can see that there is some room for improvement of CS to reduce the false positive rate, and improve the analysis efficiency.

5 Conclusions and outlook

This paper is a first proof of concept for the cluster scanning anomaly search method, which is designed to search for resonant overdensities on the distribution of an observable using clustering techniques in auxiliary observables.

We found that it outperforms the widely used bump-hunting method, which relies on the functional background fits, in several metrics relevant to an analysis. In the transition region, where the benchmark algorithm achieves 1σ to 3σ significance, CS improves the result by a factor of 2 or more for the realistic case, or by a factor of 1.5 for the idealised comparison. This reduces the number of signal events required to produce a 3σ significance by a factor of 0.61 in the realistic case and by a factor of 0.69 in the idealised case. The former factor of improvement should be expected in a real application. We also discuss the comparison of cluster scanning with other anomaly detection algorithms in Subsection 3.3, outlining its advantages and limitations.

The CS method should not be seen as a direct competitor to background fitting methods, but rather as a complementary approach that relies on a different set of assumptions about the nature of the anomaly and the background distributions, which are not well known.

There remains a large unexplored field of potential extensions and improvements to this method or synergies with other methods. Straightforward follow-up studies can explore the use of clustering methods other than k-means. One can look for other ways of selecting the anomalous clusters, alternatives to the one proposed in Subsection 3.2, that would rely on different assumptions. For example, one can require that all anomalous clusters are neighbors in the space of clustered inputs. One can also unify the assumptions of CS and functional fits to produce separate background estimates for each of the clusters separately, greatly reducing the m_{jj} dependent systematic uncertainties.

CS could benefit from using features developed by other algorithms that have already been optimised for other tasks, such as flavour tagging, or even using unsupervised learning for feature extraction.

Furthermore, since many other ML approaches to improve sensitivity in model-independent searches rely on a bump hunt for the final statistical analysis, CS could also be used to

further enhance sensitivity. This could be of particular interest when the background distribution is no longer well described by simple, smoothly decreasing functional forms.

Acknowledgments

The authors would like to acknowledge funding through the SNSF Sinergia grant CRSII5_193716 "Robust Deep Density Models for High-Energy Particle Physics and Solar Flare Analysis (RODEM)" and the SNSF project grant 200020_212127 "At the two upgrade frontiers: machine learning and the ITk Pixel detector". The research of MK is supported by the DFG under grant 396021762 - TRR 257: Particle Physics Phenomenology after the Higgs Discovery.

Code availability

The code used to produce all results presented in this paper is available at https://github.com/IvanOleksiyuk/jet_cluster_scanning .

Appendix

A Idealised fit and n-parameter fit pseudo-analysis

In general, if a distribution $\mathcal{H}_b$ of a background events is perfectly known a-priory, given a binning for this distribution, one can calculate the expected number of events in each bin. Being provided directly from the true underlying $\mathcal{H}_b$, this background estimate will on average provide the most efficient tests to discriminate samples drawn from $\mathcal{H}_b$ from samples drawn from an alternative hypothesis $\mathcal{H}_{b+s}$ with signal, compared to any other estimate of the expected background for these samples. Hence we call the expectation from $\mathcal{H}_b$ an "idealised fit".

As discussed extensively in section 3.1, estimation of the expected background is only one step of the analysis. To create a benchmark, we do pseudo-analysis on LHCO R&D dataset represented by spectrum $N_{\text{bkg,orig},b}$ using the other choices defined in section 3.1. Namely, we generate pseudo-experiments by bootstrap resampling the events from $N_{\text{bkg,orig},b}$ and add a number of signal events if needed. The "idealised" background estimation for every pseudo-experiment is equal to $N_{\text{bkg,orig},b}$ itself (as the samples were generated with these expected values). Following the discussion in subsection 3.1.3 we use MLS test statistic between this estimate and the generated pseudo-experiments, to generate null-hypothesis test statistic distribution and its value for signal contaminations and thereafter estimate the significance. Depending on the number of doped signal events, the median and quartile region significance given by this test is provided in the main text in Fig. 5b.

Unfortunately the background model is usually unknown, so for each experimental sample the background should be estimated in some less precise way relying on weaker assumptions.

As a realistic benchmark to our method we explore how sensitive the analysis is using global n-parameter functional to the kind of signal presented in LHCO R&D dataset. We use the binning with 16 bins defined in subsection 3.1.2 and count the number of background events in each bin to get an original background spectrum $N_{\text{bkg,orig},b}$.

For all the fits in this studies, we use Trust Region Reflective nonlinear least squares fitting algorithm implementation from SCIPY python package [133]. The chosen bins generally contain more than 5000 counts, so the Poisson distributions of these counts can be well approximated by a Gaussian distributions with the variances equal to the bin counts. Using variances to scale the summands in the least squares objective we make it equivalent to the maximum likelihood objective for this setup.

First, we fit our 3- and 4-parameter functions to the spectrum to see if the fit is valid. Resulting fits with 13 and 12 degrees of freedom score $\frac{\chi^2_{3-par}}{n_{dof}} \approx 1.201$ and $\frac{\chi^2_{4-par}}{n_{dof}} \approx 1.338$ that correspond to p-values of 0.275 and 0.182 which signify validity of these fits.

Unlike the CS method that doesn't generally rely on the smoothness of the background, global n-parameter takes it as the main assumption, so as $N_{\text{bkg,orig},b}$ already has some statistical fluctuations a distribution resampled from it will have even larger statistical fluctuations than the ones expected for Poisson distribution. To simulate the proper scale Poisson fluctuations in the chosen region for our pseudo-experiments we resample events not from $N_{\text{bkg,orig},b}$ but from the best possible fit. This also negates the systematic error from null-hypothesis not corresponding to the empirical functional form, so these experiments can be viewed as semi-idealised. In a more realistic cases, the space of functions given by all possible parameter values, does not contain the true form of null-hypothesis distribution and can only yield an approximation of it with limited precision. It is usual for fit functions with a small number of parameters, but with increasing number of parameters the function fit problem becomes over-defined and the function can fit the signal bump as well. Experimentally we have observed only insignificant increase in performance when comparing sampling from $N_{\text{bkg,orig},b}$ or from the best fit distributions. On top of the resampled background events we add a number of signal events from signal's original distribution when needed.

The initial parameters of the fit in each experiment are chosen to be equal to the optimal parameters of the initial fit discussed above, so that one gets an "idealised" background fit if no optimisation is done. However, because of the statistical fluctuations and/or added signal contamination, the maximisation of likelihood results in a different set of parameters for this functional form. This error of background mismodeling under its statistical fluctuations and addition of the signal is exactly the type of error we want to demonstrate with this pseudo-analysis.

The results of such analysis for different signal contamination is given in Fig. 5a. We can see that the 3-parameter fit provides a slightly better result than 4-parameter fit as the latter has more flexibility to overfit the signal and the statistical fluctuations. This is so because the samples are drawn from 3- and 4-parameter functions with fixed parameters themselves. If we were to sample from other distribution the error coming from mismatch in true end expected functional forms may switch this ordering but it will reduce both

performances. Therefore, the curves shown in Fig. 5a are upper limits of these realistic n-parameter fit analyses achievable only when the true distribution is described by one of the functions in the chosen parameterised space.

B Sparsity of the jet images

Fig. 6 show that the jet images are very sparsely populated ususally having less tha 100 non-zero pixels per 1600 pixels total.

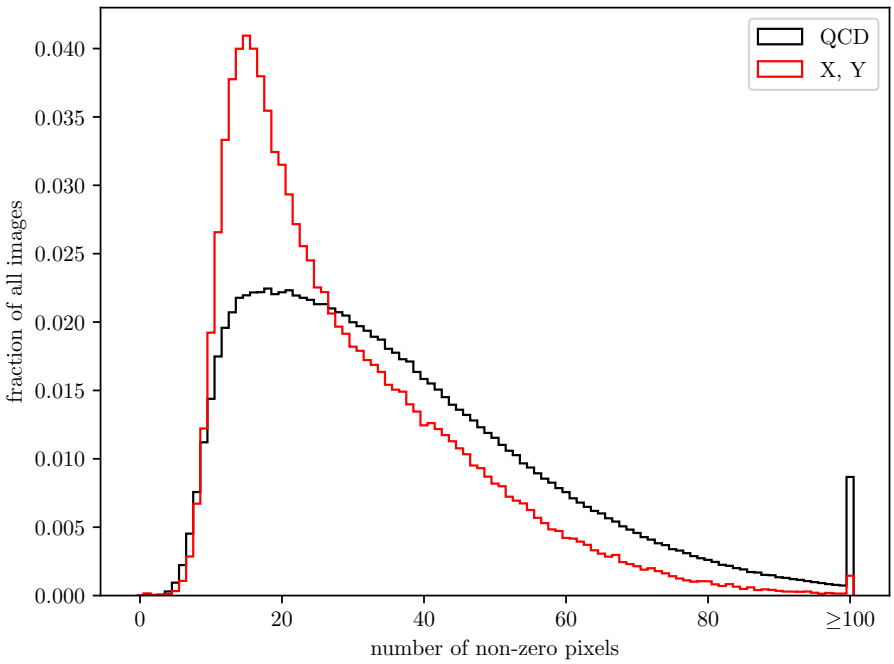


Figure 6: Distribution of images in QCD and top datasets from top-tagging task vs the number of non-0 pixels in them.

C Hyperparameter selection and motivation

In this appendix we give motivation for every not yet discussed choice of hyperparameter in our pseudo-analysis. All the hyperparameter suggestions are done in an unsupervised way coming from general assumptions about signal and background and are not optimised using the truth information from LHCO R&D data. As such the levels of significance improvements may be further increased by performing a dedicated parameter scan for a specific application, however, we recommend to follow the same reasoning when applying CS in other analyses.

Training region: Training on the full spectrum would likely result in each cluster corresponding to a specific mass region, thus the background spectrum for each cluster would not be close to the original mass spectrum. Therefore we perform clustering in a narrow

mass window [3000, 3100] GeV. We choose this window as it lies in the studied region defined in subsection 3.1.1 and has the largest statistic of all other 100 GeV windows.

Number of clusters: The most important parameter we had to choose is the number of clusters k. Two factors play the key role in this choice. On one hand, the number of clusters has to be as large as possible to better narrow down the anomaly-rich region. On the other hand, for a given number of events in the evaluation region and the binning of this region one has to take the number of clusters sufficiently small so that the least populated clusters in the smallest m_{jj} bin $\underset{i,m_{jj}}{min}(N_i(m_{jj}))$ still has enough statistics for a meaningful statistical analysis. We assume that $\underset{k,m_{jj}}{min}(N_i(m_{jj})) = O(50)$ should be sufficient. Using a coarse search, we determine, that for our choice of binning and overall statistic at hand choosing $k = 50$ gives a good trade-off as it has a median of 55 events in smallest cluster-bin and it goes below 20 only 1 time in 1000 pseudo-experiment runs, as can be seen in Fig. 7.

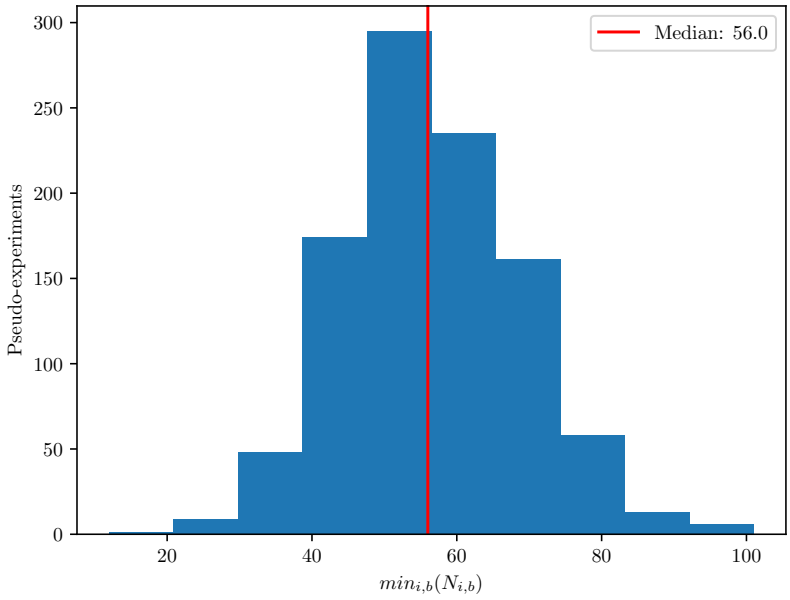


Figure 7: Distribution of the number of jet image counts in the least populated bin of the least populated cluster in each of the 1000 random background only runs of the CS algorithm on backround only data.

Batch size: Scikit-learn [124] documentation states that the parallelisation is performed on all available N_{cores} computing cores if the batch size is $N_{cores} \cdot 256$ or larger. We performed all computations with 8 core parallelisation, thus the natural choice of a batch size was 2048. It is also important to maintain the batch size much larger than the number of clusters to ensure faster convergence.

Outlier fraction: To quantify the performance of our method we introduce the signal fraction improvement score (SFI) that characterises a subset $\mathcal{S}$ of the events in evaluation set $\mathcal{E}$ by the relative increase in the signal to background ratio

$$SFI(\mathcal{S}) = \frac{N_{\text{sig}}(\mathcal{S})}{N_{\text{QCD}}(\mathcal{S})} \frac{N_{\text{QCD}}(\mathcal{E})}{N_{\text{sig}}(\mathcal{E})}. \tag{C.1}$$

Our main assumption is that the signal is distributed in clusters unevenly and there only several clusters have a significantly large SFI. To put it in numbers, we assume that not more than 20% of clusters have SFI of 2 or more. Following this assumption we choose the outlier fraction of 0.2 for outlier robust estimators. This is an ad. hoc prior assumption about the data at hand, and it has to be made prior to analysis and has no way to be validated without knowing the truth lables. Still we can show that this assumption is satisfied in our case with a margin for the pseudo-experiment shown on all the figures of section 3. 5000 signal events were giving an overdensity on the original spectrum that was not identifiable as a deviation from smooth background by human eye (without knowing the background truth), but in Fig. 2 one can easily notice two spectra with a significant bump around 3.5 TeV that stand out of the crowd of other spectra. Unsurprisingly these two spectra have SFIs of 9.1 and 8.9. Three more clusters also have a visible overdensity at this position possessing SFIs of 6.3, 5.6, 4.4. In total, exactly 8 clusters have $SFI > 2$. Still as we will see later only 3 of these clusters have a signal significant enough to be selected as anomalous, showing that our assumption is quite conservative in its limit and either the threshold SFI can be increased or the percentage of clusters to path the threshold reduced for it to still remain a valid assumption. Runs of the analysis on other (pseudo-)experiments behave in the similar manner.

Cluster selection threshold θ: First of all, we use the threshold only for positive deviations as we only search for excesses of events. Apart from the signal-rich outlier clusters the threshold can be passed by signal poor clusters, but only with an expected false positive rate of $1-(1-\text{p-value}_{\mathcal{N}(0,1)}(\theta))^{n_{bins}}$. Then for large enough thresholds the average number of false positives can be estimated as $k \cdot n_{bins} \cdot \text{p-value}_{\mathcal{N}(0,1)}$. Higher thresholds result in lower false positive and lower true positive rates. To retain the sensitivity for statistically small signal we choose to use $\theta = 3$ that will result in approximately $50 \cdot 16 \cdot 0.00135 = 1.08$ signal poor cluster being assigned a false positive label on average. Fig. 3 shows 4 clusters being chosen using this threshold. Three of them have an overdensity at 3.5 TeV and one does not, implying that it is a likely false posive.

Ensemble size: We recommend to take the ensemble size as high as possible, for given computation resource constrains to reduce the width of the test statistic distribution (see appendix F).

Parameter	value	motivation
k	50	$\min_{i,b}(N_{i,b}) = O(50)$ with binning below
mini-batch	2048	$N_{cores} \cdot 256$, must be $\gg k$
Training region	[3, 3.1] TeV	narrow mass window with high statistic
Evaluation region	[3, 4.6] TeV	the n-parameter fit is applicable excluding low statistic regions
Bin width	100 Gev	broad enough to have sufficient statistics in each bin
outlier fraction f	0.2	consistent with assumption on the maximum number of signal clusters
Cluster selection threshold θ	3 σ	low enough to let trough many true positives but high enough to filter most false positives
Test statistic	MLS	simple and specialised for local excesses
Default TS	0	minimal test statistic possible
Ensemble size	15	As large as possible realistic compute limitations

Table 1: Summary of the hyperparameters used in cluster scanning.

D Gaussianity of cluster scanning bin entries

Assumption on the Gaussianity of cluster spectra in each bin can be shown to be valid by standardising the background-only cluster counts in each bin and checking if these distributions match $\mathcal{N}(0,1)$. 50 samples are usually not enough to determine whether the distribution is Gaussian or not, but by marginalising over 16 bins we get 800 samples in total. Fig. 8 shows the said distribution for background only spectra that fits the $\mathcal{N}(0,1)$ distribution well visually and by a consensus of 3 normality tests Shapiro-Wilk, Kolmogorov-Smirnov and Jarque-Bera (note that the p-values for all the tests are high). The same distribution with added signal has some clear outliers, which reflects in lower p-values for the normality tests, but apart from these it still can be well approximated by a unit Gaussian.

E Outlier robust estimators

While searching for outliers, it is preferred to use outlier robust estimators for standard deviation (SD) and mean. We define them as follows: given a sample of observations $S = \{x_1, x_2, ... x_n\}$ we find a median $med(S)$ (which is itself an outlier robust estimator) of this sample and take a subsample $\tilde{S}_f$ that is constructed from S by discarding a fraction $0 < f < 1$ of all samples that have largest absolute distance to this median. In this way we have discarded the outliers. After that we construct estimators $\tilde{\mu}_f = mean(\tilde{S}_f)$ and $\tilde{\sigma}_f = SD(\tilde{S}_f) \cdot g(f)$. If S is a sample from $\mathcal{N}(\mu, \sigma)$ it is obvious that with $\lim_{n\to\infty} \tilde{\mu}_f = \lim_{n\to\infty} mean(S) = \mu$. If one takes S from $\mathcal{N}(0,1)$ and rescales $x_i \to \sigma x_i$, then both estimators transform as $\tilde{\sigma}_f \to \sigma\tilde{\sigma}_f$ and $SD(S) \to \sigma SD(S)$ by definition, so both estimators $\tilde{\sigma}_f$ and $SD(S)$ are proportional to a true σ of the Gaussian distribution.

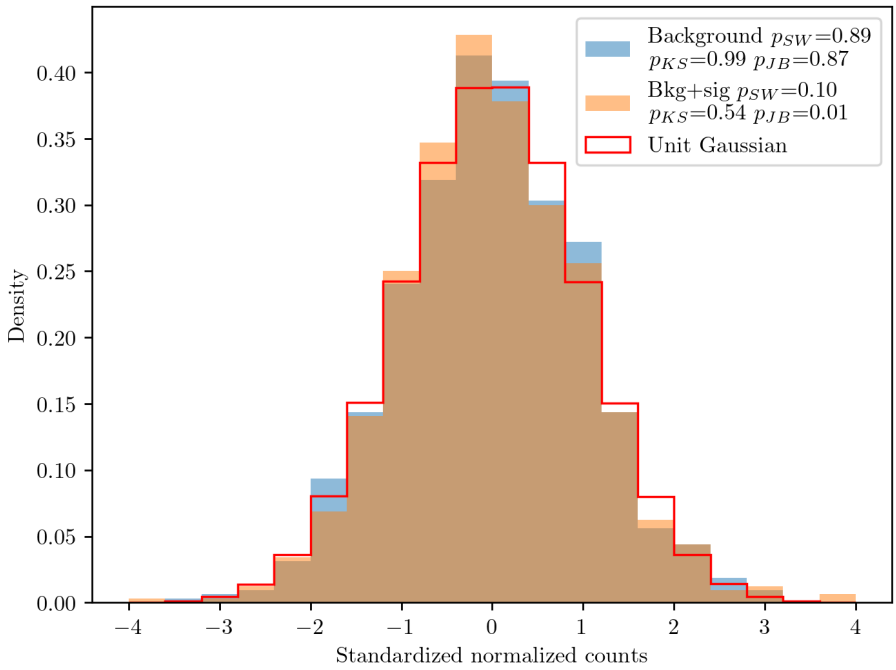


Figure 8: Histogram of deviations from the average of normalised cluster spectra measured in standard deviations of each bin. Blue histogram gives it for signal-free case and orange histogram for case with 5000 injected signal events. Red line show the expected normal distribution.

Both $\tilde{\sigma}_f$ and $SD(S)$ are independent of μ and there are no other parameters of the normal distribution for estimators to depend on, therefore for a family of Gaussian distribution estimators $\tilde{\sigma}_f$ and $SD(S)$ are proportional to each other by some constant factor $g(f)$ in the limit of infinite sample. In other words, adjusting numerically $g(f) = \frac{SD(\mathcal{N}(0,1))}{\tilde{\sigma}_f(\mathcal{N}(0,1))} = \frac{1}{\tilde{\sigma}_f(\mathcal{N}(0,1))}$ is sufficient to sattisfy $\lim\limits_{n\to\infty} \tilde{\sigma}_f = \lim\limits_{n\to\infty} SD(S) = \sigma$. So $\tilde{\mu}_f$ and $\tilde{\sigma}_f$ are unbiased estimators of μ and σ of a normal distribution, although depending on f they are less efficient than usual non-robust mean and SD.

Fig. 9 shows us the cluster spectra from Fig. 2 with subtracted normalised original spectrum (which is only needed for better visualisation as this step has no effect on the standardisation). Fig. 9 also shows the conventional and the outlier robust estimations of mean and SD of the cluster spectra values in each bin. As expected for lower m_{jj} the SD is higher as these deviations is partially caused by the Poisson fluctuations which are proportional to $\sqrt{N_{i,b}}$. We can also see the conventional estimators have a bump around 3.5 TeV that is induced by our outlier signal-rich clusters, while the robust estimators are unaffected by the outliers.

F Calibration distributions

The distribution of the test statistics given by CS without ensembling for all background only pseudo-experiments is shown in Fig. 10a as a histogram. We see that around 300

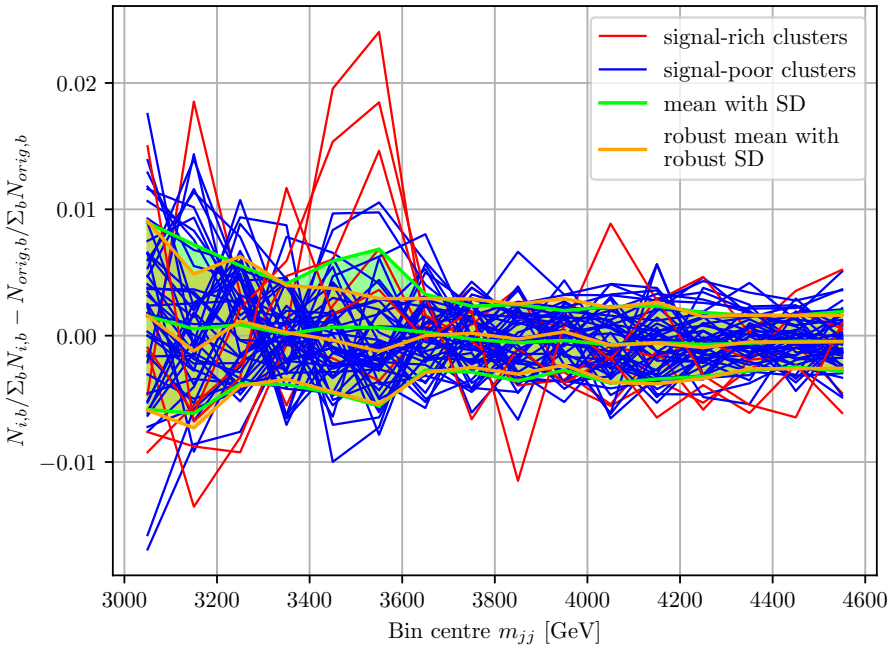


Figure 9: Normalised spectra with subtracted normalised original m_{jj} spectrum. Amount of signal is 5000. The selection of the anomalous clusters is taken from Fig. 4.

of those were assigned test statistic of 0 as they had no clusters selected as anomalous. Other cases where one or more anomalous clusters were selected form a smooth continuous distribution.

The median CS test statistic for 100 signal contaminated pseudo-experiments is represented in Fig. 10a by a vertical line, and the vertical band represent the region between the quartiles of such a test statistic sample. For each signal-doped pseudo-experiment we calculate significance as it is described in subsection 3.1.4. The median significance is quoted in the legend of the figure.

Fig. 10b shows the distribution that is analogous to the one in Fig. 10a, but with an ensemble of 15 runs of CS algorithm for each pseudo-experiment. We notice that the distribution in Fig. 10b is significantly narrower than in 10a which reduces the frequency of background only experiment having large test statistic, thus increasing the sensitivity to signal injection. An additional benefit is that the uncertainty region (between two quarterlies) for each signal doping have significantly decreased which is important for lower uncertainty in the analysis on experimental data on the excess significance or on the exclusion limits.

This motivates, that in general the ensemble size should be taken as large as reasonably possible. Our choice of ensemble size 15 together with the number of pseudo-experiments 3900 were dictated by the computing time and storage memory limits as the amount of full CS algorithm iterations is the product of those numbers (excluding the pseudo-experiments with signal injection)

Finally, Fig. 10c shows that for idealised CS without systematics introduced by mass

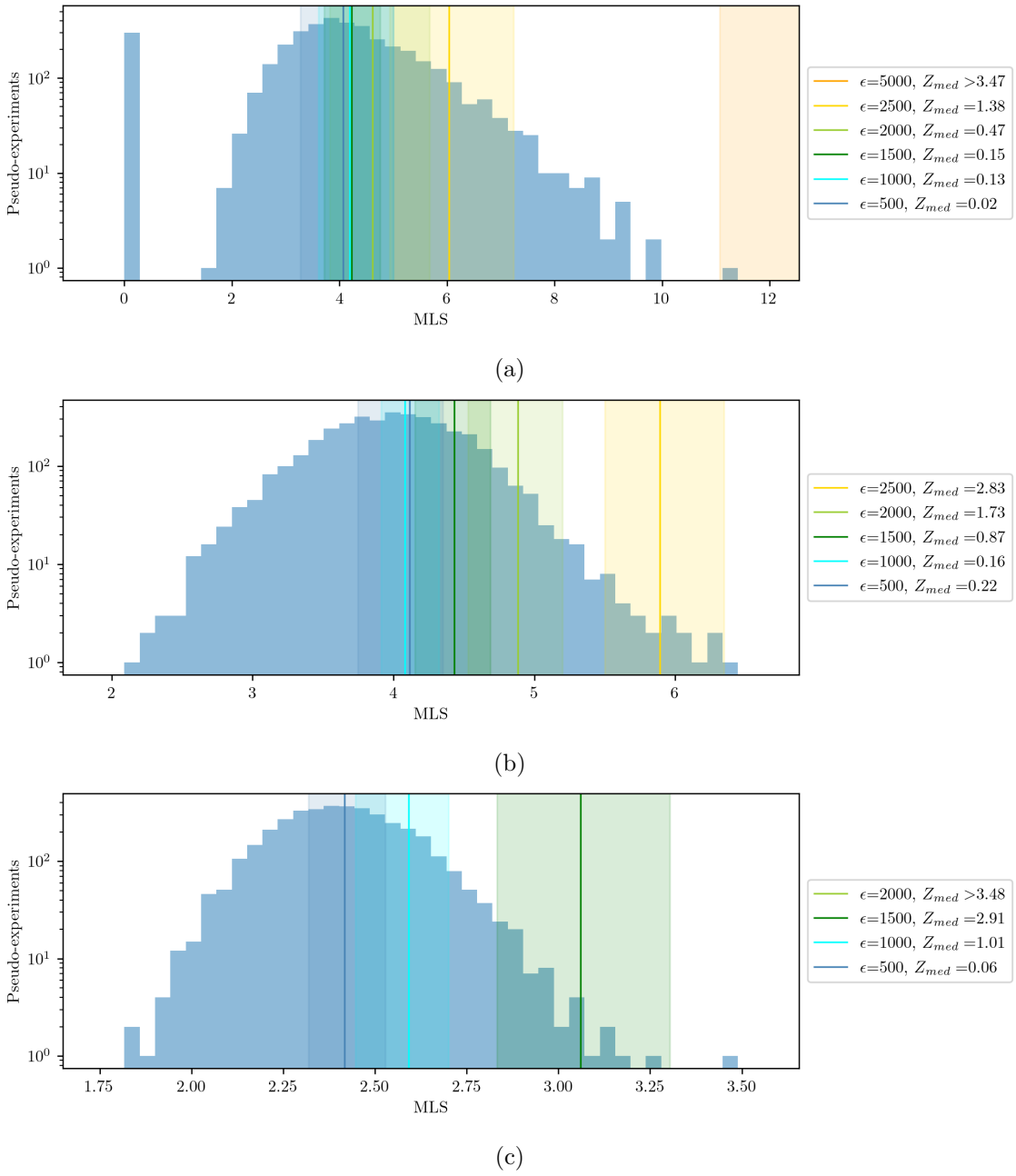


Figure 10: Histogram of the CS test statistic for pseudo-experiments with bootstrapped background only samples. Vertical lines and vertical bands show median and region between lower and higher quartiles of test statistics for pseudo-experiments with signal injection. Several signal injection levels are represented by different colours. Panel (a) shows a case with only 1 initialisation of clusters in CS per pseudo-experiment, panel (b) shows a case for ensembling 15 runs of CS with different intialisations per pseudo-experiment and panel (c) shows a case for ensembling 15 runs of idealised CS with different initialisations per pseudo-experiment.

correlations the MLS between our signal spectrum and background estimate is lower. Moreover, as expected, it improved the sensitivity of the method to the signal. Obviously this technique cannot be utilised in an actual analysis as jet labels are needed to distribute signal and background jets in a different manner. Moreover, as expected, it improved the sensitivity of the method to the signal. Obviously this technique cannot be utilised in an actual analysis as jet labels are needed to distribute signal and background jets in a different manner.

References

[1] ATLAS Collaboration, *Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012) 1–29 [1207.7214].

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* **716** (2012) 30–61 [1207.7235].

[3] R.T. D'Agnolo and A. Wulzer, *Learning New Physics from a Machine*, *Phys. Rev. D* **99** (2019) 015014 [1806.02350].

[4] R.T. D'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning multivariate new physics*, *Eur. Phys. J. C* **81** (2021) 89 [1912.12155].

[5] M. Farina, Y. Nakai and D. Shih, *Searching for New Physics with Deep Autoencoders*, *Phys. Rev. D* **101** (2020) 075021 [1808.08992].

[6] T. Heimel, G. Kasieczka, T. Plehn and J.M. Thompson, *QCD or What?*, *SciPost Phys.* **6** (2019) 030 [1808.08979].

[7] T.S. Roy and A.H. Vijay, *A robust anomaly finder based on autoencoders*, 1903.02032.

[8] O. Cerri, T.Q. Nguyen, M. Pierini, M. Spiropulu and J.-R. Vlimant, *Variational Autoencoders for New Physics Mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036 [1811.10276].

[9] A. Blance, M. Spannowsky and P. Waite, *Adversarially-trained autoencoders for robust unsupervised new physics searches*, *JHEP* **10** (2019) 047 [1905.10384].

[10] J. Hajer, Y.-Y. Li, T. Liu and H. Wang, *Novelty Detection Meets Collider Physics*, *Phys. Rev. D* **101** (2020) 076015 [1807.10261].

[11] A. De Simone and T. Jacques, *Guiding New Physics Searches with Unsupervised Learning*, *Eur. Phys. J. C* **79** (2019) 289 [1807.06038].

[12] A. Mullin, S. Nicholls, H. Pacey, M. Parker, M. White and S. Williams, *Does SUSY have friends? A new approach for LHC event analysis*, *JHEP* **02** (2021) 160 [1912.10625].

[13] B.M. Dillon, D.A. Faroughy and J.F. Kamenik, *Uncovering latent jet substructure*, *Phys. Rev. D* **100** (2019) 056002 [1904.04200].

[14] J.A. Aguilar-Saavedra, J.H. Collins and R.K. Mishra, *A generic anti-QCD jet tagger*, *JHEP* **11** (2017) 163 [1709.01087].

[15] M. Romão Crispim, N.F. Castro, R. Pedro and T. Vale, *Transferability of Deep Learning Models in Searches for New Physics at Colliders*, *Phys. Rev. D* **101** (2020) 035042 [1912.04220].

[16] M. Crispim Romão, N.F. Castro, J.G. Milhano, R. Pedro and T. Vale, *Use of a generalized energy Mover's distance in the search for rare phenomena at colliders*, *Eur. Phys. J. C* **81** (2021) 192 [2004.09360].

[17] O. Amram and C.M. Suarez, *Tag N' Train: a technique to train improved classifiers on unlabeled data*, *JHEP* **01** (2021) 153 [2002.12376].

[18] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette and T. Golling, *Variational autoencoders for anomalous jet tagging*, *Phys. Rev. D* **107** (2023) 016002 [2007.01850].

[19] C.K. Khosa and V. Sanz, *Anomaly Awareness*, *SciPost Phys.* **15** (2023) 053 [2007.14462].

[20] P. Thaprasop, K. Zhou, J. Steinheimer and C. Herold, *Unsupervised Outlier Detection in Heavy-Ion Collisions*, *Phys. Scripta* **96** (2021) 064003 [2007.15830].

[21] S. Alexander, S. Gleyzer, H. Parul, P. Reddy, M.W. Toomey, E. Usai et al., *Decoding Dark Matter Substructure without Supervision*, 2008.12731.

[22] V. Mikuni and F. Canelli, *Unsupervised clustering for collider physics*, *Phys. Rev. D* **103** (2021) 092007 [2010.07106].

[23] M. van Beekveld, S. Caron, L. Hendriks, P. Jackson, A. Leinweber, S. Otten et al., *Combining outlier analysis algorithms to identify new physics at the LHC*, *JHEP* **09** (2021) 024 [2010.07940].

[24] S.E. Park, D. Rankin, S.-M. Udrescu, M. Yunus and P. Harris, *Quasi Anomalous Knowledge: Searching for new physics with embedded knowledge*, *JHEP* **21** (2020) 030 [2011.03550].

[25] D.A. Faroughy, *Uncovering hidden new physics patterns in collider events using Bayesian probabilistic models*, *PoS* **ICHEP2020** (2021) 238 [2012.08579].

[26] T. Golling et al., *The Mass-ive Issue: Anomaly Detection in Jet Physics*, 2303.14134.

[27] P. Chakravarti, M. Kuusela, J. Lei and L. Wasserman, *Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests*, 2102.07679.

[28] J. Batson, C.G. Haaf, Y. Kahn and D.A. Roberts, *Topological Obstructions to Autoencoding*, *JHEP* **04** (2021) 280 [2102.08380].

[29] A. Blance and M. Spannowsky, *Unsupervised event classification with graphs on classical and photonic quantum computers*, *JHEP* **21** (2020) 170 [2103.03897].

[30] B. Bortolato, A. Smolkovič, B.M. Dillon and J.F. Kamenik, *Bump hunting in latent space*, *Phys. Rev. D* **105** (2022) 115009 [2103.06595].

[31] J.H. Collins, P. Martín-Ramiro, B. Nachman and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, *Eur. Phys. J. C* **81** (2021) 617 [2104.02092].

[32] B.M. Dillon, T. Plehn, C. Sauer and P. Sorrenson, *Better Latent Spaces for Better Autoencoders*, *SciPost Phys.* **11** (2021) 061 [2104.08291].

[33] T. Finke, M. Krämer, A. Morandini, A. Mück and I. Oleksiyuk, *Autoencoders for unsupervised anomaly detection in high energy physics*, *JHEP* **06** (2021) 161 [2104.09051].

[34] D. Shih, M.R. Buckley, L. Necib and J. Tamanas, *via machinae: Searching for stellar streams using unsupervised machine learning*, *Mon. Not. Roy. Astron. Soc.* **509** (2021) 5992 [2104.12789].

[35] O. Atkinson, A. Bhardwaj, C. Englert, V.S. Ngairangbam and M. Spannowsky, *Anomaly detection with convolutional Graph Neural Networks*, *JHEP* **08** (2021) 080 [2105.07988].

[36] A. Kahn, J. Gonski, I. Ochoa, D. Williams and G. Brooijmans, *Anomalous jet identification via sequence modeling*, *JINST* **16** (2021) P08012 [2105.09274].

[37] T. Dorigo, M. Fumanelli, C. Maccani, M. Mojsovska, G.C. Strong and B. Scarpa, *RanBox: anomaly detection in the copula space*, *JHEP* **01** (2023) 008 [2106.05747].

[38] S. Caron, L. Hendriks and R. Verheyen, *Rare and Different: Anomaly Scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC*, *SciPost Phys.* **12** (2022) 077 [2106.10164].

[39] E. Govorkova, E. Puljak, T. Aarrestad, M. Pierini, K.A. Woźniak and J. Ngadiuba, *LHC physics dataset for unsupervised New Physics detection at 40 MHz*, *Sci. Data* **9** (2022) 118 [2107.02157].

[40] G. Kasieczka, B. Nachman and D. Shih, *New Methods and Datasets for Group Anomaly Detection From Fundamental Physics*, in *Conference on Knowledge Discovery and Data Mining*, 7, 2021 [2107.02821].

[41] S. Volkovich, F. De Vito Halevy and S. Bressler, *A data-directed paradigm for BSM searches: the bump-hunting example*, *Eur. Phys. J. C* **82** (2022) 265 [2107.11573].

[42] E. Govorkova et al., *Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider*, *Nature Mach. Intell.* **4** (2022) 154 [2108.03986].

[43] B. Ostdiek, *Deep Set Auto Encoders for Anomaly Detection in Particle Physics*, *SciPost Phys.* **12** (2022) 045 [2109.01695].

[44] K. Fraser, S. Homiller, R.K. Mishra, B. Ostdiek and M.D. Schwartz, *Challenges for unsupervised anomaly detection in particle physics*, *JHEP* **03** (2022) 066 [2110.06948].

[45] P. Jawahar, T. Aarrestad, N. Chernyavskaya, M. Pierini, K.A. Wozniak, J. Ngadiuba et al., *Improving Variational Autoencoders for New Physics Detection at the LHC With Normalizing Flows*, *Front. Big Data* **5** (2022) 803685 [2110.08508].

[46] J. Herrero-Garcia, R. Patrick and A. Scaffidi, *A semi-supervised approach to dark matter searches in direct detection data with machine learning*, *JCAP* **02** (2022) 039 [2110.12248].

[47] J.A. Aguilar-Saavedra, *Anomaly detection from mass unspecific jet tagging*, *Eur. Phys. J. C* **82** (2022) 130 [2111.02647].

[48] R. Tombs and C.G. Lester, *A method to challenge symmetries in data with self-supervised learning*, *JINST* **17** (2022) P08024 [2111.05442].

[49] C.G. Lester and R. Tombs, *Using unsupervised learning to detect broken symmetries, with relevance to searches for parity violation in nature. (Previously: "Stressed GANs snag desserts")*, 2111.00616.

[50] V. Mikuni, B. Nachman and D. Shih, *Online-compatible unsupervised nonresonant anomaly detection*, *Phys. Rev. D* **105** (2022) 055006 [2111.06417].

[51] S. Chekanov and W. Hopkins, *Event-Based Anomaly Detection for Searches for New Physics*, *Universe* **8** (2022) 494 [2111.12119].

[52] R.T. d'Agnolo, G. Grosso, M. Pierini, A. Wulzer and M. Zanetti, *Learning new physics from an imperfect machine*, *Eur. Phys. J. C* **82** (2022) 275 [2111.13633].

[53] F. Canelli, A. de Cosa, L.L. Pottier, J. Niedziela, K. Pedro and M. Pierini, *Autoencoders for semivisible jet detection*, *JHEP* **02** (2022) 074 [2112.02864].

[54] V.S. Ngairangbam, M. Spannowsky and M. Takeuchi, *Anomaly detection in high-energy physics using a quantum autoencoder*, *Phys. Rev. D* **105** (2022) 095004 [2112.04958].

[55] L. Bradshaw, S. Chang and B. Ostdiek, *Creating simple, interpretable anomaly detectors for new physics in jet substructure*, *Phys. Rev. D* **106** (2022) 035014 [2203.01343].

[56] J.A. Aguilar-Saavedra, *Taming modeling uncertainties with mass unspecific supervised tagging*, *Eur. Phys. J. C* **82** (2022) 270 [2201.11143].

[57] T. Buss, B.M. Dillon, T. Finke, M. Krämer, A. Morandini, A. Mück et al., *What's anomalous in LHC jets?*, *SciPost Phys.* **15** (2023) 168 [2202.00686].

[58] S. Alvi, C.W. Bauer and B. Nachman, *Quantum anomaly detection for collider physics*, *JHEP* **02** (2023) 220 [2206.08391].

[59] B.M. Dillon, R. Mastandrea and B. Nachman, *Self-supervised anomaly detection for new physics*, *Phys. Rev. D* **106** (2022) 056005 [2205.10380].

[60] M. Birman, B. Nachman, R. Sebbah, G. Sela, O. Turetz and S. Bressler, *Data-directed search for new physics based on symmetries of the SM*, *Eur. Phys. J. C* **82** (2022) 508 [2203.07529].

[61] M. Letizia, G. Losapio, M. Rando, G. Grosso, A. Wulzer, M. Pierini et al., *Learning new physics efficiently with nonparametric methods*, *Eur. Phys. J. C* **82** (2022) 879 [2204.02317].

[62] C. Fanelli, J. Giroux and Z. Papandreou, *'Flux+Mutability': a conditional generative approach to one-class classification and anomaly detection*, *Mach. Learn. Sci. Tech.* **3** (2022) 045012 [2204.08609].

[63] T. Finke, M. Krämer, M. Lipp and A. Mück, *Boosting mono-jet searches with model-agnostic machine learning*, *JHEP* **08** (2022) 015 [2204.11889].

[64] R. Verheyen, *Event Generation and Density Estimation with Surjective Normalizing Flows*, *SciPost Phys.* **13** (2022) 047 [2205.01697].

[65] B.M. Dillon, L. Favaro, T. Plehn, P. Sorrenson and M. Krämer, *A Normalized Autoencoder for LHC Triggers*, 2206.14225.

[66] S. Caron, R.R. de Austri and Z. Zhang, *Mixture-of-Theories training: can we find new physics and anomalies better by mixing physical theories?*, *JHEP* **03** (2023) 004 [2207.07631].

[67] S.E. Park, P. Harris and B. Ostdiek, *Neural embedding: learning the embedding of the manifold of physics data*, *JHEP* **07** (2023) 108 [2208.05484].

[68] J.F. Kamenik and M. Szewc, *Null hypothesis test for anomaly detection*, *Phys. Lett. B* **840** (2023) 137836 [2210.02226].

[69] G. Kasieczka, R. Mastandrea, V. Mikuni, B. Nachman, M. Pettee and D. Shih, *Anomaly detection under coordinate transformations*, *Phys. Rev. D* **107** (2023) 015009 [2209.06225].

[70] J.Y. Araz and M. Spannowsky, *Quantum-probabilistic Hamiltonian learning for generative modelling & anomaly detection*, 2211.03803.

[71] J. Schuhmacher, L. Boggia, V. Belis, E. Puljak, M. Grossi, M. Pierini et al., *Unravelling physics beyond the standard model with classical and quantum anomaly detection*, *Mach. Learn. Sci. Tech.* **4** (2023) 045031 [2301.10787].

[72] S. Roche, Q. Bayer, B. Carlson, W. Ouligian, P. Serhiayenka, J. Stelzer et al., *Nanosecond anomaly detection with decision trees for high energy physics and real-time application to exotic Higgs decays*, 2304.03836.

[73] L. Vaslin, V. Barra and J. Donini, *GAN-AE: an anomaly detection algorithm for New Physics search in LHC data*, *Eur. Phys. J. C* **83** (2023) 1008 [2305.15179].

[74] ATLAS Collaboration, *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*, *Phys. Rev. D* **108** (2023) 052009 [2306.03637].

[75] S.V. Chekanov and R. Zhang, *Enhancing the hunt for new phenomena in dijet final-states using anomaly detection filter at High-Luminosity Large Hadron Collider*, 2308.02671.

[76] CMS ECal Collaboration, *Autoencoder-based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter*, 2309.10157.

[77] G. Bickendorf, M. Drees, G. Kasieczka, C. Krause and D. Shih, *Combining Resonant and Tail-based Anomaly Detection*, 2309.12918.

[78] M. Freytsis, M. Perelstein and Y.C. San, *Anomaly Detection in Presence of Irrelevant Features*, 2310.13057.

[79] E.M. Metodiev, J. Thaler and R. Wynne, *Anomaly Detection in Collider Physics via Factorized Observables*, 2312.00119.

[80] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman and D. Shih, *Machine learning in the search for new fundamental physics*, *Nature Reviews Physics* **4** (2022) 399 [2112.03769].

[81] G. Kasieczka, B. Nachman, D. Shih, O. Amram, A. Andreassen, K. Benkendorfer et al., *The LHC olympics 2020 a community challenge for anomaly detection in high energy physics*, *Reports on Progress in Physics* **84** (2021) 124201 [2101.08320].

[82] T. Aarrestad, M. van Beekveld, M. Bona, A. Boveia, S. Caron, J. Davies et al., *The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider*, *SciPost Phys.* **12** (2022) 043 [2105.14027].

[83] V. Belis, P. Odagiu and T.K. Aarrestad, *Machine learning for anomaly detection in particle physics*, *Rev. Phys.* **12** (2024) 100091 [2312.14190].

[84] J.H. Collins, P. Martín-Ramiro, B. Nachman and D. Shih, *Comparing weak- and unsupervised methods for resonant anomaly detection*, *The European Physical Journal C* **81** (2021) [2104.02092].

[85] T. Golling, G. Kasieczka, C. Krause, R. Mastandrea, B. Nachman, J.A. Raine et al., *The Interplay of Machine Learning–based Resonant Anomaly Detection Methods*, 2307.11157.

[86] ATLAS Collaboration, *Dijet resonance search with weak supervision using* $\sqrt{s} = 13$ *TeV pp collisions in the ATLAS detector*, *Physical Review Letters* **125** (2020) 131801 [2005.02983].

[87] ATLAS Collaboration, *Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, 2023. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2023-022.

[88] G. Kasieczka, B. Nachman and D. Shih, *Official Datasets for LHC Olympics 2020 Anomaly Detection Challenge (Version v6)*, 2019. 10.5281/zenodo.4536624.

[89] J.H. Kim, K. Kong, B. Nachman and D. Whiteson, *The motivation and status of two-body resonance decays after the LHC run 2 and beyond*, *Journal of High Energy Physics* **2020** (2020) [1907.06659].

[90] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[91] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., *DELPHES 3: a modular framework for fast simulation of a generic collider experiment*, *Journal of High Energy Physics* **2014** (2014) 026.

[92] A. Mertens, *New features in delphes 3*, *Journal of Physics: Conference Series* **608** (2015) 012045.

[93] M. Selvaggi, *Delphes 3: A modular framework for fast-simulation of generic collider experiments*, *Journal of Physics: Conference Series* **523** (2014) 012033.

[94] M. Cacciari, G.P. Salam and G. Soyez, *The anti-ktjet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063–063 [0802.1189].

[95] M. Cacciari, G.P. Salam and G. Soyez, *Fastjet user manual*, *The European Physical Journal C* **72** (2012) [1111.6097].

[96] E. Rodrigues et al., *The Scikit HEP Project – overview and prospects*, *EPJ Web Conf.* **245** (2020) 06028 [2007.03577].

[97] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman and A. Schwartzman, *Jet-images — deep learning edition*, *Journal of High Energy Physics* **2016** (2016) [1511.05190].

[98] G. Kasieczka, T. Plehn, A. Butter, K. Cranmer, D. Debnath, B.M. Dillon et al., *The machine learning landscape of top taggers*, *SciPost Physics* **7** (2019) [1902.09914].

[99] S. Macaluso and D. Shih, *Pulling out all the tops with computer vision and deep learning*, *Journal of High Energy Physics* **2018** (2018) [1803.00107].

[100] T. Finke, M. Krämer, A. Morandini, A. Mück and I. Oleksiyuk, *Autoencoders for unsupervised anomaly detection in high energy physics*, *JHEP* **2021** (2021) [2104.09051].

[101] A. Hallin, J. Isaacson, G. Kasieczka, C. Krause, B. Nachman, T. Quadfasel et al., *Classifying anomalies through outer density estimation*, *Phys. Rev. D* **106** (2022) 055006 [2109.00546].

[102] J.A. Raine, S. Klein, D. Sengupta and T. Golling, *CURTAINs for your sliding window: Constructing unobserved regions by transforming adjacent intervals*, *Front. Big Data* **6** (2023) 899345 [2203.09470].

[103] CMS Collaboration, *Search for narrow resonances in the b-tagged dijet mass spectrum in proton-proton collisions at s=13 TeV*, *Phys. Rev. D* **108** (2023) 012009 [2205.01835].

[104] ATLAS Collaboration, *Search for new phenomena in the dijet mass distribution using pp collision data at* $\sqrt{s} = 8$ TeV *with the atlas detector*, *Phys. Rev. D* **91** (2015) 052007 [1407.1376].

[105] ATLAS Collaboration, *Search for new phenomena in dijet mass and angular distributions from pp collisions at* $\sqrt{s} =$ *13 tev with the ATLAS detector*, *Physics Letters B* **754** (2016) 302–322 [1512.01530].

[106] CMS Collaboration, *Search for dijet resonances in proton–proton collisions at* $\sqrt{s} = 13$ *tev and constraints on dark matter and other models*, *Physics Letters B* **769** (2017) 520–542.

[107] CMS Collaboration, *Search for narrow resonances decaying to dijets in proton-proton collisions at* $\sqrt{s} = 13$ *tev*, *Physical Review Letters* **116** (2016) [1512.01224].

[108] CMS Collaboration, *Search for narrow resonances in dijet final states at* $\sqrt{s} = 8$ *tev with the novel cms technique of data scouting*, *Physical Review Letters* **117** (2016) .

[109] ATLAS Collaboration, *A search for new physics in dijet mass and angular distributions in pp collisions at* $\sqrt{s} = 7$ *tev measured with the ATLAS detector*, *New Journal of Physics* **13** (2011) 053044 [1512.01530].

[110] CMS Collaboration, *Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at* $\sqrt{s} = 8$ TeV, *Phys. Rev. D* **91** (2015) 052009 [1501.04198].

[111] CMS Collaboration, *Search for narrow resonances using the dijet mass spectrum in pp collisions at* $\sqrt{s}$=8 TeV, *Phys. Rev. D* **87** (2013) 114015 [1302.4794].

[112] CMS Colaboration, *Search for narrow resonances and quantum black holes in inclusive and b-tagged dijet mass spectra from pp collisions at* $\sqrt{s} = 7$ *tev*, *Journal of High Energy Physics* **2013** (2013) [1210.2387].

[113] ATLAS Collaboration, *Search for new physics in the dijet mass distribution using 1* fb^{-1} *of pp collision data at* $\sqrt{s} = 7$ *tev collected by the atlas detector*, *Physics Letters B* **708** (2012) 37 [1108.6311].

[114] ATLAS Collaboration, *Atlas search for new phenomena in dijet mass and angular distributions using pp collisions at* $\sqrt{s} = 7$*tev*, *Journal of High Energy Physics* **2013** (2013) [1210.1718].

[115] CMS Collaboration, *Search for resonances in the dijet mass spectrum from 7 tev pp collisions at cms*, *Physics Letters B* **704** (2011) 123–142 [1107.4771].

[116] CDF Collaboration, *Search for new particles decaying into dijets in proton-antiproton collisions at* $\sqrt{s} = 1.96$ *tev*, *Physical Review D* **79** (2009) [0812.4036].

[117] ATLAS Collaboration, *A search for new physics in dijet mass and angular distributions inppcollisions at* $\sqrt{s} = 7$ *tev measured with the atlas detector*, *New Journal of Physics* **13** (2011) 053044 [1103.3864].

[118] CMS Collaboration, *Search for dijet resonances in 7 tev pp collisions at cms*, *Phys. Rev. Lett.* **105** (2010) 211801 [1010.0203].

[119] ATLAS Collaboration, *Search for low-mass dijet resonances using trigger-level jets with the atlas detector in pp collisions at* $\sqrt{s} = 13$ TeV, *Physical Review Letters* **121** (2018) [1804.03496].

[120] K. Sekhon, R.C. Edgar and D. Amidei, *SWiFt: Sliding Window Fit Method for Resonance Searches*, Tech. Rep. ATL-COM-PHYS-2018-161, CERN, Geneva (2018).

[121] J. Alison, F.A. Di Bello, S. Chekanov, A. Coccaro, T. Golling, G. Facini et al., *Search for resonances in the di-jet mass distribution with one or two jets identified as b-jets in proton-proton collisions at* $\sqrt{s}$*=13 TeV with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2016-1561, CERN, Geneva (2016).

[122] G. Kasieczka, B. Nachman, M.D. Schwartz and D. Shih, *Automating the abcd method with machine learning*, *Phys. Rev. D* **103** (2021) 035021 [2007.14400].

[123] L. Vaslin, S. Calvet, V. Barra and J. Donini, *pybumphunter: A model independent bump hunting tool in python for high energy physics analyses*, 2208.14760.

[124] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *J. Machine Learning Res.* **12** (2011) 2825 [1201.0490].

[125] D. Arthur and S. Vassilvitskii, *K-means++: The advantages of careful seeding*, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, (USA), p. 1027–1035, Society for Industrial and Applied Mathematics, 2007.

[126] A. Andreassen, B. Nachman and D. Shih, *Simulation Assisted Likelihood-free Anomaly Detection*, *Phys. Rev. D* **101** (2020) 095004 [2001.05001].

[127] T. Golling, S. Klein, R. Mastandrea and B. Nachman, *Flow-enhanced transportation for anomaly detection*, *Phys. Rev. D* **107** (2023) 096025 [2212.11285].

[128] D. Sengupta, S. Klein, J.A. Raine and T. Golling, *CURTAINs Flows For Flows: Constructing Unobserved Regions with Maximum Likelihood Estimation*, 2305.04646.

[129] E. Buhmann, C. Ewen, G. Kasieczka, V. Mikuni, B. Nachman and D. Shih, *Full phase space resonant anomaly detection*, 2023.

[130] D. Sengupta, M. Leigh, J.A. Raine, S. Klein and T. Golling, *Improving new physics searches with diffusion models for event observables and jet constituents*, 2312.10130.

[131] M. Frate, K. Cranmer, S. Kalia, A. Vandenberg-Rodes and D. Whiteson, *Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes*, 1709.05681.

[132] K. Cranmer and I. Yavin, *RECAST: Extending the Impact of Existing Analyses*, *JHEP* **04** (2011) 038 [1010.2506].

[133] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261.

arXiv:2402.17638v1 [hep-ph] 27 Feb 2024

Dual chiral density wave induced oscillating Casimir effect

Daisuke Fujii [1,2,*] Katsumasa Nakayama [3,†] and Kei Suzuki [1,‡]

[1] *Advanced Science Research Center, Japan Atomic Energy Agency (JAEA), Tokai, 319-1195, Japan*
[2] *Research Center for Nuclear Physics, Osaka University, Ibaraki 567-0048, Japan*
[3] *RIKEN Center for Computational Science, Kobe, 650-0047, Japan*

The Casimir effect is known to be induced from photon fields confined by a small volume, and also its fermionic counterpart has been predicted in a wide range of quantum systems. Here, we investigate what types of Casimir effects can occur from quark fields in dense and thin quark matter. In particular, in the dual chiral density wave, which is a possible ground state of dense quark matter, we find that the Casimir energy oscillates as a function of the thickness of matter. This oscillating Casimir effect is regarded as an analog of that in Weyl semimetals and is attributed to the Weyl points in the momentum space of quark fields. In addition, we show that an oscillation is also induced from the quark Fermi sea, and the total Casimir energy is composed of multiple oscillations.

I. INTRODUCTION

The Casimir effect, proposed by Casimir [1], is crucially important for understanding small-volume physics in quantum field theory (see Refs. [2–6] for reviews). Casimir predicted that the decrease of the zero-point energy of photon fields by two parallel plates would cause an attractive force for the plates, which are the so-called Casimir energy and the Casimir force. The Casimir force was experimentally verified about fifty years later [7, 8].

Beyond academic interest, the engineering application of the Casimir effect to nanotechnology (Casimir engineering) has recently attracted much attention [9]. The most typical feature of the Casimir effect is an attractive force. On the other hand, when one tunes the permittivity and/or permeability of plates and medium, a repulsive force can be also realized [10–13]. In contrast to such attractive or repulsive Casimir effects, it would be interesting that the third type of Casimir effect, where "third" means that attraction or repulsion is not fixed. For example, under a setup, the sign of Casimir energy flips from attraction to repulsion as the separation distance increases, which may be called the *sign-flipping Casimir effect* (e.g., see Refs. [14–19]). As another example, the value of Casimir energy can oscillate as a function of distance, which may be called the *oscillating Casimir effect* (e.g., see Refs. [20–32]). Among them, Ref. [31] found that it occurs inside Weyl semimetals, where the origins of the oscillation are Weyl points (WPs) at finite momenta in the dispersion relations of Weyl fermions. Such new types of Casimir effects are not only of theoretical interest but also will be important for Casimir engineering.

The Casimir effect is also important for elucidating quark and gluon dynamics described by Quantum Chromodynamics (QCD) in a small volume. For example, (i) as an ab initio method for solving QCD, numerical simulations of lattice QCD are done in finite volume (e.g., in a box of a few fm size), and finite-volume effects must be understood. Since the finite-volume effect for zero-point energy is nothing but the Casimir effect, its understanding is helpful for interpretations of results in small-volume simulations. (ii) In relativistic heavy-ion collision experiments, quark-gluon plasma is produced as a fireball with a size of a few fm. Therefore, we must well understand the contribution of the Casimir effect to physics inside the fireball and near its boundary. (iii) In the interiors of neutron stars, dense quark matter as well as nuclear matter may exist. Depending on the microscopic density profile inside stars, there may be small regions of quark matter.

Under these motivations, this paper focuses on what types of Casimir effects can occur in various phases of quark matter where the thickness of z direction is extremely short (i.e., "thin" quark matter) as illustrated in Fig. 1. In particular, in this paper, we propose a QCD

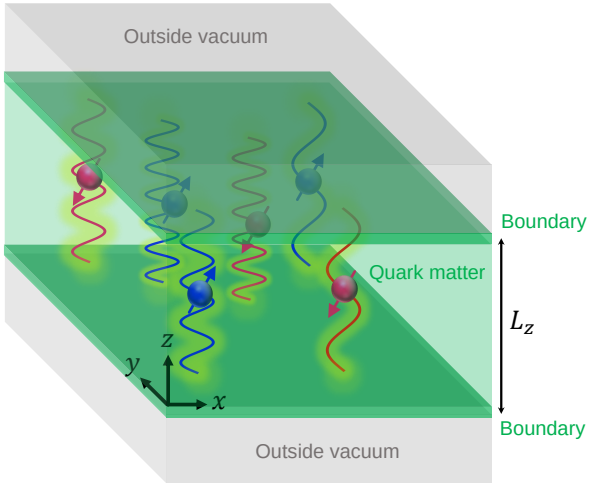


FIG. 1. Schematics of the Casimir effect for quark fields inside thin quark matter, where the quark matter is confined between boundary conditions with a thickness L_z. The "outside vacuum" should be regarded as an arbitrary vacuum consistent with the boundary conditions.

* daisuke@rcnp.osaka-u.ac.jp
† katsumasa.nakayama@riken.jp
‡ k.suzuki.2010@th.phys.titech.ac.jp

counterpart of the oscillating Casimir effect predicted in Weyl semimetals. This is realized for fermion fields in the dual chiral density wave (DCDW) phase [33] (also see Refs. [34–80] and Ref. [81] for a review). The DCDW phase is a candidate for the ground state of quark/nuclear matter in the density region of $\rho/\rho_0 = 3 \sim 6$ with the normal nuclear density $\rho_0 = 0.16\ \mathrm{fm}^{-3}$ [46, 47], and complex dispersion relations of fermions in this phase can be an origin of the oscillating Casimir effect.

In this paper, we aim to investigate types of Casimir effects induced from quark fields in dense quark matter. To accomplish this, this paper is organized as follows. In Sec. II, we formulate the zero-point energy and the Casimir energy within the Nambu–Jona-Lasinio (NJL) model [82, 83] as an effective model of QCD. In Sec. III, we show our numerical results and classify the types of Casimir effects in three density regions. Section IV is devoted to our conclusion and outlook.

II. FORMULATION

A. Dual chiral density waves

A chiral density wave is a spatially inhomogeneous chiral condensate, which is represented as a position-dependent Lorentz-scalar condensate made of the Dirac fields, ψ and $\bar{\psi} = \psi^\dagger\gamma_0$: $\langle\bar{\psi}\psi\rangle \propto \cos(\vec{q}\cdot\vec{r})$ with a wavenumber vector $\vec{q}$ and a position vector $\vec{r}$. Among various types of chiral density waves, in this study, we focus on the DCDW, where both the scalar and pseudoscalar condensates are position-dependent [33],

$$\langle\bar{\psi}\psi\rangle = \Delta\cos(\vec{q}\cdot\vec{r}), \quad \langle\bar{\psi}i\gamma_5\psi\rangle = \Delta\sin(\vec{q}\cdot\vec{r}), \tag{1}$$

where Δ is the amplitude of the DCDW and is regarded as the radius of the chiral circle, $\langle\bar{\psi}\psi\rangle^2 + \langle\bar{\psi}i\gamma_5\psi\rangle^2 = \Delta^2$.

B. NJL model

To investigate the DCDW phase of quark matter, in this work, we use the NJL model [82, 83] (see Refs. [84–87] for reviews), which was done in Refs. [41, 44, 46, 47] for early studies. The Lagrangian density of the two-flavor NJL model is written as

$$\mathcal{L}_{\mathrm{NJL}} = \bar{\psi}(i\partial\!\!\!/ + \mu\gamma_0)\psi + G[(\bar{\psi}\psi)^2 + (\bar{\psi}i\gamma_5\vec{\tau}\psi)^2], \tag{2}$$

where μ is the quark chemical potential, G is the coupling constant of the four-point interactions, and $\vec{\tau} = (\tau_1, \tau_2, \tau_3)$ is the Pauli matrix for the isospin. Here, we apply the mean-field ansatz for the DCDW,

$$\langle\bar{\psi}\psi\rangle = \Delta\cos(\vec{q}\cdot\vec{r}), \quad \langle\bar{\psi}i\gamma_5\tau_3\psi\rangle = \Delta\sin(\vec{q}\cdot\vec{r}), \tag{3}$$

$$\langle\bar{\psi}i\gamma_5\tau_1\psi\rangle = 0, \quad \langle\bar{\psi}i\gamma_5\tau_2\psi\rangle = 0, \tag{4}$$

where Δ is the amplitude of the DCDW, and $\vec{q} = (0, 0, q)$ is the wavenumber of the DCDW propagating in the z direction. Using this ansatz, the mean-field (MF) Lagrangian density is

$$\mathcal{L}_{\mathrm{MF}} = \bar{\psi}\big[i\partial\!\!\!/ + \mu\gamma_0 - M\big(\cos(qz) + i\gamma_5\tau_3\sin(qz)\big)\big]\psi - \frac{M^2}{4G}, \tag{5}$$

where $M = -2G\Delta$.

Using a local chiral transformation (which may be called a Weinberg transformation), $\psi e^{i\gamma_5\tau_3 qz/2} \to \psi_{\mathrm{W}}$, $\bar{\psi}e^{i\gamma_5\tau_3 qz/2} \to \bar{\psi}_{\mathrm{W}}$, the original quark field ψ is redefined as a new field ψ_{W}, and the position dependence of the original Lagrangian (5) is removed in the redefined quark fields. Then, from the diagonalization of the inverse quark propagator in momentum space, we obtain the four eigenvalues of (quasi-)quarks, $\omega_\pm$ (the positive-energy modes if $\mu = 0$) and $\tilde{\omega}_\pm$ (the negative-energy modes):

$$\omega_\pm = E_s - \mu, \tag{6}$$

$$\tilde{\omega}_\pm = -E_s - \mu, \tag{7}$$

where

$$E_s = \sqrt{k^2 + M^2 + q^2/4 \pm \sqrt{(qk_z)^2 + M^2q^2}}, \tag{8}$$

and $k^2 = k_x^2 + k_y^2 + k_z^2$. When $q = 0$, the dispersion relations return to those of usual massive quarks with a mass M, where the two modes with different spins are degenerate. When $q \neq 0$, the two modes are split, which is labeled by $s = \pm$. In this work, we focus on the case of $M < q/2$, which is a typical situation of the DCDW phase.[1]

Using these eigenvalues, the thermodynamic potential (per a spatial volume $V = L_xL_yL_z$) at temperature $T = 1/\beta$ is written as

$$\frac{\Omega(T)}{V} = -N_fN_c\int\frac{d^3k}{(2\pi)^3}\sum_{s=\pm}\left[E_s + \frac{1}{\beta}\ln\left(1 + e^{-\beta(E_s-\mu)}\right) + \frac{1}{\beta}\ln\left(1 + e^{-\beta(E_s+\mu)}\right)\right] + \frac{M^2}{4G}. \tag{9}$$

In this work we fix the number of flavors as $N_f = 2$ and the number of colors as $N_c = 3$.

By taking the zero-temperature limit $T \to 0$ of the thermodynamic potential (9), we obtain

$$\frac{\Omega(T\to 0)}{V} \equiv \frac{E_0}{L_z} = -N_fN_c\int\frac{d^3k}{(2\pi)^3}\sum_{s=\pm}\left[E_s + (\mu - E_s)\,\theta(\mu - E_s)\right] + \frac{M^2}{4G}. \tag{10}$$

Here, the first term is regarded as the contribution from the Dirac sea, i.e., the negative zero-point energy from

[1] The case of $M \geq q/2$ is also interesting, but the WP-induced oscillating Casimir effect does not occur.

the fermion fields. The second term with the step function is the contribution from the Fermi sea of quarks. The third term is the positive energy from the order parameter M. To provide a Casimir-effect-like picture also for the Fermi sea contribution (as well as the Dirac sea), we have denoted Eq. (10) as E_0/L_z where E_0 is the zero-point energy per unit area.

Note that in the traditional treatment of the NJL model, we determine the values of the order parameters M and q by minimizing Eq. (10) at a fixed μ. In a finite volume, the thermodynamic potential in a fixed volume is minimized, and the values of M and q should be determined. This type of analysis is important for studying the phase diagram in finite volume, but it is not the purpose of this work. The purpose of this work is to investigate the feasibility of the oscillating Casimir effect in dense quark matter.

C. Casimir energy with Lifshitz formula

The Casimir energy is defined as a finite-volume effect for the zero-point energy (10). In a finite volume, the momentum integral with respect to the three-dimensional momentum of quarks in Eq. (10) is replaced by a discrete sum. In this paper, we impose the periodic boundary conditions (PBC) on quark fields at $z = 0$ and $z = L_z$, where the z component of quark momentum is discretized as $k_z \to 2n\pi/L_z$ where $n = 0, 1, \cdots, \infty$.[2] Note that for the PBC, there is no outside vacuum as in Fig. 1, but the discretization of momentum (and corresponding eigenvalues) can induce a nonzero Casimir energy.

The zero-point energy (10) as the infinite integral and the corresponding infinite sum is divergent (unless an energy or a momentum cutoff is introduced), but the Casimir energy should be finite after using a mathematical technique. In this paper, we propose two approaches: (i) the Lifshitz formula and (ii) lattice regularizations.

Using the first approach, at zero temperature $T = 0$ and zero chemical potential $\mu = 0$, the analytical solution for the Casimir energy (per unit area) from quark fields in the DCDW phase can be written as

$$E_{\rm Cas} = -4N_f N_c \sum_{s=\pm} \int_0^\infty \frac{d\xi}{2\pi} \int \frac{dk_x dk_y}{(2\pi)^2} \ln\left[1 - e^{-L_z k_z^{[s]}}\right],$$
$$k_z^{[\pm]} = \sqrt{k_\perp^2 + M^2 + \xi^2 - \frac{q^2}{4} \mp iq\sqrt{k_\perp^2 + \xi^2}}, \tag{11}$$

where the first integral variable is the imaginary part ξ of imaginary frequency $\omega \equiv i\xi$, and $k_\perp^2 \equiv k_x^2 + k_y^2$. The overall factor of $2N_f N_c$ means the degrees of freedom for particle-antiparticle, flavors, and colors, and the spin degrees of freedom are labeled as $s = \pm$. The overall factor of -2 and the factor of $-e^{-L_z k_z^{[s]}}$ are a property of the PBC on fermion fields. For the MIT bag boundary condition, see Appendix A.

Note that Eq. (11) is one of the main findings in this paper, which is an analog of the Lifshitz formula [88] to calculate the Casimir effect for photon fields. When we substitute $M = q = 0$ and $q = 0$, Eq. (11) returns to the formulas for the usual massless and massive quarks, respectively.

Furthermore, we define a dimensionless quantity, which we call the *Casimir coefficient*, by multiplying L_z^3,

$$C_{\rm Cas}^{[3]} \equiv L_z^3 E_{\rm Cas}. \tag{12}$$

This is a convenient quantity for checking the L_z dependence of the Casimir energy.

D. Casimir energy with a lattice regularization

As another approach to calculate the Casimir energy, we use lattice regularizations [28–32, 89–96], which is regarded as a generalization of the original definition by Casimir [1]. Then, the Casimir energy (per unit area) on the three-dimensional lattice with a lattice spacing a is defined as

$$E_{\rm Cas}^{\rm Lat} = E_0^{\rm sum} - E_0^{\rm int}, \tag{13}$$
$$E_0^{\rm sum} = -\frac{N_f N_c}{a^3} \sum_{s=\pm} \int_{\rm BZ} \frac{d(ak_x)d(ak_y)}{(2\pi)^2} \times \sum_n^{\rm BZ} \left[\frac{1}{2}a|E_{s,n}^{\rm Lat} - \mu| + \frac{1}{2}a|E_{s,n}^{\rm Lat} + \mu|\right], \tag{14}$$
$$E_0^{\rm int} = -\frac{N_f N_c}{a^3} \sum_{s=\pm} \int_{\rm BZ} \frac{d(ak_x)d(ak_y)d(ak_z)}{(2\pi)^3} \times N_z\left[\frac{1}{2}a|E_s^{\rm Lat} - \mu| + \frac{1}{2}a|E_s^{\rm Lat} + \mu|\right], \tag{15}$$

where the dispersion relations $E_\pm^{\rm Lat}$ on the lattice is obtained by replacing the original momenta k_i^2 $(i = x, y, z)$ in Eqs. (6) and (7) as

$$k_i^2 \to \frac{1}{a^2}(2 - 2\cos ak_i). \tag{16}$$

The first term $E_0^{\rm sum}$ in Eq. (13) is the zero-point energy in finite volume (and at finite chemical potential) with the momentum discretized by a finite thickness $L_z = aN_z$, where N_z is the number of lattice cells. The second term $E_0^{\rm int}$ in Eq. (13) is the zero-point energy with the continuous momentum in infinite volume. The Casimir energy $E_{\rm Cas}^{\rm Lat}$ is defined as the difference between these two energies. Also, as a result of lattice regularization, the momentum integral and the momentum sum are taken within the first Brillouin zone (BZ). In the case of the

[2] The case with boundary conditions for the x or y direction is straightforward, but these boundaries do not induce the WP-induced oscillating Casimir effect since now the DCDW is along the z axis.

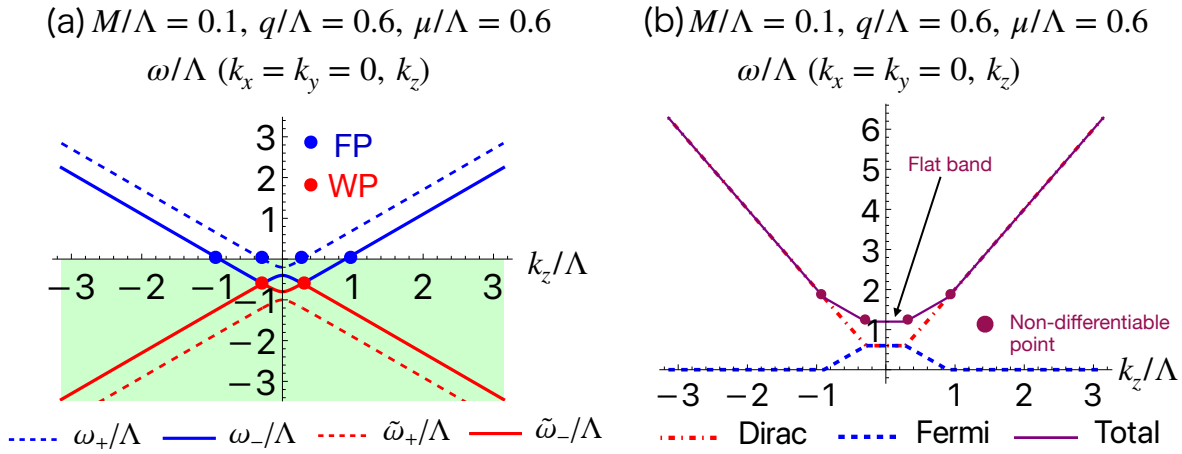


FIG. 2. (a) Dispersion relations of the four eigenmodes, Eqs. (6) and (7), at $(M/\Lambda, q/\Lambda, \mu/\Lambda) = (0.1, 0.6, 0, 6)$. (b) Each contribution from the Dirac sea (red line) or the Fermi sea (blue line) and their sum (purple line).

PBC, the sum is over $n = 0, 1, ..., N_z - 1$ (or equivalently $n = 0, 1, ..., N_z$).

Finally, we also define the dimensionless Casimir coefficient on the lattice,

$$C_{\rm Cas}^{[3]\rm Lat} \equiv L_z^3 E_{\rm Cas}^{\rm Lat} = a^3 N_z^3 E_{\rm Cas}^{\rm Lat}. \tag{17}$$

The quantities on the lattice, $E_{\rm Cas}^{\rm Lat}$ and $C_{\rm Cas}^{[3]\rm Lat}$, depend on the lattice spacing a, but its continuum limit ($a \to 0$) should coincide with Eqs. (11) and (12) in the continuum theory (if the lattice regularization is appropriate):

$$C_{\rm Cas}^{[3]} = \lim_{a\to 0} C_{\rm Cas}^{[3]\rm Lat}. \tag{18}$$

E. Dispersion relations

We remark on the dispersion relations (6) and (7) for quark fields in the DCDW phase. In Fig. 2(a), we show the four eigenmodes in the DCDW phase. Here, the parameters are fixed as $(M/\Lambda, q/\Lambda, \mu/\Lambda,) = (0.1, 0.6, 0.6)$, where each quantity is dimensionless by dividing by a dimensional parameter Λ. In this figure, the two touching points of ω_- and $\tilde{\omega}_-$ are regarded as Weyl points, and in addition, ω_- intersects the Fermi level at the Fermi points (FPs). Furthermore, ω_+ also intersects the Fermi level. Now, since we set $q = \mu$ for simplicity, the momenta of the Weyl points of ω_- and the intersection between the Fermi level and ω_+ coincide.

Figure 2(b) shows each contribution from the Dirac or Fermi sea and their sum. The Dirac-sea and Fermi-sea contributions correspond to the absolute value of the first and second terms of the integrand in Eq. (10), respectively. These dispersion relations become flat bands at low momentum and bend twice in the middle of the dispersion relation. Such non-differentiable points in the dispersion relation generally lead to an oscillating Casimir effect.

In the next section, we will show plots as shown in Fig. 2(b) for intuitively understanding the mechanism of the Casimir effect.

III. RESULTS

In this section, we show the results for the Casimir coefficients in the low-, intermediate-, and high-density regions, with the PBC (see Appendix A for the discussion with the MIT bag boundary condition). In the following, with $\Lambda = 860$ MeV, we fix the values of the order parameters, M and q, at three μ (low-, intermediate, and high-density regions) as those obtained in Ref. [46], where the authors applied the mean-field approach in the NJL model with the proper-time regularization.[3]

A. Oscillating Casimir effect (virtual parameters)

We will see in the following section that the oscillating Casimir effect is attributed to the flat band effect in the total dispersion relation caused by the presence of Weyl points. Therefore, in this subsection, we show how the flat band induces the oscillating Casimir effect by using a virtual parameter set that is not a solution to the gap equation of the NJL model but gives a typical flat band.

As typical parameters, we consider $(M/\Lambda, q/\Lambda) = (0.1, 0.6)$, as shown in Fig. 2. Then, the Casimir coefficient defined as Eq. (12) is shown in Fig. 3. The obtained results show the oscillating Casimir effect. The

[3] Note that the values of order parameters, M and q, in the NJL model depend on regularization schemes. The Casimir energy is a quantity independent of regularization schemes. Therefore, once M and q are obtained by a regularization, we can calculate the Casimir energy equivalently by any regularization (using the fixed M and q).

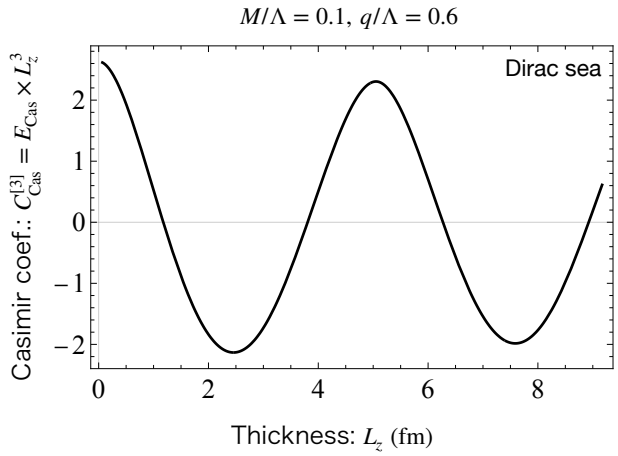


FIG. 3. An example of the Casimir coefficient for the oscillating Casimir effect from the Dirac-sea contribution, where we fix $(M/\Lambda, q/\Lambda) = (0.1, 0.6)$ as virtual parameters.

Casimir energy is obtained from the difference between the momentum integral of the dispersion relation and the infinite sum of the momentum discretized by the boundary conditions. The oscillation of Casimir energy arises from the matching of the Weyl points and the discrete levels.

In order to intuitively understand the mechanism of oscillations, we provide a graphical explanation. In the following explanation, as shown in Fig. 4, we represent $E_0^{\rm sum}$ by the sum of the blue rectangular areas $S_n^{\rm sum}$, while the corresponding $E_0^{\rm int}$ is the sum of the red areas $S_n^{\rm int}$. For example, in the low-momentum region of a massive particle, since the dispersion relation is a downward convex function, then $S_n^{\rm int} > S_n^{\rm sum}$ is satisfied, leading to a positive contribution to $N_z(S_n^{\rm sum} - S_n^{\rm int}) \propto E_{\rm Cas}^{\rm Lat}$.[4] On the other hand, in the high-momentum region, since the dispersion relation is an upward convex function due to the lattice regularization, $S_n^{\rm sum} > S_n^{\rm int}$, which leads to a negative contribution to $E_{\rm Cas}^{\rm Lat}$. In a region that can be regarded as a linear dispersion, we have $S_n^{\rm sum} = S_n^{\rm int}$.

As a crucial situation in this paper, in the flat-band region, it is also obvious that $S_n^{\rm sum} = S_n^{\rm int}$. The only exception is the case containing a Weyl point, where the corresponding areas are denoted as $S_{\rm W}^{\rm sum}$ and $S_{\rm W}^{\rm int}$. As the thickness N_z increases, the position of the Weyl point relative to a rectangle changes. The case where the endpoint of a rectangle coincides with the Weyl point is shown in Fig. 4(a). In this case, if the dispersion relation around the Weyl point is assumed to be linear, then $S_n^{\rm sum} = S_n^{\rm int}$ is exactly held. Precisely speaking, it deviates somewhat from the linear dispersion and yields minor contributions as discussed in the last paragraph. As

[4] This is a simplified situation. As a more complex case, when a dispersion relation is inverted by a Wely point or a Fermi point, the function in the high-momentum side can be an upward convex function.

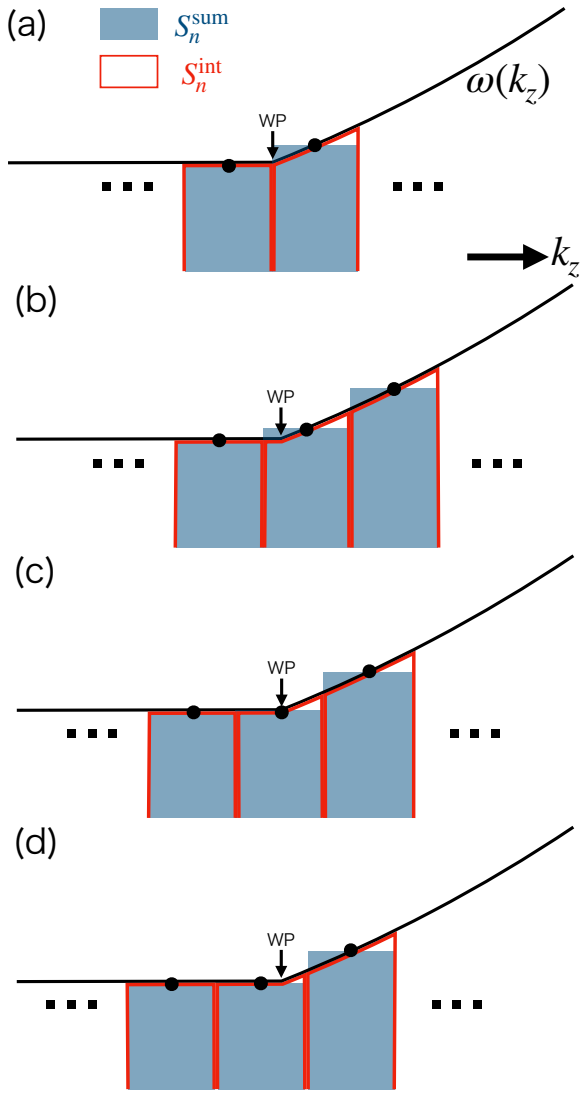


FIG. 4. Graphical explanation for the oscillating Casimir effect. The sum of the blue rectangles $S_n^{\rm sum}$ and the integral over the red areas $S_n^{\rm int}$ correspond to the zero-point energies, $E_0^{\rm sum}$ and $E_0^{\rm int}$ in finite and infinite volumes, respectively. The Casimir energy is defined as $E_0^{\rm sum} - E_0^{\rm int}$, which corresponds to $\sum_n S_n^{\rm sum} - \sum_n S_n^{\rm int}$.

N_z increases, the size and positions of rectangles change and reach the case shown in Fig. 4(b). In this case, we always have $S_n^{\rm int} > S_n^{\rm sum}$, and $E_{\rm Cas}^{\rm Lat}$ increases with N_z. As we continue to increase N_z, we eventually reach the case shown in Fig. 4(c), where the Weyl point coincides with the midpoint of the rectangle's edge. In this case, $S_{\rm W}^{\rm sum}$ goes under the flat band, which maximizes $E_{\rm Cas}^{\rm Lat}$. Furthermore, as N_z increases, the rectangle and the Weyl point reach the situation shown in Fig. 4(d). In this case, the excess of $S_{\rm W}^{\rm int}$ decreases and eventually reaches zero. Further increasing N_z, the position of the Weyl point for the rectangle returns to the case of Fig. 4(a).

Because the process from (a) to (d) occurs periodically, so that the value of $E_{\rm Cas}^{\rm Lat}$ oscillates with L_z. This is the

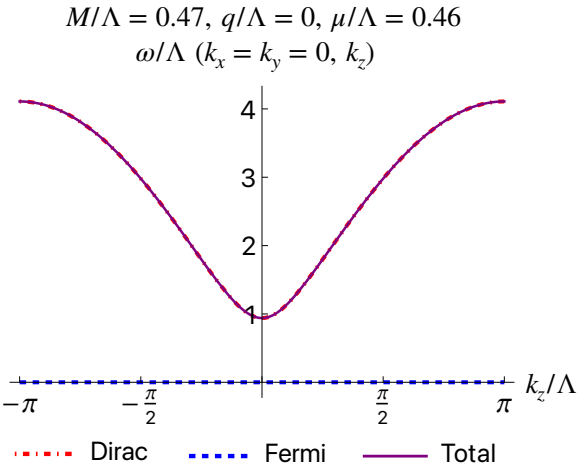


FIG. 5. Dispersion relations on the lattice with $a\Lambda = 1$ in the low-density region ($\mu/\Lambda = 0.46$).

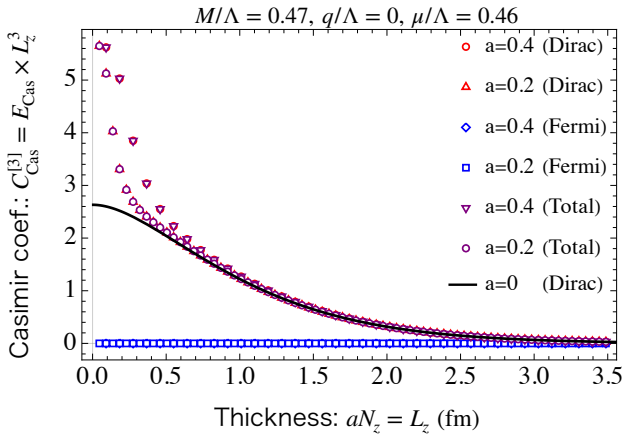


FIG. 6. Casimir coefficients in the low-density region. We denote $a\Lambda$ as a in the legends.

origin of the L_z-dependent oscillation of Casimir energy. Then, the oscillation period (for the PBC) with respect to L_z is written as [31]

$$L_z^{\rm osc} = \frac{2\pi}{k_{\rm WP}}, \tag{19}$$

where $k_{\rm WP}$ is the momentum separation between the Weyl point and the origin. For the current parameters, the period is estimated to be $L_z^{\rm osc} \sim 5.1$ fm with $k_{\rm WP}/\Lambda \sim 0.283$, which is consistent with the plot in Fig. 3. In addition, in the $3+1$ dimensions, the minor deviation from this naive estimate is caused by the integral over the $k_x \neq 0$ and $k_y \neq 0$.

B. Low-density region

First, we consider the Casimir effect in the low-chemical potential region, where we fix $\mu/\Lambda = 0.46$. In this region, we find a chiral condensed phase with $M/\Lambda = 0.47$, but the DCDW phase has not yet developed ($q = 0$).

Figure 5 shows the dispersion relation of the quark field along the momentum k_z on the lattice with $a\Lambda = 1$, where the blue line represents the contribution from the Dirac sea, the red line from the Fermi sea, and the purple line their sum. In this case, there is no mode crossing the Fermi level, which means that the total dispersion relation has no non-differentiable point. Therefore, the four dispersion relations are nothing but that of a massive Dirac fermion.

Figure 6 shows the results for the Casimir coefficient at $(M/\Lambda, q/\Lambda, \mu/\Lambda) = (0.47, 0, 0.46)$, obtained from the two approaches. The resultant Casimir effect occurs only from the contribution of the Dirac sea. On the other hand, there is no contribution to the Casimir effect from the Fermi sea, as expected. In this figure, we find that $C^{[3]}_{\rm Cas}$ approaches zero in the long-thickness region. This behavior means that the damping of the Casimir energy $E_{\rm Cas}$ is faster than $1/L_z^3$ ($1/L_z^3$ is expected for the massless fields in the three-dimensional space). Such a faster damping is well known for free massive fields [97–99].

We present results for the two cases of lattice spacing $a\Lambda = 0.2, 0.4$ in Fig. 6. We show analytical solutions for only the Dirac-sea part of the Casimir coefficient, obtained from the Lifshitz formula. We observe that in the long-thickness region of about $aN_z = L_z > 1.0$ fm, the lattice regularization works well even for $a\Lambda = 0.4$. On the other hand, in the short-thickness region, the results for $a\Lambda = 0.4$ deviate from the analytical solution due to the enhanced ultraviolet lattice cutoff effect. On the other hand, the results for $a\Lambda = 0.2$ reproduce the analytical solution well up to about $aN_z = L_z = 0.5$ fm. This result guarantees that the results from our lattice regularization are accurate up to a quite short-thickness region.

C. Intermediate-density region

As the chemical potential increases, the contribution from the quark Fermi sea causes a significant change in the behavior of the Casimir effect. Figure 7 shows the dispersion relation of the quark field along the momentum k_z for fixed parameters $(M/\Lambda, q/\Lambda, \mu/\Lambda) = (0.46, 0, 0.485)$ and $a\Lambda = 1$. In this parameter set, the DCDW phase has not yet occurred, but the $\omega_\pm$ with low momentum is below the Fermi sea. Then, there is an additional contribution to the Casimir effect from the Fermi sea as well as from the Dirac sea. Furthermore, the crossing points (Fermi points) of $\omega_\pm$ and the Fermi level lead to non-differentiable points in the total dispersion relation. Due to this, the oscillating Casimir effect occurs, by a mechanism similar to that discussed in the subsection II E.

Figure 8 shows the results for the Casimir coefficient. In the short thickness of Fig. 8, we find that the Casimir

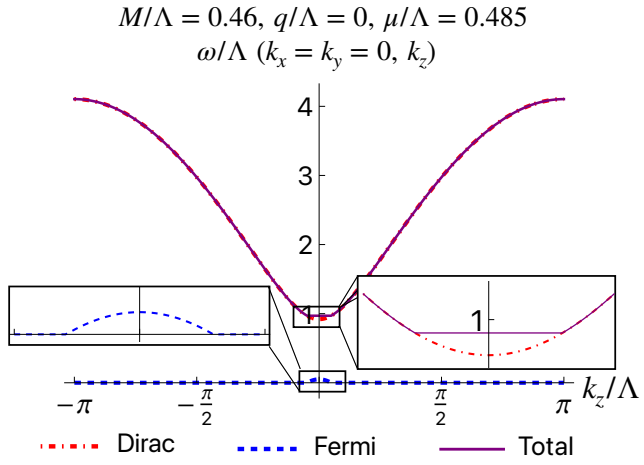


FIG. 7. Dispersion relations on the lattice with $a\Lambda = 1$ in the intermediate-density region ($\mu/\Lambda = 0.485$).

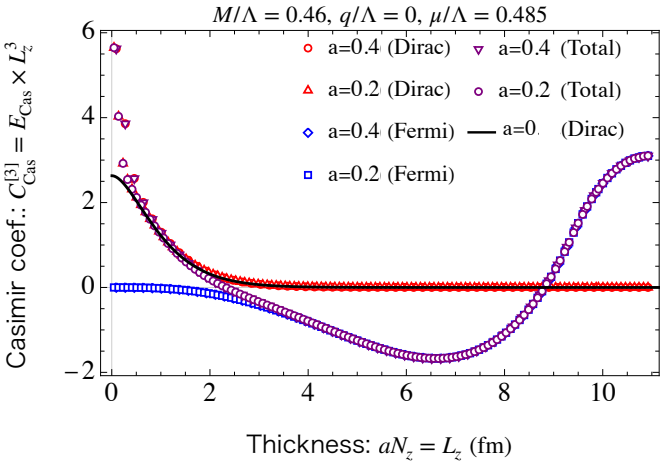


FIG. 8. Casimir coefficients in the intermediate-density region. We denote $a\Lambda$ as a in the legends.

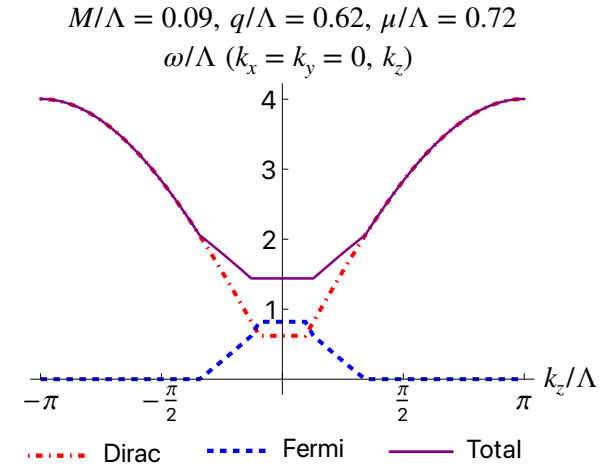


FIG. 9. Dispersion relations on the lattice with $a\Lambda = 1$ in the high-density region ($\mu/\Lambda = 0.72$).

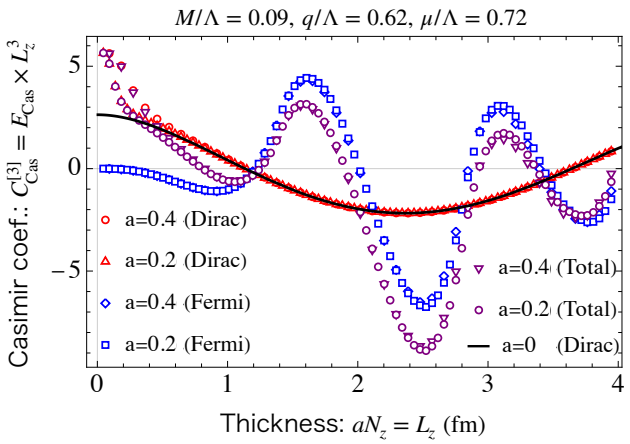


FIG. 10. Casimir coefficients in the high-density region. We denote $a\Lambda$ as a in the legends.

coefficient is dominated by the contribution from the Dirac sea, which is qualitatively the same as the result shown in the previous subsection III B. On the other hand, in the long thickness of Fig. 8, the Casimir coefficient is dominated by the contribution from the Fermi sea. As a result, the sign of the total Casimir coefficient flips around $aN_z = L_z \sim 2.3$ fm and 8.8 fm. At further long-thickness, we expect an oscillation of the Casimir coefficient. However, now we do not show the numerical results due to the limitations of calculations with sufficient accuracy. By using Eq. (19), the period of oscillation is estimated to be $L_z^{\rm osc} = 2\pi/k_{\rm FP} \sim 9.4$ fm, where $k_{\rm FP}/\Lambda \sim 0.154$ is the momentum of the Fermi point.

D. high-density region

When the chemical potential is large enough, the DCDW phase ($q \neq 0$) is realized. Figure 9 shows the dispersion relation of the quark fields along the momentum k_z for fixed parameters $(M/\Lambda, q/\Lambda, \mu/\Lambda) = (0.09, 0.62, 0.72)$ with $a\Lambda = 1$. In this case, the two touching points of ω_- and $\tilde{\omega}_-$ are regarded as the Weyl points. Furthermore, the ω_+ and ω_- cross the Fermi level. Thus, there are six non-differentiable points. In the low-momentum region, the total dispersion relation becomes a flat band due to the cancellations between ω_- and $\tilde{\omega}_-$ and between ω_+ and $\tilde{\omega}_+$, and the edges of the flat band of the total dispersion relation are non-differentiable. Furthermore, due to the four Fermi points, the total dispersion relation is non-differentiable at the four momenta.

Figure 10 shows the results for the Casimir coefficient. For the Dirac-sea effect, the Casimir effect shows oscillating behavior, as expected from the Weyl points. On the other hand, for the Fermi-sea effect as depicted by the red line, we see that the Fermi-sea part of Casimir coefficient oscillates with higher frequency than that from the Dirac sea. Its frequency is determined by the positions of

the non-differentiable points on the dispersion relation. By using Eq. (19), the oscillation periods are estimated to be $L_z^{\rm osc} = 2\pi/k_{\rm WP} \sim 4.9$ fm with $k_{\rm WP}/\Lambda \sim 0.297$ (for the Dirac sea part) and $L_z^{\rm osc} = 2\pi/k_{\rm FP} \sim 1.4$ fm with $k_{\rm FP}/\Lambda \sim 1.03$ (for the Fermi sea part of ω_-), and $L_z^{\rm osc} = 2\pi/k_{\rm FP} \sim 3.6$ fm with $k_{\rm FP}/\Lambda \sim 0.40$ (for the Fermi sea part of ω_+). The Dirac-sea and Fermi-sea contributions shown in Fig. 10 are dominated by the periods of $L_z^{\rm osc} \sim 4.9$ fm and $L_z^{\rm osc} \sim 1.4$ fm, respectively, and the total result is their superposition.

IV. CONCLUSION AND OUTLOOK

In this paper, we have discussed the Casimir effect from quark fields in some phases of quark matter. In particular, we consider the QCD counterpart of a new type of Casimir effect i.e., the oscillating Casimir effect. We have classified the typical behaviors of the Casimir effect by focusing on the three types of density regions:

1. When the quark chemical potential is small enough, the quark-number density is zero. Then, the Casimir effect for massive quark fields occurs, where the damping of the Casimir energy with increasing thickness is characterized by the dynamical masses of quarks.

2. As the quark chemical potential increases, the quark number density becomes nonzero. In this region, the contribution from the quark Fermi sea effect leads to the oscillating Casimir effect.

3. Further increasing the chemical potential eventually leads to a phase transition into the DCDW phase. In this region, in addition to the Fermi sea effect, the Weyl-point structure of quark fields in the DCDW phase results in the oscillating Casimir effect. The total Casimir energy behaves as the superposition of two types of oscillations.

To clarify terminology, one of our main findings may be called the *DCDW-induced oscillating Casimir effect*, which is a kind of the *WP-induced oscillating Casimir effect* predicted in Ref. [31]. On the other hand, the *FP-induced oscillating Casimir effect* would be a common term for general fermionic systems at finite density in finite volume.

In the following, we list some potential future studies.

- *Solving gap equation in finite volume*—In this work, we input the values of order parameters obtained from the gap equation in infinite volume, in order to understand the typical behavior of the Casimir effect for typical dispersion relations. As a more precise analysis, one can calculate the order parameters by minimizing the thermodynamic or effective potential in finite volume, and then one can understand the self-consistent relationship between the volume-dependent order parameters and the Casimir effect. This type of study would help elucidate the phase diagram of interacting fermions in finite volume (e.g., see Refs. [100–107]) from the viewpoint of the Casimir effect.

- *Lattice simulations*—Our predictions will be examined by future lattice NJL (and also QCD) simulations. Since Monte Carlo simulations at large quark chemical potential suffer from the sign-problem, sign-problem-free techniques should be applied. In particular, numerical simulations of Casimir effects for lattice gauge fields have been vigorously studied, such as the U(1) gauge [108–110], the compact U(1) gauge [111–115], the SU(2) gauge [116, 117], and the SU(3) gauge fields [118, 119], so that the interplay between quark and gluon dynamics would be interesting.

- *Finite temperature*—At finite temperature, thermal fluctuations (as well as quantum fluctuations) also induce the Casimir effect, which is the so-called thermal Casimir effect well established theoretically [88] and experimentally [120]. The Casimir energy at finite temperature can be calculated by using a straightforward modification of Eq. (11) or Eq. (13). Although the DCDW phase disappears at sufficiently high temperature, in the low-temperature region the oscillating Casimir effect can be realized.

- *Axion electrodynamics*—In this work, we consider only the Casimir effect for quark fields, which is induced by boundary conditions on quark fields. The dynamics of photons in the DCDW phase may be described by modified Maxell equations [63, 76, 121] which is the so-called axion electrodynamics [122, 123]. The Casimir energy for such modified photon fields can exhibit a sign-flipping behavior [19, 95, 124–127] when boundary conditions on photon fields are imposed. It will be interesting whether or not such a photonic Casimir effect occurs also in the DCDW phase.

- *Models with nucleons*—In this work, we consider the Casimir effect for quark degrees of freedom based on the NJL model. Apart from the NJL model, in the low-density phase of QCD, relevant degrees of freedom are hadrons, specifically nucleons. Therefore, it will be important to examine the consistency between the quark Casimir effect in the NJL model and the nucleon Casimir effect in an effective model such as the nucleon linear sigma model which can also investigate the DCDW phase.

- *Other inhomogeneous chiral phases*—We have shown the WP-induced oscillating Casimir effect in the DCDW phase, which is particular to systems with momentum-dependent Weyl points. It would be interesting what types of Casimir effects can occur in other inhomogeneous chiral phases.

For example, the real-kink crystal phase is known as a possible ground state in the 3+1 dimension NJL model [50]. In this phase, because k_z is not a conserved quantity, our approach formulated in spatial-momentum space cannot be applied, but the Casimir energy should be calculated by approaches formulated with discrete eigenenergies. Furthermore, apart from the dense quark matter, a magnetic field can also induce the DCDW phase [53], which is the so-called magnetic dual chiral density wave (MDCDW) (see Ref. [128] for a review). The Casimir effect in such a phase is also interesting.

- *Lower-dimensional models of QCD*—The WP-induced oscillating Casimir effect can occur also in lower spatial dimensions (i.e., $1+1$ or $2+1$ dimensions) [31]. Therefore, investigation of the Casimir effect in possible inhomogeneous phases of low-dimensional QCD-like models, such as the Gross-Neveu model [129–131], the chiral Gross-Neveu (NJL_2) model [132], and the 't Hooft model (QCD_2) [132–134], will be also important.

ACKNOWLEDGMENTS

The authors thank Tsutomu Ishikawa for fruitful discussions. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant No. JP20K14476).

Appendix A: Results with MIT bag boundary

In the main text, we applied the PBC to consider the Casimir effect. In this Appendix, we consider the case with the MIT bag boundary condition, which is more realistic when we consider quark matter in nature or experiments.

The MIT bag boundary condition [135] requires that the quark-current flux toward normal to the boundary is zero. For the massless Dirac field on the MIT bag boundaries, one can obtain the discretized momentum $k_z \to (n+1/2)\pi/L_z$ as an analytic solution satisfying the Dirac equation. On the other hand, for the massive case, one cannot obtain an analytic form of momentum, so that one should apply an alternative approach with no analytic form of k_z (e.g., Refs. [136–138]). In this Appendix, as an approximate estimate, we show the results using the massless solution $k_z \to (n+1/2)\pi/L_z$, which may be called "MIT bag" boundary conditions.

The Lifshitz formula for the Casimir energy from the Dirac sea with the "MIT bag" boundaries is given by

$$E_{\mathrm{Cas}}^{\text{"MIT"}} = -2N_f N_c \sum_{s=\pm} \int_0^\infty \frac{d\xi}{2\pi} \int \frac{d^2k_\perp}{(2\pi)^2} \ln\left[1+e^{-2L_z k_z^{[s]}}\right],$$
$$k_z^{[\pm]} = \sqrt{k_\perp^2 + M^2 + \xi^2 - \frac{q^2}{4} \mp iq\sqrt{k_\perp^2+\xi^2}}, \quad \text{(A1)}$$

where the first integral variable is the imaginary part ξ of imaginary frequency $\omega \equiv i\xi$, $k_\perp^2 \equiv k_x^2 + k_y^2$, and $d^2k_\perp \equiv dk_x dk_y$. The overall factor of $2N_f N_c$ means the degrees of freedom for particle-antiparticle, flavors, and colors, and the spin degrees of freedom are labeled as $s = \pm$. The overall minus sign and the factor of $+e^{-2L_z k_z^{[s]}}$ is a property of the "MIT bag" boundary on fermion fields. In the lattice rugularization, we have to replace as $\sum_n^{\mathrm{BZ}} \to \frac{1}{2}\sum_n^{\mathrm{BZ}}$ in Eq. (14), and the sum is over $n = 0, 1, ..., 2N_z - 1$.

Figure 11 shows the results of the Casimir coefficients at low-, intermediate-, and high-density regions. In the case of the "MIT bag" boundary, the number of discretized levels is twice larger than that with the PBC. As a result, the period of oscillation is half of that with the PBC: analogous to Eq. (19),

$$L_z^{\mathrm{osc}\,\text{"MIT"}} = \frac{\pi}{k_{\mathrm{WP}}}. \quad \text{(A2)}$$

Furthermore, from these results, we can also discuss the case with the antiperiodic boundary conditions (APBCs). The discrete momentum, $k_z \to (2n+1)\pi/L_z$, for the APBCs is shifted by a half period from $k_z \to 2n\pi/L_z$ for the PBCs, and hence the sign of the Casimir energy is flipped and also its absolute value changes slightly. The corresponding Lifshitz formula is written as

$$E_{\mathrm{Cas}}^{\mathrm{APBC}} = -4N_f N_c \sum_{s=\pm} \int_0^\infty \frac{d\xi}{2\pi} \int \frac{d^2k_\perp}{(2\pi)^2} \ln\left[1+e^{-L_z k_z^{[s]}}\right]. \quad \text{(A3)}$$

In the $3+1$ dimensions, the Casimir energy with $k_z \to (2n+1)\pi/L_z$ is 16 times larger than that with $k_z \to (n+1/2)\pi/L_z$. As a result, by multiplying the vertical axis obtained in Fig. 11 by 16 and the horizontal axis by 2, we can obtain the exact results for the APBCs.

[1] H. B. G. Casimir, On the Attraction Between Two Perfectly Conducting Plates, Proc. Kon. Ned. Akad. Wetensch. **51**, 793 (1948).
[2] G. Plunien, B. Müller, and W. Greiner, The Casimir Effect, Phys. Rep. **134**, 87 (1986).
[3] V. M. Mostepanenko and N. N. Trunov, The Casimir Effect and Its Applications, Sov. Phys. Usp. **31**, 965 (1988).
[4] M. Bordag, U. Mohideen, and V. M. Mostepanenko, New developments in the Casimir effect, Phys. Rep.

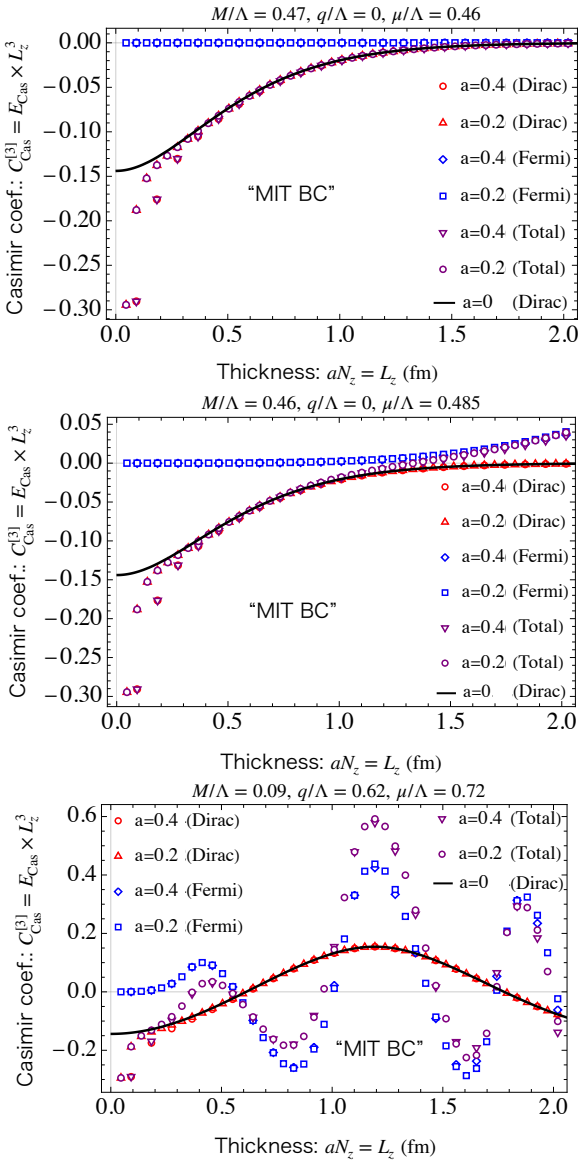


FIG. 11. Casimir coefficients in the low-density region with the "MIT bag" boundary conditions. We denote $a\Lambda$ as a in the legends.

353, 1 (2001), arXiv:quant-ph/0106045 [quant-ph].

[5] K. A. Milton, *The Casimir Effect: Physical Manifestations of Zero-Point Energy* (World Scientific, Singapore, 2001).

[6] G. L. Klimchitskaya, U. Mohideen, and V. M. Mostepanenko, The Casimir force between real materials: Experiment and theory, Rev. Mod. Phys. **81**, 1827 (2009), arXiv:0902.4022 [cond-mat.other].

[7] S. K. Lamoreaux, Demonstration of the Casimir Force in the 0.6 to 6μm Range, Phys. Rev. Lett. **78**, 5 (1997), [Erratum: Phys. Rev. Lett. **81**, 5475 (1998)].

[8] G. Bressi, G. Carugno, R. Onofrio, and G. Ruoso, Measurement of the Casimir force between parallel metallic surfaces, Phys. Rev. Lett. **88**, 041804 (2002), arXiv:quant-ph/0203002.

[9] T. Gong, M. R. Corrado, A. R. Mahbub, C. Shelden, and J. N. Munday, Recent progress in engineering the Casimir effect – applications to nanophotonics, nanomechanics, and chemistry, Nanophotonics **10**, 523 (2021).

[10] I. E. Dzyaloshinskii, E. M. Lifshitz, and L. P. Pitaevskii, The general theory of van der Waals forces, Adv. Phys. **10**, 165 (1961).

[11] T. H. Boyer, Van der Waals forces and zero-point energy for dielectric and permeable materials, Phys. Rev. A **9**, 2078 (1974).

[12] O. Kenneth, I. Klich, A. Mann, and M. Revzen, Repulsive Casimir forces, Phys. Rev. Lett. **89**, 033001 (2002), arXiv:quant-ph/0202114.

[13] J. N. Munday, F. Capasso, and V. A. Parsegian, Measured long-range repulsive Casimir–Lifshitz forces, Nature **457**, 170 (2009).

[14] C. Henkel and K. Joulain, Casimir force between designed materials: what is possible and what not, EPL **72**, 929 (2005), arXiv:quant-ph/0407153.

[15] F. S. S. Rosa, D. A. R. Dalvit, and P. W. Milonni, Casimir-Lifshitz Theory and Metamaterials, Phys. Rev. Lett. **100**, 183602 (2008), arXiv:0803.2908 [quant-ph].

[16] A. G. Grushin and A. Cortijo, Tunable Casimir Repulsion with Three Dimensional Topological Insulators, Phys. Rev. Lett. **106**, 020403 (2011), arXiv:1002.3481 [cond-mat.mtrl-sci].

[17] W.-K. Tse and A. H. MacDonald, Quantized Casimir Force, Phys. Rev. Lett. **109**, 236806 (2012), arXiv:1208.3786 [cond-mat.mes-hall].

[18] J. H. Wilson, A. A. Allocca, and V. Galitski, Repulsive casimir force between weyl semimetals, Phys. Rev. B **91**, 235115 (2015), arXiv:1501.07659 [cond-mat.mes-hall].

[19] K. Fukushima, S. Imaki, and Z. Qiu, Anomalous Casimir effect in axion electrodynamics, Phys. Rev. D **100**, 045013 (2019), arXiv:1906.08975 [hep-th].

[20] J. N. Fuchs, A. Recati, and W. Zwerger, Oscillating Casimir force between impurities in one-dimensional Fermi liquids, Phys. Rev. A **75**, 043615 (2007), arXiv:cond-mat/0610659 [cond-mat.mes-hall].

[21] P. Wächter, V. Meden, and K. Schönhammer, Indirect forces between impurities in one-dimensional quantum liquids, Phys. Rev. B **76**, 045123 (2007), arXiv:cond-mat/0703128 [cond-mat.str-el].

[22] E. B. Kolomeisky, J. P. Straley, and M. Timmins, Casimir effect in a one-dimensional gas of free fermions, Phys. Rev. A **78**, 022104 (2008), arXiv:0706.2887 [cond-mat.mes-hall].

[23] D. Zhabinskaya and E. J. Mele, Casimir interactions between scatterers in metallic carbon nanotubes, Phys. Rev. B **80**, 155405 (2009), arXiv:0909.1731 [cond-mat.mes-hall].

[24] L.-W. Chen, G.-Z. Su, J.-C. Chen, and A. Bjarne, Oscillating Casimir force between two slabs in a Fermi sea, Chin. Phys. B **21**, 010501 (2012).

[25] S. G. C. J. Chen, L., Oscillating Casimir effect of a trapped Fermi gas, J. Stat. Phys. **143**, 523 (2011).

[26] A. Han, Seyit Deniz and E. Aydiner, Oscillating Casimir force of trapped Bose gas in electromagnetic field, arXiv:1508.01040 [quant-ph].

[27] Q.-D. Jiang and F. Wilczek, Chiral Casimir forces: Repulsive, enhanced, tunable, Phys. Rev. B **99**, 125403 (2019), arXiv:1805.07994 [cond-mat.mes-hall].

[28] T. Ishikawa, K. Nakayama, and K. Suzuki, Casimir effect for lattice fermions, Phys. Lett. B **809**, 135713 (2020), arXiv:2005.10758 [hep-lat].

[29] T. Ishikawa, K. Nakayama, and K. Suzuki, Lattice-fermionic Casimir effect and topological insulators, Phys. Rev. Res. **3**, 023201 (2021), arXiv:2012.11398 [hep-lat].

[30] Y. V. Mandlecha and R. V. Gavai, Lattice fermionic Casimir effect in a slab bag and universality, Phys. Lett. B **835**, 137558 (2022), arXiv:2207.00889 [hep-lat].

[31] K. Nakayama and K. Suzuki, Dirac/Weyl-node-induced oscillating Casimir effect, Phys. Lett. B **843**, 138017 (2023), arXiv:2207.14078 [cond-mat.mes-hall].

[32] K. Nakata and K. Suzuki, Non-Hermitian Casimir Effect of Magnons, arXiv:2305.09231 [quant-ph].

[33] F. Dautry and E. M. Nyman, Pion condensation and the σ-model in liquid neutron matter, Nucl. Phys. A **319**, 323 (1979).

[34] W. Broniowski and M. Kutschera, One fermion loop contribution to the energy of the neutral pion condensate in the σ Model, Phys. Lett. B **234**, 449 (1990).

[35] M. Kutschera, W. Broniowski, and A. Kotlorz, Quark matter with neutral pion condensate, Phys. Lett. B **237**, 159 (1990).

[36] W. Broniowski and M. Kutschera, One-loop effective action in the σ model for a periodic-chiral-field ansatz, Phys. Rev. D **41**, 3800 (1990).

[37] W. Broniowski and M. Kutschera, Ambiguities in effective chiral models with cutoff, Phys. Lett. B **242**, 133 (1990).

[38] M. Kutschera, W. Broniowski, and A. Kotlorz, Quark matter with pion condensate in an effective chiral model, Nucl. Phys. A **516**, 566 (1990).

[39] M. Kato, W. Bentz, and K. Tanaka, Neutral pion condensation in quark matter including vacuum fluctuation effects, Phys. Rev. C **45**, 2445 (1992).

[40] A. Kotlorz and M. Kutschera, Quark matter inside neutron stars in an effective chiral model, Acta Phys. Polon. B **25**, 859 (1994).

[41] M. Sadzikowski and W. Broniowski, Nonuniform chiral phase in effective chiral quark models, Phys. Lett. B **488**, 63 (2000), arXiv:hep-ph/0003282.

[42] K. Takahashi and T. Tatsumi, Solution of the Dirac equation with an axial vector current in the chiral $SU(2) \otimes SU(2)$ model, Prog. Theor. Phys. **105**, 437 (2001).

[43] K. Takahashi and T. Tatsumi, $\sigma\pi^0$ condensation at finite density in the linear σ model, Phys. Rev. C **63**, 015205 (2001).

[44] M. Sadzikowski, Coexistence of pion condensation and color superconductivity in two flavor quark matter, Phys. Lett. B **553**, 45 (2003), arXiv:hep-ph/0210065.

[45] K. Takahashi, Neutral pion condensation in the chiral SU(3) $\bigotimes$ SU(3) model, Phys. Rev. C **66**, 025202 (2002).

[46] T. Tatsumi and E. Nakano, Dual chiral density wave in quark matter, arXiv:hep-ph/0408294.

[47] E. Nakano and T. Tatsumi, Chiral symmetry and density wave in quark matter, Phys. Rev. D **71**, 114006 (2005), arXiv:hep-ph/0411350.

[48] M. Sadzikowski, Comparison of the non-uniform chiral and 2SC phases at finite temperatures and densities, Phys. Lett. B **642**, 238 (2006), arXiv:hep-ph/0609186.

[49] B. Bringoltz, Chiral crystals in strong-coupling lattice QCD at nonzero chemical potential, JHEP **03**, 016, arXiv:hep-lat/0612010.

[50] D. Nickel, Inhomogeneous phases in the Nambu–Jona-Lasino and quark-meson model, Phys. Rev. D **80**, 074025 (2009), arXiv:0906.5295 [hep-ph].

[51] S. Maedan, Influence of current mass on the spatially inhomogeneous chiral condensate, Prog. Theor. Phys. **123**, 285 (2010), arXiv:0908.0594 [hep-ph].

[52] S. Carignano, D. Nickel, and M. Buballa, Influence of vector interaction and Polyakov loop dynamics on inhomogeneous chiral symmetry breaking phases, Phys. Rev. D **82**, 054009 (2010), arXiv:1007.1397 [hep-ph].

[53] I. E. Frolov, V. C. Zhukovsky, and K. G. Klimenko, Chiral density waves in quark matter within the Nambu–Jona-Lasinio model in an external magnetic field, Phys. Rev. D **82**, 076002 (2010), arXiv:1007.2984 [hep-ph].

[54] T. L. Partyka and M. Sadzikowski, Chiral density waves in quarkyonic matter, Acta Phys. Polon. B **42**, 1305 (2011), arXiv:1011.0921 [hep-ph].

[55] H. Abuki, D. Ishibashi, and K. Suzuki, Crystalline chiral condensates off the tricritical point in a generalized Ginzburg-Landau approach, Phys. Rev. D **85**, 074002 (2012), arXiv:1109.1615 [hep-ph].

[56] S. Carignano and M. Buballa, Two-dimensional chiral crystals in the Nambu-Jona-Lasinio model, Phys. Rev. D **86**, 074018 (2012), arXiv:1203.5343 [hep-ph].

[57] S. Karasawa and T. Tatsumi, Variational approach to the inhomogeneous chiral phase in quark matter, Phys. Rev. D **92**, 116004 (2015), arXiv:1307.6448 [hep-ph].

[58] D. Müller, M. Buballa, and J. Wambach, Dyson-Schwinger study of chiral density waves in QCD, Phys. Lett. B **727**, 240 (2013), arXiv:1308.4303 [hep-ph].

[59] A. Heinz, F. Giacosa, and D. H. Rischke, Chiral density wave in nuclear matter, Nucl. Phys. A **933**, 34 (2015), arXiv:1312.3244 [nucl-th].

[60] J. Moreira, B. Hiller, W. Broniowski, A. A. Osipov, and A. H. Blin, Nonuniform phases in a three-flavor Nambu-Jona-Lasinio model, Phys. Rev. D **89**, 036009 (2014), arXiv:1312.4942 [hep-ph].

[61] T. Tatsumi and T. Muto, Quark beta decay in the inhomogeneous chiral phase and cooling of compact stars, Phys. Rev. D **89**, 103005 (2014), arXiv:1403.1927 [nucl-th].

[62] S. Carignano, M. Buballa, and B.-J. Schaefer, Inhomogeneous phases in the quark-meson model with vacuum fluctuations, Phys. Rev. D **90**, 014033 (2014), arXiv:1404.0057 [hep-ph].

[63] T. Tatsumi, K. Nishiyama, and S. Karasawa, Novel Lifshitz point for chiral transition in the magnetic field, Phys. Lett. B **743**, 66 (2015), arXiv:1405.2155 [hep-ph].

[64] T. Hayata and A. Yamamoto, Inhomogeneous Polyakov loop induced by inhomogeneous chiral condensates, Phys. Lett. B **744**, 401 (2015), arXiv:1408.1905 [hep-ph].

[65] M. Harada, H. K. Lee, Y.-L. Ma, and M. Rho, Inhomogeneous quark condensate in compressed Skyrmion matter, Phys. Rev. D **91**, 096011 (2015), arXiv:1502.02508 [hep-ph].

[66] T.-G. Lee, E. Nakano, Y. Tsue, T. Tatsumi, and B. Friman, Landau-Peierls instability in a Fulde-Ferrell type inhomogeneous chiral condensed phase, Phys. Rev. D **92**, 034024 (2015), arXiv:1504.03185 [hep-ph].

[67] K. Nishiyama, S. Karasawa, and T. Tatsumi, Hybrid chiral condensate in the external magnetic field, Phys. Rev. D **92**, 036008 (2015), arXiv:1505.01928 [nucl-th].

[68] S. Carignano, E. J. Ferrer, V. de la Incera, and L. Paulucci, Crystalline chiral condensates as a component of compact stars, Phys. Rev. D **92**, 105018 (2015),

arXiv:1505.05094 [nucl-th].
[69] J. P. Carlomagno, D. Gómez Dumm, and N. N. Scoccola, Inhomogeneous phases in nonlocal chiral quark models, Phys. Rev. D **92**, 056007 (2015), arXiv:1507.01560 [hep-ph].
[70] R. Yoshiike, K. Nishiyama, and T. Tatsumi, Spontaneous magnetization of quark matter in the inhomogeneous chiral phase, Phys. Lett. B **751**, 123 (2015), arXiv:1507.02110 [hep-ph].
[71] A. Heinz, F. Giacosa, M. Wagner, and D. H. Rischke, Inhomogeneous condensation in effective models for QCD using the finite-mode approach, Phys. Rev. D **93**, 014007 (2016), arXiv:1508.06057 [hep-ph].
[72] D. Suenaga and M. Harada, Appearance of novel modes of $\overline{D}$ mesons with negative velocity in the dual chiral density wave, Phys. Rev. D **93**, 076005 (2016), arXiv:1509.08578 [hep-ph].
[73] S. Karasawa, T.-G. Lee, and T. Tatsumi, Brazovskii–Dyugaev effect on the inhomogeneous chiral transition in quark matter, PTEP **2016**, 043D02 (2016).
[74] P. Adhikari, J. O. Andersen, and P. Kneschke, Inhomogeneous chiral condensate in the quark-meson model, Phys. Rev. D **96**, 016013 (2017), [Erratum: Phys. Rev. D **98**, 099902 (2018)], arXiv:1702.01324 [hep-ph].
[75] R. Yoshiike, T.-G. Lee, and T. Tatsumi, Chiral pair fluctuations for the inhomogeneous chiral transition, Phys. Rev. D **95**, 074010 (2017), arXiv:1702.01511 [hep-ph].
[76] T. Tatsumi, R. Yoshiike, and K. Kashiwa, Anomalous Hall effect in dense QCD matter, Phys. Lett. B **785**, 46 (2018), arXiv:1803.10514 [hep-ph].
[77] S. Carignano and M. Buballa, Inhomogeneous chiral condensates in three-flavor quark matter, Phys. Rev. D **101**, 014026 (2020), arXiv:1910.03604 [hep-ph].
[78] P. Lakaschus, M. Buballa, and D. H. Rischke, Competition of inhomogeneous chiral phases and two-flavor color superconductivity in the NJL model, Phys. Rev. D **103**, 034030 (2021), arXiv:2012.07520 [hep-ph].
[79] S. M. A. Tabatabaee Mehr, Chiral symmetry breaking and phase diagram of dual chiral density wave in a rotating quark matter, Phys. Rev. D **108**, 094042 (2023), arXiv:2306.11753 [nucl-th].
[80] S. Pitsinigkos and A. Schmitt, Chiral crossover versus chiral density wave in dense nuclear matter, Phys. Rev. D **109**, 014024 (2024), arXiv:2309.01603 [nucl-th].
[81] M. Buballa and S. Carignano, Inhomogeneous chiral condensates, Prog. Part. Nucl. Phys. **81**, 39 (2015), arXiv:1406.1367 [hep-ph].
[82] Y. Nambu and G. Jona-Lasinio, Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I, Phys. Rev. **122**, 345 (1961).
[83] Y. Nambu and G. Jona-Lasinio, Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II, Phys. Rev. **124**, 246 (1961).
[84] U. Vogl and W. Weise, The Nambu and Jona Lasinio model: Its implications for hadrons and nuclei, Prog. Part. Nucl. Phys. **27**, 195 (1991).
[85] S. P. Klevansky, The Nambu–Jona-Lasinio model of quantum chromodynamics, Rev. Mod. Phys. **64**, 649 (1992).
[86] T. Hatsuda and T. Kunihiro, QCD phenomenology based on a chiral effective Lagrangian, Phys. Rept. **247**, 221 (1994), arXiv:hep-ph/9401310.
[87] M. Buballa, NJL model analysis of dense quark matter, Phys. Rept. **407**, 205 (2005), arXiv:hep-ph/0402234.
[88] E. M. Lifshitz, The theory of molecular attractive forces between solids, Sov. Phys. JETP **2**, 73 (1956).
[89] A. Actor, I. Bender, and J. Reingruber, Casimir effect on a finite lattice, Fortsch. Phys. **48**, 303 (2000), arXiv:quant-ph/9908058 [quant-ph].
[90] M. Pawellek, Finite-sites corrections to the Casimir energy on a periodic lattice, arXiv:1303.4708 [hep-th].
[91] K. Nakayama and K. Suzuki, Remnants of the nonrelativistic Casimir effect on the lattice, Phys. Rev. Res. **5**, L022054 (2023), arXiv:2204.12032 [quant-ph].
[92] K. Nakata and K. Suzuki, Magnonic Casimir Effect in Ferrimagnets, Phys. Rev. Lett. **130**, 096702 (2023), arXiv:2205.13802 [quant-ph].
[93] B. Swingle and M. Van Raamsdonk, Enhanced negative energy with a massless Dirac field, JHEP **08**, 183, arXiv:2212.02609 [hep-th].
[94] E. Flores, C. Ireland, N. Jamhour, V. Lasasso, N. Kurth, and M. Leinbach, Casimir force in discrete scalar fields I: 1D and 2D cases, arXiv:2309.00624 [quant-ph].
[95] K. Nakayama and K. Suzuki, Casimir effect in axion electrodynamics with lattice regularizations, arXiv:2310.18092 [hep-th].
[96] C. W. J. Beenakker, Topologically protected Casimir effect for lattice fermions, arXiv:2402.02477 [quant-ph].
[97] C. M. Bender and P. Hays, Zero Point Energy of Fields in a Finite Volume, Phys. Rev. D **14**, 2622 (1976).
[98] P. Hays, Vacuum fluctuations of a confined massive field in two dimensions, Annals Phys. **121**, 32 (1979).
[99] J. Ambjørn and S. Wolfram, Properties of the vacuum. I. Mechanical and thermodynamic, Annals Phys. **147**, 1 (1983).
[100] S. K. Kim, W. Namgung, K. S. Soh, and J. H. Yee, Dynamical symmetry breaking and space-time topology, Phys. Rev. D **36**, 3172 (1987).
[101] J. Braun, B. Klein, and H. J. Pirner, Volume dependence of the pion mass in the quark-meson-model, Phys. Rev. D **71**, 014032 (2005), arXiv:hep-ph/0408116.
[102] L. M. Abreu, M. Gomes, and A. J. da Silva, Finite-size effects on the phase structure of the Nambu-Jona-Lasinio model, Phys. Lett. B **642**, 551 (2006), arXiv:hep-th/0610111.
[103] D. Ebert and K. G. Klimenko, Cooper pairing and finite-size effects in a NJL-type four-fermion model, Phys. Rev. D **82**, 025018 (2010), arXiv:1005.0699 [hep-ph].
[104] A. Flachi, Strongly Interacting Fermions and Phases of the Casimir Effect, Phys. Rev. Lett. **110**, 060401 (2013), arXiv:1301.1193 [hep-th].
[105] T. Ishikawa, K. Nakayama, and K. Suzuki, Casimir effect for nucleon parity doublets, Phys. Rev. D **99**, 054010 (2019), arXiv:1812.10964 [hep-ph].
[106] T. Ishikawa, K. Nakayama, D. Suenaga, and K. Suzuki, D mesons as a probe of Casimir effect for chiral symmetry breaking, Phys. Rev. D **100**, 034016 (2019), arXiv:1905.11164 [hep-ph].
[107] T. Inagaki, Y. Matsuo, and H. Shimoji, Precise phase structure in a four-fermion interaction model on a torus, PTEP **2022**, 013B09 (2022), arXiv:2108.03583 [hep-ph].
[108] O. Pavlovsky and M. Ulybyshev, Casimir energy calculations within the formalism of the noncompact lattice QED, Int. J. Mod. Phys. A **25**, 2457 (2010), arXiv:0911.2635 [hep-lat].
[109] O. V. Pavlovsky and M. V. Ulybyshev, Casimir energy in noncompact lattice electrodynamics, Theor. Math. Phys. **164**, 1051 (2010), [Teor. Mat. Fiz. **164**, 262

(2010)].
[110] O. Pavlovsky and M. Ulybyshev, Monte-Carlo calculation of the lateral Casimir forces between rectangular gratings within the formalism of lattice quantum field theory, Int. J. Mod. Phys. A **26**, 2743 (2011), arXiv:1105.0544 [quant-ph].
[111] O. Pavlovsky and M. Ulybyshev, Casimir energy in the compact QED on the lattice, arXiv:0901.1960 [hep-lat].
[112] M. N. Chernodub, V. A. Goy, and A. V. Molochkov, Casimir effect on the lattice: U(1) gauge theory in two spatial dimensions, Phys. Rev. D **94**, 094504 (2016), arXiv:1609.02323 [hep-lat].
[113] M. N. Chernodub, V. A. Goy, and A. V. Molochkov, Nonperturbative Casimir effect and monopoles: compact Abelian gauge theory in two spatial dimensions, Phys. Rev. D **95**, 074511 (2017), arXiv:1703.03439 [hep-lat].
[114] M. N. Chernodub, V. A. Goy, and A. V. Molochkov, Casimir effect and deconfinement phase transition, Phys. Rev. D **96**, 094507 (2017), arXiv:1709.02262 [hep-lat].
[115] M. N. Chernodub, V. A. Goy, A. V. Molochkov, and A. S. Tanashkin, Casimir boundaries, monopoles, and deconfinement transition in $(3+1)$- dimensional compact electrodynamics, Phys. Rev. D **105**, 114506 (2022), arXiv:2203.14922 [hep-lat].
[116] M. N. Chernodub, V. A. Goy, A. V. Molochkov, and H. H. Nguyen, Casimir Effect in Yang-Mills Theory in $D = 2 + 1$, Phys. Rev. Lett. **121**, 191601 (2018), arXiv:1805.11887 [hep-lat].
[117] M. N. Chernodub, V. A. Goy, and A. V. Molochkov, Phase structure of lattice Yang-Mills theory on $\mathbb{T}^2 \times \mathbb{R}^2$, Phys. Rev. D **99**, 074021 (2019), arXiv:1811.01550 [hep-lat].
[118] M. Kitazawa, S. Mogliacci, I. Kolbé, and W. A. Horowitz, Anisotropic pressure induced by finite-size effects in SU(3) Yang-Mills theory, Phys. Rev. D **99**, 094507 (2019), arXiv:1904.00241 [hep-lat].
[119] M. N. Chernodub, V. A. Goy, A. V. Molochkov, and A. S. Tanashkin, Boundary states and non-Abelian Casimir effect in lattice Yang-Mills theory, Phys. Rev. D **108**, 014515 (2023), arXiv:2302.00376 [hep-lat].
[120] A. O. Sushkov, W. J. Kim, D. A. R. Dalvit, and S. K. Lamoreaux, Observation of the thermal Casimir force, Nature Phys. **7**, 230 (2011), arXiv:1011.5219 [quant-ph].
[121] E. J. Ferrer and V. de la Incera, Dissipationless Hall Current in Dense Quark Matter in a Magnetic Field, Phys. Lett. B **769**, 208 (2017), arXiv:1611.00660 [nucl-th].
[122] P. Sikivie, Experimental Tests of the Invisible Axion, Phys. Rev. Lett. **51**, 1415 (1983), [Erratum: Phys. Rev. Lett. **52**, 695 (1984)].
[123] F. Wilczek, Two Applications of Axion Electrodynamics, Phys. Rev. Lett. **58**, 1799 (1987).
[124] I. Brevik, Axion Electrodynamics and the Axionic Casimir Effect, Universe **7**, 133 (2021), arXiv:2202.11152 [hep-ph].
[125] F. Canfora, D. Dudal, T. Oosthuyse, P. Pais, and L. Rosa, The Casimir effect in chiral media using path integral techniques, JHEP **09**, 095, arXiv:2207.09175 [hep-th].
[126] T. Oosthuyse and D. Dudal, Interplay between chiral media and perfect electromagnetic conductor plates: Repulsive vs. attractive Casimir force transitions, SciPost Phys. **15**, 213 (2023), arXiv:2301.12870 [hep-th].
[127] A. M. Favitta, I. H. Brevik, and M. M. Chaichian, Axion electrodynamics: Green's functions, zero-point energy and optical activity, Annals Phys. **455**, 169396 (2023), arXiv:2302.13129 [hep-th].
[128] E. J. Ferrer and V. de la Incera, Magnetic Dual Chiral Density Wave: A Candidate Quark Matter Phase for the Interior of Neutron Stars, Universe **7**, 458 (2021), arXiv:2201.04032 [hep-ph].
[129] M. Thies, Analytical solution of the Gross-Neveu model at finite density, Phys. Rev. D **69**, 067703 (2004), arXiv:hep-th/0308164.
[130] M. Thies and K. Urlichs, Revised phase diagram of the Gross-Neveu model, Phys. Rev. D **67**, 125015 (2003), arXiv:hep-th/0302092.
[131] J. Lenz, L. Pannullo, M. Wagner, B. Wellegehausen, and A. Wipf, Inhomogeneous phases in the Gross-Neveu model in 1+1 dimensions at finite number of flavors, Phys. Rev. D **101**, 094512 (2020), arXiv:2004.00295 [hep-lat].
[132] V. Schön and M. Thies, Emergence of the Skyrme crystal in Gross-Neveu and 't Hooft models at finite density, Phys. Rev. D **62**, 096002 (2000), arXiv:hep-th/0003195.
[133] T. Kojo, A (1+1) dimensional example of Quarkyonic matter, Nucl. Phys. A **877**, 70 (2012), arXiv:1106.2187 [hep-ph].
[134] T. Hayata, Y. Hidaka, and K. Nishimura, Dense QCD_2 with matrix product states, arXiv:2311.11643 [hep-lat].
[135] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, A new extended model of hadrons, Phys. Rev. D **9**, 3471 (1974).
[136] S. G. Mamaev and N. N. Trunov, Vacuum expectation values of the energy-momentum tensor of quantized fields on manifolds with different topologies and geometries. III, Sov. Phys. J. **23**, 551 (1980).
[137] F. C. Santos and A. C. Tort, The Casimir energy of a massive fermion field revisited, (2002), arXiv:quant-ph/0201104.
[138] E. Elizalde, F. C. Santos, and A. C. Tort, The Casimir energy of a massive fermionic field confined in a (d+1)-dimensional slab bag, Int. J. Mod. Phys. A **18**, 1761 (2003), arXiv:hep-th/0206114.

arXiv:2402.17584v1 [hep-ph] 27 Feb 2024

Triangle singularity in the $J/\psi \to \phi\pi^+ a_0^-(\pi^-\eta)$, $\phi\pi^- a_0^+(\pi^+\eta)$ decays

C. W. Xiao,[1,2,3,*] J. M. Dias,[4,†] L. R. Dai,[5,‡] W. H. Liang,[1,2,§] and E. Oset[1,6,¶]

[1] *Department of Physics, Guangxi Normal University, Guilin 541004, China*

[2] *Guangxi Key Laboratory of Nuclear Physics and Technology, Guangxi Normal University, Guilin 541004, China*

[3] *School of Physics, Central South University, Changsha 410083, China*

[4] *CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

[5] *School of Science, Huzhou University, Huzhou 313000, Zhejiang, China*

[6] *Departamento de Física Teórica and IFIC, Centro Mixto Universidad de Valencia-CSIC, Institutos de Investigación de Paterna, Aptdo. 22085, 46071 Valencia, Spain*

(Dated: February 28, 2024)

Abstract

We study the $J/\psi \to \phi\pi^+ a_0(980)^-(a_0^- \to \pi^-\eta)$ decay, evaluating the double mass distribution in terms of the $\pi^-\eta$ and $\pi^+ a_0$ invariant masses. We show that the $\pi^-\eta$ mass distribution exhibits the typical cusp structure of the $a_0(980)$ seen in recent high statistics experiments, and the $\pi^+ a_0$ spectrum shows clearly a peak around $M_{\rm inv}(\pi^+ a_0) = 1420\,\text{MeV}$, corresponding to a triangle singularity. When integrating over the two invariant masses we find a branching ratio for this decay of the order of 10^{-5}, which is easily accessible in present laboratories.

* xiaochw@gxnu.edu.cn

† jorgivan.mdias@itp.ac.cn

‡ dailianrong@zjhu.edu.cn

§ liangwh@gxnu.edu.cn

¶ Oset@ific.uv.es

I. INTRODUCTION

Triangle singularities (TS), introduced in Refs. [1, 2], correspond to processes in which there is a triangle diagram in the amplitude, where the three intermediate particles can be placed simultaneously on shell, representing a reaction that can occur at the classical level [3], in which case the amplitude becomes infinite in the limit of zero width of the intermediate particles. The issue has experienced a revival in recent years, one of the reasons, among many, being the fact that after early claims by the COMPASS Collaboration of the "$a_1(1420)$" discovery [4], it was soon explained as a consequence of a triangle singularity in which the $a_1(1260)$ resonance decays into $K^*\bar{K}$, $K^* \to \pi K$, and then $K\bar{K}$ fuse to give the $f_0(980)$ resonance, the $\pi f_0(980)$ being the observed decay mode [5–8]. Earlier than that, there was an interpretation of the isospin violating decay $\eta(1405) \to f_0(980)\pi^0$ [9] also in terms of a triangle singularity [10–12], and more recently the interpretation of the $pp \to \pi^+ d$ reaction [13, 14] in terms of a triangle singularity [15]. More examples were given in the interpretation of the $J/\psi \to \eta\pi^0\phi$ reaction in Ref. [16], and in the enhanced isospin violation in $D_s^+ \to \pi^+\pi^0 a_0(980)[f_0(980)]$ or $\bar{B}_s^0 \to J/\psi\pi^0 a_0(980)[f_0(980)]$ [17, 18]. A long list of reactions showing effects of triangle singularities can be seen in Table 1 of the review paper [19]. Also, a reformulation of the TS has been provided in Ref. [20], which is at the same time pedagogical and practical.

The issue has emerged once more due to the recent BESIII paper [21], improving considerably on earlier measurements of the $J/\psi \to \eta\pi^0\phi$ reaction, where the ideas of Ref. [16] could be tested. Indeed, that reaction develops a TS in the $\pi^0\phi$ mass distribution, peaking around $M(\phi\pi^0) \sim 1400$ MeV, where a clear signal is seen in Ref. [21]. Yet, the interpretation of the peak is delicate, because, as discussed in Ref. [16], the peak is masked by a tree level contribution which has also a singular behaviour. Indeed, the TS mechanism proceeds via the diagrams of Figs. 1 (a) and (b), but the same final state is reached by the tree level diagrams of Figs. 1 (c) and (d). Then, Schmid theorem [22] comes into play because if the triangle diagrams develop a TS they can be reabsorbed into the tree level diagrams with a simple change of phase, and the decay width is not affected by the TS diagrams. A thorough review of this issue was done in Ref. [23], showing that the theorem holds in the limit of the K^* width (in the present case) going to zero, and when there are no inelastic channels. The message of Ref. [23] is that all diagrams must be calculated, and that normally the tree

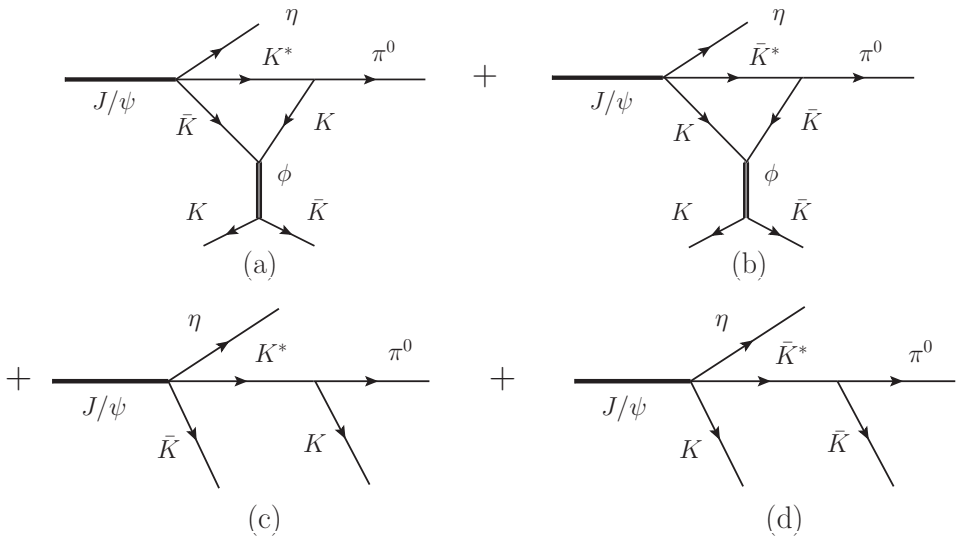


FIG. 1. Mechanisms in $J/\psi \to \eta\pi^0\phi$ ($\phi \to K\bar{K}$): (a), (b) TS; (c), (d) the tree level.

level diagrams are mostly responsible for the mass distributions, with the effects of the TS diagrams being diluted in these distributions.

In view of these problems we look now at a related reaction where a TS develops, which is not masked by the effects of tree level and Schmid theorem. The reaction is $J/\psi \to \phi\pi^- a_0(980)^+(\pi^+\eta)$, $\phi\pi^+ a_0(980)^-(\pi^-\eta)$. The mechanism is similar to that in Figs. 1(a) and 1(b), with η replaced by ϕ, but the intermediate $K\bar{K}$ produce the $a_0(980)$ resonance which decays to $\pi\eta$. Then, the tree level diagrams with $K\bar{K}$ production do not interfere with the triangle diagram, which also develops a singularity in the $\pi a_0(980)$ mass distribution. It is easy to see where we should expect the singularity applying Eq. (18) of Ref. [20] (with m_{a_0} slightly larger than $2\,m_K$), and one finds that a singularity should appear at $M_{\rm inv}(\pi a_0) \sim 1417$ MeV. The purpose of the present work is to do a detailed study of the reaction and make a realistic prediction of the shape and size of the $\pi a_0(980)$ mass distribution in that reaction.

II. FORMALISM

We look at the diagrams of Fig. 2. The diagrams of Figs. 2(a) and 2(b) show the coalescence processes where the a_0^-, a_0^+ are produced, independent on the decay channel in which the a_0 resonances are observed. The diagrams of Figs. 2(c) and 2(d) show the ex-

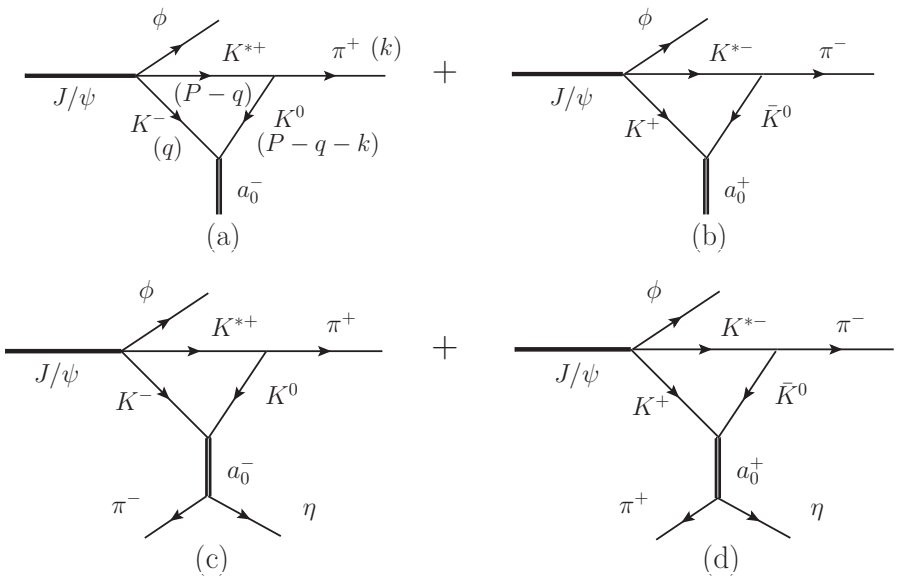


FIG. 2. Triangle diagrams for $J/\psi \to \phi\pi^+ a_0^-$ decay (a) and $J/\psi \to \phi\pi^- a_0^+$ decay (b). Figs. (c) and (d) illustrate the processes of (a) and (b) respectively, with a clear depiction of the decay channel of a_0^- and a_0^+. In Fig. (a), the momenta of the particles are shown, where $P = p_{J/\psi} - p_\phi$.

plicit reaction, with four body final state when the $a_0^{-/+}$ decay into $\pi^{-/+}\eta$, the only sizeable decay channel. We shall consider this final state, but it is practical to consider first the coalescence processes, with only three particles in the final state. We consider the reactions $J/\psi \to \phi\pi^- a_0^+$ and $J/\psi \to \phi\pi^+ a_0^-$ as different reactions and will concentrate on the first one. The same distributions would be obtained for the second reaction.

In order to be able to determine absolute rates for the $J/\psi \to \phi\pi^- a_0^+$ ($J/\psi \to \phi\pi a_0$) reaction we need information on the $J/\psi \to \phi K^{*-} K^+$ reaction, which we take from experiment. On the other hand, the dynamics of $K^* \to K\pi$ and $K\bar{K} \to a_0 \to \pi\eta$ are well known, and for the $K\bar{K} \to a_0 \to \pi\eta$ amplitudes we shall use the chiral unitary approach [24–27].

A. The $J/\psi \to \phi K^* \bar{K}$ reaction

In the Particle Data Group (PDG) [28], we have the branching ratio,

$$\mathrm{Br}(J/\psi \to \phi K^*(892)\bar{K} + c.c.) = (2.18 \pm 0.23) \times 10^{-3}. \tag{1}$$

However, we are only interested in $J/\psi \to \phi K^{*-} K^+$. It is easy to see the rate for this particular channel using isospin and C parity arguments. With the isospin convention for the

multiplets $(K^+,\, K^0)$, $(\bar{K}^0,\, -K^-)$, $(K^{*+},\, K^{*0})$, $(\bar{K}^{*0},\, -K^{*-})$, the $K^*\bar{K}$ isospin zero states demanded in the reaction of Eq. (1) are given by

$$\begin{aligned} |K^*\bar{K},\ I=0\rangle &= -\frac{1}{\sqrt{2}}\,\left(K^{*+}K^- + K^{*0}\bar{K}^0\right), \\ |\bar{K}^*K,\ I=0\rangle &= \frac{1}{\sqrt{2}}\,\left(\bar{K}^{*0}K^0 + K^{*-}K^+\right). \end{aligned} \tag{2}$$

Knowing that $CK^{*+} = -K^{*-}$, $CK^+ = K^-$ *etc*, the right combination for $J/\psi \to \phi K^*\bar{K} + c.c.$ is given by

$$J/\psi \to \phi\left(K^{*+}K^- + K^{*0}\bar{K}^0 - K^{*-}K^+ - \bar{K}^{*0}K^0\right), \tag{3}$$

which means that the branching ratio of $J/\psi \to \phi K^{*+}K^-$ will be one fourth of the one in Eq. (1),

$$\mathrm{Br}(J/\psi \to \phi K^{*+}K^-) = (0.55 \pm 0.06) \times 10^{-3}. \tag{4}$$

Furthermore, the structure of the amplitude in S-wave is given by

$$V_{J/\psi,\,\phi K^{*+}K^-} = C\ \vec{\epsilon}_{J/\psi} \cdot (\vec{\epsilon}_\phi \times \vec{\epsilon}_{K^*})\,. \tag{5}$$

We can determine C from the rate of Eq. (4) using

$$\frac{d\Gamma_{J/\psi\to\phi K^{*+}K^-}}{dM_{\rm inv}(K^{*+}K^-)} = \frac{1}{(2\pi)^3}\ \frac{1}{4M^2_{J/\psi}}\ p_\phi\, \tilde{p}_{K^-}\ \overline{\sum}\sum |t|^2, \tag{6}$$

with

$$\begin{aligned} p_\phi &= \frac{\lambda^{1/2}(M^2_{J/\psi},\, M^2_\phi,\, M^2_{\rm inv}(K^{*+}K^-))}{2M_{J/\psi}}, \\ \tilde{p}_{K^-} &= \frac{\lambda^{1/2}(M^2_{\rm inv}(K^{*+}K^-),\, M^2_{K^{*+}},\, M^2_{K^-})}{2M_{\rm inv}(K^{*+}K^-)}, \end{aligned} \tag{7}$$

and $\overline{\sum}\sum |t|^2$ stands for the average and sum over the polarizations of J/ψ, ϕ, and K^* mesons, with t defined in Eq. (5). Therefore,

$$\overline{\sum}\sum |t|^2 = 2\,C^2. \tag{8}$$

Then we find

$$\frac{C^2}{\Gamma_{J/\psi}} = \frac{\mathrm{Br}(J/\psi \to \phi K^{*+}K^-)}{\displaystyle\int \frac{2}{(2\pi)^3}\ \frac{1}{4M^2_{J/\psi}}\, p_\phi\, \tilde{p}_{K^-}\ dM_{\rm inv}(K^{*+}K^-)}, \tag{9}$$

from where we obtain

$$\frac{C^2}{\Gamma_{J/\psi}} = 1.381 \times 10^{-2} \quad (\mathrm{MeV}^{-1}), \tag{10}$$

which we will use to evaluate the strength of the triangle mechanism.

B. The $a_0^- \to K^- K^0$ coupling and $K^* \to K\pi$ vertex

The $K^{*+} \to K^0\pi^+$ coupling is easily obtained from the standard Lagrangian

$$\mathcal{L} = -ig\langle [P,\ \partial_\mu P] V^\mu \rangle, \tag{11}$$

with P and V representing the SU(3) $q\bar{q}$ matrix written in terms of pseudoscalar or vector mesons, respectively [29]. The coupling g is defined as $g = \frac{M_V}{2f}$, where $M_V = 800$ MeV and $f = 93$ MeV. This yields the vertex

$$-it = -ig\epsilon_j(K^*)(2k+q)^j, \tag{12}$$

which is evaluated in the frame where we take $\vec{P} = \vec{p}_{J/\psi} - \vec{p}_\phi = 0$. In this frame, we can neglect the ϵ^0 component of the K^* [30].

The coupling of $a_0 \to K^+K^-$ needed in the evaluation of the diagram of Fig. 2(a) without further decay of $a_0 \to \pi^-\eta$ can be accounted for in the following way. It is clear that if we evaluate the $J/\psi \to \phi\pi^+\pi^-\eta$ decay we would only need the $K^-K^0 \to \pi^-\eta$ amplitude. The coalescence decay $J/\psi \to \phi\pi^+ a_0^-$ should also be able to be calculated using the $K\bar{K}$ amplitudes and this is formally done as discussed below.

Assuming that, close to the peak of the $a_0(980)$,

$$t_{K^-K^0,\,K^-K^0}(M_{\rm inv}) = \frac{g^2_{a_0,\,K^-K^0}}{M^2_{\rm inv} - m^2_{a_0} + iM_{\rm inv}\Gamma_{a_0}}, \tag{13}$$

then, using Cauchy's integration we find immediately

$$g^2_{a_0,\,K^-K^0} = -\frac{1}{\pi}\int dM^2_{\rm inv}\ {\rm Im}\, t_{K^-K^0,\,K^-K^0}(M_{\rm inv}), \tag{14}$$

which is also trivially obtained in the limit of $\Gamma_{a_0} \to 0$ using

$${\rm Im}\left[\left(M^2_{\rm inv} - m^2_{a_0}\right) + i\epsilon\right]^{-1} = -i\pi\delta\left(M^2_{\rm inv} - m^2_{a_0}\right). \tag{15}$$

In the coalescence process we will use Eq. (14) in the evaluation of the $|t_{\rm TS}|^2$ of the triangle diagram. We will have

$$\frac{d\Gamma_{J/\psi,\to\phi\pi^+a_0^-}}{dM_{\rm inv}(\pi^+a_0^-)} = \frac{1}{(2\pi)^3}\ \frac{1}{4M^2_{J/\psi}}\ p_\phi\ \tilde{p}_{\pi^+}\ \overline{\sum}|t_{\rm TS}|^2, \tag{16}$$

where

$$p_\phi = \frac{\lambda^{1/2}(M^2_{J/\psi},\, m^2_\phi,\, M^2_{\rm inv}(\pi^+a_0^-))}{2M_{J/\psi}}, \tag{17}$$

$$\tilde{p}_{\pi^+} = \frac{\lambda^{1/2}(M^2_{\rm inv}(\pi^+a_0^-),\, m^2_{\pi^+},\, m^2_{a_0^-})}{2M_{\rm inv}(\pi^+a_0^-)}, \tag{18}$$

with $t_{\rm TS}$ corresponding to the amplitude for the process of Fig. 2(a). Since $|t_{\rm TS}|^2$ contains $g^2_{a_0,\,K^-K^0}$, given in Eq. (14), we can undo the $dM^2_{\rm inv}$ integration in Eq. (16) and write

$$\frac{d^2\Gamma_{J/\psi,\to\phi\pi^+a_0^-(\pi^-\eta)}}{dM_{\rm inv}(\pi^-\eta)\,dM_{\rm inv}(\pi^+a_0^-)} = -\frac{1}{\pi}\,2M_{\rm inv}(\pi^-\eta)\,{\rm Im}\,t_{K^-K^0,\,K^-K^0}(M_{\rm inv}(\pi^-\eta)) \times \frac{1}{(2\pi)^3}\,\frac{1}{4M^2_{J/\psi}}\,p_\phi\,\tilde{p}'_{\pi^+}\overline{\sum}|\tilde{t}_{\rm TS}|^2, \tag{19}$$

where $\tilde{p}'_{\pi^+}$ is given by Eq. (18) substituting $m_{a_0^-}$ by $M_{\rm inv}(\pi^-\eta)$, and $|\tilde{t}_{\rm TS}|^2$ is the magnitude $|t_{\rm TS}|^2$ where we remove $g^2_{a_0,\,K^-K^0}$.

Eq. (19) can be immediately reinterpreted. Indeed, in the evaluation of $|t_{\rm TS}|^2$ for the triangle diagram of Fig. 2(c), we will need $|t_{K^-K^0,\,\pi^-\eta}|^2$ along with the extra phase space for $\pi^-\eta$ with respect to Fig. 2(a). But $|t_{K^-K^0,\,\pi^-\eta}|^2$ times the phase space for $a_0^-\to\pi^-\eta$ decay is what is given by ${\rm Im}\,t_{K^-K^0,\,K^-K^0}$ via the optical theorem. The derivation done, however, has served to go from the differential mass distribution in the three body final state to the one of the four body in a simple and intuitive way.

C. The triangle amplitude

In Fig. 2(a), we show the momenta of the particles. Since we are concerned about the triangle singularity, occurring when the intermediate particles are all on shell, we can simplify the calculation (see Ref. [20]) taking the positive energy part of the propagators, which is the one that can go on shell. Hence for a K^- propagator we would write

$$\frac{1}{q^2-m_K^2+i\epsilon} = \frac{1}{2\omega(\vec{q})}\left(\frac{1}{q^0-\omega_K(\vec{q})+i\epsilon}-\frac{1}{q^0+\omega_K(\vec{q})-i\epsilon}\right), \tag{20}$$

with $\omega_K(\vec{q})=\sqrt{\vec{q}^{\,2}+m_K^2}$, and with q^0 positive only the first term of the former equation can go on shell, and we shall then keep this term alone. This simplifies the expression for the loop amplitude which reads, removing the $g_{a_0,\,K^-K^0}$ vertex:

$$\begin{aligned}-i\tilde{t}_{\rm TS} = &-i\,C\int\frac{d^4q}{(2\pi)^4}\,\epsilon_{ijl}\,\epsilon_i(J/\psi)\,\epsilon_j(\phi)\,\epsilon_l(K^*)\,(-i)g\,\epsilon_m(K^*)\,(2k+q)_m \\ &\times(-i)\,\frac{1}{2\omega_{K^-}(\vec{q})}\,\frac{1}{2\omega_{K^0}(\vec{q}+\vec{k})}\,\frac{1}{2\omega_{K^{*+}}(\vec{q})}\,\frac{i}{q^0-\omega_{K^-}(\vec{q})+i\epsilon} \\ &\times\frac{i}{P^0-q^0-\omega_{K^{*+}}(\vec{q})+i\frac{\Gamma_{K^*}}{2}}\,\frac{i}{P^0-q^0-k^0-\omega_{K^0}(\vec{q}+\vec{k})+i\epsilon},\end{aligned} \tag{21}$$

with $\omega_{K^-}(\vec{q}) = \sqrt{\vec{q}^{\,2} + m_K^2}$, $\omega_{K^{*+}}(\vec{q}) = \sqrt{\vec{q}^{\,2} + m_{K^*}^2}$, where we have explicitly taken into account the K^* width in the K^* propagator, and P^0, k^0 are given by

$$\begin{aligned} P^0 &= M_{\rm inv}(\pi^+ a_0^-), \\ k^0 &= \frac{P^{02} + m_{\pi^+}^2 - M_{\rm inv}^2(\pi^-\eta)}{2P^0}. \end{aligned} \tag{22}$$

We sum over the K^* polarizations $\sum \epsilon_l(K^*)\epsilon_m(K^*) = \delta_{lm}$, and integrating analytically over q^0 in Eq. (21) using Cauchy's residues, we obtain

$$\begin{aligned} \tilde{t}_{\rm TS} =& g\, C\, \epsilon_{ijl}\, \epsilon_i(J/\psi)\, \epsilon_j(\phi) \int \frac{d^3q}{(2\pi)^3}\, (2k+q)_l\, \frac{1}{2\omega_{K^-}(\vec{q})}\, \frac{1}{2\omega_{K^{*+}}(\vec{q})}\, \frac{1}{2\omega_{K^0}(\vec{q}+\vec{k}\,)} \\ &\times \frac{i}{P^0 - \omega_{K^-}(\vec{q}) - \omega_{K^{*+}}(\vec{q}) + i\frac{\Gamma_{K^*}}{2}}\, \frac{i}{P^0 - k^0 - \omega_{K^-}(\vec{q}) - \omega_{K^0}(\vec{q}+\vec{k}\,) + i\epsilon}. \end{aligned} \tag{23}$$

Next, considering that

$$\int d^3q\, F(\vec{q},\vec{k})\, q_l = k_l \int d^3q\; F(\vec{q},\vec{k})\, \frac{\vec{q}\cdot\vec{k}}{\vec{k}^{\,2}}, \tag{24}$$

we finally write $\tilde{t}_{\rm TS}$ as

$$\tilde{t}_{\rm TS} = g\, C\, \epsilon_{ijl}\, \epsilon_i(J/\psi)\, \epsilon_j(\phi)\, k_l\, \tilde{t}'_{\rm TS}, \tag{25}$$

with

$$\begin{aligned} \tilde{t}'_{\rm TS} =& \int \frac{d^3q}{(2\pi)^3}\, \theta(q_{max} - |\vec{q}^{\,*}|\,) \left(2 + \frac{\vec{q}\cdot\vec{k}}{\vec{k}^{\,2}}\right) \frac{1}{2\omega_{K^-}(\vec{q})}\, \frac{1}{2\omega_{K^{*+}}(\vec{q})}\, \frac{1}{2\omega_{K^0}(\vec{q}+\vec{k}\,)} \\ &\times \frac{i}{P^0 - \omega_{K^-}(\vec{q}) - \omega_{K^{*+}}(\vec{q}) + i\dfrac{\Gamma_{K^*}}{2}}\, \frac{i}{P^0 - k^0 - \omega_{K^-}(\vec{q}) - \omega_{K^0}(\vec{q}+\vec{k}\,) + i\epsilon}, \end{aligned} \tag{26}$$

and

$$\overline{\sum} \left|\tilde{t}_{\rm TS}\right|^2 = \frac{2}{3}\, \vec{k}^{\,2}\, g^2\, C^2\, \left|\tilde{t}'_{\rm TS}\right|^2. \tag{27}$$

In Eq. (26) we have introduced the factor $\theta(q_{max} - |\vec{q}^{\,*}|)$, where $\vec{q}^{\,*}$ is the K^- momentum in the $\pi^-\eta$ rest frame given by [20]

$$\vec{q}^{\,*} = \left[\left(\frac{E_{a_0}}{M_{\rm inv}(\pi^-\eta)} - 1\right)\frac{\vec{q}\cdot\vec{k}}{\vec{k}^{\,2}} + \frac{q^0}{M_{\rm inv}(\pi^-\eta)}\right]\vec{k} + \vec{q}\,, \tag{28}$$

where $E_{a_0} = \sqrt{M_{\rm inv}^2 + \vec{k}^{\,2}}$, $q^0 = \sqrt{m_K^2 + \vec{q}^{\,2}}$. This factor is demanded because in the cutoff regularization to obtain the $K\bar{K}$ scattering we have $t(\vec{q},\vec{q}\,') = t\,\theta(q_{max} - |\vec{q}|)\,\theta(q_{max} - |\vec{q}\,'|)$, with q_{max} the cutoff momentum used in the regularization of the $G_{K\bar{K}}$ function [31].

D. $K\bar{K}$ amplitudes

We need to calculate the $t_{K^-K^0,\,K^-K^0}$ amplitude, for which we use the chiral unitary approach. In this case, the $t_{K^-K^0,K^-K^0}$ amplitude is one of the T-matrix elements, that is extracted by solving the Bethe-Salpeter equation in coupled-channels,

$$T = [1 - VG]^{-1}V, \tag{29}$$

with V the kernel encoding the V_{ij} amplitudes from i- to j-channel. The channels considered are: $K\bar{K}$, $\pi\eta$, $\pi\pi$, and $\eta\eta$. The relevant V_{ij} amplitudes among all the channels in our case are taken from Ref. [32], using explicitly the $\eta - \eta'$ mixing of Ref. [33] (Eq. (A.4) of Ref. [32]). Furthermore, G is a diagonal matrix with its elements G_l corresponding to the loop function for the l-th channel. We take the G_l loop function regularized by means of a cut-off $q_{\rm max}$ in the three-monentum,

$$G_l = \int_{|\vec{q}|<q_{\rm max}} \frac{d^3q}{(2\pi)^3} \frac{\omega_1+\omega_2}{2\omega_1\,\omega_2} \frac{1}{s-(\omega_1+\omega_2)^2+i\epsilon}. \tag{30}$$

Since, only the zero charged components are considered in Ref. [32], we make use of the fact that K^-K^0 is the $I=1$, $I_3=-1$ component of $K\bar{K}$ and write,

$$t_{K^-K^0,\,K^-K^0} = \frac{1}{2}\left(t_{K^0\bar{K}^0,\,K^0\bar{K}^0} + t_{K^+K^-,\,K^+K^-} - 2\,t_{K^0\bar{K}^0,\,K^+K^-}\right). \tag{31}$$

III. RESULTS

In Fig. 3, we show the factor $A \equiv [-\frac{2}{\pi}\, M_{\rm inv}(\pi^-\eta)\, {\rm Im}\, t_{K^-K^0,K^-K^0}(M_{\rm inv}(\pi^-\eta))]$, which is involved in the double differential decay width of Eq. (19). We can see a cusp-like structure around $M_{\rm inv}(\pi^-\eta) = m_{a_0} = 980$ MeV, reflecting the spectral function of the $a_0(980)$.

Next, we discuss the amplitude $\tilde{t}'_{\rm TS}$ of Eq. (26) and $\frac{1}{\Gamma_{J/\psi}} \frac{{\rm d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{{\rm d}M_{\rm inv}(\pi^-\eta)\,{\rm d}M_{\rm inv}(\pi^+a_0)}$ of Eq. (19), which are functions of both $M_{\rm inv}(\pi^-\eta)$ and $M_{\rm inv}(\pi^+a_0)$. We will present the results in three cases: 1) fixing $M_{\rm inv}(\pi^-\eta) = m_{a_0} = 980$ MeV; 2) fixing $M_{\rm inv}(\pi^+a_0) = 1416$ MeV, where the TS occurs in the triangle loops of Fig. 2; and 3) integrating $\frac{1}{\Gamma_{J/\psi}} \frac{{\rm d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{{\rm d}M_{\rm inv}(\pi^-\eta)\,{\rm d}M_{\rm inv}(\pi^+a_0)}$ over $M_{\rm inv}(\pi^-\eta)$.

A. Fixing $M_{\rm inv}(\pi^-\eta) = m_{a_0} = 980$ MeV

In Fig. 4, we show results for ${\rm Re}\,\tilde{t}'_{\rm TS}$, ${\rm Im}\,\tilde{t}'_{\rm TS}$ and $|\tilde{t}'_{\rm TS}|$ as a function of $M_{\rm inv}(\pi^+a_0)$ while keeping $M_{\rm inv}(\pi^-\eta)$ fixed. The structure of the amplitude exhibits features typical of triangle

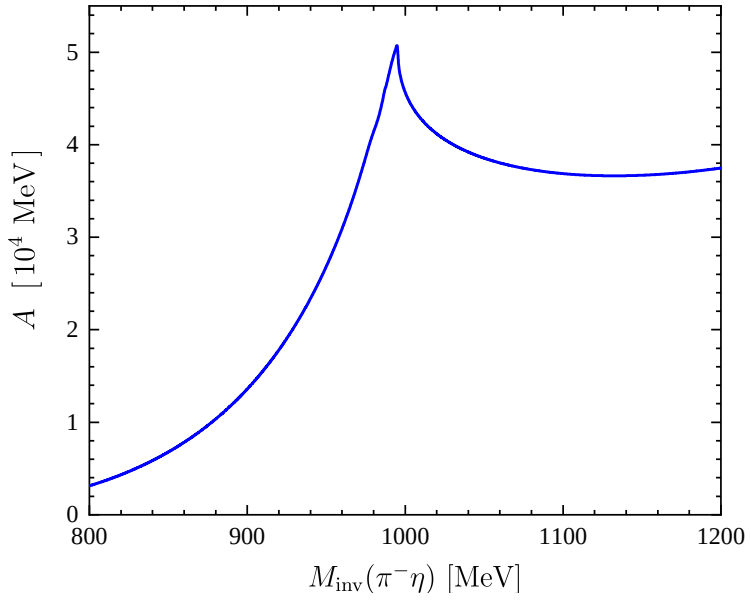


FIG. 3. Factor $A \equiv [\, -\frac{2}{\pi}\, M_{\rm inv}(\pi^-\eta)\, {\rm Im}\, t_{K^-K^0,K^-K^0}(M_{\rm inv}(\pi^-\eta))]$ as a function of $M_{\rm inv}(\pi^-\eta)$.

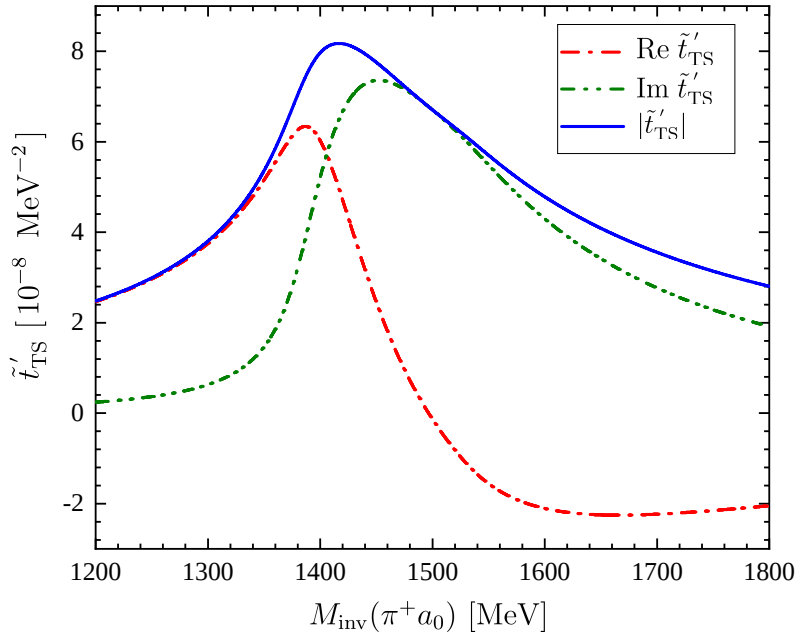


FIG. 4. $\tilde{t}'_{\rm TS}$ given by Eq. (26) as a function of $M_{\rm inv}(\pi^+a_0)$ when fixing $M_{\rm inv}(\pi^-\eta) = m_{a_0}$.

singularities observed in other cases (see Fig. 5 of Ref. [30], Fig. 4 of Ref. [34], Fig. 3 of Ref. [17], Fig. 4 of Ref. [35], Fig. 5 of Ref. [36], Fig. 4 of Ref. [37], Fig. 19 of Ref. [38]). It has the imaginary part peaking around the $M_{\rm inv}(\pi^+a_0)$, as provided by Eq. (18) of Ref. [20], and the real part changes sign around the peak of the imaginary part. It resembles much the structure of a resonance, and hence the danger to identify these peaks as genuine resonances,

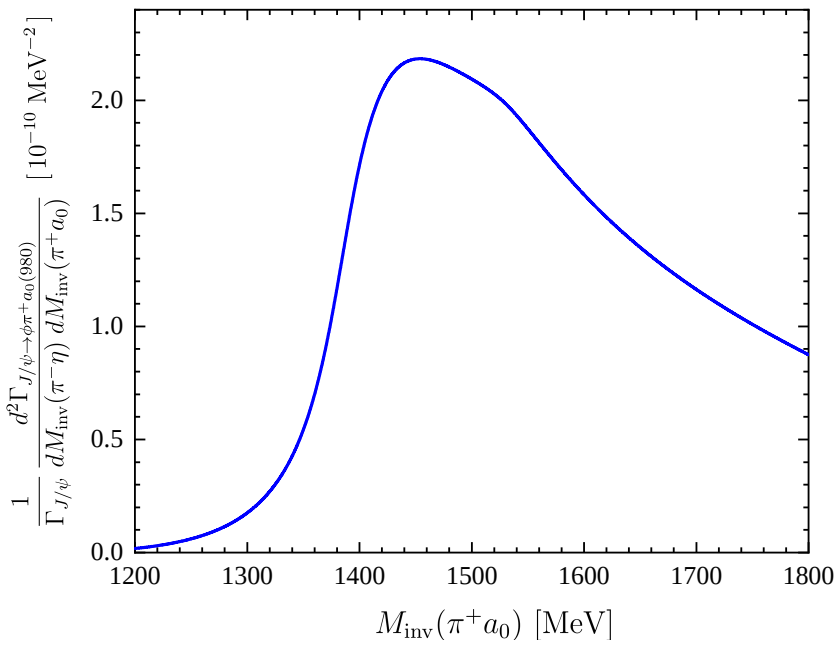


FIG. 5. $\frac{1}{\Gamma_{J/\psi}} \frac{\mathrm{d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\,\mathrm{d}M_{\rm inv}(\pi^+a_0)}$ as a function of $M_{\rm inv}(\pi^+a_0)$ when fixing $M_{\rm inv}(\pi^-\eta) = m_{a_0}$.

but as we can see, the origin of this structure simply comes from the triangle diagram and is exclusively tied to the combination of masses and invariant masses of the particles involved. It does not have its origin in the interaction of quarks or the interaction of hadrons. This is why it is referred to as a kinematical singularity. The modulus of this amplitude, $|\tilde{t}'_{\rm TS}|$, has a clear peak that should manifest in the studied reaction.

In Fig. 5, we show the double differential decay width normalized to the J/ψ width as a function of $M_{\rm inv}(\pi^+a_0)$, while fixing $M_{\rm inv}(\pi^-\eta) = m_{a_0} = 980$ MeV. We see a clear peak around $M_{\rm inv}(\pi^+a_0) = 1440$ MeV, coming from $|\tilde{t}'_{\rm TS}|$ in Eq. (19). This is what one would observe in a devoted experiment with a bin of 1 MeV for $M_{\rm inv}(\pi^+a_0)$, and 1 MeV for $M_{\rm inv}(\pi^-\eta)$. Obviously, the accumulation of events in bigger bins increases the statistics, as we shall see.

B. Fixing $M_{\rm inv}(\pi^+a_0) = 1416$ MeV

In Fig. 6, we now set $M_{\rm inv}(\pi^+a_0) = 1416$ MeV and plot $\tilde{t}'_{\rm TS}$ as a function of $M_{\rm inv}(\pi^-\eta)$. Once more, we display the real and imaginary parts of the amplitude, as well as its modulus. We see again that the imaginary part and the modulus delineate the shape of the $a_0(980)$ resonance. The real part changes sign at the peak of the $a_0(980)$, reflecting a typical resonance

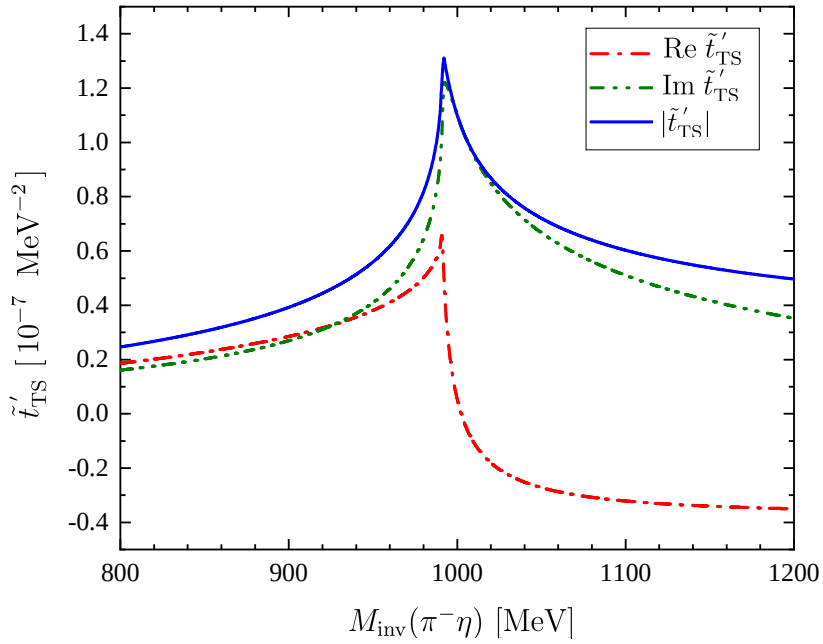


FIG. 6. $\tilde{t}'_{\rm TS}$ given by Eq. (26) as a function of $M_{\rm inv}(\pi^-\eta)$ when fixing $M_{\rm inv}(\pi^+a_0) = 1416$ MeV.

behaviour. It is interesting to see that even if the $a_0(980)$ appears as cusp, corresponding to a nearly missed bound state, or virtual state, it still exhibits the typical shape of a resonance amplitude. Such kind of behaviours for nearly missed bound states can be seen in other cases. For instance, in the $pd \to {}^3\mathrm{He}\,\eta$ reaction [39] (see Fig. 8 of that work), the amplitude exhibits a resonance structure with a peak very close to and just below the ${}^3\mathrm{He}\,\eta$ threshold. However, there is no pole below threshold, and technically, no bound state.

In Fig. 7, we show again $\frac{1}{\Gamma_{J/\psi}} \frac{\mathrm{d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\,\mathrm{d}M_{\rm inv}(\pi^+a_0)}$ as a function of $M_{\rm inv}(\pi^-\eta)$, fixing now $M_{\rm inv}(\pi^+a_0)$ at the peak of the TS amplitude. This comes from Eq. (19) and contains $|\tilde{t}'_{\rm TS}|^2$ along with the phase space. The shape of the $a_0(980)$ resonance shows up as a clear cusp structure, as seen in recent experiments with high resolution [40–42].

C. integrating $\frac{1}{\Gamma_{J/\psi}} \frac{\mathrm{d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\,\mathrm{d}M_{\rm inv}(\pi^+a_0)}$ over $M_{\rm inv}(\pi^-\eta)$

Fig. 8 shows $\frac{1}{\Gamma_{J/\psi}} \frac{\mathrm{d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\,\mathrm{d}M_{\rm inv}(\pi^+a_0)}$ as a function of $M_{\rm inv}(\pi^+a_0)$, when integrating over $M_{\rm inv}(\pi^-\eta)$ in the ranges of $m_{a_0} \pm 10$ MeV, $m_{a_0} \pm 20$ MeV, $m_{a_0} \pm 50$ MeV and $m_{a_0} \pm 100$ MeV, respectively.

In all the cases, we observe a peak corresponding to the TS. By looking at Fig. 8, we can see that integrating the double mass distribution over $M_{\rm inv}(\pi^-\eta)$ within the range $m_{a_0} \pm$

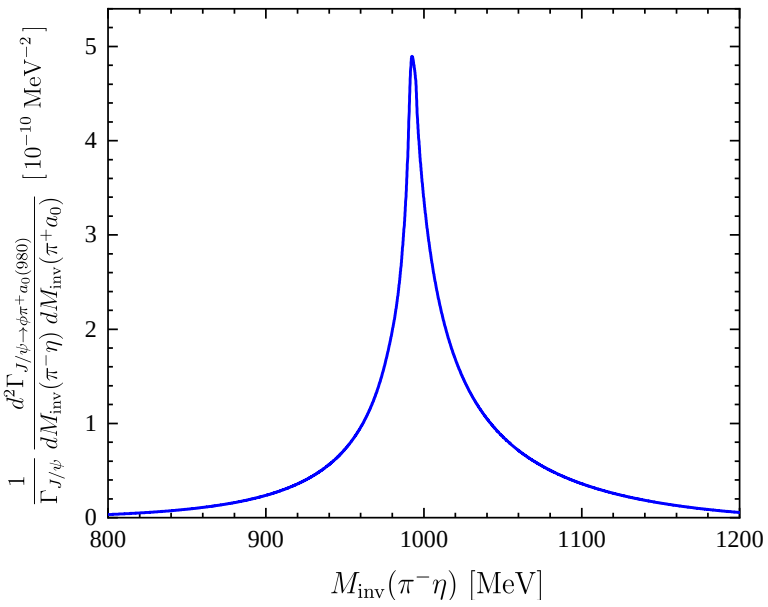


FIG. 7. $\frac{1}{\Gamma_{J/\psi}} \frac{\mathrm{d}^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\, \mathrm{d}M_{\rm inv}(\pi^+a_0)}$ as a function of $M_{\rm inv}(\pi^-\eta)$ when fixing $M_{\rm inv}(\pi^+a_0) = 1416$ MeV.

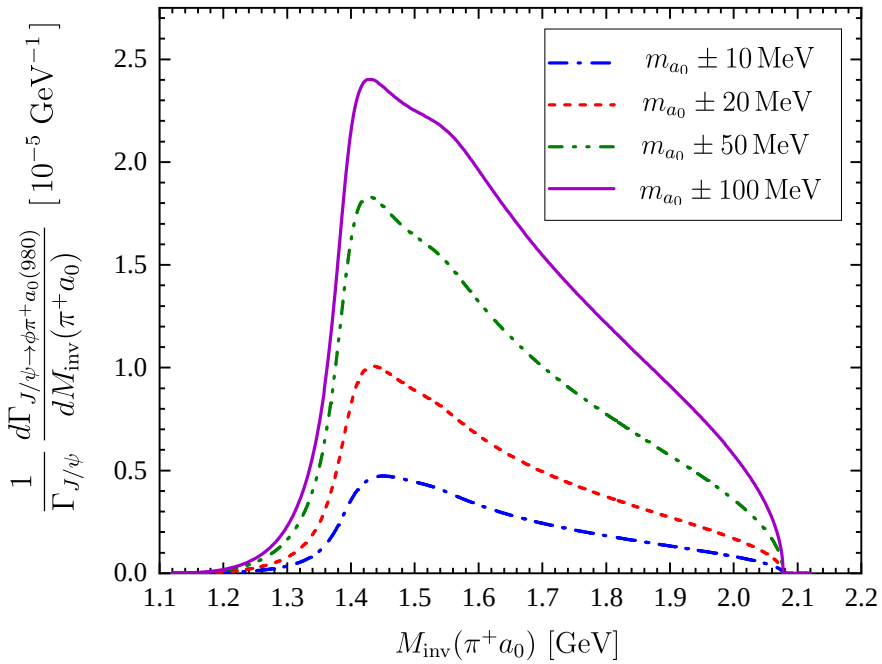


FIG. 8. $\frac{1}{\Gamma_{J/\psi}} \frac{d^2\Gamma_{J/\psi\to\phi\pi^+a_0(980)}}{\mathrm{d}M_{\rm inv}(\pi^-\eta)\, \mathrm{d}M_{\rm inv}(\pi^+a_0)}$ as a function of $M_{\rm inv}(\pi^+a_0)$ when integrating over $M_{\rm inv}(\pi^-\eta)$ in the ranges: $m_{a_0} \pm 10$ MeV, $m_{a_0} \pm 20$ MeV, $m_{a_0} \pm 50$ MeV and $m_{a_0} \pm 100$ MeV.

100 MeV accounts for the whole strength of the $a_0(980)$ resonance. And we interpret this as indicative of what should be observed in the experiments. The shape of the TS is clearly observed.

For the case where $M_{\rm inv}(\pi^-\eta) \in [m_{a_0}-100, m_{a_0}+100]$ MeV, integrating over $M_{\rm inv}(\pi^+ a_0)$ in the range $[m_{\pi^+}+m_{a_0}, M_{J/\psi}-m_\phi]$ gives the branching ratio

$$\text{Br}(J/\psi \to \phi\pi^+ a_0) = 1.07 \times 10^{-5}. \tag{32}$$

This estimate is realistic, as the only unknown magnitude required to evaluate the diagrams of Figs. 2(a) and 2(c) is the $J/\psi \to \phi K^{*+} K^-$ amplitude, which we obtained from the experimental data. This branching ratio is not small, given the copious production of J/ψ at BESIII, which allows to detect decays with branching fractions as small as 10^{-7} [28]. In view of this, we can only encourage the measurement of this reaction, which could also serve to clarify issues on the $J/\psi \to \pi^0\eta\phi\,(\phi \to K\bar{K})$ reaction [21] and its interpretation in Ref. [16] in terms of a TS.

IV. CONCLUSIONS

We have conducted a study of the $J/\psi \to \phi\pi^+ a_0(980)^-$ ($a_0^- \to \pi^-\eta$) decay, showing that it develops a triangle singularity at $M_{\rm inv}(\pi^+ a_0)$ of about 1420 MeV. The reaction proposed is motivated by the recent measurement at BESIII of the $J/\psi \to \eta\pi^0\phi$ ($\phi \to K\bar{K}$) reaction [21], that according to the work of Ref. [16] also develops a triangle singularity, however, blurred by the tree level competing mechanisms and their interconnection with the Coleman Norton theorem. In the reaction proposed there is no tree level competing mechanism and then the TS appearing can be clearly interpreted. We evaluate the mass distributions in terms of $M_{\rm inv}(\pi^-\eta)$ and $M_{\rm inv}(\pi^+ a_0)$. In the $\pi^-\eta$ mass distribution we see a clear cusp structure, as observed in recent high statistics experiments, and in the $\pi^+ a_0$ mass distribution we observe the TS peak around $M_{\rm inv}(\pi^+ a_0) = 1420$ MeV. By taking information for the needed $J/\psi \to \eta K^* \bar{K}$ amplitude from experiment, we are able to determine absolute rates for the reaction. Integrating the double mass distribution in the range of the $a_0(980)$ mass and in the range of the $\pi^+ a_0$ mass distribution, we predict a branching ratio for the reaction of the order of 10^{-5}. Given the present rates of J/ψ production at BESIII, where branching ratios of 10^{-7} can be measured, we advocate for the measurement of this reaction, that apart from showing a new example of a TS can also shed light on the interpretation of the recent BESIII measurements of the $J/\psi \to \eta\pi^0\phi$ ($\phi \to K\bar{K}$) reaction.

ACKNOWLEDGMENTS

J.M. Dias would like to express gratitude to Guangxi Normal University for the warm hospitality, as part of this work was conducted there. This work is partly supported by the National Natural Science Foundation of China under Grant Nos. 11975083, 12365019 and 11975009, and by the Central Government Guidance Funds for Local Scientific and Technological Development, China (No. Guike ZY22096024). This work is also supported partly by the Natural Science Foundation of Changsha under Grant No. kq2208257 and the Natural Science Foundation of Hunan province under Grant No. 2023JJ30647 and the Natural Science Foundation of Guangxi province under Grant No. 2023JJA110076 (CWX). J.M. Dias acknowledges the support from the Chinese Academy of Sciences under Grant No. XDB34030000, and No. YSBR-101; by the National Key R&D program of Chinese under Grant No. 2023YFA1606703; by NSFC under Grant No. 12125507, No. 12361141819, and No. 12047503. This project has received funding from the European Union Horizon 2020 research and innovation programme under the program 2020-INFRAIA-2018-1, grant agreement No. 824093 of the STRONG-2020 project.

[1] R. Karplus, C. M. Sommerfield and E. H. Wichmann, Phys. Rev. **111**, 1187-1190 (1958)

[2] L. D. Landau, Nucl. Phys. **13**, no.1, 181-192 (1959)

[3] S. Coleman and R. E. Norton, Nuovo Cim. **38**, 438-442 (1965)

[4] C. Adolph *et al.* [COMPASS], Phys. Rev. Lett. **115**, no.8, 082001 (2015)

[5] X. H. Liu, M. Oka and Q. Zhao, Phys. Lett. B **753**, 297-302 (2016)

[6] M. Mikhasenko, B. Ketzer and A. Sarantsev, Phys. Rev. D **91**, no.9, 094015 (2015)

[7] F. Aceti, L. R. Dai and E. Oset, Phys. Rev. D **94**, no.9, 096015 (2016)

[8] G. D. Alexeev *et al.* [COMPASS], Phys. Rev. Lett. **127**, no.8, 082501 (2021)

[9] M. Ablikim *et al.* [BESIII], Phys. Rev. Lett. **108**, 182001 (2012)

[10] J. J. Wu, X. H. Liu, Q. Zhao and B. S. Zou, Phys. Rev. Lett. **108**, 081803 (2012)

[11] F. Aceti, W. H. Liang, E. Oset, J. J. Wu and B. S. Zou, Phys. Rev. D **86**, 114007 (2012)

[12] X. G. Wu, J. J. Wu, Q. Zhao and B. S. Zou, Phys. Rev. D **87**, no.1, 014023 (2013)

[13] C. Richard-Serre, W. Hirt, D. F. Measday, E. G. Michaelis, M. J. M. Saltmarsh and P. Skarek, Nucl. Phys. B **20**, 413-440 (1970)

[14] Said data base website. https://gwdac.phys.gwu.edu/

[15] N. Ikeno, R. Molina and E. Oset, Phys. Rev. C **104**, no.1, 014614 (2021)

[16] H. J. Jing, S. Sakai, F. K. Guo and B. S. Zou, Phys. Rev. D **100**, no.11, 114010 (2019)

[17] S. Sakai, E. Oset and W. H. Liang, Phys. Rev. D **96**, no.7, 074025 (2017)

[18] W. H. Liang, S. Sakai, J. J. Xie and E. Oset, Chin. Phys. C **42**, no.4, 044101 (2018)

[19] F. K. Guo, X. H. Liu and S. Sakai, Prog. Part. Nucl. Phys. **112**, 103757 (2020)

[20] M. Bayar, F. Aceti, F. K. Guo and E. Oset, Phys. Rev. D **94**, no.7, 074039 (2016)

[21] M. Ablikim *et al.* [BESIII], [arXiv:2311.07043 [hep-ex]].

[22] C. Schmid, "Final-State Interactions and the Simulation of Resonances," Phys. Rev. **154**, no.5, 1363 (1967)

[23] V. R. Debastiani, S. Sakai and E. Oset, Eur. Phys. J. C **79**, no.1, 69 (2019)

[24] J. A. Oller and E. Oset, Nucl. Phys. A **620**, 438-456 (1997) [erratum: Nucl. Phys. A **652**, 407-409 (1999)]

[25] N. Kaiser, Eur. Phys. J. A **3**, 307-309 (1998)

[26] V. E. Markushin, Eur. Phys. J. A **8**, 389-399 (2000)

[27] J. Nieves and E. Ruiz Arriola, Phys. Lett. B **455**, 30-38 (1999)

[28] R. L. Workman *et al.* [Particle Data Group], PTEP **2022**, 083C01 (2022)

[29] N. Ikeno, J. M. Dias, W. H. Liang and E. Oset, Phys. Rev. D **100**, no.11, 114011 (2019)

[30] S. Sakai, E. Oset and A. Ramos, Eur. Phys. J. A **54**, no.1, 10 (2018)

[31] D. Gamermann, J. Nieves, E. Oset and E. Ruiz Arriola, Phys. Rev. D **81**, 014029 (2010)

[32] J. X. Lin, J. T. Li, S. J. Jiang, W. H. Liang and E. Oset, Eur. Phys. J. C **81**, no.11, 1017 (2021)

[33] A. Bramon, A. Grau and G. Pancheri, Phys. Lett. B **283**, 416-420 (1992)

[34] R. Pavao, S. Sakai and E. Oset, Eur. Phys. J. C **77**, no.9, 599 (2017).

[35] E. Oset and L. Roca, Phys. Lett. B **782**, 332-338 (2018).

[36] L. R. Dai, Q. X. Yu and E. Oset, Phys. Rev. D **99**, no.1, 016021 (2019).

[37] W. H. Liang, H. X. Chen, E. Oset and E. Wang, Eur. Phys. J. C **79**, no.5, 411 (2019).

[38] F. K. Guo, X. H. Liu and S. Sakai, Prog. Part. Nucl. Phys. **112**, 103757 (2020).

[39] J. J. Xie, W. H. Liang, E. Oset, P. Moskal, M. Skurzok and C. Wilkin, Phys. Rev. C **95**, no.1, 015202 (2017).

[40] P. Rubin *et al.* [CLEO], Phys. Rev. Lett. **93**, 111801 (2004).

[41] M. Ablikim *et al.* [BESIII], Phys. Rev. D **95**, no.3, 032002 (2017).

[42] M. Ablikim *et al.* [BESIII], [arXiv:2309.05760 [hep-ex]].

Quasi-Classical Gluon Fields and Low's Soft Theorem at Small x

Ming Li
Department of Physics, The Ohio State University, Columbus, OH 43210, USA
(Dated: February 28, 2024)

In the high energy limit, soft gluons can be approximately described by quasi-classical gluon fields. It is well-known that the gluon field is a pure gauge field on the transverse plane at eikonal order. We derived the complete next-to-eikonal order solutions of the classical Yang-Mills equations for soft gluons in the dense nuclear regime. Utilizing these solutions, it is shown that Low's soft theorem at small x can be obtained by considering off-diagonal matrix elements of quasi-classical chromoelectric field between single gluon states in the dilute regime. Furthermore, we extend Low's soft theorem at small x to incorporate the effects of gluon saturation in the dense regime.

arXiv:2402.17568v1 [hep-ph] 27 Feb 2024

Introduction. Using quasi-classical gluon fields to characterize soft gluons particularly in small x physics has a long history [1]. In the high energy limit, nuclear objects are highly Lorentz contracted along the longitudinal direction so that a two-dimensional shockwave picture becomes applicable in describing high energy collisions [2]. In the McLerran-Venugopalan (MV) model [3, 4] , soft gluon fields from large nuclei are solved from the classical Yang-Mills equations with color current sourced by hard gluons. This eikonal order quasi-classical gluon fields, valid in the parametrically dense regime $\mathcal{O}(1/g)$ in which nonlinear QCD interactions cannot be ignored [5, 6], was used to estimate the Weiszacker-Williams gluon distribution inside a large nucleus in the quasi-classical approximation [7]. Phenomenological applications in relativistic heavy-ion collisions [8–15] and high energy proton-nucleus collisions [16–19] using the classical field approach were actively pursued in the past three decades [20] .

Quasi-classical gluon field at eikonal order is insensitive to spin information of the nuclear object. To study spin related physical quantities particularly to understand the spin structure of proton at small x [21, 22], subeikonal order gluon field is needed. There have been a lot of efforts in recent years to identify the effective interactions at subeikonal order [23–29] and to derive small x evolution equations for polarized parton distributions inside proton/nucleus [30–35]. Notably, the spin-dependent part of subeikonal order quasi-classical gluon fields was obtained using a diagrammatic approach in covariant gauge in [36].

In this paper, we derived the complete solutions at subeikonal order for quasi-classical gluon fields including both spin-dependent and spin-independent parts in the dense nuclear regime. The solutions are presented in the covariant gauge as well as in the light-cone gauge. We found that, in addition to the piece found in [36], there exists a novel spin-dependent term induced by gluon saturation, which becomes significant only in the dense regime. The spin-independent part bears resemblance to the next-to-leading-order magnetic multipole expansion in two dimensions, as inferred from Ampere's law. Regarding the external color currents originating from hard gluons, our findings indicate that, apart from the color charge density at subeikonal order, the color spin density and transverse color current also serve as sources for the quasi-classical gluon fields.

A closely related topic to soft gluons involves Low's soft theorem [37–41], which states that at amplitude level, the radiative amplitude can be expressed as a factorized product of a soft factor and the nonradiative amplitude when taking the soft limit. In the past decades, there have been reviving interests in understanding Low's soft theorem in the Standard Model of particle physics and gravity from the perspective of asymptotic symmetries [42–44]. Subleading order Low's soft theorem was rederived from various approaches [45–48]. Our focus lies on the small x limit of the subleading order Low's soft theorem. We demonstrate that the classical field approach effectively reproduces the small x limit of Low's soft theorem up to subleading order. Specifically, the matrix element of quasi-classical chromoelectric fields between incoming and outgoing single gluon states in the dilute limit coincides with Low's soft theorem at small x. The classical field approach offers more as it allows for solutions in the dense regime, enabling the generalization of Low's soft theorem to incorporate dense gluon effects, thereby linking Low's soft theorem to gluon saturation.

In the following, we first present the details of obtaining complete next-to-eikonal order solutions by solving classical Yang-Mills equations and then establish its equivalence to Low's soft theorem at small x.

Quasi-classical Gluon Fields at Subeikonal Order. We consider a pure glue theory and solve classical Yang-Mills equations. The inclusion of quarks and solving coupled Dirac equation for soft quark fields are left for a future work. Starting from the Yang-Mills action $S = -\frac{1}{4}\int_x F^a_{\mu\nu}F^{\mu\nu,a}$ with $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + ig[A_\mu, A_\nu]^a$, one separates the full gluon field into soft gluon field $\mathcal{A}_\mu$ and hard gluon field A_μ according to their longitudinal momenta

$$A^a_\mu \to \mathcal{A}^a_\mu + A^a_\mu. \tag{1}$$

Let Λ^+ be some longitudinal momentum scale. Soft gluon fields represent modes with $k^+ \ll \Lambda^+$ while hard gluon

fields are associated with $k^+ \gg \Lambda^+$. This division of degrees of freedom aligns with the spirit of Color Glass Condensate framework [49] and the idea of rapidity factorization [50]. By implementing this separation and eliminating terms that are linear in hard gluon fields as well as terms exclusively concerning hard gluons, the effective action for soft gluons becomes

$$\begin{aligned} S = &-\frac{1}{4}\int_x \Big(\mathcal{F}_a^{\mu\nu}\mathcal{F}^a_{\mu\nu} + 4ig\mathcal{A}^\mu_a\left(\partial^\nu[A_\mu, A_\nu]^a + [A_\nu, F^{\mu\nu}]^a\right) \\ &+ 2(ig)^2\Big([\mathcal{A}^\mu, \mathcal{A}^\nu]^a[A_\mu, A_\nu]^a + [\mathcal{A}_\mu, A_\nu]^a[\mathcal{A}^\mu, A^\nu]^a \\ &- [\mathcal{A}_\mu, A_\nu]^a[\mathcal{A}^\nu, A^\mu]^a)\Big). \end{aligned} \tag{2}$$

The equations of motion for soft gluons follow

$$\mathcal{D}_\nu \mathcal{F}^{\nu\mu} = J^\mu \tag{3}$$

with the hard gluon color current

$$J^\mu = ig[A_\nu, \bar{F}^{\mu\nu}] + ig\mathcal{D}^\nu[A^\mu, A_\nu]. \tag{4}$$

The covariant derivative is defined with respect to the soft gluon field $\mathcal{D}_\nu = \partial_\nu + ig[\mathcal{A}_\nu, \ ]$. The field strength tensor for hard gluon field is defined correspondingly $\bar{F}^{\mu\nu} = \mathcal{D}^\mu A^\nu - \mathcal{D}^\nu A^\mu + ig[A^\mu, A^\nu]$. Our goal is to solve eq. (3) up to subeikonal order.

The general rule of counting the power of eikonality of gluon fields comes from their transformations under Lorentz boost [28]

$$\begin{aligned} &\mathcal{A}^+ \longrightarrow \xi^{-1}\mathcal{A}^+(\xi x^+, \xi^{-1}x^-, \mathbf{x}), \\ &\mathcal{A}^- \longrightarrow \xi\mathcal{A}^-(\xi x^+, \xi^{-1}x^-, \mathbf{x}), \\ &\mathcal{A}^i \longrightarrow \mathcal{A}^i(\xi x^+, \xi^{-1}x^-, \mathbf{x}). \end{aligned} \tag{5}$$

We used light-cone coordinates $x^\pm = (t \pm z)/\sqrt{2}$. Here $\xi = e^{-\omega}$ with ω characterizing the amount of the Lorentz boost. In high energy limit, $\xi \to 0$. Expansion in eikonality is equivalent to Taylor expansion in powers of ξ [51].

$$\begin{aligned} &\mathcal{A}^+ = \xi^{-1}\mathcal{A}^+_{(-1)} + \mathcal{A}^+_{(0)} + \xi\mathcal{A}^+_{(1)} + \dots, \\ &\mathcal{A}^- = \xi\mathcal{A}^-_{(1)} + \dots, \\ &\mathcal{A}^i = \mathcal{A}^i_{(0)} + \xi\mathcal{A}^i_{(1)} + \dots. \end{aligned} \tag{6}$$

The arguments for the leading terms in the expansions are $(0^+, \tilde{x}^-, \mathbf{x}_\perp)$, in which we have redefined $\tilde{x}^- = \xi^{-1}x^-$. Similar eikonality expansions are understood for the color current J^μ and the field strength tensor $\mathcal{F}^{\mu\nu}$ when solving eq. (3) order by order in eikonality.

Solutions in Covariant Gauge. we use "tilde" to indicate quantities in covariant gauge $\partial_\mu\widetilde{\mathcal{A}}^\mu = 0$. From the expansions in eq. (6) , $\widetilde{\mathcal{A}}^- = \widetilde{\mathcal{A}}^i = 0$ at the eikonal order. One looks for static solution $\widetilde{A}^+_{(-1)}$ that satisfies

$$\partial_\perp^2 \widetilde{A}^+_{(-1)} = -\widetilde{J}^+_{(-1)}. \tag{7}$$

The solution is formally obtained as

$$\widetilde{A}^+_{(-1)} = -\frac{1}{\partial_\perp^2}J^+_{(-1)} = \int d^2\mathbf{y}G(\mathbf{x}-\mathbf{y})J^+_{(-1)}(x^-, \mathbf{y}) \tag{8}$$

with $G(\mathbf{x}-\mathbf{y}) = -\frac{1}{2\pi}\ln(|\mathbf{x}-\mathbf{y}|\Lambda)$ regularized by some IR scale Λ.

At subeikonal order, $\widetilde{\mathcal{A}}^- = 0$ and we look for solutions $\widetilde{\mathcal{A}}^+_{(0)} \neq 0$, $\widetilde{\mathcal{A}}^i_{(0)} \neq 0$ in the background of $\widetilde{\mathcal{A}}^+_{(-1)}$. The covariant gauge condition reduces to $\partial_+\widetilde{\mathcal{A}}^+_{(0)} = -\partial_i\widetilde{\mathcal{A}}^i_{(0)}$. It is apparent that at subeikonal order $\widetilde{\mathcal{A}}^+_{(0)}$ is dependent on the light-cone time x^+ while $\widetilde{\mathcal{A}}^i_{(0)}$ is a static field. The subeikonal order fields are considered to be $\widetilde{\mathcal{A}}^+_{(0)}, \widetilde{\mathcal{A}}^i_{(0)} \sim \mathcal{O}(1)$ parametrically as compared to the eikonal order field $\widetilde{\mathcal{A}}^+_{(-1)} \sim \mathcal{O}(1/g)$. This is the parametric regime relevant to applications [36]. For the more general regime in which $\widetilde{\mathcal{A}}^+_{(-1)}, \widetilde{\mathcal{A}}^+_{(0)}, \widetilde{\mathcal{A}}^i_{(0)} \sim \mathcal{O}(1/g)$, the equations become nonlinear and it is probably more suitable to solve the equations numerically.

With these considerations, the equations of motion at subeikonal order reduce to

$$\begin{aligned} &-\partial_\perp^2\widetilde{\mathcal{A}}^+_{(0)} - 2\partial_i\widetilde{\mathcal{D}}_-\widetilde{\mathcal{A}}^i_{(0)} = J^+_{(0)}, \\ &-\partial_\perp^2\widetilde{\mathcal{A}}^i_{(0)} = \widetilde{J}^i_{(0)}. \end{aligned} \tag{9}$$

The covariant derivative here is defined as $\widetilde{\mathcal{D}}_- = \partial_- + ig[\widetilde{\mathcal{A}}^+_{(-1)}, \ ]$. The commutator part is the same order as that of the ordinary derivative because of $\mathcal{A}^+_{(-1)} \sim \mathcal{O}(1/g)$. We also consider $\widetilde{J}^+_{(0)}$ to be of order $\mathcal{O}(1)$ and it is dependent on x^+. Current conservation at eikonal order becomes $\partial_+\widetilde{J}^+_{(0)} + \partial_i\widetilde{J}^i_{(0)} = 0$. The solutions are readily obtained

$$\begin{aligned} &\widetilde{\mathcal{A}}^+_{(0)} = -\frac{1}{\partial_\perp^2}\left(2\partial_i\widetilde{\mathcal{D}}_-\widetilde{\mathcal{A}}^i_{(0)} + \widetilde{J}^+_{(0)}\right), \\ &\widetilde{\mathcal{A}}^i_{(0)} = -\frac{1}{\partial_\perp^2}\widetilde{J}^i_{(0)}. \end{aligned} \tag{10}$$

In subsequent sections, it will become evident that the transverse current $\widetilde{J}^i_{(0)}$ contains a piece that is dependent on the helicity state of hard gluons. Therefore, in the dense regime, both the transverse field $\widetilde{\mathcal{A}}^i_{(0)}$ and $\widetilde{A}^+_{(0)}$, which contains the commutator $ig[\partial_i\widetilde{A}^+_{(-1)}, \widetilde{A}^i_{(0)}]$, exhibit sensitivity to the helicity state of hard gluons. This latter helicity dependence arises solely from dense gluon effects and vanishes in the dilute regime. It is convenient to have the field strength tensor computed.

$$\begin{aligned} &\widetilde{\mathcal{F}}^{i+} = \left(\delta^{ij} - \frac{2\partial_i\partial_j}{\partial_\perp^2}\right)\widetilde{\mathcal{D}}_-\frac{1}{\partial_\perp^2}\widetilde{J}^j + \frac{\partial_i}{\partial_\perp^2}\widetilde{J}^+, \\ &\widetilde{\mathcal{F}}^{ij} = -\frac{1}{\partial_\perp^2}(\partial^i\widetilde{J}^j - \partial^j\widetilde{J}^i), \\ &\widetilde{\mathcal{F}}^{-+} = \frac{\partial_j}{\partial_\perp^2}\widetilde{J}^j, \qquad \widetilde{\mathcal{F}}^{-i} = 0. \end{aligned} \tag{11}$$

Solutions in Light-Cone Gauge. To obtain solutions in light-cone gauge $\mathcal{A}^+ = 0$, instead of directly solving the equations of motion, we make gauge transformations on the solutions from covariant gauge to light-cone gauge. We then check that the solutions thus obtained satisfy the equations of motion up to subeikonal order in light-cone gauge. At subeikonal order, the gauge transformation Wilson line becomes dependent on light-cone time x^+,

$$W \equiv \mathcal{P}\text{exp}\left\{-ig\int_{-\infty}^{x^-} dz^- \widetilde{\mathcal{A}}^+(x^+, z^-, \mathbf{x})\right\} \tag{12}$$

because $\widetilde{\mathcal{A}}^+ = \widetilde{\mathcal{A}}^+_{(-1)} + \widetilde{\mathcal{A}}^+_{(0)}$ in which $\tilde{A}^+_{(0)}$ depends on x^+. The usual eikonal Wilson line is denoted by

$$U \equiv \mathcal{P}\text{exp}\left\{-ig\int_{-\infty}^{x^-} dz^- \widetilde{\mathcal{A}}^+_{(-1)}(x^+, z^-, \mathbf{x})\right\}. \tag{13}$$

From the eikonality expansions in eq.(6), the nonvanishing fields at subeikonal order in light-cone gauge are $\mathcal{A}^i_{(0)}, \mathcal{A}^i_{(1)}$ and $\mathcal{A}^-_{(1)}$. Making gauge transformation using eq. (12), one gets the solutions in light-cone gauge,

$$\begin{aligned}\partial_-\mathcal{A}^i =& \frac{\mathcal{D}^i}{\mathcal{D}^2_\perp}(J^+_{(-1)} + J^+_{(0)}) - \left(\delta^{ij} - \frac{2\mathcal{D}^i\mathcal{D}^j}{\mathcal{D}^2_\perp}\right)\partial_-\frac{1}{\mathcal{D}^2_\perp}J^j_{(0)} \\ & - ig\left[\mathcal{F}^{i+}_{(-1)}, \frac{1}{\partial_-}\left(\frac{1}{\mathcal{D}^2_\perp}J^+_{(0)} - \frac{2\mathcal{D}_j}{\mathcal{D}^2_\perp}\partial_-\frac{1}{\mathcal{D}^2_\perp}J^j_{(0)}\right)\right]\end{aligned} \tag{14}$$

and

$$\partial_-\mathcal{A}^- = -\frac{\mathcal{D}_j}{\mathcal{D}^2_\perp}J^j_{(0)}. \tag{15}$$

The quantities without "tilde" are in light-cone gauge. They are related to their counterparts in covariant gauge by gauge transformations $J^i = U^\dagger\widetilde{J}^i U$, $J^+ = U^\dagger\widetilde{J}^+ U$. The chromoelectric field at eikonal order is formally expressed as

$$\mathcal{F}^{i+}_{(-1)} = U^\dagger\partial^i\widetilde{A}^+_{(-1)}U = -\mathcal{D}^i\frac{1}{\mathcal{D}^2_\perp}J^+_{(-1)}. \tag{16}$$

The transverse covariant derivative is defined as $\mathcal{D}^i = \partial^i + ig[\mathcal{A}^i_{(0)}, \,]$. It is straightforward to verify that the solutions given in eqs. (14) and (15) satisfy the Yang-Mills equations in light-cone gauge up to subeikonal order. We won't delve into the calculational details.

Low's Soft Theorem at Small x. In Low's subleading soft theorem, radiative amplitude can be expressed as soft factors acting on non-radiative amplitude when the radiated gluon becomes soft ($k \ll p_i$)

$$M^a(\{p_i\}; k) = \left[\mathcal{S}^{(-1)} + \mathcal{S}^{(0)}\right] gT^a_{(i)}M(\{p_i\}) \tag{17}$$

The leading soft factor $\mathcal{S}^{(-1)}$ and the subleading soft factor $\mathcal{S}^{(0)}$ have the following expressions [45–48]

$$\mathcal{S}^{(-1)} = \sum_i^n \frac{p_i\cdot\varepsilon^*_\lambda(k)}{p_i\cdot k}, \;\; \mathcal{S}^{(0)} = i\sum_i^n \frac{\varepsilon^{\mu*}_\lambda(k)k^\nu J_{i,\mu\nu}}{p_i\cdot k}. \tag{18}$$

It is considered that there are n hard gluons with momenta p_i. $T^a_{(i)}$ is the color matrix associated with gluon "i" and $\varepsilon^\mu_\lambda(k)$ is the polarization vector of the soft gluon. $J_{\mu\nu} = L_{\mu\nu} + \Sigma_{\mu\nu}$ is the total angular momentum of the hard gluons. The orbital angular momentum $L_{\mu\nu} = i\left(p_\mu\frac{\partial}{\partial p^\nu} - p_\nu\frac{\partial}{\partial p^\mu}\right)$ and the spin operator for gluons is $(\Sigma^{\mu\nu})^{\alpha\beta} = i\left(g^{\mu\alpha}g^{\nu\beta} - g^{\mu\beta}g^{\nu\alpha}\right)$. Unlike the case for photon, the gluon soft theorem receives loop corrections [52, 53]. We restrict the discussion to tree level in this paper.

Low's soft theorem up to subleading order was derived by assuming $k^+, \mathbf{k} \ll p^+, \mathbf{p}$. On the other hand, in the small x limit, one only requires that longitudinal momentum to be much smaller $k^+ \ll p^+$ while the transverse momentum is of the same order $\mathbf{k} \sim \mathbf{p}$. We expand the soft factors in Low's soft theorem to linear order in power series expansion of $z = k^+/p^+$.

The leading order soft factor becomes

$$\mathcal{S}^{(-1)} = -2\delta_{\sigma\sigma'}\left[\frac{\mathbf{k}^i}{\mathbf{k}^2} - z\left(\delta^{ij} - \frac{2\mathbf{k}^i\mathbf{k}^j}{\mathbf{k}^2}\right)\frac{\mathbf{p}^j}{\mathbf{k}^2}\right]\varepsilon^{i*}_\lambda + \mathcal{O}(z^2). \tag{19}$$

For the subleading order soft factor, the part involving spin angular momentum becomes

$$\begin{aligned}&\frac{i\varepsilon^*_{\lambda,\mu}(k)k_\nu(\Sigma^{\mu\nu})^{\alpha\beta}}{p\cdot k}\varepsilon^\alpha_\sigma(p)\varepsilon^{\beta,*}_{\sigma'}(p) \\ =&2z\sigma\delta_{\sigma\sigma'}\frac{i\epsilon^{ij}\varepsilon^{i*}_\lambda\mathbf{k}^j}{\mathbf{k}^2} + \mathcal{O}(z^2).\end{aligned} \tag{20}$$

For the part involving orbital angular momentum, only derivatives with respect to the longitudinal momentum $\partial/\partial p^+$ contribute. The nonvanishing components of the orbital angular momentum operator are $L_{-+} = ip^+\frac{\partial}{\partial p^+}$ and $L_{i+} = ip_i\frac{\partial}{\partial p^+}$. One gets

$$\frac{i\varepsilon^{\mu*}_\lambda(k)k^\nu L_{\mu\nu}}{p\cdot k} = 2\delta_{\sigma\sigma'}z\frac{\mathbf{k}^i\varepsilon^{i*}_\lambda}{\mathbf{k}^2}\left(p^+\frac{\partial}{\partial p^+}\right) + \mathcal{O}(z^2). \tag{21}$$

It is interesting to observe that in the small x limit the only components of angular momentum that contribute are $S^z = \frac{1}{2}\epsilon^{ij}\Sigma^{ij}$ for spin and L^{+-} for orbital angular momentum.

Putting all the pieces together, the soft factor up to subleading order in the small x limit is

$$\begin{aligned}\mathcal{S} =& 2\delta_{\sigma\sigma'}\Big[-\frac{\mathbf{k}^i}{\mathbf{k}^2} + z\left(\delta^{ij} - \frac{2\mathbf{k}^i\mathbf{k}^j}{\mathbf{k}^2}\right)\frac{\mathbf{p}^j}{\mathbf{k}^2} \\ &+ z\sigma\frac{i\epsilon^{ij}\mathbf{k}^j}{\mathbf{k}^2} + z\frac{\mathbf{k}^i}{\mathbf{k}^2}\left(p^+\frac{\partial}{\partial p^+}\right)\Big]\varepsilon^{i*}_\lambda + \mathcal{O}(z^2).\end{aligned} \tag{22}$$

In the Dilute Regime. We show that eq. (22) can be obtained from the quais-classical gluon field in eq. (14) by calculating its matrix element between incoming and outgoing single gluon states in the dilute regime. In the dilute regime, $\mathcal{A}^i_{(0)} \sim \mathcal{O}(1)$ and one replaces the covariant derivative $\mathcal{D}_i$ with ordinary derivative ∂_i and ignores the nonlinear commutator part in eq. (14)

$$\mathcal{F}^{+i} = \partial_- \mathcal{A}^i = \frac{\partial^i}{\partial_\perp^2} J^+ - \left(\delta^{ij} - \frac{2\partial^i\partial^j}{\partial_\perp^2}\right)\frac{\partial_-}{\partial_\perp^2} J^j. \quad (23)$$

We use mode expansions for hard gluon fields to express the color current in terms of gluon creation and annihilation operators. For $J^+_a(k^+, \mathbf{x}) = \rho^a(\mathbf{x}) + J^+_{(0)}(k^+, \mathbf{x})$

$$\rho^a(\mathbf{x}) = igf^{abc}\int_{q^+} \hat{a}_{b,\lambda}(q^+, \mathbf{x})\hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x}). \quad (24)$$

and

$$J^+_{(0)}(k^+, \mathbf{x}) = -igf^{abc}\frac{k^+}{2}\int_{q^+}\Big[\hat{a}_{b,\lambda}(q^+, \mathbf{x})\frac{\partial}{\partial q^+}\hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x}) - \frac{\partial}{\partial q^+}\hat{a}_{b,\lambda}(q^+, \mathbf{x})\hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x})\Big]. \quad (25)$$

The total transverse current can be decomposed into $J^i_a(\mathbf{x}) = -\epsilon^{il}\partial^l j_a(\mathbf{x}) + j^i_a(\mathbf{x})$ with the color spin density

$$j^a(\mathbf{x}) = igf^{abc}\int_{q^+}\frac{1}{q^+}\lambda\hat{a}_{b,\lambda}(q^+, \mathbf{x})\hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x}) \quad (26)$$

and the spin-independent transverse color current

$$j^l_a(\mathbf{x}) = gf^{abc}\int_{q^+}\frac{1}{2q^+}\Big[\partial^l\hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x})\hat{a}_{b,\lambda}(q^+, \mathbf{x}) - \hat{a}^\dagger_{c,\lambda}(q^+, \mathbf{x})\partial^l\hat{a}_{b,\lambda}(q^+, \mathbf{x})\Big]. \quad (27)$$

Unlike the case at eikonal order in which the effective degrees of freedom for hard gluons are completely determined by the color charge density $\rho^a(\mathbf{x})$, one needs the color spin density $j^a(\mathbf{x})$, the transverse color current $j^i_a(\mathbf{x})$ and the subeikonal order correction $J^+_{(0)}(k^+, \mathbf{x})$ at subeikonal order.

Using these expressions, one calculates the matrix element of chromoelectric fields between incoming and outgoing single gluon Fock states

$$\begin{aligned}&\Big\langle 0\Big|\hat{a}_{d',\sigma'}(p'^+, \mathbf{p}')\mathcal{F}^{+i}_a(k^+, \mathbf{k})\hat{a}^\dagger_{d,\sigma}(p^+, \mathbf{p})\Big|0\Big\rangle\\ =&gf^{add'}\delta_{\sigma\sigma'}2\Big[-\frac{\mathbf{k}^i}{\mathbf{k}^2} + z\frac{\mathbf{k}^i}{\mathbf{k}^2}\left(p^+\frac{\partial}{\partial p^+}\right)\\ &+\sigma z\frac{i\epsilon^{il}\mathbf{k}^l}{\mathbf{k}^2} + z\left(\delta^{ij} - \frac{2\mathbf{k}^i\mathbf{k}^j}{\mathbf{k}^2}\right)\frac{\mathbf{p}^j}{\mathbf{k}^2}\Big]\\ &\times(2\pi)2p^+\delta(p^+ - p'^+)(2\pi)^2\delta^{(2)}(\mathbf{p}' + \mathbf{k} - \mathbf{p}).\end{aligned} \quad (28)$$

Eq. (28) reproduces the soft factor at small x given in eq. (22). The overall factor 2 accounts for the inclusion of both gluon emission and absorption in the classical fields, ensuring consistency with Low's soft theorem. This alignment with Low's soft theorem at small x is anticipated to extend beyond the next-to-eikonal order. In fact, by employing the systematic expansion in eikonality, one could feasibly derive solutions for quasi-classical gluon fields up to the next-to-next-to-eikonal order and consequently provide a prediction for Low's soft theorem at next-to-next-to-eikonal order.

In the dense regime, the chromoelectric field is given by eq. (14) rather than eq. (23). Low's soft factor at small x is then related to the off-diagonal matrix element of incoming and outgoing states of dense nuclear objects such as large nuclei or protons at high energies,

$$\mathcal{S}_{\rm Low}\big|_{\rm small\, x} \sim {}_H\langle P', s', a'|\mathcal{F}^{+i}|P, s, a\rangle_H \quad (29)$$

Here $|P, s, a\rangle_H$ represents the nuclear wavefunction for hard gluons with P, s, a labeling the momentum, spin and color. The chromoelectric field $\mathcal{F}^{+i}$ for soft gluons is sourced by color currents of hard gluons eq. (4), which contain both polarized and unpolarized local field bilinears as exampled in the dilute regime by eqs. (24)-(27). The expression on the right hand side of eq. (29) is non-perturbative and further analyses in the dense regime necessitate the modeling of off-diagonal elements of color currents beyond the conventional MV model, which only concerns diagonal matrix elements between color singlets.

Conclusions. In this paper, we employed a classical field approach to derive the small x limit of Low's soft theorem up to subleading order. By solving classical Yang-Mills equations with external current, we obtained the full subeikonal order quasi-classical gluon fields. Subsequently, we established the equivalence between off-diagonal matrix elements of quasi-classical chromoelectric gluon field eq. (28) and Low's soft theorem at small x eq. (22) in the dilute regime and made a generalization to include dense gluon effects. The subeikonal order solutions can be readily used to calculate various unpolarized and polarized gluon distributions inside proton/nucleus in the quasi-classical approximation up to subeikonal order. For example, the Weizsacker-Williams gluon helicity distribution $\epsilon^{ij}\langle P, s|\mathcal{F}^{+i}\mathcal{F}^{+j}|P, s\rangle$ can be computed within the helicity generalized MV model [32]. Additionally, these solutions can aid in formulating the complete helicity generalized JIMWLK evolution equation [54]. Given the recent advancements on Low's soft theorem relating to asymptotic symmetries and color memory effects [44, 55–57], investigations into the memory effects of spin and transverse displacement at subeikonal order, particularly the role of gluon saturation, using quasi-classical gluon fields are warranted.

Acknowledgments. I thank Alex Kovner and Yuri Kovchegov for reading the manuscript and valuable comments. I am also grateful to Yuri Kovchegov for very helpful and illuminating discussions. The work is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Award Number DE-SC0004286.

[1] Y. V. Kovchegov and E. Levin, *Quantum chromodynamics at high energy*, Vol. 33 (Cambridge University Press, 2012).
[2] H. L. Verlinde and E. P. Verlinde, QCD at high-energies and two-dimensional field theory, (1993), arXiv:hep-th/9302104.
[3] L. D. McLerran and R. Venugopalan, Computing quark and gluon distribution functions for very large nuclei, Phys. Rev. D **49**, 2233 (1994), arXiv:hep-ph/9309289.
[4] L. D. McLerran and R. Venugopalan, Gluon distribution functions for very large nuclei at small transverse momentum, Phys. Rev. D **49**, 3352 (1994), arXiv:hep-ph/9311205.
[5] Y. V. Kovchegov, NonAbelian Weizsacker-Williams field and a two-dimensional effective color charge density for a very large nucleus, Phys. Rev. D **54**, 5463 (1996), arXiv:hep-ph/9605446.
[6] Y. V. Kovchegov, Quantum structure of the nonAbelian Weizsacker-Williams field for a very large nucleus, Phys. Rev. D **55**, 5445 (1997), arXiv:hep-ph/9701229.
[7] J. Jalilian-Marian, A. Kovner, L. D. McLerran, and H. Weigert, The Intrinsic glue distribution at very small x, Phys. Rev. D **55**, 5414 (1997), arXiv:hep-ph/9606337.
[8] A. Kovner, L. D. McLerran, and H. Weigert, Gluon production from nonAbelian Weizsacker-Williams fields in nucleus-nucleus collisions, Phys. Rev. D **52**, 6231 (1995), arXiv:hep-ph/9502289.
[9] A. Kovner, L. D. McLerran, and H. Weigert, Gluon production at high transverse momentum in the McLerran-Venugopalan model of nuclear structure functions, Phys. Rev. D **52**, 3809 (1995), arXiv:hep-ph/9505320.
[10] A. Krasnitz and R. Venugopalan, The Initial energy density of gluons produced in very high-energy nuclear collisions, Phys. Rev. Lett. **84**, 4309 (2000), arXiv:hep-ph/9909203.
[11] A. Krasnitz and R. Venugopalan, The Initial gluon multiplicity in heavy ion collisions, Phys. Rev. Lett. **86**, 1717 (2001), arXiv:hep-ph/0007108.
[12] I. Balitsky, Scattering of shock waves in QCD, Phys. Rev. D **70**, 114030 (2004), arXiv:hep-ph/0409314.
[13] B. Schenke, P. Tribedy, and R. Venugopalan, Fluctuating Glasma initial conditions and flow in heavy ion collisions, Phys. Rev. Lett. **108**, 252301 (2012), arXiv:1202.6646 [nucl-th].
[14] C. Gale, S. Jeon, B. Schenke, P. Tribedy, and R. Venugopalan, Event-by-event anisotropic flow in heavy-ion collisions from combined Yang-Mills and viscous fluid dynamics, Phys. Rev. Lett. **110**, 012302 (2013), arXiv:1209.6330 [nucl-th].
[15] G. Chen, R. J. Fries, J. I. Kapusta, and Y. Li, Early Time Dynamics of Gluon Fields in High Energy Nuclear Collisions, Phys. Rev. C **92**, 064912 (2015), arXiv:1507.03524 [nucl-th].
[16] B. Schenke, S. Schlichting, and R. Venugopalan, Azimuthal anisotropies in p+Pb collisions from classical Yang–Mills dynamics, Phys. Lett. B **747**, 76 (2015), arXiv:1502.01331 [hep-ph].
[17] M. Li and V. V. Skokov, First saturation correction in high energy proton-nucleus collisions. Part I. Time evolution of classical Yang-Mills fields beyond leading order, JHEP **06**, 140, arXiv:2102.01594 [hep-ph].
[18] M. Li and V. V. Skokov, First saturation correction in high energy proton-nucleus collisions. Part II. Single inclusive semi-hard gluon production, JHEP **06**, 141, arXiv:2104.01879 [hep-ph].
[19] M. Li and V. V. Skokov, First saturation correction in high energy proton-nucleus collisions. Part III. Ensemble averaging, JHEP **01**, 160, arXiv:2111.05304 [hep-ph].
[20] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, The Color Glass Condensate, Ann. Rev. Nucl. Part. Sci. **60**, 463 (2010), arXiv:1002.0333 [hep-ph].
[21] D. Adamiak, Y. V. Kovchegov, W. Melnitchouk, D. Pitonyak, N. Sato, and M. D. Sievert (Jefferson Lab Angular Momentum), First analysis of world polarized DIS data with small-x helicity evolution, Phys. Rev. D **104**, L031501 (2021), arXiv:2102.06159 [hep-ph].
[22] D. Adamiak, N. Baldonado, Y. V. Kovchegov, W. Melnitchouk, D. Pitonyak, N. Sato, M. D. Sievert, A. Tarasov, and Y. Tawabutr (Jefferson Lab Angular Momentum (JAM)), Global analysis of polarized DIS and SIDIS data with improved small-x helicity evolution, Phys. Rev. D **108**, 114007 (2023), arXiv:2308.07461 [hep-ph].
[23] E. Laenen, G. Stavenga, and C. D. White, Path integral approach to eikonal and next-to-eikonal exponentiation, JHEP **03**, 054, arXiv:0811.2067 [hep-ph].
[24] J. Jalilian-Marian, Elastic scattering of a quark from a color field: longitudinal momentum exchange, Phys. Rev. D **96**, 074020 (2017), arXiv:1708.07533 [hep-ph].
[25] G. A. Chirilli, Sub-eikonal corrections to scattering amplitudes at high energy, JHEP **01**, 118, arXiv:1807.11435 [hep-ph].
[26] Y. V. Kovchegov and M. G. Santiago, Quark sivers function at small x: spin-dependent odderon and the sub-eikonal evolution, JHEP **11**, 200, [Erratum: JHEP 09, 186 (2022)], arXiv:2108.03667 [hep-ph].
[27] T. Altinoluk and G. Beuf, Quark and scalar propagators at next-to-eikonal accuracy in the CGC through a dynamical background gluon field, Phys. Rev. D **105**, 074026 (2022), arXiv:2109.01620 [hep-ph].
[28] M. Li, Small x physics beyond eikonal approximation: an effective Hamiltonian approach, JHEP **07**, 158, arXiv:2304.12842 [hep-ph].
[29] P. Agostini, Scalar propagator in a background gluon field beyond the eikonal approximation, JHEP **11**, 099, arXiv:2307.13573 [hep-ph].
[30] Y. V. Kovchegov, D. Pitonyak, and M. D. Sievert, Helicity Evolution at Small-x, JHEP **01**, 072, [Erratum: JHEP 10, 148 (2016)], arXiv:1511.06737 [hep-ph].
[31] Y. V. Kovchegov, D. Pitonyak, and M. D. Siev-

ert, Small-x asymptotics of the quark helicity distribution, Phys. Rev. Lett. **118**, 052001 (2017), arXiv:1610.06188 [hep-ph].
[32] Y. V. Kovchegov, D. Pitonyak, and M. D. Sievert, Small-x Asymptotics of the Gluon Helicity Distribution, JHEP **10**, 198, arXiv:1706.04236 [nucl-th].
[33] Y. V. Kovchegov and M. D. Sievert, Small-x Helicity Evolution: an Operator Treatment, Phys. Rev. D **99**, 054032 (2019), arXiv:1808.09010 [hep-ph].
[34] G. A. Chirilli, High-energy operator product expansion at sub-eikonal level, JHEP **06**, 096, arXiv:2101.12744 [hep-ph].
[35] F. Cougoulic, Y. V. Kovchegov, A. Tarasov, and Y. Tawabutr, Quark and gluon helicity evolution at small x: revised and updated, JHEP **07**, 095, arXiv:2204.11898 [hep-ph].
[36] F. Cougoulic and Y. V. Kovchegov, Helicity-dependent extension of the McLerran–Venugopalan model, Nucl. Phys. A **1004**, 122051 (2020), arXiv:2005.14688 [hep-ph].
[37] F. E. Low, Scattering of light of very low frequency by systems of spin 1/2, Phys. Rev. **96**, 1428 (1954).
[38] M. Gell-Mann and M. L. Goldberger, Scattering of low-energy photons by particles of spin 1/2, Phys. Rev. **96**, 1433 (1954).
[39] F. E. Low, Bremsstrahlung of very low-energy quanta in elementary particle collisions, Phys. Rev. **110**, 974 (1958).
[40] S. Weinberg, Infrared photons and gravitons, Phys. Rev. **140**, B516 (1965).
[41] T. H. Burnett and N. M. Kroll, Extension of the low soft photon theorem, Phys. Rev. Lett. **20**, 86 (1968).
[42] A. Strominger, Asymptotic Symmetries of Yang-Mills Theory, JHEP **07**, 151, arXiv:1308.0589 [hep-th].
[43] V. Lysov, S. Pasterski, and A. Strominger, Low's Subleading Soft Theorem as a Symmetry of QED, Phys. Rev. Lett. **113**, 111601 (2014), arXiv:1407.3814 [hep-th].
[44] A. Strominger, *Lectures on the Infrared Structure of Gravity and Gauge Theory* (2017) arXiv:1703.05448 [hep-th].
[45] J. Broedel, M. de Leeuw, J. Plefka, and M. Rosso, Constraining subleading soft gluon and graviton theorems, Phys. Rev. D **90**, 065024 (2014), arXiv:1406.6574 [hep-th].
[46] Z. Bern, S. Davies, P. Di Vecchia, and J. Nohle, Low-Energy Behavior of Gluons and Gravitons from Gauge Invariance, Phys. Rev. D **90**, 084035 (2014), arXiv:1406.6987 [hep-th].
[47] C. D. White, Diagrammatic insights into next-to-soft corrections, Phys. Lett. B **737**, 216 (2014), arXiv:1406.7184 [hep-th].
[48] A. J. Larkoski, D. Neill, and I. W. Stewart, Soft Theorems from Effective Field Theory, JHEP **06**, 077, arXiv:1412.3108 [hep-th].
[49] E. Iancu, A. Leonidov, and L. McLerran, The Color glass condensate: An Introduction, in *Cargese Summer School on QCD Perspectives on Hot and Dense Matter* (2002) pp. 73–145, arXiv:hep-ph/0202270.
[50] I. Balitsky, Operator expansion for high-energy scattering, Nucl. Phys. B **463**, 99 (1996), arXiv:hep-ph/9509348.
[51] J. D. Bjorken, J. B. Kogut, and D. E. Soper, Quantum Electrodynamics at Infinite Momentum: Scattering from an External Field, Phys. Rev. D **3**, 1382 (1971).
[52] Z. Bern, S. Davies, and J. Nohle, On Loop Corrections to Subleading Soft Behavior of Gluons and Gravitons, Phys. Rev. D **90**, 085015 (2014), arXiv:1405.1015 [hep-th].
[53] S. He, Y.-t. Huang, and C. Wen, Loop Corrections to Soft Theorems in Gauge Theories and Gravity, JHEP **12**, 115, arXiv:1405.1410 [hep-th].
[54] F. Cougoulic and Y. V. Kovchegov, Helicity-dependent generalization of the JIMWLK evolution, Phys. Rev. D **100**, 114020 (2019), arXiv:1910.04268 [hep-ph].
[55] M. Pate, A.-M. Raclariu, and A. Strominger, Color Memory: A Yang-Mills Analog of Gravitational Wave Memory, Phys. Rev. Lett. **119**, 261602 (2017), arXiv:1707.08016 [hep-th].
[56] A. Ball, M. Pate, A.-M. Raclariu, A. Strominger, and R. Venugopalan, Measuring color memory in a color glass condensate at electron–ion colliders, Annals Phys. **407**, 15 (2019), arXiv:1805.12224 [hep-ph].
[57] N. Jokela, K. Kajantie, and M. Sarkkinen, Memory effect in Yang-Mills theory, Phys. Rev. D **99**, 116003 (2019), arXiv:1903.10231 [hep-th].

T-odd gluon distribution functions in a spectator model

Alessandro Bacchetta,[1,2,*] Francesco Giovanni Celiberto,[3,†] and Marco Radici[2,‡]

[1] *Dipartimento di Fisica, Università di Pavia, via Bassi 6, I-27100 Pavia*
[2] *INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy*
[3] *Universidad de Alcalá (UAH), Departamento de Física y Matemáticas, E-28805 Alcalá de Henares, Madrid, Spain*

(Dated: Wednesday 28th February, 2024, 03:00)

We present a model calculation of T-odd transverse-momentum-dependent distributions of gluons in the nucleon. The model is based on the assumption that a nucleon can emit a gluon, and what remains after the emission is treated as a single spectator particle. This spectator particle is considered to be on-shell, but its mass is allowed to take a continuous range of values, described by a spectral function. The final-state interaction that is necessary to generate T-odd functions is modeled as the exchange of a single gluon between the spectator and the outgoing parton.

PACS numbers: 12.38.-t, 12.40.-y, 14.70.Dj

arXiv:2402.17556v1 [hep-ph] 27 Feb 2024

CONTENTS

I. Introduction 2

II. The spectator model 3
A. Tree-level correlator 3
B. Additional single-gluon exchange 4
C. Spectator-gluon-spectator vertex 6
D. Gluon TMD projectors 7

III. T-odd gluon TMDs: Illustrative examples 7
A. Sivers function: g_1-vertex approximation 7
B. Sivers function: Quark-target model 9
C. Linearity function: g_1-vertex approximation 11
D. Linearity function: Quark-target model 11

IV. T-odd gluon TMDs: Results of full calculation 11

V. Summary and Outlook 17

Acknowledgments 18

A. Master integrals 18

B. Full calculation of gluon TMDs: The coefficients $C_{ijk}^{[F],l}$ 19
1. Sivers function $f_{1T}^{\perp}$ 19
2. Linearity function h_1 22
3. Propeller function $h_{1L}^{\perp}$ 25
4. Butterfly function $h_{1T}^{\perp}$ 27

References 30

* alessandro.bacchetta@unipv.it
† francesco.celiberto@uah.es
‡ marco.radici@pv.infn.it

I. INTRODUCTION

The multi-dimensional distribution of partons within a nucleon can be parametrized in terms of several sets of functions, encoding different correlations between the momentum and spin of the parton and its parent nucleon. In particular, the 3-dimensional distribution in momentum space is encoded in the so-called Transverse-Momentum-Dependent Parton Distributions (TMD PDFs or TMDs) [1]. In simple terms, TMDs extend the conventional 1-dimensional collinear Parton Distribution Functions (PDFs) into three dimensions, including also the dependence on the partonic transverse momentum.

The endeavor to constrain TMDs is a crucial step toward unraveling the multi-dimensional structure of the nucleon, and gaining deeper insight into Quantum ChromoDynamics (QCD) and color confinement. The field of TMDs has witnessed remarkable advancement in recent years, predominantly in the quark sector. Progress within the gluon sector has been relatively restrained, owing to the challenges associated with probing gluons in high-energy processes.

Gluon TMDs at leading twist, first analyzed and classified in Ref. [2], are shown in Tab. I in terms of both the polarization of the gluon and of its parent hadron. In this paper, our focus centers on (naïve) time-reversal odd (T-odd) gluon TMDs, highlighted in red in Tab. I. A notable example of a T-odd TMD is the gluon Sivers function, denoted as $f_{1T}^{\perp g}$. This function describes the distribution of unpolarized gluons in a transversely polarized nucleon and has a crucial role in the description of transverse single-spin asymmetries (see [3] and references therein). As in the case for quark TMDs, T-odd gluon TMDs are generated by the presence of initial and/or final state QCD interactions between incoming or outgoing partons and the target fragments. These interactions also underlie the peculiar process-dependence of gluon TMDs.

gluon polarization

nucleon polarization	U	circular	linear
U	f_1^g		$h_1^{\perp g}$
L		g_1^g	$h_{1L}^{\perp g}$
T	$f_{1T}^{\perp g}$	g_{1T}^g	h_1^g, $h_{1T}^{\perp g}$

Table I. Gluon TMD PDFs at twist-2. We adopt here the notation suggested in Ref. [4], similar to the quark case. U, L, T depict unpolarized, longitudinally polarized and transversely polarized nucleons. U, 'circular', 'linear' describe unpolarized, circularly polarized and linearly polarized gluons. Functions in blue are T-even. Functions in black are T-even and survive transverse-momentum integration. Functions in red are T-odd.

Experimental information on gluon TMDs is very scarce, and particularly so for T-odd ones. Ref. [5] presented the first attempt to reconstruct the unpolarized gluon TMD, f_1^g. Phenomenological studies of the T-odd gluon Sivers function were published in Refs. [6–8], but in processes where TMD factorization is not guaranteed to be applicable. An experimental measurement related to the gluon Sivers function was published by the COMPASS collaboration [9]. Several ways to experimentally access the gluon Sivers function have been discussed in the literature [10–16] and are among the primary goals of new experimental facilities [17–20].

Pioneering calculations of gluon TMD distributions [2, 21, 22] were performed using the *spectator-model* approach (see also Refs. [23, 24] for more recent versions). Originally conceived for studies in the quark-TMD sector [4, 25–29], this approach rests on the assumption that the struck nucleon emits a parton, and the residual fragments are treated as a single spectator particle, considered to be on-shell. At variance with those studies, in Ref. [30] we presented the calculation of all T-even gluon TMDs in the spectator-model approach where the spectator mass is allowed to take a continuous range of values weighted by a flexible spectral function. This modification encapsulates the effect of $q\bar{q}$ contributions, and allows to effectively reproduce both the small- and the moderate-x behavior of the TMDs.

In this paper, we extend the results of Ref. [30] by providing a systematic calculation in the same spectator-model framework of the complete set of all the four T-odd gluon TMDs at leading twist, including their process dependence.

II. THE SPECTATOR MODEL

Our model is based on the assumption that a nucleon can emit a gluon, and what remains after the emission is treated as a single spectator fermionic particle (see Fig. 1). This spectator fermion is considered to be on-shell, but its mass is allowed to take a continuous range of values, described by a spectral function. The nucleon-gluon-spectator coupling is described by an effective vertex containing two form factors, inspired by the standard nucleon form factors. Such model can effectively reproduce the known collinear (un)polarized gluon PDFs (the diagonal black entries f_1^g and g_1^g in Tab. I, that survive integration upon transverse momenta) and can be used to compute all T-even TMDs [30].

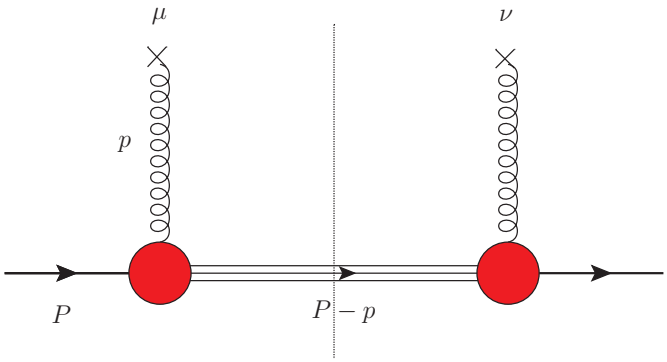


Figure 1. Tree-level cut diagram for the calculation of T-even leading-twist gluon TMDs. The triple line represents a spin-$\frac{1}{2}$ spectator. The red blob represents the nucleon-gluon-spectator vertex.

T-odd gluon TMDs vanish at tree level, because there is no residual interaction between the active parton and the spectator; equivalently, there is no interference between two competing channels producing the complex amplitude whose imaginary part gives the T-odd contribution. We can generate such structures by considering the interference between the tree-level scattering amplitude and the scattering amplitudes with an additional gluon exchange, as shown in Fig. 2. This corresponds to the one-gluon-exchange approximation of the gauge link operator. As we shall discuss in detail, the exact form of the gauge link depends on the process and in our case leads to two different types of functions.

A. Tree-level correlator

Following Ref. [30], we work in the frame where the nucleon momentum P has no transverse component:

$$P = \left[\frac{M^2}{2P^+},\, P^+,\, \mathbf{0}\right] , \tag{1}$$

where M is the nucleon mass. The parton momentum is parameterized as

$$p = \left[\frac{p^2 + \boldsymbol{p}_T^2}{2xP^+},\, xP^+,\, \boldsymbol{p}_T\right] , \tag{2}$$

where evidently $x = p^+/P^+$ is the light-cone (longitudinal) momentum fraction carried by the parton.

In the spectator-model framework one assumes that the nucleon with spin S in the state $|P,S\rangle$ can split into a gluon with momentum p and other remainders, effectively treated as a single spin-$\frac{1}{2}$ spectator particle with momentum $P-p$ and mass M_X. Similarly to Refs. [25, 30], we define a "tree-level" correlator as (see Fig. 1) [1]

$$\Phi^{\mu\nu(0)}(x,\boldsymbol{p}_T,S) = \frac{1}{(2\pi)^3\, 2\,(1-x)\,P^+}\,\mathrm{Tr}\left[(\not{P}+M)\,\frac{1+\gamma^5\not{S}}{2}\,G^{*\nu\sigma}(p,p)\mathcal{Y}_\sigma^{\dagger ab}\big(p^2\big)(\not{P}-\not{p}+M_X)\,\mathcal{Y}_\rho^{ba}(p^2)\,G^{\mu\rho}(p,p)\right], \tag{3}$$

[1] We remark that in Ref. [30] there is an error in the position of the $\mathcal{Y}$ vertices and a typo in the definition of the $G^{\mu\rho}$ propagator.

where a, b are color indices (in the adjoint representation) and

$$G^{\mu\rho}(p,q) = -\frac{i}{q^2 - m_g^2}\left(g^{\mu\rho} - \frac{q^\mu (n_-)^\rho}{p^+}\right) \tag{4}$$

is a specific Feynman rule for the gluon propagator in the definition of the correlator [31, 32], with n_-^ρ a light-like unit vector of the light-cone basis, and m_g a gluon mass regulator which will be set to zero in our calculations. We model the nucleon-gluon-spectator vertex as

$$\mathcal{Y}_\rho^{ba}(p^2) = \delta^{ba}\left[g_1(p^2)\,\gamma_\rho + g_2(p^2)\,\frac{i}{2M}\,\sigma_{\rho\nu}\,p^\nu\right], \tag{5}$$

where as usual $\sigma_{\rho\nu} = i[\gamma_\rho, \gamma_\nu]/2$, and $g_{1,2}(p^2)$ are generic form factors. In principle, the expression of $\mathcal{Y}_\rho^{ba}(p^2)$ could contain more Dirac structures. However, with our assumptions the spectator is identified with an on-shell spin-$\frac{1}{2}$ particle, much like the nucleon. Hence, we model the structure of $\mathcal{Y}_\rho^{ba}(p^2)$ similarly to the conserved electromagnetic current of a free nucleon obtained from the Gordon decomposition. The form factors $g_{1,2}(p^2)$ are formally similar to the Dirac and Pauli form factors, but obviously must not be identified with them. Similarly to our previous model description of quark TMDs [25], we use the dipolar expression

$$g_{1,2}(p^2) = \kappa_{1,2}\,\frac{p^2}{|p^2 - \Lambda_X^2|^2} = \kappa_{1,2}\,\frac{p^2\,(1-x)^2}{(\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2))^2}\ , \tag{6}$$

where $\kappa_{1,2}$ and Λ_X are normalization and cut-off parameters, respectively, and

$$L_X^2(\Lambda_X^2) = x\,M_X^2 + (1-x)\,\Lambda_X^2 - x\,(1-x)\,M^2\,. \tag{7}$$

The dipolar expression of Eq. (6) has several advantages: it cancels the singularity of the gluon propagator, it smoothly suppresses the effect of high $\boldsymbol{p}_T^2$ where the TMD formalism cannot be applied, and it compensates also the logarithmic divergences arising after integration upon $\boldsymbol{p}_T$.

In our model, the overall color prefactor at tree level is

$$C^{(0)} = \delta^{ab}\delta^{ba} = 8. \tag{8}$$

As a comparison, we will also discuss the quark-target model, which can be obtained from Eq. (3) simply by replacing

$$\mathcal{Y}_\rho^{ba} \to g_s \gamma_\rho t^a\ , \tag{9}$$

with g_s the strong coupling constant and t^a a generator of color SU(3) transformations, and by setting $M = M_X \equiv m_q$ everywhere. In this case, the overall color factor is

$$C_q^{(0)} = \frac{1}{N_C}\,\mathrm{Tr}_C[t^a t^a] = \frac{4}{3}, \tag{10}$$

where N_C is the number of colors and Tr_C indicates the trace upon color indices.

B. Additional single-gluon exchange

In general, T-odd TMDs arise only when there is a residual interaction between the active parton and the spectator. More specifically, they arise from the imaginary part of the interference between the tree-level channel and the channel describing this residual interaction. Following our model calculation for quark TMDs [25], we generate this interference by describing the residual gluon-spectator interaction through the exchange of a soft gluon (see Fig. 2). This one-gluon exchange results from the truncation at the first order in the expansion of the path-ordered exponential that defines the gauge link as the sum of infinite gluon rescatterings [33].

In the general definition of the parton-parton correlator Φ, the gauge link is a necessary ingredient to make the correlator color-gauge invariant. However, the sensitivity of TMDs to the transverse components of the gauge link introduces a process dependence, contrary to the case of collinear PDFs. While T-even quark TMDs are independent from the direction of the color flow in the involved hard scattering, T-odd quark TMDs change sign when moving

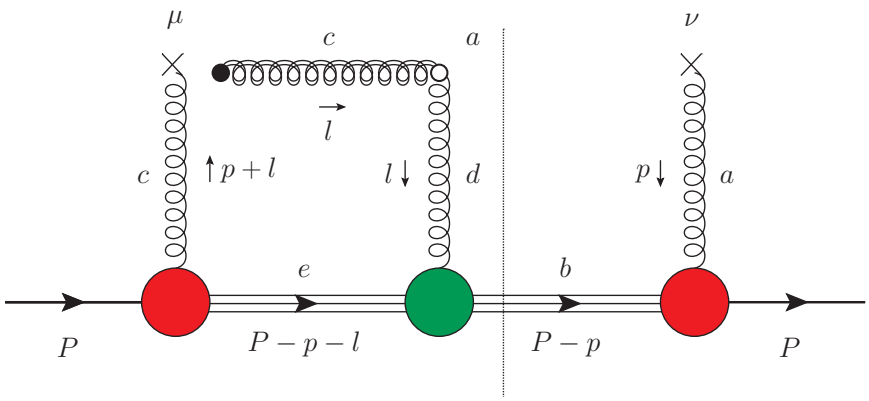


Figure 2. Diagram for the calculation of the gluon-gluon correlator including the single-gluon exchange contribution, necessary to obtain T-odd TMDs. The eikonal propagator arising from the Wilson line in the operator definition of TMDs is indicated by a gluon double line. Only the imaginary part of the box diagram on the left-hand side of the cut is relevant for the calculation of T-odd functions. The red blobs represent the nucleon-gluon-spectator vertex with color indices *ce* and *ba*, respectively, while the green blob stands for the spectator-gluon-spectator vertex with color indices *edb*. The Hermitian-conjugate diagram is not shown.

from final-state interactions with future-pointing ([+]) Wilson lines (like in Semi-Inclusive Deep-Inelastic Scattering - SIDIS) to initial-state interactions with past-pointing ([−]) Wilson lines (like in Drell-Yan processes) [34–36].

The gluon case is more intricate, due its color-octet structure, and leads to a more diversified form of modified universality with respect to the quark case. There is a gauge link with color flowing through a closed path pointing to the future, corresponding to final-state interactions between the spectator and an outgoing gluon, like in SIDIS production of two jets or heavy-quark pairs [37]. This gauge link is usually denoted with the $[+,+]$ symbol. Conversely, initial-state interactions are described by gauge links with past-pointing close Wilson lines ($[-,-]$) and occur, for example, in Higgs production via gluon fusion ($gg \to H$) [38, 39]. The gluon TMDs originating from these gauge links are called Weizsäcker–Williams (WW) gluon TMDs, or f-type gluon TMDs because the color structure of the T-odd ones involves the antisymmmetric structure constants f of the color gauge group SU(3). It turns out that T-even WW gluon TMDs are symmetric with respect to the different paths ($[+,+] = [-,-]$), while the T-odd WW ones change sign.

Moreover, color can flow through a closed path involving both initial and final states, like in photon-jet production from hadronic collisions or SIDIS [40–42]. We remark that for this class of processes TMD factorization is not expected to hold [43]; however, it is still possible to calculate the corresponding TMDs in the context of our model. Depending on the direction of color flow, we have $[+,-]$ and $[-,+]$ structures and the corresponding gluon TMDs are usually called dipole gluon TMDs, or d-type gluon TMDs because their T-odd color structure involves the symmetric structure constants d of color SU(3). [2] Similarly to the WW case, the dipole T-even gluon TMDs are symmetric with respect to different color paths ($[+,-] = [-,+]$), while T-odd dipole ones change sign. But, more importantly, WW and dipole gluon TMDs are not related to each other, and contain different physical information.

We first compute the gluon-gluon correlator corresponding to the $[+,+]$ gauge link with future-pointing closed Wilson path. The one-gluon exchange approximation of the gauge link amounts to compute the diagram depicted in Fig. 2. The double gluon line represents the struck gluon described in the eikonal approximation, following the same procedure of the quark case [25]. The Feynman rules to describe the eikonal gluon line and the eikonal vertex are written in detail in Ref. [47].

The expression of the correlator turns out to be

$$\Phi^{\mu\nu[+,+]}(x,\boldsymbol{p}_T,S) = \frac{1}{(2\pi)^3\,2\,(1-x)\,P^+}\,\mathrm{Tr}\Bigg[(\not{P}+M)\,\frac{1+\gamma^5\not{S}}{2}\,G^{*\nu\sigma}(p,p)\mathcal{Y}^{\dagger ab}_{\sigma}(p^2)(\not{P}-\not{p}+M_X)\,(g_s n_-^{\alpha} f^{dac}) \\ \times\int\frac{d^4l}{(2\pi)^4}\left(\frac{-i\mathcal{X}^{bde}_{\alpha}(l^2)}{l^2-m_g^2}\right)\left(\frac{-i}{l^+ + i\epsilon}\right)\frac{i(\not{P}-\not{p}-\not{l}+M_X)}{(P-p-l)^2-M_X^2+i\epsilon}\,\mathcal{Y}^{ec}_{\rho}\big((p+l)^2\big)\,G^{\mu\rho}(p,p+l)\Bigg]\,, \tag{11}$$

[2] Due to the connection between the T-odd TMDs at twist-2 and the collinear PDFs at twist-3, the distinction between f-type and d-type gluon TMDs appears already in the correlator of the Qiu–Sterman twist-3 collinear PDF [44–46].

where $\mathcal{X}_\alpha^{bde}$ is the spectator-gluon-spectator vertex to be defined in Section II C.

The correlator $\Phi^{\mu\nu[-,-]}$ for the $[-,-]$ past-pointing closed Wilson path can be obtained by changing the sign of the $+i\epsilon$ term in Eq. (11).

The correlator $\Phi^{\mu\nu[+,-]}$ for the $[+,-]$ gauge link (leading to $d-$type gluon TMDs [48–50]) can be simply derived by replacing in the eikonal vertex the antisymmetric color structure f^{dac} with the symmetric $-id^{dac}$ in Eq. (11):

$$\Phi^{\mu\nu[+,-]}(x,\boldsymbol{p}_T,S) = \frac{1}{(2\pi)^3\,2\,(1-x)\,P^+}\,\mathrm{Tr}\Bigg[(\not{P}+M)\,\frac{1+\gamma^5\not{S}}{2}\,G^{*\nu\sigma}(p,p)\mathcal{Y}_\sigma^{\dagger ab}(p^2)(\not{P}-\not{p}+M_X)\,(-ig_s n_-^\alpha d^{dac}) \\ \times\int\frac{d^4l}{(2\pi)^4}\left(\frac{-i\mathcal{X}_\alpha^{bde}(l^2)}{l^2-m_g^2}\right)\left(\frac{-i}{l^+ +i\epsilon}\right)\frac{i\big(\not{P}-\not{p}-\not{l}+M_X\big)}{(P-p-l)^2-M_X^2+i\epsilon}\,\mathcal{Y}_\rho^{ec}\big((p+l)^2\big)\,G^{\mu\rho}(p,p+l)\Bigg]\,. \tag{12}$$

As for the WW case, the $[-,+]$ correlator differs from the $[+,-]$ one only by the sign of the $+i\epsilon$ term in Eq. (12).

Our model agrees with the relations between gluon TMDs with different gauge link structures that have been systematically studied in [51]. For example, for the T-even unpolarized function, f_1^g, and for the T-odd gluon Sivers function, $f_{1T}^{g\perp}$, one has the following modified-universality relations [11, 40, 51]:

$$f_1^{g\,[+,+]} = f_1^{g\,[-,-]}\,, \qquad f_1^{g\,[+,-]} = f_1^{g\,[-,+]}\,, \tag{13}$$

$$f_{1T}^{\perp\,g\,[+,+]} = -f_{1T}^{\perp\,g\,[-,-]}\,, \qquad f_{1T}^{\perp\,g\,[+,-]} = -f_{1T}^{\perp\,g\,[-,+]}\,. \tag{14}$$

As it turns out, in general $f_1^{g\,[+,+]}$ *cannot* be related to $f_1^{g\,[+,-]}$, and likewise for $f_{1T}^{\perp\,g}$. They encode different information and require different extractions [3].

C. Spectator-gluon-spectator vertex

A key ingredient of our model is the spectator-gluon-spectator vertex $\mathcal{X}_\alpha^{bde}$, depicted by a green blob in Fig. 2. If the nucleon-gluon-spectator vertex $\mathcal{Y}_\mu^{ec}$ (red blob) connects a colorless initial-state (nucleon) to an octet state (gluon) and an anti-octet state (spectator), the spectator-gluon-spectator vertex $\mathcal{X}_\alpha^{bde}$ connects an anti-octet initial state (spectator) to an octet state (gluon) and an anti-octet state (spectator). Since in our model the spectator is assumed to be a spin-$\frac{1}{2}$ particle describing a collection of partons as remainders, the vertex $\mathcal{X}_\alpha^{bde}$ can in principle contain both the f^{bde} and d^{bde} color structure constants, each one multiplying a Dirac structure similar to Eq. (5):

$$\mathcal{X}_\alpha^{bde}(p^2) = f^{bde}\left[g_1^f(p^2)\,\gamma_\alpha + g_2^f(p^2)\,\frac{i}{2M}\,\sigma_{\alpha\beta}\,p^\beta\right] - i\,d^{bde}\left[g_1^d(p^2)\,\gamma_\alpha + g_2^d(p^2)\,\frac{i}{2M}\,\sigma_{\alpha\beta}\,p^\beta\right], \tag{15}$$

where $g_{1,2}^{f,d}(p^2)$ are *a priori* four different functions of p^2. In principle, they are independent from the $g_{1,2}(p^2)$ form factors entering the nucleon-gluon-spectator vertex of Eq. (5). For the sake of simplicity, we will assume $g_{1,2}^d(p^2) = g_{1,2}^f(p^2)$ and we will get d-type densities equal to the corresponding f-type ones up to a color factor:

$$C^{[+,+]} = f^{acd}f^{dca} = -2C_A^2C_F = -24\,, \tag{16}$$

$$C^{[+,-]} = (-i\,d^{acd})\,(-i\,d^{dca}) = 2\left(4-C_A^2\right)C_F = -\frac{40}{3}\,. \tag{17}$$

For this reason, in the following we will show results only for f-type gluon TMDs, and we will drop the $[+,+]$ index when not needed.

In the quark-target model, we would replace the spectator-gluon-spectator vertex by $\mathcal{X}_\alpha^{bde} \to (-ig_s\gamma_\alpha t^d)$ and obtain the color factors

$$C_q^{[+,+]} = -\frac{i}{N_C}\,\mathrm{Tr}_C[t^at^ct^d]f^{acd} = \frac{C_AC_F}{2} = -2\,, \tag{18}$$

$$C_q^{[+,-]} = -\frac{1}{N_C}\,\mathrm{Tr}_C[t^at^ct^d]d^{acd} = \frac{C_F}{2}\left(4-C_A^2\right)(C_A-2C_F) = -\frac{10}{9}\,. \tag{19}$$

Note that the ratio of the two different gauge link structures remains the same in the two cases: $[+,-]/[+,+] = 5/9$. With our simplified assumptions, therefore, the T-odd d-type functions are always about half of the f-type ones.

We further assume

$$g^d_{1,2}(p^2) = g^f_{1,2}(p^2) \equiv g_{1,2}(p^2)\ . \tag{20}$$

This means that the parameters entering our model for f-type and d-type T-odd gluon TMDs are fully determined by those ones entering the T-even gluon TMDs that contain $g_{1,2}(p^2)$ through Eq. (5). The latter parameters have been fixed by fitting the integrated T-even gluon TMDs on the known corresponding collinear PDFs [30] (see Tab. II).

D. Gluon TMD projectors

T-odd gluon TMDs can be extracted from the analytic structure of the gluon-gluon correlator by making use of suitable projectors. Using Eqs.(52-54) of Ref. [4] for the general parametrization of the gluon-gluon correlator $\Phi^{\mu\nu}(x, \boldsymbol{p}_T, S)$ for three different nucleon polarizations $S = 0, S_L, \boldsymbol{S}_T$, it is possible to show that the four T-odd gluon TMDs of Tab. I can be isolated through the following projections:

$$\begin{aligned} f_{1T}^{\perp g} &= \mathbb{P}^{\mu\nu}_{[f_{1T}^{\perp g}]}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)] \\ &= \frac{M}{2}\frac{1}{\epsilon_T^{p_T S_T}} g_T^{\mu\nu}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)]\ , \end{aligned} \tag{21}$$

$$\begin{aligned} h_{1T}^{\perp g} &= \mathbb{P}^{\mu\nu}_{[h_{1T}^{\perp g}]}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)] \\ &= \frac{M^3}{p_T^2}\frac{1}{\epsilon_T^{p_T S_T}}\left(\frac{p_T^{\{\mu} S_T^{\nu\}}}{p_T \cdot S_T} - g_T^{\mu\nu}\right)\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)]\ , \end{aligned} \tag{22}$$

$$\begin{aligned} h_1^g &= \mathbb{P}^{\mu\nu}_{[h_1^g]}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)] \\ &= \frac{M}{2\epsilon_T^{p_T S_T}}\left[\frac{4}{p_T^2} p_T^\mu p_T^\nu + \frac{p_T^{\{\mu} S_T^{\nu\}}}{p_T \cdot S_T} - 3 g_T^{\mu\nu}\right]\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, \boldsymbol{S}_T) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -\boldsymbol{S}_T)]\ , \end{aligned} \tag{23}$$

$$\begin{aligned} h_{1L}^{\perp g} &= \mathbb{P}^{\mu\nu}_{[h_{1L}^{\perp g}]}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, S_L) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -S_L)] \\ &= \frac{1}{S_L}\frac{M^2}{2p_T^4}\epsilon_T^{\{\mu}{}_\alpha p_T^{\nu\}\alpha}\ [\Phi_{\mu\nu}(x, \boldsymbol{p}_T, S_L) - \Phi_{\mu\nu}(x, \boldsymbol{p}_T, -S_L)]\ , \end{aligned} \tag{24}$$

where $\epsilon_T^{vw} \equiv \epsilon^{-+ij}\, v_i\, w_j$ with i, j transverse spatial indices and $\epsilon^{\mu\nu\alpha\beta}$ the antisymmetric Levi-Civita tensor, and

$$p_T^{\mu\nu} = p_T^\mu p_T^\nu - \frac{1}{2} p_T^2 g_T^{\mu\nu}\ , \tag{25}$$

$$g_T^{\mu\nu} = g^{\mu\nu} - n_+^{\{\mu} n_-^{\nu\}}\ , \tag{26}$$

$$v^{\{\mu} w^{\nu\}} = v^\mu w^\nu + v^\nu w^\mu\ . \tag{27}$$

III. T-ODD GLUON TMDS: ILLUSTRATIVE EXAMPLES

A. Sivers function: g_1-vertex approximation

Let us consider first the f-type Sivers function $f_{1T}^{\perp[+,+]}$ with a simpler expression for the nucleon-gluon-spectator vertex, where the term proportional to $\sigma^{\mu\nu} p_\nu$ in Eq. (5) is neglected. In other words, the $g_2(p)$ coupling is set to zero and the vertex reduces to

$$\mathcal{Y}_\rho^{ba} \to \delta^{ba}\, g_1(p^2)\, \gamma_\rho\ . \tag{28}$$

We name this the "g_1-vertex approximation". We indicate the resulting Sivers function as $f_{1T}^{\perp(g_1)}$, and similarly for all other TMDs computed in this approximation.

Using for $g_1(p^2)$ the dipolar form of Eq. (6), the corresponding projector of Eq. (21) specialized to a f-type gluon Sivers and applied to the correlator of Eq. (11), we obtain

$$
\begin{aligned}
f_{1T}^{\perp\,(g_1)}(x,\boldsymbol{p}_T) &= \mathbb{P}^{\mu\nu}_{[f_{1T}^{\perp g}]}\left[\Phi^{[+,+]\,(g_1)}_{\mu\nu}(x,\boldsymbol{p}_T,\boldsymbol{S}_T) - \Phi^{[+,+]\,(g_1)}_{\mu\nu}(x,\boldsymbol{p}_T,-\boldsymbol{S}_T)\right] \\
&= \frac{48\,g_s\kappa_1^3}{(2\pi)^3}\frac{M\,[M_X - M(1-x)]\;P^+}{(p^2-\Lambda_X^2)^2}\,2\,\mathrm{Re}\int\frac{d^4l}{(2\pi)^4}\frac{\epsilon_T^{l_T S_T}}{(l^2-\Lambda_X^2)^2}\frac{1}{l^+ + i\epsilon}\frac{1}{(l+p-P)^2 - M_X^2 + i\epsilon}\frac{1}{[(p+l)^2-\Lambda_X^2+i\epsilon]^2}\,.
\end{aligned}
\tag{29}
$$

In Eq. (29), terms proportional to $\epsilon^{ln-pP} \equiv \epsilon^{\mu\nu\alpha\beta} l_\mu n_{-\nu} p_\alpha P_\beta$ and ϵ^{lpPS} vanish because the only component of l contributing to the integral is the one parallel to p.

Similarly to the calculation of the quark Sivers TMD [25, 52–55], the non vanishing contribution to the integral of Eq. (29) comes from the poles of the two $[l^+ + i\epsilon]$ and $[(l+p-P)^2 - M_X^2 + i\epsilon]$ propagators.

Using the Cutkosky's rules, we can make the replacement

$$
\frac{1}{l^+ + i\epsilon} \to -2\pi i\,\delta(l^+)\,, \qquad \frac{1}{(l+p-P)^2 - M_X^2 + i\epsilon} \to -2\pi i\,\delta((l+p-P)^2 - M_X^2)\,. \tag{30}
$$

Moreover, we can also make use of the spectator model relation

$$
k^2 - \Lambda_X^2 = -\frac{\boldsymbol{k}_T^2 + L_X^2(\Lambda_X)}{1-x}\,, \tag{31}
$$

where k^2 generically refers to p^2, l^2 or $(p+l)^2$, and $\boldsymbol{k}_T^2$ to the corresponding euclidean transverse parts.

The final result for the WW gluon Sivers function with only g_1 coupling is

$$
f_{1T}^{\perp\,(g_1)}(x,\boldsymbol{p}_T) = -\frac{48\,g_s\kappa_1^3}{(2\pi)^3}\frac{M\,[M_X - M(1-x)]\;(1-x)^5\,P^+}{[\boldsymbol{p}_T^2 + L_X^2(\Lambda_X)]^2}\,\mathcal{D}_2(p)\,, \tag{32}
$$

where

$$
\mathcal{D}_2(p) = \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{\boldsymbol{l}_T\cdot\boldsymbol{p}_T}{\boldsymbol{p}_T^2}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2}\,. \tag{33}
$$

Introducing the Feynman parametrization, we can rewrite the integral as

$$
\mathcal{D}_2(p) = \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{\boldsymbol{l}_T\cdot\boldsymbol{p}_T}{\boldsymbol{p}_T^2}\int_0^1 d\alpha\,\frac{6\,\alpha\,(1-\alpha)}{\{\alpha\,[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)] + (1-\alpha)\,[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]\}^4}\,. \tag{34}
$$

After the change of variable $\boldsymbol{l}_T \to \boldsymbol{l}'_T = \boldsymbol{l}_T + (1-\alpha)\boldsymbol{p}_T$, we have

$$
\begin{aligned}
\mathcal{D}_2(p) &= -\frac{1}{2P^+}\int_0^1 d\alpha\int\frac{d^2\boldsymbol{l}'_T}{(2\pi)^2}\frac{6\,\alpha\,(1-\alpha)^2}{[\boldsymbol{l}_T'^2 + \alpha(1-\alpha)\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2)]^4} \\
&= -\frac{1}{4\pi P^+}\int_0^1 d\alpha\,\frac{\alpha\,(1-\alpha)^2}{[\alpha(1-\alpha)\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2)]^3} \\
&= -\frac{1}{8\pi P^+}\left[\frac{1 - 2\,L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} + 8\frac{\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2)}{|\boldsymbol{p}_T|^3\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,\tanh^{-1}\left(\frac{|\boldsymbol{p}_T|}{\sqrt{\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)}}\right)\right]\,.
\end{aligned}
\tag{35}
$$

Combining Eqs. (32) and (35) we get the final expression for our f-type Sivers function in the g_1-vertex approximation

$$
\begin{aligned}
f_{1T}^{\perp\,(g1)}(x,\boldsymbol{p}_T) &= \frac{12\,g_s\kappa_1^3}{(2\pi)^4}\frac{M\,[M_X - M(1-x)]\;(1-x)^5}{[\boldsymbol{p}_T^2 + L_X^2(\Lambda_X)]^2} \\
&\times\left[\frac{1 - 2\,L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} + 8\frac{\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2)}{|\boldsymbol{p}_T|^3\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,\tanh^{-1}\left(\frac{|\boldsymbol{p}_T|}{\sqrt{\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)}}\right)\right]\,.
\end{aligned}
\tag{36}
$$

In order to explore the effects of the g_1-vertex approximation, we fix the model parameters by simultaneously fitting the integrated unpolarized and helicity gluon TMDs onto the corresponding known collinear PDFs. Following

	PVGlue20		PVGlue20g1V	
parameter	mean	replica 11	mean	replica 11
A [GeV^{-1}]	6.1 ± 2.3	6.0	4.3 ± 1.5	4.29
a	0.82 ± 0.21	0.78	0.73 ± 0.14	0.73
b	1.43 ± 0.23	1.38	1.34 ± 0.13	1.33
C [GeV^{-1}]	371 ± 58	346	349 ± 24	350
D [GeV]	0.548 ± 0.081	0.548	0.595 ± 0.049	0.586
σ [GeV]	0.52 ± 0.14	0.50	0.42 ± 0.08	0.41
Λ_X [GeV]	0.472 ± 0.058	0.448	0.398 ± 0.035	0.384
κ_1 [GeV^2]	1.51 ± 0.16	1.46	1.33 ± 0.08	1.28
κ_2 [GeV^2]	0.414 ± 0.036	0.414	0.0	0.0

Table II. Mean values of fitted parameters with their 68% uncertainties, and corresponding values for the most representative replica 11 (see text). The original fit of Ref. [30] and the "g_1-vertex approximation" are labeled as `PVGlue20` and `PVGlue20g1V`, respectively (see text).

the methodology of Ref. [30], we first allow the spectator mass to take a continuous range of values by weighting the gluon TMDs with the spectral function described in Eqs.(16,17) of Ref. [30], which is a way to effectively take into account $q\bar{q}$ contributions. Then, we integrate the gluon TMDs upon the transverse momenta and we fix all the model parameters by fitting the unpolarized collinear PDF from `NNPDF3.1sx` [56] and the helicity collinear PDF from `NNPDFpol1.1` [57] at the indicated initial scale $Q_0 = 1.64$ GeV and in the range $10^{-3} < x < 0.7$. [3] The only exception is the parameter κ_2 in Eq. (6) that controls the strength of the g_2 coupling; here, it is systematically set to zero. Statistical uncertainties are generated using the replica method, widely used in the phenomenological extraction of quark densities from experimental data [56, 59–65].

In Tab. II, we compare the obtained values (labelled `PVGlue20g1V` in the two rightmost columns) with the original values from Ref. [30] (labelled `PVGlue20` in the second and third columns from left). The 68% uncertainties accompanying the central values are obtained by excluding the largest and smallest 16% of all 100 replica values, which would correspond to 1σ standard deviation for a Gaussian distribution. The columns labelled with "replica 11" show the parameters of the most representative replica, because in both fits its parameter values have the minimal distance from the mean values.

In Fig. 3, we show the results for the f-type unpolarized gluon TMD (upper panels) and the gluon Sivers function multiplied by $x|\boldsymbol{p}_T|/M$ (lower panels), as functions of $\boldsymbol{p}_T^2$, in the g_1-vertex approximation. [4] Left (right) plots are for TMDs calculated at $x = 10^{-3}$ ($x = 10^{-1}$) and at $Q_0 = 1.64$ GeV. As for the parameter values, the 68% uncertainty bands are formed by excluding the largest and smallest 16% of 100 computed replicas. The black solid line is the result of the most representative replica 11. Here, and in the following, the strong coupling constant is fixed to $g_s = \sqrt{\alpha_s(Q_0)} = 0.57583$. The qualitative behavior of the TMD f_1^g stays practically the same with respect to the original fit (see upper panels of Fig. 4 in Ref. [30]). The resulting gluon Sivers function decreases at low x. However, this trend can radically change when including also the g_2 vertex, as shown in Sec. IV.

B. Sivers function: Quark-target model

In a similar way, we can also derive the results for the f-type gluon Sivers function in the quark-target model. We indicate it with the superscript (q). In this approximation, the incoming proton in Fig. 2 is replaced by a quark, and similarly for the spectator remnant. Therefore, both the proton and the spectator mass, M and M_X, are set equal to the target-quark mass, m_q, and the effective nucleon-gluon-spectator vertex is replaced by a QCD quark-gluon-quark one. Starting from the expression for the $[+,+]$ gluon correlator in the quark-target model, we obtain

$$f_{1T}^{\perp\,(q)}(x, \boldsymbol{p}_T) = -\frac{4\,g_s^4}{(2\pi)^3}\frac{m_q^2\,x\,(1-x)\,P^+}{(m_q^2 x^2 + \boldsymbol{p}_T^2)}\,\mathcal{D}_q(p)\,, \tag{37}$$

where

$$\mathcal{D}_q(p) = \frac{1}{2(1-x)P^+}\int \frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\,\frac{\boldsymbol{l}_T\cdot\boldsymbol{p}_T}{\boldsymbol{p}_T^2}\,\frac{1}{\boldsymbol{l}_T^2}\,\frac{(1-x)}{(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + x^2 m_q^2}\,. \tag{38}$$

[3] The $x > 0.7$ tail was excluded to avoid large uncertainties [58] due to threshold effects and target-mass corrections, not accounted for in our model.

[4] Preliminary results for the f-type Sivers function in the g_1-vertex approximation were previously presented in Refs. [66, 67].

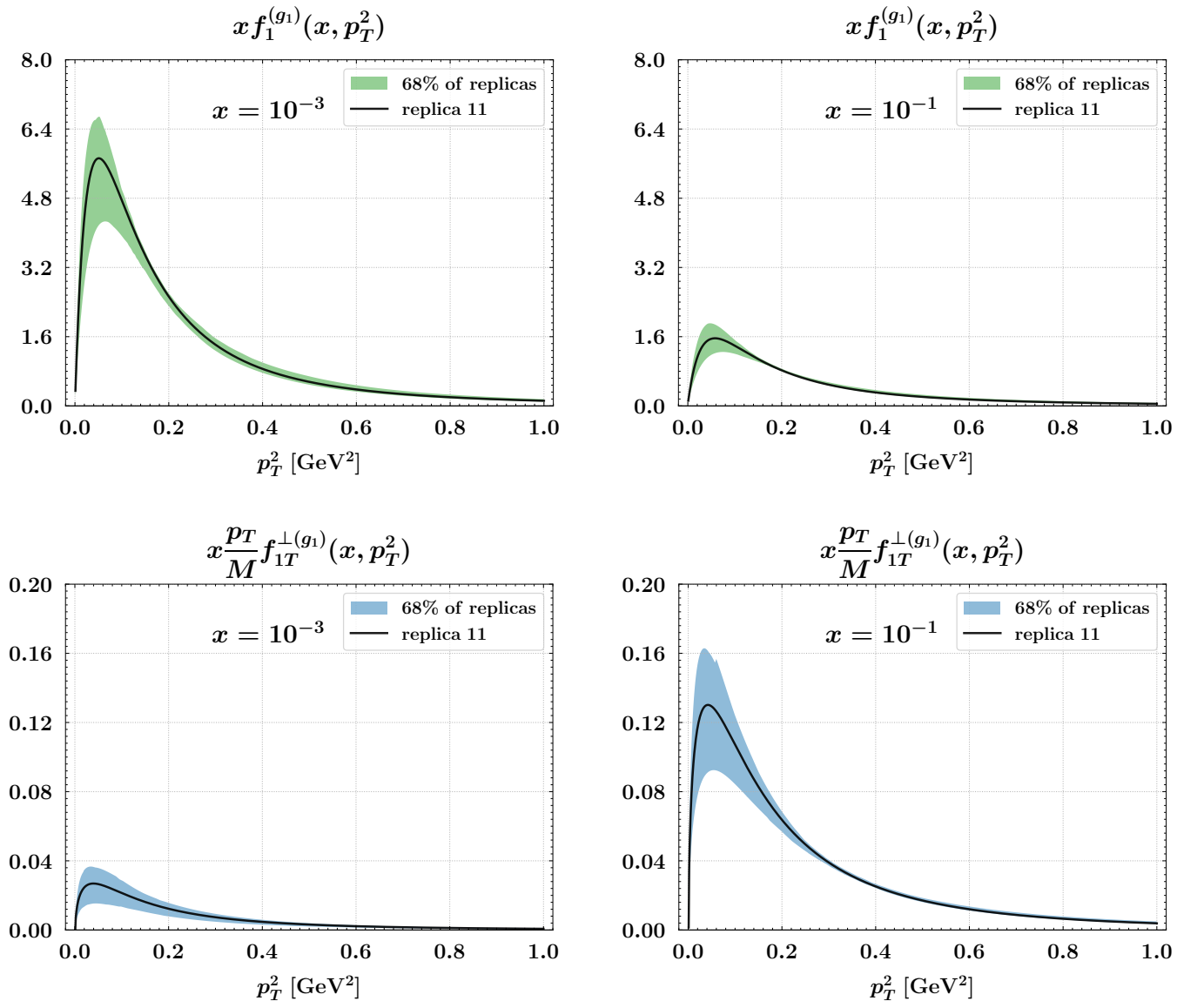


Figure 3. Transverse-momentum dependence of the $[+,+]$ unpolarized TMD (upper panel) and gluon Sivers TMD multiplied by $x|\boldsymbol{p}_T|/M$ (lower panel) calculated in the g_1-vertex approximation (see text), at $x = 10^{-3}$ (left panels) and $x = 10^{-1}$ (right panels) and at the initial scale $Q_0 = 1.64$ GeV. Uncertainty band by including the 68% of all computed replicas. Black curves refer to the most representative replica 11 (see text).

Following similar steps as in the previous case, we obtain

$$\mathcal{D}_q(p) = -\frac{1}{8\pi P^+}\frac{1}{\boldsymbol{p}_T^2}\ln\frac{\boldsymbol{p}_T^2 + x^2 m_q^2}{x^2 m_q^2}\,, \tag{39}$$

and we get the final expression for our f-type gluon Sivers function in the quark-target model

$$f_{1T}^{\perp\,(q)}(x, \boldsymbol{p}_T) = \frac{g_s^4}{(2\pi)^4}\frac{m_q^2\, x\,(1-x)}{\boldsymbol{p}_T^2(m_q^2 x^2 + \boldsymbol{p}_T^2)}\ln\frac{\boldsymbol{p}_T^2 + x^2 m_q^2}{x^2 m_q^2}\,, \tag{40}$$

which corresponds to Eq. (B12) of Ref. [4].

C. Linearity function: g_1-vertex approximation

Let us consider now the distribution of linearly polarized gluons in a transversely polarized target, denoted as h_1 in Tab. I. For simplicity, we will call it "linearity function" even if this terminology could be used for any h functions in the rightmost column of Tab. I. In spite of the similarity in notation, this function should not be confused with the analogue of the quark transversity distribution. In fact, it does not survive transverse-momentum integration and is T-odd.

In the following, we derive the f-type gluon linearity function in the g_1-vertex approximation of our spectator model. Using the corresponding projector from Eq. (23), the dipolar form for $g_1(p^2)$ as in Eq. (6), and the $[+,+]$ gluon correlator of Eq. (11), we have

$$\begin{aligned} h_1^{(g_1)}(x,\boldsymbol{p}_T) &= \mathbb{P}^{\mu\nu}_{[h_1^g]}\left[\Phi^{[+,+]\,(g_1)}_{\mu\nu}(x,\boldsymbol{p}_T,\boldsymbol{S}_T) - \Phi^{[+,+]\,(g_1)}_{\mu\nu}(x,\boldsymbol{p}_T,-\boldsymbol{S}_T)\right] \\ &= \frac{96\,g_s\kappa_1^3}{(2\pi)^3}\frac{M\,[M-M_X(1-x)]\;P^+}{(1-x)\,\boldsymbol{p}_T^2\,(p^2-\Lambda_X^2)^2} \\ &\quad\times 2\,\mathrm{Re}\int\frac{d^4l}{(2\pi)^4}\frac{l_T\cdot p_T}{(l^2-\Lambda_X^2)^2}\frac{1}{l^+ + i\epsilon}\frac{1}{(l+p-P)^2-M_X^2+i\epsilon}\frac{1}{[(p+l)^2-\Lambda_X^2+i\epsilon]^2}\,. \end{aligned} \tag{41}$$

Following the same steps described in Section III A we obtain

$$h_1^{(g_1)}(x,\boldsymbol{p}_T) = -\frac{96\,g_s\kappa_1^3}{(2\pi)^3}\frac{M\,[M-M_X(1-x)]\;(1-x)^4\,P^+}{[\boldsymbol{p}_T^2+L_X^2(\Lambda_X)]^2}\,\mathcal{D}_2(p)\,, \tag{42}$$

where $\mathcal{D}_2(p)$ is defined and computed in Eqs. (33)-(35). The final expression for our f-type gluon linearity function in the g_1-vertex approximation is

$$\begin{aligned} h_1^{(g1)}(x,\boldsymbol{p}_T) &= \frac{24 g_s\kappa_1^3}{(2\pi)^4}\frac{M\,[M-M_X(1-x)]\;(1-x)^4}{[\boldsymbol{p}_T^2+L_X^2(\Lambda_X)]^2} \\ &\quad\times\left[\frac{1-2\,L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^2} + 8\frac{\boldsymbol{p}_T^2+L_X^2(\Lambda_X^2)}{|\boldsymbol{p}_T|^3\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^{5/2}}\,\tanh^{-1}\left(\frac{|\boldsymbol{p}_T|}{\sqrt{\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)}}\right)\right]\,. \end{aligned} \tag{43}$$

Preliminary results on the f-type gluon linearity function in the g_1-vertex approximation were presented in Refs. [67, 68].

D. Linearity function: Quark-target model

In the quark-target model, following an analogous procedure to the one in Section III B, we get

$$h_1^{(q)}(x,\boldsymbol{p}_T) = -\frac{8\,g_s^4}{(2\pi)^3}\frac{m_q^2\,x\,P^+}{(m_q^2x^2+\boldsymbol{p}_T^2)}\,\mathcal{D}_q(p)\,. \tag{44}$$

Combining Eqs. (44) and (39), we get the final expression for our f-type gluon linearity function in the quark-target model

$$h_1^{(q)}(x,\boldsymbol{p}_T) = \frac{g_s^4}{(2\pi)^4}\frac{2x\,m_q^2}{\boldsymbol{p}_T^2(m_q^2x^2+\boldsymbol{p}_T^2)}\ln\frac{\boldsymbol{p}_T^2+x^2m_q^2}{x^2m_q^2}\,, \tag{45}$$

which corresponds to Eq. (B17) of Ref. [4].

IV. T-ODD GLUON TMDS: RESULTS OF FULL CALCULATION

If we include the full structure of the nucleon-gluon-spectator vertex $\mathcal{Y}_\rho^{ba}$ in Eq. (5), a given T-odd gluon TMD, generically indicated by $F(x,\boldsymbol{p}_T^2)$, can be organized as

$$F(x,\boldsymbol{p}_T^2) = \sum_{i,j,k}^{1,2} C_{ijk}^{[F]}(x,\boldsymbol{p}_T^2)\,g_s\,\kappa_i\,\kappa_j\,\kappa_k\,, \tag{46}$$

where $\kappa_{i,j,k}$ are the coupling constants encoded in the dipolar form factors of Eq. (6) with the assumption made in Eq. (20), and $C_{ijk}^{[F]}$ are related coefficients. For each T-odd gluon TMD $F(x, \boldsymbol{p}_T^2)$, the $C_{ijk}^{[F]}$ can be split in eight different contributions $C_{ijk}^{[F],l}$, $l = 1, .., 8$, and organized as linear combinations according to

$$C_{ijk}^{[F]}(x, \boldsymbol{p}_T^2) = \frac{(1-x)^4 P^+}{(2\pi)^3 \left[\boldsymbol{p}_T^2 + L_X^2(\Lambda_X)\right]^2} \sum_{l=1}^{8} \mathcal{C}_{ijk}^{[F],l}(x, \boldsymbol{p}_T^2)\, \mathcal{D}_l(x, \boldsymbol{p}_T^2)\ , \tag{47}$$

where $\mathcal{D}_l(x, \boldsymbol{p}_T^2)$ are eight different master integrals that can be found in Appendix A. The final expressions of the $C_{ijk}^{[F],l}$ coefficients for each T-odd gluon TMD F and for $l = 1, .., 8$ and $i, j, k = 1, 2$, are listed in Appendix B.

We note that both the T-odd f-type $h_{1L}^\perp$ and $h_{1T}^\perp$ vanish in the g_1-vertex approximation and in the quark-target model, because the integral describing the loop in Fig. 2 would be proportional to l^+, which is set to zero by the first of the two Cutkosky rules in Eq. (30). This result is in line with Eqs. (B16) and (B18) of Ref. [4], respectively.

In the following, we show the results of the full calculation of all the four T-odd f-type gluon TMDs that appear at leading twist (see Tab. I). We recall that in our model T-odd d-type gluon TMDs turn out to be equal to the f-type ones up to a color factor computed in Eq. (17), because in the vertices we take the same dipole-like couplings $g_{1,2}(p^2)$ for f-type and d-type functions. Moreover, the parameters of both T-odd f-type and d-type functions are fully determined by those ones entering the T-even gluon TMDs. These parameters were fixed in Ref. [30] by fitting the integrated T-even gluon TMDs onto the corresponding known collinear PDFs at the low scale $Q_0 = 1.64$ GeV; their values are listed in the columns of Tab. II labelled by `PVGlue20`.

It is convenient to start from the f-type gluon Sivers function $f_{1T}^\perp$ in order to compare with the results displayed in the lower panels of Fig. 3 using the g_1-vertex approximation.

In the upper panels of Fig. 4, we display the T-odd f-type gluon Sivers function $f_{1T}^\perp$ multiplied by $x|\boldsymbol{p}_T|/M$, as a function of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV. As in previous figures, the uncertainty band is constructed by excluding the largest and smallest 16% of all 100 computed replicas, roughly corresponding to 1σ standard deviation. The solid black line is the result of the most representative replica 11. The observed behavior in $\boldsymbol{p}_T^2$ clearly does not follow a simple Gaussian pattern, rather it shows a large flattening tail for increasing $\boldsymbol{p}_T^2$. The f-type Sivers function is regular in $\boldsymbol{p}_T^2 = 0$, as it can be realized by inspecting the coefficients of Eqs. (46),(47) listed in Tabs. III-VI and the master integrals in Appendix A. Hence, the combination $x|\boldsymbol{p}_T|/M\, f_{1T}^\perp$ vanishes at $\boldsymbol{p}_T^2 = 0$.

By comparing with the lower panels of Fig. 3 where the $x|\boldsymbol{p}_T|/M\, f_{1T}^{\perp(g_1)}$ was computed in the g_1-vertex approximation, we realize that the contribution of the g_2 coupling to the vertices $\mathcal{Y}_\rho^{ba}$ of Eq. (5) and $\mathcal{X}_\alpha^{bde}$ of Eq. (15) completely reverses the situation: the f-type gluon Sivers function now increases for decreasing x, thus supporting the statement that spin asymmetries generated by this T-odd gluon TMD could be sizable also at small-x.

In the lower panels of Fig. 4, we show the result of the full calculation of the T-odd f-type gluon linearity function h_1 multiplied by $x|\boldsymbol{p}_T|/M$, as a function of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV. Notations are the same as in previous panels. The displayed trend is similar to the f-type Sivers function. Namely, the linearity increases with decreasing x, actually having a size larger than the Sivers function. The linearity is also regular at $\boldsymbol{p}_T^2 = 0$, hence vanishes at this point when multiplied by $x|\boldsymbol{p}_T|/M$.

In Fig. 5, we show for the first time the result of the full calculation of the T-odd f-type gluon $h_{1L}^\perp$ (upper panels) and $h_{1T}^\perp$ (lower panels) functions. In particular, in the upper panel we display $x\boldsymbol{p}_T^2/M^2\, h_{1L}^\perp$ as a function of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV. Notations are the same as in previous figures. We note that the absolute size increases with decreasing x, but overall it is much smaller than the Sivers and linearity functions. Interestingly, at $x = 10^{-1}$ the $h_{1L}^\perp$ function shows a very long tail in $\boldsymbol{p}_T^2$ but changes sign having a node at $\boldsymbol{p}_T^2 \approx 0.1$ GeV^2.

In the lower panels, the $x\boldsymbol{p}_T^3/M^3\, h_{1T}^\perp$ is displayed as a function of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV. Notations are the same as in previous figures. The absolute size is one order of magnitude smaller, raising doubts on the actual possibility of ever extracting the $h_{1T}^\perp$ from a spin asymmetry measurement. However, it shows an interesting structure with a node at small $\boldsymbol{p}_T^2$ and x.

Using the T-odd gluon TMDs computed in our model, we can complete the tomographic picture of the nucleon already discussed in Ref. [30]. To this purpose, we can construct 2-dim $\boldsymbol{p}_T$-distributions of gluons at different x for various combinations of their polarization and of the nucleon spin state.

Excluding the case of a circularly polarized gluon for which no T-odd gluon TMDs occur (see Tab. I), we can have in principle six combinations: two polarization states of the gluon (unpolarized, linearly polarized) for each polarization state of the parent nucleon (unpolarized, longitudinally polarized, transversely polarized). However, the

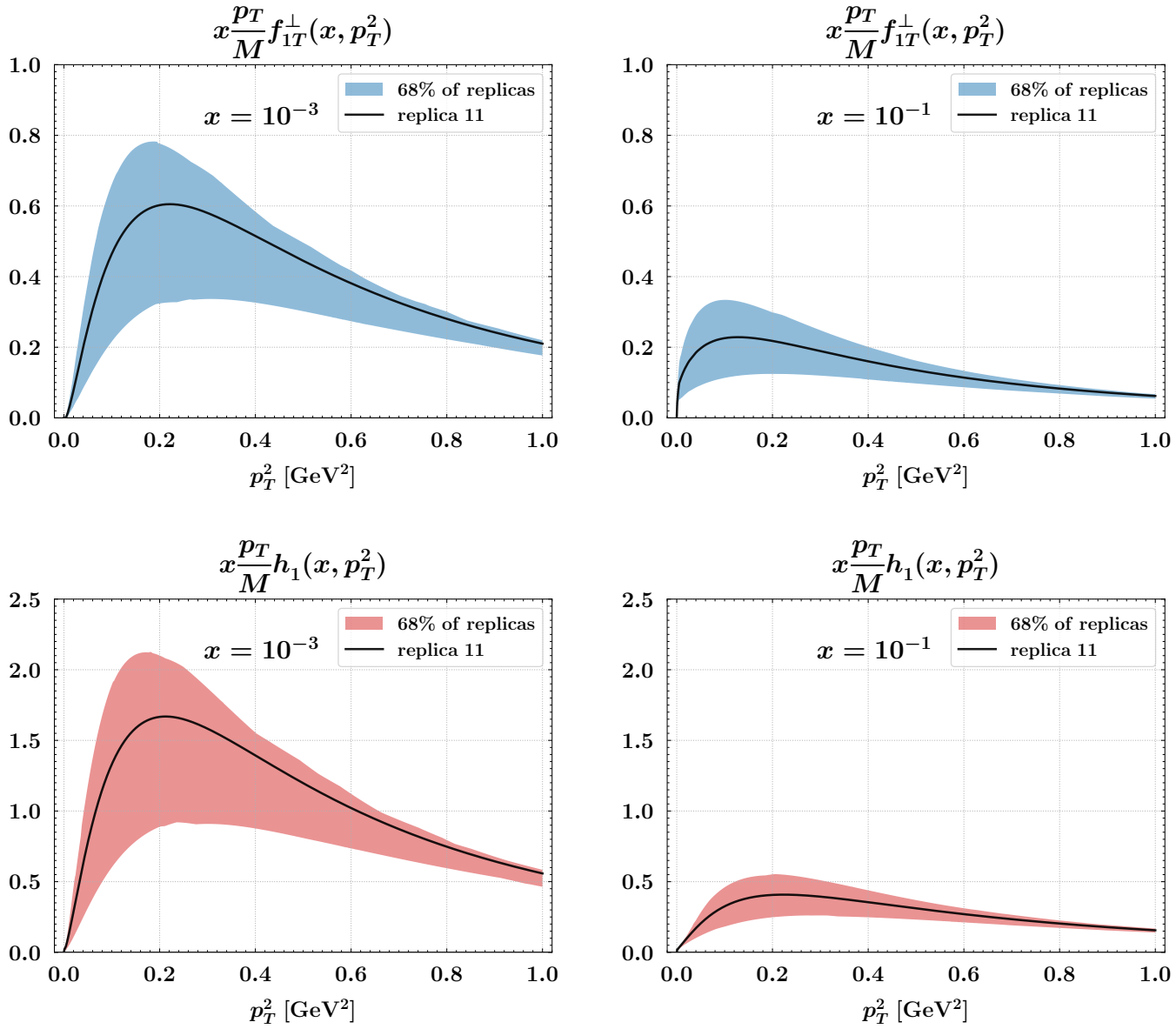


Figure 4. Transverse-momentum dependence of the $[+,+]$ gluon Sivers (upper) and linearity (lower) functions multiplied by $x|\boldsymbol{p}_T|/M$, as functions of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left), and $x = 10^{-1}$ (right) and at the initial scale $Q_0 = 1.64$ GeV. Uncertainty band from 68% of all computed replicas. Black curves refer to the most representative replica 11 (see text).

actual combinations are five, since an unpolarized gluon in a longitudinally polarized nucleon is forbidden by parity invariance (see Tab. I).

For a unpolarized gluon in a unpolarized nucleon, we identify the 2-dim density as

$$x\rho(x, p_x, p_y) = x\, f_1(x, \boldsymbol{p}_T^2)\ , \tag{48}$$

where f_1 is the leading-twist f-type unpolarized gluon TMD. The upper panels of Fig. 6 show the contour plots for the $\boldsymbol{p}_T$-distribution of $x\rho$ from replica 11 at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV, for a nucleon moving towards the reader. The color code identifies the size of the oscillations. For a better visualization, ancillary 1-dim plots are attached, representing a "slice" of $x\rho$ at $p_x = 0$ or $p_y = 0$. The 68% uncertainty band is obtained as usual by excluding the largest and smallest 16% of 100 computed replicas; the solid black line is the result of replica 11, actually corresponding to the 2-dim contour plot. Since both nucleon and gluon are unpolarized, the 2-dim density shows a perfect cylindrical symmetry around the direction of motion of the nucleon pointing towards the reader.

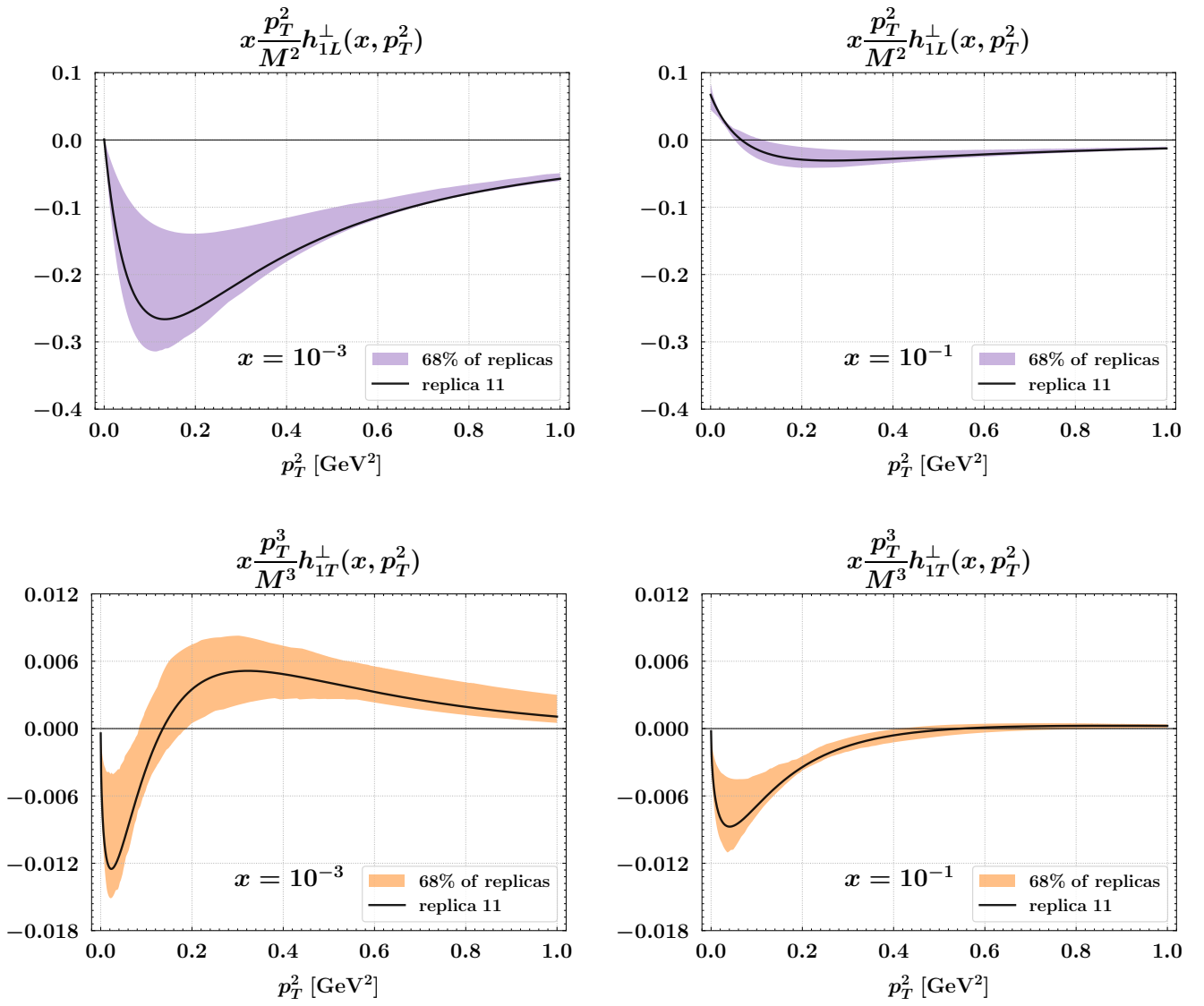


Figure 5. Transverse-momentum dependence of the $[+,+]$ gluon $h_{1L}^{\perp}$ (upper) and $h_{1T}^{\perp}$ (lower) functions multiplied by $x\boldsymbol{p}_T^2/M^2$ and $x|\boldsymbol{p}_T|^3/M^3$, respectively, as functions of $\boldsymbol{p}_T^2$ at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the initial scale $Q_0 = 1.64$ GeV. Notations as in previous figures.

For a unpolarized gluon in a nucleon transversely polarized along $\hat{x}$ ($|\boldsymbol{S}_T| = S_x$), the 2-dim density contains also the f-type gluon Sivers function:

$$x\rho_{p^\rightarrow}(x,p_x,p_y) = x\,f_1(x,\boldsymbol{p}_T^2) + x\frac{p_y}{M}\,f_{1T}^{\perp}(x,\boldsymbol{p}_T^2)\;. \tag{49}$$

The lower panels of Fig. 6 show such density in the same conditions and with the same notation as before. Since the nucleon is polarized along the $\hat{x}$ axis, the contour plot shows a distortion along the $\hat{y}$ axis. The asymmetry is clearly visible at $x = 10^{-1}$ (right panel), and it is emphasized by the ancillary 1-dim plot at $p_x = 0$. The distortion fades away for decreasing x, as shown in the left panel at $x = 10^{-3}$.

If we consider the gluon also in a linearly polarized state, then the 2-dim densities for various nucleon polarizations can become more complicated. The simplest case is for a unpolarized nucleon: the $x\rho^{\leftrightarrow}$ is a linear combination of the T-even gluon TMDs f_1 and $h_1^{\perp}$, and it has been studied in Ref. [30] (see lower panels of Fig.5 there). If the nucleon has a longitudinal polarization S_L, the 2-dim density $x\rho_{p^\odot}^{\leftrightarrow}$ is a linear combination of f_1, $h_1^{\perp}$ and $h_{1L}^{\perp}$. Finally, if the nucleon has transverse polarization S_T the 2-dim density $x\rho_{p^\rightarrow}^{\leftrightarrow}$ is a linear combination of f_1, $h_1^{\perp}$, $f_{1T}^{\perp}$, h_1 and $h_{1T}^{\perp}$, the latter two ones entering with different coefficients depending on the relative angle between the nucleon and gluon

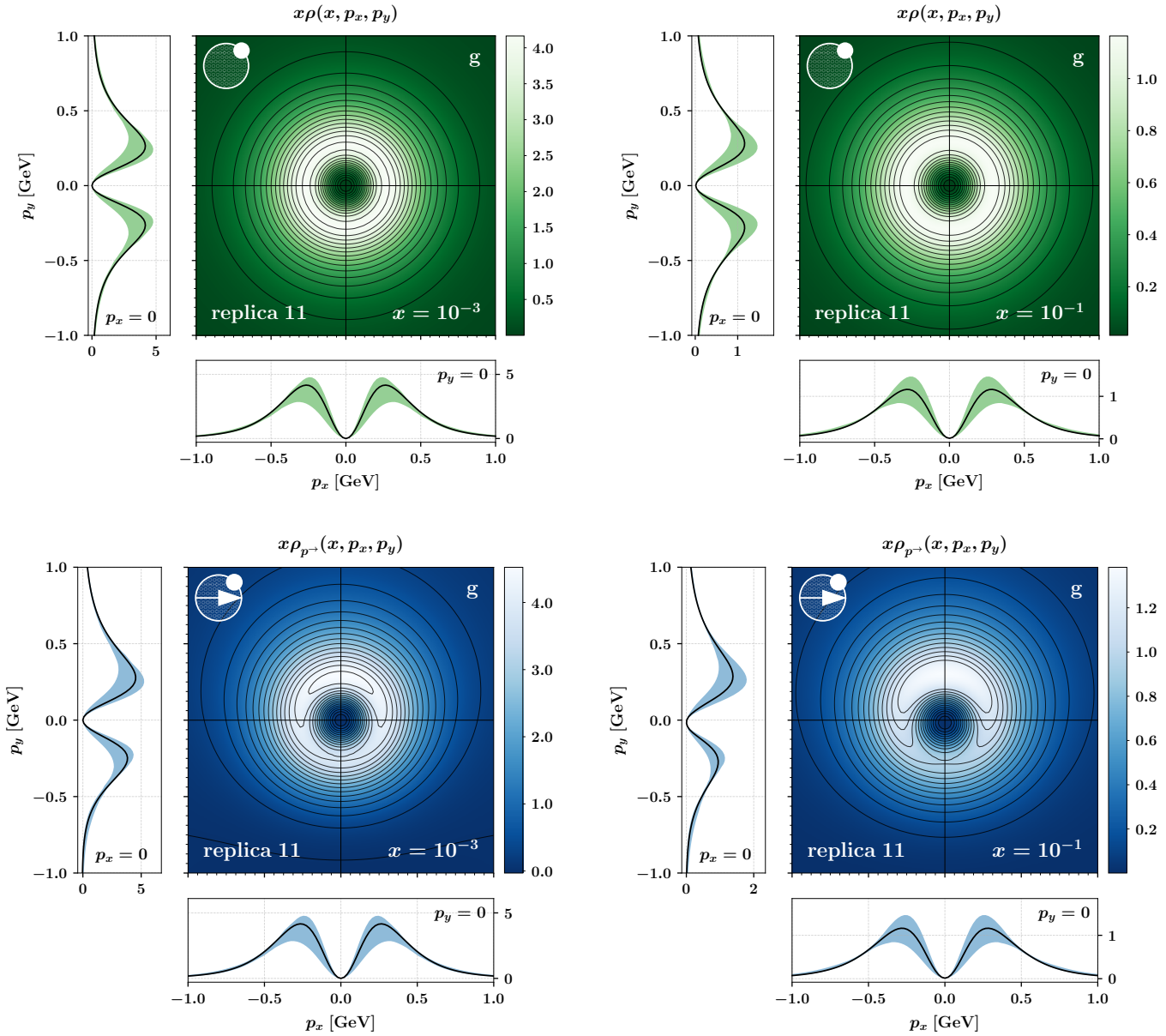


Figure 6. Unpolarized gluon density for a unpolarized (upper panels) or transversely polarized nucleon along $\hat{x}$ (lower panels) as a function of $\boldsymbol{p}_T$ at $Q_0 = 1.64$ GeV and at $x = 10^{-3}$ (left panels) and $x = 10^{-1}$ (right panels). The nucleon is virtually moving towards the reader. Results from replica 11 (see text). Ancillary 1-dim plots for the density at $p_y = 0$ and $p_x = 0$ with 68% uncertainty band. Solid black line for replica 11 (corresponding to contour plot).

polarizations. Apart for the case of unpolarized nucleon $x\rho^{\leftrightarrow}$, the other 2-dim densities are thus superpositions of three or more gluon TMDs, and their probabilistic interpretation becomes more involved.

Therefore, we prefer to isolate each T-odd TMD for linearly polarized gluons using the projectors discussed in Sec. II D, and we plot them for nucleon polarizations along specific directions.

We first select the nucleon longitudinally polarized along its direction of motion towards the reader, and the gluon linearly polarized along $\hat{x}$. Using the gluon-gluon correlator $\Phi^{xx}(S_L)$ in Eq.(54) of Ref. [4], the combination $\Phi^{xx}(S_L) - \Phi^{xx}(-S_L)$ isolates the term $h_{1L}^{\perp}\, p_x p_y/2M^2$. In the upper panels of Fig. 7, we show the contour plot for the $\boldsymbol{p}_T$-distribution of the f-type combination $h_{1L}^{\perp}\, p_x p_y/2M^2$ from replica 11 at $x = 10^{-3}$ (left) and $x = 10^{-1}$ (right) and at the scale $Q_0 = 1.64$ GeV, scaled by a factor 10^2. Because of the $p_x p_y$ weight, the contour plot shows symmetric oscillations along the $p_y = \pm p_x$ directions, emphasized in the 1-dim ancillary plots and becoming more sizeable at $x = 10^{-3}$. Sometimes in the literature, the function $h_{1L}^{\perp}$ is called "T-odd worm-gear" or "pseudo worm-gear" in

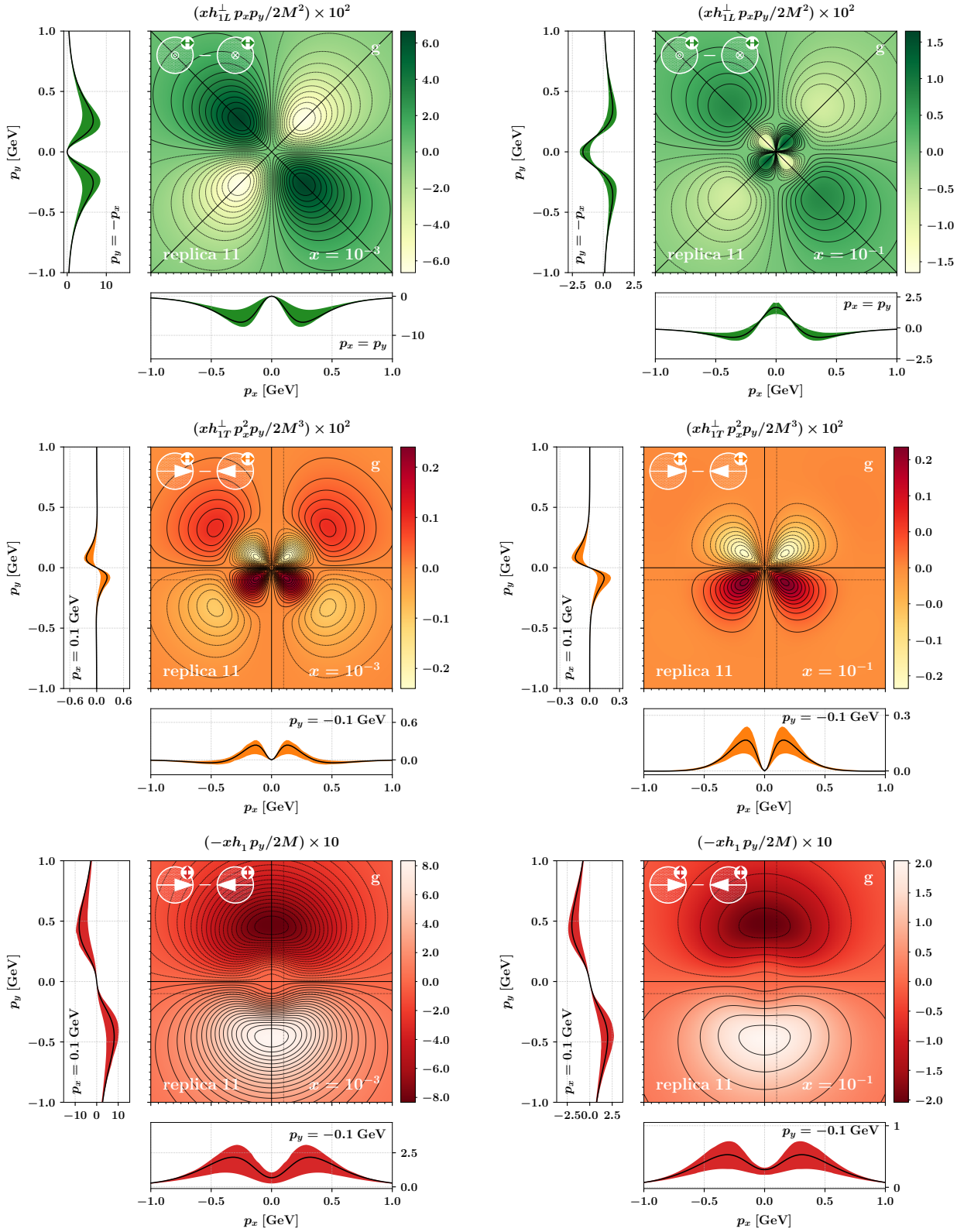


Figure 7. The 2-dim density for linearly polarized gluons in polarized nucleons as a function of $\boldsymbol{p}_T$ at $Q_0 = 1.64$ GeV and at $x = 10^{-3}$ (left panels) and $x = 10^{-1}$ (right panels). The nucleon is virtually moving towards the reader. Results from replica 11 (see text). Ancillary 1-dim plots for slices of the density at specific values of p_x or p_y, with 68% uncertainty band and solid black line for replica 11. Upper panels: gluon linear polarization along $\hat{x}$ and nucleon longitudinal polarization, proportional to $h_{1L}^{\perp}\, p_x p_y/2M^2$ scaled by 10^2. Central panels: gluon linear polarization and nucleon polarization along $\hat{x}$, proportional to $h_{1T}^{\perp}\, p_x^2 p_y/2M^3$ scaled by 10^2. Lower panels: gluon linear polarization along $\hat{y}$ and nucleon polarization along $\hat{x}$, proportional to $-h_1\, p_y/2M$ scaled by 10.

analogy with the corresponding quark function. However, we think that this nomenclature does not capture the main characteristics of this function as it emerges from the upper panels of Fig. 7. Since the nucleon is spinning around a direction pointing towards the reader, because of the displayed quadrupolar shape we propose for $h_{1L}^{\perp}$ the name of "propeller" function.

If we keep the gluon linearly polarized along $\hat{x}$ but we consider the combination $\Phi^{xx}(S_x) - \Phi^{xx}(-S_x)$, we can isolate the term $h_{1T}^{\perp}\, p_x^2 p_y/2M^3$. In the central panels of Fig. 7, we show the contour plot for the $\boldsymbol{p}_T$-distribution of the f-type combination $h_{1T}^{\perp}\, p_x^2 p_y/2M^3$, scaled by the factor 10^2 and with the same notations as before. The $p_x^2 p_y$ weight produces oscillations symmetric with respect to the $\hat{y}$ axis, emphasized in the 1-dim ancillary plots with slightly displaced slices at $p_y = -0.1$ GeV and $p_x = 0.1$ GeV. The T-odd gluon TMD $h_{1T}^{\perp}$ is sometimes referred to as "pretzelosity" in analogy with the quark case. As for $h_{1L}^{\perp}$, we think that this nomenclature is misleading. The peculiar shape of the contour plot in the lower panels of Fig. 7 suggests for $h_{1T}^{\perp}$ the name of "butterfly" function.

Finally, if we turn the gluon linear polarization along the $\hat{y}$ axis but keeping the nucleon polarization along $\hat{x}$, the combination $\Phi^{yy}(S_x) - \Phi^{yy}(-S_x)$ isolates the linearity function through the term $-h_1\, p_y/2M$. In the lower panels of Fig. 7, we show the contour plot for the $\boldsymbol{p}_T$-distribution of the f-type combination $-h_1\, p_y/2M$, scaled by the factor 10 and with the same notations as before. The p_y weight produces oscillations symmetric with respect to the $\hat{x}$ axis, emphasized in the 1-dim ancillary plots with slightly displaced slices at $p_y = -0.1$ GeV and $p_x = 0.1$ GeV.

V. SUMMARY AND OUTLOOK

In this paper, we have presented a model calculation of all four leading-twist T-odd gluon TMDs within a spectator model approach, providing insights into the complex interplay among gluon transverse momentum, gluon polarization, and nucleon spin, and offering a detailed (model-dependent) picture of the distribution of gluons in the nucleon. This paper completes our previous work [30], where we computed all the leading-twist T-even gluon TMDs in the same framework.

The model is based on the idea that a nucleon can split into a gluon and remainders that are treated as a single spectator fermion. This spectator mass is allowed to vary within a continuous range, described by a spectral function. Non vanishing T-odd structures are generated by the interference between the tree-level amplitude and an amplitude with final-state interactions, which in our model are approximated as a single-gluon exchange between the gluon and the spectator. The structure of interaction vertices reflects the nature of the involved particles. Since the spectator has spin-$\frac{1}{2}$, the vertices are modeled resembling the free nucleon electromagnetic current, replacing the Dirac and Pauli form factors with dipolar functions $g_1(p^2)$ and $g_2(p^2)$. For sake of simplicity, all model parameters have been kept the same as in our previous work on T-even gluon TMDs, where they were fixed by fitting the transverse-momentum-integrated gluon TMDs onto known parametrizations of the corresponding collinear unpolarized and helicity gluon PDFs at the lowest scale $Q_0 = 1.64$ GeV [30].

As it is well known, gluon TMDs have a more intricate dependence on the structure of the color flow (gauge link), which in turn introduces a dependence on the involved process. There are two main classes of gluon TMDs, the so-called Weizsäcker–Williams (WW) gluon TMDs (also called f-type) and the dipole gluon TMDs (also called d-type). In general, the two classes cannot be connected, as the WW and dipole gluon TMDs carry different physical information and appear in different processes. Due to the simplifying assumptions in our model, the differences between f-type and d-type gluon TMDs amount only to a calculable color factor: the size of the d-type gluon TMDs is 5/9 of the f-type ones.

We have provided analytical and numerical results for the f-type T-odd gluon TMDs using two versions of the model: a simpler version with a single form factor (g_1, taking $g_2 = 0$) for the nucleon-gluon-spectator and spectator-gluon-spectator vertices, and the full calculation with both g_1 and g_2 form factors. In the first case, we obtain nonvanishing results only for the Sivers ($f_{1T}^{\perp}$) and linearity (h_1) functions. They turn out to be much smaller than the T-even unpolarized TMD (f_1), and they show a decreasing trend for smaller values of x. In the full calculation, this trend is reversed and the size becomes comparable to f_1, suggesting that sizeable asymmetries generated by such functions could be measurable at small x. Moreover, we obtain non vanishing results also for the other two T-odd gluon TMDs: the $h_{1L}^{\perp}$ (which we name "propeller") and the $h_{1T}^{\perp}$ (which we name "butterfly"). However, both functions have a very small size, particularly the butterfly function, casting some doubts on the actual possibility of ever extracting them from measured spin asymmetries. We computed the T-odd f-type gluon TMDs also in the quark-target model. Only the Sivers and linearity functions are different from zero, and their expression matches known results in the literature.

As a final remark, the magnitude of the T-odd gluon TMDs crucially depends on the model parameters. For sake of simplicity, in this paper we have taken them equal to the model parameters of the T-even gluon TMDs [30].

However, our model is flexible enough to account for different couplings and different color structures in the interaction vertices, such that the differences between f-type and d-type gluon TMDs would not amount to a simple color factor. Only future data from the Electron-Ion Collider [17, 18, 69–71] and new-generation machines [72–77] will help us to overcome this limitations, and explore also the intriguing connections between our polarized gluon TMDs at small-x and the small-x unintegrated gluon density within a hybrid high-energy and collinear factorization framework (see, e.g., Refs. [78–80]).

ACKNOWLEDGMENTS

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 647981, 3DSPIN), by the Italian MIUR under the FARE program (code n. R16XKPHL3N, 3DGLUE), by the Atracción de Talento Grant (n. 2022-T1/TIC-24176), and by the European Union "Next Generation EU" program through the Italian PRIN 2022 grant n. 20225ZHA7W. F.G.C. thanks the Università degli Studi di Pavia for the warm hospitality.

Appendix A: Master integrals

Here below, we list the master integrals involved in the expressions of our T-odd f-type gluon TMDs. We first define

$$T_h(|\boldsymbol{p}_T|) = \tanh^{-1}\sqrt{\frac{\boldsymbol{p}_T^2}{\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)}}\,. \tag{A1}$$

Then, we have

$$\begin{aligned}\mathcal{D}_1(p) &= \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2}\\ &= \frac{1}{8\pi P^+}\left[2\frac{1 - 2L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} + 16\frac{\boldsymbol{p}_T^2 + L_X^2(\Lambda_X^2)}{|\boldsymbol{p}_T|^3\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right]\,,\end{aligned} \tag{A2}$$

$$\mathcal{D}_2(p) = \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{\boldsymbol{l}_T\cdot\boldsymbol{p}_T}{\boldsymbol{p}_T^2}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2} \equiv -\frac{1}{2}\,\mathcal{D}_1(p) \tag{A3}$$

$$\begin{aligned}\mathcal{D}_3(p) &= \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{(\boldsymbol{l}_T\cdot\boldsymbol{p}_T)^2}{|\boldsymbol{p}_T|^4}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2}\\ &= \frac{1}{8\pi P^+}\left[\frac{1 + 4L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2 + 12L_X^4(\Lambda_X^2)/\boldsymbol{p}_T^4}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} - 24\frac{[\boldsymbol{p}_T^2 + 2L_X^2(\Lambda_X^2)]\,L_X^2(\Lambda_X^2)}{|\boldsymbol{p}_T|^5\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right]\,,\end{aligned} \tag{A4}$$

$$\begin{aligned}\mathcal{D}_4(p) &= \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{\boldsymbol{l}_T^2}{\boldsymbol{p}_T^2}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2}\\ &= \frac{1}{8\pi P^+}\left[\frac{1 + 2L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2 + 4L_X^4(\Lambda_X^2)/\boldsymbol{p}_T^4}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} + 4\frac{\boldsymbol{p}_T^4 - 4L_X^4(\Lambda_X^2)}{|\boldsymbol{p}_T|^5\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right]\,,\end{aligned} \tag{A5}$$

$$\begin{aligned}\mathcal{D}_5(p) &= \frac{1}{2P^+}\int\frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{\boldsymbol{l}_T\cdot\boldsymbol{p}_T}{\boldsymbol{p}_T^2}\frac{\boldsymbol{l}_T^2}{\boldsymbol{p}_T^2}\frac{1}{[\boldsymbol{l}_T^2 + L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2 + L_X^2(\Lambda_X^2)]^2}\\ &= -\frac{1}{8\pi P^+}\left[\frac{1 + 6L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2 + 14L_X^4(\Lambda_X^2)/\boldsymbol{p}_T^4}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^2} - 2\frac{\boldsymbol{p}_T^4 + 14L_X^2(\Lambda_X^2)\,\boldsymbol{p}_T^2 + 28L_X^4(\Lambda_X^2)}{|\boldsymbol{p}_T|^5\,[\boldsymbol{p}_T^2 + 4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right]\,,\end{aligned} \tag{A6}$$

$$\begin{aligned}\mathcal{D}_6(p) &= \frac{1}{2P^+}\int \frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{|\boldsymbol{l}_T|^4}{|\boldsymbol{p}_T|^4}\frac{1}{[\boldsymbol{l}_T^2+L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2+L_X^2(\Lambda_X^2)]^2}\\ &= \frac{1}{8\pi P^+}\left[\frac{[\boldsymbol{p}_T^2+2L_X^2(\Lambda_X^2)]\,[\boldsymbol{p}_T^4+8L_X^2(\Lambda_X^2)\,\boldsymbol{p}_T^2+22L_X^4(\Lambda_X^2)]}{L_X^2(\Lambda_X^2)\,\boldsymbol{p}_T^6\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^2}\right.\\ &\quad\left.-4\frac{\boldsymbol{p}_T^6+12L_X^2(\Lambda_X^2)\,\boldsymbol{p}_T^4+40L_X^4(\Lambda_X^2)\,\boldsymbol{p}_T^2+44L_X^6(\Lambda_X^2)}{|\boldsymbol{p}_T|^7\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right],\end{aligned} \tag{A7}$$

$$\begin{aligned}\mathcal{D}_7(p) &= \frac{1}{2P^+}\int \frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{(\boldsymbol{l}_T\cdot\boldsymbol{p}_T)^2}{|\boldsymbol{p}_T|^4}\frac{|\boldsymbol{l}_T|^2}{\boldsymbol{p}_T^2}\frac{1}{[\boldsymbol{l}_T^2+L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2+L_X^2(\Lambda_X^2)]^2}\\ &= \frac{1}{8\pi P^+}\left[\frac{|\boldsymbol{p}_T|^7+10L_X^2(\Lambda_X^2)\,|\boldsymbol{p}_T|^5+36L_X^4(\Lambda_X^2)\,|\boldsymbol{p}_T|^3+36L_X^6(\Lambda_X^2)\,|\boldsymbol{p}_T|}{L_X^2(\Lambda_X^2)\,|\boldsymbol{p}_T|^7\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^2}\right.\\ &\quad\left.-6\frac{[\boldsymbol{p}_T^2+2L_X^2(\Lambda_X^2)]^2\,[\boldsymbol{p}_T^2+6L_X^2(\Lambda_X^2)]}{|\boldsymbol{p}_T|^7\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right],\end{aligned} \tag{A8}$$

$$\begin{aligned}\mathcal{D}_8(p) &= \frac{1}{2P^+}\int \frac{d^2\boldsymbol{l}_T}{(2\pi)^2}\frac{(\boldsymbol{l}_T\cdot\boldsymbol{p}_T)^3}{|\boldsymbol{p}_T|^6}\frac{1}{[\boldsymbol{l}_T^2+L_X^2(\Lambda_X^2)]^2}\frac{1}{[(\boldsymbol{l}_T+\boldsymbol{p}_T)^2+L_X^2(\Lambda_X^2)]^2}\\ &= -\frac{1}{8\pi P^+}\left[\frac{1+7L_X^2(\Lambda_X^2)/\boldsymbol{p}_T^2+18L_X^4(\Lambda_X^2)/\boldsymbol{p}_T^4}{L_X^2(\Lambda_X^2)\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^2}-4\frac{\boldsymbol{p}_T^4+10L_X^2(\Lambda_X^2)\,\boldsymbol{p}_T^2+18L_X^4(\Lambda_X^2)}{|\boldsymbol{p}_T|^5\,[\boldsymbol{p}_T^2+4L_X^2(\Lambda_X^2)]^{5/2}}\,T_h(|\boldsymbol{p}_T|)\right].\end{aligned} \tag{A9}$$

Appendix B: Full calculation of gluon TMDs: The coefficients $C_{ijk}^{[F],l}$

In the following, we list the final expressions of the $C_{ijk}^{[F],l}$ coefficients in Eqs. (46) and (47) for each T-odd gluon TMD F and for $l=1,..,8,\ i,j,k=1,2$. We note that the $C_{111}^{[F],2}$ coefficients for the f-type gluon Sivers ($[F]\equiv[f_{1T}^{\perp}]$) and linearity ($[F]\equiv[h_1]$) have already been derived when discussing the computation of these T-odd gluon TMDs in the g_1-vertex approximation (see Eqs. (32) and (42), respectively).

1. Sivers function $f_{1T}^{\perp}$

In Tabs. III-VI, we list the coefficients $C_{ijk}^{[f_{1T}^{\perp}],l}$ for $i,j,k=1,2$ and $l=1,..,8$.

Table III. Coefficients functions of the f-type gluon Sivers TMD for $ijk = \{111, 112\}$.

$C_{ijk}^{\left[f_{1T}^{\perp}\right],l}$	$ijk = 111$	$ijk = 112$
$l = 1$	0	$\frac{12}{x}\left[2x\left(M^2(2-x)x - (M-M_X)^2\right) + 2\boldsymbol{p}_T^2(2-x)\right]$
$l = 2$	$-48M(1-x)\left[M_X - M(1-x)\right]$	$-\frac{12}{x}\left[x\left(M(1-x)+M_X\right)\left(M(1-x)^2 + M_X(1+x)\right) - \boldsymbol{p}_T^2(x^2-5x+8)\right]$
$l = 3$	0	$\frac{24}{x}\,\boldsymbol{p}_T^2\,(x^2-2x+2)$
$l = 4$	0	$12\,\boldsymbol{p}_T^2\,(1-x)$
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table IV. Coefficients functions of the f-type gluon Sivers TMD for $ijk = \{121, 122\}$.

$C_{ijk}^{\left[f_{1T}^{\perp}\right],l}$	$ijk = 121$	$ijk = 122$
$l = 1$	0	0
$l = 2$	$\frac{6}{x}\left[-M^2(2-x)\,x\,(1-x)^2 + M_X^2\,(2-x)\,x + \boldsymbol{p}_T^2\,(x^2-10x+8)\right]$	$3x\,\frac{M-M_X}{M}\left[M^2(1-x)^2 - M_X^2 + \boldsymbol{p}_T^2\right]$
$l = 3$	$\frac{48\boldsymbol{p}_T^2}{x}\,(1-x)$	0
$l = 4$	$-6\boldsymbol{p}_T^2\,(2-x)$	$\frac{3\boldsymbol{p}_T^2}{M}\left[M\,(4-3x) - M_X\,(x+4)\right]$
$l = 5$	0	$\frac{12\boldsymbol{p}_T^2}{M}\left[M(1-x) - M_X\right]$
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table V. Coefficients functions of the f-type gluon Sivers TMD for $ijk = \{211, 212\}$.

$C_{ijk}^{\left[f_{1T}^{\perp}\right],l}$	$ijk = 211$	$ijk = 212$
$l = 1$	$-\frac{24}{x}\left[x\,\left(M_X^2 - M^2(1-x)^2\right) + (2-x)\,\boldsymbol{p}_T^2\right]$	0
$l = 2$	$-\frac{6}{(1-x)x}\,\Big[x^2\,\left(M_X - M(1-x)\right)\,\left(M\,(1-x)\,x + M_X\,(4-3x)\right)$ $-\boldsymbol{p}_T^2\,(x^3 - 8x + 8)\Big]$	$-\frac{3}{M(1-x)}\,\Big[M^3\,(2-x)\,(1-x)^2$ $+3M^2\,M_X\,(1-x)^2\,x^2$ $-M\,\left(M_X^2\,x\,(3x^2 - 4x + 2) + \boldsymbol{p}_T^2\,(x^2 - 6x + 4)\right)$ $+M_X^3\,x^2 + M_X\,\boldsymbol{p}_T^2\,(3x^2 - 8x + 4)\Big]$
$l = 3$	0	$\frac{12\boldsymbol{p}_T^2}{M}\,[M_X\,(1-x) - M]$
$l = 4$	$6\boldsymbol{p}_T^2\,\frac{(2-x)^2}{(1-x)}$	$\frac{3x\boldsymbol{p}_T^2}{M(1-x)}\,[M\,(2-x) - M_X\,x]$
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table VI. Coefficients functions of the f-type gluon Sivers TMD for $ijk = \{221, 222\}$.

$C_{ijk}^{\left[f_{1T}^{\perp}\right],l}$	$ijk = 221$	$ijk = 222$
$l = 1$	0	0
$l = 2$	$3x\,\frac{M^2(1-x)^2 - M_X^2 + \boldsymbol{p}_T^2}{2M(1-x)}$ $\times\left[M\,(1-x)\,(7-2x) - M_X\,(7-3x)\right]$	$\frac{3}{8M^2(1-x)x}\,\Big[x^2\,\left(M^2\,(1-x)^2 - M_X^2\right)$ $\times\left(M^2\,(2x^3+x^2-3) + 4M\,M_X\,(3-x)\,x + M_X^2\,(3-2x)\right)$ $-4x\boldsymbol{p}_T^2\,\left(M^2\,(1-x)\,(x^3-5x+6) - M\,M_X\,(3-x)x^2\right.$ $\left.+M_X^2\,(x^2+7x-6)\right)$ $+\boldsymbol{p}_T^4\,(2x^3-3x^2-24x+32)\Big]$
$l = 3$	$\frac{12\boldsymbol{p}_T^2}{M}\,\left[M_X - M\,(1-x)\right]$	$-\frac{3\boldsymbol{p}_T^2}{4M^2\,(1-x)\,x}\,\Big[M^2\,(1-x)\,x\,(2x^3+5x^2-13x+4)$ $+8M\,M_X\,(1-x)\,x^2 + M_X^2\,x\,(2x^2+11x-4)$ $-\boldsymbol{p}_T^2\,(2x^3+3x^2-28x+32)\Big]$
$l = 4$	$\frac{3x\,\boldsymbol{p}_T^2}{2M\,(1-x)}\,\left[M\,(1-x)\,(7-2x) - M_X\,(7-3x)\right]$	$-\frac{3\boldsymbol{p}_T^2}{8M^2(1-x)}\,\Big[M^2\,(1-x)\,(2x^3-5x^2-x+16)$ $+4M\,M_X\,x\,(x^2+x-4) + M_X^2\,x\,(2x^2+13x-6)$ $-\boldsymbol{p}_T^2\,(2x^2-x+8)\Big]$
$l = 5$	0	$\frac{3\boldsymbol{p}_T^2}{8M^2(1-x)}\,\Big[M^2\,(1-x)^2\,(2x^2-x-8)$ $+16M\,M_X\,(1-x)\,x + M_X^2\,(2x^2+15x-8)$ $+\boldsymbol{p}_T^2\,(6x^2-11x+32)\Big]$
$l = 6$	0	$\frac{3\boldsymbol{p}_T^4}{8M^2}\,\frac{2x^2-x+8}{1-x}$
$l = 7$	0	0
$l = 8$	0	$\frac{6\boldsymbol{p}_T^4}{M^2}\,\frac{2-x}{x}$

2. Linearity function h_1

In Tabs. VII-X, we list the coefficients $C_{ijk}^{[h_1],l}$ for $i, j, k = 1, 2$ and $l = 1, .., 8$.

Table VII. Coefficients functions of the f-type gluon linearity for $ijk = \{111, 112\}$.

$C_{ijk}^{[h_1],l}$	$ijk = 111$	$ijk = 112$
$l = 1$	0	$\frac{6}{x}\left[x\,(4-3x)\,\left(M^2\,(1-x)^2 - M_X^2\right) + \boldsymbol{p}_T^2\,(3x^2-12x+8)\right]$
$l = 2$	$-96M\,[M_X - M\,(1-x)]$	$-\frac{6}{x}\left[x\,\left(M^2\,(3x^3-6x^2+11x-8) + 8M\,M_X\,x - M_X^2\,(8-3x)\right) - \boldsymbol{p}_T^2\,(3x^2-8x+16)\right]$
$l = 3$	0	$48\boldsymbol{p}_T^2\;\frac{1-x}{x}$
$l = 4$	0	$48\boldsymbol{p}_T^2$
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table VIII. Coefficients functions of the f-type gluon linearity for $ijk = \{121, 122\}$.

$C_{ijk}^{[h_1],l}$	$ijk = 121$	$ijk = 122$
$l = 1$	0	0
$l = 2$	$\frac{6}{x}\left[x\,\left(M^2\,(1-x)^2\,(2+3x) - 16M\,M_X\,(1-x)\,x\right.\right.$ $\left.\left. -M_X^2\,(2-13x)\right) + \boldsymbol{p}_T^2\,(5x^2-2x+8)\right]$	$\frac{3}{M}\left[M^2\,M_X\,x\,(7x^2+2x-9) + M\,x\,\left(5M^2\,(1-x)^2 + 11\boldsymbol{p}_T^2\right)\right.$ $\left. -M\,M_X^2\,x\,(5-16x) + 9M_X^3\,x + M_X\,\boldsymbol{p}_T^2\,(8+x)\right]$
$l = 3$	$\frac{48\boldsymbol{p}_T^2}{x}$	$\frac{6\boldsymbol{p}_T^2}{M}\,[4M\,x + M_X\,(10+x)]$
$l = 4$	0	$\frac{6\boldsymbol{p}_T^2}{M}[4M(1-x) + M_X(2+x)]$
$l = 5$	0	$\frac{6\boldsymbol{p}_T^2}{M}\,[4M\,(1-x) + M_X\,(2-3x)]$
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	$-\frac{18\boldsymbol{p}_T^2 M_X}{M}\,(2-x)$

Table IX. Coefficients functions of the f-type gluon linearity for $ijk = \{211, 212\}$.

$C_{ijk}^{[h_1],l}$	$ijk = 211$	$ijk = 212$
$l = 1$	$-\frac{6}{x}\left[\left(M^2(1-x)^2 - M_X^2\right)(4-3x)\,x + \boldsymbol{p}_T^2(3x^3-12x+8)\right]$	$-9x\left[M^2(1-x)^2 - M_X^2 - \boldsymbol{p}_T^2\right]$
$l = 2$	$-\frac{6}{(1-x)x}\left[x^2\left(M^2(1-x)^2(7-2x) - 8M\,M_X(1-x) + M_X^2(1+2x)\right) + \boldsymbol{p}_T^2(2x^3-7x^2+16x-8)\right]$	$-\frac{3}{M(1-x)}\left[M^3(1-x)^2\,x(9-4x) + M^2\,M_X(x^3+5x-6) + 3M\,M_X^2(4-3x)\,x + M\,\boldsymbol{p}_T^2(4x^2-9x+8) + M_X(M_X^2+\boldsymbol{p}_T^2)(6-x)\right]$
$l = 3$	0	$-24\boldsymbol{p}_T^2$
$l = 4$	0	0
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table X. Coefficients functions of the f-type gluon linearity for $ijk = \{221, 222\}$.

$C_{ijk}^{[h_1],l}$	$ijk = 221$	$ijk = 222$
$l = 1$	0	0
$l = 2$	$\frac{3}{M(1-x)}\Big[2M_X\Big(M^2 x(3+2x)(1-x)^2$ $+\boldsymbol{p}_T^2(2x^2+5x-4)\Big)$ $+M(1-x)x\left(-M^2(6-x)(1-x)^2-\boldsymbol{p}_T^2(10+x)\right)$ $+M M_X^2(17x^2-23x+6)x-6M_X^3(1-2x)x\Big]$	$-\frac{3}{4M^2(1-x)x}\Big[\boldsymbol{p}_T^2 x\Big(M^2(1-x)(6x^3-31x^2+6x-4)$ $+2M M_X(x^2+21x-16)x+2M_X^2(2+x)(1+3x)\Big)$ $-x^2(M(1-x)-M_X)\left(M^3(x^2+x-2)^2-M_X^3(4-17x)\right.$ $\left.+M^2 M_X(3x^2-18x+4)(1-x)+M M_X(17x^2-x-4)\right)$ $+\boldsymbol{p}_T^4(5x^3-30x^2+52x-16)\Big]$
$l = 3$	$-12\boldsymbol{p}_T^2(2-x)$	$-\frac{3\boldsymbol{p}_T^2}{2M^2(1-x)x}\Big[M^2(2x^3-15x^2+8x+2)(1-x)x$ $-4M M_X(1-x)x^2+M_X^2(4x^2+5x-2)$ $+\boldsymbol{p}_T^2(2x^3-9x^2+30x-16)\Big]$
$l = 4$	$-12x\,\boldsymbol{p}_T^2$	$-\frac{3\boldsymbol{p}_T^2}{4M^2(1-x)}\Big[M^2(3x^3-7x^2+2x-8)(1-x)$ $-24M M_X(1-x)x-M_X^2(x^2+8x-8)$ $+\boldsymbol{p}_T^2(3x^2-12x+8)\Big]$
$l = 5$	0	$\frac{3\boldsymbol{p}_T^2}{2M^2(1-x)}\Big[M^2(1-x)^2(x^2-7x+2)+8M M_X(1-x)x$ $-M_X^2(7x^2-11x+2)-\boldsymbol{p}_T^2(x^2-x-6)\Big]$
$l = 6$	0	$\frac{3\boldsymbol{p}_T^4}{M^2}\frac{x}{1-x}$
$l = 7$	0	$-\frac{3\boldsymbol{p}_T^4}{M^2}\frac{x}{1-x}$
$l = 8$	0	$\frac{3\boldsymbol{p}_T^2}{2M^2(1-x)x}\Big[3(2-x)x\left(M_X^2(1-2x)-M^2(1-x)^2\right)$ $+\boldsymbol{p}_T^2(3x^2-18x+8)\Big]$

3. Propeller function $h_{1L}^{\perp}$

In Tabs. XI-XIV, we list the coefficients $C_{ijk}^{\left[h_{1L}^{\perp}\right],l}$ for $i, j, k = 1, 2$ and $l = 1, .., 8$.

Table XI. Coefficients functions of the f-type gluon propeller for $ijk = \{111, 112\}$.

$C_{ijk}^{\left[h_{1L}^{\perp}\right],l}$	$ijk = 111$	$ijk = 112$
$l = 1$	0	$96(1-x)\,M^2$
$l = 2$	0	$48M\,[3M\,(1-x) - M_X]$
$l = 3$	0	$96M\,[M\,(1-x) - M_X]$
$l = 4$	0	$-48M\,[M\,(1-x) - M_X]$
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table XII. Coefficients functions of the f-type gluon propeller for $ijk = \{121, 122\}$.

$C_{ijk}^{\left[h_{1L}^{\perp}\right],l}$	$ijk = 121$	$ijk = 122$
$l = 1$	0	0
$l = 2$	$96(1-x)\,M\,[M(1-x) - M_X]$	$-24\,\left[M\,(1-x)\,\left(M\,(1-x) - 2M_X\,x\right) - M_X^2\,(1-2x) + \boldsymbol{p}_T^2\right]$
$l = 3$	$96(1-x)\,M\,[M\,(1-x) - M_X]$	$\frac{48}{x}\,\left[M\,M_X\,(1-x)x^2 - M_X^2\,x^2 + \boldsymbol{p}_T^2\,(1-x)\right]$
$l = 4$	$-48(1-x)\,M\,[M\,(1-x) - M_X]$	$-\frac{24}{x}\,\left[M\,M_X\,(1-x)x^2 - M_X^2 x^2 + 2\boldsymbol{p}_T^2\right]$
$l = 5$	0	$-24\boldsymbol{p}_T^2$
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table XIII. Coefficients functions of the f-type gluon propeller for $ijk = \{211, 212\}$.

$C_{ijk}^{\left[h_{1L}^{\perp}\right],l}$	$ijk = 211$	$ijk = 212$
$l = 1$	$-96(1-x)\,M^2$	0
$l = 2$	$-48M\left[M\,(1-x) + M_X\right]$	$-48\left[M\,(1-x) - M_X\right]$
$l = 3$	0	$-\frac{48}{x}\left[x\,(M+M_X)\,\left(M\,(1-x) - M_X\right) - \boldsymbol{p}_T^2\right]$
$l = 4$	0	$\frac{24}{x}\left[x\,(M+M_X)\,\left(M\,(1-x) - M_X\right) - 2\boldsymbol{p}_T^2\right]$
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table XIV. Coefficients functions of the f-type gluon propeller for $ijk = \{221, 222\}$.

$C_{ijk}^{\left[h_{1L}^{\perp}\right],l}$	$ijk = 221$	$ijk = 222$
$l = 1$	0	0
$l = 2$	$-24\,\left[M^2\,(1-x)^2 - 2M\,M_X\,(1-x)\,x - M_X^2\,(1-2x) + \boldsymbol{p}_T^2\right]$	$\frac{24}{M}\,(M_X+M)\,\left[x\,M_X\,\left(M_X - M\,(1-x)\right) + \boldsymbol{p}_T^2\right]$
$l = 3$	$\frac{48(1-x)}{x}\,\left[x\,(M+M_X)\,\left(M_X - M\,(1-x)\right) + \boldsymbol{p}_T^2\right]$	$\frac{24M_X}{M}\,\left[x\,(M_X+M)\,\left(M_X - M\,(1-x)\right) + 2\boldsymbol{p}_T^2\right]$
$l = 4$	$-\frac{24}{x}\,\left[(1-x)\,x\,(M+M_X)\,\left(M_X - M(1-x)\right) + \boldsymbol{p}_T^2\,(2-x)\right]$	$-\frac{12}{M}\,\left[x^2\,M_X\,M\,(M_X+M) + x\,M_X\,(M_X^2 - M^2) + (M_X - 3M)\,\boldsymbol{p}_T^2\right]$
$l = 5$	0	$\frac{12\boldsymbol{p}_T^2}{M}(M+M_X)$
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

4. Butterfly function $h_{1T}^{\perp}$

In Tabs. XV-XVIII, we list the coefficients $C_{ijk}^{\left[h_{1T}^{\perp}\right],l}$ for $i, j, k = 1, 2$ and $l = 1, .., 8$.

Table XV. Coefficients functions of the f-type gluon butterfly for $ijk = \{111, 112\}$.

$C_{ijk}^{\left[h_{1T}^{\perp}\right],l}$	$ijk = 111$	$ijk = 112$
$l = 1$	0	$-\frac{M^2}{2\boldsymbol{p}_T^2(1-x)x}\left[(1-x)^2\left(M^2\,x\,(x^2-5x+4)-2M_X^2\,x+2\boldsymbol{p}_T^2\,(4-x)\right)+x\,(x^2+x-2)\,(M_X^2+\boldsymbol{p}_T^2)\right]$
$l = 2$	0	$-\frac{M^2}{2\boldsymbol{p}_T^2(1-x)x}\left[(1-x)^2\left(M^2\,x\,(x^2-5x+4)-2M_X^2\,x+2\boldsymbol{p}_T^2\,(8-x)\right)+x\,(x^2+x-2)\,(M_X^2+\boldsymbol{p}_T^2)\right]$
$l = 3$	0	$-4M^2\frac{1-x}{x}$
$l = 4$	0	0
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table XVI. Coefficients functions of the f-type gluon butterfly for $ijk = \{121, 122\}$.

$C_{ijk}^{\left[h_{1T}^{\perp}\right],l}$	$ijk = 121$	$ijk = 122$
$l = 1$	0	0
$l = 2$	$\frac{M^2}{2x\boldsymbol{p}_T^2}\left[(2-x)\,x\,(M_X^2+\boldsymbol{p}_T^2)\right.$ $\left.-(1-x)^2\,\left(M^2\,(2-x)\,x+8\boldsymbol{p}_T^2\right)\right]$	$\frac{M}{4\boldsymbol{p}_T^2}\left[M^3\,x(1-x)^2+M\left(\boldsymbol{p}_T^2\,(3x^2-8x+4)-M_X^2\,x^2\right)\right.$ $-(1-x)\,\left(M^2\,M_X\,(1-x)\,x+M\,M_X^2\,x+M\,\boldsymbol{p}_T^2\,(4-3x)\right.$ $\left.\left.+4M_X\,\boldsymbol{p}_T^2\right)+M_X^3\,x-M_X\,\boldsymbol{p}_T^2\,(4-5x)\right]$
$l = 3$	$-\frac{4M^2(1-x)^2}{x}$	$-\frac{2-3x}{2}MM_X$
$l = 4$	0	$-\frac{2-x}{2}M\,M_X$
$l = 5$	0	$-\frac{2-x}{2}M\,M_X$
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	$\frac{2-x}{2}M\,M_X$

Table XVII. Coefficients functions of the f-type gluon butterfly for $ijk = \{211, 212\}$.

$C_{ijk}^{\left[h_1^\perp T\right],l}$	$ijk = 211$	$ijk = 212$
$l = 1$	$\frac{M^2}{2x\boldsymbol{p}_T^2}\left[(1-x)\left(4M^2(1-x)^2x + 4\boldsymbol{p}_T^2(2-x)\right)\right.$ $\left.+3M^2x^2(1-x)^2 - x\left(M_X^2(4-x) + 3\boldsymbol{p}_T^2x\right)\right]$	$\frac{xM^2}{4\boldsymbol{p}_T^2}\left[M^2(1-x)^2 - M_X^2 - \boldsymbol{p}_T^2\right]$
$l = 2$	$-\frac{M^2}{\boldsymbol{p}_T^2(1-x)x}\left[(1-x)^2\left(M^2x^2 - 4\boldsymbol{p}_T^2(2-x)\right)\right.$ $\left.-x\left(M_X^2x + \boldsymbol{p}_T^2(4x^2 - 7x + 4)\right)\right]$	$\frac{M}{4(1-x)\boldsymbol{p}_T^2}\left[(1-x)^2\left(M^3x(2x^2 - 4x + 3) + 2M\boldsymbol{p}_T^2(2-x)\right)\right.$ $+2M^3x^2(1-x)^3 - M_X(2-x)(1-x)\left(2\boldsymbol{p}_T^2 + M^2x(1-x)\right)$ $-M M_X^2x(3-2x) + M\boldsymbol{p}_T^2(2x^3 - 6x^2 + 7x - 4)$ $\left.+M_X^3(2-x)x + M_X\boldsymbol{p}_T^2(x^2 + 4x - 4)\right]$
$l = 3$	0	$-2MM_X$
$l = 4$	0	0
$l = 5$	0	0
$l = 6$	0	0
$l = 7$	0	0
$l = 8$	0	0

Table XVIII. Coefficients functions of the f-type gluon butterfly for $ijk = \{221, 222\}$.

$C_{ijk}^{\left[h_{1T}^{\perp}\right],l}$	$ijk = 221$	$ijk = 222$
$l=1$	0	0
$l=2$	$\frac{M}{4(1-x)\boldsymbol{p}_T^2}\Big[M^3\,x\,(3-x)\,(1-x)^3-(1-x)\Big(2M\,M_X^2\,x^2$ $-2M\,\boldsymbol{p}_T^2\,(x^2-4x+2)-3M_X^3\,x+M_X\,\boldsymbol{p}_T^2\,(4-7x)\Big)$ $-M\,(1-x)^2\,\big(M_X\,x\,(3M_X+M\,(3-x))+\boldsymbol{p}_T^2\,(4-x)\big)$ $+2M_X^3\,x^2-2M_X\,\boldsymbol{p}_T^2\,(x^2-4x+2)\Big]$	$-\frac{1}{16(1-x)x\boldsymbol{p}_T^2}\Big[\boldsymbol{p}_T^2\,x\,\Big(M^2\,(1-x)\,(2x^3-3x^2+8x-4)$ $-2M\,M_X\,(3-x)\,x^2+2M_X^2\,(9x^2-10x+2)\Big)$ $+x^2\,\big(M^2\,(1-x)^2-M_X^2\big)\,\big(-M^2\,(1-x^3)$ $+2M\,M_X\,(3-x)\,x+M_X^2\,(1-x)\big)$ $+\boldsymbol{p}_T^4\,(x^3-3x^2+20x-16)\Big]$
$l=3$	$-2MM_X$	$-\frac{1}{8x}\Big[M^2\,x\,(2x^2-x-2)$ $-M_X^2\,x\,(1+3x)-2\boldsymbol{p}_T^2\,(8-3x)\Big]$
$l=4$	0	$\frac{1}{16(1-x)}\Big[M^2\,(1-x)\,(x^3-x^2+6x-8)$ $+M_X^2\,(5x^2-12x+8)+\boldsymbol{p}_T^2\,(x^2+8x-8)\Big]$
$l=5$	0	$-\frac{1}{16(1-x)}\Big[(1-x)\,\Big(x\,\Big(M^2\,(1-x)^2-M_X^2\Big)$ $+\boldsymbol{p}_T^2\,(8-x)\Big)+(2-x)\,\big(M^2\,(2-x)\,(1-x)^2$ $-M_X^2\,(2-5x)+\boldsymbol{p}_T^2\,(2+x)\big)\Big]$
$l=6$	0	0
$l=7$	0	0
$l=8$	0	$\frac{2-x}{8(1-x)x}\Big[M^2\,(1-x)^2\,x-M_X\,(1-2x)\,x$ $-\boldsymbol{p}_T^2\,(4-5x)\Big]$

REFERENCES

[1] R. Angeles-Martinez *et al.*, Acta Phys. Polon. B **46**, 2501 (2015), arXiv:1507.05267 [hep-ph].
[2] P. J. Mulders and J. Rodrigues, Phys. Rev. **D63**, 094021 (2001), arXiv:hep-ph/0009343 [hep-ph].
[3] D. Boer, C. Lorcé, C. Pisano, and J. Zhou, Adv. High Energy Phys. **2015**, 371396 (2015), arXiv:1504.04332 [hep-ph].
[4] S. Meissner, A. Metz, and K. Goeke, Phys. Rev. **D76**, 034002 (2007), arXiv:hep-ph/0703176 [HEP-PH].
[5] J.-P. Lansberg, C. Pisano, F. Scarpa, and M. Schlegel, Phys. Lett. B **784**, 217 (2018), [Erratum: Phys.Lett.B 791, 420–421 (2019)], arXiv:1710.01684 [hep-ph].
[6] U. D'Alesio, F. Murgia, C. Pisano, and P. Taels, Phys. Rev. **D96**, 036011 (2017), arXiv:1705.04169 [hep-ph].
[7] U. D'Alesio, C. Flore, F. Murgia, C. Pisano, and P. Taels, Phys. Rev. D **99**, 036013 (2019), arXiv:1811.02970 [hep-ph].
[8] U. D'Alesio, L. Maxia, F. Murgia, C. Pisano, and S. Rajesh, Phys. Rev. D **102**, 094011 (2020), arXiv:2007.03353 [hep-ph].
[9] C. Adolph *et al.* (COMPASS), Phys. Lett. B **772**, 854 (2017), arXiv:1701.02453 [hep-ex].
[10] F. Yuan, Phys. Rev. D **78**, 014024 (2008), arXiv:0801.4357 [hep-ph].
[11] D. Boer, P. J. Mulders, C. Pisano, and J. Zhou, JHEP **08**, 001 (2016), arXiv:1605.07934 [hep-ph].
[12] L. Zheng, E. C. Aschenauer, J. H. Lee, B.-W. Xiao, and Z.-B. Yin, Phys. Rev. **D98**, 034011 (2018), arXiv:1805.05290 [hep-ph].
[13] R. M. Godbole, A. Kaushik, A. Misra, and V. S. Rawoot, Phys. Rev. **D91**, 014005 (2015), arXiv:1405.3560 [hep-ph].

[14] A. Mukherjee and S. Rajesh, Eur. Phys. J. **C77**, 854 (2017), arXiv:1609.05596 [hep-ph].
[15] A. Bacchetta, D. Boer, C. Pisano, and P. Taels, Eur. Phys. J. C **80**, 72 (2020), arXiv:1809.02056 [hep-ph].
[16] U. D'Alesio, F. Murgia, C. Pisano, and P. Taels, Phys. Rev. **D100**, 094016 (2019), arXiv:1908.00446 [hep-ph].
[17] D. Boer *et al.*, (2011), arXiv:1108.1713 [nucl-th].
[18] A. Accardi *et al.*, Eur. Phys. J. A **52**, 268 (2016), arXiv:1212.1701 [nucl-ex].
[19] S. J. Brodsky, F. Fleuret, C. Hadjidakis, and J. P. Lansberg, Phys. Rept. **522**, 239 (2013), arXiv:1202.6585 [hep-ph].
[20] C. A. Aidala *et al.*, *Proceedings, 27th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2019): Torino, Italy, April 8-12, 2019*, PoS **DIS2019**, 233 (2019), arXiv:1901.08002 [hep-ex].
[21] Z. Lu and B.-Q. Ma, Phys. Rev. **D94**, 094022 (2016), arXiv:1611.00125 [hep-ph].
[22] J. M. Pereira-Resina-Rodrigues, *Modelling quark and gluon correlation functions*, Ph.D. thesis, Vrije Univ. Amsterdam (2001).
[23] D. Chakrabarti, P. Choudhary, B. Gurjar, R. Kishore, T. Maji, C. Mondal, and A. Mukherjee, Phys. Rev. D **108**, 014009 (2023), arXiv:2304.09908 [hep-ph].
[24] X. Xie and Z. Lu, Phys. Lett. B **842**, 137973 (2023), arXiv:2210.16532 [hep-ph].
[25] A. Bacchetta, F. Conti, and M. Radici, Phys. Rev. **D78**, 074010 (2008), arXiv:0807.0323 [hep-ph].
[26] A. Bacchetta, M. Radici, F. Conti, and M. Guagnelli, Eur. Phys. J. **A45**, 373 (2010), arXiv:1003.1328 [hep-ph].
[27] L. P. Gamberg and G. R. Goldstein, Phys. Lett. B **650**, 362 (2007), arXiv:hep-ph/0506127.
[28] L. P. Gamberg, G. R. Goldstein, and M. Schlegel, Phys. Rev. D **77**, 094016 (2008), arXiv:0708.0324 [hep-ph].
[29] R. Jakob, P. J. Mulders, and J. Rodrigues, Nucl. Phys. **A626**, 937 (1997), arXiv:hep-ph/9704335 [hep-ph].
[30] A. Bacchetta, F. G. Celiberto, M. Radici, and P. Taels, Eur. Phys. J. C **80**, 733 (2020), arXiv:2005.02288 [hep-ph].
[31] K. Goeke, S. Meissner, A. Metz, and M. Schlegel, Phys. Lett. **B637**, 241 (2006), arXiv:hep-ph/0601133 [hep-ph].
[32] J. Collins, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **32**, 1 (2011).
[33] A. Efremov and A. Radyushkin, Theor. Math. Phys. **44**, 774 (1981).
[34] S. J. Brodsky, D. S. Hwang, and I. Schmidt, Phys. Lett. **B530**, 99 (2002), arXiv:hep-ph/0201296 [hep-ph].
[35] J. C. Collins, Phys. Lett. B **536**, 43 (2002), arXiv:hep-ph/0204004.
[36] A. V. Belitsky, X. Ji, and F. Yuan, Nucl. Phys. B **656**, 165 (2003), arXiv:hep-ph/0208038.
[37] C. Pisano, D. Boer, S. J. Brodsky, M. G. Buffing, and P. J. Mulders, JHEP **10**, 024 (2013), arXiv:1307.3417 [hep-ph].
[38] D. Boer, W. J. den Dunnen, C. Pisano, and M. Schlegel, Phys. Rev. Lett. **111**, 032002 (2013), arXiv:1304.2654 [hep-ph].
[39] M. G. Echevarria, T. Kasemets, P. J. Mulders, and C. Pisano, JHEP **07**, 158 (2015), [Erratum: JHEP05,073(2017)], arXiv:1502.05354 [hep-ph].
[40] C. Bomhof, P. Mulders, and F. Pijlman, Eur. Phys. J. C **47**, 147 (2006), arXiv:hep-ph/0601171.
[41] A. Bacchetta, C. Bomhof, U. D'Alesio, P. Mulders, and F. Murgia, Phys. Rev. Lett. **99**, 212002 (2007), arXiv:hep-ph/0703153.
[42] M. G. Buffing, Z.-B. Kang, K. Lee, and X. Liu, (2018), arXiv:1812.07549 [hep-ph].
[43] T. C. Rogers, Phys. Rev. D **88**, 014002 (2013), arXiv:1304.4251 [hep-ph].
[44] J.-w. Qiu and G. F. Sterman, Phys. Rev. Lett. **67**, 2264 (1991).
[45] J.-w. Qiu and G. F. Sterman, Nucl. Phys. B **378**, 52 (1992).
[46] J.-w. Qiu and G. F. Sterman, Phys. Rev. D **59**, 014004 (1999), arXiv:hep-ph/9806356.
[47] M. G. A. Buffing, M. Diehl, and T. Kasemets, JHEP **01**, 044 (2018), arXiv:1708.03528 [hep-ph].
[48] D. Kharzeev, Y. V. Kovchegov, and K. Tuchin, Phys. Rev. D **68**, 094013 (2003), arXiv:hep-ph/0307037.
[49] F. Dominguez, B.-W. Xiao, and F. Yuan, Phys. Rev. Lett. **106**, 022301 (2011), arXiv:1009.2141 [hep-ph].
[50] F. Dominguez, C. Marquet, B.-W. Xiao, and F. Yuan, Phys. Rev. D **83**, 105005 (2011), arXiv:1101.0715 [hep-ph].
[51] M. G. A. Buffing, A. Mukherjee, and P. J. Mulders, Phys. Rev. **D88**, 054027 (2013), arXiv:1306.5897 [hep-ph].
[52] S. J. Brodsky, D. S. Hwang, and I. Schmidt, Nucl. Phys. B **642**, 344 (2002), arXiv:hep-ph/0206259.
[53] L. P. Gamberg, G. R. Goldstein, and K. A. Oganessyan, Phys. Rev. D **67**, 071504 (2003), arXiv:hep-ph/0301018.
[54] D. Boer, S. J. Brodsky, and D. S. Hwang, Phys. Rev. D **67**, 054003 (2003), arXiv:hep-ph/0211110.
[55] A. Bacchetta, A. Schaefer, and J.-J. Yang, Phys. Lett. B **578**, 109 (2004), arXiv:hep-ph/0309246.
[56] R. D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo, and L. Rottoli, Eur. Phys. J. **C78**, 321 (2018), arXiv:1710.05935 [hep-ph].
[57] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, and J. Rojo (NNPDF), Nucl. Phys. **B887**, 276 (2014), arXiv:1406.5539 [hep-ph].
[58] R. D. Ball *et al.* (NNPDF), JHEP **04**, 040 (2015), arXiv:1410.8849 [hep-ph].
[59] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, JHEP **05**, 062 (2002), arXiv:hep-ph/0204232.
[60] A. Bacchetta, F. Delcarro, C. Pisano, M. Radici, and A. Signori, JHEP **06**, 081 (2017), [Erratum: JHEP 06, 051 (2019)], arXiv:1703.10157 [hep-ph].
[61] I. Scimemi and A. Vladimirov, JHEP **06**, 137 (2020), arXiv:1912.06532 [hep-ph].
[62] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, F. Delcarro, F. Piacenza, and M. Radici, JHEP **07**, 117 (2020), arXiv:1912.07550 [hep-ph].
[63] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, M. Cerutti, F. Piacenza, M. Radici, and A. Signori (MAP), JHEP **10**, 127 (2022), arXiv:2206.07598 [hep-ph].
[64] M. Bury, F. Hautmann, S. Leal-Gomez, I. Scimemi, A. Vladimirov, and P. Zurita, JHEP **10**, 118 (2022), arXiv:2201.07114 [hep-ph].
[65] V. Moos, I. Scimemi, A. Vladimirov, and P. Zurita, (2023), arXiv:2305.07473 [hep-ph].
[66] A. Bacchetta, F. G. Celiberto, and M. Radici, PoS **EPS-HEP2021**, 376 (2022), arXiv:2111.01686 [hep-ph].

[67] A. Bacchetta, F. G. Celiberto, and M. Radici, JPS Conf. Proc. **37**, 020124 (2022), arXiv:2201.10508 [hep-ph].
[68] A. Bacchetta, F. G. Celiberto, and M. Radici, PoS **PANIC2021**, 378 (2022), arXiv:2111.03567 [hep-ph].
[69] R. Abdul Khalek *et al.*, Nucl. Phys. A **1026**, 122447 (2022), arXiv:2103.05419 [physics.ins-det].
[70] R. Abdul Khalek *et al.*, in *2022 Snowmass Summer Study* (2022) arXiv:2203.13199 [hep-ph].
[71] R. Abir *et al.*, (2023), arXiv:2305.14572 [hep-ph].
[72] A. Arbuzov *et al.*, Prog. Part. Nucl. Phys. **119**, 103858 (2021), arXiv:2011.15005 [hep-ex].
[73] E. Chapon *et al.*, Prog. Part. Nucl. Phys. **122**, 103906 (2022), arXiv:2012.14161 [hep-ph].
[74] S. Amoroso *et al.*, Acta Phys. Polon. B **53**, A1 (2022), arXiv:2203.13923 [hep-ph].
[75] L. A. Anchordoqui *et al.*, Phys. Rept. **968**, 1 (2022), arXiv:2109.10905 [hep-ph].
[76] J. L. Feng *et al.*, J. Phys. G **50**, 030501 (2023), arXiv:2203.05090 [hep-ex].
[77] A. Accardi *et al.*, (2023), arXiv:2306.09360 [nucl-ex].
[78] F. G. Celiberto, Phys. Rev. D **105**, 114008 (2022), arXiv:2204.06497 [hep-ph].
[79] M. Deak, A. van Hameren, H. Jung, A. Kusina, K. Kutak, and M. Serino, Phys. Rev. D **99**, 094011 (2019), arXiv:1809.03854 [hep-ph].
[80] F. Silvetti and M. Bonvini, Eur. Phys. J. C **83**, 267 (2023), arXiv:2211.10142 [hep-ph].

The ghost-gluon vertex in the presence of the Gribov horizon: general kinematics

N. Barrios* and M. Peláez†
Instituto de Física, Facultad de Ingeniería, Universidad de la República, J.H. y Reissig 565, 11000 Montevideo, Uruguay

M. S. Guimaraes,‡ B. W. Mintz,§ and L. F. Palhares¶
UERJ – Universidade do Estado do Rio de Janeiro,
Departamento de Física Teórica,
Rua São Francisco Xavier 524,
20550-900, Maracanã, Rio de Janeiro, Brasil
(Dated: February 28, 2024)

Correlation functions are important probes for the behavior of quantum field theories. Already at tree-level, the Refined Gribov Zwanziger (RGZ) effective action for Yang-Mills theories provides a good approximation for the gluon propagator, as compared to that calculated by nonperturbative methods such as Lattice Field Theory and Dyson-Schwinger Equations. However, the study of higher correlation functions of the RGZ theory is still at its beginning. In this work we evaluate the ghost-antighost-gluon vertex function in Landau gauge at one-loop level, in $d = 4$ space-time dimensions for the gauge groups SU(2) and SU(3). More precisely, we extend the analysis conducted in [1] for the soft-gluon limit to an arbitrary kinematic configuration. We introduce renormalization group effects by means of a toy model for the running coupling and investigate the impact of such a model in the ultraviolet tails of our results. We find that RGZ results match fairly closely those from lattice simulations, Schwinger-Dyson equations and the Curci-Ferrari model for three different kinematic configurations. This is compatible with RGZ being a feasible theory for the strong interaction in the infrared regime.

arXiv:2402.17534v1 [hep-ph] 27 Feb 2024

I. INTRODUCTION

Since the proposal of Quantum Chromodynamics (QCD) as the theory of Strong Interactions, a long path was constructed to connect the fundamental degrees of freedom – quarks and gluons – to the observed physical states and processes. At high energies, asymptotic freedom allows for a perturbative approach that, supplemented by essential nonperturbative information in Particle Distribution Functions and fragmentation phenomena, agrees with a plethora of experimental output from high-energy particle colliders. The infrared (IR) regime however is much less amenable. Monte Carlo simulations that solve the Euclidean version of QCD on a discretized space-time lattice have by now and with great effort established that this non-Abelian theory can quantitatively describe several hadronic observables [2–5]. Nevertheless, the mechanism of color confinement is still an open question, calling for the development of continuum approaches and (semi)analytical descriptions of the infrared behavior of Strong Interactions. Among the well-developed continuum methods that try to tackle this non-perturbative regime, Schwinger-Dyson equations [6–15] stand out in different hadronic applications, but also the Functional Renormalization Group [16–21] and effective models [22–26] have been employed with partial success. Other approaches, such as the Curci-Ferrari (CF) model in Landau gauge [27–37] and the screened massive expansion [38–42] have successfully described some aspects of the IR of Yang-Mills theory by employing perturbation theory.

Here we adopt another continuum approach to the nonperturbative regime of Yang-Mills theories: the Refined Gribov-Zwanziger (RGZ) theory [43, 44]. This framework, as the other continuum methods, adopts a gauge-fixed setup and is formulated from first-principles as a gauge path integral modified in the infrared by the existence of Gribov gauge copies. This idea follows the seminal work by Gribov himself [45] and the development of local actions attained by Zwanziger [46] and complemented by the emergence of dimension two condensates [43, 44, 47]. For comprehensive reviews, the reader is referred to [48–50] and references therein.

The presence of this nonperturbative background stemming from the Gribov horizon and the condensates seems to carry plenty of information from the interacting theory, so that the remaining interaction corrections might be supposed to be small, *i.e.* perturbative. Even at tree-level, the RGZ gluon propagator is compatible with the deep IR behavior observed on Landau-gauge Lattice QCD data [43], while reducing to pure Yang-Mills at large energies [51] in

* nbarrios@fing.edu.uy
† mpelaez@fing.edu.uy
‡ msguimaraes@uerj.br
§ brunomintz@gmail.com
¶ leticia.palhares@uerj.br

a fully self-consistent formulation that would include solving the Gribov gap equation and the effective potential for the different condensates. Gauge-parameter independence of gauge-invariant quantities in this setup is under control within linear covariant gauges through a nilpotent Becchi-Rouet-Stora–Tyutin symmetry, and the predictions for the gluon propagator in general covariant gauges are also compatible with available lattice results [52–55].

In this paper, the aim is to take a step forward in the direction of establishing predictions of the RGZ theory by including radiative corrections and confronting them with lattice data and other nonperturbative approaches. This is crucial for further understanding whether this theory can be a consistent model of infrared Yang-Mills and eventually QCD. Previous results on the vertices of the (R)GZ action in the Landau gauge may be found in Ref. [56]. Here, in particular, we compute the general kinematics of the gluon-ghost-antighost vertex at one loop order within the RGZ framework, extending previous work on the soft gluon limit of the same correlation function [1]. The present study enables the investigation of the role played by auxiliary fields in quantitative results in RGZ. In this specific correlation function, we show that the contribution of auxiliary fields, in the form of new vertices and mixed propagators, represents a smaller effect – less than 10% percent – as compared to that of the presence of nonzero pole masses for the gluon. Moreover, we include running effects via an infrared model, with a freezing coupling constant. Even though this is not self-consistently obtained within RGZ, this may be a reasonable approximation, because of the presence of the massive parameters from the horizon and the condensates. Overall we show that the RGZ results are fully compatible with available lattice data for SU(2) and SU(3), as well as other continuum approaches, such as the infrared-safe Curci-Ferrari model in Landau gauge [29, 35] and the Dyson-Schwinger equations [15].

This paper is organized as follows. In Sec. II, the formalism of the refined Gribov-Zwanziger framework is presented. In Sec. III, the ghost-antighost-gluon vertex is defined and its general structure in the RGZ theory is discussed. Our computation of generic momentum configurations of the one-loop ghost-gluon vertex is presented in Sec. IV, including the IR running coupling model adopted. Sec. V collects our results along with the corresponding analysis and comparisons with alternative approaches. In Sec. VI we discuss the potential influence of the IR running coupling model on the ultraviolet (UV) tails of our results. We have included perspectives and final remarks in Section VII.

II. THE REFINED GRIBOV-ZWANZIGER ACTION IN LANDAU GAUGE

The Gribov-Zwanziger theory is a framework for making sense of the gauge fixed Yang-Mills theory in the non-perturbative regime. It grew from Gribov's observation that gauge fixing in the Landau Gauge fails in the strongly coupled regime [45, 46] (see also [48–50] for reviews), in the sense that gauge copies still exist after the fixing. These are called Gribov copies and they appear in fact in any covariant gauge. The solution proposed by Gribov was to restrict the integration region of the gauge field path integral to the neighborhood containing the perturbative vacuum and bounded by the field configurations associated with the first Gribov copies, called the Gribov Horizon. In the Landau gauge $\partial_\mu A^a_\mu = 0$ the Gribov horizon is defined by the set of fields A^a_μ for which the equation $\mathcal{M}^{ab}(A)\alpha^b = 0$ has a solution, where

$$\mathcal{M}^{ab}(A)(\bullet) = -\delta^{ab}\partial^2(\bullet) + gf^{abc}\partial_\mu(A^c_\mu\bullet). \tag{1}$$

stands for the Faddeev-Popov operator in the Landau gauge.

Zwanziger [46] was able to recast Gribov's construction in the form of an effective theory whose restriction to the Gribov horizon is implemented at the level of the action. The resulting action is known as the nonlocal Gribov-Zwanziger action and defined in the Landau gauge in dimension D by

$$S_{nlGZ} = \int d^Dx\, \left(\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + ib^a\partial_\mu A^a_\mu + \bar{c}^a\mathcal{M}(A)^{ab}c^b\right) + \gamma^D H(A) - \gamma^D VD\left(N^2-1\right) \tag{2}$$

where γ has dimension of mass, V is the space-time volume (formally infinite) and $H(A)$ is known as the horizon function and given by

$$H(A) = g^2\int d^Dp\int d^Dq\left[f^{abd}\tilde{A}^d_\mu(-p)\left(\mathcal{M}^{-1}\right)^{bc}_{pq}f^{cae}\tilde{A}^e_\mu(q)\right] \tag{3}$$

The mass scale γ is not arbitrary, but fixed by a self consistent equation known as the gap equation, which can be written as a condition on a vacuum condensate. Defining

$$Z(\tilde{\lambda}) = \int \mathcal{D}A\mathcal{D}b\mathcal{D}\bar{c}\mathcal{D}c\; e^{-\int d^Dx\,\left(\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu}+ib^a\partial_\mu A^a_\mu-\bar{c}^a\mathcal{M}(A)^{ab}c^b\right)+\tilde{\lambda}H(A)-\tilde{\lambda}VD\left(N^2-1\right)} \tag{4}$$

The gap equation reads

$$\frac{\partial Z(\tilde{\lambda})}{\partial\tilde{\lambda}}\Big|_{\tilde{\lambda}=\gamma^D} = 0 \Rightarrow \langle H(A)\rangle = V D\left(N^2-1\right). \tag{5}$$

The theory as it is is highly non-local due to explicit expression of $H(A)$. In order to have a workable quantum field theory we need to put the action in a local form. This entails the introduction of new, auxiliary fields. We note that

$$\int \mathcal{D}\varphi\mathcal{D}\bar{\varphi}\mathcal{D}\omega\mathcal{D}\bar{\omega}\; e^{-\int d^4x\left(\bar{\varphi}^{ac}_\mu \mathcal{M}(A)^{ab}\varphi^{bc}_\mu - \bar{\omega}^{ac}_\mu\mathcal{M}(A)^{ab}\omega^{bc}_\mu + ig\gamma^{\frac{D}{2}}\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu)\right)} \sim e^{-\gamma^D\, H(A)} \tag{6}$$

where the symbol $\sim$ means "up to a prefactor" and $(\varphi,\bar{\varphi})$ are complex bosonic fields and $(\omega,\bar{\omega})$ are fermionic fields. Therefore we obtain the local formulation as

$$Z_{GZ} = \int \mathcal{D}A\mathcal{D}\bar{c}\mathcal{D}c\mathcal{D}b\mathcal{D}\bar{\varphi}\mathcal{D}\varphi\mathcal{D}\bar{\omega}\mathcal{D}\omega\; e^{-S_{GZ}} \tag{7}$$

where

$$\begin{aligned} S_{GZ} = &\int d^Dx\;\left(\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + ib^a\partial_\mu A^a_\mu - \bar{c}^a\mathcal{M}(A)^{ab}c^b\right)\\ &+\int d^Dx\left(\bar{\varphi}^{ac}_\mu\mathcal{M}(A)^{ab}\varphi^{bc}_\mu - \bar{\omega}^{ac}_\mu\mathcal{M}(A)^{ab}\omega^{bc}_\mu\right)\\ &+\int d^Dx\left(ig\gamma^{\frac{D}{2}}\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu) - \gamma^D D\left(N^2-1\right)\right) \end{aligned} \tag{8}$$

and γ is such that it satisfies the gap equation (5), which now reads

$$g\langle\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu)\rangle = \gamma^{\frac{D}{2}} D\left(N^2-1\right) \tag{9}$$

where translation invariance of the vacuum was used to factor out the volume $\int d^Dx\langle\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu)\rangle = V\langle\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu)\rangle$.

The vacuum defined by the gap equation is unstable and favors the formation of new condensates [43] that are related to mass scales of the gauge fields $\langle A^2\rangle \sim m^2$ and auxiliary fields $\langle\bar{\phi}\phi\rangle \sim \mu^2$. The incorporation of these condensates in the effective action formulation leads to a modified theory known as the Refined Gribov-Zwanziger theory, defined by

$$\begin{aligned} S_{RGZ} = &\int d^Dx\;\left(\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} + ib^a\partial_\mu A^a_\mu - \bar{c}^a\mathcal{M}(A)^{ab}c^b\right)\\ &+\int d^Dx\left(\bar{\varphi}^{ac}_\mu\mathcal{M}(A)^{ab}\varphi^{bc}_\mu - \bar{\omega}^{ac}_\mu\mathcal{M}(A)^{ab}\omega^{bc}_\mu\right)\\ &+\int d^Dx\left(ig\gamma^{\frac{D}{2}}\, f^{abc}A^a_\mu(\varphi^{bc}_\mu+\bar{\varphi}^{bc}_\mu) - \gamma^D D\left(N^2-1\right)\right)\\ &+\int \mathrm{d}^Dx\;\frac{m^2}{2}A^a_\mu A^a_\mu + \int \mathrm{d}^Dx\; M^2\left(\bar{\varphi}^{bc}_\mu\varphi^{bc}_\mu - \bar{\omega}^{bc}_\mu\omega^{bc}_\mu\right). \end{aligned} \tag{10}$$

It is worth pointing out the remarkable fact discovered in [54] that it is possible to recast this theory making it suitable for any linearly covariant gauge, so that the resulting RGZ theory can be made BRST invariant. This is done by replacing the gauge field A by a gauge invariant composite field A^h [54, 57]

$$A^h_\mu = h^\dagger A_\mu h + \frac{i}{g}h^\dagger\partial_\mu h\,, \tag{11}$$

with

$$h = \mathrm{e}^{ig\xi^a T^a} \equiv \mathrm{e}^{ig\xi}, \tag{12}$$

where ξ^a is the Stueckelberg field discussed in [52–55, 57, 58]. One also needs to impose a transversality constraint on A^h,

$$\partial_\mu A^{h,a}_\mu = 0\,, \tag{13}$$

so that the h field is not really independent, but is an auxiliary field. This leads to a consistent renormalizable and BRST invariant formulation. This enlarged formulation reduces to the Landau gauge if $\xi = 0$. In this work we will only consider the Landau gauge.

III. THE THREE-POINT GHOST-GLUON CORRELATION FUNCTION

Let us now start discussing the ghost-gluon correlation function and establish some notation. We first make some general remarks about the connected ghost-antighost-gluon three-point function in the RGZ theory and how the mixed propagators affect its building blocks. Next, for completeness, we quote the results obtained in [1] for the soft-gluon limit. Our original results for the ghost-antighost-gluon correlation function for arbitrary momenta are left for the next section.

A. General structure of the $A\bar{c}c$ vertex

The relevant Feynman rules for the calculation of the ghost-gluon vertex at one-loop order can be derived from the action (10) and are listed in Appendix A.

With these rules at hand, one can calculate the connected correlation function

$$\left\langle A^a_\mu(k)\,\bar{c}^b(p)\,c^c(q)\right\rangle_{q=-p-k} = \left.\frac{\delta^3 Z_c[J_A,J_{\bar{c}},J_c]}{\delta(J_{\bar{A}})^a_\mu(k)\delta J^b_{\bar{c}}(p)\delta J^c_c(q)}\right|_{q=-p-k;\quad J_i=0} \tag{14}$$

at one-loop order, where Z_c is the generator of connected correlation functions and J_i ($i=\bar{c},c,A$) are external sources linearly coupled to the fields i. As usual, the sources are taken to zero at the end of the calculation.

Before proceeding, let us remark that since the RGZ action contains bilinear couplings between fields, the theory contains mixed propagators, such as $\langle A\varphi\rangle$ and $\langle A\bar{\varphi}\rangle$. Therefore, the relation between connected and 1PI functions has to take such mixed propagators into account. This is made explicit in the Feynman diagrams of Fig. 1. Such mixed propagators and vertices involving Zwanziger's auxiliary fields φ and $\bar{\varphi}$ as well as their fermionic counterparts ω and $\bar{\omega}$, arise as a consequence of the local formulation of the Gribov horizon [1]. We give further details for the interested reader in Appendix B.

Since the tree-level mixed propagator is such that

$$\langle A^a_\mu(p)\varphi^{bc}_\nu(-p)\rangle \;=\; \frac{-ig\gamma^2 f^{abc}}{p^4+p^2(m^2+M^2)+m^2M^2+2Ng^2\gamma^4}P^\perp_{\mu\nu}(p), \tag{15}$$

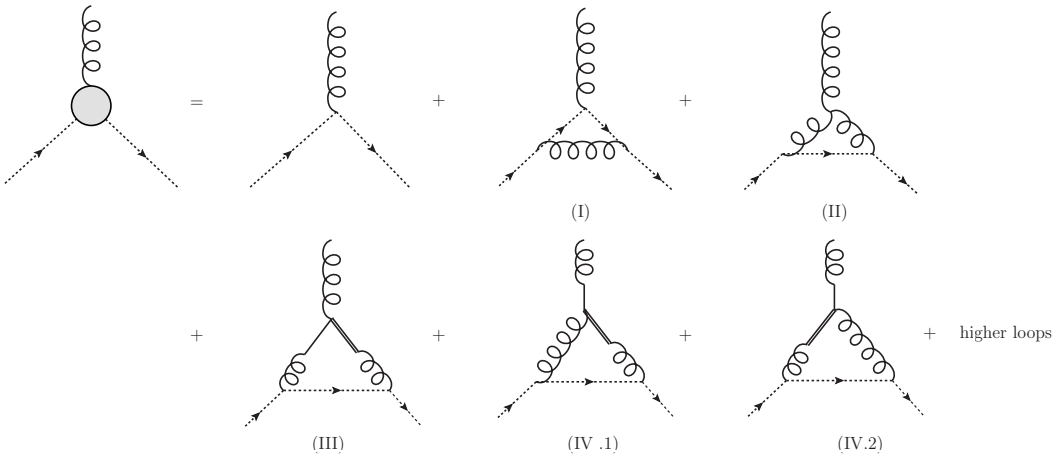


FIG. 1: Feynman diagram expansion up to one-loop order for the ghost-gluon vertex in the Refined Gribov-Zwanziger. Dashed lines represent ghosts and antighosts, while the curly lines stand for gluons. Full simple and double lines, that only appear in mixed propagators, correspond to the auxiliary fields φ and $\bar{\varphi}$, respectively. The roman numbers identifying the one-loop diagrams will be used as reference in the results section and in the appendices. The diagrams were generated by means of the JAXODRAW interface [59].

[1] Another possible formulation of the theory would include nonlocal, momentum-dependent vertices instead of the auxiliary fields. However, for the sake of using standard Quantum Field Theory techniques, we employ the local version of the theory.

the one-loop connected function (14) can then be decomposed as

$$\left\langle A^a_\nu(k)\,\bar{c}^b(p)\,c^c(q)\right\rangle = G(p)G(q)D_{AA}(k)P^\perp_{\mu\nu}(k)\left\{\Gamma_{A^a_\mu\bar{c}^bc^c}(k,p,r)+\frac{2g\gamma^2 f^{ade}}{k^2+M^2}\Gamma_{\bar{c}^bc^c\varphi^{de}_\mu}(k,p,r)\right\}_{q=-p-k}, \tag{16}$$

where

$$P^\perp_{\mu\nu}(p)=\delta_{\mu\nu}-\frac{p_\mu p_\nu}{p^2} \tag{17}$$

is the transverse projector, and the relevant 1PI functions are

$$\Gamma_{A^a_\nu\bar{c}^bc^c}(k,p,r):=\frac{\delta^3\Gamma}{\delta A^a_\nu(-k)\delta\bar{c}^b(-p)\delta c^c(-q)} \tag{18}$$

and

$$\Gamma_{\bar{c}^bc^c\varphi^{de}_\mu}(k,p,r):=\frac{\delta^3\Gamma}{\delta\bar{c}^b(-p)\delta c^c(-q)\delta\varphi^{de}_\nu(-k)}. \tag{19}$$

Note the presence of the $\Gamma_{\bar{c}c\varphi}$ and $\Gamma_{\bar{c}c\bar{\varphi}}$ vertex functions in the RGZ theory. They are zero at tree-level. In the case of $\Gamma_{\bar{c}c\varphi}$ its first nontrivial terms are given by diagrams (IV.1) and (IV.2) of fig. 1. Notice that, since $\Gamma_{\bar{c}c\varphi}=\Gamma_{\bar{c}c\bar{\varphi}}$, in fig. 1 we have included the diagrams associated with the former quantity only. The contribution of $\Gamma_{\bar{c}c\bar{\varphi}}$ to $\left\langle A^a_\nu(k)\,\bar{c}^b(p)\,c^c(q)\right\rangle$ is introduced by multiplying the contribution of $\Gamma_{\bar{c}c\varphi}$ by a factor of two, as can be observed in eq. (16).

The tensorial structure of the ghost-gluon 1PI vertex function is given by

$$\Gamma_{A^a_\mu\bar{c}^bc^c}(k,p,r)=-igf^{abc}[p_\mu B_1(k,p)+k_\mu B_2(k,p)], \tag{20}$$

where we adopt the same notation as in [15] and consider all momenta as incoming. At tree-level, the scalar structure functions are $B_1(k,p)=1$ and $B_2(k,p)=0$.

On the other hand, the contraction of the $\Gamma_{\varphi\bar{c}c}$ function with the antisymmetric color structure constant tensor can be decomposed as

$$f^{ade}\Gamma_{\bar{c}^bc^c\varphi^{de}_\mu}(k,p,r)=-igf^{abc}\left(p_\mu C_1(k,p)+k_\mu C_2(k,p)\right)-igd^{abc}\left(p_\mu C_1'(k,p)+k_\mu C_2'(k,p)\right) \tag{21}$$

Since this combination must be antisymmetric in the color indices b and c (due to the fact that the ghost fields are grassmannian), the coefficient of the totally symmetric tensor d^{abc} must vanish identically, *i.e.*, $C_1'=C_2'=0$. As a result, we may write the one-loop connected three-point function as

$$\left\langle A^a_\mu(k)\,\bar{c}^b(p)\,c^c(q)\right\rangle \;=\; -igp_\nu\, f^{abc}G(p)G(p+k)D_{AA}(k)P^\perp_{\mu\nu}(k)\left\{B_1(k,p)+\frac{2g\gamma^2}{k^2+M^2}C_1(k,p)\right\}, \tag{22}$$

where the longitudinal scalar functions B_2 and C_2 are no longer present (in the Landau gauge) due to the transversality of the gluon propagator.

In order to make contact with results from Monte Carlo lattice simulations, let us consider the scalar quantity [60]

$$\begin{aligned}G^{CCA}(p,k)&\equiv\frac{\Gamma^{abc}_{(A\bar{c}c,tree)\,\mu}(k,p,-k-p)\left\langle A^a_\mu(k)\,\bar{c}^b(p)\,c^c(-k-p)\right\rangle}{\Gamma^{abc}_{(A\bar{c}c,tree)\,\mu}(k,p,-p-k)\Gamma^{abc}_{(A\bar{c}c,tree)\,\nu}(k,p,-p-k)P_{\mu\nu}(k)G(p)G(p+k)D_{AA}(k)}\\&=B_1(k,p)+\frac{2g\gamma^2}{k^2+M^2}C_1(k,p),\end{aligned} \tag{23}$$

which from now on we denote as the ghost-gluon vertex dressing function.

There are clearly some differences between perturbative YM and RGZ calculations of the vertex function. The first of them is the modification of the gluon propagator brought about by the restriction to the Gribov horizon, which can be understood as the appearance of a pair of generally complex conjugate poles. A second difference is the presence of the tree-level $A\bar{\varphi}\varphi$ vertex, which couples the gluon to the auxiliary Zwanziger fields. This allows not only diagrams with auxiliary fields running in the internal loops, but also in the external legs, as long as the external propagator is a mixed one like, for example, $\langle A\varphi\rangle$. This possibility is realized in (16), giving rise to the contributions $\Gamma_{\varphi\bar{c}c}$ and $\Gamma_{\bar{\varphi}\bar{c}c}$, not present in perturbative YM theory. Furthermore, note that these mixed contributions only appear from one-loop

order onwards, as such vertices are absent from the classical action (10). Being finite, they do not spoil the stability of the action, in agreement with the Quantum Action Principle [61]. Finally, we note that the presence of such vertices involving the auxiliary Zwanziger fields can be thought as effective momentum-dependent gluonic interaction terms. This claim can be explicitly shown to be true, starting from the tree level action, by considering the nonlocal version of the RGZ theory (whose action can be obtained from (10) by integrating out the Zwanziger auxiliary fields) and expanding the inverse of the Faddeev-Popov operator in powers of the coupling. In this regard, we note that it is reasonable to consider the RGZ effective action not simply as an effective model with propagators given by a ratio of polynomials, but rather a renormalizable effective theory with nontrivial gluonic propagators and self-interactions, which are related to each other.

As a final remark, let us note that all diagrams contributing to the ghost-gluon vertex are finite in the RGZ theory, just as in the perturbative Yang-Mills theory. In fact, the mixed propagators present in diagrams (III) and (IV) make the diagrams even more ultraviolet convergent than the usual YM diagrams (I) and (II). This can be easily seen from the form of the $\langle A\varphi\rangle$ propagator (15).

B. The soft gluon limit

The calculation of the ghost-gluon vertex for any momentum requires the calculation of the diagrams shown in fig. 1. Since this calculation is rather long to be performed manually, in a previous work some of us considered the soft-gluon limit of $\Gamma_{A\bar{c}c}$, *i.e.*, the limit in which the gluon momentum $k \to 0$ [1].

In the soft gluon limit, each diagram takes a simplified form and furthermore, since diagram (IV) of fig. 1 is proportional to the gluon momentum (and it is finite), it vanishes for $k \to 0$. In this limit, the one-loop correction for the vertex, which has been explicitly calculated in [1], reads

$$\begin{aligned}[\Gamma^{(1)}_{A\bar{c}c}(0,p,-p)]^{abc}_{\mu} &= ig^3\frac{Nf^{abc}}{2}\Big\{R_+J_\mu(a_+;p)+R_-J_\mu(a_-;p)+2R_+^2K_\mu(a_+,a_+;p)+2R_-^2K_\mu(a_-,a_-;p)+\\ &+4R_+R_-K_\mu(a_+,a_-;p)+\frac{N}{2}\left(\frac{g\gamma^2}{a_+^2-a_-^2}\right)^2[K_\mu(a_+,a_+;p)+K_\mu(a_-,a_-;p)-\\ &-2K_\mu(a_+,a_-;p)]\Big\}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} a_+^2 &\equiv \frac{m^2+M^2+\sqrt{(m^2-M^2)^2-4\lambda^4}}{2},\\ a_-^2 &\equiv \frac{m^2+M^2-\sqrt{(m^2-M^2)^2-4\lambda^4}}{2}, \end{aligned}$$

with $\lambda^4 \equiv 2Ng^2\gamma^4$, and the integral

$$J_\mu(m_1;p) \equiv \int \frac{d^dq}{(2\pi)^d}\,\frac{1}{q^2}\frac{1}{q^2+m_1^2}\frac{p^2q^2-(p\cdot q)^2}{[(q-p)^2]^2}\,(q-p)_\mu, \tag{25}$$

is related to diagram (I) while

$$K_\mu(m_1,m_2;p) \equiv \int \frac{d^dq}{(2\pi)^d}\frac{1}{(q+p)^2}\frac{1}{q^2+m_1^2}\frac{1}{q^2+m_2^2}\left[\frac{q^2p^2-(p\cdot q)^2}{q^2}\right]q_\mu \tag{26}$$

appears in diagrams (II) and (III)[2]. The incoming antighost momentum is given by p. Therefore, the ghost momentum is $-p$, since $k=0$ in the soft-gluon limit. The massive parameters $-a_\pm^2$ are the, generally complex, poles of the RGZ gluon propagator (A1) and $R_\pm$ are their corresponding residues. It is worth pointing out that the last terms in eq. (24) come from diagram (III) in fig. 1 which is absent in standard YM theories, being proportional to the Gribov parameter.

---

[2] Analytic expressions of $J_\mu(m;p)$ and $K_\mu(m_1,m_2;p)$ are provided in appendix C.

As a last remark, let us note an important feature of the $\Gamma_{A\bar{c}c}$ vertex function in the RGZ theory: it explicitly respects the so-called Taylor kinematics, *i.e.*,

$$(\Gamma_{A\,\bar{c}\,c})^{abc}_{\mu}(p,0,-p) = 0\,, \tag{27}$$

and the so-called non-renormalization theorem of the ghost-gluon vertex

$$(\Gamma_{A\,\bar{c}\,c})^{abc}_{\mu}(-p,p,0) = -igf^{abc}p_{\mu}, \tag{28}$$

which are the same in the RGZ framework as in perturbative YM theory [62]. These are direct consequences of the Ward identities of the action (10).

IV. THE ONE-LOOP RGZ GHOST-GLUON VERTEX FOR GENERIC MOMENTUM CONFIGURATIONS

A. Feynman diagrams

It is clear from eq. (23) that the ghost-gluon vertex dressing function relates to two 1PI structures. Diagrams (I)-(III) from fig. 1 contribute to $\Gamma_{A\bar{c}c}$, while the forth diagram contributes to $\Gamma_{\varphi\bar{c}c}$.

Their expressions in terms of the Feynman rules from appendix A, following the numbering from fig. 1, are:

$$\begin{aligned}
D_{(I)} &= -ig^3\frac{N}{2}f^{abc}p_{\nu}(p+k)_{\rho}\int\frac{d^dq}{(2\pi)^d}q_{\mu}P^{\perp}_{\nu\rho}(q+p)D_{AA}(q+p)D_{\bar{c}c}(q-k)D_{\bar{c}c}(q)\,,\\
D_{(II)} &= -ig^3\frac{N}{2}f^{abc}p_{\eta}(p+k)_{\omega}\int\frac{d^dq}{(2\pi)^d}\left(-2k_{\rho}\delta_{\mu\nu}+2k_{\nu}\delta_{\mu\rho}+(2q-k)_{\mu}\delta_{\nu\rho}\right)P^{\perp}_{\nu\eta}(q)P^{\perp}_{\rho\omega}(q-k)D_{AA}(q)D_{AA}(q-k)D_{\bar{c}c}(q+p)\,,\\
D_{(III)} &= ig^3\frac{N^2}{4}f^{abc}(g\gamma^2)^2p_{\nu}(p+k)_{\rho}\int\frac{d^dq}{(2\pi)^d}(2q-k)_{\mu}P^{\perp}_{\rho\eta}(q-k)P^{\perp}_{\nu\eta}(q)D_{A\varphi}(q-k)D_{A\varphi}(q)D_{\bar{c}c}(q+p)\,,\\
f^{ade}D_{(IV.1)} &= ig^3\frac{N^2}{4}f^{abc}(g\gamma^2)p_{\eta}k_{\omega}(p+k)_{\sigma}\int\frac{d^dq}{(2\pi)^d}P^{\perp}_{\sigma\mu}(q-k)P^{\perp}_{\eta\omega}(q)D_{A\varphi}(q-k)D_{AA}(q)D_{\bar{c}c}(q+p)\,,\\
f^{ade}D_{(IV.2)} &= -ig^3\frac{N^2}{4}f^{abc}(g\gamma^2)p_{\sigma}k_{\omega}(p+k)_{\eta}\int\frac{d^dq}{(2\pi)^d}P^{\perp}_{\sigma\mu}(q)P^{\perp}_{\eta\omega}(q-k)D_{A\varphi}(q)D_{AA}(q-k)D_{\bar{c}c}(q+p)\,,
\end{aligned} \tag{29}$$

where $D_{XY}(q)$ designates the propagator of momentum q between the fields X and Y.

The above Feynman integrals depend on two external momenta, p and k, and feature only one Lorentz index which is not contracted, μ. As a result, all of them can be expressed as $F(p^2,k^2,p{\cdot}k,M^2,m^2,\lambda)p_{\mu}+H(p^2,k^2,p{\cdot}k,M^2,m^2,\lambda)k_{\mu}$, where F and H are scalar integrals. In the first stage of our computation we found the corresponding scalar integrals for each diagram[3] and evaluated the color factors by means of the ColorMath package [63].

In a second stage, we reduced the scalar integrals to one-loop master integrals [64] of the type

$$A_x \equiv \int_q G_x(q^2), \tag{30}$$

$$B_{xy}(p^2) \equiv \int_q G_x(q^2)G_y((p-q)^2), \tag{31}$$

$$C_{xyz}(p^2,k^2,p\cdot k) \equiv \int_q G_x(q^2)G_y((p-q)^2)G_z((k-q)^2), \tag{32}$$

where

$$\int_q \equiv \int 16\pi^2\mu^{2\epsilon}\frac{d^dq}{(2\pi)^d}, \tag{33}$$

[3] The contributions for the dressing function G^{CCA} come from the p_{μ} component entirely.

with $\epsilon = \frac{4-d}{2} \geq 0$ and

$$G_x(q^2) \equiv \frac{1}{q^2 + x}, \tag{34}$$

where x is a "mass" squared (which may be complex, in our case). In the case of the RGZ theory $x \in \{0, M^2, a_+^2, a_-^2\}$. The mass dimension of the gauge coupling, defined as μ^ϵ, is absorbed into the definition of the master integrals. More specifically, as the dressing function G^{CCA} is proportional to g^2, we include the respective mass dimension, $\mu^{2\epsilon}$, in the definition given by eq. (33).

We implemented the reduction to master integrals using an algorithm in MATHEMATICA based on the FIRE package [65]. The reduction for each diagram is presented in a supplementary material. The analysis is performed for an arbitrary number of colors N with the purpose of comparing with alternative approaches, both for the SU(2) and SU(3) gauge groups.

Once the reduction is finished we need to compute the one-loop master integrals. For both A and B there are well-known analytic results (see, for instance, ref. [66]). In the case of the master integral C, when needed, we evaluated it numerically.

We note here that since all one-loop diagrams are finite no renormalization factors are needed.

B. Kinematic configurations

Concerning the particular kinematic configurations we will address in this article, it is convenient to recall that three-point correlation functions depend on two external momenta. Because of translational and rotational invariance such dependence can be described by means of three independent kinematic variables. We choose the magnitudes of the antighost and gluon momentum, p and k, along with the scalar product $p \cdot k$.

In this paper we aim at extending the analysis of the ghost-gluon vertex within the RGZ framework, initially explored in [1], beyond the soft gluon limit, to include arbitrary kinematic configurations. We focus on the specific kinematic setups for which we have access to lattice simulations, namely the symmetric and orthogonal configurations.

The symmetric configuration is characterized by equal momentum magnitudes for the external antighost and gluon legs, $p = k$, forming an angle of $\theta = \pi/3$. On the other hand, the orthogonal configuration refers to external momenta perpendicular to each other, $\theta = \pi/2$. Within this work, our reference to the orthogonal configuration includes the additional condition $p = k$.

C. Numerical values of the RGZ parameters

Within the RGZ framework, there are four parameters: M and m, that are mass parameters associated with dimension two condensates, the Gribov parameter γ and the gauge coupling g. For the sake of convenience, we introduce the parameter $\lambda^4 = 2Ng^2\gamma^4$, which we will employ in place of γ.

In principle, the three massive parameters could be determined self-consistently using the Gribov gap equation and minimizing the effective potential of the theory with respect to two condensates. Here, however, we shall fix these parameters using lattice input. The parameters M, m and λ at tree-level were determined for SU(2) in [67] and for SU(3) in [68], by fitting the lattice data of the gluon propagator from YM theory with an analytic expression of the form:

$$D(p^2) = \frac{p^2 + a}{p^4 + bp^2 + c}, \tag{35}$$

where a, b and c are constants. This analytic form coincides with the tree-level gluon propagator of the RGZ framework,

$$D(p^2) = \frac{p^2 + M^2}{p^4 + (M^2 + m^2)p^2 + M^2m^2 + \lambda^4}, \tag{36}$$

allowing for the extraction of m, M and λ from eq. (35). The obtained values are displayed in table I.

Gauge group	M^2 (GeV^2)	m^2 (GeV^2)	λ (GeV)
SU(2) - Ref.[67]	2.51	-1.92	1.52
SU(3) - Ref.[68]	4.47	-3.77	2.04

TABLE I: RGZ parameters fitted from lattice results for SU(2) [67] and SU(3) [68].

D. Toy model of the running coupling

The gauge coupling g is to be determined by fitting the RGZ ghost-gluon vertex to available lattice simulations. To that end, it is useful to introduce the relative error between the RGZ output and the lattice data,

$$\chi^2 \equiv \frac{1}{N_{\rm latt}^2} \sum_i^{N_{\rm latt}} \left(\frac{G_{\rm RGZ}(p_i) - G_{\rm latt}(p_i)}{G_{\rm latt}(p_i)} \right)^2 , \tag{37}$$

where $N_{\rm latt}$ is the number of lattice data points, $G_{\rm RGZ}$ and $G_{\rm latt}$ refer to the RGZ and the numerical results, respectively.

Firstly, we simply investigate at a qualitative level the variation of the dressing function G^{CCA} as we change fixed values of g. As we will see, this is enough to reproduce either the IR or the UV but not both regimes at the same time. Most likely this is due to the presence of large logarithms in the UV domain, which spoils the validity of perturbation theory. This problem can be overcome by introducing the running of the various parameters of the RGZ framework and choosing a renormalization scale of the same order of the external momentum: $\mu \sim p$. To achieve this we would need to evaluate the two-point functions of the theory by introducing an infrared-safe scheme, of the type introduced in [29] for instance.

However, this sort of approach exceeds the scope of the present article. Instead, we keep the constant values provided in table I and make use of a toy model for the renormalization group (RG) flow of the gauge coupling, motivated by the standard Yang-Mills one loop β-function in the Modified Minimal Subtraction Scheme,

$$g^2(\mu) = \frac{g_0^2}{1 + \frac{11}{3} N \frac{g_0^2}{16\pi^2} \log(\frac{\mu^2+\Lambda^2}{\Lambda^2})}, \tag{38}$$

and choose $\mu = p$. The parameter Λ is introduced to regularize the IR[4] by freezing the running of the gauge coupling for momentum scales much smaller than Λ, where we expect the effects of the renormalization flow to be less relevant. The fits of G^{CCA} from the RGZ framework to lattice data are carried out by selecting the parameters g_0 and Λ that minimize χ.

V. RESULTS

In this section we show the results for the ghost-gluon vertex dressing function, G^{CCA}, for both the gauge groups SU(2) and SU(3) in four spacetime dimensions. We compare our results with lattice simulations and outcomes from the CF model and DSE. Additionally, we show the contributions of each Feynman diagram to the final results.

A. SU(2)

We begin by analyzing the ghost-gluon vertex dressing function G^{CCA} for the SU(2) gauge group in the symmetric and orthogonal kinematic configurations, as well as in the case of the soft gluon limit. The latter case was previously investigated in [1]. However, for completeness and because the analysis we perform in this article differs slightly from the one developed in [1], we include those results here as well.

[4] The Λ parameter is crucial so that the toy model is Landau pole free.

1. *Fitting the ghost-gluon vertex dressing function*

In fig. 2 we see the dressing function G^{CCA} for various fixed values of the gauge coupling g at one-loop order from the RGZ framework. As for the symmetric and orthogonal kinematic configurations it is clear that, although we find values of the coupling that show a very good agreement with lattice data either at the IR (pink curve) or the UV (purple curve), we are unable to reconcile both regions simultaneously. This is reasonable, as we are not taking into account the RG effects.

Interestingly, in the case of the soft gluon limit it seems that it is possible to describe the lattice data for the whole range of momenta with a constant value of gauge coupling (red curve). The quality of the fit, though, must be taken carefully due to the relatively large dispersion of the lattice data, specially at low momenta.

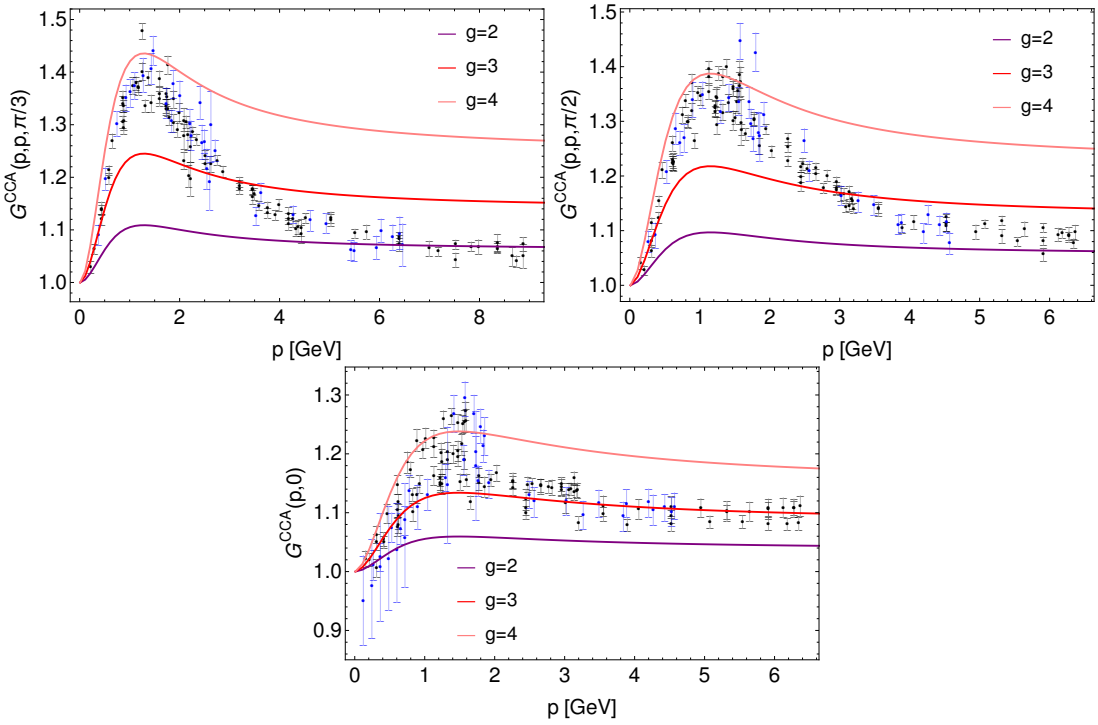


FIG. 2: The SU(2) scalar function G^{CCA} as a function of the antighost external momentum in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom) for various values of a fixed gauge coupling. Blue and black points refer to lattice data extracted from refs. [69] and [70], respectively.

To investigate to which extent the RG flow is capable of reducing the discrepancy between the lattice data and the RGZ fit, we incorporate the toy model of the running coupling shown in eq. (38). The level curves illustrating the dependence of the relative error defined in eq. (37) on the parameters g_0 and Λ are depicted in fig. 3. We notice that the regions minimizing χ are very similar for the symmetric and orthogonal configurations. These regions are also compatible with the level curves of χ observed in the soft gluon limit, though the area of minimum error appears more extensive. This discrepancy could stem from a greater uncertainty and dispersion of the lattice data in the soft gluon limit compared to the data from the symmetric and orthogonal setups.

The parameters that minimize χ in each kinematic configuration are displayed in table II. Additionally, we incorporate the values of g_0 and Λ that minimize the deviation between the lattice data and the RGZ output across all kinematic configurations simultaneously. These values are computed by minimizing the collective error, χ_{joint}, which we define as

$$\chi^2_{\text{joint}} \equiv \frac{\chi^2_{\text{sym.}} + \chi^2_{\text{orth.}} + \chi^2_{\text{soft}}}{3}, \tag{39}$$

where $\chi_{\text{sym.}}$, $\chi_{\text{orth.}}$ and χ_{soft} refer to the relative error of the symmetric, orthogonal and the soft gluon configurations, respectively.

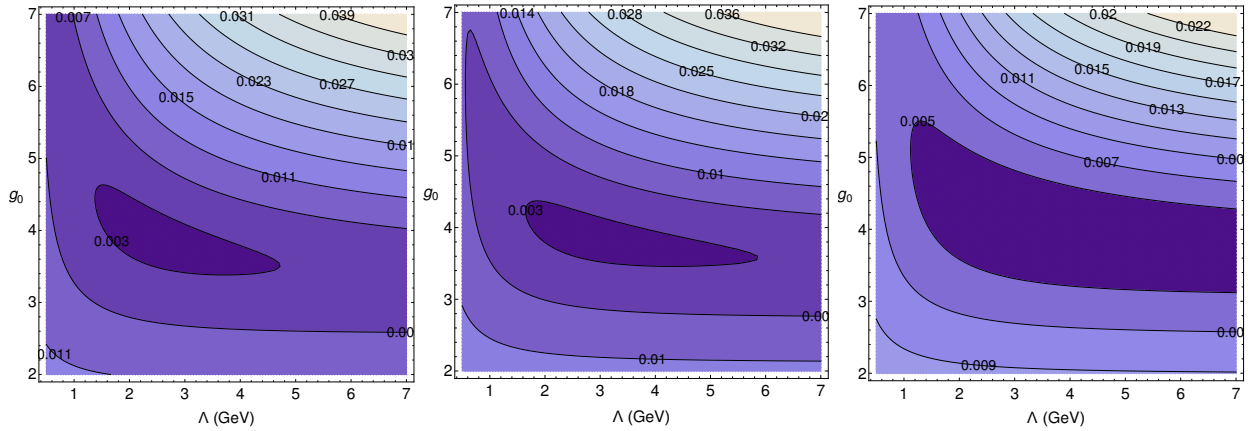


FIG. 3: Level curves of the relative error χ, defined in eq. (37), as a function of the parameters Λ and g_0, for the symmetric (left) and orthogonal (center) configurations as well as the soft gluon limit (right) for the SU(2) case.

SU(2) fits for different configurations

Kinematic configuration	g_0	Λ (GeV)	χ
Symmetric	3.90	2.40	0.0029
Orthogonal	3.85	2.90	0.0028
Soft gluon	3.95	3.90	0.0044
Sym. + Orth. + Soft	3.90	2.75	0.0036

TABLE II: Values of the parameters g_0 and Λ that minimize the discrepancy between the RGZ outcome of the function G^{CCA} and the corresponding lattice data for various kinematic configurations in the SU(2) case. The last row refers to the values that minimize the joint error χ_{joint}, defined in eq. (39).

The parameters that minimize χ_{joint} are $g_0 = 3.90$ and $\Lambda = 2.75$ GeV. Using these parameter values, we collect in table III the χ values for each kinematic configuration, which indicate that the deterioration of the individual fits for each configuration is minimal when performing the collective fit. The corresponding plots are shown in fig. 4.

SU(2) joint fit

Kinematic configuration	χ
Symmetric	0.0031
Orthogonal	0.0028
Soft gluon	0.0046

TABLE III: Values of the parameters χ of each kinematic configuration corresponding to the parameters that minimize the joint error χ_{joint}, $g_0 = 3.90$ and $\Lambda = 2.75$GeV, in the SU(2) case.

We observe that the fit of fig. 4 exhibits a strong agreement with lattice simulations across all kinematic configuration. However, the UV tails of the RGZ curves consistently fall above the lattice data points. This discrepancy could arise from the larger number of lattice points in the IR region compared to the UV domain, emphasizing the agreement between the RGZ result and lattice data in the low momentum zone at the expense of deteriorating the fit in the UV.

In fig. 4 we include also the results of the Curci-Ferrari (CF) model in Landau gauge at one- and two-loop accuracy. The outcomes from both frameworks, CF and RGZ, are consistent. Nonetheless, within this analysis, comparing both approaches must be done with care due to two reasons. Firstly, concerning the symmetric and orthogonal configurations, the CF curves of G^{CCA} are pure predictions of the model, while the RGZ curves represent a global fit of the lattice data. This renders the comparison fairer in the case of the soft-gluon limit, where the results from the CF approach come from a global fit of the two-point functions and G^{CCA}. Secondly, due to the complexity of the RGZ approach, we employ a simplified model for the RG flow of the gauge coupling, disregarding the RG effects of other parameters of the theory, which could potentially alter the final result.

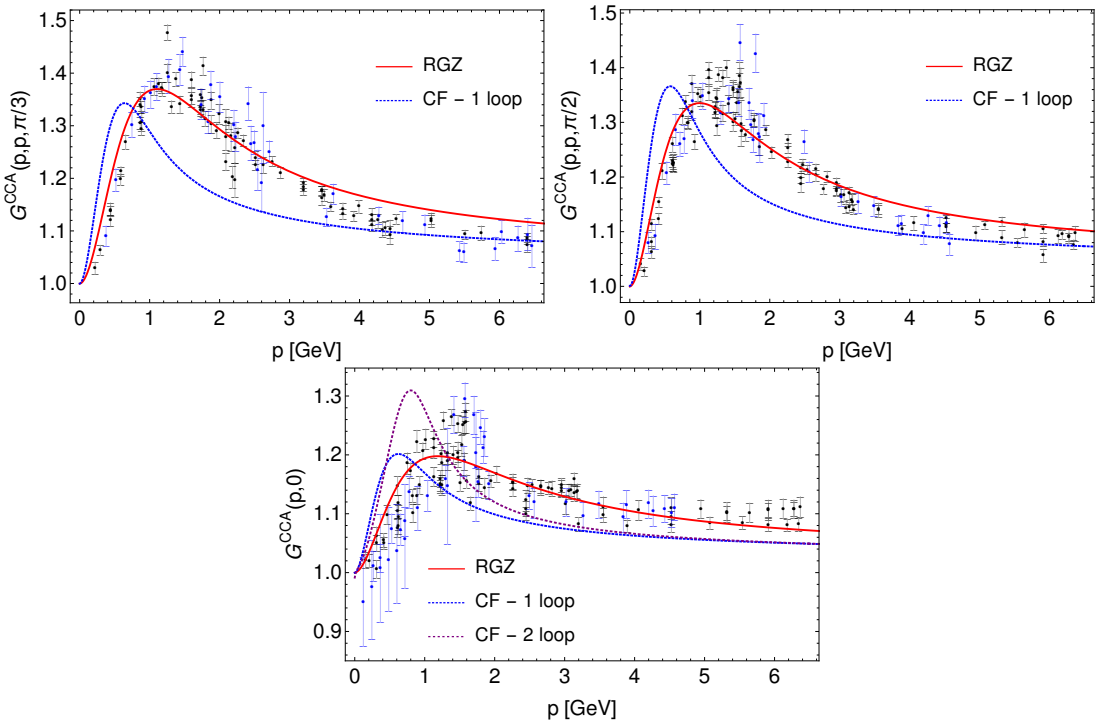


FIG. 4: The scalar function G^{CCA} for the SU(2) case as a function of the antighost external momentum in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom). The red curve represents the RGZ fit of the lattice data by using the toy model of eq. (38) for the gauge coupling. The values of the parameters, $\Lambda = 2.75$ GeV and $g_0 = 3.9$, are chosen to minimize the joint relative error χ_{joint}. Regarding the symmetric and orthogonal configurations, the blue curve is the prediction from the CF model at one-loop order [29]. In the soft gluon limit, the CF results are global fits of the gluon and ghost propagators along with the dressing function G^{CCA} at one- and two-loop order [35]. Blue and black points refer to lattice data extracted from refs. [69] and [70], respectively.

2. *Contribution of each diagram to the G^{CCA} fit*

We now examine the contribution of each diagram to the full result of G^{CCA} in the symmetric and orthogonal configurations, as well as in the soft-gluon limit. The result is presented in fig. 5, where we follow the numbering from fig. 1. As for the quantity we denote as diagram (IV), it corresponds to the full contribution $\frac{2g\gamma^2 f^{ade}}{k^2+M^2}\Gamma_{\bar{c}^b c^c \varphi^{de}_\mu}(k,p,r)$, see eq. (16).

We note that genuinely RGZ diagrams, (III) and (IV), represent a small contribution at one-loop order as compared to the largest contribution of diagram (I) and, to a lesser degree, diagram (II). This is in agreement with the results obtained by CF model [27–37] where only QCD-like diagrams with massive gluons are taken into account. Interestingly, the effect of diagrams (III) and (IV) is a reduction of the peak of G^{CCA}, at around $p = 1.5$ GeV.

B. SU(3)

In this section we replicate the analysis performed for the SU(2) gauge group but for the case of the SU(3) gauge group. Yet, the conclusions derived from this analysis will be more restricted compared to the SU(2) case due to the lack of lattice data available in kinematic configurations other than the soft gluon scenario.

In what follows we present results for the same configurations analyzed in the previous section, *i.e.* the symmetric and orthogonal configurations as well as in the soft gluon limit. Although the only configuration with available lattice data is the latter, the other configurations are valuable for comparison with alternative approaches in the continuum.

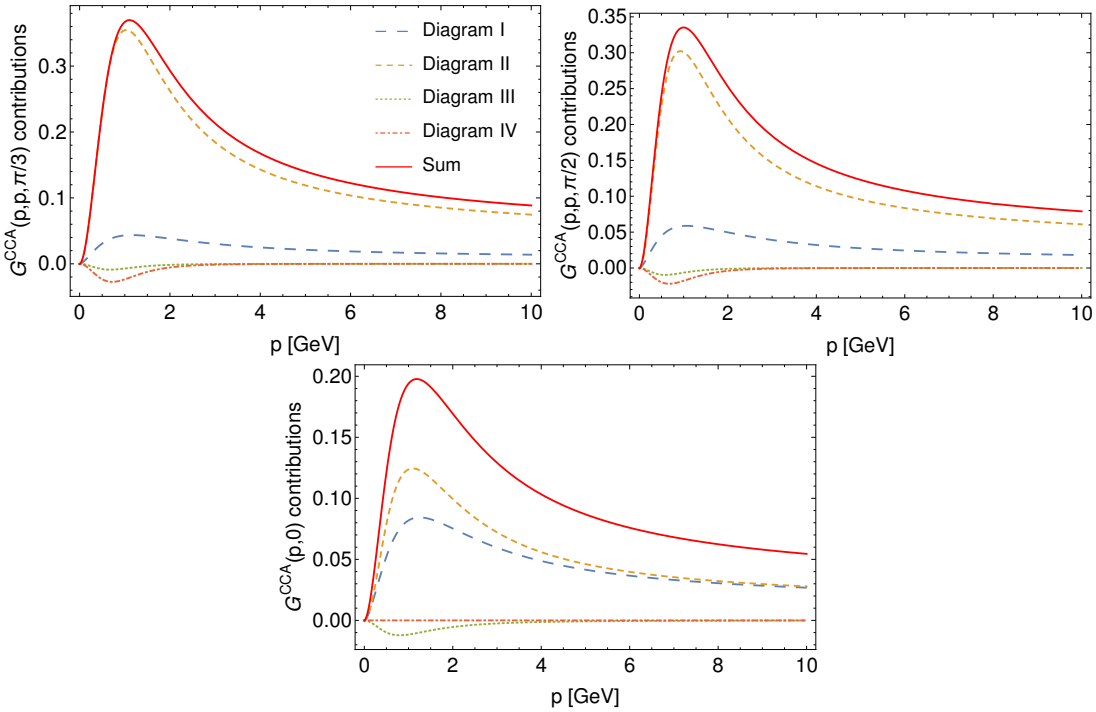


FIG. 5: Contribution of each diagram to G^{CCA} for the SU(2) case in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom), using the gauge coupling described by eq. (38). The values of the parameters, $\Lambda = 2.75$ GeV and $g_0 = 3.90$, are chosen to minimize the joint relative error χ_{joint}. The convention of colors and line-styles in the orthogonal and soft configurations is the same as the symmetric case. The red curve is the sum of all diagrams. The numbering of the diagrams is provided in fig. 1. Diagram (IV) refers to the full contribution $\frac{2g\gamma^2 f^{ade}}{k^2+M^2}\Gamma_{\bar{c}^b c^c \varphi^{de}_\mu}(k,p,r)$.

1. *Fitting the ghost-gluon vertex dressing function*

In fig. 6 we see G^{CCA} as a function of the antighost momentum for the symmetric and orthogonal configurations as well as in the soft gluon limit. Similarly to what we observed in the case of the SU(2) gauge group, as for the soft gluon limit it seems that we are capable of reproducing the lattice data by using a fixed value of the gauge coupling g, as was already pointed out in [1].

It is worth examining how the deviation between the RGZ outcome and lattice data is modified as we change the parameters Λ and g_0 when using the simplified model (eq. (38)). The level curves illustrating this dependence are depicted in fig. 7, displaying an error χ independent of the scale Λ as long as the values of g_0 remain moderate. This reinforces our previous observation that, within the soft gluon limit, the RGZ framework effectively reproduces lattice data with a fixed value of g. Nevertheless, as we aim to analyze the symmetric and orthogonal configurations as well, which are likely sensitive to RG effects, in what follows we continue employing the scale-dependent gauge coupling defined by eq. (38).

In analogy to our analysis for the SU(2) case, in order to find the best fit to lattice data, we find the parameters Λ and g_0 that minimize the relative error χ. In the soft gluon setup, the minimum of χ corresponds to $\Lambda = 0.70$ GeV and $g_0 = 2.15$, with a relative error of $\chi = 0.0040$. The resulting plots are depicted in fig. 8.

We note that our results for G^{CCA} are consistent with the outcomes of the CF model [29, 35] and DSE [15]. Furthermore, the RGZ fit shows a strong agreement with simulations, resembling the results obtained with a fixed value of the gauge coupling. Interestingly, while the UV tails of the DSE and CF curves overlap in the UV, the UV tails from the RGZ curves fall more rapidly in comparison.

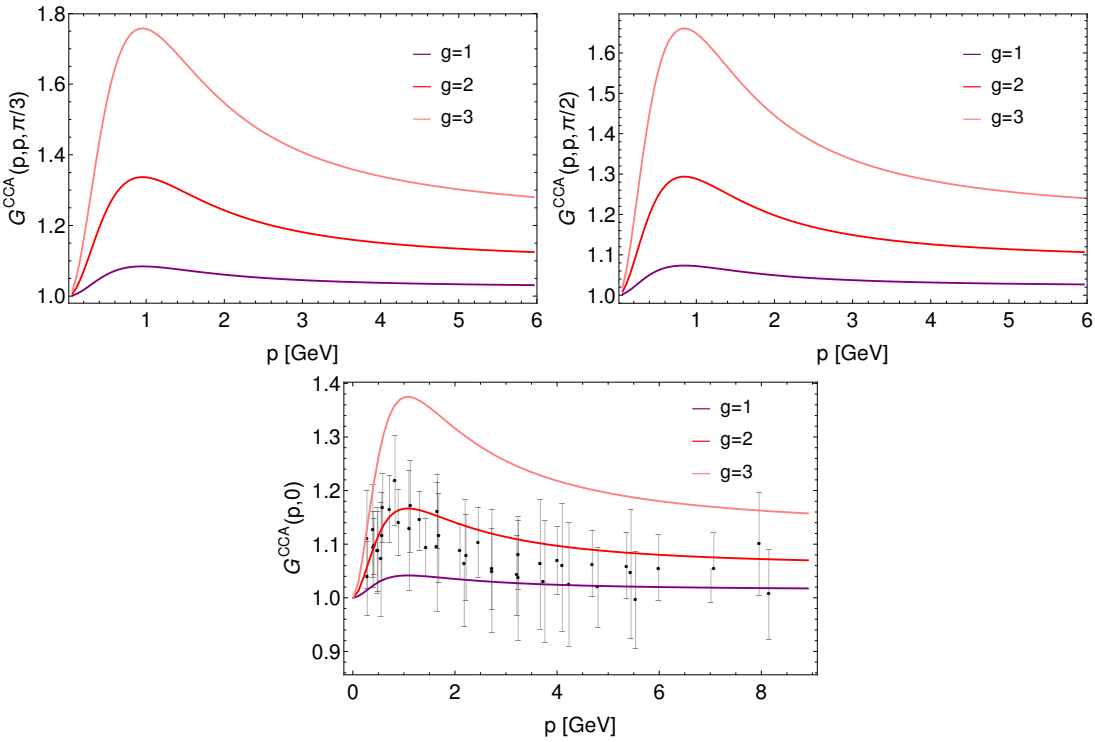


FIG. 6: The scalar function G^{CCA} for the SU(3) case as a function of the antighost external momentum in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom) for various values of a fixed gauge coupling. The lattice data were extracted from the plots in [71] using WebPlotDigitizer [72]. We estimated the error of the extraction procedure to be at most 0.8%.

.

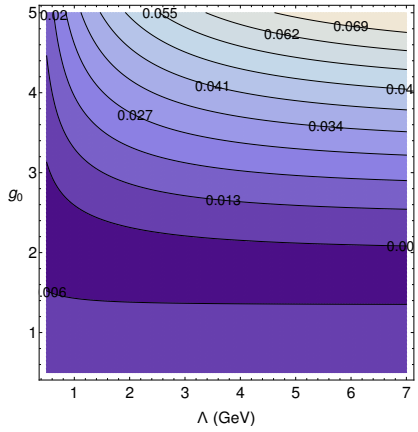


FIG. 7: Level curves of the relative error χ, defined in eq. (37), as a function of the parameters Λ and g_0, as for the soft gluon limit in the SU(3) case.

.

2. *Contribution of each diagram to the G^{CCA} fit*

To conclude this section, we investigate the contribution of each diagram to ghost-gluon vertex dressing function, which is illustrated in fig. 9. Proceeding similarly to the SU(2) case, the quantity we call diagram (IV) corresponds to the full contribution $\frac{2g\gamma^2 f^{ade}}{k^2+M^2}\Gamma_{\bar{c}^b c^c \varphi^{de}_\mu}(k,p,r)$, see eq. (16).

We observe that in the case of the symmetric and orthogonal configurations the behaviour of G^{CCA} is dominated by diagram (II), while the contribution of diagram (I) is counterbalanced by the contributions of the genuine RGZ

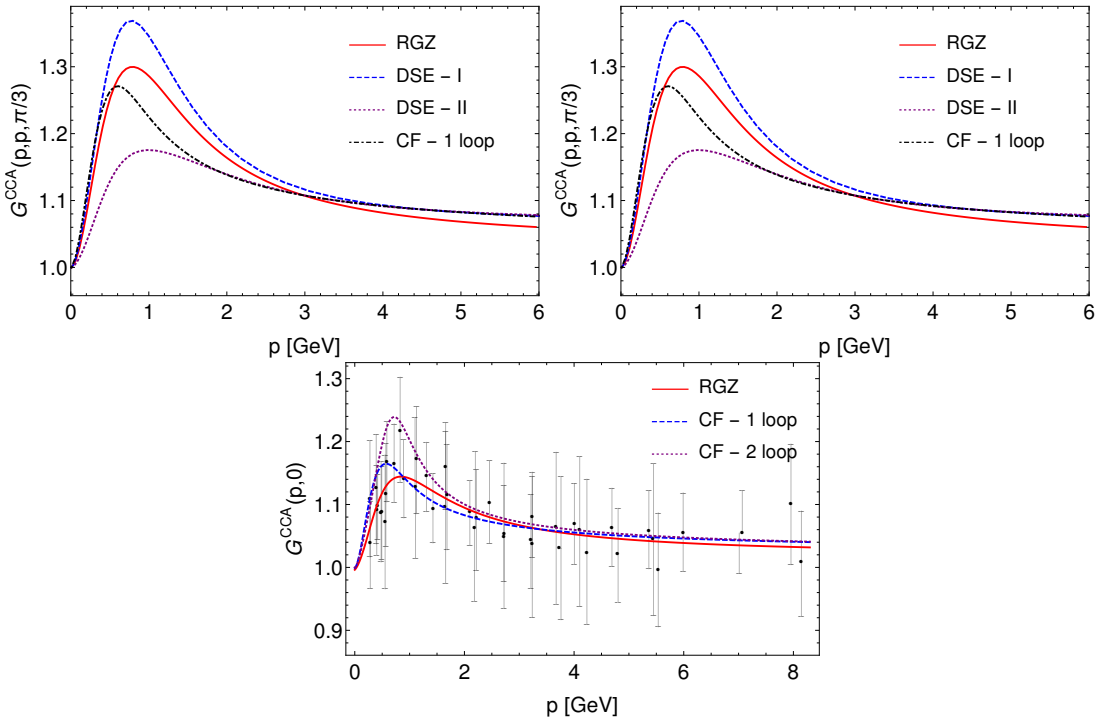


FIG. 8: The scalar function G^{CCA} for the SU(3) case as a function of the antighost external momentum in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom). The red curve corresponds to the RGZ fit of the lattice data, by utilizing the toy model of eq. (38) for the gauge coupling. Parameters $\Lambda = 0.70$ GeV and $g_0 = 2.15$ were chosen to minimize the joint relative error with the lattice data in the case of the soft gluon limit. Lattice data are extracted from [71]. We compare the results with other approaches in the continuum. Specifically with G^{CCA} obtained from DSE, considering the bare (DSE - I) and the dressed three-gluon vertex (DSE - II) [15] and G^{CCA} from the CF model at one- and two-loop order [29, 35].

diagrams (III) and (IV). Overall, the relative weights of the these contributions do not significantly differ from the SU(2) case.

VI. INFLUENCE OF THE TOY MODEL ON THE UV BEHAVIOR OF THE FUNCTION G^{CCA}

As mentioned earlier, when comparing the UV tails of the dressing function G^{CCA} obtained from the RGZ framework with other approaches, cf. figs. 4 and 8, we observe differences. Particularly striking is the observation in the case of SU(3), where a simultaneous comparison between the RGZ, the CF model and DSE results reveals a convergence in the tails from the latter approaches, starting from $p = 4$ GeV onwards. In contrast, the UV tails of the curves from the RGZ framework decrease more rapidly. It is legitimate then to ask for the potential causes of this discrepancy in the UV. In this section we explore the extent to which this difference may derive from the toy model of the RG flow, as presented in eq. (38), and its associated parameters Λ and g_0.

With that purpose we compare in fig. 10 the running of the gauge coupling employed for the CF prediction of G^{CCA} at one-loop order in figs. 4 and 8 [5], denoted as $g_{\rm CF}(\mu)$, with the running of the gauge coupling from eq. (38), $g_{\rm toy}(\mu)$, with parameters Λ and g_0, obtained from the fits for the gauge groups SU(2) and SU(3). We observe that $g_{\rm CF}(\mu)$ and $g_{\rm toy}(\mu)$ differ at all scales and, although the difference tends to diminish at high momenta, it is not negligible in the UV region.

An interesting analysis involves assessing how the differences among the UV tails of G^{CCA} vary when we minimize the discrepancy between $g_{\rm CF}(\mu)$ and $g_{\rm toy}(\mu)$ in the UV domain. For this purpose, it is convenient to introduce the quantity $\chi_{g,UV}$, as a way to estimate that discrepancy:

[5] In this section, when discussing the running coupling of the CF model for SU(2), we refer to the running coupling employed for the symmetric and orthogonal configurations, which differs from the one utilized in the soft gluon limit.

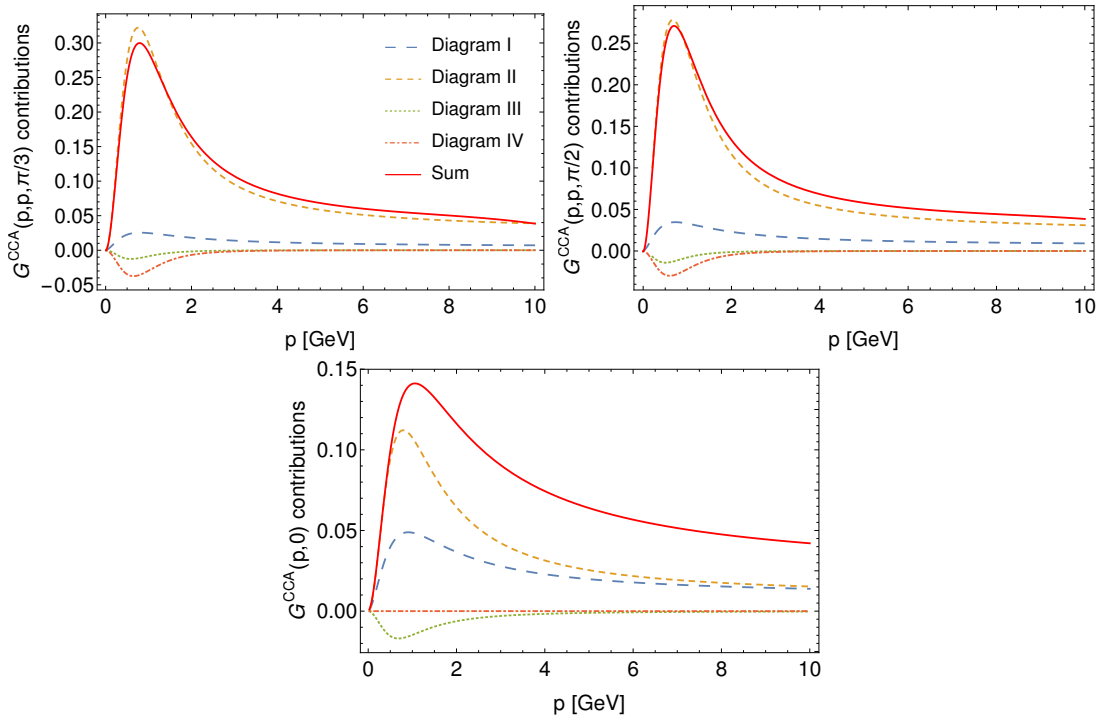


FIG. 9: Contribution of each diagram to G^{CCA} in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom), using the gauge coupling described by eq. (38). Parameters, $\Lambda = 0.70$ GeV and $g_0 = 2.15$, are chosen to minimize the relative error χ in the soft gluon limit. The convention of colors and line-styles in the orthogonal and soft configurations is the same as the symmetric case. The solid, red curve is the sum of all diagrams. The numbering of the diagrams is provided in fig. 1. Diagram (IV) refers to the full contribution $\frac{2g\gamma^2 f^{ade}}{k^2+M^2}\Gamma_{\bar{c}^b c^c \varphi^{de}_\mu}(k,p,r)$.

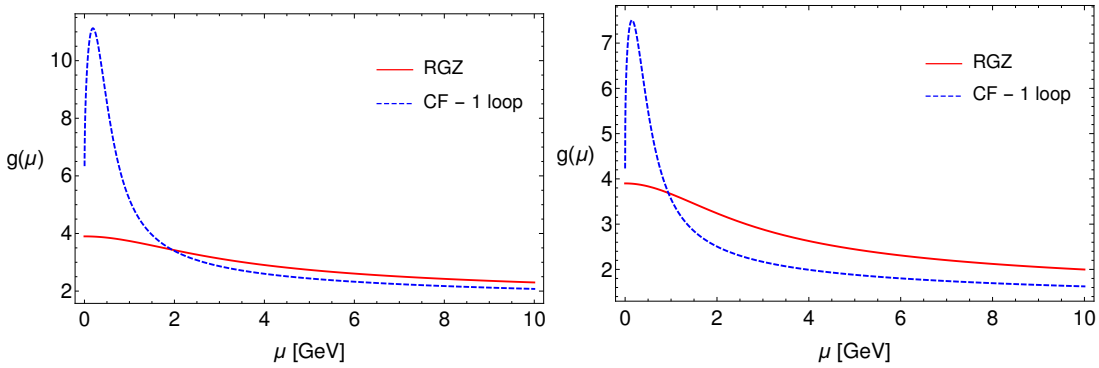


FIG. 10: The running coupling $g_{\rm toy}$ and $g_{\rm CF}$ for the SU(2) (left) and SU(3) (right) gauge groups. As for the toy model, the values of the parameters are $\Lambda = 2.4$, GeV, $g_0 = 3.9$ and $\Lambda = 0.70$ GeV, $g_0 = 2.15$, for SU(2) and SU(3), respectively.

$$\chi^2_{g,UV} \equiv \sum_{p_i=6\ \text{GeV}}^{p_i=10\ \text{GeV}} \left(\frac{g_{\rm toy}(p_i) - g_{\rm CF}(p_i)}{g_{\rm CF}(p_i)} \right)^2 . \tag{40}$$

To carry out the summation we used uniform steps of length $\Delta p = p_{n+1} - p_n = 0.5$ GeV. The parameters that minimize $\chi_{g,UV}$ are $\Lambda = 1.05$ GeV, $g_0 = 6.75$ for the SU(2) gauge group and $\Lambda = 0.90$ GeV, $g_0 = 4.85$ for SU(3). These parameters lead to an excellent agreement between $g_{\rm toy}(\mu)$ and $g_{\rm CF}(\mu)$ in the UV, as is illustrated in fig. 11 (black curve). Unfortunately, these values do not feature good compatibility with the lattice data of G^{CCA}.

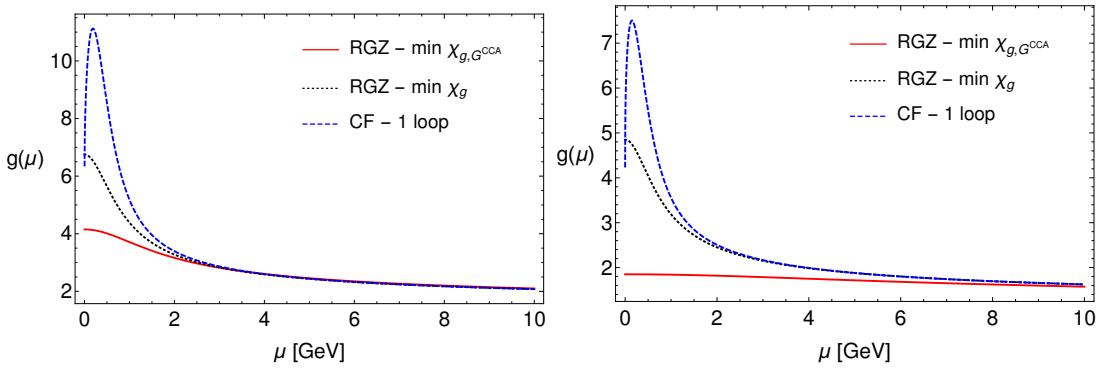


FIG. 11: The running coupling $g_{\rm toy}$ and $g_{\rm CF}$ for the SU(2) (left) and SU(3) (right) gauge groups. The black curve corresponds to the parameters that minimize $\chi^2_{g,UV}$. These are $\Lambda = 1.05$ GeV, $g_0 = 6.75$ and $\Lambda = 0.90$ GeV, $g_0 = 4.85$, for SU(2) and SU(3) respectively. The red curve corresponds to the parameters that minimize $\chi_{g,G^{CCA}}$. These are $\Lambda = 1.65$ GeV, $g_0 = 4.15$ and $\Lambda = 5.05$ GeV, $g_0 = 1.85$, for SU(2) and SU(3) respectively.

To assess the potential for enhancing agreement between the results from the CF model and DSE with the RGZ framework without significantly compromising the IR description of lattice data, we introduce the following collective error

$$\chi^2_{g,G^{CCA}} = \frac{1}{2}\Big(\chi^2_{g,UV} + \chi^2_{G^{CCA}}\Big), \tag{41}$$

where $\chi_{G^{CCA}}$ is defined according to eq. (39) and eq. (37) for the SU(2) and SU(3) gauge groups, respectively. The quantity $\chi_{g,G^{CCA}}$ takes into account both the discrepancy between $g_{\rm CF}(\mu)$ and $g_{\rm toy}(\mu)$ in the UV as well as the discrepancy between the dressing function G^{CCA} from the RGZ theory and the corresponding lattice data.

The parameters Λ and g_0, derived from minimizing $\chi_{g,G^{CCA}}$, would in principle generate a gauge coupling $g_{\rm toy}(\mu)$ that balances both the UV alignment with $g_{\rm CF}(\mu)$ and the consistency with lattice data across all momentum scales. Such running coupling is also depicted in fig. 11 (red curve). In the case of the SU(2) gauge group, we notice an enhancement in the agreement in the UV region between RGZ and CF/DSE. Notably, this improvement does not imply a significant deterioration in the IR description. This is evident when comparing fig. 12 with fig. 4.

In the case of SU(3), the minimization of $\chi_{G^{CCA}}$ leads to a running of the gauge coupling which is essentially constant. This is consistent with the plots of figs. 4 and 8, where a fixed value of g was capable of describing the lattice data for the entirety range of momentum. However, the aforementioned running does not reduce the disagreement between the results from the CF model and DSE with the RGZ framework outcomes in the UV, being the quality of the fits similar to the ones presented in figs. 4 and 8. This is reasonable, since we are essentially neglecting the dependence of g with the scale. Still, the results for SU(3) should be taken with care due to the dispersion of lattice data in the soft gluon limit and the lack of data in alternative kinematic setups.

We can conclude that the specific toy model we employed has an impact on the differences between the RGZ framework and other approaches in the UV region. However, these discrepancies probably have other sources as well such as the neglect of the RG flow of the remaining parameters of the theory and the specific choice of the renormalization scheme.

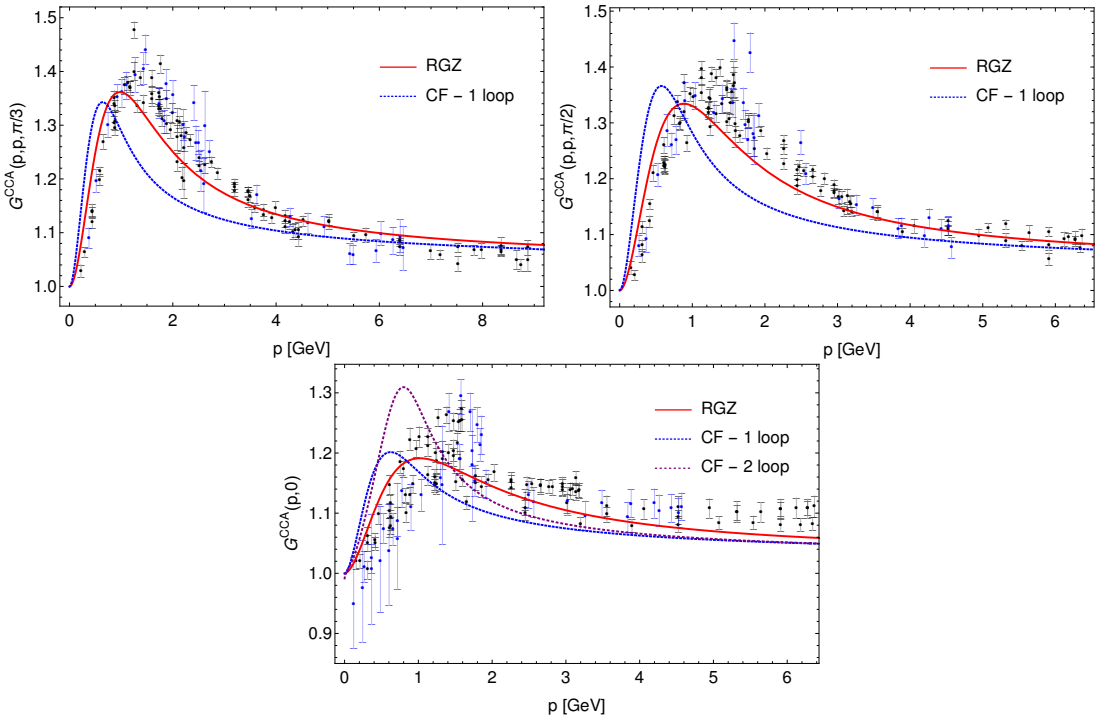


FIG. 12: The scalar function G^{CCA} as a function of the antighost external momentum in three distinct configurations: the symmetric (top left), orthogonal (top right) and the soft gluon limit (bottom) in the case of the SU(2) theory. The gauge coupling is described by the toy model of eq. (38) with parameters that minimize $\chi_{g,G^{CCA}}$, $\Lambda = 1.65$ GeV and $g_0 = 4.15$.

VII. SUMMARY AND FINAL REMARKS

In this work we evaluated the one-loop ghost-antighost-gluon correlation function within the RGZ framework for an arbitrary kinematic configuration and investigated the role of the running coupling. We performed the calculation for the pure gauge theory in a four dimensional Euclidean spacetime. With the purpose of comparing with other approaches, we specifically focused on three distinct kinematic configurations: the symmetric and orthogonal configurations and the soft gluon limit.

This calculation is important since it serves as a benchmark to assess how well the RGZ framework describes the behavior of YM theory in the low momentum regime. In this sense this paper comes to complement previous investigations on the two-point functions at tree-level order [43, 44] and the analysis of the ghost-gluon vertex in the case of the soft-gluon limit at one-loop order [1].

The only parameter of the theory we used to fit our results to the available lattice data for the ghost-antighost-gluon vertex was the gauge coupling. All the other parameters were extracted from refs. [67] and [68], where they were determined by fitting lattice results for the gluon propagator using the RGZ tree level expression. This choice reduces the number of free parameters to the minimum, making the present analysis a more strict test of the RGZ framework.

As already pointed out in [1], as for the soft gluon limit the results are compatible with lattice simulations for both SU(2) [69, 70] and SU(3) [71] by simply considering a constant value of the gauge coupling, *i.e.* independent of the momentum scale. However, this is not the case of the symmetric and orthogonal configurations, where the renormalization group effects seem to be significant. Indeed, when introducing a toy model for the running of the gauge coupling, our outcomes are quantitatively compatible with available lattice data.

Our results also display qualitative agreement with other continuum approaches, specifically the dynamical DSE solutions in two different truncation schemes [15] and the RG-improved CF model [29, 35]. Additionally, we show that quantitative differences between the RGZ results and alternative continuum approaches in the UV region can be accounted for by the particularities of the model for the RG flow of the gauge coupling.

Overall, these results support the RGZ theory as a valid description of the infrared YM dynamics. There are several ways, nevertheless, in which the present analysis can be improved. A fully dynamical calculation of the two-point functions of the RGZ framework, most likely in an infrared safe scheme, would allow us to determine the running parameters of the theory by fitting the two-point functions to lattice data. Another potential study concerning this

correlation function could investigate its dependence on the gauge parameter within the context of linear covariant gauges, utilizing the BRST invariant version of the RGZ action [54].

ACKNOWLEDGEMENTS

The authors acknowledge useful discussions with M. Tissier and N. Wschebor. The authors also would like to thank A. C. Aguilar and A. Maas for discussions and for kindly providing the data for the comparison plots in Section IV. This work has been partially supported by CNPq, CAPES, FAPERJ, PEDECIBA and the ANII-FCE-166479 project. It is a part of the project INCT-FNA Proc. 464898/2014-5. M. S. Guimaraes is a CNPq researcher under contract 310049/2020-2. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Appendix A: Relevant Feynman rules (Landau gauge)

Given its many fields and interactions, the RGZ theory has a large number of propagators and vertices. However, for the calculation of the ghost-gluon vertex at one-loop level (and in the Landau gauge), only a few of them are required. The Feynman rules corresponding to these propagators and vertices are shown below.

1. Tree-level propagators

In order to calculate the ghost-antighost-gluon 3-point function in the Refined Gribov-Zwanziger theory, only a subset of the propagators of the theory are needed. These are

$$\langle A^a_\mu(p)A^b_\nu(-p)\rangle = \delta^{ab}\left[\frac{p^2+M^2}{p^4+(m^2+M^2)p^2+m^2M^2+2Ng^2\gamma^4}P^\perp_{\mu\nu}(p)\right] \equiv \delta^{ab}P^\perp_{\mu\nu}(p)D_{AA}(p) \tag{A1}$$

$$\langle A^a_\mu(p)\varphi^{bc}_\nu(-p)\rangle = \frac{-ig\gamma^2 f^{abc}}{p^4+p^2(m^2+M^2)+m^2M^2+2Ng^2\gamma^4}P^\perp_{\mu\nu}(p) \tag{A2}$$

$$\langle A^a_\mu(p)\bar{\varphi}^{bc}_\nu(-p)\rangle = \frac{-ig\gamma^2 f^{abc}}{p^4+p^2(m^2+M^2)+m^2M^2+2Ng^2\gamma^4}P^\perp_{\mu\nu}(p) = \langle A^a_\mu(p)\varphi^{bc}_\nu(-p)\rangle \tag{A3}$$

$$\langle \bar{c}^a(p)c^b(-p)\rangle = \frac{1}{p^2}\delta^{ab} \equiv \delta^{ab}D_{\bar{c}c}(p) \tag{A4}$$

where

$$P^\perp_{\mu\nu}(p) = \delta_{\mu\nu} - \frac{p_\mu p_\nu}{p^2} \tag{A5}$$

is the transverse projector, such that $p_\mu P^\perp_{\mu\nu}(p) = p_\nu P^\perp_{\mu\nu}(p) = 0$.

It is often convenient to write the D_{AA} form factor as a sum of massive propagators, i.e.,

$$D_{AA}(p^2) = \frac{p^2+M^2}{(p^2+m^2)(p^2+M^2)+\lambda^4} \equiv \frac{R_+}{p^2+a_+^2} + \frac{R_-}{p^2+a_-^2}, \tag{A6}$$

with

$$\begin{aligned}
a_+^2 &= \frac{m^2+M^2+\sqrt{(m^2-M^2)^2-4\lambda^4}}{2},\\
a_-^2 &= \frac{m^2+M^2-\sqrt{(m^2-M^2)^2-4\lambda^4}}{2},\\
R_+ &= \frac{m^2-M^2+\sqrt{(m^2-M^2)^2-4\lambda^4}}{2\sqrt{(m^2-M^2)^2-4\lambda^4}},\\
R_- &= \frac{-m^2+M^2-\sqrt{(m^2-M^2)^2-4\lambda^4}}{2\sqrt{(m^2-M^2)^2-4\lambda^4}} = 1-R_+,
\end{aligned} \tag{A7}$$

with $\lambda^4 = 2Ng^2\gamma^4$.

Note that either the poles a_+ and a_- are complex, or the residues R_+ and R_- have opposite signs (so that one of them is negative). This property can be directly related to positivity violation of the gluon propagator in the RGZ theory.

2. Tree-level vertices

The only vertices needed for the computation of the ghost-antighost-gluon at one-loop in the RGZ theory are

$$\begin{aligned}
{}^{\rm tree}[\Gamma_{AAA}(k,p,q)]^{abc}_{\mu\nu\rho} &= -\left.\frac{\delta^3 S_{\rm tree}}{\delta A^a_\mu(k)\delta A^b_\nu(p)\delta A^c_\rho(q)}\right|_{\Phi=0} = igf^{abc}\left[(k_\nu - q_\nu)\delta_{\rho\mu} + (p_\rho - k_\rho)\delta_{\mu\nu} + (q_\mu - p_\mu)\delta_{\nu\rho}\right], \\
{}^{\rm tree}[\Gamma_{A\bar{c}c}(k,p,q)]^{abc}_{\mu} &= -\left.\frac{\delta^3 S_{\rm tree}}{\delta A^a_\mu(k)\delta \bar{c}^b(p)\delta c^c(q)}\right|_{\Phi=0} = -igf^{abc}p_\mu, \\
{}^{\rm tree}[\Gamma_{A\bar{\varphi}\varphi}(k,p,q)]^{abcde}_{\mu\nu\rho} &= -\left.\frac{\delta^3 S_{\rm tree}}{\delta A^a_\rho(k)\delta \bar{\varphi}^{bc}_\mu(p)\delta \varphi^{de}_\nu(q)}\right|_{\Phi=0} = -igf^{abd}\delta^{ce}\delta_{\nu\rho}p_\mu .
\end{aligned} \tag{A8}$$

Note that, for higher orders in the perturbative expansion, or for a general linear covariant gauge, or for other correlation functions, extra correlators and vertices will be needed.

Appendix B: On the relation between connected and 1PI correlation functions in the presence of mixed propagators

Let us denote the generating functional of connected correlation functions as $W[\vec{J}]$, where J_i are external sources associated with the different elementary fields, and let $\Gamma[\vec{\phi}]$ be the quantum action, that is, the generating functional of 1PI correlation functions. Using this notation, we start from the well-known relation

$$\left.\frac{\delta^2\Gamma[\vec{\phi}]}{\delta\phi_j\delta\phi_\ell}\right|_{\vec{\phi}=\vec{\Phi}[\vec{J}]}\frac{\delta^2 W[\vec{J}]}{\delta J_\ell \delta J_k} = -\delta_{jk} \tag{B1}$$

Taking a further derivative with respect to the source J_i, one finds

$$\frac{\delta^3 W[\vec{J}]}{\delta J_i\delta J_p\delta J_k} = -\frac{\delta^2 W[\vec{J}]}{\delta J_p\delta J_j}\left(\left.\frac{\delta^3\Gamma[\vec{\phi}]}{\delta\phi_j\delta\phi_\ell\delta\phi_m}\right|_{\vec{\phi}=\vec{\Phi}[\vec{J}]}\right)\frac{\delta^2 W[\vec{J}]}{\delta J_i\delta J_m}\frac{\delta^2 W[\vec{J}]}{\delta J_\ell\delta J_k}. \tag{B2}$$

For the present calculation of the ghost-antighost-gluon vertex, we are specifically interested in the choice

$$\begin{aligned}
i &= A^e_\mu(k) \\
p &= \bar{c}^a(p) \\
k &= c^b(q).
\end{aligned} \tag{B3}$$

Since there are no mixed propagators involving the Faddeev-Popov ghosts c and $\bar{c}$, the only nonvanishing contributions are such that $j = c$ and $\ell = \bar{c}$. Therefore, running the remaining sum for $m = A, \varphi, \bar{\varphi}$, we have

$$\begin{aligned}
\frac{\delta^3 W[\vec{J}]}{\delta J_A\delta J_{\bar{c}}\delta J_c} = -\frac{\delta^2 W[\vec{J}]}{\delta J_c\delta J_{\bar{c}}}\Bigg\{&\left(\left.\frac{\delta^3\Gamma[\vec{\phi}]}{\delta c\,\delta\bar{c}\,\delta A}\right|_{\vec{\varphi}=\vec{\Phi}[\vec{J}]}\right)\frac{\delta^2 W[\vec{J}]}{\delta J_A\delta J_A} + \\
&+\left(\left.\frac{\delta^3\Gamma[\vec{\phi}]}{\delta c\,\delta\,\bar{c}\,\delta\varphi}\right|_{\vec{\varphi}=\vec{\Phi}[\vec{J}]}\right)\frac{\delta^2 W[\vec{J}]}{\delta J_A\delta J_\varphi} + \left(\left.\frac{\delta^3\Gamma[\vec{\phi}]}{\delta c\,\delta\,\bar{c}\,\delta\bar{\varphi}}\right|_{\vec{\varphi}=\vec{\Phi}[\vec{J}]}\right)\frac{\delta^2 W[\vec{J}]}{\delta J_A\delta J_{\bar{\varphi}}}\Bigg\}\frac{\delta^2 W[\vec{J}]}{\delta J_{\bar{c}}\delta J_c},
\end{aligned} \tag{B4}$$

which, at the one-loop level, can be written as

$$\left\langle \bar{c}^a(p)\, c^b(q)\, A^c_\mu(k)\right\rangle = D_{\bar{c}c}(p)D_{\bar{c}c}(q)D_{AA}(k)P_{\mu\nu}(k)\left\{\frac{\delta^3\Gamma}{\delta c^a(-p)\delta\bar{c}^b(-q)\delta A^c_\nu(-k)} + \frac{2g\gamma^2 f^{cde}}{k^2+M^2}\frac{\delta^3\Gamma}{\delta c^a(-p)\delta\bar{c}^b(-q)\delta\varphi^{de}_\nu(-k)}\right\}, \tag{B5}$$

where we used the fact that the correlation functions obey $\langle A\varphi\rangle = \langle A\bar{\varphi}\rangle$, and also $\delta^3\Gamma/\delta c\delta\bar{c}\delta\varphi = \delta^3\Gamma/\delta c\delta\bar{c}\delta\bar{\varphi}$.

In a shorthand notation, one can conveniently write

$$\frac{\langle A\,\bar{c}\,c\rangle_c}{(\langle\bar{c}\,c\rangle_c)^2\,\langle AA\rangle_c} = \Gamma_{A\,\bar{c}\,c} + \frac{\langle A\,\varphi\rangle_c}{\langle A\,A\rangle_c}\Gamma_{\bar{c}\,c\,\varphi} + \frac{\langle A\,\bar{\varphi}\rangle_c}{\langle A\,A\rangle_c}\Gamma_{\bar{c}\,c\,\bar{\varphi}}. \tag{B6}$$

Therefore, besides the contribution $\Gamma_{A\bar{c}c}$, that would be present at pure Yang-Mills, there are also contributions from 1PI functions involving the auxiliary fields φ and $\bar{\varphi}$, as well as the respective mixed propagators. Such contributions can be thought as momentum-dependent contributions to the ghost-gluon vertex, which are present starting from the tree level of the RGZ action.

Appendix C: Analytic expression for the ghost-gluon vertex in the soft-gluon limit

In the soft-gluon limit (i.e., when the gluon momentum $k \to 0$), the expression (20) for the ghost-gluon vertex is relatively simpler than in a general kinematic regime. Given the absence of IR-divergences, the soft-gluon limit for the ghost-gluon vertex reads

$$\Gamma_{A^a_\mu\bar{C}^bC^c}(0,p,-p) = -igf^{abc}p_\mu B_1(0,p). \tag{C1}$$

The scalar function $B_1(0,p)$ has been calculated (with a slightly different notation) to one-loop order in [1]. The result can be written as[6]

$$\begin{aligned} B_1(0,p) \;=\; & 1 + \frac{Ng^2}{2}\left[R_+J(a_+;p) + R_-J(a_-;p)\right] - Ng^2\left[R_+^2K(a_+,a_+;p) + R_-^2K(a_-,a_-;p) + 2R_+R_-K(a_+,a_-;p)\right] \\ & + \frac{Ng^2}{2}\frac{Ng^2\gamma^4}{(a_+^2-a_-^2)^2}\left[K(a_+,a_+;p) + K(a_-,a_-;p) - 2K(a_+,a_-;p)\right], \end{aligned} \tag{C2}$$

where the poles $a_\pm$ and residues $R_\pm$ of the tree-level gluon propagator are given by eq. (A7), and the scalar functions J and K above are given by

$$J(m_1;p) \;=\; \frac{1}{64\pi^2}\times\frac{2m_1^2p^2(p^2+m_1^2) + p^6\log\left(1+\frac{m_1^2}{p^2}\right) - (3p^2+2m_1^2)m_1^4\log\left(1+\frac{p^2}{m_1^2}\right)}{m_1^2p^4}, \tag{C3}$$

$$\begin{aligned} K(m_1,m_2;p) \;=\; & \frac{1}{256\pi^2}\frac{1}{m_1^2m_2^2p^4(m_1^2-m_2^2)}\Bigg\{2m_1^2m_2^6p^2 - 2m_1^6m_2^2p^2 + 3m_1^2m_2^4p^4 - 3m_1^4m_2^2p^4 + \\ & + 2m_1^8m_2^2\log\left(1+\frac{p^2}{m_1^2}\right) - 2m_1^2m_2^8\log\left(1+\frac{p^2}{m_2^2}\right) + 4m_1^6m_2^2p^2\log\left(1+\frac{p^2}{m_1^2}\right) - \\ & - 4m_1^2m_2^6p^2\log\left(1+\frac{p^2}{m_2^2}\right) + 4m_1^2m_2^2p^6\log\left(\frac{p^2+m_2^2}{p^2+m_1^2}\right) + \\ & + 2m_1^2p^8\log\left(1+\frac{m_2^2}{p^2}\right) - 2m_2^2p^8\log\left(1+\frac{m_1^2}{p^2}\right)\Bigg\}, \end{aligned} \tag{C4}$$

as for $m_1 \neq m_2$, and

$$K(m,m;p) = \frac{-6p^2m^6 - 5p^4m^4 - 2p^6m^2 + \left(8p^2m^6 + 6m^8\right)\log\left(1+\frac{p^2}{m^2}\right) + 2p^8\log\left(1+\frac{m^2}{p^2}\right)}{256\pi^2p^4m^4}. \tag{C5}$$

[6] There is a typo in the corresponding result from [1], which is corrected in this expression.

[1] B. W. Mintz, L. F. Palhares, S. P. Sorella, and A. D. Pereira, Phys. Rev. D **97**, 034020 (2018), arXiv:1712.09633 [hep-th].
[2] Z. Fodor and C. Hoelbling, Rev. Mod. Phys. **84**, 449 (2012), arXiv:1203.4789 [hep-lat].
[3] S. Durr et al. (BMW), Science **322**, 1224 (2008), arXiv:0906.3599 [hep-lat].
[4] S. Borsanyi et al., Nature **593**, 51 (2021), arXiv:2002.12347 [hep-lat].
[5] S. Borsanyi et al. (BMW), Science **347**, 1452 (2015), arXiv:1406.4088 [hep-lat].
[6] A. Bashir, L. Chang, I. C. Cloet, B. El-Bennich, Y.-X. Liu, C. D. Roberts, and P. C. Tandy, Commun. Theor. Phys. **58**, 79 (2012), arXiv:1201.3366 [nucl-th].
[7] A. C. Aguilar, D. Binosi, and J. Papavassiliou, Front. Phys.(Beijing) **11**, 111203 (2016), arXiv:1511.08361 [hep-ph].
[8] A. C. Aguilar, F. De Soto, M. N. Ferreira, J. Papavassiliou, F. Pinto-Gómez, C. D. Roberts, and J. Rodríguez-Quintero, Phys. Lett. B **841**, 137906 (2023), arXiv:2211.12594 [hep-ph].
[9] A. C. Aguilar, M. N. Ferreira, and J. Papavassiliou, Phys. Rev. D **105**, 014030 (2022), arXiv:2111.09431 [hep-ph].
[10] A. C. Aguilar, C. O. Ambrósio, F. De Soto, M. N. Ferreira, B. M. Oliveira, J. Papavassiliou, and J. Rodríguez-Quintero, Phys. Rev. D **104**, 054028 (2021), arXiv:2107.00768 [hep-ph].
[11] A. C. Aguilar, F. De Soto, M. N. Ferreira, J. Papavassiliou, and J. Rodríguez-Quintero, Phys. Lett. B **818**, 136352 (2021), arXiv:2102.04959 [hep-ph].
[12] M. Q. Huber, C. S. Fischer, and H. Sanchis-Alepuz, Eur. Phys. J. C **81**, 1083 (2021), [Erratum: Eur.Phys.J.C 82, 38 (2022)], arXiv:2110.09180 [hep-ph].
[13] C. S. Fischer and M. Q. Huber, Phys. Rev. D **102**, 094005 (2020), arXiv:2007.11505 [hep-ph].
[14] M. Q. Huber, Phys. Rev. D **101**, 114009 (2020), arXiv:2003.13703 [hep-ph].
[15] A. C. Aguilar, M. N. Ferreira, C. T. Figueiredo, and J. Papavassiliou, Phys. Rev. D **99**, 034026 (2019), arXiv:1811.08961 [hep-ph].
[16] J. Berges, N. Tetradis, and C. Wetterich, Phys. Rept. **363**, 223 (2002), arXiv:hep-ph/0005122 [hep-ph].
[17] J. M. Pawlowski, Annals Phys. **322**, 2831 (2007), arXiv:hep-th/0512261 [hep-th].
[18] L. Corell, A. K. Cyrol, M. Mitter, J. M. Pawlowski, and N. Strodthoff, SciPost Phys. **5**, 066 (2018), arXiv:1803.10092 [hep-ph].
[19] A. K. Cyrol, L. Fister, M. Mitter, J. M. Pawlowski, and N. Strodthoff, Phys. Rev. D **94**, 054005 (2016), arXiv:1605.01856 [hep-ph].
[20] C. S. Fischer, A. Maas, and J. M. Pawlowski, Annals Phys. **324**, 2408 (2009), arXiv:0810.1987 [hep-ph].
[21] N. Dupuis, L. Canet, A. Eichhorn, W. Metzner, J. M. Pawlowski, M. Tissier, and N. Wschebor, Phys. Rept. **910**, 1 (2021), arXiv:2006.04853 [cond-mat.stat-mech].
[22] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961).
[23] S. P. Klevansky, Rev. Mod. Phys. **64**, 649 (1992).
[24] M. Gell-Mann and M. Levy, Nuovo Cim. **16**, 705 (1960).
[25] K. Fukushima, Physics Letters B **591**, 277 (2004), arXiv:hep-ph/0310121.
[26] B.-J. Schaefer, J. M. Pawlowski, and J. Wambach, Phys. Rev. **D76**, 074023 (2007), arXiv:0704.3234 [hep-ph].
[27] M. Tissier and N. Wschebor, Phys. Rev. D **82**, 101701 (2010), arXiv:1004.1607 [hep-ph].
[28] M. Tissier and N. Wschebor, Phys. Rev. **D84**, 045018 (2011), arXiv:1105.2475 [hep-th].
[29] M. Pelaez, M. Tissier, and N. Wschebor, Phys. Rev. **D88**, 125003 (2013), arXiv:1310.2594 [hep-th].
[30] M. Peláez, M. Tissier, and N. Wschebor, Phys. Rev. D **90**, 065031 (2014), arXiv:1407.2005 [hep-th].
[31] M. Peláez, M. Tissier, and N. Wschebor, Phys. Rev. D **92**, 045012 (2015), arXiv:1504.05157 [hep-th].
[32] J. A. Gracey, M. Peláez, U. Reinosa, and M. Tissier, Phys. Rev. D **100**, 034023 (2019), arXiv:1905.07262 [hep-th].
[33] M. Peláez, U. Reinosa, J. Serreau, M. Tissier, and N. Wschebor, Rept. Prog. Phys. **84**, 124202 (2021), arXiv:2106.04526 [hep-th].
[34] F. Figueroa and M. Peláez, Phys. Rev. D **105**, 094005 (2022), arXiv:2110.09561 [hep-th].
[35] N. Barrios, M. Peláez, U. Reinosa, and N. Wschebor, Phys. Rev. D **102**, 114016 (2020), arXiv:2009.00875 [hep-th].
[36] N. Barrios, J. A. Gracey, M. Peláez, and U. Reinosa, Phys. Rev. D **104**, 094019 (2021), arXiv:2103.16218 [hep-th].
[37] N. Barrios, M. Peláez, and U. Reinosa, Phys. Rev. D **106**, 114039 (2022), arXiv:2207.10704 [hep-ph].
[38] F. Siringo, Phys. Rev. D **107**, 016009 (2023), arXiv:2211.06448 [hep-ph].
[39] F. Siringo and G. Comitini, Phys. Rev. D **107**, 096001 (2023), arXiv:2210.11541 [hep-th].
[40] G. Comitini, D. Rizzo, M. Battello, and F. Siringo, Phys. Rev. D **104**, 074020 (2021), arXiv:2108.00417 [hep-th].
[41] G. Comitini and F. Siringo, Phys. Rev. D **102**, 094002 (2020), arXiv:2007.04231 [hep-ph].
[42] F. Siringo and G. Comitini, Phys. Rev. D **98**, 034023 (2018), arXiv:1806.08397 [hep-ph].
[43] D. Dudal, J. A. Gracey, S. P. Sorella, N. Vandersickel, and H. Verschelde, Phys. Rev. **D78**, 065047 (2008), arXiv:0806.4348 [hep-th].
[44] D. Dudal, O. Oliveira, and N. Vandersickel, Phys. Rev. **D81**, 074505 (2010), arXiv:1002.2374 [hep-lat].
[45] V. N. Gribov, Nucl. Phys. **B139**, 1 (1978).
[46] D. Zwanziger, Nucl. Phys. **B323**, 513 (1989).
[47] D. Dudal, C. P. Felix, L. F. Palhares, F. Rondeau, and D. Vercauteren, Eur. Phys. J. C **79**, 731 (2019), arXiv:1901.11264 [hep-th].
[48] N. Vandersickel and D. Zwanziger, Phys. Rept. **520**, 175 (2012), arXiv:1202.1491 [hep-th].
[49] N. Vandersickel, A Study of the Gribov-Zwanziger action: from propagators to glueballs, Ph.D. thesis, Gent U. (2011),

arXiv:1104.1315 [hep-th].
[50] R. F. Sobreiro and S. P. Sorella, in 13th Jorge Andre Swieca Summer School on Particle and Fields (2005) arXiv:hep-th/0504095.
[51] M. A. L. Capri, M. S. Guimaraes, I. Justo, L. F. Palhares, and S. P. Sorella, Eur. Phys. J. C **76**, 141 (2016), arXiv:1510.07886 [hep-th].
[52] M. A. L. Capri, D. Dudal, D. Fiorentini, M. S. Guimaraes, I. F. Justo, A. D. Pereira, B. W. Mintz, L. F. Palhares, R. F. Sobreiro, and S. P. Sorella, Phys. Rev. **D92**, 045039 (2015), arXiv:1506.06995 [hep-th].
[53] M. A. L. Capri, D. Fiorentini, M. S. Guimaraes, B. W. Mintz, L. F. Palhares, S. P. Sorella, D. Dudal, I. F. Justo, A. D. Pereira, and R. F. Sobreiro, Phys. Rev. **D93**, 065019 (2016), arXiv:1512.05833 [hep-th].
[54] M. A. L. Capri, D. Dudal, D. Fiorentini, M. S. Guimaraes, I. F. Justo, A. D. Pereira, B. W. Mintz, L. F. Palhares, R. F. Sobreiro, and S. P. Sorella, Phys. Rev. **D94**, 025035 (2016), arXiv:1605.02610 [hep-th].
[55] M. A. L. Capri, D. Dudal, A. D. Pereira, D. Fiorentini, M. S. Guimaraes, B. W. Mintz, L. F. Palhares, and S. P. Sorella, Phys. Rev. **D95**, 045011 (2017), arXiv:1611.10077 [hep-th].
[56] J. A. Gracey, Phys. Rev. **D86**, 105029 (2012), arXiv:1210.5962 [hep-th].
[57] M. Lavelle and D. McMullan, Phys. Rept. **279**, 1 (1997), arXiv:hep-ph/9509344 [hep-ph].
[58] M. A. L. Capri, D. Fiorentini, M. S. Guimaraes, B. W. Mintz, L. F. Palhares, and S. P. Sorella, Phys. Rev. **D94**, 065009 (2016), arXiv:1606.06601 [hep-th].
[59] D. Binosi, J. Collins, C. Kaufhold, and L. Theussl, Comput. Phys. Commun. **180**, 1709 (2009), arXiv:0811.4113 [hep-ph].
[60] A. Cucchieri, A. Maas, and T. Mendes, Phys. Rev. D **74**, 014503 (2006), arXiv:hep-lat/0605011.
[61] O. Piguet and S. P. Sorella, Lect. Notes Phys. Monogr. **28**, 1 (1995).
[62] J. C. Taylor, Nucl. Phys. B **33**, 436 (1971).
[63] M. Sjödahl, Eur. Phys. J. C **73**, 2310 (2013), arXiv:1211.2099 [hep-ph].
[64] G. Passarino and M. J. G. Veltman, Nucl. Phys. B **160**, 151 (1979).
[65] A. V. Smirnov and F. S. Chuharev, Comput. Phys. Commun. **247**, 106877 (2020), arXiv:1901.07808 [hep-ph].
[66] S. P. Martin, Phys. Rev. D **68**, 075002 (2003), arXiv:hep-ph/0307101.
[67] A. Cucchieri, D. Dudal, T. Mendes, and N. Vandersickel, Phys. Rev. **D85**, 094513 (2012), arXiv:1111.2327 [hep-lat].
[68] O. Oliveira and P. J. Silva, Phys. Rev. **D86**, 114513 (2012), arXiv:1207.3029 [hep-lat].
[69] A. Cucchieri, A. Maas, and T. Mendes, Phys. Rev. **D77**, 094510 (2008), arXiv:0803.1798 [hep-lat].
[70] A. Maas, SciPost Phys. **8**, 071 (2020), arXiv:1907.10435 [hep-lat].
[71] E. M. Ilgenfritz, M. Muller-Preussker, A. Sternbeck, A. Schiller, and I. L. Bogolubsky, Braz. J. Phys. **37**, 193 (2007), arXiv:hep-lat/0609043 [hep-lat].
[72] A. Rohatg, "Webplotdigitizer (version 4.3, july 2020)," `https://automeris.io/WebPlotDigitizer`.

Preprint — Submitted to Computing and Software for Big Science

A case study of sending graph neural networks back to the test bench for applications in high-energy particle physics

Emanuel Pfeffer · **Michael Waßmer** · **Yee-Ying Cung** · **Roger Wolf** · **Ulrich Husemann**

arXiv:2402.17386v1 [hep-ph] 27 Feb 2024

Abstract In high-energy particle collisions, the primary collision products usually decay further resulting in tree-like, hierarchical structures with a priori unknown multiplicity. At the stable-particle level all decay products of a collision form permutation invariant sets of final state objects. The analogy to mathematical graphs gives rise to the idea that graph neural networks (GNNs), which naturally resemble these properties, should be best-suited to address many tasks related to high-energy particle physics. In this paper we describe a benchmark test of a typical GNN against neural networks of the well-established deep fully-connected feed-forward architecture. We aim at performing this comparison maximally unbiased in terms of nodes, hidden layers, or trainable parameters of the neural networks under study. As physics case we use the classification of the final state X produced in association with top quark-antiquark pairs in proton-proton collisions at the Large Hadron Collider at CERN, where X stands for a bottom quark-antiquark pair produced either non-resonantly or through the decay of an intermediately produced Z or Higgs boson.

Keywords Graph Neural Networks · Deep Neural Networks · High-Energy Particle Physics · LHC

Emanuel Pfeffer[1] (corresponding author)
E-mail: emanuel.pfeffer@kit.edu

Michael Waßmer[1]
E-mail: michael.wassmer@kit.edu

Yee-Ying Cung
E-mail: yeeying.cung@web.de

Roger Wolf[1]
E-mail: roger.wolf@kit.edu

Ulrich Husemann[1]
E-mail: ulrich.husemann@kit.edu

[1] Karlsruhe Institute of Technology, Institute of Experimental Particle Physics, Karlsruhe, Germany

1 Introduction

The continuous rise and flourish of deep learning has significantly impacted also the community of high-energy particle physics, where modern algorithms of deep learning —mostly in the form of various neural network (NN) architectures— find applications as automation tools, for (multiclass) classification, parameter regression, or universal function approximation. The Large Hadron Collider (LHC) at CERN offers a unique test environment for such algorithms providing a large amount of independent identically distributed (i.i.d.) data from proton-proton (pp) collisions under well controlled laboratory conditions. These data feature a rich hierarchical structure, optimally suited for the application of all kinds of general methods of statistical data analysis. Moreover, the underlying physics laws and statistical models, which have emerged over many decades of research, are scrutinized to a level that allows the reliable estimation of particle properties with a relative accuracy ranging far below the per-mille level, in rare cases even below 10^{-10} [1]. This circumstance offers a toolbox for generating a large amount of perfectly known, complex, synthetic data, with a high relation to experimental observations, through the application of Monte Carlo (MC) methods [2,3]. These data are usually obtained as samples from an intractable though well-known likelihood function $\mathcal{L}$. This setup provides a unique opportunity to thoroughly benchmark any kinds of machine learning (ML) algorithms under complex, real-life laboratory conditions.

At the LHC, data analysts strive for the application of more and more sophisticated ML-models with more and more not further processed —and in this sense "raw"— input data. This strategy is fed by the belief that automated algorithms might find ways of extracting information of interest to the analyst, which are superior to selection strategies that are vulnerable to the bias of human prejudice. On the other hand, ML-algorithms should not be forced to learn already known and well-established physics principles, like symmetries inherent to the presented task. While such information can only be insufficiently passed through the necessarily finite training samples, it can be intrinsically incorporated either into the loss functions used for training, or in the NN architectures.

At the large multi-purpose LHC experiments, ATLAS [4] and CMS [5], pp collisions at a center-of-mass energy of, e.g., 13 TeV result in the creation of thousands of collision products to be recorded by the experiments. Primary collision products might decay further resulting in tree-like, hierarchical structures with a priori unknown multiplicity. The collision process can be described in a factorized approach:

During the hard scattering process, the fundamental constituents of the protons, i.e., the quarks and gluons which are also collectively referred to as partons, interact via the fundamental interactions under investigation. We refer to the result of these interactions as the partonic final state. It cannot be observed directly in an experiment. Rather, each parton undergoes a series of theoretically well-known processes, setting in at lower energy scales, resulting in stable particles. The inverse problem usually subject of high-energy particle physics is to infer the presence and properties of the stable particles and eventually the partonic final state from their observable energy deposits in the detectors.

At the stable particle level all decay products of a collision form permutation invariant sets of final state objects, which may emerge from the collision in the form

of collimated particle jets [6], forming well-suited proxies for strongly-interacting final state partons, or individual, spatially isolated particles, like leptons. From the preparation of the collision's initial state and energy and momentum conservation, physicists may infer the presence of non- or weakly-interacting particles, like neutrinos, in the collision's final state, through the principle of missing transverse energy (MET) [7]. A natural representation of this richly structured data is in the form of mathematical graphs $\mathcal{G}$, which are indeed also the basis of theoretical amplitude calculations of the quantum-mechanical wave function in the form of Feynman graphs [8].

Within the high-energy particle physics community, this observation has lead to an increased interest in NNs based on mathematical graphs (GNNs) [9,10, 11,12,13], where nodes are usually identified by particles and edges potentially by relations across particles. A comprehensive review can be found in Ref. [14]. However, on closer inspection, several features of GNNs that count for seemingly obvious advantages reveal subtle challenges:

- The mathematical model to match physics entities like stable particles to the graph $\mathcal{G}$ of the GNN, which is subject to mathematical operations, bears ambiguities. The representation of particles by the nodes of $\mathcal{G}$ appears as an obvious choice. This choice raises the question of the (potentially physical) meaning to the connecting edges, which could represent mother-daughter relations, or proximity in an arbitrarily defined space. This issue is emphasized once mathematical weights or even trainable parameters (TPs) used during node aggregation are assigned to the edges.
- Training and application imply the potentially complex and computationally time-consuming task of building $\mathcal{G}$, based on the physics inputs.
- In general, the more complex structure of the GNN compared to other NN architectures complicates the users' comprehension of how the GNN arrives at its prediction.

In this paper we describe a comparison of typical GNN architectures with NN models based on the deep fully-connected feed-forward architecture (DNN), which has been studied intensely in the past. We aim at comparisons that are to best effort unbiased in terms of expressiveness and information provided to the NN models to solve a given task. In Section 2 we give an introduction to the task that serves as benchmark for this comparison. In Section 3 the architectures and training setups of the NNs under study are described. In Section 4 we present the results of the comparison. We conclude the paper with a summary in Section 5.

2 Neural network task

2.1 Physics processes

As benchmark for the comparison we use the classification of the final state X produced in association with top (t) quark-antiquark pairs ($\mathrm{t\bar{t}}$) in pp collisions at the LHC, where X stands for a bottom (b) quark-antiquark pair ($\mathrm{b\bar{b}}$) produced either through non-resonant gluon exchange or through the decay of a massive Z or Higgs (H) boson as intermediate particle, as discussed, e.g., in Ref. [15]. Under realistic conditions, the collision of interest might be overlaid by several tens of

additional collisions, referred to as pileup. The complete detectable final state of a collision of interest, including pileup, is referred to as an event, whose feature vector $\mathbf{x}$ would be presented to the NN. An arbitrary number of such events may be generated synthetically by evaluating $\mathcal{L}$ of the full process via the MC method. In this study, we focus on the classification of the underlying hard process neglecting the effects of pileup.

The interest in the chosen classification task arose from studies of H production in association with $t\bar{t}$ in the subsequent $H \to b\bar{b}$ decay (ttH(bb)), for which $t\bar{t}$ associated Z boson production (ttZ) in the $Z \to b\bar{b}$ decay channel (ttZ(bb)) and non-resonant $b\bar{b}$ production in association with $t\bar{t}$ (ttbb) are important background processes. Exemplary Feynman diagrams of these processes in leading-order (LO) of perturbation theory are shown in Figure 1. The decay products, resembled by the outgoing lines in the diagrams, shown in Figure 1, represent the partonic final state of interest to this study, which is the same for all processes. Therefore, the processes can only be distinguished by the kinematic properties of the particles, in particular b quarks. This situation is complicated by the fact that the t quark also decays into b quarks radiating a quasi-real W boson with a branching fraction of nearly 100% [16]. The W boson subsequently decays either into quarks, which further on form jets in the detector, or leptons. For the presented study the semi-leptonic $t\bar{t}$ final state has been chosen, in which the W boson of one t decays into an electron or muon, further on referred to as ℓ, and a corresponding neutrino ν_ℓ. The other W boson decays into quarks. Due to the radiation of additional gluons and the splitting of gluons into quark-antiquark pairs additional colored particles and consequently jets might emerge from the process. This constellation implies a richly structured final state of an event with at least four b quark- and two predominantly light-quark-induced jets; an ℓ, which is spatially isolated from any other activity originating from the hard scattering process in the detector; and MET, due to the emitted ν_ℓ. The b quark-induced jets, referred to as b jets in the following, may be identified experimentally with a finite purity and efficiency, as, e.g., described in Ref. [17]; the methods of how to achieve this are not subject of this paper.

2.2 Sample preparation

Samples for all processes in question have been generated synthetically from a corresponding likelihood $\mathcal{L}$ using the MC technique. The tools used for event generation are the matrix-element generator MadGraph5_aMC@NLO [18,19] in ver-

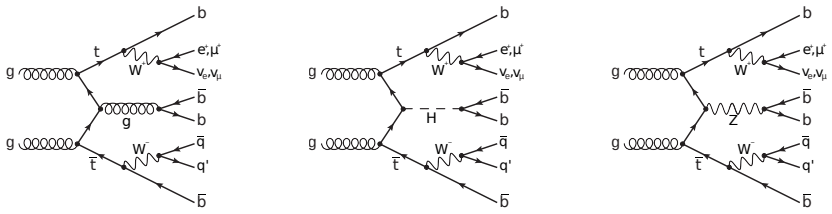


Fig. 1 Exemplary Feynman diagrams for the processes of interest to this study: (left) ttbb, (middle) ttH(bb), and (right) ttZ(bb).

Table 1 Selection requirements on the reconstructed final-state objects. The quantity $I_{\text{rel}}^{\Delta R}$ corresponds to the scalar sum of energy deposits detected within the radius ΔR around ℓ in η-ϕ divided by the p_{T} of ℓ, as defined in the text. A lower value of $I_{\text{rel}}^{\Delta R}$ implies less activity in the spacial vicinity of ℓ, indicating that ℓ originates from $\text{W} \to \ell\nu_\ell$. The variable β refers to the working points of a specific b jet identification algorithm, as described in the text.

Object	p_{T} (GeV)	$\lvert\eta\rvert$	$I_{\text{rel}}^{\Delta R}$	β
Electron	≥ 25	< 2.5	< 0.12 ($\Delta R = 0.3$)	–
Muon	≥ 25	< 2.4	< 0.25 ($\Delta R = 0.4$)	–
Jet (anti-k_{T}, R_0=0.4 [22])	≥ 20	< 2.4	–	–
b jet	≥ 20	< 2.4	–	$\beta \geq 2$

sion 2.9.9 interfaced with the Pythia event generator [20] in version 8.306 to map the partonic final state to the stable-particle level. All processes in consideration have been generated at LO in perturbative quantum chromodynamics (QCD), in the four-flavor scheme. The same setup has been used for the generation of all samples to avoid spurious differences due to the use of different generation tools.

All generated events have been passed through a simplified simulation of the CMS detector as configured during the LHC Run-2 data-taking period in the years 2016–2018, using the DELPHES simulation package [21]. For this study only reconstructed leptons, jets, and MET are considered. All detected and reconstructed final-state objects have been selected to fulfill a set of selection criteria typically used for the analysis of data collected by the CMS experiment, as summarized in Table 1. These selection criteria comprise the following observables:

- The transverse momentum p_{T} and pseudorapidity η of the reconstructed ℓ and jets.
- A variable $I_{\text{rel}}^{\Delta R}$ that corresponds to the scalar sum of energy deposits E_i detected within the radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ around ℓ, divided by the p_{T} of ℓ, where $\Delta\eta$ refers to the difference between E_i and ℓ in η and $\Delta\phi$ to the corresponding difference in azimuthal angle ϕ, based on the coordinate system deployed by CMS [5]. For ℓ originating from $\text{W} \to \ell\nu_\ell$ low values of $I_{\text{rel}}^{\Delta R}$ are expected.
- The output of a specific b jet identification algorithm represented by the discrete observable $\beta \in \{0, 1, 2, 3\}$ indicating whether an object has been identified as a b jet under a specific working point α ($\beta \geq \alpha$). The value of α represents ($\alpha = 1$) loose, ($\alpha = 2$) medium, and ($\alpha = 3$) tight selection criteria, corresponding to a rate of non-b jets wrongly identified as a b jet (false-positive rate), of approximately 10%, 1%, and 0.1%, respectively.

All events have been selected to exhibit at least six and not more than eight jets, at least four of which are assumed to be correctly identified as b jets according to $\beta \geq 2$, and exactly one ℓ, matching all selection criteria. The selection of one ℓ and six jets, of which at least four are identified as b jets, is motivated by the partonic final state under study, as depicted in Figure 1. The selection of up to two additional jets, increases the chance that the complete partonic final state can be matched to the selected jets.

For each reconstructed jet the attempt is made to assign the initiating particle of the partonic final state, based on the distance ΔR between the jet and the corresponding parton. Only partonic final state objects with transverse momentum of $p_{\text{T}} > 20\,\text{GeV}$ and $\lvert\eta\rvert < 2.4$ are considered for this assignment. From the

Table 2 Requirements for the jet-class definition, according to the matching to the partonic final state, where b_{add} corresponds to a b quark not originating from a t decay, $b_{t_{had}}$ ($b_{t_{lep}}$) to a b quark originating from $t \to bW(qq')$ ($t \to bW(\ell\nu_\ell)$), and $q_{W_{had}}$ to a quark originating from $W \to qq'$. We note that the classes ADDB and HTQ comprise two jets. Jets not assigned to any other class are assigned to the NA class and sorted by decreasing p_T. If no jet is found to fulfill the corresponding ΔR criterion, the leading jet from the NA class is re-assigned to the classes HTB, HTL, HTQ, in that order.

Class label	Assignment	Description
ADDB	$\Delta R(\text{jet}, b_{add}) < 0.4$	b jets not from t decays
HTB	$\Delta R(\text{jet}, b_{t_{had}}) < 0.4^\dagger$	b jet from t_{had}
LTB	$\Delta R(\text{jet}, b_{t_{lep}}) < 0.4^\dagger$	b jet from t_{lep}
HTQ	$\Delta R(\text{jet}, q_{W_{had}}) < 0.4^\dagger$	q jet from W
NA	No match	Additional jet

† If no assignment by ΔR is found, the leading jet from NA is assigned.

assignment five mutually exclusive jet classes are build, referring to the (ADDB) b quarks not originating from any t decay; the b quark originating from the (HTB) hadronic and (LTB) leptonic t decay, (HTQ) quarks originating from the hadronic W decay; and (NA) jets not assigned to the partonic final state.

The assignment of the partonic final state to the reconstructed jets may be incomplete, for a given event. Events for which no or only one jet is assigned to the ADDB class are discarded from the training and test samples. In all other cases, if the assignment by ΔR did not result in one jet of class LTB, one jet of class HTB, and two jets of class HTQ, the remaining not associated jets of class NA are ordered by decreasing p_T and the leading jets in p_T are re-assigned to these classes in the order of LTB, HTB, HTQ. A summary of all jet classes is given in Table 2

2.3 Task definition

The NN models are supposed to perform a classification task, in which a given event should be assigned either to ttbb, ttH(bb), or ttZ(bb). The physics process from which the event was generated constitutes the ground truth information for this task.

To simplify the task, the reconstructed jets are assumed to perfectly match the initiating partons, whenever possible, through the parton association algorithm, as described in Section 2.2. The ADDB class contains jets which stem directly from the intermediate particle to distinguish the processes under study. The features of these jets are assumed to provide the most decisive contribution to the classification.

2.4 Training setup

All NN models are subject to a supervised training. The generated samples used for this are split into a training, validation, and test sample, containing 60, 20, and 20% of the generated events, respectively. Event numbers, split by process and sample, are given in Table 3.

Each training is performed for an ensemble of ten statistically independent repetitions to obtain a rough estimate of the statistical spread of the trained models, due to random choices in the training setup. Performance measures of each model are reported as sample means μ, of which the corresponding uncertainty $\Delta\mu$ is estimated from the square root of the sample variance. All repetitions are based on the same training, validation, and test samples. Due to the large number of events in these samples, randomization through data shuffling is assumed not to change the conclusions of the studies significantly.

Each training is performed on CPUs through a distributed computing infrastructure, where each training is performed on a dedicated CPU. We ensure that each step affected by random choices is based on a different random seed. The library used to build the GNNs is PyTorch Geometric v2.0.3 [23] based on the PyTorch library [24]. The same library is used, to construct the corresponding DNNs under study. Further parameter choices of the training setups are given in the upper part of Table 4. They are the same for both NN architectures.

3 Neural networks under study

3.1 Architectures

For the GNN models, a graph representation of the final state is obtained from the reconstructed jets, ℓ, and MET in an event. Each of these objects is represented by a node i in $\mathcal{G}$. This set of nodes may be complemented by nodes for up to two more jets in the selection. Accordingly, $\mathcal{G}$ has 8–10 nodes. For each i, a vector $\mathbf{a}_i$ of attributes with $\dim(\mathbf{a}_i) = n$, $\forall i$ is defined, forming the embedding space. At initialization time, the $\mathbf{a}_i$ are initialized by the feature vectors $\mathbf{x}_i$ with $\dim(\mathbf{x}_i) = n_{\text{feat}}$, $\forall i$. The following studies also comprise configurations with $n > n_{\text{feat}}$. In these cases, attributes in the $\mathbf{a}_i$ without correspondence in $\mathbf{x}_i$ are set to zero (zero padding), at initialization time.

Table 3 Numbers of events for each process, in the training, validation, and test samples.

Process	Training	Validation	Test	Sum
ttbb	41650	13883	13803	69336
ttH(bb)	99695	33329	33229	166253
ttZ(bb)	88107	29272	29453	146832

Table 4 Common parameter choices for the NN architectures under study and their training setup.

Parameter	Setting(s)
Loss function	Binary cross-entropy
Optimizer	Adam [25] ($\gamma = 0.01$)
Mini-batch size	200
Maximum number of epochs	200
Early-stopping	Δepochs = 15, Δloss = 0.001
Use of weights and biases	Yes
Number of outputs	1 (binary)
Activation function (for hidden layers)	ReLu
Activation function (for output layer)	Sigmoid

All nodes are connected with edges, resulting in a fully-connected non-directed graph without self-loops. Relational information between nodes i and j may be assigned, in form of edge weights ω_{ij}. For this purpose three physics motivated choices are made: (i) the invariant mass m; (ii) the distance ΔR; and (iii) the reciprocal distance ΔR^{-1} of the connected final-state objects. In addition, the cases of a (one) constant, (rnd) random, and (zero) no edge connection at all are studied, resulting in a total of six variants of edge connections. An illustration of a resulting graph for an event with eight nodes and no additional jets is given in Figure 2.

The GNN algorithm to process the graph data is based on the layered GraphConv operation, as introduced and described in Ref. [26], using the sum over all i as aggregation function. After initialization, k GraphConv operations are applied, after which the resulting $\mathbf{a}_i$ are transformed into a single vector of length n, averaging over all i. A linear combination of the components of this vector, which is scaled to values between 1 (indicating ttbb as signal) and 0, for binary classification, eventually forms the output $\hat{y}$ of the GNN. A graphical illustration of this model is given in the lower part of Figure 2.

In a first study, the GNN are compared with corresponding DNN models with k hidden layers, containing n hidden nodes, each. The values of k, n, and the choice of weights, steering the exploitation of relational information by the GNN models are varied, resulting in 36 variants of parameter choices, as summarized in Table 5. We note that here, as in the following, n represents a tuple of length k. For a GNN this tuple indicates the dimension of the embedding space per GraphConv operation; for a DNN it represents the number of nodes per hidden layer. Other parameter choices related to the NN architectures under study or the setup of the NN training are made common and summarized in Table 4. Special care is taken to compare the GNN with the corresponding DNN models on the same footing, especially in terms of information about the feature space presented to them, as discussed in the following section.

3.2 Presentation of the feature space

Primary features passed to the NNs are the invariant mass (m), energy (E), η, and ϕ of each reconstructed final state object. The reconstructed final state objects comprise ℓ, MET, at least four b jets, two additional jets, all of which are

Table 5 Parameters varied for the comparison of GNN with corresponding DNN models, where n corresponds to the dimension of the embedding space during a GraphConv operation (number of nodes in a hidden layer) and k to the number of GraphConv operations (hidden layers) in the GNN (DNN) case. The choices of n are motivated by the size of the input vector n_{feat} to the GNN, as described in Section 3.2. The choices of one, rnd, and zero for the use of edge information in GNN models are compared to DNN models without relational information between individual objects. These DNN models are indicated by the label none, in corresponding figures.

Parameter	Setting(s)
k	1, 2
n	13, 26, 39
Edge weights (ω_{ij})	m, ΔR, ΔR^{-1}, one, rnd, zero

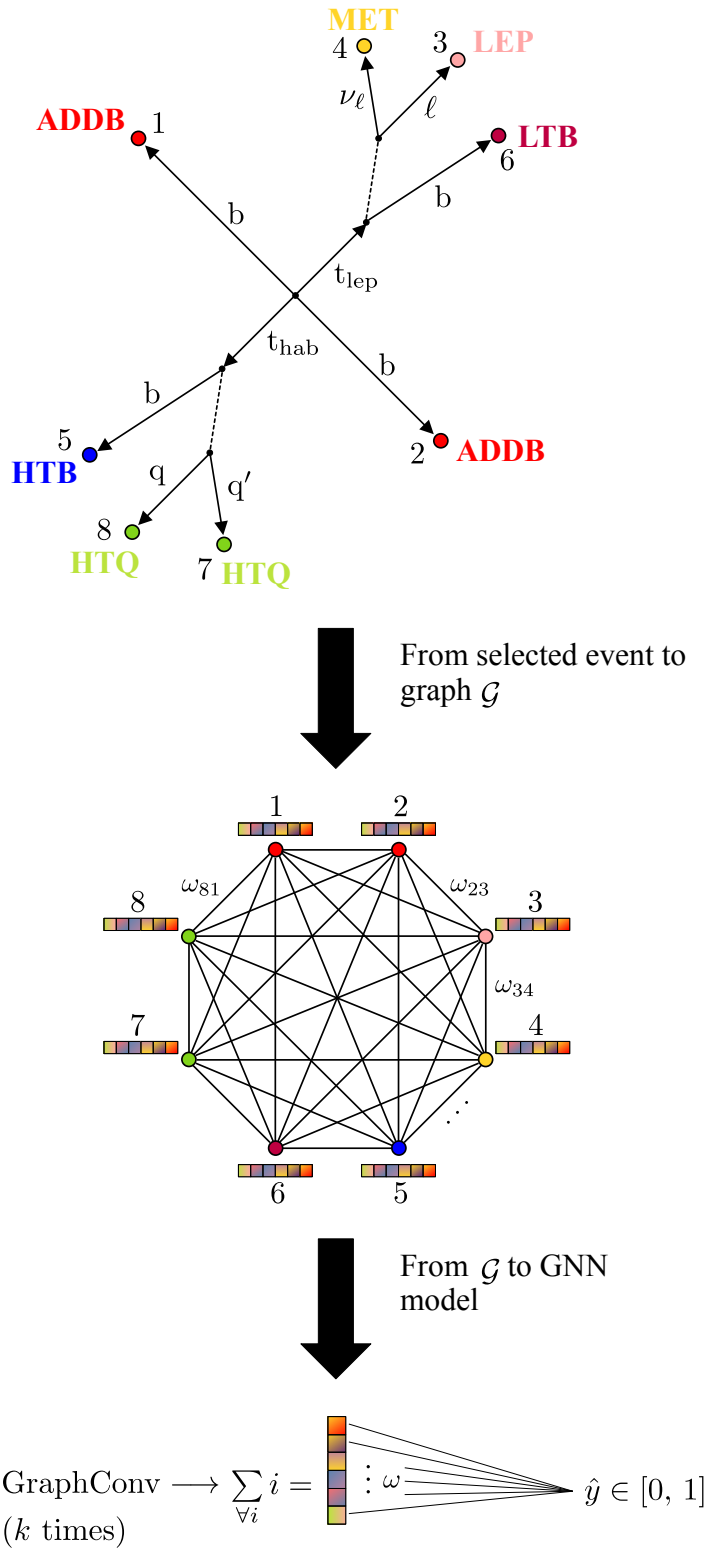


Fig. 2 Translation of an (upper part) selected event (without jets in the NA jet-class in this case) into a (middle part) graph $\mathcal{G}$ and finally into the (lower part) GNN model. For the indication of the partonic final state we do not distinguish particles from anti-particles. The individual object-classes are indicated by different colors. The nodes of $\mathcal{G}$ are labelled by $i = 1 \ldots 8$ and colored the same way as the object-classes. The boxes next to the nodes indicate the embedding space of the GNN model. The GNN output is indicated by $\hat{y}$.

associated with the partonic final state, and potentially two more non-associated jets from the event selection. For MET the features E and m are set to zero. The primary features are complemented by β. The selection requires values of $\beta \geq 2$ for identified b jets. For ℓ and MET β is set to zero.

We note that both NN architectures may profit from additional information, which is not passed explicitly through $\mathbf{x}_i$, but implicitly through the way the features are presented to the NNs. An obvious difference between the NN architectures arises from the fact that the GNN naturally supports processing of events with arbitrary object multiplicities. The $\mathbf{x}_i$ are transferred to the GNN through the $\mathbf{a}_i$, during initialization. During the GNN processing the information per i is aggregated over all nodes. In the DNN case such an aggregation step is absent. Instead, the $\mathbf{x}_i$ are concatenated into an enlarged feature vector $\mathbf{x}^{\mathrm{DNN}}$ of length $10 \times n_{\mathrm{feat}}$, comprising the $\mathbf{x}_i$ of the eight reconstructed objects, of which all jets have been associated to the partonic final state, plus potentially two additional selected jets. The order in which these objects are concatenated has been chosen to follow the association to the partonic final state. For events that contain fewer than two jets in addition to those that have been matched to the partonic final state, the corresponding entries in $\mathbf{x}^{\mathrm{DNN}}$ are filled with zeros. We point out two subtleties, which are related to these choices:

One subtlety, in favor of the GNN, arises from incorporating relational information through ω_{ij}. This advantage is compensated for by appending equivalent information to $\mathbf{x}^{\mathrm{DNN}}$. For up to ten selected objects in an event this results in up to 45 additional features. The choices of one, rnd, and zero for the use of edge information in GNN models are compared to DNN models without relational information between individual objects. These DNN models are indicated by the label none, in corresponding figures.

Another subtlety, related to the same fundamental difference, but this time in favor of the DNN, arises from the fact that through the concatenation of the $\mathbf{x}_i$ into $\mathbf{x}^{\mathrm{DNN}}$, according to the association to the partonic final state, the DNN receives extra information through the positions of the $\mathbf{x}_i$ in $\mathbf{x}^{\mathrm{DNN}}$ which is not accessible to the GNN. This advantage is compensated for by adding information about the association of the i to the partonic final state via one-hot encoding. For the five jet-classes defined in Table 2 plus one label (LEP) for ℓ and one label (MET) for MET this extension increases n_{feat} by seven, leading to the dimension of $\mathbf{x}_i$, $\forall i$ of $n_{\mathrm{feat}} = 13$.

For $\mathbf{x}^{\mathrm{DNN}}$ the $\mathbf{x}_i$ are concatenated for all i assuming two more jets in the NA class. For events with fewer than two jets in the NA class the foreseen features are initialized with zero. Together with the relational information between all potential objects in an event, this results in a dimension of $\mathbf{x}^{\mathrm{DNN}}$ of 175 features, of which up to 47 features might potentially be filled with zeros.

In this configuration the information about the association to the partonic final state is presented in the form of one-hot encoding to both NN architectures. To confirm to what extent the DNN may infer this information already from the position of the $\mathbf{x}_i$ within $\mathbf{x}^{\mathrm{DNN}}$ we also investigate configurations of the DNNs without this information in form of on-hot encoding, resulting in a reduced input vector $\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$ with dimension 105. All input features in use are listed in Table 6. All non-integer features are standardized to a mean of zero and a standard deviation of one.

Table 6 Input features used for the studies described in the text. Columns 2–4 indicate whether a given feature is continuous, discrete, or presented via one-hot encoding. The given features form the feature vectors $\mathbf{x}_i$ per object i. For the DNNs the $\mathbf{x}_i$ for potentially ten selected objects are concatenated into an extended feature vector $\mathbf{x}^{\mathrm{DNN}}$, according to their association to the partonic final state; for events with fewer than ten objects the $\mathbf{x}_i$ of the missing objects are filled with zeros. In addition, relational information between all potential objects is added to $\mathbf{x}^{\mathrm{DNN}}$. In a reduced configuration, the one-hot encoded information about the association of the objects to the partonic final state is omitted to form a reduced input vector $\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$ to the DNNs.

Input feature	Continuous	Discrete	One-hot	Comment
m	✓	–	–	Primary features
E	✓	–	–	
p_{T}	✓	–	–	
ϕ	✓	–	–	
η	✓	–	–	
β	–	✓	–	
LEP	–	–	✓	Assoc. to ℓ
MET	–	–	✓	Assoc. to MET
ADDB	–	–	✓	Assoc. to additional b quarks
HTB	–	–	✓	Assoc. to $\mathrm{t}_{\mathrm{b}_{\mathrm{had}}}$
LTB	–	–	✓	Assoc. to $\mathrm{t}_{\mathrm{b}_{\mathrm{lep}}}$
HTQ	–	–	✓	Assoc. to $\mathrm{q}_{\mathrm{W}_{\mathrm{had}}}$
NA	–	–	✓	Not associated

4 Results

To be able to draw fair conclusions from a comparison of different NN architectures (of potentially different complexity) special care has to be taken for this comparison to be based on the same ground. For this study we have focused on a common choice of non-tunable (hyper-)parameters, which are not subject to the NN training, as well as on an equal level of information primarily passed to the NNs, through training conditions and input features.

An inevitable difference remains in the organization and layering of hidden nodes, which when kept similar, may well lead to a different number of TPs and therefore a priori different expressiveness of the NN models. Vice versa, keeping the number of TPs similar, implies differences in the layering and organization of hidden nodes. Since differences of one or the other kind may not be overcome, both configurations, (i) similar layering of hidden nodes; and (ii) similar number of TPs, are studied. In any case, the enlarged size of $\mathbf{x}^{\mathrm{DNN}}$ ($\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$) with respect to $\mathbf{x}$ will give higher emphasis to the first DNN layer compared to the corresponding GNN architecture. In addition, the DNN architecture features the less complex pre-processing of the inputs, since it does not imply the creation of graphs. On the other hand, the potentially more constrained GNN may have advantages over the DNN architecture in terms of convergence properties of the training.

4.1 Comparison of neural networks with similar layering of hidden nodes

4.1.1 Neural networks with one layer of hidden nodes

A first comparison of GNN with corresponding DNN models, based on a similar layering of hidden nodes, is shown in Figure 3. The metric by which to judge the

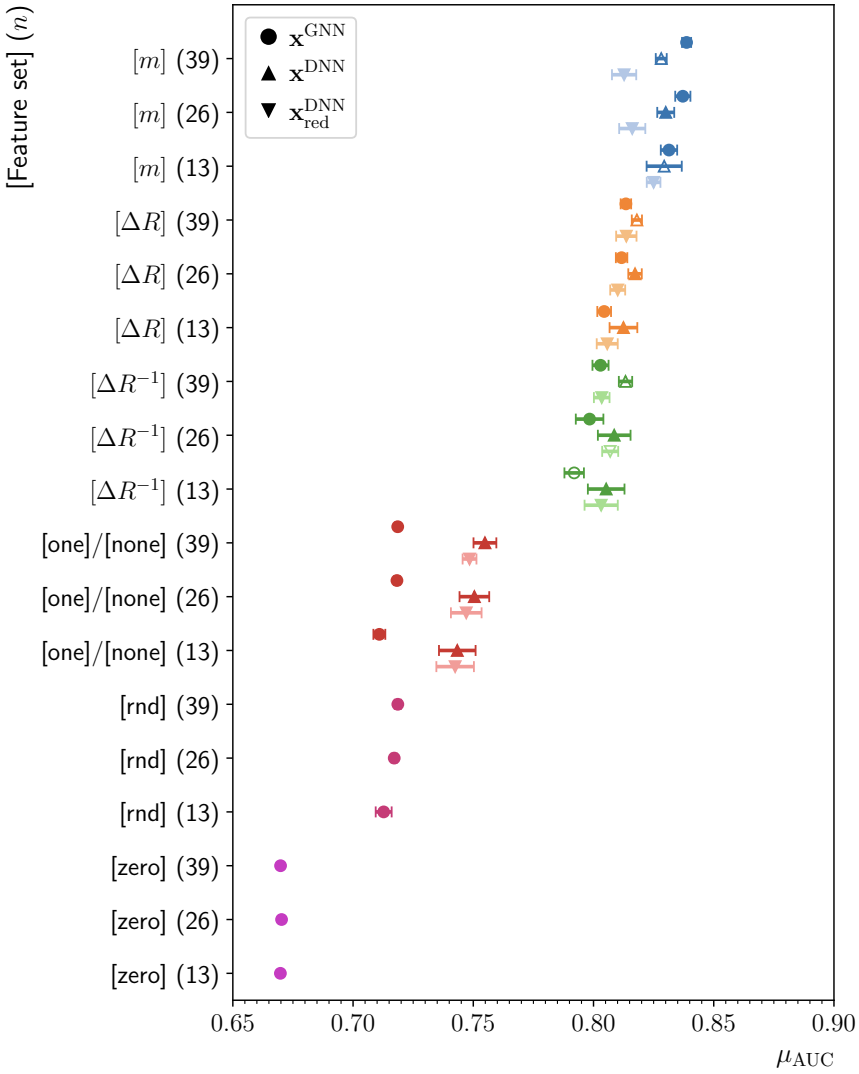


Fig. 3 Mean ROC-AUC μ_{AUC} as obtained for 18 different configurations of GNN and 24 corresponding configurations of DNN models with $k = 1$. The labels in brackets on the vertical axis indicate the use of relational information, as discussed in Section 3.2, the numbers in parentheses correspond to the choices of n. The circles refer to GNN and the upward (downward) pointing triangles to DNN models with a default (reduced) set of input features $\mathbf{x}^{\mathrm{DNN}}$ ($\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$), as discussed in Section 3.2. For better readability, markers of the same configuration are shifted vertically along the y-axis. The bars are obtained from the sample variance of an ensemble, as described in Section 2.4. Those NN architectures which belong to the same choices of varying parameters are spatially grouped and shown in the same color. Open markers indicate that significant outliers of the corresponding distribution of ROC-AUC values have been removed from the calculation of μ_{AUC} and its variance, as described in the text.

success of an NN to fulfill the task is chosen to be the mean of the ROC-AUC μ_{AUC} based on the training setup, as described in Section 2.4. The results are presented for the variation of parameters as summarized in Table 5. For the presentation in Figure 3 a simple architecture with one hidden layer for the DNN models and one GraphConv operation of the GNN models ($k = 1$) has been chosen. For the GNN architecture this implies that there is only one exchange of information across adjacent nodes of $\mathcal{G}$, i.e., each node receives information only of its nearest and not the next to nearest neighbor in $\mathcal{G}$. Since $\mathcal{G}$ has been chosen to be fully connected, there is no strict suppression of information that way, in the sense that each node i receives information from any other node $j \neq i$ in $\mathcal{G}$. The input features presented to each corresponding NN architecture are chosen, as described in Section 3.2 and summarized in Table 6.

The results for the GNN models are represented by circles, the results for the DNN models with $\mathbf{x}^{\mathrm{DNN}}$ ($\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$) as input vector by upward (downward) pointing triangles. The bars associated with the points indicate the uncertainty $\Delta\mu_{\mathrm{AUC}}$ in μ_{AUC} due to random choices in the training, as described in Section 2.4. Open markers indicate that at least one training repetition in an ensemble has been removed as an outlier from the calculations of μ_{AUC} and $\Delta\mu_{\mathrm{AUC}}$. Candidates for outliers have been identified by their μ_{AUC} values exceeding $1.5\,\Delta\mu_{\mathrm{AUC}}$, as obtained from the full ensemble. An outlier candidate has been definitely removed and the ensemble size reduced accordingly, if doing so changed $\Delta\mu_{\mathrm{AUC}}$ by a value of at least 0.0025. Following this procedure, 54 outliers have been removed from a total of 1840, which corresponds to a rate of 2.9%. Split by NN architectures, it corresponds to 17 (37) removed outliers for GNN (DNN) models from a total of 800 (1040). In no case more than two outliers have been removed from the original ten training repetitions of any individually model.

On the x-axis of the figure the corresponding values of μ_{AUC}, ranging from 0.67 to 0.83, are shown. On the y-axis the individual NN models are labelled, such that brackets indicate, what relational information between final state objects has been used, and the values in parentheses indicate the choices of n.

We conclude that all training setups have succeeded in the sense that all NN models result in values of $\mu_{\mathrm{AUC}} > 0.5$. The worst separation of signal from background we observe for the GNN models for which no relational information is exploited, indicated by three groups of NN architectures shown in pink, purple and red colors in the lower part of the figure. It is noted that the feature set zero refers to the case where the node convolution in the GraphConv operation is forcefully suppressed, and deliberately no information across nodes is exchanged at all. We anticipate that this approach counteracts the whole GNN idea. We still keep this configuration as part of the study, to gauge the effect and importance of the GraphConv operation itself. Compared to the feature set zero, the feature sets one and rnd single out cases in which information exchange across nodes takes place, but no real relational information is associated with it. Instead, the embedding spaces of the individual nodes are just mixed without particular prevalence. We note that even when allowing node convolution the GNN architecture falls significantly behind the comparable DNN architecture, even with a reduced input vector $\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$, as long as no mindful relational information across adjacent nodes i and j is provided according to ω_{ij}. This is true for assigning (one) the same or (rnd) random weights to each edge, irrespective of the expressiveness of the NNs, indicated by n. The superiority of the DNN architecture in this case cannot be

attributed to the additional information about the jet parton association, since this information is available to all NN architectures under study. In particular it is passed on to the GNNs, in the explicit form of one-hot encoding, which is not even the case for the DNNs with $\mathbf{x}_{\text{red}}^{\text{DNN}}$ as input vector. At this occasion, we note that the additional (and in fact in this case redundant) information of the parton association in form of one-hot encoding to the DNN does not lead to a significant increase of μ_{AUC}, compared to the implicit knowledge already provided by the positional information in $\mathbf{x}_{\text{red}}^{\text{DNN}}$. Hence, if the way this information is presented to the DNN were of influence, this influence is not significant in the scope of our study. We also observe that $\Delta\mu_{\text{AUC}}$ is considerably larger for the DNNs.

In conclusion, if the advantage of the GNN over the DNN architecture were that potentially excessive degrees of freedom in the DNN architecture are replaced by built-in constraints, the GNN architecture appears too confined, until these constraints are introduced mindfully. In turn the additional degrees of freedom of the DNN architecture result in a larger spread $\Delta\mu_{\text{AUC}}$ due to random choices in the training setup.

The upper part of Figure 3 reveals that, as soon as a domain-knowledge motivated ranking of information exchange across neighboring nodes is introduced, the GNN architecture significantly gains in separation power. Also here, this gain comes with an increase in $\Delta\mu_{\text{AUC}}$. The choices of (green) ΔR^{-1}, (orange) ΔR, and (blue) m as weights leads to an increase in separation power in the given order, where for each choice the values of μ_{AUC} can be grouped with a corresponding internal spread. On the other hand, a significant gain in μ_{AUC}, when increasing the expressiveness of an NN within a certain configuration group in terms of n, for the utilized NN models, is not observed.

We note that, as in the case of the positional encoding of parton association in $\mathbf{x}_{\text{red}}^{\text{DNN}}$, all choices of relational information are intrinsic to the training sample and implicitly accessible to all NNs through their feature vectors. For ΔR this is, e.g., the case through ϕ and η in the primary features of each reconstructed object. However, the information of $\Delta\phi$ and $\Delta\eta$ between pairs of reconstructed objects appears too subtle in the high-dimensional feature space, so that none of the chosen architectures could grasp it without the assistance of an accordingly conditioned representation of $\mathbf{x}$ and $\mathbf{x}^{\text{DNN}}$, even from a training sample with more than 200,000 events.

We further note that when turning the edge-weights of the GNN structure into TPs we did not obtain a separation of signal from background better than the domain-knowledge supported use of m. At the same time we observed a significantly increased spread in the achieved separation power based on random choices of the training setup.

Our physics prior assigns more physical meaning to the choice of ΔR over ΔR^{-1}, since due to causality we expect a closer relation between objects with smaller than larger spatial distance in ΔR. The observation that both choices of relational information lead to nearly similar results in μ_{AUC} we explain by a special characteristic of NN-based classification tasks in the given setup. For the NN decision, downgrading information from further-away objects is equivalent to upgrading close-by objects. The fact that corresponding DNN architectures follow the trends of the GNN architectures, as long as equipped with the same information, supports the assumption that it is this additional relational information

between objects, rather than the GNN-specific operation of mixing features across nodes that leads to the increase of μ_{AUC}.

We note that, consistently for all architectures, the highest values of μ_{AUC} are obtained with an energy-weighted distance measure like m, which again follows our prior physics intuition. It is noteworthy that in this case the DNN architecture with $\mathbf{x}_{\mathrm{red}}^{\mathrm{DNN}}$ seems to significantly lose in separation power, compared to the other architectures. In fact μ_{AUC} even decreases for increasing values of n. Also $\Delta\mu_{\mathrm{AUC}}$ appears consistently higher for all DNN compared to the GNN architectures. These observations may be interpreted as indications of the advantage of careful guidance of the NN training- and model-setup over just confronting a highly expressive NN architecture, represented by a large number of TPs, with an even excessively large training sample. This guidance may be provided through the choice and representation of input features, as well as through the choice of a more constrained NN architecture. We note that the highest value of μ_{AUC} with the smallest spread $\Delta\mu_{\mathrm{AUC}}$ is indeed obtained from the GNN architecture with highest expressiveness, given for $n = 39$.

4.1.2 Neural networks with two layers of hidden nodes

Moving on to an NN architecture with $k = 2$ introduces the ambiguity of how to choose n for each individual layer. To prevent any kinds of potential selection biases, all ways of allocating the tested values of $n = 13, 26, 39$ to the individual layers/embedding spaces are shown in Figure 4.

For the case of completely suppressed relational information for the GNN architectures (zero) μ_{AUC} remains lowest and unaffected by the choice of k, as expected for a setup in which any information transfer through a GaphConv operation is deliberately suppressed. At the same time we generally observe that no superior choice of allocating n across layers can be pointed to, throughout all tested architecture configurations. Especially the gain of using a configuration with $n = (39, 39)$ for both hidden layers/GraphConv operations over a configuration with $n = (13, 13)$ within a given architecture appears marginal.

However, the increase in k consistently mitigates the previously observed, clearly inferior separation of signal from background, of the GNN compared to the corresponding DNN architectures, for randomly (rnd) and unweighted (one) relational information. The μ_{AUC} values of these groups of GNN models start to clearly supersede the μ_{AUC} values of the one-layered DNN models with unweighted relational information, labeled by none in Figure 3, even slightly taking the lead over the two-layered DNNs of the same kind, in terms of μ_{AUC}.

We note that all two-layered NN architectures without use of relational information still result in lower values of μ_{AUC} than all tested one-layered NN models that profit from the use of relational information, as presented in Figure 3. At the same time, they are subject to an increased spread $\Delta\mu_{\mathrm{AUC}}$ compared to their one-layered counterparts, in most cases. In this sense, the wise choice of relational information outweighs the presumable advantage in expressiveness provided by $k = 2$, irrespective of the allocated values of n.

For the NN architectures including relational information, we observe no further, dramatic gains, with respect to their one-layered counterparts in μ_{AUC}, apart from a slight advantage of the GNN over the corresponding DNN architectures that seems to become more manifest. While this advantage is below the 1%-level it is

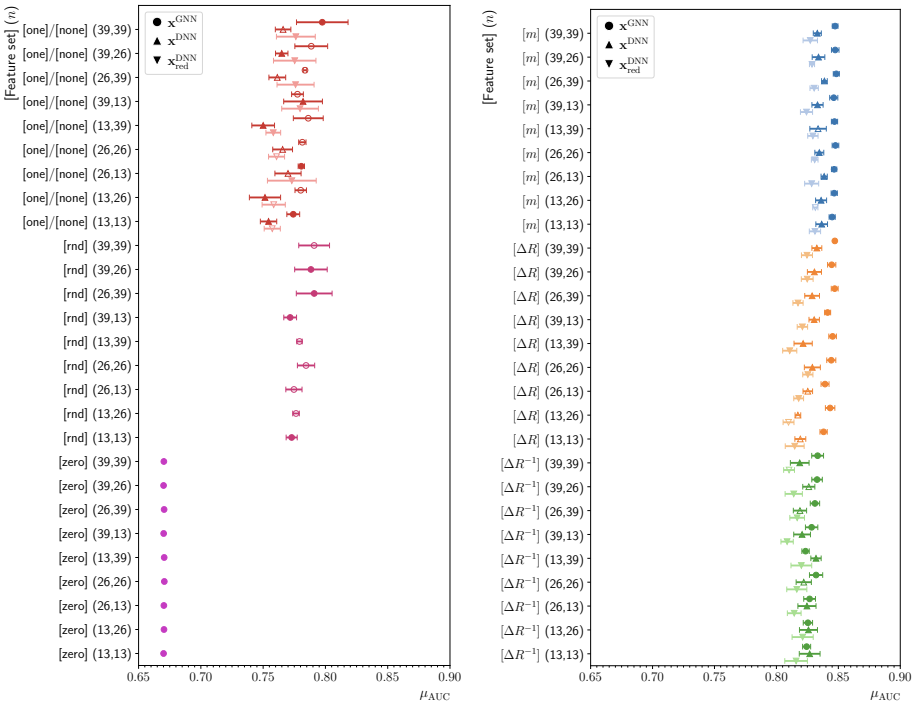


Fig. 4 Mean ROC-AUC μ_{AUC} as obtained for 54 different GNN and 72 corresponding DNN models with $k = 2$. The labels in brackets on the vertical axis indicate the use of relational information, as discussed in Section 3.2, the numbers in parentheses correspond to n. The circles refer to GNN and the upward (downward) pointing triangles to DNN models with a default (reduced) set of input features $\mathbf{x}^{\mathrm{DNN}}$ ($\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$), as discussed in Section 3.2. For better readability, markers of the same configuration are shifted vertically along the y-axis. The bars are obtained from the sample variance of an ensemble, as described in Section 2.4. NN architectures which belong to the same choices of varying parameters are spatially grouped and shown in the same color. Open markers indicate that significant outliers of the corresponding distribution of ROC-AUC values have been removed from the calculation of μ_{AUC} and its variance, as described in the text.

still significant compared to $\Delta\mu_{\mathrm{AUC}}$. A summary of the achieved values of μ_{AUC} for the one- and two-layered GNN models is shown in Figure 5. An equivalent summary for the corresponding DNN models is shown in Figure 6.

From the study we conclude that the external information of the ω_{ij} seems to give slight advantages to the GNNs with $k = 2$. We note that two subsequent GraphConv operations indeed convey more information than a DNN model with two hidden layers. Viewing ΔR^{-1}, ΔR, and m as distance measures, the first GraphConv operation conveys information about the nearest neighborhood of each i. The second GraphConv operation conveys information about the nearest neighborhood of the nearest neighbors, which is not the same as in the case of the first operation. This information is indeed not primarily accessible to the DNN architectures, but it emerges from the definition of the GraphConv operation. Along this line, once again, we conclude that not the mixing of features across neighboring nodes i during the GraphConv operation, but the additional

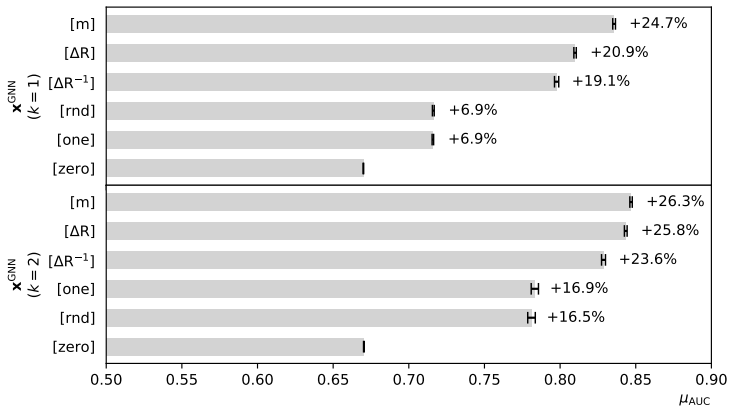


Fig. 5 Summary of the achieved values of μ_{AUC} for the GNN models with (upper half) one and (lower half) two GraphConv operations, with different use of relational information. For this summary, the associations of n with the highest values of μ_{AUC} in each group of GNN models have been used. The value of μ_{AUC} is displayed on the x-axis. Improvements relative to the least separating GNN with no relational information at all (zero) is given in numbers to the right of the bars. The use of relational information, as defined in Table 5, is indicated in brackets, on the y-axis.

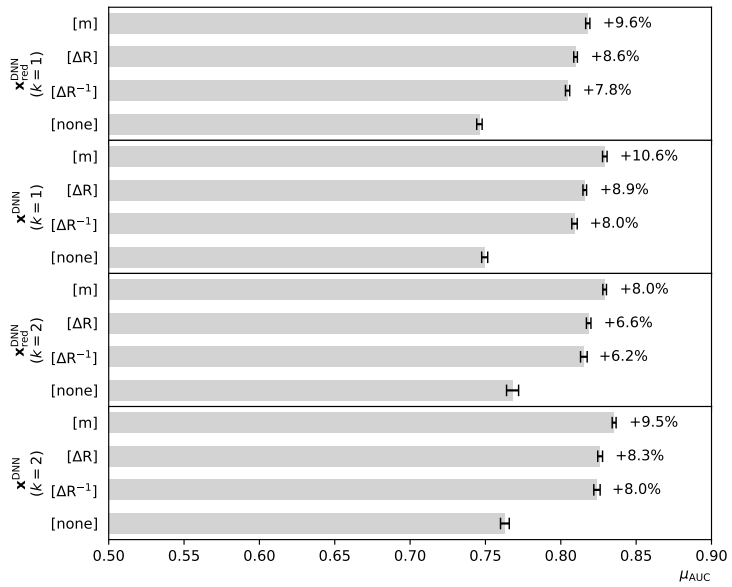


Fig. 6 Summary of the achieved values of μ_{AUC} for the DNN models with (upper half) one and (lower half) two hidden layers, with different use of relational information. For this summary, the associations of n with the highest values of μ_{AUC} in each group of DNN models have been used. The DNN configurations with $\mathbf{x}^{\mathrm{DNN}}$ and $\mathbf{x}^{\mathrm{DNN}}_{\mathrm{red}}$ are shown separately. The values of μ_{AUC} are displayed on the x-axis. Improvements relative to the least separating DNN with no relational information (none) is given in numbers to the right of the bars. The use of relational information, as defined in Table 5, is indicated in brackets, on the y-axis.

(implicit) information accessible to the GNN through this operation is the source for the slight gain in μ_{AUC}. This interpretation is supported by the observation that an increase in n does not lead to any significant improvements in μ_{AUC} despite the increase in expressiveness of the models. This also explains why the GNN models with constant (one) or randomly associated (rnd) weights suffer in their performance: Through both ways of assigning weights across the nodes i the information about any kind of distance measure across nodes, in what ever space, is omitted.

A summary of the GNN and DNN configurations with the highest values of μ_{AUC} is given in Table 7.

4.2 Neural networks with comparable numbers of trainable parameters

As stated before, two principally different NN architectures lack comparability in the sense that it might be more natural to pick up certain information from the training sample through one or the other architecture. As long as it addresses an intrinsic property of one or the other architecture, such a difference is part of the benchmark comparison. If, on the other hand, such an inequality results from withholding primary information from one or the other architecture, or potentially inappropriate advantages in terms of expressiveness, an effort should be made to study and estimate the effect of it.

We have identified and noted such an inequality, in the beginning of Section 4, in terms of the number of TPs (N_{TP}), which turns out to be naturally higher for the DNN compared to the GNN architecture, due to the usually much larger input layer. We therefore complete our study by three additional setups, for which we drop the restrictions on the layered structure of the trained NN models in favor of comparable numbers of TPs.

As shown in Table 7, the GNN model with the highest value of μ_{AUC} (labelled as $\mathrm{GNN}_{k=2}$) is based on $k = 2$ and $n = (26, 39)$, with $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2}) = 2809$ and a value of $\mu_{\mathrm{AUC}} = 0.8484 \pm 0.0008$. The DNN with the highest result of μ_{AUC}

Table 7 Summary of n, number of TPs (N_{TP}), and μ_{AUC} of the (upper part) GNN and (lower part) DNN models with the highest results in μ_{AUC}. For the DNN models a number of effective TPs $N_{\mathrm{TP}}^{\mathrm{eff}}$, as defined in the text, is also given in parentheses. Corresponding summaries are given for the cases of $k = 1$ and 2. In all cases m has been used as relational information between nodes/physics objects. Also shown are the configurations of three additional NN models discussed in Section 4.2: the $\mathrm{GNN}^{\uparrow}$ model with N_{TP} comparable within 1% with $N_{\mathrm{TP}}(\mathrm{DNN}_{k=2})$ and the $\mathrm{DNN}^{\downarrow}$ ($\mathrm{DNN}^{\downarrow}_{\mathrm{eff}}$) model with N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$) comparable within 1% with $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2})$.

NN arch.	n	N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$)	μ_{AUC}	Label
GNN	(39)	1093	0.8387±0.0006	$\mathrm{GNN}_{k=1}$
	(26, 39)	2809	0.8484±0.0008	$\mathrm{GNN}_{k=2}$
	(29, 28, 29, 29)	5829	0.8546±0.0006	$\mathrm{GNN}^{\uparrow}$
DNN	(26)	5773 (4753)	0.8300±0.0011	$\mathrm{DNN}_{k=1}$
	(26, 39)	6839 (5819)	0.8388±0.0007	$\mathrm{DNN}_{k=2}$
	(13, 14, 14)	3294 (2784)	0.8400±0.0006	$\mathrm{DNN}^{\downarrow}_{\mathrm{eff}}$
	(11, 10, 11, 11)	2816 (2385)	0.8386±0.0007	$\mathrm{DNN}^{\downarrow}$

(labelled as $\mathrm{DNN}_{k=2}$) is based on the same configuration in terms of k and n, with $N_{\mathrm{TP}}(\mathrm{DNN}_{k=2}) = 6839$ and a value of $\mu_{\mathrm{AUC}} = 0.8388 \pm 0.0007$. One may argue that for the DNN model, not the full set of TPs is really actively contributing to the solution of the task, since part of the input space is regularly filled with zeros, e.g., if fewer than eight jets are selected in an event. Therefore, we estimate, in addition to N_{TP}, a number of effective TPs ($N_{\mathrm{TP}}^{\mathrm{eff}}$) from the product of N_{TP} with the average number of nonzero input nodes in $\mathbf{x}^{\mathrm{DNN}}$, evaluated on the training dataset. This results in a value of $N_{\mathrm{TP}}^{\mathrm{eff}}(\mathrm{DNN}_{k=2}) = 5819$.

In a first approach we survey varying DNN structures with N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$) comparable to $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2})$. We do this based on the following algorithm: We allow $k \leq 4$ and any number of nodes per hidden layer (n). Of all DNN configurations for which $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2})$ is matched by N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$) within a margin of 1%, the model with the smallest spread of n across layers is selected. If no DNN configuration with N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$) within a 1% margin of the target value can be found the closest possible model is chosen. This situation occurs only in models with one hidden layer.

This procedure ensures a homogeneous structure of hidden layers. As a result, e.g., a DNN configuration with $k = 4$ and $n = (11, 10, 11, 11)$, with $N_{\mathrm{TP}} = 2816$, further on referred to as $\mathrm{DNN}^{\downarrow}$, is preferred over a model with $k = 4$ and $n = (10, 8, 14, 24)$, even though the latter results in an exact match with $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2})$. The $\mathrm{DNN}^{\downarrow}$ model achieves a value of $\mu_{\mathrm{AUC}} = 0.8386 \pm 0.0007$. A second DNN configuration with $k = 3$ and $n = (13, 14, 14)$, with $N_{\mathrm{TP}}^{\mathrm{eff}} = 2784$ within the 1% margin of $N_{\mathrm{TP}}(\mathrm{GNN}_{k=2})$ is also considered and further on referred to as $\mathrm{DNN}_{\mathrm{eff}}^{\downarrow}$. This model achieves a value of $\mu_{\mathrm{AUC}} = 0.8400 \pm 0.0006$. We observe that, although the $\mathrm{DNN}^{\downarrow}$ ($\mathrm{DNN}_{\mathrm{eff}}^{\downarrow}$) model uses only 41% (48%) of N_{TP} ($N_{\mathrm{TP}}^{\mathrm{eff}}$) of $\mathrm{DNN}_{k=2}$, this does not result in any significant loss in separation power, after training.

In a second approach we survey varying GNN structures with N_{TP} comparable to $N_{\mathrm{TP}}^{\mathrm{eff}}(\mathrm{DNN}_{k=2})$. For this purpose we exploit the same algorithm as described above, resulting in a GNN with $k = 4$ and $n = (29, 28, 29, 29)$ with $N_{\mathrm{TP}} = 5829$, further on referred to as $\mathrm{GNN}^{\uparrow}$. This model achieves a value of $\mu_{\mathrm{AUC}} = 0.8546 \pm 0.0006$. It reveals the highest value of μ_{AUC} across all tested models. The difference in μ_{AUC} with respect to other NN configurations is only at the 1%-level, but it is still significant in terms of $\Delta\mu_{\mathrm{AUC}}$.

From this finding we conclude that the GNN model with the same expressiveness as a maximally comparable DNN must have intrinsic advantages over the DNN model in extracting additional information from the given, large training sample. For the benchmark setup in use, this advantage is small but significant in the scope of the study. It emerges after external augmentation with an energy-weighted distance measure like m between the input objects/nodes, and more clearly manifests itself in the study for $k > 1$. We anticipate that this gain originates from the hierarchically structured information about nearest neighbors and the nearest neighborhood of nearest neighbors of node i, when viewing the edge weights ω_{ij} as a distance measure. This information is an intrinsic property of the GNN model and not easily accessible through the more simplistic DNN structure. An increase of TPs of the simpler DNN structure does not compensate for this informational advantage. In this interpretation the gain of $\mathrm{GNN}^{\uparrow}$ over all other configurations should mostly be attributed to the increase in k over the associa-

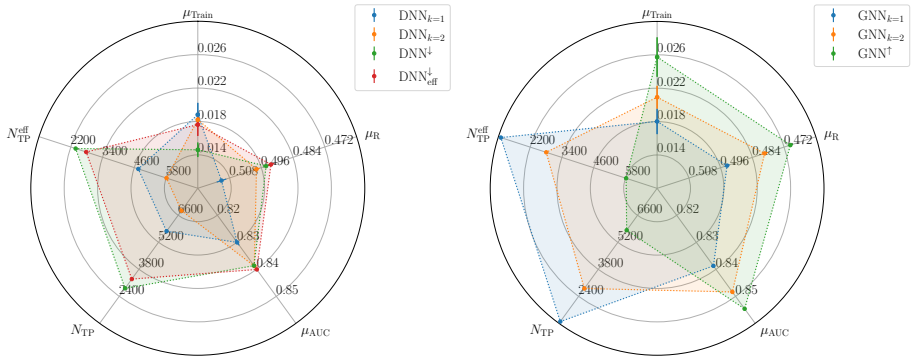


Fig. 7 Visualization of GNN and DNN models with the highest values in μ_{AUC} and $k = 1$ ($\text{GNN}_{k=1}$, $\text{DNN}_{k=1}$) and $k = 2$ ($\text{GNN}_{k=2}$, $\text{DNN}_{k=2}$). Also shown are the DNN models $\text{DNN}^{\downarrow}$ and $\text{DNN}_{\text{eff}}^{\downarrow}$ with a comparable number of (effective) TPs as the $\text{GNN}_{k=2}$ model, and the GNN model $\text{GNN}^{\uparrow}$ with a comparable number of effective TPs as the $\text{DNN}_{k=2}$ model. Five different metrics to evaluate the properties of an NN model are shown: N_{TP} ($N_{\text{TP}}^{\text{eff}}$), the mean convergence rate μ_{Train}, the mean empirical risk evaluated on the test sample μ_R, and μ_{AUC}. The spanned area in the figures indicates the capability of an NN model to fulfill the task.

tion of n per GraphConv operation. The parameter choices of the $\text{DNN}^{\downarrow}$, $\text{DNN}_{\text{eff}}^{\downarrow}$, and $\text{GNN}^{\uparrow}$ models and correspondingly achieved values of μ_{AUC} are also given in Table 7.

4.3 Convergence behavior

In this study we have investigated the capacities of GNN and DNN models to fulfill their primary target, i.e., to provide the best possible solutions to the classification task defined in Section 2.3. We anticipate that, in particular in practical life, the properties of an NN architecture may be evaluated in other terms, viz. the mean of the training speed μ_{Train}, which we evaluate as the inverse of the epoch with the highest value of the ROC-AUC on the validation dataset and the mean of the empirical risk obtained from the test dataset μ_R, which we take as a measure of the generalization property of the given NN model under study. To conclude our studies we provide a visualization of these properties and all other properties of the NN models that have been discussed throughout the paper so far, in Figure 7. In Figure 7 (left) μ_{Train}, μ_R, μ_{AUC}, N_{TP}, and $N_{\text{TP}}^{\text{eff}}$ for the $\text{DNN}_{k=1}$, $\text{DNN}_{k=2}$, $\text{DNN}^{\downarrow}$, and $\text{DNN}_{\text{eff}}^{\downarrow}$ models are shown, on five independent axes. The axes are defined such that values closer to the common origin of the figure are disfavored. This is in particular true for N_{TP}, $N_{\text{TP}}^{\text{eff}}$, and μ_R, where the values are given in descending order when moving away from the origin. In this sense a larger size of the correspondingly colored area indicates the larger capability of a given NN model to adequately solve the presented task. In Figure 7 (right) the same quantities are shown for the $\text{GNN}_{k=1}$, $\text{GNN}_{k=2}$, and $\text{GNN}^{\uparrow}$ models. The axes ranges are kept the same to ease comparison between both architectures.

In terms of μ_{Train} the DNN models usually fall behind their corresponding GNN counterparts. This finding, as well as the observation that the GNN models usually achieve a comparable or slightly larger value of μ_{AUC}, after training,

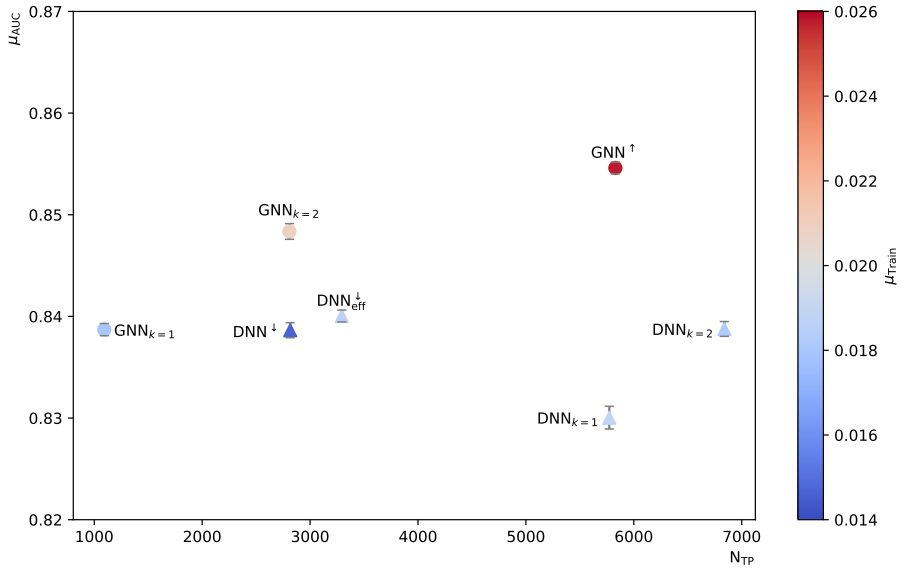


Fig. 8 Overview of the GNN and DNN models with $k = 1$ and 2, and the highest achieved values of μ_{AUC}, as well as a GNN model with an increased number of TPs ($\mathrm{GNN}^{\uparrow}$) and a DNN model $\mathrm{DNN}^{\downarrow}$ ($\mathrm{DNN}^{\downarrow}_{\mathrm{eff}}$) with a restricted (effective) number of TPs. The GNN models are indicated by circles and the DNN models by triangles. The models are shown in a three-dimensional space built from μ_{AUC}, N_{TP} and μ_{Train}. The bars in μ_{AUC} are obtained from the sample variance of a training ensemble, as described in Section 2.4. The quantity μ_{Train}, indicated by the color code of the points corresponds to the inverse of the epoch with the highest value of the ROC-AUC on the validation dataset.

indicates the assumed effect of guiding the convergence by constraints, which are built-in to the GNN architecture. This property allows the GNN architecture not only to converge to a solution that is equally good or slightly better than the solution found by corresponding DNN architectures, but also to converge faster and typically with a smaller number of TPs. We note though that a clear correlation between μ_{Train} and N_{TP} ($N^{\mathrm{eff}}_{\mathrm{TP}}$) cannot be deduced from our study. We understand this situation such that a less expressive NN model, with fewer degrees of freedom for the minimization process, may well lead to a more pronounced landscape of the expected risk and thus reduced μ_{Train}. In addition we note that also μ_R for the DNN models falls behind, compared to their GNN counterparts. Here we observe the benefit of a regularizing effect of the built-in constraints, which correlates with reduced numbers of (effective) TPs. It is obvious that an NN model with more TPs reveals a higher vulnerability to specific properties of the training sample. A summary of all NN configurations that have been discussed in this section is shown in Figure 8.

5 Summary

With this paper we have made an effort to put the comparison of graph neural network (GNN) and equivalent fully-connected feed-forward neural network (DNN)

architectures on a maximally fair ground. Under the laboratory conditions of a high-energy physics process of interest, at the CERN LHC, we have controlled the definition of the task, choice of non-tunable (hyper-)parameters of the models, which are not subject to the training, and amount of information primarily presented to the neural network models, through their input feature vectors.

Within the scope of the study we have demonstrated clear evidence that the presumable advantage of the more complex GNN over an equivalent DNN structure does not originate from an uncontrolled mixture of features in the embedding space of the graph nodes, but from the extra relational information between nodes, that we have added based on domain-knowledge. Without this extra knowledge the GNN models fall behind equivalent DNN models in terms of their capability to separate signal from background. Neither are the GNN models superior in terms of their separation power to equivalent DNN models, as long as these are equipped with the same information in the input space.

Both, the built-in permutation invariance and the circumstance that the GNN architecture a priori is not bound to a fixed number of nodes might be viewed as advantages. They might also have positive influence on the convergence behavior of the training. On the other hand it cannot be deduced that either of these properties significantly contributes to an increase, e.g., in the power to separate a given signal from background. Any advantage of the GNN over the DNN architecture that we observed in our studies could be traced back to the access to more information, which, when given to the other architecture lead to the same performance also for the other architecture.

The real advantage of the GNN over the DNN structure emerges as soon as more than one GraphConv operations are applied, during which the GNN structure naturally accesses more relational information between nodes than a DNN has access to. In the configurations investigated in our study this gain is tied to the use of relational information that can be interpreted as a distance measure to define proximity between two nodes. Apart from that we observe significant advantages of the GNN over the DNN architecture in terms of convergence and generalizability that we attribute to a level of built-in implicit constraints to the GNN model resulting in a better ratio of accessible information over trainable parameters of the model. We anticipate that these advantages might be more pronounced the more hierarchical the training data are. We assume that this property is the basis for the success of GNN structures when applied to particle physics jets, which are highly hierarchical objects. In conclusion we expect the highest gain of a GNN over a DNN structure for tasks based on hierarchically structured data, ideally based a known distance measure.

ACKNOWLEDGEMENTS

This research was supported by the German Federal Ministry of Education and Research (BMBF) under grant 05H21VKCCB.

References

1. D. Aguillard, T. Albahri, et al. Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm. *Phys. Rev. Lett.*, 131(16):161802, 2023.

2. N. Metropolis, A. Rosenbluth, et al. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087, 2004.
3. M. Rosenbluth. Genesis of the Monte Carlo Algorithm for Statistical Mechanics. *AIP Conference Proceedings*, 690(1):22, 2003.
4. G. Aad, E. Abat, et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
5. S. Chatrchyan, G. Hmayakyan, et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
6. G. Salam. Towards Jetography. *Eur.Phys.J.*, C67:637–686, 2010.
7. A. Sirunyan, A. Tumasyan, et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *JINST*, 14(07):P07004, 2019.
8. R. Feynman. The theory of positrons. *Phys. Rev.*, 76:749–759, Sep 1949.
9. F. Scarselli, M. Gori, et al. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.
10. A. Micheli. Neural network for graphs: A contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3):498–511, 2009.
11. T. Kipf and M. Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv e-prints*, page arXiv:1609.02907, September 2016.
12. W. Hamilton, R. Ying, et al. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1025–1035, Red Hook, NY, USA, 2017. Curran Associates Inc.
13. P. Veličković, G. Cucurull, et al. Graph Attention Networks. *International Conference on Learning Representations*, 2018.
14. J. Shlomi, P. Battaglia, et al. Graph Neural Networks in Particle Physics. 7 2020.
15. A. Sirunyan, A. Tumasyan, et al. Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:026, 2019.
16. R. Workman, V. Burkert, et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.
17. A. Sirunyan, A. Tumasyan, et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.
18. J. Alwall, R. Frederix, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
19. J. Alwall, M. Herquet, et al. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.
20. T. Sjöstrand, S. Ask, et al. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
21. J. de Favereau, C. Delaere, et al. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.
22. M. Cacciari, G. Salam, et al. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.
23. M. Fey and J. Lenssen. Fast graph representation learning with pytorch geometric, 2019.
24. A. Paszke, S. Gross, et al. Pytorch: An imperative style, high-performance deep learning library, 2019.
25. D. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. 12 2014.
26. C. Morris, M. Ritzert, et al. Weisfeiler and leman go neural: Higher-order graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4602–4609, Jul. 2019.

Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

ZU-TH 12/24, KA-TP-03-2024, P3H-24-011

A comparative study of flavour-sensitive observables in hadronic Higgs decays

Benjamin Campillo Aveleira[a,1,2], **Aude Gehrmann–De Ridder**[b,1,3], **Christian T Preuss**[c,1,4]

[1]Institute for Theoretical Physics, ETH, 8093 Zürich, Switzerland
[2]Institute for Theoretical Physics, KIT, 76131 Karlsruhe, Germany
[3]Department of Physics, University of Zürich, 8057 Zürich, Switzerland
[4]Department of Physics, University of Wuppertal, 42119 Wuppertal, Germany

Received: date / Accepted: date

arXiv:2402.17379v1 [hep-ph] 27 Feb 2024

Abstract Jet production from hadronic Higgs decays at future lepton colliders will have significantly different phenomenological implications than jet production via off-shell photon and Z-boson decays, owing to the fact that Higgs bosons decay to both pairs of quarks and gluons. We compute observables involving flavoured jets in hadronic Higgs decays to three partons at Born level including next-to-leading order corrections in QCD (i.e up to $\mathcal{O}(\alpha_s^2)$). The calculation is performed in the framework of an effective theory in which the Higgs boson couples directly to gluons and massless b-quarks retaining a non-vanishing Yukawa coupling. For the following flavour sensitive observables: the energy of the leading and subleading flavoured jet, the angular separation and the invariant mass of the leading b-$\bar{b}$ pair, we contrast the results obtained in both Higgs decay categories and using either of the infrared-safe flavoured jet algorithms flavour-k_T and flavour-dressing.

Keywords flavoured jets · hadronic Higgs decays · NLO QCD

1 Introduction

Precision studies of the Higgs boson discovered at LHC by CMS and ATLAS [1,2] will become possible at future lepton colliders such as [3,4], all aiming to operate as Higgs factories. In this clean experimental environment, where interactions take place at well-defined centre-of mass energies, it is expected to enable model-independent measurements of the Higgs couplings to gauge bosons and fermions at the level of a few percent. At these future lepton colliders, in particular it will become possible to have access to so-far unobserved hadronic decay channels such as Higgs decays to gluons. The latter is currently inaccessible in hadron-collider environments due to the presence of overwhelming QCD backgrounds. Only the $H \to b\bar{b}$ decay was observed to date [5,6] in associated vector-boson production where the leptonic decay signature of the vector boson helps to identify the $H \to b\bar{b}$ decay.

Hadronic Higgs decays to at least two final state hard partons proceed via two main decay modes; either as Yukawa-induced decay to a bottom-quark pair, $H \to b\bar{b}$, or as a heavy-quark-loop induced decay to two gluons, $H \to gg$. In the latter category, observables are computed in the framework of an effective field theory, in which the top-quark loop is integrated out into an effective point-like Hgg vertex.

So far these two categories of Higgs decay processes have been considered together in the computation of flavour-agnostic event-shape observables, i.e., for three-jet-like final states in [7,8,9,10,11] and for four jet-like final states in [12]. It was also recently suggested to determine branching ratios in hadronic Higgs decays via fractional energy correlators [13]. Flavour-sensitive jet observables related to the presence of a flavoured jet in the final state have so far been computed for the following LHC processes: VH production, with $H \to b\bar{b}$ or $Z + b$-jet and $Z/W + c$-jet, with the vector boson decaying leptonicallly in all cases. More precisely, parton-level predictions including up to NNLO QCD corrections using massless charm or bottom quarks at the origin of the flavoured jet have been computed most re-

[a]e-mail: benjamin.campillo@kit.edu
[b]e-mail: gehra@phys.ethz.ch
[c]e-mail: preuss@uni-wuppertal.de

cently for VH in [14,15,16,17], and for $Z/W + c/b$ in [18,19,20,21,22,23].

In a lepton collider environment, flavour-sensitive jet observables associated to Z decays have been lately considered in [24,25] while colour-sensitive observables have been used to disentangle bottom quarks stemming from Higgs decays against those stemming fror QCD background in [26].

All of these computations employ an infrared-safe procedure to define flavoured jets from massless partons. The latter procedure requires the use of an infrared-safe recombination algorithm to cluster flavourless and flavoured partons into final states including well-defined flavoured jets. Up to very recently, only the flavour-k_{T} algorithm [27] was used in theoretical computations at hadron colliders. Lately, a number of flavour-sensitive jet algorithms have been designed [24,25,28,29], with increased interest in providing modified versions of the anti-k_{T} algorithm [24,25,29], to improve the data versus theory comparisons at LHC.

All of these algorithms share the principle that, at least up to a certain order in the strong coupling, they can be proven to be infrared flavour safe, i.e., that the flavour assignment of a given jet is not spoilt by the emission of unresolved (soft or collinear) massless partons. A detailed comparison of currently available algorithms has been presented in [25].

In this paper, we compute a range of flavour-sensitive three-jet like observables in hadronic Higgs decays, including NLO QCD corrections for both decay categories, i.e related to $H \to b\bar{b}$ and $H \to gg$, as alluded to above. To define flavour-sensitive observables we employ the flavour-k_{T} algorithm [27] and the flavour-dressing algorithm [24] which can both be applied in a lepton collider environment.

The study is organised as follows: In section 2 we give a brief overview of the definition of flavoured jets using either of the two infrared flavour-safe jet algorithms before describing our computational setup in section 3. In section 4, in a first part presented in section 4.2 we compute the so-called "misindentified cross section" to check the correctness of our implementation using both infrared-safe and flavour-safe jet algorithms. In the second part of section 4, i.e., in section 4.3, we present theoretical predictions up to second order in QCD for four different flavour-sensitive observables related to both Higgs-decay categories and compare the results using both flavour-safe jet algorithms. We conclude and give an outlook on future work in section 5.

2 Flavoured jets

In experimental analyses as well as theoretical calculations, final-state particle configurations are often described by so-called jets. While the definition of jets is fixed by the choice of a sequential recombination algorithm known as a jet algorithm, the association of the jet with a given parton (or hadron) flavour is a more complicated endeavour. Naïvely, it may be tempting to define the jet flavour as the sum of the constituent flavours of each jet, i.e,

$$\text{Jet flavour} = f - \bar{f}\,, \tag{1}$$

where f is the number of particles with a given a flavour and $\bar{f}$ is the number of particles with the corresponding anti-flavour. However, this naïve approach violates an important property known as infrared flavour safety [27], which describes whether an algorithm respects physically meaningful flavour assignments when one or more particles become unresolved. The problem is that jet algorithms may in general cluster the flavoured daughters of a soft wide-angle gluon splitting $g \to q\bar{q}$ into different jets, changing their flavour association. In effect, the flavours of the two jets, and therefore any flavour-dependent observables, become explicitly dependent on the presence of a pair of unresolved particles, in violation of infrared safety. It is thus mandatory for a jet-flavour definition to respect infrared flavour safety, meaning that the presence of a pair of unresolved flavoured particles to an event must not change the flavours of the jets.

As mentioned in the introduction, several approaches to infrared-safe flavoured jet algorithms have been explored [24,25,27,28,29], of which only the flavour-k_{T} [27] and the flavour-dressing algorithm [24] explicitly describe an implementation compatible with the use of the Durham (k_{T}) algorithm. In the present study we therefore limit the discussion to these two flavoured jet algorithms, which we briefly describe below.

The flavour-k_{T} algorithm [27] achieves infrared flavour safety by modifying the Durham distance measure valid for all unflavoured partons (i, j) [30] into two cases defined depending whether the softer particles to be clustered are flavoured or not. In the flavour-k_{T} algorithm, the distance measure is given by:

$$y_{ij}^{\mathrm{F}} = \frac{2(1-\cos\theta_{ij})}{E_{\mathrm{tot}}^2}\min(E_i,E_j)^{2-\alpha}\max(E_i,E_j)^{\alpha} \tag{2}$$

if the softer of i, j is flavoured, Instead the regular Durham distance given by,

$$y_{ij}^{\mathrm{F}} = \frac{2(1-\cos\theta_{ij})}{E_{\mathrm{tot}}^2}\min(E_i,E_j)^{2}\,, \tag{3}$$

is kept if the softer of i, j is flavourless. Here, $\alpha \in (0, 2]$ is an arbitrary parameter. In this study, the value $\alpha = 2$ is chosen everywhere. The distance measure in eq. (2) ensures that soft pairs of flavoured particles are recombined first and thus avoids the previously discussed infrared-safety problem present in the standard (flavour-agnostic) Durham (k_T) algorithm. The main draw-back to the use of this algorithm is that it requires the flavour information of all particles, thus making it difficult to use in experimental analyses. Indeed, so far, it has not be used in measurements of flavour-sensitive jet observables.

The flavour-dressing algorithm [24] on the other hand does not change the underlying jet algorithm but instead alters the way flavours are assigned to jets. Instead of using the naïve jet-flavour definition, a more complex definition is invoked which ensures infrared flavour safety of the jet algorithm at hand. The basic idea is to first cluster an event into jets using a flavour-agnostic jet algorithm of choice, and, in a second step, cluster flavoured objects into so-called flavour clusters using a technique akin to soft-drop grooming [31]. In the final, "association" step the flavour clusters of step two are assigned to the jets of step one. A striking and important feature of the flavour-dressing algorithm is that it is equally applicable to theoretical calculations (and full particle-level simulations) as to experimental analyses, as discussed in [24]. It can in particular be used together with the anti-k_T algorithm, mostly used in experimental analyses at LHC but also in combination with the Durham k_T algorithm as explored in this paper.

3 Computational setup

We perform the computation of flavoured observables related to hadronic Higgs decays using the parton-level Monte-Carlo generator EERAD3 [32,33], which was originally developed to compute NNLO QCD corrections to event-shape observables in hadronic Z-decays. This generator was recently extended to compute event shapes related to hadronic Higgs decays with three- and four-jet configurations at Born level in [7,12]. In both cases, the antenna subtraction method is used to regulate infrared divergences related to real radiation contributions.

The hadronic Higgs-decay observables are computed in the framework of an effective theory including two Higgs decay categories. In the first category, the Higgs boson couples directly to gluons via an effective Higgs-gluon-gluon vertex and in the other, Standard-Model-like category, the Higgs boson decays into a massless b-quark pair retaining a non-vanishing Yukawa coupling [7,12]. This can be recasted into the following effective Lagrangian

$$\mathcal{L}_{\text{Higgs}} = -\frac{\lambda(M_t, \mu_{\text{R}})}{4} H G^a_{\mu\nu} G^{a,\mu\nu} + \frac{y_b(\mu_{\text{R}})}{\sqrt{2}} H \bar{\psi}_b \psi_b \,. \quad (4)$$

In this context, the effective Higgs-gluon-gluon coupling is given in terms of the Higgs vacuum expectation value v by

$$\lambda(M_t, \mu_{\text{R}}) = -\frac{\alpha_{\text{s}}(\mu_{\text{R}}) C(M_t, \mu_{\text{R}})}{3\pi v} \quad (5)$$

and the $Hb\bar{b}$ Yukawa coupling reads

$$y_b(\mu_{\text{R}}) = \bar{m}_b(\mu_{\text{R}}) \frac{4\pi\alpha}{\sqrt{2} M_W \sin\theta_{\text{W}}} \,. \quad (6)$$

Both λ and y_b are subject to renormalisation, which we perform at scale μ_{R} in the $\overline{\text{MS}}$ scheme using $N_{\text{F}} = 5$. The top-quark Wilson coefficient is evaluated at first order in α_{s} using the results of [34,35,36,37,38,39,40], and the running of the Yukawa mass $\bar{m}_b$ is performed using the results of [41].

It is important to highlight that the terms in eq. (4) do not interfere under the assumption of kinematically massless quarks. In particular, they do not mix under renormalisation [9]. This allows to define two separate Higgs-decay categories and to compute higher-order corrections independently for each. Throughout, we therefore consider predictions for the $H \to b\bar{b}$ and $H \to gg$ categories separately. All partonic contributions yielding three hard partons in the final state at Born level and needed for the computation presented in this paper, have been presented in detail for both categories up to $\mathcal{O}(\alpha_{\text{s}}^2)$ in [7]. In particular, it is worth mentioning that the Born-level partonic processes contributing at $\mathcal{O}(\alpha_{\text{s}})$ are: $H \to b\bar{b}g$ in the $H \to b\bar{b}$ category and $H \to ggg$, $H \to gq\bar{q}$ in the $H \to gg$ category. In the latter case, q stands for any quark with specific flavour, including the bottom quark.

For any infrared-safe observables O, the parton-level generator EERAD3 calculates the LO coefficient $\bar{A}$ and the NLO coefficient $\bar{B}$ in a perturbative expansion of the differential decay width,

$$\frac{1}{\Gamma^{(n)}_{2j}(\mu_{\text{R}})} \frac{\mathrm{d}\Gamma}{\mathrm{d}O} = \left(\frac{\alpha_{\text{s}}(\mu_{\text{R}}^2)}{2\pi}\right) \frac{\mathrm{d}\bar{A}}{\mathrm{d}O} + \left(\frac{\alpha_{\text{s}}(\mu_{\text{R}}^2)}{2\pi}\right)^2 \left(\frac{\mathrm{d}\bar{B}}{\mathrm{d}O} + \beta_0 \log\left(\frac{\mu_{\text{R}}^2}{M^2}\right) \frac{\mathrm{d}\bar{A}}{\mathrm{d}O}\right) . \quad (7)$$

Here, $\Gamma^{(n)}_{2j}$ is the partial two-body decay width to order n; specifically, $n = 0$ at LO and $n = 1$ at NLO.

In eq. (7), the LO coefficient $\bar{A}$ involves only an integration over the three-particle phase space, while the NLO contribution $\bar{B}$ involves an integration over the

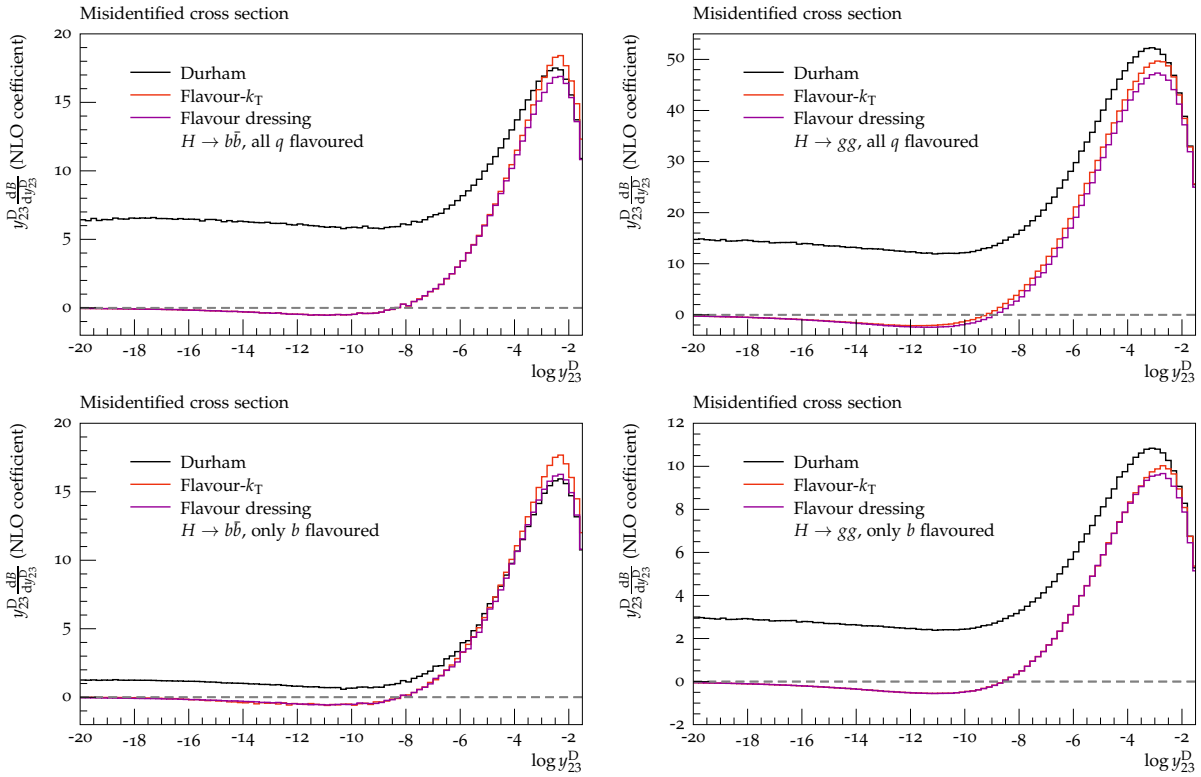


Fig. 1 Misidentified cross section in the $H \to b\bar{b}$ (left) and $H \to gg$ (right) category, considering all quarks as flavoured (top row) and only b-quarks as flavoured (bottom row). Both flavour-k_T and flavour dressing use $\alpha = 2$, as defined in the text.

four-particle phase space related to the real-radiation contribution and real subtraction terms, and an integration over the three-particle phase space, pertaining to one-loop contributions and virtual subtraction terms. Within this study, we employ the antenna subtraction scheme to construct real and virtual subtraction terms. Explicit expressions for the perturbative coefficients $\bar{A}$ and $\bar{B}$ can be found in [7].

In order to compute flavour-sensitive observables related to hadronic Higgs decays, as in this paper, on top of the ingredients needed to compute flavour-agnostic observables as described above, a new flavour layer needs to be implemented in the parton-level event generator EERAD3. This parton-level flavour-tracking procedure implemented here in EERAD3 for the first time can be summarized as follows: For each momentum configuration all contributing flavour configurations are generated and matrix elements as well as subtraction terms are evaluated separately for each flavoured parton configuration. In particular, this means that the calculation is split into different flavour contributions across all layers in eq. (7). Because the subtraction terms involve mapped configurations in which one parton is clustered into a "reduced" particle configuration, this has the consequence that the subtraction scheme is only viable if a flavour-safe jet algorithm is used. For each flavour configuration, the flavour layer then acts as an additional input to the jet algorithm and parton-level contributions needed to be considered for the evaluation of the observables.

To conclude this section, we wish to define the theoretical framework in which our predictions are valid and numerically stable. Fixed-order calculations are accurate only in phase-space regions in which hard, well-separated jets dominate. We therefore devise a resolution cut, y_{cut} on the (flavoured) jet algorithm and define three-jet states to have three particles with $\min_{i,j}(y_{ij}) \geq y_{\text{cut}}$, where y_{ij} denotes the distance measure of the respective jet algorithm. Similarly, four-jet states are defined to have four particles with $\min_{i,j}(y_{ij}) \geq y_{\text{cut}}$. Furthermore, to avoid large numerical cancellations between the real contribution and the corresponding real subtraction term in unresolved phase-space regions, we implement a technical cut-off of $y_0 = 10^{-8}$ on the smallest dimensionless two-particle invariant $y_{ij} = 2p_ip_j/s$ in real configurations. We have verified that our predictions are independent of the choice of this theoretical cut-off.

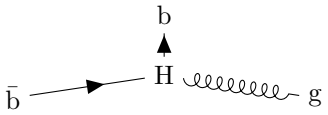


Fig. 2 Momentum-space diagram in the Higgs rest frame of a $H \to b\bar{b}g$ decay in which flavour-$k_{\rm T}$ allows for an arbitrarily soft b-quark.

4 Results

We here focus on presenting theoretical predictions for the following flavour-sensitive observables: the energy of the leading and subleading flavoured jet, the angular separation and invariant mass of the leading b-$\bar{b}$ pair. We shall present results obtained in both Higgs-decay categories and using either of the infrared-safe flavoured jet algorithms flavour-$k_{\rm T}$ and flavour-dressing. After presenting the numerical set-up in section 4.1 we divide the discussion of our results into two subsections. In section 4.2 we validate the infrared flavour safety of the flavour-$k_{\rm T}$ and flavour-dressing algorithms before presenting and comparing our predictions for flavour-sensitive observables in section 4.3.

4.1 Numerical set-up and scale-variation prescription

We consider on-shell Higgs decays with $M_H = 125$ GeV and calculate all observables in the resonance centre-of-mass frame. We use $\alpha_{\rm s}(M_Z) = 0.1179$ with either one- or two-loop running at LO and NLO, respectively. Electroweak quantities are considered constant and we use the G_μ-scheme with

$$\begin{aligned} G_{\rm F} &= 1.20495 \times 10^{-5}\ \text{G}e\text{V}^{-2}\,, \\ M_Z &= 91.1876\ \text{G}e\text{V}\,, \\ M_W &= 80.385\ \text{G}e\text{V}\,, \end{aligned} \tag{8}$$

corresponding to $\alpha = \frac{1}{128}$. To estimate theoretical uncertainties from missing higher-order corrections in our calculation, we vary the renormalisation scale around the Higgs mass, i.e., consider $\mu_{\rm R} = k_\mu M_H$ with $0.5 \leq k_\mu \leq 2$.

4.2 Infrared flavour safety

To guarantee infrared flavour safety for flavour-sensitive observables, the jet algorithms that are employed in the computation need to correctly assign flavours to jets in the deep infrared region. Correct flavour assignment is determined by the underlying two parton process, i.e., by either two flavourless jets in the $H \to gg$ category or by one flavoured and one anti-flavoured jet in the $H \to b\bar{b}$ category.

The study of infrared-flavour safety was first conducted for hadronic Z-decays in [27], in which it was highlighted that the Durham algorithm violates infrared flavour safety and a flavour-safe modification in terms of the flavour-$k_{\rm T}$ algorithm was suggested. As a measure of infrared flavour safety, the so-called misidentified cross section was defined in terms of the three-jet resolution variable y_{23}. The latter variable, measures the departure from two-jet-like into three-jet-like topologies.

In [24] the same criterion was used to validate the infrared flavour safety of the flavour-dressing algorithm. As a validation of our implementation, we thus apply the same criterion but here for hadronic Higgs decays. In this case, the misidentified cross section collects two-jet like events for which the flavour does not correspond to the flavour of the original two parton like topology in either of the Higgs categories. In fig. 1, we present the NLO coefficient $\bar{B}$ differential in the Durham resolution $y_{23}^{\rm D}$ using jet definitions according to the plain Durham, flavour-$k_{\rm T}$, and flavour-dressing algorithm. We employ two different flavour definitions; in the top row all quarks are counted as flavoured, in line with the definitions used in [24,27], while in the bottom row only b-quarks are treated as flavoured with all other quarks being flavourless. The former (in $H \to b\bar{b}$) can be used as a fundamental cross-check of our implementation with the original papers [24,27], whereas the latter is the flavour definition used in the remainder of this paper, since it allows for a straightforward experimental adaption in terms of b-tagging, see e.g. [42,43,44]. In fig. 1, to probe the deep infrared region, we have used the value of the theoretical cut y_0 to be 10^{-13}.

For an infrared-safe flavour jet algorithm, one expects the probability for flavour misidentification to vanish as the variable y_{23} tends to zero. In both Higgs-decay categories and regardless of the flavour-definition, in fig. 1 it is clearly visible that the Durham algorithm has a non-zero probability for an incorrect flavour assignment in the deep infrared region, whereas both the flavour-$k_{\rm T}$ and flavour-dressing algorithm yield a vanishing cross section for misidentified events. Confirming the findings in [24,27], both the flavour-$k_{\rm T}$ and flavour-dressing algorithms provide infrared-safe flavour jet definitions in hadronic Higgs decays, whereas the flavour-agnostic Durham algorithm does not. Thus, we shall present predictions only using these two jet algorithms in the remainder of this section.

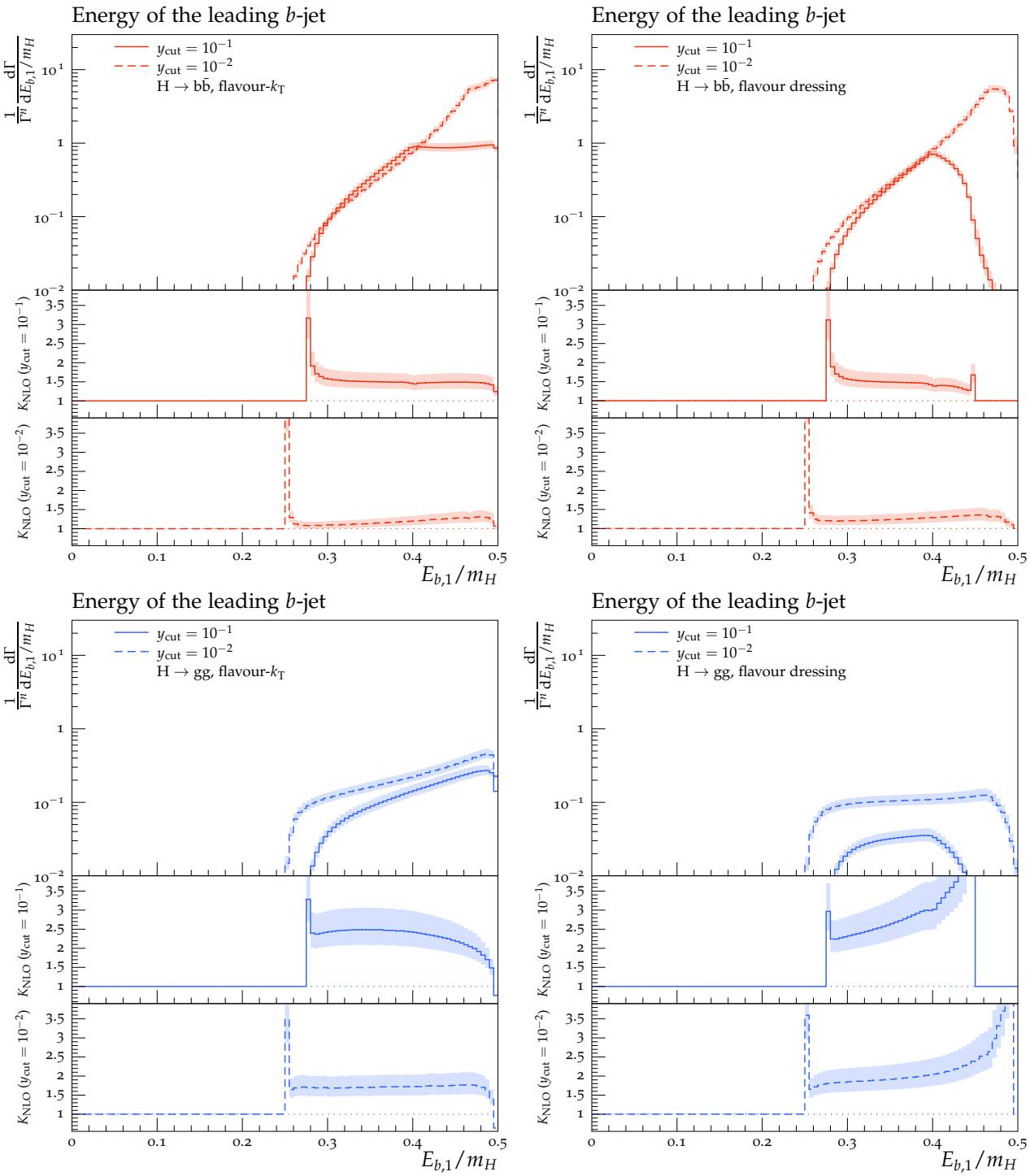


Fig. 3 Energy of the leading b-jet in the $H \to b\bar{b}$ (*top row*) and $H \to gg$ (*bottom row*) decay category using the flavour-k_{T} (*left column*) and flavour dressing (*right column*) algorithm.

4.3 Flavour-sensitive observables

We consider flavour-sensitive observables in both Higgs decay categories yielding three-jet-like configurations at Born level. For a given experimental resolution parameter y_{cut} and for each pair of final state partons i, j (quark or gluon), present in the partonic subprocesses at leading and next-to-leading order, we impose that $\min_{i,j}(y_{ij}) \geq y_{\text{cut}}$ in the respective flavoured jet algorithm. We then calculate NLO QCD, i.e ($\mathcal{O}(\alpha_s^2)$), predictions for the following observables

(a) the energy of the leading flavoured jet, $E_{b,1}$;
(b) the energy of the subleading flavoured jet, $E_{b,2}$;
(c) the angular separation of the leading b- and $\bar{b}$-jet, $\cos\theta_{b\bar{b}}$;
(d) the invariant mass of the leading b- and $\bar{b}$-jet, $m_{b\bar{b}}$.

Except for the angular separation, we only consider scaled observables, normalised to the Higgs mass m_H. In all cases, we choose $\alpha = 2$ in the flavour-k_{T} and flavour-dressing algorithm and consider only b-quarks ($\bar{b}$-quarks) as flavoured (anti-flavoured).

Theoretical predictions for the four flavour-sensitive observables defined above are shown in figs. 3 to 6 for two different jet-algorithm resolution parameter, $y_{\text{cut}} = 0.1$ (solid lines) and $y_{\text{cut}} = 0.01$ (dashed lines). In all figures, predictions in the $H \to b\bar{b}$ category are shown in the top row in red, while predictions for the $H \to gg$ category are shown in the bottom row in blue. Similarly, we show results using the flavour-k_{T} algorithm in the left-hand column and results using the flavour-dressing algorithm in the right-hand column. Each plot consists of two panels, with the upper panel showing the NLO distributions and the lower panels showing the differential K-factor NLO/LO for the two values of the jet-algorithm resolution parameter y_{cut}.

Scale variations are included by a lighter shaded band around the central predictions.

Generally, we observe larger rates in the $H \to b\bar{b}$ decay, owing to the fact that events in this category always contain at least one b-$\bar{b}$ pair, whereas most events in $H \to gg$ decays contain only unflavoured partons. We find larger NLO K-factors in the $H \to gg$ decay category, owing to the fact that our calculation employs the HEFT coupling between the Higgs and gluons, making it susceptible to large corrections. This is also reflected in the larger uncertainty band in predictions in the $H \to gg$ decay category. In both Higgs decay categories, the magnitude of the NLO corrections is comparable between the flavour-k_{T} and flavour-dressing algorithms, with somewhat larger corrections visible in the latter. It can be seen that predictions calculated with larger values of y_{cut} generally also receive larger NLO corrections. In all cases, lowering y_{cut} leads to wider distributions, owing to the larger kinematically allowed phase space.

Before moving on to discussing features specific to the individual observables, we wish to highlight a peculiar behaviour of the flavour-k_{T} algorithm that results in some interesting phenomenology. For all observables shown here, distributions in the flavour-k_{T} algorithm span a significantly larger value range than predictions employing the flavour-dressing algorithm. Although it might appear as if the flavour-k_{T} algorithm allows to probe phase-space regions with infrared sensitive configurations, this is not the case. The reason for the large allowed value range can be found in the definition of the modified distance measure of the flavour-k_{T} algorithm, eq. (2), which takes the maximum whenever the softer of the two partons is flavoured. In three-parton configurations $b\bar{b}g$ which contain a soft b-quark (or $\bar{b}$-quark), $E_b \ll E_g, E_{\bar{b}}$, cf. fig. 2, all distances involving this quark involve the maximum of the energies, $y_{bj} \propto \max(E_b, E_j)$. As no singularity is associated with a single (anti-)quark becoming soft, this definition allows for arbitrarily small (anti-)quark energies, in principle below the order of the cut-off y_{cut}, while still retaining $\min_{i,j}(y_{ij}) > y_{\text{cut}}$. In other words, the flavour-k_{T} algorithm counts three-particle configuration with a single arbitrarily soft quark as three-jet configurations, in contrast to the naïve expectation that a soft particle, regardless of its flavour, does not constitute a jet. We shall explore the consequence of this peculiar behaviour of the application of the flavour-k_{T} algorithm in analysing thorougly the shape and normalisation of the distributions presented below.

The energy of the leading flavoured jet, $E_{b,1}$:
Figure 3 shows the energy of the leading flavoured jet. In both decay categories and for both choices of the experimental resolution parameter y_{cut}, we observe similar results in the distributions using the flavour-k_{T} and flavour-dressing algorithm at lower energies. Towards the kinematical limit on the right-hand side of the plots, however, we find substantial differences between the two jet algorithms: The flavour-dressing algorithm assigns a vanishing probability to configurations with $E_{b,1} \approx m_H/2$, whereas such configurations have a non-zero probability in the flavour-k_{T} algorithm. This is related to the treatment of three-particle configurations containing a single soft quark in the flavour-k_{T} algorithm.

For the $H \to b\bar{b}$ category we find the following K-factors for the two y_{cut} values: With a y_{cut} value of 0.1 we find K-factors around $1.4 - 1.9$ for flavour-k_{T} and 1.3 to 1.9 for flavour-dressing.

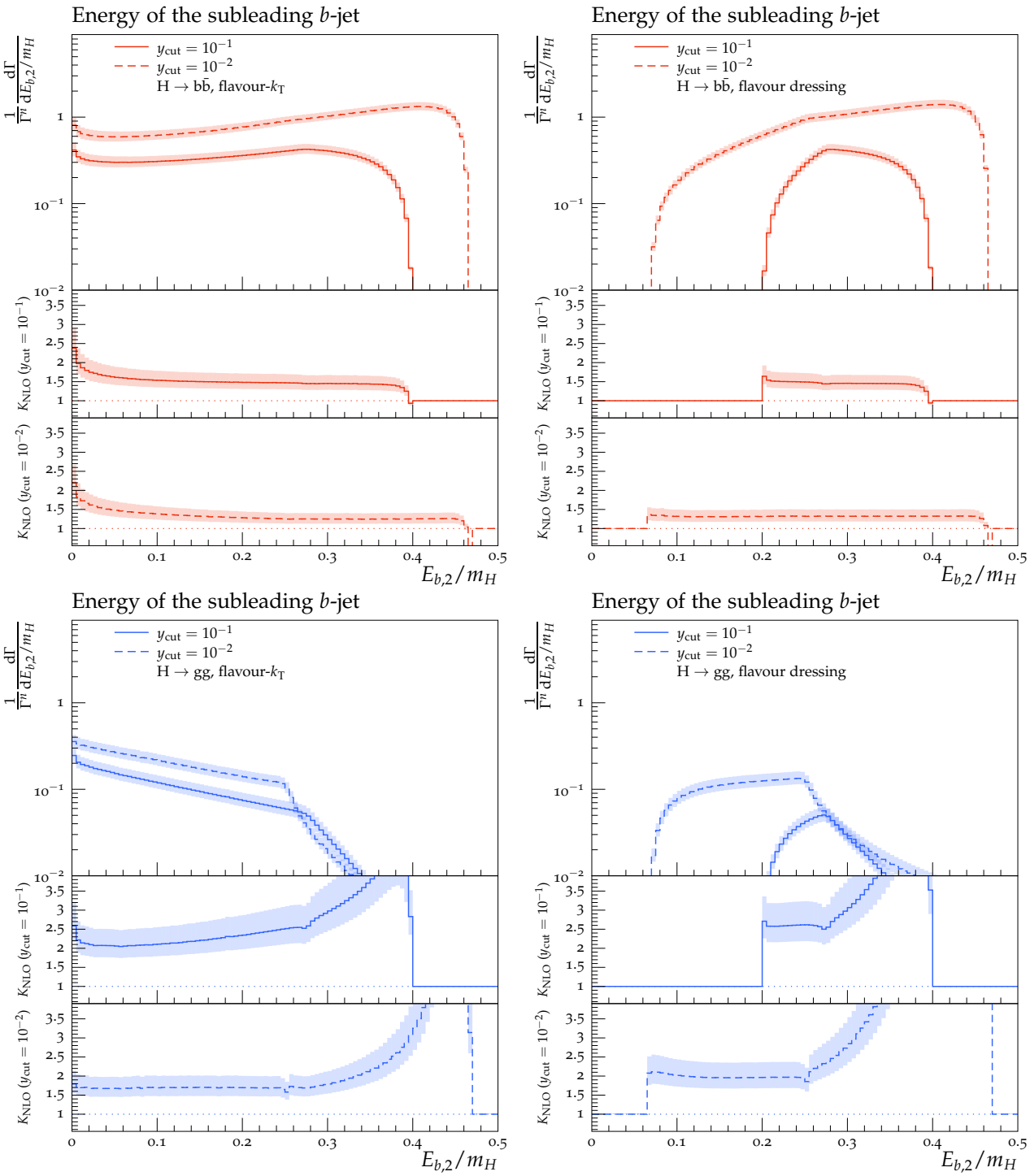


Fig. 4 Energy of the subleading b-jet in the $H \to b\bar{b}$ (*top row*) and $H \to gg$ (*bottom row*) decay category using the flavour-k_{T} (*left column*) and flavour dressing (*right column*) algorithm.

For $y_{\text{cut}} = 0.01$, we find K-factors ranging from 1.1 (1.2) to 1.3 (1.3) for flavour-k_{T} (flavour-dressing). In the $H \to gg$ category, we find higher NLO corrections, which are also reflected in the K-factors. For flavour-k_{T} and a y_{cut} of 0.1, we find K-factors from 1.5 to 2.5, where we start on the left side with a small increase from 2.4 up to 2.5 at an energy of $E_{b,1}/m_H = 0.34$, followed by a decrease to 1.5 for higher energies. The K-factors are more constant for a y_{cut} value of 0.01, varying around $1.4-1.7$. Flavour-dressing with a y_{cut} of 0.1 starts with K-factors of around 2.2 and then rapidly increases. A similar behaviour is observed for a y_{cut} value of 0.01, where we start at 1.7 and then diverge again. Flavour-k_{T} does not have these diverging NLO corrections, because of the previously discussed three-particle configurations containing a single soft quark. Analysing the situation regarding the size of the QCD corrections observed with both algorithms further, the following can be said: Already at LO, flavour-k_{T} allows one quark to have large energies (and thus the other quark has almost no energy), whereas flavour-dressing does not allow for such configurations. Only at NLO, flavoured jets are allowed to have large energies in flavour dressing, and thus the K-factors receive large contributions there.

The energy of the subleading flavoured jet, $E_{b,2}$: The distribution of the energy of the subleading flavoured jet, shown in fig. 4 can be explained by a similar reasoning as before for the leading flavoured jet distribution. Due to the peculiar behaviour for pathological $b\bar{b}g$ configurations, the flavour-k_{T} algorithm has a non-vanishing probability to find a flavoured jet with (almost) zero energy. As such however, the situation for the subleading flavoured jet is inverse to the situation of the leading flavoured jet. Indeed the corresponding subleading jet distribution now extends all the way to the maximum of $E_{b,1} \to m_H/2$, while $E_{b,2} \to 0$. For this subleading jet distribution, the gluonic decay category is subject to rather large NLO corrections at higher energies. Understanding that flavoured quarks in the $H \to gg$ decay are exclusively generated by secondary gluon decays, it becomes clear that the high-energy tail of the subleading-energy distribution receives large corrections from real corrections in which both gluons decay to a b-$\bar{b}$ pair. The K-factors for the fermionic decay category have a small increase for very low energies in the flavour-k_{T} algorithm and a decrease for high-energies for both algorithms. The ranges are from 1.3 (1.2) at to 2.0 (1.9) for flavour-k_{T} with $y_{\text{cut}} = 0.1$ ($y_{\text{cut}} = 0.01$). For flavour-dressing the ranges are from 1.3 (1.3) to 1.5 (1.4) with $y_{\text{cut}} = 0.1$ ($y_{\text{cut}} = 0.01$). The increase is for all algorithms and y_{cut} values towards lower energies. The gluonic channel has diverging K-factors for high energies, as alluded to above. Otherwise, the K-factors range for a y_{cut} of 0.1 from 2.1 to 3.0 for flavour-k_{T} and 2.6 to 3.0 for flavour-dressing, where for both the K-factor of 3.0 is obtained at an energy of $E_{b,2}/m_H = 0.3$. For a y_{cut} of 0.01 the ranges are from 1.7 to 1.8 for flavour-k_{T} and around $2.0 - 2.8$ for flavour dressing, where again the upper value is obtained at an energy of $E_{b,2}/m_H = 0.3$. Afterwards the K-factors diverge, as alluded to above.

The angular separation of the leading b- and $\bar{b}$-jet, $\cos\theta_{b\bar{b}}$: Figure 5 shows the angular separation of the leading b- and leading $\bar{b}$-jet. We see that the peak of the distributions is located at large angles ($\cos\phi \to -1$) in the $H \to b\bar{b}$ category, whereas it is located at small angles ($\cos\phi \to 1$) in $H \to gg$ decays. The reason is that in Yukawa-induced $H \to b\bar{b}$, the quark-antiquark pair directly stems from the Higgs decay and is thus expected to favour a large angular separation, whereas quark-antiquark pairs stem from gluon decays in the $H \to gg$ category and, as such, are divergent in the collinear limit. In $H \to gg$ decays, we see another peak towards $\cos\phi = -1$, which originates from decays $H \to gg \to b\bar{b}b\bar{b}$, in which the leading b and leading $\bar{b}$ originate from different gluons. Since this only happens at NLO, these obtain large NLO-corrections. The same is not visible in $H \to b\bar{b}$ decays, as the leading b-$\bar{b}$ pair almost exclusively stems from the primary Higgs-decay vertex. The peculiar behaviour of the flavour-k_{T} algorithm for the three-parton configurations with a single soft b-quark again allows the distributions to span further into the infrared region, as is visible by the numerically larger upper limit in the distributions obtained with flavour-k_{T}. Except for the peak towards $\cos\phi = -1$, the K-factors for the $H \to b\bar{b}$ channel are ranging from 1.3 (0.8) to 1.7 (1.3) for flavour-k_{T} and a y_{cut} of 0.1 (0.01), while the range for flavour-dressing with a y_{cut} of 0.1 is from 1.2 to 1.6, where the decrease is towards higher $\cos\phi$. The distribution is ended by a sudden increase in the NLO corrections. Flavour-dressing with a y_{cut} of 0.01 starts around a K-factor of 1.3 at low energies and then slowly decreases to a K-factor of 1.0 for higher energies. For $H \to gg$ we find the already discussed explosion of NLO correction towards $\cos\phi = -1$ and have otherwise K-factors in a range from 2.1 (1.5) to 2.2 (1.9) for a y_{cut} value of 0.1 (0.01), where for $y_{\text{cut}} = 0.01$ the minimal K-factor of 1.5 is obtained around $\cos\phi = 0$ and possesses a steady increase of the NLO corrections in both directions. Flavour-dressing again features huge NLO corrections towards $\cos\phi = -1$. The descent for a y_{cut} of 0.1 leads at $\cos\phi = -0.5$ to a K-factor of

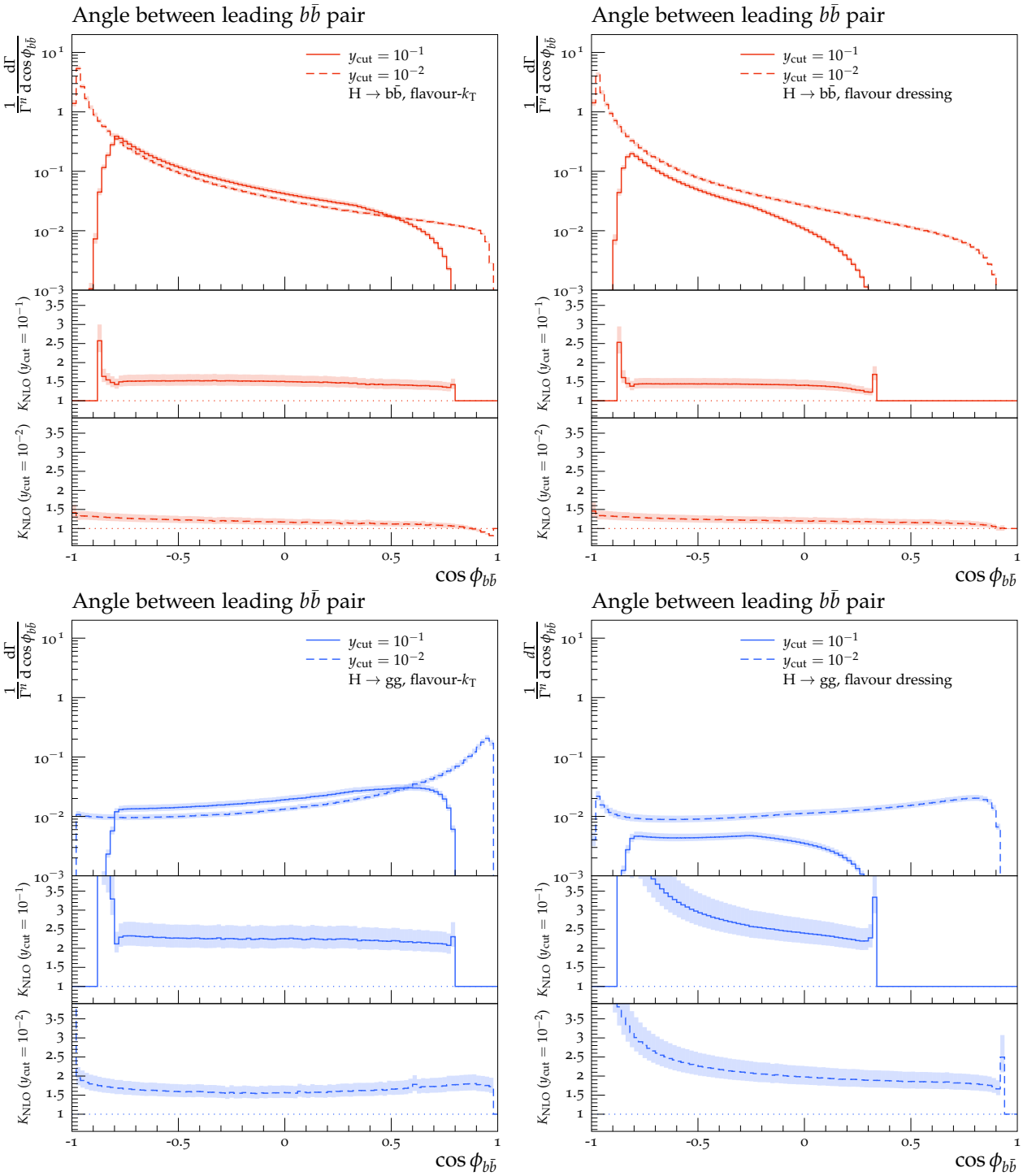


Fig. 5 Angular separation of the leading b- and $\bar{b}$-jet pair in the $H \to b\bar{b}$ (*top row*) and $H \to gg$ (*bottom row*) decay category using the flavour-k_T (*left column*) and flavour dressing (*right column*) algorithm.

3.0, and yields at $\cos\phi = 0$ a K-factor of 2.4. It then falls further to around 2.2 followed by a sharp increase again. For a y_{cut} of 0.01, the K-factors have a flatter distribution, where after the explosion on the left-hand side, it reaches at $\cos\phi = -0.5$ a K-factor of 2.3 and slowly decreases afterwards down to a K-factor of 1.7 followed again by a sharp spike.

The invariant mass of the leading b- and $\bar{b}$-jet, $m_{b\bar{b}}$:
Figure 6 shows the invariant mass of the leading b- and leading $\bar{b}$-jet, $m_{b\bar{b}}$. While Yukawa-induced decays favour quarks with a high invariant mass, $H \to gg$ decays favour quark-antiquark pairs with a small invariant mass. Again this can be understood by noting that the quarks in $H \to b\bar{b}$ decays originate directly from the primary Higgs-decay vertex, whereas they stem from secondary gluon splittings in $H \to gg$ decays and will thus have a smaller invariant mass $m_{b\bar{b}} \sim (1 - \cos\theta_{b\bar{b}})$, due to the form of the $g \to q\bar{q}$ splitting amplitude. The difference becomes more pronounced for lower values of y_{cut}, where the distribution shifts to larger angles in Yukawa-induced decays and towards smaller angles in gluonic decays. The main difference between the jet algorithms is that the flavour-k_{T} allows for smaller values of $m_{b\bar{b}}$, which again is explained by the maximum in the distance measure in equation (2). The K-factors for a y_{cut} of 0.1 in the fermionic decay category start by a peak at low invariant mass, where the peak for the flavour-dressing goes below a K-factor of 1. Both algorithms then do not have too much variation in the K-factors. For flavour-k_{T} the values range from 1.5, at higher invariant mass, to 1.6 at lower invariant mass. In flavour-dressing the increase is in the opposite direction with values ranging from 1.3, obtained around $m_{b\bar{b}} = 0.4$, to 1.5 for higher invariant mass. Both algorithms then end the distributions with another peak. Lowering the y_{cut} value to 0.01 again flattens the distribution a bit. Flavour-k_{T} starts for low invariant mass at a K-factor of 1.1 and then increases to a maximum of 1.4. Thereafter it slowly decreases to a K-factor of 1.3 for higher invariant mass. The flavour-dressing algorithm has a small K-factor of 0.2 for very low invariant mass, but then quickly increases again until reaching a K-factor of 1.2 around $m_{b\bar{b}}/m_H = 0.3$ followed by further increase to 1.3 for higher invariant mass.

The $H \to gg$ decay channel has big variations in the K-factors and need a careful analysis. Starting with flavour-k_{T} and $y_{\text{cut}} = 0.1$, we observe that the NLO corrections diverge towards negative infinity for small invariant mass, leaving thereby the region where the fixed order prediction can be trusted. Increasing the invariant mass also increases the NLO corrections, and the K-factor crosses a value of 1.0 around $m_{b\bar{b}}/m_H = 0.03$ and further rapidly increases to 1.8 around $m_{b\bar{b}}/m_H = 0.1$, 2.4 at $m_{b\bar{b}}/m_H = 0.3$, 2.7 at $m_{b\bar{b}}/m_H = 0.5$ and further to 4.0 at $m_{b\bar{b}}/m_H = 0.7$, where it then starts to diverge towards positive infinity.

The distribution is similar for flavour-k_{T} with a y_{cut} value of 0.01 with the main difference, that after reaching a peak in the K-factor of 2.3 around $m_{b\bar{b}}/m_H = 0.16$, it starts to decrease again to a K-factor of 1.7, before the NLO-corrections get larger again for higher invariant mass. For flavour-dressing the distributions are similar to flavour-k_{T}, where again for $y_{\text{cut}} = 0.1$, we start with very small NLO-corrections at low invariant mass followed by a quick increase and a divergence towards positive infinity. Similarly for $y_{\text{cut}} = 0.01$, the NLO-corrections are low at small invariant mass. After a rapid increase to a K-factor of 1.8 at $m_{b\bar{b}}/m_H = 0.2$, the NLO-corrections increases to 2.8 at $m_{b\bar{b}}/m_H = 0.6$ and then start to diverge again.

5 Summary and outlook

In this paper, for the first time, we have computed flavour-sensitive observables related to hadronic Higgs decays including both Higgs decay categories, i.e stemming from the underlying processes $H \to b\bar{b}$ and $H \to gg$ including higher order corrections up to $\mathcal{O}(\alpha_{\text{s}}^2)$ in perturbative QCD. The calculation was carried out in an effective theory in which the Higgs couples directly to gluons, while massless b-quarks retain a non-vanishing Yukawa coupling. Specifically, we considered the following observables : the energy of the leading and subleading flavoured jet, the angular separation and the invariant mass of the leading b-$\bar{b}$ pair at relative $\mathcal{O}(\alpha_{\text{s}}^2)$ in QCD. Using the antenna-subtraction framework, the computations were performed with the parton-level event generator EERAD3, extended to account for hadronic Higgs decays. A new flavour layer was implemented in EERAD3 to allow for the computation of observables with identified flavoured jets. Flavoured jets were defined using both the flavour-k_{T} and flavour-dressing algorithm. For both algorithms, infrared flavour safety was explicitly verified in section 4.2. For each observable, predictions obtained in both Higgs-decay categories, using both flavoured jet algorithms and for two values of the experimental resolution parameter y_{cut}, i.e., $y_{\text{cut}} = 0.1$ and $y_{\text{cut}} = 0.01$ were compared. In all cases, lowering the parameter y_{cut} leads to wider distributions, owing to the larger kinematically allowed phase space.

Comparing the two Higgs decay modes, we observe larger rates in the $H \to b\bar{b}$ decay category. This is related to the fact that events in this category always

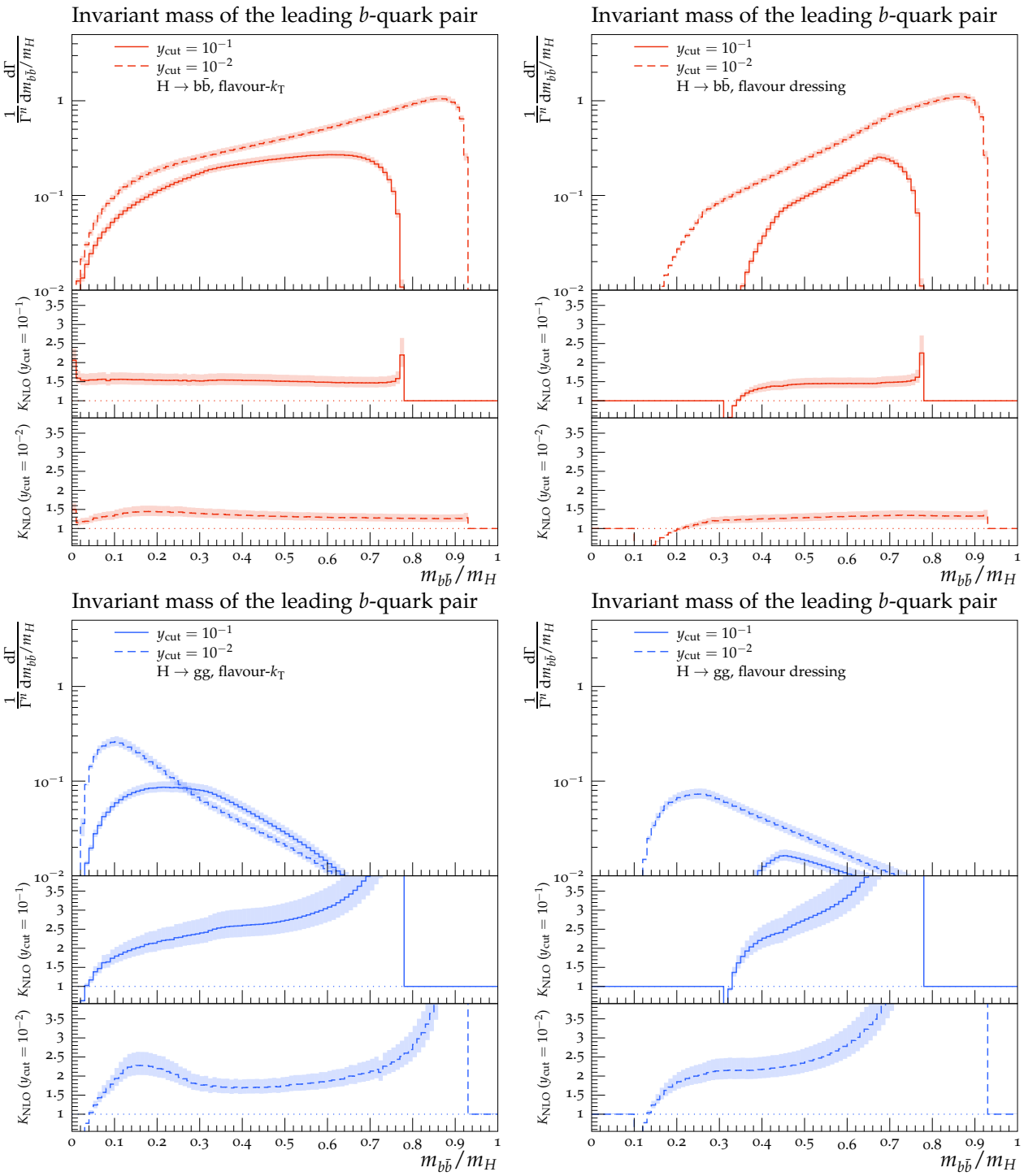


Fig. 6 Invariant mass of the leading b- and $\bar{b}$-jet pair in the $H \to b\bar{b}$ (*top row*) and $H \to gg$ (*bottom row*) decay category using the flavour-k_{T} (*left column*) and flavour dressing (*right column*) algorithm.

contain at least one b-$\bar{b}$ pair, whereas most events in $H \to gg$ decays contain only unflavoured partons. We find larger NLO K-factors in the $H \to gg$ decay category, owing to the fact that our calculation employs the HEFT coupling between the Higgs and the gluons. Comparing the results obtained using the two flavoured jet algorithms, it was highlighted that the use of the flavour-k_{T} algorithm introduces counter-intuitive phenomenological implications, owing to its treatment of pathological three-parton configurations containing a single soft flavoured quark. While states with a single soft quark do not correspond to infrared singularities in physical matrix elements, it has the effect that these configurations are identified as three-jet like despite containing an, in principle, arbitrarily soft quark.

Away from phase-space regions dominated by configurations with a single soft flavoured quark, we find qualitatively good agreement between the flavour-k_{T} and the flavour-dressing algorithm, with generally slightly larger NLO corrections in the latter. Using the flavour dressing algorithm in particular, the shape and normalisation of the individual distributions in both decay categories have characteristic fixed-order behaviour over the whole kinematical range. In particular the behaviour at the kinematical edges, i.e., the drop or peak in the cross section, can be understood considering only hard final states in both Higgs decay categories. It is worth mentioning though, that the drop or the peak of the distributions seen at the kinematical edges are systematically exchanged in one or the other Higgs decay categories. This analysis demonstrates in particular the practical applicability of the flavour-dressing algorithm to compute flavour-sensitive observables in hadronic Higgs decays including both decay modes.

Our study marks the first step towards a more complete treatment of flavour-induced effects in hadronic Higgs decays. Obtaining a solid theoretical understanding of these effects will be vital for Higgs precision studies at future lepton colliders. Among these efforts, two avenues for future work are particularly worth mentioning: the study of flavour-tagged event shapes and the analysis of the phenomenological impact gained by the inclusion of NNLO-type corrections to flavour-sensitive observables in hadronic Higgs decays. We anticipate to return to both avenues in the future.

Acknowledgements We would like to thank Giovanni Stagnitto for useful discussions and a careful reading of the manuscript. AG and CTP acknowledge support by the Swiss National Science Foundation (SNF) under contract 200021-197130 and by the Swiss National Supercomputing Centre (CSCS) under project ID ETH5f. BCA is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 - TRR257. Parts of the computations were carried out on the PLEIADES cluster at the University of Wuppertal, supported by the Deutsche Forschungsgemeinschaft (DFG, grant No. INST 218/78-1 FUGG) and the Bundesministerium für Bildung und Forschung (BMBF).

References

1. G. Aad, et al., Phys. Lett. B **716**, 1 (2012). DOI 10.1016/j.physletb.2012.08.020
2. S. Chatrchyan, et al., Phys. Lett. B **716**, 30 (2012). DOI 10.1016/j.physletb.2012.08.021
3. A. Abada, et al., Eur. Phys. J. ST **228**(2), 261 (2019). DOI 10.1140/epjst/e2019-900045-4
4. A. Arbey, et al., Eur. Phys. J. C **75**(8), 371 (2015). DOI 10.1140/epjc/s10052-015-3511-9
5. M. Aaboud, et al., Phys. Lett. B **786**, 59 (2018). DOI 10.1016/j.physletb.2018.09.013
6. A.M. Sirunyan, et al., Phys. Rev. Lett. **121**(12), 121801 (2018). DOI 10.1103/PhysRevLett.121.121801
7. G. Coloretti, A. Gehrmann-De Ridder, C.T. Preuss, JHEP **06**, 009 (2022). DOI 10.1007/JHEP06(2022)009
8. J. Gao, JHEP **01**, 038 (2018). DOI 10.1007/JHEP01(2018)038
9. J. Gao, Y. Gong, W.L. Ju, L.L. Yang, JHEP **03**, 030 (2019). DOI 10.1007/JHEP03(2019)030
10. M.X. Luo, V. Shtabovenko, T.Z. Yang, H.X. Zhu, JHEP **06**, 037 (2019). DOI 10.1007/JHEP06(2019)037
11. J. Gao, V. Shtabovenko, T.Z. Yang, JHEP **02**, 210 (2021). DOI 10.1007/JHEP02(2021)210
12. A. Gehrmann-De Ridder, C.T. Preuss, C. Williams, (2023)
13. M. Knobbe, F. Krauss, D. Reichelt, S. Schumann, Eur. Phys. J. C **84**(1), 83 (2024). DOI 10.1140/epjc/s10052-024-12430-4
14. R. Gauld, A. Gehrmann-De Ridder, E.W.N. Glover, A. Huss, I. Majer, JHEP **10**, 002 (2019). DOI 10.1007/JHEP10(2019)002
15. W. Bizoń, E. Re, G. Zanderighi, JHEP **06**, 006 (2020). DOI 10.1007/JHEP06(2020)006
16. S. Zanoli, M. Chiesa, E. Re, M. Wiesemann, G. Zanderighi, JHEP **07**, 008 (2022). DOI 10.1007/JHEP07(2022)008
17. A. Behring, W. Bizoń, F. Caola, K. Melnikov, R. Röntsch, Phys. Rev. D **101**(11), 114012 (2020). DOI 10.1103/PhysRevD.101.114012
18. R. Gauld, A. Gehrmann-De Ridder, E.W.N. Glover, A. Huss, I. Majer, Phys. Rev. Lett. **125**(22), 222002 (2020). DOI 10.1103/PhysRevLett.125.222002
19. M. Czakon, A. Mitov, M. Pellen, R. Poncelet, JHEP **06**, 100 (2021). DOI 10.1007/JHEP06(2021)100
20. M. Czakon, A. Mitov, M. Pellen, R. Poncelet, JHEP **02**, 241 (2023). DOI 10.1007/JHEP02(2023)241
21. R. Gauld, A. Gehrmann-De Ridder, E.W.N. Glover, A. Huss, A.R. Garcia, G. Stagnitto, Eur. Phys. J. C **83**(4), 336 (2023). DOI 10.1140/epjc/s10052-023-11530-x
22. A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, A. Huss, A.R. Garcia, G. Stagnitto, (2023)
23. O. Fedkevych, C.K. Khosa, S. Marzani, F. Sforza, Phys. Rev. D **107**(3), 034032 (2023). DOI 10.1103/PhysRevD.107.034032
24. R. Gauld, A. Huss, G. Stagnitto, Phys. Rev. Lett. **130**(16), 161901 (2023). DOI 10.1103/PhysRevLett.130.161901

25. F. Caola, R. Grabarczyk, M.L. Hutt, G.P. Salam, L. Scyboz, J. Thaler, Phys. Rev. D **108**(9), 094010 (2023). DOI 10.1103/PhysRevD.108.094010
26. L. Cavallini, A. Coccaro, C.K. Khosa, G. Manco, S. Marzani, F. Parodi, D. Rebuzzi, A. Rescia, G. Stagnitto, Eur. Phys. J. C **82**(5), 493 (2022). DOI 10.1140/epjc/s10052-022-10447-1
27. A. Banfi, G.P. Salam, G. Zanderighi, Eur. Phys. J. C **47**, 113 (2006). DOI 10.1140/epjc/s2006-02552-4
28. S. Caletti, A.J. Larkoski, S. Marzani, D. Reichelt, Eur. Phys. J. C **82**(7), 632 (2022). DOI 10.1140/epjc/s10052-022-10568-7
29. M. Czakon, A. Mitov, R. Poncelet, JHEP **04**, 138 (2023). DOI 10.1007/JHEP04(2023)138
30. S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, Phys. Lett. B **269**, 432 (1991). DOI 10.1016/0370-2693(91)90196-W
31. A.J. Larkoski, S. Marzani, G. Soyez, J. Thaler, JHEP **05**, 146 (2014). DOI 10.1007/JHEP05(2014)146
32. A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, G. Heinrich, JHEP **12**, 094 (2007). DOI 10.1088/1126-6708/2007/12/094
33. A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, G. Heinrich, Comput. Phys. Commun. **185**, 3331 (2014). DOI 10.1016/j.cpc.2014.07.024
34. T. Inami, T. Kubota, Y. Okada, Z. Phys. C **18**, 69 (1983). DOI 10.1007/BF01571710
35. A. Djouadi, J. Kalinowski, P.M. Zerwas, Z. Phys. C **54**, 255 (1992). DOI 10.1007/BF01566654
36. K.G. Chetyrkin, B.A. Kniehl, M. Steinhauser, Phys. Rev. Lett. **79**, 353 (1997). DOI 10.1103/PhysRevLett.79.353
37. K.G. Chetyrkin, B.A. Kniehl, M. Steinhauser, Nucl. Phys. B **510**, 61 (1998). DOI 10.1016/S0550-3213(97)00649-4
38. K.G. Chetyrkin, J.H. Kuhn, C. Sturm, Nucl. Phys. B **744**, 121 (2006). DOI 10.1016/j.nuclphysb.2006.03.020
39. Y. Schroder, M. Steinhauser, JHEP **01**, 051 (2006). DOI 10.1088/1126-6708/2006/01/051
40. P.A. Baikov, K.G. Chetyrkin, J.H. Kühn, Phys. Rev. Lett. **118**(8), 082002 (2017). DOI 10.1103/PhysRevLett.118.082002
41. J.A.M. Vermaseren, S.A. Larin, T. van Ritbergen, Phys. Lett. B **405**, 327 (1997). DOI 10.1016/S0370-2693(97)00660-6
42. A.M. Sirunyan, et al., JINST **13**(05), P05011 (2018). DOI 10.1088/1748-0221/13/05/P05011
43. G. Aad, et al., JINST **11**(04), P04008 (2016). DOI 10.1088/1748-0221/11/04/P04008
44. R. Aaij, et al., JINST **10**(06), P06013 (2015). DOI 10.1088/1748-0221/10/06/P06013

Impact of strong magnetic field, baryon chemical potential, and medium anisotropy on polarization and spin alignment of hadrons

Bhagyarathi Sahoo,* Captain R. Singh,† and Raghunath Sahoo‡
Department of Physics, Indian Institute of Technology Indore, Simrol, Indore 453552, India
(Dated: February 28, 2024)

The recent observation of global polarization of Λ ($\bar{\Lambda}$) hyperons and spin alignment of ϕ and K^{*0} vector mesons create remarkable interest in investigating the particle polarization in the relativistic fluid produced in heavy-ion collisions at GeV/TeV energies. Among other sources of polarization, the Debye mass of a medium plays a crucial role in particle polarization. Any modification brought to the effective mass due to the temperature, strong magnetic field (eB), baryonic chemical potential (μ_B), and medium anisotropy (ξ), vorticity, etc., certainly affects the particle polarization. In this work, we explore the global hyperon polarization and the spin alignment of vector mesons corresponding to the strong magnetic field, baryonic chemical potential, and medium anisotropy. We find that the degree of polarization is flavor-dependent for hyperons. Meanwhile, vector meson spin alignment depends on the hadronization mechanisms of initially polarized quarks and anti-quarks. Medium anisotropy significantly changes the degree of polarization in comparison with the magnetic field and baryon chemical potential.

arXiv:2402.17344v1 [hep-ph] 27 Feb 2024

I. INTRODUCTION

So far, the thermalized state of strongly interacting partons, called quark-gluon plasma (QGP), is probed in heavy-ion collisions through a baseline, the *pp* collisions. It was assumed that QGP existence in *pp* collisions is next to impossible because it lacks the necessary conditions for QGP to be formed. On the contrary, in recent LHC events, ultra- relativistic *pp* collision experiments have reported behavior similar to heavy-ion collisions, e.g., collective flow, strangeness enhancement, etc [1, 2]. However, other studies suggest similar phenomena may arise due to the other QCD processes [3**?** , 4]. These studies raise a question concerning the present QGP signatures and a need for the next generation of probes. In the quest for such a baseline-independent probe, polarization comes into picture and can have implications for understanding hot QCD matter. The particle production mechanisms lead to a finite polarization of the light/heavy baryons and vector mesons [5]. However, there are various sources by which hadrons can get polarized in ultra-relativistic collisions. One such primary source could be the initial state polarization, which arises due to the motion and spin of the constituent quarks of the colliding nucleons. This initial state polarization can be transmitted to the quarks participating in the collision process. The magnetic field produced by the charged spectator protons in such collisions interacts with the electric charge and spin of the quarks, which may lead to quark polarization. A topological charge imbalance in the presence of an external magnetic field leads to the charge separation in the direction of the magnetic field. This phenomenon is known as the chiral magnetic effect (CME) [6, 7]. If QGP exhibits chiral symmetry restoration, then CME could lead to quark polarization along the direction of the magnetic field. A similar phenomenon called chiral vortical effect (CVE) is expected due to the non-zero local vorticity, which also contributes to the hadron polarization [8, 9]. The hydrodynamic behavior or collective motion of QGP can also induce polarization in the quark distribution due to the anisotropic expansion [10, 11]. Determining the precise sources and types of polarization in ultra-relativistic collisions requires sophisticated theoretical studies with corresponding experimental observations.

In 2005, Liang and Wang predicted in non-central heavy-ion collisions, the orbital angular momentum (OAM) of the partonic system polarizes the quarks and anti-quarks through spin-orbit coupling. They asserted that the initial partons created in the collisions could generate a longitudinal fluid shear distribution representing the local relative OAM in the same direction as global OAM at a finite impact parameter. This quark polarization manifests the polarization of the hadrons (with finite spin) along the direction of OAM during the process of hadronization [12–15]. Apart from global polarization of hadrons, such global quark polarization has many observable consequences, such as left-right asymmetry in hadron spectra and global transverse polarization of thermal photons, dileptons, etc. [16, 17]. They have studied the global polarization of hyperons [12] and spin alignment of vector mesons [13] in different hadronization scenarios. Following, in 2013, Becattini et al. predicted the global spin polarization of Λ hyperons due to the OAM-manifested thermal vorticity [18]. Various theoretical predictions of global Λ hyperons polarization by different hydrodynamic and transport models are well agreed with the experimental results available at Relativistic Heavy Ion Collider (RHIC) [19–27]. These hydrodynamic and transport

* Bhagyarathi.Sahoo@cern.ch
† captainriturajsingh@gmail.com
‡ Raghunath.Sahoo@cern.ch (Corresponding Author)

models considered the thermal equilibrated distribution of spin degrees of freedom for polarization studies [28–35].

In addition to Λ hyperons, the averaged global polarization measurement of Ω and Ξ hyperons are obtained using the polarization transfer method in STAR [36]. Similarly, the polarization of quarkonium states is obtained by measuring the angular distribution of dileptons [5, 37–40]. Furthermore, the global spin alignment of ϕ, K^{*0}, K^0_S mesons is observed at the LHC and RHIC from the angular distribution of decay daughters in the mother particles's rest frame [41–46]. Recent spin alignment measurements for ϕ vector mesons by the STAR collaboration initiate a spark in the theoretical communities to investigate the source of large positive spin alignment for ϕ vector meson, while K^{*0} shows no such spin-alignment within uncertainities [41].

Some theoretical studies claim that in addition to vorticity and electromagnetic fields, a strong vector meson force field is responsible for hyperon polarization [47] as well as spin alignment of vector mesons [48, 49]. Recently, Sheng et al. [49] explain the large spin alignment of ϕ vector meson in a quark polarization mechanism by a strong force field, in which quarks interact with the dense medium through a strong force and become polarized. The strong force field is proposed to polarize the vector meson with the hidden flavor quantum number. Apart from these sources, there could be various other possible sources for vector meson spin-alignment, which need further investigation. Moreover, recent observation of longitudinal local polarization of Λ hyperons at LHC and RHIC mismatches with the theoretical predictions [11], which opens a new door for the heavy-ion community to have intense investigations on this topic. However, various studies have been developed to solve this puzzle through different aspects [50, 51]. Similar to hyperons, the local polarization of quarks and anti-quarks due to anisotropic expansion of the medium contributes to the spin alignment of vector meson [52]. The authors have argued that the deviation of ρ_{00} from 1/3 is not solely coming from global spin alignment but also has the contribution of quarks and/or anti-quarks, which are locally polarized. However, the contribution of local quark polarization requires the measurement of off-diagonal elements of the spin density matrix.

In the experiment, the methods used to measure polarization for hyperons and vector mesons are different. However, both depend on the angular distribution of decay daughters in the mother particles' rest frame. The hyperons undergo parity-violating weak decays, while the vector mesons mainly decay through the parity, conserving strong decay. The Λ hyperon polarization is estimated by the projection of the daughter proton's momentum along the OAM (or vorticity axis) direction. On the other hand, the spin alignment of vector meson has been studied by measuring the ρ_{00} element of the spin density matrix. The detailed explanation of the spin density matrix is discussed in section II.

Ultra-peripheral heavy-ion collisions have several consequences; one such example is that they produce a strong magnetic field. This magnetic field decays with time and can, in principle, affect the QGP medium evolution, thermodynamic, and transport properties [53–55]. The magnetic field is believed to affect the Λ polarization, specifically on the polarization splitting of Λ and $\bar{\Lambda}$ at the lower center of mass energies. The effect of magnetic field on polarization and spin alignment of hadrons is discussed in Ref. [56–64]. Recent observation of global polarization of Λ hyperons indicate that hyperon polarization decreases with the increasing center of mass energy [19–21]. So, at lower energies, the probability of particles getting polarized is large, which could be a consequence of baryon stopping. This high baryon chemical potential may affect the evolution process and particle polarization. Further, due to the rapid longitudinal expansion of the fireball created in heavy-ion collisions, a high degree of momentum-space anisotropy is produced in the QGP medium in its local rest frame [65]. This momentum-space anisotropy has many novel consequences; for example, it is essential in modeling heavy-quark dynamics, making the inter-quark potentials spatially anisotropic. This anisotropy in the QGP medium may affect the particle yield, flow coefficients, particle polarization, etc.

In this work, we investigate the effect of static strong magnetic field, baryon chemical potential, and momentum-space anisotropy parameter on the polarization of hyperons and spin alignment of vector mesons through Debye screening mass. Please note that in the rest of the paper, we shall refer to Debye screening mass as Debye mass. The individual and combined effect of eB, μ_B and ξ on the polarization and spin-alignment of hadrons is studied. It is important to mention that the magnetic field may generate an additional anisotropy in the medium, which is not explicitly considered in this study. We consider two different hadronization mechanisms, i.e., the recombination and fragmentation processes, for both hyperons and vector mesons.

The structure of this work is as follows. The section II gives a detailed overview of the formalism used in this work. We briefly examine the results in section III and provide a summary in section IV

II. FORMULATION

The global quark polarization is obtained through parton elastic scattering within the framework of an effective static potential model. This model uses effective mass or Debye mass to incorporate the parton scattering, which induces the quark polarization [12–15], given as,

$$P_q = -\frac{\pi}{2}\frac{m_D p}{E(E+m_q)} \tag{1}$$

where $E = \sqrt{p^2+m^2}$, p and m_q are the energy, momentum, and mass of quark in the center of mass frame of the parton scattering, m_D is the Debye mass.

Assuming $P_u = P_d \equiv P_q$, recent studies [56] calculated the polarization of hyperons using a spin density matrix. The following study is consistent with the previous findings [12], where the authors have calculated the polarization of hyperons using recombination and fragmentation processes.

This global quark polarization leads to the global hyperons polarization at the hadronization stage. The global polarization due to exclusive recombination process for Λ, $\Sigma^{\pm}$, Ξ^{-}, Ω^{-}, and Δ^{++} hyperons are given as [12, 56];

$$P_{\Lambda}^{rec} = P_s, \qquad P_{\Sigma^{\pm}}^{rec} = \frac{4P_q - P_s - 3P_sP_q^2}{3-4P_qP_s+P_q^2} \tag{2}$$

$$P_{\Xi^{-}}^{rec} = \frac{4P_s - P_q - 3P_qP_s^2}{3-4P_qP_s+P_s^2}, \qquad P_{\Omega^{-}}^{rec} = \frac{P_s(5+P_s^2)}{1+P_s^2} \tag{3}$$

$$P_{\Delta^{++}}^{rec} = \frac{P_u(5+P_u^2)}{1+P_u^2} \tag{4}$$

The hyperon polarization from inclusive recombination is difficult to estimate. However, at the extreme limit of inclusive recombination (i.e., fragmentation), the hyperon polarization due to polarized quarks is estimated in Ref. [12]. The fragmentation of hyperons is given by;

$$P_{\Lambda}^{frag} = \frac{n_sP_s}{n_s+2f_s}, \qquad P_{\Sigma^{\pm}}^{frag} = \frac{4f_sP_q - n_sP_s}{3(n_s+2f_s)} \tag{5}$$

$$P_{\Xi^{-}}^{frag} = \frac{4n_sP_s - f_sP_q}{3(2n_s+f_s)}, \; P_{\Omega^{-}}^{frag} = \frac{P_s}{3}, \; P_{\Delta^{++}}^{frag} = \frac{P_u}{3} \tag{6}$$

Such global polarization of quarks and anti-quarks also describes the spin alignment of vector mesons through different hadronization scenarios [13].

Unlike spin-polarization for hyperons, the spin alignment of vector meson is described by the spin density matrix $\rho_{m,m'}$, where m and $m^{'}$ label the spin component along the quantization axis. The spin alignment for vector meson is mostly measured by the 00^{th} element of the spin density matrix, i.e., ρ_{00}. It can be measured experimentally through the angular distribution of the decay products of the particles rest frame [43]. The deviation of ρ_{00} from 1/3 (i.e $\rho_{00} - 1/3$) determines the degree of alignment of vector meson. In Ref. [13], the authors have described the spin alignment of vector mesons in different hadronization scenarios.

In this work, we consider two scenarios of the hadronization process; the first one is due to the recombination of polarized quarks and anti-quarks. The second one is due to the recombination of polarized quarks and anti-quarks, in which either of them is created in the fragmentation process. The 00^{th} element of spin density matrix for ρ, ϕ and K^* mesons in recombination and fragmentation processes in which both quarks and anti-quarks are polarized are obtained as,

$$\rho_{00}^{\rho(rec)} = \frac{1-P_q^2}{3+P_q^2}, \; \rho_{00}^{\phi(rec)} = \frac{1-P_s^2}{3+P_s^2}, \; \rho_{00}^{K^{*(rec)}} = \frac{1-P_qP_s}{3+P_qP_s} \tag{7}$$

$$\rho_{00}^{\rho(frag)} = \frac{1+\beta P_q^2}{3-\beta P_q^2}, \;\; \rho_{00}^{\phi(frag)} = \frac{1+\beta P_s^2}{3-\beta P_s^2} \tag{8}$$

$$\rho_{00}^{K^{*(frag)}} = \frac{f_s}{n_s+f_s}\frac{1+\beta P_q^2}{3-\beta P_q^2} + \frac{n_s}{n_s+f_s}\frac{1+\beta P_s^2}{3-\beta P_s^2} \tag{9}$$

where $\beta \simeq 0.5$. n_s and f_s are the strange quark abundances relative to up and down quarks in QGP and quark fragmentation, respectively.

Further, we have briefly outlined the formalism to explore the effect of magnetic field, baryon chemical potential, and anisotropic parameters on the polarization of hyperons and mesons.

A. Debye Mass under Strong Magnetic Field

The magnetic field produced in the peripheral heavy-ion collisions is believed to decay very quickly. So, it is expected to only affect the partonic phase during the early stage of the evolution process. Magnetic fields tend to align charged particles in a specific direction, resulting in the polarization of the system. Given that quarks possess a fractional electric charge, it is anticipated that a magnetic field could potentially modify the orientation of quarks within the QGP phase. Consequently, in the presence of a strong magnetic field, the total Debye mass for fermions or quarks in the static limit has been calculated through the temporal component of one loop vacuum self-energy diagram and has the following form [66–69];

$$m_D^2(T,eB) = 4\pi\alpha_s(\Lambda^2,|eB|)\left[T^2\frac{N_c}{3} + \sum_f \frac{|q_f eB|}{4\pi^2}\right] \tag{10}$$

here, $\alpha_s(\Lambda^2,|eB|)$ is the running coupling constant, which depends on the magnetic field and temperature. It can be written as [68],

$$\alpha_s(\Lambda^2, |eB|) = \frac{g_s^2}{4\pi} = \frac{\alpha_s(\Lambda^2)}{1 + b_1\alpha_s(\Lambda^2)\ln\left(\frac{\Lambda^2}{\Lambda^2+|eB|}\right)} \quad (11)$$

with

$$\alpha_s(\Lambda^2) = \frac{1}{b_1 \ln\left(\frac{\Lambda^2}{\Lambda^2_{MS}}\right)} \quad (12)$$

where $b_1 = \frac{11N_c-2N_f}{12\pi}$, $\Lambda_{MS} = 0.176$ GeV for $N_f = 3$ and renormalization scale $\Lambda = 2\pi T$ [70].

B. Debye Mass with Baryon Chemical Potential

Finite μ_B dictates the dominance of the number of quarks over anti-quarks, such an imbalance affects the final production yield of hadrons. In nature, any order parameter or disparity ultimately leads to the polarization of the system. Likewise, a non-zero baryon chemical potential can be the source of a polarized QGP medium. Any change brought to the QGP affects its effective mass, as a consequence, μ_B is incorporated in the Debye mass. In a hot and high baryon density medium (T, $\mu_B >> m_q$, with m_q as the light quark mass), the leading order HTL/HDL calculation predicts resultant Debye mass which has the following form [71–74];

$$m_D^2(T, \mu_B) = 4\pi\alpha_s(\Lambda^2)\left[T^2\left(\frac{N_c}{3} + \frac{N_f}{6}\right) + \sum_f \frac{\mu_B^2}{18\pi^2}\right] \quad (13)$$

Here, the strong coupling constant $\alpha_s(\Lambda^2)$ takes the form given in Eq. 12, and in the presence of baryon chemical potential, the renormalization scale is modified into

$$\Lambda = 2\pi\sqrt{T^2 + \frac{\mu_B^2}{\pi^2}} \quad (14)$$

C. Effect of Anisotropy on Debye Mass

Momentum space anisotropy originated due to the collision geometry causing differential expansions of the fireball along the longitudinal and transverse directions. As a consequence, it also alters the effective mass of the medium. The effect of momentum-space anisotropy of the medium on Debye mass is given as [65, 75–79];

$$m_D(T, \xi) = m_D(T)\Big[f_1(\xi)\cos^2\theta + f_2(\xi)\Big]^{-\frac{1}{4}} \quad (15)$$

and,

$$m_D(T) = \sqrt{4\pi\alpha_s(\Lambda^2)T^2\left(\frac{N_c}{3} + \frac{N_f}{6}\right)} \quad (16)$$

The parameter ξ measures the degree of momentum-space anisotropy, defined as;

$$\xi = \frac{1}{2}\frac{<\mathbf{k}_\perp^2>}{<k_z^2>} - 1 \quad (17)$$

where $k_z \equiv \mathbf{k}\cdot\mathbf{n}, \mathbf{k}_\perp \equiv \mathbf{k} - \mathbf{n}(\mathbf{k}\cdot\mathbf{n})$ correspond to the particle momenta along and perpendicular to the direction of anisotropy, and θ is the angle between quark pair alignment ($\mathbf{r}$) with respect to direction of anisotropy ($\mathbf{n}$). The $f_1(\xi)$ and $f_2(\xi)$ are parameterized in such a way that Eq. 15 remains always true under small and large values of ξ,

$$f_1(\xi) = \frac{9\xi(1+\xi)^{\frac{3}{2}}}{2\sqrt{3+\xi}(3+\xi^2)}\frac{\pi^2(\sqrt{2} - (1+\xi)^{\frac{1}{8}}) + 16(\sqrt{3+\xi} - \sqrt{2}))}{(\sqrt{6}-\sqrt{3})\pi^2 - 16(\sqrt{6}-3)} \quad (18)$$

‘

$$f_2(\xi) = \xi\left(\frac{16}{\pi^2} - \frac{\sqrt{2}(\frac{16}{\pi^2} - 1) + (1+\xi)^{\frac{1}{8}}}{\sqrt{3+\xi}}\right)\left(1 - \frac{(1+\xi)^{\frac{3}{2}}}{1+\frac{\xi^2}{3}}\right) + f_1(\xi) + 1 \quad (19)$$

D. Net Debye Mass

In previous sections, we have discussed the Debye mass for individual cases now combining all together (eB, μ_B, and ξ), we have obtained the resultant Debye mass;

$$m_D(T, \mu_B, B, \xi) = m_D(T, \mu_B, B)\Big[f_1(\xi)\cos^2\theta + f_2(\xi)\Big]^{-\frac{1}{4}} \quad (20)$$

and

$$m_D(T,\mu_B,B)=\sqrt{4\pi\alpha_s(\Lambda^2,|eB|)\left[T^2\frac{N_c}{3}+\sum_f\frac{|q_f eB|}{4\pi^2T}\int_0^\infty dp_z\{n^+(E)(1-n^+(E))+n^-(E)(1-n^-(E))\}\right]} \quad (21)$$

with

$$n^{\pm}(E)=\left[\exp\frac{(E\mp\mu_B)}{T}+1\right]^{-1} \quad (22)$$

The $n^+(E)$ and $n^-(E)$ are the phase-space distribution functions of quarks and anti-quarks, respectively.

III. RESULTS AND DISCUSSION

Following the discussed formulation, we studied the influence of magnetic field (eB), baryon chemical potential (μ_B), and anisotropic parameter (ξ) on the polarization of hyperons and spin alignment of vector mesons. As quark-antiquark polarization leads to global hadron polarization, based on that, we have predicted the global polarization of various hadrons. Further, The spin alignment of vector mesons and polarization of hyperons are obtained corresponding to two different hadronization processes: (i) recombination and (ii) fragmentation. Firstly, the hyperon polarization prediction is given and discussed. Subsequently, the spin-alignment of vector mesons is presented. But before that, the change in the Debye mass due to magnetic field, baryon chemical potential, and the anisotropic parameter is depicted in Fig 1 to ensure the validity of our results. It shows the variation of effective Debye mass with the magnetic field (left panel), baryon chemical potential (middle panel), and anisotropic parameter (right panel) for three different values of temperature, i.e., T = 200 MeV, T = 250 MeV, and T = 300 MeV.

The left panel of Fig. 1 shows that the Debye mass increases with magnetic field and temperature. However, the impact of the external magnetic field on the rate of change in Debye mass is stronger at lower temperatures. The observed enhancement in the Debye mass with the external magnetic field is in qualitative agreement with the findings reported in Refs. [80–82]. The middle panel of Fig. 1 reveals a modest decrease in the Debye mass with increasing baryon chemical potential in a baryon-rich medium. Although its decrease is very small in magnitude, the effect of baryon chemical potential is a bit prominent at lower temperatures compared to higher temperatures. In the right panel of Fig. 1, we found that the Debye mass decreases significantly as the anisotropy parameter increases. The decrease in the Debye mass with the momentum anisotropy agrees well with the finding reported in Refs. [65, 75].

Based on these observations, it can be said that the medium effects are encoded in the Debye mass. It is more sensitive to the anisotropy parameter than the magnetic field and baryon chemical potential. The Debye mass of the medium represents the change in the degree of freedom of the systems, if the Debye mass of the system is very small, then the color degrees of freedom would reduce to the hadron degree of freedom. The Debye mass depends on two things: the medium's temperature (energy) and the coupling strength of the medium. The magnetic field in the medium gives an additional boost to the energy level and strengthens the coupling constant. Due to this, a positive change in Debye mass with increasing field can be seen in the left panel of Fig. 1. On the other hand, finite baryon chemical potential also provides additional energy to the medium but reduces the coupling strengths. Consequently, in Fig. 1, we observe the negative change in the Debye mass with increasing baryon chemical potential. Meanwhile, the anisotropy of the medium does not affect the coupling strength of the medium. However, an uneven distribution of the partons in the medium measured in terms of ξ reduces the net effective mass of the medium. Therefore, a visible reduction in the Debye mass with large values of ξ is observed in Fig. 1. The combined effect of the magnetic field, baryon chemical potential, and medium anisotropy on Debye mass can be estimated using Eq. 21, which is not explicitly illustrated in this work. But one can easily infer the combined effect of these parameters on Debye mass from Fig. 1.

Furthermore, it is noteworthy to mention that in the quark recombination model, the hyperon polarization and spin alignment of vector meson from polarized quarks and anti-quarks does not include momentum dependence explicitly. This momentum dependence is introduced in a non-relativistic quark-coalescence model through the spin-density matrix [56], with which the polarization of vector meson and baryon of spin-1/2 and spin-3/2 is calculated systematically. However, the conventional quark coalescence or recombination models do not include the spin degrees of freedom. Again, the quark coalescence model is improved in Ref. [57].

A. Polarization of hyperons

To understand the polarization dynamics of hyperons, it is crucial to understand how these polarized quarks recombine to produce the final polarized hyperons. The polarization of hyperons depends on the process of hadronization. Here, we estimate the individual and combined effect of magnetic field, baryon chemical potential, and medium anisotropy on the global polar-

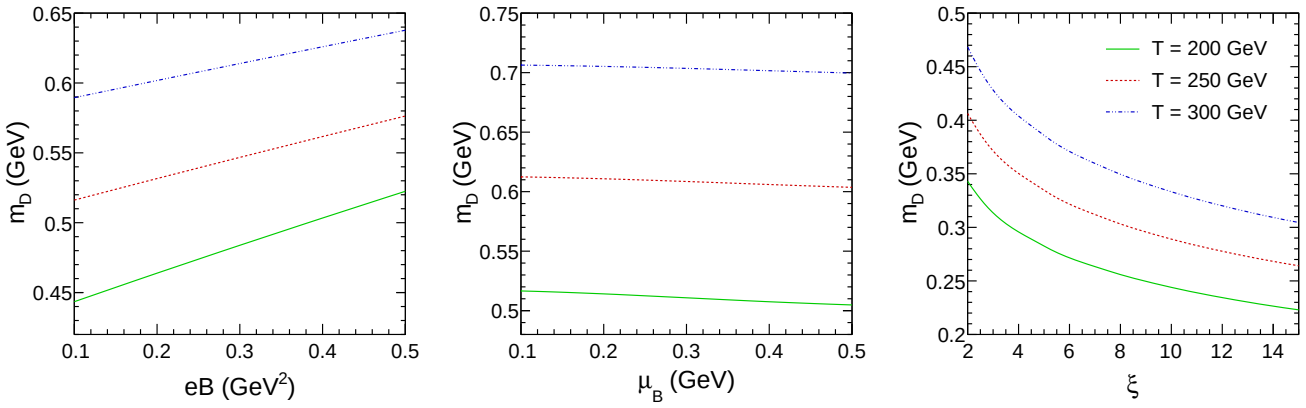


FIG. 1: (Color online) The variation of Debye screening mass (m_D) as a function of magnetic field (eB)(left panel), baryon chemical potential (μ_B)(middle panel) and anisotropic parameter (ξ)(right panel) for three different values of temperatures, i.e T = 200 MeV, T=250 MeV, and T = 300 MeV.

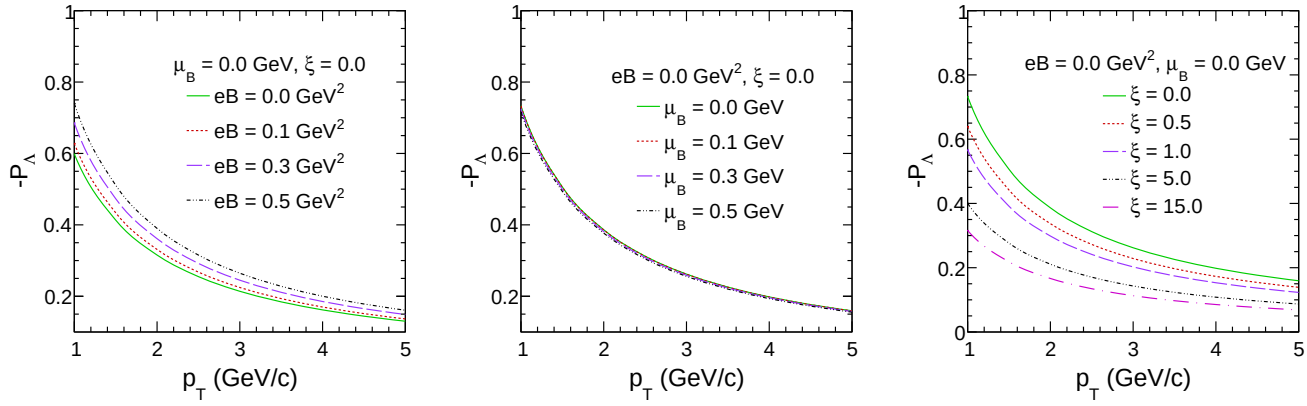


FIG. 2: (Color online) The variation of Λ hyperon polarization as a function of transverse momentum due to recombination process in the presence of magnetic field (eB)(left panel), baryon chemical potential (μ_B)(middle panel) and anisotropic parameter (ξ)(right panel) at temperature T = 200 MeV.

ization of hyperons.

The polarization of Λ, $\Sigma^{\pm}$, Ξ^{-}, Δ^{++}, and Ω^{-} hyperons in quark recombination process is calculated using the Eqs. 2, 3, 4. The left panel of Fig. 2 shows the variation of Λ hyperon polarization as a function of the transverse momentum (p_T) for different values of the magnetic field at $\mu_B = 0$ GeV and $\xi = 0$ in recombination process of hadronization. We consider three finite values of the magnetic field, i.e., eB = 0.1, 0.3, and 0.5 GeV^2. Figure 2 indicates the polarization of Λ hyperon decreases as a function of the p_T for all values of the magnetic field. It is observed that the increase in polarization of the Λ hyperon with a magnetic field can be attributed to the corresponding increase in the Debye mass. Apart from that, the magnetic field helps to polarize the particle and anti-particle through their magnetic moments. The Λ hyperon has a negative magnetic moment, whereas the $\bar{\Lambda}$ hyperon has a positive magnetic moment. In the presence of a magnetic field, the spin of the $\bar{\Lambda}$ is strongly aligned along the direction of the field, while the Λ spin is relatively poorly aligned against the direction of the field. This argument could qualitatively describe the splitting of Λ and $\bar{\Lambda}$ hyperon with $P_{\bar{\Lambda}} > P_{\Lambda}$. However, current formalism predicts the same polarization for Λ and $\bar{\Lambda}$. Moreover, various phenomenological studies suggest that hyperon polarization is directly proportional to the magnetic field and plays a crucial role in particle polarization [56, 58].

During the exclusive recombination process of hadronization, it has been observed that the polarization

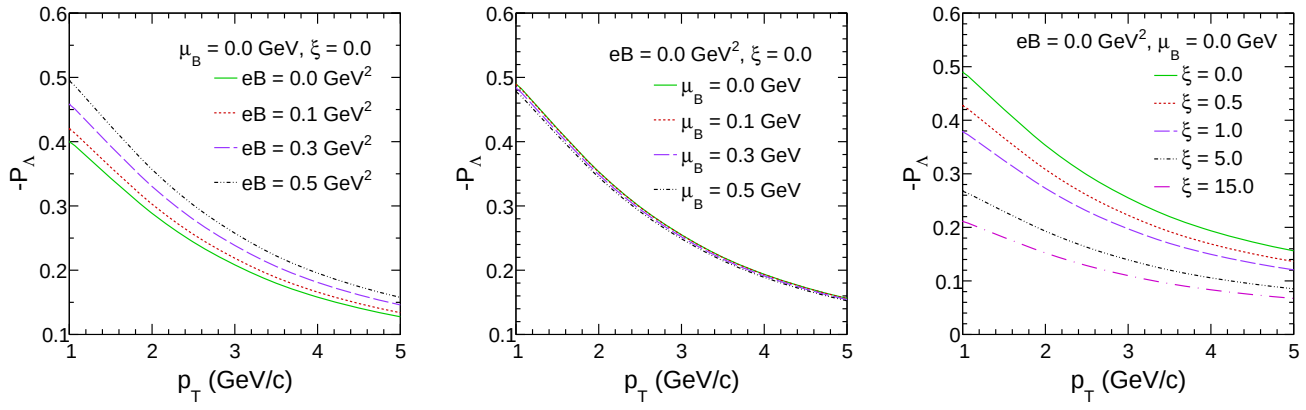


FIG. 3: (Color online) The variation of Λ hyperon polarization as a function of transverse momentum due to fragmentation process in the presence of magnetic field (eB)(left panel), baryon chemical potential (μ_B)(middle panel) and anisotropic parameter (ξ)(right panel) at temperature T = 200 MeV.

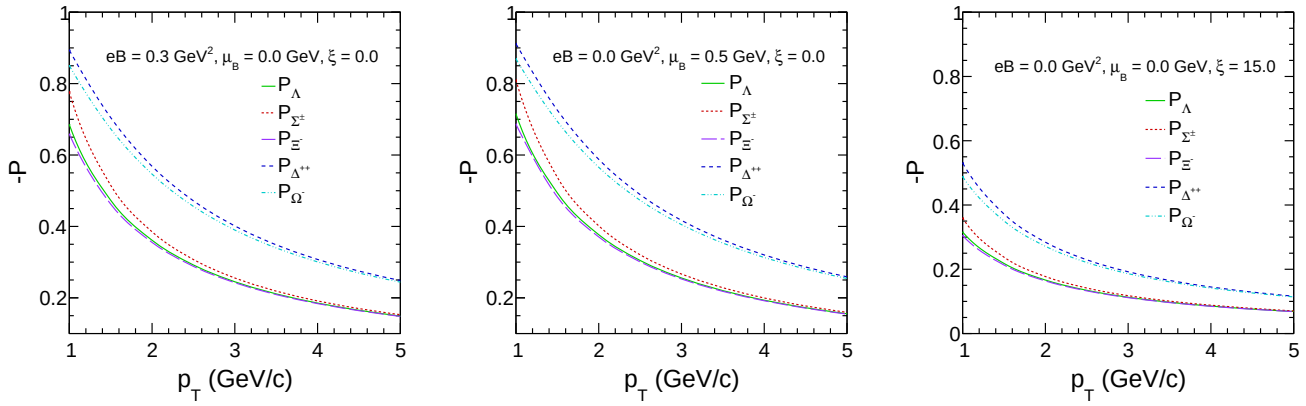


FIG. 4: (Color online)The variation of baryon polarization as a function of transverse momentum in the recombination process due to magnetic field (eB)(left panel), baryon chemical potential (μ_B)(middle panel) and anisotropic parameter (ξ)(right panel) at temperature T = 200 MeV.

of Λ hyperons aligns with the polarization of the initial strange quark. The initial strange quarks with lower momentum have higher values of Λ polarization after recombining to Λ hyperons and vice-versa. The middle panel of Fig. 2 illustrates the dependence of Λ hyperon polarization as a function of p_T for various values of chemical potential in the recombination scenario of hadronization. Three finite values of chemical potential are considered with μ_B = 0.1, 0.3, and 0.5 GeV. It can be seen here that there is no notable effect of baryon chemical potential on the polarization of Λ. The right panel of Fig. 2 depicts the variation of Λ polarization as a function of p_T for various values of the anisotropic parameter. Anisotropy in the medium is believed to restrict the alignment of particle spins along a specific direction. Here, we considered four different values of anisotropic parameters, i.e., ξ = 0.5, 1.0, 5.0, 15.0, along with the isotropic case, i.e., ξ = 0. It is observed that the Λ hyperon polarization decreases with the anisotropy parameter. It is an intriguing observation that the medium anisotropy plays a significant role in influencing the polarization of Λ hyperons.

Figure 3 depicts the change in the Λ polarization with p_T corresponding to the fragmentation process of hadronization. The polarization of Λ, $\Sigma^{\pm}$, Ξ^{-}, Δ^{++}, and Ω^{-} hyperons due to fragmentation is obtained using Eqs. 5, 6. The polarization corresponding to the fragmentation process is obtained using the n_s and f_s parameters. The values of these parameters

are extracted from Ref. [83]. Similar to Fig. 2, the polarization is shown for the same values of eB (left panel), μ_B (middle panel), and ξ (right panel) in Fig. 3. The p_T dependence of Λ hyperon polarization in the presence of an external magnetic field, baryon chemical potential, and medium anisotropy in the fragmentation process follow the same trend as the recombination process, except a change in the slope is observed at low p_T. However, the magnitude of Λ polarization due to fragmentation is comparably less than the recombination process.

Further, in Fig. 4, the variation in the polarization because of recombination is shown as a function of p_T at $eB = 0.3$ GeV^2, $\mu_B = 0$ GeV and $\xi = 0$ for Λ, $\Sigma^{\pm}$, Ξ^-, Δ^{++}, and Ω^- hyperons. Moreover, we observe that the hyperon spin-polarization is dominant at low p_T. The middle and right panel of Fig. 4 shows Λ, $\Sigma^{\pm}$, Ξ^-, Δ^{++}, and Ω^- polarization with p_T for baryon chemical potential $\mu_B = 0.5$ GeV and anisotropic parameter $\xi = 15$, respectively. The role of baryon chemical potential on particle polarization is quite similar to the magnetic field. Fig. 4 indicates a quark flavor-dependent polarization of hyperons in the heavy-ion collisions. We found that there is grouping among the particles of baryon octet with spin-1/2 and baryon decouplet with spin-3/2 during the spin-orbit coupling. Particles having the same quark flavor, for example, Ω^- (sss) and Δ^{++} (uuu) have similar spin-orientation along the direction of OAM as compared to particles having mixed quark flavor, e.g., Λ (uds), $\Sigma^{\pm}$ (uus/dds), and Ξ^- (dss). Therefore, particles of the same species possibly have some correlation among themselves, which drives the spin polarization. This correlation might be absent in mixed species particles; such an observation is recently reported for ϕ meson in Ref. [49]. In addition, the baryon-decouplet of spin-3/2 has higher polarization as compared to the baryon-octet of spin-1/2. Therefore, particle spin might play a role in its polarization. However, to confirm the importance of spin in the global polarization picture and to better understand the polarization mechanism in heavy-ion collisions, the experimental verification of global polarization for different particles with different spin and/or magnetic moments is required. Although the recent measurement global polarization of Ξ and Ω hyperons at STAR hint at a possible hierarchy in the global polarization, i.e., $P_\Lambda < P_\Xi < P_\Omega$ [36]. The large systematic and statistical uncertainties raise a question as to whether the difference is either due to the mass, lifetime, strangeness quantum number, differential freeze-out, etc.

Similarly, Fig. 5 shows the dependence of Λ, $\Sigma^{\pm}$, Ξ^-, Δ^{++}, and Ω^- hyperon polarization as a function of p_T in the presence of external magnetic field (left panel), baryon chemical potential (middle panel) and medium anisotropy parameters (right panel) in the fragmentation process of hadronization. In the fragmentation process, the polarization of Δ^{++} and Ω^- is 1/3 of the polarization due to the u and s quarks, respectively. Therefore, baryons having the same quark flavor, each quark contributes equally to the final state baryon polarization. Meanwhile, baryons of mixed quark flavors have linear combinations due to each flavor. As consequence, a significant decrease in the Δ^{++}, and Ω^- polarization is observed compared to Λ, $\Sigma^{\pm}$, and Ξ^- in Fig 5. However, the polarization in the fragmentation process is always less compared to recombination for all baryons. The change in the polarization due to fragmentation in the presence of the magnetic field, baryon chemical potential, and medium anisotropy follow the same trend of the polarization obtained from the recombination process.

The left panel of Fig. 6 shows the combined effect of the magnetic field, baryon chemical potential, and medium anisotropy on Λ hyperon polarization. The interplay of these three effects is examined in Fig 6. We found that the medium anisotropy has a prevalent impact on the polarization observable as compared to the magnetic field and baryon chemical potential. The observed feature of medium anisotropy on polarization is due to the influence of anisotropy on Debye mass. Medium anisotropy plays a crucial role in controlling the spin-polarization of hyperon, which needs an extensive investigation from a theoretical standpoint. The right panel of Fig. 6 shows change in the polarization for Λ along with $\Sigma^{\pm}$, Ξ^-, Ω^-, and Δ^{++} baryons corresponding to the combined effect of the magnetic field, baryon chemical potential, and medium momentum-space anisotropy.

B. Spin alignment of vector mesons

Similar to the global polarization of hyperon, the spin-alignment of vector meson is also based on the same mechanism of spin-orbit coupling. Since a vector meson with spin one can have three different spin orientations, the probability for its spin to align in a given direction is 1/3. Any value of the spin alignment differs from 1/3, which means the polarization of the vector meson is along that direction. With the recent large spin alignment measurement of ϕ vector meson at RHIC [41, 42] and LHC [44], it becomes a challenging and active area of research for the theoretical community to know the possible sources of spin alignment for ϕ vector meson. The spin alignment of vector meson comes from both constituent quark and anti-quark and hence its polarization is quadratic in vorticity [56].

The left panel of Fig. 7 shows the effect of magnetic field on the spin alignment of ϕ meson for two different hadronization scenarios as described in formalism section II. The curves below line 1/3 depict the spin alignment of ϕ-meson due to the recombination of polarized

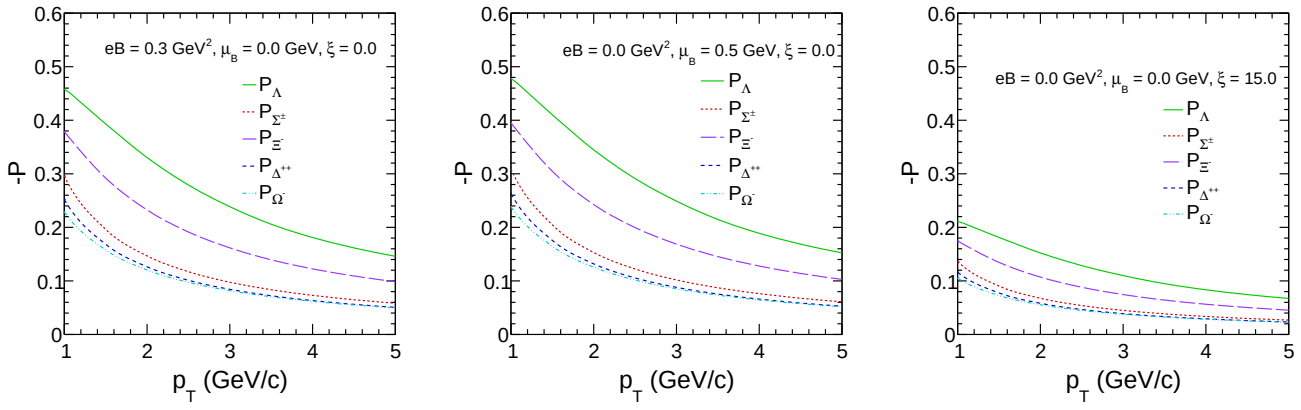


FIG. 5: (Color online) The variation of hyperon polarization as a function of transverse momentum due to the fragmentation process in the presence of magnetic field (eB)(left panel), baryon chemical potential (μ_B)(middle panel) and anisotropic parameter (ξ)(right panel) at temperature T = 200 MeV.

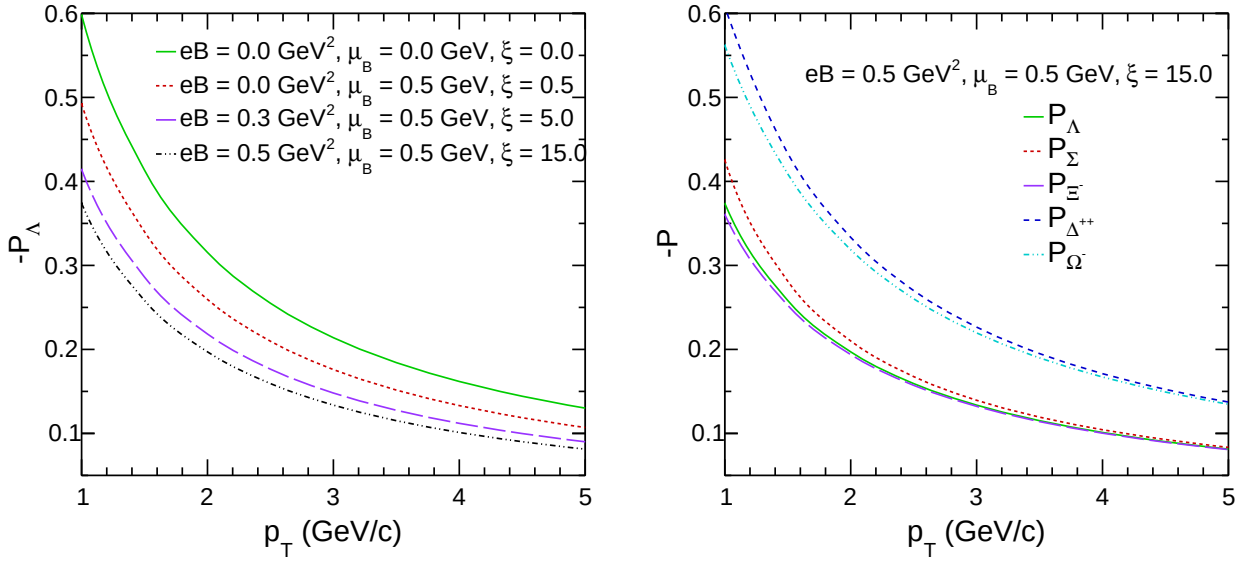


FIG. 6: (Color online)The variation of hyperon polarization as a function of transverse momentum due to magnetic field (left panel), baryon chemical potential (middle panel), and anisotropic parameter (right panel) at temperature T = 200 MeV.

quarks and anti-quarks. The curves above line 1/3 show the hadronization process of quark and anti-quark pairs in which either of them is created in the fragmentation process. We compare the obtained results with the STAR Collaboration data for Au+Au collisions for various centers of mass energies in (20-60)% bin [41]. The choice of center of mass energy selection of STAR data for Au+Au collisions in Fig 7 and Fig. 8 is just for illustration purposes, and it has nothing specific relation with the magnetic field and anisotropy. It is found that the absolute deviation of ρ_{00} from 1/3 ($|\rho_{00} - 1/3|$) for ϕ vector meson increases with magnetic field for both cases of hadronization processes. The magnetic field aligns the particle spin in a specific direction; hence, it helps to enhance the degree of spin alignment. It is evident that the spin alignment of vector meson is directly proportional to

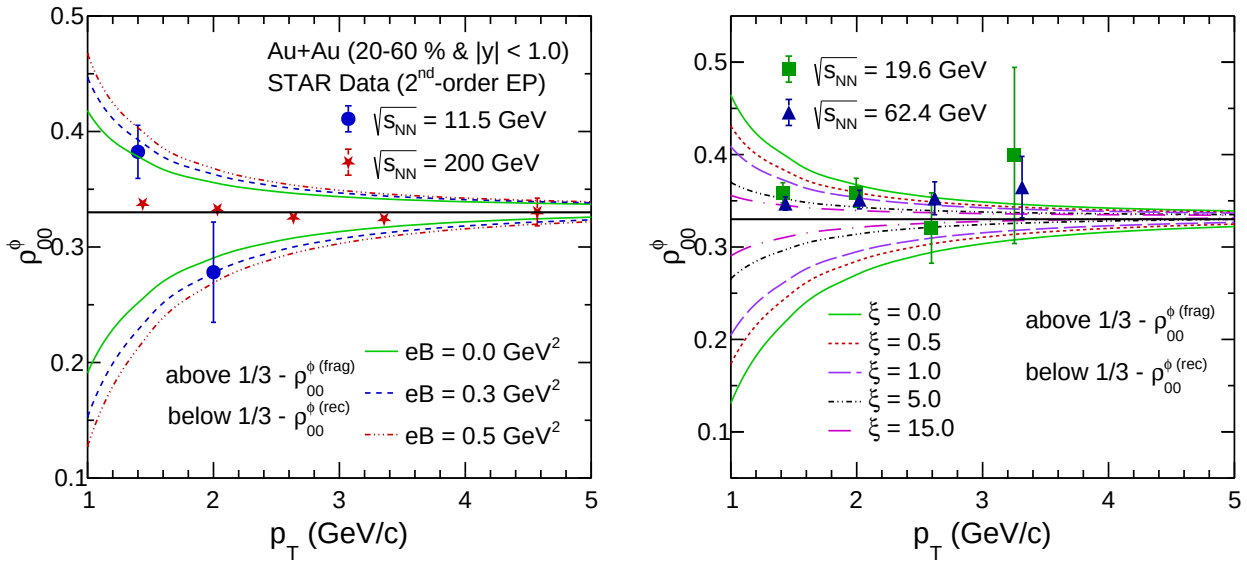


FIG. 7: (Color online) The variation of spin alignment of ϕ vector meson as a function of transverse momentum due to the magnetic field (left panel) and anisotropic parameter (right panel) at temperature T = 200 MeV. The results are compared with STAR measurements taken from Ref. [41].

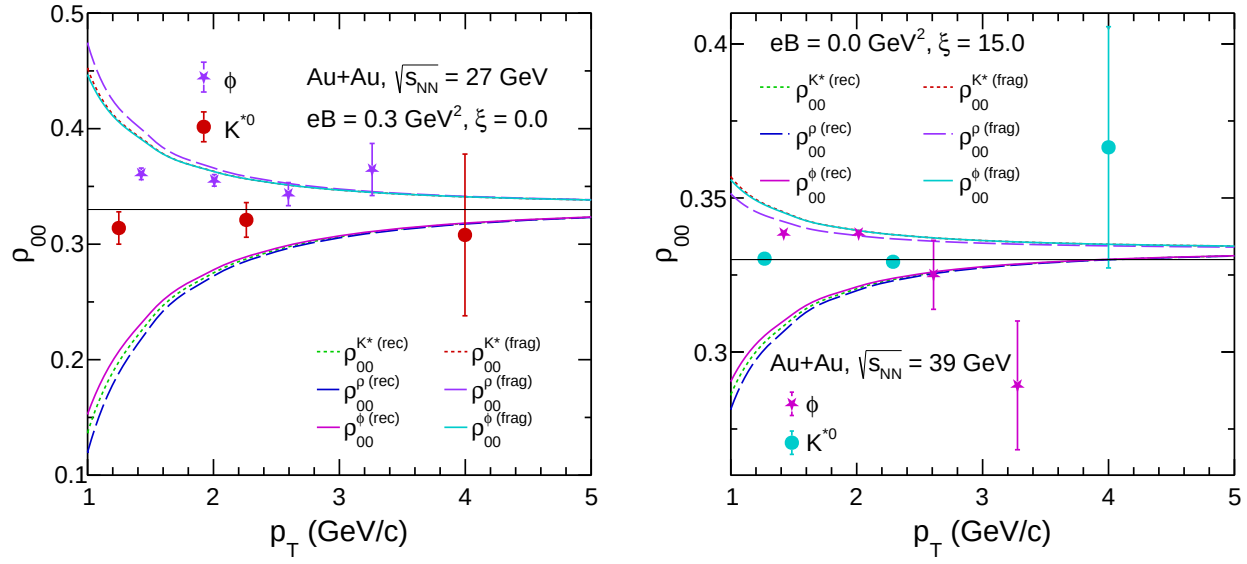


FIG. 8: (Color online) The variation of spin alignment of ϕ, K^{*0}, ρ vector meson as a function of transverse momentum due to the magnetic field (left panel), and anisotropic parameter (right panel) at temperature T = 200 MeV. The results are compared with STAR measurements taken from Ref. [41].

the square of the magnetic field. Since baryon chemical potential does not have a substantial effect on hyperon polarization, we have neglected the effect of baryon chemical potential on the spin alignment measurement of vector meson. The right panel of Fig. 7 shows the effect of ξ on the spin alignment of ϕ meson with a set of ξ values. From the figure, it is evident that medium anisotropy significantly modifies the vector meson spin alignment. The left panel of Fig. 8 shows the spin alignment of ϕ, K^*, and ρ meson at $eB = 0.5$ GeV^2 for corresponding to fragmentation and recombination hadronization scenarios. The right panel of Fig. 8 shows the spin alignment of ϕ, K^*, and ρ meson for $\xi = 15$. Figure 7 and 8 depicts the absolute deviation of ρ_{00} from 1/3 is more at low p_T for the recombination process as compared to the fragmentation. This implies that the degree of spin alignment is more at low p_T in the recombination process for ϕ, K^*, and ρ vector mesons. The value of ρ_{00} is close to 1/3 towards high p_T showing zero spin alignment for ϕ, K^*, and ρ mesons. In the quark polarization model, the polarization of a quark is inversely proportional to the square of its mass, which suggests a mass ordering in the spin alignment effect of vector mesons due to their constituent quark composition. Figure 8 depicts $\rho_{00}^{\rho} > \rho_{00}^{K^*} > \rho_{00}^{\phi}$ at low p_T in the recombination scenario.

IV. CONCLUSION

In summary, we explored the global polarization of hyperons and spin-alignment of vector mesons in the presence of magnetic field, baryon chemical potential, and medium momentum-space anisotropy. The individual and combined effect of these parameters on hyperons and vector mesons polarization have been studied in the recombination and fragmentation process of hadronization. We found that the polarization of hyperons and spin-alignment of vector mesons increases with the increase in magnetic field. In this study, the baryon chemical potential has negligible dependence on hyperon polarization. The medium anisotropy significantly affects the polarization and spin alignment of hadrons. It is interesting to note that the recombination process dominates at low p_T in comparison with the fragmentation process. The Debye mass-based polarization may have a substantial implication on particle production yield. Therefore, for a comprehensive study, one may include such an effect to study the hadron production in heavy-ion collisions.

The hyperon polarization obtained through the global quark polarization method is comparably higher than the experimentally measured hyperon polarization in heavy-ion collisions at RHIC. In addition to the magnetic field, baryon chemical potential, and medium momentum-space anisotropy used in this work, there may be various other sources that may affect the hyperon polarization. Such sources could be medium rotation, hadronic interactions, anisotropic flows in the medium, and fluctuation of other force fields such as vorticity, temperature gradient, shear tensor, strong force field, etc. However, the contribution of these sources on hyperon polarization has yet to be studied collectively. The qualitative prediction of global polarization is an open challenge and requires further investigation.

ACKNOWLEDGEMENT

Bhagyarathi Sahoo acknowledges the financial aid from CSIR, Government of India. The authors gratefully acknowledge the DAE-DST, Government of India funding under the mega-science project "Indian Participation in the ALICE experiment at CERN" bearing Project No.SR/MF/PS-02/2021-IITI (E-37123).

[1] J. Adam *et al.* [ALICE Collaboration], Nature Phys. **13**, 535 (2017).
[2] V. Khachatryan *et al.* [CMS Collaboration], Phys. Lett. B **765**, 193 (2017).
[3] A. Ortiz Velasquez, P. Christiansen, E. Cuautle Flores, I. Maldonado Cervantes and G. Paić, Phys. Rev. Lett. **111**, 042001 (2013).
[4] C. Zhang, A. Behera, S. Bhatta and J. Jia, Phys. Lett. B **822**, 136702 (2021).
[5] B. Sahoo, D. Sahu, S. Deb, C. R. Singh and R. Sahoo, [arXiv:2308.15151 [hep-ph]].
[6] K. Fukushima, D. E. Kharzeev and H. J. Warringa, Phys. Rev. D **78**, 074033 (2008).
[7] D. E. Kharzeev, L. D. McLerran and H. J. Warringa, Nucl. Phys. A **803** 227 (2008).
[8] O. Rogachevsky, A. Sorin and O. Teryaev, Phys. Rev. C **82**, 054910 (2010).
[9] D. Kharzeev and A. Zhitnitsky, Nucl. Phys. A **797**, 67 (2007).
[10] F. Becattini and I. Karpenko, Phys. Rev. Lett. **120**, 012302 (2018).
[11] J. Adam *et al.* [STAR Collaboration], Phys. Rev. Lett. **123**, 132301 (2019).

[12] Z. T. Liang and X. N. Wang, Phys. Rev. Lett. **94**, 102301 (2005). [erratum: Phys. Rev. Lett. **96** (2006), 039901]
[13] Z. T. Liang and X. N. Wang, Phys. Lett. B **629**, 20 (2005).
[14] Z. t. Liang, J. Phys. G **34**, S323-330 (2007).
[15] J. H. Gao, S. W. Chen, W. t. Deng, Z. T. Liang, Q. Wang and X. N. Wang, Phys. Rev. C **77**, 044902 (2008).
[16] D. L. Adams *et al.* [FNAL-E704], Phys. Lett. B **264**, 462(1991).
[17] A. Ipp, A. Di Piazza, J. Evers and C. H. Keitel, Phys. Lett. B **666**, 315 (2008).
[18] F. Becattini, V. Chandra, L. Del Zanna and E. Grossi, Annals Phys. **338** 32 (2013).
[19] B. I. Abelev *et al.* [STAR Collaboration], Phys. Rev. C **76**, 024915 (2007).
[20] L. Adamczyk *et al.* [STAR Collaboration], Nature **548** 62 (2017).
[21] J. Adam *et al.* [STAR Collaboration], Phys. Rev. C **98**, 014910 (2018).
[22] S. Alzhrani, S. Ryu and C. Shen, Phys. Rev. C **106**, 014905 (2022).
[23] O. Vitiuk, L. V. Bravina and E. E. Zabrodin, Phys. Lett. B **803**, 135298 (2020).
[24] Y. Jiang, Z. W. Lin and J. Liao, Phys. Rev. C **94** (2016) no.4, 044910 [erratum: Phys. Rev. C **95** (2017) no.4, 049904]
[25] Y. Sun and C. M. Ko, Phys. Rev. C **96**, 024906 (2017).
[26] Y. B. Ivanov, V. D. Toneev and A. A. Soldatov, Phys. Rev. C **100**, 014908 (2019).
[27] K. K. Pradhan, B. Sahoo, D. Sahu and R. Sahoo, [arXiv:2304.05190 [hep-ph]].
[28] L. P. Csernai, V. K. Magas and D. J. Wang, Phys. Rev. C **87**, 034906 (2013).
[29] L. P. Csernai, D. J. Wang, M. Bleicher and H. Stöcker, Phys. Rev. C **90**, 021904 (2014).
[30] L. G. Pang, H. Petersen, Q. Wang and X. N. Wang, Phys. Rev. Lett. **117**, 192301 (2016).
[31] H. Li, L. G. Pang, Q. Wang and X. L. Xia, Phys. Rev. C **96**, 054908 (2017).
[32] B. Fu, K. Xu, X. G. Huang and H. Song, Phys. Rev. C **103**, 024903 (2021).
[33] I. Karpenko and F. Becattini, Nucl. Phys. A **967**, 764 (2017).
[34] F. Becattini and M. A. Lisa, Ann. Rev. Nucl. Part. Sci. **70**, 395 (2020).
[35] Xie, Y., Wang, D. and Csernai, L.P., Eur. Phys. J. C **80**, 39 (2020).
[36] J. Adam *et al.* [STAR Collaboration], Phys. Rev. Lett. **126**, 162301 (2021).
[37] S. Acharya *et al.*, [ALICE Collaboration], Phys. Lett. B **815**, 136146 (2021).
[38] S. Acharya *et al.*, [ALICE Collaboration], Phys. Rev. Lett. **131**, 042303 (2023).
[39] L. Adamczyk *et al.*, [STAR Collaboration], Phys. Lett. B **739**, 180 (2014).
[40] J. Adam *et al.*, [STAR Collaboration], Phys. Rev. D **102**, 092009 (2020).
[41] M. S. Abdallah *et al.* [STAR Collaboration], Nature **614** 244 (2023).
[42] S. Singha [STAR Collaboration], Nucl. Phys. A **1005**, 121733 (2021).
[43] B. I. Abelev *et al.* [STAR Collaboration], Phys. Rev. C **77**, 061902 (2008).
[44] S. Acharya *et al.* [ALICE Collaboration], Phys. Rev. Lett. **125**, 012301 (2020).
[45] R. Singh [ALICE Collaboration], Nucl. Phys. A **982**, 515 (2019).
[46] B. Mohanty, S. Kundu, S. Singha and R. Singh, Mod. Phys. Lett. A **36**, 2130026 (2021).
[47] L. P. Csernai, J. I. Kapusta and T. Welle, Phys. Rev. C **99**, 021901 (2019).
[48] X. L. Sheng, L. Oliva and Q. Wang, Phys. Rev. D **101**, 096005 (2020). [erratum: Phys. Rev. D **105**, 099903 (2022).]
[49] X. L. Sheng, L. Oliva, Z. T. Liang, Q. Wang and X. N. Wang, Phys. Rev. Lett. **131**, 042304 (2023).
[50] F. Becattini, M. Buzzegoli, G. Inghirami, I. Karpenko and A. Palermo, Phys. Rev. Lett. **127**, 272302 (2021).
[51] H. Z. Wu, L. G. Pang, X. G. Huang and Q. Wang, Phys. Rev. Research. **1**, 033058 (2019).
[52] X. L. Xia, H. Li, X. G. Huang and H. Zhong Huang, Phys. Lett. B **817** (2021), 136325
[53] V. Skokov, A. Y. Illarionov and V. Toneev, Int. J. Mod. Phys. A **24**, 5925 (2009).
[54] B. Sahoo, C. R. Singh, D. Sahu, R. Sahoo and J. e. Alam, Eur. Phys. J. C **83**, 873 (2023).
[55] B. Sahoo, K. K. Pradhan, D. Sahu and R. Sahoo, Phys. Rev. D **108**, 074028 (2023).
[56] Y. G. Yang, R. H. Fang, Q. Wang and X. N. Wang, Phys. Rev. C **97**, 034917 (2018).
[57] X. L. Sheng, Q. Wang and X. N. Wang, Phys. Rev. D **102**, 056013 (2020).
[58] F. Becattini, I. Karpenko, M. Lisa, I. Upsal and S. Voloshin, Phys. Rev. C **95**, 054902 (2017).
[59] S. Wu and Y. Xie, Eur. Phys. J. A **59**, 108 (2023).
[60] Z. Z. Han and J. Xu, Phys. Lett. B **786**, 255 (2018).
[61] K. Xu, F. Lin, A. Huang and M. Huang, Phys. Rev. D **106**, L071502 (2022).
[62] A. Ayala, M. A. Ayala Torres, E. Cuautle, I. Domínguez, M. A. Fontaine Sanchez, I. Maldonado, E. Moreno-Barbosa, P. A. Nieto-Marín, M. Rodríguez-Cahuantzi and J. Salinas, *et al.* Phys. Lett. B **810**, 135818 (2020).
[63] X. Guo, J. Liao and E. Wang, Sci. Rep. **10**, 2196 (2020).
[64] J. H. Gao, Z. T. Liang, S. Pu, Q. Wang and X. N. Wang, Phys. Rev. Lett. **109**, 232301 (2012).
[65] L. Dong, Y. Guo, A. Islam and M. Strickland, Phys. Rev. D **104**, 096017 (2021).
[66] B. Karmakar, A. Bandyopadhyay, N. Haque and M. G. Mustafa, Eur. Phys. J. C **79**, 658 (2019).
[67] H. X. Zhang, J. W. Kang and B. W. Zhang, Eur. Phys. J. C **81**, 623 (2021).
[68] A. Bandyopadhyay, B. Karmakar, N. Haque and M. G. Mustafa, Phys. Rev. D **100**, 034031 (2019).
[69] B. Singh, L. Thakur and H. Mishra, Phys. Rev. D **97**, 096011 (2018).
[70] N. Haque, A. Bandyopadhyay, J. O. Andersen, M. G. Mustafa, M. Strickland and N. Su, JHEP **05**, 027 (2014).
[71] J. Zhao and B. Chen, Phys. Rev. C **107**, 044909 (2023).
[72] D. Lafferty and A. Rothkopf, Phys. Rev. D **101**, 056010 (2020).
[73] G. Huang and P. Zhuang, Phys. Rev. D **104**, 074001 (2021).
[74] M. Le Bellac, Thermal Field Theory (Cambridge University Press, Cambridge, England, 1996).
[75] L. Dong, Y. Guo, A. Islam, A. Rothkopf and M. Strickland, JHEP **09**, 200 (2022).

[76] M. Strickland and D. Bazow, Nucl. Phys. A **879**, 25-58 (2012).
[77] P. Romatschke and M. Strickland, Phys. Rev. D **68**, 036004 (2003).
[78] P. Romatschke and M. Strickland, Phys. Rev. D **70**, 116006 (2004).
[79] M. Strickland, Anisotropic hydrodynamics: Three lectures, Acta Phys. Pol. B 45, 2355 (2014).
[80] C. Bonati, M. D'Elia, M. Mariti, M. Mesiti, F. Negro, A. Rucci and F. Sanfilippo, Phys. Rev. D **95**, 074515 (2017).
[81] A. Bandyopadhyay, C. A. Islam and M. G. Mustafa, Phys. Rev. D **94**, 114034 (2016).
[82] A. Bandyopadhyay, R. L. S. Farias, B. S. Lopes and R. O. Ramos, Phys. Rev. D **100**, 076021 (2019).
[83] J. w. Zhang, H. h. Li, F. l. Shao and J. Song, Chin. Phys. C **44**, 014101 (2020).

Prepared for submission to JHEP

Constraining Asymmetric Dark Matter using Colliders and Direct Detection

Arnab Roy [a,b], **Basudeb Dasgupta** [a] and **Monoranjan Guchait** [a]

[a] *Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India*
[b] *School of Physics and Astronomy, Monash University, Wellington Road, Clayton, Victoria 3800, Australia*

E-mail: arnab.roy1@monash.edu, bdasgupta@theory.tifr.res.in, guchait@tifr.res.in

Abstract: We reappraise the viability of asymmetric dark matter (ADM) realized as a Dirac fermion coupling dominantly to the Standard Model fermions. Treating the interactions of such a DM particle with quarks/leptons in an effective-interactions framework, we derive updated constraints using mono-jet searches from the Large Hadron Collider (LHC) and mono-photon searches at the Large Electron-Positron (LEP) collider. We carefully model the detectors used in these experiments, which is found to have significant impact. The constraint of efficient annihilation of the symmetric part of the ADM, as well as other observational constraints are synthesized to produce a global picture. Consistent with previous work, we find that ADM with mass in the range 1–100 GeV is strongly constrained, thus ruling out its best motivated mass range. However, we find that leptophilic ADM remains allowed for $\gtrsim 10$ GeV DM, including bounds from colliders, direct detection, and stellar heating. We forecast that the Future Circular Collider for electron-positron collisions (FCC-ee) will improve sensitivity to DM-lepton interactions by almost an order of magnitude.

arXiv:2402.17265v1 [hep-ph] 27 Feb 2024

Contents

1 Introduction **1**

2 Dark Matter Interactions and Production **3**
- 2.1 Effective interactions 3
- 2.2 Relic density and the asymmetry criterion 4

3 Constraining ADM-Quark Interactions **6**
- 3.1 Mono-jet searches at hadron colliders 6
- 3.2 Direct detection experiments 7
- 3.3 Constraints on the $\Lambda - m_\chi$ plane 8

4 Constraining ADM-Lepton Interactions **11**
- 4.1 Mono-photon searches at LEP 11
- 4.2 Discovery potential of ADM at FCC-ee 14
- 4.3 Constraints from compact stars and direct detection 15
- 4.4 Constraints and discovery potential for leptophilic ADM 16

5 Summary and Outlook **18**

A Expressions of $\langle\sigma v\rangle$ for different operators **21**

1 Introduction

Despite the significant asymmetry between the cosmological abundances of matter and antimatter, we observe comparable abundances of ordinary matter, dominated by baryons (b), and of dark matter (DM), namely $\Omega_{\rm b} \simeq \Omega_{\rm DM}/5$ [1]. This observation has motivated the Asymmetric Dark Matter (ADM) hypothesis [2–5]. Here, the similar baryon and DM abundances are a consequence of both densities arising from an asymmetry of the corresponding Dirac fermion over its anti-fermion, with the two asymmetries related to each other in some way [6–11]. Similar to models of weakly interacting massive particle type DM, where the DM is produced via thermal freeze-out and traditionally though of as having mass near the electroweak scale, the models for ADM are also testable at particle colliders because the expected mass-range for ADM particles is around the GeV scale or close to it. This is because ADM models predict that the number abundances, $n_{\rm DM}$ and $n_{\rm b}$, are similar and the ratio of observed mass abundances then predicts ADM particles at the GeV-scale [12–21].

The ADM hypothesis therefore presents a theoretical alternative to WIMPs, retaining almost all of its strengths. These models are well-motivated from the ultraviolet considerations, such as GUTs and additional confined gauge sectors [22–31]. They are conceptually

minimal, employing the mathematical structures and physical mechanisms that are believed to already underlie the Standard Model, e.g., Dirac fermions, particle-antiparticle asymmetry, gauge interactions, anomalies, and sphalerons. And importantly, they predict GeV-scale masses for the DM particle, with observable signals not only at direct detection (DD) experiments and colliders, but also constraints from astrophysics and cosmology.

While considerable attention has been devoted to the theories of ADM, its detectability in particle colliders remains relatively less well-explored. A previous study by March-Russell et al. (2012) [32] presented collider and DD constraints on ADM, together with the constraints imposed by the model itself. The key idea being that efficient annihilation of the symmetric part of the ADM relic density requires that the coupling between DM and SM fermions cannot be too weak. This is under the usual simplifying assumption that the dark matter annihilates into and scatters off SM fermions, and these processes are mediated by the same elementary interaction. Quite remarkably, already in this early work, it was found that collider data and direct detection searches practically rule out ADM with mass all the way up to 100 GeV or more.

Developments over the past decade motivate a reappraisal. Continued data-taking at the LHC as well as at the significantly more sensitive and diverse set of DD searches now allow a more comprehensive study of ADM couplings to SM fermions. Further, there is a renewed interest in ADM, especially in connection to their dramatic impact on compact stars [33–52]. And finally, several new DM experiments [53], as well as experiments at the HL-LHC, aim to search widely for DM, which motivate an updated idea of *'where to look for new physics'*.

In this study, we revisit collider constraints on ADM and provide new and updated bounds on the interactions between ADM and quarks/leptons. Our analysis includes a careful incorporation of detector efficiencies, corresponding to different phase space regions of the final state particles. For DM-quark interactions, our collider limits are based on the outcomes of mono-jet searches conducted at the LHC. Further, by combining other DD and collider bounds on DM with the prerequisites of the ADM hypothesis, we find that all the operators we consider are ruled out for DM mass up to a few hundred GeV; for some operators the exclusion stretches almost to TeV masses. Prima facie the basic ADM scenario appears strongly disfavored.

We then turn our attention to a leptophilic variant of the ADM scenario, where ADM couples to SM leptons but not to quarks. Such ADM models may be motivated by theory, and are poised to take advantage of the upcoming experiments. In some ADM models, the DM asymmetry fundamentally arises due to new sphalerons that transform the lepton asymmetry created by leptogenesis into a DM asymmetry. Here, a chiral gauge interaction shared by DM and leptons directly leads to the above phenomenology and does not necessarily need to involve quarks. On the observational side, DM interactions with leptons (electrons, in particular) has gained interest over the past decade. Unsurprisingly, DM-lepton interactions have been studied in depth [13, 54–68] and a gamut of experiments have already produced interesting limits on leptophilic DM [69–72]. More importantly, there are now several novel proposals that promise to make significant progress in this direction in the near future [73–89]. Limits on DM-lepton interactions from the heating of neutron stars

(NS) [90] and white dwarfs (WD) [91] happen to provide some of the strongest constraints. On the collider front, mono-photon searches at LEP have been set bounds on four effective operators (three of which are independent) [92], which are more stringent than the NS or WD bounds for DM masses in the 1-100 GeV range. Investigations have also focused on the prospects of identifying leptophilic dark matter through future lepton colliders, mainly the International Linear Collider (ILC) [93–95]. In the latter part of this paper, we update the bounds on DM-lepton interactions using the LEP mono-photon searches, compare with other constraints as for the quark case, and appraise the viability of leptophilic ADM. We also forecast discovery prospects at the prospective FCC-ee.

The structure of this paper is as follows. In Section 2, we start by establishing the EFT framework and provide a brief introduction to the ADM hypothesis and review the existing constraints on interactions between ADM and quarks. In Section 3, we calculate new bounds on the ADM-quark interactions from the mono-jet searches at the LHC and present the constraints on ADM-quark interactions for all EFT operators. Additionally, in Section 4, we investigate the DM-lepton interactions. Here, a comprehensive methodology to calculate the LEP bounds are presented, followed by the FCC-ee discovery prospects. The combined pictures of the allowed (and excluded) regions of ADM-lepton interactions from all relevant searches are presented at the end of section 4. Lastly, Section 5 summarizes our findings and concludes.

2 Dark Matter Interactions and Production

2.1 Effective interactions

We assume the DM is a Dirac fermion and the mediators responsible for its interactions with the SM particles are heavier than the DM itself. The interactions are thus encoded in higher-dimensional operators within the framework of an effective field theory (EFT) involving SM and DM degrees of freedom. However, we will ignore the dimension-5 couplings of DM to the SM Higgs, which is already very well studied and constrained. See, for example, the stringent upper limits of $\mathrm{H_{SM}} \to$ invisible branching fractions [96, 97].[1] With these considerations, in Table 1 we write down all allowed EFT operators at dimension-6. These operators are of the form

$$\mathcal{O}_{\Gamma\Gamma'} = \frac{1}{\Lambda^2}(\bar{\psi}\Gamma\psi)(\bar{\chi}\Gamma'\chi), \tag{2.1}$$

where ψ and χ are Dirac fermion SM and DM particles, respectively, whereas $\Gamma^{(\prime)}$s are strings of gamma matrices,

$$\Gamma = \{1, \gamma^5, \gamma^\mu, \gamma^\mu\gamma^5, \sigma^{\mu\nu}, \sigma^{\mu\nu}\gamma^5\}. \tag{2.2}$$

[1] Note that the Higgs portal for fermionic DM of mass smaller than ∼20 GeV, consistent with $\mathrm{BR}(\mathrm{H_{SM}} \to$ invisible$) < 10\%$, leads to DM-nucleon scattering cross-section below the "neutrino floor" [98–100], which is uninteresting from DD perspective [101]. Alternatively, though a higher mass region of DM allows parameter regions above the neutrino floor, they are excluded by the DD limits [101].

Table 1: List of dimension-6 operators connecting SM fermions to Dirac DM with flavor-universal couplings, along with the dependence of the scattering cross section on the spin of the SM target (and the leading dependence on q or v), as well as the leading partial-wave contribution to the corresponding annihilation rate; q and v denote the magnitude of transferred momentum and the relative velocity between the target and the DM, respectively.

Operator	Definition	Scattering		Annihilation
$\mathcal{O}_{ss}$	$\bar{\psi}\psi\bar{f}f$	SI	(1)	p-wave
$\mathcal{O}_{pp}$	$\bar{\psi}\gamma^5\psi\bar{f}\gamma^5 f$	SD	(q^2)	s-wave
$\mathcal{O}_{sp}$	$\bar{\psi}\psi\bar{f}i\gamma^5 f$	SD	(q)	p-wave
$\mathcal{O}_{ps}$	$\bar{\psi}i\gamma^5\psi\bar{f}f$	SI	(q)	s-wave
$\mathcal{O}_{vv}$	$\bar{\psi}\gamma^\mu\psi\bar{f}\gamma_\mu f$	SI	(1)	s-wave
$\mathcal{O}_{aa}$	$\bar{\psi}\gamma^\mu\gamma^5\psi\bar{f}\gamma_\mu\gamma^5 f$	SD	(1)	s-wave $\propto m_q^2/m_\chi^2$
$\mathcal{O}_{va}$	$\bar{\psi}\gamma^\mu\psi\bar{f}\gamma_\mu\gamma^5 f$	SD	(v)	s-wave
$\mathcal{O}_{av}$	$\bar{\psi}\gamma^\mu\gamma^5\psi\bar{f}\gamma_\mu f$	SD	(v)	p-wave
$\mathcal{O}_{tt}$	$\bar{\psi}\sigma^{\mu\nu}\psi\bar{f}\sigma_{\mu\nu}f$	SD	(1)	s-wave
$\mathcal{O}_{pt}$	$\bar{\psi}i\sigma^{\mu\nu}\gamma^5\psi\bar{f}\sigma_{\mu\nu}f$	SI	(q)	s-wave

Among the tensor operators, $\bar{\psi}\sigma^{\mu\nu}\psi\bar{f}i\sigma_{\mu\nu}\gamma^5 f$ is rewritable as $\bar{\psi}i\sigma^{\mu\nu}\gamma^5\psi\bar{f}\sigma_{\mu\nu}f \equiv \mathcal{O}_{pt}$ and $\bar{\psi}i\sigma^{\mu\nu}\gamma^5\psi\bar{f}i\sigma_{\mu\nu}\gamma^5 f$ as $\bar{\psi}\sigma^{\mu\nu}\psi\bar{f}\sigma_{\mu\nu}f \equiv \mathcal{O}_{tt}$. So they are not separately considered. Moreover, operators enter the Lagrangian along with respective Wilson coefficients (WC). In this work, we will consider that the WCs are absorbed into the Λ.

Using available data, we will determine the viability of each of these operators as a function of the DM mass m_χ and the EFT expansion scale Λ. Note that the operators, $\mathcal{O}_{ss}$, $\mathcal{O}_{vv}$, $\mathcal{O}_{aa}$ and $\mathcal{O}_{tt}$ present velocity-unsuppressed interactions with nucleon, while the other interactions are suppressed. On the other hand, operators marked as SI are spin independent (i.e., do not depend on the spin of the target SM fermion), and thus allow more sensitive searches that can take advantage of coherent enhancement. For a detailed discussion of these non-relativistic dependencies of DM, see Ref. [102]. The annihilation rate, $\langle\sigma v\rangle$, also depends on the velocity of DM. While s-wave annihilation is independent of velocity, p-wave annihilation is velocity-dependent and thus suppressed. As a result, $\mathcal{O}_{ss}$, $\mathcal{O}_{sp}$, and $\mathcal{O}_{av}$ lead to a velocity-suppressed annihilation rate. It is worth noting that, although annihilation for $\mathcal{O}_{aa}$ occurs through s-wave, it is suppressed by a factor proportional to m_q^2/m_χ^2.

2.2 Relic density and the asymmetry criterion

The asymmetric dark matter paradigm assumes that both DM (χ) and the anti-DM ($\bar{\chi}$) are present up to some era of the evolution of the Universe, as in the case of baryonic matter. Then by some mechanism, an asymmetry is generated between DM and anti-DM sectors, which gets frozen out with the departure from thermal equilibrium, and the symmetric part

gets annihilated out. In this study we remain agnostic about the asymmetry generation and sharing mechanisms.

The total dark-abundance can be written as the sum of the yields of DM and anti-DM, i.e., $Y_\chi + Y_{\overline{\chi}}$, where the yield $Y = n/s$ is defined as the number density scaled by the entropy density of the early universe. This, in turn can be expressed as a symmetric and an asymmetric component:

$$Y_{\rm tot} = Y_\chi + Y_{\overline{\chi}} = (Y_\chi - Y_{\overline{\chi}}) + 2Y_{\overline{\chi}} = Y_{\rm asy} + Y_{\rm sym}, \tag{2.3}$$

where, $Y_{\rm asy}$ is the asymmetric yield, and $Y_{\rm sym}$ is the yield of the symmetric component.

For the symmetric part, following [103], we take

$$Y_{\rm sym} = 2Y_{\overline{\chi}} = \frac{2Y_{\rm asy}}{\exp\left[Y_{\rm asy}\lambda\left(\frac{a}{x_F} + \frac{3b}{x_F^2}\right)\right] - 1}, \tag{2.4}$$

where a and b are coefficients in the partial wave expansion of $\langle\sigma v\rangle$, λ is a function of DM mass:

$$\lambda = \frac{4\pi}{\sqrt{90}} m_\chi M_{\rm pl}\sqrt{g_*}, \tag{2.5}$$

with reduced Planck mass $M_{\rm pl} = 2.4 \times 10^{18}$ GeV, and $g_* \sim 100$ being the number of available relativistic degrees of freedom at freeze-out. Also, the shifted coordinate x_F is given by [103]

$$x_{\rm F} = x_{\rm F_0}\left(1 + 0.285\frac{a\lambda Y_{\rm asy}}{x_{\rm F_0}^3} + 1.35\frac{b\lambda Y_{\rm asy}}{x_{\rm F_0}^4}\right), \tag{2.6}$$

with $x_{\rm F_0} = m_\chi/T_{\rm F}$ being the usual coordinate at freeze-out. The difference between $x_{\rm F}$ and $x_{\rm F_0}$ encodes the ADM correction to the $\bar{\chi}$ abundance, which is otherwise under-predicted by the standard WIMP treatment. The expressions of the annihilation cross-sections times velocity for each of the operators in Table 1 are given in Appendix A.

The constraint imposed in ADM model is that the symmetric part, as in Eq. 2.4, must be decidedly sub-leading to the asymmetric part. We implement this by restricting $Y_{\rm sym}$ to contribute $\leq 1\%$ (say) to $\Omega_{\rm DM}h^2$ [32]:

$$Y_{\rm sym} \leq \frac{1}{100} \times \frac{\Omega_{\rm DM}h^2}{2.76 \times 10^8}\left(\frac{\rm GeV}{m_\chi}\right). \tag{2.7}$$

This is to be achieved by annihilating out the symmetric component ($Y_{\rm sym} \to 0$) into SM fermions using the operators presented in Table 1. When there is a stronger interaction between DM and SM particles, it becomes easier to annihilate the symmetric part. However, it also makes more challenging to pass the constraints set by various experiments. For instance, at a specific value of the DM mass m_χ, a large Λ implies less effective scattering interactions (SI or SD) with quarks, but at the same time, it also indicates less efficient annihilation in the early universe. Consequently, the requirement for efficient annihilation sets an upper limit on Λ (say, $\Lambda < \Lambda_1$) via Eq. 2.7. In contrast, null results from DD

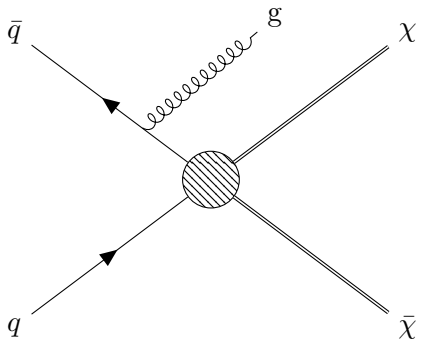


Figure 1: Representative Feynman diagram of mono-jet production along with DM.

experiments establish a lower limit ($\Lambda > \Lambda_2$). The space that lies in-between, if it exists (i.e., $\Lambda_1 > \Lambda_2$), defines the permissible range for ADM scenarios. However, if one finds $\Lambda_1 \leq \Lambda_2$ in a particular mass range, it rules out the feasibility of an ADM scenario within that region. Operators that lead to p-wave annihilation, such as $\mathcal{O}_{ss}$, $\mathcal{O}_{sp}$, and $\mathcal{O}_{av}$, require a lower value of Λ to satisfy the ADM requirement, resulting in stronger exclusions.

3 Constraining ADM-Quark Interactions

In this section, we first revisit the constraints on the DM-quark effective interactions and then find their implications on ADM scenario. The DM-quark interactions are mainly constrained by the DD experiments and the results of mono-jet searches at the LHC, where direct interaction of DM with quarks is present. These constraints were already shown in Ref. [32]. More extensive experimental results and several new searches have come up since, which are expected to impose stronger and more robust constraints. In this section, we update the bounds derived from mono-jet searches at LHC using latest data and by taking relevant detector effects into account. We then present the big picture for ADM-quark interactions, comparing with constraints from the asymmetry, and including other important constraints, especially those from recent DD experiments.

3.1 Mono-jet searches at hadron colliders

Any non-vanishing effective interactions between DM and quarks, listed in Table 1, should pair-produce DM through p-p collision at the LHC, if its mass is within the reach of LHC energy, $\sqrt{s} = 13$ TeV. However, production of solely DM particles does not yield any detectable signals within the detector, since they do not interact with the detector. Only if the DM particles are accompanied by some visible particles, then the imbalance of the transverse momentum (p_T) can be measured by detecting the visible particles, such as mono-jet, where one extra jet is produced along with the DM (see Fig. 1 for a representative Feynman diagram), and results in signals of energetic jets with missing p_T ($\not{E}_T$),

$$pp \to \chi\bar{\chi}j \to \not{E}_T j. \tag{3.1}$$

At the LHC Run-II experiment, both ATLAS and CMS experiments have reported their results of mono-jet searches [104, 105]. Ref. [104] looks for jets with high p_T with

Table 2: 95% CL upper limits on the visible mono-jet production cross-section at the LHC ($\sqrt{s} = 13$ TeV) [104].

Selection criteria	σ^{95}(fb)
$p_T^j > 200$ GeV	736
$p_T^j > 1200$ GeV	0.3

different selection criteria on p_T of the jet. Since the mono-jet cross-section drops with the p_T of jet, a lower p_T selection includes a larger fraction of events in the low p_T region, but fewer events in the high p_T region. Contrarily, a higher p_T cut includes the fraction of events from the higher p_T phase space only. As the scaling of the cross-section of SM backgrounds with p_T may not necessarily be identical to the DM signal, mono-jet analyses with different p_T cuts result in different levels of exclusion on the DM-quark interactions. In order to interpret results of mono-jet searches within the framework of effective interactions considered here, we closely follow analysis strategy described in Ref. [104], as shown in Table 2.

The hard-scattering matrix elements are generated in `MG5aMC_atNLO-3.3.0` [106] using `Feynrules` UFO [107] package, which is required to include the effective interactions for a given set of parameters (Λ, m_χ). Events are produced up to one jet in Madgraph, and MLM-matching is used to avoid double counting, which is essential in reproducing a mono-jet analysis. The showering and hadronization are performed using `PYTHIA8` [108, 109]. Unlike the previous study [32], we consider the ATLAS detector effects using `Delphes` [110] ATLAS-card. After applying specific cuts on p_T^j (see Table 2), we can determine the acceptance rate of the signal. It is obtained by dividing the number of events that meet the criteria on p_T^j (N) by the total number of simulated events (N_0), i.e., $\mathcal{A} = N/N_0$. However, it is important to note that various detector efficiencies (ϵ) are already applied during the simulation, and their impacts are already included in the value of N. Therefore, for a given parameter space, taking into account the effects of DM-EFT, we calculate the measurable cross-section, $\sigma_{\rm mes} = \sigma \times \mathcal{A}$. If $\sigma_{\rm mes}$ is greater than the experimentally determined 95% upper limit, then we claim that the given choice of $\{\Lambda, m_\chi\}$ is excluded. Repeating this procedure for various values of $\{\Lambda, m_\chi\}$, finally we obtain the excluded region in the $\Lambda - m_\chi$ plane due to the mono-jet searches, as shown in Figs. 2-6.

3.2 Direct detection experiments

DM can interact with the nucleon via the effective interactions listed in the Table 1. The strength $(1/\Lambda)$ of the interactions affect the DM-nucleon scattering cross-section. Thus, the bounds on the DM-nucleon scattering cross-section can be recast as constraints on these interactions in terms of Λ and mass of DM (m_χ). Notice that, the spin-dependence of the scattering cross-section significantly impacts the bounds placed on each effective operator in Table 1. Some of the interactions get suppressed by the powers of DM velocity (v). For

example, the vector current singles out the temporal component of a spinor and the axial-vector current picks up the spatial component. Consequently, the combinations of these two ($\mathcal{O}_{av}$ or $\mathcal{O}_{va}$) are velocity suppressed, and eventually leading to weaker bounds from the DD experiments. On the other hand, spin-independent DM-nucleon scattering cross-section often turn out to be restricted severely. Considering universal coupling to quarks, the scattering cross-section for each of the non-suppressed spin-independent interactions listed in Table 1 are given by [32],

$$\sigma_{\rm SI}^{\mathcal{O}_{ss}} \sim \frac{1}{\pi\Lambda^4}\mu_p^2 f_p^2, \tag{3.2}$$

$$\sigma_{\rm SI}^{\mathcal{O}_{vv}} \sim \frac{9}{\pi\Lambda^4}\mu_p^2, \tag{3.3}$$

where, $\mu_p = m_\chi m_p/(m_\chi + m_p)$ is the reduced mass of the nucleon-DM system and f_p is the DM effective coupling to protons. Here we use $f_p = 0.3$ [111]. The non-suppressed scattering cross-section for the spin-dependent operators listed in Table 1 is [32]:

$$\sigma_{\rm SD}^{\mathcal{O}_{tt}} \sim 4 \times \sigma_{\rm SD}^{\mathcal{O}_{aa}} \sim \frac{16}{\pi\Lambda^4}\mu_p^2 \left(\sum_q \Delta_q^p\right)^2 , \tag{3.4}$$

where Δ_q^p accounts for the spin content of the nucleon and $(\sum_q \Delta_q^p)^2 \simeq 0.32$ [112].

The LUX-ZEPLIN (LZ) experiment currently provides the most stringent constraint on the cross-section for SI scattering between DM and nucleons. Specifically, they reported a limit of 6.5×10^{-48} cm^2 on $\sigma_{\rm SI}$ at 90% CL for 19 GeV DM mass [113]. In parallel, the DarkSide-50 experiment excluded a cross-section of 10^{-41} cm^2 for DM-nucleon SI interactions with a 90% CL, for a DM particle mass of 1.8 GeV [114]. The constraints on the SD scattering cross-sections of dark matter with protons and neutrons have also improved for large ranges of DM particle masses. Notably, the most stringent limits on DM-neutron interactions originate from the XENON-nT experiment [115], having a minimum of 6.3×10^{-42}cm^2 for DM mass 30 GeV at 90% CL. Whereas, for DM-proton interactions the PICO-60 experiment [116] sets the strongest constraints, e.g., 2.5×10^{-41}cm^2 for DM mass 25 GeV at 90% CL.

To interpret these findings in the EFT context, we have converted the limits on DM-nucleon scattering cross-sections to the plane defined by the dark matter particle mass (m_χ) and the EFT scale (Λ), using Eqns.. (3.3) and (3.4). Additionally, for the sake of comparison against the constraints outlined in Ref. [32], we have included the bounds predicted by the XENON-100 experiment [117]. These comparisons are illustrated in Figs. 2 and 3.

3.3 Constraints on the $\Lambda - m_\chi$ plane

In Figs. 2-6, the exclusion on Λ as a function of ADM mass are shown along with the results due to various DD experiments, such as LZ, Xenon, Darkside. The bounds from the mono-jet searches are presented for two jet-p_T selection criteria, $p_T(j) > 200$ GeV (LHC-200) and $p_T(j) > 1200$ GeV (LHC-1200). The current DD exclusions are also presented in the same plot to gather an idea of the overall picture of the exclusions.

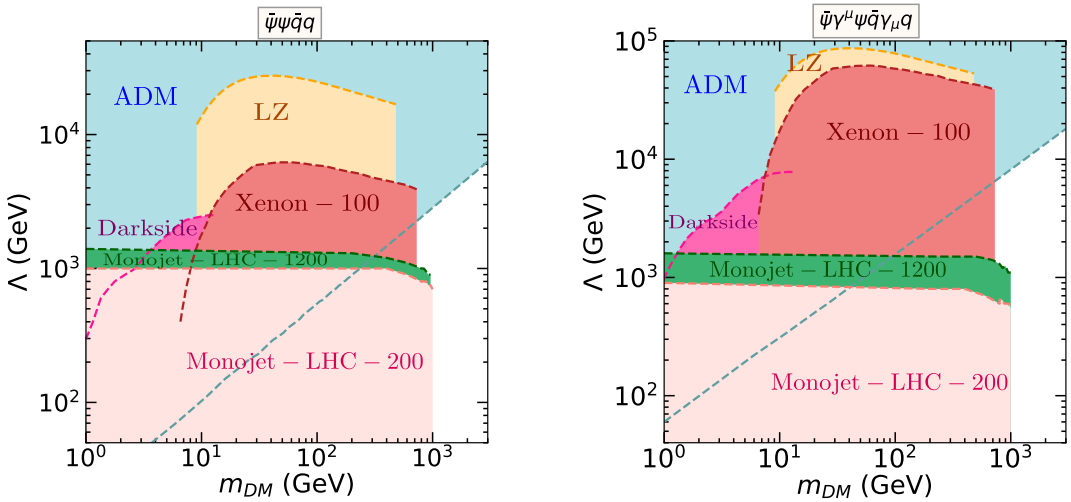


Figure 2: Current exclusion limits on the scale Λ at different DM mass (m_{DM}) for the scalar ($\mathcal{O}_{ss} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{q}q$) (left) and vector ($\mathcal{O}_{vv} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^{\mu}\psi\bar{q}\gamma_{\mu}q$) (right) type interaction of DM with quarks from different experimental observations and ADM considerations, labeled on the respective regions with darker shades. The experimental exclusions extends up to the bottom of the plots and overlapping regions are implicit.

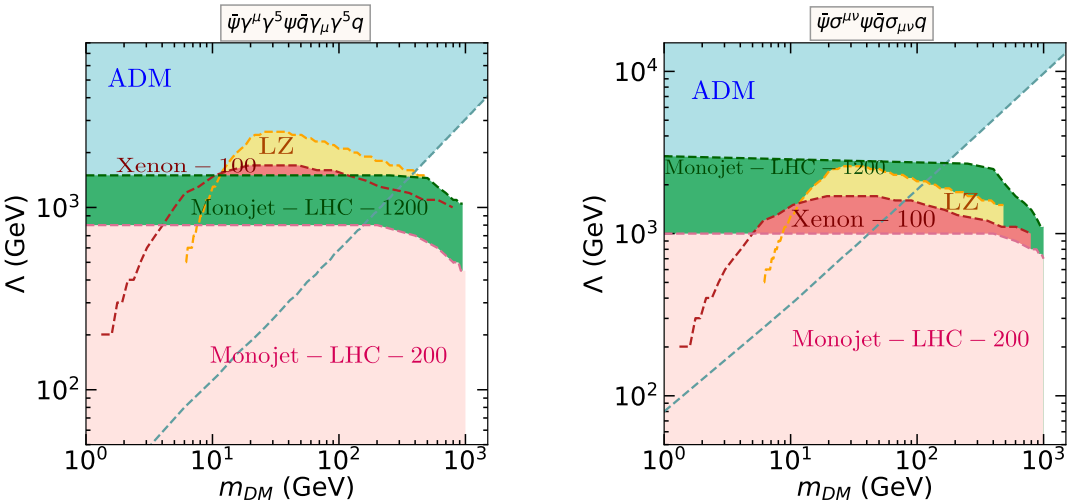


Figure 3: Same as Fig. 2 but for axial-vector ($\mathcal{O}_{aa} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^{\mu}\gamma^{5}\psi\bar{q}\gamma_{\mu}\gamma^{5}q$) (left) and tensor ($\mathcal{O}_{tt} \equiv \frac{1}{\Lambda^2}\bar{\psi}\sigma^{\mu\nu}\psi\bar{q}\sigma_{\mu\nu}q$) (right) type interaction of DM with quarks.

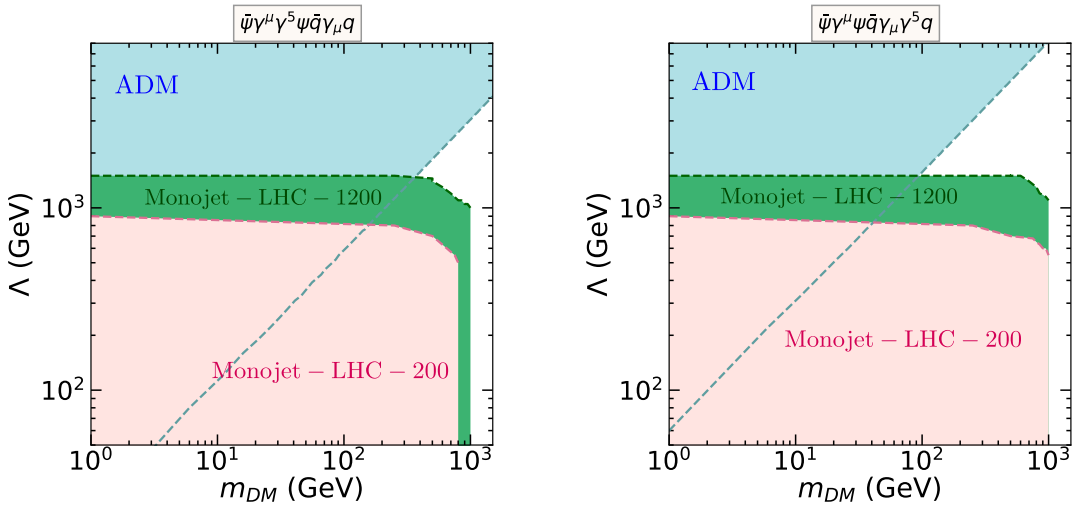


Figure 4: Same as Fig. 2 but for axial-vector−vector ($\mathcal{O}_{av} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\gamma^5\psi\bar{q}\gamma_\mu q$)(left) and vector−axial-vector ($\mathcal{O}_{va} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\psi\bar{q}\gamma_\mu\gamma^5 q$) (right) type interaction of DM with quarks.

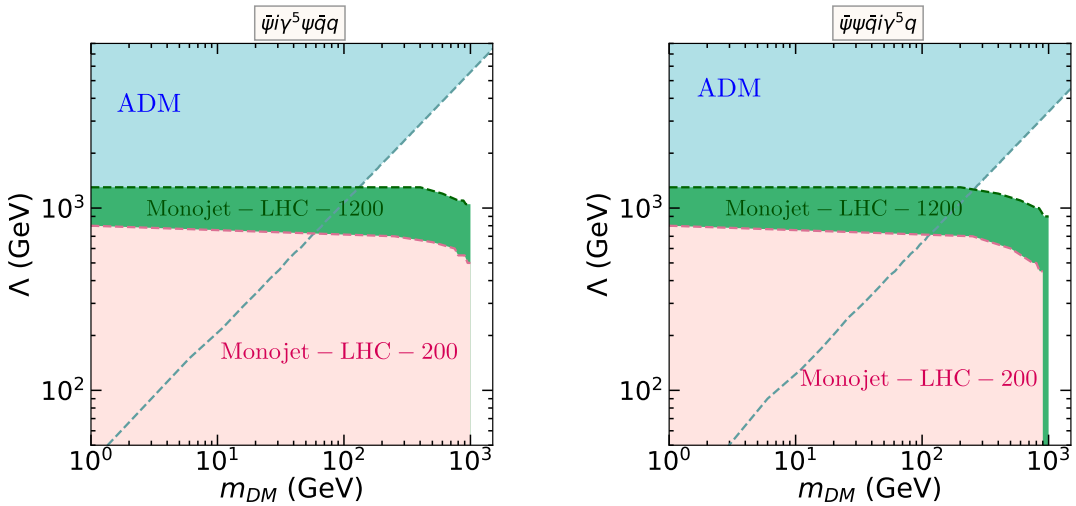


Figure 5: Same as Fig. 2 but for pseudo-scalar−scalar ($\mathcal{O}_{ps} \equiv \frac{1}{\Lambda^2}\bar{\psi}i\gamma^5\psi\bar{q}q$) (left) and scalar−pseudo-scalar ($\mathcal{O}_{sp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{q}i\gamma^5 q$) (right) type interaction of DM with quarks.

Clearly, the current DD limits are found to be much stronger compared to those presented in the Ref. [32]. For large part of the mass region, the DD bounds dominate the exclusion on the DM-quark interactions. However, in low DM-mass regions, even in case of the non-suppressed interactions, we found that the bounds coming from the mono-jet searches supersede the DD limits. Especially for the $\mathcal{O}_{aa}$ and $\mathcal{O}_{tt}$ operator, the current limits exclude a good fraction of region. Indeed, the DD measurements are found to be not sensitive for operators with suppressed interactions. In these cases the mono-jet searches put the dominant constraints for these interactions (see Figs. 4-6).

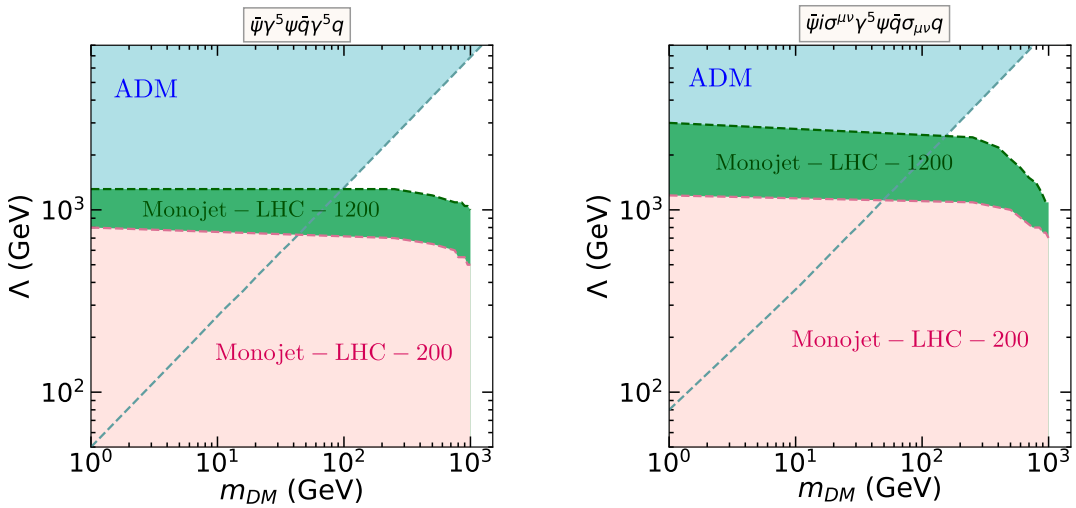


Figure 6: Same as Fig. 2 but for pseudo-scalar−pseudo-scalar ($\mathcal{O}_{pp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^5\psi\bar{q}\gamma^5 q$) (left) and pseudo-tensor−tensor ($\mathcal{O}_{pt} \equiv \frac{1}{\Lambda^2}\bar{\psi}i\sigma^{\mu\nu}\gamma^5\psi\bar{q}\sigma_{\mu\nu}q$) (right) type interaction of DM with quarks.

4 Constraining ADM-Lepton Interactions

In this section we consider the case of leptophilic DM. Note in these models DM-quark interactions occur only at the loop level. Consequently, the limits obtained based on the mono-jet searches at hadron colliders and nuclear recoil dependent DD experiments become much weaker in this case. Moreover, the LHC being a hadron collider, the production of DM, which is not coupled to quarks or gluons, is suppressed. On the other hand, leptophilic DM attracts bounds from monophoton production at lepton colliders, such as LEP, or may be searched at future lepton colliders, e.g., ILC, FCC-ee. Interestingly, studies of effects of DM capture in compact stars help constrain DM-lepton interactions strongly, which we will discuss in Section 4.3.

4.1 Mono-photon searches at LEP

Similar to the mono-jet production in association with DM at the LHC, in an electron-position collider, e.g., at LEP, DM can be produced accompanied by a photon radiated off from the electron or positron (see Fig. 7 (left)), namely,

$$e^+e^- \to \chi\bar{\chi}\gamma \to \not{E}_T\gamma. \tag{4.1}$$

The SM counterpart of it involves production of neutrinos in place of DM (see Fig. 7 (right)), such as,

$$e^+e^- \to \nu\bar{\nu}\gamma \to \not{E}_T\gamma. \tag{4.2}$$

These processes give rise to a mono-photon signal at the detector, a potentially robust way to look for DM in e^+e^- collider. In Ref. [92], the authors studied the mono-photon

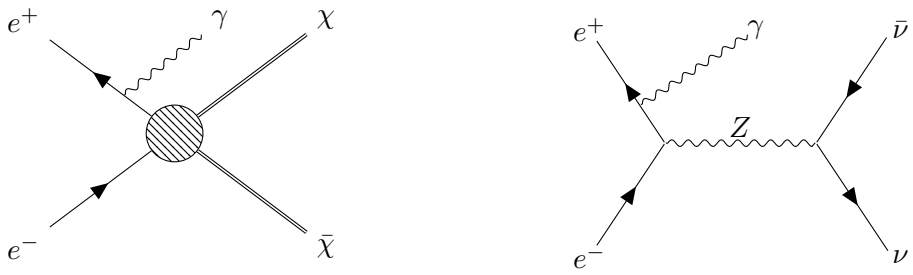


Figure 7: (Left) Feynman diagram of DM production from e^+e^- through effective interaction with additional photon from ISR; (right) Feynman diagram for the SM mono-photon production process.

production in the e^+e^- collider and predicted the excluded region in the $\Lambda - m_\chi$ plane for four effective operators (three of which are independent) of the DM-lepton interactions, considering the mono-photon analysis of the DELPHI experiment at LEP [118], including a simulation of the effects of the DELPHI detector.

In this section, we revisit this LEP exclusions including all possible DM-lepton contact interactions (same as Table 1, with $f = e$). Moreover, while presenting these LEP exclusions, we also include the limits from DD or indirect detection experiments, to review the overall picture of the current limits on leptophilic DM. Using these limits, we also infer the exclusions for the case of ADMs. This approach permits us to delineate the parameter spaces for viable ADM scenarios, similar to the analysis in the previous section.

In the following, we document our methodology to find the exclusions due to the LEP data. To start with, we performed a Monte-Carlo (MC) simulation of the SM mono-photon process, as in Eq. (4.2). This simulation is tailored to closely replicate the experimental data. We factor in all the relevant aspects such as detector efficiencies, resolutions, and effects of smearing. Then we validated our MC simulation with the experimental data. Post-validation, we proceeded to employ the same configuration to generate events associated with dark matter for a wide range of parameter values, Λ and m_χ. Subsequently, we evaluated the degree of excess that could potentially manifest at LEP as a result of these dark matter events. Finally, this allows us to exclude regions in the Λ-m_χ plane.

We closely follow the original LEP analysis Ref. [118], describing the measurements of the mono-photon events with the DELPHI detector corresponding to luminosity $\mathcal{L} = 650\ \text{pb}^{-1}$. These measurements are presented in bins of,

$$X_\gamma = \frac{E_\gamma}{E_{\text{beam}}}, \tag{4.3}$$

where E_γ and E_{beam} are the energies of the photon and the colliding beams respectively. The DM events are generated using Madgraph5-aMC@NLO-3.3.0[106] including the ISR photon, using a UFO model file for leptophilic DM produced using the Feynrules package [107]. Finally, mono-photon events are passed through PYTHIA8 [108, 109] for showering. It is important to emphasize here that accurate measurements of photons arising from initial state radiation (ISR) is crucial to model a mono-photon process. In our analysis, we ensure that all final state photons pass the requirements for identification with

Table 3: Efficiencies and energy resolutions of various components of the DELPHI detector. Trigger and acceptance efficiencies increase linearly, with specified values at the given energies. Energy unis are in GeV.

Detector	Angular coverage	Overall efficiency	Trigger efficiency	Acceptance efficiency	Angular requirement	Energy resolution
HPC	$45^\circ < \theta < 135^\circ$	$X_\gamma > 0.6$	52% at $E_\gamma = 6$; 72% at $E_\gamma = 30$; 84% at $E_\gamma = 100$ and above	41% at $E_\gamma = 6$; 78% at $E_\gamma = 80$, and above	-	$0.043 \oplus 0.32/\sqrt{E}$
FEMC	$12^\circ < \theta < 32^\circ$, or $(148^\circ < \theta < 168^\circ)$	$X_\gamma > 0.1$	93% at $E_\gamma = 10$; 100% at $E_\gamma = 15$; and above	57% at $E_\gamma = 10$; 75% at $E_\gamma = 100$; $\times$ const. factor 89%	$\theta > 28^\circ - 80^\circ X_\gamma$	$0.03 \oplus 0.12/\sqrt{E} \oplus 0.11/E$
STIC	$3.8^\circ < \theta < 8^\circ$, or $(172^\circ < \theta < 176.2^\circ)$	$X_\gamma > 0.3$	Const. 48% all over	-	$\theta > 9.2^\circ - 9^\circ X_\gamma$	$0.0152 \oplus 0.135/\sqrt{E}$

the DELPHI detector. For completeness and clarity, we provide a brief overview of the detector efficiencies and resolutions in the following.

The DELPHI detector had three parts, namely, High-Density Projection Chamber (HPC), Forward Electromagnetic Calorimeter (FEMC), and Small Angle Tile Calorimeter (STIC). We present the energy resolutions and different efficiencies of various sub-detector in Table 3. There is an overall efficiency for each sub-detector components which depends on X_γ, followed by energy-dependent trigger and acceptance efficiencies. The trigger and acceptance efficiencies increase linearly with energy. The slope of the increase can be obtained using a few fixed values at certain energies, which are presented in Table 3. To reduce the large background of radiative Bhabha events in FEMC, and due to beam-gas interactions in STIC, an energy-dependent angular requirement is imposed. The energy resolution adds an energy-dependent smearing for each of the photon that have passed all the selection criteria. Following Ref. [92], we have added a Lorentzian smearing with a width of $0.052E_\gamma$ GeV, which is necessary to obtain a more realistic distribution.

In order to demonstrate whether our simulated data correctly mimics the LEP measurements, we compare the p-value of X_γ distribution corresponding to the SM mono-photon production (Eq. 4.2) with the DELPHI measurements. We obtain the data-MC agreement for X_γ as shown in Fig. 8. Some erroneous data can be found in the last bin, most likely due to inaccurate detector resolution modeling. We remove this bin from the χ^2-analysis. The χ^2/dof is $\sim 14/19$ for 19 d.o.f corresponding to yields in 19 bins. A good agreement is found, confirming our simulation methodology. In the same figure, the distribution for the DM signal is also plotted considering $m_\chi = 10$ GeV and $\Lambda = 300$ GeV, where the lower X_γ bins show deviations due to DM effects.

Now we use this setup to generate the mono-photon energy distribution for the mono-photon production for DM (Eq. 4.1) for each choice of m_χ and Λ. Calculation of χ^2 between the photon energy distributions for SM-MC and the DM, it is possible to predict the allowed and ruled out DM parameter space. Results are shown for different operators in Figs. 10-14. Notice that, these bounds are the strongest one for leptophilic dark matter in the 1-100 GeV mass range, which are discussed later.

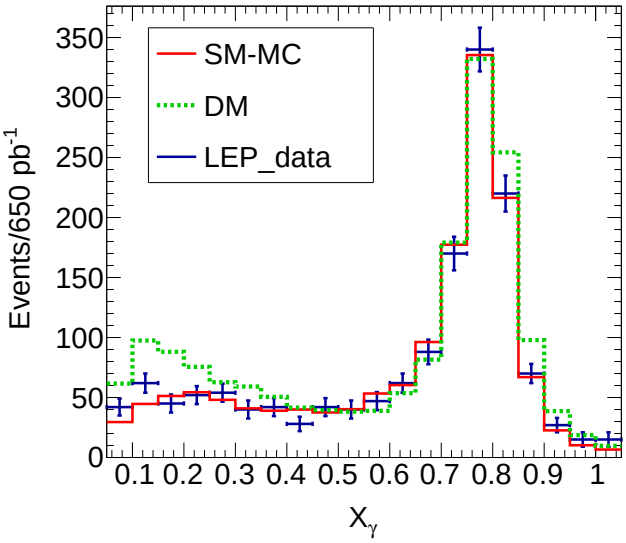


Figure 8: Distribution of $X_\gamma = E_\gamma/E_{\rm beam}$ for single photon event, where the agreement of the SM Monte-Carlo (red histogram) with the DELPHI-data (blue points with error bar) can be observed. The DM signal (green histogram), with vector-like interaction with $\Lambda = 300$ GeV and $m_\chi = 10$ GeV, shows excess for the lower values of X_γ and otherwise consistent with the SM.

4.2 Discovery potential of ADM at FCC-ee

In this section we discuss the discovery potential of leptophilic DM in the FCC-ee (Future Circular Collider - electron-positron) corresponding to the allowed region of parameter space. The FCC-ee is a proposed e^+e^- collider with relatively higher center of mass (COM) energies and luminosities, measuring observables at unprecedented precision [119]. The FCC-ee collider is proposed to operate in several stages, with different COM energies, such as $\sqrt{s} = 91, 160, 240$ and 365 GeV. In our study, we have used the highest energy option ($\sqrt{s} = 365$ GeV) to calculate the discovery reach of ADM scenarios. It is to be noted that, this sensitivity study is applicable to any leptophilic DM with contact interactions of Table 1, and not limited to ADM.

The FCC-ee experiment should produce mono-photon events similar to LEP. However, the new detector set-up with increased COM energy is expected to probe wider ranges of parameters (Λ, m_χ).

The production cross-sections of both the signal and backgrounds depend highly on the polarization of the e^+ and e^- beams. For the sake of demonstration, we compute the cross-section of the SM mono-photon background using Madgraph5-aMC@NLO-3.4.1[106], varying the polarization of the beams. The variation of the cross-section with different choices of beam polarizations are presented in Fig. 9 (left). As the beam polarization setting of FCC-ee is not fixed yet, we used an unpolarized (50%-L and 50%-R combination) beam for representative purposes.

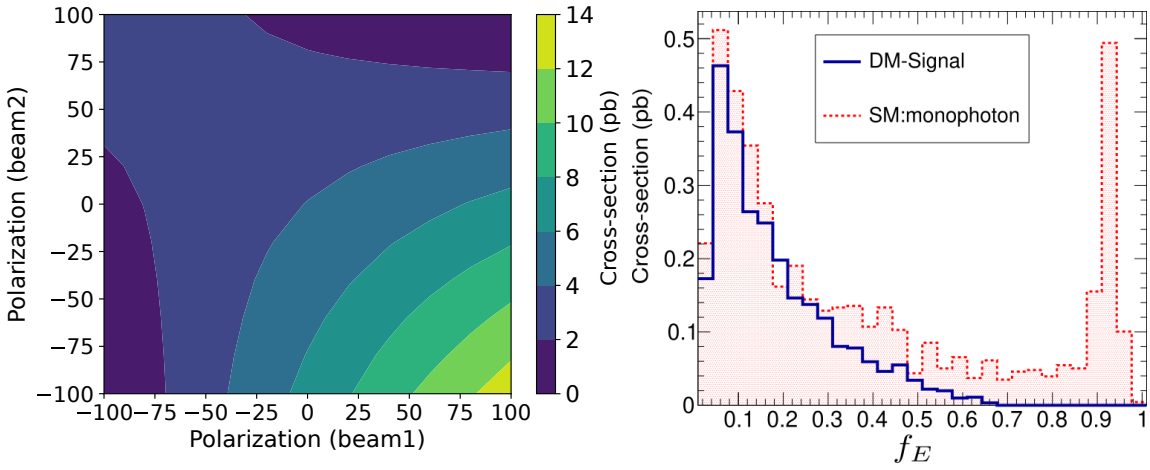


Figure 9: (Left) Variation of the SM mono-photon production cross-section ($e^+e^- \to \nu\nu\gamma$) at FCC-ee with variation of beam-polarization; (right) distribution of the energy fraction of photon in the mono-photon events, for DM-signal (choosing $\Lambda = 400$ GeV, $m_\chi = 100$ GeV, and vector-like interaction) and SM-background.

For the signal, we essentially have one detectable object, namely the photon in the final state. All other kinematic variables such as $\not{E}_{\rm T}$ and the transverse mass, $M_T(\gamma, \not{E}_T)$, are strongly correlated to the energy of the photon. Hence a selection for photon energy results in the effective discrimination between signal and background. Hence, a selection as

$$f_E = E_\gamma / E_{\rm beam} < 0.3, \tag{4.4}$$

is found to be very powerful to suppress backgrounds as shown in Fig. 9 (right panel). The resulting signal significances, $S/\sqrt{S+B}$, are estimated at $\mathcal{L} = 340$ fb^{-1} for each of the given parameter spaces $\{m_\chi, \Lambda\}$. Repeating this analysis for different choices of $\{m_\chi, \Lambda\}$, we present the region for which $S/\sqrt{S+B} > 3.0$, i.e., the 3-σ reach. The results are presented in the following subsection.

4.3 Constraints from compact stars and direct detection

As seen in the previous sections, lepton-colliders are sensitive to a relatively smaller range of leptophilic ADM masses. Comparatively, DM capture in compact objects can be used to set bounds on a wider range of DM mass [48, 120–125]. Especially, neutron stars (NS) [90, 91, 124, 126–133] and white dwarfs (WD) [91, 123, 134–137] constrains DM-lepton interactions strongly for a MeV to TeV range of DM mass. Due to their high density, NS can capture dark matter very efficiently and can lead to the heating of NS [128, 132]. Since old, isolated NS can naturally cool to temperatures below 1000 K, this heating can be utilized to establish constraints on the strength of DM interactions. For instance, Refs. [90, 91] have calculated these limits in detail for DM-electron interactions.

A comparable scenario can arise with white dwarfs (WDs) as well. The core of a WD is primarily composed of a degenerate electron gas. Consequently, if there is interaction

between DM and electrons, it would likely be more pronounced at the core of the WD. The DM can then scatter of the electrons and can heat up the WD. Also, a DM-dense WD core may result in DM annihilation, another source of rising the temperature of WD. The presence of DM would then prevent the natural cooling of the old, isolated WDs. Conversely, the absence of anomalous cooling in WDs can lead to stringent constraints on the strength of the interaction between DM and electrons. Following this, Ref. [91] calculated the limits on DM-lepton interactions from WDs.

On the other hand, among different DD experiments, Damic [76, 138] and SuperCDMS [139] put the strongest constraint on DM-lepton interactions for MeV to GeV range of DM mass. However, these DD bounds are generally not competitive in the preferred mass range for ADM. Nevertheless, the upcoming DAMIC-M-1kg-yr experiment is expected to have a promising sensitivity [138], and we consider this projected constraint for illustrative purposes. Finally, we present an overall picture of the constraints on DM-lepton interactions from all possible sources.

4.4 Constraints and discovery potential for leptophilic ADM

In this subsection, we discuss the constraints on the ADM-lepton interactions imposed by the experimental measurements, as discussed in section 4.1. In order to achieve a comprehensive perspective on the global exclusions associated with these interactions, we also incorporate important constraints originating from both indirect and direct searches. These additional constraints, gathered from Refs. [90, 91], are presented alongside the new collider-based bounds. It is worth noting that the constraints arising from dark matter-nucleon scattering experiments are relatively weak. This is because these interactions can only occur at the loop-level. Conversely, experiments focusing on dark matter-electron scattering probe a range of much lower dark matter masses, a region that is not highly motivated by the ADM hypothesis.

We present the overall picture of the allowed and excluded parameter spaces for DM-electron interaction in Figs. 10 -14. Each of the shaded regions within the presented figures signify exclusion due to specific hypotheses. For instance, the blue shaded region eliminates the possibility of dark matter being asymmetric, while the dark red portion is excluded based on the outcomes of the LEP mono-photon searches, and so on. Interestingly, for $\mathcal{O}_{ss}, \mathcal{O}_{sp}, \mathcal{O}_{aa}, \mathcal{O}_{av}$, where $\chi\bar{\chi} \to e^+e^-$ occurs through velocity suppressed p-wave (s-wave annihilation is either absent or has m_q^2 suppression), the ADM requirement excludes a larger parameter space in the $\Lambda - m_\chi$ plane. It is important to note here that the other exclusions, which are not restricted to the ADM hypothesis, hold for any DM with these effective interactions, and all the regions above the shaded region (except blue) are allowed. Our study reveals that, though the constraints from the NS or WDs on the DM-lepton interactions are overall strong, the bounds coming from the mono-photon searches at the LEP experiments surpasses all those in the sensitive region, $m_\chi \sim 1-100$ GeV. Considering the ADM hypothesis, we find that most of the operators fail to provide an ADM solution for any value of Λ in the region $m_\chi \sim 1-100$ GeV, sensitive to the LEP experiment. Only for operators $\mathcal{O}_{vv}$, $\mathcal{O}_{av}$, and $\mathcal{O}_{ps}$, the bounds are somewhat weak ~ 50 GeV, while for $\mathcal{O}_{va}$, $\mathcal{O}_{pt}$, $\mathcal{O}_{sp}$, and $\mathcal{O}_{pp}$ they are much weaker $\sim$10 GeV. Notice that, for the interactions $\mathcal{O}_{ss}$,

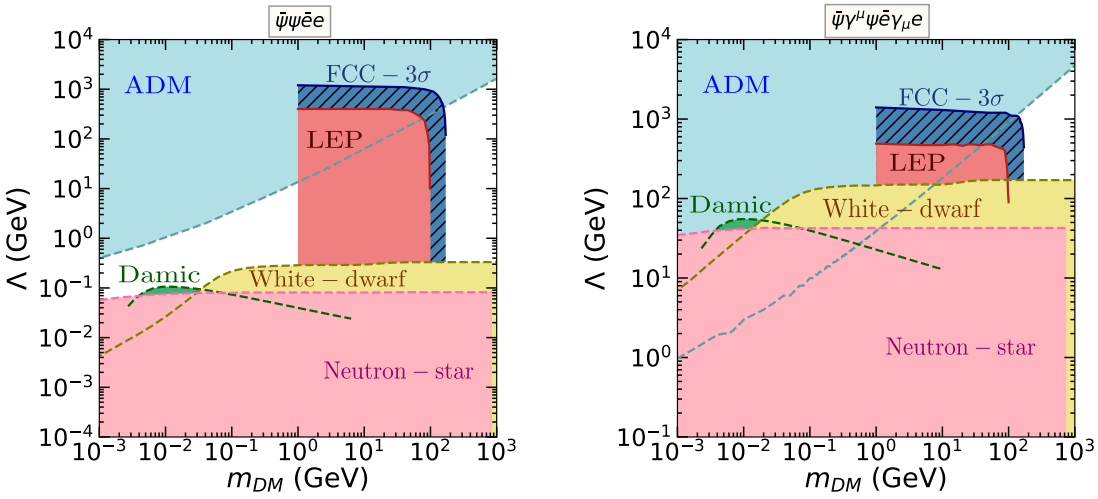


Figure 10: Limits on Λ_e for scalar ($\mathcal{O}_{ss} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{e}e$)(left) and vector ($\mathcal{O}_{vv} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^{\mu}\psi\bar{e}\gamma_{\mu}e$) (right) type interaction of DM with electrons from different experimental observations and ADM scenario, labeled on the respective regions with darker shades. The striped dark blue region corresponds to the FCC-3σ reach. The experimental exclusions/reach extends up to the bottom of the plots and overlapping regions are implicit.

$\mathcal{O}_{sp}$, $\mathcal{O}_{ps}$, and $\mathcal{O}_{pp}$ the LEP bounds are the only ones that can completely exclude ADM possibility of a certain mass.

The blue striped region in Figs. 10 -14 presents the 3-σ reach of the upcoming FCC-ee experiment in the mono-photon search channel in the $m_{\chi} - \Lambda$ plane. The pattern of the probing parameter region is similar to the LEP mono-photon bounds, as expected. However, owing to higher COM energy, the reach extends to $m_{\chi} \sim 175$ GeV and constraints may strengthen by a factor of 2 or so.

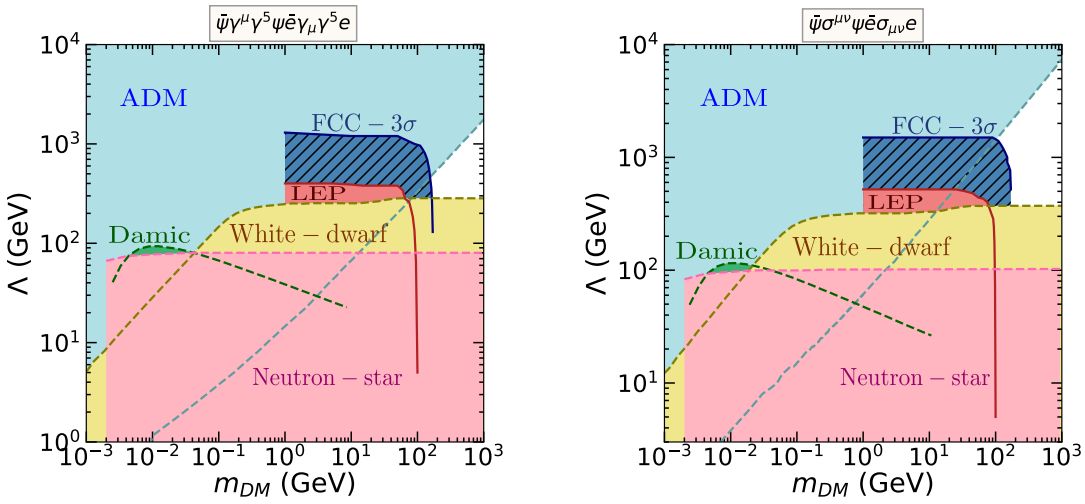


Figure 11: Same as Fig. 10 but for axial-vector ($\mathcal{O}_{aa} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\gamma^5\psi\bar{e}\gamma_\mu\gamma^5 e$) (left) and tensor ($\mathcal{O}_{tt} \equiv \frac{1}{\Lambda^2}\bar{\psi}\sigma^{\mu\nu}\psi\bar{e}\sigma_{\mu\nu}e$) (right) type interaction of DM with electrons.

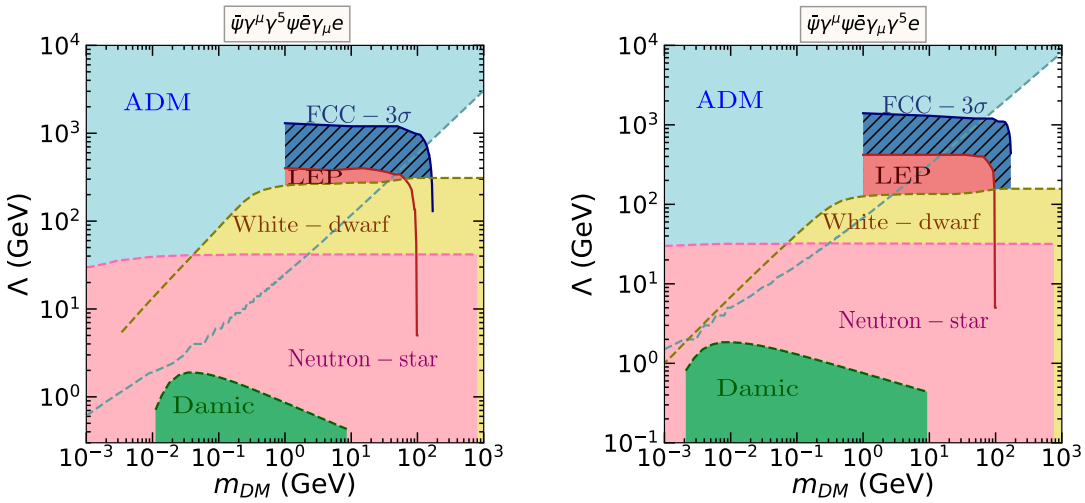


Figure 12: Same as Fig. 10 but for axial-vector−vector ($\mathcal{O}_{av} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\gamma^5\psi\bar{e}\gamma_\mu e$) (left) and vector−axial-vector ($\mathcal{O}_{va} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\psi\bar{e}\gamma_\mu\gamma^5 e$) (right) type interaction of DM with electrons.

5 Summary and Outlook

In this paper, we present a detailed study of the current status of the effective interactions of ADM with quarks and leptons, with a future projection of the sensitivity of FCC-ee in probing leptophilic ADM. We start with the ADM-quark interactions, where the DD constraints often become most stringent. Considering the strongest exclusions from different DD experiments, we find the excluded and allowed region corresponding to each interaction in the $\Lambda - m_\chi$ plane. However, for several effective interactions, the DM-nucleon scattering cross-sections are suppressed. In these cases, the constraints from monojet

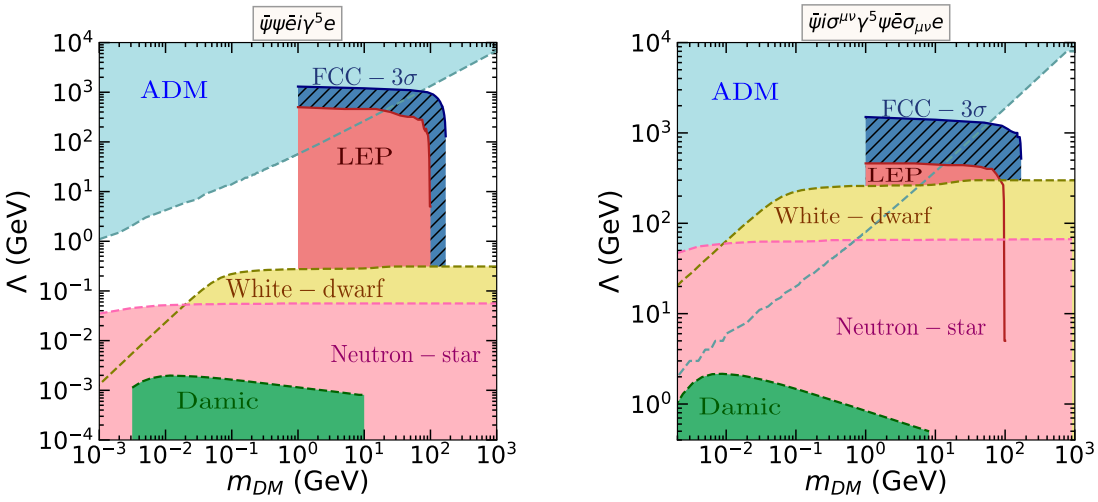


Figure 13: Same as Fig. 10 but for scalar−pseudo-scalar ($\mathcal{O}_{sp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{e}i\gamma^5 e$) (left) and pseudo-tensor−tensor ($\mathcal{O}_{pt} \equiv \frac{1}{\Lambda^2}\bar{\psi}i\sigma^{\mu\nu}\gamma^5\psi\bar{e}\sigma_{\mu\nu}e$) (right) type interaction of DM with electrons.

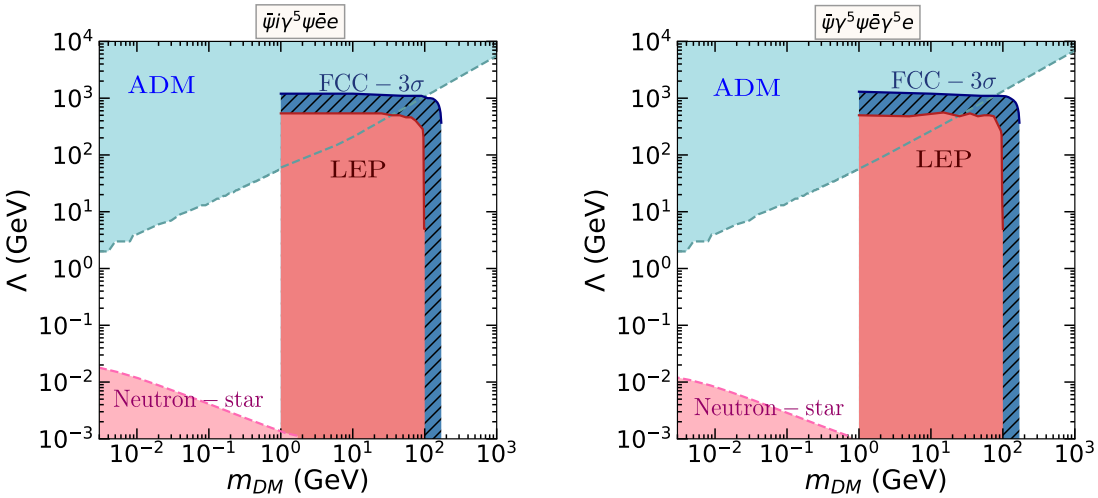


Figure 14: Same as Fig. 10 but for pseudo-scalar−scalar ($\mathcal{O}_{ps} \equiv \frac{1}{\Lambda^2}\bar{\psi}i\gamma^5\psi\bar{e}e$) (left) and pseudo-scalar−pseudo-scalar ($\mathcal{O}_{pp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^5\psi\bar{e}\gamma^5 e$) (right) type interaction of DM with electrons.

searches are dominant. We calculate the exclusions from the monojet searches using the most recent LHC measurements and including detector effects. We find that, for scalar ($\mathcal{O}_{ss} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{e}e$) and vector ($\mathcal{O}_{vv} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^\mu\psi\bar{e}\gamma_\mu e$) type of interactions, the possibility of ADM is ruled out almost up to ~ 1 TeV. Whereas, for other types of interactions, ADM of mass larger than a few hundred GeV are still allowed in a narrow range of Λ. None of the searches exclude the viability of ADM above ~ 1 TeV. Note that the exclusion regions from the experimental searches are not restricted to the ADM but are also applicable to any DM with the interactions in Table 1.

Next, we concentrate on the ADM-lepton interactions. In this case, the strongest bounds generally come from NS or WD studies. However, we found that the constraints from the monophoton searches at the LEP experiment, which we have calculated in detail for each interaction, overshoot those from the NS or WD for all types of interactions in the 1-100 GeV mass range. According to our study, the possibility of a leptophilic ADM is ruled out by the LEP experiment for $\sim$ 1-100 GeV mass range of the ADM, except for the vector$-$axial-vector operator ($\mathcal{O}_{va} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^{\mu}\psi\bar{e}\gamma_{\mu}\gamma^5 e$), the pseudo-tensor$-$tensor operator ($\mathcal{O}_{pt} \equiv \frac{1}{\Lambda^2}\bar{\psi}\sigma^{\mu\nu}i\gamma^5\psi\bar{e}\sigma_{\mu\nu}e$), the scalar$-$pseudo-scalar operator ($\mathcal{O}_{sp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\psi\bar{e}i\gamma^5 e$), and the pseudo-scalar$-$pseudo-scalar operator ($\mathcal{O}_{pp} \equiv \frac{1}{\Lambda^2}\bar{\psi}\gamma^5\psi\bar{e}\gamma^5 e$) mediated interactions, where the constraints are loose $\sim \mathcal{O}(10)$ GeV. Again, the constraints apply to any DM having the considered effective interactions (Table 1).

As a possible successor of e^+e^- colliders, the FCC-ee experiment will be interesting to probe the still viable regions of the ADM-lepton interactions. In our analysis, we calculate the region of parameter spaces where the monophoton searches at the FCC-ee will obtain the 3-σ sensitivity. This study indicates that FCC-ee can probe up to $\sim$ 200 GeV mass range of any general kind of DM for EFT expansion scale up to $\sim$1 TeV, which, though not a significant improvement, seems attractive from the aspect of possible FCC-ee physics goals.

Acknowledgements

The authors are thankful to Dr. Nishita Desai for useful discussions and suggestions. The authors acknowledge support from the Department of Atomic Energy (DAE), Government of India, under Project Identification No. RTI4002. BD acknowledges partial support through a Swarnajayanti Fellowship of the Department of Science and Technology (DST), Government of India. AR thanks the support of Monash University, where he moved at the end of the writing stage of this manuscript, for support towards completing this work.

A Expressions of $\langle\sigma v\rangle$ for different operators

The expressions of the annihilation rates for the operators in Table 1 can be obtained as,

$$\sigma^{\mathcal{O}_{ss}} v = \frac{3m_\chi^2}{8\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{3/2} v^2 \tag{A.1}$$

$$\sigma^{\mathcal{O}_{pp}} v = \frac{3m_\chi^2}{\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[1+\frac{v^2}{8}\left(\frac{2m_\chi^2-m_q^2}{m_\chi^2-m_q^2}\right)\right] \tag{A.2}$$

$$\sigma^{\mathcal{O}_{sp}} v = \frac{3m_\chi^2}{4\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2} v^2 \tag{A.3}$$

$$\sigma^{\mathcal{O}_{ps}} v = \frac{3m_\chi^2}{4\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{3/2}\left[1+v^2\left(\frac{2m_q^2-m_\chi^2}{m_\chi^2-m_q^2}\right)\right] \tag{A.4}$$

$$\sigma^{\mathcal{O}_{vv}} v = \frac{3m_\chi^2}{2\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[\left(2+\frac{m_q^2}{m_\chi^2}\right)+v^2\left(\frac{8m_\chi^4-4m_q^2m_\chi^2+5m_q^4}{24m_\chi^2(m_\chi^2-m_q^2)}\right)\right] \tag{A.5}$$

$$\sigma^{\mathcal{O}_{aa}} v = \frac{3m_\chi^2}{2\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[\frac{m_q^2}{m_\chi^2}+v^2\left(\frac{8m_\chi^4-22m_q^2m_\chi^2+17m_q^4}{24m_\chi^2(m_\chi^2-m_q^2)}\right)\right] \tag{A.6}$$

$$\sigma^{\mathcal{O}_{va}} v = \frac{3m_\chi^2}{\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[\left(1-\frac{m_q^2}{m_\chi^2}\right)+\frac{v^2}{24}\left(4+5\frac{m_q^2}{m_\chi^2}\right)\right] \tag{A.7}$$

$$\sigma^{\mathcal{O}_{av}} v = \frac{m_\chi^2}{4\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left(2+\frac{m_q^2}{m_\chi^2}\right) v^2 \tag{A.8}$$

$$\sigma^{\mathcal{O}_{tt}} v = \frac{3m_\chi^2}{\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[\left(1+2\frac{m_q^2}{m_\chi^2}\right)+v^2\left(\frac{4m_\chi^4-11m_q^2m_\chi^2+16m_q^4}{24m_\chi^2(m_\chi^2-m_q^2)}\right)\right] \tag{A.9}$$

$$\sigma^{\mathcal{O}_{pt}} v = \frac{6m_\chi^2}{\pi\Lambda^4}\sum_q \left(1-\frac{m_q^2}{m_\chi^2}\right)^{1/2}\left[\left(1-\frac{m_q^2}{m_\chi^2}\right)+\frac{v^2}{24}\left(4+11\frac{m_q^2}{m_\chi^2}\right)\right] \tag{A.10}$$

References

[1] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, 1807.06209.

[2] S. Nussinov, *TECHNOCOSMOLOGY: COULD A TECHNIBARYON EXCESS PROVIDE A 'NATURAL' MISSING MASS CANDIDATE?*, *Phys. Lett. B* **165** (1985) 55.

[3] G.B. Gelmini, L.J. Hall and M.J. Lin, *What Is the Cosmion?*, *Nucl. Phys. B* **281** (1987) 726.

[4] K.M. Zurek, *Asymmetric Dark Matter: Theories, Signatures, and Constraints*, *Phys. Rept.* **537** (2014) 91 [1308.0338].

[5] K. Petraki and R.R. Volkas, *Review of asymmetric dark matter*, *Int. J. Mod. Phys. A* **28** (2013) 1330028 [1305.4939].

[6] S.M. Barr, R.S. Chivukula and E. Farhi, *Electroweak Fermion Number Violation and the Production of Stable Particles in the Early Universe*, *Phys. Lett. B* **241** (1990) 387.

[7] S.M. Barr, *Baryogenesis, sphalerons and the cogeneration of dark matter*, *Phys. Rev. D* **44** (1991) 3062.

[8] D.B. Kaplan, *A Single explanation for both the baryon and dark matter densities*, *Phys. Rev. Lett.* **68** (1992) 741.

[9] S.B. Gudnason, C. Kouvaris and F. Sannino, *Towards working technicolor: Effective theories and dark matter*, *Phys. Rev. D* **73** (2006) 115003 [hep-ph/0603014].

[10] S.B. Gudnason, C. Kouvaris and F. Sannino, *Dark Matter from new Technicolor Theories*, *Phys. Rev. D* **74** (2006) 095008 [hep-ph/0608055].

[11] G.D. Kribs, T.S. Roy, J. Terning and K.M. Zurek, *Quirky Composite Dark Matter*, *Phys. Rev. D* **81** (2010) 095001 [0909.2034].

[12] D.E. Kaplan, M.A. Luty and K.M. Zurek, *Asymmetric Dark Matter*, *Phys. Rev. D* **79** (2009) 115016 [0901.4117].

[13] T. Cohen and K.M. Zurek, *Leptophilic Dark Matter from the Lepton Asymmetry*, *Phys. Rev. Lett.* **104** (2010) 101301 [0909.2035].

[14] T. Cohen, D.J. Phalen, A. Pierce and K.M. Zurek, *Asymmetric Dark Matter from a GeV Hidden Sector*, *Phys. Rev. D* **82** (2010) 056001 [1005.1655].

[15] H. An, S.-L. Chen, R.N. Mohapatra and Y. Zhang, *Leptogenesis as a Common Origin for Matter and Dark Matter*, *JHEP* **03** (2010) 124 [0911.4463].

[16] E.J. Chun, *Leptogenesis origin of Dirac gaugino dark matter*, *Phys. Rev. D* **83** (2011) 053004 [1009.0983].

[17] J. Shelton and K.M. Zurek, *Darkogenesis: A baryon asymmetry from the dark matter sector*, *Phys. Rev. D* **82** (2010) 123512 [1008.1997].

[18] H. Davoudiasl, D.E. Morrissey, K. Sigurdson and S. Tulin, *Hylogenesis: A Unified Origin for Baryonic Visible Matter and Antibaryonic Dark Matter*, *Phys. Rev. Lett.* **105** (2010) 211304 [1008.2399].

[19] N. Haba and S. Matsumoto, *Baryogenesis from Dark Sector*, *Prog. Theor. Phys.* **125** (2011) 1311 [1008.2487].

[20] P.-H. Gu, M. Lindner, U. Sarkar and X. Zhang, *WIMP Dark Matter and Baryogenesis*, *Phys. Rev. D* **83** (2011) 055008 [1009.2690].

[21] M. Blennow, B. Dasgupta, E. Fernandez-Martinez and N. Rius, *Aidnogenesis via Leptogenesis and Dark Sphalerons*, *JHEP* **03** (2011) 014 [1009.3159].

[22] K. Agashe and G. Servant, *Baryon number in warped GUTs: Model building and (dark matter related) phenomenology*, *JCAP* **02** (2005) 002 [hep-ph/0411254].

[23] R. Kitano and I. Low, *Grand unification, dark matter, baryon asymmetry, and the small scale structure of the universe*, hep-ph/0503112.

[24] N. Cosme, L. Lopez Honorez and M.H.G. Tytgat, *Leptogenesis and dark matter related?*, *Phys. Rev. D* **72** (2005) 043505 [hep-ph/0506320].

[25] D. Suematsu, *Nonthermal production of baryon and dark matter*, *Astropart. Phys.* **24** (2006) 511 [hep-ph/0510251].

[26] M.H.G. Tytgat, *Relating leptogenesis and dark matter*, in *41st Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 320–323, 6, 2006 [hep-ph/0606140].

[27] T. Banks, S. Echols and J.L. Jones, *Baryogenesis, dark matter and the Pentagon*, *JHEP* **11** (2006) 046 [hep-ph/0608104].

[28] V. Page, *Non-thermal right-handed sneutrino dark matter and the Omega(DM)/Omega(b) problem*, *JHEP* **04** (2007) 021 [hep-ph/0701266].

[29] D. Hooper, J. March-Russell and S.M. West, *Asymmetric sneutrino dark matter and the Omega(b) / Omega(DM) puzzle*, *Phys. Lett. B* **605** (2005) 228 [hep-ph/0410114].

[30] R. Kitano and I. Low, *Dark matter from baryon asymmetry*, *Phys. Rev. D* **71** (2005) 023510 [hep-ph/0411133].

[31] A. Belyaev, M.T. Frandsen, S. Sarkar and F. Sannino, *Mixed dark matter from technicolor*, *Phys. Rev. D* **83** (2011) 015007 [1007.4839].

[32] J. March-Russell, J. Unwin and S.M. West, *Closing in on Asymmetric Dark Matter I: Model independent limits for interactions with quarks*, *JHEP* **08** (2012) 029 [1203.4854].

[33] S.D. McDermott, H.-B. Yu and K.M. Zurek, *Constraints on Scalar Asymmetric Dark Matter from Black Hole Formation in Neutron Stars*, *Phys. Rev. D* **85** (2012) 023519 [1103.5472].

[34] C. Kouvaris and P. Tinyakov, *Excluding Light Asymmetric Bosonic Dark Matter*, *Phys. Rev. Lett.* **107** (2011) 091301 [1104.0382].

[35] C. Kouvaris, *Limits on Self-Interacting Dark Matter*, *Phys. Rev. Lett.* **108** (2012) 191301 [1111.4364].

[36] C. Kouvaris and P. Tinyakov, *(Not)-constraining heavy asymmetric bosonic dark matter*, *Phys. Rev. D* **87** (2013) 123537 [1212.4075].

[37] J. Bramante, K. Fukushima and J. Kumar, *Constraints on bosonic dark matter from observation of old neutron stars*, *Phys. Rev. D* **87** (2013) 055012 [1301.0036].

[38] N.F. Bell, A. Melatos and K. Petraki, *Realistic neutron star constraints on bosonic asymmetric dark matter*, *Phys. Rev. D* **87** (2013) 123507 [1301.6811].

[39] I. Goldman, R.N. Mohapatra, S. Nussinov, D. Rosenbaum and V. Teplitz, *Possible Implications of Asymmetric Fermionic Dark Matter for Neutron Stars*, *Phys. Lett. B* **725** (2013) 200 [1305.6908].

[40] J. Bramante, K. Fukushima, J. Kumar and E. Stopnitzky, *Bounds on self-interacting fermion dark matter from observations of old neutron stars*, *Phys. Rev. D* **89** (2014) 015010 [1310.3509].

[41] J. Bramante and T. Linden, *Detecting Dark Matter with Imploding Pulsars in the Galactic Center*, *Phys. Rev. Lett.* **113** (2014) 191301 [1405.1031].

[42] J. Bramante and F. Elahi, *Higgs portals to pulsar collapse*, *Phys. Rev. D* **91** (2015) 115001 [1504.04019].

[43] C. Kouvaris and N.G. Nielsen, *Asymmetric Dark Matter Stars*, *Phys. Rev. D* **92** (2015) 063526 [1507.00959].

[44] J. Bramante, T. Linden and Y.-D. Tsai, *Searching for dark matter with neutron star mergers and quiet kilonovae*, *Phys. Rev. D* **97** (2018) 055016 [1706.00001].

[45] C. Kouvaris, P. Tinyakov and M.H.G. Tytgat, *NonPrimordial Solar Mass Black Holes*, *Phys. Rev. Lett.* **121** (2018) 221102 [1804.06740].

[46] M.I. Gresham and K.M. Zurek, *Asymmetric Dark Stars and Neutron Star Stability*, *Phys. Rev. D* **99** (2019) 083008 [1809.08254].

[47] R. Garani, Y. Genolini and T. Hambye, *New Analysis of Neutron Star Constraints on Asymmetric Dark Matter*, *JCAP* **05** (2019) 035 [1812.08773].

[48] B. Dasgupta, A. Gupta and A. Ray, *Dark matter capture in celestial objects: light mediators, self-interactions, and complementarity with direct detection*, *JCAP* **10** (2020) 023 [2006.10773].

[49] B. Dasgupta, R. Laha and A. Ray, *Low Mass Black Holes from Dark Core Collapse*, *Phys. Rev. Lett.* **126** (2021) 141105 [2009.01825].

[50] N. Rutherford, G. Raaijmakers, C. Prescod-Weinstein and A. Watts, *Constraining bosonic asymmetric dark matter with neutron star mass-radius measurements*, *Phys. Rev. D* **107** (2023) 103051 [2208.03282].

[51] S. Bhattacharya, B. Dasgupta, R. Laha and A. Ray, *Can LIGO Detect Nonannihilating Dark Matter?*, *Phys. Rev. Lett.* **131** (2023) 091401 [2302.07898].

[52] J. Bramante and N. Raj, *Dark matter in compact stars*, *Phys. Rept.* **1052** (2024) 1 [2307.14435].

[53] S. Cebrián, *Review on dark matter searches*, *J. Phys. Conf. Ser.* **2502** (2023) 012004 [2205.06833].

[54] E.A. Baltz and L. Bergstrom, *Detection of leptonic dark matter*, *Phys. Rev. D* **67** (2003) 043516 [hep-ph/0211325].

[55] C.-R. Chen and F. Takahashi, *Cosmic rays from Leptonic Dark Matter*, *JCAP* **02** (2009) 004 [0810.4110].

[56] P.J. Fox and E. Poppitz, *Leptophilic Dark Matter*, *Phys. Rev. D* **79** (2009) 083528 [0811.0399].

[57] Q.-H. Cao, E. Ma and G. Shaughnessy, *Dark Matter: The Leptonic Connection*, *Phys. Lett. B* **673** (2009) 152 [0901.1334].

[58] A. Ibarra, A. Ringwald, D. Tran and C. Weniger, *Cosmic Rays from Leptophilic Dark Matter Decay via Kinetic Mixing*, *JCAP* **08** (2009) 017 [0903.3625].

[59] X.-J. Bi, X.-G. He and Q. Yuan, *Parameters in a class of leptophilic models from PAMELA, ATIC and FERMI*, *Phys. Lett. B* **678** (2009) 168 [0903.0122].

[60] J. Kopp, V. Niro, T. Schwetz and J. Zupan, *DAMA/LIBRA and leptonically interacting Dark Matter*, *Phys. Rev. D* **80** (2009) 083502 [0907.3159].

[61] P. Ko and Y. Omura, *Supersymmetric U(1)B x U(1)L model with leptophilic and leptophobic cold dark matters*, *Phys. Lett. B* **701** (2011) 363 [1012.4679].

[62] W. Chao, *Pure Leptonic Gauge Symmetry, Neutrino Masses and Dark Matter*, *Phys. Lett. B* **695** (2011) 157 [1005.1024].

[63] D. Schmidt, T. Schwetz and T. Toma, *Direct Detection of Leptophilic Dark Matter in a Model with Radiative Neutrino Masses*, *Phys. Rev. D* **85** (2012) 073009 [1201.0906].

[64] M. Das and S. Mohanty, *Leptophilic dark matter in gauged $L_\mu - L_\tau$ extension of MSSM*, *Phys. Rev. D* **89** (2014) 025004 [1306.4505].

[65] P. Schwaller, T.M.P. Tait and R. Vega-Morales, *Dark Matter and Vectorlike Leptons from Gauged Lepton Number*, *Phys. Rev. D* **88** (2013) 035001 [1305.1108].

[66] Y. Bai and J. Berger, *Lepton Portal Dark Matter*, *JHEP* **08** (2014) 153 [1402.6696].

[67] P. Agrawal, Z. Chacko and C.B. Verhaaren, *Leptophilic Dark Matter and the Anomalous Magnetic Moment of the Muon*, *JHEP* **08** (2014) 147 [1402.7369].

[68] J. Kopp, L. Michaels and J. Smirnov, *Loopy Constraints on Leptophilic Dark Matter and Internal Bremsstrahlung*, *JCAP* **04** (2014) 022 [1401.6457].

[69] XENON100 collaboration, *Exclusion of Leptophilic Dark Matter Models using XENON100 Electronic Recoil Data*, *Science* **349** (2015) 851 [1507.07747].

[70] LZ collaboration, *Projected sensitivities of the LUX-ZEPLIN experiment to new physics via low-energy electron recoils*, *Phys. Rev. D* **104** (2021) 092009 [2102.11740].

[71] C.-Y. Chen, J. Kozaczuk and Y.-M. Zhong, *Exploring leptophilic dark matter with NA64-μ*, *JHEP* **10** (2018) 154 [1807.03790].

[72] L. Marsicano, M. Battaglieri, A. Celentano, R. De Vita and Y.-M. Zhong, *Probing Leptophilic Dark Sectors at Electron Beam-Dump Facilities*, *Phys. Rev. D* **98** (2018) 115022 [1812.03829].

[73] CRESST collaboration, *Results on MeV-scale dark matter from a gram-scale cryogenic calorimeter operated above ground*, *Eur. Phys. J. C* **77** (2017) 637 [1707.06749].

[74] SuperCDMS collaboration, *Light Dark Matter Search with a High-Resolution Athermal Phonon Detector Operated Above Ground*, *Phys. Rev. Lett.* **127** (2021) 061801 [2007.14289].

[75] SENSEI collaboration, *SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper-CCD*, *Phys. Rev. Lett.* **125** (2020) 171802 [2004.11378].

[76] DAMIC-M collaboration, *First Constraints from DAMIC-M on Sub-GeV Dark-Matter Particles Interacting with Electrons*, *Phys. Rev. Lett.* **130** (2023) 171003 [2302.02372].

[77] A. Aguilar-Arevalo, F.A. Bessia, N. Avalos, D. Baxter, X. Bertou, C. Bonifazi et al., *The oscura experiment*, *arXiv preprint arXiv:2202.10518* (2022) .

[78] M.C. Marshall, M.J. Turner, M.J.H. Ku, D.F. Phillips and R.L. Walsworth, *Directional detection of dark matter with diamond*, *Quantum Sci. Technol.* **6** (2021) 024011 [2009.01028].

[79] S.A. Hertel, A. Biekert, J. Lin, V. Velan and D.N. McKinsey, *Direct detection of sub-GeV dark matter using a superfluid ^{4}He target*, *Phys. Rev. D* **100** (2019) 092007 [1810.06283].

[80] SPICE, HeRALD collaboration, *Applying Superfluid Helium to Light Dark Matter Searches: Demonstration of the HeRALD Detector Concept*, 2307.11877.

[81] G.-B. Kim, *Self-absorption and Phonon Pulse Shape Discrimination in Scintillating Bolometers*, *Journal of Low Temperature Physics* **199** (2020) 1004.

[82] Y. Hochberg, Y. Zhao and K.M. Zurek, *Superconducting Detectors for Superlight Dark Matter*, *Phys. Rev. Lett.* **116** (2016) 011301 [1504.07237].

[83] Y. Hochberg, I. Charaev, S.-W. Nam, V. Verma, M. Colangelo and K.K. Berggren, *Detecting Sub-GeV Dark Matter with Superconducting Nanowires*, *Phys. Rev. Lett.* **123** (2019) 151802 [1903.05101].

[84] R.M. Geilhufe, F. Kahlhoefer and M.W. Winkler, *Dirac Materials for Sub-MeV Dark Matter Detection: New Targets and Improved Formalism*, *Phys. Rev. D* **101** (2020) 055005 [1910.02091].

[85] Y. Kahn and T. Lin, *Searches for light dark matter using condensed matter systems*, *Rept. Prog. Phys.* **85** (2022) 066901 [2108.03239].

[86] C. Blanco, R. Essig, M. Fernandez-Serra, H. Ramani and O. Slone, *Dark matter direct detection with quantum dots*, *Phys. Rev. D* **107** (2023) 095035 [2208.05967].

[87] Z. Huang, C. Lane, S.E. Grefe, S. Nandy, B. Fauseweh, S. Paschen et al., *Dark Matter Detection with Strongly Correlated Topological Materials: Flatband Effect*, 2305.19967.

[88] A. Das, J. Jang and H. Min, *Sub-MeV Dark Matter Detection with Bilayer Graphene*, 2312.00866.

[89] B. Carew, A.R. Caddell, T.N. Maity and C.A.J. O'Hare, *The neutrino fog for dark matter-electron scattering experiments*, 2312.04303.

[90] N.F. Bell, G. Busoni and S. Robles, *Capture of Leptophilic Dark Matter in Neutron Stars*, *JCAP* **06** (2019) 054 [1904.09803].

[91] N.F. Bell, G. Busoni, M.E. Ramirez-Quezada, S. Robles and M. Virgato, *Improved treatment of dark matter capture in white dwarfs*, *JCAP* **10** (2021) 083 [2104.14367].

[92] P.J. Fox, R. Harnik, J. Kopp and Y. Tsai, *LEP Shines Light on Dark Matter*, *Phys. Rev. D* **84** (2011) 014028 [1103.0240].

[93] S. Dutta, D. Sachdeva and B. Rawat, *Signals of Leptophilic Dark Matter at the ILC*, *Eur. Phys. J. C* **77** (2017) 639 [1704.03994].

[94] B. Barman, S. Bhattacharya, S. Girmohanta and S. Jahedi, *Effective Leptophilic WIMPs at the e+e- collider*, *JHEP* **04** (2022) 146 [2109.10936].

[95] S. Kundu, A. Guha, P.K. Das and P.S.B. Dev, *EFT analysis of leptophilic dark matter at future electron-positron colliders in the mono-photon and mono-Z channels*, *Phys. Rev. D* **107** (2023) 015003 [2110.06903].

[96] CMS collaboration, *Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at s=13 TeV*, *Phys. Rev. D* **105** (2022) 092007 [2201.11585].

[97] ATLAS collaboration, *Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using 139 fb^{-1} of proton-proton data recorded by the ATLAS experiment*, *JHEP* **08** (2022) 104 [2202.07953].

[98] L.E. Strigari, *Neutrino Coherent Scattering Rates at Direct Dark Matter Detectors*, *New J. Phys.* **11** (2009) 105011 [0903.3630].

[99] J. Billard, L. Strigari and E. Figueroa-Feliciano, *Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments*, *Phys. Rev. D* **89** (2014) 023524 [1307.5458].

[100] J. Monroe and P. Fisher, *Neutrino Backgrounds to Dark Matter Searches*, *Phys. Rev. D* **76** (2007) 033007 [0706.3019].

[101] G. Arcadi, A. Djouadi and M. Kado, *The Higgs-portal for dark matter: effective field theories versus concrete realizations*, *Eur. Phys. J. C* **81** (2021) 653 [2101.02507].

[102] J. Fan, M. Reece and L.-T. Wang, *Non-relativistic effective theory of dark matter direct detection*, *JCAP* **11** (2010) 042 [1008.1591].

[103] H. Iminniyaz, M. Drees and X. Chen, *Relic Abundance of Asymmetric Dark Matter*, *JCAP* **07** (2011) 003 [1104.5548].

[104] ATLAS collaboration, *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, *Phys. Rev. D* **103** (2021) 112006 [2102.10874].

[105] CMS collaboration, *Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, *JHEP* **11** (2021) 153 [2107.13021].

[106] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[107] A. Alloul, N.D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250 [1310.1921].

[108] T. Sjostrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175].

[109] T. Sjostrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [0710.3820].

[110] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [1307.6346].

[111] G. Belanger, F. Boudjema, A. Pukhov and A. Semenov, *Dark matter direct detection rate in a generic model with micrOMEGAs 2.2*, *Comput. Phys. Commun.* **180** (2009) 747 [0803.2360].

[112] COMPASS collaboration, *The Polarised Valence Quark Distribution from semi-inclusive DIS*, *Phys. Lett. B* **660** (2008) 458 [0707.4077].

[113] LZ collaboration, *First Dark Matter Search Results from the LUX-ZEPLIN (LZ) Experiment*, 2207.03764.

[114] DarkSide collaboration, *Low-Mass Dark Matter Search with the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018) 081307 [1802.06994].

[115] XENON collaboration, *Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T*, *Phys. Rev. Lett.* **122** (2019) 141301 [1902.03234].

[116] PICO collaboration, *Dark Matter Search Results from the Complete Exposure of the PICO-60* C_3F_8 *Bubble Chamber*, *Phys. Rev.* **D100** (2019) 022001 [1902.04031].

[117] XENON100 collaboration, *Dark Matter Results from 100 Live Days of XENON100 Data*, *Phys. Rev. Lett.* **107** (2011) 131302 [1104.2549].

[118] DELPHI collaboration, *Photon events with missing energy in e+ e- collisions at s**(1/2) = 130-GeV to 209-GeV*, *Eur. Phys. J. C* **38** (2005) 395 [hep-ex/0406019].

[119] FCC collaboration, *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2*, *Eur. Phys. J. ST* **228** (2019) 261.

[120] W.H. Press and D.N. Spergel, *Capture by the sun of a galactic population of weakly interacting, massive particles*, *APJ* **296** (1985) .

[121] A. Gould, *Weakly Interacting Massive Particle Distribution in and Evaporation from the Sun*, *APJ* **321** (1987) 560.

[122] A. Gould, *Resonant Enhancements in Weakly Interacting Massive Particle Capture by the Earth*, *APJ* **321** (1987) 571.

[123] G. Bertone and M. Fairbairn, *Compact Stars as Dark Matter Probes*, *Phys. Rev. D* **77** (2008) 043515 [0709.1485].

[124] J. Bramante, A. Delgado and A. Martin, *Multiscatter stellar capture of dark matter*, *Phys. Rev. D* **96** (2017) 063002 [1703.04043].

[125] B. Dasgupta, A. Gupta and A. Ray, *Dark matter capture in celestial objects: Improved treatment of multiple scattering and updated constraints from white dwarfs*, *JCAP* **08** (2019) 018 [1906.04204].

[126] C. Kouvaris, *WIMP Annihilation and Cooling of Neutron Stars*, *Phys. Rev. D* **77** (2008) 023006 [0708.2362].

[127] C. Kouvaris and P. Tinyakov, *Can Neutron stars constrain Dark Matter?*, *Phys. Rev. D* **82** (2010) 063531 [1004.0586].

[128] M. Baryakhtar, J. Bramante, S.W. Li, T. Linden and N. Raj, *Dark Kinetic Heating of Neutron Stars and An Infrared Window On WIMPs, SIMPs, and Pure Higgsinos*, *Phys. Rev. Lett.* **119** (2017) 131801 [1704.01577].

[129] N.F. Bell, G. Busoni and S. Robles, *Heating up Neutron Stars with Inelastic Dark Matter*, *JCAP* **09** (2018) 018 [1807.02840].

[130] C.-S. Chen and Y.-H. Lin, *Reheating neutron stars with the annihilation of self-interacting dark matter*, *JHEP* **08** (2018) 069 [1804.03409].

[131] J.F. Acevedo, J. Bramante, R.K. Leane and N. Raj, *Warming Nuclear Pasta with Dark Matter: Kinetic and Annihilation Heating of Neutron Star Crusts*, *JCAP* **03** (2020) 038 [1911.06334].

[132] N. Raj, P. Tanedo and H.-B. Yu, *Neutron stars at the dark matter direct detection frontier*, *Phys. Rev. D* **97** (2018) 043006 [1707.09442].

[133] A. de Lavallaz and M. Fairbairn, *Neutron Stars as Dark Matter Probes*, *Phys. Rev. D* **81** (2010) 123521 [1004.0629].

[134] M. McCullough and M. Fairbairn, *Capture of Inelastic Dark Matter in White Dwarves*, *Phys. Rev. D* **81** (2010) 083520 [1001.2737].

[135] D. Hooper, D. Spolyar, A. Vallinotto and N.Y. Gnedin, *Inelastic Dark Matter As An Efficient Fuel For Compact Stars*, *Phys. Rev. D* **81** (2010) 103531 [1002.0005].

[136] J.F. Acevedo and J. Bramante, *Supernovae Sparked By Dark Matter in White Dwarfs*, *Phys. Rev. D* **100** (2019) 043020 [1904.11993].

[137] P.W. Graham, R. Janish, V. Narayan, S. Rajendran and P. Riggins, *White Dwarfs as Dark Matter Detectors*, *Phys. Rev. D* **98** (2018) 115027 [1805.07381].

[138] DAMIC-M collaboration, *The DAMIC-M experiment: Status and first results*, *SciPost Phys. Proc.* **12** (2023) 014 [2210.12070].

[139] SuperCDMS collaboration, *First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector*, *Phys. Rev. Lett.* **121** (2018) 051301 [1804.10697].

arXiv:2402.17162v1 [hep-ph] 27 Feb 2024

Gluon GTMDs at nonzero skewness and impact parameter dependent parton distributions

Chentao Tan[1] and Zhun Lu[1,*]

[1] School of Physics, Southeast University, Nanjing 211189, China

We investigate the leading twist generalized transverse momentum dependent parton distributions (GTMDs) of the unpolarized and longitudinally polarized gluons in the nucleon. We adopt a light-front gluon-triquark model for the nucleon motivated by soft-wall AdS/QCD. The gluon GTMDs are defined through the off-forward gluon-gluon generalized correlator and are expressed as the overlap of light-cone wave functions. The GTMDs can be employed to provide the generalized parton distributions (GPDs) by integrating out the transverse momentum. The Fourier transform of the GPDs encodes the parton distributions in the transverse position space, namely, the impact parameter dependent parton distributions (IPDs). We also calculate the three gluon IPDs corresponding to the GPDs H^g, E^g and $\widetilde{H}^g$, and present their dependence on x and $b_\perp$, respectively.

I. INTRODUCTION

An outstanding goal in hadron physics is to understand the structure of hadrons in terms of quarks and gluons. The deep inelastic scattering (DIS) is among the key tools to reveal hadronic structure, because one can extract the parton distribution functions (PDFs) [1–4] from such process. The PDFs are functions of the longitudinal momentum fraction, which encode the distributions of longitudinal momentum and polarizations of partons. A more comprehensive picture about the nucleon can be revealed by the transverse momentum dependent parton distributions (TMDs) [5, 6], which encode transverse motion of partons inside the nucleon. TMDs can be extracted from semi-inclusive reactions such as the semi-inclusive deep inelastic scatting (SIDIS) and the Drell-Yan process [7–11]. Besides the TMDs, in the off-forward region a new type of nucleon structure–the so-call generalized parton distributions (GPDs) [12–20]–emerge. It is the extension of the ordinary PDF from the forward scattering region to the off-forward scattering region. The GPDs appear in the description of hard exclusive reactions, such as the deeply virtual Compton scattering(DVCS) and the deeply virtual meson production(DVMP) [13, 16, 17, 19]. Furthermore, the transverse position of partons is encoded in the impact parameter dependent parton distributions (IPDs) [21, 22], which are the Fourier transform of the GPDs [17] at zero skewness with respect to the transverse momentum transfer of the hadron.

The most complete structural information of hadrons is contained in the so-called generalized transverse momentum dependent parton distributions (GTMDs) [23, 24], which are often considered as the "mother distributions", since several GTMDs can project to the TMDs and the GPDs in certain kinematical limits. The quark GTMDs may be measurable in the exclusive double Drell-Yan process [25], while the feasibility to measure the gluon GTMDs in the diffractive dijet production has been studied [26–29].

The first complete classification of various parton distributions and their relations with each other has been discussed in Refs. [23, 24]. There are sixteen twist-2 GTMDs for the quark and the gluon in the nucleon, respectively. These GTMDs encode the information of the distributions of the unpolarized and polarized partons. They are characterized by revealing the nucleon spin structure, for example, $F_{1,4}$ and $G_{1,1}$ play an important role in describing the canonical orbit angular momentum (OAM) [30–33] and the spin-orbit correlations [34, 35] of partons, respectively. The quark GTMDs for the nucleon have been calculated in various models, such as the light-cone constituent quark model [30, 36, 37], the light-front dressed quark model [38–40], the light-cone spectator model [41], the light-front quark-diquark model [42–47], the chiral soliton model [30, 36] and the quark target model [48], and so have these distributions for the pion [49–52]. In addition, the importance of the IPDs lies in their physical interpretation as a probability density in the impact parameter space. The quark IPDs for hadrons have been studied in Refs. [22, 53]. However, the theoretical study on the gluon GTMDs and IPDs is still not sufficient.

It should be noted that most of the previous calculations for the GTMDs are made by assuming that the skewness ξ is 0. However, most of the data probed in the experiments is at $\xi \neq 0$. Therefore, a more detailed investigation of the GTMDs at nonzero skewness is necessary. In this work, we investigate the leading twist GTMDs of the unpolarized and longitudinally polarized gluons in the nucleon using the soft-wall AdS/QCD model [54, 55]. This model has been widely applied to the calculations of the PDFs, form factors and mass spectrum [56]. In the description of hadronic form factors at large Q^2 [56–59], the soft wall AdS/QCD embodies a main advantage, namely, the analytical

* zhunlu@seu.edu.cn

implementation of quark counting rules [60]. Same as the light-cone spectator model [61], the minimum Fock state for the nucleon containing a gluon here is also a two-body composite system of a struck gluon and a three-quark spectator. We calculate the gluon GTMDs by writing the contracted correlators in the overlap representation in terms of the light-cone wave functions. In order to understand the gluon distributions in the transverse position space, we also calculate the three IPDs corresponding to the GPDs H^g, E^g and $\widetilde{H}^g$.

The rest of the paper is organized as follows. In Sec. II, we provide the definitions of the gluon GTMDs and IPDs for the nucleon. In Sec. III, we write down the overlap representations of the contracted gluon correlators using the light-cone wave functions in the soft-wall AdS/QCD model. In Sec. IV, we present not only the three-dimensional plots of the GTMDs, but also the dependence of the IPDs on x and $b_\perp$, respectively. Summary of the paper is given in Sec. V.

II. DEFINITIONS OF GLUON GTMDS AT NONZERO SKEWNESS AND IPDS

A complete set of the gluon GTMDs for the nucleon has been presented in Ref. [24, 62]. In the fixed light-cone time $z^+ = 0$, they can be defined through the off-forward gluon-gluon generalized correlator

$$W^{g[ij]}_{\lambda',\lambda}(x,P,\Delta,\boldsymbol{k}_\perp) = \int \frac{dz^- d^2\boldsymbol{z}_\perp}{(2\pi)^3 P^+} e^{ik\cdot z}\left\langle p',\lambda'\middle| F^{+i}_a(-\frac{z}{2})\mathcal{W}_{ab}(-\frac{z}{2},\frac{z}{2})F^{+j}_b(\frac{z}{2})\middle|p,\lambda\right\rangle\Big|_{z^+=0}, \tag{1}$$

where $p = P - \Delta/2$ ($p' = P + \Delta/2$) and λ (λ') represent the four-momentum and the helicity of the initial (final) nucleon, respectively. This correlator depends on four arguments. Among them, $P = (p + p')/2$ and $\Delta = p' - p$ denote the average four-momentum and the momentum transfer of the nucleon, respectively, $x = k^+/P^+$ denotes the average longitudinal gluon momentum fraction, and $\boldsymbol{k}_\perp$ denotes the average transverse gluon momentum. In Eq. (1), the gluon field strength tensor $F^{\mu\nu}_a$ has a general form

$$F^{\mu\nu}_a = \partial^\mu A^\nu_a - \partial^\nu A^\mu_a + g f_{abc} A^\mu_b A^\nu_c, \tag{2}$$

where f_{abc} and g are the structure constants and the coupling constant of SU(3), respectively. The Wilson line

$$\mathcal{W}_{ab}\Big(-\frac{z}{2},\frac{z}{2}\Big)\Big|_{z^+=0} = \Big[0^+,-\frac{z^-}{2},-\frac{\boldsymbol{z}_\perp}{2};0^+,\frac{z^-}{2},\frac{\boldsymbol{z}_\perp}{2}\Big]_{ab} = \mathcal{P}\exp\Big[-g\int_{-\frac{z^-}{2}}^{\frac{z^-}{2}} dy^- \int_{-\frac{\boldsymbol{z}_\perp}{2}}^{\frac{\boldsymbol{z}_\perp}{2}} d\boldsymbol{y}_\perp f_{abc} A^+_c(0^+,y^-,\boldsymbol{y}_\perp)\Big] \tag{3}$$

ensures the color gauge invariance of the bi-local gluon operator, where $\mathcal{P}$ denotes all possible ordered paths followed by the gluon field A. From now on, we choose the light-front gauge ($A^+ = 0$) and take the gauge link to be unit.

In leading twist, there are sixteen gluon GTMDs for the nucleon [62]. In this work, we concentrate on

$$\begin{aligned}
W^g_{\lambda',\lambda} =& \delta^{ij}_\perp W^{g[ij]}_{\lambda',\lambda} \\
=& \frac{1}{2M}\bar{u}(p',\lambda')\Big[F^g_{1,1} + \frac{i\sigma^{i+}k^i_\perp}{P^+}F^g_{1,2} + \frac{i\sigma^{i+}\Delta^i_\perp}{P^+}F^g_{1,3} + \frac{i\sigma^{ij}k^i_\perp\Delta^j_\perp}{M^2}F^g_{1,4}\Big]u(p,\lambda) \\
=& \frac{1}{M\sqrt{1-\xi^2}}\Big\{\Big[M\delta_{\lambda',\lambda} - \frac{1}{2}(\lambda\Delta^1_\perp + i\Delta^2_\perp)\delta_{-\lambda',\lambda}\Big]F^g_{1,1} + (1-\xi^2)(\lambda k^1_\perp + ik^2_\perp)\delta_{-\lambda',\lambda}F^g_{1,2} \\
&+ (1-\xi^2)(\lambda\Delta^1_\perp + i\Delta^2_\perp)\delta_{-\lambda',\lambda}F^g_{1,3} + \frac{i\epsilon^{ij}_\perp k^i_\perp\Delta^j_\perp}{M^2}\Big[\lambda M\delta_{\lambda',\lambda} - \frac{\xi}{2}(\Delta^1_\perp + i\lambda\Delta^2_\perp)\delta_{-\lambda',\lambda}\Big]F^g_{1,4}\Big\}
\end{aligned} \tag{4}$$

and

$$\begin{aligned}
\widetilde{W}^g_{\lambda',\lambda} =& -i\epsilon^{ij}_\perp W^{g[ij]}_{\lambda',\lambda} \\
=& \frac{1}{2M}\bar{u}(p',\lambda')\Big[-\frac{i\epsilon^{ij}_\perp k^i_\perp\Delta^j_\perp}{M^2}G^g_{1,1} + \frac{i\sigma^{i+}\gamma_5 k^i_\perp}{P^+}G^g_{1,2} + \frac{i\sigma^{i+}\gamma_5\Delta^i_\perp}{P^+}G^g_{1,3} + i\sigma^{+-}\gamma_5 G^g_{1,4}\Big]u(p,\lambda) \\
=& \frac{1}{M\sqrt{1-\xi^2}}\Big\{-\frac{i\epsilon^{ij}_\perp k^i_\perp\Delta^j_\perp}{M^2}\Big[M\delta_{\lambda',\lambda} - \frac{1}{2}(\lambda\Delta^1_\perp + i\Delta^2_\perp)\delta_{-\lambda',\lambda}\Big]G^g_{1,1} + (1-\xi^2)(k^1_\perp + i\lambda k^2_\perp)\delta_{-\lambda',\lambda}G^g_{1,2} \\
&+ (1-\xi^2)(\Delta^1_\perp + i\lambda\Delta^2_\perp)\delta_{-\lambda',\lambda}G^g_{1,3} + \Big[\lambda M\delta_{\lambda',\lambda} - \frac{\xi}{2}(\Delta^1_\perp + i\lambda\Delta^2_\perp)\delta_{-\lambda',\lambda}\Big]G^g_{1,4}\Big\},
\end{aligned} \tag{5}$$

where M denotes the nucleon mass, the skewness $\xi = (p^+ - p'^+)/(p^+ + p'^+) = -\Delta^+/(2P^+)$ denotes the longitudinal momentum fraction transfer of the nucleon. In Eqs. (4,5), two tensors $\delta_\perp^{ij} = -g_\perp^{ij}$ and $\epsilon_\perp^{ij} = \epsilon^{+-ij}$ have been introduced, where $g^{\mu\nu}$ is the metric tensor and $\epsilon^{\alpha\beta\rho\sigma}$ is an antisymmetric tensor with $\epsilon^{+-12} = 1$. Four F-type GTMDs describe the distributions of unpolarized gluons, while four G-type GTMDs describe the gluon helicity distributions. For transversely polarized gluons, one needs another set of eight leading twist H-type GTMDs. In general, the GTMDs are complex-valued functions [23, 24], and they depend on the variables x, ξ, $\boldsymbol{\Delta}_\perp^2$, $\boldsymbol{k}_\perp^2$ and $\boldsymbol{\Delta}_\perp \cdot \boldsymbol{k}_\perp$. Then we express the gluon GTMDs in terms of the correlators contracted with two tensors as follows:

$$W^g_{+,+}(x,P,\Delta,\boldsymbol{k}_\perp) + W^g_{+,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} M F^g_{1,1}, \tag{6}$$

$$W^g_{+,+}(x,P,\Delta,\boldsymbol{k}_\perp) - W^g_{+,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} \frac{i\epsilon_\perp^{ij} k_\perp^i \Delta_\perp^j}{M^2} M F^g_{1,4}, \tag{7}$$

$$W^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) + W^g_{-,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} \left\{ -\frac{1}{2}(\Delta_\perp^1 + i\Delta_\perp^2) F^g_{1,1} + (1-\xi^2)(\Delta_\perp^1 + i\Delta_\perp^2) F^g_{1,3} \right\}, \tag{8}$$

$$\begin{aligned} W^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) - W^g_{-,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = &\frac{2}{M\sqrt{1-\xi^2}} \left\{ (1-\xi^2)(k_\perp^1 + ik_\perp^2) F^g_{1,2} \right. \\ &\left. + \frac{i\epsilon_\perp^{ij} k_\perp^i \Delta_\perp^j}{M^2} \left[-\frac{\xi}{2}(\Delta_\perp^1 + i\Delta_\perp^2) \right] F^g_{1,4} \right\}, \end{aligned} \tag{9}$$

for F-type GTMDs, and

$$\widetilde{W}^g_{+,+}(x,P,\Delta,\boldsymbol{k}_\perp) + \widetilde{W}^g_{+,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} M G^g_{1,4}, \tag{10}$$

$$\widetilde{W}^g_{+,+}(x,P,\Delta,\boldsymbol{k}_\perp) - \widetilde{W}^g_{+,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} \left[-\frac{i\epsilon_\perp^{ij} k_\perp^i \Delta_\perp^j}{M^2} M G^g_{1,1} \right], \tag{11}$$

$$\widetilde{W}^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) + \widetilde{W}^g_{-,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = \frac{2}{M\sqrt{1-\xi^2}} \left\{ (1-\xi^2)(\Delta_\perp^1 + i\Delta_\perp^2) G^g_{1,3} - \frac{\xi}{2}(\Delta_\perp^1 + i\Delta_\perp^2) G^g_{1,4} \right\}, \tag{12}$$

$$\begin{aligned} \widetilde{W}^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) - \widetilde{W}^g_{-,+}(x,P,\Delta,-\boldsymbol{k}_\perp) = &\frac{2}{M\sqrt{1-\xi^2}} \left\{ -\frac{i\epsilon_\perp^{ij} k_\perp^i \Delta_\perp^j}{M^2} \left[-\frac{1}{2}(\Delta_\perp^1 + i\Delta_\perp^2) \right] G^g_{1,1} \right. \\ &\left. + (1-\xi^2)(k_\perp^1 + ik_\perp^2) G^g_{1,2} \right\}, \end{aligned} \tag{13}$$

for G-type GTMDs, where the subscript + (−) denotes that the helicity of the nucleon is +1/2 (−1/2).

There are four leading twist gluon GPDs that can be expressed in terms of the GTMDs in Eqs. (4,5) as [63]

$$H^g(x,\xi,\boldsymbol{\Delta}_\perp^2) = \int d^2\boldsymbol{k}_\perp \left[F^g_{1,1} + 2\xi^2 \left(\frac{\boldsymbol{\Delta}_\perp \cdot \boldsymbol{k}_\perp}{\boldsymbol{\Delta}_\perp^2} F^g_{1,2} + F^g_{1,3} \right) \right], \tag{14}$$

$$E^g(x,\xi,\boldsymbol{\Delta}_\perp^2) = \int d^2\boldsymbol{k}_\perp \left[-F^g_{1,1} + 2(1-\xi^2) \left(\frac{\boldsymbol{\Delta}_\perp \cdot \boldsymbol{k}_\perp}{\boldsymbol{\Delta}_\perp^2} F^g_{1,2} + F^g_{1,3} \right) \right], \tag{15}$$

$$\widetilde{H}^g(x,\xi,\boldsymbol{\Delta}_\perp^2) = \int d^2\boldsymbol{k}_\perp \left[2\xi \left(\frac{\boldsymbol{\Delta}_\perp \cdot \boldsymbol{k}_\perp}{\boldsymbol{\Delta}_\perp^2} G^g_{1,2} + G^g_{1,3} \right) + G^g_{1,4} \right], \tag{16}$$

$$\widetilde{E}^g(x,\xi,\boldsymbol{\Delta}_\perp^2) = \int d^2\boldsymbol{k}_\perp \left[\frac{2(1-\xi^2)}{\xi} \left(\frac{\boldsymbol{\Delta}_\perp \cdot \boldsymbol{k}_\perp}{\boldsymbol{\Delta}_\perp^2} G^g_{1,2} + G^g_{1,3} \right) - G^g_{1,4} \right]. \tag{17}$$

The IPDs are defined as the Fourier transform of the GPDs at zero skewness with respect to $\boldsymbol{\Delta}_\perp$. At $\xi = 0$, $\widetilde{E}^g$ will not show up, and the other three GPDs provide the corresponding IPDs $\mathcal{H}^g$, $\mathcal{E}^g$ and $\widetilde{\mathcal{H}}^g$ through the Fourier transform

$$\mathcal{X}(x,\boldsymbol{b}_\perp) = \int \frac{d^2\boldsymbol{\Delta}_\perp}{(2\pi)^2} e^{-i\boldsymbol{\Delta}_\perp \cdot \boldsymbol{b}_\perp} X(x,0,\boldsymbol{\Delta}_\perp^2), \tag{18}$$

where $\boldsymbol{b}_\perp$ is a two-dimensional vector in the impact parameter space conjugate to $\boldsymbol{\Delta}_\perp$. The IPD $\mathcal{H}^g$ is related to the transverse position distribution of unpolarized gluons in an unpolarized nucleon, $\mathcal{E}^g$ is related to the transverse position distribution of unpolarized gluons in a transversely polarized nucleon, and $\widetilde{\mathcal{H}}^g$ is related to the transverse position distribution of longitudinally polarized gluons in a longitudinally polarized nucleon.

III. GLUON CORRELATORS IN SOFT-WALL ADS/QCD MODEL

In this section, we express the contracted gluon correlators as the overlap of the light-cone wave functions in the soft-wall AdS/QCD model. The light-cone formalism has been widely used in different theoretical approaches to reveal the structural information of hadrons [40, 64, 65]. On the other hand, the soft-wall AdS/QCD model has also been successfully employed to describe the structural information of nucleons, e.g., various parton distributions, electromagnetic form factors, spin asymmetries, etc., [55, 63]. The authors in Ref. [63] presented the quark GTMDs at nonzero skewness for the nucleon in this model, then the investigation of these distributions for gluons will be equally important. Similarly, $|qqqg\rangle$ is considered as the minimum Fock state for a nucleon containing a gluon. We treat this four-body system as a two-body composite system of a spin-1 gluon g and a spin-1/2 three-quark spectator X for simplicity [61]:

$$|p;S\rangle \to |g_{s_g} X_{s_X}(qqq)\rangle, \tag{19}$$

where s_g and s_X denote the gluon and spectator spins, respectively. We introduce the notation of the light-cone wave function $\psi^{\lambda}_{\lambda_g\lambda_X}(x,\boldsymbol{k}_\perp)$, where λ, λ_g and λ_X denote the helicity of the nucleon, the gluon and the three-quark spectator, respectively. Then the Fock-state expansion for a nucleon with helicity $\lambda=+1/2$ (we use the shorten notation $+$) is given by

$$\begin{aligned}
|\Psi^{+}_{\text{two particle}}(p^+,\boldsymbol{p}_\perp=\boldsymbol{0}_\perp)\rangle = \int \frac{d^2\boldsymbol{k}_\perp dx}{16\pi^3\sqrt{x(1-x)}} \\
\times\Big[\psi^{+}_{+1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|+1,+\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle+\psi^{+}_{+1-\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|+1,-\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle \\
+\psi^{+}_{-1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|-1,+\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle+\psi^{+}_{-1-\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|-1,-\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle\Big],
\end{aligned} \tag{20}$$

where

$$\begin{aligned}
\psi^{+}_{+1+\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= \frac{k^1-ik^2}{M}\phi^{(2)}(x,\boldsymbol{k}_\perp^2), \\
\psi^{+}_{+1-\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= \phi^{(1)}(x,\boldsymbol{k}_\perp^2), \\
\psi^{+}_{-1+\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= -\frac{k^1+ik^2}{M}(1-x)\phi^{(2)}(x,\boldsymbol{k}_\perp^2), \\
\psi^{+}_{-1-\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= 0.
\end{aligned} \tag{21}$$

The Fock-state expansion for a nucleon with helicity $\lambda=-1/2$ (we use the shorten notation $-$) has the form

$$\begin{aligned}
|\Psi^{-}_{\text{two particle}}(p^+,\boldsymbol{p}_\perp=\boldsymbol{0}_\perp)\rangle = \int \frac{d^2\boldsymbol{k}_\perp dx}{16\pi^3\sqrt{x(1-x)}} \\
\times\Big[\psi^{-}_{+1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|+1,+\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle+\psi^{-}_{+1-\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|+1,-\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle \\
+\psi^{-}_{-1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|-1,+\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle+\psi^{-}_{-1-\frac{1}{2}}(x,\boldsymbol{k}_\perp)\Big|-1,-\frac{1}{2};xp^+,\boldsymbol{k}_\perp\Big\rangle\Big],
\end{aligned} \tag{22}$$

where

$$\begin{aligned}
\psi^{-}_{+1+\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= 0, \\
\psi^{-}_{+1-\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= -[\psi^{+}_{-1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)]^* \\
&= \frac{k^1-ik^2}{M}(1-x)\phi^{(2)}(x,\boldsymbol{k}_\perp^2), \\
\psi^{-}_{-1+\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= +[\psi^{+}_{+1-\frac{1}{2}}(x,\boldsymbol{k}_\perp)]^* \\
&= \phi^{(1)}(x,\boldsymbol{k}_\perp^2), \\
\psi^{-}_{-1-\frac{1}{2}}(x,\boldsymbol{k}_\perp) &= -[\psi^{+}_{+1+\frac{1}{2}}(x,\boldsymbol{k}_\perp)]^* \\
&= -\frac{k^1+ik^2}{M}\phi^{(2)}(x,\boldsymbol{k}_\perp^2).
\end{aligned} \tag{23}$$

The functions $\phi^{(1)}(x,\boldsymbol{k}_\perp^2)$ and $\phi^{(2)}(x,\boldsymbol{k}_\perp^2)$ in Eqs. (21,23) can be expressed in terms of the gluon PDFs

$$G^{\pm}(x) =\frac{G(x)\pm\Delta G(x)}{2}, \tag{24}$$

where

$$\begin{aligned} G(x) =&f_1^g(x) = \int d^2\boldsymbol{k}_\perp f_1^g(x,\boldsymbol{k}_\perp^2),\\ \Delta G(x) =&g_{1L}^g(x) = \int d^2\boldsymbol{k}_\perp g_{1L}^g(x,\boldsymbol{k}_\perp^2), \end{aligned} \tag{25}$$

as

$$\begin{aligned} \phi^{(1)}(x,\boldsymbol{k}_\perp^2) =&\frac{4\pi}{\kappa}\sqrt{G^+(x)}\beta(x)\sqrt{D_g(x)}\exp\Big[-\frac{\boldsymbol{k}_\perp^2}{2\kappa^2}D_g(x)\Big],\\ \frac{1}{M}\phi^{(2)}(x,\boldsymbol{k}_\perp^2) =&\frac{4\pi}{\kappa^2}\sqrt{G^-(x)}\frac{D_g(x)}{1-x}\exp\Big[-\frac{\boldsymbol{k}_\perp^2}{2\kappa^2}D_g(x)\Big], \end{aligned} \tag{26}$$

where

$$\beta(x) =\sqrt{1-\frac{G^-(x)}{G^+(x)(1-x)^2}}. \tag{27}$$

The parameter κ is related to the nucleon mass, as $M=2\sqrt{2}\kappa$ [57, 66]. The profile function $D_g(x)$ can be connected to the profile function $f_g(x)$ as [67]

$$D_g(x) =-\frac{\log[1-[f_g(x)]^{2/5}(1-x)^2]}{(1-x)^2}, \tag{28}$$

where $f_g(x)$ is the solution of the following differential equation with respect to x:

$$[-f_g(x)(1-x)^5]' = xG(x). \tag{29}$$

Based on the above light-cone wave functions, the overlap representations of the contracted correlators in Eqs. (6-13) can be expressed as

$$W^g_{\lambda,\lambda}(x,P,\Delta,\boldsymbol{k}_\perp) = N\sum_{\lambda_g,\lambda_X}\psi^{\lambda\star}_{\lambda_g,\lambda_X}(x^{\rm out},\boldsymbol{k}_\perp^{\rm out})\psi^{\lambda}_{\lambda_g,\lambda_X}(x^{\rm in},\boldsymbol{k}_\perp^{\rm in})(\epsilon^1_{\lambda_g}\epsilon^{1*}_{\lambda_g}+\epsilon^2_{\lambda_g}\epsilon^{2*}_{\lambda_g}), \tag{30}$$

$$W^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) = -N\sum_{\lambda_g,\lambda_X}\psi^{-\star}_{\lambda_g,\lambda_X}(x^{\rm out},\boldsymbol{k}_\perp^{\rm out})\psi^{+}_{\lambda_g,\lambda_X}(x^{\rm in},\boldsymbol{k}_\perp^{\rm in})(\epsilon^1_{\lambda_g}\epsilon^{1*}_{\lambda_g}+\epsilon^2_{\lambda_g}\epsilon^{2*}_{\lambda_g}), \tag{31}$$

$$\widetilde{W}^g_{\lambda,\lambda}(x,P,\Delta,\boldsymbol{k}_\perp) = iN\sum_{\lambda_g,\lambda_X}\psi^{\lambda\star}_{\lambda_g,\lambda_X}(x^{\rm out},\boldsymbol{k}_\perp^{\rm out})\psi^{\lambda}_{\lambda_g,\lambda_X}(x^{\rm in},\boldsymbol{k}_\perp^{\rm in})(\epsilon^1_{\lambda_g}\epsilon^{2*}_{\lambda_g}-\epsilon^2_{\lambda_g}\epsilon^{1*}_{\lambda_g}), \tag{32}$$

$$\widetilde{W}^g_{-,+}(x,P,\Delta,\boldsymbol{k}_\perp) = -iN\sum_{\lambda_g,\lambda_X}\psi^{-\star}_{\lambda_g,\lambda_X}(x^{\rm out},\boldsymbol{k}_\perp^{\rm out})\psi^{+}_{\lambda_g,\lambda_X}(x^{\rm in},\boldsymbol{k}_\perp^{\rm in})(\epsilon^1_{\lambda_g}\epsilon^{2*}_{\lambda_g}-\epsilon^2_{\lambda_g}\epsilon^{1*}_{\lambda_g}), \tag{33}$$

where $N=1/(2(2\pi)^3)$. The gluon polarization vectors $\epsilon^\mu_{\lambda_g}$ read

$$\epsilon^\mu_{\pm} = (0,0,\boldsymbol{\epsilon}_\pm) = \frac{1}{\sqrt{2}}(0,0,\mp1,-i). \tag{34}$$

The initial and final transverse momenta of the struck gluon are given by

$$\boldsymbol{k}_\perp^{\rm in} =\boldsymbol{k}_\perp-(1-x^{\rm in})\frac{\boldsymbol{\Delta}_\perp}{2}, \qquad \text{with} \qquad x^{\rm in}=\frac{x+\xi}{1+\xi}, \tag{35}$$

$$\boldsymbol{k}_\perp^{\rm out} =\boldsymbol{k}_\perp+(1-x^{\rm out})\frac{\boldsymbol{\Delta}_\perp}{2}, \qquad \text{with} \qquad x^{\rm out}=\frac{x-\xi}{1-\xi}, \tag{36}$$

respectively.

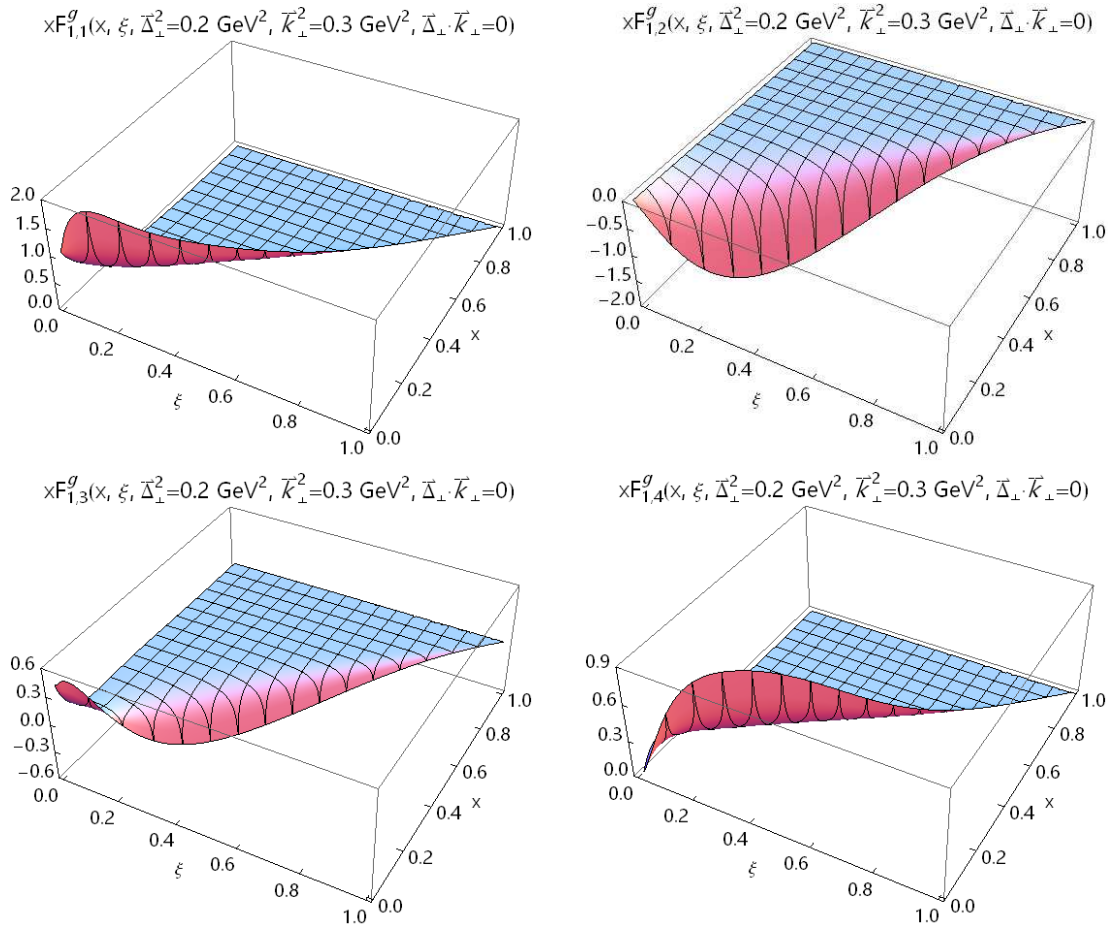


FIG. 1. The GTMDs (timed with x) as functions of x and ξ for an unpolarized gluon at fixed $\boldsymbol{\Delta}_\perp^2 = 0.2$ GeV^2 and $\boldsymbol{k}_\perp^2 = 0.3$ GeV^2.

IV. NUMERICAL RESULTS

In this section, we present the numerical results of the gluon GTMDs and IPDs. By substituting the light-cone wave functions into the overlap representation of the correlators, we get the analytical results for the gluon GTMDs in terms of the unpolarized gluon PDF $G(x)$ and the longitudinally polarized gluon PDF $\Delta G(x)$, which can be taken from world data analysis. In Ref. [68], the authors provide three parametrizations for the gluon PDFs based on the QCD formalism developed in Refs. [69, 70] and the NNPDF global analysis [71, 72]. They are almost equivalent and we choose the "ansatz-3" parametrization, which reads [68]

$$\begin{aligned} xG^+(x) =& x^\alpha[A(1-x)^{4+\beta} + B(1-x)^{5+\beta}](1+\gamma\sqrt{x}+\delta x), \\ xG^-(x) =& x^\alpha[A(1-x)^{6+\beta'} + B(1-x)^{7+\beta'}](1+\gamma'\sqrt{x}+\delta' x), \end{aligned} \tag{37}$$

where the parameters are fixed as

$$\begin{aligned} &\alpha = 0.034, \quad \beta = 0.54, \quad \gamma = -2.63, \quad \delta = 2.54, \\ &\beta' = 2.00, \qquad \gamma' = -2.80, \quad \delta' = 2.25. \end{aligned} \tag{38}$$

The normalization parameters A and B are determined from the moments of the gluon PDFs $\langle x_g \rangle$ and ΔG, where the former (the second-x moment of $G(x)$) represents the total gluon momentum fraction and the latter (the first-x moment of $\Delta G(x)$) represents the total gluon helicity. We take $A = 14.09$ and $B = -11.24$ by globally analysing the unpolarized gluon PDF in Ref. [68]. Finally, in view of the good fit between the model results and the experimental data in Ref. [55], the parameter κ is fixed as 0.383 GeV.

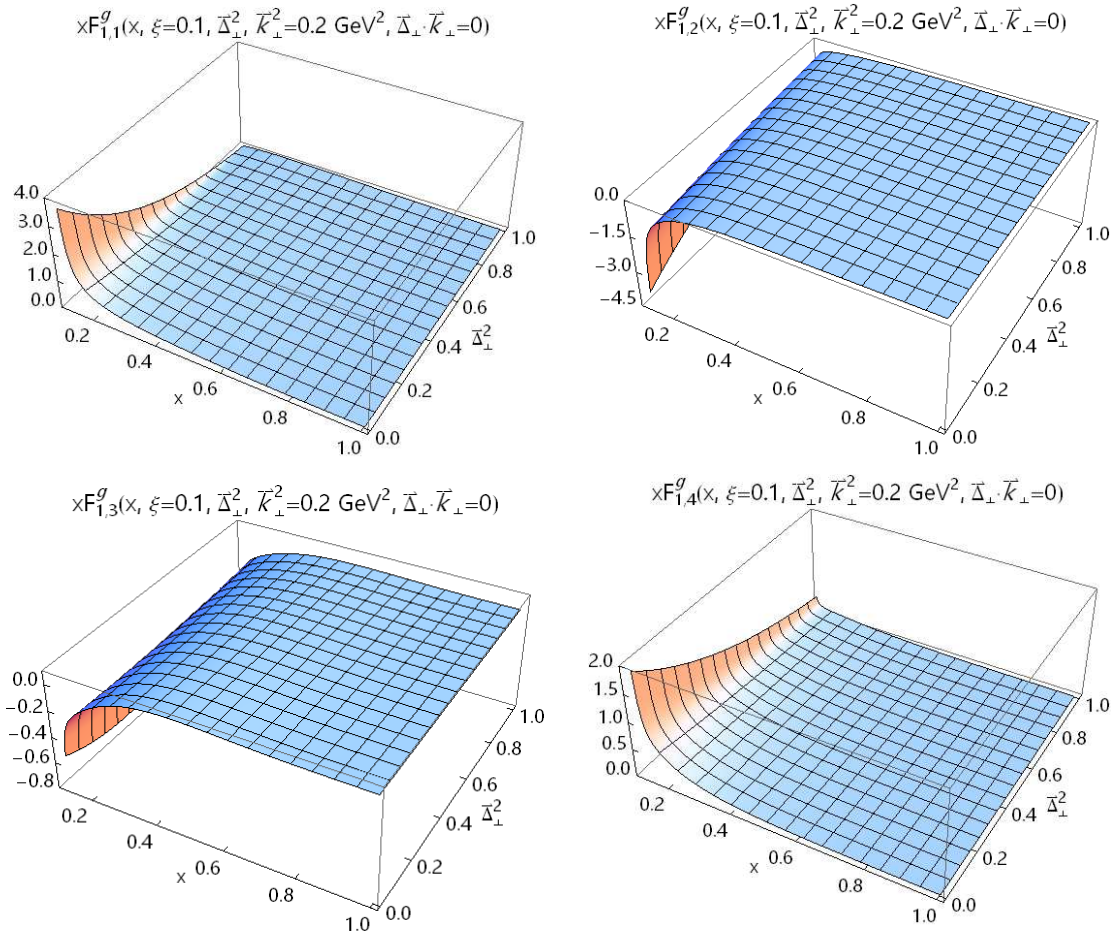


FIG. 2. The GTMDs (timed with x) as functions of x and $\boldsymbol{\Delta}_\perp^2$ for an unpolarized gluon at fixed $\xi = 0.1$ and $\boldsymbol{k}_\perp^2 = 0.2$ GeV^2.

A. GTMDs for an unpolarized gluon

In Fig. 1, we present the dependence of the GTMDs (timed with x) for an unpolarized gluon in a nucleon on x and ξ at fixed $\boldsymbol{\Delta}_\perp^2 = 0.2$ GeV^2 and $\boldsymbol{k}_\perp^2 = 0.3$ GeV^2 with $\boldsymbol{\Delta}_\perp \perp \boldsymbol{k}_\perp$. We observe that all the distributions exhibit the accessibility in the DGLAP region $x > \xi$, where $xF^g_{1,1}$ and $xF^g_{1,4}$ show the positive distributions, $xF^g_{1,2}$ shows the negative distribution, while $xF_{1,3}$ is positive in the lower-$x(< 0.1)$ region and is negative in the higher-x region. Referring to the definition of $F_{1,3}$ in Eq. (8), we find that $F^g_{1,1}$ dominates at low-x, which leads to the positive distribution of $xF^g_{1,3}$ at $x < 0.1$. As the longitudinal momentum transfer to the nucleon Δ^+ increases, the peaks of the GTMDs (timed with x), which are always at the limit of $x \to \xi$, shift toward the higher values of x, and the magnitudes decrease from the low-x until they disappear. Our model results are obtained in the light-cone gauge, so the canonical gluon OAM l^g_z can be defined through $F^g_{1,4}$ at $\xi = 0$ and $\boldsymbol{\Delta}_\perp = 0$ as

$$l^g_z = -\int dx d^2\boldsymbol{k}_\perp \frac{\boldsymbol{k}_\perp^2}{M^2} F^g_{1,4}(x, 0, 0, \boldsymbol{k}_\perp^2, 0), \tag{39}$$

which gives the intrinsic gluon OAM, independent from spectator interactions, and encodes the correlation between the gluon OAM and the nucleon spin. The positive distribution of $F^g_{1,4}$ indicates that $l^g_z < 0$, that is, the total gluon OAM will reduce the contribution of the gluon angular momentum to the nucleon spin. On the other hand, this result also means that the gluon OAM and the nucleon spin tend to be antialigned, which is consistent with the result of the light-cone spectator model [73].

Fig. 2 shows the dependence of the unpolarized gluon GTMDs (timed with x) on x and $\boldsymbol{\Delta}_\perp^2$ at fixed $\xi = 0.1$ and $\boldsymbol{k}_\perp^2 = 0.2$ GeV^2 with $\boldsymbol{\Delta}_\perp \perp \boldsymbol{k}_\perp$. We note that the overall shape of all the plots is similar. The peaks of these distributions are always at the limit of $x \to 0$ and the magnitudes decrease with increasing $\boldsymbol{\Delta}_\perp^2$. On the other hand, with the increase of the light-cone fraction x, which indicates that the gluon carries a larger portion of the nucleon

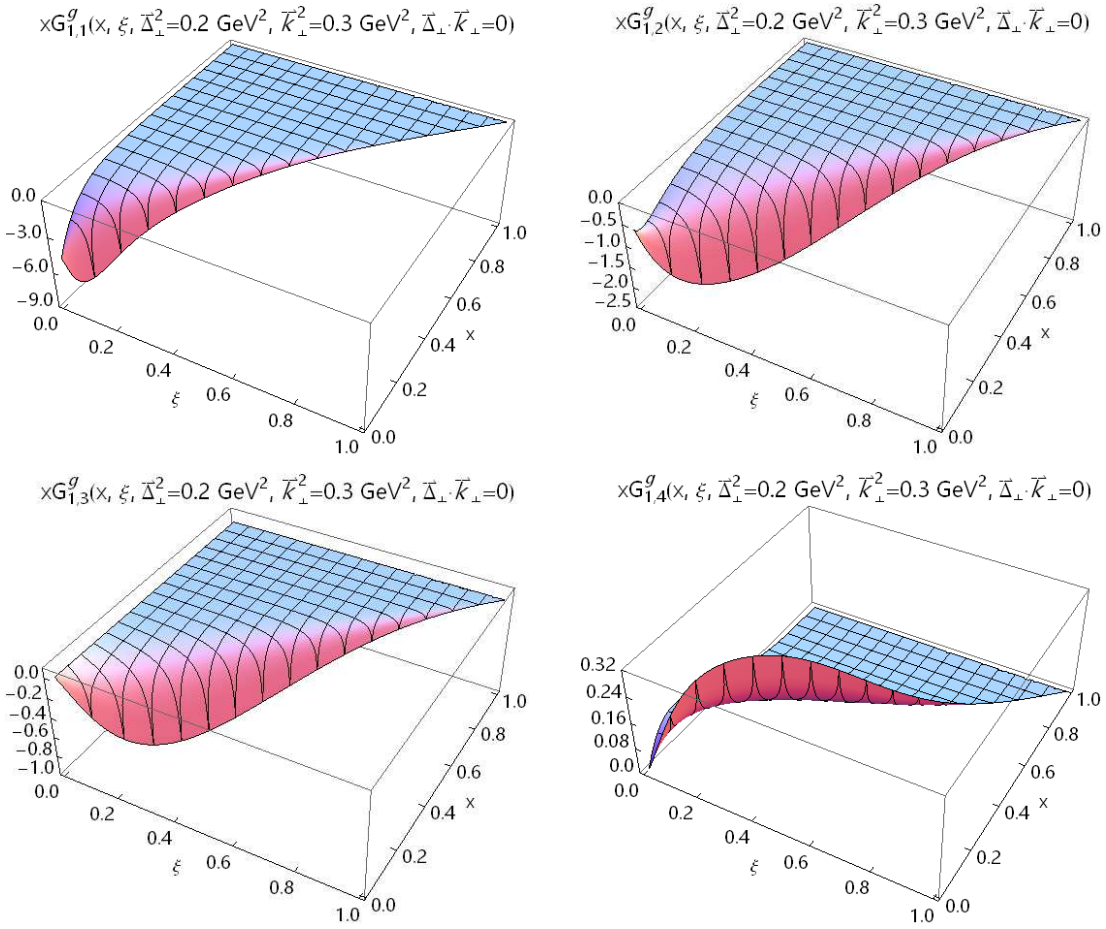


FIG. 3. The GTMDs (timed with x) as functions of x and ξ for a longitudinally polarized gluon at fixed $\boldsymbol{\Delta}_\perp^2 = 0.2$ GeV^2 and $\boldsymbol{k}_\perp^2 = 0.3$ GeV^2.

momentum, the magnitudes also decrease, but at a much higher rate. Therefore, the main contributions of the gluon GTMDs are concentrated at low-x, and this is also a common property of the gluon PDFs, TMDs and GPDs [6, 55]. In the TMD limit ($\Delta = 0$), the T-even part of $F^g_{1,1}$ projects to the unpolarized gluon TMD $f_1^g(x, \boldsymbol{k}_\perp^2)$, and the T-odd part of $F^g_{1,2}$ is linked to the gluon Sivers function $-f_{1T}^{\perp g}(x, \boldsymbol{k}_\perp^2)$ [24, 74]. In the GPD limit (integrating over $\boldsymbol{k}_\perp$), the projection from the GTMDs $F^g_{1,1}$, $F^g_{1,2}$ and $F^g_{1,3}$ to the GPDs H^g and E^g have been shown in Eqs. (14,15).

B. GTMDs for a longitudinally polarized gluon

In Fig. 3, we present the dependence of the GTMDs (timed with x) for a longitudinally polarized gluon in a nucleon on x and ξ at fixed $\boldsymbol{\Delta}_\perp^2 = 0.2$ GeV^2 and $\boldsymbol{k}_\perp^2 = 0.3$ GeV^2 with $\boldsymbol{\Delta}_\perp \perp \boldsymbol{k}_\perp$. As mentioned earlier, all the distributions are only accessible in the DGLAP region $x > \xi$. We observe that $xG^g_{1,1}$ and $xG^g_{1,2}$ are negative, $xG^g_{1,4}$ is positive, while $xG^g_{1,3}$ is negative at lower-x(< 0.5) and it is slightly positive at higher-x. Referring to the definition of $G_{1,3}$ in Eq. (12), we find that the contribution of $G^g_{1,4}$ is largely suppressed in the whole $x > \xi$ region. Moreover, the qualitative behaviors of the peaks and magnitudes of these distributions with increasing Δ^+ are the same as those of the unpolarized GTMDs. Similarly, the canonical gluon spin-orbit correlations C_z^g can be defined through $G^g_{1,1}$ at $\Delta = 0$ as

$$C_z^g = \int dx d^2\boldsymbol{k}_\perp \frac{\boldsymbol{k}_\perp^2}{M^2} G^g_{1,1}(x, 0, 0, \boldsymbol{k}_\perp^2, 0). \tag{40}$$

The negative $G^g_{1,1}$ implies that the gluon spin and OAM tend to be antialigned in our model, which is also consistent with the result of the light-cone spectator model [73]. In the TMD limit, the T-even part of $G^g_{1,4}$ reduces to the gluon

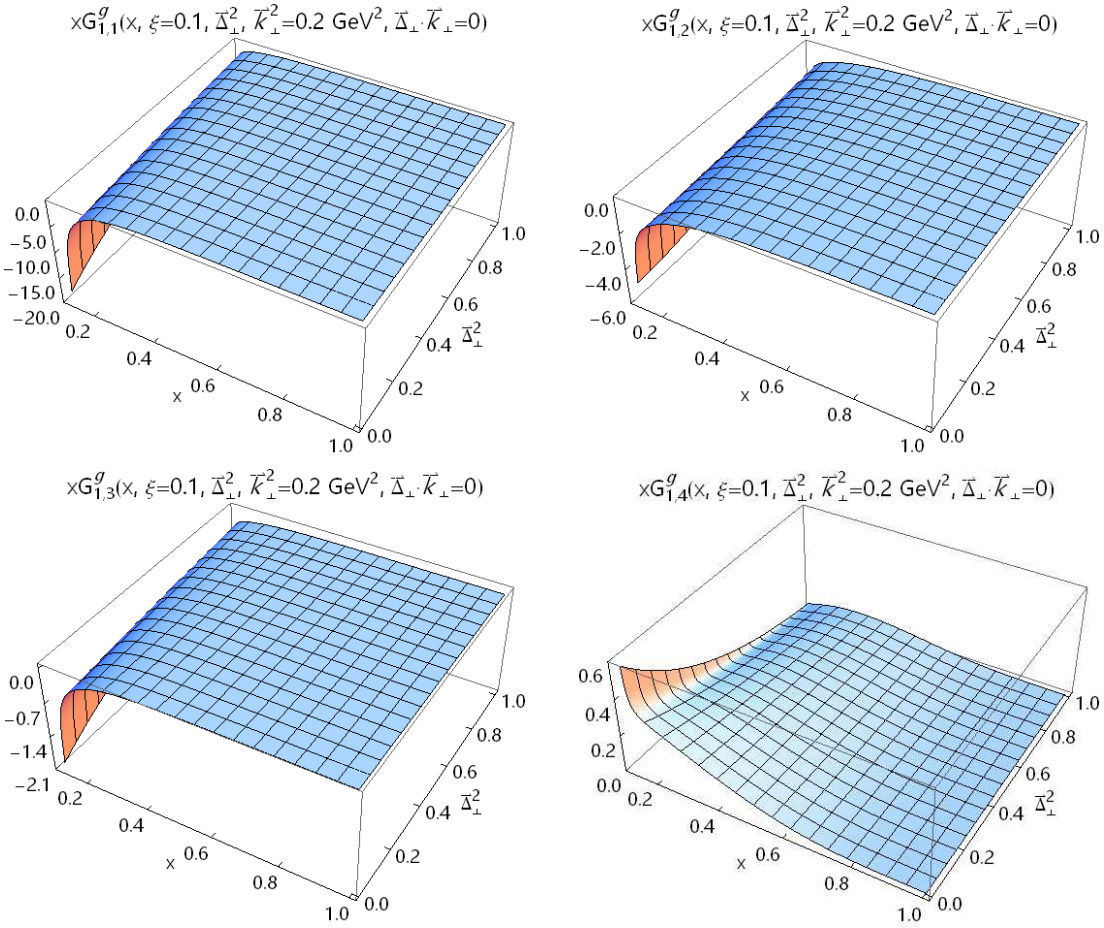


FIG. 4. The GTMDs (timed with x) as functions of x and $\boldsymbol{\Delta}_\perp^2$ for a longitudinally polarized gluon at fixed $\xi = 0.1$ and $\boldsymbol{k}_\perp^2 = 0.2$ GeV^2.

helicity TMD $g_{1L}^g(x, \boldsymbol{k}_\perp^2)$, and the T-even part of $G_{1,2}^g$ reduces to the gluon worm-gear TMD $g_{1T}^g(x, \boldsymbol{k}_\perp^2)$ [24, 74]. In the GPD limit, the projection from the GTMDs $G_{1,2}^g$, $G_{1,3}^g$ and $G_{1,4}^g$ to the GPDs $\tilde{H}^g$ and $\tilde{E}^g$ have been shown in Eqs. (16,17).

In Fig. 4, we present the dependence of the longitudinally polarized gluon GTMDs (timed with x) on x and $\boldsymbol{\Delta}_\perp^2$ at fixed $\xi = 0.1$ and $\boldsymbol{k}_\perp^2 = 0.2$ GeV^2 with $\boldsymbol{\Delta}_\perp \perp \boldsymbol{k}_\perp$. We find that the qualitative properties of these distributions are similar to those of the unpolarized GTMDs.

C. IPDs $\mathcal{H}^g$, $\mathcal{E}^g$ and $\widetilde{\mathcal{H}}^g$

In Fig. 5(a-c), we present the model results of the gluon IPDs $\mathcal{H}^g(x, \boldsymbol{b}_\perp)$, $\mathcal{E}^g(x, \boldsymbol{b}_\perp)$ and $\widetilde{\mathcal{H}}^g(x, \boldsymbol{b}_\perp)$ as functions of $b_\perp$ for $x = 0.1$, 0.2 and 0.3, respectively. We notice that these distributions have the same trends. On the one hand, they approach to 0 with the increase of $b_\perp$. On the other hand, the magnitudes decrease with increasing x. $\mathcal{E}^g$ shows the negative distribution, while $\widetilde{\mathcal{H}}^g$ is positive and it decreases monotonically. Furthermore, $\mathcal{H}^g$ is also positive and it has a peak at $b_\perp > 0$ when $x < 0.2$.

Fig. 5(d-f) show $x\mathcal{H}^g(x, \boldsymbol{b}_\perp)$, $x\mathcal{E}^g(x, \boldsymbol{b}_\perp)$ and $x\widetilde{\mathcal{H}}^g(x, \boldsymbol{b}_\perp)$ as functions of x for $b_\perp = 0.1$, 0.2 and 0.3 fm, respectively. The overall shape of all the plots is similar, and the polarity of each distribution is the same as that of the corresponding distribution in Fig. 5(a-c). We observe that all the distributions vanish at $x = 1$ and also have the tendency to approach to 0 at lower-x(< 0.001). The peaks of these distributions appear at low-x and shift toward the lower values of x with the increase of $b_\perp$. The magnitudes of $x\mathcal{E}^g$ and $x\widetilde{\mathcal{H}}^g$ decrease as $b_\perp$ increases, while $x\mathcal{H}^g$ has this trend only when $x > 0.2$, which is consistent with the feature shown in Fig. 5(a).

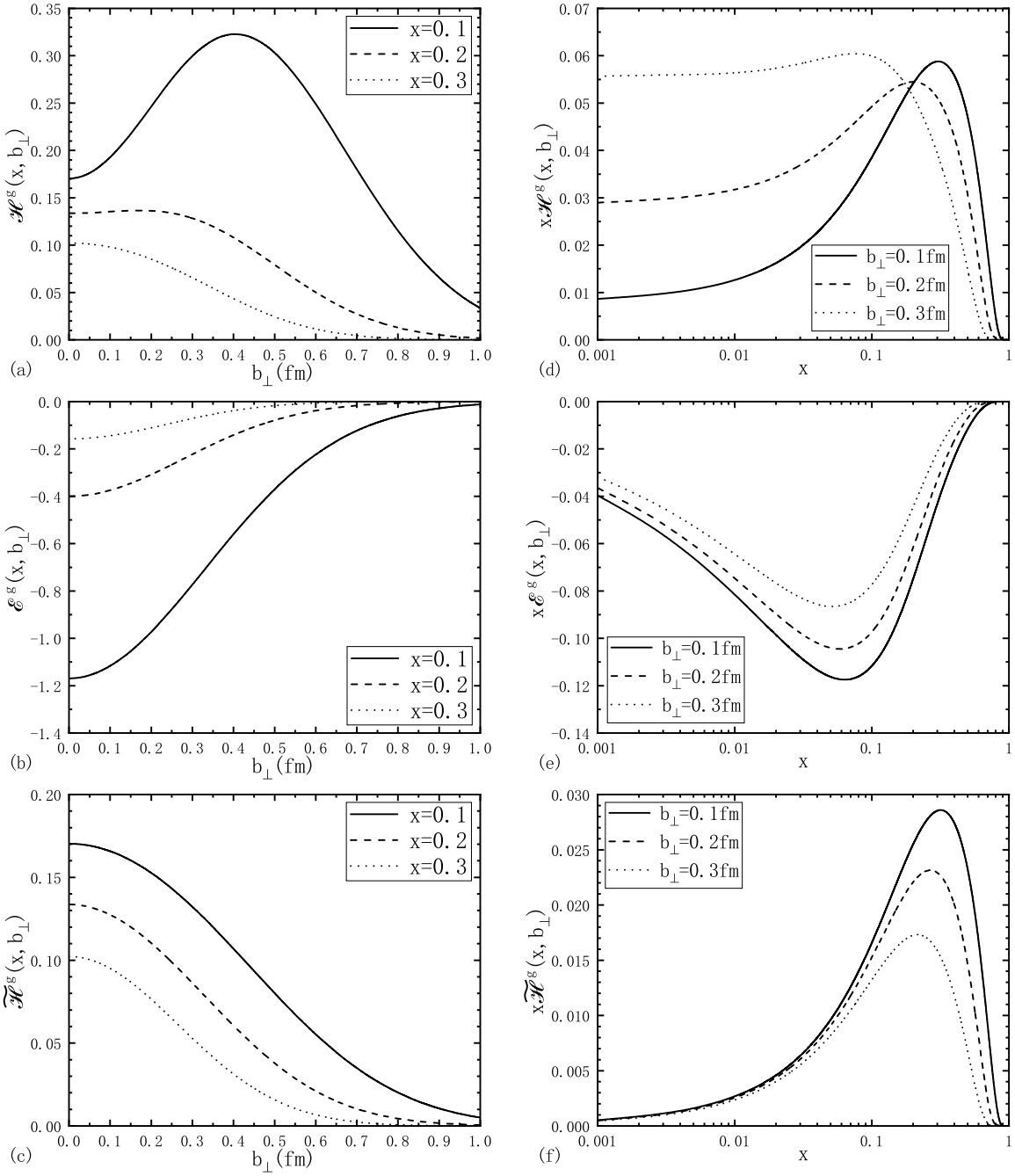


FIG. 5. The left plane: The gluon IPDs $\mathcal{H}^g(x, \boldsymbol{b}_\perp)$, $\mathcal{E}^g(x, \boldsymbol{b}_\perp)$ and $\widetilde{\mathcal{H}}^g(x, \boldsymbol{b}_\perp)$ as functions of positive $b_\perp$ for $x = 0.1$, 0.2 and 0.3. The right plane: The gluon IPDs (timed with x) $x\mathcal{H}^g(x, \boldsymbol{b}_\perp)$, $x\mathcal{E}^g(x, \boldsymbol{b}_\perp)$ and $x\widetilde{\mathcal{H}}^g(x, \boldsymbol{b}_\perp)$ as functions of x for $b_\perp = 0.1$, 0.2 and 0.3 fm.

V. CONCLUSIONS

We investigated the leading-twist GTMDs of the unpolarized and longitudinally polarized gluons inside the nucleon by applying a light-front gluon-triquark model for the nucleon motivated by soft-wall AdS/QCD. We obtained the numerical results for the GTMDs, showing that all the distributions are accessible in the DGLAP region $x > \xi$. The peaks of these distributions appear at the limit of $x \to \xi$ and shift toward higher values of x with the increase of Δ^+, and the magnitudes decrease as x increases. An observation is that the main contributions of the gluon GTMDs are concentrated at low-x. In addition, we found that the gluon OAM and the nucleon spin tend to be antialigned in our

model, so is the gluon spin and OAM. The former also implies that the total gluon OAM may reduce the total angular momentum contribution of the gluon to the nucleon spin. Finally, we studied the gluon IPDs $\mathcal{H}^g(x, \boldsymbol{b}_\perp)$, $\mathcal{E}^g(x, \boldsymbol{b}_\perp)$ and $\widetilde{\mathcal{H}}^g(x, \boldsymbol{b}_\perp)$. The model results showed that $\mathcal{H}^g$ and $\widetilde{\mathcal{H}}^g$ are positive, while $\mathcal{E}^g$ is negative. The peaks of the IPDs (timed with x) appear at low-x and shift toward lower values of x with the increase of $b_\perp$, and the magnitudes of the IPDs decrease with the increase of x. This study can provide better understanding on the information of nucleon structure in transverse momentum/position space in terms of gluon.

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China under grant number 12150013.

[1] J. C. Collins and D. E. Soper, Nucl. Phys. B **194**, 445-492 (1982)
[2] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Eur. Phys. J. C **4**, 463-496 (1998) [arXiv:hep-ph/9803445 [hep-ph]].
[3] M. Gluck, E. Reya and A. Vogt, Z. Phys. C **67**, 433-448 (1995)
[4] M. Glück, E. Reya and A. Vogt, Eur. Phys. J. C **5**, 461-470 (1998) [arXiv:hep-ph/9806404 [hep-ph]].
[5] P. J. Mulders and J. Rodrigues, Phys. Rev. D **63**, 094021 (2001) [arXiv:hep-ph/0009343 [hep-ph]].
[6] S. Meissner, A. Metz and K. Goeke, Phys. Rev. D **76**, 034002 (2007) [arXiv:hep-ph/0703176 [hep-ph]].
[7] P. J. Mulders and R. D. Tangerman, Nucl. Phys. B **461**, 197-237 (1996) [erratum: Nucl. Phys. B **484**, 538-540 (1997)] [arXiv:hep-ph/9510301 [hep-ph]].
[8] A. Bacchetta, M. Diehl, K. Goeke, A. Metz, P. J. Mulders and M. Schlegel, JHEP **02**, 093 (2007) [arXiv:hep-ph/0611265 [hep-ph]].
[9] V. Barone, A. Drago and P. G. Ratcliffe, Phys. Rept. **359**, 1-168 (2002) [arXiv:hep-ph/0104283 [hep-ph]].
[10] S. J. Brodsky, D. S. Hwang and I. Schmidt, Phys. Lett. B **530**, 99-107 (2002) [arXiv:hep-ph/0201296 [hep-ph]].
[11] A. Bacchetta, F. Delcarro, C. Pisano, M. Radici and A. Signori, JHEP **06**, 081 (2017) [erratum: JHEP **06**, 051 (2019)] [arXiv:1703.10157 [hep-ph]].
[12] D. Müller, D. Robaschik, B. Geyer, F. M. Dittes and J. Hořejši, Fortsch. Phys. **42** (1994), 101-141 [arXiv:hep-ph/9812448 [hep-ph]].
[13] X. D. Ji, Phys. Rev. D **55**, 7114-7125 (1997) [arXiv:hep-ph/9609381 [hep-ph]].
[14] A. V. Radyushkin, Phys. Rev. D **56**, 5524 (1997) [arXiv:hep-ph/9704207].
[15] D. Mueller, D. Robaschik, B. Geyer, F. M. Dittes and J. Horejsi, Fortsch. Phys. **42**, 101 (1994) [arXiv:hep-ph/9812448].
[16] K. Goeke, M. V. Polyakov and M. Vanderhaeghen, Prog. Part. Nucl. Phys. **47**, 401-515 (2001) [arXiv:hep-ph/0106012 [hep-ph]].
[17] M. Diehl, Phys. Rept. **388**, 41-277 (2003) [arXiv:hep-ph/0307382 [hep-ph]].
[18] X. Ji, Ann. Rev. Nucl. Part. Sci. **54** (2004), 413-450 doi:10.1146/annurev.nucl.54.070103.181302
[19] A. V. Belitsky and A. V. Radyushkin, Phys. Rept. **418**, 1-387 (2005) [arXiv:hep-ph/0504030 [hep-ph]].
[20] S. Boffi and B. Pasquini, Riv. Nuovo Cim. **30** (2007) no.9, 387-448 doi:10.1393/ncr/i2007-10025-7 [arXiv:0711.2625 [hep-ph]].
[21] M. Burkardt, Phys. Rev. D **62**, 071503 (2000) [erratum: Phys. Rev. D **66**, 119903 (2002)] [arXiv:hep-ph/0005108 [hep-ph]].
[22] M. Burkardt, Int. J. Mod. Phys. A **18**, 173-208 (2003) [arXiv:hep-ph/0207047 [hep-ph]].
[23] S. Meissner, A. Metz, M. Schlegel and K. Goeke, JHEP **08**, 038 (2008) [arXiv:0805.3165 [hep-ph]].
[24] S. Meissner, A. Metz and M. Schlegel, JHEP **08**, 056 (2009) [arXiv:0906.5323 [hep-ph]].
[25] S. Bhattacharya, A. Metz and J. Zhou, Phys. Lett. B **771**, 396-400 (2017) [erratum: Phys. Lett. B **810**, 135866 (2020)] [arXiv:1702.04387 [hep-ph]].
[26] Y. Hagiwara, Y. Hatta, R. Pasechnik, M. Tasevsky and O. Teryaev, Phys. Rev. D **96**, no.3, 034009 (2017) [arXiv:1706.01765 [hep-ph]].
[27] X. Ji, F. Yuan and Y. Zhao, Phys. Rev. Lett. **118**, no.19, 192004 (2017) [arXiv:1612.02438 [hep-ph]].
[28] Y. Hatta, Y. Nakagawa, F. Yuan, Y. Zhao and B. Xiao, Phys. Rev. D **95**, no.11, 114032 (2017) [arXiv:1612.02445 [hep-ph]].
[29] S. Bhattacharya, R. Boussarie and Y. Hatta, Phys. Rev. Lett. **128**, no.18, 182002 (2022) [arXiv:2201.08709 [hep-ph]].
[30] C. Lorce and B. Pasquini, Phys. Rev. D **84**, 014015 (2011) [arXiv:1106.0139 [hep-ph]].
[31] Y. Hatta, Phys. Lett. B **708**, 186-190 (2012) [arXiv:1111.3547 [hep-ph]].
[32] X. Ji, X. Xiong and F. Yuan, Phys. Rev. Lett. **109**, 152005 (2012) [arXiv:1202.2843 [hep-ph]].
[33] C. Lorce, Phys. Lett. B **719**, 185-190 (2013) [arXiv:1210.2581 [hep-ph]].
[34] C. Lorcé, Phys. Lett. B **735**, 344-348 (2014) [arXiv:1401.7784 [hep-ph]].
[35] C. Tan and Z. Lu, Phys. Rev. D **105**, no.3, 034004 (2022) [arXiv:2110.08502 [hep-ph]].
[36] C. Lorce, B. Pasquini, X. Xiong and F. Yuan, Phys. Rev. D **85**, 114006 (2012) [arXiv:1111.4827 [hep-ph]].
[37] C. Lorce, B. Pasquini and M. Vanderhaeghen, JHEP **05**, 041 (2011) [arXiv:1102.4704 [hep-ph]].
[38] A. Mukherjee, S. Nair and V. K. Ojha, Phys. Rev. D **90**, no.1, 014024 (2014) [arXiv:1403.6233 [hep-ph]].

[39] A. Mukherjee, S. Nair and V. K. Ojha, Phys. Rev. D **91**, no.5, 054018 (2015) [arXiv:1501.03728 [hep-ph]].
[40] J. More, A. Mukherjee and S. Nair, Phys. Rev. D **95**, no.7, 074039 (2017) [arXiv:1701.00339 [hep-ph]].
[41] T. Liu and B. Q. Ma, Phys. Rev. D **91**, 034019 (2015) [arXiv:1501.07690 [hep-ph]].
[42] D. Chakrabarti, T. Maji, C. Mondal and A. Mukherjee, Eur. Phys. J. C **76**, no.7, 409 (2016) [arXiv:1601.03217 [hep-ph]].
[43] D. Chakrabarti, T. Maji, C. Mondal and A. Mukherjee, Phys. Rev. D **95**, no.7, 074028 (2017) [arXiv:1701.08551 [hep-ph]].
[44] D. Chakrabarti, N. Kumar, T. Maji and A. Mukherjee, Eur. Phys. J. Plus **135**, no.6, 496 (2020) [arXiv:1902.07051 [hep-ph]].
[45] T. Gutsche, V. E. Lyubovitskij and I. Schmidt, Eur. Phys. J. C **77**, no.2, 86 (2017) [arXiv:1610.03526 [hep-ph]].
[46] S. Kaur and H. Dahiya, Adv. High Energy Phys. **2020**, 9429631 (2020) [arXiv:1906.04662 [hep-ph]].
[47] N. Kumar and C. Mondal, Nucl. Phys. B **931**, 226-249 (2018) [arXiv:1705.03183 [hep-ph]].
[48] K. Kanazawa, C. Lorcé, A. Metz, B. Pasquini and M. Schlegel, Phys. Rev. D **90**, no.1, 014028 (2014) [arXiv:1403.5226 [hep-ph]].
[49] Z. L. Ma and Z. Lu, Phys. Rev. D **98**, no.5, 054024 (2018) [arXiv:1808.00140 [hep-ph]].
[50] S. Kaur and H. Dahiya, Phys. Rev. D **100**, no.7, 074008 (2019) [arXiv:1908.01939 [hep-ph]].
[51] N. Kaur and H. Dahiya, Eur. Phys. J. A **56**, no.6, 172 (2020) [arXiv:1909.10146 [hep-ph]].
[52] J. L. Zhang and J. L. Ping, Eur. Phys. J. C **81**, no.9, 814 (2021)
[53] W. Broniowski and E. Ruiz Arriola, Phys. Lett. B **574**, 57-64 (2003) [arXiv:hep-ph/0307198 [hep-ph]].
[54] V. E. Lyubovitskij and I. Schmidt, Phys. Rev. D **104**, no.1, 014001 (2021) [arXiv:2105.07842 [hep-ph]].
[55] V. E. Lyubovitskij and I. Schmidt, Phys. Rev. D **103**, no.9, 094017 (2021) [arXiv:2012.01334 [hep-ph]].
[56] S. J. Brodsky, G. F. de Teramond, H. G. Dosch and J. Erlich, Phys. Rept. **584**, 1-105 (2015) [arXiv:1407.8131 [hep-ph]].
[57] T. Gutsche, V. E. Lyubovitskij and I. Schmidt, Nucl. Phys. B **952**, 114934 (2020) [arXiv:1906.08641 [hep-ph]].
[58] S. J. Brodsky and G. F. de Teramond, Phys. Rev. D **77**, 056007 (2008) [arXiv:0707.3859 [hep-ph]].
[59] Z. Abidin and C. E. Carlson, Phys. Rev. D **79**, 115003 (2009) [arXiv:0903.4818 [hep-ph]].
[60] S. J. Brodsky and G. R. Farrar, Phys. Rev. Lett. **31**, 1153-1156 (1973)
[61] Z. Lu and B. Q. Ma, Phys. Rev. D **94**, no.9, 094022 (2016) [arXiv:1611.00125 [hep-ph]].
[62] C. Lorcé and B. Pasquini, JHEP **09**, 138 (2013) [arXiv:1307.4497 [hep-ph]].
[63] T. Maji, C. Mondal and D. Kang, Phys. Rev. D **105**, no.7, 074024 (2022) [arXiv:2202.08635 [hep-ph]].
[64] A. Bacchetta, F. Conti and M. Radici, Phys. Rev. D **78**, 074010 (2008) [arXiv:0807.0323 [hep-ph]].
[65] S. J. Brodsky, M. Diehl and D. S. Hwang, Nucl. Phys. B **596**, 99-124 (2001) [arXiv:hep-ph/0009254 [hep-ph]].
[66] T. Gutsche, V. E. Lyubovitskij, I. Schmidt and A. Vega, Phys. Rev. D **85**, 076003 (2012) [arXiv:1108.0346 [hep-ph]].
[67] V. E. Lyubovitskij and I. Schmidt, Phys. Rev. D **102**, no.3, 034011 (2020) [arXiv:2005.10163 [hep-ph]].
[68] R. S. Sufian, T. Liu and A. Paul, Phys. Rev. D **103**, no.3, 036007 (2021) [arXiv:2012.01532 [hep-ph]].
[69] S. J. Brodsky and I. A. Schmidt, Phys. Lett. B **234**, 144-150 (1990)
[70] S. J. Brodsky, M. Burkardt and I. Schmidt, Nucl. Phys. B **441**, 197-214 (1995) [arXiv:hep-ph/9401328 [hep-ph]].
[71] R. D. Ball *et al.* [NNPDF], Eur. Phys. J. C **77**, no.10, 663 (2017) [arXiv:1706.00428 [hep-ph]].
[72] E. R. Nocera *et al.* [NNPDF], Nucl. Phys. B **887**, 276-308 (2014) [arXiv:1406.5539 [hep-ph]].
[73] C. Tan and Z. Lu, [arXiv:2312.07997 [hep-ph]].
[74] S. Bhattacharya, A. Metz, V. K. Ojha, J. Y. Tsai and J. Zhou, Phys. Lett. B **833**, 137383 (2022) [arXiv:1802.10550 [hep-ph]].

arXiv:2402.17140v1 [hep-ph] 27 Feb 2024

Direct Detection of Dark Photon Dark Matter with the James Webb Space Telescope

Haipeng An,[1,2,3,4,*] Shuailiang Ge,[3,5,†] Jia Liu,[5,3,‡] and Zhiyao Lu[5,§]

[1]*Department of Physics, Tsinghua University, Beijing 100084, China*

[2]*Center for High Energy Physics, Tsinghua University, Beijing 100084, China*

[3]*Center for High Energy Physics, Peking University, Beijing 100871, China*

[4]*Frontier Science Center for Quantum Information, Beijing 100084, China*

[5]*School of Physics and State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing 100871, China*

(Dated: February 28, 2024)

Abstract

In this study, we propose an investigation into dark photon dark matter (DPDM) within the infrared frequency band, utilizing highly sensitive infrared light detectors commonly integrated into space telescopes, such as the James Webb Space Telescope (JWST). The presence of DPDM induces electron oscillations in the reflector of these detectors. Consequently, these oscillating electrons can emit monochromatic electromagnetic waves with a frequency almost equivalent to the mass of DPDM. By employing the stationary phase approximation, we can demonstrate that when the size of the reflector significantly exceeds the wavelength of the electromagnetic wave, the contribution to the electromagnetic wave field at a given position primarily stems from the surface unit perpendicular to the relative position vector. This simplification results in the reduction of electromagnetic wave calculations to ray optics. By applying this concept to JWST, our analysis of observational data demonstrates the potential to establish constraints on the kinetic mixing between the photon and dark photon within the range [10, 500] THz. Despite JWST not being optimized for DPDM searches, our findings reveal constraints comparable to those obtained from the XENON1T experiment in the laboratory, as well as astrophysical constraints from solar emission. Additionally, we explore strategies to optimize future experiments specifically designed for DPDM searches.

* anhp@mail.tsinghua.edu.cn

† sge@pku.edu.cn

‡ jialiu@pku.edu.cn

§ 2000011457@stu.pku.edu.cn

CONTENTS

I. Introduction 2

II. High Frequency Approximation 4

A. Monochromatic DPDM 5

B. Non-monochromatic DPDM 9

III. The Optical Telescope Element (OTE) of JWST 11

IV. Calculating the Equivalent Flux Density 13

V. Constraints from JWST Observation Data 16

VI. Summary and Outlook 19

Acknowledgement 20

A. Detailed derivation of Eq. (21) 20

B. A Brief Proof of Eq. (8) 21

C. Ray Transfer Matrix Analysis 23

D. Data Analysis Method 26

References 28

I. INTRODUCTION

The elusive nature of dark matter has thus far eluded detection through various non-gravitational search efforts. The scope of potential candidates has expanded beyond traditional weakly interacting massive particles (WIMPs) to encompass a broad range of mass scales. One intriguing category comprises ultralight bosonic dark matter, which has garnered significant attention as the lightest dark matter candidate. Among these, the dark photon is a notable ultralight vector dark matter candidate due to its kinetic mixing marginal operator coupling with the photon field, serving as one of the simplest extensions of the Standard

Model [1–6]. The kinetic mixing dark photon could have been generated in the early universe and holds promise as a viable dark matter candidate [7–10]. Various mechanisms facilitate its plausibility, including the misalignment mechanism coupled with a non-minimal Ricci scalar coupling [8, 9, 11–13], inflationary fluctuations [10, 14–23], parametric resonances [24–29], or the decay of cosmic strings [30].

Due to the vast range of unknown mass of the dark photon dark matter (DPDM), there are various detecting methods accordingly [31, 32]. The relevant searches for dark photon DM are haloscope experiments [33–38], dish antenna experiments [39–42], plasma telescopes [43], CMB spectrum distortion [9, 44] and radio telescopes [45–47]. The searches include direct detection of local DPDM in laboratories and observation on its impact in the early universe.

Recently, we proposed to search for Dark Photon Dark Matter (DPDM) conversion, specifically $A' \to \gamma$, locally using radio telescopes such as FAST and LOFAR [46]. For instance, the FAST radio telescope, equipped with a large dish antenna, converts DPDM into a regular photon field. In each small surface area of FAST, an oscillating electric dipole is generated by the DPDM field, with a frequency matching the DPDM mass. Summing up the contributions from each surface area yields the total converted electromagnetic field. The original proposal utilized a spherical reflector, causing the electromagnetic field to constructively focus on the spherical center [39, 41, 48]. This concept has been previously employed in shielded room-sized experiments by various works, utilizing variations such as plane/parabolic reflectors or dipole antennas [42, 49–55].

In this article, we would like to explore this idea with the recent new telescope, the James Webber Science Telescope (JWST), which is running in space. JWST covers the frequency range of 10–500 THz for infrared astronomy. Our searches for DPDM benefit from JWST's high frequency resolution $R = f/\Delta f$, which ranges from 4 to 3000, depending on different obervation modes [56]. Also, since it works in space, we expect it to have a much lower noise background than terrestrial facilities. The Near Infrared Spectrograph (NISpec) and the Mid-Infrared Instrument (MIRI), two instruments carried by JWST, are especially useful for searching for DPDM. By analyzing data collected by the two instruments, we get the upper limits for the DPDM-photon coupling constant, $\epsilon \sim 10^{-10} - 10^{-12}$ in the frequency range $10 - 500$ THz at 95% confidence level (C.L.).

The Lagrangian of the dark photon model in this work is a vector boson that couples to the SM particles through its kinetic mixing with photon, and the Lagrangian is described

by the equation

$$\mathcal{L} = -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}m^2_{A'}A'_\mu A'^\mu - \frac{1}{2}\epsilon F_{\mu\nu}F'^{\mu\nu} \ , \tag{1}$$

where F and F' are the dark photon and photon field strength, ϵ is the kinematic mixing. After proper rotation and redefinition, one can eliminate the kinematic mixing term and arrive at the interaction Lagrangian for A', the SM photon A, and the electromagnetic current $j^\mu_{\rm em}$,

$$\mathcal{L}_{\rm int} = e j^\mu_{\rm em}(A_\mu - \epsilon A'_\mu) \ , \tag{2}$$

where e is the electromagnetic coupling. Therefore, the DPDM electric field, $\boldsymbol{E}' = -\dot{\boldsymbol{A}}' - \nabla A'^0$, can accelerate the charge carriers in a reflector and thus be converted to SM electromagnetic waves.

This work is structured as follows: In Section I, we present the background of DPDM conversion. In Section II, we offer a mathematical proof of the reduction from complex electroweak wave calculation to ray optics using the stationary phase approximation. Section III applies the simplified calculation to JWST, obtaining the equivalent flux density of the electromagnetic field in Section IV. Section V establishes limits on the kinetic mixing parameter of the dark photon using JWST observational data. Finally, in Section VI, we draw our conclusions.

II. HIGH FREQUENCY APPROXIMATION

In order to calculate the EM signals induced by DPDM on a metal reflector plate, the most direct way is to divide the reflector into many small patches and sum up the induced EM signals over all of them. The length of each small patch is required to be much smaller than the wavelength λ of the induced EW signal while at the same time much larger than the reflector thickness. Consequently, to ensure enough accuracy, the simulation mesh must be fine enough for the distance between mesh points to be smaller than the wavelength. This method works well in the case of FAST telescope [57]. The FAST detects radio photons around 1 GHz and has a reflector roughly 500 m in diameter. Therefore, we only need to divide the FAST reflector into $\sim 10^6$ patches for an accurate simulation. On the other hand, JWST works at a much higher frequency range, $10 - 500$ THz (i.e., the photon wavelength $\lambda \sim 0.6 - 30$ μm), and the diameter of the JWST's primary mirror is $D = 6.6$ meter. To achieve an acceptable level of accuracy in simulating the JWST case, we require over 10^{15}

patches, which is finer than the FAST case by orders of magnitude. However, such a fine mesh imposes an immense computational burden, rendering it unfeasible to simulate signal strength using available computer resources.

Fortunately, as we will see, we can theoretically demonstrate that calculating the induced EM signals becomes considerably more straightforward in the high-frequency regime, $D \gg \lambda$. Similar to the case of reaching the ray optics at the high-frequency limit of the wave optics, we can establish that the strength of the DPDM-induced signal can be obtained using a ray-optical method. This simplification ultimately leads to a set of algebraic equations that can be easily calculated, even manually without the need of a fine-mesh simulation. The physical interpretation of such a simplification is that, in the high-frequency regime, interferences between different patches on the reflector have a negligible impact on the final result.

To be more specific, this simplification works primarily due to two key factors. Firstly, the phases of the electric fields contributed by different patches on a plate vary significantly, while their strengths remain relatively the same. This results in significant cancellations between the electric fields generated by different patches. Secondly, one significant parameter is the coherence length of DPDM. The DPDM and the induced photon have the same energy, but the coherent length of DPDM, λ', is much larger than the wavelengh of the induced photons, λ. This is due to the non-relativistic nature of dark matter with a low speed $v_{\rm DM} \sim 10^{-3}$, which gives $\lambda' \sim \lambda/v_{\rm DM}$. Importantly, λ' is still significantly smaller than the diameter of the JWST's mirrors. As a result, the interferences between different patches are dampened by incoherence.

In the following, we are going to prove that the ray optics is indeed applicable here in calculating the EM signals induced by DPDM for the JWST case which locates in the high-frequency regime, $D/\lambda \sim 10^6$ and $D/\lambda' \sim 10^3$. Finally, a formula for computing the strength of the induced signal will be introduced.

A. Monochromatic DPDM

Firstly, we consider a simplified case that DPDM is monochromatic in frequency. Then, in the next subsection, we discuss the more realistic case where the velocity distribution of dark matter is included. Under the effect of the DPDM's dark electric field, each small

patch on a reflector plate can be treated as an electric dipole,

$$\boldsymbol{p} = 2\epsilon \boldsymbol{A'_\tau} \Delta S, \tag{3}$$

where $\boldsymbol{A'_\tau}$ is the component of $\boldsymbol{A'}$ parallel with the patch and ΔS is the area of the patch [57]. The dipole is oscillating as a result of the oscillation of the DPDM field. Then, summing up the EM radiations from all the dipoles, we arrive at the expressions of the induced EM fields at position $\boldsymbol{r}$ [57],

$$\boldsymbol{E}(\boldsymbol{r}) = -\frac{\epsilon m_{A'}^2 |\boldsymbol{A'}|}{2\pi} \int [\boldsymbol{\tau} \times (\boldsymbol{r} - \boldsymbol{r'})] \times (\boldsymbol{r} - \boldsymbol{r'}) \frac{e^{i m_{A'} |\boldsymbol{r} - \boldsymbol{r'}| + i \boldsymbol{k'} \cdot (\boldsymbol{r'} - \boldsymbol{r})}}{|\boldsymbol{r} - \boldsymbol{r'}|^3} \mathrm{d}S', \tag{4}$$

$$\boldsymbol{B}(\boldsymbol{r}) = -\frac{\epsilon m_{A'}^2 |\boldsymbol{A'}|}{2\pi} \int \boldsymbol{\tau} \times (\boldsymbol{r} - \boldsymbol{r'}) \frac{e^{i m_{A'} |\boldsymbol{r} - \boldsymbol{r'}| + i \boldsymbol{k'} \cdot (\boldsymbol{r'} - \boldsymbol{r})}}{|\boldsymbol{r} - \boldsymbol{r'}|^2} \mathrm{d}S'. \tag{5}$$

$\boldsymbol{k'}$ is the wave vector of DPDM. $\boldsymbol{A'_\tau} = \boldsymbol{\tau} |\boldsymbol{A'}|$ where the tangent vector $\boldsymbol{\tau}$ can be calculated as $\boldsymbol{\tau} \equiv \boldsymbol{n}_0 - (\boldsymbol{n}_0 \cdot \boldsymbol{n}) \boldsymbol{n}$. The two unit vectors, $\boldsymbol{n}_0$ and $\boldsymbol{n}$, represent the direction of $\boldsymbol{A'}$ and the normal direction of $\mathrm{d}S$, respectively. The magnitude of oscillations, $|\boldsymbol{A'}|$, is determined by the dark matter energy density ρ_{DM}, that is,

$$\rho_{\mathrm{DM}} = \frac{1}{2} m_{A'}^2 |\boldsymbol{A'}|^2 = \frac{1}{2} |\boldsymbol{E'}|^2. \tag{6}$$

Then, we can calculate the energy flux density,

$$\langle \boldsymbol{S'} \rangle_t = \frac{1}{2} \mathrm{Re}(\boldsymbol{E} \times \boldsymbol{B}^*). \tag{7}$$

$\langle ... \rangle_t$ here means the average over time. In principle, for any reflectors, we can numerically simulate the DPDM-induced EM waves using these formulas. However, as we discussed above, such a simulation requires a very fine mesh that is hard to realize in computers for the JWST case. As we are going to see below, we figure out a more analytical method applicable in the high-frequency regime.

The key to simplifying our formulas in the high-frequency regime is to use the stationary phase approximation. In general, the stationary phase approximation works in solving the following integral as α tends to infinity [58],

$$\begin{aligned} &\int_{\mathbb{R}^n} g(\boldsymbol{x}) e^{i\alpha f(\boldsymbol{x})} \mathrm{d}^n \boldsymbol{x} \\ &= \sum_{\boldsymbol{x}_0 \in \Sigma} e^{i\alpha f(\boldsymbol{x}_0)} |\det(\mathrm{Hess}(f(\boldsymbol{x}_0)))|^{-1/2} e^{\frac{i\pi}{4} \mathrm{sgn}(\mathrm{Hess}(f(\boldsymbol{x}_0)))} \left(\frac{2\pi}{\alpha}\right)^{\frac{n}{2}} g(\boldsymbol{x}_0) + o(\alpha^{-\frac{n}{2}}), \ \alpha \to \infty. \end{aligned} \tag{8}$$

where the function $g(\boldsymbol{x})$ is either zero or exponentially suppressed when $\boldsymbol{x}$ is large, (the condition where $g(\boldsymbol{x})$ is zero when $\boldsymbol{x}$ is large, can be more accurately described by the mathematical terminology of "compactly supported"). Σ is the set of points where $\nabla f = 0$. $\mathrm{Hess}(f(x_0))$ is the Hessian of f, and $\mathrm{sgn}(\mathrm{Hess}(f(\boldsymbol{x}_0)))$ is the signature of the Hessian,

$$\mathrm{Hess}(f(\boldsymbol{x}_0))_{ij} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}, \tag{9}$$

$$\mathrm{sgn}(\mathrm{Hess}(f(\boldsymbol{x}_0))) = \sharp(\text{positive eigenvalues}) - \sharp(\text{negative eigenvalues}). \tag{10}$$

Note that Eq. (8) is only valid when we assume $\nabla f = 0$ has only discrete solutions, otherwise $\mathrm{Hess}(f(x_0))$ is non-degenerate at $x_0 \in \Sigma$.

This stationary-phase approximation (8) can be intuitively understood in the following way: when k is large, the exponential oscillates rapidly with a small change of $\boldsymbol{x}$, while $g(\boldsymbol{x})$ changes very little. Therefore, the integral vanishes unless we are considering a small patch in $\mathbb{R}^n$ around $\nabla f = 0$, which is the stationary point of the phase factor. A rigorous proof of Eq. (8) is provided in Appendix B.

In the JWST case, the phase factor is $m_{A'}|\boldsymbol{r} - \boldsymbol{r}'| + \boldsymbol{k}' \cdot (\boldsymbol{r}' - \boldsymbol{r})$. Given that the dark photon's wave vector is approximately 10^{-3} times its frequency, the phase factor is dominated by the first term. By rewriting the first term as $m_{A'} D \times (|\boldsymbol{r} - \boldsymbol{r}'|/D)$, with D being the characteristic length of JWST optical elements, the second factor becomes an $\mathcal{O}(1)$ function of spacial coordinates, and $m_{A'} D \gg 1$ by assumption. This is equivalent to the case $\alpha \gg 1$ in Eq. (8). Consequently, we can apply the stationary-phase approximation (8) to calculate Eq. (4). The process of calculating Eq. (5) is the same, so it will not be shown here for the sake of conciseness. In the most general setup, a conductor is a closed surface. In Eq. (4), the domain of integration is the whole conductor surface. By assumption, this integral can be split into several integrals in compact subsets of $\mathbb{R}^2$, with the surface element $\mathrm{d}S$ re-expressed in the form $J(\boldsymbol{r}')\mathrm{d}u\mathrm{d}v$ where J is the Jacobian determined by the equation of the surface. We study each of these integrals separately, and denote by Ω the domain of integration. Define

$$g(\boldsymbol{r}') = \begin{cases} -\dfrac{\epsilon m_{A'}^2 |\boldsymbol{A}'|}{2\pi} \dfrac{[\boldsymbol{\tau}(\boldsymbol{r}') \times (\boldsymbol{r} - \boldsymbol{r}')] \times (\boldsymbol{r} - \boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|^3} J(\boldsymbol{r}') & \boldsymbol{r}' \in \Omega \\ 0 & \boldsymbol{r}' \notin \Omega \end{cases}. \tag{11}$$

Clearly, $g(\boldsymbol{r}')$ is compactly supported. Applying the stationary phase approximation, the

only contribution to the integral comes from points where

$$\frac{\partial}{\partial u}|\boldsymbol{r}-\boldsymbol{r'}(u,v)|=0,\quad \frac{\partial}{\partial v}|\boldsymbol{r}-\boldsymbol{r'}(u,v)|=0, \tag{12}$$

or equivalently,

$$(\boldsymbol{r}-\boldsymbol{r'})\cdot\frac{\partial \boldsymbol{r'}}{\partial u}=0,\quad (\boldsymbol{r}-\boldsymbol{r'})\cdot\frac{\partial \boldsymbol{r'}}{\partial v}=0. \tag{13}$$

u and v can be understood as two parameters describing a certain patch of the surface, so $\partial\boldsymbol{r'}/\partial u$ and $\partial\boldsymbol{r'}/\partial v$ are tangent vectors at $\boldsymbol{r'}$. Therefore, Eq. (13) tells us that $\boldsymbol{r}-\boldsymbol{r'}$ is perpendicular to the tangent plane at $\boldsymbol{r'}$. This result can be interpreted in a more intuitive way. Considering a conductor surface, at each point $\boldsymbol{r'}$ on the surface, a light ray is only emitted in the normal direction, then the signal received at the position $\boldsymbol{r}$ is the sum of the light rays passing through the position $\boldsymbol{r}$. Therefore, we see the calculations in the JWST case can be accomplished within the framework of ray optics. The key difference with conventional ray optics is that a light ray induced by DPDM is always perpendicular to the local surface from which it is emitted, regardless of the direction in which the DPDM is incident.

We denote by $\hat{\boldsymbol{r}}_j$ the jth point on the conductor surface such that $\boldsymbol{r}-\hat{\boldsymbol{r}}_j$ is perpendicular to the tangent plane at $\hat{\boldsymbol{r}}_j$. Using the stationary-phase method, Eq. (4) becomes

$$\boldsymbol{E}(\boldsymbol{r})=\sum_j ie^{im_{A'}|\boldsymbol{r}-\hat{\boldsymbol{r}}_j|}\epsilon m_{A'}\hat{\boldsymbol{A}}'(\hat{\boldsymbol{r}}_j) \tag{14}$$

where $\hat{\boldsymbol{A}}'(\hat{\boldsymbol{r}}_j)=|\boldsymbol{A}'|\boldsymbol{\tau}(\hat{\boldsymbol{r}}_j)$ is the projection of $\boldsymbol{A}'$ onto the tangent plane at $\hat{\boldsymbol{r}}_j$. The expression for $\boldsymbol{B}$ is similar which is not present here for the purpose of conciseness. Putting everything together, we get

$$\langle\boldsymbol{S}(\boldsymbol{r})\rangle_t=\sum_j\frac{1}{2}\epsilon^2 m_{A'}^2\hat{\boldsymbol{A}}'(\hat{\boldsymbol{r}}_j)^2\hat{\boldsymbol{n}}(\hat{\boldsymbol{r}}_j)+\text{interference terms} \tag{15}$$

where $\hat{\boldsymbol{n}}(\hat{\boldsymbol{r}}_j)$ is the out-pointing normal direction of the conductor surface at $\hat{\boldsymbol{r}}_j$. The interference term comes from the cross products of contributions from $\hat{\boldsymbol{r}}_j$ and $\hat{\boldsymbol{r}}_k$, with $j\neq k$. If there is only one $\hat{\boldsymbol{r}}_j$, there is only one term in the summation and the interference term drops out. As a consistency check, one can compare (15) with the result of infinitely large metal plate given in appendix I-A of Ref. [57].

In the most extreme case, the conductor is a sphere and the detector is placed at the center of the sphere. Then, for the detector, the phase is stationary everywhere on the

sphere and no simplification can be made to Eqs. (4) and (5). Note that this is not the case with JWST, because the reflectors in JWST are not spherical, and the points $\hat{\boldsymbol{r}}_j$ are indeed discrete. In addition, in the frequency range of JWST, the correlation length is much smaller than the characteristic length of the reflectors, and the interference between different patches is further suppressed. This will be discussed in more detail in the following section.

B. Non-monochromatic DPDM

According to the Standard Halo Model, dark matter has a truncated Maxwellian distribution in the momentum space. Consequently, we should take into consideration the effect of finite coherence length of DPDM. The dark photon field at a location $\boldsymbol{x}$ can be expressed as,

$$\boldsymbol{E'}(\boldsymbol{x},t)=\int_{<k_{\rm esc}}\frac{\mathrm{d}^3\boldsymbol{k'}}{(2\pi)^3}be^{-\frac{\boldsymbol{k'}^2}{k_0^2}}\times\boldsymbol{E'_0}e^{i(\boldsymbol{k'}\cdot\boldsymbol{x}-\omega t+\theta(\boldsymbol{k'}))} \tag{16}$$

where $\theta(\boldsymbol{k'})$ is a random phase associated with the $\boldsymbol{k'}$ mode and $\omega=\sqrt{\boldsymbol{k'}^2+m_{A'}^2}$ is the energy. b is a normalization factor. We have $k_0=m_{A'}v_0$ and $k_{\rm esc}=m_{A'}v_{\rm esc}$, where $v_0\approx 235\,\mathrm{km/s}$ is the most probable velocity and $v_{\rm esc}$ is the escape velocity of leaving the Galaxy at the position of the solar system which is about 500 km/s [59, 60]. Due to randomness, we assume that there is no correlation between different momentum modes,

$$\langle e^{i(\theta(\boldsymbol{k}'_1)-\theta(\boldsymbol{k}'_2))}\rangle_t=a(2\pi)^3\delta^3(\boldsymbol{k}'_1-\boldsymbol{k}'_2) \tag{17}$$

where a is a dimensionful constant. Then, analogous to Eqs. (4) and (5), the full expressions for the induced electric and magnetic fields read

$$\boldsymbol{E}=-\int_{<k_{\rm esc}}\frac{\mathrm{d}^3\boldsymbol{k'}}{(2\pi)^3}\frac{\epsilon m_{A'}|\boldsymbol{E'_0}|}{2\pi}be^{-\frac{\boldsymbol{k'}^2}{k_0^2}}\int\mathrm{d}S'[\boldsymbol{\tau}(\boldsymbol{r'})\times(\boldsymbol{r}-\boldsymbol{r'})]\times(\boldsymbol{r}-\boldsymbol{r'})\frac{e^{i(\omega|\boldsymbol{r}-\boldsymbol{r'}|+\boldsymbol{k'}\cdot\boldsymbol{r'}+\theta(\boldsymbol{k'}))}}{|\boldsymbol{r}-\boldsymbol{r'}|^3}, \tag{18}$$

$$\boldsymbol{B}=-\int_{<k_{\rm esc}}\frac{\mathrm{d}^3\boldsymbol{k'}}{(2\pi)^3}\frac{\epsilon m_{A'}|\boldsymbol{E'_0}|}{2\pi}be^{-\frac{\boldsymbol{k'}^2}{k_0^2}}\int\mathrm{d}S'\boldsymbol{\tau}(\boldsymbol{r'})\times(\boldsymbol{r}-\boldsymbol{r'})\frac{e^{i(\omega|\boldsymbol{r}-\boldsymbol{r'}|+\boldsymbol{k'}\cdot\boldsymbol{r'}+\theta(\boldsymbol{k'}))}}{|\boldsymbol{r}-\boldsymbol{r'}|^2}. \tag{19}$$

One can further obtain the full expression for the energy flux density,

$$\begin{aligned}\langle\boldsymbol{S}\rangle_t=\frac{1}{2}\int_{<k_{\rm esc}}\frac{\mathrm{d}^3\boldsymbol{k'}}{(2\pi)^3}\left(\frac{\epsilon m_{A'}|\boldsymbol{E'_0}|}{2\pi}\right)^2ab^2e^{-\frac{2\boldsymbol{k'}^2}{k_0^2}}\int\mathrm{d}S'\mathrm{d}S''\,\{[\boldsymbol{\tau}(\boldsymbol{r'})\times(\boldsymbol{r}-\boldsymbol{r'})]\times(\boldsymbol{r}-\boldsymbol{r'})\}\\ \times[\boldsymbol{\tau}(\boldsymbol{r''})\times(\boldsymbol{r}-\boldsymbol{r''})]\mathrm{Re}\left(\frac{e^{im_{A'}(|\boldsymbol{r}-\boldsymbol{r'}|-|\boldsymbol{r}-\boldsymbol{r''}|)+i\boldsymbol{k'}\cdot(\boldsymbol{r'}-\boldsymbol{r''})}}{|\boldsymbol{r}-\boldsymbol{r'}|^3|\boldsymbol{r}-\boldsymbol{r''}|^2}\right).\end{aligned} \tag{20}$$

This expression can be simplified after some computation, the detail is given in appendix A.

$$\langle \boldsymbol{S} \rangle_t = \rho_{\rm DM} \left(\frac{\epsilon}{\lambda}\right)^2 \int {\rm d}S' {\rm d}S'' \left\{ [\boldsymbol{\tau}(\boldsymbol{r'}) \times (\boldsymbol{r} - \boldsymbol{r'})] \times (\boldsymbol{r} - \boldsymbol{r'}) \right\} \times [\boldsymbol{\tau}(\boldsymbol{r''}) \times (\boldsymbol{r} - \boldsymbol{r''})] \times e^{-\frac{1}{8}k_0^2 |\boldsymbol{r'} - \boldsymbol{r''}|^2} {\rm Re} \left(\frac{e^{i m_{A'} (|\boldsymbol{r} - \boldsymbol{r'}| - |\boldsymbol{r} - \boldsymbol{r''}|)}}{|\boldsymbol{r} - \boldsymbol{r'}|^3 |\boldsymbol{r} - \boldsymbol{r''}|^2} \right) \tag{21}$$

If we choose the reflector to be spherical, some analytic results can be derived. In the $k_0 \to 0$ limit, i.e. the infinite correlation length limit, the result is

$$\frac{\langle \boldsymbol{S} \rangle_t}{\rho_{\rm DM}} = \frac{1}{3} \pi^2 \epsilon^2 \frac{R^2}{\lambda^2} s_\gamma s_{\theta_0}^2 \sqrt{c_\gamma^2 (2 - 3c_{\theta_0} + c_{\theta_0}^3)^2 + 4 s_\gamma^2 (c_{\theta_0}^3 - 1)^2} \tag{22}$$

where γ is the angle between the polarization vector $\boldsymbol{n}_0$ and z-direction, θ_0 describes how large the spherical surface is, with $\theta_0 = 0$ for the surface shrinking to a point and $\theta_0 = \pi$ for the surface becoming a full sphere. This is the same as what we obtained in [57]. We are also interested in the $k_0 \to \infty$ limit, i.e. the zero correlation length limit, the result is

$$\frac{\langle \boldsymbol{S} \rangle_t}{\rho_{\rm DM}} = \frac{\epsilon^2}{2 v_0^2} s_{\theta_0}^2 \sqrt{4 s_\gamma^2 c_\gamma^2 s_{\theta_0}^4 + (2(1 + c_{\theta_0}^2) s_\gamma^2 + (1 + c_\gamma^2) s_{\theta_0}^2)^2} \tag{23}$$

Interestingly, the flux in the high frequency limit doesn't depend on the radius of the sphere, but it does depend on θ_0. Note that (23) only applies when the radius of the reflector is much larger than the dark photon wavelength, which is 10^3 times the same-frequency EM wavelength. Naively, larger reflectors produce stronger signal, but (23) tells us that the signal saturates when the reflector is much larger than the dark photon wavelength.

Coming back to the JWST case, we apply again the stationary phase approximation. In order that the integration is not suppressed, $\boldsymbol{r'}$ has to satisfy two conditions:

$$\boldsymbol{r} - \boldsymbol{r'} \perp \text{tangent plane at } \boldsymbol{r'}, \quad \boldsymbol{r'} = \boldsymbol{r''} \tag{24}$$

This means that due to finite correlation length, the contribution from interference terms completely vanishes. We denote by $\hat{\boldsymbol{r}}_i$ the ith solution to the perpendicular condition (24), and the total flux density is,

$$\langle \boldsymbol{S}(\boldsymbol{r}) \rangle_t = \sum_i \frac{1}{2} \epsilon^2 m_{A'}^2 \hat{\boldsymbol{A'}}^2(\hat{\boldsymbol{r}}_i) \hat{\boldsymbol{n}}(\hat{\boldsymbol{r}}_i) \tag{25}$$

Again, $\hat{\boldsymbol{n}}(\hat{\boldsymbol{r}}_i)$ is the out-directed normal vector at $\hat{\boldsymbol{r}}_i$, and $\hat{\boldsymbol{A'}}(\hat{\boldsymbol{r}}_i) = |\boldsymbol{A'}| \boldsymbol{\tau}(\hat{\boldsymbol{r}}_i)$. Note that if the set of solutions to (24) is not discrete, we should change the sum into an integration. If one wishes to average over all possible polarization, the result is

$$\langle \boldsymbol{S}(\boldsymbol{r}) \rangle_t = \sum_i \frac{2}{3} \epsilon^2 \rho_{\rm DM}(\hat{\boldsymbol{r}}_i) \hat{\boldsymbol{n}}(\hat{\boldsymbol{r}}_i) \tag{26}$$

Component	RoC(mm)	Surface	Conic	V_1(mm)	V_2(mm)	V_3(mm)	Size(mm)
Primary	15879.7	concave	-0.9967	0	0	0	6605.2(diameter)
Secondary	1778.9	convex	-1.6598	7169.0	0	0	738(diameter)
Tertiary	3016.2	concave	-0.6595	-796.3	0	-0.19	728(length)×517(width)
Fine Steering Mirror		flat		1047.8	0	-2.36	172.5(diameter)

Table I: Parameters for the optical elements of JWST. The data can be found in JWST documentation [61].

In the JWST case, all approximation conditions applied in this section are satisfied, so one may directly use (26) to compute the dark photon flux density.

III. THE OPTICAL TELESCOPE ELEMENT (OTE) OF JWST

The Optical Telescope Element (OTE) of the James Webb Space Telescope (JWST) comprises a primary mirror, a secondary mirror, a tertiary mirror, and a fine steering mirror. A sketch of the mirror system is shown in Figure 1. Detailed parameters for these optical components are available in JWST documentation [61], and we have also summarized them in Table I. In the table, 'RoC' denotes the radius of curvature, and 'conic' denotes the conic constant K which can be related to eccentricity $e_{\rm cone}$ as

$$K = -e_{\rm cone}^2. \tag{27}$$

Table I provides insights into the optical characteristics of the JWST's optical elements. The primary and tertiary mirrors exhibit elliptical shapes, while the secondary mirror is hyperbolic and the fine steering mirror is flat. The primary, secondary and fine steering mirrors are rounded, while the tertiary mirror is rectangular[62]. Their sizes are listed in Table I. V_1, V_2 and V_3 are the spacial displacements of the mirrors, as shown in Fig. 1.

The EM energy flux density induced by DPDM originates from several sources within the optical system. DPDM can interact with different components of the optical train, including the primary mirror. The light emitted from the primary mirror is then sequentially reflected by the secondary mirror, the tertiary mirror, and the fine steering mirror before reaching the detector. Additionally, DPDM can also directly interact with the secondary mirror, leading

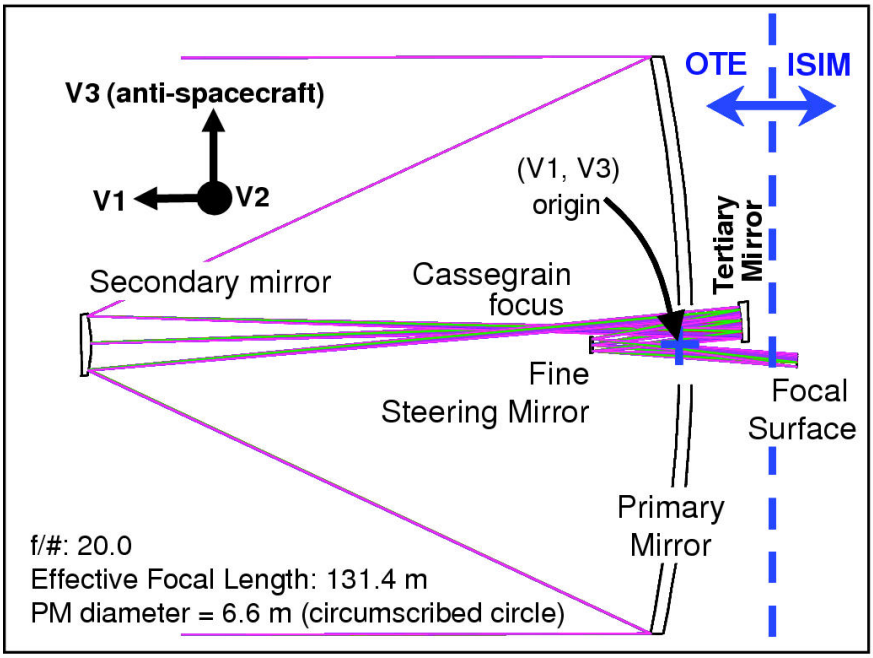


Figure 1: The OTE of JWST. Everything is drawn to scale. Figure taken from Ref. [56].

to a 'double' reflection before being detected. The same principle applies to interactions with the tertiary and fine steering mirrors. However, it is important to note that not all the lights induced by DPDM on the mirrors will reach the detector. Some of them may stray and cannot be focused on the detector after multiple reflections between the mirrors. After all, the JWST mirror system is designed to focus parallel lights from distant sources, while our light rays are always perpendicular to the mirror surface at which they are induced. To quantify the amount of the induced flux that can reach the detector, a detailed analysis of light propagation within the mirror system is necessary.

In the following section, we are going to carry out such an analysis based on the ray-transfer-matrix method. This method is applicable when the paraxial condition is satisfied, which means that the rays should be within a small angle to the optical axis throughout the system. We will show that the paraxial condition is indeed met in our case. It's important to note that while the mirrors' absolute sizes may not be significantly smaller than their radii of curvature, their effective sizes are small so that can satisfy the paraxial condition. Here, 'effective' refers to the portion of the mirror surface that make contributions to the flux finally detected.

IV. CALCULATING THE EQUIVALENT FLUX DENSITY

In this section, we are going to use the ray-transfer-matrix method to calculate the induced flux that can finally be detected by the JWST detector. A technical review of the ray-transfer-matrix method is shown in Appendix C. Firstly, we can alter the direction of light rays to make them move to the right for convenience, while at the same time we replace each reflector with a corresponding type of lens as depicted in Figure 2. The radii of curvature for the first three lenses are denoted by ρ_1, ρ_2, and ρ_3, respectively. In this section, we will use subscripts P, S, T and F for the shorthand of "primary mirror", "secondary mirror", "tertiary mirror" and "fine steering mirror" respectively.

In ray-transfer-matrix method, a ray is described with a 2-component vector X, the first component being the angle between the ray and the optical axis, the second being its vertical displacement from the optical axis. Each optical operation that a ray undergoes—such as free travel or refraction through a lens—is represented by a 2×2 matrix. Specifically, within this section, free travel over a distance L_i will be symbolized by the matrix U_i, while refraction on the primary mirror will be denoted by the matrix U_P, and likewise for other mirrors in the optical system.

We can write out the transition matrix of each lens and interval,

$$U_P = \begin{bmatrix} 1 & -\frac{2}{\rho_1} \\ 0 & 1 \end{bmatrix} \qquad U_S = \begin{bmatrix} 1 & \frac{2}{\rho_2} \\ 0 & 1 \end{bmatrix} \qquad U_T = \begin{bmatrix} 1 & -\frac{2}{\rho_3} \\ 0 & 1 \end{bmatrix}, \tag{28}$$

$$U_1 = \begin{bmatrix} 1 & 0 \\ L_1 & 1 \end{bmatrix} \qquad U_2 = \begin{bmatrix} 1 & 0 \\ L_2 & 1 \end{bmatrix} \qquad U_3 = \begin{bmatrix} 1 & 0 \\ L_3 & 1 \end{bmatrix}. \tag{29}$$

Light is emitted from each reflector, and the corresponding vectors are

$$X_P = \begin{bmatrix} -\frac{y}{\rho_1} \\ y \end{bmatrix}, \qquad X_S = \begin{bmatrix} \frac{y}{\rho_2} \\ y \end{bmatrix}, \qquad X_T = \begin{bmatrix} -\frac{y}{\rho_3} \\ y \end{bmatrix}, \qquad X_F = \begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{30}$$

Here y represents the height of the emission point relative to the optical axis.

We require that the light emitted from the primary mirror can reach the other three mirrors. For example, to check whether the light emitted from the primary mirror can reach the fine steering mirror, we use the following vector,

$$X = U_3 U_T U_2 U_S U_1 X_P. \tag{31}$$

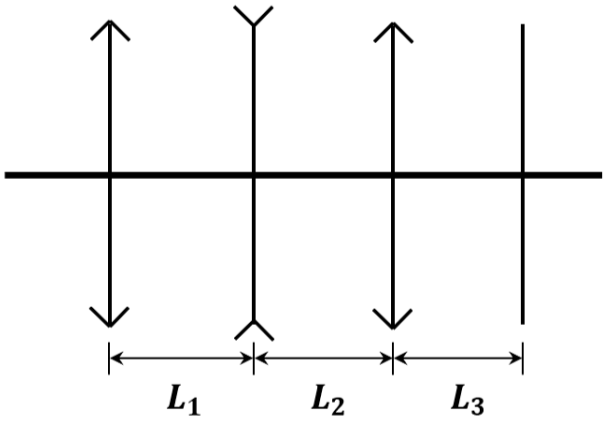


Figure 2: The OTE of JWST can be simplified to a set of lenses. The configuration comprises two convex lenses, positioned as the first and third lenses, with the second lens being concave and the fourth being flat.

If the light can be received by the fine steering mirror, then the absolute value of the second component of vector X has to be smaller than the radius of the fine steering mirror, which yields an inequality restricting the possible values of y. Therefore, by requiring that the light emitted by the primary mirror can be reflected three times and finally get into the detector, we have three inequalities which cut out an effective region on the surface of the primary mirror. Direct calculations show that this effective region is a rectangle with the length as 146.795 mm and the width as 104.248 mm. However, the primary mirror of JWST is hollowed in the center. The hollowed region is a hexagon with the side length as 762mm, which entirely encloses the effective region. Therefore, EM waves emitted by the primary mirror can't be received by the detector.

We can perform a similar analysis for the the other three mirrors. The effective region is also rectangular, with the length and width as

$$a_S = 132.904, \qquad b_S = 94.3834. \tag{32}$$

The effective region of the tertiary mirror is rounded, with the diameter as

$$d_T = 443.9 \tag{33}$$

In addition, the whole area of the fine steering mirror is effective.

We also need to analyze the focusing ability. After a direct calculation, we find that

$$X_S \Rightarrow \begin{bmatrix} -0.00307001y \\ -6.20097y \end{bmatrix}, \qquad X_T \Rightarrow \begin{bmatrix} -0.000331543y \\ -0.261223y \end{bmatrix}, \qquad X_F \Rightarrow \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{34}$$

where the unit of y is millimeter. Let us take X_s for example. After a series of reflections, the light emitted at the height y goes into the detector at height $6.20097y$. This means that the energy emitted by a disk with radius y is redistributed into a disk with radius $6.20097y$, and thus the energy flux density is $(6.20097)^{-2}$ the flux density emitted.

To sum up, the flux density inducedd by dark photon can be written as

$$I_{DM} = \frac{2}{3}\epsilon^2 \rho_{\rm DM} \times (1 + 6.20097^{-2} + 0.261223^{-2}) \tag{35}$$

In order to compare our calculated result of the DPDM-induced EM signal with the real data recorded by JWST, we need to translate the induced signal into the equivalent flux density of the incoming astronomical EM signals. This means that we need to further calculate the focusing ability for the case of the incoming planar EM waves. The incoming plane waves can be described by a vector

$$X_0 = \begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{36}$$

We again use the transition matrices to calculate the light at the receiver,

$$X_0 \Rightarrow \begin{bmatrix} 7.52946 \times 10^{-6} y \\ -0.0080392y \end{bmatrix} \tag{37}$$

Noting that the first entry is very small, we conclude that indeed the incoming planar wave is transformed into another planar wave. From the second entry, we get the enhancement factor $(0.0080392)^{-2}$. Thus, we can calculate the equivalent flux density by

$$I_{\rm eqv} \times (0.0080392)^{-2} = I_{DM} \tag{38}$$

which gives

$$I_{\rm eqv} = 6.75618 \times 10^{-4} \epsilon^2 \rho_{\rm DM}. \tag{39}$$

However, there is one more subtlety. We have calculated the effective range of the secondary mirror and the tertiary mirror, the light emitted from the secondary mirror can indeed

cover the whole receiver, but light emitted from the tertiary mirror is concentrated in a smaller region, and can only be observed with a well-chosen observation angle. To be very conservative, we simply assume that the light emitted by the tertiary mirror cannot go into the detector, and we have

$$I'_{\rm eqv} = 4.42064 \times 10^{-5} \epsilon^2 \rho_{\rm DM}. \tag{40}$$

V. CONSTRAINTS FROM JWST OBSERVATION DATA

The James Webb Space Telescope (JWST) stands as the cutting-edge space telescope, equipped with various detectors and versatile observation modes. In our study, we harnessed data from 972 distinct observation projects to establish constraints on the Dark Photon-Dark Matter coupling constant. Out of these projects, 713 relied on the Near-Infrared Spectrograph (NIRSpec) [cite], while 259 made use of the Mid-Infrared Instrument (MIRI) [cite]. It is worth noting that the data selected for analysis excludes background subtraction, ensuring its suitability for our research.

The JWST data we collected from the Mikulski Archive for Space Telescopes (MAST) database [63] includes two crucial parameters: the measured spectral flux density, denoted as $\bar{O}_i$, and the associated statistical uncertainty, denoted as $\sigma_{\bar{O}_i}$. In our effort to establish upper limits on the coupling of Dark Photon with the Standard Model electromagnetic current, denoted as ϵ, we followed the data analysis approach detailed in previous works [46, 57, 64].

To provide a concise overview, we summarize the key aspects of our method here, while reserving more detailed information for Appendix A. Our analysis begins by applying a local polynomial function to model the background surrounding a selected frequency bin, i_0, while considering neighboring bins. We estimate systematic uncertainties by comparing data deviations to the background fit. Next, we introduce a hypothetical Dark Photon Dark Matter signal with a strength denoted as S at the specific bin, i_0. This allows us to construct a likelihood function, L, that incorporates S into the comparison between data and the background function. Nuisance parameters are introduced to account for the coefficients of the background polynomial function.

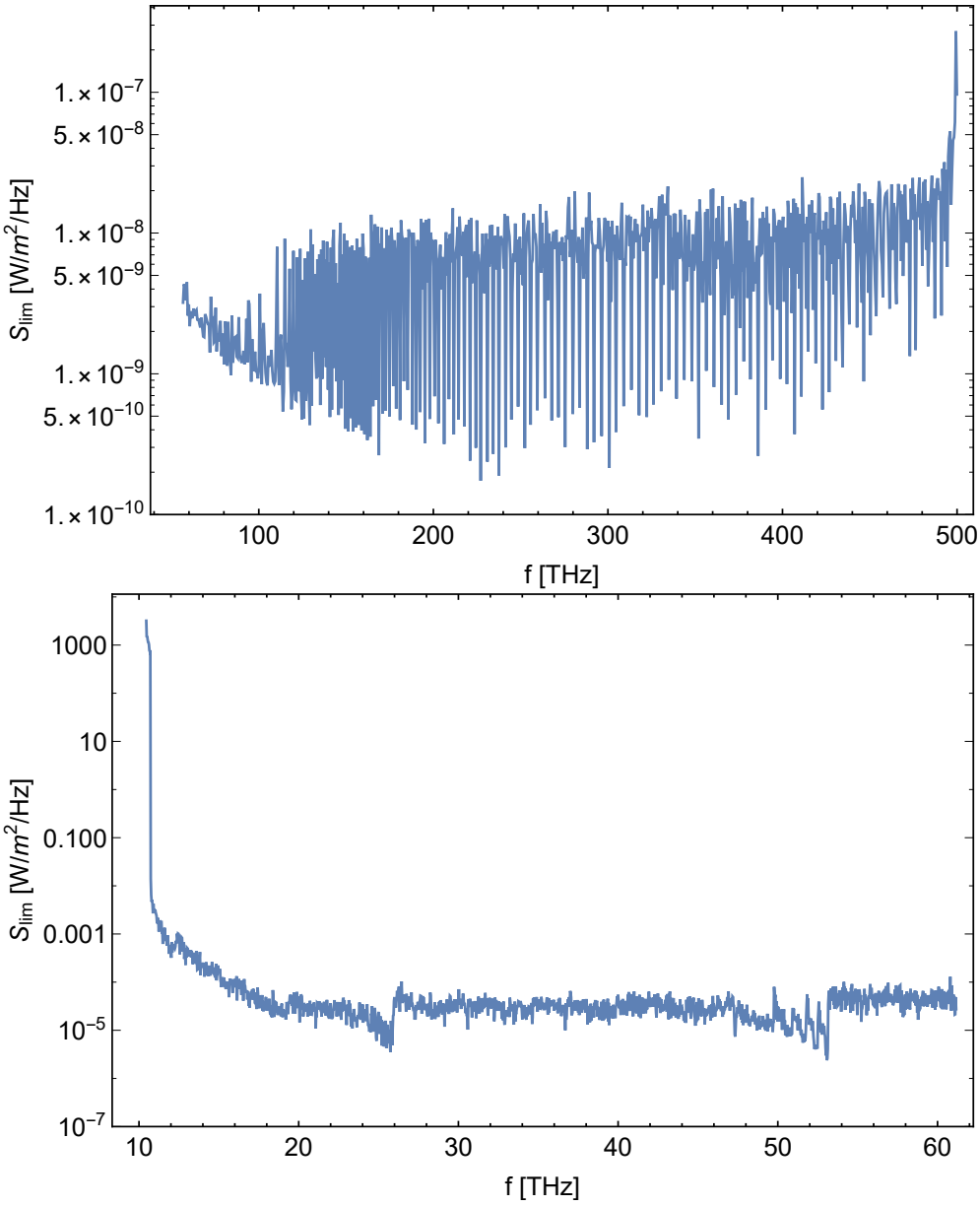


Figure 3: Model-independent 95% C.L. upper limits on a constant monochromatic signal from JWST data. The first figure corresponds to NIRSpec observation data, and the lower one MIRI. It shows the strongest limit from the all projects at each frequency bin.

Following the statistical method developed in Ref. [64], we compute the ratio, λ_S, between the maximized likelihood under two conditions: first, when only the nuisance parameters are varied to maximize L while keeping S constant, and second, when both the nuisance parameters and S are varied to maximize L. The test statistic, $-2\log(\lambda_S)$, follows a half-χ^2 distribution [64]. This analysis allows us to derive the 95% confidence level upper limit, $S_{\rm lim}$,

for a constant monochromatic signal. The results are illustrated in Fig. 3.

We establish upper limits on the mixing parameter ϵ as $S_{\rm lim} = S'_{\rm eqv}$, where $S'_{\rm eqv}$ represents the signal strength from theoretical calculations for Dark Photon Dark Matter. Different datasets from NIRSpec and MIRI yield varying constraints on the signal strength coupling ϵ, and we select the most stringent among them. The constraint on ϵ and its comparison with previous experiments are illustrated in Fig. 4.

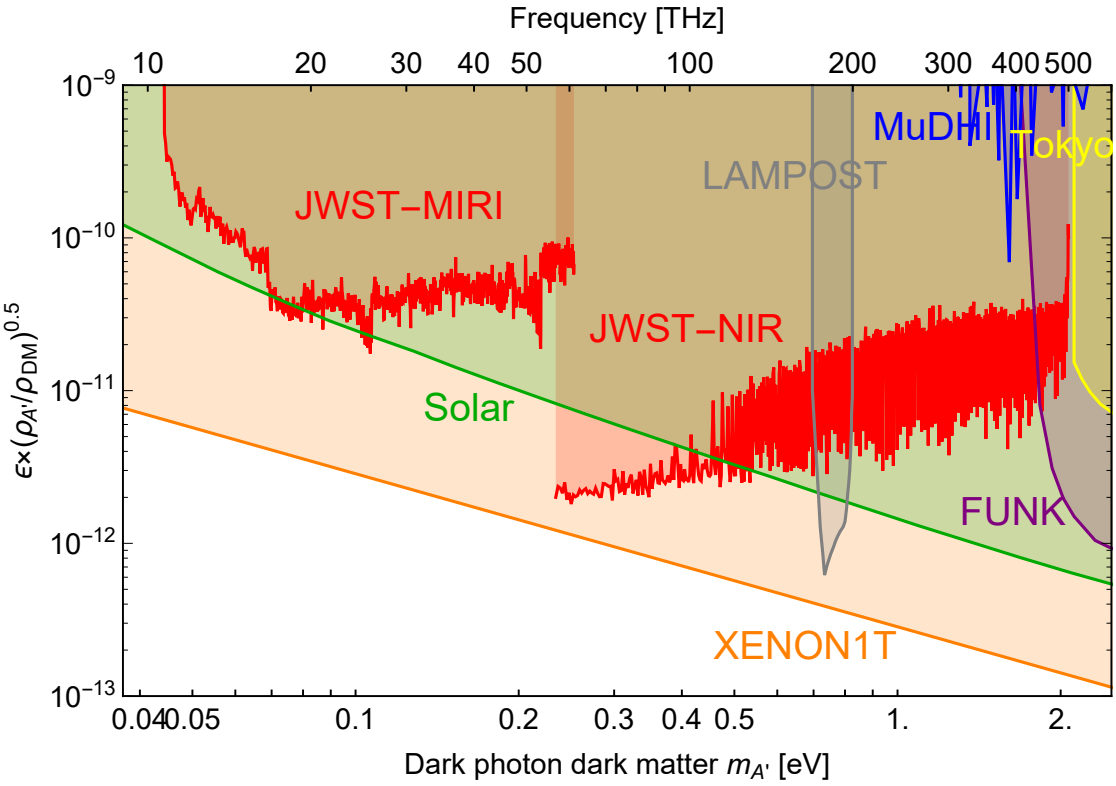


Figure 4: Constraints on the kinematic mixing parameter ϵ between Dark Photon Dark Matter (DPDM) and photons in the randomized polarization scheme. The solid red curve represents the 95% confidence level (C.L.) exclusion limit for DPDM using JWST data. The left and right sections display constraints derived from NIRSpec and MIRI observation data, respectively. Additionally, we provide a comparison with existing limits, including those from Solar [65], XENON1T [66], Lampost [67], Mudhi [68], Funk [54], and Tokyo [50].

VI. SUMMARY AND OUTLOOK

In this study, we conducted a direct detection search for dark photon dark matter (DPDM) using a haloscope setup. Our approach involved converting local DPDM into a normal electromagnetic field at the mirrors of the James Webb Space Telescope (JWST), which was then detected by its receiver. Due to the frequency match between the electromagnetic field and the DPDM mass, the resulting signal took the form of a monochromatic electromagnetic wave. Typically, accounting for the contribution of each surface area on the mirror to the electromagnetic field is necessary. The final electromagnetic energy received by the receiver encompasses the interference from each surface unit. We demonstrated that for high-mass DPDM, the contribution to the electromagnetic wave field at a given position primarily stems from the surface unit perpendicular to the separation vector, allowing us to calculate the electromagnetic flux at the receiver using ray-optics.

We utilized data from JWST observations to search for a monochromatic signal within the continuous background in the 10-500 terahertz (THz) range. Both the JWST Mid-Infrared Instrument (MIRI) and Near Infrared Spectrograph (NIRSpec) data were employed. Our analysis enabled us to establish limits on the DPDM kinetic mixing coupling at approximately $\epsilon \sim 3 \times 10^{-11}$ and $\epsilon \sim \mathcal{O}(10^{-12})$ respectively. This broadband search for DPDM yielded valuable lower-frequency constraints, complementing other experiments conducted in room-sized laboratories, such as Lampost, Mudhi, FUNK, and TOKYO. However, our results indicated a coupling weaker by about one order of magnitude compared to the XENON1T results, which utilized the potential dark photon flux generated by the Sun, without assuming that the dark photon is the dark matter. Our results are also comparable with the astrophysical bound from the solar emission of the dark photon particles.

While our results are slightly weaker than XENON1T constraints, there exists potential for improvement given that JWST is not specifically designed for the direct detection of DPDM. JWST boasts an outstanding receiver capable of detecting THz signals. By incorporating a spherical mirror as the reflector, as proposed in Refs. [39, 41, 48], or adopting a flat reflector along with a parabolic collection mirror, akin to the design in the TOKYO experiment [42, 51] and the BRASS-p experiment [55], it is possible to enhance sensitivity by orders of magnitude due to the right focus on the DPDM signal.

ACKNOWLEDGEMENT

JL would like to thank Le Hoang Nguyen for the helpful discussions. The work of HA is supported in part by the National Key R&D Program of China under Grant No. 2023YFA1607104 and 2021YFC2203100, the NSFC under Grant No. 11975134, and the Tsinghua University Initiative Scientific Research Program. The work of SG is supported by NSFC under Grant No. 12247147, the International Postdoctoral Exchange Fellowship Program, and the Boya Postdoctoral Fellowship of Peking University. The work of JL is supported by NSFC under Grant No. 12075005, 12235001.

Appendix A: Detailed derivation of Eq. (21)

Starting from equation (20), the integration over $\boldsymbol{k}'$ can be done analytically,

$$\int_{<k_{\rm esc}} \frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3} e^{-\frac{2\boldsymbol{k}'^2}{k_0^2}} e^{i\boldsymbol{k}'\cdot(\boldsymbol{r}'-\boldsymbol{r}'')} = \frac{k_0^2 e^{-2z^2} e^{-iyz}}{32\sqrt{2}\pi^2\Delta r}\left[2\sqrt{2}i(-1+e^{2iyz}) + \sqrt{\pi}ye^{-(y-4iz)^2/8}\left(\mathrm{erf}\left(\frac{4z+iy}{2\sqrt{2}}\right)+\mathrm{erf}\left(\frac{4z-iy}{2\sqrt{2}}\right)\right)\right] \tag{A1}$$

where $\Delta r = |\boldsymbol{r}'-\boldsymbol{r}''|$, $z = k_{\rm esc}/k_0$, and $y = k_0\Delta r$. When z or $k_{\rm esc}$ is large, the expression above can be expanded as

$$\int_{<k_{\rm esc}} \frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3} e^{-\frac{2\boldsymbol{k}'^2}{k_0^2}} e^{i\boldsymbol{k}'\cdot(\boldsymbol{r}'-\boldsymbol{r}'')} \approx \frac{k_0^3}{16\sqrt{2}\pi^{3/2}}\left(e^{-\frac{1}{8}y^2} - \frac{2\sqrt{2}}{\sqrt{\pi}y}e^{-2z^2}\sin(yz)\right). \tag{A2}$$

The next-to-leading order term is suppressed by e^{-2z^2}, and can be neglected. In addition, note that that

$$\rho_{\rm DM} = \frac{1}{2}\langle \boldsymbol{E}'(\boldsymbol{r},t)\boldsymbol{E}'^*(\boldsymbol{r},t)\rangle = \frac{ab^2}{64\pi^2}k_0^3|\boldsymbol{E}_0'|^2\left(\sqrt{2\pi}\mathrm{erf}(\sqrt{2}z) - 4ze^{-2z^2}\right) \tag{A3}$$

which, under the large z approximation, becomes

$$\rho_{\rm DM} = \frac{ab^2}{32\sqrt{2}\pi^{3/2}}k_0^3|\boldsymbol{E}_0'|^2. \tag{A4}$$

So the expression for $\langle \boldsymbol{S}\rangle_t$ reads

$$\langle \boldsymbol{S}\rangle_t = \rho_{\rm DM}\left(\frac{\epsilon}{\lambda}\right)^2 \int \mathrm{d}S'\mathrm{d}S''\left\{[\boldsymbol{\tau}(\boldsymbol{r}')\times(\boldsymbol{r}-\boldsymbol{r}')]\times(\boldsymbol{r}-\boldsymbol{r}')\right\}\times[\boldsymbol{\tau}(\boldsymbol{r}'')\times(\boldsymbol{r}-\boldsymbol{r}'')] \times e^{-\frac{1}{8}k_0^2|\boldsymbol{r}'-\boldsymbol{r}''|^2}\mathrm{Re}\left(\frac{e^{im_{A'}(|\boldsymbol{r}-\boldsymbol{r}'|-|\boldsymbol{r}-\boldsymbol{r}''|)}}{|\boldsymbol{r}-\boldsymbol{r}'|^3|\boldsymbol{r}-\boldsymbol{r}''|^2}\right) \tag{A5}$$

Appendix B: A Brief Proof of Eq. (8)

We demonstrate the one-dimensional case; extending to higher dimensions is straightforward.

We wish to evaluate

$$\lim_{\alpha\to\infty}\sqrt{\alpha}\int_{-\infty}^{\infty}g(x)e^{i\alpha f(x)}\mathrm{d}x \tag{B1}$$

in the large α limit. In this analysis, we assume that $f(x)$ is second order differentiable, $g(x)$ is continuous. We also assume that $g(x)$ is either compactly supported or exhibits exponential decay and that $f'(x)=0$ has only a discrete set of solutions. Here, we will focus on proving the case where $g(x)$ is compactly supported, noting that a similar approach can be applied when $g(x)$ has exponential decay.

Denote by Σ the set of points where $f'(x)=0$. Define

$$\Delta=\min\{x_i-x_j|x_i,x_j\in\Sigma\} \tag{B2}$$

Denote by I the set on which $g(x)$ is supported. Define a set A as follows,

$$A=\left\{[a,b]\,\middle|\,[a,b]=\left[x_0-\frac{\delta}{2},x_0+\frac{\delta}{2}\right]\cap I\ ,\ x_0\in\Sigma\right\} \tag{B3}$$

where δ is a positive real number that satisfies $\delta<\min\{\Delta,\alpha^{-1/2+\epsilon}\}$, with ϵ being a real number in the range $0<\epsilon<1/8$. Due to the continuity of $f'(x)$, it follows that $f(x)$ is monotonic between any two adjacent points in Σ. Consequently, we can divide the interval I into a finite set of closed intervals, each of which is either an element of A or an interval on which $f(x)$ is strictly monotonic.

Let's begin by examining the integral over intervals where $f(x)$ is monotonic.

$$\sqrt{\alpha}\int_{c}^{d}g(x)e^{i\alpha f(x)}\mathrm{d}x \tag{B4}$$

Where c and d are two real numbers. Since $f(x)$ is monotonic, it has an inverse, here denoted by $f^{-1}(x)$. We define $y=f^{-1}(x)$, and the integral can be expressed as

$$\sqrt{\alpha}\int_{c'}^{d'}\tilde{g}(y)e^{i\alpha y}\mathrm{d}y \tag{B5}$$

where $c'=f(c)$, $d'=f(d)$, and $\tilde{g}(y)=g(f^{-1}(y))(f^{-1})'(y)$. It can be shown that this expression goes to zero as $\alpha\to\infty$. By dividing the interval $[c',d']$ into a set of intervals with length $2\pi/\alpha$, integrating on each small interval contributes a result of order $\mathcal{O}(\alpha^{-3/2})$.

Summing over all intervals gives an additional factor of α, resulting in the overall order $\mathcal{O}(\alpha^{-1/2})$. Consequently, (B5) vanishes in the large α limit.

Next consider the integral on elements of A:

$$\sqrt{\alpha}\int_{x_0-a}^{x_0+b} g(x)e^{i\alpha f(x)}\mathrm{d}x \tag{B6}$$

where $x_0 \in \Sigma$ and $\delta \geq a,\ b \geq 0$. Notice that

$$\sqrt{\alpha}\left|\int_{x_0-a}^{x_0+b}(g(x)-g(x_0))e^{i\alpha f(x)}\mathrm{d}x\right| \leq \sqrt{\alpha}(a+b)\sup_{x_0-a\leq x\leq x_0+b}\{g(x)-g(x_0)\} = \mathcal{O}(\alpha^{-1/2+2\epsilon}) \tag{B7}$$

As α approaches infinity, the expression above tends to zero. Therefore, the integral we aim to evaluate can be replaced by the following expression,

$$\sqrt{\alpha}\int_{x_0-a}^{x_0+b} g(x_0)e^{i\alpha f(x)}\mathrm{d}x \tag{B8}$$

In the vicinity of x_0, $f(x)$ can be Taylor-expanded as

$$f(x) = f(x_0) + \frac{1}{2}f''(x_0)(x-x_0)^2 + R_2(x) \tag{B9}$$

where $R_2(x)$ represents the remainder term. Divide the integral into two parts

$$\begin{aligned}\sqrt{\alpha}\int_{x_0-a}^{x_0+b} g(x_0)e^{i\alpha f(x)}\mathrm{d}x = \sqrt{\alpha}\int_{x_0-a}^{x_0+b} g(x_0)\exp\left[i\alpha\left(f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2\right)\right]\mathrm{d}x \\ + \sqrt{\alpha}\int_{x_0-a}^{x_0+b} g(x_0)[\exp(i\alpha R_2(x))-1]\exp\left[i\alpha\left(f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2\right)\right]\mathrm{d}x\end{aligned} \tag{B10}$$

As $\alpha R_2(x) = \mathcal{O}(\alpha^{-1/2+3\epsilon})$, the second term is of order $\mathcal{O}(\alpha^{-1/2+4\epsilon})$, and therefore vanishes in the large α limit. Furthermore, in the large α limit, it can be demonstrated that the first term is equal to

$$\sqrt{\alpha}\int_{-\infty}^{\infty} g(x_0)\exp\left[i\alpha\left(f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2\right)\right]\mathrm{d}x \tag{B11}$$

We just have to prove that both

$$\sqrt{\alpha}\int_{-\infty}^{x_0-a} g(x_0)\exp\left[i\alpha\left(f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2\right)\right]\mathrm{d}x \tag{B12}$$

and

$$\sqrt{\alpha}\int_{x_0+b}^{\infty} g(x_0)\exp\left[i\alpha\left(f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2\right)\right]\mathrm{d}x \tag{B13}$$

tends to zero as $\alpha \to \infty$. Take (B13) for example. Let's define $t = \alpha(x - x_0)^2$, substitute x with t, and (B13) becomes

$$g(x_0)e^{i\alpha f(x_0)} \int_{\alpha b^2}^{\infty} \frac{1}{2\sqrt{t}} \exp\left[\frac{1}{2} i f''(x_0) t\right] \mathrm{d}t \tag{B14}$$

When α is sufficiently large, we have $\alpha b^2 = \alpha^{2\epsilon}$. Therefore, as α approaches infinity, (B13) tends to zero.

To sum up, we have

$$\int_{-\infty}^{\infty} g(x)e^{i\alpha f(x)}\mathrm{d}x = \sum_{x_0 \in \Sigma} \int_{-\infty}^{\infty} g(x_0) \exp\left[i\alpha\left(f(x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2\right)\right] \mathrm{d}x \tag{B15}$$

The integral on the right-hand side is nothing but a Gaussian integral, so the result is

$$\int_{-\infty}^{\infty} g(x)e^{i\alpha f(x)}\mathrm{d}x = \sum_{x_0 \in \Sigma} \sqrt{\frac{2\pi}{\alpha}} |f''(x_0)|^{-1/2} e^{\frac{i\pi}{4}\mathrm{sgn}(f''(x_0))} g(x_0) e^{i\alpha f(x_0)} \tag{B16}$$

It is straight forward to generalize to higher dimensional case. We can use the same method to prove that

$$\int_{\mathbb{R}^n} g(\boldsymbol{x})e^{i\alpha f(\boldsymbol{x})}\mathrm{d}^n\boldsymbol{x} = \sum_{\boldsymbol{x}_0 \in \Sigma} \int_{\mathbb{R}^n} g(\boldsymbol{x}_0) \exp\left[i\alpha\left(f(\boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \mathrm{Hess}(f(\boldsymbol{x}_0))(\boldsymbol{x} - \boldsymbol{x}_0)\right)\right] \mathrm{d}^n\boldsymbol{x} \tag{B17}$$

Here Σ is defined to be the set of points where $\nabla f = 0$, and the Hessian matrix is defined to be

$$\mathrm{Hess}(f(\boldsymbol{x})) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{B18}$$

To evaluate the multidimensional Gaussian integral, we diagonalize the Hessian matrix. This transforms the integral into the product of n one dimensional Gaussian integrals, ultimately leading to equation (8).

Appendix C: Ray Transfer Matrix Analysis

In systems satisfying paraxial condition, we can utilize "ray transfer matrix analysis" to simplify the calculations. A beam of light can be characterized by two parameters:

the angle(counterclockwise) between the light and the optical axis, and the vertical distance(upward) between the light and the optical axis. These two parameters can be organized into a column vector

$$\boldsymbol{v} = \begin{bmatrix} \theta \\ y \end{bmatrix} \tag{C1}$$

Here we assume that the light is travelling from left to right.

We can perform various operations on the light. First, it can travel a distance L through free space, as depicted in Figure 5. During free travel, the angle θ remains unchanged, while the height y increases by θL. This process can be described by the left multiplication of a matrix

$$\boldsymbol{v} \Rightarrow \boldsymbol{v}' = \begin{bmatrix} 1 & 0 \\ L & 1 \end{bmatrix} \boldsymbol{v} \tag{C2}$$

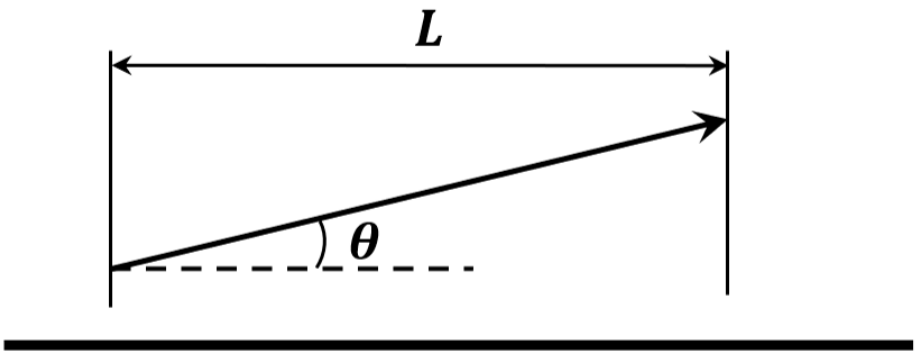


Figure 5: A beam of light travelling in free space

We can also represent the effects of reflectors using matrices. Reflecting changes the direction of the light from right-going to left-going, which can introduce complications. To simplify matters, we reflect the direction of the light, ensuring that it always travels rightward, as shown in figure 6.

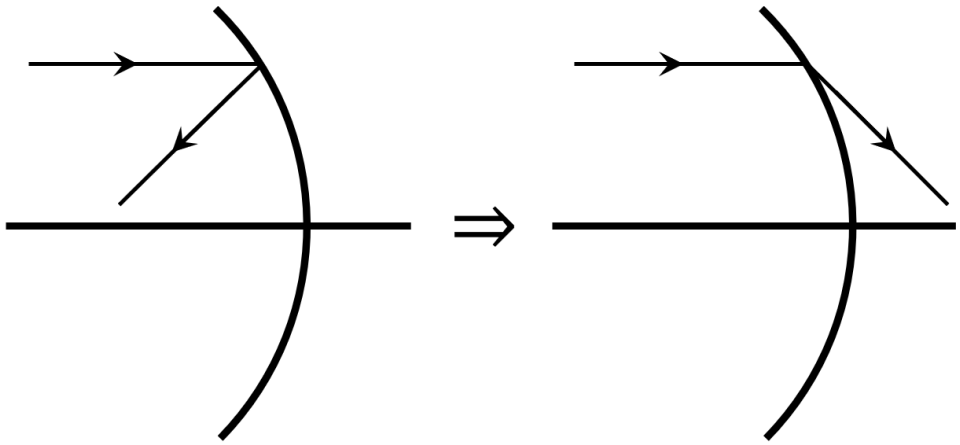


Figure 6: Reflecting the light so that it is always right-going

We are now prepared to analyze the effect of a reflector. When a beam of light is reflected by the reflector, its height remains unchanged while the direction θ is altered. Given that the paraxial condition is met, the reflector can be approximated by a spherical mirror, with its radius equal to the radius of curvature. Through direct analysis, we find that the effect of a reflector can be described by the following matrices

$$\begin{bmatrix} 1 & -\frac{2}{\rho} \\ 0 & 1 \end{bmatrix} \text{(concave)} \qquad \begin{bmatrix} 1 & \frac{2}{\rho} \\ 0 & 1 \end{bmatrix} \text{(convex)} \tag{C3}$$

In the JWST setup, the axis of symmetry of the reflectors may differ from the optical axis, as shown in Figure 7 and Figure 8. Fortunately, the system is linear, so both of these effects simply add an overall constant to the beams of light we are considering.

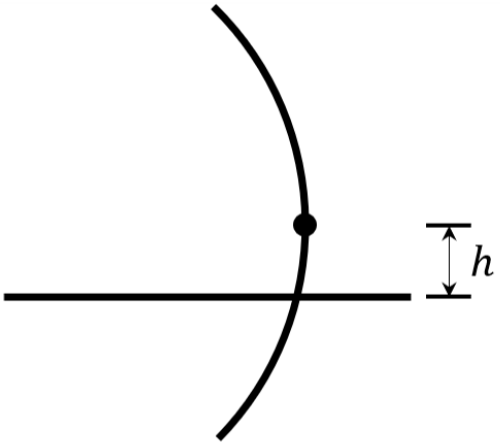


Figure 7: The center of the reflector is higher than the optical axis.

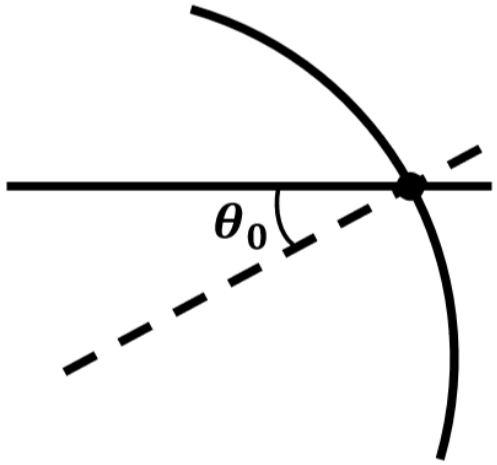


Figure 8: The axis of symmetry is not parallel to the optical axis.

First, if the center of the mirror is higher than the optical axis, the effect of the mirror is

$$\begin{bmatrix} \theta \\ y \end{bmatrix} \Rightarrow \begin{bmatrix} \theta - \frac{2y}{\rho} + \frac{2h}{\rho} \\ y \end{bmatrix} \tag{C4}$$

So there is an overall angle $2h/\rho$, which can only cause overall vertical or angular displacement.

In the second case, where the axis of symmetry is not parallel to the optical axis, we have

$$\begin{bmatrix} \theta \\ y \end{bmatrix} \Rightarrow \begin{bmatrix} \theta - \frac{2y}{\rho} - 2\theta_0 \\ y \end{bmatrix} \tag{C5}$$

Similarly, the $2\theta_0$ term can only cause overall vertical or angular displacements. As a result, the displacement of mirrors has no effect on the signal strength, allowing us to disregard the displacements and permitting light to pass through some mirrors if necessary.

Appendix D: Data Analysis Method

The data analysis method adopted in the present work follows that in [46, 57, 64]. Utilizing JWST observation [63], we have a dataset of spectral flux density $\bar{O}_i$ (mean value) along with the associated statistical error $\sigma_{\bar{O}_i}$ at a series of frequency bins indexed by i. To model the local flux background around bin i_0, We apply a polynomial function $B(a, f)$, fitting the data from bin $i_0 - k$ to bin $i_0 + k$,

$$B(a, f) = a_0 + a_1 f + a_2 f^2 + ... + a_n f^n. \tag{D1}$$

$a = \{a_0, a_1, a_2, ..., a_n\}$ are the coefficients of the polynomial terms. The weighted sum of squared residuals,

$$\sum_{i=i_0-k}^{i_0+k} \frac{1}{\sigma_{\bar{O}_i}^2}[B(a, f_i) - \bar{O}_i]^2, \tag{D2}$$

is minimized at $a = \tilde{a}$. The deviations of the data points from the background fitting result, $\delta_i \equiv B(\tilde{a}, f_i) - \bar{O}_i$, can be modeled as a systematic error at bin i_0, that is

$$\sigma_{i_0}^{\rm sys} = \sqrt{\frac{1}{2k-1}\sum_{i=i_0-k}^{i_0+k}(\delta_i - \bar{\delta})^2}. \tag{D3}$$

$\bar{\delta}$ is the average of the list δ_i. Note that in computing Eqs. (D2)-(D3), we do not include the bin i_0 in the calculations. Additionally, for practical purposes, we set $n = 3$ and $k = 5$. By adding these two kinds of uncertainties in quadrature, we get the total uncertainty at bin i_0,

$$\sigma_{i_0}^{\rm sys} = \sqrt{(\sigma_{i_0}^{\rm sys})^2 + \sigma_{\bar{O}_{i0}}^2}. \tag{D4}$$

Next, to set upper limits on the coupling of DPDM with photon, we employ a likelihood-based statistical method [64]. A likelihood function is constructed around bin i_0 as follows,

$$L(S, a) = \prod_{i=i_0-k}^{i_0+k} \frac{1}{\sqrt{2\pi}\sigma_i^{\rm tot}} \exp\left[-\frac{1}{2}\left(\frac{B(a, f_i) + S\delta_{ii_0} - \bar{O}_i}{\sigma_i^{\rm tot}}\right)^2\right]. \tag{D5}$$

Here, we consider the parameter a's as nuisance parameters. S represents the DPDM-induced signal, and we assume its location to be in bin i_0. It's worth noting that the frequency dispersion of DPDM is $\mathcal{B}_{\rm DPDM} \sim 0.15\ {\rm kHz} \times (m_{A'}/\mu{\rm eV})$. This is much smaller than the instrumental spectral resolution which ranges from 10 GHz to 40 THz, depending on different observation modes [56], so the DPDM-induced signal can be safely confined within a single frequency bin.

Then, we build the test statistic as

$$q_S = \begin{cases} -2\ln\frac{L(S,\hat{\hat{a}})}{L(\hat{S},\hat{a})}, & \hat{S} \le S \\ 0, & \hat{S} > S \end{cases}. \tag{D6}$$

L is maximized at $a = \hat{a}$ and $S = \hat{S}$; it is conditionally maximized at $a = \hat{\hat{a}}$ for a fixed S. As has been demonstrated in Ref. [64], the test statistic q_S satisfies the half-chi-squared distribution,

$$f(q_S|S) = \frac{1}{2}\delta(q_S) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_S}}\exp(-q_S/2), \tag{D7}$$

the cumulative distribution of which is labeled as $\Phi(\sqrt{q_S})$. Then, we define the p-value function as $p_S = [1 - \Phi(\sqrt{q_S})]/[1 - \Phi(\sqrt{q_0})]$ which measures the deviation of the assumed signal S to the null $S = 0$. We set $p_S = 5\%$ and then determine the value of S corresponding to this p_S, which we denote as $S_{\rm lim}$. Consequently, if an assumed signal has a strength $S > S_{\rm lim}$, we can exclude it at the 95% confidence level.

---

[1] Bob Holdom, "Two U(1)'s and Epsilon Charge Shifts," Phys. Lett. **166B**, 196–198 (1986).

[2] Keith R. Dienes, Christopher F. Kolda, and John March-Russell, "Kinetic mixing and the supersymmetric gauge hierarchy," Nucl. Phys. B **492**, 104–118 (1997), arXiv:hep-ph/9610479.

[3] S. A. Abel and B. W. Schofield, "Brane anti-brane kinetic mixing, millicharged particles and SUSY breaking," Nucl. Phys. B **685**, 150–170 (2004), arXiv:hep-th/0311051.

[4] S. A. Abel, M. D. Goodsell, J. Jaeckel, V. V. Khoze, and A. Ringwald, "Kinetic Mixing of the Photon with Hidden U(1)s in String Phenomenology," JHEP **07**, 124 (2008), arXiv:0803.1449 [hep-ph].

[5] Steven A. Abel, Joerg Jaeckel, Valentin V. Khoze, and Andreas Ringwald, "Illuminating the Hidden Sector of String Theory by Shining Light through a Magnetic Field," Phys. Lett. B **666**, 66–70 (2008), arXiv:hep-ph/0608248.

[6] Mark Goodsell, Joerg Jaeckel, Javier Redondo, and Andreas Ringwald, "Naturally Light Hidden Photons in LARGE Volume String Compactifications," JHEP **11**, 027 (2009), arXiv:0909.0515 [hep-ph].

[7] Javier Redondo and Marieke Postma, "Massive hidden photons as lukewarm dark matter," JCAP **02**, 005 (2009), arXiv:0811.0326 [hep-ph].

[8] Ann E. Nelson and Jakub Scholtz, "Dark Light, Dark Matter and the Misalignment Mechanism," Phys. Rev. **D84**, 103501 (2011), arXiv:1105.2812 [hep-ph].

[9] Paola Arias, Davide Cadamuro, Mark Goodsell, Joerg Jaeckel, Javier Redondo, and Andreas Ringwald, "WISPy Cold Dark Matter," JCAP **1206**, 013 (2012), arXiv:1201.5902 [hep-ph].

[10] Peter W. Graham, Jeremy Mardon, and Surjeet Rajendran, "Vector Dark Matter from Inflationary Fluctuations," Phys. Rev. **D93**, 103520 (2016), arXiv:1504.02102 [hep-ph].

[11] Gonzalo Alonso-Álvarez, Thomas Hugle, and Joerg Jaeckel, "Misalignment & Co.: (Pseudo-)scalar and vector dark matter with curvature couplings," (2019), arXiv:1905.09836 [hep-ph].

[12] Kazunori Nakayama, "Vector Coherent Oscillation Dark Matter," JCAP **1910**, 019 (2019), arXiv:1907.06243 [hep-ph].

[13] Kazunori Nakayama, "Constraint on Vector Coherent Oscillation Dark Matter with Kinetic Function," JCAP **08**, 033 (2020), arXiv:2004.10036 [hep-ph].

[14] Yohei Ema, Kazunori Nakayama, and Yong Tang, "Production of Purely Gravitational Dark Matter: The Case of Fermion and Vector Boson," JHEP **07**, 060 (2019), arXiv:1903.10973 [hep-ph].

[15] Edward W. Kolb and Andrew J. Long, "Completely dark photons from gravitational particle production during the inflationary era," JHEP **03**, 283 (2021), arXiv:2009.03828 [astro-ph.CO].

[16] Borna Salehian, Mohammad Ali Gorji, Hassan Firouzjahi, and Shinji Mukohyama, "Vector dark matter production from inflation with symmetry breaking," Phys. Rev. D **103**, 063526 (2021), arXiv:2010.04491 [hep-ph].

[17] Aqeel Ahmed, Bohdan Grzadkowski, and Anna Socha, "Gravitational production of vector dark matter," JHEP **08**, 059 (2020), arXiv:2005.01766 [hep-ph].

[18] Yuichiro Nakai, Ryo Namba, and Ziwei Wang, "Light Dark Photon Dark Matter from Inflation," JHEP **12**, 170 (2020), arXiv:2004.10743 [hep-ph].

[19] Kazunori Nakayama and Yong Tang, "Gravitational Production of Hidden Photon Dark Matter in Light of the XENON1T Excess," Phys. Lett. B **811**, 135977 (2020), arXiv:2006.13159 [hep-ph].

[20] Hassan Firouzjahi, Mohammad Ali Gorji, Shinji Mukohyama, and Borna Salehian, "Dark photon dark matter from charged inflaton," JHEP **06**, 050 (2021), arXiv:2011.06324 [hep-ph].

[21] Mar Bastero-Gil, Jose Santiago, Lorenzo Ubaldi, and Roberto Vega-Morales, "Dark photon dark matter from a rolling inflaton," JCAP **02**, 015 (2022), arXiv:2103.12145 [hep-ph].

[22] Hassan Firouzjahi, Mohammad Ali Gorji, Shinji Mukohyama, and Alireza Talebian, "Dark matter from entropy perturbations in curved field space," Phys. Rev. D **105**, 043501 (2022), arXiv:2110.09538 [gr-qc].

[23] Takanori Sato, Fuminobu Takahashi, and Masaki Yamada, "Gravitational production of dark photon dark matter with mass generated by the Higgs mechanism," (2022), arXiv:2204.11896 [hep-ph].

[24] Raymond T. Co, Aaron Pierce, Zhengkang Zhang, and Yue Zhao, "Dark Photon Dark Matter

Produced by Axion Oscillations," (2018), arXiv:1810.07196 [hep-ph].

[25] Jeff A. Dror, Keisuke Harigaya, and Vijay Narayan, "Parametric Resonance Production of Ultralight Vector Dark Matter," (2018), arXiv:1810.07195 [hep-ph].

[26] Mar Bastero-Gil, Jose Santiago, Lorenzo Ubaldi, and Roberto Vega-Morales, "Vector dark matter production at the end of inflation," (2018), arXiv:1810.07208 [hep-ph].

[27] Prateek Agrawal, Naoya Kitajima, Matthew Reece, Toyokazu Sekiguchi, and Fuminobu Takahashi, "Relic Abundance of Dark Photon Dark Matter," (2018), arXiv:1810.07188 [hep-ph].

[28] Raymond T. Co, Keisuke Harigaya, and Aaron Pierce, "Gravitational waves and dark photon dark matter from axion rotations," JHEP **12**, 099 (2021), arXiv:2104.02077 [hep-ph].

[29] Kazunori Nakayama and Wen Yin, "Hidden photon and axion dark matter from symmetry breaking," JHEP **10**, 026 (2021), arXiv:2105.14549 [hep-ph].

[30] Andrew J. Long and Lian-Tao Wang, "Dark Photon Dark Matter from a Network of Cosmic Strings," (2019), arXiv:1901.03312 [hep-ph].

[31] Marco Fabbrichesi, Emidio Gabrielli, and Gaia Lanfranchi, "The Dark Photon," (2020), 10.1007/978-3-030-62519-1, arXiv:2005.01515 [hep-ph].

[32] Andrea Caputo, Alexander J. Millar, Ciaran A. J. O'Hare, and Edoardo Vitagliano, "Dark photon limits: A handbook," Phys. Rev. D **104**, 095029 (2021), arXiv:2105.04565 [hep-ph].

[33] S. De Panfilis, A. C. Melissinos, B. E. Moskowitz, J. T. Rogers, Y. K. Semertzidis, Walter Wuensch, H. J. Halama, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, "Limits on the Abundance and Coupling of Cosmic Axions at 4.5-Microev ¡ m(a) ¡ 5.0-Microev," Phys. Rev. Lett. **59**, 839 (1987).

[34] Walter Wuensch, S. De Panfilis-Wuensch, Y. K. Semertzidis, J. T. Rogers, A. C. Melissinos, H. J. Halama, B. E. Moskowitz, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, "Results of a Laboratory Search for Cosmic Axions and Other Weakly Coupled Light Particles," Phys. Rev. **D40**, 3153 (1989).

[35] C. Hagmann, P. Sikivie, N. S. Sullivan, and D. B. Tanner, "Results from a search for cosmic axions," Phys. Rev. **D42**, 1297–1300 (1990).

[36] Stephen J. Asztalos *et al.* (ADMX), "Large scale microwave cavity search for dark matter axions," Phys. Rev. **D64**, 092003 (2001).

[37] S. J. Asztalos *et al.* (ADMX), "A SQUID-based microwave cavity search for dark-matter axions," Phys. Rev. Lett. **104**, 041301 (2010), arXiv:0910.5914 [astro-ph.CO].

[38] Le Hoang Nguyen, Andrei Lobanov, and Dieter Horns, "First results from the WISPDMX radio frequency cavity searches for hidden photon dark matter," JCAP **1910**, 014 (2019), arXiv:1907.12449 [hep-ex].

[39] Dieter Horns, Joerg Jaeckel, Axel Lindner, Andrei Lobanov, Javier Redondo, and Andreas Ringwald, "Searching for WISPy Cold Dark Matter with a Dish Antenna," JCAP **04**, 016 (2013), arXiv:1212.2970 [hep-ph].

[40] Joerg Jaeckel and Javier Redondo, "An antenna for directional detection of WISPy dark matter," JCAP **11**, 016 (2013), arXiv:1307.7181 [hep-ph].

[41] Joerg Jaeckel and Stefan Knirck, "Directional Resolution of Dish Antenna Experiments to Search for WISPy Dark Matter," JCAP **01**, 005 (2016), arXiv:1509.00371 [hep-ph].

[42] Stefan Knirck, Takayuki Yamazaki, Yoshiki Okesaku, Shoji Asai, Toshitaka Idehara, and Toshiaki Inada, "First results from a hidden photon dark matter search in the meV sector using a plane-parabolic mirror system," JCAP **1811**, 031 (2018), arXiv:1806.05120 [hep-ex].

[43] Graciela B. Gelmini, Alexander J. Millar, Volodymyr Takhistov, and Edoardo Vitagliano, "Probing dark photons with plasma haloscopes," Phys. Rev. D **102**, 043003 (2020), arXiv:2006.06836 [hep-ph].

[44] Samuel D. McDermott and Samuel J. Witte, "Cosmological evolution of light dark photon dark matter," Phys. Rev. **D101**, 063030 (2020), arXiv:1911.05086 [hep-ph].

[45] Haipeng An, Fa Peng Huang, Jia Liu, and Wei Xue, "Radio-frequency Dark Photon Dark Matter across the Sun," Phys. Rev. Lett. **126**, 181102 (2021), arXiv:2010.15836 [hep-ph].

[46] Haipeng An, Xingyao Chen, Shuailiang Ge, Jia Liu, and Yan Luo, "Searching for Ultralight Dark Matter Conversion in Solar Corona using LOFAR Data," (2023), arXiv:2301.03622 [hep-ph].

[47] Haipeng An, Shuailiang Ge, and Jia Liu, "Solar Radio Emissions and Ultralight Dark Matter," Universe **9**, 142 (2023), arXiv:2304.01056 [hep-ph].

[48] Joerg Jaeckel and Javier Redondo, "Resonant to broadband searches for cold dark matter consisting of weakly interacting slim particles," Phys. Rev. D **88**, 115002 (2013), arXiv:1308.1103 [hep-ph].

[49] Jun'ya Suzuki, Yoshizumi Inoue, Tomoki Horie, and Makoto Minowa, "Hidden photon CDM search at Tokyo," in *11th Patras Workshop on Axions, WIMPs and WISPs* (2015) pp. 145–148, arXiv:1509.00785 [hep-ex].

[50] J. Suzuki, T. Horie, Y. Inoue, and M. Minowa, "Experimental Search for Hidden Photon CDM in the eV mass range with a Dish Antenna," JCAP **09**, 042 (2015), arXiv:1504.00118 [hep-ex].

[51] Nozomu Tomita, Shugo Oguri, Yoshizumi Inoue, Makoto Minowa, Taketo Nagasaki, Jun'ya Suzuki, and Osamu Tajima, "Search for hidden-photon cold dark matter using a K-band cryogenic receiver," JCAP **09**, 012 (2020), arXiv:2006.02828 [hep-ex].

[52] Benjamin Godfrey *et al.*, "Search for dark photon dark matter: Dark E field radio pilot experiment," Phys. Rev. D **104**, 012013 (2021), arXiv:2101.02805 [physics.ins-det].

[53] Pierre Brun, Laurent Chevalier, and Christophe Flouzat, "Direct Searches for Hidden-Photon Dark Matter with the SHUKET Experiment," Phys. Rev. Lett. **122**, 201801 (2019), arXiv:1905.05579 [hep-ex].

[54] A. Andrianavalomahefa *et al.* (FUNK Experiment), "Limits from the Funk Experiment on the Mixing Strength of Hidden-Photon Dark Matter in the Visible and Near-Ultraviolet Wavelength Range," Phys. Rev. D **102**, 042001 (2020), arXiv:2003.13144 [astro-ph.CO].

[55] Fayez Bajjali *et al.*, "First results from BRASS-p broadband searches for hidden photon dark matter," JCAP **08**, 077 (2023), arXiv:2306.05934 [hep-ex].

[56] Jonathan P. Gardner *et al.*, "The James Webb Space Telescope," Space Sci. Rev. **123**, 485 (2006), arXiv:astro-ph/0606175.

[57] Haipeng An, Shuailiang Ge, Wen-Qing Guo, Xiaoyuan Huang, Jia Liu, and Zhiyao Lu, "Direct detection of dark photon dark matter using radio telescopes," Phys. Rev. Lett. **130**, 181001 (2023).

[58] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals*, Dover Books on Mathematics Series (Dover Publications, 1986).

[59] A. K. Drukier, Katherine Freese, and D. N. Spergel, "Detecting Cold Dark Matter Candidates," Phys. Rev. D **33**, 3495–3508 (1986).

[60] N. Wyn Evans, Ciaran A. J. O'Hare, and Christopher McCabe, "Refinement of the standard halo model for dark matter searches in light of the Gaia Sausage," Phys. Rev. D **99**, 023012 (2019), arXiv:1810.11468 [astro-ph.GA].

[61] Paul Lightsey, Charles Atkinson, Mark Clampin, and Lee Feinberg, "James webb space telescope: Large deployable cryogenic telescope in space," Optical Engineering **51**, 1003– (2012).

[62] Paul A. Lightsey, J. Scott Knight, and Gary Golnik, "Status of the optical performance for the James Webb Space Telescope," in *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, edited by Jr. Oschmann, Jacobus M., Mark Clampin, Giovanni G. Fazio, and Howard A. MacEwen (2014) p. 914304.

[63] MAST 2021, JWST Archive Manual, eds. R.A. Shaw, B. Cherinka, P. Forshay, J. Yoon (Baltimore: STScI).

[64] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, "Asymptotic formulae for likelihood-based tests of new physics," The European Physical Journal C **71** (2011), 10.1140/epjc/s10052-011-1554-0.

[65] Shao-Ping Li and Xun-Jie Xu, "Production rates of dark photons and z' in the sun and stellar cooling bounds," (2023), arXiv:2304.12907 [hep-ph].

[66] E. Aprile *et al.* (XENON), "Emission of single and few electrons in XENON1T and limits on light dark matter," Phys. Rev. D **106**, 022001 (2022), arXiv:2112.12116 [hep-ex].

[67] Jeff Chiles *et al.*, "New Constraints on Dark Photon Dark Matter with Superconducting Nanowire Detectors in an Optical Haloscope," Phys. Rev. Lett. **128**, 231802 (2022), arXiv:2110.01582 [hep-ex].

[68] Laura Manenti *et al.*, "Search for dark photons using a multilayer dielectric haloscope equipped with a single-photon avalanche diode," Phys. Rev. D **105**, 052010 (2022), arXiv:2110.10497 [hep-ex].

Estimation of the electromagnetic field in intermediate-energy heavy-ion collisions

Hidetoshi Taya,[1,*] Toru Nishimura,[2,3] and Akira Ohnishi[3]

[1] *RIKEN iTHEMS, RIKEN, Wako 351-0198, Japan*
[2] *Department of Physics, Osaka University, Toyonaka, Osaka, 560-0043, Japan*
[3] *Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto, 606-8317, Japan*

We estimate the spacetime profile of the electromagnetic field in head-on heavy-ion collisions at intermediate collision energies $\sqrt{s_{\rm NN}} = \mathcal{O}(3\text{ - }10\text{ GeV})$. Using a hadronic cascade model (JAM; Jet AA Microscopic transport model), we numerically demonstrate that the produced field has strength $eE = \mathcal{O}((30\text{ - }60\text{ MeV})^2)$, which is supercritical to the Schwinger limit of QED and is non-negligibly large compared even to the hadron/QCD scale, and survives for a long time $\tau = \mathcal{O}(10\text{ fm}/c)$ due to the baryon stopping. We show that the produced field is nonperturbatively strong in the sense that the nonperturbativity parameters (e.g., the Keldysh parameter) are sufficiently large, which is in contrast to high-energy collisions $\sqrt{s_{\rm NN}} \gtrsim 100$ GeV, where the field is merely perturbative. Our results imply that the electromagnetic field may have phenomenological impacts on hadronic/QCD processes in intermediate-energy heavy-ion collisions and that heavy-ion collisions can be used as a new tool to explore strong-field physics in the nonperturbative regime.

arXiv:2402.17136v1 [hep-ph] 27 Feb 2024

I. INTRODUCTION

Super dense matter, such as that realized inside a neutron star or even denser, can be produced on Earth by colliding heavy ions at intermediate collision energies $\sqrt{s_{\rm NN}} = \mathcal{O}(3\text{ - }10\text{ GeV})$. Such collision experiments have been performed in the Beam Energy Scan program at RHIC [1] and are planned worldwide (e.g., FAIR [2], NICA [3], HIAF [4], J-Parc-HI [5]) to reveal the extreme form of matter in the dense limit and to develop a better understanding of strong interaction, or quantum chromodynamics (QCD). These experimental programs have motivated various theoretical studies, which are mainly aimed at investigating the consequences of the high-density matter and the dynamics of how it can be created during the collisions, e.g., novel phases of QCD at finite density (see Ref. [6] for a review) and the development of various transport models to simulate the real-time collision dynamics such as RQMD [7], UrQMD [8, 9], JAM [10], and SMASH [11].

The purpose of this paper is, rather than pursuing the high-density physics as previously discussed, to point out that a strong electromagnetic field can be created in intermediate-energy heavy-ion collisions. The generation of such a strong electromagnetic field is of interest not only to hadron/QCD physics but also to the area of strong-field physics. For hadron/QCD physics, electromagnetic observables such as di-lepton yields [12–15], which are promising probes of nontrivial processes induced by the high-density matter, are naturally affected by the presence of a strong electromagnetic field. A correct estimation of the electromagnetic-field profile (and also its implementation into transport-model simulations; cf. Ref. [16, 17]) is, therefore, important when extracting/interpreting signals of the high-density matter from the actual experimental data. As for strong-field physics, the generation of a strong electromagnetic field would provide a unique and novel opportunity to study quantum electrodynamics (QED) in the nonperturbative regime beyond the Schwinger limit $eE_{\rm cr} := m_e^2 = (0.511\text{ MeV})^2$ (with $e = |e|$ being the elementary electric charge, E electric field strength, and m_e the electron mass). Currently, strong-field physics is driven mainly by high-power lasers (see, e.g., Refs. [18, 19] for reviews). The focused laser intensity of $I = 1\times10^{23}$ W/cm^2 (corresponding to $E \approx 10^{-3}\,E_{\rm cr}$) is the current world record [20], which is envisaged to be surpassed by the latest and future facilities such as Extreme Light Infrastructure (ELI) $I = \mathcal{O}(10^{25}$ W/cm$^2)$ [21]. Although the laser intensity is growing rapidly, it is and will remain, for at least the next decade, several orders of magnitude below the Schwinger limit $E_{\rm cr}$. Therefore, it is difficult to study strong-field phenomena with current lasers. This means that a novel method or physical system to realize a strong electromagnetic field is highly demanded.

There exist a number of studies on the generation of a strong electromagnetic field at low- and high-energies both theoretically and experimentally. Let us briefly review them, so as to clarify our motivation to go to the intermediate energy. At low energies, due to the baryon stopping (i.e., the Landau picture [22, 23]), the collided ions stick together at the collision point and form up a gigantic ion with large atomic number $Z = \mathcal{O}(100)$, and thereby creates a strong Coulomb electric field of the order of $eE \sim (e^2/4\pi)Z/R^2 = \mathcal{O}((20\text{ MeV})^2)$, where $R = \mathcal{O}(10\text{ fm})$ is the typical radius of the gigantic ion. The produced field is weak compared to the hadron/QCD scale but is far surpassing the Schwinger limit of QED. Thus, it is expected to induce intriguing nonlinear QED processes such as the vacuum decay (see Ref. [24] for a recent analysis), experimental investigation of which has been done around 1980s but is not conclusive yet (see, e.g., Ref. [25] for possible interpretations of the experimental results). On the other hand, at high energies, the Bjorken picture [26] is valid rather than the Landau picture. The colliding ions penetrate with each other

* hidetoshi.taya@riken.jp

without sticking, and hence they do not form up a gigantic ion with large Z, unlike the low-energy case. Nevertheless, high-energy heavy-ion collisions are able to produce a very strong electromagnetic field with a different mechanism. Namely, due to a strong Lorentz contraction, the charge density ρ of the incident ions are enhanced by the Lorentz factor $\gamma \approx (\sqrt{s_{\rm NN}}/1\ {\rm GeV})/2$ as $\rho \to \gamma\rho \approx \gamma e\rho_0/2$, where $\rho_0 \approx 0.168\ {\rm fm}^{-3}$ is the nuclear saturation density and we halved it because roughly the half of the ions is composed of charged nucleons (i.e., proton). Due to this enhancement, at central collision events, a strong Coulomb electric field with peak strength $eE \sim e\rho R$, which can exceed the hadron/QCD scale at the RHIC/LHC energy scale $\sqrt{s_{\rm NN}} = \mathcal{O}(100\text{ - }1000\ {\rm GeV})$ and can achieve $eE = \mathcal{O}(100\text{ - }1000\ {\rm MeV})$ [27], is produced at the instant of when the colliding ions maximally overlap with each other [and also a strong magnetic field can be produced in non-central events at the moment when the ions pass through each other due to the Ampère law (see Ref. [28] for a review), which is experimentally utilized to test intriguing nonlinear QED processes such as the photon-photon scattering [29] and the linear Breit-Wheeler process [30, 31]].

Besides the field strength, there is a crucial difference between the low- and high-energy cases: the lifetime of the produced field. The lifetime of the field is long at low energies due to the baryon stopping. For example, it has been shown that the lifetime τ can reach $\tau = \mathcal{O}(100\text{ - }1000\ {\rm fm}/c)$ for collision energies close to the Coulomb barrier [32–34]. In contrast, the lifetime is extremely short at high energies. The produced field can survive only for the instance that the colliding ions pass through each other, which is strongly suppressed by the Lorentz factor as $\tau \sim R/\gamma = \mathcal{O}(0.1\text{ - }0.01\ {\rm fm}/c)$ at the RHIC/LHC energy.

The shortness of the lifetime τ significantly affects the *nonperturbativity* of the physics induced by the strong field. Namely, no matter how strong a field is, the physics has to be perturbative if it is short-lived. The low-order perturbation theory becomes sufficient in such a limit, and therefore the physics becomes "trivial" in the sense that it is not very different from the usual electromagnetic processes in the vacuum. Intuitively, this is simply because there is no time for the finite field to have multiple interactions with a particle. To be more quantitative, suppose we have, as an example, an electric field with peak strength E_0 and lifetime τ, and consider the vacuum pair production by such a strong electric field, vacuum $+ E \to e^+ + e^-$. For this case, the interplay between the nonperturbative and perturbative pair production is controlled by two dimensionless quantities [35–44],

$$\xi(m) := \frac{eE_0\tau}{m} \quad \text{and} \quad \nu := eE_0\tau^2\ . \tag{1}$$

Note that the parameter ξ is known as the Keldysh parameter [45] and, depending on the context, is also called the classical nonlinearity parameter [19]. The mass m is the mass of the particle to be produced. For enough strong and long-lived fields such that $\xi, \nu \gg 1$, the pair production becomes nonperturbative in the sense that the rate of the pair production acquires a non-analytic dependence in e and E_0 as $\propto \exp[-(\text{const.}) \times (m^2/eE_0)]$. In the opposite limit, $\xi, \nu \ll 1$, it becomes purely perturbative, i.e., the low-order perturbative treatment becomes sufficient, meaning that the rate only has the power dependence in e and E_0 as $\propto (eE_0/m^2)^n$ $(n \in \mathbb{N})$. This example of the vacuum pair production clearly demonstrates that the supercriticality $eE_0 \gtrsim m^2$ is not sufficient to guarantee the nonperturbativity of strong-field processes. One must, thus, pay attention to the magnitude of the lifetime τ, or the resulting ξ and ν, as well, in addition to the strength E_0. In terms of these nonperturbativity parameters (1), high-energy heavy-ion collisions at the typical RHI/LHC scale correspond to $\nu = \mathcal{O}(0.1)$ and $\xi = \mathcal{O}(100)$ for electron $m = m_e$ and $\mathcal{O}(1)$ for pion $m = m_\pi \approx 140$ MeV. This means, although ξ can be nonperturbatively large for the QED scale $m = m_e$ due to the largeness of the field strength, ν always remains perturbatively small due to the shortness of the lifetime. For the hadron/QCD scale $m = m_\pi$, neither ξ nor ν can be nonperturbative. Therefore, in either case of QED or QCD, the pair production cannot be nonperturbative, implying that strong-field physics in the nonperturbative regime cannot be explored with the field produced in high-energy heavy-ion collisions. This naive argument, based on the "order parameters" of the vacuum pair production, is consistent with the actual experimental results. There have been observations of next-leading-order QED processes such as light-by-light scattering [29] and linear Breit-Wheeler pair production [30], in which the experimental results are consistent with perturbative QED calculations and no signatures of higher-order nonlinear effects have been detected.

We are now in a position to address the advantage of going to intermediate energies. It is natural to expect that the electromagnetic field generated in intermediate-energy heavy-ion collisions should have characteristics between the low- and high-energy cases. Namely, although the field strength would be weaker than that in the high-energy case, it should be stronger than that in the low-energy case. This means that the produced field will remain much stronger than the Schwinger limit of QED and may still be comparable to the hadron/QCD scale. As for lifetime, we can naturally expect that the field produced at intermediate energies should survive longer than the high-energy field. If this is true, it overcomes the problem of the short lifetime in the high energy limit, while maintaining a sufficiently large field strength. Hence, it enables us to access the strong-field physics in the nonperturbative regime. It is therefore worthwhile to investigate the generation of a strong electromagnetic field in intermediate-energy heavy-ion collisions and to make a realistic estimate to clarify whether this is the case, and, if so, how important it is.

Based on the motivations explained thus far, we would like to make a realistic estimate of the electromagnetic

field in intermediate-energy heavy-ion collisions based on a hadron-transport-model simulation. Indeed, neither the Landau nor the Bjorken picture is complete at intermediate energies (cf. Ref. [46] as a related for vorticity estimation). Accordingly, the generation mechanism of the electromagnetic field is more involved compared to the low- and high-energy limits, and hence it is impossible to carry out a simple analytic estimate at intermediate energies. In particular, dynamical effects become more important than the other energy regimes. For example, both elastic- and inelastic- multiple collision processes among the constituent nucleons are important to determine the sticking or non-sticking of the colliding ions. Hadrons are produced or decay during the time evolution through the inelastic processes, meaning that the simple but widely-used modeling based the Liénard-Wiechert potential of the electromagnetism is not applicable[1]. The Lorentz contraction comes into play, unlike at the low energy limit, but is not as strong as that at the high energy, and therefore hadrons consisting of the colliding ions need to be treated as anisotropic objects with finite size. All of those dynamical effects can be taken into account systematically by using established hadron-transport models of heavy-ion collisions. In this work, we adopt JAM (Jet AA Microscopic transport model) [10, 47]. JAM is a hadronic cascade model to simulate the realtime dynamics of heavy-ion collisions. In addition to elastic hadron-hadron scatterings, inelastic processes are implemented in JAM via resonance production, soft-string excitation, and multiple mini-jet production, so that it can be applied to a wide range of collision energies from $\sqrt{s_{\rm NN}} = \mathcal{O}(2\ {\rm GeV})$ up to $\mathcal{O}(100\ {\rm GeV})$. More details of JAM, including its comparison with other models, can be found in, e.g., Ref. [48].

This paper is organized as follows. We first explain our numerical setup and how we calculate the electromagnetic field with JAM in Sec. II. We then present the numerical results in Sec. III that include the spacetime profiles of the charge density (Sec. III A) and the resulting electromagnetic field (Sec. III B), the time evolution of the peak strength (Sec. III C), and the corresponding nonperturbativity parameters ξ and ν to demonstrate that the field is nonperturbatively strong (Sec. III D).

Notation: Our metric is the mostly minus, i.e., $g^{\mu\nu} := {\rm diag}(+1,-1,-1,-1)$. Spatial three vectors are indicated by the bold letters, e.g., $x^\mu := (t, \boldsymbol{x})$ for spacetime coordinates and $A^\mu = (A^0, \boldsymbol{A})$ for the four-vector potential. We take the z-axis along the beam direction and write $\boldsymbol{x} =: (x, y, z) =: (\boldsymbol{x}_\perp, z)$.

[1] The Liénard-Wiechert potential is explicitly dependent on time-derivatives of the trajectory of a charged particle $\boldsymbol{r}(t)$. Therefore, it is valid only if $\boldsymbol{r}(t)$ does not have any singularities, which means the particle cannot decay nor be produced during the time evolution.

II. NUMERICAL RECIPE

A. Calculation program

We wish to calculate the electric $\boldsymbol{E}$ and magnetic $\boldsymbol{B}$ fields in heavy-ion collisions at intermediate collision energies. The electromagnetic fields are given by

$$\boldsymbol{E} := -\partial_t \boldsymbol{A} - \boldsymbol{\nabla} A^0 \quad \text{and} \quad \boldsymbol{B} := \boldsymbol{\nabla} \times \boldsymbol{A}\ , \tag{2}$$

where A^μ is the retarded vector potential,

$$A^\mu(t, \boldsymbol{x}) := \frac{1}{4\pi} \int {\rm d}^3\boldsymbol{x}' \frac{J^\mu(t - |\boldsymbol{x} - \boldsymbol{x}'|, \boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|}\ , \tag{3}$$

with J^μ being the total electric current carried by the charged particles in the collisions. In the present paper, we focus on the event-averaged values of $O = \boldsymbol{E}, \boldsymbol{B}$ and take the event averaging of Eq. (2) as

$$O := \frac{1}{N} \sum_{i=1}^{N} O_i\ , \tag{4}$$

where O_i is the event-by-event result and N is the total number of the events. Note that event-by-event fluctuations may be important in actual experiments (see also Ref. [27] for the high-energy case). For example, it has been argued in Refs. [49–51] that the maximum value of the baryon density can fluctuate event-by-event by more than 30 % at intermediate energies. Such large fluctuations of the baryon density $\rho_{\rm B}$ imply that the charge density $\rho/e \approx \rho_{\rm B}/2$ should also fluctuate, and so does the resulting electromagnetic field. We shall report the event-by-event fluctuations of the electromagnetic fields in intermediate-energy heavy-ion collisions in a forthcoming paper.

We evaluate the electromagnetic field (2) numerically as follow. We first calculate the electric current J^μ with JAM. JAM calculates the phase-space distributions of hadrons in heavy-ion collisions $(\boldsymbol{x}_n(t), p_n^\mu(t))$, where the subscript n labels the hadrons in a collision event. The electric current J^μ at a position $(t', \boldsymbol{x}')$ is then given in terms of the JAM phase-space distribution as

$$J^\mu(t', \boldsymbol{x}') := \sum_{n \in \text{all hadrons}} \frac{p_n^\mu(t')}{p_n^0(t')} \rho_n(t', \boldsymbol{x}')\ , \tag{5}$$

where ρ_n is the charge density of a single charged hadron. In JAM, hadrons are treated as if they are point-like, and hence the charge density of each hadron is localized strictly at $\boldsymbol{x}_n$. Hadrons are finite-sized in reality, motivated by which we smear the charge density by using the (relativistic) Gaussian distribution:

$$\rho_n(t', \boldsymbol{x}') := q_n \frac{\gamma_n(t')}{(\sqrt{2\pi}\sigma)^3} \times \exp\left[-\frac{|\boldsymbol{x}' - \boldsymbol{x}_n(t')|^2 + \gamma_n^2(t') \left(\boldsymbol{v}_n(t') \cdot (\boldsymbol{x}' - \boldsymbol{x}_n(t'))\right)^2}{2\sigma^2} \right]\ ,$$

where q_n is the electric charge, σ the smearing width, and $\gamma := \sqrt{1-\boldsymbol{v}_n^2} := \sqrt{1-(\boldsymbol{p}_n/p_n^0)^2}$ the Lorentz factor for the n-th hadron.

Once the electric current J^μ is obtained with JAM, can we calculate the retarded vector potential $A^\mu(t,\boldsymbol{x})$ by carrying out the $\boldsymbol{x}'$ integration according to Eq. (3) at each spacetime point $(t,\boldsymbol{x})$. This can be done with the standard numerical integration schemes, e.g., the Gauss-Legendre quadrature method, which we have adopted in this work.

In the default setting of JAM, the simulation starts from some initial time $t_{\rm in} =: 0$, around which the incident ions start to collide. The JAM hadron phase-space data is, thus, available only later than $t_{\rm in}$, so is the electric current J^μ (5). To perform the integration (3), the information of the electric current at time $t' = t-|\boldsymbol{x}-\boldsymbol{x}'|$, which can be earlier than $t_{\rm in}$ (when $|\boldsymbol{x}-\boldsymbol{x}'| > t - t_{\rm in}$), is needed. To get the electric current before the initial time $t_{\rm in}$, we extrapolate the phase-space data at the initial time $t = t_{\rm in}$ by assuming that the hadrons that constitutes the incident ions go straight trajectories in the phase-space without any interactions:

$$\begin{aligned} \boldsymbol{x}_n(t<t_{\rm in}) &:= \boldsymbol{x}_n(t_{\rm in}) + (t-t_{\rm in})\frac{\boldsymbol{p}_n(t_{\rm in})}{p_n^0(t_{\rm in})} \,, \\ p_n^\mu(t<t_{\rm in}) &:= p_n^\mu(t_{\rm in}) \,. \end{aligned} \tag{6}$$

Having calculated the retarded vector potential A^μ (3), can we obtain the electromagnetic fields via the differentiation (2). We in this work have adopted the standard central difference method with discretizing the $\boldsymbol{x}$ coordinates with uniform meshes.

B. Numerical parameters

We summarize the numerical parameters adopted in the present work in Table I.

Using JAM, we simulate head-on (i.e., the impact parameter b is strictly fixed at $b = 0$) collisions of gold ions at the intermediate energies $\sqrt{s_{\rm NN}} = 2.4, 3.0, 3.5, 3.9, 4.5, 5.2, 6.2, 7.2, 7.7, 9.2,$ and 11.5. Several advanced options are available in JAM (such as the mean-field potential and hybrid hydrodynamic simulation), which we switch off and run JAM with the default setting to carry out the most basic hadronic cascade simulation. Note that, although we have concentrated on gold collisions as a first step to study the electromagnetic-field physics in intermediate-energy heavy-ion collisions, it is an interesting question to consider other ion species. In particular, deformed ions such as uranium and also asymmetric collisions such as Au + Cu (see, e.g., Refs. [52–54] at high energies) would be of interest, in which case the resulting electromagnetic fields should have a preferred direction due to the collision geometry, leading to observable effects such as a charged v_1 flow. The impact parameter dependence is also an interesting issue. With finite impact parameters magnetic field shall dominate, while electric field does so in head-on events as we shall demonstrate below. We shall report the interplay of the dominance between the electric and magnetic fields in a future publication.

We have five numerical parameters in the evaluation of the electromagnetic field outlined in Sec. II A: the event number N, the smearing width σ, and the discretizations of $t, \boldsymbol{x}$, and $\boldsymbol{x}'$. To reduce the numerical cost, we are reluctant to take a relatively small event number $N = 100$, although the residual event-by-event fluctuations do not look very significant after the event averaging as we show later. We shall use the particular value of the smearing width $\sigma = 1$ fm for all hadron species, which is physically motivated by that the typical size of a nucleon is of the order of 1 fm and is a common choice in the literature. We have checked that the numerical results are less sensitive to the value of σ after taking the event averaging. We have also carefully checked that the lattice volumes of $\boldsymbol{x}$ and $\boldsymbol{x}'$ and the mesh sizes $\Delta t, \Delta\boldsymbol{x}$, and $\Delta\boldsymbol{x}'$ are sufficiently large and fine, respectively, in such a way that the numerical results are not sensitive to them. Note that, for the present parameters, a single numerical simulation (per energy) consumes $\mathcal{O}(20$ GB) RAM and takes about a few days with a standard GPU computer available on the market.

III. NUMERICAL RESULTS

We present the numerical results obtained with the recipe outlined in Sec. II. The main result is that an electric field which is non-perturbatively strong and long-lived can be created in intermediate-energy heavy-ion collisions; see Sec. III D, in particular Fig. 4. Before getting there, we discuss the generation of such a strong electric field step by step. We first demonstrate in Sec. III A that a highly-charged matter $\rho/e = \mathcal{O}(0.5\ \text{fm}^{-3})$, as a reminiscent of the dense baryon matter, is realized at intermediate energies. The highly-charged matter sources a strong electric field, whose spacetime profile shall be discussed in Sec. III B. We shall discuss the intensity and the lifetime of the produced electric field in detail in Sec. III C and show in Sec. III D that it can be nonperturbatively strong and long-lived, in contrast to what is realized at high energies.

A. Spacetime profile of the charge density

Fig. 1 shows the spacetime profile of the charge density. The key observation is that intermediate-energy heavy-ion collisions can realize a high charge density, typically of the order of $\rho/e = \mathcal{O}(0.5\ \text{fm}^{-3})$, for a relatively long time $\tau = \mathcal{O}(10\ \text{fm}/c)$. The magnitude of the charge density is roughly $\mathcal{O}(5)$ times greater than that of an usual charged ion at rest $\rho/e \approx \rho_0/2 \approx 0.08\ \text{fm}^{-3}$. This highly-charged matter sources a strong electric field, as we shall demonstrate later.

Ion species	^{197}Au
Collision energy $\sqrt{s_{\rm NN}}$	2.4, 3.0, 3.5, 3.9, 4.5, 5.2, 6.2, 7.2, 7.7, 9.2, and 11.5 GeV
Impact parameter b	0 fm
Gaussian smearing width σ	1 fm
Total event number N	100
Discretization of time t	$t \in [0, 30]$ fm/c with $\Delta t = 0.5$ fm/c
Discretization of $\boldsymbol{x} = (x, y, z)$	$x, y, z \in [-30, +30]$ fm with $(n_x, n_y, n_z) = (121, 121, 121)$ sites
	(a uniform 3D lattice with $\Delta x = \Delta y = \Delta z = 0.5$ fm)
Discretization of $\boldsymbol{x}' = (x', y', z')$	$x', y', z' \in [-30, +30]$ fm with $(n_{x'}, n_{y'}, n_{z'}) = (120, 120, 120)$ sites
	(a non-uniform 3D lattice for the Gauss-Legendre quadrature method)

TABLE I. Parameters adopted in the numerical simulation.

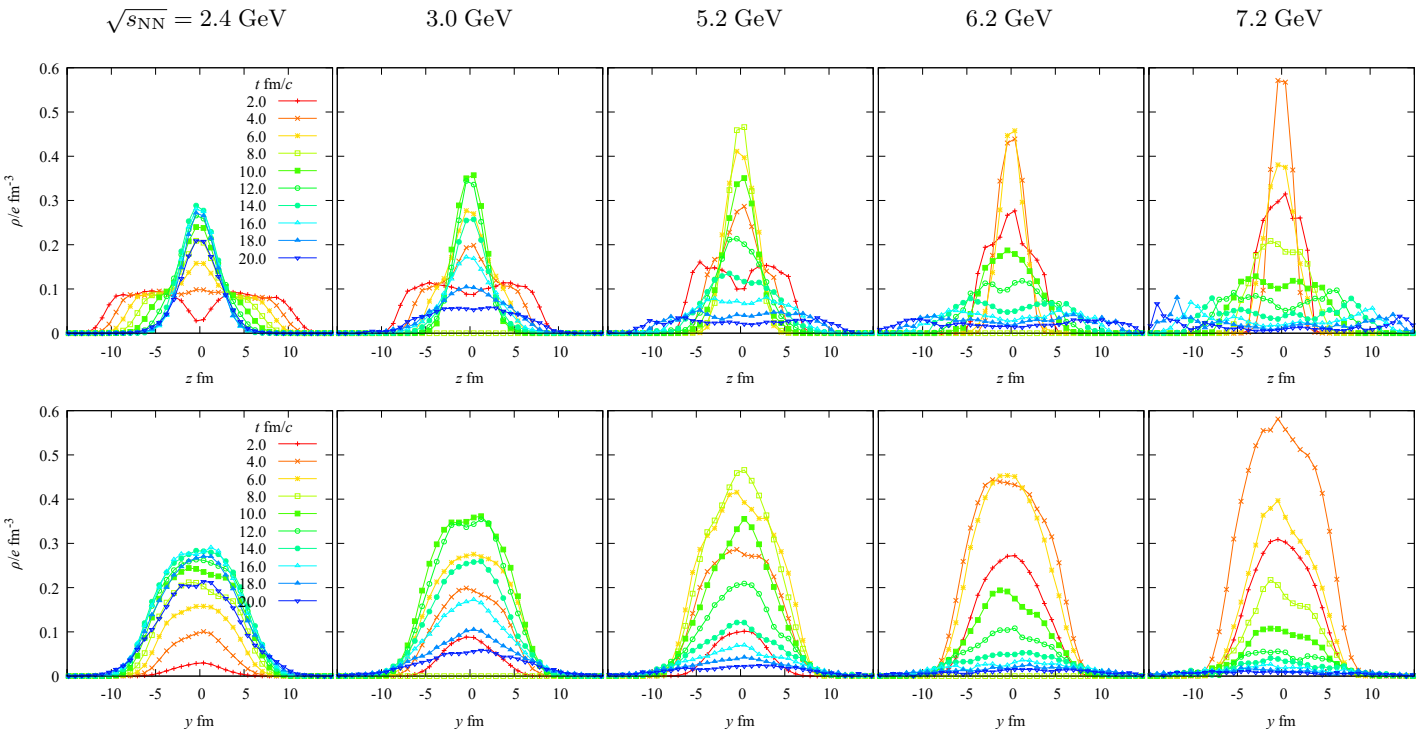


FIG. 1. The spacetime profile of electric charge density ρ/e at some selected collision energies ($\sqrt{s_{\rm NN}} = 2.4, 3.0, 5.2, 6.2$, and 7.2 GeV). The different colors indicate different time slices, ranging from 2 fm/c (red) to 20 fm/c (blue) with the increment of 2 fm/c. The top and bottom panels correspond to the distributions in the beam z direction at fixed $x = y = 0$ fm and in the transverse y direction at fixed $x = z = 0$ fm, respectively.

There are two important physics that contribute to the creation of the highly-charged matter. The first one is the Lorentz contraction. Due to the Lorentz contraction in the beam direction, the incident ions are not spherical, but anisotropic, and the longitudinal size is shortened as $R \to R/\gamma \approx (13\ \text{fm})/(\sqrt{s_{\rm NN}}/1\ \text{GeV})$, where $R \approx (1.1\ \text{fm}) \times A^{1/3} \approx 6.4$ fm is the radius of a gold ion $A = 197$ at rest and we used $2\gamma \approx (\sqrt{s_{\rm NN}}/1\ \text{GeV})$. Accordingly, the charge density is enhanced from that at rest by the Lorentz factor γ and hence becomes larger with collision energy as $\rho/e \approx \gamma \times \rho_0/2 \approx (0.04\ \text{fm}^{-3}) \times (\sqrt{s_{\rm NN}}/1\ \text{GeV})$. The collision dynamics can roughly be understood as an overlapping process of these two Lorentz-contracted ions. The maximum charge density is then achieved when the two ions maximally overlap with each other. As a zero-th order approximation, we may neglect the interaction between the ions and hence assume that they just pass through each other during a collision. Then, the maximum can be estimated as $\max \rho/e \approx 2 \times \rho/e \approx (0.08\ \text{fm}^{-3}) \times (\sqrt{s_{\rm NN}}/1\ \text{GeV})$ and the time of which achieved as $t_{\rm max} \approx 2R/\gamma \approx (26\ \text{fm}/c)/(\sqrt{s_{\rm NN}}/1\ \text{GeV})$. These estimates are in rough agreement with the numbers shown Fig. 1.

The second is the interaction (i.e., the baryon stop-

ping), which is especially important in determining the lifetime of the highly-charged matter. The importance of the interaction is evident in Fig. 1. If it is non-interacting, the charge density should exhibit a two-peak structure after the maximum, i.e., the collided ions pass through each other and leave nothing in the mid-rapidity $z \approx 0$ (the Bjorken picture [26]). This is not the case in Fig. 1, which shows that the density is single-peaked at the mid-rapidity after the collision. In other words, the interaction makes it difficult for the colliding ions to penetrate each other, instead causing them to merge at the point of collision (the Landau picture [22, 23]). The interaction makes the lifetime τ of the highly-charged matter longer, compared to what is naively expected from the Bjorken picture $\tau \approx 2R/\gamma \approx (26\ \text{fm}/c)/(\sqrt{s_{\rm NN}}/1\ \text{GeV})$. For example, at $\sqrt{s_{\rm NN}} = 3.0$ GeV, a relatively dense charge state $\rho/e \gtrsim 0.1\ \text{fm}^{-3}$ appears around $t \approx 2$ fm/c and survives until $t \approx 18$ fm/c, whose lifetime $\tau \approx 18 - 2$ fm/c = 16 fm/c is much longer than the naive Bjorken estimate $\tau \approx 26/3$ fm/$c \approx 9$ fm/c. The interaction effect is non-negligible and hence a highly-charged matter with a relatively long lifetime is realized in the intermediate energy values $\sqrt{s_{\rm NN}} = \mathcal{O}(3\text{ - }10\ \text{GeV})$.

B. Spacetime profile of the produced field

The highly-charged matter created in intermediate-energy heavy-ion collisions (see Fig. 1) produces a strong electric field. To show this, we plot, in Fig. 2, an electromagnetic Lorentz invariant,

$$\mathcal{F} := \boldsymbol{E}^2 - \boldsymbol{B}^2 \,. \tag{7}$$

Note that the other invariant $\mathcal{G} := \boldsymbol{E} \cdot \boldsymbol{B}$ is found to be vanishing (up to residual event-by-event fluctuations). This is reasonable, since we have no electric current $\boldsymbol{J}$ to produce a macroscopic magnetic field in central collision events and hence $\boldsymbol{B} \approx \boldsymbol{0}$. This also means

$$\mathcal{F} \approx \boldsymbol{E}^2 \,. \tag{8}$$

That is, the produced field is electric, which is also evident in Fig. 2 that the sign of $\mathcal{F}$ is always positive. It is thus convenient to introduce an *effective* electric field strength $E_{\rm eff}$ such that

$$E_{\rm eff} := \sqrt{\mathcal{F}} \,. \tag{9}$$

Fig. 2 shows that the typical strength of the produced electric field is $e^2\mathcal{F} = \mathcal{O}(10^{6\text{ - }7}\ \text{MeV}^4)$, corresponding to $eE_{\rm eff} = \mathcal{O}((30\text{ - }60\ \text{MeV})^2)$. The created field is strong in the sense that it is far beyond the Schwinger limit of QED, $eE_{\rm cr} = m_e^2 = (0.511\ \text{MeV})^2$. It is not supercritical to the hadron/QCD scale, $m_\pi \approx 140$ MeV, but still is non-negligibly strong $eE_{\rm eff} = \mathcal{O}((30\ \% \times m_\pi)^2)$, implying that it can affect hadron/QCD processes. The produced field also has a sufficiently large spacetime volume $\tau \times V = \mathcal{O}(10\ \text{fm}/c) \times \mathcal{O}((10\ \text{fm})^3)$ because of the baryon stopping and of that the highly-charged matter has a macroscopically large spatial volume $\propto R^3/\gamma$.

Let us have a closer look at the spacetime structure of the produced field. First, the produced field has a donuts-like spatial structure, with peaks located around where the charge density vanishes (see Fig. 1). Namely, the effective field strength $E_{\rm eff}$ is zero at the collision point $\boldsymbol{x} \approx \boldsymbol{0}$, and, as going away from the collision point, $E_{\rm eff}$ increases almost linearly $\propto |\boldsymbol{x}|$ and then decreases quadratically $\propto 1/|\boldsymbol{x}|^2$ after it gets peaked. These are the direct consequences of the Gauss law $\text{div}\,\boldsymbol{E} = \rho$ (and is similar to the textbook exercise of the electric field produced by a uniformly charged sphere; see, e.g., Jackson [55]).

Second, the field strength in the transverse plane (the bottom panel in Fig. 2) increases with collision energy $\sqrt{s_{\rm NN}}$, while that along the beam direction (the top panel in Fig. 2) is less sensitive to $\sqrt{s_{\rm NN}}$ and is weaker than that in the transverse plane. This is a reminiscent of the anisotropic structure of the charged matter due to the Lorentz contraction (see Fig. 1) and can be understood simply with the Gauss law. For simplicity, let us model the matter as a uniformly charged short cylinder, with radius $|\boldsymbol{x}_\perp| \le R$ and length $|z| \le R/\gamma$. Then, the Gauss law tells us that the field strength is maximized along the longitudinal direction at the ends of the charged cylinder $|z| = R/\gamma$ and the corresponding strength is $E = \rho R/\gamma$. Similarly, the maximum in the transverse plane is $E = \rho R/2$ at $|\boldsymbol{x}_\perp| = R$. Thus, the field strength in the transverse plane is enhanced by the anisotropy factor γ compared to that along the longitudinal axis.

Note that the charge density ρ can roughly be estimated as a simple overlapping of two Lorentz-contracted ions $\rho/e \approx \gamma\rho_0$ (see Sec. III A). Therefore, the maximum field strength over the space can be estimated, by using the simple modeling in the last paragraph, as $eE \approx e^2\gamma\rho_0 R/2 \approx ((\sqrt{s_{\rm NN}}/1\ \text{GeV})^{1/2} \times 31\ \text{MeV})^2$. This simple estimate is in rough agreement with the actual numbers shown in Fig. 2 (albeit a bit overestimating, which we discuss in Sec. III C).

Third, the produced field expands faster with increasing collision energy. For lower energies $\sqrt{s_{\rm NN}} \lesssim 5$ GeV, the expansion is negligible and the field is roughly staying around the original position created. In contrast, for higher energies $\sqrt{s_{\rm NN}} \gtrsim 5$ GeV, the expansion becomes significant and the field moves almost at the speed of light. For the longitudinal expansion, this is a natural consequence that the Bjorken picture begins to be effective for higher energies. Although the Landau picture dominates in the intermediate-energy regime as we have explained in Sec. III A, it is also true that the Bjorken picture is becoming effective for higher energies, which can be observed in Fig. 1 from that the charged matter expands in the beam direction after created and begins to exhibit a two-peak structure at, e.g., $\sqrt{s_{\rm NN}} = 7.2$ GeV. The expansion also becomes faster in the transverse direction, simply because the typical kinetic energy of the particles in the collision system increases with collision

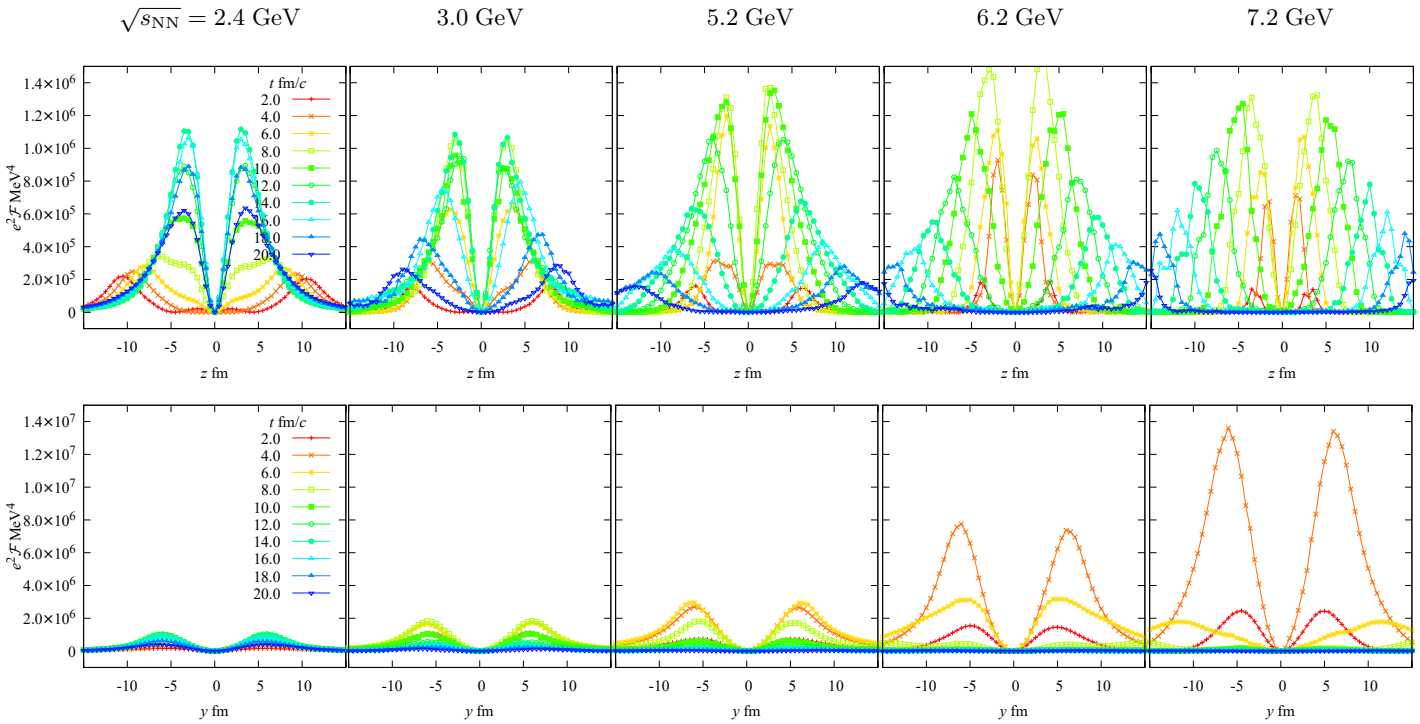


FIG. 2. The typical spacetime profile of the Lorentz invariant $\mathcal{F} = \boldsymbol{E}^2 - \boldsymbol{B}^2$ in intermediate-energy heavy-ion collisions. The plotting style is the same as Fig. 1: the time slices are indicated by the colors, and the top and bottom panels correspond to the distributions in the z and y directions, respectively. Note that the vertical scale is ten times greater in the bottom than in the top.

energy, allowing them to move faster.

C. Field strength

Having observed the generation of a strong electric field in intermediate-energy heavy-ion collisions, we turn to discuss more quantitative aspects of the produced field, in particular the peak field strength over the space.

We first discuss the time evolution of the field strength; see Fig. 3 (i). The strength gets maximized when the incident ions overlap with each other the most and shows up a sharp peak at around $t \approx (26\ \text{fm}/c)/(\sqrt{s_{\rm NN}}/1\ \text{GeV})$ (within the non-interacting estimate; see Sec. III A). After the peak, the field strength decays rather slowly, as the collision remnant can survive in the mid-rapidity region due to the baryon stopping, and consequently the field can maintain its strength for a relatively long time. For example, a single ion at rest has a field strength of $eE \lesssim e^2(\rho_0/2)R/3 \approx 25$ MeV, above which strength is kept for more than $\mathcal{O}(15\ \text{fm}/c)$ in intermediate-energy heavy-ion collisions.

The peak in Fig. 3 (i) becomes larger and sharper with increasing collision energy. This is also evident in Fig. 3 (ii) and (iii), which show, respectively, the maximum strength of the peak, $\max E_{\rm eff}$, and the full width at the half maximum of $E_{\rm eff}$ (FWHM),

$$\tau := \int_{E_{\rm eff}(t) > \max E_{\rm eff}/2} \mathrm{d}t\ . \tag{10}$$

Note that the FWHM (10) can naturally be regarded as the lifetime of the field produced, and so we use Eq. (10) to *define* the lifetime in what follows.

The maximum peak field strength increases with collision energy [see Fig. 3 (ii)], as anticipated with the simple estimate $eE \approx e^2\gamma\rho_0 R/2 \approx ((\sqrt{s_{\rm NN}}/1\ \text{GeV})^{1/2} \times 31\ \text{MeV})^2$ that we made in Sec. III B. The key point is the enhancement of the charge density by the Lorentz contraction, which yields to the energy dependence. Although the simple estimate can capture the qualitative feature of the peak field strength, it fails at the quantitative level mainly because the modeling with a uniformly-charged short cylinder is crude. In particular, the actual charge distribution (see Fig. 1) is tailed and thus the charge extends more than the rigid cylinder distribution, which makes the actual peak field strength smaller than the simple estimate. Indeed, we find

$$\max \sqrt{eE_{\rm eff}} = \left(\frac{\sqrt{s_{\rm NN}}}{1\ \text{GeV}}\right)^{1/2} \times 22.38\ \text{MeV}\ , \tag{11}$$

which has the smaller coefficient (31 MeV $\to$ 22 MeV), can fit the numerical result well.

The lifetime (10) decreases with collision energy [see Fig. 3 (ii)], which is essentially due to the Lorentz

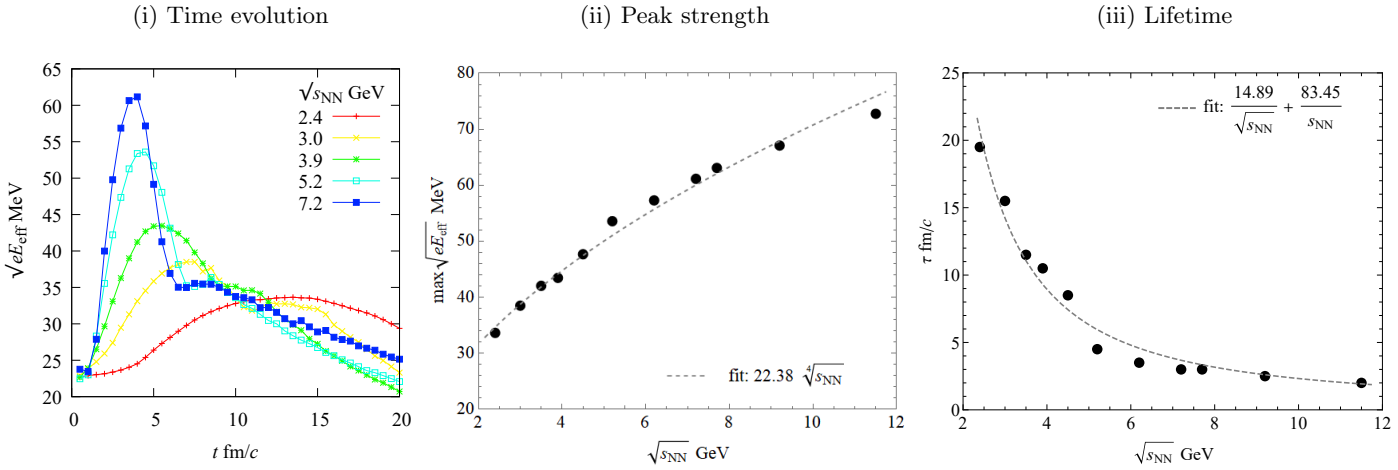


FIG. 3. The effective electric-field strength $E_{\rm eff}$ in intermediate-energy heavy-ion collisions. (i) The time evolution of $E_{\rm eff}$ for some selected collision energies. (ii) The maximum peak strength of $E_{\rm eff}$, extracted from (i), plotted against collision energy $\sqrt{s_{\rm NN}}$. (iii) The lifetime of the field, defined as FWHM of $E_{\rm eff}$ (10). The dashed lines in (ii) and (iii) are the numerical fittings of the obtained results.

contraction. The typical lifetime of the field is determined by that of the highly-charged matter. Therefore, $\tau \approx 2R/\gamma \approx (26\ \text{fm}/c)/(\sqrt{s_{\rm NN}}/1\ \text{GeV})$ in the non-interacting limit (see Sec. III A). In reality, however, the interaction is important, especially for lower energies $\sqrt{s_{\rm NN}} \lesssim 5$ GeV, and thus the naive non-interacting estimate only gives a poor fit at the quantitative level. The interaction makes the lifetime considerably longer. We numerically find that such an interaction effect can be reproduced well with a higher-order term $\propto s_{\rm NN}^{-1}$, added to the naive $s_{\rm NN}^{-1/2}$ dependence, and the best fitting curve is found to be

$$\tau = \left(\frac{\sqrt{s_{\rm NN}}}{1\ \text{GeV}}\right)^{-1} \times 14.89\ \text{fm}/c + \left(\frac{\sqrt{s_{\rm NN}}}{1\ \text{GeV}}\right)^{-2} \times 83.45\ \text{fm}/c\ . \qquad (12)$$

Note that the fitting curve (12) does not reproduce the coefficient of the naive estimate, 26 fm/c, even in the limit of $\sqrt{s_{\rm NN}} \to \infty$. This can be understood as an indication that the Bjorken picture cannot be complete and the interaction is important at intermediate energies.

We emphasize that the lifetime is affected significantly by the interaction and made considerably longer, while the peak strength is less affected. Accordingly, the lifetime is more dependent on collision energy than the peak strength is. As we see shortly below, this shall be the essence why we can have a nonperturbatively strong electric field in intermediate-energy heavy-ion collisions, in particular for $\sqrt{s_{\rm NN}} \lesssim 5$ fm/c, where the interaction becomes more important.

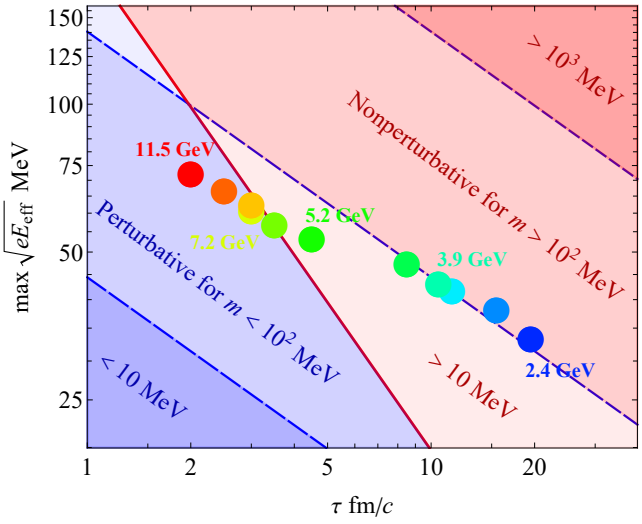


FIG. 4. Sensitivity plot for nonperturbativity of the produced electric field in intermediate-energy heavy-ion collisions. The dots represent the characteristics of the field $(\tau, \max\sqrt{eE_{\rm eff}})$ extracted from Fig. 3 at each collision energy $\sqrt{s_{\rm NN}}$, ranging from $\sqrt{s_{\rm NN}} = 2.4$ GeV (blue), 3.0, 3.5, 3.9, 4.5, 5.2, 6.2, 7.2, 7.7, 9.2, to 11.5 GeV (red). The lines represent the nonperturbativity parameters (1): $1 = \xi(10\ \text{MeV})$ (bottom blue dashed), $1 = \xi(10^2\ \text{MeV})$ (middle blue dashed), $1 = \xi(10^3\ \text{MeV})$ (top blue dashed), and $1 = \nu$ (red). Those lines set "phase boundaries" of the nonperturbativity (of the vacuum pair production). The red regions $\xi(m), \nu > 1$ are nonperturbative (for mass scales m), while it is perturbative in the blue regions $\xi(m), \nu < 1$.

D. Nonperturbativity

As one of the possible "order parameters" for the nonperturbativity of a strong field, we calculate the non-

perturbativity parameters ξ and ν (1) and discuss the sensitivity region in intermediate-energy heavy-ion collisions. The result is shown in Fig. 4. It clearly shows that intermediate-energy heavy-ion collisions, in particular with $\sqrt{s_{\rm NN}} \lesssim 5$ fm/c, can access the nonperturbative regime. Namely, although higher collision energies $\sqrt{s_{\rm NN}} \gtrsim 5$ fm/c are advantageous in that the field strength is strong, which can be comparable even to the hadron/QCD scale, it is disadvantageous in that the lifetime gets extremely short and accordingly the physics has to be purely perturbative. On the other hand, although the achievable field strength is weaker in lower collision energies $\sqrt{s_{\rm NN}} \lesssim 5$ fm/c, it is still much stronger than the critical strength of QED and is non-negligible to the hadron/QCD scale, and also the lifetime can be very long, allowing us to enter the nonperturbative regime.

Using the fitting results, Eq. (11) for $\max\sqrt{eE_{\rm eff}}$ and Eq. (12) for τ, we find that the nonperturbativity parameters ξ and ν can be parametrized as

$$\xi = \frac{37.27\ \text{MeV} + \left(\frac{\sqrt{s_{\rm NN}}}{1\ \text{GeV}}\right)^{-1} \times 208.93\ \text{MeV}}{m}\,, \qquad (13)$$
$$\nu = \left(9.34 \times s_{\rm NN}^{-3/4} + 1.67 \times s_{\rm NN}^{-1/4}\right)^2 \approx 87.18 \times s_{\rm NN}^{-3/2}\,.$$

It is clear that both ξ and ν increase with decreasing collision energy $\sqrt{s_{\rm NN}}$, i.e., lowering the energy is more beneficial for the nonperturbativity. The interaction effect in the Landau stopping regime gives a significant contribution here, as the dominant $s_{\rm NN}^{-1/2}$ and $s_{\rm NN}^{-3/2}$ dependencies in ξ and ν, respectively, arise from the $s_{\rm NN}^{-1}$ term in the lifetime τ (12), which we have added to account for the interaction effect.

For clarity, we remark that the meaning of the "nonperturbative" in Fig. 4 is just that the physical observable (or, to be precise, the number of particle and antiparticle pairs with mass m produced from the vacuum by a strong electric field E [35–44]) acquires a nonperturbative dependence $\propto \exp[-(\text{const})\times m^2/eE]$. In other words, being in the nonperturbative regime does not necessarily guarantee the significance of the nonperturbative effect and/or its detectability in actual experiments. In fact, unless we have a field strength comparable to or exceeding the mass scale $eE \gtrsim m^2$, the effect is strongly suppressed by the exponential $\propto \exp[-(\text{const}) \times m^2/eE]$ and is simply negligible. For example, the lowest energy $\sqrt{s_{\rm NN}} = 2.4$ GeV is nonperturbative for $m \approx 100$ MeV, but the field strength is at most $eE = \mathcal{O}((30\ \text{MeV})^2)$ and therefore the corresponding nonperturbative effect is suppressed as $\propto \exp[-(100/30)^2] = \mathcal{O}(10^{-5})$. The exponential suppression becomes mild for $m^2 \lesssim eE$ and hence $m \lesssim 30$ MeV is the mass region for $\sqrt{s_{\rm NN}} = 2.4$ GeV such that the nonperturbative effect becomes significant and/or would be detectable in actual experiments. In this sense of the significance/detectability, the highest collision energy closest to the phase boundary is the most advantageous to observe a nonperturbative effect with the largest m. According to Fig. 4, such a collision energy is $\sqrt{s_{\rm NN}} \approx 5$ GeV and the corresponding mass region is $m \lesssim 50$ MeV.

IV. SUMMARY AND DISCUSSION

We have studied the generation of a strong electric field in head-on collisions of gold ions at intermediate energies $\mathcal{O}$(3 - 10 GeV), based on a hadron transport-model simulation JAM. Our main statement is that intermediate-energy heavy-ion collisions are useful not only to study the densest matter on Earth but also to study strong-field physics in the nonperturbative regime. Namely, we have shown that the produced electric field is as strong as $eE_{\rm eff} = \mathcal{O}((30\text{ - }60\ \text{MeV})^2)$ (see Figs. 2 and 3), which is supercritical to the Schwinger limit of QED and is still non-negligibly strong compared to the hadron/QCD scale. The produced field is sufficiently long-lived and enables us to explorer the nonperturbative regime of strong-field physics up to mass scale of $m \lesssim 50$ MeV (see Fig. 4), which is unaccessible with any other experiments at the present such as high-power lasers and high-energy heavy-ion collisions.

The present work is just a first step, e.g., towards realistic estimations of nonperturbative strong-field effects and/or modifications to hadron/QCD dynamics in intermediate-energy heavy-ion collisions. Let us briefly discuss possible directions and outlook below.

First, it is a very exciting possibility that we can study strong-field physics by using intermediate-energy heavy-ion collisions, since such a truly strong-field regime above the QED critical field strength cannot be achieved with any other experiments at the present. One of the most intriguing targets is the Schwinger effect, i.e., the nonperturbative vacuum pair production, which was predicted more than seventy years ago [56, 57] and applied to various contexts, including the Hawking radiation in a black hole [58], but has never been observed in actual experiments yet. As we have shown, intermediate-energy heavy-ion collisions generate a nonperturbatively strong electric field that is able to induce the Schwinger effect, and therefore can in principle be used as a new experimental setup to test it. It is, therefore, an important task to predict possible experimental signatures of the Schwinger effect in heavy-ion collisions. Naively, we can expect a thermal-like excess in electron/positron yields (and possibly muon/antimuon as well) in the very low-momentum regime $|\boldsymbol{p}| \lesssim \sqrt{eE}$, where $eE \approx \mathcal{O}((50\ \text{MeV})^2)$ for intermediate energies, as the celebrated Schwinger formula [57] says that the momentum spectrum of the produced pairs is $\propto \exp[-\pi(m^2+\boldsymbol{p}^2)/eE]$. It is worthwhile to quantify this excess, the zero-th order evaluation of which can be done with the Schwinger formula, or the locally-constant-field approximation [43, 59–63]. For a more realistic estimation, it would be necessary to include various modifications due to, e.g., spatial inhomogeneities, polarization of the field, and event-by-event fluctuations in actual ex-

periments. Meanwhile, such a low-momentum spectrum would naturally be contaminated by hadronic processes such as the Dalitz decay, $\pi^0 \to \gamma + e^- + e^+$, and therefore it poses an experimental challenge how to eliminate those backgrounds to test the Schwinger effect with heavy-ion collisions.

Second, from the standpoint of studying the densest matter and/or the QCD phase diagram with intermediate-energy heavy-ion collisions, the existence of the strong electric field is a noise that needs to be tamed, since electromagnetic observables such as charged flow and di-lepton yields are naturally affected. For example, it has been argued in Ref. [14] that di-lepton yields are considerably enhanced due to the existence of the QCD critical point in the low invariant-mass regime $\lesssim$ 100 MeV and that such an enhancement can be used as an experimental signature of the QCD critical point. Meanwhile, it is natural to expect additional contributions to such a low-energy di-lepton spectrum due to strong-field effects, including the Schwinger effect that we have mentioned in the last paragraph and other effects such as the nonlinear Breit-Wheeler process $\gamma + E \to e^+ + e^-$, but they have never been estimated. Considering the fact that so far no clear signals of the critical point, or novel phases of QCD, have been found in the Beam Energy Scan program at RHIC [1], it is reasonable to assume that the signals, if they exist, should be small. Therefore, possible contaminations must be removed beforehand as much as possible, which of course applies to the strong-field effects.

Third, our result suggests that intermediate-energy heavy-ion collisions may provide a unique opportunity to study QCD in a new extreme condition characterized by a strong electric field. Although it would be difficult to have something very nontrivial for the hadronic scale $m = \mathcal{O}(100$ - 1000 MeV), which is greater than the achievable field strength $eE = \mathcal{O}((30$ - 60 MeV$)^2)$, it is very reasonable to expect nonperturbative changes in the deconfined phase of QCD, where the typical mass scale m is the current quark mass $\mathcal{O}(1$ MeV). There are a number of studies of QCD in a strong magnetic field and it has been predicted that the QCD phase diagram is modified significantly (see, e.g., Ref. [64]). In contrast, only a few exist for a strong electric field [65, 66] and there has been no consensus on what would happen, meaning that it requires further theoretical study.

ACKNOWLEDGMENTS

The authors thank Kensuke Homma, Asanosuke Jinno, Masakiyo Kiatazawa, and Yasushi Nara for enlightening discussions. This work is supported by JSPS KAKENHI under grant No. 22K14045 (HT), the RIKEN special postdoctoral researcher program (HT), JST SPRING (grant No. JPMJSP2138) (TN), and Multidisciplinary PhD Program for Pioneering Quantum Beam Application (TN).

[1] A. Aparin (STAR Collaboration), STAR Experiment Results From Beam Energy Scan Program, Phys. Atom. Nucl. **86**, 758 (2023).
[2] https://fair-center.de/.
[3] https://nica.jinr.ru/.
[4] https://english.imp.cas.cn/research/facilities/HIAF/.
[5] https://asrc.jaea.go.jp/soshiki/gr/hadron/jparc-hi/index.html.
[6] K. Fukushima and T. Hatsuda, The phase diagram of dense QCD, Rept. Prog. Phys. **74**, 014001 (2011), arXiv:1005.4814 [hep-ph].
[7] H. Sorge, Flavor production in Pb (160-A/GeV) on Pb collisions: Effect of color ropes and hadronic rescattering, Phys. Rev. C **52**, 3291 (1995), arXiv:nucl-th/9509007.
[8] S. A. Bass *et al.*, Microscopic models for ultrarelativistic heavy ion collisions, Prog. Part. Nucl. Phys. **41**, 255 (1998), arXiv:nucl-th/9803035.
[9] M. Bleicher *et al.*, Relativistic hadron hadron collisions in the ultrarelativistic quantum molecular dynamics model, J. Phys. G **25**, 1859 (1999), arXiv:hep-ph/9909407.
[10] Y. Nara, N. Otuka, A. Ohnishi, K. Niita, and S. Chiba, Study of relativistic nuclear collisions at AGS energies from p + Be to Au + Au with hadronic cascade model, Phys. Rev. C **61**, 024901 (2000), arXiv:nucl-th/9904059.
[11] J. Weil *et al.* (SMASH), Particle production and equilibrium properties within a new hadron transport approach for heavy-ion collisions, Phys. Rev. C **94**, 054905 (2016), arXiv:1606.06642 [nucl-th].
[12] C. Sasaki, Signatures of chiral symmetry restoration in dilepton production, Phys. Lett. B **801**, 135172 (2020), arXiv:1906.05077 [hep-ph].
[13] O. Savchuk, A. Motornenko, J. Steinheimer, V. Vovchenko, M. Bleicher, M. Gorenstein, and T. Galatyuk, Enhanced dilepton emission from a phase transition in dense matter, J. Phys. G **50**, 125104 (2023), arXiv:2209.05267 [nucl-th].
[14] T. Nishimura, M. Kitazawa, and T. Kunihiro, Enhancement of dilepton production rate and electric conductivity around the QCD critical point, PTEP **2023**, 053D01 (2023), arXiv:2302.03191 [hep-ph].
[15] T. Nishimura, Y. Nara, and J. Steinheimer, Enhanced Dilepton production near the color superconducting phase and the QCD critical point, (2023), arXiv:2311.14135 [hep-ph].
[16] Y. Sun, Y. Wang, Q. Li, and F. Wang, Effect of internal magnetic field on collective flow in heavy ion collisions at intermediate energies, Phys. Rev. C **99**, 064607 (2019), arXiv:1905.12492 [nucl-th].
[17] G.-F. Wei, C. Liu, X.-W. Cao, Q.-J. Zhi, W.-J. Xiao, C.-Y. Long, and Z.-W. Long, Necessity of self-consistent calculations for the electromagnetic field in probing the nuclear symmetry energy using pion observables in heavy-ion collisions, Phys. Rev. C **103**, 054607 (2021), arXiv:2105.01866 [nucl-th].

[18] A. Di Piazza, C. Muller, K. Z. Hatsagortsyan, and C. H. Keitel, Extremely high-intensity laser interactions with fundamental quantum systems, Rev. Mod. Phys. **84**, 1177 (2012), arXiv:1111.3886 [hep-ph].

[19] A. Fedotov, A. Ilderton, F. Karbstein, B. King, D. Seipt, H. Taya, and G. Torgrimsson, Advances in QED with intense background fields, Phys. Rept. **1010**, 1 (2023), arXiv:2203.00019 [hep-ph].

[20] J. W. Yoon, Y. G. Kim, I. W. Choi, J. H. Sung, H. W. Lee, S. K. Lee, and C. H. Nam, Realization of laser intensity over 10^{23} w/cm^2, Optica **8**, 630 (2021).

[21] The white paper of ELI can be find at `https://www.eli-np.ro/whitebook.php`.

[22] E. Fermi, High Energy Nuclear Events, Progress of Theoretical Physics **5**, 570 (1950), https://academic.oup.com/ptp/article-pdf/5/4/570/5430247/5-4-570.pdf.

[23] L. D. Landau, On the multiparticle production in high-energy collisions, Izv. Akad. Nauk Ser. Fiz. **17**, 51 (1953).

[24] I. A. Maltsev, V. M. Shabaev, R. V. Popov, Y. S. Kozhedub, G. Plunien, X. Ma, T. Stöhlker, and D. A. Tumakov, How to observe the vacuum decay in low-energy heavy-ion collisions, Phys. Rev. Lett. **123**, 113401 (2019), arXiv:1903.08546 [physics.atom-ph].

[25] J. Rafelski, J. Kirsch, B. Müller, J. Reinhardt, and W. Greiner, Probing QED Vacuum with Heavy Ions, FIAS Interdisc. Sci. Ser. , 211 (2017), arXiv:1604.08690 [nucl-th].

[26] J. D. Bjorken, Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region, Phys. Rev. D **27**, 140 (1983).

[27] W.-T. Deng and X.-G. Huang, Event-by-event generation of electromagnetic fields in heavy-ion collisions, Phys. Rev. C **85**, 044907 (2012), arXiv:1201.5108 [nucl-th].

[28] K. Hattori and X.-G. Huang, Novel quantum phenomena induced by strong magnetic fields in heavy-ion collisions, Nucl. Sci. Tech. **28**, 26 (2017), arXiv:1609.00747 [nucl-th].

[29] M. Aaboud *et al.* (ATLAS), Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC, Nature Phys. **13**, 852 (2017), arXiv:1702.01625 [hep-ex].

[30] J. Adam *et al.* (STAR), Measurement of e^+e^- Momentum and Angular Distributions from Linearly Polarized Photon Collisions, Phys. Rev. Lett. **127**, 052302 (2021), arXiv:1910.12400 [nucl-ex].

[31] J. D. Brandenburg, J. Seger, Z. Xu, and W. Zha, Report on progress in physics: observation of the Breit–Wheeler process and vacuum birefringence in heavy-ion collisions, Rept. Prog. Phys. **86**, 083901 (2023), arXiv:2208.14943 [hep-ph].

[32] T. Maruyama, A. Bonasera, M. Papa, and S. Chiba, Formation and decay of superheavy systems, Eur. Phys. J. A **14**, 191 (2002), arXiv:nucl-th/0107021.

[33] V. I. Zagrebaev, Y. T. Oganessian, M. G. Itkis, and W. Greiner, Superheavy nuclei and quasi-atoms produced in collisions of transuranium ions, Phys. Rev. C **73**, 031602 (2006).

[34] J. Tian, X. Wu, K. Zhao, Y. Zhang, and Z. Li, Properties of the composite systems formed in the reactions of U-238 + U-238 and Th-232 + Cf-250, Phys. Rev. C **77**, 064603 (2008).

[35] V. S. Popov, Production of e^+e^- Pairs in an Alternating External Field, JETP Lett. **13**, 185 (1971).

[36] V. S. Popov, Pair production in a variable external field (quasiclassical approximation), Zh. Eksp. Teor. Fiz. **61**, 1334 (1971).

[37] E. Brezin and C. Itzykson, Pair Production in Vacuum by an Alternating Field, Phys. Rev. D **2**, 1191 (1970).

[38] G. V. Dunne and C. Schubert, Worldline instantons and pair production in inhomogeneous fields, Phys. Rev. D **72**, 105004 (2005), arXiv:hep-th/0507174.

[39] G. V. Dunne, Q.-h. Wang, H. Gies, and C. Schubert, Worldline instantons. II. The Fluctuation prefactor, Phys. Rev. D **73**, 065028 (2006), arXiv:hep-th/0602176.

[40] T. Oka, Nonlinear doublon production in a Mott insulator: Landau-Dykhne method applied to an integrable model, Phys. Rev. B **86**, 075148 (2012), arXiv:1105.3145 [cond-mat.str-el].

[41] H. Taya, H. Fujii, and K. Itakura, Finite pulse effects on e^+e^- pair creation from strong electric fields, Phys. Rev. D **90**, 014039 (2014), arXiv:1405.6182 [hep-ph].

[42] F. Gelis and N. Tanji, Schwinger mechanism revisited, Prog. Part. Nucl. Phys. **87**, 1 (2016), arXiv:1510.05451 [hep-ph].

[43] I. A. Aleksandrov, G. Plunien, and V. M. Shabaev, Locally-constant field approximation in studies of electron-positron pair production in strong external fields, Phys. Rev. D **99**, 016020 (2019), arXiv:1811.01419 [hep-ph].

[44] H. Taya, T. Fujimori, T. Misumi, M. Nitta, and N. Sakai, Exact WKB analysis of the vacuum pair production by time-dependent electric fields, JHEP **03**, 082, arXiv:2010.16080 [hep-th].

[45] L. V. Keldysh, Ionization in the Field of a Strong Electromagnetic Wave, J. Exp. Theor. Phys. **20**, 1307 (1965).

[46] X.-G. Deng, X.-G. Huang, Y.-G. Ma, and S. Zhang, Vorticity in low-energy heavy-ion collisions, Phys. Rev. C **101**, 064908 (2020), arXiv:2001.01371 [nucl-th].

[47] The latest version of JAM (JAM2) is publically available at `https://gitlab.com/transportmodel/jam2`.

[48] 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].

[49] A. Ohnishi, Physics of heavy-ion collisions at JHF, in *Workshop on Nuclear Physics at JHF* (2002) pp. 56–65.

[50] A. Ohnishi, Approaches to qcd phase diagram; effective models, strong-coupling lattice qcd, and compact stars, in *Journal of Physics: Conference Series*, Vol. 668 (IOP Publishing, 2016) p. 012004.

[51] A. Jinno, M. Kiatazawa, Y. Nara, T. Nishimura, and H. Taya, Spacetime volume of the dense baryon matter in intermediate-energy heavy-ion collisions, .

[52] Y. Hirono, M. Hongo, and T. Hirano, Estimation of electric conductivity of the quark gluon plasma via asymmetric heavy-ion collisions, Phys. Rev. C **90**, 021903 (2014), arXiv:1211.1114 [nucl-th].

[53] V. Voronyuk, V. D. Toneev, S. A. Voloshin, and W. Cassing, Charge-dependent directed flow in asymmetric nuclear collisions, Phys. Rev. C **90**, 064903 (2014), arXiv:1410.1402 [nucl-th].

[54] L. Adamczyk *et al.* (STAR), Charge-dependent directed flow in Cu+Au collisions at $\sqrt{s_{NN}} = 200$ GeV, Phys. Rev. Lett. **118**, 012301 (2017), arXiv:1608.04100 [nucl-ex].

[55] J. D. Jackson, *Classical Electrodynamics* (Wiley, 1998).

[56] F. Sauter, Uber das Verhalten eines Elektrons im homogenen elektrischen Feld nach der relativistischen Theorie

Diracs, Z. Phys. **69**, 742 (1931).
[57] J. S. Schwinger, On gauge invariance and vacuum polarization, Phys. Rev. **82**, 664 (1951).
[58] S. W. Hawking, Particle Creation by Black Holes, Commun. Math. Phys. **43**, 199 (1975), [Erratum: Commun.Math.Phys. 46, 206 (1976)].
[59] S. S. Bulanov, N. B. Narozhny, V. D. Mur, and V. S. Popov, On e+ e- pair production by a focused laser pulse in vacuum, Phys. Lett. A **330**, 1 (2004), arXiv:hep-ph/0403163.
[60] H. Gies and F. Karbstein, An Addendum to the Heisenberg-Euler effective action beyond one loop, JHEP **03**, arXiv:1612.07251 [hep-th].
[61] S. P. Gavrilov and D. M. Gitman, Vacuum instability in slowly varying electric fields, Phys. Rev. D **95**, 076013 (2017), arXiv:1612.06297 [hep-th].
[62] F. Karbstein, Heisenberg-Euler effective action in slowly varying electric field inhomogeneities of Lorentzian shape, Phys. Rev. D **95**, 076015 (2017), arXiv:1703.08017 [hep-ph].
[63] D. G. Sevostyanov, I. A. Aleksandrov, G. Plunien, and V. M. Shabaev, Total yield of electron-positron pairs produced from vacuum in strong electromagnetic fields: validity of the locally constant field approximation, (2020), arXiv:2012.10751 [hep-ph].
[64] J. O. Andersen, W. R. Naylor, and A. Tranberg, Phase diagram of QCD in a magnetic field: A review, Rev. Mod. Phys. **88**, 025001 (2016), arXiv:1411.7176 [hep-ph].
[65] A. Yamamoto, Lattice QCD with strong external electric fields, Phys. Rev. Lett. **110**, 112001 (2013), arXiv:1210.8250 [hep-lat].
[66] G. Endrodi and G. Marko, QCD phase diagram and equation of state in background electric fields, Phys. Rev. D **109**, 034506 (2024), arXiv:2309.07058 [hep-lat].

PROCEEDINGS OF SCIENCE

Theoretical Highlights of CP Violation in B Decays

Eleftheria Malami[a,b,*]

[a] *Center for Particle Physics Siegen (CPPS), Theoretische Physik 1, Universität Siegen, D-57068 Siegen, Germany*

[b] *Nikhef, Science Park 105, 1098 XG Amsterdam, Netherlands*

E-mail: Eleftheria.Malami@uni-siegen.de, emalami@nikhef.nl

In this presentation, we discuss recent key topics in theoretical analyses of CP violation in benchmarks decays of the B meson. We provide the most updated values of the mixing phases and discuss the importance of including the penguin contributions in their studies. Exploring intriguing patterns in purely tree decays, interesting new methodologies can be developed and applied. New data related to the CP asymmetries of key modes like $B_d^0 \to \pi^0 K_S$ and $B_s^0 \to K^+K^-$ lead to interesting results. The new $R_{K^{(*)}}$ measurement, compatible with the Standard Model, can still allow for electron-muon symmetry violation through new sources of CP violation.

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:2402.16976v1 [hep-ph] 26 Feb 2024

1. First Few Words

The concept of CP violation in the B–meson system is important in order to test the Standard Model (SM) as well as to search for hints of New Physics (NP). Central role in the studies of CP violation plays the Cabibbo–Kobayashi–Maskawa (CKM) matrix [1, 2], and an important topic linked to this matrix is the Unitarity triangle (UT). We provide recent highlights in the studies of CP violation and discuss benchmark modes.

2. Topic 1: Mixing Phases and Penguin Contributions

An important feature of neutral B mesons is the oscillation between B_q^0 and $\bar{B}_q^0$, which may lead to interference effects if both B_q^0 and $\bar{B}_q^0$ decay into the same final state. These interference effects give rise to CP-violating asymmetries. Associated to the B_q^0–$\bar{B}_q^0$ mixing phenomenon are the CP-violating mixing phases ϕ_d and ϕ_s for the B_d and B_s systems, respectively. Benchmark decays for determining these phases are the $B_d^0 \to J/\psi K_s^0$ and $B_s^0 \to J/\psi\phi$ modes. The theoretical precision is limited by doubly-Cabibbo suppressed penguin contributions, which are difficult to calculate. So, instead of calculating these penguin topologies, we use control channels to determine them, exploiting the SU(3) flavour symmetry of strong interactions. The corresponding analysis is given in Ref. [3] and allows us to extract the values of the mixing phases, taking the penguin contributions into account. The experimental input for the phases is:

$$\phi_s^{\rm eff} = (-4.1 \pm 1.3)^\circ, \quad \phi_d^{\rm eff} = (43.6 \pm 1.4)^\circ, \tag{1}$$

where $\phi_q^{\rm eff} = \phi_q + \Delta\phi_q$ with $\Delta\phi_q$ indicating the hadronic phase shift. The extracted penguin parameters $a_{(V)}$ and $\theta_{(V)}$ and mixing phases are the following [4]:

$$\text{vector-pseudoscalar states} \quad a = 0.14^{+0.17}_{-0.11}\,, \qquad \theta = \left(173^{+35}_{-45}\right)^\circ\,, \qquad \phi_d = \left(44.4^{+1.6}_{-1.5}\right)^\circ\,, \tag{2}$$

$$\text{vector-vector states} \quad a_V = 0.044^{+0.085}_{-0.038}\,, \quad \theta_V = \left(306^{+\ 48}_{-112}\right)^\circ\,, \quad \phi_s = (-4.2 \pm 1.4)^\circ\,. \tag{3}$$

We note that updated measured values for $\phi_s^{\rm eff}$ and $\phi_d^{\rm eff}$ have recently been provided by the LHCb Collaboration and the Belle II experiment in Refs. [5–7], respectively. In future analyses, achieving much higher precision, it is important that the penguin contributions are properly included.

Having provided the mixing phases, we move on to the UT and the determination of its apex. For this aim, special attention needs to be given to the determination of the CKM input parameters. As presented in Ref. [8], one way of determining the UT is through the angle γ and the side R_b. Concerning the angle γ, it is measured by the LHCb collaboration through $B \to DK$ modes. An alternative way of obtaining this angle is via the isospin analysis of $B \to \pi\pi, \rho\pi, \rho\rho$ decays, yielding the UT angle α. Utilising the ϕ_d phase, the value of α is converted into γ. The two results agree with each other, thus with the current precision, we can make an average of these γ values:

$$\gamma_{\rm avg} = (68.4 \pm 3.3)^\circ. \tag{4}$$

Regarding R_b, tensions arise between the inclusive and exclusive determinations of the $|V_{ub}|$ and $|V_{cb}|$ matrix elements. The essential point is to avoid making averages between these values but to

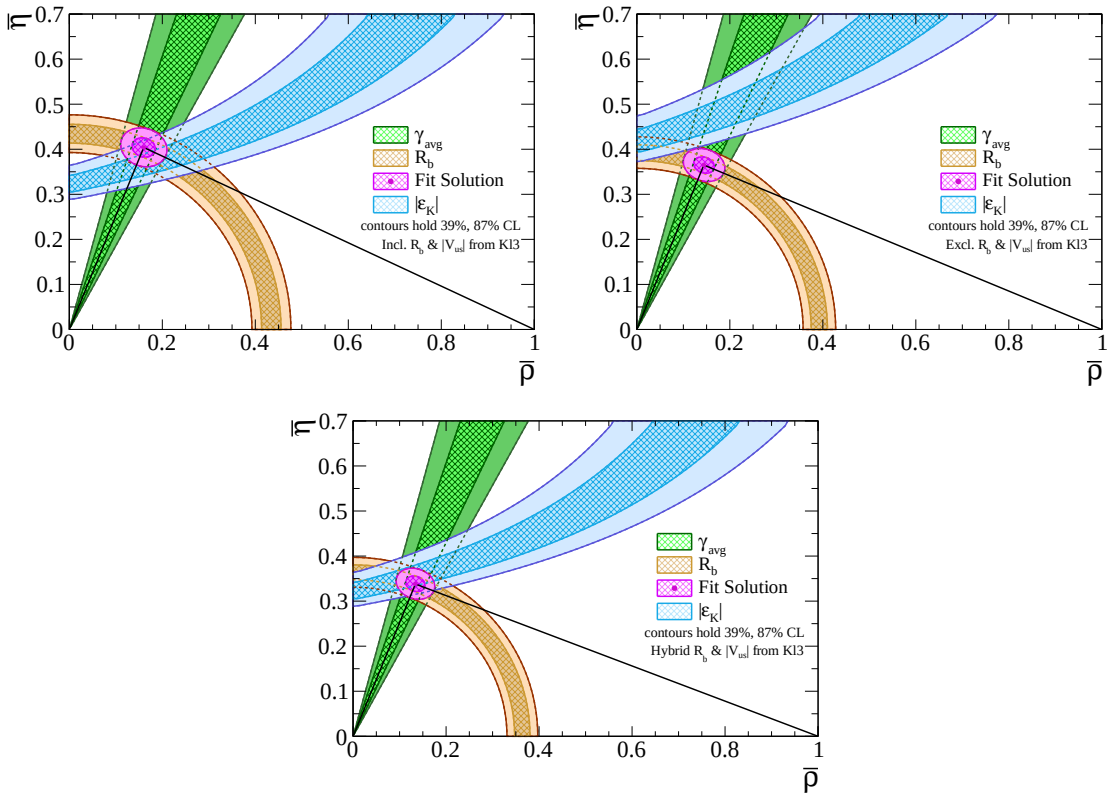


Figure 1: Determination of the UT apex. Top left: inclusive case, top right: exclusive case and bottom plot: hybrid case with excl. $|V_{ub}|$ and incl. $|V_{cb}|$ [3].

perform separate analysis for the two different approaches. On top of that, we can explore a hybrid option; the case of exclusive $|V_{ub}|$ and inclusive $|V_{cb}|$, as presented in [8]. An illustration of the UT for these three cases is given in Fig. 1. In the same figure, the ε_K hyperbola (blue contour), coming from indirect CP violation in the neutral kaon system, is also shown.

The studies of the UT are a key input for obtaining SM predictions of the B_q mixing parameters and eventually, exploring the corresponding space for NP left through the current data. As discussed in [8], the corresponding results have interesting applications in rare leptonic decays. In particular, we can minimise the impact of the CKM parameters when constraining NP in $B_s^0 \to \mu^+\mu^-$ by creating ratios of the branching fraction of this decay and the mass difference Δm_s [9, 10]:

$$R_{s\mu} = \bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)/\Delta m_s, \tag{5}$$

where the CKM elements drop out in the SM.

3. Topic 2: Puzzles in Tree Decays

The pure tree decays $\bar{B}_s^0 \to D_s^+ K^-$ and $B_s^0 \to D_s^+ K^-$ (and their CP conjugates) offer a powerful probe for testing the SM description of CP violation [11–13]. Intriguing puzzles arise in the angle γ of the UT and the individual branching ratios, which complement each other. A strategy that

allows us to study these anomalies and search for hints of NP is discussed in Refs. [14, 15]. Let us summarise the key points of this methodology.

As a first step, we explore CP violation. Due to B_q^0–$\bar{B}_q^0$ mixing, interference effects arise between the $\bar{B}_s^0 \to D_s^+ K^-$ and $B_s^0 \to D_s^+ K^-$ channels. These interference effects lead to a time-dependent CP asymmetry, which yields the observables C, S, $\mathcal{A}_{\Delta\Gamma}$ and their CP conjugates. A measure of the strength of the interference effects is given by the quantities ξ and $\bar{\xi}$. Therefore, one can use the observables C, S, $\mathcal{A}_{\Delta\Gamma}$ and $\bar{C}$, $\bar{S}$, $\bar{\mathcal{A}}_{\Delta\Gamma}$, in order to determine ξ and $\bar{\xi}$, respectively, from the experimental data in an unambiguous way. Within the SM, in the product $\xi \times \bar{\xi}$ hadronic matrix elements cancel out allowing a theoretically clean extraction to $(\phi_s + \gamma)$. In the presence of NP, the generalisation of this relation takes the following form:

$$\xi \times \bar{\xi} = \sqrt{1 - 2\left[\frac{C + \bar{C}}{(1 + C)\,(1 + \bar{C})}\right]}\, e^{-i[2(\phi_s + \gamma_{\rm eff})]}, \tag{6}$$

where again hadronic uncertainties cancel. In particular, here it is possible that $C + \bar{C}$ is not equal to 0, as in the SM. The above expression leads to a theoretically clean determination of the angle:

$$\gamma_{\rm eff} \equiv \gamma + \gamma_{\rm NP}, \tag{7}$$

where $\gamma_{\rm NP}$ is a function of the NP parameters ρ, φ, δ, and $\bar{\rho}$, $\bar{\varphi}$, $\bar{\delta}$ (for the CP conjugate case). Here, $\rho = \left[A(\bar{B}_s^0 \to D_s^+ K^-)_{\rm NP}/A(\bar{B}_s^0 \to D_s^+ K^-)_{\rm SM}\right]$ measures the strength of NP, while δ and φ denote the CP-conserving and CP-violating phases, and similarly for $\bar{\rho}$, $\bar{\varphi}$, $\bar{\delta}$. Using information on γ [16] from other processes, we extract $\gamma_{\rm NP}$.

The second step corresponds to information from the branching ratios. We create ratios by combining the branching fractions of the non-leptonic decays we study with differential branching ratios of their semi-leptonic partner channels. These ratios with the semileptonic decays minimize the dependence on the CKM matrix elements and the hadronic form factors. Therefore, they provide a useful setup which permits the extraction of the colour factors $|a_1|$ from the data in the theoretically cleanest possible manner. Comparing these experimental results with theoretical predictions, we find tensions even up to the 4.8 σ level. This intriguing pattern is in line with what we expect from the puzzling situation with γ. In order to interpret these $|a_1|$ deviations, we introduce the quantities:

$$\bar{b} \equiv \frac{\langle \mathcal{B}(\bar{B}_s^0 \to D_s^+ K^-)_{\rm th} \rangle}{\mathcal{B}(\bar{B}_s^0 \to D_s^+ K^-)_{\rm th}^{\rm SM}} = 1 + 2\,\bar{\rho}\cos\bar{\delta}\cos\bar{\varphi} + \bar{\rho}^2, \tag{8}$$

$$b \equiv \frac{\langle \mathcal{B}(B_s^0 \to D_s^- K^+)_{\rm th} \rangle}{\mathcal{B}(B_s^0 \to D_s^- K^+)_{\rm th}^{\rm SM}} = 1 + 2\,\rho\cos\delta\cos\varphi + \rho^2, \tag{9}$$

where now we use as input the theoretical expectation of $|a_1|$. The extracted values of b and $\bar{b}$ deviate from the SM. We highlight that making use of other control channels, we are able to constrain the contributions from exchange diagrams and no anomalous enhancement is observed due to these topologies.

Last but not least, we explore how much room there is for NP utilising all three $\gamma_{\rm eff}$, b and $\bar{b}$. More specifically, we obtain correlations between the NP parameters $\rho(\varphi)$ and $\bar{\rho}(\bar{\varphi})$, assuming that the strong phases equal to 0. Constraining these NP parameters, we find that it is possible to accommodate the current data with new contributions of moderate size.

We note that for the numerical analysis in Refs. [14, 15] the values presented by the LHCb Collaboration in Ref. [17] have been used. A new standalone measurement by LHCb using only Run II has recently been reported [18], which is interesting to explore further.

4. Topic 3: $B_d^0 \to \pi^0 K_S$ and $B_s \to K^+K^-$

Another powerful probe for CP violation studies is given by charmless two-body B decays, such as $B \to \pi K$, $B_{(s)} \to KK$ and $B \to \pi\pi$ [19–21]. Here, we will present updates related to two of these channels, $B_d^0 \to \pi^0 K_S$ and $B_s \to K^+K^-$, which are dominated by penguin topologies.

The $B_d^0 \to \pi^0 K_S$ decay is one of the most interesting $B \to \pi K$ channels, as this is the only one exhibiting mixing-induced CP violation. Therefore, it is important to measure CP violation with the highest precision in this system, and especially the mixing-induced CP violation. This mode is extensively studied in Refs. [22–25].

Following the analysis in Ref. [22], utilising an isospin relation and complementing it with a minimal SU(3) input, we obtain correlations between the CP asymmetries, given by the following expression [26]:

$$S_{\pi^0 K_s} = \sqrt{1 - A^2_{\pi^0 K_s}}\, \sin(\phi_d - \phi_{00}). \tag{10}$$

Here, $S_{\pi^0 K_s}$ is the mixing-induced and $A_{\pi^0 K_s}$ the direct CP asymmetry, ϕ_d is the $B_d^0 - \bar{B}_d^0$ mixing phase and ϕ_{00} denotes the angle between the decay amplitude $B_d^0 \to \pi^0 K^0$ and its CP-conjugate $\bar{B}_d^0 \to \pi^0 \bar{K}^0$. Interestingly, tensions arise with the SM picture. Thus, we need to explore how this puzzle can be resolved. Two are the options: either the data should change or NP physics contributions might enter the penguin sector.

An update on the time-dependent CP violation in $B_d^0 \to \pi^0 K_S$ was recently provided by Belle II. The new new results for the mixing induced and direct CP asymmetries are the following [27]:

$$A^{\text{Belle II}}_{\pi^0 K_s} = 0.04^{+0.15}_{-0.14} \pm 0.05, \quad S^{\text{Belle II}}_{\pi^0 K_s} = 0.75^{+0.20}_{-0.23} \pm 0.04. \tag{11}$$

These results can be compared with the current world average:

$$A^{\text{world average}}_{\pi^0 K_s} = -0.01 \pm 0.10, \quad S^{\text{world average}}_{\pi^0 K_s} = 0.57 \pm 0.17. \tag{12}$$

The new Belle II data become competitive with the world's most precise measurements. In comparison with the theoretical results for the CP asymmetries derived from Eq. 10, this new measurement has been shifted towards the theory predictions, showing a better agreement within the uncertainties. This is an interesting point as it can play a key role in resolving the longstanding $B \to \pi K$ puzzle.

The second interesting channel we discuss is the $B_s \to K^+K^-$ [28], where the first observation of CP violation in this decay was recently reported by the LHCb collaboration [29]. The new LHCb measurements reveal surprising differences between the direct CP asymmetries in the following modes [29]:

$$\mathcal{A}^{\text{dir}}_{\text{CP}}(B_s^0 \to K^-K^+) - \mathcal{A}^{\text{dir}}_{\text{CP}}(B_d^0 \to \pi^-K^+) = 0.089 \pm 0.031\,, \tag{13}$$

$$\mathcal{A}^{\text{dir}}_{\text{CP}}(B_d^0 \to \pi^-\pi^+) - \mathcal{A}^{\text{dir}}_{\text{CP}}(B_s^0 \to K^-\pi^+) = -0.095 \pm 0.040. \tag{14}$$

These decays differ only via the spectator quark thus, it is unlikely that this pattern indicates NP. An analysis performed in Ref. [30] shows that these differences can be accommodated in the SM through exchange and penguin-annihilation topologies, that are sizeable – at the level of 20%.

On top of that, the analysis in Ref. [30] provides an interesting way of extracting the angle γ of the UT. The proposed strategy relies only on CP asymmetries without requiring information on branching ratios. The result:

$$\gamma = (65^{+11}_{-7})^\circ, \tag{15}$$

agrees excellently with the γ values from the $B \to DK$ decays, which are pure tree transitions. As an a alternative, the B_s^0–$\bar{B}_s^0$ mixing phase ϕ_s can also be determined if now one uses the value of γ as an input. For this purpose, the methodology of using ratios of non-leptonic and semileptonic $B_{(s)}$ decay rates is utilisied, providing a clean way of obtaining ϕ_s.

5. Topic 4: CP Violation in Rare Decays. What about $R_{K^{(*)}}$ and the Electron-Muon Symmetry Violation?

The new results for $R_{K^{(*)}}$ presented by the LHCb collaboration in 2022 [31, 32]:

$$\langle R_K \rangle = 0.949 \pm 0.05, \quad \text{for momentum transfer } q^2 \in [1.1, 6.0] \text{ GeV}, \tag{16}$$

brought new perspectives for testing the electron–muon universality. These results agree with Lepton Flavour Univerasality (LFU). The differential rates for $B \to K\mu^+\mu^-$ are small compared to the SM predictions, deviating at the $3.5\,\sigma$ level, thereby still indicating possible NP through these decays. How much electron-muon universality violation is possibly left for this NP, now constrained by R_K? As shown in the analysis in Ref. [33] due to new CP-violating effects, there is still significant room for violation of the electron–muon universality.

More specifically, this analysis explores the CP-violating effects in the NP studies of rare decays making use of the muonic Wilson coefficients. The experimental branching ratio and the direct CP asymmetry of the $B^- \to K^-\mu^+\mu^-$ mode constrains the corresponding muonic Wilson coefficients $C_{i\mu}$. Combining this constrain with the new $\langle R_K \rangle$ measurement allows the determination of the Wilson coefficient C_{ie} in the electronic sector. Having determined C_{ie}, the electronic direct and mixing-induced CP asymmetry are also obtained. The following conclusions are drawn:

i) NP Wilson coefficients entering the electronic modes can strongly differ from the corresponding ones entering the muonic channels and

ii) CP violating phases can be significantly different, therefore also the CP asymmetries between the electronic and muonic modes, which are the observables that they experimentalists should utilize in order to test the violation of LFU.

Therefore, it is still possible to have significant electron–muon universality violation, if NP effects are associated with new sources of CP violation, which are encoded in Wilson coefficients. In the era of high-precision B physics, it is important to perform experimental searches focusing on differences in CP asymmetries between the $b \to se^+e^-$ and $b \to s\mu^+\mu^-$ transitions. These studies will be essential for further testing LFU.

6. Epilogue

A lot of progress has been achieved over the last years in the studies of CP violation, which was possible through the synergy between theorists and experimentalists. There are exciting new perspectives to further explore CP violation. Moving towards the high precision era of B physics and monitoring the evolution of the data will lead to a much sharper picture.

Acknowledgements

I would like to thank the organisers of the HQL2023 for the invitation and the opportunity to attend such a great conference.

References

[1] N. Cabibbo, Phys. Rev. Lett. **10** (1963), 531-533 doi:10.1103/PhysRevLett.10.531

[2] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973), 652-657 doi:10.1143/PTP.49.652

[3] M. Z. Barel, K. De Bruyn, R. Fleischer and E. Malami, J. Phys. G **48** (2021) no.6, 065002

[4] M. Z. Barel, K. De Bruyn, R. Fleischer and E. Malami, PoS **CKM2021** (2023), 111 doi:10.22323/1.411.0111 [arXiv:2203.14652 [hep-ph]].

[5] R. Aaij *et al.* [LHCb], [arXiv:2308.01468 [hep-ex]].

[6] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **132** (2024) no.2, 021801

[7] I. Adachi *et al.* [Belle-II], [arXiv:2302.12898 [hep-ex]].

[8] K. De Bruyn, R. Fleischer, E. Malami and P. van Vliet, J. Phys. G **50** (2023) no.4, 045003

[9] A. J. Buras, Phys. Lett. B **566** (2003), 115-119

[10] A. J. Buras and E. Venturini, Acta Phys. Polon. B **53** no.6, 6-A1

[11] R. Aleksan, I. Dunietz and B. Kayser, Z. Phys. C **54** (1992), 653-660

[12] R. Fleischer, Nucl. Phys. B **671** (2003), 459-482

[13] K. De Bruyn, R. Fleischer, R. Knegjens, M. Merk, M. Schiller and N. Tuning, Nucl. Phys. B **868** (2013), 351-367

[14] R. Fleischer and E. Malami, Eur. Phys. J. C **83** (2023) no.5, 420

[15] R. Fleischer and E. Malami, Phys. Rev. D **106** (2022) no.5, 056004

[16] Y. S. Amhis *et al.* [HFLAV], Eur. Phys. J. C **81**, no.3, 226 (2021)

[17] R. Aaij *et al.* [LHCb Collaboration], JHEP **1803**, 059 (2018)

[18] [LHCb], LHCb-CONF-2023-004.

[19] M. Neubert and J. L. Rosner, Phys. Lett. B **441** (1998), 403-409

[20] A. J. Buras and R. Fleischer, Eur. Phys. J. C **11** (1999), 93-109

[21] T. Huber and G. Tetlalmatzi-Xolocotzi, Eur. Phys. J. C **82** (2022) no.3, 210

[22] R. Fleischer, R. Jaarsma, E. Malami and K. K. Vos, Eur. Phys. J. C **78** (2018) no.11, 943

[23] R. Fleischer, R. Jaarsma and K. K. Vos, Phys. Lett. B **785** (2018), 525-529

[24] R. Fleischer, S. Jager, D. Pirjol and J. Zupan, Phys. Rev. D **78** (2008), 111501

[25] A. J. Buras, R. Fleischer, S. Recksiegel and F. Schwab, Nucl. Phys. B **697** (2004), 133-206

[26] S. Faller, R. Fleischer and T. Mannel, Phys. Rev. D **79** (2009), 014005

[27] M. Veronesi [BELLE II], [arXiv:2305.09153 [hep-ex]].

[28] R. Fleischer, Phys. Lett. B **459** (1999), 306-320

[29] R. Aaij *et al.* [LHCb], JHEP **03** (2021), 075

[30] R. Fleischer, R. Jaarsma and K. K. Vos, JHEP **02** (2023), 081

[31] R. Aaij *et al.* [LHCb], Phys. Rev. Lett. **131** (2023) no.5, 051803

[32] R. Aaij *et al.* [LHCb], Phys. Rev. D **108** (2023) no.3, 032002

[33] R. Fleischer, E. Malami, A. Rehult and K. K. Vos, JHEP **06** (2023), 033

arXiv:2402.16963v1 [hep-ph] 26 Feb 2024

The symmetry approach to quark and lepton masses and mixing

Gui-Jun Ding[a] *, José W.F. Valle[b] †

[a] *Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*

[b] *AHEP Group, Instituto de Física Corpuscular - CSIC/Universitat de València, Parque Científico C/Catedratico José Beltrán, 2, E-46980 Paterna (València) - SPAIN*

Abstract

The Standard Model lacks an organizing principle to describe quark and lepton "flavours". We review the impact of neutrino oscillation experiments, which show that leptons mix very differently from quarks, placing a major challenge, but also providing a key input to the flavour puzzle. We briefly sketch the seesaw and "scotogenic" approaches to neutrino mass, the latter including also WIMP dark matter. We discuss the limitations of popular neutrino mixing patterns and examine the possibility that they arise from symmetry, giving a bottom-up approach to residual flavour and CP symmetries. We show how family and/or CP symmetries can generate novel viable and predictive mixing patterns. We review the model-independent ways to predict lepton mixing and test both mixing predictions as well as mass sum rules. We also discuss UV-complete flavour theories in four and more space-time dimensions, and their predictions. Benchmarks given include an A_4 scotogenic construction with trimaximal mixing pattern TM2. Higher-dimensional completions are also reviewed, such as 5-D warped flavordynamics. We present a T' warped flavordynamics theory with TM1 mixing pattern, detectable neutrinoless double beta decay rates and providing a very good fit of flavour observables, including quarks. We also review how 6-D orbifolds offer a way to determine the structure of the 4-D family symmetry from the symmetries between the extra-D branes. We describe a scotogenic A_4 orbifold predicting the "golden" quark-lepton mass relation, large neutrino mass with normal ordering, higher atmospheric octant, restricted reactor angle, and an excellent global flavour fit, including quark observables. Finally, we discuss promising recent progress in tackling the flavor issue through the use of modular symmetries.

Keywords: Fermion mixing, CP violation, generalized CP, flavor and modular symmetry, orbifolds, warped-flavordynamics.

*E-mail: dinggj@ustc.edu.cn

†E-mail: valle@ific.uv.es

Contents

1 Introduction **4**

1.1 Quark masses and mixing . . . 6
1.2 Lepton masses and mixing . . . 7
1.3 Neutrino oscillation recap . . . 9
1.4 Neutrinoless double beta decay . . . 12

2 Origin of neutrino masses and flavour puzzle **16**

2.1 Effective neutrino masses . . . 16
2.2 The seesaw paradigm . . . 16
2.3 Dark matter as the source of neutrino mass . . . 18
2.4 Missing partner seesaw and dark matter: the scoto-seesaw . . . 18
2.5 The low-scale inverse and linear seesaw mechanisms . . . 19
2.6 Dark inverse and linear seesaw mechanisms . . . 20
2.7 Neutrinos and the flavour puzzle . . . 21

3 Lepton mixing patterns **24**

3.1 Tri-bimaximal mixing (TBM) . . . 24
3.2 Generalizations of tri-bimaximal mixing . . . 25
3.3 Golden ratio mixing pattern . . . 26
3.4 Bi-maximal mixing pattern . . . 27

4 Flavour and CP symmetries from the bottom-up **29**

4.1 Residual symmetries of leptons . . . 29
4.2 Reconstructing lepton mixing from remnant CP symmetry . . . 33
4.3 Residual symmetries of quarks . . . 36

5 Viable lepton mixing patterns **39**

5.1 Revamped TBM mixing . . . 39
5.1.1 Case a: G_1 flavour and X_1, X_4 CP symmetries . . . 41
5.1.2 Case b: G_2 flavour and X_2, X_4 CP symmetries . . . 43
5.2 Revamped Golden-Ratio mixing scheme . . . 44
5.3 Bi-large mixing . . . 46
5.3.1 Bi-large mixing from abelian family symmetry . . . 47
5.3.2 Confronting bi-large mixing with oscillation data . . . 51

6 Lepton mixing from flavour and CP symmetry **54**

6.1 Lepton mixing from flavour symmetry alone . . . 54
6.1.1 Fully preserved residual symmetry $G_\nu = K_4$ . . . 56
6.1.2 Partially preserved residual symmetry $G_\nu = Z_2$ . . . 57

6.2 Combining flavour and CP symmetry . . . 59
6.2.1 Mathematical consistency . . . 60
6.2.2 Implications of residual flavour and CP symmetries . . . 63
6.3 Predictive scenarios with CP symmetry . . . 64
6.3.1 $\mathcal{G}_l = Z_n$ $(n \geq 3)$, $\mathcal{G}_\nu = Z_2 \times CP$ . . . 65
6.3.2 $\mathcal{G}_l = Z_2 \times CP$, $\mathcal{G}_\nu = Z_2 \times CP'$ . . . 72
6.4 Quark and lepton mixing from a common flavour group . . . 75
6.5 Geometrical CP violation . . . 79

7 Testing flavor and CP symmetries **81**
7.1 Testing mixing predictions . . . 81
7.2 Testing mass sum rules . . . 84
7.3 Flavor symmetry toolkit . . . 88

8 Benchmark UV-complete models in 4-D **89**
8.1 A_4 family symmetry in a scotogenic model . . . 89
8.1.1 Charged lepton masses . . . 90
8.1.2 Dark fermion masses . . . 91
8.1.3 Scotogenic neutrino masses . . . 92
8.2 A benchmark model with both flavour and CP symmetries . . . 95
8.2.1 Flavon superpotential . . . 96
8.2.2 The charged lepton sector . . . 97
8.2.3 The neutrino sector . . . 98

9 Family symmetry in 5-D models with a warped extra dimension **102**
9.1 Warped flavourdynamics with the T' family group . . . 102
9.1.1 Lepton masses and mixing . . . 103
9.1.2 Quark masses and CKM matrix . . . 107
9.2 Global flavour fit . . . 108

10 Family symmetry from 6-D orbifolds **111**
10.1 General preliminaries . . . 111
10.2 Scotogenic orbifold . . . 113
10.3 Symmetry breaking and fermion masses . . . 115
10.3.1 Quark and lepton masses . . . 116
10.3.2 Scotogenic neutrino masses . . . 117
10.4 Global fit of flavour observables . . . 118
10.4.1 Preliminaries . . . 118
10.4.2 Fit procedure . . . 119
10.5 Flavour predictions of the scotogenic orbifold model . . . 119

10.5.1 Golden quark-lepton mass relation . . . 119
10.5.2 Neutrino oscillation predictions . . . 120
10.5.3 Neutrinoless double beta decay predictions . . . 121

11 Recent progress: modular symmetry **123**
11.1 The modular group . . . 123
11.2 Modular invariance . . . 125
11.3 Generalized CP in modular symmetry . . . 126
11.4 Modular symmetry from top-down . . . 127
11.5 Quark-lepton mass relations from modular symmetry . . . 127
11.6 Modular versus traditional flavor symmetry . . . 128

12 Summary and outlook **129**

A The A_4 group **133**

B The S_4 group **135**

C The dihedral group **138**

D Diagonalization of a 2×2 complex symmetric matrix **141**

1 Introduction

Elucidating the spontaneous mechanism of symmetry breaking [1] within the Standard Model (SM) [2–5], i.e. the existence of a physical Higgs boson [6–8], has so far been the main accomplished goal of the successful LHC programme. This was achieved, at least partially, with the discovery of a scalar particle with properties closely resembling those of the SM Higgs, by the ATLAS [9] and CMS [10] experiments at CERN. More is expected from future studies, for example, at the upcoming FCC facility [11, 12] and other complementary lepton accelerators [13–17].

Another major milestone in elementary particle physics has been the discovery of neutrino oscillations in solar and atmospheric neutrino experiments [18, 19]. These were followed by reactor- and accelerator-based studies, e.g. [20–23] which, altogether, imply nonzero neutrino masses, as well as large mixing angles in the lepton sector [24, 25] [1]. This comes as a surprise, when compared with the pattern seen in the Cabibbo-Kobayashi-Maskawa matrix describing quark mixing and CP violation [28]. It also means that, even after the discovery of the Higgs boson, the arquitecture of particle physics is still quite far from "complete".

The origin of neutrino masses is one of the deepest secrets of modern particle physics [29]. The most general neutrino mass generation template is given by the $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge theory framework characterizing the SM [30–32]. In this review we will be mainly concerned with bottom-up

[1] There is a fairly good agreement with the other determinations, by the Bari group [26] and NuFit [27].

approaches, for which this is the most appropriate choice. However the seesaw can also been discussed in terms of left-right theories and SO(10) [33–38].

Following the gauge principle that underlies the SM construction, neutrinos should also get mass through spontaneous symmetry breaking (SSB). Small neutrino masses would be understood dynamically through new vacuum expectation values (VEVs). A specially interesting case is that of spontaneous violation of lepton number [32, 39]. Besides the high-scale seesaw, we stress that one can also have dynamical generation of neutrino masses at low scales [40–42].

Within the SM picture three of the fundamental interactions of nature (electromagnetic, weak and strong) all have a gauge description. It would be very appealing if these could also have a common origin at very high energies [34, 43–47], though no hard evidence for this beautiful idea has ever been found. Despite their potential in providing an all-encompassing unified description including also gravity [48–50], superstring theories have so-far also failed to provide a phenomenologically convincing roadmap.

Another major drawback of the SM construction is that it fails to explain family replication, fermion mass hierarchies and mixing pattern. The discovery of oscillations has only exacerbated this fact. The disparity observed in the pattern of quark and lepton mixing parameters appears to us unlikely to be the result of pure chance. As a result, we will not examine the possibility of neutrino mixing anarchy [51,52]. Although viable, we find this hypothesis theoretically unsatisfactory. Instead, our main common thread in this review will be the symmetry approach to the "flavour problem".

Here we will be mainly concerned with explaining the detailed pattern of the weak interactions of quarks and leptons within family-symmetry extended theories based on the $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge group, with special emphasis on the case of leptons. Early attempts to understand the lepton mixing pattern starting from the quark sector have now become obsolete, since the discovery of neutrino oscillations. A fully successful flavour theory should explain not only the observed large mixing angles in the lepton sector but also the CKM mixing pattern governing quark mixing and CP violation. Likewise, it should also account for the pattern of quark and lepton masses. In this review we will illustrate how this may, at least partially, be achieved either within 4-dimensional renormalizable gauge field theories [53–69] or in the context of theories with extra spacetime dimensions [70–84]. To set up notation for the following chapters here we start with some preliminaries on the gauge-theoretic description of quark and lepton mixing, followed by a very brief critique of the SM drawbacks.

We should stress that there are already several excellent reviews on the discrete flavor symmetry approach to address the SM flavor puzzle [85–93]. Apart from providing an update to these, the present review will focus on how the residual flavor and CP symmetries can constrain the fermion mixing angles and CP violation phases independently of the details of a specific implementation. We discuss in detail both theory and phenomenological predictions. In addition, we discuss several other topics and predictive benchmark flavor model examples, both in four dimensions and extra dimensions. The latter include warped 5D flavordynamics as well as orbifold-based scenarios, which are reviewed here for the first time. We discuss extensively the predictions of a broad class of symmetry-based theories of flavor. Finally, we also give a brief discussion of theories based on modular symmetries and how these may help with the

vacuum alignment problem.

1.1 Quark masses and mixing

We recall that in a gauge field theory like the Standard Model all the fermion masses arise from spontaneous symmetry breaking. The most general quark Yukawa interactions with the Higgs doublet H allowed by symmetry are given by [94–96],

$$\mathcal{L}^q_{\rm Yuk} = -(y_D)_{ij}\overline{D^i_R}H^\dagger Q^j_L - (y_U)_{ij}\overline{U^i_R}\widetilde{H}^\dagger Q^j_L + \text{h.c.}\,, \tag{1.1}$$

where $\widetilde{H} = i\tau_2 H^*$ and $Q^i_L = \left(U^i_L, D^i_L\right)^T$. The Yukawa matrices y_D and y_U are arbitrary complex matrices in flavour space. Upon electroweak symmetry breaking, the Higgs vacuum expectation value (VEV) $\langle H\rangle = (0, v/\sqrt{2})^T$ with $v \simeq 246$GeV gives us the quark mass terms

$$\mathcal{L}^q_{\rm mass} = -\overline{U}_R m_U U_L - \overline{D}_R m_D D_L + \text{h.c.}\,, \tag{1.2}$$

where $U_L \equiv (u_L, c_L, t_L)^T$, $D_L \equiv (d_L, s_L, b_L)^T$, $U_R \equiv (u_R, c_R, t_R)^T$ and $D_R \equiv (d_R, s_R, b_R)^T$ denote the three generations of left and right-handed up- and down-type quark fields, respectively. The mass matrices are determined by the Yukawa couplings and the Higgs VEV as follows

$$m_U = \frac{v}{\sqrt{2}}y_U, \quad m_D = \frac{v}{\sqrt{2}}y_D\,. \tag{1.3}$$

Gauge invariance does not constrain the flavour structure of the Yukawa couplings y_D and y_U and therefore m_U and m_D are arbitrary complex matrices. These matrices can be brought into diagonal form by separate unitary transformations on the left and right fermions, i.e.

$$W_u^\dagger m_U V_u = \text{diag}(m_u, m_c, m_t), \quad W_d^\dagger m_D V_d = \text{diag}(m_d, m_s, m_b)\,, \tag{1.4}$$

which implies

$$V_u^\dagger m_U^\dagger m_U V_u = \text{diag}\left(m_u^2, m_c^2, m_t^2\right)\,, \tag{1.5}$$

$$V_d^\dagger m_D^\dagger m_D V_d = \text{diag}\left(m_d^2, m_s^2, m_b^2\right)\,. \tag{1.6}$$

By performing such unitary transformations $U_L \to V_u U_L$, $D_L \to V_d D_L$, $U_R \to W_u U_R$ and $D_R \to W_d D_R$ one can go to the physical mass eigenstates. The resulting flavour-changing quark charged current weak interaction reads

$$\mathcal{L}^q_{CC} = \frac{g}{\sqrt{2}}\overline{U}_L\gamma^\mu V_{CKM} D_L W^+_\mu + \text{h.c.}\,, \tag{1.7}$$

where $V_{CKM} \equiv V_u^\dagger V_d$ is the Cabibbo-Kobayshi-Maskawa (CKM) mixing matrix [97, 98]. In what follows we adopt the standard form for the CKM matrix describing quark mixing, i.e.

$$V_{CKM} = \begin{pmatrix} c_{12}^q c_{13}^q & s_{12}^q c_{13}^q & s_{13}^q e^{-i\delta^q} \\ -s_{12}^q c_{23}^q - c_{12}^q s_{13}^q s_{23}^q e^{i\delta^q} & c_{12}^q c_{23}^q - s_{12}^q s_{13}^q s_{23}^q e^{i\delta^q} & c_{13}^q s_{23}^q \\ s_{12}^q s_{23}^q - c_{12}^q s_{13}^q c_{23}^q e^{i\delta^q} & -c_{12}^q s_{23}^q - s_{12}^q s_{13}^q c_{23}^q e^{i\delta^q} & c_{13}^q c_{23}^q \end{pmatrix} , \tag{1.8}$$

with $c_{ij}^q \equiv \cos\theta_{ij}^q$, $s_{ij}^q \equiv \sin\theta_{ij}^q$. This parameterization is the one adopted by the Review of Particle Physics of the Particle Data Group (PDG) [28], and supplements the original proposal in Ref. [30] by specifying a convenient factor ordering. Notice that only one physical phase remains after the allowed quark phase redefinitions. The CKM-phase-parameter δ_q is expressed in a neat and rephasing-invariant way as [99],

$$\delta^q = \phi_{13}^q - \phi_{12}^q - \phi_{23}^q \ , \tag{1.9}$$

in terms of the fundamental ϕ_{ij}^q phases. These are the relative phases between up- and down-type diagonalization matrices. From Ref. [28] we extract the following allowed ranges [100, 101] for the CKM parameters,

$$V_{CKM} = \begin{pmatrix} 0.97431 \pm 0.00012 & 0.22514 \pm 0.00055 & (0.00365 \pm 0.00010) e^{i(-66.8 \pm 2.0)^\circ} \\ (-0.22500 \pm 0.00054) e^{i(0.0351 \pm 0.0010)^\circ} & (0.97344 \pm 0.00012) e^{i(-0.001880 \pm 0.000052)^\circ} & 0.04241 \pm 0.00065 \\ (0.00869 \pm 0.00014) e^{i(-22.23 \pm 0.63)^\circ} & (-0.04124 \pm 0.00056) e^{i(1.056 \pm 0.032)^\circ} & 0.999112 \pm 0.000024 \end{pmatrix} . \tag{1.10}$$

1.2 Lepton masses and mixing

It is a characteristic feature of the Standard Model that spontaneous gauge symmetry breaking leaves neutrinos as massless fermions. However, the discovery of neutrino oscillations [18, 19] implies non-zero neutrino masses and neutrino mixing. One can introduce three right-handed neutrino fields ν_R^i for $i = 1, 2, 3$ as full $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ singlets. In analogy with the quark sector the neutrino and charged lepton masses are described by the Yukawa interactions,

$$\mathcal{L}_{\mathrm{Yuk}}^l = -(y_l)_{ij} \overline{l_R^i} H^\dagger L_L^j - (y_\nu)_{ij} \overline{\nu_R^i} \widetilde{H}^\dagger L_L^j + \mathrm{h.c.} \, , \tag{1.11}$$

where $L_L^i = \left(\nu_L^i, l_L^i\right)^T$ are the lepton doublet fields. In this case neutrinos would be Dirac particles. This requires the *ad hoc* imposition of lepton number symmetry to forbid the right-handed Majorana mass term $M_{ij} \overline{\nu_R^i} \, (\nu_R^j)^c$ allowed by SM gauge invariance. Moreover, the Yukawa coulping y_ν should be of order 10^{-11} in order to accommodate the neutrino masses below the eV scale.

Under the assumption of lepton number conservation one gets, after electroweak symmetry breaking, the following charged lepton and neutrino mass terms,

$$\mathcal{L}_{\mathrm{mass}}^l = -\bar{l}_R m_l l_L - \overline{\nu}_R m_\nu \nu_L + \mathrm{h.c.} \, . \tag{1.12}$$

However, without *a priori* assumptions, gauge theories suggest that neutrinos are Majorana particles [30] [2]. However, this important issue of the nature of neutrinos must be settled by experiment. A prime example is to search for neutrinoless double beta decay [106–109]. A positive detection would imply, by the black-box theorem, the Majorana nature of at least one of the neutrinos [110].

Majorana neutrino masses can be effectively described by the non-renormalizable Weinberg operator $(y_\nu)_{ij}\left(\overline{(L_L^i)^c}\, i\tau_2 H\right)\left(H^T i\tau_2 L_L^j\right)/(2\Lambda)$ [29]. As a result the charged lepton and neutrino mass terms take of the following form

$$\mathcal{L}^l_{\rm mass} = -\bar{l}_R m_l l_L - \frac{1}{2}\overline{\nu_L^c} m_\nu \nu_L + \text{h.c.}\,, \tag{1.13}$$

where $m_\nu = y_\nu v^2/(2\Lambda)$. The smallness of neutrino mass may be ascribed to the large new physics scale Λ. There are also attractive low-scale realizations of the seesaw mechanism, as we comment below. Both charged lepton and neutrino mass matrices are diagonalized through unitary transformations as follows

$$U_l^\dagger m_l^\dagger m_l U_l = \text{diag}\left(m_e^2, m_\mu^2, m_\tau^2\right)\,, \tag{1.14}$$

and

$$\begin{aligned} &U_\nu^T m_\nu U_\nu = \text{diag}\left(m_1, m_2, m_3\right), && \text{for Majorana neutrinos}\,, \\ &U_\nu^\dagger m_\nu^\dagger m_\nu U_\nu = \text{diag}\left(m_1^2, m_2^2, m_3^2\right), && \text{for Dirac neutrinos}\,. \end{aligned} \tag{1.15}$$

where the light neutrino masses $m_{1,2,3}$ are real and non-negative. The nonzero mass squared differences measured in neutrino oscillation experiments imply a non-degenerate mass spectrum $m_1 \neq m_2 \neq m_3$. As a consequence, we can express the lepton mass matrices in terms of U_l, U_ν and the mass eigenvalues as

$$m_l^\dagger m_l = U_l\, \text{diag}\left(m_e^2, m_\mu^2, m_\tau^2\right) U_l^\dagger\,, \tag{1.16}$$

$$m_\nu = U_\nu^* \text{diag}\left(m_1, m_2, m_3\right) U_\nu^\dagger, \qquad \text{for Majorana neutrinos}\,, \tag{1.17}$$

$$m_\nu^\dagger m_\nu = U_\nu \text{diag}\left(m_1^2, m_2^2, m_3^2\right) U_\nu^\dagger, \qquad \text{for Dirac neutrinos}\,. \tag{1.18}$$

Transforming to the charged lepton and neutrino mass eigenstates, one obtains the leptonic charged-current weak interaction as [30]

$$\mathcal{L}^l_{CC} = \frac{g}{\sqrt{2}} \bar{l}_L \gamma^\mu U_l^\dagger U_\nu \nu_L W_\mu^- + h.c.\,, \tag{1.19}$$

where the combination $U_l^\dagger U_\nu$ defines the lepton mixing matrix, i.e.

$$U = U_l^\dagger U_\nu\,. \tag{1.20}$$

[2] Note that the most general description of neutrinos is in terms of two-component spinors, Dirac fermions being just a particular case of Majorana [102]. Although the two-component approach is universal and often more insightful, in this review we will adopt the more familiar four-component formalism used in most textbooks, for example [103–105], in which Majorana and Dirac neutrinos appear as separate "cases".

Like the CKM matrix describing quark mixing, the lepton mixing matrix U arises from the mismatch between the diagonalizations of charged leptons and neutrinos. Assuming unitarity, the matrix U is characterized by three angles and three physical CP phases [30],

$$U = \begin{pmatrix} c_{12}^{\ell} c_{13}^{\ell} & s_{12}^{\ell} c_{13}^{\ell} e^{-i\phi_{12}} & s_{13}^{\ell} e^{-i\phi_{13}} \\ -s_{12}^{\ell} c_{23}^{\ell} e^{i\phi_{12}} - c_{12}^{\ell} s_{13}^{\ell} s_{23}^{\ell} e^{-i(\phi_{23}-\phi_{13})} & c_{12}^{\ell} c_{23}^{\ell} - s_{12}^{\ell} s_{13}^{\ell} s_{23}^{\ell} e^{-i(\phi_{23}+\phi_{12}-\phi_{13})} & c_{13}^{\ell} s_{23}^{\ell} e^{-i\phi_{23}} \\ s_{12}^{\ell} s_{23}^{\ell} e^{i(\phi_{23}+\phi_{12})} - c_{12}^{\ell} s_{13}^{\ell} c_{23}^{\ell} e^{i\phi_{13}} & -c_{12}^{\ell} s_{23}^{\ell} e^{i\phi_{23}} - s_{12}^{\ell} s_{13}^{\ell} c_{23}^{\ell} e^{-i(\phi_{12}-\phi_{13})} & c_{13}^{\ell} c_{23}^{\ell} \end{pmatrix} , \tag{1.21}$$

with the abbreviation $c_{ij}^{\ell} \equiv \cos\theta_{ij}^{\ell}$ and $s_{ij}^{l} \equiv \sin\theta_{ij}^{l}$. The above universal symmetrical presentation for the matrix U in Eq. (1.21) is very convenient both for the description of quark as well as lepton mixing [30]. Notice that there is an important difference between leptons and quarks concerning CP violation, namely, all three CP phases in Eq. (1.21) are physical parameters. One of them is the lepton analoque of the CKM phase in Eq. (1.9), also called *Dirac phase*.

$$\delta^{\ell} = \phi_{13} - \phi_{12} - \phi_{23} \ . \tag{1.22}$$

The others are the so-called *Majorana phases*, which can not be eliminated by field redefinitions [30] if neutrinos are Majorana particles. Only the Dirac phase affects conventional neutrino oscillation probabilities, while the Majorana phases can only affect lepton number violating processes [111, 112]. Note that in what follows we will also use the notation δ_{CP} for the lepton Dirac CP phase δ^{ℓ}. It will be an experimental challenge to obtain robust information on their magnitudes.

Current neutrino oscillation data restrict the elements of the lepton mixing matrix. One finds, at 3σ, the following ranges [24, 25]

$$U^{\rm NO} = \begin{pmatrix} 0.7838 \to 0.8442 & 0.5133 \to 0.6004 & (-0.1568 \to 0.1489) + i(-0.1182 \to 0.1520) \\ (-0.4831 \to -0.2394) + i(-0.0749 \to 0.0963) & (0.4635 \to 0.6749) + i(-0.0521 \to 0.0668) & 0.6499 \to 0.7719 \\ (0.3068 \to 0.5391) + i(-0.0643 \to 0.0933) & (-0.6897 \to -0.4821) + i(-0.0446 \to 0.0644) & 0.6161 \to 0.7434 \end{pmatrix} \tag{1.23}$$

for the case of normal neutrino mass-ordering. On the other hand, for inverted-ordered neutrino masses one has [24, 25],

$$U^{\rm IO} = \begin{pmatrix} 0.7835 \to 0.8440 & 0.5133 \to 0.6005 & (-0.1423 \to 0.1490) + i(0.0191 \to 0.1553) \\ (-0.4806 \to -0.2682) + i(0.0114 \to 0.0990) & (0.4546 \to 0.6395) + i(0.0074 \to 0.0695) & 0.6493 \to 0.7711 \\ (0.3102 \to 0.5133) + i(0.0094 \to 0.0947) & (-0.6956 \to -0.5248) + i(0.0057 \to 0.0654) & 0.6171 \to 0.7436 \end{pmatrix} \tag{1.24}$$

Current determinations of the mass-ordering and the atmospheric octant are not yet fully robust. On the other hand we still struggle with a very poor determination of the magnitude of the Dirac CP phase (see below).

1.3 Neutrino oscillation recap

We now give a "drone view" of the current status of neutrino oscillation parameters. The basic discovery made in solar and atmospheric studies was soon followed by reactor and accelerator-based experiments that have not only provided independent confirmation, but also improved parameter determination. Current experimental data mainly converge towards a consistent global picture – the three-neutrino paradigm – in which the oscillation parameters are determined as shown in figure 1.

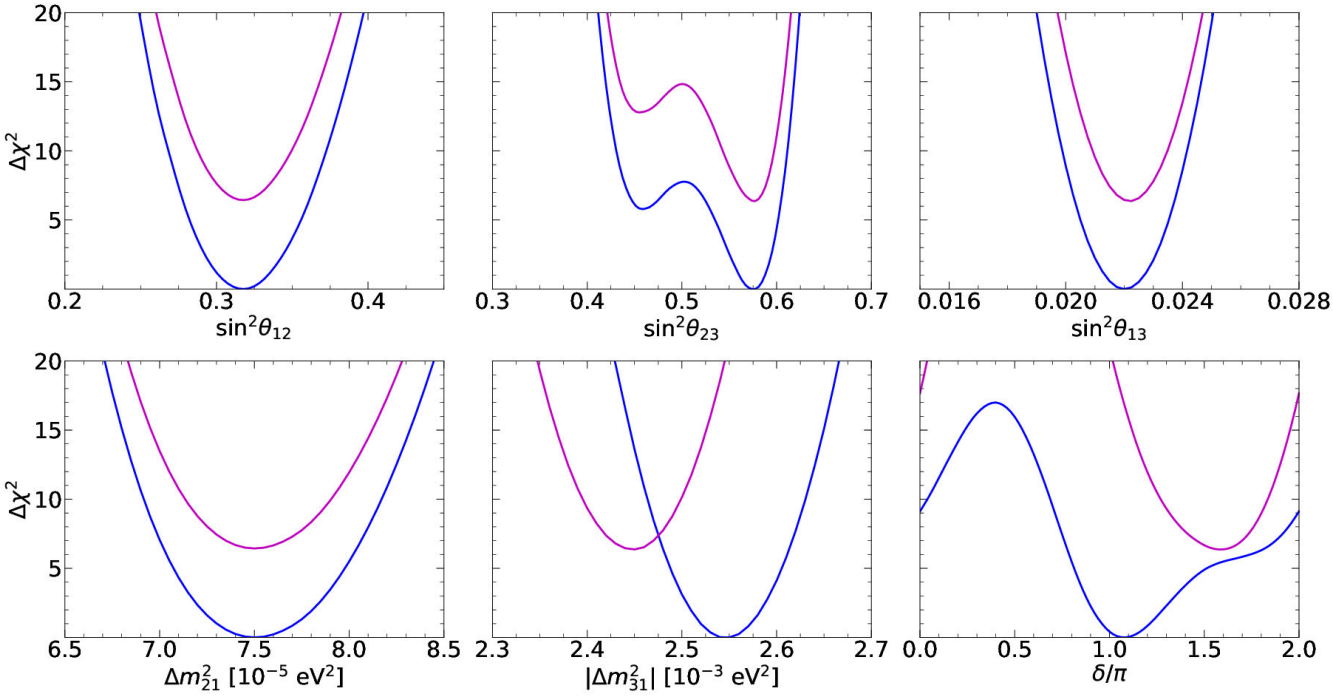


Figure 1: Current summary of neutrino oscillation parameters, where $\delta \equiv \delta_{CP}$. From [24, 25].

The top-three panels show that two mixing angles are fairly large, at odds with the corresponding mixing angles observed in the quark sector. In fact, it is the smallest lepton mixing angle, θ_{13}, that lies intriguingly close in magnitude to the largest of the quark mixing angles, i.e. the Cabibbo angle. Currently we have no explanation for this fact.

The global oscillation parameter determinations can also be shown as the "matrix" in figure 2. Besides the individual oscillation parameter determinations, given by the "diagonal" entries, the "off-diagonal" entries of figure 2 show all pairwise parameter correlations [3], very useful for model-builders. Indeed, any flavour model leads to predictions for various entries of the above triangular matrix. The zenodo website in [25] will be used extensively in this review in order to examine the allowed parameter regions in different theoretical setups.

At this point one should stress that, although the overall picture provided by the "three-neutrino paradigm" is mostly robust, there are still loose ends. As already mentioned, the determination of the neutrino spectrum and the octant of the atmospheric angle is not fully robust yet, while the precise value of the leptonic CP phase also remains an open challenge. A robust CP determination will be a key objective of the Deep Underground Neutrino Experiment (DUNE) [113]. The experiment will have two detector systems placed along Fermilab's Long Baseline Neutrino Facility (LBNF) beam. One of them will be near the beam source, while a much larger one will be placed underground 1300 km away at the Sanford Underground Research Laboratory in South Dakota, in the same mine where Ray Davis

[3] Numerical tables for the relevant χ^2 profiles can be downloaded from Ref. [25], including all pairwise correlations. These tables can be used to test various neutrino mixing patterns predicted by different flavour theories, such as those in Refs. [53–56, 59–84].

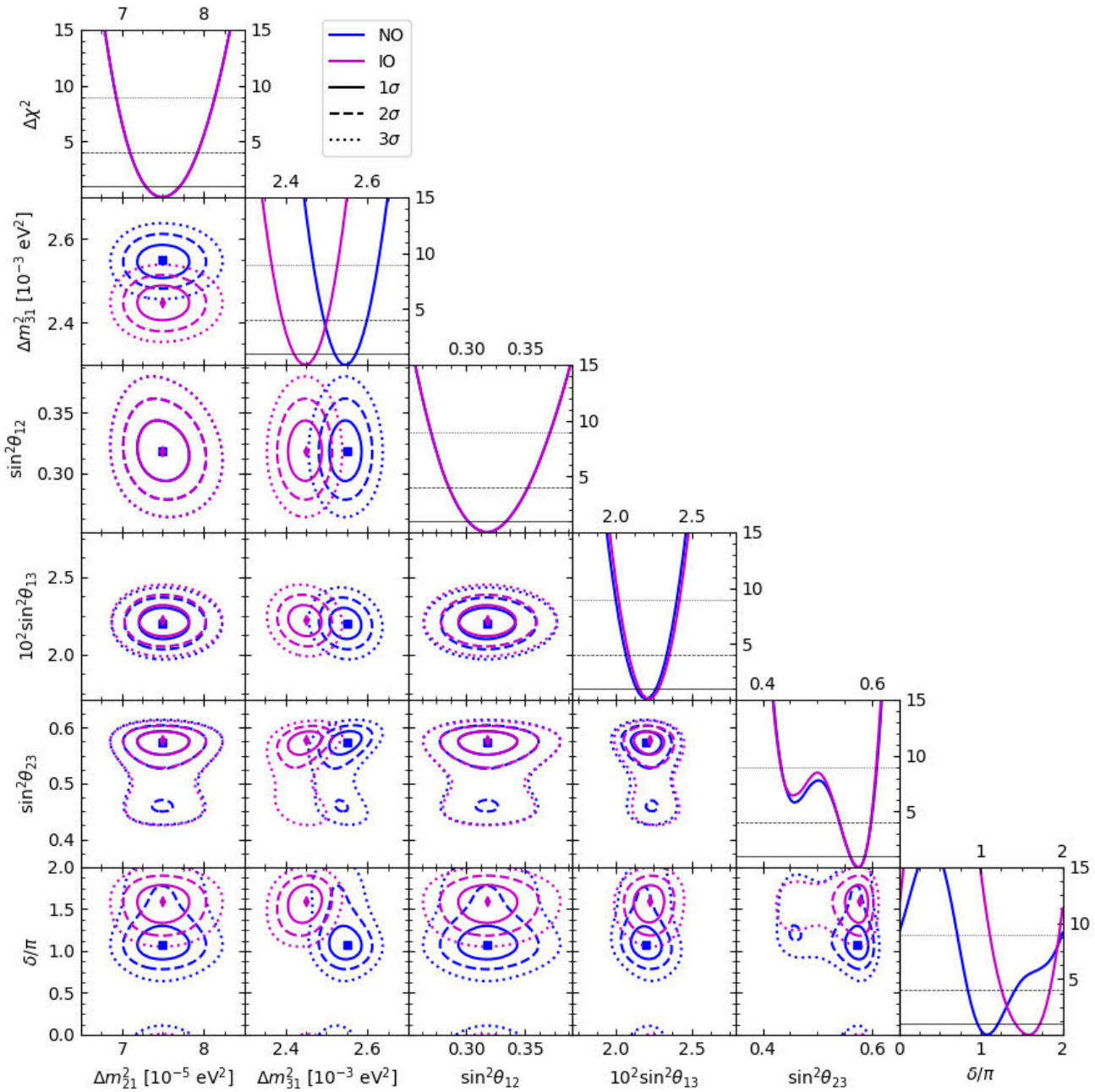


Figure 2: Neutrino oscillation parameters pair-correlations. As before $\delta \equiv \delta_{CP}$. From [24, 25].

pioneered neutrino oscillation studies in the sixties. Further improvements are expected at the ambitious T2HK proposal in Japan [114]. Reaching a precise CP phase determination will be a challenge for the coming years. Likewise, underpinning the mass ordering [115]. Octant resolution may be harder, if the preferred θ_{23} value lies close to maximality. See for example Refs. [116, 117] for details and related references. Altogether, octant discrimination, underpinning the mass-ordering and performing a precise CP determination constitute the target of the next generation of oscillation searches [118].

Last, but not least, we mention that there are also some experimental anomalies in neutrino physics

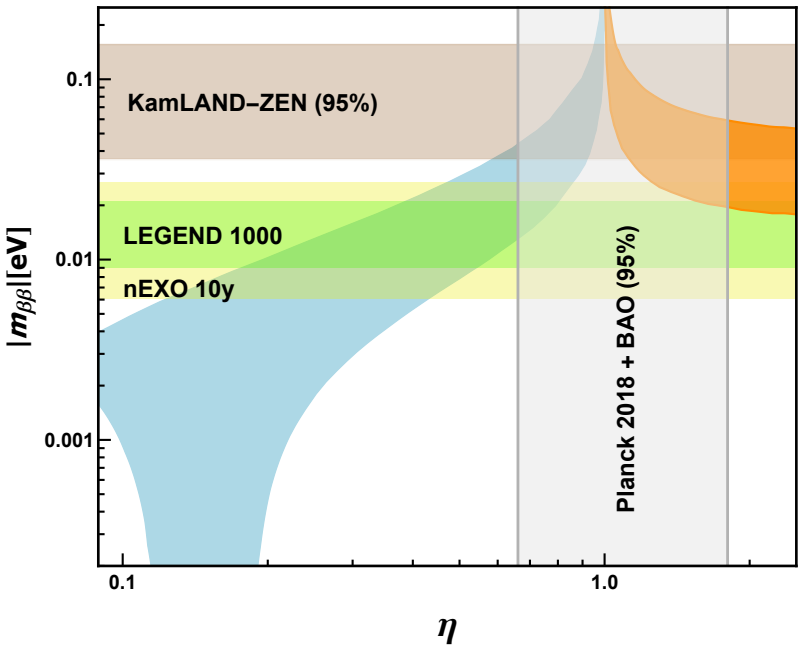


Figure 3: The $0\nu\beta\beta$ decay amplitude in a generic three-neutrino scheme versus the degeneracy parameter η. The curved bands are the normal and inverted ordering branches allowed by neutrino oscillations respectively [124]. The current experimental bound $|m_{\beta\beta}| < (36-156)$ meV at 90% confidence level (C.L.) from KamLAND-Zen [125] and the future sensitivity ranges $|m_{\beta\beta}| < (9.0-21)$ meV from LEGEND-1000 [126] and $|m_{\beta\beta}| < (6.1-27)$ meV from nEXO [127] are indicated by light brown, light yellow and light green horizontal bands respectively. The vertical grey band is excluded by the 95% C.L. limit $\Sigma_i m_i < 0.120\,\text{eV}$ from Planck [128, 129].

that challenge the simple picture provided by the "three-neutrino paradigm" suggesting, perhaps, a four-neutrino oscillation scenario [119]. They will not be discussed in this review, the interested readers can see Refs. [120–123].

1.4 Neutrinoless double beta decay

Neutrinoless double-beta decay (or $0\nu\beta\beta$ for short) is the prime lepton number violating process [106–109]. Given that neutrinos are observed to be massive fermions, and expected to be Majorana-type [30], it follows that $0\nu\beta\beta$ decay should take place, as a consequence of neutrino exchange. In this case, the effective mass parameter characterizing the amplitude is given as

$$|m_{\beta\beta}| = \left|\sum_{j=1}^{3} U_{ej}^2 m_j\right| = \left|c_{12}^{\ell\,2} c_{13}^{\ell\,2} m_1 + s_{12}^{\ell\,2} c_{13}^{\ell\,2} m_2 e^{-2i\phi_{12}} + s_{13}^{\ell\,2} m_3 e^{-2i\phi_{13}}\right| \,. \tag{1.25}$$

Here we note that, in contrast with the parametrization adopted by the PDG, the symmetrical form of the lepton mixing matrix provides a conceptually transparent description of $0\nu\beta\beta$ in which, as it should, only Majorana phases appear [99]. Altogether, the original symmetrical form of the lepton mixing matrix, Eq. (1.21), provides an insightful description both for the Dirac phase, Eq. (1.22), as well as the Majorana phases, Eq. (1.25).

An important point to notice from Eq. (1.25) is that, thanks to the Majorana phases, the $0\nu\beta\beta$ amplitude can vanish due to possible destructive interference among the three individual neutrino amplitudes. This is clearly seen in the blue branch of figure 3, taken from Ref. [124], which shows the $0\nu\beta\beta$ decay amplitute versus the degeneracy parameter η. When $\eta \to 1$ neutrino masses become degenerate, and the two bands correspond to the two possible mass orderings. The most favorable case for $0\nu\beta\beta$ decay detectability happens when neutrinos are nearly degenerate, as predicted in some UV-complete flavour theories [53–55]. The $\eta \equiv 1$ case corresponds to the idealized limit where neutrinos would be strictly degenerate. In order to generate oscillations neutrino masses must deviate from exact degeneracy. This can happen in two ways, corresponding to the two curved branches seen in the figure. The normal-ordered (NO) neutrino region is indicated in the left (blue) band, while inverted ordering (IO) gives the upper-right (orange) branch. One sees that, thanks to the presence of the Majorana phases [30], the $0\nu\beta\beta$ amplitude can vanish for normal ordering (but not for inverted). The horizontal band denotes the current KamLAND-Zen limit [125], while the vertical one is excluded [124] by cosmological observations [128,129] e.g. by the Planck collaboration. Altogether, one sees current data strongly disfavour nearly degenerate neutrinos.

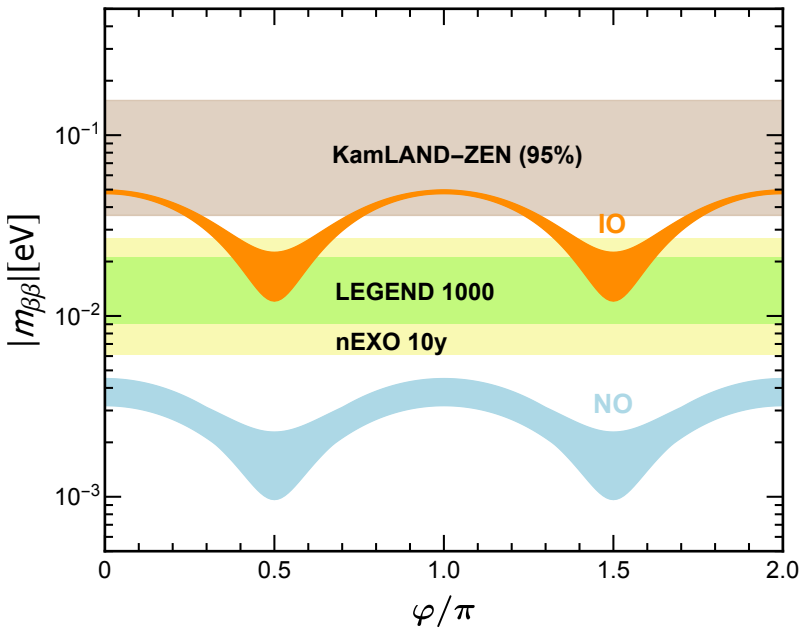


Figure 4: $0\nu\beta\beta$ decay amplitude when one neutrino is massless. The light-blue and light-orange bands are the current 3σ C.L. regions for normal and inverted mass-ordering. The current bound from KamLAND-ZEN [125] and projected sensitivities of LEGEND 1000 [126] and nEXO [127] are indicated.

Another interesting situation happens if one (or two) of the three neutrinos is massless or nearly so, as in the "missing partner" seesaw mechanism [30] and other "incomplete-multiplet" schemes [4]. The minimal viable tree-level type-I seesaw has only two right-handed neutrinos [131–133]. In such a scheme no cancellation is possible, even for normal-ordering [134–138]. The resulting regions are the two periodic

[4] A massless neutrino may also arise from the presence of anti-symmetric Yukawa couplings, see e.g. [130].

bands seen in figure 4, which are expressed in terms of the only free parameter available, namely, the relative neutrino Majorana phase φ [5]. The colored horizontal bands show current experimental limits of KamLAND-ZEN and the expected sensitivities of LEGEND 1000 and nEXO, where the width of the bands reflects nuclear matrix element uncertainties [108, 109]. Notice in this case the existence of a predicted theoretical lower bound on $|m_{\beta\beta}|$. Taking into account the sensitivities expected at upcoming $0\nu\beta\beta$ experiments, one sees that this lower bound for NO lies below detectability in the foreseable future.

In contrast, inverse mass-ordering provides a lower bound that lies higher than the one generically expected for the IO three-massive-neutrino case. This implies a guaranteed discovery in the next round of experiments [139, 140]. In fact, for this case, the recent KamLAND-Zen limit [125] already probes the Majorana phase, as seen by the magenta band in figure 4. In short, for the one-massless-neutrino case there is a chance, perhaps, that one could be able to extract the value of the relevant Majorana phase from experiment.

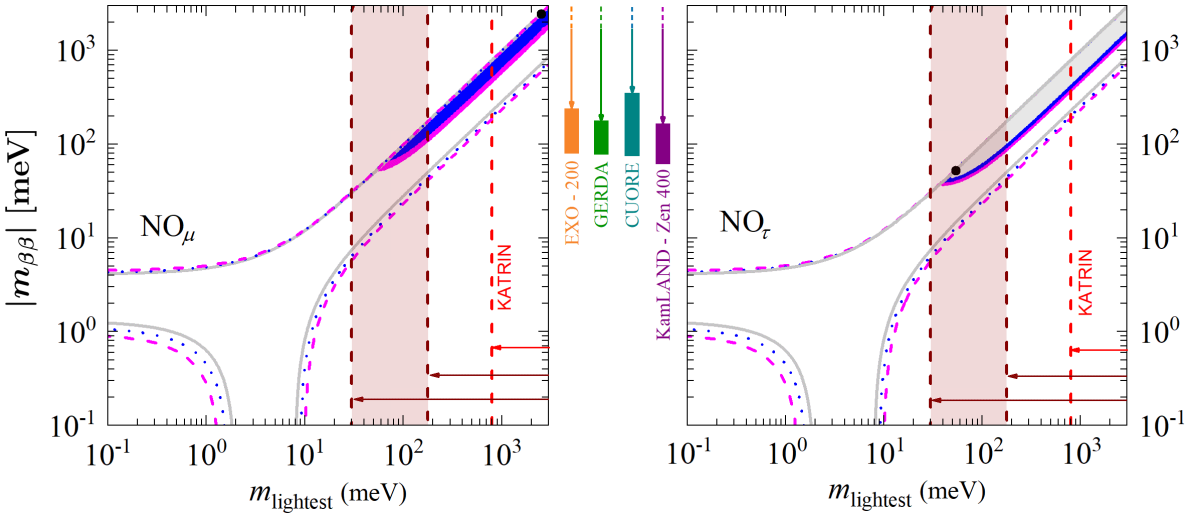


Figure 5: The solid (dotted) [dashed] lines delimit the 1σ (2σ) [3σ] $|m_{\beta\beta}|$ regions allowed by oscillations. Predictions of two normal-ordered $\mathcal{Z}_8$ schemes are given [148]. Vertical bars in mid-panels indicate current 95%CL $|m_{\beta\beta}|$ upper bounds from KamLAND-Zen 400 [149], GERDA [150], CUORE [151] and EXO-200 [152].

We now turn to the general three-massive-neutrino case. Although there is no garanteed minimum value for $|m_{\beta\beta}|$ in this case, there can still be a lower-bound, even for normal mass-ordering, provided the cancellation of the amplitudes is prevented by the structure of the leptonic weak interaction vertex [66, 68, 69, 141–144]. This can happen as a result of the imposition of a family symmetry to account for the mixing pattern seen in oscillations. As an example, figure 5 shows the predictions of a $\mathcal{Z}_8$ family symmetry scheme. The hollow solid (dotted) [dashed] lines delimit the 1σ (2σ) [3σ] $|m_{\beta\beta}|$ regions allowed in the general three-neutrino case. The sub-regions shown in gray, blue and magenta, respectively, are the flavour-model predictions. The black dots correspond to best-fits. The vertical shaded band indicates the current sensitivity of cosmological data [128]. The vertical dashed red line corresponds to the KATRIN

[5] For vanishing lightest neutrino mass the relevant Majorana phase is $\varphi = \phi_{12} - \phi_{13}$ for NO and $\varphi = \phi_{12}$ for IO.

tritium beta decay [145] upper limit $m_\beta < 0.8$ eV (90% CL) [146,147]. The heights of the bars shown in the mid-part of figure 5 reflect the uncertainties in the nuclear matrix elements relevant for the computation of the decay rates. The flavour predictions for $|m_{\beta\beta}|$ illustrate our point. For example, one sees how the preferred flavour-predicted point in the right panel sits right inside the cosmologically interesting band, and close to the current $0\nu\beta\beta$ limits, as indicated in between the panels. Similar predictions for $0\nu\beta\beta$ amplitudes occur in other family symmetry models, some of which will be discussed in this review, see for instance discussions given in Chapters 8 and 10.

In short, oscillations have left an important imprint upon neutrinoless double beta decay studies. The results of the negative searches conducted so-far imply that we must consider both the possibilities of Dirac or Majorana neutrinos, as we do in this review. Nonetheless, there is a reasonable chance that, perhaps, $0\nu\beta\beta$ could be seen in the coming round of experiments. This would constitute a major breakthrough. Indeed, a positive $0\nu\beta\beta$ decay detection would imply, as a consequence of the black box theorem, that at least one of the neutrinos has Majorana nature [110]. The argument is illustrated in figure 6. Note that the black-box argument holds irrespective of the underlying physics responsible for generating the process [153, 154].

In some cases, as we saw in figure 4, the discovery of $0\nu\beta\beta$ decay might allow us to underpin also the value of the relevant Majorana phase. Note however that, although a positive $0\nu\beta\beta$ discovery would imply that at least one of the neutrinos is a Majorana particle, a negative result would not imply that neutrinos are Dirac-type, as the amplitude can be suppressed even for Majorana-type neutrinos, due to the effect of the Majorana phases. It has been argued that if a null $0\nu\beta\beta$ decay signal was accompanied by a positive $0\nu4\beta$ quadruple beta decay signal [155, 156], then at least one neutrino should be a Dirac particle [157].

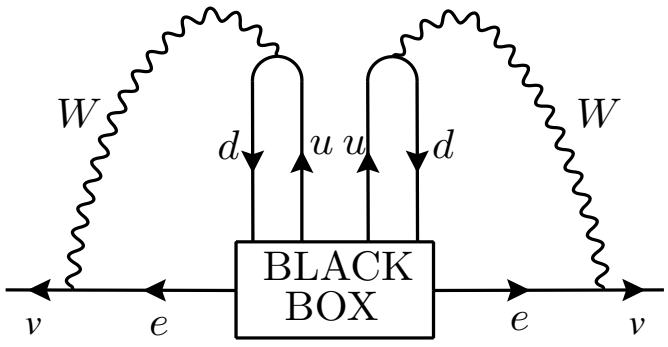


Figure 6: The observation of $0\nu\beta\beta$ decay implies that at least one neutrino is a Majorana fermion [110].

2 Origin of neutrino masses and flavour puzzle

Despite its amazing success in describing three out of the four known fundamental forces of nature, there are many drawbacks and open issues in the Standard Model. Altogether, they imply that the theory of particle physics needs a completion beyond its current form. Here we start with one of the most important issues, i.e. the lack of neutrino masses.

2.1 Effective neutrino masses

Although the Standard Model lacks neutrino masses, these can arise effectively from a unique dimension-five operator characterizing lepton number non-conservation [29], as illustrated in figure 7. In this case neutrinos are Majorana fermions, as generally expected in gauge theories. Indeed, on general grounds it was argued in [30], within the $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ setup, that neutrinos are expected to be Majorana fermions.

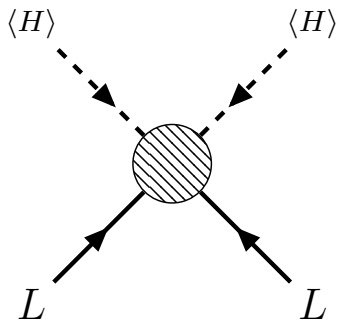


Figure 7: Majorana neutrino mass generation from the dimension-five operator.

However, neutrinos could also be Dirac fermions, this is currently an open experimental question. Dirac neutrinos can indeed emerge in the presence of extra symmetries, that could be discrete or continuous, global or local. For example, the imposition of $U(1)_{B-L}$ symmetry, where B and L are baryon and lepton numbers respectively, would forbid the Majorana mass terms of the right-handed neutrinos [158–161]. Alternatively, as mentioned below, the Dirac nature of neutrinos could result from the presence of an underlying Peccei-Quinn symmetry [162,163] or be associated with dark matter stability [164–171]. Small Dirac-type neutrino masses as in Eq. (1.11), could arise from effective dimension-five [172–175] as well as dimension-six operators [176,177], that have by now been classified.

2.2 The seesaw paradigm

An attractive ultraviolet completion of the dimension-five operator is provided by the seesaw mechanism. It is specially interesting, as it gives a simple *dynamical* understanding of small neutrino masses by minimizing the Higgs potential, as well as the possibility of having a stable electroweak vacuum [178–183].

It has become usual to distinguish three main seesaw varieties, namely type-I [30–35], type-II [30–32, 36–38] and type-III [184], illustrated in figure 8. These generate neutrino masses at tree level from the mediation of new heavy singlet fermion (ν_R), triplet scalar (Δ) and triplet fermion (Σ_R), respectively. In the simplest seesaw neutrinos acquire masses through the exchange of heavy scalars, called type-I in [30] and today called type-II. Such seesaw realization allows one to reconstruct the parameters characterizing the neutrino sector from various experiments [185], in particular those from high energy colliders [186,187]. Moreover, the type-II seesaw opens the really tantalizing possibility of probing the absolute neutrino mass and mass ordering in collider experiments [188].

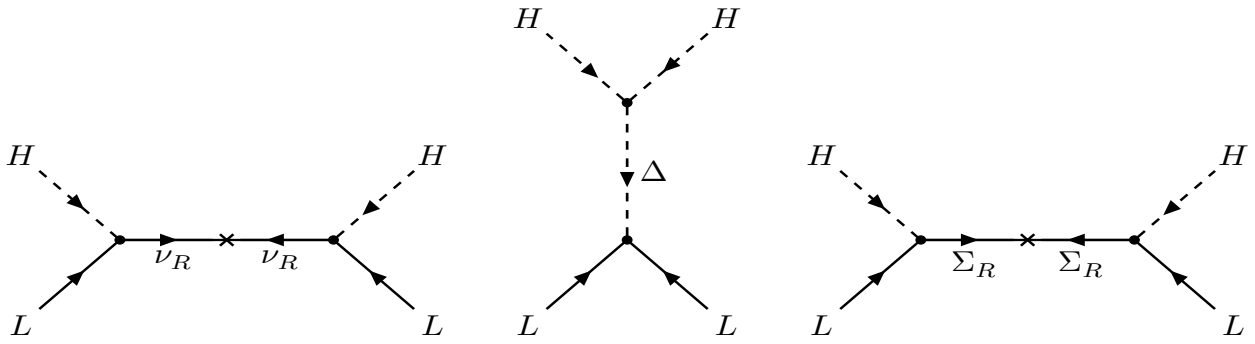


Figure 8: Feynman diagrams for the conventional seesaw types, where the mediators ν_R and Σ_R are SM singlet and triplet fermions respectively, while Δ is a SM triplet scalar.

Here we focus on the type-I seesaw mechanism, where neutrinos acquire masses through the exchange of heavy gauge-singlet fermions, as illustrated in the left panel of figure 8. Seesaw mediators were originally thought to lie at a high mass scale, associated to SO(10) unification or left-right symmetry. The associated physics has been covered in several textbooks [102–105] and reviews [189]. Note that, following Ref. [30], here we do not assume left-right symmetry in the seesaw mechanism, but simply the minimal well-tested $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge structure. Notice that this seesaw description of neutrino mass generation can be made fully dynamical in the presence of a singlet vacuum expectation value (VEV) driving the spontaneous violation of lepton number symmetry. The theory is then accompanied by a Nambu-Goldstone boson, dubbed Majoron [32, 39].

Notice that the most general realization of the seesaw mechanism leads most naturally to Majorana neutrinos. However, in the presence of adequate extra symmetries, Dirac neutrinos can emerge from the seesaw [190]. Indeed, there may be good reasons for neutrinos to be Dirac-type. For example, the Dirac nature of neutrinos could also be associated to the existence of an underlying Peccei-Quinn symmetry [162, 163]. Moreover, it could signal the stability of dark matter [164–171]. Finally, there are interesting scenarios where the Dirac nature of neutrinos is associated with a family symmetry [191, 192].

2.3 Dark matter as the source of neutrino mass

It is well-known that about 85% of the matter in the universe is "dark" in the sense that it does not appear to interact with the electromagnetic field and is, as a consequence, very hard to detect. It is not our purpose to provide a comprehensive discussion of the observational aspects of cosmological dark matter; for a review see Ref. [193]. The first point we note is that the existence of such cosmological Dark matter is totally unaccounted for within the Standard Model. A very interesting possibility is that dark matter is made up of a novel weakly interacting massive particle, dubbed WIMP, typically present in the case of supersymmetric theories with conserved R-parity [194].

Underpinning the mechanism yielding small neutrino masses and/or dark matter is of paramount importance in particle physics, as it may select which new physics to expect as the next step. This is a very broad subject which we will not try to review. Rather, in this review we just comment on the issue of particle dark matter candidates, and how they may be simply related to the mechanism of neutrino mass generation. We focus on the very interesting possibility that the dark matter sector mediates neutrino mass generation, as postulated within the scotogenic picture [195–201]. An interesting twist is to imagine the existence of a universal "hidden" dark matter sector that *seeds* neutrino mass generation, which proceeds *a la seesaw*, as in the dark inverse seesaw [202] mechanism. Alternatively, one can envisage a dark linear seesaw mechanism [203, 204].

Both dark inverse and dark linear seesaw realizations will be described below. In either case dark matter will be WIMP-like. Instead of being related to supersymmetry, WIMP dark matter in these models is closely related to neutrino physics, either as mediator or as seed of neutrino mass generation. Flavored scotogenic dark matter may also be implemented [205], providing an attractive way to reconcile low-scale radiative neutrino mass generation with dark matter, while addressing the flavour problem at the same time, see Sec. 2.7.

Last, but not least, we mention that particle physics theories where neutrino masses arise from the spontaneous breaking of a continuous global lepton number symmetry have a natural dark matter candidate [206, 207], namely the associated Nambu-Goldstone boson, dubbed majoron [32, 39]. Indeed, the majoron is likely to pick up a mass from gravitational instanton effects, that explicitly violate global symmetries [208]. The majoron necessarily decays to neutrinos, with an amplitude proportional to their tiny mass, which typically gives it cosmologically long lifetimes [32]. The associated restrictions on the decaying warm dark matter picture coming from the CMB [209] as well as mono-energetic photon emission in astrophysics have been examined in detail [210]. Using N-body simulations it has also been been shown that the warm majoron dark matter model provides a viable alternative to the ΛCDM scenario, with predictions that can differ substantially on small scales [211].

2.4 Missing partner seesaw and dark matter: the scoto-seesaw

In contrast to a left-right symmetric SO(10)-based seesaw mechanism, where the number of "left" and "right" neutrinos must match as a consequence of gauge invariance, in the most general SM-based seesaw mechanism [30] one can have any number (m) of "right-handed" neutrino mediators, since they are gauge

singlets. Theories with $m < 3$ in the classification of [30] could simplify substantially the form of the lepton mixing matrix [6].

Such "missing partner" seesaw schemes with less "right" than "left" neutrinos, imply that some of the active "left" neutrinos remain massless [30, 131, 132]. The (3,2) choice corresponds to the minimal viable type-I seesaw, in which there are only two mass parameters corresponding to the experimentally measured solar and atmospheric splittings. The lightest neutrino is massless, leading to the $0\nu\beta\beta$ prediction discussed in Sec. 1.4.

On the other hand, the minimal (3,1) seesaw leads to the same $0\nu\beta\beta$ prediction as (3,2), but has only one neutrino mass parameter [30]. Although such setup is not consistent with current neutrino data, as it lacks solar neutrino oscillations, the later may arise from a completion in which the degeneracy between the lowest-mass neutrinos is lifted by some other mechanism, for example, as a result of radiative corrections.

Indeed, the (3,1) setup offers a template to reconcile the seesaw paradigm and the WIMP dark matter paradigm within a minimal hybrid construction called "scoto-seesaw" mechanism. Such simplest (3,1) "scoto-seesaw" scheme provides a comprehensive theory of neutrino mass generation as well as WIMP dark matter, in which the relative magnitudes of solar and atmospheric oscillation lengths are explained due to a loop factor [136, 137].

2.5 The low-scale inverse and linear seesaw mechanisms

As a final example, one may consider having more "right-" than "left-handed" neutrinos in the seesaw mechanism, for example, two isosinglets for each "active" isodoublet family. This can be implemented with explicit [213] as well as spontaneous violation of lepton number [40]. In the lepton-number-conserving limit one finds that the three light neutrinos remain massless, as in the Standard Model [214–216]. In contrast to the Standard Model, however, lepton flavour and lepton CP symmetries can be substantially violated [7]. This shows that flavour and CP violation can exist in the leptonic weak interaction despite the masslessness of neutrinos, implying that such processes need not be suppressed by the small neutrino masses, and hence can have large rates [214–216, 222, 223] [8]. Such "(3,6)" setup [30], where the 3 doublet neutrinos are accompanied by 6 heavy singlet neutral leptons, provides the template for building genuine "low-scale" realizations of the seesaw mechanism in which lepton number is restored at low values of the lepton number violation scale. The models are natural in *t'Hooft* sense, leading to small, symmetry-protected neutrino masses. Realizations of such genuine "low-scale" seesaw mechanisms include the inverse [40, 213] as well as the linear seesaw [229–231]. Last, but not least, we notice that a dynamical realization of the seesaw mechanism involving the spontaneous violation of lepton number symmetry can improve the consistency properties of the electroweak vacuum [178–183].

If realized at low-scale, the type-I seesaw mechanism may also lead to signals at high-energy colliders. Already back in the LEP days, it was suggested that the neutrino mass mediators could be produced

[6] For an early example, with lepton number symmetry conservation, see [212].
[7] General discussions of leptonic flavour and CP violation are given in [217–221].
[8] For generic references on lepton flavour violation and seesaw schemes see, for example [224–228].

at high energy colliders [232–237]. This proposal was indeed taken up by subsequent experiments, for example the ATLAS and CMS experiments at the LHC [238–240] as well as future proposals [12,241–243]. These have also taken into account the possibility of having displaced vertices [244–246] [9] coming from the fact that the couplings responsible for the mediator decays can be neutrino-mass-suppressed.

2.6 Dark inverse and linear seesaw mechanisms

The symmetry protection provided by the above schemes can be upgraded into a "double protection", by having the seed of lepton number violation to arise radiatively. Indeed, a very interesting possibility has been suggested, namely the dark inverse [202] and also the dark linear seesaw mechanism [203, 204]. In the former case there is a universal gauge-singlet or "hidden" dark matter sector that *seeds* neutrino mass generation, which proceeds *a la seesaw*. The same idea can be used to promote the linear seesaw mechanism into a mechanism sourced by a dark sector. The latter is not unique, interesting examples were given in [203] and [204].

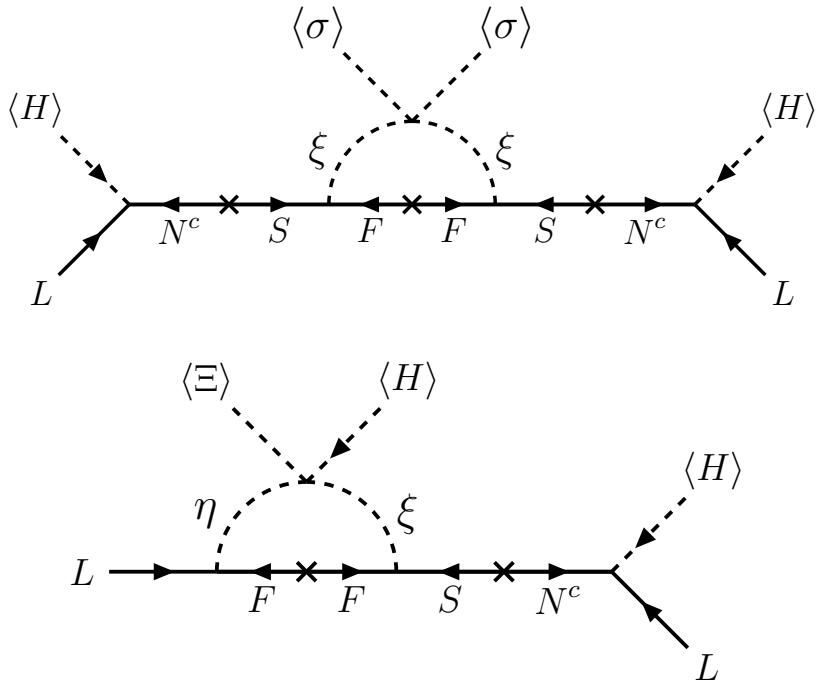


Figure 9: Feynman diagrams for "dark" inverse seesaw (upper panel) [202], and the "dark" linear seesaw mechanism (lower one) [203]. For another realization of the linear seesaw see [204].

The dark inverse seesaw is illustrated in the upper panel of figure 9. The lepton number violation loop that "seeds" neutrino mass is mediated by the "dark" gauge singlet fermion F and dark singlet scalar ξ [202]. This provides an elegant way to reconcile the seesaw and dark matter paradigms, providing an interesting dynamical seesaw theory where dark matter and neutrino mass generation are closely

[9] Other neutrino mass mediator searches can also lead to displaced vertices [247–249].

inter-connected. Scenarios have also been proposed where a dark sector seeds neutrino mass generation radiatively within the linear seesaw mechanism [203, 204], as illustrated in the lower panel. Neutrino masses are also calculable, since tree-level contributions are forbidden by symmetry. By having the seesaw realized at low-scale [40, 213], as indicated in figure 9, in either case one has charged lepton flavour violating processes e.g. $\mu \to e\gamma$ with accessible rates [202]. We stress that these can be large, despite the tiny neutrino masses. Interesting dark-matter and collider physics implications have also been discussed.

2.7 Neutrinos and the flavour puzzle

We now turn to the issue of understanding the pattern of the weak interactions of quarks and leptons from first principles. Such "flavour problem" constitutes a major challenge in modern particle physics. Why three families of quarks and leptons? How to explain their mass hierarchies e.g. why, though otherwise so similar, the muon is about 200 times heavier than the electron? Why is the top quark mass so large compared with the masses of the other fermions? On the other hand, we also face the question of explaining why do the fermion mixing matrices follow the peculiar pattern observed? In particular, why do leptons mix so differently from the way quarks do? All of these shortcomings pose a real challenge on unified descriptions of nature.

As it stands, the Standard Model of particle physics lacks an organizing principle in terms of which to understand the flavour problem. Hence it can not be a complete theory of nature. Although the Standard Model suffers from many other drawbacks, in this review we focus mainly on explaining the fermion mixing pattern and CP violation, though we also discuss some ideas to account for the fermion mass hierarchies. A "flavour completion" of the SM would have potentially important tests in the laboratory. Moreover, a deep understanding of the flavour problem and CP violation may prove crucial in understanding the baryon asymmetry of the universe [250, 251].

In this review we cover some recent attempts to account for the pattern of neutrino mixing indicated by oscillation experiments. Our key ingredient in formulating a theory of flavour is the imposition of an extra symmetry G_f relating the families. The leading Lagrangian for the lepton masses should be invariant under both the SM gauge symmetry as well as the flavour symmetry, and it can be generally written as

$$\mathcal{L}_m^l = -[y_l(\Phi_l)]_{ij}\overline{l_R^i}H^\dagger L_L^j - \frac{1}{2\Lambda}[y_\nu(\Phi_\nu)]_{ij}\overline{(L_L^i)^c}\, i\sigma_2 H H^T i\sigma_2 L_L^j + \text{h.c.}\,, \tag{2.1}$$

where $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ is Pauli's conjugation matrix, $y(\Phi)$ generically denotes the Yukawa couplings, which are determined as functions of the flavons, and the neutrino masses are described by the Weinberg operator [29]. Here Φ_l and Φ_ν denote the flavon fields with vacuum expectation values that break G_f down to the residual subgroups G_l and G_ν in the charged lepton and neutrino sectors respectively. In concrete models the higher-order terms can lead to subleading corrections. As a result, the mismatch of the residual symmetries G_l and G_ν allows one to make model-independent symmetry predictions for the lepton mixing matrix $U = U_l^\dagger U_\nu$, as illustrated in figure 10.

Within the bottom-up approach one may consider remnant flavour symmetries and/or remnant CP

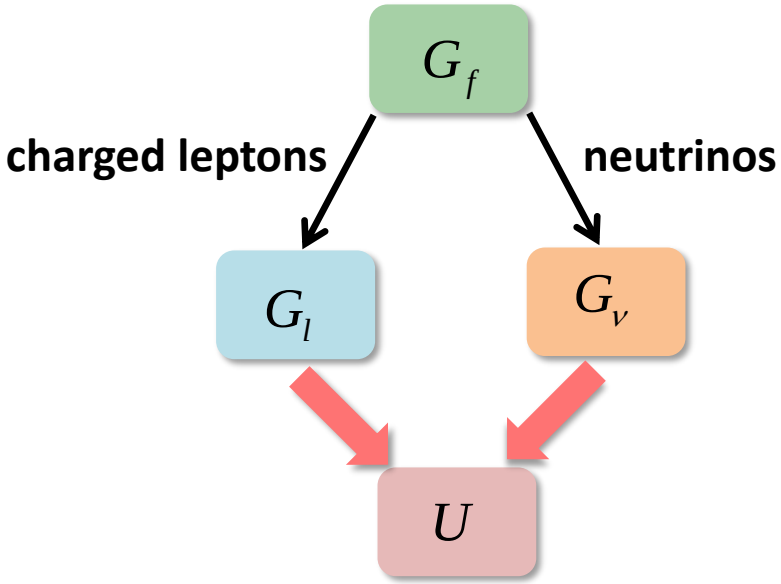


Figure 10: Predicting lepton flavour mixing from the flavour group G_f breaking to different subgroups.

symmetries, such as generalizations of the $\mu - \tau$ reflection symmetry [252–258], as a way to restrict the lepton mass matrices, irrespective of the details of the underlying flavour symmetry. This can be used in turn to constrain the lepton flavour mixing parameters, such as mixing angles [259–263] and especially the CP phases [264–268]. In what follows we call this approach "residual symmetry method" [269–272]. We notice that such $\mu-\tau$ reflection symmetry can emerge within UV-complete theories, e.g. [54, 264]. Indeed, one may seek to constrain the flavour mixing parameters within a complete flavour theory. Attaining a satisfactory theory that explains both quark and lepton sectors together remains an open challenge [273, 274], especially within a quark-lepton unified framework [34, 43–47, 275]. This difficulty [85–93] follows mainly due to the disparity between quark and lepton mixing angles. A possible idea is that lepton mixing angles are large with respect to quark mixing angles because neutrino masses are degenerate in the family symmetry limit [54]. However, nearly degenerate neutrinos are now strongly disfavored by cosmological restrictions [124].

There have been many efforts to approach the flavour problem in the quark sector in terms of an underlying family symmetry, either within a bottom-up or top-down approach [276–286]. For example, the Cabibbo mixing angle can be described by a flavour symmetry [263, 270, 278, 281, 287]. Similarly, quark mixing angles and CP violation may also be accounted for in the framework of generalized CP symmetries [282–286].

Turning now to mass predictions, we note that some quark-lepton relations may emerge naturally from the imposition of family symmetries, even in the absence of a genuine unification group. A prime example is the "golden" formula,

$$\frac{m_\tau}{\sqrt{m_\mu m_e}} \approx \frac{m_b}{\sqrt{m_s m_d}}\,, \tag{2.2}$$

relating charged-lepton and down-type quark masses. This successful relation may constitute, perhaps, part of any complete theory of flavour. The golden formula could be associated to an underlying Peccei-Quinn symmetry [287]. Alternatively, it could arise within UV-complete family-symmetry-based theories such as [288–290] or [65, 66]. Finally, we note that distinct variants of the golden formula can arise from modular symmetries [291].

We stress that the golden-type relations involve mass ratios, stable under renormalization-group evolution, and allow us to relate quarks and leptons without invoking a genuine unified gauge group such as SU(5) or SO(10). More ambitious approaches to the flavour problem have been proposed within extra spacetime dimensions. For example, 5-D warped schemes were proposed in which mass hierarchies are accounted for by adequate choices of the bulk mass parameters, while quark and lepton mixing angles are restricted by the imposition of a flavour symmetry [70–84]. Predictions for neutrino mixing, CP violation and $0\nu\beta\beta$ decay, as well as a good global fit of flavour observables emerge in warped flavordynamics [78,79].

On the other hand, 6-dimensional orbifold compactification has been suggested as a promising way to determine the structure of the family symmetry in four dimensions. In this approach the 4-dimensional flavour group emerges from the symmetries between the branes in extra dimensions [292, 293]. Interesting flavour and CP predictions have been obtained from 6-dimensional orbifold compactification schemes, including neutrino oscillation and $0\nu\beta\beta$ decay predictions, as well as a successful global description of the flavour observables [143, 144, 294, 295].

3 Lepton mixing patterns

The most striking implications of the oscillation experiments is that leptons mix quite differently from quarks. Here we consider the interpretation of the oscillation results in terms of phenomenological neutrino mixing patterns. Several of these were suggested, largely motivated by the desire to shed light on the neutrino oscillation parameters. The most salient features of the oscillation phenomenon are captured by the so-called tri-bimaximal (TBM) mixing pattern [296–298]. However there are other interesting mixing patterns, some of which we survey below [10].

3.1 Tri-bimaximal mixing (TBM)

The Tri-bimaximal lepton mixing pattern embodies bi-maximal mixing of atmospheric neutrinos, tri-maximal solar mixing, no reactor mixing and hence no CP violation. It has been taken very seriously until the Daya Bay collaboration [21,303,304] provided a robust measurement of the angle θ_{13} in 2012. The latest and most precise measurement of Daya Bay gives $\sin^2 2\theta_{13} = 0.0851 \pm 0.0024$ [305,306]. Nevertheless, the TBM pattern still remains as an interesting first step towards a full description of neutrino mixing. Its mixing matrix is given by

$$U_{TBM} = \begin{pmatrix} \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{3}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \end{pmatrix} , \tag{3.1}$$

which gives

$$\theta_{23} = 45^\circ, \quad \theta_{12} = \arctan \frac{1}{\sqrt{2}} \simeq 35.26^\circ, \quad \theta_{13} = 0^\circ . \tag{3.2}$$

Under the assumption of Majorana neutrinos, the most general form of the associated neutrino mass matrix in the charged lepton diagonal basis is

$$\begin{aligned} m_\nu^{TBM} &= U_{TBM} \mathrm{diag}(m_1, m_2, m_3) U_{TBM}^T \\ &= \frac{1}{6} \begin{pmatrix} 4m_1 + 2m_2 & -2m_1 + 2m_2 & -2m_1 + 2m_2 \\ -2m_1 + 2m_2 & m_1 + 2m_2 + 3m_3 & m_1 + 2m_2 - 3m_3 \\ -2m_1 + 2m_2 & m_1 + 2m_2 - 3m_3 & m_1 + 2m_2 + 3m_3 \end{pmatrix} . \end{aligned} \tag{3.3}$$

It is easy to check that the above neutrino mass matrix is invariant under the following residual flavour transformations

$$G_1^{TBM} = \frac{1}{3} \begin{pmatrix} 1 & -2 & -2 \\ -2 & -2 & 1 \\ -2 & 1 & -2 \end{pmatrix} , \; G_2^{TBM} = \frac{1}{3} \begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix} , \; G_3^{TBM} = - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} , \tag{3.4}$$

[10] Note that, within a UV-complete theory framework, there can be radiative corrections to the TBM predictions [299–302].This issue lies beyond the scope of the model-independent approach followed here.

up to an overall sign. In other words, the neutrino mass matrix m_ν^{TBM} fulfills

$$(G_i^{TBM})^T m_\nu^{TBM} G_i^{TBM} = m_\nu^{TBM} \,. \tag{3.5}$$

The symmetry transformation G_3^{TBM} exchanges the second and third columns as well as the second and third rows of the neutrino mass matrix, it is the well-known $\mu-\tau$ permutation symmetry [307–311], and it enforces maximal θ_{23} and vanishing θ_{13}. This symmetry emerges is some UV-complete theories, e.g. [54].

The second transformation G_2^{TBM} requires that the sum of the entries in each row of the light neutrino mass matrix should be equal to that of the corresponding column, the so-called magic symmetry [255,312]. The invariance under G_2^{TBM} determines one column of the lepton mixing matrix to be $(1,1,1)^T/\sqrt{3}$. The residual symmetry transformation G_e of the charged lepton mass matrix is a generic diagonal phase matrix in the flavour basis [11]. The simplest choice for G_e which can distinguish the three charged leptons is

$$G_e^{TBM} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix} \tag{3.6}$$

where ω is the cubic root of one, i.e. $\omega = e^{2\pi i/3}$.

The group generated by G_1^{TBM}, G_2^{TBM}, G_3^{TBM} and G_e^{TBM} turns out to be S_4 group. Consequently, the minimal flavour symmetry group capable of yielding the tri-bimaximal mixing is S_4 [269, 270]. However, in the absence of symmetry breaking by flavons in the $\mathbf{1}'$ and $\mathbf{1}''$ representations the A_4 flavor symmetry can also give rise to the tri-bimaximal mixing pattern [53, 54, 59, 70, 313–315].

3.2 Generalizations of tri-bimaximal mixing

Soon after the tri-bimaximal mixing ansatz was proposed, possible deviations were considered. A simple generalization is obtained by assuming that just one given row or column of the lepton mixing matrix takes the same form as for tri-bimaximal [252]. If the first column of the tri-bimaximal mixing is retained, this is known as the TM1 mixing pattern [316–318], and the associated mixing matrix can be parameterized as

$$U_{TM1} = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & \sqrt{2}\cos\theta & \sqrt{2}\, e^{-i\delta}\sin\theta \\ -1 & \sqrt{2}\cos\theta - \sqrt{3}\, e^{i\delta}\sin\theta & \sqrt{3}\cos\theta + \sqrt{2}\, e^{-i\delta}\sin\theta \\ -1 & \sqrt{2}\cos\theta + \sqrt{3}\, e^{i\delta}\sin\theta & -\sqrt{3}\cos\theta + \sqrt{2}\, e^{-i\delta}\sin\theta \end{pmatrix} . \tag{3.7}$$

The predictions for the lepton mixing angles and leptonic Jarlskog invariant are expressed in terms of just two free parameters δ and θ as

$$\sin^2\theta_{13} = \frac{1}{3}\sin^2\theta, \qquad \sin^2\theta_{12} = \frac{1+\cos 2\theta}{5+\cos 2\theta},$$

[11] If G_e is a non-abelian subgroup, the charged lepton mass spectrum would be completely or partially degenerate. Thus G_e should be a cyclic group Z_n with the index $n \geq 3$ or a product of cyclic groups such as $Z_2 \times Z_2$ in order to distinguish among the generations.

$$\sin^2\theta_{23} = \frac{1}{2} + \frac{\sqrt{6}\,\cos\delta\sin 2\theta}{5+\cos 2\theta}, \quad J_{CP} = \frac{1}{6\sqrt{6}}\sin\delta\sin 2\theta\,. \tag{3.8}$$

One finds that the lepton mixing angles and Dirac CP phases are correlated as follows

$$3\cos^2\theta_{12}\cos^2\theta_{13} = 2, \quad \tan 2\theta_{23}\cos\delta^{\ell} = \frac{5\sin^2\theta_{13}-1}{2\sin\theta_{13}\sqrt{2-6\sin^2\theta_{13}}}\,. \tag{3.9}$$

We notice that this TM1 pattern emerges as a prediction of the UV-complete models proposed in Refs. [64, 79, 319, 320].

On the other hand, by retaining the second column of the tri-bimaximal mixing matrix one obtains the TM2 [316–318] pattern, a particular case of trimaximal mixing [252, 321–323], given as

$$U_{TM2} = \frac{1}{\sqrt{6}}\begin{pmatrix} 2\cos\theta & \sqrt{2} & 2e^{-i\delta}\sin\theta \\ -\cos\theta - \sqrt{3}\,e^{i\delta}\sin\theta & \sqrt{2} & \sqrt{3}\,\cos\theta - e^{-i\delta}\sin\theta \\ -\cos\theta + \sqrt{3}\,e^{i\delta}\sin\theta & \sqrt{2} & -\sqrt{3}\,\cos\theta - e^{-i\delta}\sin\theta \end{pmatrix}\,. \tag{3.10}$$

All the lepton mixing parameters depend on just two free parameters δ and θ,

$$\begin{aligned} &\sin^2\theta_{13} = \frac{2}{3}\sin^2\theta, \quad \sin^2\theta_{12} = \frac{1}{2+\cos 2\theta}\,, \\ &\sin^2\theta_{23} = \frac{1}{2} - \frac{\sqrt{3}\cos\delta\sin 2\theta}{2(2+\cos 2\theta)}\,, \quad J_{CP} = \frac{1}{6\sqrt{3}}\sin\delta\sin 2\theta\,, \end{aligned} \tag{3.11}$$

which implies the following correlations

$$3\sin^2\theta_{12}\cos^2\theta_{13} = 1, \qquad \tan 2\theta_{23}\cos\delta^{\ell} = \frac{\cos 2\theta_{13}}{\sin\theta_{13}\sqrt{2-3\sin^2\theta_{13}}}\,. \tag{3.12}$$

Here we note also that this TM2 neutrino mixing pattern emerges within the UV-complete models described in Refs. [61, 62, 78, 205, 323–325].

3.3 Golden ratio mixing pattern

Within the standard golden ratio (GR) mixing pattern [12], the lepton mixing angles are given by $\theta_{23} = 45^{\circ}$, $\theta_{13} = 0^{\circ}$ and $\cot\theta_{12} = \phi_g$, where $\phi_g = (1+\sqrt{5})/2$ is the golden ratio [328, 329]. The lepton mixing matrix is of the form

$$U_{GR} = \begin{pmatrix} c_{12} & s_{12} & 0 \\ -\frac{s_{12}}{\sqrt{2}} & \frac{c_{12}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{s_{12}}{\sqrt{2}} & \frac{c_{12}}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \equiv \frac{1}{\sqrt{2\sqrt{5}\phi_g}}\begin{pmatrix} \sqrt{2}\phi_g & \sqrt{2} & 0 \\ -1 & \phi_g & \sqrt{\sqrt{5}\phi_g} \\ -1 & \phi_g & -\sqrt{\sqrt{5}\phi_g} \end{pmatrix}\,. \tag{3.13}$$

[12] Another proposal for the golden ratio mixing has the solar mixing angle given as $\cos\theta_{12} = \phi_g/2$ [326, 327].

In this case, the neutrino mass matrix is given by

$$m_\nu^{GR}=\frac{m_1}{2\sqrt{5}}\begin{pmatrix}2\phi_g & -\sqrt{2} & -\sqrt{2}\\ -\sqrt{2} & -1/\phi_g & -1/\phi_g\\ -\sqrt{2} & -1/\phi_g & -1/\phi_g\end{pmatrix}+\frac{m_2}{2\sqrt{5}}\begin{pmatrix}2/\phi_g & \sqrt{2} & \sqrt{2}\\ \sqrt{2} & \phi_g & \phi_g\\ \sqrt{2} & \phi_g & \phi_g\end{pmatrix}+\frac{m_3}{2}\begin{pmatrix}0 & 0 & 0\\ 0 & 1 & -1\\ 0 & -1 & 1\end{pmatrix}\,. \tag{3.14}$$

The residual flavour symmetry transformations of m_ν^{GR} take the form

$$\begin{aligned}&G_1^{GR}=\frac{1}{\sqrt{5}}\begin{pmatrix}1 & -\sqrt{2} & -\sqrt{2}\\ -\sqrt{2} & -\phi_g & 1/\phi_g\\ -\sqrt{2} & 1/\phi_g & -\phi_g\end{pmatrix}\,,\quad G_2^{GR}=\frac{1}{\sqrt{5}}\begin{pmatrix}-1 & \sqrt{2} & \sqrt{2}\\ \sqrt{2} & -1/\phi_g & \phi_g\\ \sqrt{2} & \phi_g & -1/\phi_g\end{pmatrix}\,,\\ &G_3^{GR}=-\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}\,.\end{aligned} \tag{3.15}$$

The minimal flavour symmetry that can produce the golden ratio mixing pattern is the A_5 group [330–332]. Accordingly, a finite Z_5 subgroup is preserved by the charged lepton mass term with

$$G_e=\begin{pmatrix}1 & 0 & 0\\ 0 & e^{2\pi i/5} & 0\\ 0 & 0 & e^{-2\pi i/5}\end{pmatrix}\,. \tag{3.16}$$

In order to be phenomenologically viable, the golden-ratio pattern would certainly require a revamping, along the lines considered in section 5.

3.4 Bi-maximal mixing pattern

For the bi-maximal mixing, both solar angle and atmospheric mixing angles are maximal $\theta_{12}=\theta_{23}=45^\circ$ while the reactor angle is vanishing [333]. In the basis where the charged lepton mass matrix is diagonal, the corresponding neutrino mass matrix is given by

$$m_\nu^{BM}=\frac{1}{4}\begin{pmatrix}2\,(m_1+m_2) & \sqrt{2}\,(m_2-m_1) & \sqrt{2}\,(m_2-m_1)\\ \sqrt{2}\,(m_2-m_1) & m_1+m_2+2m_3 & m_1+m_2-2m_3\\ \sqrt{2}\,(m_2-m_1) & m_1+m_2-2m_3 & m_1+m_2+2m_3\end{pmatrix}\,. \tag{3.17}$$

The residual flavour symmetry transformations of the above neutrino mass matrix are

$$G_1^{BM}=\frac{1}{2}\begin{pmatrix}0 & -\sqrt{2} & -\sqrt{2}\\ -\sqrt{2} & -1 & 1\\ -\sqrt{2} & 1 & -1\end{pmatrix}\,,\quad G_2^{BM}=\frac{1}{2}\begin{pmatrix}0 & \sqrt{2} & \sqrt{2}\\ \sqrt{2} & -1 & 1\\ \sqrt{2} & 1 & -1\end{pmatrix}\,,\quad G_3^{BM}=-\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}\,. \tag{3.18}$$

We can choose the residual flavour symmetry of the charged lepton mass matrix to be a Z_4 subgroup with

$$G_e^{BM} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix} . \tag{3.19}$$

The group generated by G_1^{BM}, G_2^{BM}, G_3^{BM} and G_e^{BM} is also the S_4 group [334, 335]. Therefore the S_4 flavour symmetry can also be used to produce the bimaximal mixing.

Note, however, that the solar angle θ_{12} and reactor angle θ_{13} would have to undergo very large corrections in order to be compatible with current neutrino oscillation data [24, 25], making this pattern very problematic. As a result, Bi-maximal mixing should be discarded or generalized in a radical way. One of its possible radical generalizations is the Bi-large mixing pattern discussed in section 5.3.

Besides the tri-bimaximal, golden ratio and bi-maximal mixing patterns, there are other constant mixing patterns compatible with experimental data with nonzero θ_{13} [259,336–338]. They could be derived from the breaking of large flavor symmetry groups such as $\Delta(96)$, $\Delta(384)$ etc. No sizable corrections are necessary for these mixing patterns, however the required vacuum configuration and symmetry breaking are more complicated.

As shown above, the neutrino mass matrices corresponding to the simple tri-bimaximal, golden ratio and bi-maximal mixing patterns obey certain residual flavour symmetries. In the following section, we will show that the neutrino and charged lepton mass matrices generally can have both residual flavour and residual CP symmetries, and also that a residual flavour symmetry can be generated by a residual CP symmetry. In particular, residual symmetries can provide a method to revamp all mixing patterns discussed above. Indeed, in the next section we will show how to revamp them in a systematic manner by using the residual symmetry method. The generalized patterns are not only phenomenologically viable, but also predictive, since the form of the resulting lepton mixing matrices can be restricted.

4 Flavour and CP symmetries from the bottom-up

In this section, we will examine quark and lepton mass matrices that have both remnant flavour and CP symmetry. Remnant flavour symmetries can be generated by performing two remnant CP transformations in succession, with explicit forms of the remnant CP symmetry derived from the experimentally measured mixing matrix. On the other hand, the fermion mixing matrices can be constructed from the postulated residual CP transformations of the quark and lepton mass matrices. In the following, we present the remnant flavour and CP symmetries of the quark and lepton mass matrices, their parametrization and the master formula to construct the mixing matrix from the remnant CP symmetry.

4.1 Residual symmetries of leptons

In the absence of a fundamental theory of flavour we study the effect of possible remnant symmetries G_ν and G_l of the neutrino and charged lepton mass terms. Their existence may provide a model-independent approach towards underpinning the ultimate nature of the underlying theory. Let us now focus on this point.

The lepton masses terms are given in Eq. (1.12) and Eq. (1.13) for Dirac and Majorana neutrinos respectively. Since in the SM the only physical mixing matrix relates to left-handed fermions, we are concerned with the hermitian mass matrix $m_l^\dagger m_l$ connecting left-handed charged leptons on both sides. Under generic unitary transformations of the left-handed lepton fields l_L and ν_L,

$$l_L \to G_l l_L, \quad \nu_L \to G_\nu \nu_L \,, \tag{4.1}$$

where both G_l and G_ν are three-dimensional unitary matrices, the charged lepton and neutrino mass matrices transform as

$$m_l^\dagger m_l \to G_l^\dagger m_l^\dagger m_l G_l \,, \tag{4.2}$$

$$m_\nu \to G_\nu^T m_\nu G_\nu, \qquad \text{for Majorana neutrinos}\,, \tag{4.3}$$

$$m_\nu^\dagger m_\nu \to G_\nu^\dagger m_\nu^\dagger m_\nu G_\nu, \quad \text{for Dirac neutrinos}\,. \tag{4.4}$$

In order for the symmetry to hold, the mass matrices must satisfy

$$G_l^\dagger m_l^\dagger m_l G_l = m_l^\dagger m_l \,, \tag{4.5}$$

$$G_\nu^T m_\nu G_\nu = m_\nu, \qquad \text{for Majorana neutrinos}\,, \tag{4.6}$$

$$G_\nu^\dagger m_\nu^\dagger m_\nu G_\nu = m_\nu^\dagger m_\nu, \quad \text{for Dirac neutrinos}\,. \tag{4.7}$$

Applying these invariance conditions to Eqs. (1.16, 1.17, 1.18) and assuming no mass eigenvalue vanishes we obtain

$$U_l^\dagger G_l U_l = \text{diag}\left(e^{i\alpha_e}, e^{i\alpha_\mu}, e^{i\alpha_\tau}\right) \,, \tag{4.8}$$

$$U_\nu^\dagger G_\nu U_\nu = \mathrm{diag}\left(\pm 1, \pm 1, \pm 1\right), \qquad \text{for Majorana neutrinos}, \tag{4.9}$$

$$U_\nu^\dagger G_\nu U_\nu = \mathrm{diag}\left(e^{i\alpha_{\nu_e}}, e^{i\alpha_{\nu_\mu}}, e^{i\alpha_{\nu_\tau}}\right), \quad \text{for Dirac neutrinos}, \tag{4.10}$$

where $\alpha_{e,\mu,\tau}$ and $\alpha_{\nu_e,\nu_\mu,\nu_\tau}$, are arbitrary real parameters. It follows that the residual flavour symmetry transformations G_l and G_ν are of the following form [263, 265]:

$$G_l = U_l \mathrm{diag}\left(e^{i\alpha_e}, e^{i\alpha_\mu}, e^{i\alpha_\tau}\right) U_l^\dagger\,, \tag{4.11}$$

$$G_\nu = U_\nu \mathrm{diag}\left(\pm 1, \pm 1, \pm 1\right) U_\nu^\dagger, \qquad \text{for Majorana neutrinos}, \tag{4.12}$$

$$G_\nu = U_\nu \mathrm{diag}\left(e^{i\alpha_{\nu_e}}, e^{i\alpha_{\nu_\mu}}, e^{i\alpha_{\nu_\tau}}\right) U_\nu^\dagger, \quad \text{for Dirac neutrinos}. \tag{4.13}$$

One sees that the charged lepton mass term generically admits a $U(1) \times U(1) \times U(1)$ remnant flavour symmetry. For the neutrino mass matrix the remnant flavour symmetry depends on the nature of neutrinos. For the case of Dirac neutrinos it is also $U(1) \times U(1) \times U(1)$ [263]. For Majorana neutrinos, the eight possible choices of G_ν in Eq. (4.12) correspond to $Z_2 \times Z_2 \times Z_2$. Notice that G_ν and $-G_\nu$ should be identified as the same residual flavour symmetry transformation, since the minus sign can be absorbed as a neutrino field redefinition. Indeed, they both lead to the same constraint on the neutrino mass matrix. We are therefore left with four possible solutions for G_ν, which can be chosen as [269–272],

$$G_i = U_\nu\, d_i U_\nu^\dagger, \qquad i = 1, 2, 3, 4\,, \tag{4.14}$$

where the d_i are given as

$$\begin{aligned} d_1 &= \mathrm{diag}\left(1, -1, -1\right), \qquad & d_2 &= \mathrm{diag}\left(-1, 1, -1\right), \\ d_3 &= \mathrm{diag}\left(-1, -1, 1\right), \qquad & d_4 &= \mathrm{diag}\left(1, 1, 1\right)\,. \end{aligned} \tag{4.15}$$

One sees that G_4 is simply the trivial identity matrix, and we can further check that

$$G_i^2 = 1, \qquad G_i G_j = G_j G_i = G_k \quad \text{with} \quad i \neq j \neq k \neq 4. \tag{4.16}$$

It follows that the residual flavour symmetry of the Majorana neutrino mass matrix is a Klein group isomorphic to $Z_2 \times Z_2$. In the flavour basis where the charged lepton mass matrix m_l is diagonal, U_l would be trivial, so that the lepton mixing comes just from the neutrino sector, i.e. $U = U_\nu$. Hence the residual symmetries of the neutrino and charged lepton mass matrices are determined in terms of the mixing angles and CP violation phases contained in the mixing matrix.

An important observation is that, besides the above residual flavour symmetry, the lepton mass matrices can have residual CP (or charge conjugation parity) symmetry. The CP transformation properties of the left-handed neutrino and charged lepton fields are given by

$$l_L(x) \overset{CP}{\longmapsto} i X_l \gamma^0 C \bar{l}_L^T(\mathcal{P}x), \quad \nu_L(x) \overset{CP}{\longmapsto} i X_\nu \gamma^0 C \bar{\nu}_L^T(\mathcal{P}x)\,, \tag{4.17}$$

where $\mathcal{P}x = (t, -\vec{x})$, C is the charge-conjugation matrix [94–96], X_l and X_ν are 3×3 unitary matrices acting in family space. Notice that the matrices X_l and X_ν generalize the conventional CP transformation prescription. As a result, they are referred to in the literature as generalized CP transformations [264, 339–344].

Requiring that X_l and X_ν are symmetries of the lepton mass terms in Eqs. (1.12, 1.13) implies that the lepton mass matrices m_l and m_ν should satisfy [265, 345]

$$X_l^\dagger m_l^\dagger m_l X_l = \left(m_l^\dagger m_l\right)^* , \tag{4.18}$$

$$X_\nu^T m_\nu X_\nu = m_\nu^* , \qquad \text{for Majorana neutrinos} , \tag{4.19}$$

$$X_\nu^\dagger m_\nu^\dagger m_\nu X_\nu = \left(m_\nu^\dagger m_\nu\right)^* , \quad \text{for Dirac neutrinos} . \tag{4.20}$$

Plugging Eqs. (1.16, 1.17, 1.18) into the above invariance conditions, one sees that the unitary transformations U_l and U_ν must be subject to the following conditions

$$U_l^\dagger X_l U_l^* = \text{diag}\left(e^{i\beta_e}, e^{i\beta_\mu}, e^{i\beta_\tau}\right) , \tag{4.21}$$

$$U_\nu^\dagger X_\nu U_\nu^* = \text{diag}\left(\pm 1, \pm 1, \pm 1\right) , \qquad \text{for Majorana neutrinos} , \tag{4.22}$$

$$U_\nu^\dagger X_\nu U_\nu^* = \text{diag}\left(e^{i\beta_{\nu_e}}, e^{i\beta_{\nu_\mu}}, e^{i\beta_{\nu_\tau}}\right) , \quad \text{for Dirac neutrinos} , \tag{4.23}$$

where $\beta_{e,\mu,\tau}$ and $\beta_{\nu_e,\nu_\mu,\nu_\tau}$ are real free parameters. Hence the residual CP transformations X_l and X_ν should be of the form [263–268]

$$X_l = U_l \text{diag}\left(e^{i\beta_e}, e^{i\beta_\mu}, e^{i\beta_\tau}\right) U_l^T , \tag{4.24}$$

$$X_\nu = U_\nu \text{diag}\left(\pm 1, \pm 1, \pm 1\right) U_\nu^T , \qquad \text{for Majorana neutrinos} , \tag{4.25}$$

$$X_\nu = U_\nu \text{diag}\left(e^{i\beta_{\nu_e}}, e^{i\beta_{\nu_\mu}}, e^{i\beta_{\nu_\tau}}\right) U_\nu^T , \quad \text{for Dirac neutrinos} . \tag{4.26}$$

Clearly, both X_l and X_ν are unitary and symmetric matrices [264, 265]

$$X_l = X_l^T , \quad X_\nu = X_\nu^T , \quad X_l X_l^\dagger = X_\nu X_\nu^\dagger = 1 . \tag{4.27}$$

From the expressions of remnant flavor symmetry in Eqs. (4.11, 4.12, 4.13) and remnant CP transformations in Eqs. (4.24, 4.25, 4.26), one can check that the residual flavour and CP symmetries satisfy the following restricted consistency conditions

$$\begin{aligned} & X_l G_l^* X_l^{-1} = G_l^{-1} , \\ & X_\nu G_\nu^* X_\nu^{-1} = G_\nu , \quad \text{for Majorana neutrinos} , \\ & X_\nu G_\nu^* X_\nu^{-1} = G_\nu^{-1} , \quad \text{for Dirac neutrinos} . \end{aligned} \tag{4.28}$$

If we successively perform two CP transformations on the left-handed charged lepton fields, characterized

by $X_l = U_l \text{diag}\left(e^{i\beta_e}, e^{i\beta_\mu}, e^{i\beta_\tau}\right) U_l^T$ and $X_l' = U_l \text{diag}\left(e^{i\beta'_e}, e^{i\beta'_\mu}, e^{i\beta'_\tau}\right) U_l^T$ we obtain [13]

$$l_L(x) \overset{CP}{\longmapsto} iX_l\gamma^0 C\bar{l}_L^T(\mathcal{P}x) \overset{CP}{\longmapsto} X_l X_l'^* l_L(x)\,, \tag{4.29}$$

with

$$X_l X_l'^* = U_l \text{diag}\left(e^{i(\beta_e-\beta'_e)}, e^{i\left(\beta_\mu-\beta'_\mu\right)}, e^{i(\beta_\tau-\beta'_\tau)}\right) U_l^\dagger\,. \tag{4.30}$$

Moreover, from the invariance condition of $m_l^\dagger m_l$ under the residual CP symmetry, Eq. (4.18), it is easy to show that

$$X_l'^T X_l^\dagger m_l^\dagger m_l X_l X_l'^* = m_l^\dagger m_l\,. \tag{4.31}$$

This means that performing two CP transformations in succession is equivalent to a flavour symmetry transformation $X_l X_l'^* \equiv G_l$ [265]. The same conclusion also holds true for neutrinos if they are Dirac particles.

For the case of Majorana neutrinos, there are eight possibilities for X_ν. In this case one should take X_ν and $-X_\nu$ as the same CP transformation, since the minus sign can be absorbed into the neutrino fields since the mass term involves a product of two neutrino fields. As a result only four of them are relevant. Without loss of generality they can be chosen to be [265–267]

$$X_i = U_\nu d_i U_\nu^T, \quad i = 1, 2, 3, 4\,, \tag{4.32}$$

with d_i given in Eq. (4.15). The remaining four can be obtained from the above by multiplying an overall -1 factor. It is straightforward to check that the neutrino mass matrix m_ν satisfies

$$X_j^\dagger X_i^T m_\nu X_i X_j^* = m_\nu\,. \tag{4.33}$$

As a result, remnant flavour symmetries can be generated by remnant CP symmetries as well. Explicitly, we have the following relations [265]:

$$\begin{aligned}
&X_2X_3^* = X_3X_2^* = X_4X_1^* = X_1X_4^* = G_1,\\
&X_1X_3^* = X_3X_1^* = X_4X_2^* = X_2X_4^* = G_2,\\
&X_1X_2^* = X_2X_1^* = X_4X_3^* = X_3X_4^* = G_3,\\
&X_1X_1^* = X_2X_2^* = X_3X_3^* = X_4X_4^* = G_4 = 1\,.
\end{aligned} \tag{4.34}$$

As a result, once we impose a set of generalized CP transformations, there is always an associated flavour symmetry. Furthermore, Eq. (4.34) implies that any residual CP transformation can be expressed in terms of the remaining ones as follows [265],

$$X_i = X_j X_m^* X_n, \qquad i \neq j \neq m \neq n\,. \tag{4.35}$$

[13] Notice that an overall minus sign is dropped in the last step, as it can be absorbed into the lepton field.

In other words, only three of the four remnant CP transformations are independent. In the flavour basis where m_l is diagonal, the residual CP transformation X_l is an arbitrary diagonal phase matrix, while $X_\nu = U\text{diag}\left(\pm 1, \pm 1, \pm 1\right) U^T$. Hence the remnant CP symmetry can be constructed from the neutrino mixing matrix, and its explicit form can be determined more precisely with the improved measurement of the mixing angles and CP phases [24, 25].

Here we have so far assumed that the three light neutrino masses are non-vanishing. If the lightest neutrino is massless one can analyze the residual symmetry of the neutrino mass matrix in the same way. This interesting situation occurs generically in the "missing partner" seesaw mechanism [30] in which there are less "right" than "left"-handed neutrinos [14] (a massless neutrino also emerges in theories where anti-symmetric Yukawa couplings are involved in generating neutrino mass [130]). Note that an incomplete fermion multiplet structure [15] can also be interesting in cosmology [134–138].

With one neutrino massless one has $m_1 = 0$ for normal ordered (NO) neutrino mass spectrum, and $m_3 = 0$ for the inverted ordered (IO) spectrum. If neutrinos have Dirac nature, one finds that the residual CP transformation X_ν would still be given by Eq. (4.26). On the other hand, if neutrinos are Majorana particles, one must use Eq. (4.25) replacing the diagonal entry "± 1" in the position (11) for NO and (33) for IO with an arbitrary phase factor [348], i.e.

$$X_\nu = \begin{cases} U_\nu \text{diag}\left(e^{i\beta}, \pm 1, \pm 1\right) U_\nu^T, & \text{for NO}\,, \\ U_\nu \text{diag}\left(\pm 1, \pm 1, e^{i\beta}\right) U_\nu^T, & \text{for IO}\,, \end{cases} \tag{4.36}$$

where β is real. Similarly the residual flavour symmetry G_ν becomes [265, 348]

$$G_\nu = \begin{cases} U_\nu \text{diag}\left(e^{i\alpha}, \pm 1, \pm 1\right) U_\nu^\dagger, & \text{for NO}\,, \\ U_\nu \text{diag}\left(\pm 1, \pm 1, e^{i\alpha}\right) U_\nu^\dagger, & \text{for IO}\,, \end{cases} \tag{4.37}$$

with real α. Hence the generalized CP and residual flavour symmetry groups have the structure $U(1) \times Z_2$ (instead of $Z_2 \times Z_2$) modulo a possible overall factor -1 in this case. On the other hand, this kind of residual symmetry can enforce a massless Majorana neutrino. The breaking of finite discrete groups into this form of G_ν in the neutrino sector was analyzed in [349–351], where the angle α was a rational multiple of π.

4.2 Reconstructing lepton mixing from remnant CP symmetry

As shown in above, residual CP symmetries can be derived from the mixing matrix, and conversely, the lepton mixing matrix can be constructed from the remnant CP symmetries in the neutrino and the charged lepton sectors. In concrete models, we can start from a set of CP transformations $\mathcal{X}_{\mathcal{CP}}$ respected by the Lagrangian at some high energy scale. Subsequently $\mathcal{X}_{\mathcal{CP}}$ is spontaneously broken by some scalar fields into different remnant symmetries in the neutrino and the charged lepton sectors. The misalignment

[14] The present neutrino data allows the possibility that the lightest neutrino is massless, consequently at least two right-handed neutrinos are necessary in type-I seesaw [131–133].

[15] For a discussion of seesaw leptogenesis with two and three right-handed neutrinos see [346, 347].

between the two remnant symmetries is responsible for the mismatch of the rotations which diagonalize the neutrino and charged lepton matrices, leading to the lepton mixing matrix. We now present the general parametrization for the remnant CP symmetries of the neutrino and charged lepton sector, and the corresponding restrictions on the unitary transformations U_ν and U_l.

We start from the simplest nontrivial case in which a single remnant CP transformation X_R is preserved by the neutrino mass matrix. As shown in section 4.1, X_R should be a symmetric unitary matrix, otherwise the light neutrino masses would be degenerate. Thus X_R can be parameterized as follows [265]:

$$X_R = e^{i\kappa_1} v_1 v_1^T + e^{i\kappa_2} v_2 v_2^T + e^{i\kappa_3} v_3 v_3^T \,, \tag{4.38}$$

where the phases κ_1, κ_2 and κ_3 can be taken in the range of 0 and 2π without loss of generality, v_1, v_2 and v_3 are mutually orthogonal vectors with

$$\begin{aligned} v_1 &= \begin{pmatrix} \cos\varphi \\ \sin\varphi\cos\phi \\ \sin\varphi\sin\phi \end{pmatrix} \,, \quad v_2 = \begin{pmatrix} \sin\varphi\cos\rho \\ -\sin\phi\sin\rho - \cos\varphi\cos\phi\cos\rho \\ \cos\phi\sin\rho - \cos\varphi\sin\phi\cos\rho \end{pmatrix} \,, \\ v_3 &= \begin{pmatrix} \sin\varphi\sin\rho \\ \sin\phi\cos\rho - \cos\varphi\cos\phi\sin\rho \\ -\cos\phi\cos\rho - \cos\varphi\sin\phi\sin\rho \end{pmatrix} \,. \end{aligned} \tag{4.39}$$

Invariance of the neutrino mass matrix under X_R implies that U_ν should be subject to the constraint in Eq. (4.22) and Eq. (4.23) for Majorana and Dirac neutrinos. As a consequence, U_ν is fixed to be [265]

$$U_\nu = (v_1, v_2, v_3)\,\mathrm{diag}\big(e^{i\frac{\kappa_1}{2}}, e^{i\frac{\kappa_2}{2}}, e^{i\frac{\kappa_3}{2}}\big) O_3(\theta_1,\theta_2,\theta_3) Q_\nu \,, \tag{4.40}$$

where O_3 is a generic real orthogonal matrix,

$$O_3(\theta_1,\theta_2,\theta_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & \sin\theta_1 \\ 0 & -\sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} \cos\theta_3 & \sin\theta_3 & 0 \\ -\sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \,, \tag{4.41}$$

where the rotation angles $\theta_{1,2,3}$ are free. The matrix Q_ν is diagonal, and it entries are ± 1 and $\pm i$ which encode the CP parity of the neutrinos, while it is unphysical for Dirac neutrinos.

If two remnant CP transformations X_{R1} and X_{R2} out of the original CP symmetry are preserved in the neutrino sector, they can generally written as [265]

$$\begin{aligned} X_{R1} &= e^{i\kappa_1} v_1 v_1^T + e^{i\kappa_2} v_2 v_2^T + e^{i\kappa_3} v_3 v_3^T \,, \\ X_{R2} &= e^{i\kappa_1} v_1 v_1^T - e^{i\kappa_2} v_2 v_2^T - e^{i\kappa_3} v_3 v_3^T \,, \end{aligned} \tag{4.42}$$

for Majorana neutrinos [16]. A remnant flavour transformation G_R can be obtained from X_{R1} and X_{R2} as

$$G_R = X_{R1}X_{R2}^* = X_{R2}X_{R1}^* = 2v_1v_1^T - 1\,, \tag{4.43}$$

which satisfies $G_R^2 = 1$. Hence a remnant Z_2 flavor symmetry generated by G_R is induced and it fixes one column of U_ν to be v_1. Besides the parameters characterizing the remnant CP symmetry, U_ν is determined just by a free rotation angle θ [265],

$$U_\nu = (v_1, v_2, v_3)\,\mathrm{diag}\big(e^{i\frac{\kappa_1}{2}}, e^{i\frac{\kappa_2}{2}}, e^{i\frac{\kappa_3}{2}}\big)R_{23}(\theta)P_\nu Q_\nu\,, \tag{4.44}$$

where $R_{23}(\theta)$ denotes a rotation matrix through an angle θ in the (23)-plane with $0 \leq \theta < \pi$,

$$R_{23}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}\,. \tag{4.45}$$

Since the remnant symmetry can not constrain the ordering of the light neutrino mass eigenvalues, U_ν is determined up to independent column permutations, and consequently P_ν is a generic permutation matrix which can take six possible forms 1, P_{12}, P_{13}, P_{23}, $P_{23}P_{12}$, $P_{23}P_{13}$ with

$$P_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\,, \quad P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\,, \quad P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\,. \tag{4.46}$$

We see that the unitary transformations U_ν for the two permutations P_ν and $P_{23}P_\nu$ are related by the redefinitions $\theta \to \theta - \pi/2$ and $Q_\nu \to P_\nu^T\mathrm{diag}(1,1,-1)P_\nu Q_\nu$. Hence only three inequivalent permutations of the columns are relevant in this case. In other words, the fixed vector $(\cos\varphi, \sin\varphi\cos\phi, \sin\varphi\sin\phi)^T$ can be the first column, the second column or the third column of the matrix U_ν.

For the case of Majorana neutrinos, we consider the scenario that all independent remnant CP transformations are preserved by the neutrino mass matrix. The remnant CP transformations can be parameterized as [267]

$$\begin{aligned} X_{R1} &= e^{i\kappa_1}v_1v_1^T + e^{i\kappa_2}v_2v_2^T + e^{i\kappa_3}v_3v_3^T\,, \\ X_{R2} &= e^{i\lambda_1}v_1v_1^T + e^{i\lambda_2}w_2w_2^T + e^{i\lambda_3}w_3w_3^T\,, \\ X_{R3} &= e^{i\lambda_1}v_1v_1^T - e^{i\lambda_2}w_2w_2^T - e^{i\lambda_3}w_3w_3^T\,, \\ X_{R4} &= e^{i\kappa_1}v_1v_1^T - e^{i\kappa_2}v_2v_2^T - e^{i\kappa_3}v_3v_3^T\,, \end{aligned} \tag{4.47}$$

[16] For Dirac neutrinos, the neutrino mixing matrix U_ν would be completely fixed up to column permutations, if the order of the remnant flavour transformation $G_R = X_{R1}X_{R2}^* = X_{R2}X_{R1}^*$ is greater than or equal to three, so as to distinguish the three families.

where v_1, w_2 and w_3 also form another set of real orthonormal vectors with

$$w_2 = \cos\xi v_2 - \sin\xi v_3, \qquad w_3 = \sin\xi v_2 + \cos\xi v_3\,, \tag{4.48}$$

and the phases $e^{i\lambda_1}$, $e^{i\lambda_2}$ and $e^{i\lambda_3}$ are given by

$$e^{i\lambda_1} = -e^{i\kappa_1}, \quad e^{i\lambda_2} = -\frac{e^{i\kappa_2}\cos^2\xi + e^{i\kappa_3}\sin^2\xi}{\left|e^{i\kappa_2}\cos^2\xi + e^{i\kappa_3}\sin^2\xi\right|}, \quad e^{i\lambda_3} = \frac{e^{i\kappa_2}\sin^2\xi + e^{i\kappa_3}\cos^2\xi}{\left|e^{i\kappa_2}\sin^2\xi + e^{i\kappa_3}\cos^2\xi\right|}\,. \tag{4.49}$$

A remnant Klein four flavour symmetry $K_4 \equiv \{1, G_{R1}, G_{R2}, G_{R3}\}$ can be generated by performing two CP transformations, and the three nontrivial residual flavour symmetry transformations G_{Ri} for $i = 1, 2, 3$ can be expressed as

$$\begin{aligned}
G_{R1} &= X_{R1}X^*_{R4} = v_1 v_1^T - v_2 v_2^T - v_3 v_3^T\,,\\
G_{R2} &= X_{R1}X^*_{R3} = -v_1 v_1^T - c_{22} v_2 v_2^T - c_{33} v_3 v_3^T - c_{23} v_2 v_3^T - c_{32} v_3 v_2^T\,,\\
G_{R3} &= X_{R1}X^*_{R2} = -v_1 v_1^T + c_{22} v_2 v_2^T + c_{33} v_3 v_3^T + c_{23} v_2 v_3^T + c_{32} v_3 v_2^T\,,
\end{aligned} \tag{4.50}$$

with

$$c_{22} = -c_{33} = -\frac{\cos 2\xi}{\left|e^{i\kappa_2}\cos^2\xi + e^{i\kappa_3}\sin^2\xi\right|}\,, \quad c_{23} = c^*_{32} = \frac{\cos\left(\frac{\kappa_2-\kappa_3}{2}\right) e^{i\frac{\kappa_2-\kappa_3}{2}}\sin 2\xi}{\left|e^{i\kappa_2}\cos^2\xi + e^{i\kappa_3}\sin^2\xi\right|}\,. \tag{4.51}$$

In this case, the neutrino mixing matrix U_ν is completely determined by the remnant CP transformations without additional free parameters [267],

$$U_\nu = (v_1, v_2, v_3)\,\mathrm{diag}\big(e^{i\frac{\kappa_1}{2}}, e^{i\frac{\kappa_2}{2}}, e^{i\frac{\kappa_3}{2}}\big) R_{23}(\chi) P_\nu Q_\nu\,, \tag{4.52}$$

where the angle χ fulfills

$$\tan 2\chi = \cos\left(\frac{\kappa_2 - \kappa_3}{2}\right)\tan 2\xi\,. \tag{4.53}$$

Notice that the lepton mixing angles and the three CP violating phases are completely fixed by the remnant CP symmetry in this case. Moreover, note that the master *formulae* of Eqs. (4.40, 4.44, 4.52) for U_ν hold irrespective of how the remnant CP symmetry is dynamically realized.

4.3 Residual symmetries of quarks

In this section, we turn to the residual flavour and CP symmetries of the quark mass matrices. The Lagrangian for the quark masses is given in Eq. (1.2). One can easily check that the hermitian combinations $m_U^\dagger m_U$ and $m_D^\dagger m_D$ are invariant under the following unitary transformations [263],

$$U_L \to G_u U_L, \quad D_L \to G_d D_L\,, \tag{4.54}$$

with

$$G_u = V_u \mathrm{diag}\left(e^{i\alpha_u}, e^{i\alpha_c}, e^{i\alpha_t}\right) V_u^\dagger, \quad G_d = V_d \mathrm{diag}\left(e^{i\alpha_d}, e^{i\alpha_s}, e^{i\alpha_b}\right) V_d^\dagger\,, \tag{4.55}$$

where α_q $(q = u, d, c, s, t, b)$ are arbitrary phase parameters. Hence the following equalities are satisfied,

$$G_u^\dagger m_U^\dagger m_U G_u = m_U^\dagger m_U, \quad G_d^\dagger m_D^\dagger m_D G_d = m_D^\dagger m_D \,. \tag{4.56}$$

In other words, both the up-quark mass matrix m_U and down-quark mass matrix m_D have a residual $U(1) \times U(1) \times U(1)$ flavour symmetry [263]. Notice that the same conclusion holds true for any Dirac fermion mass matrix, hence it also applies for the charged lepton mass term.

Let us now turn to the discussion of residual CP symmetries of the quark mass term. Assuming the left-handed quarks U_L and D_L transform as

$$U_L(x) \overset{CP}{\longmapsto} iX_u\gamma^0 C\overline{U}_L^T(\mathcal{P}x), \quad D_L(x) \overset{CP}{\longmapsto} iX_d\gamma^0 C\overline{D}_L^T(\mathcal{P}x)\,, \tag{4.57}$$

this is a symmetry of the quark mass matrices m_U and m_D if and only if $m_U^\dagger m_U$ and $m_D^\dagger m_D$ fulfill the conditions

$$X_u^\dagger m_U^\dagger m_U X_u = (m_U^\dagger m_U)^\star, \quad X_d^\dagger m_D^\dagger m_D X_d = (m_D^\dagger m_D)^\star \,. \tag{4.58}$$

Similarly to the charged lepton sector, one can show that X_u and X_d must take the following form

$$X_u = V_u \mathrm{diag}\left(e^{i\beta_u}, e^{i\beta_c}, e^{i\beta_t}\right) V_u^T, \quad X_d = V_d \mathrm{diag}\left(e^{i\beta_d}, e^{i\beta_s}, e^{i\beta_b}\right) V_d^T \,, \tag{4.59}$$

where β_q $(q = u, d, c, s, t, b)$ are real. It is easy to see that both X_u and X_d are unitary symmetric matrices, and that residual flavour symmetries G_u and G_d will be generated by X_u and X_d, respectively. In the basis where the up-quark mass matrix is diagonal, V_u is diagonal, so that V_d coincides with the CKM matrix, $V_d = V_{CKM}$. As a result, the explicit form of the remnant flavour symmetries G_u, G_d and the remnant CP symmetries X_u, X_d can be fixed in terms of the measured values of the CKM matrix elements [28]. As in section 4.2 for lepton sector, the quark mixing matrix can also be fixed by the remnant CP symmetry.

So far we have shown that lepton and quark mass terms generically admit residual flavour and CP symmetries. These are associated to the charged current flavour mixing matrix. One may ask about the origin of these residual symmetries. It is well known that the spontaneous breaking of the $\mathrm{SU(3)_c \otimes SU(2)_L \otimes U(1)_Y}$ gauge symmetry preserves the U(1) gauge symmetry associated to electromagnetism. Likewise, one may conjecture that at some high energy scale the true underlying theory has a flavour and CP symmetry which is subsequently broken down to the residual symmetry or a subgroup of it. This gives a further motivation for introducing flavour and CP symmetries, i.e. to explain flavour mixing and CP violation in a model-independent manner. We also saw that the residual flavour symmetry of the charged leptons and quarks should be contained in $U(1) \times U(1) \times U(1)$. The same holds for neutrinos if the are Dirac particles. In contrast, the residual flavour symmetry of the neutrino sector should be a subgroup of $Z_2 \times Z_2 \times Z_2$ if they are Majorana fermions.

In a pioneering work [276], Froggatt and Nielsen originally took the $G_f = U(1)$ flavor symmetry in order to explain the quark mass ratios and the CKM mixing angles, which are expressed as powers of small G_f breaking parameters. This is the so-called FN mechanism. Concerning the lepton sector,

broken flavor symmetries based on non-abelian discrete groups were found to reproduce certain interesting mixing patterns, such as TBM as a first order approximation [53, 54, 70, 313]. Continuous Lie groups $U(2)$ [277,352–354], $SO(3)$ [287,355,356] and $SU(3)$ [293,357–359] as flavor symmetries were also studied to address the flavor puzzles. Typically, they must be strongly broken.

Besides the use of full-fledged flavor symmetries, there are more phenomenological approaches to the flavor puzzle involving, e.g., the use of texture zeros in the fermion mass matrices [360–363]. Alternatively the assumption of lepton mixing "anarchy" [51,52]. In the present review, we shall focus on the approach of discrete flavor symmetry as well as generalized CP symmetry.

5 Viable lepton mixing patterns

The "constant" lepton mixing patterns discussed in section 3, such as the tri-bimaximal, golden ratio and bi-maximal mixing, characterized by numerical predictions for the mixing angles and phase, are all ruled out by neutrino oscillation data [24, 25], especially by the precise measurement of the "reactor angle" θ_{13} [305, 306]. They all need to be "revamped" in order to be compatible with experimental data and to enable meaningful theoretical predictions for CP violation.

In this section we assume that neutrinos are Majorana particles and show how the imposition of residual flavour and CP symmetries [257, 258, 265, 364–366] can be used to produce systematic generalizations of the patterns discussed in the section 3. Indeed, imposing the residual symmetries G_i in Sec. 4 fixes the i-th column of the mixing matrix. In this way one can obtain generalized patterns which can be not only viable, but also predictive, in which the mixing matrix is described by just a few parameters. This model-independent approach of predicting new mixing patterns holds irrespective of how the relevant mass matrices arise from first principles. We now describe some examples.

5.1 Revamped TBM mixing

Working in the charged lepton diagonal basis, we start our discussion with the "complex TBM" matrix (cTBM) [367], which is given by

$$U_{cTBM} = \begin{pmatrix} \sqrt{\frac{2}{3}} & \frac{e^{-i\rho}}{\sqrt{3}} & 0 \\ -\frac{e^{i\rho}}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{e^{-i\sigma}}{\sqrt{2}} \\ \frac{e^{i(\rho+\sigma)}}{\sqrt{6}} & -\frac{e^{i\sigma}}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{pmatrix} . \tag{5.1}$$

This cTBM mixing matrix predicts the same mixing angles as the usual real TBM pattern in Eq. (3.2), though the Majorana phases are non-vanishing. Within the symmetrical parametrization of the lepton mixing matrix in Eqs. (1.21) and (1.22) they are given by

$$\phi_{12} = \rho, \qquad \phi_{23} = \sigma \, . \tag{5.2}$$

The TBM matrix in Eq. (3.1) corresponds to zero Majorana phases $\rho = \sigma = 0$ and hence we call it real TBM [17]. From Eq. (4.32), we know the four CP symmetry matrices $X_{1,2,3,4}$ associated with the cTBM mixing pattern are

$$X_i = U_{cTBM} d_i U_{cTBM}^T \, , \tag{5.3}$$

[17] The minus sign of the third row is absorbed into the charged leptons.

where $d_{1,2,3,4}$ are diagonal matrices with entries ± 1 given in Eq. (4.15). Thus the above four CP symmetries are given in matrix form as

$$
\begin{aligned}
X_1 &= \frac{1}{6}\begin{pmatrix} 4-2e^{-2i\rho} & -2e^{-i\rho}-2e^{i\rho} & 2e^{i(\rho+\sigma)}+2e^{-i(\rho-\sigma)} \\ -2e^{-i\rho}-2e^{i\rho} & -2+e^{2i\rho}-3e^{-2i\sigma} & -3e^{-i\sigma}-e^{i(2\rho+\sigma)}+2e^{i\sigma} \\ 2e^{i(\rho+\sigma)}+2e^{-i(\rho-\sigma)} & -3e^{-i\sigma}-e^{i(2\rho+\sigma)}+2e^{i\sigma} & -3+e^{2i(\rho+\sigma)}-2e^{2i\sigma} \end{pmatrix}, \\
X_2 &= \frac{1}{6}\begin{pmatrix} -4+2e^{-2i\rho} & 2e^{-i\rho}+2e^{i\rho} & -2e^{i(\rho+\sigma)}-2e^{-i(\rho-\sigma)} \\ 2e^{-i\rho}+2e^{i\rho} & 2-e^{2i\rho}-3e^{-2i\sigma} & -3e^{-i\sigma}+e^{i(2\rho+\sigma)}-2e^{i\sigma} \\ -2e^{i(\rho+\sigma)}-2e^{-i(\rho-\sigma)} & -3e^{-i\sigma}+e^{i(2\rho+\sigma)}-2e^{i\sigma} & -3-e^{2i(\rho+\sigma)}+2e^{2i\sigma} \end{pmatrix}, \\
X_3 &= \frac{1}{6}\begin{pmatrix} -4-2e^{-2i\rho} & -2e^{-i\rho}+2e^{i\rho} & -2e^{i(\rho+\sigma)}+2e^{-i(\rho-\sigma)} \\ -2e^{-i\rho}+2e^{i\rho} & -2-e^{2i\rho}+3e^{-2i\sigma} & 3e^{-i\sigma}+e^{i(2\rho+\sigma)}+2e^{i\sigma} \\ -2e^{i(\rho+\sigma)}+2e^{-i(\rho-\sigma)} & 3e^{-i\sigma}+e^{i(2\rho+\sigma)}+2e^{i\sigma} & 3-e^{2i(\rho+\sigma)}-2e^{2i\sigma} \end{pmatrix}, \\
X_4 &= \frac{1}{6}\begin{pmatrix} 4+2e^{-2i\rho} & 2e^{-i\rho}-2e^{i\rho} & 2e^{i(\rho+\sigma)}-2e^{-i(\rho-\sigma)} \\ 2e^{-i\rho}-2e^{i\rho} & 2+e^{2i\rho}+3e^{-2i\sigma} & 3e^{-i\sigma}-e^{i(2\rho+\sigma)}-2e^{i\sigma} \\ 2e^{i(\rho+\sigma)}-2e^{-i(\rho-\sigma)} & 3e^{-i\sigma}-e^{i(2\rho+\sigma)}-2e^{i\sigma} & 3+e^{2i(\rho+\sigma)}+2e^{2i\sigma} \end{pmatrix}. \qquad (5.4)
\end{aligned}
$$

The CP symmetries corresponding to the "standard" real TBM matrix of Eq. (3.1) are obtained simply by taking the limit of $\rho, \sigma \to 0$ in Eq. (5.4). These CP symmetries are therefore given by

$$
\begin{aligned}
X_1 &= \frac{1}{3}\begin{pmatrix} 1 & -2 & 2 \\ -2 & -2 & -1 \\ 2 & -1 & -2 \end{pmatrix}, \qquad X_2 = \frac{1}{3}\begin{pmatrix} -1 & 2 & -2 \\ 2 & -1 & -2 \\ -2 & -2 & -1 \end{pmatrix}, \\
X_3 &= \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad X_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (5.5)
\end{aligned}
$$

As shown in Eq. (4.34), the residual flavour symmetry can be generated by the CP transformations,

$$
\begin{aligned}
&G_1 = X_2 X_3^* = X_3 X_2^* = X_4 X_1^* = X_1 X_4^*, \quad G_2 = X_1 X_3^* = X_3 X_1^* = X_4 X_2^* = X_2 X_4^*, \\
&G_3 = X_1 X_2^* = X_2 X_1^* = X_4 X_3^* = X_3 X_4^*, \quad G_4 = X_1 X_1^* = X_2 X_2^* = X_3 X_3^* = X_4 X_4^*. \qquad (5.6)
\end{aligned}
$$

It is our goal here to obtain generalized but restricted forms for the mixing matrices starting from the "original" ones by exploiting residual flavour and CP symmetries. Notice that only three of the four CP and flavour symmetries are really independent [265, 267]. If any three of the four CP symmetries in Eq. (5.4) are imposed simultaneously, the neutrino mixing matrix would be the cTBM matrix in Eq. (5.1) with $\theta_{13} = 0$. Therefore, we will impose only two or only one of these CP symmetries, so that realistic mixing patterns with non-vanishing θ_{13} and CP violation are obtained.

5.1.1 Case a: G_1 flavour and X_1, X_4 CP symmetries

The requirement that the CP transformations X_1 and X_4 are symmetries of the neutrino mass matrix m_ν implies that the G_1 flavour symmetry is preserved and m_ν satisfies

$$X_1^T m_\nu X_1 = m_\nu^*, \qquad X_4^T m_\nu X_4 = m_\nu^* \,. \tag{5.7}$$

Consequently the light neutrino mass matrix is of the following form

$$m'_\nu = U_{cTBM}^T m_\nu U_{cTBM} = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & \delta m \\ 0 & \delta m & m_3 \end{pmatrix} , \tag{5.8}$$

where the parameters m_1, m_2, m_3 and δm are real. The mass matrix m'_ν can be diagonalized by a real orthogonal matrix $R_{23}(\theta)$ given by

$$R_{23}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \quad \text{with} \quad \tan 2\theta = \frac{2\,\delta m}{m_3 - m_2} \,. \tag{5.9}$$

As a result, in this case the lepton mixing matrix is given as

$$\begin{aligned} U &= U_{cTBM}\, R_{23}\, Q_\nu \\ &= \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & \sqrt{2}e^{-i\rho}\cos\theta & \sqrt{2}e^{-i\rho}\sin\theta \\ -e^{i\rho} & \sqrt{2}\cos\theta - \sqrt{3}e^{-i\sigma}\sin\theta & \sqrt{2}\sin\theta + \sqrt{3}e^{-i\sigma}\cos\theta \\ e^{i(\rho+\sigma)} & -\sqrt{3}\sin\theta - \sqrt{2}e^{i\sigma}\cos\theta & \sqrt{3}\cos\theta - \sqrt{2}e^{i\sigma}\sin\theta \end{pmatrix} Q_\nu \,, \end{aligned} \tag{5.10}$$

where $Q_\nu = \mathrm{diag}(e^{ik_1\pi/2}, e^{ik_2\pi/2}, e^{ik_3\pi/2})$ is a diagonal unitary matrix with $k_{1,2,3} = 0, 1, 2, 3$. The entries ± 1 and $\pm i$ encode the CP parities of the neutrino states and render the neutrino mass eigenvalues non-negative. From Eq. (5.10) one can then extract the expressions of lepton mixing angles and CP violating phases as follows,

$$\begin{aligned} &\sin^2\theta_{13} = \frac{\sin^2\theta}{3}, \quad \sin^2\theta_{12} = \frac{\cos^2\theta}{\cos^2\theta + 2}, \quad \sin^2\theta_{23} = \frac{1}{2} + \frac{\sqrt{6}\sin 2\theta\cos\sigma}{2\cos^2\theta + 4} \,, \\ &\sin\delta^\ell = -\frac{\mathrm{sign}(\sin 2\theta)(\cos^2\theta + 2)\sin\sigma}{\sqrt{(\cos^2\theta+2)^2 - 6\sin^2 2\theta\cos^2\sigma}} \,, \quad \tan\delta^\ell = \frac{2 + \cos^2\theta}{2 - 5\cos^2\theta}\tan\sigma \,, \\ &\phi_{12} = \rho + \frac{(k_1 - k_2)\pi}{2} \,, \qquad \phi_{13} = \rho + \frac{(k_1 - k_3)\pi}{2} \,. \end{aligned} \tag{5.11}$$

Notice that in the symmetric parametrization the CP violating phase characterizing neutrino oscillations is given by the invariant combination $\delta^\ell = \phi_{13} - \phi_{12} - \phi_{23}$ [99], see Eq. (1.22). We see that the first column of the lepton mixing matrix in Eq. (5.10) is $(2, -e^{i\rho}, e^{i(\rho+\sigma)})^T/\sqrt{6}$ which is in common with that of the cTBM mixing pattern. This arises from the preserved G_1 symmetry. Eliminating the parameters

θ and σ in Eq. (5.11), we see that the lepton mixing angles and CP phases are correlated with each other according to

$$\cos^2\theta_{12}\cos^2\theta_{13} = \frac{2}{3}\,, \qquad \tan 2\theta_{23}\cos\delta^\ell = \frac{5\sin^2\theta_{13}-1}{2\sin\theta_{13}\sqrt{2-6\sin^2\theta_{13}}}\,. \tag{5.12}$$

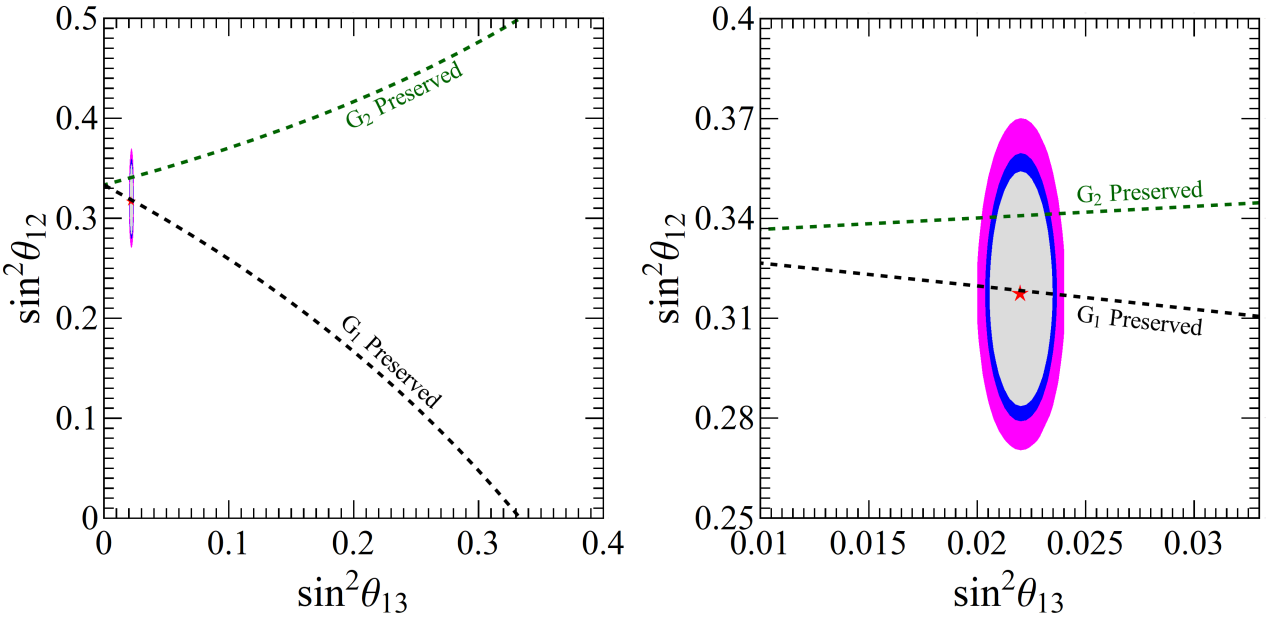


Figure 11: Predicted correlation between $\sin^2\theta_{12}$ and $\sin^2\theta_{13}$ in the revamped TBM scheme. The black-dashed line corresponds to the case where G_1 is preserved by the neutrino sector (Eq. (5.12), left), while the green-dashed one refers to the case where G_2 is preserved (Eq. (5.18), left). The right panel is a zoom of the left one. The global fit regions correspond to 90%, 95% and 99% confidence levels [24, 25].

The first equation in Eq. (5.12) relates the solar and the reactor angles while, for given values of the latter, the second equation correlates the CP phase δ^ℓ and the atmospheric angle. These correlations can be used to test the mixing matrix of Eq. (5.10) at current and future oscillation experiments. Notice that these correlations are a generic feature of mass matrices which preserve the G_1 symmetry. These are displayed in figure 11 and figure 12. In the limit of $\rho, \sigma \to 0$, we see that the mixing angles θ_{12} and θ_{13} remain the same, while the Dirac CP phase vanishes $\sin\delta^\ell \to 0$, so that CP would be conserved in neutrino oscillations. Notice that both Majorana phases [111] become some integer multiples of $\pi/2$ and therefore they correspond to just CP signs [368, 369].

When the two CP symmetries X_2 and X_3 are imposed, the neutrino mass matrix preserves the flavour symmetry $G_1 = X_2 X_3^* = X_3 X_2^*$ as well. The resulting predictions for lepton mixing parameters are obtained from Eq. (5.11) by redefining $\rho \to \rho + \pi/2$ and $\sigma \to \sigma - \pi/2$.

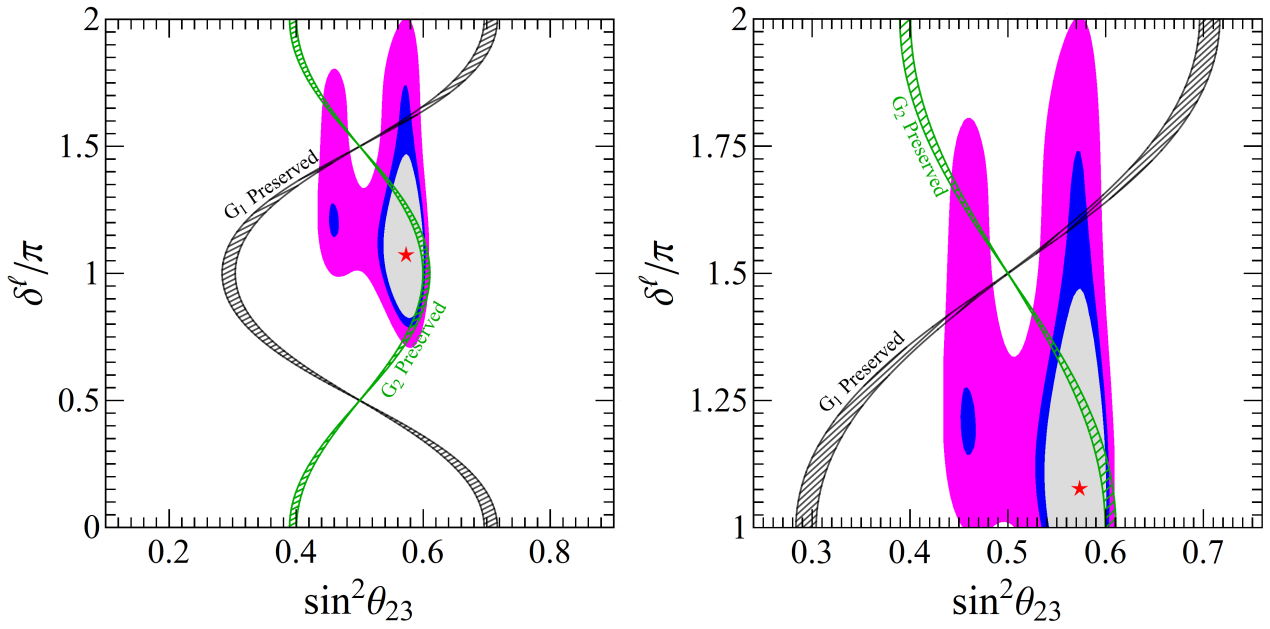


Figure 12: Predicted correlation between δ^ℓ and $\sin^2\theta_{23}$ in the revamped TBM scheme. The black region corresponds to the case where G_1 is preserved by the neutrino sector, as given by Eq. (5.12), right. The green region corresponds to the case where G_2 is preserved, as given by Eq. (5.18), right. The right panel is a zoom of the left one. The global fit regions correspond to 90%, 95% and 99% confidence levels [24,25].

5.1.2 Case b: G_2 flavour and X_2, X_4 CP symmetries

The combination of X_2 and X_4 leads to the conservation of $G_2 = X_2 X_4^* = X_4 X_2^*$ flavour symmetry. The invariance of neutrino mass matrix under the action of X_2 and X_4 requires

$$X_2^T m_\nu X_2 = m_\nu^* , \qquad X_4^T m_\nu X_4 = m_\nu^* , \tag{5.13}$$

from which we can determine the neutrino mass matrix to be of the following form

$$m'_\nu = U_{cTBM}^T m_\nu U_{cTBM} = \begin{pmatrix} m_1 & 0 & \delta m \\ 0 & m_2 & 0 \\ \delta m & 0 & m_3 \end{pmatrix} , \tag{5.14}$$

where $m_{1,2,3}$ and δm are generic real parameters. The matrix m'_ν can be diagonalized by a rotation matrix $R_{13}(\theta)$ in the (13)-plane,

$$R_{13}(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad \text{with} \quad \tan 2\theta = \frac{2\,\delta m}{m_3 - m_1} . \tag{5.15}$$

Consequently the residual CP symmetries X_2 and X_4 fix the lepton mixing matrix to be

$$
\begin{aligned}
U &= U_{cTBM} R_{13}(\theta) Q_\nu \\
&= \frac{1}{\sqrt{6}} \begin{pmatrix} 2\cos\theta & \sqrt{2}e^{-i\rho} & 2\sin\theta \\ -e^{i\rho}\cos\theta - \sqrt{3}e^{-i\sigma}\sin\theta & \sqrt{2} & -e^{i\rho}\sin\theta + \sqrt{3}e^{-i\sigma}\cos\theta \\ e^{i(\rho+\sigma)}\cos\theta - \sqrt{3}\sin\theta & -\sqrt{2}e^{i\sigma} & e^{i(\rho+\sigma)}\sin\theta + \sqrt{3}\cos\theta \end{pmatrix} Q_\nu \,.
\end{aligned} \tag{5.16}
$$

Note that the second column of the mixing matrix is $(e^{-i\rho}, 1, -e^{i\sigma})^T/\sqrt{3}$ which matches with that of the cTBM mixing pattern. We can extract the mixing angles and the predicted CP violation phases from Eq. (5.16) in the usual way, leading to

$$
\begin{aligned}
&\sin^2\theta_{13} = \frac{2\sin^2\theta}{3}\,, \quad \sin^2\theta_{12} = \frac{1}{2\cos^2\theta+1}\,, \quad \sin^2\theta_{23} = \frac{1}{2} - \frac{\sqrt{3}\sin 2\theta\cos(\rho+\sigma)}{4\cos^2\theta+2}\,, \\
&\sin\delta^\ell = -\frac{\mathrm{sign}(\sin 2\theta)(2\cos^2\theta+1)\sin(\rho+\sigma)}{\sqrt{(2\cos^2\theta+1)^2 - 3\cos^2(\rho+\sigma)\sin^2 2\theta}}\,, \quad \tan\delta^\ell = \frac{(2\cos^2\theta+1)\tan(\rho+\sigma)}{1-4\cos^2\theta}\,, \\
&\phi_{12} = \rho + \frac{(k_1-k_2)\pi}{2}\,, \quad \phi_{13} = \frac{(k_1-k_3)\pi}{2}\,.
\end{aligned} \tag{5.17}
$$

The mixing parameters are again correlated with each other, as follows

$$
\sin^2\theta_{12}\cos^2\theta_{13} = \frac{1}{3}\,, \qquad \tan 2\theta_{23}\cos\delta^\ell = \frac{\cos 2\theta_{13}}{\sin\theta_{13}\sqrt{2-3\sin^2\theta_{13}}}\,. \tag{5.18}
$$

These correlations lead to predictions for the oscillations parameters, given in figures 11 and 12.

For the case that the CP symmetries X_1 and X_3 are preserved, the flavour symmetry $G_2 = X_1 X_3^* = X_3 X_1^*$ would be preserved as well [18]. The resulting predictions for lepton mixing matrix and mixing parameters can be obtained from Eqs. (5.16, 5.17) by redefining $\rho \to \rho + \pi/2$, $\sigma \to \sigma - \pi$.

Finally, if a single CP symmetry is preserved by the neutrino mass matrix, the lepton mixing matrix is determined up to a three dimensional orthogonal matrix. The resulting lepton flavour mixing predictions can be analyzed in a similar fashion [365]. The simple TBM mixing matrix can also be revamped by exploiting the generalized CP symmetries of the charged lepton mass matrix [370].

5.2 Revamped Golden-Ratio mixing scheme

In a way analogous to what we did for the TBM mixing matrix, we can also revamp the Golden-Ratio (GR) mixing pattern. We start from the complex GR mixing matrix, in the charged lepton diagonal basis,

$$
U_{cGR} = \frac{1}{\sqrt{2\sqrt{5}\,\phi_g}} \begin{pmatrix} \sqrt{2}\phi_g & \sqrt{2}e^{-i\rho} & 0 \\ -e^{i\rho} & \phi_g & \sqrt{\sqrt{5}\,\phi_g}\,e^{-i\sigma} \\ -e^{i(\rho+\sigma)} & \phi_g e^{i\sigma} & -\sqrt{\sqrt{5}\,\phi_g} \end{pmatrix}\,, \tag{5.19}
$$

[18] The imposition of G_3 is uninteresting here, as it leads to $\theta_{13} = 0$.

which reduces to the real GR mixing matrix of Eq. (3.13) in the limit of $\rho = \sigma = 0$. The four CP symmetry matrices $X_i = U_{cGR}\, d_i\, U^T_{cGR}$ associated with the cGR mixing pattern are of the following form,

$$
\begin{aligned}
X_1 &= \frac{1}{\sqrt{5}}\begin{pmatrix} 1+\frac{2i\sin\rho}{\phi_g}e^{-i\rho} & -\sqrt{2}\cos\rho & -\sqrt{2}\,e^{i\sigma}\cos\rho \\ -\sqrt{2}\cos\rho & -\frac{1+\sqrt{5}e^{-2i\sigma}}{2}+\frac{i\sin\rho}{\phi_g}e^{i\rho} & \frac{-e^{i\sigma}+\sqrt{5}e^{-i\sigma}}{2}+\frac{i\sin\rho}{\phi_g}e^{i(\rho+\sigma)} \\ -\sqrt{2}e^{i\sigma}\cos\rho & \frac{-e^{i\sigma}+\sqrt{5}e^{-i\sigma}}{2}+\frac{i\sin\rho}{\phi_g}e^{i(\rho+\sigma)} & -\frac{e^{2i\sigma}+\sqrt{5}}{2}+\frac{i\sin\rho}{\phi_g}e^{i(\rho+2\sigma)} \end{pmatrix}, \\
X_2 &= \frac{1}{\sqrt{5}}\begin{pmatrix} -1-\frac{2i\sin\rho}{\phi_g}e^{-i\rho} & \sqrt{2}\cos\rho & \sqrt{2}\,e^{i\sigma}\cos\rho \\ \sqrt{2}\cos\rho & \frac{1-\sqrt{5}e^{-2i\sigma}}{2}-\frac{i\sin\rho}{\phi_g}e^{i\rho} & \frac{e^{i\sigma}+\sqrt{5}e^{-i\sigma}}{2}-\frac{i\sin\rho}{\phi_g}e^{i(\rho+\sigma)} \\ \sqrt{2}\,e^{i\sigma}\cos\rho & \frac{e^{i\sigma}+\sqrt{5}\,e^{-i\sigma}}{2}-\frac{i\sin\rho}{\phi_g}e^{i(\rho+\sigma)} & \frac{e^{2i\sigma}-\sqrt{5}}{2}-\frac{i\sin\rho}{\phi_g}e^{i(\rho+2\sigma)} \end{pmatrix}, \\
X_3 &= \frac{1}{\sqrt{5}}\begin{pmatrix} -1-\frac{2\cos\rho}{\phi_g}e^{-i\rho} & \sqrt{2}\,i\sin\rho & \sqrt{2}\,ie^{i\sigma}\sin\rho \\ \sqrt{2}\,i\sin\rho & \frac{-1+\sqrt{5}\,e^{-2i\sigma}}{2}-\frac{\cos\rho}{\phi_g}e^{i\rho} & -\frac{e^{i\sigma}+\sqrt{5}\,e^{-i\sigma}}{2}-\frac{\cos\rho}{\phi_g}e^{i(\rho+\sigma)} \\ \sqrt{2}\,ie^{i\sigma}\sin\rho & -\frac{e^{i\sigma}+\sqrt{5}\,e^{-i\sigma}}{2}-\frac{\cos\rho}{\phi_g}e^{i(\rho+\sigma)} & \frac{-e^{2i\sigma}+\sqrt{5}}{2}-\frac{\cos\rho}{\phi_g}e^{i(\rho+2\sigma)} \end{pmatrix}, \\
X_4 &= \frac{1}{\sqrt{5}}\begin{pmatrix} 1+\frac{2\cos\rho}{\phi_g}e^{-i\rho} & -\sqrt{2}\,i\sin\rho & -\sqrt{2}\,ie^{i\sigma}\sin\rho \\ -\sqrt{2}\,i\sin\rho & \frac{1+\sqrt{5}\,e^{-2i\sigma}}{2}+\frac{\cos\rho}{\phi_g}e^{i\rho} & \frac{e^{i\sigma}-\sqrt{5}\,e^{-i\sigma}}{2}+\frac{\cos\rho}{\phi_g}e^{i(\rho+\sigma)} \\ -\sqrt{2}\,ie^{i\sigma}\sin\rho & \frac{e^{i\sigma}-\sqrt{5}\,e^{-i\sigma}}{2}+\frac{\cos\rho}{\phi_g}e^{i(\rho+\sigma)} & \frac{e^{2i\sigma}+\sqrt{5}}{2}+\frac{\cos\rho}{\phi_g}e^{i(\rho+2\sigma)} \end{pmatrix}. \qquad (5.20)
\end{aligned}
$$

Taking the limit of $\rho, \sigma \to 0$, we obtain

$$
\begin{aligned}
&X_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & -\phi_g & 1/\phi_g \\ -\sqrt{2} & 1/\phi_g & -\phi_g \end{pmatrix}, \quad X_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} -1 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -1/\phi_g & \phi_g \\ \sqrt{2} & \phi_g & -1/\phi_g \end{pmatrix}, \\
&X_3 = -\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (5.21)
\end{aligned}
$$

The relations between residual flavour and CP symmetries in Eq. (5.6) are fulfilled. If all the four remnant CP transformations in Eq. (5.20) are preserved by the neutrino mass matrix, the complex GR mixing pattern with vanishing θ_{13} would be produced.

Similarly to section 5.1, we consider the scenario of partially preserved remnant CP symmetries. If the CP symmetries X_1, X_4 or X_2, X_3 are preserved in the neutrino sector, the remnant flavour symmetry $G_1 = X_1X_4^* = X_4X_1^* = X_2X_3^* = X_3X_2^*$ would be conserved as well. As a consequence, the first column of the lepton mixing matrix would be determined to be $(\sqrt{2}\phi_g, -e^{i\rho}, -e^{i(\rho+\sigma)})^T/\sqrt{2\sqrt{5}\,\phi_g}$, the same as in the cGR mixing. It follows that the relation $\cos^2\theta_{12}\cos^2\theta_{13} = \frac{\phi_g}{\sqrt{5}}$ is fulfilled. Using the 3σ allowed range $2.000\times10^{-2} \leq \sin^2\theta_{13} \leq 2.405\times10^{-2}$ [24,25], we find the solar mixing angle must lie in the region $0.2586 \leq \sin^2\theta_{12} \leq 0.2616$ which doesn't overlap with the experimental 3σ range of θ_{12} [24,25].

The phenomenologically viable case is found when the CP symmetries X_2, X_4 or X_1, X_3 are preserved by the neutrino mass matrix. The mixing parameters for the latter CP transformation can be obtained from those of the former by redefining $\rho \to \rho + \pi/2, \sigma \to \sigma - \pi$. Without loss of generality, we shall focus on preserved CP transformations X_2, X_4 which leads to conservation of the flavour symmetry generated

by the $G_2 = X_2 X_4^* = X_4 X_2^*$ transformation [19]. The lepton mixing matrix is found to be

$$\begin{aligned} U &= U_{cGR} R_{13}(\theta) Q_\nu \\ &= \frac{1}{\sqrt{2\sqrt{5}\phi_g}} \begin{pmatrix} \sqrt{2}\phi_g \cos\theta & \sqrt{2} e^{-i\rho} & \sqrt{2}\phi_g \sin\theta \\ -e^{i\rho}\cos\theta - 5^{1/4} e^{-i\sigma}\sqrt{\phi_g}\sin\theta & \phi_g & -e^{i\rho}\sin\theta + 5^{1/4} e^{-i\sigma}\sqrt{\phi_g}\cos\theta \\ -e^{i(\rho+\sigma)}\cos\theta + 5^{1/4}\sqrt{\phi_g}\sin\theta & e^{i\sigma}\phi_g & -e^{i(\rho+\sigma)}\sin\theta - 5^{1/4}\sqrt{\phi_g}\cos\theta \end{pmatrix} Q_\nu \,. \end{aligned} \tag{5.22}$$

The mixing angles and CP violating phases read as

$$\begin{aligned} &\sin^2\theta_{13} = \frac{\phi_g \sin^2\theta}{\sqrt{5}} \,, \quad \sin^2\theta_{12} = \frac{1}{1+\phi_g^2\cos^2\theta} \,, \quad \sin^2\theta_{23} = \frac{1}{2} - \frac{\sqrt{\sqrt{5}\phi_g}\sin 2\theta\cos(\rho+\sigma)}{2\phi_g^2\cos^2\theta + 2} \,, \\ &\sin\delta^\ell = -\frac{\mathrm{sign}(\sin 2\theta)(\phi_g^2\cos^2\theta+1)\sin(\rho+\sigma)}{\sqrt{(\phi_g^2\cos^2\theta+1)^2 - \sqrt{5}\phi_g\sin^2 2\theta\cos^2(\rho+\sigma)}} \,, \\ &\tan\delta^\ell = -\frac{(\phi_g^2\cos^2\theta+1)\tan(\rho+\sigma)}{\cos 2\theta + \phi_g^2\cos^2\theta} \,, \quad \phi_{12} = \rho + \frac{(k_1-k_2)\pi}{2} \,, \quad \phi_{13} = \frac{(k_1-k_3)\pi}{2} \,. \end{aligned} \tag{5.23}$$

As a consequence, we can derive the following exact relations among the mixing parameters

$$\sin^2\theta_{12}\cos^2\theta_{13} = \frac{1}{\sqrt{5}\phi_g} \,, \qquad \tan 2\theta_{23}\cos\delta^\ell = \frac{\phi_g^2\cot^2\theta_{13} - 2}{2\sqrt{\phi_g^2\cot^2\theta_{13} - 1}} \,. \tag{5.24}$$

For the best fit value $\sin^2\theta_{13} = 2.200 \times 10^{-2}$ [24, 25], we find the solar mixing angle $\sin^2\theta_{12} = 0.2826$ which is within the 3σ range. The correlations of Eq. (5.24) are displayed in figure 13.

5.3 Bi-large mixing

It is unlikely that any revamping procedure can make the bi-maximal mixing pattern of section 3.4 consistent with the oscillation data, since the measured value of the solar angle θ_{12} deviates too much from the maximal value [24, 25]. The bi-large pattern is phenomenologically motivated by the fact that the smallest lepton mixing angle θ_{13} is similar in magnitude to the largest of the elements of the quark mixing matrix, the Cabibbo angle. It suggests that the latter may act as the universal seed for quark and lepton mixings [371–375]. Bi-large neutrino mixing implies that the lepton and quark sectors may be related with each other, a possible new strategy in the quest for quark-lepton symmetry and unification.

Within the simplest bi-large mixing hypothesis, the solar and atmospheric mixing angles are expressed as [371],

$$\sin\theta_{13} = \lambda, \quad \sin\theta_{12} = s\lambda, \quad \sin\theta_{23} = a\lambda \,, \tag{5.25}$$

where the small parameter λ is the reactor angle, s and a are free parameters of order one. Using the best fit values of the mixing angles [24, 25], one finds $\lambda \simeq 0.148$, $s \simeq 3.802$ and $a \simeq 5.108$ for normal neutrino mass ordering. Since the bi-large approach describes the structure of the lepton mixing matrix in terms of θ_{13} as input, no revamping is needed.

[19] Notice again that imposing G_3 is uninteresting here, as it leads to $\theta_{13} = 0$.

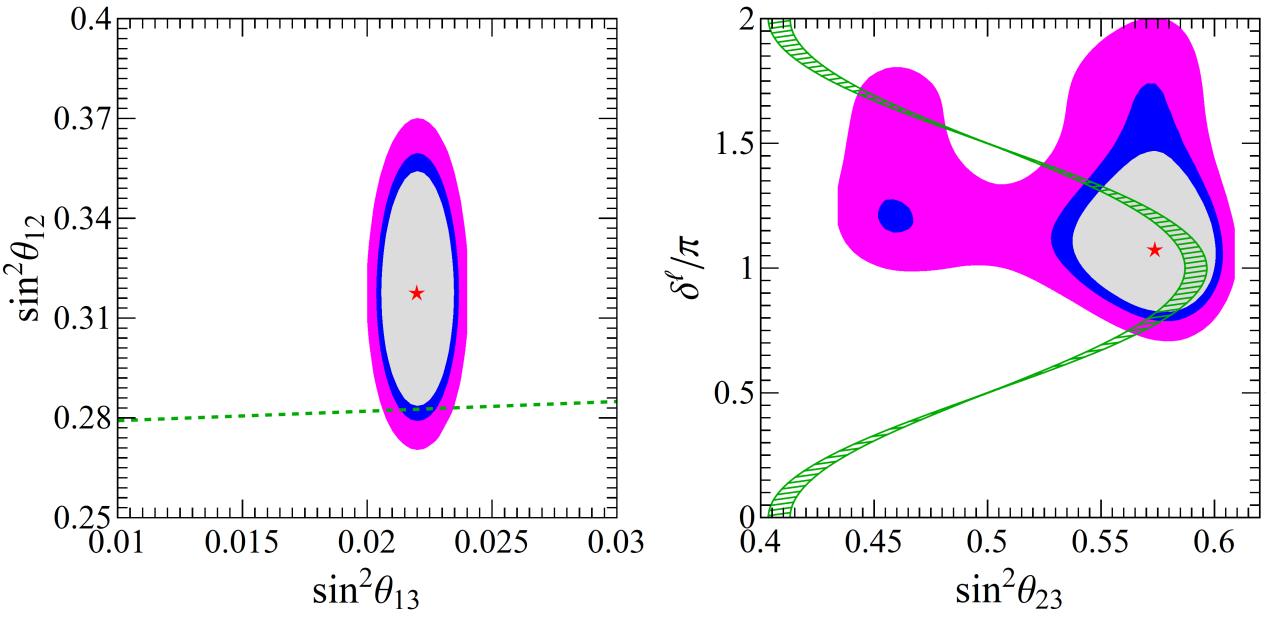


Figure 13: Predicted oscillation parameter correlations in the revamped GR scheme. The "generic" global-fit allowed regions corresponding to 90%, 95% and 99% confidence level are displayed [24, 25].

5.3.1 Bi-large mixing from abelian family symmetry

As shown above, the bi-large mixing ansatz assumes that the three lepton mixing angles are of the same order of magnitude, to first approximation,

$$\texttt{BL}_1: \quad \sin\theta_{12} \sim \lambda_c, \quad \sin\theta_{13} \sim \lambda_c, \quad \sin\theta_{23} \sim \lambda_c\,, \tag{5.26}$$

where $\lambda_c \simeq 0.23$ is the Cabibbo angle [28], and the notation "$\sim$" implies that the above relations contain unknown factors of order one. The freedom in these factors can be used to obtain an adequate description of neutrino mixing. Although global analyses of neutrino oscillation data show preference for the second octant of the atmospheric mixing angle θ_{23} [24, 25], the significance is not yet overwhelming. For θ_{23} in the preferred higher octant, it has been proposed [372] that the bi-large mixing ansatz could be

$$\texttt{BL}_2: \quad \sin\theta_{12} \sim \lambda_c, \quad \sin\theta_{13} \sim \lambda_c, \quad \sin\theta_{23} \sim 1-\lambda_c\,. \tag{5.27}$$

Here we will show how the above variants of the bi-large mixing patterns can be achieved in a Froggatt-Nielsen-type scenario [276] with an Abelian flavour symmetry. Assuming the presence of two Higgs doublets $H_{u,d}$, and that neutrino masses are described by an effective Weinberg operator we generalize the well-known $U(1)$ flavour symmetry to a larger $U(1) \times Z_m \times Z_n \subset U(1) \times U(1)' \times U(1)''$ family symmetry. In order to break the $U(1)$, Z_m and Z_n flavour groups we assume three flavons Θ_1, Θ_2 and Θ_3 with horizontal charges $\Theta_1 : (-1,0,0)$, $\Theta_2 : (0,-1,0)$, $\Theta_3 : (0,0,-1)$, respectively, and their VEVs $\langle\Theta_{1,2,3}\rangle/\Lambda$ scaled by the cutoff Λ are of order λ_c. We assume the horizontal charges of $H_{u,d}$ to be zero.

Fermion masses are then described by the following effective Yukawa couplings [372],

$$\begin{aligned}\mathcal{W} = &(y_u)_{ij} Q_i U_j^c H_u \left(\frac{\Theta_1}{\Lambda}\right)^{F(Q_i)+F(U_j^c)} \left(\frac{\Theta_2}{\Lambda}\right)^{\left[Z_m(Q_i)+Z_m(U_j^c)\right]} \left(\frac{\Theta_3}{\Lambda}\right)^{\left[Z_n(Q_i)+Z_n(U_j^c)\right]} \\ &+(y_d)_{ij} Q_i D_j^c H_d \left(\frac{\Theta_1}{\Lambda}\right)^{F(Q_i)+F(D_j^c)} \left(\frac{\Theta_2}{\Lambda}\right)^{\left[Z_m(Q_i)+Z_m(D_j^c)\right]} \left(\frac{\Theta_3}{\Lambda}\right)^{\left[Z_n(Q_i)+Z_n(D_j^c)\right]} \\ &+(y_e)_{ij} L_i E_j^c H_d \left(\frac{\Theta_1}{\Lambda}\right)^{F(L_i)+F(E_j^c)} \left(\frac{\Theta_2}{\Lambda}\right)^{\left[Z_m(L_i)+Z_m(E_j^c)\right]} \left(\frac{\Theta_3}{\Lambda}\right)^{\left[Z_n(L_i)+Z_n(E_j^c)\right]} \\ &+(y_\nu)_{ij} \frac{1}{\Lambda} L_i L_j H_u H_u \left(\frac{\Theta_1}{\Lambda}\right)^{F(L_i)+F(L_j)} \left(\frac{\Theta_2}{\Lambda}\right)^{\left[Z_m(L_i)+Z_m(L_j)\right]} \left(\frac{\Theta_3}{\Lambda}\right)^{\left[Z_n(L_i)+Z_n(L_j)\right]} , \end{aligned} \tag{5.28}$$

where $F(\psi)$ denotes the $U(1)$ charge of the field ψ, $Z_{m,n}(\psi)$ stand for the $Z_{m,n}$ charge of ψ, and the brackets $[\ldots]$ around the exponents denote that we are modding out by m (n) according to the Z_m (Z_n) addition rule, i.e., $\Big[Z_m(Q_i) + Z_m(U_j^c)\Big] = Z_m(Q_i) + Z_m(U_j^c)$ (mod m). Hence fermion mass matrices are expressed in terms of the horizontal charges as follows,

$$\begin{aligned}(M_u)_{ij} &= (y_u)_{ij} \lambda_c^{F(Q_i)+F(U_j^c)+\left[Z_m(Q_i)+Z_m(U_j^c)\right]+\left[Z_n(Q_i)+Z_n(U_j^c)\right]} v_u \,, \\ (M_d)_{ij} &= (y_d)_{ij} \lambda_c^{F(Q_i)+F(D_j^c)+\left[Z_m(Q_i)+Z_m(D_j^c)\right]+\left[Z_n(Q_i)+Z_n(D_j^c)\right]} v_d \,, \\ (M_e)_{ij} &= (y_e)_{ij} \lambda_c^{F(L_i)+F(E_j^c)+\left[Z_m(L_i)+Z_m(E_j^c)\right]+\left[Z_n(L_i)+Z_n(E_j^c)\right]} v_d \,, \\ (M_\nu)_{ij} &= (y_\nu)_{ij} \lambda_c^{F(L_i)+F(L_j)+\left[Z_m(L_i)+Z_m(L_j)\right]+\left[Z_n(L_i)+Z_n(L_j)\right]} \frac{v_u^2}{\Lambda} \,. \end{aligned} \tag{5.29}$$

If all the horizontal charges are positive, the hierarchical structure of the mass matrices allows a simple order-of-magnitude estimate of the various mass ratios and mixing angles. For instance, the entries of the CKM matrix are estimated to be,

$$(V_{CKM})_{ij} \sim \lambda_c^{F_{\rm eff}(Q_i) - F_{\rm eff}(Q_j) \pm \alpha m \pm \beta n} \,, \tag{5.30}$$

where $F_{\rm eff}(\psi) = F(\psi) + Z_m(\psi) + Z_n(\psi)$, and $\alpha, \beta = 0, 1$ depends on the charge assignment under Z_m and Z_n. Likewise for the lepton sector, one obtains

$$\sin\theta_{ij} \sim \lambda_c^{F_{\rm eff}(L_i) - F_{\rm eff}(L_j) \pm \alpha m \pm \beta n} \,. \tag{5.31}$$

Notice that the mixing angles can be enhanced or suppressed by $\lambda^{\pm m \pm n}$ relative to the scaling predictions obtained with the continuous $U(1) \times U(1)' \times U(1)''$ family symmetry. Moreover, $U(1) \times Z_m \times Z_n$ reduces to $U(1) \times Z_m$ if $n = 1$, and to the usual $U(1)$ if $m = n = 1$.

- Model for $\texttt{BL}_1$ mixing

  The family symmetry group is $U(1) \times Z_3 \times Z_4$, and we assign the lepton fields to transform under the

flavour symmetry as follows [372],

$$\begin{aligned} &L_1: \ (4,1,3), \qquad L_2: \ (3,2,2), \qquad L_3: \ (1,1,1), \\ &E_1^c: \ (3,2,2), \qquad E_2^c: \ (1,2,2), \qquad E_3^c: \ (0,0,0)\,. \end{aligned} \tag{5.32}$$

One can then read out the pattern of charged lepton and neutrino mass matrices,

$$M_e \sim \begin{pmatrix} \lambda_c^8 & \lambda_c^6 & \lambda_c^8 \\ \lambda_c^7 & \lambda_c^5 & \lambda_c^7 \\ \lambda_c^7 & \lambda_c^5 & \lambda_c^3 \end{pmatrix} v_d, \qquad M_\nu \sim \begin{pmatrix} \lambda_c^{12} & \lambda_c^8 & \lambda_c^7 \\ \lambda_c^8 & \lambda_c^7 & \lambda_c^7 \\ \lambda_c^7 & \lambda_c^7 & \lambda_c^6 \end{pmatrix} \frac{v_u^2}{\Lambda}\,. \tag{5.33}$$

These give rise to the following mass ratios and lepton mixing angles,

$$\begin{aligned} &\frac{m_e}{m_\mu} \sim \lambda_c^3, \qquad \frac{m_\mu}{m_\tau} \sim \lambda_c^2\,, \\ &m_1 \sim \lambda_c^8 \frac{v_u^2}{\Lambda}, \quad m_2 \sim \lambda_c^7 \frac{v_u^2}{\Lambda}, \quad m_3 \sim \lambda_c^6 \frac{v_u^2}{\Lambda}\,, \\ &\sin\theta_{12} \sim \lambda_c, \qquad \sin\theta_{13} \sim \lambda_c, \qquad \sin\theta_{23} \sim \lambda_c\,. \end{aligned} \tag{5.34}$$

This way we obtain the $\texttt{BL}_1$ mixing pattern. A simple numerical analysis with different seed procedures for the order-one Yukawa coefficients, leads to the θ_{23} distributions given in figure 14. One sees that $\sin^2\theta_{23} < 1/2$ (first octant) is preferred in this case.

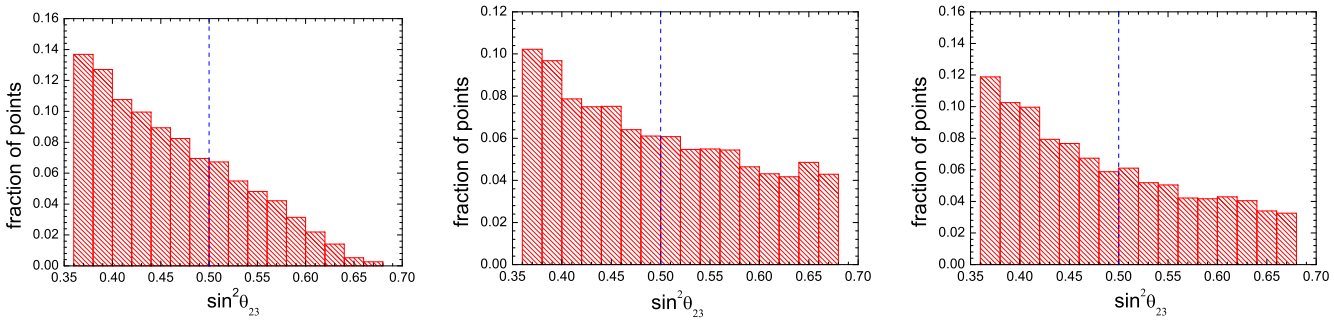


Figure 14: Distributions of the atmospheric neutrino mixing angle θ_{23} in the $\texttt{BL}_1$ model. The left, middle and right panels are obtained with flat, exponential and Gaussian seed procedures. From [372].

We now extend the model to the quark sector along the lines of $SU(5)$ unification. Each family of standard quarks and leptons is embedded in $SU(5)$ multiplets $\mathbf{10} = (Q, U^c, E^c)$ and $\bar{\mathbf{5}} = (D^c, L)$. The fields within a single SU(5) multiplet transform in the same way under the family symmetry. Therefore the lepton assignment in Eq. (5.32) implies that the quark charges under the flavour group $U(1) \times Z_3 \times Z_4$ should be

$$Q_1: \ (3,2,2), \qquad Q_2: \ (1,2,2), \qquad Q_3: \ (0,0,0)\,,$$

$$U_1^c:\ (3,2,2), \qquad U_2^c:\ (1,2,2), \qquad U_3^c:\ (0,0,0)\,,$$
$$D_1^c:\ (4,1,3), \qquad D_2^c:\ (3,2,2), \qquad D_3^c:\ (1,1,1)\,. \tag{5.35}$$

Consequently, the up-type and down-type quark mass matrices are given as

$$M_u \sim \begin{pmatrix} \lambda_c^7 & \lambda_c^5 & \lambda_c^7 \\ \lambda_c^5 & \lambda_c^3 & \lambda_c^5 \\ \lambda_c^7 & \lambda_c^5 & 1 \end{pmatrix} v_u, \qquad M_d \sim \begin{pmatrix} \lambda_c^8 & \lambda_c^7 & \lambda_c^7 \\ \lambda_c^6 & \lambda_c^5 & \lambda_c^5 \\ \lambda_c^8 & \lambda_c^7 & \lambda_c^3 \end{pmatrix} v_d\,, \tag{5.36}$$

The resulting quark masses and CKM mixing matrix are determined to follow the pattern

$$\begin{aligned} &m_u \sim \lambda_c^7 v_u, \qquad && m_c \sim \lambda_c^3 v_u, \qquad && m_t \sim v_u, \\ &m_d \sim \lambda_c^8 v_d, \qquad && m_s \sim \lambda_c^5 v_d, \qquad && m_b \sim \lambda_c^3 v_d\,, \\ &|V_{us}| \sim \lambda_c^2, \qquad && |V_{cb}| \sim \lambda_c^2, \qquad && |V_{ub}| \sim \lambda_c^4\,, \end{aligned} \tag{5.37}$$

which are in agreement with experimental data except $|V_{us}|$, for which an accidental enhancement of $\mathcal{O}(\lambda^{-1})$ amongst the free order-one coefficients is needed so as to reproduce the correct Cabibbo angle.

- Model for $\texttt{BL}_2$ mixing

  Here the flavour symmetry is $U(1) \times Z_2$, and the charge assignments for the quarks and leptons are taken as [372]

$$D_1^c, L_1:\ (3,0), \qquad D_2^c, L_2:\ (3,1), \qquad D_3^c, L_3:\ (2,0),$$
$$Q_1, U_1^c, E_1^c:\ (4,0), \qquad Q_2, U_2^c, E_2^c:\ (2,1), \qquad Q_3, U_3^c, E_3^c:\ (0,1)\,. \tag{5.38}$$

  Consequently the neutrino and charged lepton mass matrices take the following form

$$M_e \sim \begin{pmatrix} \lambda_c^7 & \lambda_c^6 & \lambda_c^4 \\ \lambda_c^8 & \lambda_c^5 & \lambda_c^3 \\ \lambda_c^6 & \lambda_c^5 & \lambda_c^3 \end{pmatrix} v_d, \qquad M_\nu \sim \begin{pmatrix} \lambda_c^6 & \lambda_c^7 & \lambda_c^5 \\ \lambda_c^7 & \lambda_c^6 & \lambda_c^6 \\ \lambda_c^5 & \lambda_c^6 & \lambda_c^4 \end{pmatrix} \frac{v_u^2}{\Lambda}\,. \tag{5.39}$$

  The charged lepton mass matrix M_e has a "lopsided" structure, and can give the correct order-of-magnitude for the charged lepton masses and a large 2-3 mixing. Combining neutrino and charged lepton diagonalization matrices, the resulting $\texttt{BL}_2$ lepton mixing pattern is determined as,

$$\sin\theta_{12} \sim \lambda_c, \qquad \sin\theta_{13} \sim \lambda_c \qquad \sin\theta_{23} \sim 1\,. \tag{5.40}$$

Using the master formula of Eq. (5.29) we can easily read off the quark mass matrices as

$$M_u \sim \begin{pmatrix} \lambda_c^8 & \lambda_c^7 & \lambda_c^5 \\ \lambda_c^7 & \lambda_c^4 & \lambda_c^2 \\ \lambda_c^5 & \lambda_c^2 & 1 \end{pmatrix} v_u, \qquad M_d \sim \begin{pmatrix} \lambda_c^7 & \lambda_c^8 & \lambda_c^6 \\ \lambda_c^6 & \lambda_c^5 & \lambda_c^5 \\ \lambda_c^4 & \lambda_c^3 & \lambda_c^3 \end{pmatrix} v_d \,. \tag{5.41}$$

This leads to the following pattern of CKM matrix elements and quark mass ratios,

$$\begin{gathered} |V_{us}| \sim \lambda_c, \qquad |V_{cb}| \sim \lambda_c^2, \qquad |V_{ub}| \sim \lambda_c^3 \,, \\ \frac{m_u}{m_c} \sim \lambda_c^4, \qquad \frac{m_c}{m_t} \sim \lambda_c^4, \qquad \frac{m_d}{m_s} \sim \lambda_c^2, \qquad \frac{m_s}{m_b} \sim \lambda_c^2, \qquad \frac{m_b}{m_t} \sim \lambda_c^3 \,, \end{gathered} \tag{5.42}$$

in very good qualitative agreement with observed values.

5.3.2 Confronting bi-large mixing with oscillation data

We now make the bi-large mixing ansatz more predictive. We assume a CP conserving neutrino diagonalization matrix U_ν, with its three angles related to the Cabibbo angle in a simple manner, and a CKM-like charged lepton diagonalization. Under these assumptions we illustrate the predictive power of bi-large mixing [374].

- Constraining bi-large mixing: pattern I

  Here the neutrino mixing angles are assumed to be related to the Cabibbo angle as follows [374]

$$\sin\theta_{13} = \lambda_c, \quad \sin\theta_{12} = 2\lambda_c, \quad \sin\theta_{23} = 1 - \lambda_c \,. \tag{5.43}$$

  The Dirac CP phase is taken as $\delta^\nu_{CP} = \pi$ with vanishing Majorana phases. The resulting neutrino diagonalization matrix is given by

$$U_\nu \simeq \begin{pmatrix} 1 - \frac{5\lambda_c^2}{2} & 2\lambda_c & -\lambda_c \\ \lambda_c - 2\sqrt{2}\lambda_c^{3/2} & \sqrt{2\lambda_c} - \frac{\lambda_c^{3/2}}{2\sqrt{2}} & 1 - \lambda_c - \frac{\lambda_c^2}{2} \\ 2\lambda_c + \sqrt{2}\lambda_c^{3/2} & -1 + \lambda_c & \sqrt{2\lambda_c} - \frac{\lambda_c^{3/2}}{2\sqrt{2}} \end{pmatrix} . \tag{5.44}$$

  Motivated by $SO(10)$, we take a CKM-type charged lepton diagonalization matrix

$$U_l = R_{23}(\theta_{23}^{CKM})\, \Phi\, R_{12}(\theta_{12}^{CKM})\, \Phi^\dagger \simeq \begin{pmatrix} 1 - \frac{\lambda_c^2}{2} & \lambda_c e^{-i\phi} & 0 \\ -\lambda_c e^{i\phi} & 1 - \frac{\lambda_c^2}{2} & A\lambda_c^2 \\ A\lambda_c^3 e^{i\phi} & -A\lambda_c^2 & 1 \end{pmatrix} , \tag{5.45}$$

  where $\sin\theta_{12}^{\rm CKM} = \lambda_c$, $\sin\theta_{23}^{\rm CKM} = A\lambda_c^2$, $\lambda_c = 0.22453 \pm 0.00044$ and $A = 0.836 \pm 0.015$ are the Wolfenstein parameters [28], and $\Phi = {\rm diag}(e^{-i\phi/2}, e^{i\phi/2}, 1)$ where ϕ is a free phase parameter. One sees that the lepton mixing matrix $U = U_l^\dagger U_\nu$ only depends on a single free phase parameter ϕ, which can in general

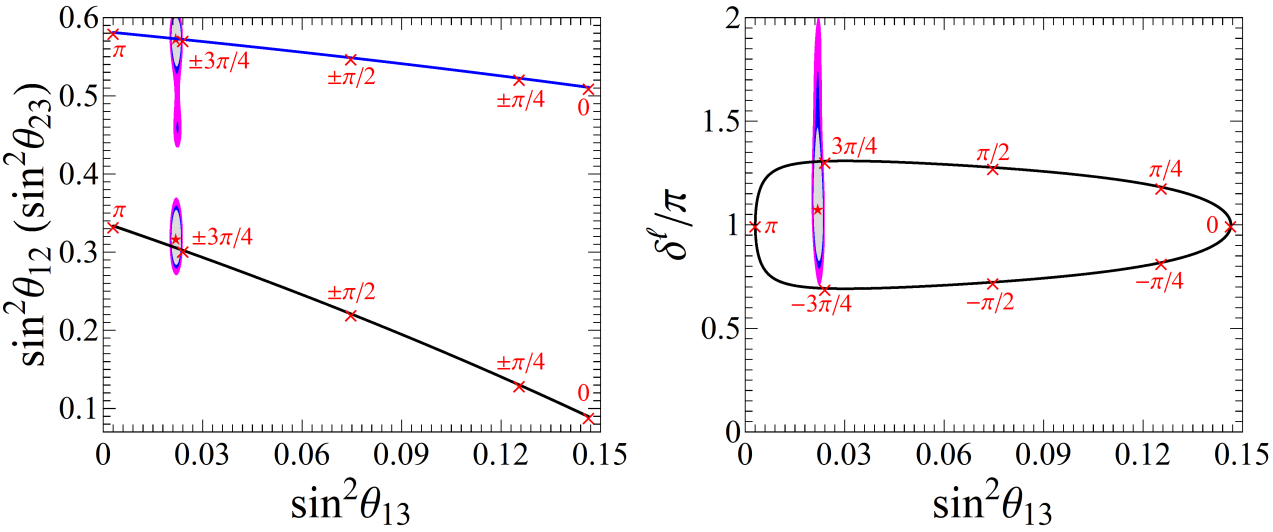


Figure 15: Predicting the solar and atmospheric mixing parameters $\sin^2\theta_{12}$ (lower, black line) and $\sin^2\theta_{23}$ (upper, blue line) versus $\sin^2\theta_{13}$ in the first kind of constrained bi-large mixing scheme. The right panel shows δ^ℓ versus $\sin^2\theta_{13}$. The allowed 90%, 95% and 99% CL regions of the global oscillation fit are displayed [24, 25]. The crosses correspond to $\phi = 0, \pi, \pm\pi/4, \pm\pi/2, \pm 3\pi/4$.

take values between $-\pi$ and π. The leptonic mixing angles and Jarlskog invariant are found to be

$$\begin{aligned}
\sin^2\theta_{13} &\simeq 4\lambda_c^2(1-\lambda_c)\cos^2\frac{\phi}{2}\,,\\
\sin^2\theta_{12} &\simeq 2\lambda_c^2\left(2-2\sqrt{2\lambda_c}\cos\phi+\lambda_c\right),\\
\sin^2\theta_{23} &\simeq (1-\lambda_c)^2-2\sqrt{2}A\lambda_c^{\frac{5}{2}}-2\lambda_c^3(1+2\cos\phi)\,,\\
J_{CP} &\simeq -2\left(\sqrt{2}+\sqrt{\lambda}\right)\lambda_c^{5/2}\sin\phi\,.
\end{aligned} \tag{5.46}$$

We show these correlations in figure 15. They show how the precise measurement of the reator angle can be promoted to sharp predictions for solar and atmospheric mixing angles. Similarly, the Dirac CP phase is also predicted, up to a two-fold degeneracy.

- Constraining bi-large mixing: pattern II

  In this case the bi-large ansatz for the neutrino mixing angles is [374]

$$\sin\theta_{13}^\nu = \lambda_c\,, \quad \sin\theta_{12}^\nu = 2\lambda_c\,, \quad \sin\theta_{23}^\nu = 3\lambda_c\,, \tag{5.47}$$

  with $\delta_{CP}^\nu = \pi$ and vanishing Majorana phases. Motivated by $SU(5)$ unification, we take the charged-lepton diagonalization matrix to be of the form

$$U_l = \Phi^\dagger R_{12}^T(\,\theta_{12}^{CKM}\,)\Phi R_{23}^T(\,\theta_{23}^{CKM}\,) \simeq \begin{pmatrix} 1-\frac{1}{2}\lambda_c^2 & -\lambda_c\, e^{i\phi} & A\lambda_c^3 e^{i\phi} \\ \lambda_c\, e^{-i\phi} & 1-\frac{1}{2}\lambda_c^2 & -A\lambda_c^2 \\ 0 & A\lambda_c^2 & 1 \end{pmatrix}. \tag{5.48}$$

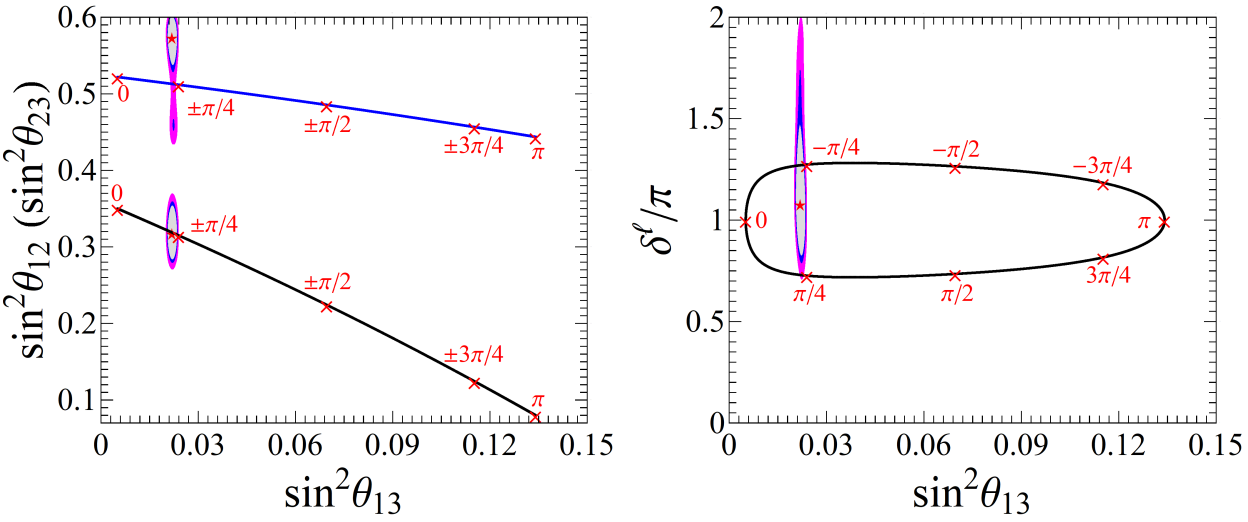


Figure 16: Predicting the solar and atmospheric mixing angles and Dirac phase in the second constrained bi-large mixing scheme. We adopt the same convention as in figure 15.

Taking into account both charged and neutral diagonalizations we can extract the following results for the lepton mixing angles and leptonic Jarlskog invariant,

$$\begin{aligned}
\sin^2\theta_{13} &\simeq \lambda_c^2 - 6\lambda_c^3\cos\phi + 8\lambda_c^4\,, \\
\sin^2\theta_{12} &\simeq \lambda_c^2(5+4\cos\phi) - 2\lambda_c^4(8+13\cos\phi)\,, \\
\sin^2\theta_{23} &\simeq 9\lambda_c^2 + 6\lambda_c^3(A+\cos\phi) - \lambda_c^4(8-2A\cos\phi-A^2)\,, \\
J_{CP} &\simeq -\left[3+(16+A)\lambda_c\right]\lambda_c^3\sin\phi\,.
\end{aligned} \tag{5.49}$$

The resulting correlations between mixing parameters are displayed in figure 16. As before, requiring $\sin^2\theta_{13}$ in the 3σ range [24, 25], we find that the other oscillation parameters θ_{12}, θ_{23} and δ_{CP} vary in the following regions

$$\begin{aligned}
&0.02000 \leq \sin^2\theta_{13} \leq 0.02405\,, \qquad 0.314824 \leq \sin^2\theta_{12} \leq 0.322459\,, \\
&0.511755 \leq \sin^2\theta_{23} \leq 0.513969\,, \qquad 1.26643 \leq \delta_{CP}/\pi \leq 1.27402\,.
\end{aligned} \tag{5.50}$$

6 Lepton mixing from flavour and CP symmetry

Symmetries have been widely used to address the family puzzle. Early attempts employed a $U(1)$ flavour symmetry [276], spontaneously broken by the vacuum expectation value of a singlet scalar field. This would qualitatively account for the small quark mass ratios, as well as mixing angles. With the discovery of neutrino oscillations the idea of flavour symmetry was substantially extended, with many symmetry groups and breaking patterns studied. It has been found that finite discrete flavour groups [376] are particularly suitable to reproduce the large lepton mixing angles and provide non-trivial predictions. The basic theory is assumed to be invariant under a flavour group G_f, but G_f is spontaneously broken into different subgroups G_ν and G_l in the neutrino and charged lepton sectors respectively. This is achieved by the VEVs of a set of scalar fields, and the misalignment between neutrino and charged lepton mass matrices arises from the non-trivial breaking pattern of the flavour symmetry.

Neutrino model building using non-abelian discrete flavor symmetries has been surveyed in a number of dedicated reviews [85–93]. A prime non-abelian discrete flavour symmetry is A_4 [53, 54, 70, 313] which, under some circumstances, leads to the celebrated tri-bimaximal mixing pattern. Although no longer compatible with the experimental measurement of the reactor angle $\theta_{13} \simeq 8.5^\circ$ [21, 22, 303, 305, 306, 377], the simplest A_4 symmetry can, as we saw, be revamped so as to produce viable and predictive patterns of neutrino mixing.

All in all, the idea that lepton mixing emerges from the mismatch of the embedding of the two residual subgroups G_ν and G_l into of the flavour group G_f is still viable, interesting and predictive. In particular, if combining flavour symmetry with the generalized CP symmetry, one can not only accommodate lepton mixing angles but also predict leptonic CP violating phases in terms of few free parameters. In what follows, we shall review the possible schemes of predicting lepton mixing from residual symmetry, emphasizing the role of generalized CP symmetry. The results only depend on the assumed symmetry breaking pattern and are independent of the details of the residual symmetry and the particle content of the flavor symmetry breaking sector, or possible additional symmetries of the theory.

6.1 Lepton mixing from flavour symmetry alone

Within a top-down approach, one imposes a certain flavour symmetry group G_f at some high energy scale. The full Lagrangian is invariant under G_f, which is subsequently broken down to different subgroups G_ν and G_l in the neutrino and charged lepton sectors respectively. The residual symmetry G_l can be any abelian subgroup of G_f, and the same holds for G_ν if neutrinos are Dirac particles. On the other hand G_ν can only be the Klein group K_4 (or a subgroup) for the case of Majorana neutrinos. Throughout this review, we shall focus on the scenario in which G_f is a non-abelian finite discrete group. In general, the three families of left-handed lepton doublets are assigned to an irreducible faithful three-dimensional representation $\rho_{\mathbf{3}}$ of G_f. We assume that the charged lepton and neutrino mass matrices are invariant under the action of the elements of G_l and G_ν, i.e.

$$\rho^\dagger_{\mathbf{3}}(g_l) m_l^\dagger m_l \rho_{\mathbf{3}}(g_l) = m_l^\dagger m_l, \qquad g_l \in G_l\,, \tag{6.1}$$

and

$$\begin{aligned}
&\rho^{\dagger}_{\mathbf{3}}(g_\nu) m^{\dagger}_{\nu} m_\nu \rho_{\mathbf{3}}(g_\nu) = m^{\dagger}_{\nu} m_\nu, \quad g_\nu \in G_\nu, \quad \text{for Dirac neutrinos}\,, \\
&\rho^{T}_{\mathbf{3}}(g_\nu) m_\nu \rho_{\mathbf{3}}(g_\nu) = m_\nu, \quad g_\nu \in G_\nu, \qquad \text{for Majorana neutrinos}\,.
\end{aligned} \tag{6.2}$$

Notice that it is sufficient to impose the conditions in Eqs. (6.1) and (6.2) on the generators of G_l and G_ν. Moreover, we see that these conditions imply

$$\left[\rho_{\mathbf{3}}(g_l), m^{\dagger}_{l} m_l\right] = 0, \qquad \left[\rho_{\mathbf{3}}(g_\nu), m^{\dagger}_{\nu} m_\nu\right] = 0\,. \tag{6.3}$$

As $\rho_{\mathbf{3}}(g_l)$ and $m^{\dagger}_{l} m_l$ commute with each other, the unitary transformation U_l that diagonalizes $m^{\dagger}_{l} m_l$ also diagonalizes $\rho_{\mathbf{3}}(g_l)$ up to permutations and phases of columns

$$U^{\dagger}_{l} \rho_{\mathbf{3}}(g_l) U_l = \text{diag}(e^{i\alpha_e}, e^{i\alpha_\mu}, e^{i\alpha_\tau})\,. \tag{6.4}$$

Since g_l is an element of the discrete flavour symmetry group G_f, the $e^{i\alpha_{e,\mu,\tau}}$ are all roots of unity. Analogously $\rho_{\mathbf{3}}(g_\nu)$ and $m^{\dagger}_{\nu} m_\nu$ (or m_ν) are diagonalized by the same matrix U_ν,

$$U^{\dagger}_{\nu} \rho_{\mathbf{3}}(g_\nu) U_\nu = \begin{cases} \text{diag}(e^{i\beta_e}, e^{i\beta_\mu}, e^{i\beta_\tau}), & \text{for Dirac neutrinos,} \\ \text{diag}(\pm 1, \pm 1, \pm 1), & \text{for Majorana neutrinos}\,, \end{cases} \tag{6.5}$$

where $\beta_{e,\mu,\tau}$ are rational multiples of π. Then the lepton mixing matrix U is determined as

$$U = U^{\dagger}_{l} U_\nu \tag{6.6}$$

up to independent row and column permutations. In short, given a family symmetry group G_f and the residual subgroups G_l and G_ν, the unitary transformations U_l and U_ν as well as the lepton matrix U can be obtained by diagonalizing the representation matrices of the generators of G_l and G_ν, as illustrated in the figure 17. In practice, we only need to find the eigenvectors of $\rho_{\mathbf{3}}(g_l)$ and $\rho_{\mathbf{3}}(g_\nu)$ that form the column vectors of U_l and U_ν. In this approach, it is not necessary to construct the explicit form of the mass matrices $m^{\dagger}_{l} m_l$ and $m^{\dagger}_{\nu} m_\nu$ (or m_ν) although this can be accomplished in a straightforward manner. Since the lepton masses cannot be predicted in this framework, in particular the neutrino mass spectrum can have either normal ordering or inverted ordering, the unitary matrices U_l and U_ν are uniquely fixed up to permutations and phases of their column vectors. As a consequence, the lepton mixing matrix U is determined up to independent row and column permutations and arbitrary phase matrices multiplied from the left and right sides. Therefore the Majorana CP phases are not constrained by the residual flavour symmetry.

Notice that if we switch the roles of the subgroups G_ν and G_l, the lepton matrix U would become its Hermitian conjugate. If two pairs of subgroups $\{G_l, G_\nu\}$ and $\{G'_l, G'_\nu\}$ are conjugate [20], both residual

[20] The two pairs of groups are conjugate if their generators g_e, g_ν and g'_e, g'_ν are related by group conjugacy, i.e., $g'_l = h g_l h^{-1}$,

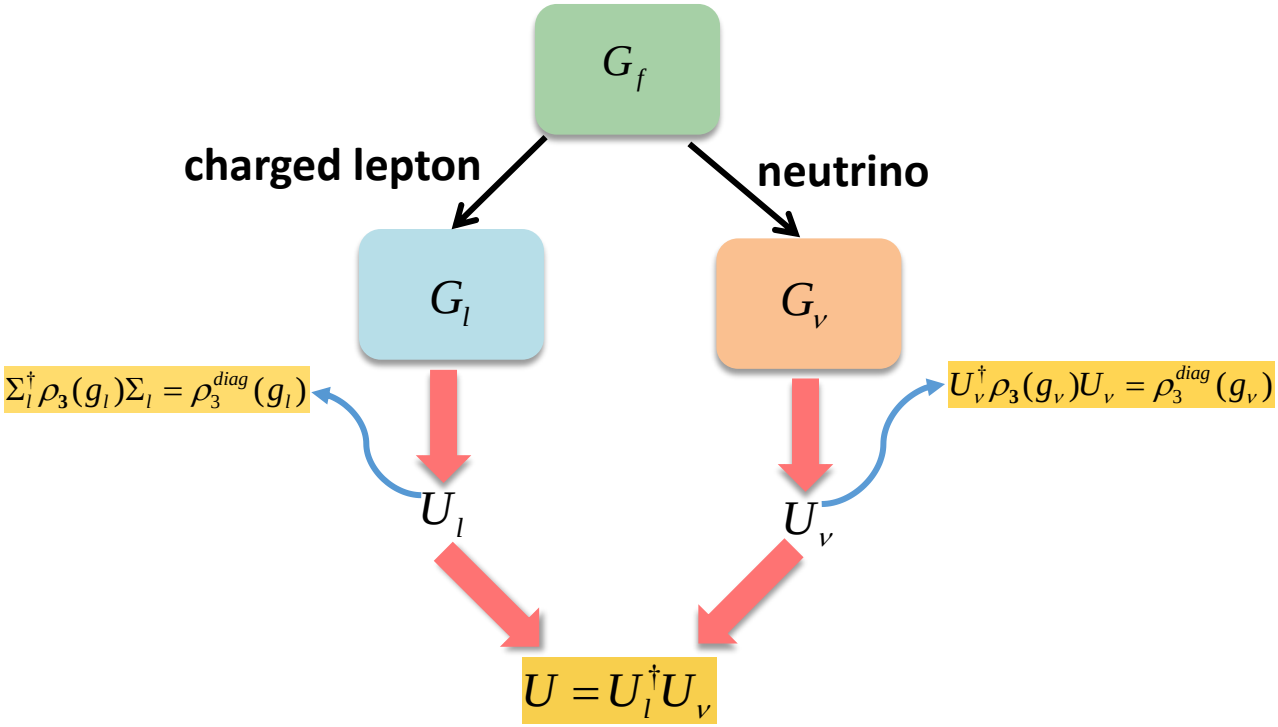


Figure 17: Model-independent predictions for the lepton mixing matrix from the flavour symmetry G_f broken to different residual subgroups G_l and G_ν in the neutrino and charged lepton sectors. Here g_l and g_ν are G_l and G_ν generators, respectively. See text for details.

symmetries would lead to the same result for the lepton mixing matrix. The reason is that if the generators of G_l and G_ν are diagonalized by U_l and U_ν respectively, $\rho_{\mathbf{3}}(h)U_l$ and $\rho_{\mathbf{3}}(h)U_\nu$ would diagonalize those of G_l' and G_ν'. On the other hand, if the left-handed lepton fields are assigned to the complex-conjugate of the triplet representation, the lepton mixing matrix would be complex conjugated, so that the predictions for the lepton mixing angles are unchanged and the signs of the CP violation phases are inverted.

6.1.1 Fully preserved residual symmetry $G_\nu = K_4$

As already mentioned, if the flavour group G_f is of finite order and the residual symmetries G_l and G_ν distinguish the three families of charged leptons and neutrinos then the lepton mixing matrix would be completely determined by the residual symmetries, regardless of the lepton masses or any other parameter of the underlying theory [259–263]. This is the so-called direct model building approach [88, 378]. This happens, for instance, if G_ν is the Klein group K_4 and G_l is a cyclic group Z_n with $n \geq 3$ or the product of several cyclic groups. A complete classification of all resulting mixing matrices has been performed by using theorems on roots of unity [261]. One finds that the lepton mixing matrix can take 17 sporadic patterns plus one infinite series, and only the latter could be compatible with the experimental data,

$g_\nu' = h g_\nu h^{-1}$, where h is an element of G_f.

given as [261, 263]

$$U = \frac{1}{\sqrt{3}} \begin{pmatrix} \sqrt{2}\cos\theta & 1 & -\sqrt{2}\sin\theta \\ -\sqrt{2}\cos\left(\theta - \frac{\pi}{3}\right) & 1 & \sqrt{2}\sin\left(\theta - \frac{\pi}{3}\right) \\ -\sqrt{2}\cos\left(\theta + \frac{\pi}{3}\right) & 1 & \sqrt{2}\sin\left(\theta + \frac{\pi}{3}\right) \end{pmatrix}, \tag{6.7}$$

where θ is a rational multiple of π, with its exact value completely determined by group-theoretical considerations. This result is consistent with comprehensive scanning over the finite groups and possible symmetry breaking patterns [262, 263, 379]. We obtain the following expressions for the lepton mixing angles

$$\sin^2\theta_{13} = \frac{2}{3}\sin^2\theta, \quad \sin^2\theta_{12} = \frac{1}{1+2\cos^2\theta}, \quad \sin^2\theta_{23} = \frac{2\sin^2(\theta - \frac{\pi}{3})}{1+2\cos^2\theta}\,. \tag{6.8}$$

Combining the above equations one sees that the three mixing angles depend on a single parameter θ, consequently the following relations must be satisfied,

$$3\sin^2\theta_{12}\cos^2\theta_{13} = 1, \qquad \sin^2\theta_{23} = \frac{1}{2} \pm \frac{1}{2}\tan\theta_{13}\sqrt{2-\tan^2\theta_{13}}\,. \tag{6.9}$$

Using the best fit value of the reactor mixing angle $\sin^2\theta_{13} \simeq 0.022$ for NO [24, 25] we get

$$\sin^2\theta_{12} \simeq 0.341, \quad \sin^2\theta_{23} \simeq 0.605 \text{ or } 0.395\,. \tag{6.10}$$

The value of solar mixing angle is within the experimentally preferred 2σ region, the atmospheric angle $\sin^2\theta_{23} \simeq 0.605$ lies in the 3σ region, while another possible value $\sin^2\theta_{23} \simeq 0.395$ is disfavored by the present data [24, 25]. As regards the CP violation phases, the Majorana phases are unconstrained [21] and the Dirac CP phase vanishes

$$\sin\delta_{CP} = 0\,. \tag{6.11}$$

Notice that current measurements still do not provide a fully robust global CP determination [24, 25].

In order to reproduce the experimentally favored mixing angles the minimal flavour symmetry group is $G_f = (Z_{18} \times Z_6) \rtimes S_3$ [260, 263] for Majorana neutrinos, while for Dirac neutrinos the minimal flavour group is $G_f = (Z_9 \times Z_3) \rtimes S_3$ [263]. Hence the order of the flavour symmetry group should be at least 648 and 162 for Majorana and Dirac neutrinos, respectively, leading to a realistic value of $\theta = \pi/18$. In concrete models, the residual symmetries G_l and G_ν are spoiled by higher order terms involving flavon fields. These induce corrections suppressed by the flavon VEVs with respect to the flavor scale.

6.1.2 Partially preserved residual symmetry $G_\nu = Z_2$

The idea of partially preserved residual symmetry was proposed [381–384] in order to accommodate a non-zero Dirac CP violation phase δ^ℓ and degrade the order of the flavour symmetry. In this approach, part of the residual symmetry of the neutrino mass matrix arises from the assumed flavour symmetry at

[21] One can predict the values of Majorana phases by including the generalized CP symmetry. For instance, for the flavor symmetry $\Delta(6n^2) = (Z_n \times Z_n) \rtimes S_3$ in combination with CP, the Majorana phase ϕ_{12} can take several discrete values for each n, namely ϕ_{13} is a multiple of $\pi/2$ [380] if residual K_4 flavor symmetry and CP symmetry are preserved by neutrino mass matrix.

high the energy scale. A widely studied scenario is that a Z_2 (instead of K_4) subgroup is preserved in the neutrino sector, i.e. the residual group is $G_\nu = Z_2^{g_\nu}$, where g_ν refers to the generator of G_ν. Note that the presentation rule of the cyclic group Z_n^g is $Z_n^g \equiv \left\{1, g, g^2, \ldots, g^{n-1}\right\}$.

The invariance of the neutrino mass matrix under $Z_2^{g_\nu}$ requires that Eq. (6.2) holds. Consequently the residual flavour symmetry imposes the following restriction on the unitary transformation

$$U_\nu^\dagger \rho_{\mathbf{3}}(g_\nu) U_\nu = \pm P_\nu^T \mathrm{diag}(1,-1,-1) P_\nu \,, \tag{6.12}$$

where P_ν is a permutation matrix, and we have taken into account that the eigenvalues of $\rho_{\mathbf{3}}(g_\nu)$ is $+1$ or -1, since g_ν is of order two. Let us denote U_0 as a diagonalization matrix of $\rho_{\mathbf{3}}(g_\nu)$ with

$$U_0^\dagger \rho_{\mathbf{3}}(g_\nu) U_0 = \pm \mathrm{diag}(1,-1,-1) \,, \tag{6.13}$$

Then the unitary transformation U_ν would be of the form

$$U_\nu = U_0 U_{23}(\theta, \delta) P_\nu \,, \tag{6.14}$$

where $U_{23}(\theta, \delta)$ is a block diagonal complex unitary rotation,

$$U_{23}(\theta,\delta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta e^{-i\delta} \\ 0 & -\sin\theta e^{i\delta} & \cos\theta \end{pmatrix} . \tag{6.15}$$

Notice that, since the residual Z_2 flavour symmetry can not fully distinguish the three neutrino families in this case, only one column of U_ν is numerically fixed. In the charged lepton sector, a residual subgroup G_l is preserved, so that the unitary transformation U_l obeys the condition $U_l^\dagger \rho_{\mathbf{3}}(g_l) U_l = \mathrm{diag}(e^{i\alpha_e}, e^{i\alpha_\mu}, e^{i\alpha_\tau})$ as shown in Eq. (6.4), where $\alpha_{e,\mu,\tau}$ are rational multiples of π. Thus the assumed residual symmetry allows us to determine the lepton mixing matrix in terms of two free parameters as

$$U = U_l^\dagger U_0 U_{23}(\theta, \delta) P_\nu \,. \tag{6.16}$$

Hence only one column is fixed by residual symmetry in this case, and this is dubbed semi-direct approach in [88]. For example, if the flavour group is S_4 and the residual symmetries are chosen as $G_\nu = Z_2^{SU}$ and $G_l = Z_3^T$, then $U_0 = U_{TBM}$ and the lepton mixing is

$$U = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & \sqrt{2}\cos\theta & \sqrt{2}\, e^{-i\delta}\sin\theta \\ -1 & \sqrt{2}\cos\theta - \sqrt{3}\, e^{i\delta}\sin\theta & \sqrt{3}\cos\theta + \sqrt{2}\, e^{-i\delta}\sin\theta \\ -1 & \sqrt{2}\cos\theta + \sqrt{3}\, e^{i\delta}\sin\theta & -\sqrt{3}\cos\theta + \sqrt{2}\, e^{-i\delta}\sin\theta \end{pmatrix} . \tag{6.17}$$

We see that the first column is $(2/\sqrt{6}, -1/\sqrt{6}, -1/\sqrt{6})^T$ which is in common with the TBM mixing pattern [296–298]. In fact, this is exactly the TM1 lepton mixing matrix in Eq. (3.7) [316–318]. In

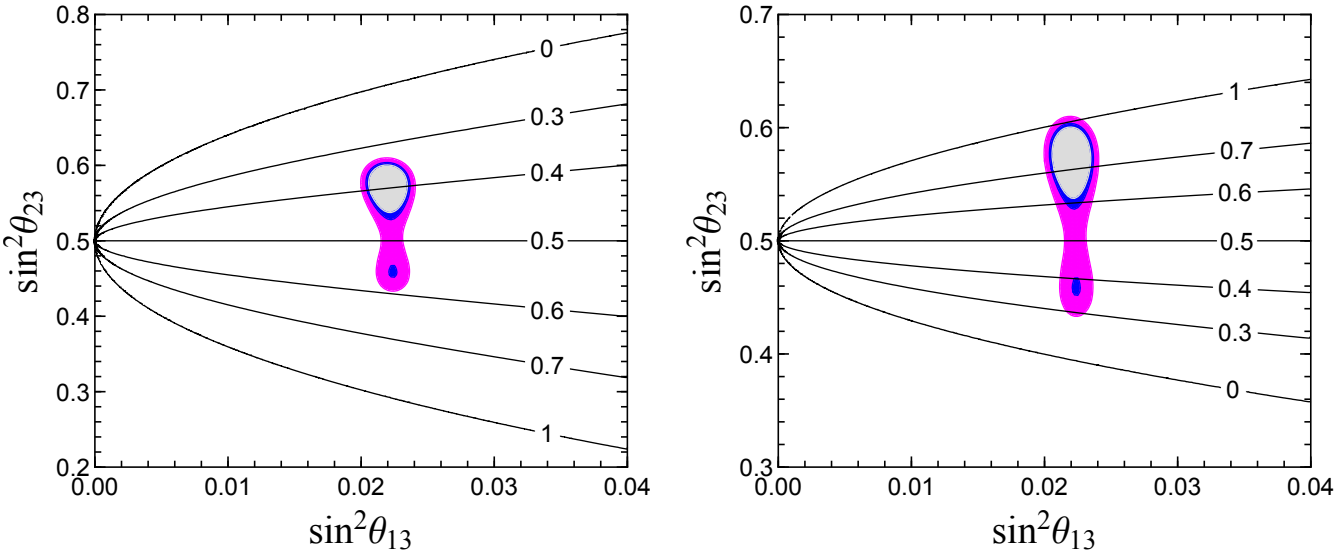


Figure 18: Contours of $|\delta_{CP}/\pi|$ in the $\sin^2\theta_{23}-\sin^2\theta_{13}$ plane. The left panel is for the case that the fixed column is $(2/\sqrt{6}, -1/\sqrt{6}, -1/\sqrt{6})^T$ (TM1), and the right panel is for $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^T$ (TM2). The global-fit allowed oscillation regions corresponding to 90%, 95% and 99% confidence level are shown [24,25]

figure 18 we plot the contour of $|\delta_{CP}|$ in the plane $\sin^2\theta_{23}$ versus $\sin\theta_{13}$. The value of the Dirac CP phase δ_{CP} is not too constrained, as long as the atmospheric mixing angle θ_{23} remains poorly measured.

On the other hand, if the residual groups are $G_\nu = Z_2^S$ and $G_l = Z_3^T$ for either A_4 or S_4 flavour symmetry, one column of the lepton mixing matrix is enforced to be $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})^T$ with,

$$U = \frac{1}{\sqrt{6}} \begin{pmatrix} 2\cos\theta & \sqrt{2} & 2e^{-i\delta}\sin\theta \\ -\cos\theta - \sqrt{3}\, e^{i\delta}\sin\theta & \sqrt{2} & \sqrt{3}\,\cos\theta - e^{-i\delta}\sin\theta \\ -\cos\theta + \sqrt{3}\, e^{i\delta}\sin\theta & \sqrt{2} & -\sqrt{3}\,\cos\theta - e^{-i\delta}\sin\theta \end{pmatrix}, \tag{6.18}$$

which is exactly the tri-maximal mixing pattern TM2 in Eq. (3.10) [252, 321–323]. The predictions for the Dirac phase δ_{CP} are shown in figure 18. The figure refers to the case of normal mass ordering, with very similar results for the case of inverse-ordered masses.

6.2 Combining flavour and CP symmetry

We saw how a non-vanishing Dirac CP phase can be obtained if a Z_2 rather than the Klein group K_4 is the residual flavour symmetry preserved by the neutrino mass matrix. However, δ_{CP} can vary within a wide region. Moreover, the Majorana phases can not be predicted just from the flavour symmetry. In order to understand the CP violating phases, one can impose a generalized CP (gCP) symmetry [341–343], see [264,344,385] for discussions of gCP to the lepton flavor mixing problem. A simple example is the $\mu-\tau$ reflection symmetry [252–256], see Ref. [386] for review. The $\mu-\tau$ reflection symmetry is a combination of the canonical CP transformation with the $\mu-\tau$ exchange symmetry, which exchanges a muon (tau)

neutrino with a tau (muon) antineutrino,

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} \mapsto \begin{pmatrix} \nu_e^c \\ \nu_\tau^c \\ \nu_\mu^c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \nu_e^c \\ \nu_\mu^c \\ \nu_\tau^c \end{pmatrix} . \tag{6.19}$$

Notice that the generalized CP transformation matrix is not diagonal in family space. In the charged lepton mass basis, the neutrino mass matrix invariant under $\mu - \tau$ reflection has the following form

$$m_\nu = \begin{pmatrix} a & b & b^* \\ b & c & d \\ b^* & d & c^* \end{pmatrix} , \tag{6.20}$$

where a and d are real, b and c are complex numbers. As a result, both atmospheric mixing angle θ_{23} and Dirac CP phase would be maximal, while Majorana phases take CP conserving values. These should be contrasted with present neutrino oscillation data [24, 25]. Deviations of θ_{23} and δ_{CP} from maximal values can easily arise from a more general $\mu - \tau$ reflection symmetry acting on the neutrino sector. Such CP transformation is of the following form [257],

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} \mapsto \begin{pmatrix} \nu_e^c \\ \cos\Theta\, \nu_\mu^c + i \sin\Theta\, \nu_\tau^c \\ \cos\Theta\, \nu_\tau^c + i \sin\Theta\, \nu_\mu^c \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Theta & i\sin\Theta \\ 0 & i\sin\Theta & \cos\Theta \end{pmatrix} \begin{pmatrix} \nu_e^c \\ \nu_\mu^c \\ \nu_\tau^c \end{pmatrix} , \tag{6.21}$$

where the angle Θ characterizes the CP transformation. It reduces to the $\mu - \tau$ reflection symmetry in the limit of $\Theta = \pi/2$. The atmospheric mixing angle θ_{23} and Dirac phase δ_{CP} are predicted to be strongly correlated as follows [257],

$$\sin^2 \delta_{\rm CP} \sin^2 2\theta_{23} = \sin^2 \Theta \, , \tag{6.22}$$

while the Majorana phases take on trivial CP conserving values. The correlation in Eq. (6.22) allows us to predict the range of the Dirac CP violating phase $|\sin\delta_{CP}|$ as a function of the parameter Θ as shown in figure 19. Note that the sign of δ_{CP} can not be fixed.

6.2.1 Mathematical consistency

Family symmetries can be generated if one successively performs two generalized CP transformations. Hence generalized CP symmetries can be thought of as associated to some underlying flavour symmetry. A convenient strategy for defining such CP transformations is to start from a flavour symmetry group G_f and find all possible CP transformations H_{CP} which can generate the given flavour group transformations. The purpose of this section is to determine the restricted lepton mixing matrices that can be obtained from discrete flavour and CP symmetries.

It is highly non-trivial to define a CP transformation consistently in the presence of a family symmetry G_f [264, 344, 385, 387]. Let us consider a set of fields φ in a generic irreducible representation $\mathbf{r}$ of G_f,

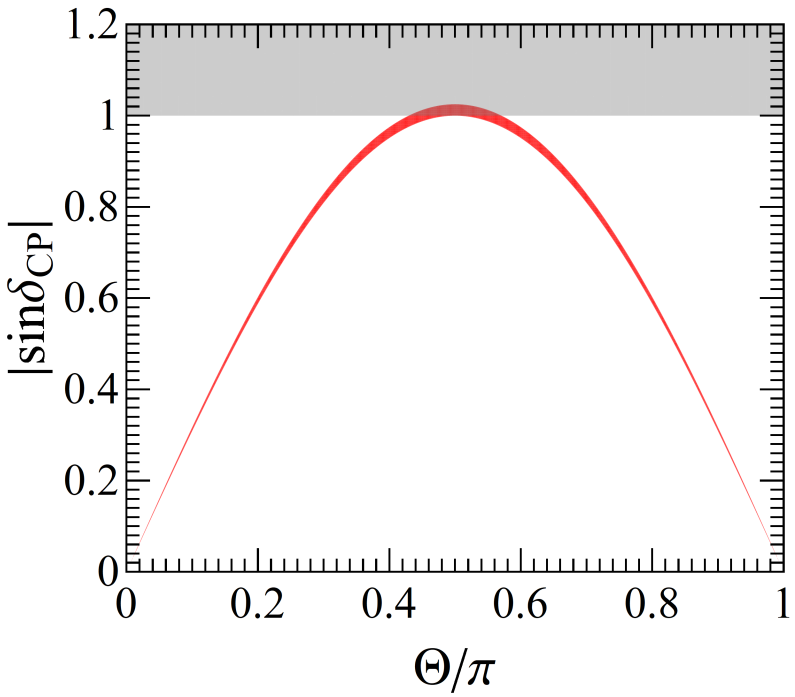


Figure 19: $|\sin\delta_{CP}|$ regions versus Θ characterizing the generalized $\mu-\tau$ reflection, where θ_{23} is required to lie in its 3σ allowed range [24, 25].

transforming under the action of G_f as

$$\varphi(x) \xrightarrow{G_f} \rho_{\mathbf{r}}(g)\varphi(x), \qquad g \in G_f \ , \tag{6.23}$$

where $\rho_{\mathbf{r}}(g)$ denotes the representation matrix for any element g in the irreducible representation $\mathbf{r}$. The generalized CP acts on φ as

$$\varphi(x) \xrightarrow{\mathcal{CP}} X_{\mathbf{r}}\, \varphi^*(x_{\mathcal{P}})\,, \tag{6.24}$$

where $x_{\mathcal{P}} = (t, -\vec{x})$, and $X_{\mathbf{r}}$ is the CP transformation matrix in flavor space, assumed to be a unitary matrix, so as to leave the kinetic term invariant [22]. A physical CP transformation should map each field $\varphi(x)$ in any irreducible representation $\mathbf{r}$ of G_f into its complex conjugate $\varphi^*(x_{\mathcal{P}})$ in the complex representation $\mathbf{r}^*$ [387]. If we first perform a CP transformation, then apply a flavour symmetry transformation, and subsequently an inverse CP transformation we obtain

$$\varphi(x) \xrightarrow{\mathcal{CP}} X_{\mathbf{r}}\, \varphi^*(x_{\mathcal{P}}) \xrightarrow{G_f} X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g)\varphi^*(x_{\mathcal{P}}) \xrightarrow{\mathcal{CP}^{-1}} X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g)X^{-1}_{\mathbf{r}}\varphi(x)\,. \tag{6.25}$$

As shown in figure 20, the theory should still be invariant since it is invariant under each transformation individually. To make the theory consistent the resulting net transformation should be equivalent to a flavour symmetry transformation $\rho_{\mathbf{r}}(g')$ of some flavour group element g', i.e.

$$X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g)X^{-1}_{\mathbf{r}} = \rho_{\mathbf{r}}(g'), \qquad g, g' \in G_f \tag{6.26}$$

[22] When φ denotes a spinor, the obvious action of CP on the spinor indices is understood.

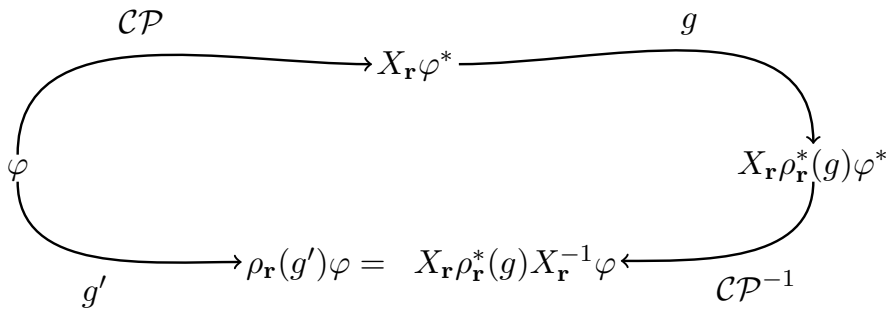


Figure 20: The consistency condition for flavour and generalized CP symmetries.

where the elements g and g' are independent of the representation $\mathbf{r}$. Eq. (6.26) is the important consistency condition which must be fulfilled for all irreducible representations of G_f in order to ensure generalized CP and flavour symmetry invariance simultaneously. If the condition Eq. (6.26) is not fulfilled, the group G_f is not the full symmetry group of the Lagrangian, and one would have to consider a larger group, which closes under CP transformations. The allowed form of the generalized CP transformations is strongly restricted by the consistency condition of Eq. (6.26).

In practice, it suffices to consider the consistency conditions for the generators of G_f. It is remarkable that both $e^{i\theta}X_{\mathbf{r}}$ and $\rho_{\mathbf{r}}(h)X_{\mathbf{r}}$ satisfy the consistency condition Eq. (6.26) for a generalized CP transformation $X_{\mathbf{r}}$, where θ is real and h is any element of G_f. For a well-defined CP transformation $X_{\mathbf{r}}$, $\rho_{\mathbf{r}}(h)X_{\mathbf{r}}$ is also a viable CP transformation for any $h \in G_f$. The two CP transformations $X_{\mathbf{r}}$ and $\rho_{\mathbf{r}}(h)X_{\mathbf{r}}$ differ by a flavour symmetry transformation $\rho_{\mathbf{r}}(h)$. Since the latter is certainly a symmetry of the Lagrangean, these two CP transformations are indistinguishable. Moreover, Eq. (6.26) implies that the generalized CP transformation $X_{\mathbf{r}}$ maps the group element g into g' and the flavour group multiplication is preserved under this mapping, i.e. $X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g_1g_2)X^{-1}_{\mathbf{r}} = X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g_1)X^{-1}_{\mathbf{r}}X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g_2)X^{-1}_{\mathbf{r}}$. Therefore the CP transformation $X_{\mathbf{r}}$ defines a homomorphism $\mathfrak{u} : g \to g'$, $g, g' \in G_f$ of the family symmetry group G_f. Note that the homomorphism $\mathfrak{u}'$ associated with the CP transformation $\rho_{\mathbf{r}}(h)X_{\mathbf{r}}$ for any $h \in G_f$ is related to $\mathfrak{u}$ by conjugation:

$$\mathfrak{u}'(g) = h\mathfrak{u}(g)h^{-1}, \quad \forall\ h, g \in G_f\,. \tag{6.27}$$

Notice that when $\rho_{\mathbf{r}}$ is a faithful representation, the elements g and g' have the same order, the mapping defined in Eq. (6.26) is bijective, and thus the associated CP transformation becomes an automorphism, see Ref. [344] for a more formal treatment. Furthermore, taking trace on both sides of the consistency condition in Eq. (6.26), we find that the group characters $\chi_{\mathbf{r}}$ fulfill

$$\chi_{\mathbf{r}}(g') = \mathrm{tr}[\rho_{\mathbf{r}}(g')] = \mathrm{tr}[X_{\mathbf{r}}\rho^*_{\mathbf{r}}(g)X^{-1}_{\mathbf{r}}] = \mathrm{tr}[\rho^*_{\mathbf{r}}(g)] = \mathrm{tr}[\rho^\dagger_{\mathbf{r}}(g)] = \chi_{\mathbf{r}}(g^{-1})\,. \tag{6.28}$$

Hence g' and g^{-1} should be in the same conjugacy class, that is to say the CP transformation corresponds to a class-inverting automorphism of G_f. As a consequence, when determining the generalized CP transformation compatible with a flavour symmetry group G_f it is sufficient to focus on the class-inverting automorphisms. Because g and g' in Eq. (6.26) are generally different group elements, flavour transfor-

mations and CP transformations in general do not commute. Hence the mathematical structure of the group comprising G_f and generalized CP is in general a semi-direct product, and the full symmetry is $G_f \rtimes H_{CP}$ [264], where H_{CP} is the group generated by generalized CP transformations. The semi-direct product would reduce to a direct product in the case $g = g'$.

6.2.2 Implications of residual flavour and CP symmetries

The presence of generalized CP allows for more symmetry breaking patterns than flavor symmetry alone. In this approach, the parent flavour and CP symmetries are broken down to different residual subgroups $G_l \rtimes H^l_{CP}$ and $G_\nu \rtimes H^\nu_{CP}$ in the charged lepton and neutrino sectors respectively. The mismatch between the residual symmetries $G_l \rtimes H^l_{CP}$ and $G_\nu \rtimes H^\nu_{CP}$ gives rise to a certain lepton mixing pattern. It is remarkable that the lepton flavour mixing is fixed by the group structure of $G_f \rtimes H_{CP}$ and the residual symmetries [265,267]. The details of the breaking mechanisms realizing the assumed residual symmetries are irrelevant. In the following, we present the formalism to determine the residual symmetries on the charged lepton and neutrino mass matrices. We shall assume neutrinos to be Majorana particles, thus the residual flavour symmetry G_ν is a Z_2 or Klein subgroup K_4 (Dirac neutrinos can be discussed in a very similar way). As usual the three generations of left-handed leptons are assumed to transform as a faithful irreducible triplet $\mathbf{3}$ of the G_f symmetry.

For the residual symmetry $G_l \rtimes H^l_{CP}$ to hold, the Hermitian combination $m_l^\dagger m_l$ of the charged lepton mass matrix should be invariant under the action of $G_l \rtimes H^l_{CP}$, i.e.,

$$\rho^\dagger_{\mathbf{3}}(g_l) m_l^\dagger m_l \rho_{\mathbf{3}}(g_l) = m_l^\dagger m_l, \qquad g_l \in G_l\,, \tag{6.29}$$

$$X^\dagger_{l\mathbf{3}} m_l^\dagger m_l X_{l\mathbf{3}} = (m_l^\dagger m_l)^*, \qquad X_{l\mathbf{3}} \in H^l_{CP}\,, \tag{6.30}$$

where the charged lepton mass matrix m_l is given in the convention in which the left-handed (right-handed) fields are on the right-hand (left-hand) side of m_l. We denote the unitary diagonalization matrix of $m_l^\dagger m_l$ as U_l which satisfies $U_l^\dagger m_l^\dagger m_l U_l = \mathrm{diag}(m_e^2, m_\mu^2, m_\tau^2)$. Once the residual symmetry $G_l \rtimes H^l_{CP}$ and the triplet representation $\mathbf{3}$ are specified, the explicit form of $m_l^\dagger m_l$ can be constructed from Eqs. (6.29, 6.30) in a straightforward way, and thus U_l can be determined. In fact, one can also directly extract the restrictions on U_l from Eqs. (6.29,6.30) without working out the explicit form of $m_l^\dagger m_l$ as follows

$$U_l^\dagger \rho_{\mathbf{3}}(g_l) U_l = \rho^{diag}_{\mathbf{3}}(g_l)\,, \qquad U_l^\dagger X_{l\mathbf{3}} U_l^* = X^{diag}_{l\mathbf{3}}\,, \tag{6.31}$$

where both $\rho^{diag}_{\mathbf{3}}(g_l)$ and $X^{diag}_{l\mathbf{3}}$ are diagonal phase matrices.

The first identity in Eq. (6.31) comes from imposing the residual flavour symmetry G_l, and the latter arises from the residual CP symmetry H^l_{CP}. We see that both $\rho_{\mathbf{3}}(g_l)$ and $m_l^\dagger m_l$ can be diagonalized by the same unitary matrix U_l, and the residual CP transformation $X_{l\mathbf{3}} = U_l X^{diag}_{l\mathbf{3}} U_l^T$ should be a symmetric unitary matrix. Moreover, Eq. (6.31) implies that the residual flavour and CP symmetries should satisfy

the following restricted consistency conditions:

$$X_{l\mathbf{r}}\rho^*_{\mathbf{r}}(g_l)X^{-1}_{l\mathbf{r}}=\rho_{\mathbf{r}}(g_l^{-1}),\qquad g_l\in G_l\,, \tag{6.32}$$

If the residual flavour symmetry G_l distinguishes the three families, the inclusion of generalized CP H^l_{CP} would not add any information as far as lepton mixing is concerned [335,388]. Likewise, the requirement that $G_\nu \rtimes H^\nu_{CP}$ is preserved in the neutrino sector implies that the neutrino mass matrix m_ν must be invariant under the action of $G_\nu \rtimes H^\nu_{CP}$, i.e.,

$$\begin{aligned}&\rho^T_{\mathbf{3}}(g_\nu)m_\nu\rho_{\mathbf{3}}(g_\nu)=m_\nu,\qquad g_\nu\in G_\nu\,,\\&X^T_{\nu\mathbf{3}}m_\nu X_{\nu\mathbf{3}}=m^*_\nu,\qquad X_{\nu\mathbf{3}}\in H^\nu_{CP}\,.\end{aligned} \tag{6.33}$$

Since G_ν is a Z_2 or K_4 subgroup for Majorana neutrinos, the eigenvalues of the representation matrix $\rho_{\mathbf{3}}(g_\nu)$ can only be $+1$ or -1. The diagonalization matrix of m_ν is denoted as U_ν and satisfies $U^T_\nu m_\nu U_\nu = \mathrm{diag}(m_1,m_2,m_3)$. From Eq. (6.33), we find that the residual symmetry $G_\nu \rtimes H^\nu_{CP}$ leads to the following restrictions on the unitary transformation U_ν,

$$U^\dagger_\nu\rho_{\mathbf{3}}(g_\nu)U_\nu=\mathrm{diag}(\pm1,\pm1,\pm1)\,,\qquad U^\dagger_\nu X_{\nu\mathbf{3}}U^*_\nu=\mathrm{diag}(\pm1,\pm1,\pm1)\,. \tag{6.34}$$

Hence the residual CP transformation $X_{\nu\mathbf{3}}$ should be a symmetric matrix [264,265,267] and the restricted consistency condition in the neutrino sector is

$$X_{\nu\mathbf{r}}\rho^*_{\mathbf{r}}(g_\nu)X^{-1}_{\nu\mathbf{r}}=\rho_{\mathbf{r}}(g_\nu),\quad g_\nu\in G_\nu,\quad X_{\nu\mathbf{r}}\in H^\nu_{CP}\,, \tag{6.35}$$

which implies that the residual flavour symmetry G_ν and residual CP symmetry H^ν_{CP} in the neutrino sector commute with each other. Consequently, for Majorana neutrinos the mathematical structure of the residual symmetry is a direct product $G_\nu \times H^\nu_{CP}$. Solving the constraints Eqs. (6.31, 6.34) imposed by the residual symmetry, one can fix the unitary transformations U_l, U_ν and hence the lepton mixing matrix, as

$$U=U^\dagger_l U_\nu\,. \tag{6.36}$$

In the following, we present several different choices for the residual scheme and the corresponding predictions for the lepton mixing matrix.

6.3 Predictive scenarios with CP symmetry

Assuming a generalized CP symmetry, the Majorana CP violation phases can be predicted and we have more choices for the possible residual symmetries. In particular, some scenarios are quite predictive, with the lepton mixing matrix elements depending only on few free parameters, as shown in table 1. For notational simplicity, we denote $\mathcal{G}_l \equiv G_l \rtimes H^l_{CP}$ and $\mathcal{G}_\nu \equiv G_\nu \rtimes H^\nu_{CP}$ which are the residual symmetries

of the charged lepton and neutrino sectors respectively [23]. We have assumed that the residual flavour symmetry Z_n with $n \geq 3$ is sufficient to distinguish the three generations of charged leptons, otherwise it can be taken to be a product of cyclic groups. The unitary transformations Σ_l and Σ_ν are the Takagi factorizations of the residual CP transformations H^l_{CP} and H^ν_{CP} respectively, in the presence of residual CP. They should also diagonalize the residual flavour symmetry transformations G_l and G_ν respectively. Moreover, $R_{23}(\theta)$ is a rotation matrix in the (23)-plane, Eq. (4.45), while $O_3(\theta_1, \theta_2, \theta_3)$ is a general three-dimensional rotation matrix, Eq. (4.41). Both P_l and P_ν are permutation matrices since the neutrino and charged lepton masses are not constrained by the residual symmetry. Furthermore, Q_l and Q_ν are diagonal phase matrices, Q_l can be absorbed into the charged lepton fields, the non-zero elements of Q_ν encoding the CP parity of neutrinos are equal to ± 1 and $\pm i$ and they can shift the Majorana phases ϕ_{12} and ϕ_{13} by $\pm\pi/2$ or $\pm\pi$. In what follows, we present two predictive scenarios for illustration.

$\mathcal{G}_l$	$\mathcal{G}_\nu$	U	number of parameters
Z_n	$K_4 \times CP$	$Q_l^\dagger P_l^T \Sigma_l^\dagger \Sigma_\nu P_\nu Q_\nu$	0
Z_n	$Z_2 \times CP$	$Q_l^\dagger P_l^T \Sigma_l^\dagger \Sigma_\nu R_{23}(\theta) P_\nu Q_\nu$	1
$Z_2 \times CP$	$K_4 \times CP'$	$Q_l^\dagger P_l^T R_{23}^T(\theta_l) \Sigma_l^\dagger \Sigma_\nu P_\nu Q_\nu$	
$Z_2 \times CP$	$Z_2 \times CP'$	$Q_l^\dagger P_l^T R_{23}^T(\theta_l) \Sigma_l^\dagger \Sigma_\nu R_{23}(\theta_\nu) P_\nu Q_\nu$	2
Z_2	$K_4 \times CP$	$Q_l^\dagger P_l^T U_{23}^\dagger(\theta_l, \delta_l) \Sigma_l^\dagger \Sigma_\nu P_\nu Q_\nu$	
Z_n	CP	$Q_l^\dagger P_l^T \Sigma_l^\dagger \Sigma_\nu O_3(\theta_1, \theta_2, \theta_3) Q_\nu$	3
CP	$K_4 \times CP'$	$Q_l^\dagger O_3^T(\theta_1, \theta_2, \theta_3) \Sigma_l^\dagger \Sigma_\nu Q_\nu$	
Z_2	$Z_2 \times CP$	$Q_l^\dagger P_l^T U_{23}^\dagger(\theta_l, \delta_l) \Sigma_l^\dagger \Sigma_\nu R_{23}(\theta_\nu) P_\nu Q_\nu$	

Table 1: Possible choices of residual symmetries and corresponding predictions for lepton mixing matrix, when the latter depends on up to three free parameters. If the residual CP is absent from $\mathcal{G}_\nu$, the Majorana CP phases are not restricted.

6.3.1 $\mathcal{G}_l = Z_n$ $(n \geq 3)$, $\mathcal{G}_\nu = Z_2 \times CP$

We consider the scenario where the three families of left-handed leptons transform inequivalently as one-dimensional representations under the residual flavour symmetry $G_l = Z_n$ with $n \geq 3$, so that they are distinguished by the abelian subgroup G_l. The representation matrix $\rho_3(g_l)$ can be diagonalized by a unitary matrix Σ_l fulfilling $\Sigma_l^\dagger \rho_{\mathbf{3}}(g_l) \Sigma_l = \rho_{\mathbf{3}}^{diag}(g_l)$, where the three columns of Σ_l are formed by the three eigenvectors of $\rho_{\mathbf{3}}(g_l)$ and Σ_l is determined up to an arbitrary diagonal unitary matrix Q_l and a permutation matrix P_l. Eq. (6.31) implies that the charged lepton diagonalization matrix U_l coincides with Σ_l, i.e.,

$$U_l = \Sigma_l P_l Q_l \,. \tag{6.37}$$

Concerning the neutrino sector, since $X_{\nu\mathbf{3}}$ is a symmetric and unitary matrix, by performing the Takagi factorization $X_{\nu\mathbf{3}}$ can be written as

$$X_{\nu\mathbf{3}} = \Sigma_\nu \Sigma_\nu^T \,, \quad \text{with} \quad \Sigma_\nu^\dagger \rho_{\mathbf{3}}(g_\nu) \Sigma_\nu = \pm \text{diag}(1, -1, -1) \,. \tag{6.38}$$

[23] In concrete model building, one could possibly have more branches of residual symmetry [389–392].

The procedure of obtaining the unitary matrix Σ_ν has been given in [345]. The neutrino diagonalization matrix U_ν satisfying the conditions in Eq. (6.34) is determined to be of the following form [264, 345]

$$U_\nu = \Sigma_\nu R_{23}(\theta) P_\nu Q_\nu \,, \tag{6.39}$$

where $R_{23}(\theta)$ stands for a rotation matrix through an angle θ in the (23)-plane and it takes the form of Eq. (4.45), and P_ν is a generic permutation matrix. Here Q_ν in Eq. (6.39) is a diagonal matrix with elements equal to ± 1 and $\pm i$. Hence, without loss of generality it can be given as,

$$Q_\nu = \begin{pmatrix} 1 & 0 & 0 \\ 1 & i^{k_1} & 0 \\ 0 & 0 & i^{k_2} \end{pmatrix} \tag{6.40}$$

with $k_{1,2} = 0, 1, 2, 3$. Hence the lepton mixing matrix U is of the form [264, 265, 345]

$$U = U_l^\dagger U_\nu = Q_l^\dagger P_l^T \Sigma_l^\dagger \Sigma_\nu R_{23}(\theta) P_\nu Q_\nu \,. \tag{6.41}$$

Notice that, as usual, the phase matrix Q_l can be absorbed into the charged lepton fields [30], while the effect of Q_ν is to shift the Majorana CP phases ϕ_{12} and ϕ_{13} by integral multiples of $\pi/2$. It is remarkable that the lepton mixing matrix is constrained to only depend on a single free parameter θ, whose value can be fixed by the precisely measured reactor mixing angle θ_{13}. We can therefore predict the values of the other two mixing angles θ_{12}, θ_{23} as well as the CP violation phases. The lepton mixing matrix is determined up to possible permutations of rows and columns, since both neutrino and charged lepton masses are not constrained in this approach. Moreover, the Z_2 residual flavour symmetry can only distinguish one neutrino family from the other two generations, so that only one column of the lepton mixing matrix is fixed through the choice of residual symmetry, as can be seen from Eq. (6.41). Comparing with the oscillation data in Eqs. (1.23, 1.24), one can determine the phenomenologically allowed permutation matrices P_l and P_ν.

It follows from the above that, for any postulated residual symmetry subgroups G_l and $Z_2 \times CP$, the lepton mixing matrix can be extracted by using Eq. (6.41) in a simple manner, the recipe is summarized in figure 21. The predicted lepton mixing matrix only depends on the structure of the symmetry group $G_f \rtimes H_{CP}$ and the assumed residual symmetry, regardless of the underlying dynamics which breaks $G_f \rtimes H_{CP}$ down to the residual subgroup.

As an illustration, we consider the S_4 flavor symmetry in combination with the generalized CP symmetry. The group theory and representation of S_4 are given in B. The outer automorphism group of S_4 is trivial [320,344], all the automorphisms of S_4 are inner automorphisms which are group conjugations. Without loss of generality it is sufficient to consider the representative automorphism $\mathfrak{u} : (s, t, u) \to (s, t^{-1}, u)$, where s, t and u are the generators of S_4, the corresponding generalized CP transformation $X^0_{\mathbf{r}}$ is determined

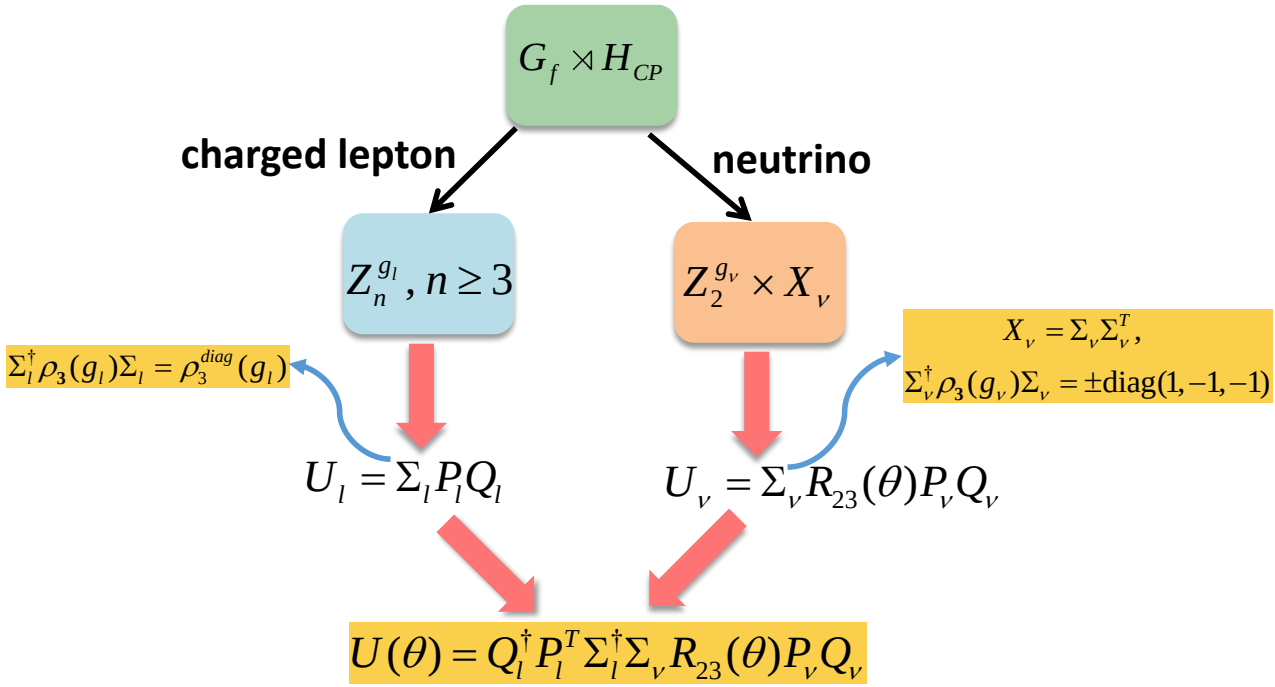


Figure 21: The model independent predictions for lepton mixing matrix for the preserved residual symmetry abelian subgroup G_l and $Z_2 \times CP$ in the charged lepton and neutrino sectors respectively.

by the following consistency equations

$$\begin{aligned} X^0_{\mathbf{r}} \rho^*_{\mathbf{r}}(s) \left(X^0_{\mathbf{r}}\right)^{-1} &= \rho_{\mathbf{r}}(\mathfrak{u}(s)) = \rho_{\mathbf{r}}(s)\,, \\ X^0_{\mathbf{r}} \rho^*_{\mathbf{r}}(t) \left(X^0_{\mathbf{r}}\right)^{-1} &= \rho_{\mathbf{r}}(\mathfrak{u}(t)) = \rho_{\mathbf{r}}(t^2)\,, \\ X^0_{\mathbf{r}} \rho^*_{\mathbf{r}}(u) \left(X^0_{\mathbf{r}}\right)^{-1} &= \rho_{\mathbf{r}}(\mathfrak{u}(u)) = \rho_{\mathbf{r}}(u)\,. \end{aligned} \tag{6.42}$$

From the explicit form of the representation matrices in Eqs. (B.3, B.4, B.5), it follows that $X^0_{\mathbf{r}}$ is the unit matrix up to an arbitrary overall phase,

$$X^0_{\mathbf{r}} = 1\,. \tag{6.43}$$

Including the family symmetry transformation, the generalized CP transformation consistent with the S_4 symmetry is given by [320, 325, 335]

$$X_{\mathbf{r}} = \rho_{\mathbf{r}}(g) X^0_{\mathbf{r}} = \rho_{\mathbf{r}}(g), \quad g \in S_4\,. \tag{6.44}$$

It was found that three sets of residual symmetries are compatible with the experimental data [264]. Notice that a fourth residual symmetry with $G_l = Z_3^t$, $G_\nu = Z_2^{su}$, $X_\nu = 1$ is not phenomenologically viable, since the reactor and atmospheric angles can not be accommodated simultaneously.

(i) $G_l = Z_3^t$, $G_\nu = Z_2^s$, $X_\nu = 1$

In this case, the unitary transformation Σ_l is the unit matrix since the representation matrix $\rho_{\mathbf{3}}(t)$ is diagonal, and the Takagi factorization matrix Σ_ν is found to be

$$\Sigma_\nu = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & 2 & 0 \\ \sqrt{2} & -1 & \sqrt{3} \\ \sqrt{2} & -1 & -\sqrt{3} \end{pmatrix} . \tag{6.45}$$

Hence the lepton mixing matrix is determined to be

$$U = \frac{1}{\sqrt{6}} \begin{pmatrix} 2\cos\theta & \sqrt{2} & 2\sin\theta \\ -\cos\theta - \sqrt{3}\sin\theta & \sqrt{2} & \sqrt{3}\cos\theta - \sin\theta \\ -\cos\theta + \sqrt{3}\sin\theta & \sqrt{2} & -\sqrt{3}\cos\theta - \sin\theta \end{pmatrix} Q_\nu \,, \tag{6.46}$$

where we have taken $P_\nu = P_{12}$ such that $(1,1,1)^T/\sqrt{3}$ is in the second column, in order to be consistent with experimental data. We can extract the lepton mixing parameters as follows,

$$\begin{aligned} &\sin^2\theta_{13} = \frac{2}{3}\sin^2\theta \,, \quad \sin^2\theta_{23} = \frac{1}{2} - \frac{\sqrt{3}\sin 2\theta}{2(2+\cos 2\theta)} = \frac{1}{2} \pm \frac{1}{2}\tan\theta_{13}\sqrt{2-\tan^2\theta_{13}} \,, \\ &\sin^2\theta_{12} = \frac{1}{2+\cos 2\theta} = \frac{1}{3\cos^2\theta_{13}}, \quad \sin 2\phi_{12} = \sin 2\phi_{13} = \sin\delta_{CP} = 0 \,. \end{aligned} \tag{6.47}$$

Using the experimental best fit value $(\sin^2\theta_{13})^{\rm bf} = 0.022$ [24,25], we find the solar mixing angle $\sin^2\theta_{12} \simeq 0.341$ and the atmospheric angle

$$\sin^2\theta_{23} \simeq \begin{cases} 0.426 & \text{for } \theta < \pi/2 \\ 0.574 & \text{for } \theta > \pi/2 \end{cases} . \tag{6.48}$$

The Dirac and Majorana CP phases are determined to take on CP conserving values in this case.

(ii) $G_l = Z_3^t$, $G_\nu = Z_2^s$, $X_\nu = u$

This case differs from the previous one in the residual CP X_ν. The unitary matrix Σ_ν is

$$\Sigma_\nu = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2}\,i & 2i & 0 \\ \sqrt{2}\,i & -i & \sqrt{3} \\ \sqrt{2}\,i & -i & -\sqrt{3} \end{pmatrix} . \tag{6.49}$$

Consequently, for $P_\nu = P_{12}$, the lepton mixing matrix is fixed to be

$$U = \frac{1}{\sqrt{6}} \begin{pmatrix} 2\cos\theta & \sqrt{2} & 2\sin\theta \\ -\cos\theta + i\sqrt{3}\sin\theta & \sqrt{2} & -\sin\theta - i\sqrt{3}\cos\theta \\ -\cos\theta - i\sqrt{3}\sin\theta & \sqrt{2} & -\sin\theta + i\sqrt{3}\cos\theta \end{pmatrix} Q_\nu \,. \tag{6.50}$$

We can extract the lepton mixing parameters as

$$\sin^2\theta_{13}=\frac{2}{3}\sin^2\theta\,,\quad \sin^2\theta_{12}=\frac{1}{2+\cos 2\theta}=\frac{1}{3\cos^2\theta_{13}}\,,$$
$$\sin^2\theta_{23}=\frac{1}{2}\,,\quad |\sin\delta_{CP}|=1\,,\quad \sin 2\phi_{12}=\sin 2\phi_{13}=0\,. \tag{6.51}$$

Because the $\mu-\tau$ reflection symmetry $X_\nu=u$ is imposed on the neutrino mass matrix [252–255], both atmospheric angle θ_{23} and Dirac CP phase δ_{CP} are maximal, while the Majorana CP phases ϕ_{12} and ϕ_{13} are trivial.

(iii) $G_l=Z_3^t$, $G_\nu=Z_2^{su}$, $X_\nu=u$

The Takagi factorization matrix Σ_ν is determined to be

$$\Sigma_\nu=\frac{1}{\sqrt{6}}\begin{pmatrix}2i & \sqrt{2}\,i & 0\\ -i & \sqrt{2}\,i & -\sqrt{3}\\ -i & \sqrt{2}\,i & \sqrt{3}\end{pmatrix}\,. \tag{6.52}$$

Using the mater formula Eq. (6.41), we obtain the lepton mixing matrix as

$$U=\frac{1}{\sqrt{6}}\begin{pmatrix}2 & \sqrt{2}\cos\theta & \sqrt{2}\sin\theta\\ -1 & \sqrt{2}\cos\theta-i\sqrt{3}\sin\theta & \sqrt{2}\sin\theta+i\sqrt{3}\cos\theta\\ -1 & \sqrt{2}\cos\theta+i\sqrt{3}\sin\theta & \sqrt{2}\sin\theta-i\sqrt{3}\cos\theta\end{pmatrix}Q_\nu\,. \tag{6.53}$$

The predictions for mixing angles and CP violation phases are

$$\sin^2\theta_{13}=\frac{1}{3}\sin^2\theta\,,\quad \sin^2\theta_{12}=\frac{1+\cos 2\theta}{5+\cos 2\theta}=\frac{1-3\sin^2\theta_{13}}{3\cos^2\theta_{13}}\,,$$
$$\sin^2\theta_{23}=\frac{1}{2}\,,\quad |\sin\delta_{CP}|=1\,,\quad \sin 2\phi_{12}=\sin 2\phi_{13}=0\,. \tag{6.54}$$

For $(\sin^2\theta_{13})^{\rm bf}=0.022$ [24, 25] we find, from Eq. (6.54), the solar mixing angle as $\sin^2\theta_{12}\simeq 0.318$. Similar to case ii, the $\mu-\tau$ reflection symmetry $X_\nu=u$ implies maximal values of θ_{23} and δ_{CP}, and CP conserving values of ϕ_{12},ϕ_{13}.

The values of the effective $|m_{\beta\beta}|$ parameter characterizing the $0\nu\beta\beta$ decay amplitude can be determined for the viable cases associated to the lepton mixing matrices in Eqs. (6.46, 6.50, 6.53). This effective mass parameter depends on the CP parity encoded in Q_ν. Since the rotation angle θ is strongly constrained by the reactor angle θ_{13} and the Majorana phases can be predicted, the effective mass $|m_{\beta\beta}|$ is severely restricted, as shown in figure 22. The narrow width of each band comes from varying θ_{13} and the neutrino mass squared differences over their allowed 3σ ranges [24].

For inverted neutrino mass ordering, the term proportional to m_3 in $|m_{\beta\beta}|$ is suppressed by the small values of m_3 and $\sin^2\theta_{13}$, consequently the value of k_2 is almost irrelevant. For $(k_1,k_2)=(0,0),(0,1)$, the effective mass $m_{\beta\beta}$ is close to the upper boundary of the IO region obtained by using the 3σ global

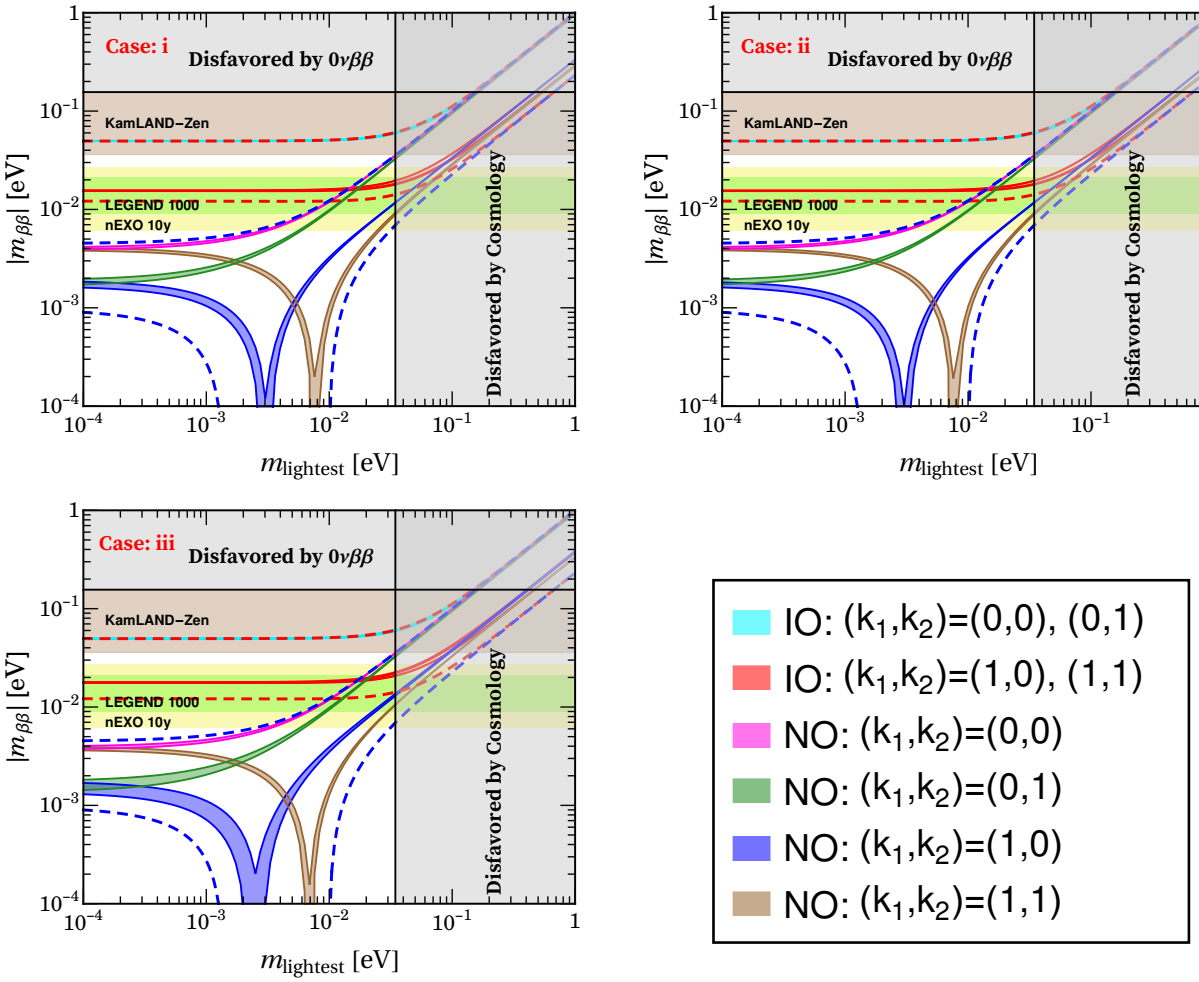


Figure 22: Predicted effective Majorana neutrino mass $|m_{\beta\beta}|$ for three viable mixing patterns when the S_4 flavour symmetry and CP symmetry are broken to an Abelian subgroup and $Z_2 \times CP$ in the charged lepton sector and neutrino sector respectively. The red (blue) dashed lines indicate the most general allowed regions for IO (NO) neutrino mass ordering obtained by varying the mixing parameters in their 3σ ranges [24, 25]. The current experimental bound $|m_{\beta\beta}| < (36 - 156)\,\text{meV}$ at 90% C.L. from KamLAND-Zen [125] and the future sensitivity ranges $|m_{\beta\beta}| < (9.0 - 21)$ meV from LEGEND-1000 [126] and $|m_{\beta\beta}| < (6.1 - 27)$ meV from nEXO [127] are indicated by light brown, light yellow and light green horizontal bands respectively. The vertical grey exclusion band represents the bound $\sum_i m_i < 0.120\text{eV}$ from Planck at 95% C.L. [128, 129].

data, while $|m_{\beta\beta}|$ is close to the lower boundary of IO for $(k_1, k_2) = (1, 0), (1, 1)$.

Symmetry breaking patterns of abelian subgroups in the charged lepton sector and $Z_2 \times CP$ in the neutrino sector have also been studied for many flavour symmetry groups combined with generalized CP, such as A_4 [264, 388, 393], S_4 [264, 320, 325, 335, 394–396], $\Delta(27)$ [397, 398], $\Delta(48)$ [399, 400], A_5 [401–405], $\Delta(96)$ [406] and the infinite group series $\Delta(3n^2) = (Z_n \times Z_n) \rtimes Z_3$ [407, 408], $\Delta(6n^2) = (Z_n \times Z_n) \rtimes S_3$ [407, 409] and $D^{(1)}_{9n,3n} = (Z_{9n} \times Z_{3n}) \rtimes S_3$ [410]. For popular flavour symmetries A_4, S_4 and A_5, the Dirac CP phase is predicted to take simple values $\delta_{CP} = 0, \pi, \pm\pi/2$, and the Majorana phases take on CP conserving values for the experimentally viable mixing patterns. Moreover, the values of the Dirac

CP phase and atmospheric mixing angle are correlated in flavor groups A_4, S_4 and A_5, the atmospheric angle is maximal for $\delta_{CP}=\pm\pi/2$ and non-maximal for $\delta_{CP}=0,\pi$. Both Dirac and Majorana CP phases can depend nontrivially on the parameter θ for larger flavour symmetry groups such as $\Delta(96)$ [406]. A systematical classification of the possible mixing patterns resulting from the pair of residual symmetry subgroups $\{G_l, Z_2\times CP\}$ has been performed [345]. The neutrino mixing patterns compatible with oscillation data are given as follows [345]:

$$
\begin{aligned}
U^{I(a)} &= \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2}\sin\varphi_1 & e^{i\varphi_2} & \sqrt{2}\cos\varphi_1 \\ \sqrt{2}\cos\left(\varphi_1-\frac{\pi}{6}\right) & -e^{i\varphi_2} & -\sqrt{2}\sin\left(\varphi_1-\frac{\pi}{6}\right) \\ \sqrt{2}\cos\left(\varphi_1+\frac{\pi}{6}\right) & e^{i\varphi_2} & -\sqrt{2}\sin\left(\varphi_1+\frac{\pi}{6}\right) \end{pmatrix} R_{23}(\theta)Q_\nu\,, \\
U^{I(b)} &= \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2}\cos\varphi_1 & e^{i\varphi_2} & \sqrt{2}\sin\varphi_1 \\ -\sqrt{2}\sin\left(\varphi_1-\frac{\pi}{6}\right) & -e^{i\varphi_2} & \sqrt{2}\cos\left(\varphi_1-\frac{\pi}{6}\right) \\ -\sqrt{2}\sin\left(\varphi_1+\frac{\pi}{6}\right) & e^{i\varphi_2} & \sqrt{2}\cos\left(\varphi_1+\frac{\pi}{6}\right) \end{pmatrix} R_{12}(\theta)Q_\nu\,, \\
U^{II} &= \frac{1}{\sqrt{3}}\begin{pmatrix} e^{i\varphi_1} & 1 & e^{i\varphi_2} \\ \omega e^{i\varphi_1} & 1 & \omega^2 e^{i\varphi_2} \\ \omega^2 e^{i\varphi_1} & 1 & \omega e^{i\varphi_2} \end{pmatrix} R_{13}(\theta)Q_\nu\,, \\
U^{III} &= \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{2}e^{i\varphi_1}\sin\varphi_2 & 1 & \sqrt{2}e^{i\varphi_1}\cos\varphi_2 \\ \sqrt{2}e^{i\varphi_1}\cos\left(\varphi_2+\frac{\pi}{6}\right) & 1 & -\sqrt{2}e^{i\varphi_1}\sin\left(\varphi_2+\frac{\pi}{6}\right) \\ -\sqrt{2}e^{i\varphi_1}\cos\left(\varphi_2-\frac{\pi}{6}\right) & 1 & \sqrt{2}e^{i\varphi_1}\sin\left(\varphi_2-\frac{\pi}{6}\right) \end{pmatrix} R_{13}(\theta)Q_\nu\,, \\
U^{IV(a)} &= \frac{1}{\sqrt{2\sqrt{5}\,\phi_g}}\begin{pmatrix} -\sqrt{2}\,\phi_g & \sqrt{2} & 0 \\ 1 & \phi_g & -\sqrt{\sqrt{5}\,\phi_g} \\ 1 & \phi_g & \sqrt{\sqrt{5}\,\phi_g} \end{pmatrix} R_{13}(\theta)Q_\nu\,, \\
U^{IV(b)} &= \frac{1}{\sqrt{2\sqrt{5}\,\phi_g}}\begin{pmatrix} -\sqrt{2}\,\phi_g i & \sqrt{2} & 0 \\ i & \phi_g & -\sqrt{\sqrt{5}\,\phi_g} \\ i & \phi_g & \sqrt{\sqrt{5}\,\phi_g} \end{pmatrix} R_{13}(\theta)Q_\nu\,, \\
U^{V} &= \frac{1}{2}\begin{pmatrix} \phi_g & 1 & \phi_g-1 \\ \phi_g-1 & -\phi_g & 1 \\ 1 & 1-\phi_g & -\phi_g \end{pmatrix} R_{23}(\theta)Q_\nu\,, \\
U^{VI} &= \frac{1}{2\sqrt{3}}\begin{pmatrix} (\sqrt{3}-1)e^{i\varphi} & 2 & -(\sqrt{3}+1)e^{i\left(\varphi+\frac{3\pi}{4}\right)} \\ -(\sqrt{3}+1)e^{i\varphi} & 2 & (\sqrt{3}-1)e^{i\left(\varphi+\frac{3\pi}{4}\right)} \\ 2e^{i\varphi} & 2 & 2e^{i\left(\varphi+\frac{3\pi}{4}\right)} \end{pmatrix} R_{13}(\theta)Q_\nu\,, \\
U^{VII} &= \frac{1}{2\sqrt{6}}\begin{pmatrix} -\frac{\sqrt{3}}{s_3} & 2\sqrt{2} & \frac{s_2-s_1}{s_1 s_2} \\ \frac{\sqrt{3}}{s_2} & 2\sqrt{2} & -\frac{s_1+s_3}{s_1 s_3} \\ \frac{\sqrt{3}}{s_1} & 2\sqrt{2} & \frac{s_2+s_3}{s_2 s_3} \end{pmatrix} R_{23}(\theta)Q_\nu\,,
\end{aligned}
\tag{6.55}
$$

up to row and column permutations, where $s_n\equiv\sin(2n\pi/7)$ with $n=1,2,3$ and $R_{ij}(\theta)$ is the rotation

matrix through an angle θ in the (ij)-plane. The parameters φ_1, φ_2 and φ are group theoretical indices characterizing the flavour group and the residual symmetry and they can only take some discrete values for a given flavour symmetry. The possible values of φ_1, φ_2 and φ for different finite flavour group G_f up to order 2000 and the corresponding predictions for lepton mixing parameters are given in [345]. It is remarkable that the above mixing patterns can be obtained from the flavour groups $\Delta(6n^2)$, $D^{(1)}_{9n,3n}$, A_5 and $\Sigma(168)$ in combination with generalized CP symmetry.

As shown in table 1, if the residual symmetries $K_4 \times CP$ and $Z_2 \times CP'$ are preserved in the neutrino and charged lepton sectors respectively, the lepton mixing matrix would only depends on a single real parameter θ as well. In this case, one row of the mixing matrix is completely fixed by residual symmetry, regardless of the free parameter θ.

It turns out that only one type of mixing pattern can accommodate the oscillation data [345]:

$$U^{VIII} = \frac{1}{2} R^T_{13}(\theta) \begin{pmatrix} \sqrt{2}e^{i\varphi_1} & -\sqrt{2}e^{i\varphi_1} & 0 \\ 1 & 1 & -\sqrt{2}e^{i\varphi_2} \\ 1 & 1 & \sqrt{2}e^{i\varphi_2} \end{pmatrix} Q_\nu \,, \tag{6.56}$$

where φ_1 and φ_2 take discrete values determined by the choice of the residual symmetry and the flavour symmetry group.

6.3.2 $\mathcal{G}_l = Z_2 \times CP$, $\mathcal{G}_\nu = Z_2 \times CP'$

The residual subgroups in both the neutrino and charged lepton sectors are of the structure $Z_2 \times CP$ in this scheme [282, 284, 395]. Since the neutrino sector still preserves the residual symmetry $Z_2^{g_\nu} \times X_\nu$, the unitary transformation U_ν is given by Eq. (6.39). The residual symmetry of the charged lepton sector is denoted as $Z_2^{g_l} \times X_l$ in this scheme, and the constrained consistency condition in Eq. (6.32) should be fulfilled, i.e., $X_{l\mathbf{r}} \rho^*_{\mathbf{r}}(g_l) X^{-1}_{l\mathbf{r}} = \rho_{\mathbf{r}}(g_l)$. The three families of left-handed lepton doublets are assigned to transform as a faithful triplet ρ of the flavour group G_f, where the representation ρ can be either irreducible or reducible. The residual symmetry $Z_2^{g_l} \times X_l$ of the charged lepton sector requires that $m_l^\dagger m_l$ is invariant under $Z_2^{g_l} \times X_l$, where m_l is the charged lepton mass matrix. Thus the unitary transformation U_l of the charged leptons should fulfill

$$U_l^\dagger \rho(g_l) U_l = \text{diag}(\pm 1, \pm 1, \pm 1) \,, \qquad U_l^\dagger X_l U_l^* = \text{diag}\left(e^{-i\alpha_e}, e^{-i\alpha_\mu}, e^{-i\alpha_\tau}\right) \equiv Q_l^{\dagger 2} \,, \tag{6.57}$$

where $Q_l = \text{diag}\left(e^{i\alpha_e/2}, e^{i\alpha_\mu/2}, e^{i\alpha_\tau/2}\right)$ is a diagonal phase matrix with $\alpha_{e,\mu,\tau}$ real. Notice that the eigenvalue of $\rho(g_l)$ is $+1$ or -1 because the generator g_l is of order 2. The residual CP transformation X_l should be a symmetric matrix, otherwise the charged lepton masses would be partially degenerate. Likewise, for the neutrino sector we can perform the Takagi factorization for X_l as follows

$$X_l = \Sigma_l \Sigma_l^T \,, \qquad \Sigma_l^\dagger \rho(g_l) \Sigma_l = \pm \text{diag}(1, -1, -1) \,. \tag{6.58}$$

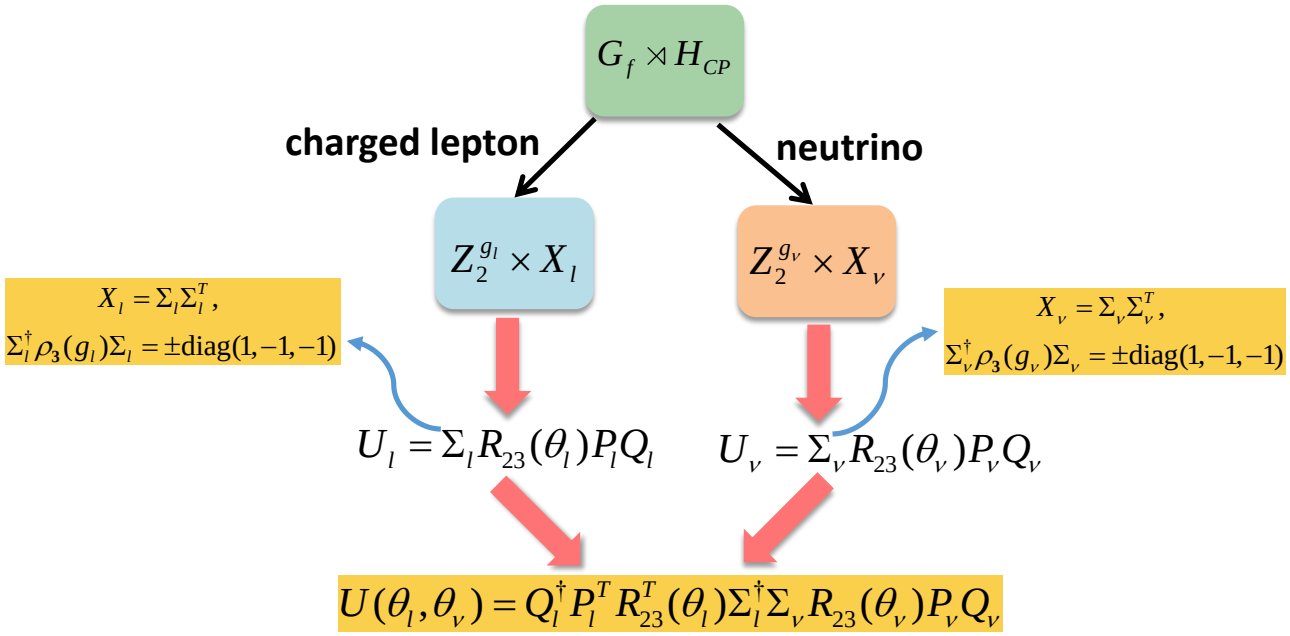


Figure 23: Model-independent predictions for the lepton mixing matrix when the residual symmetry has the structure $Z_2 \times CP$ in both neutrino and charged lepton sectors.

Then the residual symmetry constrains U_l to be

$$U_l = \Sigma_l R_{23}(\theta_l) P_l Q_l \tag{6.59}$$

where P_l is a generic three-dimensional permutation matrix, since the charged lepton masses are not predicted in this scheme. Hence the lepton mixing matrix is determined to be of the following form

$$U = U_l^\dagger U_\nu = Q_l^\dagger P_l^T R_{23}^T(\theta_l) \Sigma_l^\dagger \Sigma_\nu R_{23}(\theta_\nu) P_\nu Q_\nu \,, \tag{6.60}$$

where the phase matrix Q_l can be absorbed into the charged lepton fields. In this scenario, the recipe for extracting the predicted lepton mixing matrix from the residual symmetry is summarized in figure 23. One sees that the resulting lepton mixing matrix only depends on two free rotation angles θ_l and θ_ν, lying in the range $0 \leq \theta_{l,\nu} < \pi$. The precisely measured reactor angle θ_{13} and the solar angle θ_{12} can be accommodated for certain values of θ_l and θ_ν, leading to relations for the other mixing parameters. Notice that in this scheme only one element of the mixing matrix is fixed to a constant value by the residual subgroups.

A comprehensive study of the lepton mixing patterns which can arise from the breaking of S_4 and CP symmetries into two distinct $Z_2 \times CP$ subgroups in the neutrino and charged lepton sectors leads to eighteen phenomenologically viable cases [395]. In the following, we give one typical example which predicts non-trivial CP violation phases. The three families of left-handed leptons are embedded in a triplet **3** of S_4, and the residual symmetries of the neutrino and charged lepton mass matrices are $Z_2^{st^2su} \times X_l$ and $Z_2^s \times X_\nu$ respectively, with $X_l = t^2$ and $X_\nu = su$. The Takagi factorization Σ_l and Σ_ν are

found to be

$$\Sigma_l = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & 0 & -\sqrt{2} \\ e^{\frac{i\pi}{3}} & -\sqrt{3}\, e^{\frac{i\pi}{3}} & \sqrt{2}\, e^{\frac{i\pi}{3}} \\ e^{-\frac{i\pi}{3}} & \sqrt{3}\, e^{-\frac{i\pi}{3}} & \sqrt{2}\, e^{-\frac{i\pi}{3}} \end{pmatrix}, \quad \Sigma_\nu = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2}\, i & 0 & -2 \\ \sqrt{2}\, i & -\sqrt{3}\, i & 1 \\ \sqrt{2}\, i & \sqrt{3}\, i & 1 \end{pmatrix} \tag{6.61}$$

The lepton mixing matrix can be easily obtained by using the master formula of Eq. (6.60), one of its elements is fixed to be $1/\sqrt{2}$. In order to be compatible with experimental data, the fixed element can be either the (23) or (33) entry, so that we can take the permutation matrices $(P_l, P_\nu) = (P_{12}, P_{13})$ or $(P_{13}.P_{12}, P_{13})$. For the first case, $(P_l, P_\nu) = (P_{12}, P_{13})$, we find that the lepton mixing parameters are

$$\sin^2\theta_{13} = \frac{1}{2}\cos^2\theta_l, \quad \sin^2\theta_{12} = \frac{1}{2} + \frac{(1-3\cos 2\theta_l)\sin 2\theta_\nu}{6-2\cos 2\theta_l}, \quad \sin^2\theta_{23} = \frac{1}{2-\cos^2\theta_l}\,, \tag{6.62}$$

while the CP invariants are determined as

$$J_{CP} = \frac{\sin 2\theta_l \cos 2\theta_\nu}{8\sqrt{2}}, \quad I_1 = \frac{(2\sin 2\theta_l - 3\sin 4\theta_l)\cos 2\theta_\nu}{16\sqrt{2}}, \quad I_2 = \frac{\sin\theta_l \cos^3\theta_l \cos 2\theta_\nu}{2\sqrt{2}}\,. \tag{6.63}$$

The Jarlskog invariant J_{CP} of neutrino oscillations is related to δ^ℓ in Eq. (1.22), while the other two invariants I_1 and I_2 are given in terms of the basic Majorana phases ϕ_{12}, ϕ_{13} involved in $0\nu\beta\beta$ decay, see section 1. One finds the following expressions [219, 411–413],

$$\begin{aligned} J_{CP} &= \mathrm{Im}\,(U_{11}U_{33}U_{13}^*U_{31}^*) = \frac{1}{8}\sin 2\theta_{12}\sin 2\theta_{13}\sin 2\theta_{23}\cos\theta_{13}\sin\delta^\ell\,, \\ I_1 &= \mathrm{Im}\,\left(U_{11}^{2*}U_{12}^2\right) = -\sin^2\theta_{12}\cos^2\theta_{12}\cos^4\theta_{13}\sin 2\phi_{12}, \\ I_2 &= \mathrm{Im}\,\left(U_{11}^{2*}U_{13}^2\right) = -\cos^2\theta_{12}\cos^2\theta_{13}\sin^2\theta_{13}\sin 2\phi_{13}\,. \end{aligned} \tag{6.64}$$

The lepton mixing matrices corresponding to the two kinds of permutations $(P_l, P_\nu) = (P_{13}.P_{12}, P_{13})$ and $(P_l, P_\nu) = (P_{12}, P_{13})$ are related to each other by the exchange of the second and third rows. Thus the atmospheric angle θ_{23} and Dirac CP violation phase δ^ℓ become $\pi/2 - \theta_{23}$ and $\delta^\ell + \pi$ respectively, while the other mixing angles θ_{12}, θ_{13} and the Majorana CP phases ϕ_{12} and ϕ_{13} remain unchanged.

The results of the χ^2 analysis are presented in table 2. Moreover, a numerical analysis is performed, with both θ_l and θ_ν varying freely in the range of 0 to π, requiring all the three lepton mixing angles to lie in the experimental 3σ regions [24, 25]. In figure 24 we display the 3σ contour regions for $\sin^2\theta_{12}$, $\sin^2\theta_{13}$ and $\sin^2\theta_{23}$, as well as their experimental best fit values in the $\theta_l - \theta_\nu$ plane. Dashed (solid) lines are the best fit mixing-angle values [24, 25]. Left (right) panels are for $(P_l, P_\nu) = (P_{12}, P_{13})$ and $(P_l, P_\nu) = (P_{13}.P_{12}, P_{13})$, respectively. One sees that the lepton mixing angles can be accommodated in the small regions around the best fit points.

In figure 25 we show the contour plots of the CP violation phases δ^ℓ, ϕ_{12} and ϕ_{13} in the plane θ_ν versus θ_l. The black areas denote the regions where the lepton mixing angles are compatible with oscillation data within 3σ. These will be testable at forthcoming long baseline neutrino oscillation experiments. Since all the lepton mixing angles and CP phases are predicted to lie in narrow regions, we also have tight predictions for the effective Majorana mass of neutrinoless double beta decay, as shown in figure 26.

	(P_l, P_ν)	$\chi^2_{\rm min}$	$(\theta_l^{\rm bf}, \theta_\nu^{\rm bf})/\pi$	$\sin^2\theta_{13}$	$\sin^2\theta_{12}$	$\sin^2\theta_{23}$	δ^ℓ/π	ϕ_{12}/π (mod 1/2)	ϕ_{13}/π (mod 1/2)
NO	(P_{12}, P_{13})	20.078	(0.433,0.938)	0.022	0.318	0.511	0.539	0.397	0.468
			(0.567,0.562)						
			(0.433,0.562)				1.461	0.103	0.032
			(0.567,0.938)						
	$(P_{13}.P_{12}, P_{13})$	37.057	(0.433,0.938)	0.022	0.318	0.489	1.539	0.397	0.468
			(0.567,0.562)						
			(0.433,0.562)				0.461	0.103	0.032
			(0.567,0.938)						
IO	(P_{12}, P_{13})	15.352	(0.432,0.938)	0.022	0.318	0.511	0.539	0.396	0.468
			(0.568,0.562)						
			(0.432,0.562)				1.461	0.104	0.032
			(0.568,0.938)						
	$(P_{13}.P_{12}, P_{13})$	27.629	(0.432,0.938)	0.022	0.318	0.489	1.539	0.396	0.468
			(0.568,0.562)						
			(0.432,0.562)				0.461	0.104	0.032
			(0.568,0.938)						

Table 2: χ^2 analysis for the residual symmetries $Z_2^{st^2su} \times X_l$ in the charged lepton sector and $Z_2^s \times X_\nu$ in the neutrino sector with $X_l = t^2$ and $X_\nu = su$. We give the best fit values $\theta_l^{\rm bf}$ and $\theta_\nu^{\rm bf}$ for θ_l and θ_ν corresponding to $\chi^2_{\rm min}$. We also list the mixing angles and CP violating phases at the best fit point.

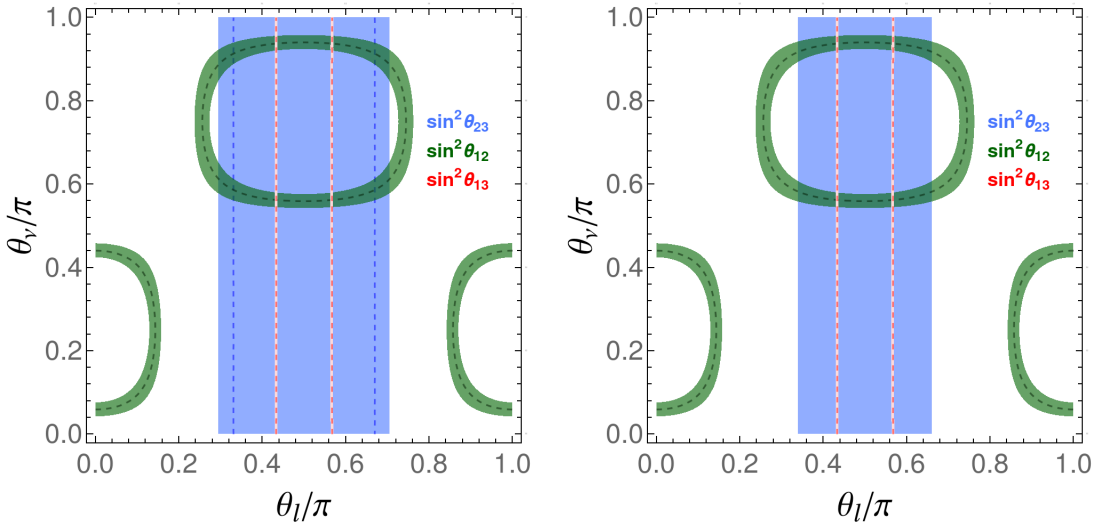


Figure 24: Contour plots of $\sin^2\theta_{ij}$ in the plane θ_ν versus θ_l for the residual symmetries $Z_2^{st^2su} \times X_l$ in the charged lepton sector and $Z_2^s \times X_\nu$ in the neutrino sector with $X_l = t^2$ and $X_\nu = su$. The red, blue and green areas are the 3σ regions of $\sin^2\theta_{13}, \sin^2\theta_{23}$ and $\sin^2\theta_{12}$ respectively. The dashed lines correspond to the best fit mixing angle values taken from [24, 25].

6.4 Quark and lepton mixing from a common flavour group

Discrete flavour symmetries are particularly suitable to account for the large lepton mixing angles. As discussed in [259, 263, 270, 278–281] they can also address the quark mixing pattern. Assuming that the flavour group of quarks is broken down to different residual subgroups in the up- and down-quark sectors, only the Cabibbo mixing between the first two quark families can be generated with only flavour symmetry [263, 270, 278, 280, 281]. It is remarkable that the hierarchical quark mixing angles and CP

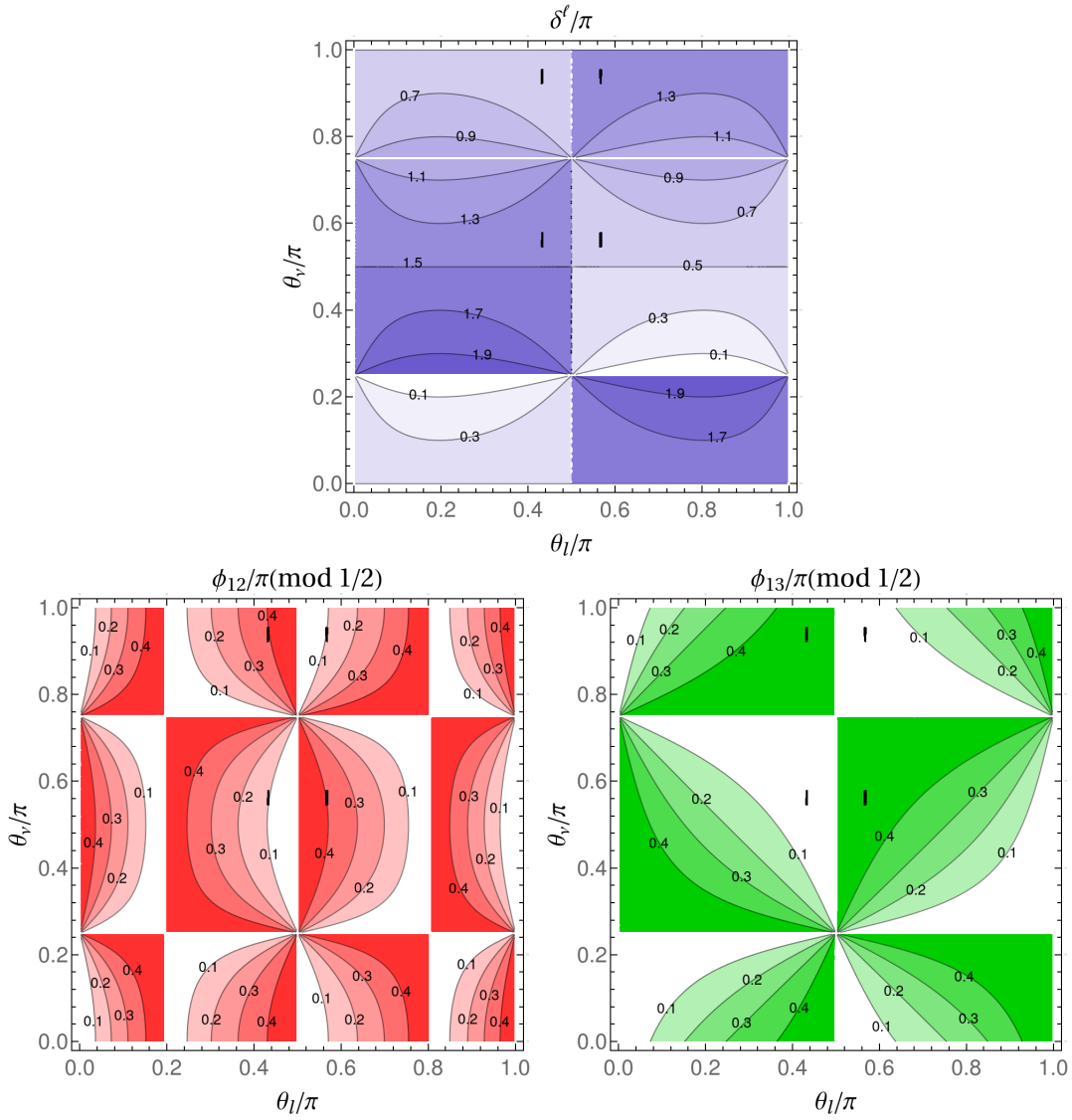


Figure 25: Contour plots of the CP violation phases δ^ℓ, ϕ_{12} and ϕ_{13} in the $\theta_l - \theta_\nu$ plane, where the residual symmetries are $Z_2^{st^2su} \times X_l$ in the charged lepton sector and $Z_2^s \times X_\nu$ in the neutrino sector with $X_l = t^2$ and $X_\nu = su$. In the black areas all three lepton mixing angles lie within their experimental 3σ ranges. Here we choose the row and column permutations $(P_l, P_\nu) = (P_{12}, P_{13})$, the Dirac CP phase δ^ℓ changes to $\delta^\ell + \pi$ while the Majorana phases ϕ_{12} and ϕ_{13} are invariant for $(P_l, P_\nu) = (P_{13}.P_{12}, P_{13})$.

violation can be explained if the flavour symmetry is extended with CP symmetry and the residual subgroups of the up- and down-quark sectors are $Z_2^{g_u} \times X_u$ and $Z_2^{g_d} \times X_d$ respectively [282, 284], where both CP transformations X_u and X_d are unitary and symmetric. The diagonalization matrices U_u and U_d would be restricted to be of the following form

$$U_u = \Sigma_u R_{23}(\theta_u) P_u Q_u \,, \quad U_d = \Sigma_d R_{23}(\theta_d) P_d Q_d \tag{6.65}$$

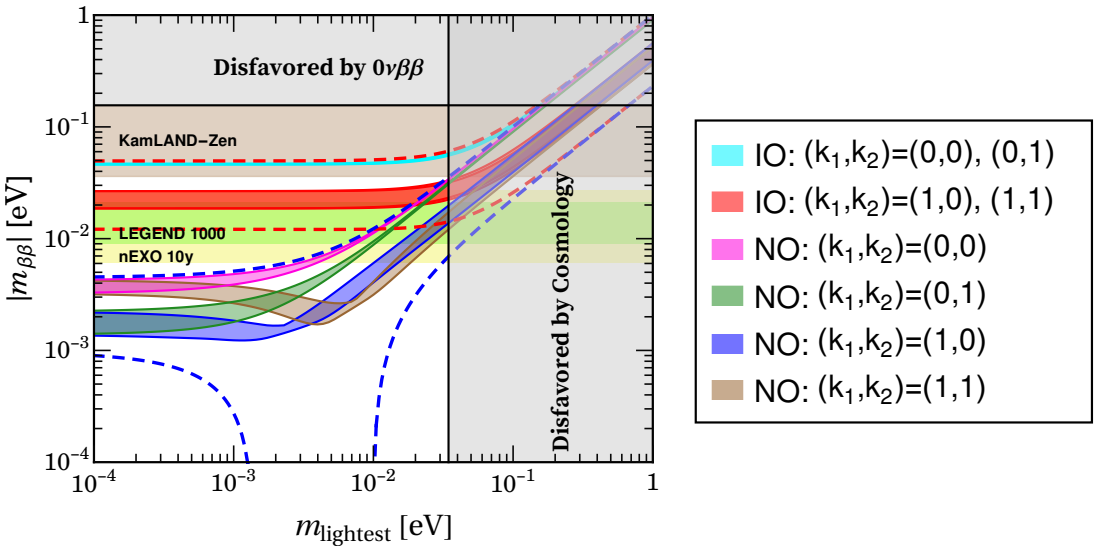


Figure 26: The effective Majorana neutrino mass $m_{\beta\beta}$ versus the lightest neutrino mass, for residual symmetries $Z_2^{st^2su} \times X_l$ in the charged lepton and $Z_2^s \times X_\nu$ in the neutrino sector, with $X_l = t^2$ and $X_\nu = su$. Here we adopt the same convention for the different bands as used in figure 22.

where $P_{u,d}$ are three dimensional permutation matrices, $Q_{u,d}$ are generic phase matrices, and $\Sigma_{u,d}$ are the Takagi factorizations of $X_{u,d}$, fulfilling

$$\begin{aligned} X_u &= \Sigma_u \Sigma_u^T \,, \qquad \Sigma_u^\dagger \rho(g_u) \Sigma_u = \pm \text{diag}(1,-1,-1) \,, \\ X_d &= \Sigma_d \Sigma_d^T \,, \qquad \Sigma_d^\dagger \rho(g_d) \Sigma_d = \pm \text{diag}(1,-1,-1) \,. \end{aligned} \tag{6.66}$$

As a consequence, the CKM mixing matrix is predicted as

$$V_{CKM} = U_u^\dagger U_d = Q_u^\dagger P_u^T R_{23}^T(\theta_u) \Sigma_u^\dagger \Sigma_d R_{23}(\theta_d) P_d Q_d \,, \tag{6.67}$$

which only depends on two free rotation angles θ_u and θ_d, limited in the range $0 \leq \theta_{u,d} < \pi$. Notice that both Q_u and Q_d are unphysical, as they can be absorbed into quark fields. We take the flavour symmetry as the dihedral group D_n which has only one- and two-dimensional irreducible representations, as shown in C.

Since the top quark is much heavier than the others, the first two families of left-handed quarks are assigned to a doublet of D_n, while the third generation is a D_n singlet. Without loss of generality, we can take

$$\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \sim \mathbf{2}_1, \quad Q_3 \sim \mathbf{1}_1 \,, \tag{6.68}$$

where $Q_1 \equiv (u_L, d_L)^T$, $Q_2 \equiv (c_L, s_L)^T$, and $Q_3 \equiv (t_L, b_L)^T$. The dihedral group and CP symmetry are broken down to $Z_2^{SR^{z_u}} \times X_u$ in up-quark sector and $Z_2^{SR^{z_d}} \times X_d$ in down-quark sector with $X_u = R^{-z_u+x_u}$ and $X_d = R^{-z_d+x_d}$, where $z_u, z_d = 0, 1, \ldots, n-1$, $x_u = x_d = 0$ for odd n and $x_u, x_d = 0, n/2$ if the group

index n is even. The Takagi factorization of the residual symmetry $Z_2^{SR^z} \times R^{-z+x}$ is determined to be

$$\Sigma = \frac{i^{2x/n}}{\sqrt{2}} \begin{pmatrix} -e^{-\frac{i\pi z}{n}} & 0 & e^{-\frac{i\pi z}{n}} \\ e^{\frac{i\pi z}{n}} & 0 & e^{\frac{i\pi z}{n}} \\ 0 & \sqrt{2}\, i^{-2x/n} & 0 \end{pmatrix} . \tag{6.69}$$

From our master formula in Eq. (6.67), we find that the quark mixing matrix takes the following form,

$$V_{CKM} = \begin{pmatrix} \cos\varphi_1 & -c_d \sin\varphi_1 & s_d \sin\varphi_1 \\ c_u \sin\varphi_1 & s_d s_u e^{i\varphi_2} + c_d c_u \cos\varphi_1 & c_d s_u e^{i\varphi_2} - s_d c_u \cos\varphi_1 \\ -s_u \sin\varphi_1 & s_d c_u e^{i\varphi_2} - c_d s_u \cos\varphi_1 & c_d c_u e^{i\varphi_2} + s_d s_u \cos\varphi_1 \end{pmatrix} , \tag{6.70}$$

up to row and column permutations with

$$\varphi_1 = \frac{(z_u - z_d)\pi}{n}, \quad \varphi_2 = \frac{(x_u - x_d)\pi}{n} \tag{6.71}$$

and $c_d \equiv \cos\theta_d$, $s_d \equiv \sin\theta_d$, $c_u \equiv \cos\theta_u$, $s_u \equiv \sin\theta_u$. The parameters φ_1 and φ_2 depend on the choice of residual symmetry, and they can take the following discrete values

$$\begin{aligned} &\varphi_1 \ (\text{mod}\ 2\pi) = 0, \frac{1}{n}\pi, \frac{2}{n}\pi, \ldots, \frac{2n-1}{n}\pi\,, \\ &\varphi_2 \ (\text{mod}\ 2\pi) = 0, \frac{1}{2}\pi, \frac{3}{2}\pi\,. \end{aligned} \tag{6.72}$$

One can straightforwardly extract the quark mixing parameters from Eq. (6.70). Eliminating the free parameters θ_u and θ_d, one obtains the following correlations among the quark mixing angles and CP phase [284],

$$\cos^2\theta_{13}^q \cos^2\theta_{12}^q = \cos^2\varphi_1, \quad \sin\delta^q \simeq \frac{\sin 2\varphi_1 \sin\varphi_2}{\sin 2\theta_{12}^q \cos^2\theta_{13}^q \cos\theta_{23}^q} . \tag{6.73}$$

The experimental data on the CKM matrix can be well accommodated for $\varphi_1 = \pi/14$, $\varphi_2 = \pi/2$, which can be achieved from the D_{14} flavour group with the residual symmetry indices $z_u = 1$, $z_d = 0$, $x_u = 7$, $x_d = 0$. The best-fit values of $\theta_{u,d}$ and mixing parameters are determined to be,

$$\begin{aligned} &\theta_u = 0.01237\pi, \quad \theta_d = 0.99473\pi, \quad \sin\theta_{12}^q = 0.22249\,, \\ &\sin\theta_{13}^q = 0.00369, \quad \sin\theta_{23}^q = 0.04206, \quad J_{CP}^q = 3.104 \times 10^{-5}\,. \end{aligned} \tag{6.74}$$

Notice that $\sin\theta_{13}^q$, $\sin\theta_{23}^q$ and J_{CP}^q are consistent with the global fit results of the UTfit collaboration [101]. The mixing angle $\sin\theta_{12}^q$ is about 1% smaller than its measured value, so that higher-order corrections in a concrete model are needed to reconcile it with the data.

The D_{14} flavour group can also explain the lepton flavour structure if it is broken down to $Z_2^{SR^{z_l}} \times X_l$ and $Z_2^{SR^{z_\nu}} \times X_\nu$ in the charged lepton and neutrino sector, respectively, where $X_l = R^{-z_l+x_l}$, $X_\nu = R^{-z_\nu+x_\nu}$ with $z_{l,\nu} = 0, 1, \ldots, 13$ and $x_{l,\nu} = 0, 7$. The lepton mixing matrix has the same form as Eq. (6.70), the rotation angles θ_u and θ_d should be replaced with θ_l and θ_ν respectively. Choosing the residual

symmetry indices $z_l = 4$, $z_\nu = 0$, $x_l = 7$ and $x_\nu = 0$, we have $\varphi_1 = 2\pi/7$ and $\varphi_2 = \pi/2$. Choosing the permutations as $P_l = P_{12}P_{23}$ and $P_\nu = P_{13}$, the lepton mixing angles can be accommodated for certain values of the free parameters $\theta_{l,\nu}$:

$$\begin{aligned}&\theta_e^{\rm bf} = 0.439\pi,\ \theta_\nu^{\rm bf} = 0.811\pi,\ \chi^2_{\rm min} = 4.147,\ \sin^2\theta_{13} = 0.0220,\ \sin^2\theta_{12} = 0.318\,,\\ &\sin^2\theta_{23} = 0.603,\ \delta^\ell/\pi = 1.530,\ \phi_{12}/\pi = -0.082\ (\text{mod}\ 1/2),\ \phi_{13}/\pi = 1.474\ (\text{mod}\ 1/2)\,.\end{aligned} \tag{6.75}$$

In summary, one sees how the dihedral group as well as the residual symmetry $Z_2 \times CP$ provide an interesting opportunity for model building. Indeed, we saw how the D_{14} flavour symmetry can provide a unified description of flavour mixing for both quarks and leptons.

If both left-handed quarks and leptons are assigned as irreducible triplets of the flavour symmetry and the residual symmetry is $Z_2 \times CP$, one finds that $\Delta(294)$ is the minimal flavour group that can generate realistic quark and lepton flavour mixing patterns [283]. In contrast, the singlet plus doublet assignment seems better than the triplet assignment. Once the CP symmetry is included, the order of the flavour symmetry group can be reduced considerably, i.e. 28 versus 294 in this scheme.

There are also other schemes to explain the mixing patterns of quarks and leptons using flavour and CP symmetries. For instance, quark and lepton mixing patterns can arise from the stepwise breaking of these symmetries to different residual subgroups in different sectors of the theory [285, 286], with charged fermion mass hierarchies generated by operators with different numbers of flavons. For a concrete model with $\Delta(384)$ flavour symmetry see [286].

6.5 Geometrical CP violation

It is well-known that CP symmetry is broken by complex Yukawa couplings in the SM, which leads to the CP violation in charged current interactions through the CKM matrix. However, the origin of CP violation is still a mystery. Analogous to the electroweak symmetry, the CP symmetry could be spontaneously broken by the VEVs of some scalar fields [414]. In models of spontaneous CP violation, the Lagrangian is invariant under the CP symmetry so that all parameters of the scalar potential are real in a certain basis. Spontaneous CP violation is achieved through complex VEVs for the Higgs multiplets which also break the gauge symmetry. Usually the phases of the fields depend on the coupling constants in the scalar potential.

The phases of the Higgs multiplets could have geometrical values, independently of the potential parameters if there is an additional (accidental) CP symmetry of the potential. The resulting CP breaking vacua lead to the so-called geometrical CP violation or calculable phases [415]. It has been shown that more than two Higgs doublets and non-abelian symmetry relating the Higgs multiplets are necessary conditions in order to realize the geometrical CP violation. It turns out that $\Delta(27)$ and $\Delta(54)$ are the smallest groups which lead to calculable phases [415, 416].

If one assigns three Higgs doublets $H \equiv (H_1, H_2, H_3)$ to a triplet of $\Delta(27)$ or $\Delta(54)$, the scalar potential

has only one relevant phase dependent term, i.e.

$$V = V_0 + \left[\lambda_4 \sum_{i \neq j \neq k} (H_i^\dagger H_j)(H_i^\dagger H_k) + \text{h.c.} \right] . \tag{6.76}$$

The traditional CP transformation $H_i \xrightarrow{\mathcal{CP}} H_i^*$ forces the coupling λ_4 to be real. Then one can obtain the following two possible vacua with calculable phases [415]

$$\begin{aligned} \langle H \rangle &= \frac{v}{\sqrt{3}} \begin{pmatrix} 1 \\ \omega \\ \omega^2 \end{pmatrix} , \qquad \lambda_4 < 0 \, , \\ \langle H \rangle &= \frac{v}{\sqrt{3}} \begin{pmatrix} \omega^2 \\ 1 \\ 1 \end{pmatrix} , \qquad \lambda_4 > 0 \, . \end{aligned} \tag{6.77}$$

The same scalar potential as Eq. (6.76) and calculable phases in Eq. (6.77) can be obtained from other non-abelian symmetry groups such as $\Delta(3n^2)$ [417] and $\Delta(6n^2)$ [418], where n is a multiple of 3.

The observed fermion masses and flavor mixings could possibly be accommodated if one properly assigns the transformations of fermion fields under the symmetry group [416]. In short, the geometrical CP violation arises from the correct interplay among the scalar content, non-abelian symmetry group and CP symmetry.

7 Testing flavor and CP symmetries

As shown above, discrete flavor and generalized CP symmetries allow us to predict the lepton mixing matrix in terms of few parameters. Eliminating free parameters generally leads to correlations among the lepton mixing angles and CP violation phases. Such predictions are often called lepton mixing *sum rules* in the literature, though they are not always strictly so. For example, for the residual symmetry $\mathcal{G}_l = Z_n$ ($n \geq 3$), $\mathcal{G}_\nu = Z_2 \times CP$ discussed in section 6.3.1, the lepton mixing matrix depends on a unique real parameter θ, as shown in Eq. (6.41). The parameter θ becomes determined in terms of the precisely measured reactor mixing angle θ_{13}, leading to sharp predictions for the leptonic mixing angles and CP violation phases that can be tested at current and future neutrino oscillation experiments.

Ultimately they could be used to distinguish different symmetry-based flavor models. It is remarkable that, besides the mixing angles, flavor symmetry in combination with generalized CP symmetry allows us to predict both the Dirac and Majorana leptonic CP violation phases. Hence the effective neutrino mass $|m_{\beta\beta}|$ is constrained to lie within narrow regions, as can be seen from figures 22 and 26.

As a result one could also test flavor and CP symmetries by confronting with the data of the current and forthcoming $0\nu\beta\beta$ experiments. In fact, some flavor models relate the Majorana phases to the neutrino masses, which could also lead to restrictions on the effective mass $|m_{\beta\beta}|$, and a very powerful tool to test and discriminate flavour models.

7.1 Testing mixing predictions

Neutrino physics has entered the precision era, providing a good opportunity for probing different flavor models. The reactor angle θ_{13} is the best-measured leptonic mixing parameter. The precise measurement of nonzero θ_{13} by Daya Bay [21, 419], Double Chooz [420, 421], and RENO [22, 422] has excluded many flavor models predicting θ_{13} close to zero. The long baseline neutrino oscillation experiments NOνA [423] and T2K [424] give measurements of the leptonic mixing angle θ_{23} and also the first hint of leptonic CP violation [425–427] associated to the leptonic Dirac CP violation phase $\delta_{\rm CP}$. However, the value of δ_{CP} has not been significantly constrained by neutrino oscillation experiments. The latest data of T2K favor near-maximal Dirac CP violation phase, the atmospheric mixing angle $\theta_{23} > 45^\circ$ and normal mass ordering [428]. The data of T2K constrain the CP phase δ_{CP} in the range $\delta_{CP} = -1.97^{+0.97}_{-0.70}$, and $\delta_{CP} = 0, \pi$ is excluded at more than 90% confidence level [428]. In comparison with T2K, the data of NOνA exclude the CP phase in the vicinity of $\delta_{CP} = \pi/2$ at more than 3σ for the inverted mass ordering, and the values around $\delta_{CP} = 3\pi/2$ in the normal ordering are disfavored at 2σ confidence [429]. Improved CP measurements constitute the target of upcoming experiments such as DUNE. The future long baseline neutrino oscillation experiments DUNE [113, 430] and T2HK [431] should measure θ_{23} and $\delta_{\rm CP}$ with very good precision. It is expected that DUNE can observe the signal of lepton CP violation with 5σ significance after about 7 years if $\delta_{CP} = -\pi/2$ and after about 10 years for 50% of δ_{CP} values, and CP violation can be observed with 3σ significance for 75% of δ_{CP} values after about 13 years of running [113, 430]. T2HK has shown that it can expect a discovery of CP violation over 76%(58%) of the parameter space at $3\sigma(5\sigma)$ [431]. The forthcoming medium baseline reactor neutrino experiment JUNO

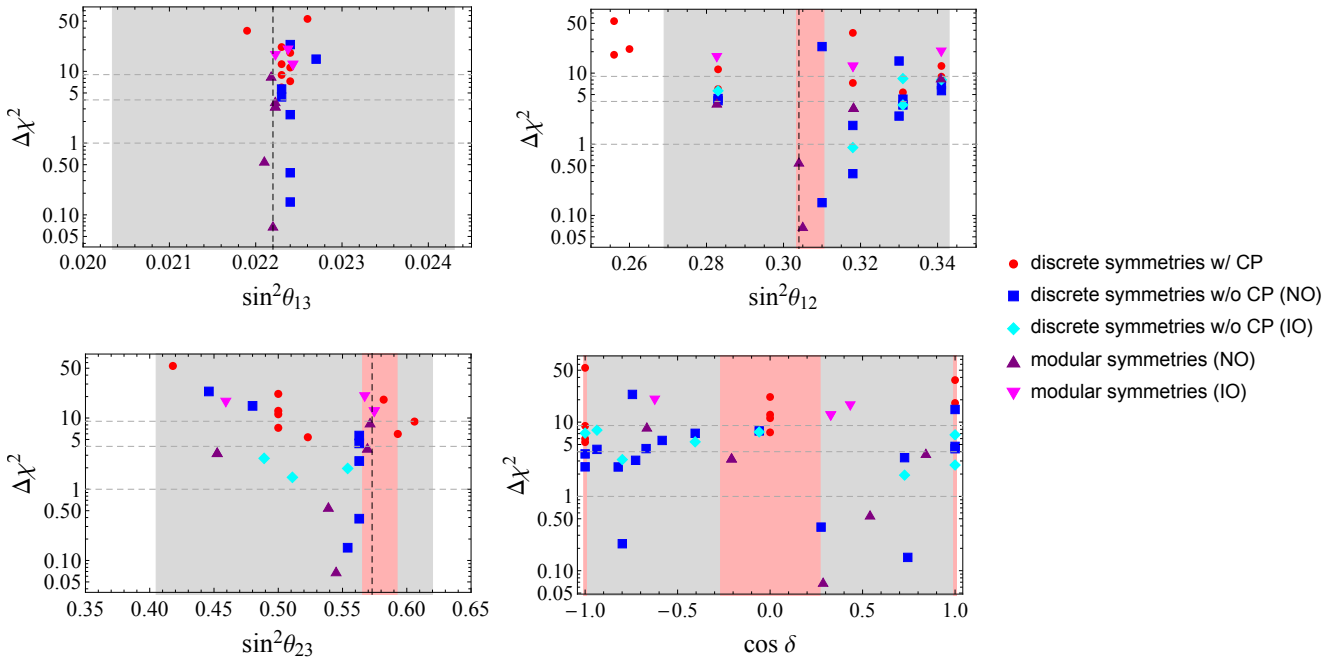


Figure 27: Best-fit predictions of the models based on discrete symmetries broken to certain residual symmetries of the lepton mass matrices [432,433], and the models based on modular symmetries discussed in [434–436]. The gray regions are the 3σ ranges and the dashed line is the best fit value for normal-ordered neutrino masses. The red band in the panel of $\sin^2\theta_{12}$ is the prospective 3σ range after 6 years of JUNO running [115], and the red regions in the panels of $\sin^2\theta_{23}$ and $\cos\delta$ are the prospective 3σ ranges after 15 years of DUNE running [430]. This figure is taken from [437].

can make very precise measurement of the solar angle θ_{12}, and the error of $\sin^2\theta_{12}$ can be reduced to the level of $0.5\% - 0.7\%$ [115].

We show in figure 27 the best fit predictions for the mixing parameters in some typical models based on discrete flavor symmetry with or without generalized CP symmetry and modular symmetry. We see that the synergy between JUNO and long baseline neutrino experiments DUNE and T2HK will be extremely powerful for testing the huge number of flavor symmetry models.

As an illustration, we give an extensively discussed prediction for lepton mixing angles as follows [438, 439],

$$\cos\delta_{CP} = \frac{\tan\theta_{23}}{\sin 2\theta_{12}\,\sin\theta_{13}}\left[\cos 2\theta^{\nu}_{12} + \left(\sin^2\theta_{12} - \cos^2\theta^{\nu}_{12}\right)\left(1 - \cot^2\theta_{23}\,\sin^2\theta_{13}\right)\right] , \tag{7.1}$$

which generally arises when considering charged-lepton corrections to neutrino mixing matrices which are completely fixed by symmetry with the neutrino angle $\theta^{\nu}_{13} = 0$ and charged lepton rotation angle $\theta^{e}_{13} = 0$. This prediction gives $\cos\delta_{CP}$ as a function of the measured lepton mixing angles and one fixed parameter θ^{ν}_{12} determined by the underlying discrete symmetry. It is noticeable that Eq. (7.1) is specified by fixing the value of just one parameter, the angle θ^{ν}_{12}. Thus one can enumerate the viable models of this type by deriving the values of θ^{ν}_{12} from symmetry considerations. This leads us the following well-motivated sum rules characterized by specific values of θ^{ν}_{12}, namely the one based on TBM mixing with

	Ranges of δ_{CP} obtained by varying			
	θ_{12} in 3σ	θ_{23} in 3σ	θ_{13} in 3σ	$\theta_{12}\&\theta_{23}\&\theta_{13}$ in 3σ
TBM	$\pm[240.54^\circ, 289.19^\circ]$	$\pm[264.58^\circ, 268.06^\circ]$	$\pm[264.86^\circ, 265.89^\circ]$	$\pm[235.27^\circ, 291.21^\circ]$
GRA	$\pm[271.01^\circ, 322.93^\circ]$	$\pm[289.72^\circ, 295.36^\circ]$	$\pm[292.95^\circ, 294.82^\circ]$	$\pm[270.22^\circ, 332.89^\circ]$
GRB	$\pm[233.08^\circ, 283.21^\circ]$	$\pm[258.11^\circ, 263.51^\circ]$	$\pm[258.57^\circ, 260.20^\circ]$	$\pm[226.19^\circ, 284.41^\circ]$
HG	$\pm[284.70^\circ, 360^\circ]$	$\pm[300.64^\circ, 312.11^\circ]$	$\pm[307.15^\circ, 310.98^\circ]$	$\pm[283.21^\circ, 360^\circ]$

Table 3: The prediction of the sum rule Eq. (7.1) for the allowed ranges of $|\delta_{CP}|$ due to the present 3σ uncertainties in the values of the neutrino mixing angles. Here we vary at least one lepton mixing angle in its corresponding 3σ intervals for the NO spectrum [24].

$\theta^\nu_{12} = \arcsin(1/\sqrt{3}) \approx 35.26^\circ$ [296–298], the one based on BM mixing with $\theta^\nu_{12} = 45^\circ$ [333], the one based on the type A golden ratio mixing (GRA) with $\theta^\nu_{12} = \arctan(1/\phi_g) \approx 31.72^\circ$ [328, 329], the one based on the type B golden ratio mixing (GRB) with $\theta^\nu_{12} = \arccos(\phi_g/2) = 36^\circ$ [326, 327], and the one based on hexagonal (HEX) mixing with $\theta^\nu_{12} = 30^\circ$ [317, 440].

Using the sum rule of Eq. (7.1) and plugging into the best fit values of the neutrino mixing angles for NO [24], one can straightforwardly determine the value of $\cos\delta_{CP}$ and the CP phase δ_{CP} for the above mentioned values of θ^ν_{12},

$$\begin{aligned}
\text{TBM}: &\quad \cos\delta_{CP} \approx -0.08, \quad \delta_{CP} \approx \pm 94.65^\circ\,, \\
\text{GRA}: &\quad \cos\delta_{CP} \approx 0.41, \quad \delta_{CP} \approx \pm 66.09^\circ\,, \\
\text{GRB}: &\quad \cos\delta_{CP} \approx -0.18, \quad \delta_{CP} \approx \pm 100.65^\circ\,, \\
\text{HG}: &\quad \cos\delta_{CP} \approx 0.63, \quad \delta_{CP} \approx \pm 50.90^\circ\,,
\end{aligned} \tag{7.2}$$

Notice that there are two values of δ_{CP} of opposite sign for each value of $\cos\delta_{CP}$. The sum rule of Eq. (7.1) for the BM case, $\theta^\nu_{12} = 45^\circ$, is not compatible with the current best fit values of the lepton mixing angles [24], and will be dropped hereafter. We also take into account the experimental uncertainties of the three lepton mixing angles by varying θ_{12}, θ_{23}, and θ_{13} in their 3σ experimentally allowed regions. The ranges of the CP violation $\delta_{\rm CP}$ obtained from Eq. (7.1) are summarized in table 3. We see that out of all mixing angles, the 3σ allowed range of θ_{12} causes the largest uncertainty in δ_{CP} resulting from Eq. (7.1).

Future long baseline experiments DUNE and T2HK will be able to make precision measurement of δ_{CP} and θ_{23}. The combination of DUNE and T2HK provides better sensitivity to δ_{CP} than either of these two experiments in isolation. The prospective DUNE+T2HK data should allow one to test the predictions for $\cos\delta_{CP}$, as shown in figure 28. Given the 3σ range $\delta_{CP} \in [127.80^\circ, 358.20^\circ]$ of the Dirac phase δ_{CP} from the latest global analysis [24], only δ_{CP} values in the interval of 180° to 360° are considered in figure 28. We see that a significant part of the true value of δ_{CP} gets disfavoured at more than 3σ for each symmetry predicted value of the angle θ^ν_{12}. Furthermore, detailed analysis showed that future facilities DUNE and T2HK in combination with JUNO could distinguish the different cases of θ^ν_{12} [439, 441].

In summary, the existence of mixing predictions is a characteristic feature of flavor symmetry models, as explicitly shown in Eqs. (5.12, 5.18, 5.24, 6.9, 6.22, 6.47, 6.51, 6.54, 6.73). These relations highlight

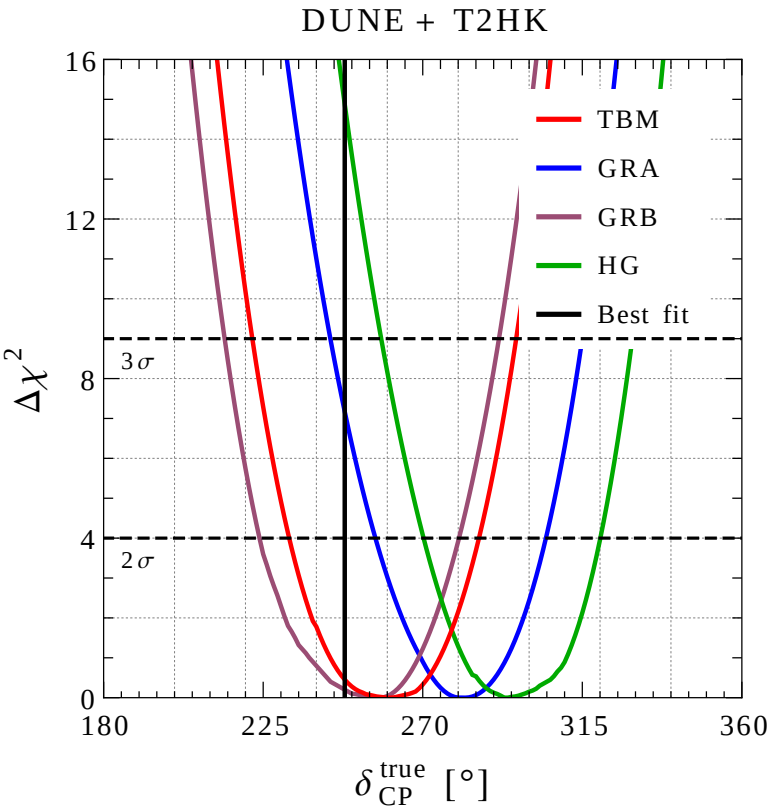


Figure 28: Compatibility of the prediction in Eq. (7.1) with any potentially true value of the Dirac CP phase δ_{CP} in the range $\delta_{CP} \in [180°, 360°]$. Four values of θ_{12}^{ν} are considered: $\theta_{12}^{\nu} = \arcsin(1/\sqrt{3}) \approx 35.26°$ for TBM, $\theta_{12}^{\nu} = \arctan(1/\phi_g) \approx 31.72°$ for GRA, $\theta_{12}^{\nu} = 36°$ for GRB, and $\theta_{12}^{\nu} = 30°$ for HG. From [441].

the testability of flavor symmetry, and there are many other possible relations among the lepton mixing parameters in the literature [355, 438, 439, 442–445]. The phenomenological implications of these sum rules and the prospects of testing them at precision neutrino facilities have been extensively discussed [433, 439, 441, 444, 446, 447].

7.2 Testing mass sum rules

The neutrino mass matrix leading to the three neutrino masses in discrete flavor symmetry models typically involves a reduced set of parameters. Indeed, several discrete flavor-symmetry-based models yield a constrained neutrino mass matrix, leading to specific neutrino mass sum-rules. This is turn is phenomenologically interesting, as it can lead to predictions for the effective neutrinoless double beta mass parameter [141]. For example, one can have a $0\nu\beta\beta$ lower bound on even for normal ordered neutrino spectrum [142].

If the light neutrino mass matrix depends on two complex parameters [448] one can extract a relation between the three complex neutrino mass eigenvalues, leading to a neutrino mass sum rule. Typically neutrino mass sum rules can arise from any neutrino mass generation mechanism in which the structure of the mass matrix is generated by two flavons. Prime examples neutrino mass sum rules are $2\tilde{m}_2 + \tilde{m}_3 = \tilde{m}_1$ and $2\tilde{m}_2^{-1} + \tilde{m}_3^{-1} = \tilde{m}_1^{-1}$ predicted by early A_4 models [70, 313, 449]. For systematic categorisation and

consequences of neutrino mass sum rules in beta decay, $0\nu\beta\beta$ decay and cosmology see Refs. [141,450–453]. A sample of the flavor models in the literature, includes the following twelve different neutrino mass sum rules [451]:

$$
\begin{aligned}
&\tilde{m}_1+\tilde{m}_2=\tilde{m}_3, && \tilde{m}_1+\tilde{m}_3=2\tilde{m}_2\,,\\
&2\tilde{m}_2+\tilde{m}_3=\tilde{m}_1, && \tilde{m}_1+\tilde{m}_2=2\tilde{m}_3\,,\\
&\tilde{m}_1+\frac{\sqrt{3}+1}{2}\tilde{m}_3=\frac{\sqrt{3}-1}{2}\tilde{m}_2, && \tilde{m}_1^{-1}+\tilde{m}_2^{-1}=\tilde{m}_3^{-1}\,,\\
&2\tilde{m}_2^{-1}+\tilde{m}_3^{-1}=\tilde{m}_1^{-1}, && \tilde{m}_1^{-1}+\tilde{m}_3^{-1}=2\tilde{m}_2^{-1}\,,\\
&\tilde{m}_3^{-1}\pm 2i\tilde{m}_2^{-1}=\tilde{m}_1^{-1}, && \tilde{m}_1^{1/2}-\tilde{m}_3^{1/2}=2\tilde{m}_2^{1/2}\,,\\
&\tilde{m}_1^{1/2}+\tilde{m}_3^{1/2}=2\tilde{m}_2^{1/2}, && \tilde{m}_1^{-1/2}+\tilde{m}_2^{-1/2}=2\tilde{m}_3^{-1/2}\,,
\end{aligned}
\tag{7.3}
$$

where $\tilde{m}_i$ stand for the complex neutrino mass eigenvalues, which can be expressed in terms of the Majorana phases $\phi_i \in [0, 2\pi)$ and the physical mass eigenvalues $m_i \geq 0$ as $\tilde{m}_i = m_i e^{-i\phi_i}$ with ϕ_3 chosen to be unphysical. These sum rules appeared in models based on A_4, S_4, A_5, T', T_7, $\Delta(54)$ and $\Delta(96)$ flavor symmetry, when the three neutrino mass eigenvalues can be described by two model parameters only, see Ref. [451] for a very good discussion and references on models predicting these sum rules. In addition, five new mass sum rules have been identified in flavor models based on modular symmetries with residual symmetries [434]. From Eq. (7.3) one sees that all neutrino mass sum rules can be parametrized in the following manner [451, 452]:

$$
c_1\left(m_1 e^{-i\phi_1}\right)^d e^{i\Delta\chi_{13}} + c_2\left(m_2 e^{-i\phi_2}\right)^d e^{i\Delta\chi_{23}} + m_3^d \;= 0\,,
\tag{7.4}
$$

with $c_1, c_2 > 0$, where ϕ_1 and ϕ_2 are the Majorana phases. The parameters c_1, c_2, d, $\Delta\chi_{13}$, and $\Delta\chi_{23}$ characterize the sum rule, and they can be straightforwardly read out for any of the above twelve known sum rules. Interpreting the complex numbers as vectors in the complex plane, the neutrino mass sum rule can be geometrically understood as a sum of three vectors which form a triangle in the complex plane [450]. Since the sum rule in Eq. (7.4) is a complex equation, it requires both real part and imaginary parts to be vanishing, i.e.,

$$
\begin{aligned}
&c_1 m_1^d\cos\beta + c_2 m_2^d\cos\alpha + m_3^d = 0\,,\\
&c_1 m_1^d\sin\beta + c_2 m_2^d\sin\alpha = 0\,,
\end{aligned}
\tag{7.5}
$$

with the angles $\alpha \equiv -d\phi_2 + \Delta\chi_{23}$ and $\beta \equiv -d\phi_1 + \Delta\chi_{13}$. Eq. (7.5) allows to express α and β in terms of the parameters of the sum rule,

$$
\cos\alpha = \frac{c_1^2 m_1^{2d} - c_2^2 m_2^{2d} - m_3^{2d}}{2c_2 m_2^d m_3^d}\,, \qquad \cos\beta = \frac{c_2^2 m_2^{2d} - c_1^2 m_1^{2d} - m_3^{2d}}{2c_1 m_1^d m_3^d}\,,
$$

$$\sin\alpha = \pm\frac{\sqrt{4c_2^2 m_2^{2d} m_3^{2d} - (c_1^2 m_1^{2d} - c_2^2 m_2^{2d} - m_3^{2d})^2}}{2c_2 m_2^d m_3^d}\,,$$
$$\sin\beta = \mp\frac{\sqrt{4c_2^2 m_2^{2d} m_3^{2d} - (c_1^2 m_1^{2d} - c_2^2 m_2^{2d} - m_3^{2d})^2}}{2c_1 m_1^d m_3^d}\,, \tag{7.6}$$

which relate Majorana CP phases with light neutrino masses. Given that $|\cos\alpha| \leq 1$, $|\sin\alpha| \leq 1$, $|\cos\beta| \leq 1$ and $|\sin\beta| \leq 1$, it follows that the validity of the sum rule requires the three neutrino masses to satisfy the following triangle inequalities,

$$\begin{aligned} & |c_1 m_1^d e^{i\beta}| \leq |c_2 m_2^d e^{i\alpha}| + |c_3 m_3^d|, \\ \text{and} \quad & |c_2 m_2^d e^{i\alpha}| \leq |c_1 m_1^d e^{i\beta}| + |c_3 m_3^d|, \\ \text{and} \quad & |c_3 m_3^d| \leq |c_1 m_1^d e^{i\beta}| + |c_2 m_2^d e^{i\alpha}|\,, \end{aligned} \tag{7.7}$$

which implies a triangle formed out of the three vectors $c_1 m_1^d e^{i\beta}$, $c_2 m_2^d e^{i\alpha}$ and $c_3 m_3^d$ in the complex plane. Given the solar and atmospheric neutrino mass squared differences $\Delta m^2_{\rm sol} \equiv m_2^2 - m_1^2 = 7.50^{+0.22}_{-0.20} \times 10^{-5}$ eV^2 and $\Delta m^2_{\rm atm} \equiv |m_3^2 - m_1^2| = 2.55\,(2.45)^{+0.02}_{-0.03} \times 10^{-3}$ eV^2 [24] for NO (IO), the light neutrino masses are related to the smallest mass m_1 (m_3) as follows,

$$\begin{aligned} & \text{NO}: \;\; m_2 = \sqrt{\Delta m^2_{\rm sol} + m_1^2}, \quad m_3 = \sqrt{\Delta m^2_{\rm atm} + m_1^2}\,, \\ & \text{IO}: \;\; m_1 = \sqrt{\Delta m^2_{\rm atm} + m_3^2}, \quad m_2 = \sqrt{\Delta m^2_{\rm atm} + \Delta m^2_{\rm sol} + m_3^2}\,. \end{aligned} \tag{7.8}$$

Thus mass rules usually lead to a lower limit on the lightest neutrino mass through the triangle inequalities in Eq. (7.7), and some sum rules may only allow for a certain mass ordering. For instance, neutrino masses can only be normal ordered for the sum rule $2\tilde{m}_2 + \tilde{m}_3 = \tilde{m}_1$ since the triangle inequality $m_3 + m_1 \geq 2m_2$ cannot be fulfilled for IO. Solving the inequality $m_3 \leq m_1 + 2m_2$ in the case of NO, one obtain the lower limit on m_1,

$$m_1 \geq \sqrt{\frac{\Delta m^2_{\rm atm}}{8}}\,(1 - 3r) \simeq 0.016 \text{ eV}\,, \tag{7.9}$$

where the exact result has been expanded in terms of the small ratio $r \equiv \Delta m^2_{\rm sol}/\Delta m^2_{\rm atm}$. Another benchmark sum rule $2\tilde{m}_2^{-1} + \tilde{m}_3^{-1} = \tilde{m}_1^{-1}$ works for both mass orderings, and the limits on the lightest neutrino mass from the triangle inequalities are found to be

$$\begin{aligned} & \text{NO}: \;\; 0.0043 \text{ eV} \simeq \sqrt{\frac{\Delta m^2_{\rm sol}}{3}}\left(1 - \frac{4\sqrt{3r}}{9}\right) \leq m_1 \leq \sqrt{\frac{\Delta m^2_{\rm sol}}{3}}\left(1 + \frac{4\sqrt{3r}}{9}\right) \simeq 0.0057 \text{ eV}\,, \\ & \text{IO}: \;\; m_3 \geq \sqrt{\frac{\Delta m^2_{\rm atm}}{8}}\left(1 + \frac{r}{3}\right) \simeq 0.018 \text{ eV}\,. \end{aligned} \tag{7.10}$$

One sees that certain mass sum rules predict a value of the lightest neutrino mass close to the current upper limit from cosmology and therefore they could be probed in the near future. Moreover, the lower

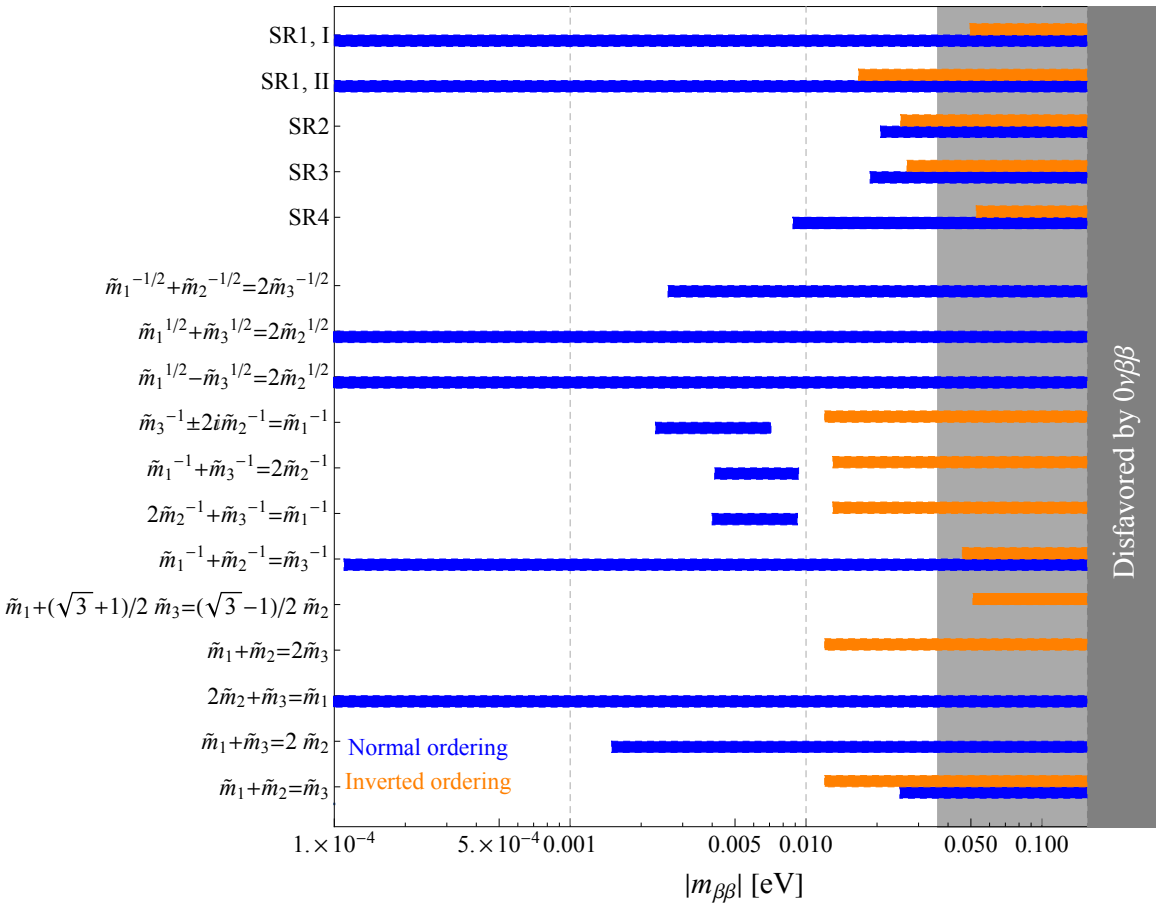


Figure 29: Predictions for $|m_{\beta\beta}|$ from different neutrino mass sum rules. The upper five mass sum rules have been derived in models based on modular symmetries [434], the lower twelve mass sum rules come from models based on discrete symmetries [451]. The grey regions show the constraints on $|m_{\beta\beta}|$ from Ref. [125] using commonly-adopted nuclear matrix-element calculations. Taken from [107].

bound on the lightest neutrino mass obtained in this way combined with the predictions for Majorana phases in Eq. (7.6) can be used to estimate the lower bound for the effective mass $|m_{\beta\beta}|$ from the general expression in Eq. (1.25). A detailed description of the analysis procedure can be found in Refs. [450, 451]. The predictions for the effective Majorana neutrino mass $|m_{\beta\beta}|$ for the different sum rules are displayed in figure 29.

One sees that certain sum rules can be fully or partially probed by the next generation of $0\nu\beta\beta$ decay experiments even taking into account the large uncertainties in the nuclear matrix element calculations. The exact mass sum rule could be violated by the higher-order terms resulting from flavour symmetry breaking or by the renormalization group evolution effects [452]. It was shown that the predictions of the sum rules are still valid at least qualitatively [452, 453]. To sum up one can say that, precision measurements of mixing angles, an observation of neutrinoless double beta decay can also provide insights into underlying flavor symmetries.

7.3 Flavor symmetry toolkit

One can check the validity of flavor symmetry models and constrain their predicted mixing parameters by comparing them those extracted from global oscillation fit results [24–27]. Nevertheless, numerical simulation of neutrino experiments and statistical analysis are often necessary to test and discriminate theoretical models in the current and future neutrino oscillation experiments such as NOνA, T2K, JUNO, DUNE and T2HK. The public available software such as `GLoBES` [454, 455], `Prob3++` [456], `nuCRAFT` [457] and `nuSQuIDS` [458] have been widely used to simulate the experimental characteristics of neutrino oscillation experiments. The compatibility between the experimental data and the expected outcome of a given neutrino experiment is frequently evaluated by the chi-square test. The involved simulation and analysis with `GLoBES` is quite involved and time-consuming. In order to facilitate the analysis of leptonic flavour symmetry models in neutrino oscillation experiments, a dedicated package `FaSE-GLoBES` has been recently developed [459] and is available via the link `https://github.com/tcwphy/FASE_GLoBES`. `FaSE-GLoBES` is a supplemental tool for `GLoBES`, written in `c/c++` language, and allows users to assign any flavour symmetry model and analyze how it can be constrained and tested by the simulated neutrino oscillation experiments.

8 Benchmark UV-complete models in 4-D

In addition to the model-independent approaches described in previous chapters, the flavour problem may be tackled in a more complete, UV-complete manner, by guessing the structure of the underlying family symmetry and building explicit models on a case-by-case basis, for reviews see, e.g. Refs. [86–91]. In this chapter we present two simple extensions of the Standard Model implementing family symmetry within the renormalisable $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge field theoretic framework.

8.1 A_4 family symmetry in a scotogenic model

Here A_4 is used as flavour symmetry within the scotogenic picture [196] in which the neutrino masses are generated radiatively, and the lightest of the mediators is identified with the dark matter particle. We adopt the singlet-triplet extension [197–201] of the original model [196], making it substantially richer in the associated phenomenology [138]. We employ the Ma-Rajasekaran basis [53], the representation matrices of the generators and the Clebsch-Gordon coefficients are listed in table 12.

We now present a flavored extension of the theory proposed in [205]. The basic fields and their symmetry tranformation properties are summarized in table 4. The left-handed leptons form an A_4 triplet, while right-handed ones come in as inequivalent singlets. Besides the SM fields, the original singlet-triplet scotogenic model [197] contains new weak iso-triplet and iso-singlet fermions Σ and F. Together with the Higgs scalars ϕ and Ω, these transform as triplets under the action of the family group A_4, while the dark scalar η is a flavour singlet.

A characteristic prediction of this model is the lower bound for the $0\nu\beta\beta$ decay amplitute, discussed in section 1.4. This follows from its incomplete fermion multiplet nature, see figure 4. In contrast to its original form, both charged lepton and neutrino mass matrices will now have a non-trivial structure, predicting trimaximal neutrino mixing. As usual, the $\mathbb{Z}_2$ parity is imposed in order to ensure the stability of the dark matter candidate and the radiative nature of neutrino mass generation. With the fields and

	Standard Model		dark fermions		scalars		
	L	e_R, μ_R, τ_R	Σ	F	ϕ	η	Ω
multiplicity	3	3	3	3	3	1	3
$\mathrm{SU(3)_c}$	1	1	1	1	1	1	1
$\mathrm{SU(2)_L}$	2	1	3	1	2	2	3
$\mathrm{U(1)_Y}$	-1	-2	0	0	1	1	0
$\mathbb{Z}_2$	1	1	-1	-1	1	-1	1
A_4	$\mathbf{3}$	$\mathbf{1}, \mathbf{1'}, \mathbf{1''}$	$\mathbf{3}$	$\mathbf{3}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{3}$

Table 4: Transformation properties of the fields in the scotogenic model with A_4 family symmetry [205].

symmetry assignments in table 4, we can read out the Yukawa terms relevant to fermion masses as follows,

$$\begin{aligned}\mathcal{L}_{\mathcal{Y}} \supset & -y_e(\overline{L}\phi)_{\mathbf{1}}e_R - y_\mu(\overline{L}\phi)_{\mathbf{1''}}\mu_R - y_\tau(\overline{L}\phi)_{\mathbf{1'}}\tau_R - Y_F\left(\overline{L}F\right)_{\mathbf{1}}\tilde{\eta} - Y_\Sigma(\overline{L}\tilde{\Sigma}^c)_{\mathbf{1}}\tilde{\eta} \\ & -Y_{\Omega,1}\left(\mathrm{Tr}\left[\left(\overline{\Sigma}\Omega\right)_{\mathbf{3}_S}F\right]\right)_{\mathbf{1}} - Y_{\Omega,2}\left(\mathrm{Tr}\left[\left(\overline{\Sigma}\Omega\right)_{\mathbf{3}_A}F\right]\right)_{\mathbf{1}} + \mathrm{h.c.}\,. \end{aligned} \tag{8.1}$$

In addition we have the bare mass terms

$$\mathcal{L}_{\mathcal{M}} \supset -\frac{1}{2}M_{\Sigma}\mathrm{Tr}\big((\overline{\Sigma}\tilde{\Sigma}^{c})_{\mathbf{1}}\big) - \frac{1}{2}M_{F}\left(\overline{F^{c}}F\right)_{\mathbf{1}} + \mathrm{h.c.}\,, \tag{8.2}$$

where $\tilde{\eta} = i\sigma_2\eta^*$. The $SU(2)_L$ triplets Σ and Ω are written in 2×2 matrix notation as

$$\Omega = \begin{pmatrix} \Omega^0/\sqrt{2} & \Omega^+ \\ \Omega^- & -\Omega^0/\sqrt{2} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma^0/\sqrt{2} & \Sigma^+ \\ \Sigma^- & -\Sigma^0/\sqrt{2} \end{pmatrix} \tag{8.3}$$

with $\tilde{\Sigma}^c \equiv i\sigma_2\Sigma^c i\sigma_2$. The scalar triplet Ω is assumed to be real.

The A_4 flavour symmetry is broken by the VEVs of the scalar fields ϕ and Ω, with the following VEV alignment in flavour space

$$\langle\phi\rangle = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} v_{\phi}, \quad \langle\Omega\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} v_{\Omega}, \quad \langle\eta\rangle = 0\,, \tag{8.4}$$

which can be a global minimum of the A_4-invariant scalar potential in certain regions of parameter space [205]. Notice that the VEVs of ϕ and Ω break the A_4 flavour symmetry down to the subgroups Z_3^t and Z_2^s respectively, where the superscripts denote the generators of the subgroups. The ρ parameter constrains the VEV v_Ω to be small, and the current electroweak precision tests lead to [28]

$$v_{\Omega} \leq 4.5 \text{ GeV} \quad \text{at } 3\sigma \text{ CL} \tag{8.5}$$

8.1.1 Charged lepton masses

The first three terms in the Yukawa Lagrangian are responsible for the charged lepton masses. Inserting the VEV of ϕ and using the multiplication law for the contraction of two triplets in table 12, one can straightforwardly read out the charged lepton mass matrix as

$$M_{\ell} = \begin{pmatrix} y_e & y_\mu & y_\tau \\ y_e & \omega y_\mu & \omega^2 y_\tau \\ y_e & \omega^2 y_\mu & \omega y_\tau \end{pmatrix} v_{\phi}\,. \tag{8.6}$$

The matrix $M_\ell M_\ell^\dagger$ can be diagonalized to $\mathrm{diag}(3|y_e v_\phi|^2, 3|y_\mu v_\phi|^2, 3|y_\tau v_\phi|^2)$ by means of the unitary transformation

$$U_{\ell} = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \tag{8.7}$$

which is a constant matrix.

8.1.2 Dark fermion masses

With the alignment of Ω in Eq. (8.4), we find the mass matrix of the dark fermions F and Σ^0 is of the following form

$$M_\chi = \begin{pmatrix} M_\Sigma & 0 & 0 & 0 & 0 & 0 \\ 0 & M_F & 0 & 0 & 0 & 0 \\ 0 & 0 & M_\Sigma & (Y_{\Omega,1}-Y_{\Omega,2})v_\Omega & 0 & 0 \\ 0 & 0 & (Y_{\Omega,1}-Y_{\Omega,2})v_\Omega & M_F & 0 & 0 \\ 0 & 0 & 0 & 0 & M_\Sigma & (Y_{\Omega,1}+Y_{\Omega,2})v_\Omega \\ 0 & 0 & 0 & 0 & (Y_{\Omega,1}+Y_{\Omega,2})v_\Omega & M_F \end{pmatrix} \tag{8.8}$$

in the convention of $-\frac{1}{2}(\overline{\Sigma_1^0}, \overline{F_1^c}, \overline{\Sigma_2^0}, \overline{F_3^c}, \overline{\Sigma_3^0}, \overline{F_2^c}) M_\chi \left(\Sigma_1^{0c}, F_1, \Sigma_2^{0c}, F_3, \Sigma_3^{0c}, F_2\right)^T$. The last block in Eq. (8.8) determines the masses of the dark Majorana fermions F and Σ. The symmetric complex 6×6 matrix M_χ can be diagonalized by a 6×6 block-diagonal matrix V:

$$V^T M_\chi V = \mathrm{diag}(m_{\chi_1^0}, m_{\chi_2^0}, m_{\chi_3^0}, m_{\chi_4^0}, m_{\chi_5^0}, m_{\chi_6^0})\,, \tag{8.9}$$

with

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\theta_1\, e^{i(\phi_1+\varrho_1)/2} & \sin\theta_1\, e^{i(\phi_1+\sigma_1)/2} & 0 & 0 \\ 0 & 0 & -\sin\theta_1\, e^{i(-\phi_1+\varrho_1)/2} & \cos\theta_1\, e^{i(-\phi_1+\sigma_1)/2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\theta_2\, e^{i(\phi_2+\varrho_2)/2} & \sin\theta_2\, e^{i(\phi_2+\sigma_2)/2} \\ 0 & 0 & 0 & 0 & -\sin\theta_2\, e^{i(-\phi_2+\varrho_2)/2} & \cos\theta_2\, e^{i(-\phi_2+\sigma_2)/2} \end{pmatrix}. \tag{8.10}$$

Following the method described in D, one finds that the rotation angles θ_1 and θ_2 are given as

$$\tan 2\theta_1 = \frac{\Delta_{34}}{M_F^2 - M_\Sigma^2}\,, \qquad \tan 2\theta_2 = \frac{\Delta_{56}}{M_F^2 - M_\Sigma^2}\,, \tag{8.11}$$

with

$$\begin{aligned} \Delta_{34} &= 2Y_-\sqrt{M_\Sigma^2 + M_F^2 + 2M_\Sigma M_F \cos 2\phi_{34}}\,, \\ \Delta_{56} &= 2Y_+\sqrt{M_\Sigma^2 + M_F^2 + 2M_\Sigma M_F \cos 2\phi_{56}}\,, \\ Y_- &\equiv |(Y_{\Omega,1}-Y_{\Omega,2})v_\Omega|\,, \qquad \phi_{34} \equiv \arg((Y_{\Omega,1}-Y_{\Omega,2})v_\Omega)\,, \\ Y_+ &\equiv |(Y_{\Omega,1}+Y_{\Omega,2})v_\Omega|\,, \qquad \phi_{56} \equiv \arg((Y_{\Omega,1}+Y_{\Omega,2})v_\Omega)\,. \end{aligned} \tag{8.12}$$

The eigenvalues $M_{\chi^0_{1,2,3,4,5,6}}$ are given by

$$m_{\chi_1^0} = M_\Sigma\,, \qquad m_{\chi_2^0} = M_F\,,$$

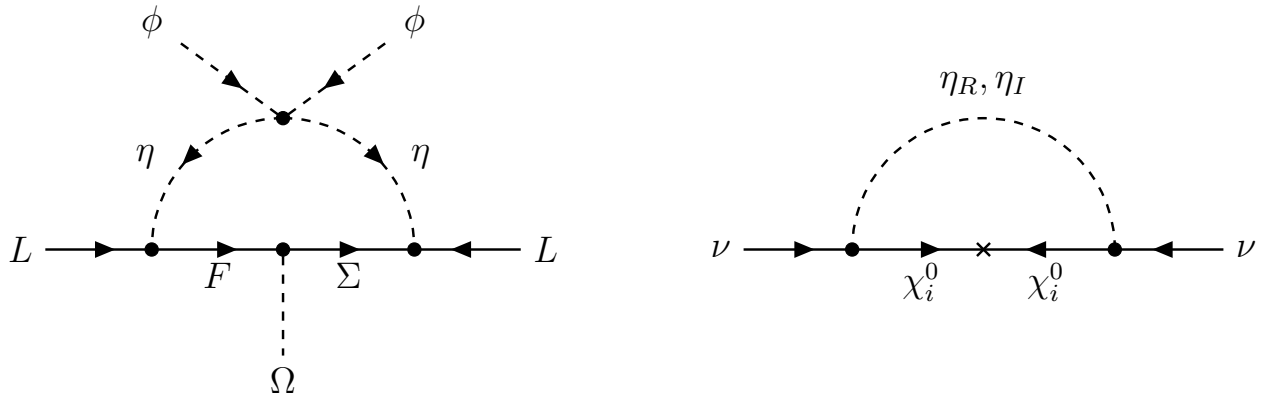


Figure 30: Feynman diagrams for scotogenic neutrino mass generation in the weak- (left panel) and mass-eigenstate basis (right panel).

$$
\begin{aligned}
m^2_{\chi^0_3} &= \frac{1}{2}\left(M^2_\Sigma + M^2_F + 2Y^2_- - \sqrt{(M^2_F - M^2_\Sigma)^2 + \Delta^2_{34}}\right) , \\
m^2_{\chi^0_4} &= \frac{1}{2}\left(M^2_\Sigma + M^2_F + 2Y^2_- + \sqrt{(M^2_F - M^2_\Sigma)^2 + \Delta^2_{34}}\right) , \\
m^2_{\chi^0_5} &= \frac{1}{2}\left(M^2_\Sigma + M^2_F + 2Y^2_+ - \sqrt{(M^2_F - M^2_\Sigma)^2 + \Delta^2_{56}}\right) , \\
m^2_{\chi^0_6} &= \frac{1}{2}\left(M^2_\Sigma + M^2_F + 2Y^2_+ + \sqrt{(M^2_F - M^2_\Sigma)^2 + \Delta^2_{56}}\right) ,
\end{aligned} \tag{8.13}
$$

The Majorana fermion mass eigenstates $\chi^0_{1,2,3,4,5,6}$ are related with Σ^c_0 and F by the unitary transformation V via

$$
\begin{pmatrix} \Sigma^{0c}_1 \\ F_1 \\ \Sigma^{0c}_2 \\ F_3 \\ \Sigma^{0c}_3 \\ F_2 \end{pmatrix} = V \begin{pmatrix} \chi^0_1 \\ \chi^0_2 \\ \chi^0_3 \\ \chi^0_4 \\ \chi^0_5 \\ \chi^0_6 \end{pmatrix} . \tag{8.14}
$$

8.1.3 Scotogenic neutrino masses

We now turn to neutrino masses. These arise radiatively, at the one-loop level, as shown in figure 30, mediated by a "dark sector", within a scotogenic setup. In contrast to the "flavour-blind" singlet-triplet scotogenic model, the dark fermions now transform as A_4 triplets and all of the six dark fermions mediate the one-loop diagrams. The interactions contributing to neutrino mass generation arise from the Yukawa terms involving Y_F and Y_Σ.

In the mass-eigenstate basis of dark Majorana fermions, the relevant Lagrangian is of the following

form

$$\mathcal{L}_\nu = -\frac{1}{\sqrt{2}} h_{\alpha i} \overline{\nu_\alpha} \eta_R \chi_i^0 + \frac{i}{\sqrt{2}} h_{\alpha i} \overline{\nu_\alpha} \eta_I \chi_i^0 - \frac{1}{\sqrt{2}} h^*_{\alpha i} \overline{\chi_i^0} \eta_R \nu_\alpha - \frac{i}{\sqrt{2}} h^*_{\alpha i} \overline{\chi_i^0} \eta_I \nu_\alpha \,, \tag{8.15}$$

where α is the family index and the rectangular matrix h is given by

$$h = \begin{pmatrix} \frac{Y_\Sigma}{\sqrt{2}} & Y_F & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{Y_\Sigma}{\sqrt{2}} & 0 & 0 & Y_F \\ 0 & 0 & 0 & Y_F & \frac{Y_\Sigma}{\sqrt{2}} & 0 \end{pmatrix} V \,. \tag{8.16}$$

Calculating the one-loop diagram in figure 30, we find that the radiatively generated neutrino mass matrix is given by

$$(m_\nu)_{\alpha\beta} = i \sum_i \frac{h_{\alpha i} h_{\beta i}}{32\pi^2} m_{\chi_i^0} \left(- \frac{m^2_{\eta_R} \ln(\frac{m^2_{\eta_R}}{m^2_{\chi_i^0}})}{m^2_{\eta_R} - m^2_{\chi_i^0}} + \frac{m^2_{\eta_I} \ln(\frac{m^2_{\eta_I}}{m^2_{\chi_i^0}})}{m^2_{\eta_I} - m^2_{\chi_i^0}} \right) , \tag{8.17}$$

where m_{η_R} and m_{η_I} denote the masses of η_R and η_I respectively. Note that η_R and η_I are the real and imaginary parts of the neutral field $\eta^0 = (\eta_R + i\eta_I)/\sqrt{2}$. It is notable that the light neutrino mass matrix is predicted to be block-diagonal,

$$m_\nu = \begin{pmatrix} \times & 0 & 0 \\ 0 & \times & \times \\ 0 & \times & \times \end{pmatrix} , \tag{8.18}$$

where the symbol "$\times$" indicates a non-vanishing element. The reason is that only the flavon Ω is involved in the neutrino sector and its VEV preserves the Z_2^s subgroup, so that the light neutrino mass matrix is invariant under the action of the A_4 generator s, i,e., $\rho^\dagger_{\mathbf{3}}(s) m_\nu \rho^*_{\mathbf{3}}(s) = m_\nu$. This implies the form in Eq. (8.18). The corresponding neutrino diagonalization matrix is of the form

$$U_\nu = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_\nu & \sin\theta_\nu e^{i\delta_\nu} \\ 0 & -\sin\theta_\nu e^{-i\delta_\nu} & \cos\theta_\nu \end{pmatrix} , \tag{8.19}$$

which satisfies $U_\nu^\dagger \mathcal{M}_\nu U_\nu^* = \mathrm{diag}(m_1, m_2, m_3)$.

One sees that at this level there is no solar mixing. However, including the contribution from the charged lepton sector one obtains a realistic lepton mixing matrix given as

$$U = \frac{1}{\sqrt{3}} \begin{pmatrix} \cos\theta_\nu - \sin\theta_\nu e^{-i\delta_\nu} & 1 & \cos\theta_\nu + \sin\theta_\nu e^{i\delta_\nu} \\ \omega^2 \cos\theta_\nu - \omega \sin\theta_\nu e^{-i\delta_\nu} & 1 & \omega \cos\theta_\nu + \omega^2 \sin\theta_\nu e^{i\delta_\nu} \\ \omega \cos\theta_\nu - \omega^2 \sin\theta_\nu e^{-i\delta_\nu} & 1 & \omega^2 \cos\theta_\nu + \omega \sin\theta_\nu e^{i\delta_\nu} \end{pmatrix} . \tag{8.20}$$

with non-vanishing solar mixing angle. One sees the full lepton mixing matrix is predicted in terms of two free parameters θ_ν and δ_ν. Without loss of generality, these can be taken in the regions $0 \leq \theta_\nu \leq \pi$ and $0 \leq \delta_\nu \leq \pi$.

One also notes that the lepton mixing matrix U has the so-called trimaximal **TM2** form [205], since the second column is fixed to be $\frac{1}{\sqrt{3}}(1,1,1)^T$ [316–318]. From the mixing matrix (8.20) one can express the lepton mixing angles and the leptonic Jarlskog invariant in terms of just two free parameters

$$\sin^2\theta_{13} = \frac{1+\sin 2\theta_\nu \cos\delta_\nu}{3}\,, \quad \sin^2\theta_{12} = \frac{1}{2-\sin 2\theta_\nu\cos\delta_\nu}\,,$$
$$\sin^2\theta_{23} = \frac{1}{2} - \frac{\sqrt{3}\sin 2\theta_\nu \sin\delta_\nu}{4-2\sin 2\theta_\nu\cos\delta_\nu}\,, \quad J_{CP} = \frac{\cos 2\theta_\nu}{6\sqrt{3}}\,.$$

One predicts the following relations amongst the mixing angles and the CP phase,

$$\cos^2\theta_{13}\sin^2\theta_{12} = \frac{1}{3}\,, \;\; \cos\delta_{CP} = \frac{2(3\cos^2\theta_{12}\cos^2\theta_{23} + 3\sin^2\theta_{12}\sin^2\theta_{13}\sin^2\theta_{23} - 1)}{3\sin 2\theta_{23}\sin 2\theta_{12}\sin\theta_{13}}\,. \tag{8.21}$$

Using the 3σ range of the reactor angle $2.000\times10^{-2} \leq \sin^2\theta_{13} \leq 2.405\times10^{-2}$ for NO and $2.018\times10^{-2} \leq \sin^2\theta_{13} \leq 2.424\times10^{-2}$ for IO [24, 25], we predict narrow ranges for the solar mixing angle [205],

$$\text{NO:}\;\; 0.3401 \leq \sin^2\theta_{12} \leq 0.3415\,, \quad \text{IO:}\;\; 0.3402 \leq \sin^2\theta_{12} \leq 0.3416\,, \tag{8.22}$$

These predictions are very close to the 1σ upper limits from the general neutrino oscillation global fit [24, 25] and should be testable in forthcoming neutrino oscillation experiments.

To sum up we have obtained the three lepton mixing angles and the Dirac CP phase in terms of just two parameters θ_ν and δ_ν as given in figure 31. The resulting predictions for the two most poorly determined oscillation parameters $\sin^2\theta_{23}$ and δ_{CP} are given in figure 32, where the star and dot stand

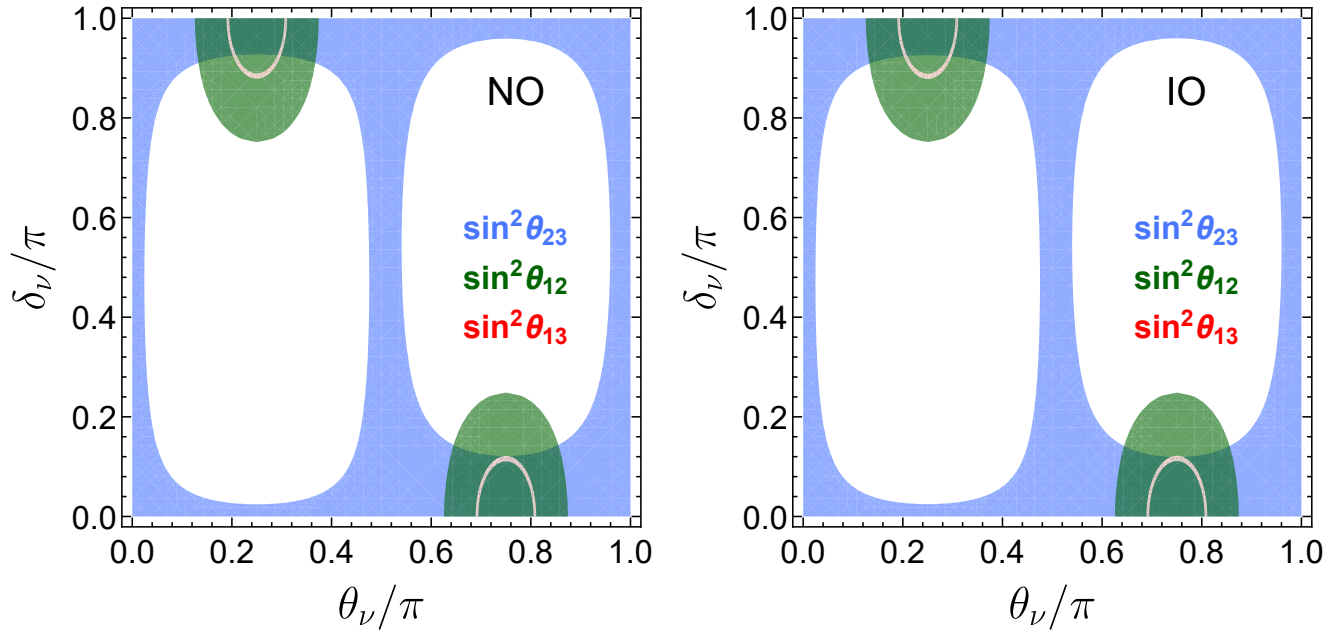


Figure 31: Contour plots of $\sin^2\theta_{12}$, $\sin^2\theta_{13}$, and $\sin^2\theta_{23}$ in the $\theta_\nu - \delta_\nu$ plane. The red, green and blue areas denote the allowed 3σ regions of $\sin^2\theta_{13}$, $\sin^2\theta_{12}$ and $\sin^2\theta_{23}$ respectively.

for the global best fit points for NO and IO respectively.

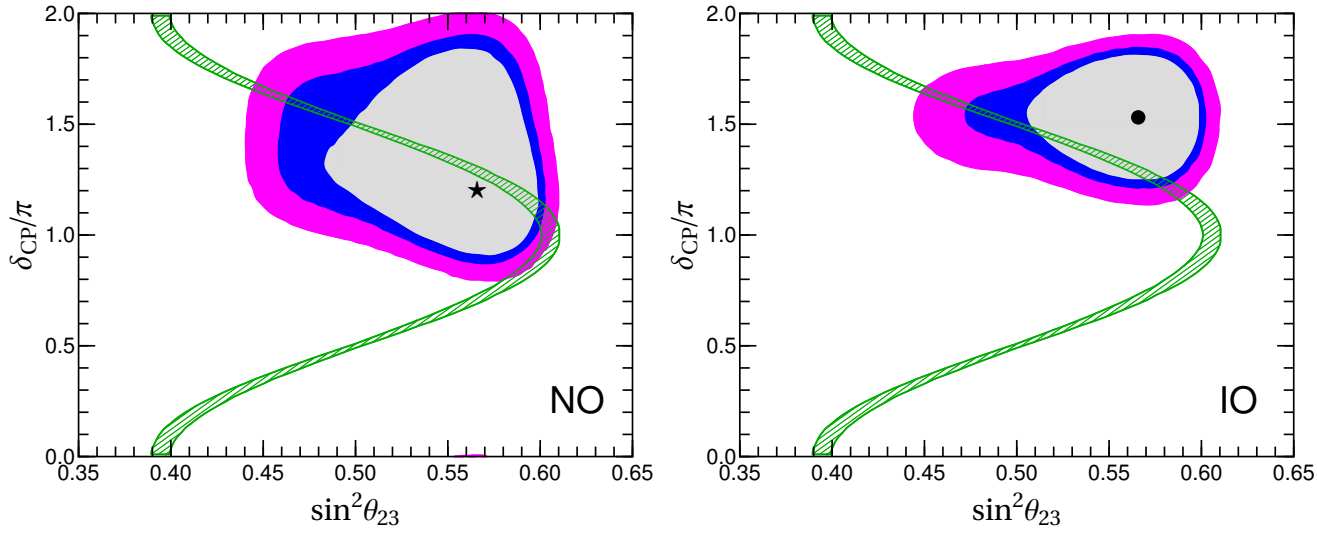


Figure 32: The hatched green bands indicate the predicted correlations between δ_{CP} and $\sin^2\theta_{23}$ for both neutrino mass orderings. The undisplayed parameters $\sin^2\theta_{13}$ and $\sin^2\theta_{12}$ are required to lie within their 3σ regions from the current oscillation global fit [24, 25]. The generic 1σ, 2σ and 3σ regions from the current neutrino oscillation global fit are indicated by the shaded areas [24, 25].

One sees also that the CP phase δ_{CP} is predicted to lie in a range narrower than that obtained in the generic global fits of neutrino oscillations. We also mention that in this model the smallness of the reactor angle θ_{13} and CP violation parameter J_{CP} has a dynamical origin, given by the small ratio $v_\Omega/(M_F+M_\Sigma)$ involving the triple Higgs VEV, see [205] for details. Finally, as already mentioned, the incomplete fermion multiplet structure implies that one of the neutrinos is massless, leading to the $0\nu\beta\beta$ decay predictions in figure 4. All in all, this construction offers a serious benchmark theory for neutrino oscillations and dark matter.

8.2 A benchmark model with both flavour and CP symmetries

In this section, we describe a model implementing both the S_4 flavour symmetry and the generalized CP symmetry [325]. See Refs. [320, 394] for alternative models. It realizes the breaking patterns of flavour and CP symmetry analyzed in subsection 6.3.1. We adopt a supersymmetric (SUSY) formulation of the model in four dimensional space-time. The model is renormalizable at high energies, it gives rise to tri-bimaximal neutrino mixing at leading order, and the next-to-leading order contributions break the tri-bimaximal to a trimaximal mixing pattern. We introduce the auxiliary symmetry $Z_3 \times Z_4$ so as to generate charged lepton mass hierarchies and forbid unwanted operators.

In order to construct the superpotential responsible for the alignment of the flavon vacuum expectation values, we use the standard supersymmetric driving field mechanism [313]. We assume the existence of an R-symmetry $U(1)_R$ containing the usual R-parity, under which the Higgs and flavon fields are uncharged, while matter fields carry a $+1$ R-charge. In addition, the so-called driving fields are necessary and they

carry two units of R-charge. Therefore all terms in the superpotential should be bilinear in the matter superfields or linear in the driving field.

8.2.1 Flavon superpotential

The matter and flavon fields and their transformation properties under the flavour symmetry are summarized in table 5. On the other hand the driving fields in our model and their transformation rules under

Field	L	N^c	e^c	μ^c	τ^c	$H_{u,d}$	φ_T	η	φ_S	ϕ	ξ	Δ
S_4	$\mathbf{3}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{1'}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{3}$	$\mathbf{2}$	$\mathbf{3'}$	$\mathbf{2}$	$\mathbf{1}$	$\mathbf{1'}$
Z_3	ω	ω^2	ω^2	1	ω	1	ω	ω	ω^2	ω^2	ω^2	1
Z_4	1	1	i	-1	$-i$	1	i	i	1	1	1	1
$U(1)_R$	1	1	1	1	1	0	0	0	0	0	0	0

Table 5: Transformation properties of matter, Higgs and flavon fields.

the flavour symmetry group are listed in table 6. The most general driving superpotential $\mathcal{W}_d$ invariant under $S_4 \times Z_3 \times Z_4$ with $R = 2$ is given by [325]

$$\begin{aligned}\mathcal{W}_d = {} & g_1 \left(\varphi_T^0 \left(\varphi_T \varphi_T\right)_{\mathbf{3'}}\right)_{\mathbf{1}} + g_2 \left(\varphi_T^0 \left(\eta \varphi_T\right)_{\mathbf{3'}}\right)_{\mathbf{1}} + g_3 \zeta^0 \left(\varphi_T \varphi_T\right)_{\mathbf{1}} + g_4 \zeta^0 \left(\eta\eta\right)_{\mathbf{1}} \\ & + f_1 \left(\varphi_S^0 \left(\varphi_S \varphi_S\right)_{\mathbf{3'}}\right)_{\mathbf{1}} + f_2 \left(\varphi_S^0 \left(\phi \varphi_S\right)_{\mathbf{3'}}\right)_{\mathbf{1}} + f_3 \left(\varphi_S^0 \varphi_S\right)_{\mathbf{1}} \xi + f_4 \left(\widetilde{\varphi}_S^0 \left(\phi \varphi_S\right)_{\mathbf{3}}\right)_{\mathbf{1}} \\ & + f_5 \xi^0 \left(\varphi_S \varphi_S\right)_{\mathbf{1}} + f_6 \xi^0 \left(\phi\phi\right)_{\mathbf{1}} + f_7 \xi^0 \xi^2 + M^2 \Delta^0 + f_8 \Delta^0 \Delta^2 \,. \end{aligned} \tag{8.23}$$

Since we impose CP symmetry on the theory in the unbroken phase, all couplings in w_d are real. In the limit of unbroken supersymmetry, the minimum of the scalar potential is determined by vanishing $F-$terms for the driving fields. As a result, the VEVs of the flavons are aligned as follows [325]

$$\begin{aligned} \langle \varphi_T \rangle &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} v_T, \qquad \langle \eta \rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} v_\eta, \qquad \langle \Delta \rangle = v_\Delta \,, \\ \langle \varphi_S \rangle &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} v_S, \qquad \langle \phi \rangle = \begin{pmatrix} 1 \\ 1 \end{pmatrix} v_\phi, \qquad \langle \xi \rangle = v_\xi \,, \end{aligned} \tag{8.24}$$

Field	φ_T^0	ζ^0	φ_S^0	$\widetilde{\varphi}_S^0$	ξ^0	Δ^0	Ω_1	Ω_1^c	Ω_2	Ω_2^c	Ω_3	Ω_3^c	Σ	Σ^c
S_4	$\mathbf{3'}$	$\mathbf{1}$	$\mathbf{3'}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{2}$	$\mathbf{2}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{3}$	$\mathbf{3'}$	$\mathbf{3'}$
Z_3	ω	ω	ω^2	ω^2	ω^2	1	1	1	ω	ω^2	ω^2	ω	ω^2	ω
Z_4	-1	-1	1	1	1	1	-1	-1	$-i$	i	1	1	1	1
$U(1)_R$	2	2	2	2	2	2	1	1	1	1	1	1	1	1

Table 6: Transformation rules for driving and messenger fields under the $S_4 \times Z_4 \times Z_3$ and $U(1)_R$ symmetries.

with

$$v_T=\frac{g_2}{2g_1}v_\eta,\quad v_S^2=-\frac{1}{6f_2^2f_5}\left(f_3^2f_6+2f_2^2f_7\right)v_\xi^2,\quad v_\phi=-\frac{f_3}{2f_2}v_\xi,\quad v_\Delta^2=-M^2/f_8\,, \tag{8.25}$$

where v_η and v_ξ are undetermined, as they are related to a flat direction. The phase of v_η can be absorbed into the lepton fields, thus we can take the VEVs v_η and v_T real without loss of generality. The common phase of v_S, v_ϕ and v_ξ does not affect neutrino masses and mixing, because it can be factored out in the light neutrino mass matrix. As a consequence, v_ϕ and v_ξ can be considered real while v_S is real or purely imaginary depending on the coefficient $-\left(f_3^2f_6+2f_2^2f_7\right)/(f_2^2f_5)$ being positive or negative. Furthermore, the VEV v_Δ is real for $f_8<0$ and purely imaginary for $f_8>0$.

8.2.2 The charged lepton sector

The renormalizable Yukawa superpotential terms involving charged leptons is obtained by integrating out the three pairs of messengers Ω_i and Ω_i^c $(i=1,2,3)$. These are chiral superfields with non-vanishing hypercharge $+2(-2)$ for Ω_i (Ω_i^c). Given the field content and the symmetry assignments in tables 5 and 6, one can read off the superpotential for the charged leptons as follows,

$$\begin{aligned}\mathcal{W}_\ell=\;&z_1\left(L\Omega_3\right)_{\mathbf{1}}H_d+z_2\left(\Omega_3^c\varphi_T\right)_{\mathbf{1}}\tau^c+z_3\left(\left(\Omega_3^c\varphi_T\right)_{\mathbf{2}}\Omega_2\right)_{\mathbf{1}}+z_4\left(\Omega_2^c\eta\right)_{\mathbf{1}'}\mu^c\\&+z_5\left(\left(\Omega_2^c\eta\right)_{\mathbf{2}}\Omega_1\right)_{\mathbf{1}}+z_6\left(\Omega_1^c\eta\right)_{\mathbf{1}}e^c+M_{\Omega_1}\left(\Omega_1\Omega_1^c\right)_{\mathbf{1}}+z_7\Delta\left(\Omega_1\Omega_1^c\right)_{\mathbf{1}'}\\&+M_{\Omega_2}\left(\Omega_2\Omega_2^c\right)_{\mathbf{1}}+z_8\Delta\left(\Omega_2\Omega_2^c\right)_{\mathbf{1}'}+M_{\Omega_3}\left(\Omega_3\Omega_3^c\right)_{\mathbf{1}}\,,\end{aligned} \tag{8.26}$$

where CP invariance requires all coupling constants z_i and messenger masses M_{Ω_1}, M_{Ω_2} and M_{Ω_3} to be real. The terms $\Delta\left(\Omega_1\Omega_1^c\right)_{\mathbf{1}'}$ and $\Delta\left(\Omega_2\Omega_2^c\right)_{\mathbf{1}'}$ lead to corrections to the Ω_1 and Ω_2 masses respectively. The mass scales of the messenger fields are much larger than the flavon VEVs, hence the contributions of these two operators can be safely neglected. Integrating out the messenger pairs Ω_i and Ω_i^c as shown in figure 33, we obtain the effective superpotential for the charged lepton masses as

$$\mathcal{W}_\ell^{eff}=-\frac{z_1z_2}{M_{\Omega_3}}\left(L\varphi_T\right)_{\mathbf{1}}H_d\tau^c+\frac{z_1z_3z_4}{M_{\Omega_2}M_{\Omega_3}}\left(\left(L\varphi_T\right)_{\mathbf{2}}\eta\right)_{\mathbf{1}'}\mu^c-\frac{z_1z_3z_5z_6}{M_{\Omega_1}M_{\Omega_2}M_{\Omega_3}}\left(\left(L\varphi_T\right)_{\mathbf{2}}\left(\eta\eta\right)_{\mathbf{2}}\right)_{\mathbf{1}}H_de^c\,. \tag{8.27}$$

Using the flavon VEVs in Eq. (8.24), we find a diagonal charged lepton mass matrix with

$$m_\tau=-z_1z_2\frac{v_T}{M_{\Omega_3}}v_d,\quad m_\mu=z_1z_3z_4\frac{v_Tv_\eta}{M_{\Omega_2}M_{\Omega_3}}v_d,\quad m_e=-z_1z_3z_5z_6\frac{v_Tv_\eta^2}{M_{\Omega_1}M_{\Omega_2}M_{\Omega_3}}v_d\,, \tag{8.28}$$

where $v_d=\langle H_d\rangle$ is the VEV of the Higgs field H_d. We see that the electron, muon and tau masses are suppressed by different powers of v_T and v_η, so that the mass hierarchies among the charged leptons are naturally reproduced.

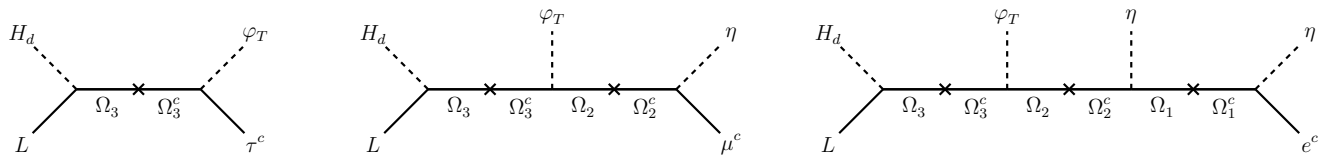


Figure 33: Diagrams for charged-lepton-mass effective operators, crosses indicating fermion mass insertions.

8.2.3 The neutrino sector

The renormalizable superpotential responsible for generating the light neutrino masses can be written as the sum of the leading-order terms and the relevant messenger terms:

$$\mathcal{W}_\nu = \mathcal{W}_\nu^{LO} + \mathcal{W}_\nu^{\Sigma} \ , \tag{8.29}$$

where

$$\begin{aligned} \mathcal{W}_\nu^{LO} &= y \left(L N^c\right)_{\mathbf{1}} H_u + y_1 \left(\left(N^c N^c\right)_{\mathbf{3'}} \varphi_S\right)_{\mathbf{1}} + y_2 \left(N^c N^c\right)_{\mathbf{1}} \xi + y_3 \left(\left(N^c N^c\right)_{\mathbf{2}} \phi\right)_{\mathbf{1}} \ , \\ \mathcal{W}_\nu^{\Sigma} &= x_1 \left(\left(N^c \Sigma\right)_{\mathbf{3'}} \varphi_S\right)_{\mathbf{1}} + x_2 \left(\left(N^c \Sigma\right)_{\mathbf{2}} \phi\right)_{\mathbf{1}} + x_3 \left(N^c \Sigma^c\right)_{\mathbf{1'}} \Delta + M_\Sigma \left(\Sigma \Sigma^c\right)_{\mathbf{1}} \ . \end{aligned} \tag{8.30}$$

where the couplings x_i and y_i are real due to CP invariance and the messenger field Σ (Σ^c) is a chiral superfield with vanishing hypercharge. The first term of $\mathcal{W}_\nu^{LO}$ gives rise to a very simple form for the Dirac neutrino mass matrix,

$$m_D = y v_u \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \ , \tag{8.31}$$

with $v_u = \langle H_u \rangle$. On the other hand, the last three terms of $\mathcal{W}_\nu^{LO}$ lead to the mass matrix m_M^{LO} of the right-handed neutrinos. Given the alignment of φ_S, ϕ and ξ shown in Eq. (8.24), we have

$$m_M^{LO} = y_1 v_s \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} + y_2 v_\xi \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} + y_3 v_\phi \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} . \tag{8.32}$$

The effective light neutrino mass matrix is given by the simplest type-I seesaw formula [32]

$$m_\nu^{LO} = -m_D (m_M^{LO})^{-1} m_D^T$$

which is exactly diagonalized by the tri-bimaximal mixing matrix U_{TBM},

$$U_{TBM}^T m_\nu^{LO} U_{TBM} = \text{diag}\left(m_1^{LO}, m_2^{LO}, m_3^{LO}\right) \ , \tag{8.33}$$

so we obtain

$$m_1^{LO}=-\frac{y^2v_u^2}{3y_1v_S+y_2v_\xi-y_3v_\phi},\ m_2^{LO}=-\frac{y^2v_u^2}{y_2v_\xi+2y_3v_\phi},\ m_3^{LO}=-\frac{y^2v_u^2}{3y_1v_S-y_2v_\xi+y_3v_\phi}\,.$$

Note that the tri-bimaximal mixing pattern is produced at leading order. This follows from the fact that the VEVs of the flavons φ_S, ϕ and ξ involved in the neutrino mass term are invariant under the action of the Klein subgroup generated by the tri-bimaximal s and u generators.

The leading-order and next-to-leading order (NLO) contributions to the right-handed neutrino masses are depicted in figure 34.

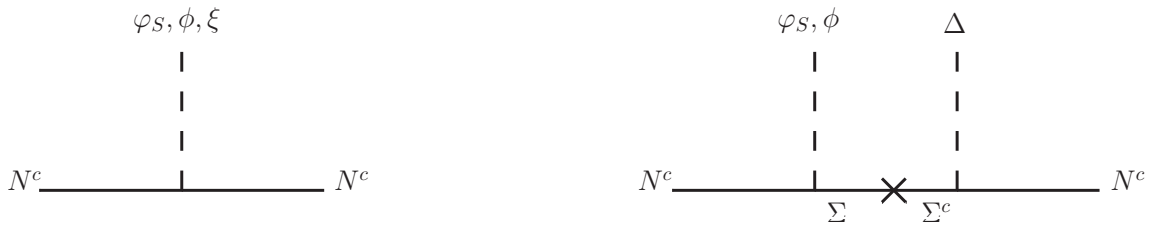


Figure 34: The diagrams for the RH neutrino masses, where the cross indicates a fermionic mass insertion.

Integrating out the messenger fields Σ and Σ^c, we obtain the following NLO effective operator

$$\mathcal{W}_\nu^{NLO}=-\frac{x_2x_3}{M_\Sigma}\Delta\left((N^cN^c)_{\mathbf{2}}\,\phi\right)_{\mathbf{1'}}\,. \tag{8.34}$$

Notice that the VEV of the flavon Δ breaks the residual Klein symmetry down to a Z_2 subgroup generated by s at NLO. Consequently, the corrected right-handed neutrino mass matrix is given as

$$m_M=a\begin{pmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{pmatrix}+b\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}+c\begin{pmatrix}0&1&1\\1&1&0\\1&0&1\end{pmatrix}+d\begin{pmatrix}0&1&-1\\1&-1&0\\-1&0&1\end{pmatrix}\,, \tag{8.35}$$

with

$$a=y_1v_S,\qquad b=y_2v_\xi,\qquad c=y_3v_\phi,\qquad d=x_2x_3\frac{v_\Delta v_\phi}{M_\Sigma}\,. \tag{8.36}$$

Therefore, up to an overall factor $y^2v_u^2$, the corrected light neutrino mass matrix obtained from the seesaw formula has the form

$$m_\nu=\alpha\begin{pmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{pmatrix}+\beta\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}+\gamma\begin{pmatrix}0&1&1\\1&1&0\\1&0&1\end{pmatrix}+\epsilon\begin{pmatrix}0&1&-1\\1&-1&0\\-1&0&1\end{pmatrix}\,, \tag{8.37}$$

where the parameters α, β, γ and ϵ are given by,

$$\alpha=\frac{a}{-9a^2+(b-c)^2+3d^2}\,,\qquad \beta=-\frac{1}{3(b+2c)}+\frac{2(b-c)}{3\left[9a^2-(b-c)^2-3d^2\right]}\,,$$

$$\gamma = -\frac{1}{3(b+2c)} - \frac{b-c}{3\left[9a^2-(b-c)^2-3d^2\right]}\,, \qquad \epsilon = \frac{d}{-9a^2+(b-c)^2+3d^2}\,, \tag{8.38}$$

The first three terms in the light neutrino matrix of Eq. (8.37) preserve the tri-bimaximal mixing form. The last term, which is proportional to ϵ, violates it. The associated parameter ϵ is induced by the NLO contributions suppressed by v_Δ/M_Σ with respect to α, β and γ. This naturally accounts for the small reactor mixing angle θ_{13} and the small deviation from the maximal atmospheric mixing.

As shown in section 8.2.1, the VEVs v_ϕ and v_ξ can be assumed real, while v_S and v_Δ can be real or purely imaginary due to the generalized CP symmetry. If v_S and v_Δ are real, the vacuum alignment of the flavons φ_S, ϕ, ξ and Δ is invariant under the CP transformation $X_{\mathbf{r}} = 1$ and the residual flavour symmetry Z_2^s. The parameters α, β, γ and ϵ are all real, the lepton mixing matrix has the form of Eq. (6.46) with the rotation angle θ given by

$$\tan 2\theta = \frac{\sqrt{3}\epsilon}{\gamma-\beta}\,. \tag{8.39}$$

The lepton mixing angles and CP violation phases are given in Eq. (6.47). The light neutrino masses are

$$\begin{aligned} m_1 &= \left|3\alpha - \text{sign}\left(\epsilon \sin 2\theta\right)\sqrt{(\gamma-\beta)^2+3\epsilon^2}\right|\,, \\ m_2 &= \left|\beta+2\gamma\right|\,, \\ m_3 &= \left|3\alpha + \text{sign}\left(\epsilon \sin 2\theta\right)\sqrt{(\gamma-\beta)^2+3\epsilon^2}\right|\,. \end{aligned} \tag{8.40}$$

This model allows for both neutrino mass orderings, either NO or IO.

Moreover, if the VEV v_S is real while v_Δ is pure imaginary, the neutrino sector would preserve the residual flavour symmetry Z_2^s and CP symmetry $X_{\mathbf{r}} = u$ which corresponds to the $\mu-\tau$ refection symmetry. The parameters α, β, γ in Eq. (8.37) are real, and ϵ is pure imaginary. We find that the lepton mixing matrix is given by Eq. (6.50) with the rotation angle θ given as

$$\tan 2\theta = \frac{i\epsilon}{\sqrt{3}\alpha}\,. \tag{8.41}$$

Both the atmospheric angle θ_{23} and the CP violation phase δ_{CP} are maximal, as shown in Eq. (6.51). The light neutrino masses are given by

$$\begin{aligned} m_1 &= \left|\beta-\gamma + \text{sign}\left(\alpha \cos 2\theta\right)\sqrt{9\alpha^2-3\epsilon^2}\right|\,, \\ m_2 &= \left|\beta+2\gamma\right|\,, \\ m_3 &= \left|\beta-\gamma - \text{sign}\left(\alpha \cos 2\theta\right)\sqrt{9\alpha^2-3\epsilon^2}\right|\,, \end{aligned} \tag{8.42}$$

Again, this is consistent with both neutrino mass orderings, either NO or IO.

All in all, this model implements the first two symmetry breaking patterns analyzed in detail in section 6.3.1. The reader is addressed to that section for a discussion of the corresponding neutrino

mixing, CP violation and $0\nu\beta\beta$ decay predictions. See Ref. [320] for an S_4 model realizing the third breaking pattern and the lepton mixing matrix of Eq. (6.53).

9 Family symmetry in 5-D models with a warped extra dimension

Extra dimensions [460] provide an interesting way to address the so-called hierarchy problem [461–463], by making the fundamental scale of gravity exponentially reduced from the Planck mass down to the TeV scale. This follows as a result of having the Standard Model fields localized near the boundary of the extra dimensions. Here we stress that the fermion mass hierarchies can also be addressed through the localization of fermion profiles which are fixed by the bulk mass parameters.

As we have shown in the previous section, the lepton and quark mixing angles can be well explained by using a discrete flavour symmetry within concrete 4-D models. One can also implement flavour symmetry in the context of extra dimensions, so that the structure of both mass hierarchies as well as mixing angles can be addressed in a clear manner. There has been intense activity using discrete family symmetries to build UV-complete 4-D gauge theories [85–93], describing the masses and mixing matrices of leptons and quarks. The pattern of fermion mixing could also arise from the imposition of family symmetries in extra dimensions, including holographic models [70–84].

In this section we consider warped flavourdynamics schemes, of which there have been two recent proposals in the literature [78, 79]. The first warped flavourdynamics model is based on the $\Delta(27)$ family symmetry, with neutrinos as Dirac fermions, and a predicted TM2 mixing pattern [78]. An alternative proposal for a warped flavourdynamics scheme is based on the T' family group [79]. Here neutrinos are Majorana fermions with a predicted TM1 mixing pattern. In what follows we develop the key features of this second example, and refer the interested reader to the original work in [78] for the other case.

9.1 Warped flavourdynamics with the T' family group

Here we present our benchmark extra-dimensional model with T' flavour symmetry [79]. The T' group is the double covering of A_4. The relation between T' and A_4 is quite similar to the familiar relation between SU(2) and SO(3). Although SU(2) and SO(3) possess the same Lie algebra, SO(3) has only odd-dimensional representations, while SU(2) possesses both even and odd-dimensional representations. Likewise T' has three doublet representations $\mathbf{2}$, $\mathbf{2}'$ and $\mathbf{2}''$ besides the singlets $\mathbf{1}$, $\mathbf{1}'$, $\mathbf{1}''$ and triplet $\mathbf{3}$ of A_4.

We formulate our model in the framework of the Randall-Sundrum model [462]. The bulk geometry is described by the following metric

$$ds^2 = e^{-2ky}\eta_{\mu\nu}dx^\mu dx^\nu - dy^2\,. \tag{9.1}$$

This extra dimension y is compactified, and the two 3-branes with opposite tension are located at $y = 0$, the UV brane, and the infra-red (IR) brane at $y = L$.

In order to comply with electroweak precision measurement constraints, the electroweak symmetry of the model is promoted to $\mathrm{G_{bulk}} = \mathrm{SU(2)_L} \otimes \mathrm{SU(2)_R} \otimes \mathrm{U(1)_{B-L}}$ [464, 465] and it is broken down to the standard model gauge group $\mathrm{G_{SM}} = \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ on the UV brane by the boundary conditions (BCs) of the gauge bosons. The Higgs field lives in the bulk, and it is in the $(\mathbf{2}, \mathbf{2})$ bi-doublet representation of

$\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{SU}(2)_\mathrm{R}$. The Kaluza-Klein (KK) Higgs field decomposition is [466]

$$H(x^\mu, y) = H(x^\mu)\frac{f_H(y)}{\sqrt{L}} + \text{heavy KK Modes}\,, \tag{9.2}$$

where $f_H(y)$ is the zero mode profile. For an adequate choice of BCs, we have

$$f_H(y) = \sqrt{\frac{2kL(1-\beta)}{1-e^{-2(1-\beta)kL}}}e^{kL}e^{(2-\beta)k(y-L)}\,, \tag{9.3}$$

with $\beta = \sqrt{4+m_H^2/k^2}$ and m_H is the bulk mass of the Higgs field.

The three families of leptons and quarks transform under $\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{SU}(2)_\mathrm{R}$ in the following way,

$$\Psi_{\ell_i} = \begin{pmatrix} \nu_i^{[++]} \\ e_i^{[++]} \end{pmatrix} \sim (\mathbf{2},\mathbf{1})\,, \quad \Psi_{e_i} = \begin{pmatrix} \tilde{\nu}_i^{[+-]} \\ e_i^{[--]} \end{pmatrix} \sim (\mathbf{1},\mathbf{2})\,, \quad \Psi_{\nu_i} = \begin{pmatrix} N_i^{[--]} \\ \tilde{e}_i^{[+-]} \end{pmatrix} \sim (\mathbf{1},\mathbf{2})\,, \tag{9.4}$$

$$\Psi_{Q_i} = \begin{pmatrix} u_i^{[++]} \\ d_i^{[++]} \end{pmatrix} \sim (\mathbf{2},\mathbf{1})\,, \quad \Psi_{d_i} = \begin{pmatrix} \tilde{u}_i^{[+-]} \\ d_i^{[--]} \end{pmatrix} \sim (\mathbf{1},\mathbf{2})\,, \quad \Psi_{u_i} = \begin{pmatrix} u_i^{[--]} \\ \tilde{d}_i^{[+-]} \end{pmatrix} \sim (\mathbf{1},\mathbf{2})\,. \tag{9.5}$$

where the two signs in the bracket indicate Neumann $(+)$ or Dirichlet $(-)$ boundary conditions for the left-handed component of the corresponding field on UV and IR branes respectively. The Kaluza-Klein decomposition of a 5D fermion for the two different BCs are

$$\psi^{[++]}(x^\mu, y) = \frac{e^{2ky}}{\sqrt{L}}\Big\{\psi_L(x^\mu)f_L(y, c_L) + \text{heavy KK modes}\Big\}\,, \tag{9.6}$$

$$\psi^{[--]}(x^\mu, y) = \frac{e^{2ky}}{\sqrt{L}}\Big\{\psi_R(x^\mu)f_R(y, c_R) + \text{heavy KK modes}\Big\}\,. \tag{9.7}$$

The 5D fields with $[++]$ BCs only have left-handed zero modes, while those with $[--]$ BCs only have right-handed ones. The functions $f_L(y, c_L)$ and $f_R(y, c_R)$ are the zero mode profiles [467, 468]

$$f_L(y, c_L) = \sqrt{\frac{(1-2c_L)kL}{e^{(1-2c_L)kL}-1}}e^{-c_L ky}\,, \quad f_R(y, c_R) = \sqrt{\frac{(1+2c_R)kL}{e^{(1+2c_R)kL}-1}}e^{c_R ky}\,, \tag{9.8}$$

where c_L and c_R denote the bulk mass of the 5D fermions in units of the AdS_5 curvature k. As usual, we adopt the zero mode approximation which identifies the standard model fields with the zero modes of corresponding 5D fields.

9.1.1 Lepton masses and mixing

The field content and the symmetry assignment are given in table 7. The zero mode of Ψ_L is the left-handed lepton doublet, and the zero modes of $\Psi_{e,\mu,\tau}$ and Ψ_ν are the right-handed charged leptons and neutrinos respectively. The left-handed lepton fields are assumed to transform as a triplet under flavour symmetry group.

Field	Ψ_l	Ψ_e	Ψ_μ	Ψ_τ	Ψ_ν	H	$\varphi_l(IR)$	$\sigma_l(IR)$	$\varphi_\nu(UV)$	$\rho_\nu(UV)$
$SU(2)_L \times SU(2)_R \times U(1)_{B-L}$	$(2,1,-1)$	$(1,2,-1)$	$(1,2,-1)$	$(1,2,-1)$	$(1,2,-1)$	$(2,2,0)$	$(1,1,0)$	$(1,1,0)$	$(1,1,0)$	$(1,1,0)$
T'	$\mathbf{3}$	$\mathbf{1'}$	$\mathbf{1''}$	$\mathbf{1}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{1''}$	$\mathbf{3}$	$\mathbf{3}$
Z_3	ω^2	1	1	1	ω^2	1	ω	ω	ω	ω
Z_4	i	i	i	i	i	1	-1	-1	i	$-i$

Table 7: Lepton and flavon fields under the $\mathrm{SU(2)_L} \times \mathrm{SU(2)_R} \times \mathrm{U(1)_{B-L}}$ gauge group and the $T' \times Z_3 \times Z_4$ family symmetry, with $\omega = e^{2\pi i/3}$. Flavons φ_l, σ_l and φ_ν, ρ_ν lie on the IR and UV branes respectively.

Note that in order to forbid dangerous terms, besides the flavour group T', we introduce the auxiliary symmetry $Z_3 \times Z_4$. Four flavons φ_l, σ_l, φ_ν and ρ_ν are introduced to break the T' family symmetry. The flavons φ_l and σ_l couple to the charged lepton sector, and are localized on the IR brane, while φ_ν and ρ_ν are localized on the UV brane. The vacuum expectation values of these flavons are aligned along the following directions

$$\langle \varphi_l \rangle = (1,0) v_{\varphi_l} \,, \quad \langle \sigma_l \rangle = v_{\sigma_l} \,, \quad \langle \varphi_\nu \rangle = (1, -2\omega^2, -2\omega) v_{\varphi_\nu} \,, \quad \langle \rho_\nu \rangle = (1, -2\omega, -2\omega^2) v_{\rho_\nu} \,, \tag{9.9}$$

with $\omega = e^{2i\pi/3}$. The above vacuum alignment can be the global minimum of the scalar potential in certain regions of parameters [79]. At leading order, the lepton mass terms respecting both gauge symmetry as well as the flavour symmetry $T' \times Z_3 \times Z_4$ take the following form

$$\mathcal{L}_Y^l = \frac{\sqrt{G}}{\Lambda'^3} \Big[y_e (\varphi_l^2 \overline{\Psi}_l)_{\mathbf{1''}} H \Psi_e + y_\mu (\varphi_l^2 \overline{\Psi}_l)_{\mathbf{1'}} H \Psi_\mu + y_\tau (\varphi_l^2 \overline{\Psi}_l)_{\mathbf{1}} H \Psi_\tau \Big] \delta(y - L) + \text{h.c.} \,, \tag{9.10}$$

$$\begin{aligned} \mathcal{L}_Y^\nu = {} & y_{\nu_1} \frac{\sqrt{G}}{\Lambda'} (\overline{\Psi}_l H \Psi_\nu)_{\mathbf{1}} \delta(y - L) + \frac{1}{2} \frac{\sqrt{G}}{\Lambda^2} \Big[y_{\nu_2} (\overline{N^C} N)_{\mathbf{1}} (\varphi_\nu^2)_{\mathbf{1}} + y_{\nu_3} (\overline{N^C} N)_{\mathbf{1}} (\rho_\nu^2)_{\mathbf{1}} \\ & + y_{\nu_4} \Big((\overline{N^C} N)_{\mathbf{3_S}} (\varphi_\nu^2)_{\mathbf{3_S}} \Big)_{\mathbf{1}} + y_{\nu_5} \Big((\overline{N^C} N)_{\mathbf{3_S}} (\rho_\nu^2)_{\mathbf{3_S}} \Big)_{\mathbf{1}} \Big] \delta(y) + \text{h.c.} \,, \end{aligned} \tag{9.11}$$

where $G = e^{-8ky}$ is the determinant of the 5D metric.

For the vacuum configuration in Eq. (9.9), the charged lepton mass matrix is diagonal and the three charged lepton masses are

$$m_e = \tilde{y}_e \frac{v_{\varphi_l}^2}{\Lambda'^2} v \,, \quad m_\mu = \tilde{y}_\mu \frac{v_{\varphi_l}^2}{\Lambda'^2} v \,, \quad m_\tau = \tilde{y}_\tau \frac{v_{\varphi_l}^2}{\Lambda'^2} v \,, \tag{9.12}$$

with

$$\tilde{y}_{e,\mu,\tau} = \frac{y_{e,\mu,\tau}}{L\Lambda'} f_L(L, c_\ell) f_R(L, c_{e,\mu,\tau}) \,. \tag{9.13}$$

Neutrino masses are generated by the type-I seesaw mechanism, and the large seesaw scale arises naturally, since the Majorana mass terms of the right-handed neutrinos are UV–localized. The first term in Eq. (9.11) leads to a diagonal Dirac neutrino mass matrix $m_D = \tilde{y}_{\nu_1} v \mathbb{1}_3$ where $\tilde{y}_{\nu_1} = \frac{y_{\nu_1}}{L\Lambda'} f_L(L, c_\ell) f_R(L, c_\nu)$ and $\mathbb{1}_3$ is the 3×3 unit matrix. The last four are the Majorana mass terms for the right-handed neutrinos, leading

to the mass matrix

$$m_N = \left(\tilde{y}_{\nu_2}\frac{v_{\varphi_\nu}^2}{\Lambda}+\tilde{y}_{\nu_3}\frac{v_{\rho_\nu}^2}{\Lambda}\right)\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}+\tilde{y}_{\nu_4}\frac{v_{\varphi_\nu}^2}{\Lambda}\begin{pmatrix}2&2\omega&2\omega^2\\2\omega&-4\omega^2&-1\\2\omega^2&-1&-4\omega\end{pmatrix}$$
$$+\tilde{y}_{\nu_5}\frac{v_{\rho_\nu}^2}{\Lambda}\begin{pmatrix}2&2\omega^2&2\omega\\2\omega^2&-4\omega&-1\\2\omega&-1&-4\omega^2\end{pmatrix}, \tag{9.14}$$

where

$$\tilde{y}_{\nu_{2,3,4,5}} = \frac{y_{\nu_{2,3,4,5}}}{L\Lambda}f_R^2(0,c_\nu)\,. \tag{9.15}$$

The light neutrino mass matrix is given by the simple type-I seesaw formula

$$m_\nu = -m_D m_N^{-1} m_D^T$$

After performing a tri-bimaximal transformation on the neutrino fields, m_ν acquires block diagonal form,

$$m'_\nu = U_{TBM}^\dagger m_\nu U_{TBM}^* = m_0\begin{pmatrix}\frac{-1}{1+3(y_4+y_5)}&0&0\\0&\frac{1-3(y_4+y_5)}{18(y_4-y_5)^2+3(y_4+y_5)-1}&\frac{3\sqrt{2}i(y_4-y_5)}{18(y_4-y_5)^2+3(y_4+y_5)-1}\\0&\frac{3\sqrt{2}i(y_4-y_5)}{18(y_4-y_5)^2+3(y_4+y_5)-1}&\frac{-1}{18(y_4-y_5)^2+3(y_4+y_5)-1}\end{pmatrix}, \tag{9.16}$$

with $m_0 = \frac{\tilde{y}_{\nu_1}^2\Lambda v^2}{\tilde{y}_{\nu_2}v_{\varphi_\nu}^2+\tilde{y}_{\nu_3}v_{\rho_\nu}^2}$, $y_4 = \frac{\tilde{y}_{\nu_4}v_{\varphi_\nu}^2}{\tilde{y}_{\nu_2}v_{\varphi_\nu}^2+\tilde{y}_{\nu_3}v_{\rho_\nu}^2}$ and $y_5 = \frac{\tilde{y}_{\nu_5}v_{\rho_\nu}^2}{\tilde{y}_{\nu_2}v_{\varphi_\nu}^2+\tilde{y}_{\nu_3}v_{\rho_\nu}^2}$.

One sees from Eq. (9.16) that the light neutrino mass matrix only depends on two complex parameters y_4 and y_5 and on the overall scale m_0. As a result one expects sharp predictions both for neutrino masses as well as mixing parameters. As shown in D, the above block-diagonal matrix is exactly diagonalized as

$$U_\nu'^\dagger m'_\nu U'^*_\nu = \mathrm{diag}(m_1,m_2,m_3), \qquad U'_\nu = \begin{pmatrix}1&0&0\\0&\cos\theta_\nu&\sin\theta_\nu e^{i\delta_\nu}\\0&-\sin\theta_\nu e^{-i\delta_\nu}&\cos\theta_\nu\end{pmatrix}, \tag{9.17}$$

The charged lepton mass matrix m_l is already in diagonal form, consequently the lepton mixing matrix is determined to be

$$U = U_{TBM}U'_\nu = \frac{1}{\sqrt{6}}\begin{pmatrix}2&\sqrt{2}\cos\theta_\nu&\sqrt{2}\sin\theta_\nu e^{i\delta_\nu}\\-1&\sqrt{2}\cos\theta_\nu-\sqrt{3}\sin\theta_\nu e^{-i\delta_\nu}&\sqrt{3}\cos\theta_\nu+\sqrt{2}\sin\theta_\nu e^{i\delta_\nu}\\-1&\sqrt{2}\cos\theta_\nu+\sqrt{3}\sin\theta_\nu e^{-i\delta_\nu}&-\sqrt{3}\cos\theta_\nu+\sqrt{2}\sin\theta_\nu e^{i\delta_\nu}\end{pmatrix}. \tag{9.18}$$

We notice that the first column of the lepton mixing matrix is fixed to be $(2,-1,-1)^T/\sqrt{6}$ which coincides with the first column of the TBM mixing pattern. In other words, the lepton mixing matrix has trimaximal TM1 form [316–318].

From Eq. (9.18) one can extract the mixing angles and the leptonic Jarlskog invariant in the usual way, to find

$$\sin^2\theta_{13} = \frac{1}{3}\sin^2\theta_\nu\,, \quad \sin^2\theta_{12} = \frac{1+\cos 2\theta_\nu}{5+\cos 2\theta_\nu}\,,$$
$$\sin^2\theta_{23} = \frac{1}{2} + \frac{\sqrt{6}\sin 2\theta_\nu\cos\delta_\nu}{5+\cos 2\theta_\nu}\,, \quad J_{CP} = -\frac{\sin 2\theta_\nu \sin\delta_\nu}{6\sqrt{6}}\,. \tag{9.19}$$

One sees from Eq. (9.19) that all the three mixing angles and the Dirac CP phase δ_{CP} are given in terms of just two free parameters δ_ν and θ_ν. The above relations imply two predicted correlations amongst the mixing angles and the Dirac CP violation phase,

$$\cos^2\theta_{12}\cos^2\theta_{13} = \frac{2}{3}\,, \qquad \cos\delta_{CP} = \frac{(3\cos 2\theta_{12}-2)\cos 2\theta_{23}}{3\sin 2\theta_{23}\sin 2\theta_{12}\sin\theta_{13}}\,. \tag{9.20}$$

that characterize the TM1 mixing pattern.

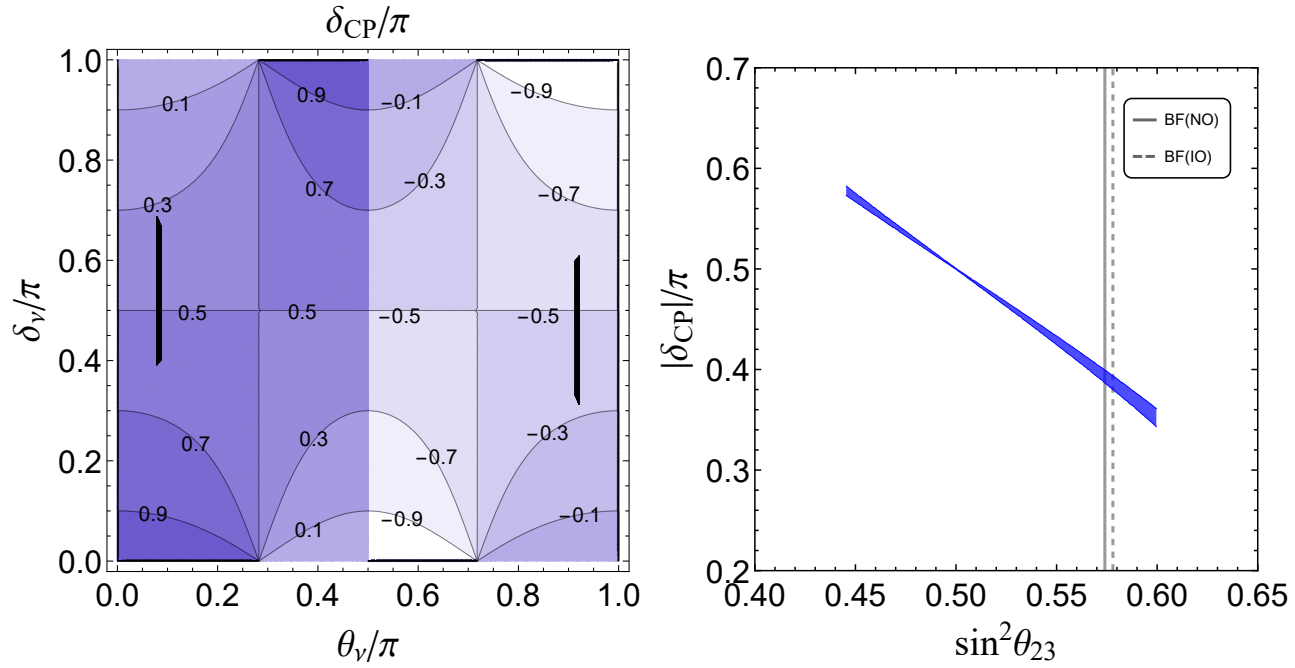


Figure 35: Contour plots of δ_{CP} in the $\theta_\nu - \delta_\nu$ plane (left) and predicted correlation between $|\delta_{CP}|$ and $\sin^2\theta_{23}$ (right). The vertical solid/dashed lines in the right panel are the best-fit $\sin^2\theta_{23}$ values for NO/IO spectra, respectively [24, 25].

In the left panel of figure 35 we display the contour plot of δ_{CP} in the plane δ_ν versus θ_ν. The small black areas in the left panel indicate the regions in which all three lepton mixing angles lie within their experimentally allowed 3σ ranges. The right panel of figure 35 shows a very tight correlation between $|\delta_{CP}|$ and the magnitude of the atmospheric angle θ_{23}. One sees that both octants are consistent, and that the CP phase parameter is predicted to lie in the restricted range $[0.325\pi, 0.592\pi]$. Upcoming long-baseline experiments will be able to test these predictions for θ_{23} and δ_{CP} [113, 114, 118].

We now turn to neutrinoless double beta decay. In figure 36 we display the expected values for the

effective Majorana neutrino mass $|m_{\beta\beta}|$ characterizing the $0\nu\beta\beta$ decay amplitude. If the neutrino mass spectrum is inverted-ordered (IO), the effective Majorana masss has a lower limit $|m_{\beta\beta}| \geq 0.0162$ eV, while the lightest neutrino mass satisfies $m_{\rm lightest} \geq 0.0133$ eV. For the case of normal-ordering (NO), the effective mass $|m_{\beta\beta}|$ lies in the narrow interval $[5.2{\rm meV}, 9.6{\rm meV}]$, and the allowed range of $m_{\rm lightest}$ is $[4.8{\rm meV}, 7.2{\rm meV}]$. Clearly, as indicated in the figure, the predicted regions for the lightest neutrino mass and $|m_{\beta\beta}|$ are quite restricted. The existing experimental bounds as well as the estimated experimental sensitivities are also indicated by the horizontal bands [125–127] in figure 36. The predicted decay amplitudes do suggest a guaranteed $0\nu\beta\beta$ discovery at the forthcoming round of experiments [107].

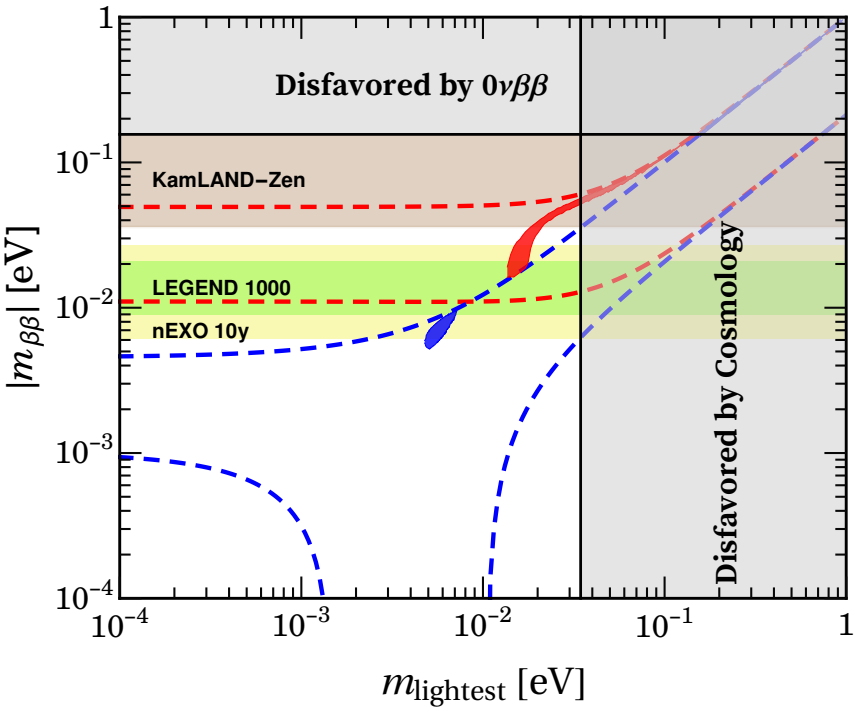


Figure 36: Predicted mass parameter characterizing the $0\nu\beta\beta$ decay amplitude; red and blue regions are for IO and NO neutrino mass spectra, respectively. Here we adopt the same convention as figure 22 for different bands and boundaries.

9.1.2 Quark masses and CKM matrix

This model can be extended to include quarks, the transformation properties of the quark fields under the family symmetry $T' \times Z_3 \times Z_4$ are listed in table 8. Note that no new scalars are needed, beyond the flavons φ_l and σ_l characterizing the lepton sector.

Field	Ψ_{UC}	Ψ_T	Ψ_u	Ψ_c	Ψ_t	Ψ_{ds}	Ψ_b	H	$\varphi_l(IR)$	$\sigma_l(IR)$
$SU(2)_L \times SU(2)_R \times U(1)_{B-L}$	$(2,1,1/3)$	$(2,1,1/3)$	$(1,2,1/3)$	$(1,2,1/3)$	$(1,2,1/3)$	$(1,2,1/3)$	$(1,2,1/3)$	$(2,2,0)$	$(1,1,0)$	$(1,1,0)$
T'	$\mathbf{2}$	$\mathbf{1}$	$\mathbf{1'}$	$\mathbf{1''}$	$\mathbf{1'}$	$\mathbf{2'}$	$\mathbf{1''}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{1''}$
Z_3	ω^2	ω	1	ω^2	1	ω	ω^2	1	ω	ω
Z_4	1	-1	1	-1	-1	1	-1	1	-1	-1

Table 8: Classification of the quark fields under the bulk gauge group $\mathrm{SU(2)_L \times SU(2)_R \times U(1)_{B-L}}$ and the flavour symmetry $T' \times Z_3 \times Z_4$.

The quark Yukawa interactions are localized on the IR brane and constrained by the T' flavour symmetry to be of the following form,

$$\begin{aligned}
\mathcal{L}_Y^d &= \frac{\sqrt{G}}{\Lambda'^3}\Big[y_{ds_1}(\overline{\Psi}_{UC}H\Psi_{ds})_{\mathbf{3}}\varphi_l^{*\,2} + y_{ds_2}(\overline{\Psi}_{UC}H\Psi_{ds})_{\mathbf{1}'}\sigma_l^{*\,2} + y'_b(\overline{\Psi}_T H\Psi_b)_{\mathbf{1}''}\sigma_l^2\Big]\delta(y-L) + \text{h.c.} + \cdots \\
\mathcal{L}_Y^u &= \frac{\sqrt{G}}{\Lambda'^3}\Big[y'_u\Lambda'(\overline{\Psi}_T H\Psi_u)_{\mathbf{1}'}\sigma_l + y_t\Lambda'(\overline{\Psi}_{UC}H\Psi_t)_{\mathbf{2}}\varphi_l^* + y_u(\overline{\Psi}_{UC}H\Psi_u)_{\mathbf{2}}\varphi_l\sigma_l \\
&\qquad + y'_c(\overline{\Psi}_T H\Psi_c)_{\mathbf{1}''}\sigma_l^2 + y'_t(\overline{\Psi}_T H\Psi_t)_{\mathbf{1}'}\sigma_l^{*2}\Big]\delta(y-L) + \text{h.c.} + \cdots\,,
\end{aligned} \tag{9.21}$$

for the down-type and up-type quark masses respectively. One can read out the mass matrices for the zero modes of the quark fields as

$$m^d = v\begin{pmatrix} \tilde{y}_{ds_2}v_{\sigma_l}^{*2}/\Lambda'^2 & 0 & 0 \\ \tilde{y}_{ds_1}v_{\varphi_l}^{*2}/\Lambda'^2 & \tilde{y}_{ds_2}v_{\sigma_l}^{*2}/\Lambda'^2 & 0 \\ 0 & 0 & \tilde{y}'_b v_{\sigma_l}^2/\Lambda'^2 \end{pmatrix}, \tag{9.22}$$

$$m^u = v\begin{pmatrix} \tilde{y}_u v_{\varphi_l}v_{\sigma_l}/\Lambda'^2 & 0 & 0 \\ 0 & 0 & \tilde{y}_t v_{\varphi_l}^*/\Lambda' \\ \tilde{y}'_u v_{\sigma_l}/\Lambda' & \tilde{y}'_c v_{\sigma_l}^2/\Lambda'^2 & \tilde{y}'_t v_{\sigma_l}^{*2}/\Lambda'^2 \end{pmatrix}. \tag{9.23}$$

where the parameters with tilde are given by

$$\tilde{y}_{u,t,ds_{1,2}} = \frac{y_{u,t,ds_{1,2}}}{L\Lambda'}f_L(L,c_{UC})f_R(L,c_{u,t,ds})\,, \quad \tilde{y}'_{u,c,t,b} = \frac{y'_{u,c,t,b}}{L\Lambda'}f_L(L,c_T)f_R(L,c_{u,c,t,b})\,. \tag{9.24}$$

Notice that the down-quark mass matrix is block diagonal and the (11) and (22) entries are exactly equal. Note also that the up-quark sector gives a negligible contribution to the Cabibbo angle θ_c but is responsible for generating V_{ub} and V_{cb}. As a result this model gives rise to the well-known Gatto-Sartori relation $m_d/m_s \simeq \tan^2\theta_c$ [469].

9.2 Global flavour fit

Let us now perform a global fit of the masses and flavour mixing parameters of both quarks and leptons within this model. The fundamental 5D scales on the UV and IR branes are taken to be $\Lambda \simeq k \simeq 2.44\times10^{18}$ GeV and $\Lambda' = ke^{-kL} \simeq 1.5$ TeV respectively. The vacuum expectation value of the Higgs field is fixed to its SM value $v \simeq 174$ GeV, and we choose the flavon VEVs as $v_{\varphi_l}/\Lambda' = v_{\sigma_l}/\Lambda' = v_{\varphi_\nu}/\Lambda = v_{\rho_\nu}/\Lambda = 0.2$. In what follows we give a typical set of values for the free parameters. The bulk mass parameters and the Yukawa coupling constants of the charged lepton and quarks are given by

$$\begin{aligned}
&c_l = 0.460, \quad c_e = -0.725, \quad c_\mu = -0.553, \quad c_\tau = -0.117, \\
&c_{UC} = 0.587, \quad c_T = -0.980, \quad c_u = -0.516, \quad c_c = -0.555, \\
&c_t = 0.966, \quad c_{ds} = -0.503, \quad c_b = -0.532, \\
&y_e = y_\mu = y_\tau = 1.0, \quad y_u = 6.321, \quad y_t = 6.20, \quad y'_u = 4.00,
\end{aligned}$$

$$y_c' = 1.00, \quad y_t' = 8.30, \quad y_{ds1} = 4.00, \quad y_{ds2} = 0.892, \quad y_b' = 4.00\,. \tag{9.25}$$

The values of the parameters in the neutrino sector depend on the neutrino mass ordering,

$$\begin{aligned} &\text{NO}: c_\nu = -0.404, \quad y_{\nu 1} = y_{\nu 2} = y_{\nu 3} = 1, \quad y_{\nu 4} = 0.235 + 0.0770i, \quad y_{\nu 5} = 0.340 + 0.0710i, \\ &\text{IO}: c_\nu = -0.383, \quad y_{\nu 1} = y_{\nu 2} = y_{\nu 3} = 1, \quad y_{\nu 4} = -0.354 + 0.275i, \quad y_{\nu 5} = -0.562 + 0.270i\,. \end{aligned} \tag{9.26}$$

The resulting predictions for flavour observables such as lepton and quark mass and mixing parameters are all listed in table 9. One sees that all Standard Model fermion masses and mixings can be very well reproduced.

parameters	best-fit $\pm$ 1σ	predictions
$\sin\theta_{12}^q$	0.22500±0.00100	0.22503
$\sin\theta_{13}^q$	0.003675±0.000095	0.003668
$\sin\theta_{23}^q$	0.04200±0.00059	0.04205
$\delta_{CP}^q/^\circ$	66.9±2	68.2
m_u [MeV]	$2.16^{+0.49}_{-0.26}$	2.16
m_c [GeV]	1.27±0.02	1.27
m_t [GeV]	172.9±0.4	172.90
m_d [MeV]	$4.67^{+0.48}_{-0.17}$	4.21
m_s [MeV]	93^{+11}_{-5}	93.00
m_b [GeV]	$4.18^{+0.03}_{-0.02}$	4.18
$\sin^2\theta_{12}^l/10^{-1}$ (NO)	$3.20^{+0.20}_{-0.16}$	3.19
$\sin^2\theta_{12}^l/10^{-1}$ (IO)		3.18
$\sin^2\theta_{23}^l/10^{-1}$ (NO)	$5.47^{+0.20}_{-0.30}$	5.47
$\sin^2\theta_{23}^l/10^{-1}$ (IO)	$5.51^{+0.18}_{-0.30}$	5.51
$\sin^2\theta_{13}^l/10^{-2}$ (NO)	$2.160^{+0.083}_{-0.069}$	2.160
$\sin^2\theta_{13}^l/10^{-2}$ (IO)	$2.220^{+0.074}_{-0.076}$	2.220
δ_{CP}^l/π (NO)	$1.32^{+0.21}_{-0.15}$	1.567
δ_{CP}^l/π (IO)	$1.56^{+0.13}_{-0.15}$	1.571
m_e [MeV]	$0.511 \pm 3.1 \times 10^{-9}$	0.511
m_μ [MeV]	$105.658 \pm 2.4 \times 10^{-6}$	105.658
m_τ [MeV]	1776.86 ± 0.12	1776.86
Δm_{21}^2 $[10^{-5}\text{eV}^2]$ (NO)	$7.55^{+0.20}_{-0.16}$	7.55
Δm_{21}^2 $[10^{-5}\text{eV}^2]$ (IO)		
$\lvert\Delta m_{31}^2\rvert$ $[10^{-3}\text{eV}^2]$ (NO)	2.50±0.03	2.50
$\lvert\Delta m_{31}^2\rvert$ $[10^{-3}\text{eV}^2]$ (IO)	$2.42^{+0.03}_{-0.04}$	2.42
χ^2 (NO)	–	7.65
χ^2 (IO)		7.66

Table 9: Global warped flavordynamics fit: neutrino oscillation parameters are taken from the global analysis in [24, 25], while the quark parameters are taken from the Review of Particle Physics [28].

All in all the model provides a consistent scenario for the flavour problem, in which fermion mass hierarchies are accounted for by adequate choices of the bulk mass parameters, while quark and lepton mixing angles are restricted by the assumed T' flavour symmetry. Note that in this model neutrinos are

Majorana particles, the tiny neutrino masses are generated by the type-I seesaw mechanism with "right-handed" neutrino masses in the range of $[10^{12}, 10^{13}]$ GeV and relatively sizeable rates for $0\nu\beta\beta$ decay, accessible within the next round of experiments. For an alternative warped flavourdynamics construction along the similar lines, see Ref. [78]. In that case the model uses the $\Delta(27)$ family symmetry, neutrinos are Dirac fermions, and the predicted neutrino mixing pattern is TM2.

10 Family symmetry from 6-D orbifolds

Underpinning the nature of the underlying family symmetry of particle physics amongst the huge plethora of possibilities constitutes a formidable task. As already seen in the previous section, a promising approach to the flavour problem is to imagine the existence of new dimensions in spacetime. In the present section we illustrate how the existence of extra dimensions may shed light on the nature of the family symmetry in four dimensions. The idea was suggested within six-dimensional setups compactified on a torus [292,293].

The resulting theories share a realistic A_4 family symmetry, featuring the "golden" quark-lepton mass formula

$$\frac{m_\tau}{\sqrt{m_\mu m_e}} \approx \frac{m_b}{\sqrt{m_s m_d}}\,. \tag{10.1}$$

This formula was proposed in [288], and emerges also in other 4-D flavour schemes such as those in Refs. [289, 290] and [65, 66], as well as in implementations of the Peccei-Quinn symmetry [287]. It is remarkable that it also comes out neatly in scenarios where the family symmetry arises from the compactification of 6-dimensional orbifolds, as proposed in Refs. [143, 144, 294, 295] and considered next. Fermions are nicely arranged in terms of the A_4 family symmetry. Different setups can be identified, with very interesting phenomenology. Indeed, they bring in the possibility of predicting neutrino mixing angles and CP phases, as well as providing a good global description of flavour observables.

10.1 General preliminaries

Here we summarize the theory framework and survey its main features. First of all we have a 6-dimensional version of the standard $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge symmetry, supplemented with orbifold compactification, as outlined below. In the full six-dimensional theory, the spacetime manifold $\mathcal{M}$ is identified as the direct product $\mathcal{M} = \mathbb{M}^4 \times (T^2/\mathbb{Z}_2)$, where $\mathbb{M}^4$ is the four-dimensional Minkowski spacetime, and $T^2/\mathbb{Z}_2$ is a one-parameter family (given by θ) of 2-D toroidal orbifolds defined by the following relations satisfied by the extra-dimensional coordinates

$$\left(x^5, x^6\right) = \left(x^5 + 2\pi R_1, x^6\right), \tag{10.2}$$

$$\left(x^5, x^6\right) = \left(x^5 + 2\pi R_2 \cos\theta, x^6 + 2\pi R_2 \sin\theta\right). \tag{10.3}$$

$$\left(x^5, x^6\right) = \left(-x^5, -x^6\right), \tag{10.4}$$

The first two equations define a torus, with θ describing its twist angle, and the third equation defines the $\mathbb{Z}_2$ orbifolding. For simplicity we assume that the characteristic radii of the compact extra dimensions are similar, i.e.

$$R_1 \sim R_2 \sim 1/M_c, \tag{10.5}$$

in terms of the compactification scale M_c. Moreover, the twist angle is assumed to be $\theta = 2\pi/3$. To simplify the analysis we define the scaled complex coordinate $z = M_c(x_5 + ix_6)/(2\pi)$ and rewrite Eqs. (10.2)-(10.4) as

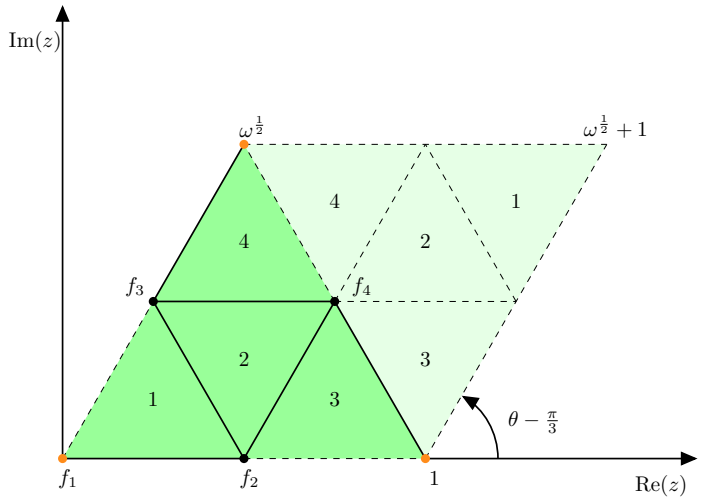


Figure 37: The fundamental domain of the $T^2/\mathbb{Z}_2$ orbifold is the darkest region, obtained after the compactification of the corresponding domain of the twisted torus, which includes the lightest region. The resulting space is reminiscent of a tetrahedron, and can be visualized by identifying the three orange dots into a single vertex. The fixed points of the orbifold are located at the vertices of the tetrahedron.

$$z = z + 1, \tag{10.6}$$

$$z = z + \omega, \tag{10.7}$$

$$z = -z, \tag{10.8}$$

where ω is the cubic root of unity

$$\omega \equiv e^{i\theta} = e^{i2\pi/3}. \tag{10.9}$$

A key feature of orbifolds is that they have singular points. In our case there are four of these, located at the points that remain fixed by the transformations in Eqs. (10.6)-(10.8), namely

$$f_1 = 0, \quad f_2 = \frac{1}{2}, \quad f_3 = \frac{\omega}{2}, \quad f_4 = \frac{1+\omega}{2}. \tag{10.10}$$

These fixed points define the location of 4-dimensional branes embedded in the 6-dimensional space $\mathcal{M}$. In figure 37 we display both the fundamental domain of the twisted torus T^2 (light shaded green), as well as the fundamental domain of the $T^2/\mathbb{Z}_2$ orbifold (dark shaded green). After compactification, the continuous Poincaré symmetry of the two extra dimensions is broken, leaving a residual A_4 symmetry of the branes [314]. The appearance of the discrete A_4 symmetry can be understood as the invariance under permutations displayed by the four fixed points of the orbifold. Any of these can be written in terms of two independent transformations

$$S : z \longrightarrow z + 1/2, \quad T : z \longrightarrow \omega^2 z. \tag{10.11}$$

These can be also expressed as elements of the permutation group S_4.

$$S = (12)(34), \quad T = (1)(243), \tag{10.12}$$

This way S and T are related to the generators of the A_4 group, satisfying

$$S^2 = T^3 = (ST)^3 = 1\,, \tag{10.13}$$

which are exactly the multiplication rules of the A_4 group in Eq. (A.1).

The model is based on this remnant A_4 as a family symmetry. Same-charge fields located on the four different branes would transform into each other by the remnant A_4 transformations. These four branes transform as the reducible representation **4**, which decomposes as a sum of irreducible representations $\mathbf{4} \to \mathbf{3} + \mathbf{1}$. Thus, the brane-localized fields must transform under the flavour group as A_4 triplets or singlets, so the family symmetry is spontaneously broken. Below we show how this can provide a realistic pattern for the three families of leptons and quarks in a rather predictive and economical way.

Notice that the assumption of extra dimensions implies the existence of infinitely many 4-D fields associated with every bulk field, called a Kaluza-Klein (KK) tower. Their masses $(n^2 + m^2)M_c$ are determined by positive integers n, m. In our case, the fields in the bulk are the SM gauge fields g_μ, W_μ, B_μ, the right-handed quarks u_i^c and the gauge singlet scalar σ.

The tower of massive KK modes from the vector $\mathrm{SU(2)_L}$ triplets can affect the Peskin-Takeuchi oblique parameters S, T and U in an important way. The current experimental bound for our setup (2 non-universal extra dimensions) is [470, 471]

$$M_c > 2.1 \text{ TeV}.$$

For a compactification scale sufficiently close to 2 TeV, the electroweak precision tests could in principle probe the extra dimensions at the High Luminosity LHC run.

10.2 Scotogenic orbifold

Our basic setup is a 6-dimensional extension of the Standard $\mathrm{SU(3)_c} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ Model, featuring the orbifold compactification described in the previous section, and inheriting the A_4 discrete family symmetry in a natural manner. Its simplest model-realization includes three right-handed neutrinos, mediating neutrino mass generation through the type-I seesaw mechanism [143, 144]. Instead of pursuing such an approach, however, here we focus on a more complete, yet equally economical, scotogenic variant that also provides a WIMP dark matter candidate [294, 295].

The field content and transformation properties of our benchmark scotogenic variant under the various symmetry groups are shown in table 10. Note that all fermionic fields, except for the right-handed quarks, transform as flavour triplets and are localized in the orbifold branes.

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$	$\mathbb{Z}_4$	A_4	Location
L	**1**	**2**	-1	1	**3**	Brane
d^c	**3**	**1**	$2/3$	1	**3**	Brane
e^c	**1**	**1**	2	1	**3**	Brane
Q	**3**	**2**	$1/3$	1	**3**	Brane
$u^c_{1,2,3}$	**3**	**1**	$-4/3$	-1	$\mathbf{1''}, \mathbf{1'}, \mathbf{1}$	Bulk
F	**1**	**1**	0	i	**3**	Brane
H_u	**1**	**2**	1	-1	**3**	Brane
H_d	**1**	**2**	-1	1	**3**	Brane
η	**1**	**2**	1	$-i$	**1**	Brane
σ	**1**	**1**	0	-1	**3**	Bulk

Table 10: Field representation content and symmetries of the scotogenic orbifold model [295].

The model assumes an auxiliary lepton quarticity symmetry [165, 167]. This $\mathbb{Z}_4$ is spontaneously broken to a residual $\mathbb{Z}_2$ symmetry, defining the "dark sector" which comprises the dark fermion F and scalar η. Both transform non-trivially under the "dark symmetry" ensuring stability of the lightest $\mathbb{Z}_2$-charged field. This makes it a potentially viable dark matter candidate, whose stability is directly related to the radiative origin of neutrino masses, Fig. 38.

The Higgs sector consists of two flavour-triplet weak iso-doublets, H_u and H_d and a SM singlet scalar σ driving the spontaneous breaking of both lepton number [32, 39], as well as family symmetry, a sort of "flavoured" Majoron scheme.

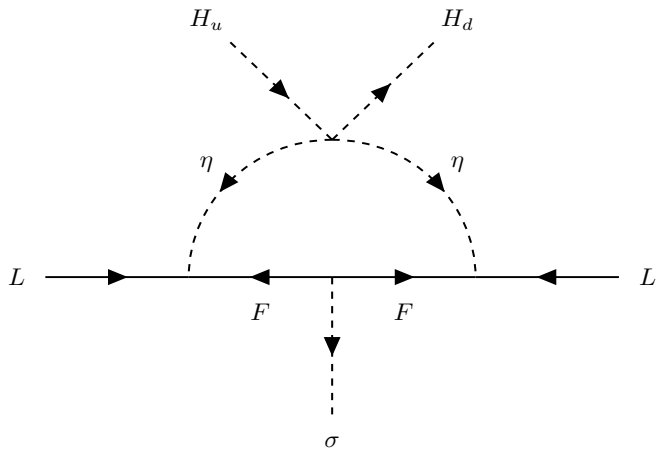


Figure 38: One-loop diagram for Majorana neutrino masses, mediated by the "dark sector" particles [295].

Given the defining symmetries of the model, one can write the most general effective Yukawa Lagrangian below the compactification scale. The Yukawa interaction terms of down-type quarks and charged leptons have the same structure, given by

$$\mathcal{L}^{\text{Yukawa}}_{H_d} = y^e_1 \left(L H_d e^c\right)_{\mathbf{1}_1} + y^e_2 \left(L H_d e^c\right)_{\mathbf{1}_2} + y^d_1 \left(Q H_d d^c\right)_{\mathbf{1}_1} + y^d_2 \left(Q H_d d^c\right)_{\mathbf{1}_2} + \text{H.c.}, \tag{10.14}$$

while the transformation properties of the up-type quark fields under A_4 yield

$$\mathcal{L}_{H_u}^{\text{Yukawa}} = y_1^u \left(QH_u\right)_{\mathbf{1'}} u_1^c + y_2^u \left(QH_u\right)_{\mathbf{1''}} u_2^c + y_3^u \left(QH_u\right)_{\mathbf{1}} u_3^c + \text{H.c.} \tag{10.15}$$

Notice that the bold subscripts in each term indicate its transformation properties under the remnant A_4 family symmetry. The explicit expressions for the invariant multiplet products are given in A.

The dark fermion triplet F couples to the scalar field σ. The latter acquires a vacuum expectation value which drives the spontaneous breaking of lepton number symmetry, $\mathbb{Z}_4$, and the A_4 family symmetry, giving rise to Majorana mass terms for the dark fermions,

$$\mathcal{L}_{\sigma}^{\text{Yukawa}} = y^{\sigma} \left(F^T F \sigma\right)_{\mathbf{1}_1} + \text{H.c.} \tag{10.16}$$

The dark scalar η plays a crucial role in the model, as it couples with both dark fermions and neutrinos

$$\mathcal{L}_{\eta}^{\text{Yukawa}} = y_1^{\eta} \left(L\eta F\right)_{\mathbf{1}} + \text{H.c.} \tag{10.17}$$

In the following, we will assume that all Yukawa couplings are real, and therefore that the model preserves a trivial CP symmetry.

10.3 Symmetry breaking and fermion masses

The scalar potential $V(H_d, H_u, \eta, \sigma)$ comprises all terms up to quartic interactions consistent with all the symmetries. Apart from the Higgs scalars H_d, H_u, σ it contains also the dark scalar η which does not develop a vacuum expectation value (VEV). The symmetry breakdown of the model proceeds in two steps. At high energies the electroweak singlet scalar field σ develops a VEV compatible with the extra-dimensional boundary conditions. Subsequently, at lower energies, the electroweak Higgs doublets H_u, H_d acquire VEVs according to the minimization of the scalar potential.

In order to describe the high-scale A_4 symmetry breaking produced by σ we introduce a boundary condition P, consistent with the orbifold construction. It defines a non-trivial gauge/Poincaré twist of the orbifold, and must be a symmetry of the Lagrangian. We assume that the transformation P acts trivially on the A_4 singlet bulk fields. Hence the only bulk field transforming non-trivially under P is the flavour triplet scalar σ, obeying the boundary condition

$$\sigma(x, z) = P\sigma(x, -z). \tag{10.18}$$

so as to be consistent with Eq. (10.8). The invariance of the kinetic term of the σ field in the 6-D Lagrangian implies that $P \in SU(3)$, while the condition in Eq. (10.18) ensures that the matrix P will leave invariant the interactions of fields in the brane. Thus the boundary condition matrix must satisfy

$$P \in SU(3), \qquad P^2 = \mathbf{1}_{3\times 3}, \quad P^{\dagger} = P. \tag{10.19}$$

The boundary condition on the σ field applies also to its VEV alignment. As a result, the masses of the dark fermions F are a direct outcome of the boundary condition of the bulk field σ in the two extra dimensions

$$\langle \sigma \rangle = P \langle \sigma \rangle \,. \tag{10.20}$$

The boundary condition matrix P is a property of the orbifold, its form is given explicitly in [295]. Here we adopt the most general VEV alignment consistent with the spontaneous breaking of lepton number and the A_4 family symmetry, expressed as

$$\langle \sigma \rangle = v_\sigma \begin{pmatrix} \epsilon_1^\sigma \, e^{i\varphi} \\ \epsilon_2^\sigma \\ 1 \end{pmatrix}, \qquad \text{with} \qquad v_\sigma, \epsilon_1^\sigma, \epsilon_2^\sigma \in \mathbb{R} \quad \text{and} \quad 0 \leq \varphi < \pi. \tag{10.21}$$

Here we will not present a dedicated analysis of the scalar potential, except to stress the importance of the λ_5 term for the neutrino mass generation mechanism, namely

$$V(H_u, H_d, \eta) \supset \frac{1}{2}\lambda_5 \left[\left({H_d}^T (i\sigma_2)\, \eta \right)_{\mathbf{3}} \left(H_u^\dagger \eta \right)_{\mathbf{3}} \right]_{\mathbf{1}} + \text{H.c.} \tag{10.22}$$

Here λ_5 is a coupling constant and σ_2 is the second Pauli matrix. This term lifts the degeneracy of the mass eigenstates of the neutral components of η, denoted as $\sqrt{2}\,\text{Re}(\eta^0)$ and $\sqrt{2}\,\text{Im}(\eta^0)$, playing a key role in the scotogenic generation of neutrino masses at the one-loop level, as illustrated in Fig. 38. One can show that there is enough freedom in parameter space to drive the spontaneous breaking of the gauge symmetries down to $\text{U}(1)_{\text{EM}}$.

10.3.1 Quark and lepton masses

Since A_4 breaks spontaneously at the v_σ scale, in the second stage of spontaneous symmetry breaking, we assume that the weak iso-doublets H_u and H_d obtain the most general A_4 breaking VEVs consistent with trivial CP symmetry. For real $v_u, v_d, \epsilon_{1,2}^{u,d}$ it is given as

$$\langle H_u \rangle = v_u \begin{pmatrix} \epsilon_1^u \\ \epsilon_2^u \\ 1 \end{pmatrix}, \quad \langle H_d \rangle = v_d \begin{pmatrix} \epsilon_1^d \\ \epsilon_2^d \\ 1 \end{pmatrix}. \tag{10.23}$$

After spontaneous symmetry breaking, the up-quark mass matrix becomes

$$M_u = v_u \begin{pmatrix} y_1^u \epsilon_1^u & y_2^u \epsilon_1^u & y_3^u \epsilon_1^u \\ y_1^u \epsilon_2^u \omega^2 & y_2^u \epsilon_2^u \omega & y_3^u \epsilon_2^u \\ y_1^u \omega & y_2^u \omega^2 & y_3^u \end{pmatrix}, \tag{10.24}$$

while the down-quark and charged lepton mass matrices take the form

$$M_d = v_d \begin{pmatrix} 0 & y_1^d \epsilon_1^d & y_2^d \epsilon_2^d \\ y_2^d \epsilon_1^d & 0 & y_1^d \\ y_1^d \epsilon_2^d & y_2^d & 0 \end{pmatrix}, \qquad M_e = v_d \begin{pmatrix} 0 & y_1^e \epsilon_1^d & y_2^e \epsilon_2^d \\ y_2^e \epsilon_1^d & 0 & y_1^e \\ y_1^e \epsilon_2^d & y_2^e & 0 \end{pmatrix}. \tag{10.25}$$

All Yukawa couplings in the last two equations are assumed to be real due to our imposition of trivial CP symmetry.

10.3.2 Scotogenic neutrino masses

The A_4 flavour symmetry structure of the Yukawa term in Eq. (10.16) implies that the Majorana mass matrix of the dark fermions M_F must have the following structure

$$M_F = y_\sigma v_\sigma \begin{pmatrix} 0 & 1 & \epsilon_2^\sigma \\ 1 & 0 & \epsilon_1^\sigma e^{i\varphi} \\ \epsilon_2^\sigma & \epsilon_1^\sigma e^{i\varphi} & 0 \end{pmatrix}. \tag{10.26}$$

In order to describe our one-loop scotogenic mechanism for neutrino masses we write the dark fermion F fields in the mass eigenstate basis ($\tilde{F}$) by performing the singular value decomposition of the dark fermion mass matrix M_F. Since the latter is symmetric, only one unitary matrix V is needed in the Takagi decomposition [30],

$$y^\sigma \left(F^T F \sigma\right)_{\mathbf{1}_1} = F^T M_F F = F^T V^T D V F = (VF)^T D (VF) \equiv \tilde{F}^T D \tilde{F}, \tag{10.27}$$

where $D = \mathrm{diag}(m_{F_1}, m_{F_2}, m_{F_3})$ and $\tilde{F} \equiv VF$ denotes the dark fermion triplet expressed in the mass eigenstate basis. We can then rewrite Eq. (10.17) as

$$\mathcal{L}_\eta^{\text{Yukawa}} = y_1^\eta \eta \left(L V^\dagger \tilde{F}\right) + \text{H.c.} \tag{10.28}$$

As already mentioned, neutrino masses are forbidden at tree-level due to the auxiliary $\mathbb{Z}_4$ symmetry. However, thanks to the mediation of the dark fields η and F, neutrino masses emerge at one-loop through the diagram depicted in Fig. 38, which has the basic scotogenic structure [196].

Defining $y_1^\eta V^\dagger \equiv h$ in Eq. (10.27) we can write the expression for the one-loop neutrino mass matrix M_ν as

$$(M_\nu)_{ij} = \sum_k^3 \frac{h_{ik}(h^T)_{kj}}{16\pi^2} S(m_{F_k}), \tag{10.29}$$

where $S(m_{F_k})$ is for the loop factor

$$S(m_{F_k}) = m_{F_k}\left(\frac{m_R^2}{m_R^2 - m_{F_k}^2}\ln\frac{m_R^2}{m_{F_k}^2} - \frac{m_I^2}{m_I^2 - m_{F_k}^2}\ln\frac{m_I^2}{m_{F_k}^2}\right), \tag{10.30}$$

with $m_R = m(\sqrt{2}\,\mathrm{Re}\,\eta^0)$, $m_I = m(\sqrt{2}\,\mathrm{Im}\,\eta^0)$ with

$$m_R^2 - m_I^2 \equiv 2\lambda_5\left(\langle H_u\rangle_{\mathbf{3}}\,\langle H_d\rangle_{\mathbf{3}}\right)_{\mathbf{1}}. \tag{10.31}$$

Neutrino masses are not only loop-suppressed, but also symmetry-protected, as they vanish in the limit $\lambda_5 \to 0$, see Eq. (10.22).

After spontaneous symmetry breaking, the auxiliary $\mathbb{Z}_4$ breaks down to a residual $\mathbb{Z}_2$ that stabilizes the lightest dark particle. There are two possible dark matter candidates: the lightest state in the complex neutral scalar η, or the lightest Majorana fermion in the flavour triplet F. In either case, the phenomenology of dark matter is qualitatively similar to that of other scotogenic scenarios, which has been extensively discussed, see for example Ref. [138] and references therein.

10.4 Global fit of flavour observables

Before describing the global fit of flavour observables, some preliminaries are necessary. The bottom line is that, due to the reduced number of parameters, the model makes strong flavour predictions.

10.4.1 Preliminaries

Here we adopt the symmetrical parametrization of the quark and lepton mixing matrices [30], described in Sect. 1. For the case of quarks, choosing the PDG ordering prescription, the symmetrical form leads to the standard Cabibbo-Kobayashi-Maskawa (CKM) matrix in Eq. (1.8). On the other hand, for the mixing of leptons, it is described by Eq. (1.21) [24]. In both cases the mixing matrix description is supplemented by the PDG factor ordering convention. However, in the leptonic case the symmetrical form provides a neater description of CP violation than the PDG form. In this case the Dirac CP phase that enters in neutrino oscillations is the leptonic analogue of the quark Jarlskog parameter, identified with the "rephasing-invariant" combination

$$\delta^\ell = \phi_{13} - \phi_{12} - \phi_{23}.$$

Notice that this phase must not be present in the effective mass parameter $\langle m_{\beta\beta}\rangle$ characterizing the amplitude for neutrinoless double beta decay, which involves only the two Majorana phases, Eq. (1.25). Hence the symmetrical presentation is more transparent for describing the $0\nu\beta\beta$ decay amplitude [99].

[24] In both cases we assume unitarity, neglecting therefore possible mixing with exotic fermions that could be relevant, say, within a low-scale seesaw scheme [472].

10.4.2 Fit procedure

The above model is characterized by 16 independent parameters in the flavour sector, identified as follows: 8 real Yukawa couplings $y^{e,d}_{1,2}$, $y^{u}_{1,2,3}$, y^{η}_{1}, 6 real VEV ratios $\epsilon^{u,d}_{1,2}$, $\epsilon^{\sigma}_{1,2}$, one quartic coupling λ_5 and one CP violating phase φ contained in $\langle\sigma\rangle$. These parameters describe 22 observables, namely, 12 masses, 4 CKM matrix elements, plus 6 lepton mixing matrix parameters including the 2 Majorana phases.

In order to explore the predictivity of the model, we perform a global flavour fit to the available experimental data, by minimizing the chi-square function defined as

$$\chi^2 = \sum (\mu_{\rm exp} - \mu_{\rm model})^2/\sigma^2_{\rm exp}, \tag{10.32}$$

where the sum runs through the 19 measured physical parameters, i.e. 6 quark masses, 3 CKM mixing angles, 1 CKM CP phase, 3 charged lepton masses, 3 lepton mixing angles, 1 lepton CP violating phase and 2 neutrino squared mass splittings. Note that we have only limits on the lightest neutrino mass from experiment and no direct information on the Majorana $0\nu\beta\beta$ phases. The fit is performed by scanning the values of the 16 independent model parameters that provide a description of the above 22 flavour observables.

In our chi-square minimization with respect to the 16 independent model parameters, all quark and charged-lepton masses were evaluated at the same energy scale, chosen as M_Z [473]. This assumption was shown to be consistent with the golden mass relation [143]. Coming to the neutrino oscillation parameters, these were extracted from the global fit in [24,25], neglecting the effect of running to M_Z [473,474]. The remaining observables were taken from the PDG [28]. In order to extract the flavour observables from the mass matrices in Eqs. (10.24) and (10.25) we use the Mathematica Mixing Parameter Tools package [475].

The results of our flavour fit are summarized in table 11. The minimum at $\chi^2 \approx 2$ shows that the model reproduces the observed pattern of fermion masses and mixing rather well. From Table 11 one can read directly the predictions of the model concerning the mass of the lightest neutrino and the values of the CP phases characterizing the lepton sector.

10.5 Flavour predictions of the scotogenic orbifold model

We now discuss in more detail the flavour predictions of our scotogenic orbifold model, both in the quark and lepton sectors. These follow directly from the A_4 family symmetry that results from the orbifold compactification of the extra dimensions.

10.5.1 Golden quark-lepton mass relation

This is a key feature of the model that results from the down-type quarks and charged lepton assignments under the A_4 flavour symmetry. Indeed, after spontaneous symmetry breaking these obtain masses from the same common Higgs doublet H_d, leading to the mass matrices in Eq.(10.25). After diagonalization

to physical mass-eigenstates, one obtains the golden quark-lepton mass relation

$$\frac{m_\tau}{\sqrt{m_\mu m_e}} \approx \frac{m_b}{\sqrt{m_s m_d}}, \tag{10.33}$$

This relation emerges in models with SO(3) family symmetry implementing a Peccei-Quinn symmetry [287], and in the theories proposed in Refs. [65, 66] and [288–290].

In our case the golden relation is a common feature of the models with A_4 family symmetry arising from the compactification of 6-Dimensional orbifolds, proposed in Refs. [143, 144] and further discussed in [294, 295]. One can show that, given the current experimental measurements of the relevant masses, the golden relation holds with good precision, see for example, Fig.1 in [143]. Besides, it constitutes a very robust prediction under the renormalization group evolution, as it involves only fermion mass ratios.

Parameter	Value
$y_1^e v_d/\text{GeV}$	-1.745
$y_2^e v_d/(10^{-1}\text{GeV})$	1.021
$y_1^d v_d/(10^{-2}\text{GeV})$	-5.039
$y_2^d v_d/\text{GeV}$	2.852
$y_1^u v_u/(10^{-1}\text{GeV})$	6.074
$y_2^u v_u/(10^{2}\text{GeV})$	1.712
$y_3^u v_u/\text{GeV}$	7.157
$\epsilon_1^u/10^{-4}$	7.055
$\epsilon_2^u/10^{-2}$	-5.044
$\epsilon_1^d/10^{-3}$	-2.814
$\epsilon_2^d/10^{-3}$	5.833
ϵ_1^σ	1.501
ϵ_2^σ	-0.654
φ	3.527
$(y_1^\eta)^2 y_\sigma v_\sigma/(\text{KeV})$	1.813
$2\lambda_5 \langle H_u\rangle \langle H_d\rangle/(\text{KeV})^2$	0.012

Observable	Data		Model best fit
	Central value	1σ range	
θ_{12}^ℓ /°	34.3	$33.3 \to 35.3$	33.0
θ_{13}^ℓ /°	8.45	$8.31 \to 8.61$	8.52
θ_{23}^ℓ /°	49.26	$48.47 \to 50.05$	50.44
δ^ℓ /°	194	$172 \to 218$	192
m_e /MeV	0.486	$0.486 \to 0.486$	0.486
m_μ /GeV	0.102	$0.102 \to 0.102$	0.102
m_τ /GeV	1.745	$1.743 \to 1.747$	1.745
$\Delta m^2_{21}/(10^{-5}\,\text{eV}^2)$	7.50	$7.30 \to 7.72$	7.50
$\Delta m^2_{31}/(10^{-3}\,\text{eV}^2)$	2.55	$2.52 \to 2.57$	2.54
m_1 /meV			135.35
m_2 /meV			135.63
m_3 /meV			144.43
ϕ_{12} /°			87.01
ϕ_{13} /°			190.30
ϕ_{23} /°			271.05
θ_{12}^q /°	13.04	$12.99 \to 13.09$	13.04
θ_{13}^q /°	0.20	$0.19 \to 0.22$	0.20
θ_{23}^q /°	2.38	$2.32 \to 2.44$	2.38
δ^q /°	68.75	$64.25 \to 73.25$	60.23
m_u /MeV	1.28	$0.76 \to 1.55$	1.28
m_c /GeV	0.626	$0.607 \to 0.645$	0.626
m_t /GeV	171.6	$170 \to 173$	171.6
m_d /MeV	2.74	$2.57 \to 3.15$	2.49
m_s /MeV	54	$51 \to 57$	54
m_b /GeV	2.85	$2.83 \to 2.88$	2.85
χ^2			1.96

Table 11: Flavour parameters and observables: measured versus predicted values for the best fit point.

10.5.2 Neutrino oscillation predictions

In order to identify the predictions of the model concerning the oscillation parameters we have randomly varied the parameters around the global best fit point in Table 11, while requiring compatibility with all measured flavour observables at 3σ. The results of the analysis are given in Fig 39, where the blue contours represent the 90, 95, and 99% C.L. profiles from the Valencia global oscillation fit in [24, 25], while the purple dots indicate regions compatible at 3σ with all experimental data. The best fit point of the global oscillation fit is marked with a black star, while that of the global flavour fit is indicated by a white cross. One sees from Fig. 39 how the predicted values of the leptonic Dirac CP phase are restricted

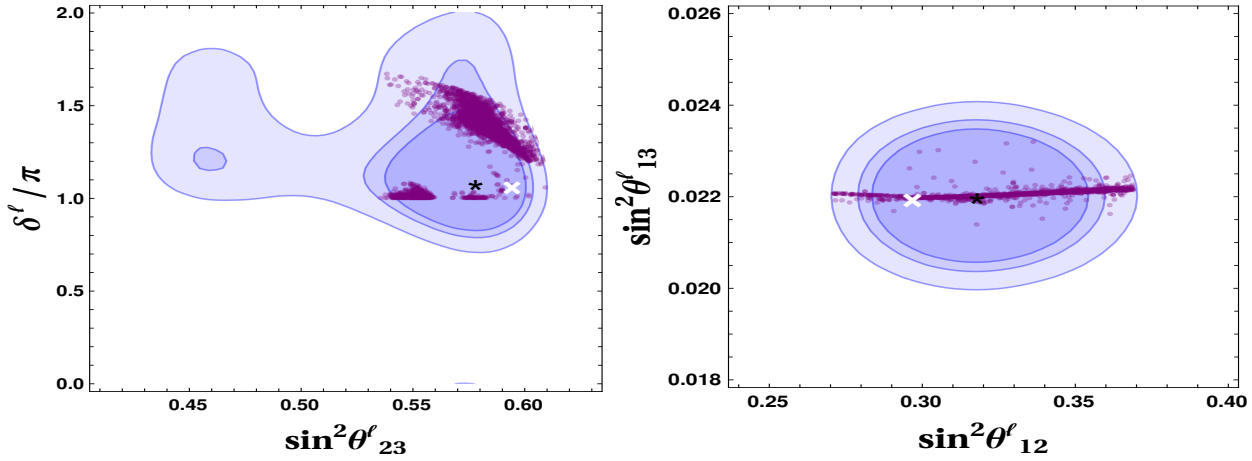


Figure 39: Allowed values for the mixing angles and the leptonic Dirac CP phase. The purple points are compatible at 3σ with all flavour observables, while the blue shades are the generic 90, 95 and 99% C.L. regions of the global oscillation fit in [24, 25]. The black star is the central point of the global oscillation fit, while the white cross stands for the best fit point in Table 11.

to the range $\delta^\ell \geq \pi$, while the atmospheric angle θ^ℓ_{23} is required to lie in the higher octant. Besides, one sees from the right panel in Fig. 39 a sharp prediction for the reactor angle θ^ℓ_{13}, see also table 11.

A scan of the parameter space of the model consistent at 3σ with all current experiments reveals that only the Normal Ordered (NO) neutrino mass spectrum is possible. Indeed, the best fit point in table 11 has positive Δm^2_{31}, corresponding to NO, and a rather high absolute scale for the neutrino masses.

10.5.3 Neutrinoless double beta decay predictions

Concerning neutrinoless double beta decay, using the Majorana phase and neutrino mass predictions of table 11 in Eq. (1.25), one finds the preferred effective amplitude parameter:

$$|m_{\beta\beta}| = 58.08\,\text{meV}. \tag{10.34}$$

A detailed analysis is presented in figure 40, where we show in purple the region of predicted $|m_{\beta\beta}|$ values as a function of the lightest neutrino mass m_1. To be conservative, we randomly varied the model parameters within the allowed 3σ range. The best fit point from table 11 is marked in red. One sees that the predicted central value of $|m_{\beta\beta}|$ lies inside the current exclusion band of Kamland-Zen ($36 - 156$ meV) [125].

shown in orange in Fig. 40. This will also be probed by cosmological observations and possibly by future beta decay endpoint studies. We also display the projected sensitivities of the next round of $0\nu\beta\beta$ experiments LEGEND [476] and nEXO [477] as the colored horizontal bands.

Notice that the central value of the lightest neutrino mass m_1 obtained from the global fit is disfavored by the latest results of the Planck collaboration on the sum of light neutrino masses [128]. This tension is further enhanced by the addition of Baryon Acoustic Oscillations (BAO) data [124], see vertical band in lighter gray in figure 40. Beyond the central prediction, however, there is a broad parameter region consistent both with measured flavour observables as well as with the cosmological bounds.

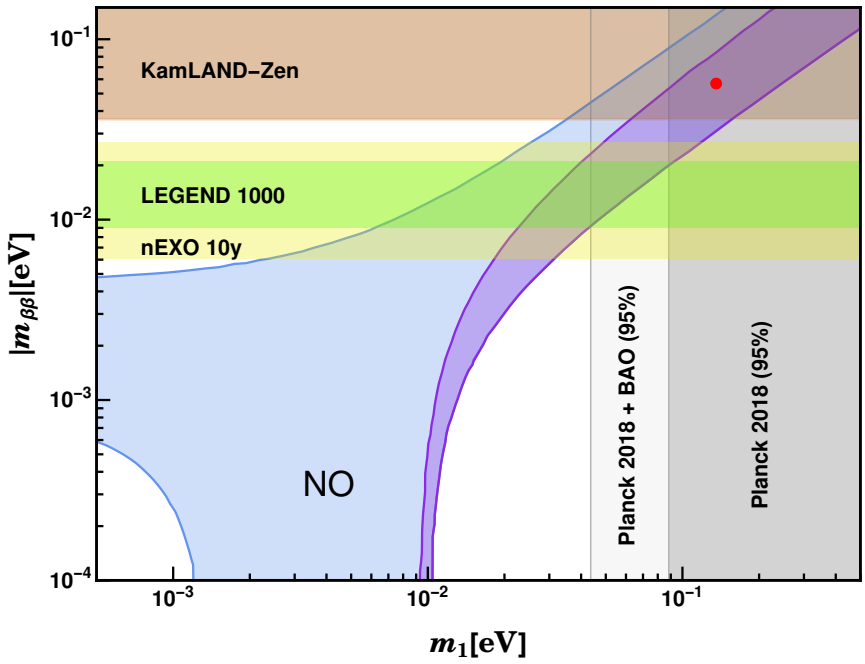


Figure 40: Effective $0\nu\beta\beta$ decay amplitude versus the lightest neutrino mass m_1. From the global fit one finds that only normal ordering is allowed. The blue region is the generic one consistent with oscillations at 2σ. The purple region is the one allowed at 3σ around the global best fit point in table 11, marked in red. The current KamLAND-Zen limit is shown in brown, and the projected sensitivities of future experiments LEGEND-1000 and nEXO are indicated with light yellow and light green horizontal bands respectively. The vertical gray bands represent the current sensitivity of cosmological data from the Planck collaboration (dark gray), and in combination with BAO data (light gray) [128, 129]

.

11 Recent progress: modular symmetry

In flavour symmetry models, complicated vacuum alignment assumptions are often required to break the flavor symmetry. The vacuum expectation values of the associated flavons should be conveniently oriented in family space. Moreover, higher-dimensional operators with flavon insertions and unknown coefficients are often present, affecting the resulting model predictions. In order to alleviate these shortcomings modular invariance as flavor symmetry has been recently proposed [478]. In what follows we briefly sketch recent work, for dedicated reviews on modular symmetries see Refs. [91, 479, 480].

11.1 The modular group

The modular group $SL(2,\mathbb{Z})$ is the group of 2×2 integer matrices with determinant one,

$$SL(2,\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| a,b,c,d \in \mathbb{Z}, ad-bc=1 \right\} . \tag{11.1}$$

The modular group $SL(2,\mathbb{Z})$ acts on the complex variable τ in the upper-half plane as the linear fractional transformation [481]

$$\gamma\tau = \frac{a\tau+b}{c\tau+d} \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \quad \text{and} \quad \mathrm{Im}(\tau) > 0\,. \tag{11.2}$$

One sees that γ and $-\gamma$ induce the same linear fraction transformation.

The modular group can be generated by the two generators S and T,

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} , \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} , \tag{11.3}$$

which lead to duality and shift symmetries of τ as follows,

$$\tau \xrightarrow{S} -\frac{1}{\tau}, \qquad \tau \xrightarrow{S} \tau+1\,. \tag{11.4}$$

Any value of τ in the upper half complex plane can be shifted in to the region of $-\frac{1}{2} \leq \mathrm{Re}(\tau) < \frac{1}{2}$ by multiple T transformations. Moreover, it can be mapped to the region $|\tau| \leq 1$ by an S transformation. As a consequence, the complex modulus τ could be restricted to the fundamental domain $\mathcal{F}$ of the modular group,

$$\mathcal{F} = \left\{ \tau \middle| -\frac{1}{2} \leq \mathrm{Re}(\tau) < \frac{1}{2}, \mathrm{Im}(\tau) > 0, |\tau| > 1 \right\} \cup \left\{ \tau \middle| -\frac{1}{2} \leq \mathrm{Re}(\tau) \leq 0, \mathrm{Im}(\tau) > 0, |\tau| = 1 \right\} , \tag{11.5}$$

which is displayed in figure 41. Every complex modulus τ can be mapped into the fundamental domain by a modular transformation of Eq. (11.2), and no two points in $\mathcal{F}$ can be related by modular transformations. Notice the right half boundary of $\mathcal{F}$ is not included into the fundamental domain, since it can related

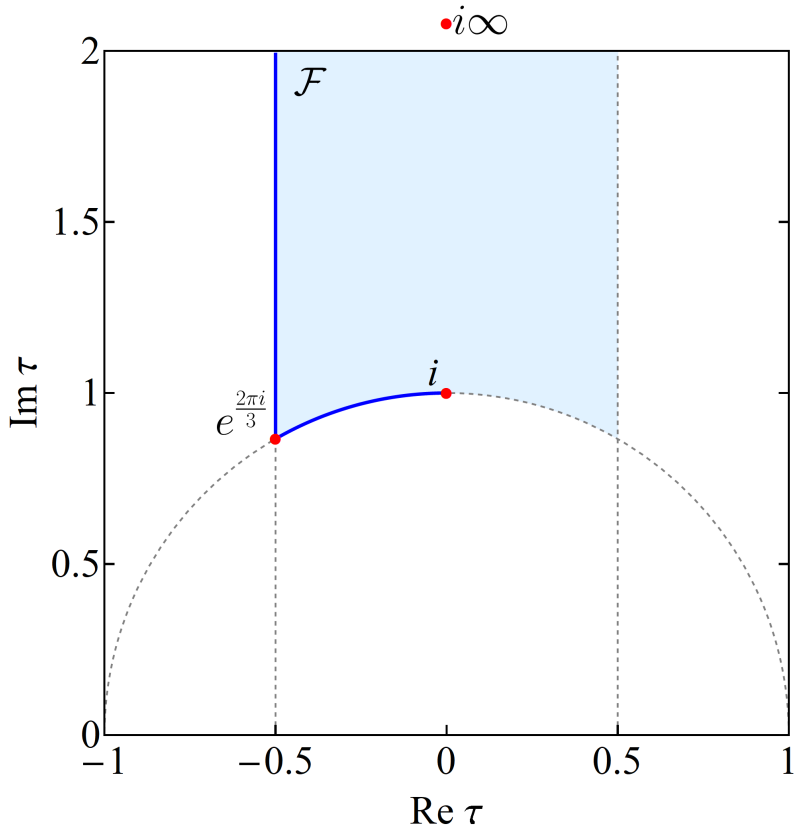


Figure 41: The fundamental domain $\mathcal{F}$ of the modular group, where the red points denote the three fixed points $\tau_0 = i, e^{2\pi i/3}, i\infty$ which preserve a residual modular symmetry.

to the left half boundary by some modular transformation. Moreover, no value of τ is left invariant by the whole modular group action of Eq. (11.2). In the fundamental domain, there are only three fixed points $\tau_0 = i, e^{2\pi i/3}, i\infty$ which break the modular group $SL(2,\mathbb{Z})$ partially. These are invariant under the modular transformations S, ST and T respectively [481–483]. Notice that each τ is trivially invariant under S^2 which is a negative identity matrix.

Modular symmetry provides a novel origin of discrete flavor symmetry through the quotient of $SL(2,\mathbb{Z})$ by the principal congruence subgroup of level N which is defined as,

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \middle| a, d = 1 \,(\text{mod } N), b, c = 0 \,(\text{mod } N) \right\} , \tag{11.6}$$

which is a normal subgroup of finite index in $SL(2,\mathbb{Z})$. The finite modular groups are the quotient groups $\Gamma_N = SL(2,\mathbb{Z})/\pm\Gamma(N)$ which play the role of discrete flavor symmetry. Remarkably, for $N \leq 5$ the finite modular groups Γ_N are isomorphic to the permutation groups [259, 478]

$$\Gamma_2 \cong S_3, \quad \Gamma_3 \cong A_4, \quad \Gamma_4 \cong S_4, \quad \Gamma_5 \cong A_5 \,, \tag{11.7}$$

which have been widely used as traditional flavor symmetries.

11.2 Modular invariance

A key concept of the theory of modular symmetries [478] is that of modular forms of level N and weight $2k$, denoted $Y(\tau)$. These are holomorphic functions of the modulus τ with the following transformation property

$$Y(\gamma\tau) = (c\tau + d)^{2k} Y(\tau), \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(N)\,, \tag{11.8}$$

where the integer $k \geq 0$ determines the weight $2k$.

There are only a finite number of linearly independent modular forms, denoted as $Y_i(\tau)$, and one can always choose a basis such that the transformation of the modular forms can be described by a unitary representation ρ of Γ_N:

$$Y_i(\gamma\tau) = (c\tau + d)^{2k} \rho_{ij}(\gamma) Y_j(\tau)\,, \tag{11.9}$$

where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ refers to a representative element of Γ_N. Likewise, the modular transformation properties of matter fields $\varphi^{(I)}$ are completely specified by the weight k_I and the unitary representation $\rho^{(I)}$ of the finite modular group Γ_N,

$$\varphi^I \to (c\tau + d)^{k_I} \rho^{(I)} \varphi^{(I)}\,. \tag{11.10}$$

Note that the modular flavor symmetry requires supersymmetry to preserve the holomorphicity of the modular form. Then the superpotential can be expanded in a power series of the matter fields $\varphi^{(I)}$ as follows,

$$\mathcal{W} = \sum_n Y_{I_1 I_2 \ldots I_n}(\tau) \varphi^{(I_1)} \varphi^{(I_2)} \ldots \varphi^{(I_n)}\,. \tag{11.11}$$

Modular invariance requires $Y_{I_1 I_2 \ldots I_n}(\tau)$ to be a modular form of weight k_Y and level N transforming in the representation ρ of Γ_N,

$$Y_{I_1 I_2 \ldots I_n}(\gamma\tau) = (c\tau + d)^{k_Y} \rho(\gamma) Y_{I_1 I_2 \ldots I_n}(\tau)\,, \tag{11.12}$$

with the conditions

$$k_Y + k_{I_1} + k_{I_2} + \ldots + + k_{I_n} = 0, \qquad \rho \otimes \rho^{(I_1)} \otimes \rho^{(I_2)} \otimes \ldots \rho^{(I_n)} \supset \mathbf{1}\,, \tag{11.13}$$

where $\mathbf{1}$ denotes the invariant singlet representation of Γ_N. One sees that the modular flavor symmetry requires the Yukawa couplings to be modular forms $Y_{I_1 I_2 \ldots I_n}(\tau)$. In the simplest implementation, flavons are not necessary and the complex modulus τ is the unique symmetry breaking parameter, greatly simplifying the alignment problem. Moreover, modular invariance requires the Yukawa couplings and fermion mass matrices to be combinations of modular forms, holomorphic functions of τ, all higher-dimensional operators are unambiguously determined in the limit of unbroken supersymmetry. As a result, flavour models with modular invariance depend on fewer parameters, enhancing their predictive power.

The above formalism of modular flavor symmetry has been extended to the double covers [435,484–487] or the metaplectic covers [487, 488] of the finite group Γ_N, then the modular weights can take integer values or more general rational values.

11.3 Generalized CP in modular symmetry

The interplay of modular symmetry and generalized CP symmetry (gCP) has been studied [489–491]. The consistency between the modular symmetry and gCP fixes the CP transformation of the modulus τ to be [489, 492–495]

$$\tau \xrightarrow{CP} -\tau^* \tag{11.14}$$

up to modular transformations. For any modular transformation

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}),$$

one can check that the action of the transformation chain $CP \to \gamma \to CP^{-1}$ on τ is,

$$\tau \xrightarrow{CP} -\tau^* \xrightarrow{\gamma} -\frac{a\tau^*+b}{c\tau^*+d} \xrightarrow{CP^{-1}} \frac{a\tau-b}{-c\tau+d}\,. \tag{11.15}$$

This implies that the CP transformation corresponds to an automorphism of the modular group, and it maps any modular transformation γ to another modular transformation $u(\gamma)$,

$$u(\gamma) \equiv CP \circ \gamma \circ CP^{-1} = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}. \tag{11.16}$$

In particular, one has

$$u(S) = CP \circ S \circ CP^{-1} = S^{-1}, \quad u(T) = CP \circ T \circ CP^{-1} = T^{-1}\,. \tag{11.17}$$

We proceed to consider the gCP transformation of an arbitrary chiral superfield multiplet $\varphi(x)$ in the representation $\mathbf{r}$ of the finite modular group Γ_N,

$$\varphi(x) \xrightarrow{CP} X_{\mathbf{r}}\,\overline{\varphi}(x_P)\,, \tag{11.18}$$

where a bar denotes the Hermitian conjugate superfield, and where $x = (t, \mathbf{x})$, $x_P = (t, -\mathbf{x})$ and $X_{\mathbf{r}}$ is a unitary matrix acting on flavour space. Applying the consistency condition chain of Eq. (11.16) to the matter multiplet φ, one find that the CP transformation matrix $X_{\mathbf{r}}$ has to satisfy the following constraint [489],

$$X_{\mathbf{r}}\,\rho^*_{\mathbf{r}}(\gamma) X^{-1}_{\mathbf{r}} = \rho_{\mathbf{r}}(u(\gamma))\,. \tag{11.19}$$

It is sufficient to just consider the modular generators $\gamma = S, T$, namely

$$X_{\mathbf{r}}\, \rho^*_{\mathbf{r}}(S) X^{-1}_{\mathbf{r}} = \rho_(S^{-1}), \quad X_{\mathbf{r}}\, \rho^*_{\mathbf{r}}(T) X^{-1}_{\mathbf{r}} = \rho_(T^{-1})\,, \tag{11.20}$$

which fixes $X_{\mathbf{r}}$ up to an overall phase by Schur's lemma, for each irreducible representation $\mathbf{r}$. In the basis where both S and T are represented by symmetric and unitary matrices, $X_{\mathbf{r}} = \mathbb{1}_{\mathbf{r}}$ solves the consistency condition. As a consequence, CP symmetry imposition enforces all coupling constants to be real in such basis. The VEV of τ is the unique source breaking both modular and gCP symmetries. The combination of modular and CP symmetries allows one to construct quite predictive flavor models. Indeed, it is remarkable that all the lepton masses, mixing angles and CP violation phases could be described in terms of only four real couplings plus the complex modulus τ in the minimal modular model of Refs. [496, 497]. The gCP symmetry can also be consistently implemented in the context of multiple moduli [491].

11.4 Modular symmetry from top-down

The idea of modular flavor symmetry is inspired by top-down considerations from string theory. The modular symmetry can naturally appear in orbifold compactifications of the heterotic string [492, 498] and magnetized toroidal compactification [499–505]. The double covers and metaplectic covers of the finite modular groups could be reproduced from top-down constructions. Given the fact that string compactification generally yields several moduli, the modular invariance has been extended to involve multiple moduli based on the direct product of several $SL(2,\mathbb{Z})$ [506] or the Symplectic modular group $Sp(2g,\mathbb{Z})$ [507], where g is a generic positive integer called genus. Note that $Sp(2g,\mathbb{Z})$ can arise as the duality group in string Calabi-Yau compactifications [508–510], and the group $Sp(2,\mathbb{Z})$ is isomorphic to the modular group $SL(2,\mathbb{Z})$. Moreover, it has been found that the modular symmetry and traditional flavor symmetry appear together in top-down constructions. This leads to the concept of eclectic flavor group [511, 512], a maximal extension of the traditional flavor group by finite modular group. It is more predictive than the finite modular group and also the traditional flavor group by itself, combining the advantages of both approaches. In particular, the interplay of modular symmetry and traditional flavor symmetry can restrict the Kähler potential more severely than the modular symmetry by itself. The possibility of eclectic flavor group can also be consistently combined with the gCP symmetry [511].

11.5 Quark-lepton mass relations from modular symmetry

Quite generally, the problem of understanding the pattern of fermion masses and mixings presents a two-fold challenge. While predicting fermion mixings through the imposition of flavor symmetries is relativeley straightforward, formulating a convincing theory of fermion masses seems tougher.

It has recently been shown that realistic fermion mass relations can arise naturally in modular invariant models [436, 513–515], without relying on *ad hoc* flavon alignments. As an example, Ref. [291] gave a set of viable fermion mass relations based on the $\Gamma_4 \cong S_4$ symmetry. The new versions exhibit calculable deviations from the usual Golden Mass Relation in Eq. (10.1). They were derived from modular flavor

symmetry in a rather general manner, relying only on the modular flavor group and its vector-valued modular forms, rather than *ad hoc* flavon alignments [25]. The new relations were shown to be viable and experimentally testable, distinguishing modular models from the conventional flavon-based prediction in Eq. (10.1).

The method is largely model-independent, and can adapted also to obtain predictions in the up-quark sector and neutrinos. It may also prove useful for more comprehensive modular invariant models and for top-down constructions.

11.6 Modular versus traditional flavor symmetry

We now comment on the predictive power of traditional flavor symmetry and modular symmetry. As we have shown in section 6, traditional flavor symmetry in combination with generalized CP symmetry allows one to predict lepton mixing angles and CP violation phases in terms of just one or two real free parameters when certain residual symmetry is preserved. In general, however, lepton masses are unconstrained by the residual symmetry.

In the case of modular symmetry, the fermion mass matrices exhibit certain symmetry-determined flavor structure through the modular forms which are functions of the modulus τ. One finds that both fermion masses as well as flavor mixing parameters can be determined within specific models. However, one must resort to numerical analysis to reveal the possible correlations amongst the flavor parameters, since a residual modular symmetry with a single modulus has not led to phenomenologically viable models. Concerning vacuum alignment, it is notoriously difficult to dynamically realize the alignment required by the residual symmetry within traditional flavor symmetry constructions. On the other hand, determining the VEV of the modulus τ from a dynamical principle is also an open question, some possible schemes have been proposed to stabilize the modular VEV [516–524]. In modular flavor models, the VEV of τ is usually treated as free parameter, and its value is determined by confronting the model predictions with experimental data.

In short, modular flavor symmetries may shed light on the ultimate symmetry underlying the flavor problem. They have been studied from both the bottom-up and top-down approaches. There are still some unsolved problems such as modulus stabilization and the Kähler potential problem. However, invoking modular symmetry seems a useful tool towards formulating theories of flavor. A comprehensive description of these recent developments lies outside the scope of the present work, see Refs. [91, 479, 480] dedicated modular symmetry reviews.

[25] They are determined by the Clebsch-Gordan coefficients of the given finite modular group as well as the expansion coefficients of its modular forms.

12 Summary and outlook

The flavour problem constitutes perhaps one of the major open issues and deepest challenges of particle physics. Even more so after the revolutionary discovery and confirmation of neutrino oscillations. The main legacy of these experiments has been to show that leptons mix rather differently from quarks, providing a key input on the flavour issue. Indeed, it seems unlikely that the peculiar pattern of neutrino mixing angles extracted from experiment is just a random coincidence. Rather, it should have its roots in some, perhaps subtle, underlying symmetry of nature.

By itself, the Standard Model lacks an organizing principle in terms of which to understand the "flavours" of the fundamental building blocks of matter. First of all, the flavour challenge comprises an understanding of family replication, i.e. why nature repeats itself three times. Moreover, it also requires an explanation for the pattern of fermion masses and mixings. This review examined the possibility that the flavour puzzle has a symmetry explanation.

To set up the stage we started with an introductory section in which we gave basic preliminaries on the description of lepton and quark mixing, as well as a recap on the status of neutrino oscillations, Fig. 1 and Fig. 2, and neutrinoless double beta decay, e.g. Fig. 3. We discussed ways to have detectable $0\nu\beta\beta$ decay rates, in particular theories where the lightest neutrino is massless or nearly so, see Fig. 4, as occurs, for example, in the missing partner seesaw mechanism, where there are less "right" than "left"-handed neutrinos. We also commented on the detectability of $0\nu\beta\beta$ decay in schemes with family symmetries, e.g. Fig. 5, and also on the significance of a possible $0\nu\beta\beta$ discovery for particle physics, Fig. 6.

In Section 2 we briefly discussed the origin of neutrino masses, both from point of view of effective theories, e.g. the Weinberg operator in Fig. 7, as well as, for instance, the type-I realization of the seesaw paradigm, Fig. 8, or the "scotogenic" paradigm. In the latter dark matter mediates neutrino mass generation through radiative corrections. We also briefly discussed the idea that dark matter *seeds* neutrino mass generation that proceeds *a la seesaw*, as in the recent recently proposed low-scale seesaw variants dubbed "dark inverse" and "dark linear seesaw" Fig. 9. Finally, we mentioned the *flavour problem*, a major SM drawback and main thread of this review, and how it may be approached by symmetry, as illustrated in Fig. 10.

In Sect. 3 we described various phenomenological lepton mixing patterns, showing how most of them are at odds with current experimental data, especially from Daya Bay. In Sect. 4 we gave a bottom-up description of residual flavour and CP symmetries, both for the case of leptons as well as quarks. Indeed, the residual CP and flavor symmetries of quark and lepton mass terms are determined in terms of the experimentally measured mixing matrix, as given in Eqs. (4.11, 4.12, 4.13, 4.24, 4.25, 4.26) and Eqs. (4.55, 4.59). On the other hand, they may arise from the breaking of flavour and CP symmetries at high energies. It is remarkable that one can fix the quark and lepton mixing matrices from the structure of the flavour symmetry group and the residual symmetries, irrespective of the flavour symmetry breaking dynamics. Residual symmetries can restrict the mixing matrix as summarized in table 1. Moreover, remnant symmetries are quite useful as a guide to construct concrete flavour models. They can also

provide adequate revampings of various neutrino mixing patterns at odds with experiment so as to yield generalizations that are not only viable, but also predictive, as discussed in Sect. 5. For example, viable and predictive generalizations of the TBM mixing pattern are given in Figs. 11 and 12. On the other hand, similar generalizations of the Golden-ratio mixing pattern are given in Fig. 13. Soon after the first Daya-Bay results indicating a nonzero value of θ_{13} the bi-large mixing pattern was suggested, in which solar and atmspheric mixing angles as well as the Dirac CP phase are determined in terms of θ_{13}, as seen in Figs. 15 and 16.

Within the family symmetry paradigm, the symmetry group is usually broken down to different subgroups in the neutrino and charged lepton sectors, the lepton mixing matrix arises from the mismatch of the symmetry breaking patterns, as shown in Fig. 17 in Sect.6. If a Klein subgroup is preserved by the (Majorana) neutrino mass term and the residual abelian subgroup of the charged lepton sector distinguishes the three generations, the lepton mixing matrix would be completely fixed from group theory up to row and column permutations. On the other hand, the Dirac CP phase δ_{CP} is predicted to take on CP conserving values, while the Majorana phases cannot be constrained. Instead, if a Z_2 subgroup is preserved in the neutrino sector, then only one column of the lepton mixing matrix is determined and δ_{CP} can lie in a relative large region, e.g. see Fig. 18. Generalized CP symmetries are very powerful to constrain the CP phase. For instance, the $\mu-\tau$ reflection symmetry on the neutrino fields enforces δ_{CP} to be maximal, while the predicted δ_{CP} values for the generalized $\mu-\tau$ reflection are displayed in Fig. 19.

A flavour symmetry transformation can be generated by performing two CP transformations, the interplay of flavour symmetry and generalized CP symmetry is highly nontrivial, requiring that the consistency condition in Fig. 20 must hold. The inclusion of generalized CP symmetry in flavour symmetry provides richer symmetry breaking patterns. For example, one usually assumes that the flavour and CP symmetry are broken to an abelian subgroup in the charged lepton sector and $Z_2 \times CP$ in the neutrino sector. In this case all the lepton mixing parameters depend only on a single real rotation angle θ in Fig. 21. The value of θ is fixed by the precisely measured reactor angle θ_{13}, leading to a determination of the other lepton mixing angles and CP violation phases, see, e.g. Eq. (6.47). This also implies definite predictions for the $0\nu\beta\beta$ decay amplitude, as shown in Fig. 22. Another possibility is that the remnant subgroup preserved by both neutrino and charged lepton mass terms has the structure $Z_2 \times CP$. In this case the lepton mixing matrix depends on two free rotation angles θ_l and θ_ν, see Fig. 23. All the mixing angles and CP violation phases as well as the effective $0\nu\beta\beta$ mass parameter are predicted to lie in very restricted regions, as shown in Figs. 24, 25, 26. It is remarkable that this scheme can be extended to the quark sector, both quark and lepton mixings can be accommodated using a single flavour group, in terms of total four free parameters. Generalized flavour and CP symmetries may also be employed for the case of Dirac neutrinos. Similar results hold true, except that the remnant flavour symmetry of neutrino sector can be an arbitrary abelian subgroup. In particular, the master formulas in Figs. 21 and 23 also hold for Dirac neutrinos.

In section 7 we gave a panoramic view illustrating possible tests of flavor symmetry models. The

existence of mixing predictions is a characteristic feature of flavor symmetry models. This follows from the fact that typically they have less free parameters than physical observables, therefore correlations between light neutrino observables are predicted. This is explicitly shown in Eqs. (5.12, 5.18, 5.24, 6.9, 6.22, 6.47, 6.51, 6.54, 6.73). For example the ranges of the CP violation $\delta_{\rm CP}$ obtained from Eq. (7.1) are summarized in table 3. The predictions of some typical flavor models are displayed in figure 27. One can see that a precise measurement of the lepton mixing parameters at forthcoming neutrino facilities JUNO, DUNE and T2HK should provide a test of flavor models, see also figure 28. On the other hand, neutrino mass sum rules are generally parameterized as Eq. (7.4), providing another type of correlation. It relates the three complex light neutrino mass eigenvalues amongst each other and thus also relate the Majorana CP phases to the neutrino masses. The mass sum rules lead to strong restrictions on the lightest neutrino mass and to distinct predictions for the effective mass $|m_{\beta\beta}|$ probed in $0\nu\beta\beta$ decay, as shown in figure 29. Finally, we commented on a recently developed toolkit to contrast flavor models with upcoming oscillation experiments.

So far we have confined ourselves to model-independent approaches to the flavour puzzle, in which the underlying theory is unspecified and only the predictive power of symmetry is explored. As a next step we turned to various UV-complete approaches to the flavour puzzle. In Sect. 8 we gave two benchmarks for UV-complete constructions in 4 dimensions. The first incorporates dark matter in a "scotogenic" manner, Fig. 30 together with a successful family symmetry leading to the so-called trimaximal pattern TM2. The neutrino oscillation predictions are illustrated in Figs. 31 and 32. An alternative example given puts together family and CP symmetry within the same construction.

More ambitious approaches to the flavor problem have been proposed within extra spacetime dimensions. For example, in Sect. 9 we described a five-dimensional warped flavordynamics scenario in which mass hierarchies are accounted for by adequate choices of the bulk mass parameters, while quark and lepton mixing angles are restricted by the imposition of a family symmetry. We presented a T' model leading to Majorana neutrinos, with the TM1 mixing pattern and tight neutrino oscillation correlations, given in Fig. 35. The $0\nu\beta\beta$ decay rates lie within the sensitivities of the next round of experiments, as indicated in Fig. 36. Finally, one has a good global fit of all flavour observables, including quarks, see table 9.

Finally, in Sect. 10 we described orbifold compactification as a promising way to determine the structure of the family symmetry in four dimensions. The construction is illustrated in Fig. 37. We illustrated the idea with a benchmark 6-dimensional scotogenic (see Fig. 38) orbifold scenario, in which a 4-dimensional A_4 flavour group emerges from the symmetries between the branes in extra dimensions. Predictions include the "golden" quark-lepton mass relation, Eq. (10.33), and a very good global description of all flavour observables, including quarks, table 11. Concerning neutrino oscillations, the mass ordering and atmospheric octant are predicted, together with the reactor angle, see Fig. 39. The lightest neutrino mass can be probed in neutrinoless double beta decay searches as well as cosmology, Fig. 40.

We have also stressed that adequate vacuum alignment is required in flavour symmetry models. Generally one must introduce additional large shaping symmetries and many new fields so as to cleverly design the flavon potential needed to obtain the correct vacuum alignment. In Sec. 11 we discussed how

modular symmetry provides an interesting way to overcome this drawback [478]. In this case the role of flavor symmetry is played by the modular invariance, and the complex modulus τ could be the unique source of modular symmetry breaking. Modular invariance requires the Yukawa couplings to be modular forms, thus a small number of free parameters are involved in modular models. Such modular symmetry approach has been comprehensively reviewed elsewhere [91, 479, 480] within the bottom-up as well as top-down approaches.

Recent progress in the formulation of flavor models involving the use of modular symmetry is discussed in Sec. 11. This idea opens the door to a simpler description and deeper understanding of the flavor symmetry breaking required for obtaining viable flavor predictions without the need to invoke *flavons* in an *ad hoc* manner.

All in all, the legacy of the oscillation program over the past two decades has been a tremendous progress in the field, bringing neutrinos to the center of the particle physics stage. Indeed, addressing the dynamical origin of small neutrino masses touches the heart of the electroweak theory, such as the consistency of symmetry breaking. Moreover, the precise measurement of the neutrino mixing parameters could shed light into the flavor problem. One might have expected that this would bring a decisive boost towards the formulation of a comprehensive theory of fermion masses and mixings. It is somewhat frustrating, however, that so far no decisive flavor road map has emerged emerged. One can reproduce the observations in many different ways, within a wide range of models that go all the way from anarchy to discrete family symmetries. While the latter seems intellectually more appealing, we have not yet been able to underpin a convincing final theory of flavor. Despite many interesting ideas and the formulation of a plethora elegant models, the structure of the three families of fermions remains mysterious.

From the experimental viewpoint in the coming decade we expect a vibrant period for oscillation studies, within and beyond the minimum paradigm. Current neutrino facilities as well as future ones, such as JUNO, DUNE and T2HK should be capable of measuring the solar angle θ_{12}, the atmospheric angle θ_{23} and the Dirac CP phase δ_{CP} with high sensitivity. The next generation of ton-scale $0\nu\beta\beta$ decay experiments will probe the Majorana nature of neutrinos, exploring the whole region associated with the inverted ordering spectrum. These experiments should provide important insights into the mysteries behind flavor mixing, fermion mass hierarchies and CP violation.

A The A_4 group

A_4 is the even permutation group of four objects, and it is isomorphic to the symmetry group of a regular tetrahedron. The A_4 group can be generated by two generators s and t which satisfy the following multiplication rules [26],

$$s^2 = t^3 = (st)^3 = 1\,. \tag{A.1}$$

The 12 elements of A_4 belong to four conjugacy classes:

$$\begin{aligned} &1C_1 = \{1\}\,, && 3C_2 = \{s, tst^2, t^2st\}\,, \\ &4C_3 = \{t, st, ts, sts\}\,, && 4C_3' = \{t^2, st^2, t^2s, st^2s\}\,, \end{aligned} \tag{A.2}$$

where the conjugacy class is denoted by kC_n, k is the number of elements belonging to it, and the subscript n is the order of the elements contained in it. A_4 has a unique Klein group $K_4^{(s,tst^2)}$ and three Z_3 subgroups generated by t, st, ts and sts respectively. The A_4 group has four inequivalent irreducible representations: three singlets $\mathbf{1}$, $\mathbf{1}'$, $\mathbf{1}''$ and a triplet $\mathbf{3}$. Two different bases are used in the literature, the Ma-Rajasekaran (MR) basis [53] and the Altarelli-Feruglio (AF) basis [313]. They differ in the three-dimensional irreducible representation $\mathbf{3}$, the representation matrix of the generator t is real in the MR basis [53] while it is complex and diagonal in AF basis [313]. The explicit form of the representation matrices are summarized in table 12. The two bases are related through a unitary transformation,

$$s_{AF} = V^\dagger s_{MR} V, \quad t_{AF} = V^\dagger t_{MR} V\,, \tag{A.3}$$

where

$$V = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}\,. \tag{A.4}$$

If we have two triplets $\alpha \sim (\alpha_1, \alpha_2, \alpha_3)$ and $\beta \sim (b_1, b_2, b_3)$, their product decomposes as the sum

$$\mathbf{3} \otimes \mathbf{3} = \mathbf{1} \oplus \mathbf{1}' \oplus \mathbf{1}'' \oplus \mathbf{3}_\mathrm{s} \oplus \mathbf{3}_\mathrm{a}\,, \tag{A.5}$$

where $\mathbf{3}_\mathrm{s}$ and $\mathbf{3}_\mathrm{a}$ denote the symmetric and the antisymmetric triplet combinations respectively. The results for the contractions in the above two bases are summarized in table 12.

[26] We use small s, t, u for the generators of A_4 and S_4 flavour groups in order to avoid confusion with oblique parameters S, T, U.

Representation matrices of A_4 generators				
	Ma-Rajasekaran basis		Altarelli-Feruglio basis	
	s	t	s	t
$\mathbf{1}$	1	1	1	1
$\mathbf{1}'$	1	ω	1	ω
$\mathbf{1}''$	1	ω^2	1	ω^2
$\mathbf{3}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	$\frac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}$

Tensor products of two A_4 triplets		
	Ma-Rajasekaran basis	Altarelli-Feruglio basis
$(\alpha \otimes \beta)_{\mathbf{1}}$	$\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3$	$\alpha_1\beta_1 + \alpha_2\beta_3 + \alpha_3\beta_2$
$(\alpha \otimes \beta)_{\mathbf{1}'}$	$\alpha_1\beta_1 + \omega^2\alpha_2\beta_2 + \omega\alpha_3\beta_3$	$\alpha_3\beta_3 + \alpha_1\beta_2 + \alpha_2\beta_1$
$(\alpha \otimes \beta)_{\mathbf{1}''}$	$\alpha_1\beta_1 + \omega\alpha_2\beta_2 + \omega^2\alpha_3\beta_3$	$\alpha_2\beta_2 + \alpha_3\beta_1 + \alpha_1\beta_3$
$(\alpha \otimes \beta)_{\mathbf{3}_\mathrm{s}}$	$\begin{pmatrix} \alpha_2\beta_3 + \alpha_3\beta_2 \\ \alpha_3\beta_1 + \alpha_1\beta_3 \\ \alpha_1\beta_2 + \alpha_2\beta_1 \end{pmatrix}$	$\begin{pmatrix} 2\alpha_1\beta_1 - \alpha_2\beta_3 - \alpha_3\beta_2 \\ 2\alpha_3\beta_3 - \alpha_1\beta_2 - \alpha_2\beta_1 \\ 2\alpha_2\beta_2 - \alpha_3\beta_1 - \alpha_1\beta_3 \end{pmatrix}$
$(\alpha \otimes \beta)_{\mathbf{3}_\mathrm{a}}$	$\begin{pmatrix} \alpha_2\beta_3 - \alpha_3\beta_2 \\ \alpha_3\beta_1 - \alpha_1\beta_3 \\ \alpha_1\beta_2 - \alpha_2\beta_1 \end{pmatrix}$	$\begin{pmatrix} \alpha_2\beta_3 - \alpha_3\beta_2 \\ \alpha_1\beta_2 - \alpha_2\beta_1 \\ \alpha_3\beta_1 - \alpha_1\beta_3 \end{pmatrix}$

Table 12: The representation matrices of the A_4 generators s and t in the different irreducible representations. We also give the tensor product rule of two A_4 triplets $\alpha = (\alpha_1, \alpha_2, \alpha_3) \sim \mathbf{3}$ and $\beta = (\beta_1, \beta_2, \beta_3) \sim \mathbf{3}$, where $\omega = e^{i2\pi/3} = -1/2 + i\sqrt{3}/2$ is a cubic root of unity.

B The S_4 group

S_4 is the permutation group of four distinct objects; geometrically it is the symmetry group of a regular octahedron. Its generators s, t and u obey the following multiplication rules [325, 395, 525],

$$s^2 = t^3 = u^2 = (st)^3 = (su)^2 = (tu)^2 = (stu)^4 = 1\,. \tag{B.1}$$

Note that the generators s and t alone generate the group A_4, while the generators t and u alone generate the group S_3. The S_4 group elements can be divided into 5 conjugacy classes

$$\begin{aligned}
1C_1 &= \{1\}\,,\\
3C_2 &= \left\{s, tst^2, t^2st\right\},\\
6C_2' &= \left\{u, tu, su, ut, stsu, st^2su\right\},\\
8C_3 &= \left\{t, st, ts, sts, t^2, st^2, t^2s, st^2s\right\},\\
6C_4 &= \left\{stu, tsu, t^2su, st^2u, tst^2u, t^2stu\right\},
\end{aligned} \tag{B.2}$$

The group structure of S_4 has been studied in detail in [526], it has thirty proper subgroups of orders 1, 2, 3, 4, 6, 8, 12 or 24. Since the number of inequivalent irreducible representations is equal to the number of conjugacy classes and the sum of the squares of the dimensions of the irreducible representations must be equal to the order of the group, it is easy to see that S_4 has two singlet irreducible representations **1** and **1′**, one doublet representation **2** and two triplet representations **3** and **3′**. In the singlet representations **1** and **1′**, we have

$$\begin{aligned}
&\mathbf{1}\ :\ s = t = u = 1\,,\\
&\mathbf{1'}\ :\ s = t = 1, \quad u = -1\,.
\end{aligned} \tag{B.3}$$

For the doublet representation **2**, the generators are represented by

$$s = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad t = \begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}, \quad u = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{B.4}$$

with $\omega = e^{2\pi i/3}$. In the triplet representations **3** and **3′**, the generators are

$$\begin{aligned}
\mathbf{3}: s &= \frac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}, \quad t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix}, \quad u = -\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\\
\mathbf{3'}: s &= \frac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}, \quad t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix}, \quad u = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.
\end{aligned} \tag{B.5}$$

Classes	$1C_1$	$3C_2$	$6C_2'$	$8C_3$	$6C_4$
$\mathbf{1}$	1	1	1	1	1
$\mathbf{1'}$	1	1	-1	1	-1
$\mathbf{2}$	2	2	0	-1	0
$\mathbf{3}$	3	-1	-1	0	1
$\mathbf{3'}$	3	-1	1	0	-1

Table 13: Character table of the S_4 group.

Notice that the representations $\mathbf{3}$ and $\mathbf{3'}$ differ in the overall sign of the generator u. The character of an element is the trace of its representation matrix, then we can straightforwardly obtain the character table of S_4, as shown in table 13. Moreover, the decompositions of the tensor product of the S_4 irreducible representations are as follows,

$$\begin{aligned}
&\mathbf{1}\otimes\mathbf{r}=\mathbf{r}, \qquad \mathbf{1'}\otimes\mathbf{1'}=\mathbf{1}, \qquad \mathbf{1'}\otimes\mathbf{2}=\mathbf{2}, \qquad \mathbf{1'}\otimes\mathbf{3}=\mathbf{3'}, \qquad \mathbf{1'}\otimes\mathbf{3'}=\mathbf{3},\\
&\mathbf{2}\otimes\mathbf{2}=\mathbf{1}\oplus\mathbf{1'}\oplus\mathbf{2}, \qquad \mathbf{2}\otimes\mathbf{3}=\mathbf{2}\otimes\mathbf{3'}=\mathbf{3}\otimes\mathbf{3'},\\
&\mathbf{3}\otimes\mathbf{3}=\mathbf{3'}\otimes\mathbf{3'}=\mathbf{1}\oplus\mathbf{2}\oplus\mathbf{3}\oplus\mathbf{3'}, \qquad \mathbf{3}\otimes\mathbf{3'}=\mathbf{1'}\oplus\mathbf{2}\oplus\mathbf{3}\oplus\mathbf{3'}\,,
\end{aligned} \tag{B.6}$$

where $\mathbf{r}$ stands for any irreducible representation of S_4. We proceed to present the Clebsch-Gordan (CG) coefficients in the above basis. The entries of the two multiplets in the tensor product are denoted by α_i and β_i respectively. For the product of the singlet $\mathbf{1'}$ with a doublet or a triplet, we have [325, 395, 525]

$$\mathbf{1'}\otimes\mathbf{2}=\mathbf{2}=\begin{pmatrix}\alpha\beta_1\\-\alpha\beta_2\end{pmatrix}, \quad \mathbf{1'}\otimes\mathbf{3}=\mathbf{3'}=\begin{pmatrix}\alpha\beta_1\\\alpha\beta_2\\\alpha\beta_3\end{pmatrix}, \quad \mathbf{1'}\otimes\mathbf{3'}=\mathbf{3}=\begin{pmatrix}\alpha\beta_1\\\alpha\beta_2\\\alpha\beta_3\end{pmatrix}. \tag{B.7}$$

The CG coefficients for the products involving the doublet representation $\mathbf{2}$ are found to be

$$\mathbf{2}\otimes\mathbf{2}=\mathbf{1}\oplus\mathbf{1'}\oplus\mathbf{2}, \qquad \text{with} \quad \begin{cases}\mathbf{1} &= \alpha_1\beta_2+\alpha_2\beta_1\\ \mathbf{1'} &= \alpha_1\beta_2-\alpha_2\beta_1\\ \mathbf{2} &= \begin{pmatrix}\alpha_2\beta_2\\\alpha_1\beta_1\end{pmatrix}\end{cases} \tag{B.8}$$

$$\mathbf{2}\otimes\mathbf{3}=\mathbf{3}\oplus\mathbf{3'}, \qquad \text{with} \quad \begin{cases}\mathbf{3} &= \begin{pmatrix}\alpha_1\beta_2+\alpha_2\beta_3\\\alpha_1\beta_3+\alpha_2\beta_1\\\alpha_1\beta_1+\alpha_2\beta_2\end{pmatrix}\\ \mathbf{3'} &= \begin{pmatrix}\alpha_1\beta_2-\alpha_2\beta_3\\\alpha_1\beta_3-\alpha_2\beta_1\\\alpha_1\beta_1-\alpha_2\beta_2\end{pmatrix}\end{cases} \tag{B.9}$$

$$\mathbf{2}\otimes\mathbf{3}'=\mathbf{3}\oplus\mathbf{3}', \qquad \text{with} \quad \begin{cases} \mathbf{3} = \begin{pmatrix} \alpha_1\beta_2-\alpha_2\beta_3 \\ \alpha_1\beta_3-\alpha_2\beta_1 \\ \alpha_1\beta_1-\alpha_2\beta_2 \end{pmatrix} \\ \mathbf{3}' = \begin{pmatrix} \alpha_1\beta_2+\alpha_2\beta_3 \\ \alpha_1\beta_3+\alpha_2\beta_1 \\ \alpha_1\beta_1+\alpha_2\beta_2 \end{pmatrix} \end{cases} \tag{B.10}$$

Similarly, for the tensor products among the triplet representations $\mathbf{3}$ and $\mathbf{3}'$, we have

$$\mathbf{3}\otimes\mathbf{3}=\mathbf{3}'\otimes\mathbf{3}'=\mathbf{1}\oplus\mathbf{2}\oplus\mathbf{3}\oplus\mathbf{3}', \text{ with } \begin{cases} \mathbf{1} = \alpha_1\beta_1+\alpha_2\beta_3+\alpha_3\beta_2 \\ \mathbf{2} = \begin{pmatrix} \alpha_2\beta_2+\alpha_1\beta_3+\alpha_3\beta_1 \\ \alpha_3\beta_3+\alpha_1\beta_2+\alpha_2\beta_1 \end{pmatrix} \\ \mathbf{3} = \begin{pmatrix} \alpha_2\beta_3-\alpha_3\beta_2 \\ \alpha_1\beta_2-\alpha_2\beta_1 \\ \alpha_3\beta_1-\alpha_1\beta_3 \end{pmatrix} \\ \mathbf{3}' = \begin{pmatrix} 2\alpha_1\beta_1-\alpha_2\beta_3-\alpha_3\beta_2 \\ 2\alpha_3\beta_3-\alpha_1\beta_2-\alpha_2\beta_1 \\ 2\alpha_2\beta_2-\alpha_1\beta_3-\alpha_3\beta_1 \end{pmatrix} \end{cases} \tag{B.11}$$

$$\mathbf{3}\otimes\mathbf{3}'=\mathbf{1}'\oplus\mathbf{2}\oplus\mathbf{3}\oplus\mathbf{3}', \text{ with } \begin{cases} \mathbf{1}' = \alpha_1\beta_1+\alpha_2\beta_3+\alpha_3\beta_2 \\ \mathbf{2} = \begin{pmatrix} \alpha_2\beta_2+\alpha_1\beta_3+\alpha_3\beta_1 \\ -(\alpha_3\beta_3+\alpha_1\beta_2+\alpha_2\beta_1) \end{pmatrix} \\ \mathbf{3} = \begin{pmatrix} 2\alpha_1\beta_1-\alpha_2\beta_3-\alpha_3\beta_2 \\ 2\alpha_3\beta_3-\alpha_1\beta_2-\alpha_2\beta_1 \\ 2\alpha_2\beta_2-\alpha_1\beta_3-\alpha_3\beta_1 \end{pmatrix} \\ \mathbf{3}' = \begin{pmatrix} \alpha_2\beta_3-\alpha_3\beta_2 \\ \alpha_1\beta_2-\alpha_2\beta_1 \\ \alpha_3\beta_1-\alpha_1\beta_3 \end{pmatrix} \end{cases} \tag{B.12}$$

C The dihedral group

The dihedral group D_n is the symmetry group of an n-sided regular polygon, which includes rotations and reflections. For instance, D_3 is the symmetry group of the regular triangle, and consequently it is isomorphic to S_3. A regular polygon with n sides is invariant under a rotation of multiples of $2\pi/n$ about the origin. If n is odd, each reflection axis connects the midpoint of one side to the opposite vertex. If n is even, there are $n/2$ reflection axes connecting the midpoints of opposite sides and $n/2$ axes of symmetry connecting opposite vertices. Consequently the group D_n has $2n$ elements. The group D_n can be generated by a rotation R of order n and a reflection S of order 2, and they satisfy the following multiplication rules

$$R^n = S^2 = (RS)^2 = 1\,, \tag{C.1}$$

where R denotes a rotation of $2\pi/n$ about the origin, and S is a reflection across n lines through the origin. Notice that D_1 is Z_2 group generated by S and D_2 is isomorphic to $Z_2 \times Z_2$. All the group elements of D_n can be expressed as

$$g = S^\alpha R^\beta \tag{C.2}$$

where $\alpha = 0, 1$ and $\beta = 0, 1, \ldots, n-1$. Then it is straightforward to determine the conjugacy classes of the dihedral group. Depending on whether the group index n is even or odd, the $2n$ group elements of D_n can be classified into three or five types of conjugacy classes.

- odd n

$$\begin{aligned} 1C_1 &= \{1\}\,, \\ 2C_m^{(\rho)} &= \{R^\rho, R^{-\rho}\}\,, \quad \text{with} \quad \rho = 1, \ldots, \frac{n-1}{2}\,, \\ nC_2 &= \{S, SR, SR^2, \ldots, SR^{n-1}\}\,, \end{aligned} \tag{C.3}$$

  where m refers to the order of the element R^ρ.

- even n

$$\begin{aligned} 1C_1 &= \{1\}\,, \\ 1C_2 &= \{R^{n/2}\}\,, \\ 2C_m^{(\rho)} &= \{R^\rho, R^{-\rho}\}\,, \quad \text{with} \quad \rho = 1, \ldots, \frac{n-2}{2}\,, \\ \frac{n}{2}C_2 &= \{S, SR^2, SR^4, \ldots, SR^{n-4}, SR^{n-2}\}\,, \\ \frac{n}{2}C_2 &= \{SR, SR^3, \ldots, SR^{n-3}, SR^{n-1}\}\,. \end{aligned} \tag{C.4}$$

The subgroups of D_n are either dihedral groups or cyclic groups, and they are given by

$$\begin{aligned} Z_j &= < R^{\frac{n}{j}} > \quad \text{with} \quad j|n\,, \\ Z_2^{(m)} &= < SR^m > \quad \text{with} \quad m = 0, 1, \ldots, n-1\,, \\ D_j^{(m)} &= < R^{\frac{n}{j}}, SR^m > \quad \text{with} \quad j|n,\ m = 0, 1, \ldots, \frac{n}{j} - 1\,, \end{aligned} \tag{C.5}$$

where the elements inside the angle brackets denote the generators of the subgroups. We see that the total number of dihedral subgroups is the sum of positive divisors of n.

The group D_n is a subgroup of $SO(2)$, and it has only one-dimensional and two-dimensional irreducible representations. The representations of D_n crucially depend on the value of the group index n.

- odd n

  If n is an odd integer, the group D_n has two singlet representations $\mathbf{1}_i$ and $\frac{n-1}{2}$ doublet representations $\mathbf{2}_j$, where the subscripts i and j take values $i = 1, 2$ and $j = 1, \ldots, \frac{n-1}{2}$. The sum of the squares of the dimensions of the irreducible representations is

  $$1^2 + 1^2 + 2^2 \times \frac{n-1}{2} = 2n\,, \tag{C.6}$$

  which is exactly the number of elements of the D_n group. In the singlet representations, the generators R and S are represented by

  $$\mathbf{1}_1 :\ R = S = 1\,, \quad \mathbf{1}_2 :\ R = 1,\ S = -1\,. \tag{C.7}$$

  For the doublet representations, we have

  $$\mathbf{2}_j :\ R = \begin{pmatrix} e^{2\pi i \frac{j}{n}} & 0 \\ 0 & e^{-2\pi i \frac{j}{n}} \end{pmatrix}\,, \quad S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \tag{C.8}$$

  with $j = 1, \ldots, \frac{n-1}{2}$.

- even n

  For the case that the index n an even integer, the group D_n has four singlet representations $\mathbf{1}_i$ with $i = 1, 2, 3, 4$ and $\frac{n}{2} - 1$ doublet representations $\mathbf{2}_j$ with $j = 1, \ldots, \frac{n}{2} - 1$. It is straightforward to obtain the generators R and S in the singlet representations

  $$\begin{aligned} &\mathbf{1}_1 :\ R = S = 1\,, && \mathbf{1}_2 :\ R = 1,\ S = -1\,, \\ &\mathbf{1}_3 :\ R = -1,\ S = 1\,, && \mathbf{1}_4 :\ R = S = -1\,. \end{aligned} \tag{C.9}$$

  The explicit forms of these generators in the irreducible two-dimensional representations are

  $$\mathbf{2}_j :\ R = \begin{pmatrix} e^{2\pi i \frac{j}{n}} & 0 \\ 0 & e^{-2\pi i \frac{j}{n}} \end{pmatrix}\,, \quad S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \tag{C.10}$$

  with $j = 1, \ldots, \frac{n}{2} - 1$. Notice that the doublet representation $\mathbf{2}_j$ and the complex conjugate $\bar{\mathbf{2}}_j$ are equivalent, and they are related through a similarity transformation, i.e., $R^* = VRV^{-1}$ and $S^* = VSV^{-1}$ with $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Hence all the doublet representations of D_n are real, although the representation matrix of R is complex in our basis. Moreover, if $a = (a_1, a_2)^T$ is a doublet

transforming as $\mathbf{2}_j$, then $(a_2^*, a_1^*)^T$ transform as $\mathbf{2}_j$ under D_n as well.

The D_n group has a class-inverting outer automorphism $\mathfrak{u}$, and its action on the generators R and S is

$$R \overset{\mathfrak{u}}{\longmapsto} R^{-1}\,, \quad S \overset{\mathfrak{u}}{\longmapsto} S\,. \tag{C.11}$$

The CP transformation corresponding to $\mathfrak{u}$ is denoted by $X^0_{\mathbf{r}}$, its concrete form is determined by the following consistency conditions,

$$\begin{aligned} X^0_{\mathbf{r}}\rho^*_{\mathbf{r}}(R)X^{0\dagger}_{\mathbf{r}} &= \rho_{\mathbf{r}}\left(\mathfrak{u}\left(R\right)\right) = \rho_{\mathbf{r}}\left(R^{-1}\right)\,, \\ X^0_{\mathbf{r}}\rho^*_{\mathbf{r}}(S)X^{0\dagger}_{\mathbf{r}} &= \rho_{\mathbf{r}}\left(\mathfrak{u}\left(S\right)\right) = \rho_{\mathbf{r}}\left(S\right)\,. \end{aligned} \tag{C.12}$$

Since both $\rho_{\mathbf{r}}(R)$ and $\rho_{\mathbf{r}}(S)$ are symmetric and unitary, $X^0_{\mathbf{r}}$ coincides the canonical CP transformation up to an overall irrelevant phase, i.e.,

$$X^0_{\mathbf{r}} = \mathbb{1}\,. \tag{C.13}$$

From $X^0_{\mathbf{r}}$ and the flavor symmetry transformations of D_n, we can obtain other generalized CP transformations $X_{\mathbf{r}} = \rho_{\mathbf{r}}(g)X^0_{\mathbf{r}} = \rho_{\mathbf{r}}(g),\ g \in D_n$, yet they don't impose any new restrictions.

D Diagonalization of a 2×2 complex symmetric matrix

A generic complex symmetric 2×2 matrix can be written as

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix} , \tag{D.1}$$

where a, b and c are complex. It can be diagonalized by a two dimensional unitary matrix U via

$$U^T M U = \mathrm{diag}(\lambda_1, \lambda_2), \quad U = \begin{pmatrix} \cos\theta & e^{i\phi}\sin\theta \\ -e^{-i\phi}\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{-i\alpha} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} . \tag{D.2}$$

The eigenvalues $\lambda_{1,2}$ are non-negative with

$$\begin{aligned} \lambda_1^2 &= \frac{1}{2}\left[|a|^2 + |b|^2 + 2|c|^2 - \sqrt{(|b|^2 - |a|^2)^2 + 4|a^* c + b c^*|^2}\right] \\ \lambda_2^2 &= \frac{1}{2}\left[|a|^2 + |b|^2 + 2|c|^2 + \sqrt{(|b|^2 - |a|^2)^2 + 4|a^* c + b c^*|^2}\right] \end{aligned} \tag{D.3}$$

Without loss of generality, the rotation angle θ can be limited in the region $0 \leq \theta \leq \pi/2$ and it satisfies

$$\tan 2\theta = \frac{2|a^* c + b c^*|}{|b|^2 - |a|^2} . \tag{D.4}$$

The expressions of the phases ϕ, α and β are

$$\begin{aligned} \phi &= \arg(a^* c + b c^*) , \\ \alpha &= \frac{1}{2}\arg\left[a(|b|^2 - |\lambda_1|^2) - b^* c^2\right] , \\ \beta &= \frac{1}{2}\arg\left[b(|\lambda_2|^2 - |a|^2) + a^* c^2\right] . \end{aligned} \tag{D.5}$$

The general case of an arbitrary dimension is given in Ref. [30].

Acknowledgements

This review is based on the fruitful collaboration with many friends and colleagues worldwide. We have also benefited from useful discussions with many others. We express our sincere gratitude to all of them. This work was supported by the Spanish grants PID2020-113775GB-I00 (AEI/10.13039/501100011033) and Prometeo CIPROM/2021/054 and by the National Natural Science Foundation of China under Nos. 12375104, 11975224, 11835013.

References

[1] C.-N. Yang, R. L. Mills, Conservation of Isotopic Spin and Isotopic Gauge Invariance, Phys. Rev. 96 (1954) 191–195. doi:10.1103/PhysRev.96.191.

[2] S. L. Glashow, Partial Symmetries of Weak Interactions, Nucl. Phys. 22 (1961) 579–588. doi:10.1016/0029-5582(61)90469-2.

[3] S. Weinberg, A Model of Leptons, Phys. Rev. Lett. 19 (1967) 1264–1266. doi:10.1103/PhysRevLett.19.1264.

[4] A. Salam, Weak and Electromagnetic Interactions, Conf. Proc. C 680519 (1968) 367–377. doi:10.1142/9789812795915_0034.

[5] S. L. Glashow, J. Iliopoulos, L. Maiani, Weak Interactions with Lepton-Hadron Symmetry, Phys. Rev. D 2 (1970) 1285–1292. doi:10.1103/PhysRevD.2.1285.

[6] G. S. Guralnik, C. R. Hagen, T. W. B. Kibble, Global Conservation Laws and Massless Particles, Phys. Rev. Lett. 13 (1964) 585–587. doi:10.1103/PhysRevLett.13.585.

[7] F. Englert, R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, Phys. Rev. Lett. 13 (1964) 321–323. doi:10.1103/PhysRevLett.13.321.

[8] P. W. Higgs, Spontaneous Symmetry Breakdown without Massless Bosons, Phys. Rev. 145 (1966) 1156–1163. doi:10.1103/PhysRev.145.1156.

[9] G. Aad, et al., Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B 716 (2012) 1–29. arXiv:1207.7214, doi:10.1016/j.physletb.2012.08.020.

[10] S. Chatrchyan, et al., Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC, Phys. Lett. B 716 (2012) 30–61. arXiv:1207.7235, doi:10.1016/j.physletb.2012.08.021.

[11] A. Abada, et al., HE-LHC: The High-Energy Large Hadron Collider: Future Circular Collider Conceptual Design Report Volume 4, Eur. Phys. J. ST 228 (5) (2019) 1109–1382. doi:10.1140/epjst/e2019-900088-6.

[12] A. Abada, et al., FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1, Eur. Phys. J. C 79 (6) (2019) 474. doi:10.1140/epjc/s10052-019-6904-3.

[13] T. K. Charles, et al., The Compact Linear Collider (CLIC) - 2018 Summary Report 2/2018. arXiv:1812.06018, doi:10.23731/CYRM-2018-002.

[14] P. Bambade, et al., The International Linear Collider: A Global Projectarxiv:1903.01629.

[15] The International Linear Collider Technical Design Report - Volume 2: PhysicsarXiv:1306.6352.

[16] Physics and Detectors at CLIC: CLIC Conceptual Design ReportarXiv:1202.5940, doi:10.5170/CERN-2012-003.

[17] M. Dong, et al., CEPC Conceptual Design Report: Volume 2 - Physics & DetectorarXiv:1811.10545.

[18] A. B. McDonald, Nobel Lecture: The Sudbury Neutrino Observatory: Observation of flavor change for solar neutrinos, Rev. Mod. Phys. 88 (3) (2016) 030502. doi:10.1103/RevModPhys.88.030502.

[19] T. Kajita, Nobel Lecture: Discovery of atmospheric neutrino oscillations, Rev. Mod. Phys. 88 (3) (2016) 030501. doi:10.1103/RevModPhys.88.030501.

[20] K. Eguchi, et al., First results from KamLAND: Evidence for reactor anti-neutrino disappearance, Phys. Rev. Lett. 90 (2003) 021802. arXiv:hep-ex/0212021, doi:10.1103/PhysRevLett.90.021802.

[21] F. P. An, et al., Observation of electron-antineutrino disappearance at Daya Bay, Phys. Rev. Lett. 108 (2012) 171803. arXiv:1203.1669, doi:10.1103/PhysRevLett.108.171803.

[22] J. K. Ahn, et al., Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment, Phys. Rev. Lett. 108 (2012) 191802. arXiv:1204.0626, doi:10.1103/PhysRevLett.108.191802.

[23] K. Abe, et al., Indication of Electron Neutrino Appearance from an Accelerator-produced Off-axis Muon Neutrino Beam, Phys. Rev. Lett. 107 (2011) 041801. arXiv:1106.2822, doi:10.1103/PhysRevLett.107.041801.

[24] P. F. de Salas, D. V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C. A. Ternes, M. Tórtola, J. W. F. Valle, 2020 global reassessment of the neutrino oscillation picture, JHEP 02 (2021) 071. arXiv:2006.11237, doi:10.1007/JHEP02(2021)071.

[25] P. F. De Salas, et al., Chi2 profiles from Valencia neutrino global fit, http://globalfit.astroparticles.es/ (2021). doi:10.5281/zenodo.4726908.
URL https://doi.org/10.5281/zenodo.4726908

[26] F. Capozzi, E. Di Valentino, E. Lisi, A. Marrone, A. Melchiorri, A. Palazzo, Unfinished fabric of the three neutrino paradigm, Phys. Rev. D 104 (8) (2021) 083031. arXiv:2107.00532, doi:10.1103/PhysRevD.104.083031.

[27] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, A. Zhou, The fate of hints: updated global analysis of three-flavor neutrino oscillations, JHEP 09 (2020) 178. arXiv:2007.14792, doi:10.1007/JHEP09(2020)178.

[28] R. L. Workman, Others, Review of Particle Physics, PTEP 2022 (2022) 083C01. doi:10.1093/ptep/ptac097.

[29] S. Weinberg, Baryon and Lepton Nonconserving Processes, Phys. Rev. Lett. 43 (1979) 1566–1570. doi:10.1103/PhysRevLett.43.1566.

[30] J. Schechter, J. W. F. Valle, Neutrino Masses in SU(2) x U(1) Theories, Phys. Rev. D 22 (1980) 2227. doi:10.1103/PhysRevD.22.2227.

[31] T. P. Cheng, L.-F. Li, Neutrino Masses, Mixings and Oscillations in SU(2) x U(1) Models of Electroweak Interactions, Phys. Rev. D 22 (1980) 2860. doi:10.1103/PhysRevD.22.2860.

[32] J. Schechter, J. W. F. Valle, Neutrino Decay and Spontaneous Violation of Lepton Number, Phys. Rev. D 25 (1982) 774. doi:10.1103/PhysRevD.25.774.

[33] P. Minkowski, $\mu \to e\gamma$ at a Rate of One Out of 10^9 Muon Decays?, Phys. Lett. B 67 (1977) 421–428. doi:10.1016/0370-2693(77)90435-X.

[34] M. Gell-Mann, P. Ramond, R. Slansky, Complex Spinors and Unified Theories, Conf. Proc. C 790927 (1979) 315–321. arXiv:1306.4669.

[35] R. N. Mohapatra, G. Senjanovic, Neutrino Mass and Spontaneous Parity Nonconservation, Phys. Rev. Lett. 44 (1980) 912. doi:10.1103/PhysRevLett.44.912.

[36] M. Magg, C. Wetterich, Neutrino Mass Problem and Gauge Hierarchy, Phys. Lett. B 94 (1980) 61–64. doi:10.1016/0370-2693(80)90825-4.

[37] G. Lazarides, Q. Shafi, C. Wetterich, Proton Lifetime and Fermion Masses in an SO(10) Model, Nucl. Phys. B 181 (1981) 287–300. doi:10.1016/0550-3213(81)90354-0.

[38] R. N. Mohapatra, G. Senjanovic, Neutrino Masses and Mixings in Gauge Models with Spontaneous Parity Violation, Phys. Rev. D 23 (1981) 165. doi:10.1103/PhysRevD.23.165.

[39] Y. Chikashige, R. N. Mohapatra, R. D. Peccei, Are There Real Goldstone Bosons Associated with Broken Lepton Number?, Phys. Lett. B 98 (1981) 265–268. doi:10.1016/0370-2693(81)90011-3.

[40] M. C. Gonzalez-Garcia, J. W. F. Valle, Fast Decaying Neutrinos and Observable Flavor Violation in a New Class of Majoron Models, Phys. Lett. B 216 (1989) 360–366. doi:10.1016/0370-2693(89)91131-3.

[41] S. M. Boucenna, S. Morisi, J. W. F. Valle, The low-scale approach to neutrino masses, Adv. High Energy Phys. 2014 (2014) 831598. arXiv:1404.3751, doi:10.1155/2014/831598.

[42] Y. Cai, J. Herrero-García, M. A. Schmidt, A. Vicente, R. R. Volkas, From the trees to the forest: a review of radiative neutrino mass models, Front. in Phys. 5 (2017) 63. arXiv:1706.08524, doi:10.3389/fphy.2017.00063.

[43] H. Georgi, S. L. Glashow, Unity of All Elementary Particle Forces, Phys. Rev. Lett. 32 (1974) 438–441. doi:10.1103/PhysRevLett.32.438.

[44] H. Georgi, H. R. Quinn, S. Weinberg, Hierarchy of Interactions in Unified Gauge Theories, Phys. Rev. Lett. 33 (1974) 451–454. doi:10.1103/PhysRevLett.33.451.

[45] H. Fritzsch, P. Minkowski, Unified Interactions of Leptons and Hadrons, Annals Phys. 93 (1975) 193–266. doi:10.1016/0003-4916(75)90211-0.

[46] H. Georgi, C. Jarlskog, A New Lepton - Quark Mass Relation in a Unified Theory, Phys. Lett. B 86 (1979) 297–300. doi:10.1016/0370-2693(79)90842-6.

[47] S. Dimopoulos, H. Georgi, Softly Broken Supersymmetry and SU(5), Nucl. Phys. B 193 (1981) 150–162. doi:10.1016/0550-3213(81)90522-8.

[48] M. B. Green, J. H. Schwarz, E. Witten, SUPERSTRING THEORY. VOL. 1: INTRODUCTION, Cambridge Monographs on Mathematical Physics, 1988.

[49] J. Polchinski, String theory. Vol. 2: Superstring theory and beyond, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2007. doi:10.1017/CBO9780511618123.

[50] S. T. a. Dieter Lust, Lectures on String Theory, Lecture Notes in Physics 346, Springer Berlin Heidelberg, 1989.
URL http://gen.lib.rus.ec/book/index.php?md5=a63659a291b19c9e6abc9856b960f536

[51] L. J. Hall, H. Murayama, N. Weiner, Neutrino mass anarchy, Phys. Rev. Lett. 84 (2000) 2572–2575. arXiv:hep-ph/9911341, doi:10.1103/PhysRevLett.84.2572.

[52] A. de Gouvea, H. Murayama, Neutrino Mixing Anarchy: Alive and Kicking, Phys. Lett. B 747 (2015) 479–483. arXiv:1204.1249, doi:10.1016/j.physletb.2015.06.028.

[53] E. Ma, G. Rajasekaran, Softly broken A4 symmetry for nearly degenerate neutrino masses, Phys. Rev. D 64 (2001) 113012. arXiv:hep-ph/0106291, doi:10.1103/PhysRevD.64.113012.

[54] K. S. Babu, E. Ma, J. W. F. Valle, Underlying A4 symmetry for the neutrino mass matrix and the quark mixing matrix, Phys. Lett. B 552 (2003) 207–213. arXiv:hep-ph/0206292, doi:10.1016/S0370-2693(02)03153-2.

[55] M. Hirsch, J. C. Romao, S. Skadhauge, J. W. F. Valle, A. Villanova del Moral, Phenomenological tests of supersymmetric A(4) family symmetry model of neutrino mass, Phys. Rev. D 69 (2004) 093006. arXiv:hep-ph/0312265, doi:10.1103/PhysRevD.69.093006.

[56] E. Ma, H. Sawanaka, M. Tanimoto, Quark Masses and Mixing with A4 Family Symmetry, Phys. Lett. B 641 (2006) 301–304. arXiv:hep-ph/0606103, doi:10.1016/j.physletb.2006.08.062.

[57] M. Hirsch, S. Morisi, J. W. F. Valle, Modelling tri-bimaximal neutrino mixing, Phys. Rev. D 79 (2009) 016001. arXiv:0810.0121, doi:10.1103/PhysRevD.79.016001.

[58] M. Hirsch, S. Morisi, J. W. F. Valle, Tri-bimaximal neutrino mixing and neutrinoless double beta decay, Phys. Rev. D 78 (2008) 093007. arXiv:0804.1521, doi:10.1103/PhysRevD.78.093007.

[59] M. Hirsch, S. Morisi, J. W. F. Valle, A4-based tri-bimaximal mixing within inverse and linear seesaw schemes, Phys. Lett. B 679 (2009) 454–459. arXiv:0905.3056, doi:10.1016/j.physletb.2009.08.003.

[60] G.-J. Ding, D. Meloni, A Model for Tri-bimaximal Mixing from a Completely Broken A_4, Nucl. Phys. B 855 (2012) 21–45. arXiv:1108.2733, doi:10.1016/j.nuclphysb.2011.10.001.

[61] S. F. King, C. Luhn, Trimaximal neutrino mixing from vacuum alignment in A4 and S4 models, JHEP 09 (2011) 042. arXiv:1107.5332, doi:10.1007/JHEP09(2011)042.

[62] Y. Shimizu, M. Tanimoto, A. Watanabe, Breaking Tri-bimaximal Mixing and Large θ_{13}, Prog. Theor. Phys. 126 (2011) 81–90. arXiv:1105.2929, doi:10.1143/PTP.126.81.

[63] M. S. Boucenna, M. Hirsch, S. Morisi, E. Peinado, M. Taoso, J. W. F. Valle, Phenomenology of Dark Matter from A_4 Flavor Symmetry, JHEP 05 (2011) 037. arXiv:1101.2874, doi:10.1007/JHEP05(2011)037.

[64] I. de Medeiros Varzielas, L. Lavoura, Flavour models for TM_1 lepton mixing, J. Phys. G 40 (2013) 085002. arXiv:1212.3247, doi:10.1088/0954-3899/40/8/085002.

[65] S. Morisi, M. Nebot, K. M. Patel, E. Peinado, J. W. F. Valle, Quark-Lepton Mass Relation and CKM mixing in an A4 Extension of the Minimal Supersymmetric Standard Model, Phys. Rev. D 88 (2013) 036001. arXiv:1303.4394, doi:10.1103/PhysRevD.88.036001.

[66] S. F. King, S. Morisi, E. Peinado, J. W. F. Valle, Quark-Lepton Mass Relation in a Realistic A_4 Extension of the Standard Model, Phys. Lett. B 724 (2013) 68–72. arXiv:1301.7065, doi:10.1016/j.physletb.2013.05.067.

[67] S. Morisi, D. V. Forero, J. C. Romão, J. W. F. Valle, Neutrino mixing with revamped A_4 flavor symmetry, Phys. Rev. D 88 (1) (2013) 016003. arXiv:1305.6774, doi:10.1103/PhysRevD.88.016003.

[68] M. Hirsch, A. Villanova del Moral, J. W. F. Valle, E. Ma, Predicting neutrinoless double beta decay, Phys. Rev. D 72 (2005) 091301, [Erratum: Phys.Rev.D 72, 119904 (2005)]. arXiv:hep-ph/0507148, doi:10.1103/PhysRevD.72.091301.

[69] M. Hirsch, A. S. Joshipura, S. Kaneko, J. W. F. Valle, Predictive flavour symmetries of the neutrino mass matrix, Phys. Rev. Lett. 99 (2007) 151802. arXiv:hep-ph/0703046, doi:10.1103/PhysRevLett.99.151802.

[70] G. Altarelli, F. Feruglio, Tri-bimaximal neutrino mixing from discrete symmetry in extra dimensions, Nucl. Phys. B 720 (2005) 64–88. arXiv:hep-ph/0504165, doi:10.1016/j.nuclphysb.2005.05.005.

[71] G. Altarelli, F. Feruglio, C. Hagedorn, A SUSY SU(5) Grand Unified Model of Tri-Bimaximal Mixing from A_4, JHEP 03 (2008) 052. arXiv:0802.0090, doi:10.1088/1126-6708/2008/03/052.

[72] C. Csaki, C. Delaunay, C. Grojean, Y. Grossman, A Model of Lepton Masses from a Warped Extra Dimension, JHEP 10 (2008) 055. arXiv:0806.0356, doi:10.1088/1126-6708/2008/10/055.

[73] M.-C. Chen, K. T. Mahanthappa, F. Yu, A Viable Randall-Sundrum Model for Quarks and Leptons with T-prime Family Symmetry, Phys. Rev. D 81 (2010) 036004. arXiv:0907.3963, doi:10.1103/PhysRevD.81.036004.

[74] T. J. Burrows, S. F. King, A(4) Family Symmetry from SU(5) SUSY GUTs in 6d, Nucl. Phys. B 835 (2010) 174–196. arXiv:0909.1433, doi:10.1016/j.nuclphysb.2010.04.002.

[75] A. Kadosh, E. Pallante, An A(4) flavor model for quarks and leptons in warped geometry, JHEP 08 (2010) 115. arXiv:1004.0321, doi:10.1007/JHEP08(2010)115.

[76] A. Kadosh, E. Pallante, CP violation and FCNC in a warped A_4 flavor model, JHEP 06 (2011) 121. arXiv:1101.5420, doi:10.1007/JHEP06(2011)121.

[77] A. Kadosh, $\Theta_1 3$ and charged Lepton Flavor Violation in "warped" A_4 models, JHEP 06 (2013) 114. arXiv:1303.2645, doi:10.1007/JHEP06(2013)114.

[78] P. Chen, G.-J. Ding, A. D. Rojas, C. A. Vaquera-Araujo, J. W. F. Valle, Warped flavor symmetry predictions for neutrino physics, JHEP 01 (2016) 007. arXiv:1509.06683, doi:10.1007/JHEP01(2016)007.

[79] P. Chen, et al., Predictions from warped flavor dynamics based on the T' family group, Phys. Rev. D 102 (9) (2020) 095014. arXiv:2003.02734, doi:10.1103/PhysRevD.102.095014.

[80] F. del Aguila, A. Carmona, J. Santiago, Neutrino Masses from an A4 Symmetry in Holographic Composite Higgs Models, JHEP 08 (2010) 127. arXiv:1001.5151, doi:10.1007/JHEP08(2010)127.

[81] C. Hagedorn, M. Serone, Leptons in Holographic Composite Higgs Models with Non-Abelian Discrete Symmetries, JHEP 10 (2011) 083. arXiv:1106.4021, doi:10.1007/JHEP10(2011)083.

[82] G.-J. Ding, Y.-L. Zhou, Dirac Neutrinos with S_4 Flavor Symmetry in Warped Extra Dimensions, Nucl. Phys. B 876 (2013) 418–452. arXiv:1304.2645, doi:10.1016/j.nuclphysb.2013.08.011.

[83] C. Hagedorn, M. Serone, General Lepton Mixing in Holographic Composite Higgs Models, JHEP 02 (2012) 077. arXiv:1110.4612, doi:10.1007/JHEP02(2012)077.

[84] A. E. Cárcamo Hernández, I. de Medeiros Varzielas, N. A. Neill, Novel Randall-Sundrum model with S_3 flavor symmetry, Phys. Rev. D 94 (3) (2016) 033011. arXiv:1511.07420, doi:10.1103/PhysRevD.94.033011.

[85] H. Ishimori, T. Kobayashi, H. Ohki, Y. Shimizu, H. Okada, M. Tanimoto, Non-Abelian Discrete Symmetries in Particle Physics, Prog. Theor. Phys. Suppl. 183 (2010) 1–163. arXiv:1003.3552, doi:10.1143/PTPS.183.1.

[86] G. Altarelli, F. Feruglio, Discrete Flavor Symmetries and Models of Neutrino Mixing, Rev. Mod. Phys. 82 (2010) 2701–2729. arXiv:1002.0211, doi:10.1103/RevModPhys.82.2701.

[87] S. Morisi, J. W. F. Valle, Neutrino masses and mixing: a flavour symmetry roadmap, Fortsch. Phys. 61 (2013) 466–492. arXiv:1206.6678, doi:10.1002/prop.201200125.

[88] S. F. King, C. Luhn, Neutrino Mass and Mixing with Discrete Symmetry, Rept. Prog. Phys. 76 (2013) 056201. arXiv:1301.1340, doi:10.1088/0034-4885/76/5/056201.

[89] S. F. King, A. Merle, S. Morisi, Y. Shimizu, M. Tanimoto, Neutrino Mass and Mixing: from Theory to Experiment, New J. Phys. 16 (2014) 045018. arXiv:1402.4271, doi:10.1088/1367-2630/16/4/045018.

[90] S. F. King, Unified Models of Neutrinos, Flavour and CP Violation, Prog. Part. Nucl. Phys. 94 (2017) 217–256. arXiv:1701.04413, doi:10.1016/j.ppnp.2017.01.003.

[91] F. Feruglio, A. Romanino, Lepton flavor symmetries, Rev. Mod. Phys. 93 (1) (2021) 015007. arXiv:1912.06028, doi:10.1103/RevModPhys.93.015007.

[92] Z.-z. Xing, Flavor structures of charged fermions and massive neutrinos, Phys. Rept. 854 (2020) 1–147. arXiv:1909.09610, doi:10.1016/j.physrep.2020.02.001.

[93] G. Chauhan, P. S. B. Dev, I. Dubovyk, B. Dziewit, W. Flieger, K. Grzanka, J. Gluza, B. Karmakar, S. Zieba, Phenomenology of Lepton Masses and Mixing with Discrete Flavor Symmetries, arXiv:2310.20681.

[94] M. E. Peskin, D. V. Schroeder, An Introduction to quantum field theory, Addison-Wesley, Reading, USA, 1995.

[95] C. P. Burgess, G. D. Moore, The standard model: A primer, Cambridge University Press, 2006.

[96] J. F. Donoghue, E. Golowich, B. R. Holstein, Dynamics of the standard model, Vol. 2, CUP, 2014. doi:10.1017/CBO9780511524370.

[97] N. Cabibbo, Unitary Symmetry and Leptonic Decays, Phys. Rev. Lett. 10 (1963) 531–533. doi:10.1103/PhysRevLett.10.531.

[98] M. Kobayashi, T. Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, Prog. Theor. Phys. 49 (1973) 652–657. doi:10.1143/PTP.49.652.

[99] W. Rodejohann, J. W. F. Valle, Symmetrical Parametrizations of the Lepton Mixing Matrix, Phys. Rev. D 84 (2011) 073011. arXiv:1108.3484, doi:10.1103/PhysRevD.84.073011.

[100] M. Bona, et al., The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of Delta m(s) and BR(B —> tau nu(tau)), JHEP 10 (2006) 081. arXiv:hep-ph/0606167, doi:10.1088/1126-6708/2006/10/081.

[101] M. Bona, et al., Utfit website.
URL http://www.utfit.org/UTfit/

[102] J. W. F. Valle, J. C. Romao, Neutrinos in high energy and astroparticle physics, Physics textbook, Wiley-VCH, Weinheim, 2015.
URL https://www.wiley.com/WileyCDA/WileyTitle/productCd-3527411976,descCd-buy.html

[103] M. Fukugita, T. Yanagida, Physics of neutrinos and applications to astrophysics, 2003.

[104] C. Giunti, C. W. Kim, Fundamentals of Neutrino Physics and Astrophysics, 2007.

[105] Z.-z. Xing, S. Zhou, Neutrinos in particle physics, astronomy and cosmology, 2011.

[106] B. J. P. Jones, The Physics of Neutrinoless Double Beta Decay: A Primer, in: Theoretical Advanced Study Institute in Elementary Particle Physics: The Obscure Universe: Neutrinos and Other Dark Matters, 2021. arXiv:2108.09364.

[107] V. Cirigliano, et al., Neutrinoless Double-Beta Decay: A Roadmap for Matching Theory to ExperimentarXiv:2203.12169.

[108] M. J. Dolinski, A. W. P. Poon, W. Rodejohann, Neutrinoless Double-Beta Decay: Status and Prospects, Ann. Rev. Nucl. Part. Sci. 69 (2019) 219–251. arXiv:1902.04097, doi:10.1146/annurev-nucl-101918-023407.

[109] J. D. Vergados, H. Ejiri, F. Simkovic, Theory of Neutrinoless Double Beta Decay, Rept. Prog. Phys. 75 (2012) 106301. arXiv:1205.0649, doi:10.1088/0034-4885/75/10/106301.

[110] J. Schechter, J. W. F. Valle, Neutrinoless Double beta Decay in SU(2) x U(1) Theories, Phys. Rev. D 25 (1982) 2951. doi:10.1103/PhysRevD.25.2951.

[111] J. Schechter, J. W. F. Valle, Neutrino Oscillation Thought Experiment, Phys. Rev. D 23 (1981) 1666. doi:10.1103/PhysRevD.23.1666.

[112] S. M. Bilenky, J. Hosek, S. T. Petcov, On Oscillations of Neutrinos with Dirac and Majorana Masses, Phys. Lett. B 94 (1980) 495–498. doi:10.1016/0370-2693(80)90927-2.

[113] R. Acciarri, et al., Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNFarXiv:1512.06148.

[114] K. Abe, et al., Hyper-Kamiokande Design ReportarXiv:1805.04163.

[115] F. An, et al., Neutrino Physics with JUNO, J. Phys. G 43 (3) (2016) 030401. arXiv:1507.05613, doi:10.1088/0954-3899/43/3/030401.

[116] S. S. Chatterjee, P. Pasquini, J. W. F. Valle, Resolving the atmospheric octant by an improved measurement of the reactor angle, Phys. Rev. D 96 (1) (2017) 011303. arXiv:1703.03435, doi:10.1103/PhysRevD.96.011303.

[117] R. Srivastava, C. A. Ternes, M. Tórtola, J. W. F. Valle, Zooming in on neutrino oscillations with DUNE, Phys. Rev. D 97 (9) (2018) 095025. arXiv:1803.10247, doi:10.1103/PhysRevD.97.095025.

[118] M. S. Athar, et al., Status and perspectives of neutrino physics, Prog. Part. Nucl. Phys. 124 (2022) 103947. arXiv:2111.07586, doi:10.1016/j.ppnp.2022.103947.

[119] M. Maltoni, T. Schwetz, M. A. Tortola, J. W. F. Valle, Status of global fits to neutrino oscillations, New J. Phys. 6 (2004) 122. arXiv:hep-ph/0405172, doi:10.1088/1367-2630/6/1/122.

[120] M. Drewes, et al., A White Paper on keV Sterile Neutrino Dark Matter, JCAP 01 (2017) 025. arXiv:1602.04816, doi:10.1088/1475-7516/2017/01/025.

[121] C. Giunti, T. Lasserre, eV-scale Sterile Neutrinos, Ann. Rev. Nucl. Part. Sci. 69 (2019) 163–190. arXiv:1901.08330, doi:10.1146/annurev-nucl-101918-023755.

[122] S. Böser, C. Buck, C. Giunti, J. Lesgourgues, L. Ludhova, S. Mertens, A. Schukraft, M. Wurm, Status of Light Sterile Neutrino Searches, Prog. Part. Nucl. Phys. 111 (2020) 103736. arXiv:1906.01739, doi:10.1016/j.ppnp.2019.103736.

[123] B. Dasgupta, J. Kopp, Sterile Neutrinos, Phys. Rept. 928 (2021) 1–63. arXiv:2106.05913, doi:10.1016/j.physrep.2021.06.002.

[124] M. Lattanzi, M. Gerbino, K. Freese, G. Kane, J. W. F. Valle, Cornering (quasi) degenerate neutrinos with cosmology, JHEP 10 (2020) 213. arXiv:2007.01650, doi:10.1007/JHEP10(2020)213.

[125] S. Abe, et al., Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen, Phys. Rev. Lett. 130 (5) (2023) 051801. arXiv:2203.02139, doi:10.1103/PhysRevLett.130.051801.

[126] N. Abgrall, et al., The Large Enriched Germanium Experiment for Neutrinoless $\beta\beta$ Decay: LEGEND-1000 Preconceptual Design Reportarxiv:2107.11462.

[127] G. Adhikari, et al., nEXO: neutrinoless double beta decay search beyond 10^{28} year half-life sensitivity, J. Phys. G 49 (1) (2022) 015104. arXiv:2106.16243, doi:10.1088/1361-6471/ac3631.

[128] N. Aghanim, et al., Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. 641 (2020) A6, [Erratum: Astron.Astrophys. 652, C4 (2021)]. arXiv:1807.06209, doi:10.1051/0004-6361/201833910.

[129] M. Gerbino, et al., Synergy between cosmological and laboratory searches in neutrino physics, Phys. Dark Univ. 42 (2023) 101333. arXiv:2203.07377, doi:10.1016/j.dark.2023.101333.

[130] J. Leite, et al., A theory for scotogenic dark matter stabilised by residual gauge symmetry, Phys. Lett. B 802 (2020) 135254. arXiv:1909.06386, doi:10.1016/j.physletb.2020.135254.

[131] S. F. King, Large mixing angle MSW and atmospheric neutrinos from single right-handed neutrino dominance and U(1) family symmetry, Nucl. Phys. B 576 (2000) 85–105. arXiv:hep-ph/9912492, doi:10.1016/S0550-3213(00)00109-7.

[132] P. H. Frampton, S. L. Glashow, T. Yanagida, Cosmological sign of neutrino CP violation, Phys. Lett. B 548 (2002) 119–121. arXiv:hep-ph/0208157, doi:10.1016/S0370-2693(02)02853-8.

[133] M. Raidal, A. Strumia, Predictions of the most minimal seesaw model, Phys. Lett. B 553 (2003) 72–78. arXiv:hep-ph/0210021, doi:10.1016/S0370-2693(02)03124-6.

[134] M. Reig, D. Restrepo, J. W. F. Valle, O. Zapata, Bound-state dark matter with Majorana neutrinos, Phys. Lett. B 790 (2019) 303–307. arXiv:1806.09977, doi:10.1016/j.physletb.2019.01.023.

[135] D. M. Barreiros, R. G. Felipe, F. R. Joaquim, Combining texture zeros with a remnant CP symmetry in the minimal type-I seesaw, JHEP 01 (2019) 223. arXiv:1810.05454, doi:10.1007/JHEP01(2019)223.

[136] N. Rojas, R. Srivastava, J. W. F. Valle, Simplest Scoto-Seesaw Mechanism, Phys. Lett. B 789 (2019) 132–136. arXiv:1807.11447, doi:10.1016/j.physletb.2018.12.014.

[137] S. Mandal, R. Srivastava, J. W. F. Valle, The simplest scoto-seesaw model: WIMP dark matter phenomenology and Higgs vacuum stability, Phys. Lett. B 819 (2021) 136458. arXiv:2104.13401, doi:10.1016/j.physletb.2021.136458.

[138] I. M. Ávila, et al., Phenomenology of scotogenic scalar dark matter, Eur. Phys. J. C 80 (10) (2020) 908. arXiv:1910.08422, doi:10.1140/epjc/s10052-020-08480-z.

[139] M. Agostini, et al., Probing Majorana neutrinos with double-β decay, Science 365 (2019) 1445. arXiv:1909.02726, doi:10.1126/science.aav8613.

[140] M. Agostini, G. Benato, J. A. Detwiler, J. Menéndez, F. Vissani, Toward the discovery of matter creation with neutrinoless $\beta\beta$ decay, Rev. Mod. Phys. 95 (2) (2023) 025002. arXiv:2202.01787, doi:10.1103/RevModPhys.95.025002.

[141] L. Dorame, D. Meloni, S. Morisi, E. Peinado, J. W. F. Valle, Constraining Neutrinoless Double Beta Decay, Nucl. Phys. B 861 (2012) 259–270. arXiv:1111.5614, doi:10.1016/j.nuclphysb.2012.04.003.

[142] L. Dorame, S. Morisi, E. Peinado, J. W. F. Valle, A. D. Rojas, A new neutrino mass sum rule from inverse seesaw, Phys. Rev. D 86 (2012) 056001. arXiv:1203.0155, doi:10.1103/PhysRevD.86.056001.

[143] F. J. de Anda, J. W. Valle, C. A. Vaquera-Araujo, Flavour and CP predictions from orbifold compactification, Phys.Lett. B801 (2020) 135195. arXiv:1910.05605, doi:10.1016/j.physletb.2019.135195.

[144] F. J. de Anda, et al., Probing the predictions of an orbifold theory of flavor, Phys. Rev. D 101 (11) (2020) 116012. arXiv:2004.06735, doi:10.1103/PhysRevD.101.116012.

[145] J. A. Formaggio, A. L. C. de Gouvêa, R. G. H. Robertson, Direct Measurements of Neutrino Mass, Phys. Rept. 914 (2021) 1–54. arXiv:2102.00594, doi:10.1016/j.physrep.2021.02.002.

[146] M. Aker, et al., Direct neutrino-mass measurement with sub-electronvolt sensitivity, Nature Phys. 18 (2) (2022) 160–166. arXiv:2105.08533, doi:10.1038/s41567-021-01463-1.

[147] M. Aker, et al., Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN, Phys. Rev. Lett. 123 (22) (2019) 221802. arXiv:1909.06048, doi:10.1103/PhysRevLett.123.221802.

[148] D. M. Barreiros, F. R. Joaquim, R. Srivastava, J. W. F. Valle, Minimal scoto-seesaw mechanism with spontaneous CP violation, JHEP 04 (2021) 249. arXiv:2012.05189, doi:10.1007/JHEP04(2021)249.

[149] A. Gando, et al., Search for Majorana Neutrinos near the Inverted Mass Hierarchy Region with KamLAND-Zen, Phys. Rev. Lett. 117 (8) (2016) 082503, [Addendum: Phys.Rev.Lett. 117, 109903 (2016)]. arXiv:1605.02889, doi:10.1103/PhysRevLett.117.082503.

[150] M. Agostini, et al., Final Results of GERDA on the Search for Neutrinoless Double-β Decay, Phys. Rev. Lett. 125 (25) (2020) 252502. arXiv:2009.06079, doi:10.1103/PhysRevLett.125.252502.

[151] D. Q. Adams, et al., Improved Limit on Neutrinoless Double-Beta Decay in ^{130}Te with CUORE, Phys. Rev. Lett. 124 (12) (2020) 122501. arXiv:1912.10966, doi:10.1103/PhysRevLett.124.122501.

[152] G. Anton, et al., Search for Neutrinoless Double-β Decay with the Complete EXO-200 Dataset, Phys. Rev. Lett. 123 (16) (2019) 161802. arXiv:1906.02723, doi:10.1103/PhysRevLett.123.161802.

[153] M. Duerr, M. Lindner, A. Merle, On the Quantitative Impact of the Schechter-Valle Theorem, JHEP 06 (2011) 091. arXiv:1105.0901, doi:10.1007/JHEP06(2011)091.

[154] L. Gráf, M. Lindner, O. Scholer, Unraveling the $0\nu\beta\beta$ decay mechanisms, Phys. Rev. D 106 (3) (2022) 035022. arXiv:2204.10845, doi:10.1103/PhysRevD.106.035022.

[155] J. Heeck, W. Rodejohann, Neutrinoless Quadruple Beta Decay, EPL 103 (3) (2013) 32001. arXiv:1306.0580, doi:10.1209/0295-5075/103/32001.

[156] R. Arnold, et al., Search for neutrinoless quadruple-β decay of ^{150}Nd with the NEMO-3 detector, Phys. Rev. Lett. 119 (4) (2017) 041801. arXiv:1705.08847, doi:10.1103/PhysRevLett.119.041801.

[157] M. Hirsch, R. Srivastava, J. W. F. Valle, Can one ever prove that neutrinos are Dirac particles?, Phys. Lett. B 781 (2018) 302–305. arXiv:1711.06181, doi:10.1016/j.physletb.2018.03.073.

[158] R. E. Marshak, R. N. Mohapatra, Quark - Lepton Symmetry and B-L as the U(1) Generator of the Electroweak Symmetry Group, Phys. Lett. B 91 (1980) 222–224. doi:10.1016/0370-2693(80)90436-0.

[159] R. N. Mohapatra, R. E. Marshak, Local B-L Symmetry of Electroweak Interactions, Majorana Neutrinos and Neutron Oscillations, Phys. Rev. Lett. 44 (1980) 1316–1319, [Erratum: Phys.Rev.Lett. 44, 1643 (1980)]. doi:10.1103/PhysRevLett.44.1316.

[160] C. Wetterich, Neutrino Masses and the Scale of B-L Violation, Nucl. Phys. B 187 (1981) 343–375. doi:10.1016/0550-3213(81)90279-0.

[161] J. Leite, et al., Scotogenic dark matter and Dirac neutrinos from unbroken gauged B − L symmetry, Phys. Lett. B 807 (2020) 135537. arXiv:2003.02950, doi:10.1016/j.physletb.2020.135537.

[162] E. Peinado, M. Reig, R. Srivastava, J. W. F. Valle, Dirac neutrinos from Peccei–Quinn symmetry: A fresh look at the axion, Mod. Phys. Lett. A 35 (21) (2020) 2050176. arXiv:1910.02961, doi:10.1142/S021773232050176X.

[163] A. G. Dias, J. Leite, J. W. F. Valle, C. A. Vaquera-Araujo, Reloading the axion in a 3-3-1 setup, Phys. Lett. B 810 (2020) 135829. arXiv:2008.10650, doi:10.1016/j.physletb.2020.135829.

[164] Y. Farzan, E. Ma, Dirac neutrino mass generation from dark matter, Phys. Rev. D 86 (2012) 033007. arXiv:1204.4890, doi:10.1103/PhysRevD.86.033007.

[165] S. Centelles Chuliá, E. Ma, R. Srivastava, J. W. F. Valle, Dirac Neutrinos and Dark Matter Stability from Lepton Quarticity, Phys.Lett. B767 (2017) 209–213. arXiv:1606.04543, doi:10.1016/j.physletb.2017.01.070.

[166] C. Bonilla, E. Ma, E. Peinado, J. W. F. Valle, Two-loop Dirac neutrino mass and WIMP dark matter, Phys. Lett. B 762 (2016) 214–218. arXiv:1607.03931, doi:10.1016/j.physletb.2016.09.027.

[167] R. Srivastava, C. A. Ternes, M. Tórtola, J. W. F. Valle, Testing a lepton quarticity flavor theory of neutrino oscillations with the DUNE experiment, Phys. Lett. B 778 (2018) 459–463. arXiv:1711.10318, doi:10.1016/j.physletb.2018.01.014.

[168] S. Centelles Chuliá, R. Srivastava, J. W. F. Valle, Generalized Bottom-Tau unification, neutrino oscillations and dark matter: predictions from a lepton quarticity flavor approach, Phys. Lett. B 773 (2017) 26–33. arXiv:1706.00210, doi:10.1016/j.physletb.2017.07.065.

[169] M. Reig, D. Restrepo, J. W. F. Valle, O. Zapata, Bound-state dark matter and Dirac neutrino masses, Phys. Rev. D 97 (11) (2018) 115032. arXiv:1803.08528, doi:10.1103/PhysRevD.97.115032.

[170] C. Bonilla, S. Centelles-Chuliá, R. Cepedello, E. Peinado, R. Srivastava, Dark matter stability and Dirac neutrinos using only Standard Model symmetries, Phys. Rev. D 101 (3) (2020) 033011. arXiv:1812.01599, doi:10.1103/PhysRevD.101.033011.

[171] C. Alvarado, C. Bonilla, J. Leite, J. W. F. Valle, Phenomenology of fermion dark matter as neutrino mass mediator with gauged B-L, Phys. Lett. B 817 (2021) 136292. arXiv:2102.07216, doi:10.1016/j.physletb.2021.136292.

[172] M. Roncadelli, D. Wyler, Naturally Light Dirac Neutrinos in Gauge Theories, Phys. Lett. B 133 (1983) 325–329. doi:10.1016/0370-2693(83)90156-9.

[173] P.-H. Gu, H.-J. He, Neutrino Mass and Baryon Asymmetry from Dirac Seesaw, JCAP 12 (2006) 010. arXiv:hep-ph/0610275, doi:10.1088/1475-7516/2006/12/010.

[174] C.-Y. Yao, G.-J. Ding, Systematic analysis of Dirac neutrino masses from a dimension five operator, Phys. Rev. D 97 (9) (2018) 095042. arXiv:1802.05231, doi:10.1103/PhysRevD.97.095042.

[175] S. Centelles Chuliá, R. Srivastava, J. W. F. Valle, Seesaw roadmap to neutrino mass and dark matter, Phys. Lett. B 781 (2018) 122–128. arXiv:1802.05722, doi:10.1016/j.physletb.2018.03.046.

[176] C.-Y. Yao, G.-J. Ding, Systematic Study of One-Loop Dirac Neutrino Masses and Viable Dark Matter Candidates, Phys. Rev. D 96 (9) (2017) 095004, [Erratum: Phys.Rev.D 98, 039901 (2018)]. arXiv:1707.09786, doi:10.1103/PhysRevD.96.095004.

[177] S. Centelles Chuliá, R. Srivastava, J. W. F. Valle, Seesaw Dirac neutrino mass through dimension-six operators, Phys. Rev. D 98 (3) (2018) 035009. arXiv:1804.03181, doi:10.1103/PhysRevD.98.035009.

[178] J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, A. Riotto, A. Strumia, Higgs mass implications on the stability of the electroweak vacuum, Phys. Lett. B 709 (2012) 222–228. arXiv:1112.3022, doi:10.1016/j.physletb.2012.02.013.

[179] C. Bonilla, R. M. Fonseca, J. W. F. Valle, Consistency of the triplet seesaw model revisited, Phys. Rev. D 92 (7) (2015) 075028. arXiv:1508.02323, doi:10.1103/PhysRevD.92.075028.

[180] C. Bonilla, R. M. Fonseca, J. W. F. Valle, Vacuum stability with spontaneous violation of lepton number, Phys. Lett. B 756 (2016) 345–349. arXiv:1506.04031, doi:10.1016/j.physletb.2016.03.037.

[181] S. Mandal, R. Srivastava, J. W. F. Valle, Consistency of the dynamical high-scale type-I seesaw mechanism, Phys. Rev. D 101 (11) (2020) 115030. doi:10.1103/PhysRevD.101.115030.

[182] S. Mandal, R. Srivastava, J. W. F. Valle, Electroweak symmetry breaking in the inverse seesaw mechanism, JHEP 03 (2021) 212. arXiv:2009.10116, doi:10.1007/JHEP03(2021)212.

[183] S. Mandal, J. C. Romão, R. Srivastava, J. W. F. Valle, Dynamical inverse seesaw mechanism as a simple benchmark for electroweak breaking and Higgs boson studies, JHEP 07 (2021) 029. arXiv:2103.02670, doi:10.1007/JHEP07(2021)029.

[184] R. Foot, H. Lew, X. G. He, G. C. Joshi, Seesaw Neutrino Masses Induced by a Triplet of Leptons, Z. Phys. C 44 (1989) 441. doi:10.1007/BF01415558.

[185] S. Mandal, O. G. Miranda, G. Sanchez Garcia, J. W. F. Valle, X.-J. Xu, Toward deconstructing the simplest seesaw mechanism, Phys. Rev. D 105 (9) (2022) 095020. arXiv:2203.06362, doi:10.1103/PhysRevD.105.095020.

[186] P. Fileviez Perez, T. Han, G.-y. Huang, T. Li, K. Wang, Neutrino Masses and the CERN LHC: Testing Type II Seesaw, Phys. Rev. D 78 (2008) 015018. arXiv:0805.3536, doi:10.1103/PhysRevD.78.015018.

[187] Y. Cai, T. Han, T. Li, R. Ruiz, Lepton Number Violation: Seesaw Models and Their Collider Tests, Front. in Phys. 6 (2018) 40. arXiv:1711.02180, doi:10.3389/fphy.2018.00040.

[188] S. Mandal, O. G. Miranda, G. S. Garcia, J. W. F. Valle, X.-J. Xu, High-energy colliders as a probe of neutrino properties, Phys. Lett. B 829 (2022) 137110. arXiv:2202.04502, doi:10.1016/j.physletb.2022.137110.

[189] J. W. F. Valle, Neutrino physics overview, J. Phys. Conf. Ser. 53 (2006) 473–505. arXiv:hep-ph/0608101, doi:10.1088/1742-6596/53/1/031.

[190] J. W. F. Valle, C. A. Vaquera-Araujo, Dynamical seesaw mechanism for Dirac neutrinos, Phys. Lett. B 755 (2016) 363–366. arXiv:1601.05237, doi:10.1016/j.physletb.2016.02.031.

[191] A. Aranda, C. Bonilla, S. Morisi, E. Peinado, J. W. F. Valle, Dirac neutrinos from flavor symmetry, Phys. Rev. D 89 (3) (2014) 033001. arXiv:1307.3553, doi:10.1103/PhysRevD.89.033001.

[192] D. Borah, B. Karmakar, A_4 flavour model for Dirac neutrinos: Type I and inverse seesaw, Phys. Lett. B 780 (2018) 461–470. arXiv:1712.06407, doi:10.1016/j.physletb.2018.03.047.

[193] G. Bertone, D. Hooper, J. Silk, Particle dark matter: Evidence, candidates and constraints, Phys. Rept. 405 (2005) 279–390. arXiv:hep-ph/0404175, doi:10.1016/j.physrep.2004.08.031.

[194] G. Jungman, M. Kamionkowski, K. Griest, Supersymmetric dark matter, Phys. Rept. 267 (1996) 195–373. arXiv:hep-ph/9506380, doi:10.1016/0370-1573(95)00058-5.

[195] Z.-j. Tao, Radiative seesaw mechanism at weak scale, Phys. Rev. D 54 (1996) 5693–5697. arXiv:hep-ph/9603309, doi:10.1103/PhysRevD.54.5693.

[196] E. Ma, Verifiable radiative seesaw mechanism of neutrino mass and dark matter, Phys. Rev. D 73 (2006) 077301. arXiv:hep-ph/0601225, doi:10.1103/PhysRevD.73.077301.

[197] M. Hirsch, R. A. Lineros, S. Morisi, J. Palacio, N. Rojas, J. W. F. Valle, WIMP dark matter as radiative neutrino mass messenger, JHEP 10 (2013) 149. arXiv:1307.8134, doi:10.1007/JHEP10(2013)149.

[198] A. Merle, M. Platscher, N. Rojas, J. W. F. Valle, A. Vicente, Consistency of WIMP Dark Matter as radiative neutrino mass messenger, JHEP 07 (2016) 013. arXiv:1603.05685, doi:10.1007/JHEP07(2016)013.

[199] M. A. Díaz, N. Rojas, S. Urrutia-Quiroga, J. W. F. Valle, Heavy Higgs Boson Production at Colliders in the Singlet-Triplet Scotogenic Dark Matter Model, JHEP 08 (2017) 017. arXiv:1612.06569, doi:10.1007/JHEP08(2017)017.

[200] S. Choubey, S. Khan, M. Mitra, S. Mondal, Singlet-Triplet Fermionic Dark Matter and LHC Phenomenology, Eur. Phys. J. C 78 (4) (2018) 302. arXiv:1711.08888, doi:10.1140/epjc/s10052-018-5785-1.

[201] D. Restrepo, A. Rivera, Phenomenological consistency of the singlet-triplet scotogenic model, JHEP 04 (2020) 134. arXiv:1907.11938, doi:10.1007/JHEP04(2020)134.

[202] S. Mandal, N. Rojas, R. Srivastava, J. W. F. Valle, Dark matter as the origin of neutrino mass in the inverse seesaw mechanism, Phys. Lett. B 821 (2021) 136609. arXiv:1907.07728, doi:10.1016/j.physletb.2021.136609.

[203] A. E. Cárcamo Hernández, V. K. N., J. W. F. Valle, Linear seesaw mechanism from dark sector, JHEP 09 (2023) 046. arXiv:2305.02273, doi:10.1007/JHEP09(2023)046.

[204] A. Batra, H. B. Câmara, F. R. Joaquim, Dark linear seesaw mechanism, Phys. Lett. B 843 (2023) 138012. arXiv:2305.01687, doi:10.1016/j.physletb.2023.138012.

[205] G.-J. Ding, J.-N. Lu, J. W. F. Valle, Trimaximal neutrino mixing from scotogenic A_4 family symmetry, Phys. Lett. B 815 (2021) 136122. arXiv:2009.04750, doi:10.1016/j.physletb.2021.136122.

[206] V. Berezinsky, J. W. F. Valle, The KeV majoron as a dark matter particle, Phys. Lett. B 318 (1993) 360–366. arXiv:hep-ph/9309214, doi:10.1016/0370-2693(93)90140-D.

[207] M. Lattanzi, R. A. Lineros, M. Taoso, Connecting neutrino physics with dark matter, New J. Phys. 16 (12) (2014) 125012. arXiv:1406.0004, doi:10.1088/1367-2630/16/12/125012.

[208] S. R. Coleman, Why There Is Nothing Rather Than Something: A Theory of the Cosmological Constant, Nucl. Phys. B 310 (1988) 643–668. doi:10.1016/0550-3213(88)90097-1.

[209] M. Lattanzi, J. W. F. Valle, Decaying warm dark matter and neutrino masses, Phys. Rev. Lett. 99 (2007) 121301. arXiv:0705.2406, doi:10.1103/PhysRevLett.99.121301.

[210] M. Lattanzi, S. Riemer-Sorensen, M. Tortola, J. W. F. Valle, Updated CMB and x- and γ-ray constraints on Majoron dark matter, Phys. Rev. D 88 (6) (2013) 063528. arXiv:1303.4685, doi:10.1103/PhysRevD.88.063528.

[211] J.-L. Kuo, M. Lattanzi, K. Cheung, J. W. F. Valle, Decaying warm dark matter and structure formation, JCAP 12 (2018) 026. arXiv:1803.05650, doi:10.1088/1475-7516/2018/12/026.

[212] J. Schechter, J. W. F. Valle, Comment on the Lepton Mixing Matrix, Phys. Rev. D 21 (1980) 309. doi:10.1103/PhysRevD.21.309.

[213] R. N. Mohapatra, J. W. F. Valle, Neutrino Mass and Baryon Number Nonconservation in Superstring Models, Phys. Rev. D 34 (1986) 1642. doi:10.1103/PhysRevD.34.1642.

[214] J. Bernabeu, A. Santamaria, J. Vidal, A. Mendez, J. W. F. Valle, Lepton Flavor Nonconservation at High-Energies in a Superstring Inspired Standard Model, Phys. Lett. B 187 (1987) 303–308. doi:10.1016/0370-2693(87)91100-2.

[215] G. C. Branco, M. N. Rebelo, J. W. F. Valle, Leptonic CP Violation With Massless Neutrinos, Phys. Lett. B 225 (1989) 385–392. doi:10.1016/0370-2693(89)90587-X.

[216] N. Rius, J. W. F. Valle, Leptonic CP Violating Asymmetries in Z^0 Decays, Phys. Lett. B 246 (1990) 249–255. doi:10.1016/0370-2693(90)91341-8.

[217] H. Nunokawa, S. J. Parke, J. W. F. Valle, CP Violation and Neutrino Oscillations, Prog. Part. Nucl. Phys. 60 (2008) 338–402. arXiv:0710.0554, doi:10.1016/j.ppnp.2007.10.001.

[218] A. de Gouvea, B. Kayser, R. N. Mohapatra, Manifest CP Violation from Majorana Phases, Phys. Rev. D 67 (2003) 053004. arXiv:hep-ph/0211394, doi:10.1103/PhysRevD.67.053004.

[219] G. C. Branco, R. G. Felipe, F. R. Joaquim, Leptonic CP Violation, Rev. Mod. Phys. 84 (2012) 515–565. arXiv:1111.5332, doi:10.1103/RevModPhys.84.515.

[220] A. Vicente, Lepton flavor violation beyond the MSSM, Adv. High Energy Phys. 2015 (2015) 686572. arXiv:1503.08622, doi:10.1155/2015/686572.

[221] A. Abada, J. Kriewald, A. M. Teixeira, On the role of leptonic CPV phases in cLFV observables, Eur. Phys. J. C 81 (11) (2021) 1016. arXiv:2107.06313, doi:10.1140/epjc/s10052-021-09754-w.

[222] F. Deppisch, J. W. F. Valle, Enhanced lepton flavor violation in the supersymmetric inverse seesaw model, Phys. Rev. D 72 (2005) 036001. arXiv:hep-ph/0406040, doi:10.1103/PhysRevD.72.036001.

[223] F. Deppisch, T. S. Kosmas, J. W. F. Valle, Enhanced mu- - e- conversion in nuclei in the inverse seesaw model, Nucl. Phys. B 752 (2006) 80–92. arXiv:hep-ph/0512360, doi:10.1016/j.nuclphysb.2006.06.032.

[224] A. Ilakovac, A. Pilaftsis, Flavor violating charged lepton decays in seesaw-type models, Nucl. Phys. B 437 (1995) 491. arXiv:hep-ph/9403398, doi:10.1016/0550-3213(94)00567-X.

[225] E. Arganda, M. J. Herrero, A. M. Teixeira, mu-e conversion in nuclei within the CMSSM seesaw: Universality versus non-universality, JHEP 10 (2007) 104. arXiv:0707.2955, doi:10.1088/1126-6708/2007/10/104.

[226] A. Abada, V. De Romeri, S. Monteil, J. Orloff, A. M. Teixeira, Indirect searches for sterile neutrinos at a high-luminosity Z-factory, JHEP 04 (2015) 051. arXiv:1412.6322, doi:10.1007/JHEP04(2015)051.

[227] A. Abada, M. Lucente, Looking for the minimal inverse seesaw realisation, Nucl. Phys. B 885 (2014) 651–678. arXiv:1401.1507, doi:10.1016/j.nuclphysb.2014.06.003.

[228] A. Abada, V. De Romeri, A. M. Teixeira, Impact of sterile neutrinos on nuclear-assisted cLFV processes, JHEP 02 (2016) 083. arXiv:1510.06657, doi:10.1007/JHEP02(2016)083.

[229] E. K. Akhmedov, M. Lindner, E. Schnapka, J. W. F. Valle, Dynamical left-right symmetry breaking, Phys. Rev. D 53 (1996) 2752–2780. arXiv:hep-ph/9509255, doi:10.1103/PhysRevD.53.2752.

[230] E. K. Akhmedov, M. Lindner, E. Schnapka, J. W. F. Valle, Left-right symmetry breaking in NJL approach, Phys. Lett. B 368 (1996) 270–280. arXiv:hep-ph/9507275, doi:10.1016/0370-2693(95)01504-3.

[231] M. Malinsky, J. C. Romao, J. W. F. Valle, Novel supersymmetric SO(10) seesaw mechanism, Phys. Rev. Lett. 95 (2005) 161801. arXiv:hep-ph/0506296, doi:10.1103/PhysRevLett.95.161801.

[232] W.-Y. Keung, G. Senjanovic, Majorana Neutrinos and the Production of the Right-handed Charged Gauge Boson, Phys. Rev. Lett. 50 (1983) 1427. doi:10.1103/PhysRevLett.50.1427.

[233] M. Dittmar, A. Santamaria, M. C. Gonzalez-Garcia, J. W. F. Valle, Production Mechanisms and Signatures of Isosinglet Neutral Heavy Leptons in Z^0 Decays, Nucl. Phys. B 332 (1990) 1–19. doi:10.1016/0550-3213(90)90028-C.

[234] M. C. Gonzalez-Garcia, A. Santamaria, J. W. F. Valle, Isosinglet Neutral Heavy Lepton Production in Z Decays and Neutrino Mass, Nucl. Phys. B 342 (1990) 108–126. doi:10.1016/0550-3213(90)90573-V.

[235] J. A. Aguilar-Saavedra, F. Deppisch, O. Kittel, J. W. F. Valle, Flavour in heavy neutrino searches at the LHC, Phys. Rev. D 85 (2012) 091301. arXiv:1203.5998, doi:10.1103/PhysRevD.85.091301.

[236] S. P. Das, F. F. Deppisch, O. Kittel, J. W. F. Valle, Heavy Neutrinos and Lepton Flavour Violation in Left-Right Symmetric Models at the LHC, Phys. Rev. D 86 (2012) 055006. arXiv:1206.0256, doi:10.1103/PhysRevD.86.055006.

[237] F. F. Deppisch, N. Desai, J. W. F. Valle, Is charged lepton flavor violation a high energy phenomenon?, Phys. Rev. D 89 (5) (2014) 051302. arXiv:1308.6789, doi:10.1103/PhysRevD.89.051302.

[238] G. Aad, et al., Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector, JHEP 10 (2019) 265. arXiv:1905.09787, doi:10.1007/JHEP10(2019)265.

[239] A. M. Sirunyan, et al., Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. Lett. 120 (22) (2018) 221801. arXiv:1802.02965, doi:10.1103/PhysRevLett.120.221801.

[240] A. Tumasyan, et al., Inclusive nonresonant multilepton probes of new phenomena at $\sqrt{s}$=13 TeV, Phys. Rev. D 105 (11) (2022) 112007. arXiv:2202.08676, doi:10.1103/PhysRevD.105.112007.

[241] A. Abada, et al., FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2, Eur. Phys. J. ST 228 (2) (2019) 261–623. doi:10.1140/epjst/e2019-900045-4.

[242] J. L. Feng, et al., The Forward Physics Facility at the High-Luminosity LHC, J. Phys. G 50 (3) (2023) 030501. arXiv:2203.05090, doi:10.1088/1361-6471/ac865e.

[243] A. M. Abdullahi, et al., The present and future status of heavy neutral leptons, J. Phys. G 50 (2) (2023) 020501. arXiv:2203.08039, doi:10.1088/1361-6471/ac98f9.

[244] J. C. Helo, M. Hirsch, S. Kovalenko, Heavy neutrino searches at the LHC with displaced vertices, Phys. Rev. D 89 (2014) 073005, [Erratum: Phys.Rev.D 93, 099902 (2016)]. arXiv:1312.2900, doi:10.1103/PhysRevD.89.073005.

[245] M. Drewes, J. Hajer, Heavy Neutrinos in displaced vertex searches at the LHC and HL-LHC, JHEP 02 (2020) 070. arXiv:1903.06100, doi:10.1007/JHEP02(2020)070.

[246] R. Beltrán, G. Cottin, J. C. Helo, M. Hirsch, A. Titov, Z. S. Wang, Long-lived heavy neutral leptons at the LHC: four-fermion single-N_R operators, JHEP 01 (2022) 044. arXiv:2110.15096, doi:10.1007/JHEP01(2022)044.

[247] A. Datta, B. Mukhopadhyaya, F. Vissani, Tevatron signatures of an R-parity violating supersymmetric theory, Phys. Lett. B 492 (2000) 324–330. arXiv:hep-ph/9910296, doi:10.1016/S0370-2693(00)01104-7.

[248] W. Porod, M. Hirsch, J. Romao, J. W. F. Valle, Testing neutrino mixing at future collider experiments, Phys. Rev. D 63 (2001) 115004. arXiv:hep-ph/0011248, doi:10.1103/PhysRevD.63.115004.

[249] F. de Campos, O. J. P. Eboli, M. B. Magro, W. Porod, D. Restrepo, J. W. F. Valle, Probing neutrino mass with displaced vertices at the Tevatron, Phys. Rev. D 71 (2005) 075001. arXiv:hep-ph/0501153, doi:10.1103/PhysRevD.71.075001.

[250] A. D. Sakharov, Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe, Pisma Zh. Eksp. Teor. Fiz. 5 (1967) 32–35. doi:10.1070/PU1991v034n05ABEH002497.

[251] M. Fukugita, T. Yanagida, Baryogenesis Without Grand Unification, Phys. Lett. B 174 (1986) 45–47. doi:10.1016/0370-2693(86)91126-3.

[252] P. F. Harrison, W. G. Scott, Symmetries and generalizations of tri - bimaximal neutrino mixing, Phys. Lett. B 535 (2002) 163–169. arXiv:hep-ph/0203209, doi:10.1016/S0370-2693(02)01753-7.

[253] P. F. Harrison, W. G. Scott, mu - tau reflection symmetry in lepton mixing and neutrino oscillations, Phys. Lett. B 547 (2002) 219–228. arXiv:hep-ph/0210197, doi:10.1016/S0370-2693(02)02772-7.

[254] W. Grimus, L. Lavoura, A Nonstandard CP transformation leading to maximal atmospheric neutrino mixing, Phys. Lett. B 579 (2004) 113–122. arXiv:hep-ph/0305309, doi:10.1016/j.physletb.2003.10.075.

[255] P. F. Harrison, W. G. Scott, The Simplest neutrino mass matrix, Phys. Lett. B 594 (2004) 324–332. arXiv:hep-ph/0403278, doi:10.1016/j.physletb.2004.05.039.

[256] Y. Farzan, A. Y. Smirnov, Leptonic CP violation: Zero, maximal or between the two extremes, JHEP 01 (2007) 059. arXiv:hep-ph/0610337, doi:10.1088/1126-6708/2007/01/059.

[257] P. Chen, G.-J. Ding, F. Gonzalez-Canales, J. W. F. Valle, Generalized $\mu-\tau$ reflection symmetry and leptonic CP violation, Phys. Lett. B 753 (2016) 644–652. arXiv:1512.01551, doi:10.1016/j.physletb.2015.12.069.

[258] P. Chen, S. Centelles Chuliá, G.-J. Ding, R. Srivastava, J. W. F. Valle, Neutrino Predictions from Generalized CP Symmetries of Charged Leptons, JHEP 07 (2018) 077. arXiv:1802.04275, doi:10.1007/JHEP07(2018)077.

[259] R. de Adelhart Toorop, F. Feruglio, C. Hagedorn, Finite Modular Groups and Lepton Mixing, Nucl. Phys. B 858 (2012) 437–467. arXiv:1112.1340, doi:10.1016/j.nuclphysb.2012.01.017.

[260] M. Holthausen, K. S. Lim, M. Lindner, Lepton Mixing Patterns from a Scan of Finite Discrete Groups, Phys. Lett. B 721 (2013) 61–67. arXiv:1212.2411, doi:10.1016/j.physletb.2013.02.047.

[261] R. M. Fonseca, W. Grimus, Classification of lepton mixing matrices from finite residual symmetries, JHEP 09 (2014) 033. arXiv:1405.3678, doi:10.1007/JHEP09(2014)033.

[262] J. Talbert, [Re]constructing Finite Flavour Groups: Horizontal Symmetry Scans from the Bottom-Up, JHEP 12 (2014) 058. arXiv:1409.7310, doi:10.1007/JHEP12(2014)058.

[263] C.-Y. Yao, G.-J. Ding, Lepton and Quark Mixing Patterns from Finite Flavor Symmetries, Phys. Rev. D 92 (9) (2015) 096010. arXiv:1505.03798, doi:10.1103/PhysRevD.92.096010.

[264] F. Feruglio, C. Hagedorn, R. Ziegler, Lepton Mixing Parameters from Discrete and CP Symmetries, JHEP 07 (2013) 027. arXiv:1211.5560, doi:10.1007/JHEP07(2013)027.

[265] P. Chen, C.-C. Li, G.-J. Ding, Lepton Flavor Mixing and CP Symmetry, Phys. Rev. D 91 (2015) 033003. arXiv:1412.8352, doi:10.1103/PhysRevD.91.033003.

[266] L. L. Everett, T. Garon, A. J. Stuart, A Bottom-Up Approach to Lepton Flavor and CP Symmetries, JHEP 04 (2015) 069. arXiv:1501.04336, doi:10.1007/JHEP04(2015)069.

[267] P. Chen, C.-Y. Yao, G.-J. Ding, Neutrino Mixing from CP Symmetry, Phys. Rev. D 92 (7) (2015) 073002. arXiv:1507.03419, doi:10.1103/PhysRevD.92.073002.

[268] L. L. Everett, A. J. Stuart, Lepton Sector Phases and Their Roles in Flavor and Generalized CP Symmetries, Phys. Rev. D 96 (3) (2017) 035030. arXiv:1611.03020, doi:10.1103/PhysRevD.96.035030.

[269] C. S. Lam, Determining Horizontal Symmetry from Neutrino Mixing, Phys. Rev. Lett. 101 (2008) 121602. arXiv:0804.2622, doi:10.1103/PhysRevLett.101.121602.

[270] C. S. Lam, Symmetry of Lepton Mixing, Phys. Lett. B 656 (2007) 193–198. arXiv:0708.3665, doi:10.1016/j.physletb.2007.09.032.

[271] C. S. Lam, Group Theory and Dynamics of Neutrino Mixing, Phys. Rev. D 83 (2011) 113002. arXiv:1104.0055, doi:10.1103/PhysRevD.83.113002.

[272] C. S. Lam, Horizontal Symmetry: Bottom Up and Top DownarXiv:1105.5166.

[273] S. Antusch, I. de Medeiros Varzielas, V. Maurer, C. Sluka, M. Spinrath, Towards predictive flavour models in SUSY SU(5) GUTs with doublet-triplet splitting, JHEP 09 (2014) 141. arXiv:1405.6962, doi:10.1007/JHEP09(2014)141.

[274] F. Björkeroth, F. J. de Anda, I. de Medeiros Varzielas, S. F. King, Towards a complete $A_4\times$ SU(5) SUSY GUT, JHEP 06 (2015) 141. arXiv:1503.03306, doi:10.1007/JHEP06(2015)141.

[275] M. Reig, J. W. F. Valle, C. A. Vaquera-Araujo, F. Wilczek, A Model of Comprehensive Unification, Phys. Lett. B 774 (2017) 667–670. arXiv:1706.03116, doi:10.1016/j.physletb.2017.10.038.

[276] C. D. Froggatt, H. B. Nielsen, Hierarchy of Quark Masses, Cabibbo Angles and CP Violation, Nucl. Phys. B 147 (1979) 277–298. doi:10.1016/0550-3213(79)90316-X.

[277] R. Barbieri, G. R. Dvali, L. J. Hall, Predictions from a U(2) flavor symmetry in supersymmetric theories, Phys. Lett. B 377 (1996) 76–82. arXiv:hep-ph/9512388, doi:10.1016/0370-2693(96)00318-8.

[278] A. Blum, C. Hagedorn, M. Lindner, Fermion Masses and Mixings from Dihedral Flavor Symmetries with Preserved Subgroups, Phys. Rev. D 77 (2008) 076004. arXiv:0709.3450, doi:10.1103/PhysRevD.77.076004.

[279] M. Holthausen, K. S. Lim, Quark and Leptonic Mixing Patterns from the Breakdown of a Common Discrete Flavor Symmetry, Phys. Rev. D 88 (2013) 033018. arXiv:1306.4356, doi:10.1103/PhysRevD.88.033018.

[280] T. Araki, H. Ishida, H. Ishimori, T. Kobayashi, A. Ogasahara, CKM matrix and flavor symmetries, Phys. Rev. D 88 (2013) 096002. arXiv:1309.4217, doi:10.1103/PhysRevD.88.096002.

[281] I. de Medeiros Varzielas, R. W. Rasmussen, J. Talbert, Bottom-Up Discrete Symmetries for Cabibbo Mixing, Int. J. Mod. Phys. A 32 (06n07) (2017) 1750047. arXiv:1605.03581, doi:10.1142/S0217751X17500476.

[282] C.-C. Li, J.-N. Lu, G.-J. Ding, Toward a unified interpretation of quark and lepton mixing from flavor and CP symmetries, JHEP 02 (2018) 038. arXiv:1706.04576, doi:10.1007/JHEP02(2018)038.

[283] J.-N. Lu, G.-J. Ding, Quark and lepton mixing patterns from a common discrete flavor symmetry with a generalized CP symmetry, Phys. Rev. D 98 (5) (2018) 055011. arXiv:1806.02301, doi:10.1103/PhysRevD.98.055011.

[284] J.-N. Lu, G.-J. Ding, Dihedral flavor group as the key to understand quark and lepton flavor mixing, JHEP 03 (2019) 056. arXiv:1901.07414, doi:10.1007/JHEP03(2019)056.

[285] C. Hagedorn, J. König, Lepton and quark mixing from a stepwise breaking of flavor and CP, Phys. Rev. D 100 (7) (2019) 075036. arXiv:1811.07750, doi:10.1103/PhysRevD.100.075036.

[286] C. Hagedorn, J. König, Lepton and quark masses and mixing in a SUSY model with $\Delta(384)$ and CP, Nucl. Phys. B 953 (2020) 114953. arXiv:1811.09262, doi:10.1016/j.nuclphysb.2020.114953.

[287] M. Reig, J. W. F. Valle, F. Wilczek, SO(3) family symmetry and axions, Phys. Rev. D 98 (9) (2018) 095008. arXiv:1805.08048, doi:10.1103/PhysRevD.98.095008.

[288] S. Morisi, E. Peinado, Y. Shimizu, J. W. F. Valle, Relating quarks and leptons without grand-unification, Phys. Rev. D 84 (2011) 036003. arXiv:1104.1633, doi:10.1103/PhysRevD.84.036003.

[289] C. Bonilla, S. Morisi, E. Peinado, J. W. F. Valle, Relating quarks and leptons with the T_7 flavour group, Phys. Lett. B 742 (2015) 99–106. arXiv:1411.4883, doi:10.1016/j.physletb.2015.01.017.

[290] C. Bonilla, J. M. Lamprea, E. Peinado, J. W. F. Valle, Flavour-symmetric type-II Dirac neutrino seesaw mechanism, Phys. Lett. B 779 (2018) 257–261. arXiv:1710.06498, doi:10.1016/j.physletb.2018.02.022.

[291] M.-C. Chen, S. F. King, O. Medina, J. W. F. Valle, Quark-lepton mass relations from modular flavor symmetryarXiv:2312.09255.

[292] F. J. de Anda, S. F. King, An $S_4 \times SU(5)$ SUSY GUT of flavour in 6d, JHEP 07 (2018) 057. arXiv:1803.04978, doi:10.1007/JHEP07(2018)057.

[293] F. J. de Anda, S. F. King, $SU(3) \times SO(10)$ in 6d, JHEP 10 (2018) 128. arXiv:1807.07078, doi:10.1007/JHEP10(2018)128.

[294] F. J. de Anda, I. Antoniadis, J. W. F. Valle, C. A. Vaquera-Araujo, Scotogenic dark matter in an orbifold theory of flavor, JHEP 10 (2020) 190. arXiv:2007.10402, doi:10.1007/JHEP10(2020)190.

[295] F. J. de Anda, O. Medina, J. W. F. Valle, C. A. Vaquera-Araujo, Scotogenic Majorana neutrino masses in a predictive orbifold theory of flavor, Phys. Rev. D 105 (5) (2022) 055030. arXiv:2110.06810, doi:10.1103/PhysRevD.105.055030.

[296] P. F. Harrison, D. H. Perkins, W. G. Scott, Tri-bimaximal mixing and the neutrino oscillation data, Phys. Lett. B 530 (2002) 167. arXiv:hep-ph/0202074, doi:10.1016/S0370-2693(02)01336-9.

[297] Z.-z. Xing, Nearly tri bimaximal neutrino mixing and CP violation, Phys. Lett. B 533 (2002) 85–93. arXiv:hep-ph/0204049, doi:10.1016/S0370-2693(02)01649-0.

[298] X. G. He, A. Zee, Some simple mixing and mass matrices for neutrinos, Phys. Lett. B 560 (2003) 87–90. arXiv:hep-ph/0301092, doi:10.1016/S0370-2693(03)00390-3.

[299] S. Luo, Z.-z. Xing, Generalized tri-bimaximal neutrino mixing and its sensitivity to radiative corrections, Phys. Lett. B 632 (2006) 341–348. arXiv:hep-ph/0509065, doi:10.1016/j.physletb.2005.10.068.

[300] F. Plentinger, W. Rodejohann, Deviations from tribimaximal neutrino mixing, Phys. Lett. B 625 (2005) 264–276. arXiv:hep-ph/0507143, doi:10.1016/j.physletb.2005.08.092.

[301] M. Hirsch, E. Ma, J. C. Romao, J. W. F. Valle, A. Villanova del Moral, Minimal supergravity radiative effects on the tri-bimaximal neutrino mixing pattern, Phys. Rev. D 75 (2007) 053006. arXiv:hep-ph/0606082, doi:10.1103/PhysRevD.75.053006.

[302] P. Wilina, M. S. Singh, N. N. Singh, Deviations from tribimaximal and golden ratio mixings under radiative corrections of neutrino masses and mixings, Int. J. Mod. Phys. A 37 (25) (2022) 2250156. arXiv:2205.01936, doi:10.1142/S0217751X22501561.

[303] F. P. An, et al., Improved Measurement of Electron Antineutrino Disappearance at Daya Bay, Chin. Phys. C 37 (2013) 011001. arXiv:1210.6327, doi:10.1088/1674-1137/37/1/011001.

[304] F. P. An, et al., Spectral measurement of electron antineutrino oscillation amplitude and frequency at Daya Bay, Phys. Rev. Lett. 112 (2014) 061801. arXiv:1310.6732, doi:10.1103/PhysRevLett.112.061801.

[305] K. Luk, Daya-bay talk on final 3158-day data (International Conference on Neutrino Physics and Astrophysics, 2022).
URL https://neutrino2022.org/

[306] F. P. An, et al., Precision Measurement of Reactor Antineutrino Oscillation at Kilometer-Scale Baselines by Daya Bay, Phys. Rev. Lett. 130 (16) (2023) 161802. arXiv:2211.14988, doi:10.1103/PhysRevLett.130.161802.

[307] T. Fukuyama, H. Nishiura, Mass matrix of Majorana neutrinosarXiv:hep-ph/9702253.

[308] E. Ma, M. Raidal, Neutrino mass, muon anomalous magnetic moment, and lepton flavor nonconservation, Phys. Rev. Lett. 87 (2001) 011802, [Erratum: Phys.Rev.Lett. 87, 159901 (2001)]. arXiv:hep-ph/0102255, doi:10.1103/PhysRevLett.87.011802.

[309] K. R. S. Balaji, W. Grimus, T. Schwetz, The Solar LMA neutrino oscillation solution in the Zee model, Phys. Lett. B 508 (2001) 301–310. arXiv:hep-ph/0104035, doi:10.1016/S0370-2693(01)00532-9.

[310] C. S. Lam, A 2-3 symmetry in neutrino oscillations, Phys. Lett. B 507 (2001) 214–218. arXiv:hep-ph/0104116, doi:10.1016/S0370-2693(01)00465-8.

[311] W. Grimus, L. Lavoura, Softly broken lepton numbers and maximal neutrino mixing, JHEP 07 (2001) 045. arXiv:hep-ph/0105212, doi:10.1088/1126-6708/2001/07/045.

[312] C. S. Lam, Magic neutrino mass matrix and the Bjorken-Harrison-Scott parameterization, Phys. Lett. B 640 (2006) 260–262. arXiv:hep-ph/0606220, doi:10.1016/j.physletb.2006.08.007.

[313] G. Altarelli, F. Feruglio, Tri-bimaximal neutrino mixing, A4 and the modular symmetry, Nucl. Phys. B 741 (2006) 215–235. arXiv:hep-ph/0512103, doi:10.1016/j.nuclphysb.2006.02.015.

[314] G. Altarelli, F. Feruglio, Y. Lin, Tri-bimaximal neutrino mixing from orbifolding, Nucl. Phys. B 775 (2007) 31–44. arXiv:hep-ph/0610165, doi:10.1016/j.nuclphysb.2007.03.042.

[315] G. Altarelli, F. Feruglio, L. Merlo, Tri-Bimaximal Neutrino Mixing and Discrete Flavour Symmetries, Fortsch. Phys. 61 (2013) 507–534. arXiv:1205.5133, doi:10.1002/prop.201200117.

[316] C. H. Albright, W. Rodejohann, Comparing Trimaximal Mixing and Its Variants with Deviations from Tri-bimaximal Mixing, Eur. Phys. J. C 62 (2009) 599–608. arXiv:0812.0436, doi:10.1140/epjc/s10052-009-1074-3.

[317] C. H. Albright, A. Dueck, W. Rodejohann, Possible Alternatives to Tri-bimaximal Mixing, Eur. Phys. J. C 70 (2010) 1099–1110. arXiv:1004.2798, doi:10.1140/epjc/s10052-010-1492-2.

[318] X.-G. He, A. Zee, Minimal Modification to Tri-bimaximal Mixing, Phys. Rev. D 84 (2011) 053004. arXiv:1106.4359, doi:10.1103/PhysRevD.84.053004.

[319] C. Luhn, Trimaximal TM_1 neutrino mixing in S_4 with spontaneous CP violation, Nucl. Phys. B 875 (2013) 80–100. arXiv:1306.2358, doi:10.1016/j.nuclphysb.2013.07.003.

[320] C.-C. Li, G.-J. Ding, Generalised CP and trimaximal TM_1 lepton mixing in S_4 family symmetry, Nucl. Phys. B 881 (2014) 206–232. arXiv:1312.4401, doi:10.1016/j.nuclphysb.2014.02.002.

[321] J. D. Bjorken, P. F. Harrison, W. G. Scott, Simplified unitarity triangles for the lepton sector, Phys. Rev. D 74 (2006) 073012. arXiv:hep-ph/0511201, doi:10.1103/PhysRevD.74.073012.

[322] X.-G. He, A. Zee, Minimal modification to the tri-bimaximal neutrino mixing, Phys. Lett. B 645 (2007) 427–431. arXiv:hep-ph/0607163, doi:10.1016/j.physletb.2006.11.055.

[323] W. Grimus, L. Lavoura, A Model for trimaximal lepton mixing, JHEP 09 (2008) 106. arXiv:0809.0226, doi:10.1088/1126-6708/2008/09/106.

[324] W. Grimus, L. Lavoura, A. Singraber, Trimaximal lepton mixing with a trivial Dirac phase, Phys. Lett. B 686 (2010) 141–145. arXiv:0911.5120, doi:10.1016/j.physletb.2010.02.032.

[325] G.-J. Ding, S. F. King, C. Luhn, A. J. Stuart, Spontaneous CP violation from vacuum alignment in S_4 models of leptons, JHEP 05 (2013) 084. arXiv:1303.6180, doi:10.1007/JHEP05(2013)084.

[326] W. Rodejohann, Unified Parametrization for Quark and Lepton Mixing Angles, Phys. Lett. B 671 (2009) 267–271. arXiv:0810.5239, doi:10.1016/j.physletb.2008.12.010.

[327] A. Adulpravitchai, A. Blum, W. Rodejohann, Golden Ratio Prediction for Solar Neutrino Mixing, New J. Phys. 11 (2009) 063026. arXiv:0903.0531, doi:10.1088/1367-2630/11/6/063026.

[328] A. Datta, F.-S. Ling, P. Ramond, Correlated hierarchy, Dirac masses and large mixing angles, Nucl. Phys. B 671 (2003) 383–400. arXiv:hep-ph/0306002, doi:10.1016/j.nuclphysb.2003.08.026.

[329] Y. Kajiyama, M. Raidal, A. Strumia, The Golden ratio prediction for the solar neutrino mixing, Phys. Rev. D 76 (2007) 117301. arXiv:0705.4559, doi:10.1103/PhysRevD.76.117301.

[330] L. L. Everett, A. J. Stuart, Icosahedral (A(5)) Family Symmetry and the Golden Ratio Prediction for Solar Neutrino Mixing, Phys. Rev. D 79 (2009) 085005. arXiv:0812.1057, doi:10.1103/PhysRevD.79.085005.

[331] F. Feruglio, A. Paris, The Golden Ratio Prediction for the Solar Angle from a Natural Model with A5 Flavour Symmetry, JHEP 03 (2011) 101. arXiv:1101.0393, doi:10.1007/JHEP03(2011)101.

[332] G.-J. Ding, L. L. Everett, A. J. Stuart, Golden Ratio Neutrino Mixing and A_5 Flavor Symmetry, Nucl. Phys. B 857 (2012) 219–253. arXiv:1110.1688, doi:10.1016/j.nuclphysb.2011.12.004.

[333] V. D. Barger, S. Pakvasa, T. J. Weiler, K. Whisnant, Bimaximal mixing of three neutrinos, Phys. Lett. B 437 (1998) 107–116. arXiv:hep-ph/9806387, doi:10.1016/S0370-2693(98)00880-6.

[334] G. Altarelli, F. Feruglio, L. Merlo, Revisiting Bimaximal Neutrino Mixing in a Model with S(4) Discrete Symmetry, JHEP 05 (2009) 020. arXiv:0903.1940, doi:10.1088/1126-6708/2009/05/020.

[335] C.-C. Li, G.-J. Ding, Deviation from bimaximal mixing and leptonic CP phases in S_4 family symmetry and generalized CP, JHEP 08 (2015) 017. arXiv:1408.0785, doi:10.1007/JHEP08(2015)017.

[336] R. de Adelhart Toorop, F. Feruglio, C. Hagedorn, Discrete Flavour Symmetries in Light of T2K, Phys. Lett. B 703 (2011) 447–451. arXiv:1107.3486, doi:10.1016/j.physletb.2011.08.013.

[337] G.-J. Ding, TFH Mixing Patterns, Large θ_{13} and $\Delta(96)$ Flavor Symmetry, Nucl. Phys. B 862 (2012) 1–42. arXiv:1201.3279, doi:10.1016/j.nuclphysb.2012.04.002.

[338] C. Hagedorn, A. Meroni, L. Vitale, Mixing patterns from the groups $\Sigma(n\phi)$, J. Phys. A 47 (2014) 055201. arXiv:1307.5308, doi:10.1088/1751-8113/47/5/055201.

[339] G. Ecker, W. Grimus, W. Konetschny, Quark Mass Matrices in Left-right Symmetric Gauge Theories, Nucl. Phys. B 191 (1981) 465–492. doi:10.1016/0550-3213(81)90309-6.

[340] G. Ecker, W. Grimus, H. Neufeld, Spontaneous CP Violation in Left-right Symmetric Gauge Theories, Nucl. Phys. B 247 (1984) 70–82. doi:10.1016/0550-3213(84)90373-0.

[341] J. Bernabeu, G. C. Branco, M. Gronau, CP Restrictions on Quark Mass Matrices, Phys. Lett. B 169 (1986) 243–247. doi:10.1016/0370-2693(86)90659-3.

[342] G. Ecker, W. Grimus, H. Neufeld, A Standard Form for Generalized CP Transformations, J. Phys. A 20 (1987) L807. doi:10.1088/0305-4470/20/12/010.

[343] H. Neufeld, W. Grimus, G. Ecker, Generalized CP Invariance, Neutral Flavor Conservation and the Structure of the Mixing Matrix, Int. J. Mod. Phys. A 3 (1988) 603–616. doi:10.1142/S0217751X88000254.

[344] M. Holthausen, M. Lindner, M. A. Schmidt, CP and Discrete Flavour Symmetries, JHEP 04 (2013) 122. arXiv:1211.6953, doi:10.1007/JHEP04(2013)122.

[345] C.-Y. Yao, G.-J. Ding, CP Symmetry and Lepton Mixing from a Scan of Finite Discrete Groups, Phys. Rev. D 94 (7) (2016) 073006. arXiv:1606.05610, doi:10.1103/PhysRevD.94.073006.

[346] A. Abada, G. Arcadi, V. Domcke, M. Drewes, J. Klaric, M. Lucente, Low-scale leptogenesis with three heavy neutrinos, JHEP 01 (2019) 164. arXiv:1810.12463, doi:10.1007/JHEP01(2019)164.

[347] M. Drewes, Y. Georis, J. Klarić, Mapping the Viable Parameter Space for Testable Leptogenesis, Phys. Rev. Lett. 128 (5) (2022) 051801. arXiv:2106.16226, doi:10.1103/PhysRevLett.128.051801.

[348] C.-C. Li, G.-J. Ding, Implications of residual CP symmetry for leptogenesis in a model with two right-handed neutrinos, Phys. Rev. D 96 (7) (2017) 075005. arXiv:1701.08508, doi:10.1103/PhysRevD.96.075005.

[349] A. S. Joshipura, K. M. Patel, Horizontal symmetries of leptons with a massless neutrino, Phys. Lett. B 727 (2013) 480–487. arXiv:1306.1890, doi:10.1016/j.physletb.2013.11.003.

[350] A. S. Joshipura, K. M. Patel, A massless neutrino and lepton mixing patterns from finite discrete subgroups of U(3), JHEP 04 (2014) 009. arXiv:1401.6397, doi:10.1007/JHEP04(2014)009.

[351] S. F. King, P. O. Ludl, Direct and Semi-Direct Approaches to Lepton Mixing with a Massless Neutrino, JHEP 06 (2016) 147. arXiv:1605.01683, doi:10.1007/JHEP06(2016)147.

[352] R. Barbieri, L. J. Hall, S. Raby, A. Romanino, Unified theories with U(2) flavor symmetry, Nucl. Phys. B 493 (1997) 3–26. arXiv:hep-ph/9610449, doi:10.1016/S0550-3213(97)00134-X.

[353] R. Barbieri, L. J. Hall, A. Romanino, Consequences of a U(2) flavor symmetry, Phys. Lett. B 401 (1997) 47–53. arXiv:hep-ph/9702315, doi:10.1016/S0370-2693(97)00372-9.

[354] M. Linster, R. Ziegler, A Realistic $U(2)$ Model of Flavor, JHEP 08 (2018) 058. arXiv:1805.07341, doi:10.1007/JHEP08(2018)058.

[355] S. F. King, Predicting neutrino parameters from SO(3) family symmetry and quark-lepton unification, JHEP 08 (2005) 105. arXiv:hep-ph/0506297, doi:10.1088/1126-6708/2005/08/105.

[356] S. F. King, M. Malinsky, Towards a Complete Theory of Fermion Masses and Mixings with SO(3) Family Symmetry and 5-D SO(10) Unification, JHEP 11 (2006) 071. arXiv:hep-ph/0608021, doi:10.1088/1126-6708/2006/11/071.

[357] I. de Medeiros Varzielas, G. G. Ross, SU(3) family symmetry and neutrino bi-tri-maximal mixing, Nucl. Phys. B 733 (2006) 31–47. arXiv:hep-ph/0507176, doi:10.1016/j.nuclphysb.2005.10.039.

[358] S. Antusch, S. F. King, M. Malinsky, Solving the SUSY Flavour and CP Problems with SU(3) Family Symmetry, JHEP 06 (2008) 068. arXiv:0708.1282, doi:10.1088/1126-6708/2008/06/068.

[359] F. Bazzocchi, S. Morisi, M. Picariello, E. Torrente-Lujan, Embedding A(4) into SU(3) x U(1) flavor symmetry: Large neutrino mixing and fermion mass hierarchy in SO(10) GUT, J. Phys. G 36 (2009) 015002. arXiv:0802.1693, doi:10.1088/0954-3899/36/1/015002.

[360] S. Weinberg, The Problem of Mass, Trans. New York Acad. Sci. 38 (1977) 185–201. doi:10.1111/j.2164-0947.1977.tb02958.x.

[361] H. Fritzsch, Calculating the Cabibbo Angle, Phys. Lett. B 70 (1977) 436–440. doi:10.1016/0370-2693(77)90408-7.

[362] H. Fritzsch, Weak Interaction Mixing in the Six - Quark Theory, Phys. Lett. B 73 (1978) 317–322. doi:10.1016/0370-2693(78)90524-5.

[363] P. H. Frampton, S. L. Glashow, D. Marfatia, Zeroes of the neutrino mass matrix, Phys. Lett. B 536 (2002) 79–82. arXiv:hep-ph/0201008, doi:10.1016/S0370-2693(02)01817-8.

[364] P. Chen, G.-J. Ding, F. Gonzalez-Canales, J. W. F. Valle, Classifying CP transformations according to their texture zeros: theory and implications, Phys. Rev. D 94 (3) (2016) 033002. arXiv:1604.03510, doi:10.1103/PhysRevD.94.033002.

[365] P. Chen, S. Centelles Chuliá, G.-J. Ding, R. Srivastava, J. W. F. Valle, CP symmetries as guiding posts: revamping tri-bi-maximal mixing. Part I, JHEP 03 (2019) 036. arXiv:1812.04663, doi:10.1007/JHEP03(2019)036.

[366] S. Centelles Chulia, A. Trautner, Asymmetric tri-bi-maximal mixing and residual symmetries, Mod.Phys.Lett.A 35 (35) (2020) 2050292. arXiv:1911.12043, doi:10.1142/S0217732320502922.

[367] P. Chen, S. Centelles Chuliá, G.-J. Ding, R. Srivastava, J. W. F. Valle, Realistic tribimaximal neutrino mixing, Phys. Rev. D 98 (5) (2018) 055019. arXiv:1806.03367, doi:10.1103/PhysRevD.98.055019.

[368] J. Schechter, J. W. F. Valle, Majorana Neutrinos and Magnetic Fields, Phys. Rev. D 24 (1981) 1883–1889, [Erratum: Phys.Rev.D 25, 283 (1982)]. doi:10.1103/PhysRevD.25.283.

[369] L. Wolfenstein, CP Properties of Majorana Neutrinos and Double beta Decay, Phys. Lett. B 107 (1981) 77–79. doi:10.1016/0370-2693(81)91151-5.

[370] P. Chen, S. Centelles Chuliá, G.-J. Ding, R. Srivastava, J. W. F. Valle, CP symmetries as guiding posts: Revamping tribimaximal mixing. II., Phys. Rev. D 100 (5) (2019) 053001. arXiv:1905.11997, doi:10.1103/PhysRevD.100.053001.

[371] S. M. Boucenna, S. Morisi, M. Tortola, J. W. F. Valle, Bi-large neutrino mixing and the Cabibbo angle, Phys. Rev. D 86 (2012) 051301. arXiv:1206.2555, doi:10.1103/PhysRevD.86.051301.

[372] G.-J. Ding, S. Morisi, J. W. F. Valle, Bilarge neutrino mixing and Abelian flavor symmetry, Phys. Rev. D 87 (5) (2013) 053013. arXiv:1211.6506, doi:10.1103/PhysRevD.87.053013.

[373] S. Roy, S. Morisi, N. N. Singh, J. W. F. Valle, The Cabibbo angle as a universal seed for quark and lepton mixings, Phys. Lett. B 748 (2015) 1–4. arXiv:1410.3658, doi:10.1016/j.physletb.2015.06.052.

[374] P. Chen, G.-J. Ding, R. Srivastava, J. W. F. Valle, Predicting neutrino oscillations with "bi-large" lepton mixing matrices, Phys. Lett. B 792 (2019) 461–464. arXiv:1902.08962, doi:10.1016/j.physletb.2019.04.022.

[375] S. Roy, K. Sashikanta Singh, J. Borah, Revamped Bi-Large neutrino mixing with Gatto-Sartori-Tonin like relation, Nucl. Phys. B 960 (2020) 115204. arXiv:2001.07401, doi:10.1016/j.nuclphysb.2020.115204.

[376] H. Ishimori, T. Kobayashi, H. Ohki, H. Okada, Y. Shimizu, M. Tanimoto, An introduction to non-Abelian discrete symmetries for particle physicists, Vol. 858, 2012. doi:10.1007/978-3-642-30805-5.

[377] Y. Abe, et al., Reactor electron antineutrino disappearance in the Double Chooz experiment, Phys. Rev. D 86 (2012) 052008. arXiv:1207.6632, doi:10.1103/PhysRevD.86.052008.

[378] S. F. King, C. Luhn, On the origin of neutrino flavour symmetry, JHEP 10 (2009) 093. arXiv:0908.1897, doi:10.1088/1126-6708/2009/10/093.

[379] S. F. King, T. Neder, A. J. Stuart, Lepton mixing predictions from $\Delta(6n^2)$ family Symmetry, Phys. Lett. B 726 (2013) 312–315. arXiv:1305.3200, doi:10.1016/j.physletb.2013.08.052.

[380] S. F. King, T. Neder, Lepton mixing predictions including Majorana phases from $\Delta(6\mathrm{n}^2)$ flavour symmetry and generalised CP, Phys. Lett. B 736 (2014) 308–316. arXiv:1403.1758, doi:10.1016/j.physletb.2014.07.043.

[381] S.-F. Ge, D. A. Dicus, W. W. Repko, $\mathbb{Z}_2$ Symmetry Prediction for the Leptonic Dirac CP Phase, Phys. Lett. B 702 (2011) 220–223. arXiv:1104.0602, doi:10.1016/j.physletb.2011.06.096.

[382] S.-F. Ge, D. A. Dicus, W. W. Repko, Residual Symmetries for Neutrino Mixing with a Large θ_{13} and Nearly Maximal δ_D, Phys. Rev. Lett. 108 (2012) 041801. arXiv:1108.0964, doi:10.1103/PhysRevLett.108.041801.

[383] D. Hernandez, A. Y. Smirnov, Lepton mixing and discrete symmetries, Phys. Rev. D 86 (2012) 053014. arXiv:1204.0445, doi:10.1103/PhysRevD.86.053014.

[384] D. Hernandez, A. Y. Smirnov, Discrete symmetries and model-independent patterns of lepton mixing, Phys. Rev. D 87 (5) (2013) 053005. arXiv:1212.2149, doi:10.1103/PhysRevD.87.053005.

[385] W. Grimus, M. N. Rebelo, Automorphisms in gauge theories and the definition of CP and P, Phys. Rept. 281 (1997) 239–308. arXiv:hep-ph/9506272, doi:10.1016/S0370-1573(96)00030-0.

[386] Z.-z. Xing, Z.-h. Zhao, A review of μ-τ flavor symmetry in neutrino physics, Rept. Prog. Phys. 79 (7) (2016) 076201. arXiv:1512.04207, doi:10.1088/0034-4885/79/7/076201.

[387] M.-C. Chen, M. Fallbacher, K. T. Mahanthappa, M. Ratz, A. Trautner, CP Violation from Finite Groups, Nucl. Phys. B 883 (2014) 267–305. arXiv:1402.0507, doi:10.1016/j.nuclphysb.2014.03.023.

[388] G.-J. Ding, S. F. King, A. J. Stuart, Generalised CP and A_4 Family Symmetry, JHEP 12 (2013) 006. arXiv:1307.4212, doi:10.1007/JHEP12(2013)006.

[389] S. F. King, Littlest Seesaw, JHEP 02 (2016) 085. arXiv:1512.07531, doi:10.1007/JHEP02(2016)085.

[390] G.-J. Ding, S. F. King, C.-C. Li, Tri-Direct CP in the Littlest Seesaw Playground, JHEP 12 (2018) 003. arXiv:1807.07538, doi:10.1007/JHEP12(2018)003.

[391] G.-J. Ding, S. F. King, C.-C. Li, Lepton mixing predictions from S_4 in the tridirect CP approach to two right-handed neutrino models, Phys. Rev. D 99 (7) (2019) 075035. arXiv:1811.12340, doi:10.1103/PhysRevD.99.075035.

[392] P.-T. Chen, G.-J. Ding, S. F. King, C.-C. Li, A New Littlest Seesaw Model, J. Phys. G 47 (6) (2020) 065001. arXiv:1906.11414, doi:10.1088/1361-6471/ab7e8d.

[393] C.-C. Li, J.-N. Lu, G.-J. Ding, A_4 and CP symmetry and a model with maximal CP violation, Nucl. Phys. B 913 (2016) 110–131. arXiv:1608.01860, doi:10.1016/j.nuclphysb.2016.09.005.

[394] F. Feruglio, C. Hagedorn, R. Ziegler, A realistic pattern of lepton mixing and masses from S_4 and CP, Eur. Phys. J. C 74 (2014) 2753. arXiv:1303.7178, doi:10.1140/epjc/s10052-014-2753-2.

[395] J.-N. Lu, G.-J. Ding, Alternative Schemes of Predicting Lepton Mixing Parameters from Discrete Flavor and CP Symmetry, Phys. Rev. D 95 (1) (2017) 015012. arXiv:1610.05682, doi:10.1103/PhysRevD.95.015012.

[396] J. T. Penedo, S. T. Petcov, A. V. Titov, Neutrino mixing and leptonic CP violation from S_4 flavour and generalised CP symmetries, JHEP 12 (2017) 022. arXiv:1705.00309, doi:10.1007/JHEP12(2017)022.

[397] G. C. Branco, I. de Medeiros Varzielas, S. F. King, Invariant approach to CP in family symmetry models, Phys. Rev. D 92 (3) (2015) 036007. arXiv:1502.03105, doi:10.1103/PhysRevD.92.036007.

[398] G. C. Branco, I. de Medeiros Varzielas, S. F. King, Invariant approach to $\mathcal{CP}$ in unbroken $\Delta(27)$, Nucl. Phys. B 899 (2015) 14–36. arXiv:1505.06165, doi:10.1016/j.nuclphysb.2015.07.024.

[399] G.-J. Ding, Y.-L. Zhou, Predicting lepton flavor mixing from $\Delta(48)$ and generalized CP symmetries, Chin. Phys. C 39 (2) (2015) 021001. arXiv:1312.5222, doi:10.1088/1674-1137/39/2/021001.

[400] G.-J. Ding, Y.-L. Zhou, Lepton mixing parameters from $\Delta(48)$ family symmetry and generalised CP, JHEP 06 (2014) 023. arXiv:1404.0592, doi:10.1007/JHEP06(2014)023.

[401] C.-C. Li, G.-J. Ding, Lepton Mixing in A_5 Family Symmetry and Generalized CP, JHEP 05 (2015) 100. arXiv:1503.03711, doi:10.1007/JHEP05(2015)100.

[402] A. Di Iura, C. Hagedorn, D. Meloni, Lepton mixing from the interplay of the alternating group A_5 and CP, JHEP 08 (2015) 037. arXiv:1503.04140, doi:10.1007/JHEP08(2015)037.

[403] P. Ballett, S. Pascoli, J. Turner, Mixing angle and phase correlations from A5 with generalized CP and their prospects for discovery, Phys. Rev. D 92 (9) (2015) 093008. arXiv:1503.07543, doi:10.1103/PhysRevD.92.093008.

[404] J. Turner, Predictions for leptonic mixing angle correlations and nontrivial Dirac CP violation from A_5 with generalized CP symmetry, Phys. Rev. D 92 (11) (2015) 116007. arXiv:1507.06224, doi:10.1103/PhysRevD.92.116007.

[405] A. Di Iura, M. L. López-Ibáñez, D. Meloni, Neutrino masses and lepton mixing from $A_5 \times$ CP, Nucl. Phys. B 949 (2019) 114794. arXiv:1811.09662, doi:10.1016/j.nuclphysb.2019.114794.

[406] G.-J. Ding, S. F. King, Generalized CP and $\Delta(96)$ family symmetry, Phys. Rev. D 89 (9) (2014) 093020. arXiv:1403.5846, doi:10.1103/PhysRevD.89.093020.

[407] C. Hagedorn, A. Meroni, E. Molinaro, Lepton mixing from $\Delta(3n^2)$ and $\Delta(6n^2)$ and CP, Nucl. Phys. B 891 (2015) 499–557. arXiv:1408.7118, doi:10.1016/j.nuclphysb.2014.12.013.

[408] G.-J. Ding, S. F. King, Generalized CP and $\Delta(3n^2)$ Family Symmetry for Semi-Direct Predictions of the PMNS Matrix, Phys. Rev. D 93 (2016) 025013. arXiv:1510.03188, doi:10.1103/PhysRevD.93.025013.

[409] G.-J. Ding, S. F. King, T. Neder, Generalised CP and $\Delta(6n^2)$ family symmetry in semi-direct models of leptons, JHEP 12 (2014) 007. arXiv:1409.8005, doi:10.1007/JHEP12(2014)007.

[410] C.-C. Li, C.-Y. Yao, G.-J. Ding, Lepton Mixing Predictions from Infinite Group Series $D^{(1)}_{9n,3n}$ with Generalized CP, JHEP 05 (2016) 007. arXiv:1601.06393, doi:10.1007/JHEP05(2016)007.

[411] C. Jarlskog, Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation, Phys. Rev. Lett. 55 (1985) 1039. doi:10.1103/PhysRevLett.55.1039.

[412] G. C. Branco, L. Lavoura, M. N. Rebelo, Majorana Neutrinos and CP Violation in the Leptonic Sector, Phys. Lett. B 180 (1986) 264–268. doi:10.1016/0370-2693(86)90307-2.

[413] J. F. Nieves, P. B. Pal, Minimal Rephasing Invariant CP Violating Parameters With Dirac and Majorana Fermions, Phys. Rev. D 36 (1987) 315. doi:10.1103/PhysRevD.36.315.

[414] T. D. Lee, A Theory of Spontaneous T Violation, Phys. Rev. D 8 (1973) 1226–1239. doi:10.1103/PhysRevD.8.1226.

[415] G. C. Branco, J. M. Gerard, W. Grimus, Geometrical T Violation, Phys. Lett. B 136 (1984) 383–386. doi:10.1016/0370-2693(84)92024-0.

[416] I. de Medeiros Varzielas, D. Emmanuel-Costa, Geometrical CP Violation, Phys. Rev. D 84 (2011) 117901. arXiv:1106.5477, doi:10.1103/PhysRevD.84.117901.

[417] C. Luhn, S. Nasri, P. Ramond, The Flavor group Delta($3n^2$), J. Math. Phys. 48 (2007) 073501. arXiv:hep-th/0701188, doi:10.1063/1.2734865.

[418] J. A. Escobar, C. Luhn, The Flavor Group Delta(6n**2), J. Math. Phys. 50 (2009) 013524. arXiv:0809.0639, doi:10.1063/1.3046563.

[419] D. Adey, et al., Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay, Phys. Rev. Lett. 121 (24) (2018) 241805. arXiv:1809.02261, doi:10.1103/PhysRevLett.121.241805.

[420] Y. Abe, et al., Indication of Reactor $\bar{\nu}_e$ Disappearance in the Double Chooz Experiment, Phys. Rev. Lett. 108 (2012) 131801. arXiv:1112.6353, doi:10.1103/PhysRevLett.108.131801.

[421] Y. Abe, et al., Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector, JHEP 10 (2014) 086, [Erratum: JHEP 02, 074 (2015)]. arXiv:1406.7763, doi:10.1007/JHEP02(2015)074.

[422] G. Bak, et al., Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO, Phys. Rev. Lett. 121 (20) (2018) 201801. arXiv:1806.00248, doi:10.1103/PhysRevLett.121.201801.

[423] D. S. Ayres, et al., NOvA: Proposal to Build a 30 Kiloton Off-Axis Detector to Study $\nu_\mu \to \nu_e$ Oscillations in the NuMI Beamlinearxiv:hep-ex/0503053.

[424] K. Abe, et al., The T2K Experiment, Nucl. Instrum. Meth. A 659 (2011) 106–135. arXiv:1106.1238, doi:10.1016/j.nima.2011.06.067.

[425] K. Abe, et al., Search for CP Violation in Neutrino and Antineutrino Oscillations by the T2K Experiment with 2.2×10^{21} Protons on Target, Phys. Rev. Lett. 121 (17) (2018) 171802. arXiv:1807.07891, doi:10.1103/PhysRevLett.121.171802.

[426] M. A. Acero, et al., First Measurement of Neutrino Oscillation Parameters using Neutrinos and Antineutrinos by NOvA, Phys. Rev. Lett. 123 (15) (2019) 151803. arXiv:1906.04907, doi:10.1103/PhysRevLett.123.151803.

[427] K. Abe, et al., Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations, Nature 580 (7803) (2020) 339–344, [Erratum: Nature 583, E16 (2020)]. arXiv:1910.03887, doi:10.1038/s41586-020-2177-0.

[428] K. Abe, et al., Measurements of neutrino oscillation parameters from the T2K experiment using 3.6×10^{21} protons on target, Eur. Phys. J. C 83 (9) (2023) 782. arXiv:2303.03222, doi:10.1140/epjc/s10052-023-11819-x.

[429] M. A. Acero, et al., Improved measurement of neutrino oscillation parameters by the NOvA experiment, Phys. Rev. D 106 (3) (2022) 032004. arXiv:2108.08219, doi:10.1103/PhysRevD.106.032004.

[430] B. Abi, et al., Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physicsarxiv:2002.03005.

[431] K. Abe, et al., Physics potential of a long-baseline neutrino oscillation experiment using a J-PARC neutrino beam and Hyper-Kamiokande, PTEP 2015 (2015) 053C02. arXiv:1502.05199, doi:10.1093/ptep/ptv061.

[432] S. T. Petcov, A. V. Titov, Assessing the Viability of A_4, S_4 and A_5 Flavour Symmetries for Description of Neutrino Mixing, Phys. Rev. D 97 (11) (2018) 115045. arXiv:1804.00182, doi:10.1103/PhysRevD.97.115045.

[433] M. Blennow, M. Ghosh, T. Ohlsson, A. Titov, Testing Lepton Flavor Models at ESSnuSB, JHEP 07 (2020) 014. arXiv:2004.00017, doi:10.1007/JHEP07(2020)014.

[434] J. Gehrlein, M. Spinrath, Leptonic Sum Rules from Flavour Models with Modular Symmetries, JHEP 03 (2021) 177. arXiv:2012.04131, doi:10.1007/JHEP03(2021)177.

[435] P. P. Novichkov, J. T. Penedo, S. T. Petcov, Double cover of modular S_4 for flavour model building, Nucl. Phys. B 963 (2021) 115301. arXiv:2006.03058, doi:10.1016/j.nuclphysb.2020.115301.

[436] P. P. Novichkov, J. T. Penedo, S. T. Petcov, Fermion mass hierarchies, large lepton mixing and residual modular symmetries, JHEP 04 (2021) 206. arXiv:2102.07488, doi:10.1007/JHEP04(2021)206.

[437] J. Gehrlein, S. Petcov, M. Spinrath, A. Titov, Testing neutrino flavor models, in: Snowmass 2021, 2022. arXiv:2203.06219.

[438] D. Marzocca, S. T. Petcov, A. Romanino, M. C. Sevilla, Nonzero $|U_{e3}|$ from Charged Lepton Corrections and the Atmospheric Neutrino Mixing Angle, JHEP 05 (2013) 073. arXiv:1302.0423, doi:10.1007/JHEP05(2013)073.

[439] P. Ballett, S. F. King, C. Luhn, S. Pascoli, M. A. Schmidt, Testing solar lepton mixing sum rules in neutrino oscillation experiments, JHEP 12 (2014) 122. arXiv:1410.7573, doi:10.1007/JHEP12(2014)122.

[440] J. E. Kim, M.-S. Seo, Quark and lepton mixing angles with a dodeca-symmetry, JHEP 02 (2011) 097. arXiv:1005.4684, doi:10.1007/JHEP02(2011)097.

[441] S. K. Agarwalla, S. S. Chatterjee, S. T. Petcov, A. V. Titov, Addressing Neutrino Mixing Models with DUNE and T2HK, Eur. Phys. J. C 78 (4) (2018) 286. arXiv:1711.02107, doi:10.1140/epjc/s10052-018-5772-6.

[442] I. Masina, A Maximal atmospheric mixing from a maximal CP violating phase, Phys. Lett. B 633 (2006) 134–140. arXiv:hep-ph/0508031, doi:10.1016/j.physletb.2005.10.097.

[443] S. Antusch, S. F. King, Charged lepton corrections to neutrino mixing angles and CP phases revisited, Phys. Lett. B 631 (2005) 42–47. arXiv:hep-ph/0508044, doi:10.1016/j.physletb.2005.09.075.

[444] P. Ballett, S. F. King, C. Luhn, S. Pascoli, M. A. Schmidt, Testing atmospheric mixing sum rules at precision neutrino facilities, Phys. Rev. D 89 (1) (2014) 016016. arXiv:1308.4314, doi:10.1103/PhysRevD.89.016016.

[445] I. Girardi, S. T. Petcov, A. V. Titov, Predictions for the Leptonic Dirac CP Violation Phase: a Systematic Phenomenological Analysis, Eur. Phys. J. C 75 (2015) 345. arXiv:1504.00658, doi:10.1140/epjc/s10052-015-3559-6.

[446] M. Blennow, M. Ghosh, T. Ohlsson, A. Titov, Probing Lepton Flavor Models at Future Neutrino Experiments, Phys. Rev. D 102 (11) (2020) 115004. arXiv:2005.12277, doi:10.1103/PhysRevD.102.115004.

[447] G.-J. Ding, Y.-F. Li, J. Tang, T.-C. Wang, Confronting tridirect CP -symmetry models with neutrino oscillation experiments, Phys. Rev. D 100 (5) (2019) 055022. arXiv:1905.12939, doi:10.1103/PhysRevD.100.055022.

[448] J. Gehrlein, M. Spinrath, Neutrino Mass Sum Rules and Symmetries of the Mass Matrix, Eur. Phys. J. C 77 (5) (2017) 281. arXiv:1704.02371, doi:10.1140/epjc/s10052-017-4817-6.

[449] E. Ma, Aspects of the tetrahedral neutrino mass matrix, Phys. Rev. D 72 (2005) 037301. arXiv:hep-ph/0505209, doi:10.1103/PhysRevD.72.037301.

[450] J. Barry, W. Rodejohann, Neutrino Mass Sum-rules in Flavor Symmetry Models, Nucl. Phys. B 842 (2011) 33–50. arXiv:1007.5217, doi:10.1016/j.nuclphysb.2010.08.015.

[451] S. F. King, A. Merle, A. J. Stuart, The Power of Neutrino Mass Sum Rules for Neutrinoless Double Beta Decay Experiments, JHEP 12 (2013) 005. arXiv:1307.2901, doi:10.1007/JHEP12(2013)005.

[452] J. Gehrlein, A. Merle, M. Spinrath, Renormalisation Group Corrections to Neutrino Mass Sum Rules, JHEP 09 (2015) 066. arXiv:1506.06139, doi:10.1007/JHEP09(2015)066.

[453] J. Gehrlein, A. Merle, M. Spinrath, Predictivity of Neutrino Mass Sum Rules, Phys. Rev. D 94 (9) (2016) 093003. arXiv:1606.04965, doi:10.1103/PhysRevD.94.093003.

[454] P. Huber, M. Lindner, W. Winter, Simulation of long-baseline neutrino oscillation experiments with GLoBES (General Long Baseline Experiment Simulator), Comput. Phys. Commun. 167 (2005) 195. arXiv:hep-ph/0407333, doi:10.1016/j.cpc.2005.01.003.

[455] P. Huber, J. Kopp, M. Lindner, M. Rolinec, W. Winter, New features in the simulation of neutrino oscillation experiments with GLoBES 3.0: General Long Baseline Experiment Simulator, Comput. Phys. Commun. 177 (2007) 432–438. arXiv:hep-ph/0701187, doi:10.1016/j.cpc.2007.05.004.

[456] R. G. Calland, A. C. Kaboth, D. Payne, Accelerated Event-by-Event Neutrino Oscillation Reweighting with Matter Effects on a GPU, JINST 9 (2014) P04016. arXiv:1311.7579, doi:10.1088/1748-0221/9/04/P04016.

[457] M. Wallraff, C. Wiebusch, Calculation of oscillation probabilities of atmospheric neutrinos using nuCraft, Comput. Phys. Commun. 197 (2015) 185–189. arXiv:1409.1387, doi:10.1016/j.cpc.2015.07.010.

[458] C. A. Argüelles, J. Salvado, C. N. Weaver, nuSQuIDS: A toolbox for neutrino propagation, Comput. Phys. Commun. 277 (2022) 108346. arXiv:2112.13804, doi:10.1016/j.cpc.2022.108346.

[459] J. Tang, T.-C. Wang, Flavour Symmetry Embedded - GLoBES (FaSE-GLoBES), Comput. Phys. Commun. 263 (2021) 107899. arXiv:2006.14886, doi:10.1016/j.cpc.2021.107899.

[460] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali, New dimensions at a millimeter to a Fermi and superstrings at a TeV, Phys. Lett. B 436 (1998) 257–263. arXiv:hep-ph/9804398, doi:10.1016/S0370-2693(98)00860-0.

[461] N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali, The Hierarchy problem and new dimensions at a millimeter, Phys. Lett. B 429 (1998) 263–272. arXiv:hep-ph/9803315, doi:10.1016/S0370-2693(98)00466-3.

[462] L. Randall, R. Sundrum, A Large mass hierarchy from a small extra dimension, Phys. Rev. Lett. 83 (1999) 3370–3373. arXiv:hep-ph/9905221, doi:10.1103/PhysRevLett.83.3370.

[463] N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali, Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity, Phys. Rev. D 59 (1999) 086004. arXiv:hep-ph/9807344, doi:10.1103/PhysRevD.59.086004.

[464] K. Agashe, A. Delgado, M. J. May, R. Sundrum, RS1, custodial isospin and precision tests, JHEP 08 (2003) 050. arXiv:hep-ph/0308036, doi:10.1088/1126-6708/2003/08/050.

[465] K. Agashe, R. Contino, L. Da Rold, A. Pomarol, A Custodial symmetry for $Zb\bar{b}$, Phys. Lett. B 641 (2006) 62–66. arXiv:hep-ph/0605341, doi:10.1016/j.physletb.2006.08.005.

[466] G. Cacciapaglia, C. Csaki, G. Marandella, J. Terning, The Gaugephobic Higgs, JHEP 02 (2007) 036. arXiv:hep-ph/0611358, doi:10.1088/1126-6708/2007/02/036.

[467] Y. Grossman, M. Neubert, Neutrino masses and mixings in nonfactorizable geometry, Phys. Lett. B 474 (2000) 361–371. arXiv:hep-ph/9912408, doi:10.1016/S0370-2693(00)00054-X.

[468] T. Gherghetta, A. Pomarol, Bulk fields and supersymmetry in a slice of AdS, Nucl. Phys. B 586 (2000) 141–162. arXiv:hep-ph/0003129, doi:10.1016/S0550-3213(00)00392-8.

[469] R. Gatto, G. Sartori, M. Tonin, Weak Selfmasses, Cabibbo Angle, and Broken SU(2) x SU(2), Phys. Lett. B 28 (1968) 128–130. doi:10.1016/0370-2693(68)90150-0.

[470] N. Deutschmann, T. Flacke, J. S. Kim, Current LHC Constraints on Minimal Universal Extra Dimensions, Phys. Lett. B 771 (2017) 515–520. arXiv:1702.00410, doi:10.1016/j.physletb.2017.06.004.

[471] N. Ganguly, A. Datta, Exploring non minimal Universal Extra Dimensional Model at the LHC, JHEP 10 (2018) 072. arXiv:1808.08801, doi:10.1007/JHEP10(2018)072.

[472] F. J. Escrihuela, D. V. Forero, O. G. Miranda, M. Tortola, J. W. F. Valle, On the description of nonunitary neutrino mixing, Phys. Rev. D 92 (5) (2015) 053009, [Erratum: Phys.Rev.D 93, 119905 (2016)]. arXiv:1503.08879, doi:10.1103/PhysRevD.92.053009.

[473] S. Antusch, V. Maurer, Running quark and lepton parameters at various scales, JHEP 11 (2013) 115. arXiv:1306.6879, doi:10.1007/JHEP11(2013)115.

[474] Z.-z. Xing, H. Zhang, S. Zhou, Updated Values of Running Quark and Lepton Masses, Phys. Rev. D 77 (2008) 113016. arXiv:0712.1419, doi:10.1103/PhysRevD.77.113016.

[475] S. Antusch, J. Kersten, M. Lindner, M. Ratz, M. A. Schmidt, Running neutrino mass parameters in see-saw scenarios, JHEP 03 (2005) 024. arXiv:hep-ph/0501272, doi:10.1088/1126-6708/2005/03/024.

[476] N. Abgrall, et al., The Large Enriched Germanium Experiment for Neutrinoless Double Beta Decay (LEGEND), AIP Conf. Proc. 1894 (1) (2017) 020027. arXiv:1709.01980, doi:10.1063/1.5007652.

[477] J. B. Albert, et al., Sensitivity and Discovery Potential of nEXO to Neutrinoless Double Beta Decay, Phys. Rev. C 97 (6) (2018) 065503. arXiv:1710.05075, doi:10.1103/PhysRevC.97.065503.

[478] F. Feruglio, Are neutrino masses modular forms?, 2019, pp. 227–266. arXiv:1706.08749, doi:10.1142/9789813238053_0012.

[479] T. Kobayashi, M. Tanimoto, Modular flavor symmetric models, 2023. arXiv:2307.03384.

[480] G.-J. Ding, S. F. King, Neutrino Mass and Mixing with Modular Symmetry arXiv:2311.09282.

[481] H. Cohen, F. Strömberg, Modular Forms: A Classical Approach, Graduate studies in mathematics, American Mathematical Society, 2017.
URL https://books.google.com.hk/books?id=1MmctQEACAAJ

[482] P. P. Novichkov, S. T. Petcov, M. Tanimoto, Trimaximal Neutrino Mixing from Modular A4 Invariance with Residual Symmetries, Phys. Lett. B 793 (2019) 247–258. arXiv:1812.11289, doi:10.1016/j.physletb.2019.04.043.

[483] G.-J. Ding, S. F. King, X.-G. Liu, J.-N. Lu, Modular S_4 and A_4 symmetries and their fixed points: new predictive examples of lepton mixing, JHEP 12 (2019) 030. arXiv:1910.03460, doi:10.1007/JHEP12(2019)030.

[484] X.-G. Liu, G.-J. Ding, Neutrino Masses and Mixing from Double Covering of Finite Modular Groups, JHEP 08 (2019) 134. arXiv:1907.01488, doi:10.1007/JHEP08(2019)134.

[485] X.-G. Liu, C.-Y. Yao, G.-J. Ding, Modular invariant quark and lepton models in double covering of S_4 modular group, Phys. Rev. D 103 (5) (2021) 056013. arXiv:2006.10722, doi:10.1103/PhysRevD.103.056013.

[486] X. Wang, B. Yu, S. Zhou, Double covering of the modular A_5 group and lepton flavor mixing in the minimal seesaw model, Phys. Rev. D 103 (7) (2021) 076005. arXiv:2010.10159, doi:10.1103/PhysRevD.103.076005.

[487] C.-Y. Yao, X.-G. Liu, G.-J. Ding, Fermion masses and mixing from the double cover and metaplectic cover of the A_5 modular group, Phys. Rev. D 103 (9) (2021) 095013. arXiv:2011.03501, doi:10.1103/PhysRevD.103.095013.

[488] X.-G. Liu, C.-Y. Yao, B.-Y. Qu, G.-J. Ding, Half-integral weight modular forms and application to neutrino mass models, Phys. Rev. D 102 (11) (2020) 115035. arXiv:2007.13706, doi:10.1103/PhysRevD.102.115035.

[489] P. P. Novichkov, J. T. Penedo, S. T. Petcov, A. V. Titov, Generalised CP Symmetry in Modular-Invariant Models of Flavour, JHEP 07 (2019) 165. arXiv:1905.11970, doi:10.1007/JHEP07(2019)165.

[490] T. Kobayashi, Y. Shimizu, K. Takagi, M. Tanimoto, T. H. Tatsuishi, H. Uchida, CP violation in modular invariant flavor models, Phys. Rev. D 101 (5) (2020) 055046. arXiv:1910.11553, doi:10.1103/PhysRevD.101.055046.

[491] G.-J. Ding, F. Feruglio, X.-G. Liu, CP symmetry and symplectic modular invariance, SciPost Phys. 10 (6) (2021) 133. arXiv:2102.06716, doi:10.21468/SciPostPhys.10.6.133.

[492] A. Baur, H. P. Nilles, A. Trautner, P. K. S. Vaudrevange, Unification of Flavor, CP, and Modular Symmetries, Phys. Lett. B 795 (2019) 7–14. arXiv:1901.03251, doi:10.1016/j.physletb.2019.03.066.

[493] B. S. Acharya, D. Bailin, A. Love, W. A. Sabra, S. Thomas, Spontaneous breaking of CP symmetry by orbifold moduli, Phys. Lett. B 357 (1995) 387–396, [Erratum: Phys.Lett.B 407, 451–451 (1997)]. arXiv:hep-th/9506143, doi:10.1016/0370-2693(95)00945-H.

[494] T. Dent, CP violation and modular symmetries, Phys. Rev. D 64 (2001) 056005. arXiv:hep-ph/0105285, doi:10.1103/PhysRevD.64.056005.

[495] J. Giedt, CP violation and moduli stabilization in heterotic models, Mod. Phys. Lett. A 17 (2002) 1465–1473. arXiv:hep-ph/0204017, doi:10.1142/S0217732302007879.

[496] G.-J. Ding, X.-G. Liu, C.-Y. Yao, A minimal modular invariant neutrino model, JHEP 01 (2023) 125. arXiv:2211.04546, doi:10.1007/JHEP01(2023)125.

[497] G.-J. Ding, X.-G. Liu, J.-N. Lu, M.-H. Weng, Modular binary octahedral symmetry for flavor structure of Standard Model, JHEP 11 (2023) 083. arXiv:2307.14926, doi:10.1007/JHEP11(2023)083.

[498] A. Baur, H. P. Nilles, A. Trautner, P. K. S. Vaudrevange, A String Theory of Flavor and $\mathcal{CP}$, Nucl. Phys. B 947 (2019) 114737. arXiv:1908.00805, doi:10.1016/j.nuclphysb.2019.114737.

[499] T. Kobayashi, S. Nagamoto, S. Uemura, Modular symmetry in magnetized/intersecting D-brane models, PTEP 2017 (2) (2017) 023B02. arXiv:1608.06129, doi:10.1093/ptep/ptw184.

[500] T. Kobayashi, S. Nagamoto, S. Takada, S. Tamba, T. H. Tatsuishi, Modular symmetry and non-Abelian discrete flavor symmetries in string compactification, Phys. Rev. D 97 (11) (2018) 116002. arXiv:1804.06644, doi:10.1103/PhysRevD.97.116002.

[501] T. Kobayashi, S. Tamba, Modular forms of finite modular subgroups from magnetized D-brane models, Phys. Rev. D 99 (4) (2019) 046001. arXiv:1811.11384, doi:10.1103/PhysRevD.99.046001.

[502] Y. Kariyazono, T. Kobayashi, S. Takada, S. Tamba, H. Uchida, Modular symmetry anomaly in magnetic flux compactification, Phys. Rev. D 100 (4) (2019) 045014. arXiv:1904.07546, doi:10.1103/PhysRevD.100.045014.

[503] H. Ohki, S. Uemura, R. Watanabe, Modular flavor symmetry on a magnetized torus, Phys. Rev. D 102 (8) (2020) 085008. arXiv:2003.04174, doi:10.1103/PhysRevD.102.085008.

[504] S. Kikuchi, T. Kobayashi, S. Takada, T. H. Tatsuishi, H. Uchida, Revisiting modular symmetry in magnetized torus and orbifold compactifications, Phys. Rev. D 102 (10) (2020) 105010. arXiv:2005.12642, doi:10.1103/PhysRevD.102.105010.

[505] Y. Almumin, M.-C. Chen, V. Knapp-Pérez, S. Ramos-Sánchez, M. Ratz, S. Shukla, Metaplectic Flavor Symmetries from Magnetized Tori, JHEP 05 (2021) 078. arXiv:2102.11286, doi:10.1007/JHEP05(2021)078.

[506] I. de Medeiros Varzielas, S. F. King, Y.-L. Zhou, Multiple modular symmetries as the origin of flavor, Phys. Rev. D 101 (5) (2020) 055033. arXiv:1906.02208, doi:10.1103/PhysRevD.101.055033.

[507] G.-J. Ding, F. Feruglio, X.-G. Liu, Automorphic Forms and Fermion Masses, JHEP 01 (2021) 037. arXiv:2010.07952, doi:10.1007/JHEP01(2021)037.

[508] K. Ishiguro, T. Kobayashi, H. Otsuka, Spontaneous CP violation and symplectic modular symmetry in Calabi-Yau compactifications, Nucl. Phys. B 973 (2021) 115598. arXiv:2010.10782, doi:10.1016/j.nuclphysb.2021.115598.

[509] A. Baur, M. Kade, H. P. Nilles, S. Ramos-Sanchez, P. K. S. Vaudrevange, Siegel modular flavor group and CP from string theory, Phys. Lett. B 816 (2021) 136176. arXiv:2012.09586, doi:10.1016/j.physletb.2021.136176.

[510] H. P. Nilles, S. Ramos-Sanchez, A. Trautner, P. K. S. Vaudrevange, Orbifolds from Sp(4,Z) and their modular symmetries, Nucl. Phys. B 971 (2021) 115534. arXiv:2105.08078, doi:10.1016/j.nuclphysb.2021.115534.

[511] H. P. Nilles, S. Ramos-Sánchez, P. K. S. Vaudrevange, Eclectic Flavor Groups, JHEP 02 (2020) 045. arXiv:2001.01736, doi:10.1007/JHEP02(2020)045.

[512] H. P. Nilles, S. Ramos-Sanchez, P. K. S. Vaudrevange, Lessons from eclectic flavor symmetries, Nucl. Phys. B 957 (2020) 115098. arXiv:2004.05200, doi:10.1016/j.nuclphysb.2020.115098.

[513] F. Feruglio, V. Gherardi, A. Romanino, A. Titov, Modular invariant dynamics and fermion mass hierarchies around $\tau = i$, JHEP 05 (2021) 242. arXiv:2101.08718, doi:10.1007/JHEP05(2021)242.

[514] F. Feruglio, Universal Predictions of Modular Invariant Flavor Models near the Self-Dual Point, Phys. Rev. Lett. 130 (10) (2023) 101801. arXiv:2211.00659, doi:10.1103/PhysRevLett.130.101801.

[515] F. Feruglio, Fermion masses, critical behavior and universality, JHEP 03 (2023) 236. arXiv:2302.11580, doi:10.1007/JHEP03(2023)236.

[516] M. Cvetic, A. Font, L. E. Ibanez, D. Lust, F. Quevedo, Target space duality, supersymmetry breaking and the stability of classical string vacua, Nucl. Phys. B 361 (1991) 194–232. doi:10.1016/0550-3213(91)90622-5.

[517] E. Gonzalo, L. E. Ibáñez, A. M. Uranga, Modular symmetries and the swampland conjectures, JHEP 05 (2019) 105. arXiv:1812.06520, doi:10.1007/JHEP05(2019)105.

[518] T. Kobayashi, Y. Shimizu, K. Takagi, M. Tanimoto, T. H. Tatsuishi, A_4 lepton flavor model and modulus stabilization from S_4 modular symmetry, Phys. Rev. D 100 (11) (2019) 115045, [Erratum: Phys.Rev.D 101, 039904 (2020)]. arXiv:1909.05139, doi:10.1103/PhysRevD.100.115045.

[519] K. Ishiguro, T. Kobayashi, H. Otsuka, Landscape of Modular Symmetric Flavor Models, JHEP 03 (2021) 161. arXiv:2011.09154, doi:10.1007/JHEP03(2021)161.

[520] P. P. Novichkov, J. T. Penedo, S. T. Petcov, Modular flavour symmetries and modulus stabilisation, JHEP 03 (2022) 149. arXiv:2201.02020, doi:10.1007/JHEP03(2022)149.

[521] J. M. Leedom, N. Righi, A. Westphal, Heterotic de Sitter beyond modular symmetry, JHEP 02 (2023) 209. arXiv:2212.03876, doi:10.1007/JHEP02(2023)209.

[522] V. Knapp-Perez, X.-G. Liu, H. P. Nilles, S. Ramos-Sanchez, M. Ratz, Matter matters in moduli fixing and modular flavor symmetries, Phys. Lett. B 844 (2023) 138106. arXiv:2304.14437, doi:10.1016/j.physletb.2023.138106.

[523] S. Kikuchi, T. Kobayashi, K. Nasu, Y. Yamada, Moduli trapping mechanism in modular flavor symmetric models, JHEP 08 (2023) 081. arXiv:2307.13230, doi:10.1007/JHEP08(2023)081.

[524] S. F. King, X. Wang, Modulus stabilisation in the multiple-modulus frameworkarXiv:2310.10369.

[525] C. Hagedorn, S. F. King, C. Luhn, A SUSY GUT of Flavour with S4 x SU(5) to NLO, JHEP 06 (2010) 048. arXiv:1003.4249, doi:10.1007/JHEP06(2010)048.

[526] G.-J. Ding, Fermion Masses and Flavor Mixings in a Model with S(4) Flavor Symmetry, Nucl. Phys. B 827 (2010) 82–111. arXiv:0909.2210, doi:10.1016/j.nuclphysb.2009.10.021.

arXiv:2402.16961v1 [hep-ph] 26 Feb 2024

Resolving the Ultracollinear Paradox with Effective Field Theory

Matthew Baumgart[1] and Panagiotis Christeas[1]

[1] *Department of Physics, Arizona State University, Tempe, AZ 85287, USA*

Abstract

Naive intuition about scale decoupling breaks down in the presence of fermion masses. Kinematic enhancements can greatly extend the range where one needs to keep a finite mass in calculations to obtain a correct result at even the $\mathcal{O}(1)$ level. Treating a light fermion as massive though, leads to a known but somewhat obscure paradox, a seeming leading-order sensitivity to an arbitrarily small mass. We show how a proper formulation in effective field theory not only resolves the physical conundrum, but answers the very practical question of when fermion masses are required. This has important implications for the development of shower Monte Carlo above the weak scale.

1 Introduction

Mathematically, the fact that $\int_0^\infty \frac{m^2}{(k^2+m^2)^2} dk^2 = 1$ is utterly unremarkable. The physicist can recognize in this equation though, the seeds of catastrophe, even paradox. The integrand resembles a fermion propagator squared, with the numerator projected onto the terms from the "m" piece of $\not{p} + m$. The integration arises easily enough from loop momenta or phase space. Standard lore tells us to drop the masses of any fermions that are well below the energy scale of interest, unless one cares about the details of small power corrections. However, this simple integral reveals a strong discrepancy between zeroing out m before or after the calculation in question.

This noncommutation of integration and the limit $m \to 0$ was recognized long ago by Smilga and dubbed a "quasiparadox" of massless QED [1]. In fact, the issue plagues general theories of relativistic fermions. One can push to an absurd limit by imagining a particle with $m = 10^{-10^{10^{10}}}$ GeV. That would be significantly below current limits on the photon, and yet keeping the m piece of its propagator numerator seemingly leads to an unsuppressed perturbative contribution. By similar logic, one could imagine using the LHC to determine the Standard Model neutrino masses.

The physical resolution arises from thinking about the production of light fermions at high energies. In the immediate aftermath of the collision, the fermion is not in an approximately on-shell, asymptotic state (outside of a tiny corner of phase space). Instead, it begins life as a wavepacket that evolves in time, eventually decaying to the physical fermion plus a cloud of radiation. It is these emissions of other particles that introduce terms of the form $\frac{m^2}{(k^2+m^2)^2}$. Thus, if we want to experimentally determine the mass then we must observe this radiation. Sensitivity to the enhanced part of the integrand $\frac{m^2}{(k^2+m^2)^2}$, where the mass effects are important, requires $k_\perp^2 \lesssim m^2$. As we will detail in calculations below, what enters the denominators in our calculations of interest provides a condition on transverse momentum. However, as Smilga points out, generating such "ultracollinear" radiation takes time, and for a particle with energy E and flight path of length L, there is a lower cutoff on the emission angle of $\theta \sim 1/\sqrt{E\,L}$. We therefore pick up a contribution

$$\int_{k_{\perp\min}} \frac{m^2}{(k_\perp^2 + m^2)^2} dk_\perp^2 \sim \frac{m^2}{E^2\theta^2} \sim \frac{m^2\,L}{E}, \tag{1}$$

with $k_{\perp\min} \sim \sqrt{E/L} \gg m$. This resolves the strong form of the paradox as sensitivity to particle masses now depends sensibly on the size of the experiment. Independent of detailed technical considerations, this places a lower limit on fermion mass measurable by the LHC, $m_{\mathrm{LHC,\,min}} \sim \mathcal{O}(100)$ eV.

We have avoided a deep physical conundrum, and yet the situation remains unsatisfying. Taken at face value, the result of Eq. 1 implies that we still must track fermion masses at scales substantially below those of our process's center of mass energy. For computations at the LHC, one might worry that every neglected charged-lepton mass results in a sizable error. It is the purpose of the present paper to allay these fears, in general, but also to clarify the energy regime where ultracollinear effects are important, providing a systematic treatment of how to include them. This will show where masses must be kept in a practical computation.

We see from Eq. 1 that the relevant question concerns the scale of $k_{\perp\min}$. The physical size of the experiment provides *an* IR cutoff, but it is not necessarily *the* IR cutoff for the problem at hand. For example, light fermions produced well above the weak scale have the ability to radiate on-shell electroweak bosons. Thus, in this regime, particle virtualities are regulated in the IR by something at least at the order of m_W, which dominates all fermion masses except the top quark. Below this scale, with weak particles integrated out, but still well above the fermion mass, m_f, one must nonetheless account for the vanishingly small phase space to produce nearly on-shell light particles in high-energy processes. This is quantified by the ubiquitous Sudakov factors found in observables that quantify the (non)showering of radiation.

Unlike collisions (or explosions) of macroscopic objects, which tend to spray debris in all directions, the scattering of high-energy particles produces well-collimated jets of radiation. One finds an explanation for this phenomenon in the structure of Feynman diagrams. The internal propagators of the showering particles provide kinematic enhancement if they are nearly on-shell, as is typically the case for soft or collinear emissions. This provides a perturbative origin for jets, but it leaves a further puzzle: *why then isn't every emission as collinear as possible?*. Instead, the process is dominated by "strong-ordering," as implemented in shower Monte Carlo like Pythia by either angular or virtuality ordering, with the widest or most off-shell emissions coming first [2]. The differential cross section one gets by demanding the absence of real emissions must still account for kinematically-enhanced virtual corrections to all orders of perturbation theory. This gives the overall Sudakov factor, which schematically modifies a Born-level cross section $d\sigma_0 \to d\sigma_0 \exp\left[-C\frac{\alpha}{\pi}\log\left(Q^2/m_{\mathrm{IR}}^2\right)\right]$ in the toy Yukawa theory we study in this work of interest, where C is an $\mathcal{O}(1)$ number and m_{IR} is parametrically near the scale of observation, typically a particle mass or kinematic cut. Gauge theories famously have stronger "double-log" soft-collinear enhancements, giving $d\sigma_0 \to d\sigma_0 \exp\left[-C\frac{\alpha}{\pi}\log\left(Q^2/m_{\mathrm{IR}}^2\right)^2\right]$. These overall terms thus give the "no-split" probability, as nicely summarized in [3]. We thus see that for particles with masses $m_f \ll Q$, the rate to produce them from the high-energy collision exactly on-shell gets highly suppressed.

The Sudakov factors show that even in the absence of unambiguous lower physical scales like masses, the physics of jets dynamically produces a soft scale, the "jet mass" (m_J), that can be parametrically large compared to m_f. We propose in this work, a threshold of $\exp\left[-C\frac{\alpha}{\pi}\log\left(Q^2/m_{\mathrm{IR}}^2\right)^a\right] < 50\%$, where $a = 1, 2$ depending on the theory. If the threshold is crossed for $m_{\mathrm{IR}} > m_f$, then the particle is more likely to have split than not, and one then works in an effective field theory (EFT) where the fermion is treated as massless, which we take to be SCET1 for the appropriate particles (*cf.* [4–7] for the foundational papers on SCET, or Soft-Collinear Effective Theory). Once one gets to a jet scale where $\exp\left(-C\frac{\alpha}{\pi}\log\left(m_J^2/m_f^2\right)\right) > 50\%$, then at m_J one matches to a SCET2 theory where that particular fermion is massive. Lastly, if one drops below m_f, then one integrates out the particle.[1] This approach thus systematically improves the crude approx-

[1] We leave it for future exploration whether data is best explained by some other threshold criteria, like having a no-split probability $< 32\%$, 5 %, *etc.* Regardless of the details, we necessarily have a regime where fermions need to be included as propagating degrees of freedom, but with their finite masses accounted for.

imation of taking a particle to either be massless at high-energies or integrated out at low energies. Letting the Sudakov factor decide when to keep m_f in a calculation can greatly modify the naive estimate that once $m_f^2/Q^2 \ll 1$, we are effectively massless. In particular, we see that the strength of the interaction must be accounted for. For example, if it is very feeble, then producing a light particle in a high-energy process can lead directly to on-shell states since the amplitude to radiate anything would then be highly suppressed.

The natural arena for the correct treatment of ultracollinear physics is the regime of high-energy multi-particle emissions in colliders and astroparticle systems. In [8], the authors developed a general-purpose electroweak shower Monte Carlo, emphasizing the role that ultracollinear corrections play in the proper handling of both next-to-leading-log (NLL) resummation and next-to-leading-order (NLO) matching. They found that the (im)proper treatment of fermion masses affects observables at the the few-% level for $\mathcal{O}(10-100)$ TeV colliders, which is well within the scope of current theoretical precision. Furthermore, the discrepancy can grow with energy scale, meaning that astroparticle studies at even higher energies are especially sensitive. Since the focus of [8] is a practical realistic shower algorithm, they did not generalize the conceptual issues and questions that surround ultracollinear enhancement. This is our focus. We even find in a simple Yukawa theory that finite-mass contributions can even lead to larger, $\mathcal{O}(1)$ numerical shifts to observables.

The calculation of showering radiation is done by Monte Carlo in practice. It is a useful question for future study to what extent finite mass effects and the ultracollinear regime are receiving proper treatment. It is not the case that practical shower routines have manifestly & universally accounted for this physics. *Its correct implementation would certainly be in keeping though, with the present sophistication of modern parton showers*. For example, recent efforts of the community include further QCD and EW corrections along with their resummation [9], [10], [11], [12], [13], non-perturbative physics such as hadronization, FSR (fragmentation functions) and ISR (parton distribution functions) modeling [10], [11], [13], and taking into account mass effects of particles [12], [13], [14]. Additionally, by going beyond next-to-leading-log (NLL) resummation in observables dominated by soft emissions, [15] fills an important gap in building general NLL parton showers. When fragmentation functions are considered, especially in EW theory, the massless case is strictly considered so far [9,10]. There is also the matter of better precision and higher accuracy in the computation which is related to developing the computational tools and algorithms used for the analysis. More specialized approaches have been examined that include SCET [14], the complex mass scheme to treat finite width effects [12], [14], or alternative jet techniques [14], [16]. Good agreement with experimental results have been achieved so far [13], [14]. *A priori*, except for [8], the explicit discussions of mass in these references are different from our ultracollinear concern. Along with finite-width effects, we can treat them systematically within EFT, where fragmentation functions enter as particular operators.

The aim of the present work is to focus on the subtle physics of having $m_f^2/(k_\perp^2+m_f^2)^2$ contributions to observables, disentangling them from the many other technical complications that arise in properly showering the full Standard Model. To this end, we work with a "simplest nontrivial example," a Yukawa theory consisting of one massless scalar ("higgs"), one massless fermion ("bottom"), and one fermion ("top") with mass M_t. We use these particle names hereafter, with the understanding that they are *not* strictly the Standard Model versions. This allows for three different regimes of jet physics which depend on the energy scale relative to M_t (*cf.* Fig. 1). We will see that properly accounting for the mass in the intermediate-regime theory requires inclusion of the top width, Γ_t, in its propagator. This shifts splitting functions at $\mathcal{O}(\alpha_y)$ and is generally more important numerically and more universal (since this effect occurs for massive particles of any spin) than the ultracollinear terms with which we motivated our study. However, we continue to emphasize the ultracollinear regime as its treatment is more subtle and forces one into very careful thinking about the physics of massive particles in high-energy processes.

The structure of the paper is as follows, in Section 2 we review the calculational setup; Appendix A provides further details of the fragmentation formalism, which allows us to calculate observables in the presence of multiple emissions. Section 2 also gives the specific implementation of SCET

in our Yukawa theory. This EFT allows for efficient computation of the fragmentation quantities across a wide variety of scales and the resummation of large logarithms that results. The treatment of SCET soft modes is discussed in Appendix B. In Section 3, we sketch some of the calculational ingredients, leaving details for Appendix C. Having obtained our quantities of interest, we combine them at leading-log (LL) and a subset of NLL corrections with NLO matching in Section 4 and Appendix D before presenting numerical results & concluding in Section 5.

2 A Toy Theory for Ultracollinear Physics

We begin by specifying our observable of interest. Thanks to factorization, we are able to write our normalized differential decay rate as a convolution of two separate functions, the Wilson Coefficient C_i and the Fragmentation Function D_{ij} where we capture the high-, and low-energy physics, respectively,

$$\frac{1}{\Gamma_0}\frac{d\Gamma_{j+X}}{dz} = \int_z^1 \frac{dx}{x} C_i\left(x/z\right) D_{ij}\left(z\right). \tag{2}$$

We are thus interested in the differential rate to observe particle j, carrying a fraction, z, of its mother's plus-momentum, where $p^+ = E + p^z$. Though we have written it as the decay rate of some heavy particle, it is trivial to replace $\Gamma \to \sigma$ and discuss instead a normalized differential cross section. The normalization factor Γ_0 is typically determined from the total decay rate for a process of interest or sometimes its tree-level value. It needs specification for a particular calculation. In our case, Γ_0 is the all-orders two-body decay rate of the heavy particle.

In Appendix A, we provide a brief recap of the fragmentation formalism and its implementation in SCET. In particular, the D_{ij} fragmentation functions are computed as operator matrix elements. When subject to only leading-log running in the massless case, they give the probability for i partons to split into j partons. However, we will see that this interpretation breaks down at higher perturbative order and in the presence of finite particle width. By matching our calculation at some scale, we determine the Wilson coefficients, C_i for each species of parton. This gives a boundary condition for renormalization group (RG) evolution. The usual invariance of the observable, LHS, with change in renormalization scale gives us equations for C_i or D_{ij} that we can use to run to other scales, resumming large logarithms in the process. This is further detailed in Section 4.

Our toy theory consists of three particles inspired by their Standard Model namesakes: 1) one massive Dirac fermion, *top*, that is in a doublet of an SU(2) global symmetry; 2) one massless SU(2)-singlet Dirac fermion, *bottom*; 3) one masslesss SU(2)-doublet complex scalar, $higgs$. They have one flavor off-diagonal Yukawa interaction,

$$\mathcal{L} = \bar{t}\left(i\partial\!\!\!/ - M_t\right) t + \bar{b}\left(i\partial\!\!\!/\right) b + (\partial^\mu h)^\dagger(\partial_\mu h) + y_t\, h\, \bar{t}\, b + h.c. \tag{3}$$

To initiate the fragmentation process, we produce tops at high-energy. Though the fragmentation functions and splitting functions that evolve them are agnostic to the production details, for concreteness, we assume a scalar, X, with $M_X \gg M_t$ that decays exclusively to $t\bar{t}$ via,

$$\mathcal{L}_X = y_X\, X\, \bar{t}\, t. \tag{4}$$

At a qualitative level, because the X decays into a significantly lighter sector, what it ultimately produces are jets of $t,\, b\, h$ particles. To accurately describe their contents, we work in four different energy regimes (*cf.* Figure 1): 1) A full theory consisting of eqs. 3 & 4, along with X kinetic terms; matching to this determines the high-energy boundary conditions for $C_{t,b,h}$. 2) Integrating out the X, the kinematics of its top-particle decay products at very high energies is dominated by collinear splittings. We can easily resum the resulting large-logs by working in SCET1 with all massless particles. 3) As the jets evolve and particle fragmentation decreases energies, we enter a regime where

we can no longer neglect M_t. This is described by SCET2 with a soft scale $\sim M_t$. 4) Lastly, passing below M_t, we integrate out the top, leaving a theory of $b,\ h$ particles that only interact via irrelevant operators.

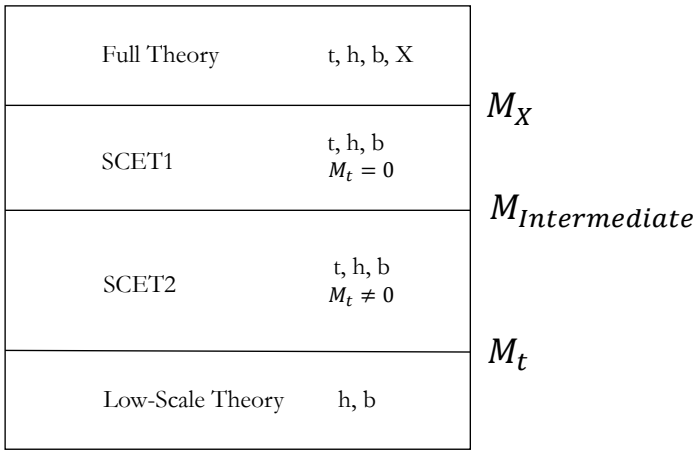
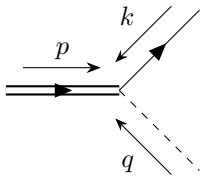


Figure 1: Schematic description of our sequence of theories

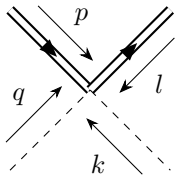
Our interest is only the collinear sectors of SCET1 and SCET2. We write here the SCET2 collinear lagrangian, from which one can easily get the SCET1 collinear sector by setting $M_t = 0$.

$$\begin{aligned}
\mathcal{L}_{\text{SCET}} = &\bar{t}_n \left\{ i\,(n\cdot\partial) + i\partial\!\!\!/_\perp \frac{1}{\bar{n}\cdot\mathcal{P}} i\partial\!\!\!/_\perp \right\} \frac{\bar{n}\!\!\!/}{2} t_n - M_t^2\, \bar{t}_n \frac{1}{\bar{n}\cdot\mathcal{P}} t_n + \\
&+ \bar{b}_n \left\{ i\,(n\cdot\partial) + i\partial\!\!\!/_\perp \frac{1}{\bar{n}\cdot\mathcal{P}} i\partial\!\!\!/_\perp \right\} \frac{\bar{n}\!\!\!/}{2} b_n + (\partial_\mu h_n)^\dagger (\partial^\mu h_n) \\
&+ y\,\bar{t}_n \left\{ (i\partial\!\!\!/_\perp - M_t) \frac{1}{\bar{n}\cdot\mathcal{P}} h_n - h_n \frac{1}{\bar{n}\cdot\mathcal{P}} i\partial\!\!\!/_\perp \right\} \frac{\bar{n}\!\!\!/}{2} b_n \\
&+ y\,\bar{b}_n \left\{ i\partial\!\!\!/_\perp \frac{1}{\bar{n}\cdot\mathcal{P}} h_n^\dagger - h_n^\dagger \frac{1}{\bar{n}\cdot\mathcal{P}} (i\partial\!\!\!/_\perp + M_t) \right\} \frac{\bar{n}\!\!\!/}{2} t_n \\
&- y^2 \left\{ h_n \bar{t}_n \frac{1}{\bar{n}\cdot\mathcal{P}} \frac{\bar{n}\!\!\!/}{2} t_n h_n^\dagger + h_n^\dagger \bar{b}_n \frac{1}{\bar{n}\cdot\mathcal{P}} \frac{\bar{n}\!\!\!/}{2} b_n h_n \right\}. \qquad (5)
\end{aligned}$$

By the usual convention the label momentum operator $\bar{n}\cdot\mathcal{P}$ only acts on fields to its right. Also by standard SCET practice, the two lightcone vectors $n^\mu, \bar{n}^\mu$ are defined as $n^2 = \bar{n}^2 = 0$ and $n\cdot\bar{n} = 2$. They form a lightcone basis where we can express any (momentum) vector as $p^\mu = (\bar{n}\cdot p, n\cdot p, p_\perp) = (p^+, p^-, p_\perp)$. This lagrangian gives the following Feynman rules:



$$= (+iy)\left[\left(\frac{\not{k}_\perp}{k^+} - \frac{\not{p}_\perp}{p^+}\right) - \frac{M_t}{p^+}\right]\frac{\not{n}}{2}$$



$$= (-iy^2)\left(\frac{1}{p^+ + k^+}\right)\frac{\not{n}}{2}.$$

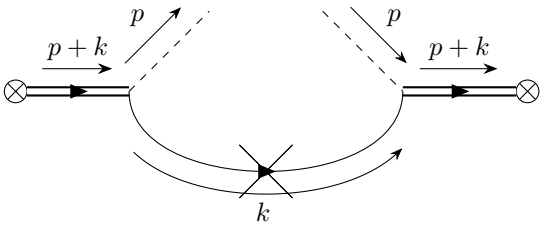
The double-line indicates a t, the single line a b, and the dashed line an h. One should note that these Feynman rules assume all incoming momenta. In practice, we will only need the three-point vertex above. To the order we are working, the four-point vertex[2] could only contribute in a virtual process, but as we show in Section 3.1, it does not affect the fragmentation functions. One might also worry about the need for soft interactions. However, as we demonstrate in Appendix B, in SCET1, where we can neglect M_t, the soft limit of the collinear interactions given in Eq. 5 itself reproduces the soft limit of the full-theory contributions to fragmentation amplitudes[3]. In SCET2, the soft-collinear vertex that would contribute is power-suppressed. Thus, we are left with a straightforward set of all-collinear calculations we detail in Section 3 and Appendix C.

3 Finite Mass Effects at LO & Beyond

In Section 2, we discussed the fragmentation function, D_{ij}, our main calculational quantity of interest. We provide here an overview of its computation in SCET1/SCET2, and how finite particle masses affect results beyond mere power corrections. In using the fragmentation formalism to compute particle production at a variety of energy scales, it becomes necessary to run the operators whose matrix elements give D_{ij}. As we will show in Section 4, the anomalous dimension, γ_{ij}, for D_{ij} is

$$\gamma_{ij} = -\frac{dD_{ik}}{d\log(\mu)}\,(D)^{-1}_{kj}. \tag{6}$$

Fragmentation Functions are calculated through a fixed order diagram based on the definitions in Eq. 50. For example, we can take the D_{th} computation at $\mathcal{O}(\alpha_y)$:

$D_{th} =$ 

(7)

[2]There is also another four-point vertex that is the same as the one written above, where the t is substituted by b.
[3]Zero-bin contribution is equal to the soft contribution.

By construction, this process consists of an energetic mother parton fragmenting to a daughter. In the SCET, the mother is created by a collinear field. We can investigate various contributions for the other states from the soft and collinear sectors by setting their momenta in these regions. By energy conservation, one or both of the daughter particles is necessarily collinear. In SCET2 the soft emission is power suppressed while in SCET1 the soft contribution is cancelled out by the zero-bin, as shown in B. Therefore, only the all-collinear process is the leading one of interest. As a consequence of the power counting in SCET1, we drop M_t as a scale below that of theory's soft scale. To regulate IR divergences, we impose a fixed μ_{IR} cutoff. Our SCET1 splitting functions just reproduce the standard massless splitting functions [17].

Once the scale is sufficiently close to M_t that on-shell top production is likely, we match to the intermediate-scale theory, SCET2. We can no longer neglect M_t, and its presence modifies our results at LO because of its finite width, and at NLO through the ultracollinear contributions. When the shower reaches the scale M_t, we integrate out the top. In this Low-Energy Theory, there are only trivial fragmentation functions for the remaining particles, $\sim \delta(1-z)$.

3.1 Outline of the FF Computation

The computation of a real-emission fragmentation function will be outlined with an example of D_{th} in SCET2.[4] Starting from the definition of the Fragmentation Function given in Eq. 50,

$$D_{th}(z) = \frac{z}{4Q}\,\mathrm{Tr}\int d\Pi_X\,\delta\left(1-\frac{p_X^+ + p^+}{Q}\right)\langle 0|\frac{\bar{n}\!\!\!/}{2}t(p+p_X)|h(p)\,X\rangle\langle h(p)\,X|\bar{t}(p+p_X)|0\rangle\Big|_{p_\perp=0}, \tag{8}$$

where we have used $N_c = 2$ since the initiating t is a doublet of SU(2).[5] At $\mathcal{O}(\alpha_y)$ we need to compute the graph shown at the start of this section. As we can see, the graph is cut on the observed daughter and recoiling particle lines, which puts these particles on shell. The unintegrated momentum p is thus on-shell by construction, and we note that we have rotated to the frame with $p_\perp$ = 0. This also introduces a δ-function setting $k^2 = 0$. After imposing all δ-function constraints, one straightforward integral is left, for which we use dimensional regularization to remove its UV divergence,

$$D_{th}(z) = \alpha_y \int \frac{d^{2-2\epsilon}k_\perp}{(2\pi)^2}\,\frac{z\,k_\perp^2 + \frac{(1-z)^2}{z}M_t^2}{\left(k_\perp^2 - \left(\frac{1-z}{z}\right)M_t^2\right)^2 + \left(\frac{1-z}{z}\right)^2 M_t^2\,\Gamma_t^2}. \tag{9}$$

We recognize that the M_t^2 term in the numerator, combined with the denominator, provides an ultracollinear correction to the LL result. Additionally, the kinematics allow the top quark propagator to hit the pole at $(p+k)^2 = M_t^2$. By working in the complex mass scheme [18], [19], and deforming $M_t \to M_t - i\,\Gamma_t$, we avoid a divergence, but do pick up a contribution enhanced by $1/\Gamma_t$. At LO, $\Gamma_t = \frac{\alpha_y M_t}{4}$. The physics of such a large correction makes sense since we are necessarily in the region of on-shell top production. In addition to the usual perturbative splitting process that creates b, h particles, the top can also decay to them, which it does with a probability that approaches 100%.[6] Upon performing the integral in Eq. 9 and renormalizing in $\overline{\mathrm{MS}}$, we reach the

[4] One can obtain the analogous SCET1 D_{th} result by setting $M_t = 0$ in the SCET2 integrand and imposing a fixed IR cutoff on perp momentum, μ_{IR}. For notational simplicity though, we always call the IR scale in the log argument M_t.

[5] The X in Eq. 8 is meant to represent a generic final state of the SCET2 theory. It is not the heavy X particle we used in the full theory as a source of $t\bar{t}$.

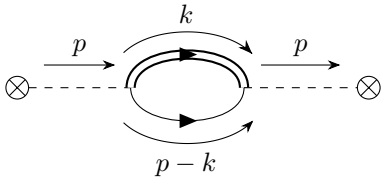
[6] If we were interested in the details of top decay product distributions near its mass pole, then it may be necessary to resum the large logs of Γ_t/M_t. The Unstable Particle Effective Theory of [20] is one straightforward formalism for tackling this problem. However, we take the observables from our processes of interest, jet production from energetic collider or astroparticle processes, as sufficiently inclusive that this resummation is beyond our scope. It is also pointed out in Ref. [21] that including finite widths breaks gauge invariance. Gauge theories are beyond the scope of this work, but are an important future direction.

final expression,

$$D_{th}(z) = \frac{\alpha_y}{4\pi}\left[z\,\log\left(\frac{\mu^2}{M_t^2}\right) - z\,\log\left(\frac{1-z}{z}\right) - 1 + \frac{\pi\,M_t}{\Gamma_t}\right], \tag{10}$$

where μ is the renormalization scale. The NLL terms, with no explicit $\log(\mu^2/M_t^2)$ or $1/\Gamma_t$ are due to the presence of finite M_t in both numerator and denominator. These are the ultracollinear contributions. Lastly, the $1/\Gamma_t$ term modifies the running at LO from the D^{-1} contribution to γ_{ij} (*cf.* Eq. 6).

It is also necessary to obtain the self-fragmentation functions and their running. These are just the wavefunction renormalization factors for the fields in our theory.[7] Since these computations take a different form, we give an example here with D_{hh} ($= Z_h\,\delta(1-z)$, where Z_h is our $\overline{\text{MS}}$ wavefunction renormalization).



(11)

For the self-fragmentation function, the contribution of this graph to the residue to $\mathcal{O}(\alpha_y)$ is calculated as

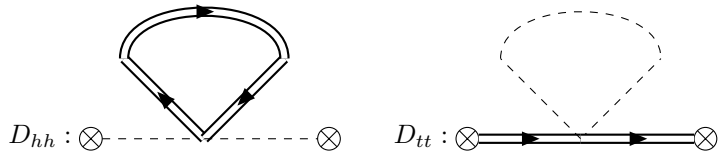
$$Z_h = \left(1 - \frac{1}{p^+}\frac{d}{dp^-}\Sigma(p)\Big|_{p^2=0}\right)^{-1}, \tag{12}$$

where $\Sigma(p)$ is the expression of the loop. We again compute in the SCET2 theory with dimensional regularization and $\overline{\text{MS}}$, obtaining

$$D_{hh}(z) = \left(1 - \frac{\alpha_y}{2\pi}\left[\log\left(\frac{\mu^2}{M_t^2}\right) + \frac{1}{2}\right]\right)\delta(1-z) \tag{13}$$

Similar to the real emission graph, the log coefficient is the leading-log term. The other $\mathcal{O}(\alpha_y)$ contribution arises from the finite M_t that we track in the SCET2 theory. The calculations for the other real & virtual-emission fragmentation functions proceed similarly. In Appendix C, one can find more calculational details for real and virtual fragmentation functions. We collect the results for the entire theory in Table 1 below.

In principle, there are other diagrams contributing to D_{hh} and D_{tt} of the following form[8]:



(14)

These graphs however, have trivial contribution at $\mathcal{O}(\alpha_y)$ since $\dfrac{d}{dp^-}\Sigma(p)$ would give 0.

[7] Since the top is unstable, its pole and residue are complex valued. Thus, the factor that enters the fragmentation function is $|\sqrt{Z_t}|^2$. The imaginary part of $\sqrt{Z_t}$ only enters at $\mathcal{O}(\alpha_y)$, and as we only work to that order, in practice we can drop it.

[8] There are also two extra graphs, one contributing to D_{hh} and one to D_{bb}, where they are the same but t is substituted by b.

3.2 Summary of Results

We tabulate here all results that contribute to matching of Wilson coefficients at $\mathcal{O}(\alpha_y)$ as we run from scales with $Q \gg M_t$ to energies below M_t, where the top is integrated out of the theory. This includes shifts to the LO results from the finite top width, Γ_t. We also include those parts of the anomalous dimensions that arise from the finite M_t in the SCET2 regime and contribute to running at NLL order. These are ultracollinear contributions in the sense that their existence requires the inclusion of nonzero M_t in the initial phase of the calculation, but they persist if one takes $M_t \to 0$ at the end of it. We dub these terms "NLL-UC". We stress that we have not performed a full NLL resummation, as this requires two-loop results in the toy theory that are beyond our scope. Nonetheless, in presenting numerical results in Section 5, we show the shifts incurred in the Wilson coefficients by including the NLL-UC terms. This is to demonstrate the size of the contributions that would be missed under a more naive treatment that keeps the top massless until it is integrated out.

F.F.	Leading-Log	Finite M_t corrections
$D_{tb}(z) =$	$\frac{\alpha_y}{4\pi}(1-z)\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{4\pi}\left((1-z)\log\left(\frac{z}{1-z}\right)-1+\pi\frac{M_t}{\Gamma_t}\right)$
$D_{th}(z) =$	$\frac{\alpha_y}{4\pi}z\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{4\pi}\left(z\log\left(\frac{z}{1-z}\right)-1+\pi\frac{M_t}{\Gamma_t}\right)$
$D_{bt}(z) =$	$\frac{\alpha_y}{2\pi}(1-z)\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{2\pi}\left((1-z)\log\left(\frac{z^2}{1-z}\right)+z\right)$
$D_{bh}(z) =$	$\frac{\alpha_y}{2\pi}z\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{2\pi}\left(z\log(z)+1-z\right)$
$D_{hb}(z) =$	$\frac{\alpha_y}{2\pi}\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{2\pi}\left(\log(z)-1+z\right)$
$D_{ht}(z) =$	$\frac{\alpha_y}{2\pi}\log\left(\frac{\mu^2}{M_t^2}\right)$	$+\frac{\alpha_y}{2\pi}\left(\log\left(\frac{z^2}{1-z}\right)-z\right)$
$D_{tt}(z) =$	$\left[1-\frac{\alpha_y}{8\pi}\log\left(\frac{\mu^2}{M_t^2}\right)\right]\delta(1-z)$	0
$D_{hh}(z) =$	$\left[1-\frac{\alpha_y}{2\pi}\log\left(\frac{\mu^2}{M_t^2}\right)\right]\delta(1-z)$	$-\frac{\alpha_y}{2\pi}\left(\frac{1}{2}\right)\delta(1-z)$
$D_{bb}(z) =$	$\left[1-\frac{\alpha_y}{4\pi}\log\left(\frac{\mu^2}{M_t^2}\right)\right]\delta(1-z)$	$-\frac{\alpha_y}{4\pi}\left(\frac{3}{2}\right)\delta(1-z)$

Table 1: The fragmentation functions to $\mathcal{O}(\alpha_y)$. In the high-scale theory (SCET1), we only get the results in the LL column, and M_t is strictly an IR cutoff on perp momentum. In the intermediate-energy theory (SCET2), we pickup full M_t dependence. Since $\Gamma_t = \frac{\alpha_y M_t}{4}$, we see that all entries in the "Finite M_t corrections" column survive in the $M_t \to 0$ limit, even though we need to keep $M_t \neq 0$ for their initial determination.

As we will further see in Section 4, it is useful to write the Fragmentation Functions in matrix

form:

$$D_{ij} = \begin{pmatrix} D_{hh} & D_{hb} & D_{ht} \\ D_{bh} & D_{bb} & D_{bt} \\ D_{th} & D_{tb} & D_{tt} \end{pmatrix} \tag{15}$$

In the high-scale SCET1 theory, where all particles are treated as massless, the LO "fragmentation matrix" is just $\delta(1-z)$ terms on the diagonal. As discussed above, after passing to the intermediate-scale SCET2 theory with finite M_t & Γ_t, the Fragmentation Functions at LO are modified to the following expression:

$$D_{ij}^{LO} = \begin{pmatrix} \delta(1-z) & 0 & 0 \\ 0 & \delta(1-z) & 0 \\ 1 & 1 & \delta(1-z) \end{pmatrix} \tag{16}$$

The off-diagonal, z-independent terms arise from the near-certain decay of the unstable tops. We give the full contributions up to NLO in $\mathcal{O}(\alpha_y)$ in Table 1. In Mellin Space, the total contribution in the intermediate-mass theory (SCET2) to NLO is

$$\begin{aligned} D_{ij} &= D_{ij}^{\text{LO}} + D_{ij}^{M_t=0,\,\text{NLO}} \log\left(\frac{\mu^2}{M_t^2}\right) + D_{ij}^{M_t\neq 0,\,\text{NLO}} = \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{s} & \frac{1}{s} & 1 \end{pmatrix} + \\ &+ \frac{\alpha_y}{2\pi} \begin{pmatrix} -1 & \frac{1}{s} & \frac{1}{s} \\ \frac{1}{(s+1)} & -\frac{1}{2} & \frac{1}{s(s+1)} \\ \frac{1}{2(s+1)} & \frac{1}{2s(s+1)} & -\frac{1}{4} \end{pmatrix} \log\left(\frac{\mu^2}{M_t^2}\right) + \\ &+ \frac{\alpha_y}{2\pi} \begin{pmatrix} -\frac{1}{2} & \frac{-2s-1}{s^2(s+1)} & \frac{H(s)}{s} - \frac{2}{s^2} - \frac{1}{s+1} \\ \frac{1}{s(s+1)^2} & -\frac{3}{4} & \frac{s(s+1)H(s)+s^3-4s-2}{s^2(s+1)^2} \\ \frac{H(s+1)}{2(s+1)} - \frac{1}{2(s+1)^2} - \frac{1}{2s} & \frac{sH(s)-(s+1)^2}{2s^2(s+1)} & 0 \end{pmatrix}, \end{aligned} \tag{17}$$

where $H(s)$ is the generalized harmonic number. One can easily obtain the high-scale (SCET1) fragmentation matrix from Eq. 17 by dropping the off-diagonal elements in D_{ij}^{LO}, along with the $D_{ij}^{M_t\neq 0,\,\text{NLO}}$ term. As we discuss next, the running and log resummation necessary for showering particles in the different energy regimes of our theory is a 'turn-the-crank' extension of the work described thus far.

4 Running and Matching

The explicit μ-dependence we see in the renormalized fragmentation functions of Table 1 lets us use RG evolution to run our Wilson coefficients down from their UV-determined boundary conditions to IR values. In doing so, we automatically sum the LL contributions to our observable to all orders. Indeed, one of the primary benefits of calculating in SCET is the straightforward way that this resummation follows from the RGE of its operators. Recall from Figure 1 that our calculation has four regimes: full Yukawa theory, SCET1 with $M_t = 0$, SCET2 with finite M_t, and lastly a theory below M_t with the top integrated out. The full description thus involves three matching steps and two stages of RG running.

The full-theory/SCET1 matching occurs at a high-scale we call M_X, thinking of it as the mass of a very particle that decays exlusively to tops. The SCET1 running is standard and has no top mass

effects by construction. Then, at some scale $M_{\text{intermediate}}$, top particles are more likely than not to be produced on-shell and we must account for finite M_t. Since M_t is now comparable to our soft scale, we have a SCET2 theory. The matching of Wilson coefficients from SCET1 is trivial. However, for all scales below $M_{\text{intermediate}}$, they run with anomalous dimensions modified by finite M_t.[9] Lastly, in matching at M_t, one necessarily accounts for the finite-width and ultracollinear contributions due to M_t in the fragmentation functions.

4.1 RG Evolution for Wilson Coefficients and Fragmentation Functions

Although it is standard, we review here the evolution equation derivation for fragmentation functions and their corresponding Wilson coefficients. Our discussion largely follows that in [22]. We will find that for an unstable particle, a careful accounting is needed even for a correct treatment at $\mathcal{O}(1)$. If we take an observable that factorizes like our differential decay rate of interest (*cf.* Eq. 2), we can write it as follows,

$$\sigma'_j = C_i(\mu) D_{ij}(\mu), \tag{18}$$

where σ'_j is a differential rate involving particle j. Since this quantity must be μ-independent,

$$\frac{d\sigma'_j}{d\log(\mu)} = 0 \quad \Rightarrow \quad \frac{dC_i(\mu)}{d\log(\mu)} D_{ij}(\mu) + C_i(\mu) \frac{dD_{ij}(\mu)}{d\log(\mu)} = 0 \tag{19}$$

It is useful to first examine the fragmentation function; one can write the bare $D_{0,ij}$ in terms of the renormalized D_{kj},

$$D_{0,ij} = Z_{ik} D_{kj}, \tag{20}$$

where Z_{ij} are the renormalization factors. This gives us an additional constraint to setup the RG equation since bare quantities are necessarily independent of the renormalization scale,

$$\frac{dD_{0,ij}}{d\log(\mu)} = 0 \quad \Rightarrow \quad \frac{dZ_{ik}}{d\log(\mu)} D_{kj} + Z_{ik} \frac{dD_{kj}}{d\log(\mu)} = 0 \tag{21}$$

Solving for $\dfrac{dD_{kj}}{d\log(\mu)}$ gives

$$\frac{dD_{mj}}{d\log(\mu)} = -Z^{-1}_{mi} \frac{dZ_{ik}}{d\log(\mu)} D_{kj}. \tag{22}$$

We thus recognize the anomalous dimension matrix, $\gamma_{mk} \equiv Z^{-1}_{mi} \dfrac{dZ_{ik}}{d\log(\mu)}$, which gives

$$\begin{aligned} \frac{dD_{mj}}{d\log(\mu)} &= -\gamma_{mk} D_{kj} \Rightarrow \\ \frac{dD}{d\log(\mu)} &= -\gamma \cdot D \end{aligned} \tag{23}$$

[9] The SCET1/SCET2 transition is most simply handled by matching our SCET1 of interest to a fictitious SCET2 where the top mass is tuned by hand to be parametrically below its soft scale, and thus dropped. Having obtained the Wilson coefficients of this artificial theory, we then tune M_t to be larger until it coincides with the soft scale. This necessarily modifies operator dimensions, but it cannot change the Wilson coefficients, which are set by UV physics and are thus decoupled from the soft scale, including M_t effects.

Since we can compute the renormalized fragmentation functions D_{ij} straightforwardly, it is useful to invert the last equation to get

$$\boxed{\gamma = -\frac{dD}{d\log(\mu)} D^{-1}} \tag{24}$$

This form shows its utility in deriving the evolution equation for the Wilson Coefficients. Rewriting the equation for μ-independence of the observable (*cf.* Eq. 19), one gets

$$\frac{dC_i}{d\log(\mu)} D_{ij} = -C_i \frac{dD_{ij}}{d\log(\mu)}, \tag{25}$$

which allows us to introduce the anomalous dimension,

$$\frac{dC_i}{d\log(\mu)} D_{ij} = -C_i \left(-\gamma_{ik} D_{kj}\right). \tag{26}$$

We can then eliminate D_{ij} to get a flow equation for the Wilson coefficients phrased purely in terms of C and γ,

$$\begin{gathered} \frac{d\,C_l}{d\log(\mu)} = C_i \gamma_{il} \Rightarrow \\ \boxed{\frac{d\,C}{d\log(\mu)} = \gamma^T \cdot C} \end{gathered} \tag{27}$$

As mentioned above, we perform our running in two distinct theories. Even though they have the same particle content, their γs differ at the $\mathcal{O}(1)$ level due to the different scales they inhabit and whether or not one has a finite M_t. In "z-space" where we work explicitly with plus-momentum fractions, we construct our observable as a convolution,

$$\frac{1}{\Gamma_0} \frac{d\Gamma_{j+X}}{dz} = \int_z^1 \frac{dx}{x} C_i\left(x/z\right) D_{ij}\left(z\right). \tag{28}$$

The running equations for Wilson Coefficients are then given by

$$\frac{dC_l(z)}{d\log(\mu)} = \int_z^1 \frac{dx}{x} C_i(x/z)\, \gamma_{il}(z) \tag{29}$$

We can reduce this integro-differential equation to a mere differential equation by means of a Mellin Transform, defined by the following relations,

$$\left\{\mathcal{M} f\right\}(s) = \int_0^\infty z^{s-1} f(z) \qquad \text{(Mellin Transform)} \tag{30}$$

$$\left\{\mathcal{M}^{-1} F\right\}(z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} z^{-s} F(s) \qquad \text{(Inverse Mellin Transform)}, \tag{31}$$

with the property needed for simplifying convolution:

$$\left\{\mathcal{M} \int \frac{dx}{x} \quad f(x/z) g(z)\right\}(s) = F(s) G(s) \tag{32}$$

Thus Eq. 29 transforms into a simpler algebraic equation,

$$\frac{d\,C_l(s)}{d\log(\mu)} = C_i(s) \gamma_{il}(s). \tag{33}$$

For simplicity, we use the same notation for functions and their (inverse) Mellin transforms.

4.2 Running & Matching in the Toy Yukawa Theory

We can now apply this general approach to evolving fragmentation functions and resumming large logs to our specific theory of interest. We already have what we need in the Mellin space fragmentation function matrices given in Eq. 17. To compute our low-scale observables, we now just need to apply the formalism of Section 4.1. We first compute the SCET2 LL anomalous dimensions, which are just the $\mathcal{O}(\alpha_y)$ contributions. Using the notation of Eq. 17, we have

$$\begin{aligned}\gamma_{ij}^{\text{LL}} &= -\frac{dD_{ik}}{d\log(\mu)} D_{kj}^{-1}\Big|_{\mathcal{O}(\alpha_y)} = -2D_{ik}^{m=0,\,\text{NLO}} \left(D_{kj}^{\text{LO}}\right)^{-1} = \\ &= \frac{\alpha_y}{\pi} \begin{pmatrix} -\delta(1-z) & 1 & 1 \\ z & -\frac{1}{2}\delta(1-z) & 1-z \\ \frac{z}{2} & \frac{(1-z)}{2} & -\frac{1}{4}\delta(1-z) \end{pmatrix} \begin{pmatrix} \delta(1-z) & 0 & 0 \\ 0 & \delta(1-z) & 0 \\ 1 & 1 & \delta(1-z) \end{pmatrix}^{-1}.\end{aligned} \tag{34}$$

A simple way to handle the inverse matrix is to convert everything to Mellin space, and then transform back to z space. It is straightforward to get the LL SCET1 anomalous dimesnions from Eq. 34. We just need to zero out the off-diagonal entries in $D_{kj}^{\text{LO}\,-1}$. These terms arise from the Γ_t that follows from having finite M_t. The operator running thus undergoes an $\mathcal{O}(1)$ shift once our scale is sufficiently low that we evolve nearly on-shell particles.

Running at NLL just requires us to calculate γ_{ij} from Eq. 24 to $\mathcal{O}(\alpha_y^2)$. However, this result requires a full accounting of two-loop diagrams (equivalently, three-body phase space for real emissions) and is beyond our scope. The "ultracollinear" terms in D_{ij}, those that are M_t independent, but whose existence depends on keeping M_t in the initial calculation, do contribute to γ_{ij} at $\mathcal{O}(\alpha_y^2)$. When we present results in Section 5 we show Wilson coefficients that have undergone "NLL-UC" running, which includes just the ultracollinear contributions at subleading order. This is necessarily an incomplete result, but it shows the size of the effect that *would be missed* if one were to drop fermion masses in a regime where one should account for them.[10]

The $\mathcal{O}(\alpha_y^2)$ contribution to the anomalous dimension matrix that captures the ultracollinear terms is:

$$\begin{aligned}\gamma_{ij}^{\text{NLL-UC}} &= -\frac{dD_{ik}}{d\log(\mu)} D_{kj}^{-1}\Big|_{\mathcal{O}(\alpha_y^2),(\alpha_y^2\log(\mu))\to 0} \\ &= -\left(\frac{d\alpha_y}{d\log\mu}\frac{\partial D_{ik}^{m\neq 0,\text{NLO}}}{\partial\alpha_y} + 2D_{ik}^{m=0,\text{NLO}}\right)\left(D_{ik}^{\text{LO}} + D_{ik}^{m\neq 0,\text{NLO}}\right)^{-1}\Big|_{\mathcal{O}(\alpha_y^2)} \\ &= -\left(\frac{d\alpha_y}{d\log\mu}\frac{\partial D_{ik}^{m\neq 0,\text{NLO}}}{\partial\alpha_y}\cdot D_{kj}^{-1\,\text{LO}} - 2D_{ik}^{m=0,\text{NLO}}\cdot D_{kj}^{m\neq 0,\text{NLO}}\right),\end{aligned} \tag{35}$$

where we again classify terms using the terminology of Eq. 17. We recognize in Eq. 35 factors of the β-function of our coupling, which is just

$$\beta(\alpha_y) \;=\; \frac{7\alpha_y^2}{4\pi}. \tag{36}$$

[10] Naively computing γ_{ij} to $\mathcal{O}(\alpha_y^2)$ from our one-loop results in Table 1 generates some terms of the form $\alpha_y^2\log(\mu)$. In running the Wilson coefficients, these would integrate to $\alpha_y^2\log^2$ contributions. However, that is an order that one gets from exponentiating the LL result. If it were to survive, it would imply that the LL resummation fails. Since that is not the case as it would violate the μ-independence of our observable, a complete calculation of γ_{ij} to $\mathcal{O}(\alpha_y^2)$ must cancel any contributions of the form $\alpha_y^2\log$. Thus, we drop any such terms in our NLL-UC anomalous dimensions.

The Mellin transform of the matrix given by Eq. 35 is a complicated expression that we give in Appendix D. Nonetheless, it is straightforward to construct.

Having obtained what we need to calculate the effects due to running, we can proceed to consider matching. The basic idea is that our observable, $\frac{1}{\Gamma_0}\frac{d\Gamma_j}{dz}$ is the same in both theories at the matching scale.

$$\frac{1}{\Gamma_0}\frac{d\Gamma_j(\mu=\mu_M+\epsilon)}{dz} = \frac{1}{\Gamma_0}\frac{d\Gamma'_j(\mu=\mu_M-\epsilon)}{dz} \tag{37}$$

where j labels the observed particle, the LHS represents a higher-scale theory and the RHS is the lower-scale one. Using the factorized form of our differential decay rate (*cf.* 2) and going to Mellin space, we have

$$C_i(s,\mu_M)D_{ij}(s,\mu_M) = C'_k(s,\mu_M)D'_{kj}(s,\mu_M) \tag{38}$$

We note that this equation holds j by j, and thus that particle must necessarily exist in both theories. Solving for C'_k, we get

$$C'_k(s,\mu_M) = C_i(s,\mu_M)D_{ij}(s,\mu_M)\left(D'_{jk}(s,\mu_M)\right)^{-1}, \tag{39}$$

matching order by order in α_y.

In the tower of theories that takes us from the creation of our massive fermion, the top, to scales below its mass, where we integrate it out, we pass through three matching scales (*cf.* Fig. 1). We provide here an overview of the issues with each of these. In Section 5, we present a specific numerical implementation.

- Matching between the Full Theory and SCET1 at M_X:

  The main nontrivialites at this stage are the complications that arise when passing from a conventional quantum field theory to SCET. However, this is now a standard procedure involving a well-established effective field theory [4–7]. One simplification relative to the original QCD SCET is that we have no issues with gauge invariance in a pure Yukawa theory. By construction, the kinematically-enhanced fragmentation region of a collinear splitting of the top falls out of the leading-power SCET Feynman rule. It is this matching into SCET that sets the UV boundary conditions for our Wilson coefficients that will undergo further running. At LO, only tops arise from the X decay. However, at NLO it becomes possible to produce bottom or higgs as a hard, three-body decay, a production mechanism not captured the fragmentation of a top with the SCET vertex. Thus, matching to our observable, $(1/\Gamma_0)\,d\Gamma/dz$, we find

$$\begin{aligned} C_t(z,\mu=M_X) &= \delta(1-z) \\ C_b(z,\mu=M_X) &= \alpha_y\left(\frac{3z-2}{8\pi}\right) \\ C_h(z,\mu=M_X) &= \alpha_y\frac{z\left(-3z^2+z+2\right)+2\log(1-z)}{8\pi z^2}. \end{aligned} \tag{40}$$

- Matching between SCET1 and SCET2 at $M_{\text{Intermediate}}$:

  In principle, there are many shifts that can occur in passing from a SCET1 to a SCET2 theory. The former has a hierarchy between the virtualities of its soft and collinear sectors, and the latter does not. Additionally, the change in power counting of the soft sector also shuffles the (ir)relevance of interactions. Furthermore, the SCET2 theory is sensitive to the the finite value

of M_t by construction. Nevertheless, the C_i Wilson coefficients of our fragmentation operators enjoy many simplifications. As discussed above, the UV physics that sets C_i is decoupled from M_t. In practice, we match to a fictitious SCET2 with $M_t = 0$, and then tune M_t up to the SCET2 soft scale, which can have no effect on the C_i we obtained. Additionally, our fragmentation functions are computed from collinear fields alone. The SCET1 soft contributions are removed by zero-bin subtraction [23], while the SCET2 soft vertex is power-suppressed. This means that we are matching D_{ij} computed in the collinear regime of both theories with $M_t = 0$. Thus, we have

$$C_i(s, M_{\text{Intermediate}}) D_{ij}(s, M_{\text{Intermediate}}) = C'_k(s, M_{\text{Intermediate}}) D'_{kj}(s, M_{\text{Intermediate}}), \tag{41}$$

with the same D-matrices on both sides. Thus, we get $C_i = C'_i$, *i.e.* a trivial matching. It is the case though, that the NLL running in the SCET2 theory will be immediately different from that in SCET1 due to the finite M_t effects.

- Matching between SCET2 and the Low-Scale Theory at M_t:

  In some sense, this last matching is quite conventional; we integrate out a particle below its mass. In the context of fragmentation, our matching must account for the fact that the t had been a propagating source of h, b particles, and now is gone from the theory. This is where the finite-width contributions to D_{th} and D_{tb} come in (*cf.* Table 1). In matching to the Low Scale Theory with no t, they give one last, $\mathcal{O}(1)$ injection of h, b as we account for every t particle's decay into them. Once in the low-scale theory (LST), interactions only proceed through irrelevant operators. We thus get no leading-power contributions to fragmentation. Thus, we have a simple fragmentation matrix,

$$D_{ij\,\text{LST}}(s, \mu = M_t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{42}$$

  Since we only need to match h, b final states, we only need the SCET2 fragmentation matrix with the h, b columns. To NLO in Mellin space, these are:

$$D_{ij,2}(z)(s, \mu = M_t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{s} & \frac{1}{s} \end{pmatrix} + \frac{\alpha_y}{8\pi} \begin{pmatrix} -2 & -4\frac{2s+1}{s^3+s^2} \\ \frac{4}{s(s+1)^2} & -3 \\ 2\left(\frac{H(s+1)}{s+1} - \frac{1}{s} - \frac{1}{(s+1)^2}\right) & 2\left(\frac{\frac{sH(s)}{s+1} - s - 1}{s^2}\right) \end{pmatrix}, \tag{43}$$

  The triviality of $D_{ij\,\text{LST}}$ lets us easily compute the $C_{i\,\text{LST}}$ from Eq. 38.

5 Numerical Results & Conclusions

We have seen the subtle and important physics of tracking a fermion mass as one propagates it down in scale from its production to its eventual decay. It is nevertheless useful to provide a numerical example to see how these effects shift results quantitatively. In keeping with our Standard Model inspiration, we take M_t = 173 GeV and we set $y_t(M_t) = 1$. We make the b and h massless. For describing fragmentation, at scales immediately above M_t, we work in SCET2.

We can ask though, at what scale we can stop tracking M_t and instead pass to the SCET1 theory with $M_t = 0$. As follows from our discussion in Section 1, our condition to account for the fermion mass is the scale below which its "no-split" probability, *i.e.* the odds that we have produced a physical, on-shell particle are greater than 50%. We compute the Sudakov (no-split) factor for our t following

the presentation in [3]. That reference defines the splitting function, P_{ij}, as the z-space anomalous dimension with α_y/π stripped off. Inverse Mellin transforming $\gamma_{th}^{\rm LL}$ from Table 3, we get $P_{th} = 1/4 + z/2$. Thus, the Sudakov factor is

$$\Delta(Q, M_t) = \exp\left(-\int_{M_t^2}^{Q^2} \frac{\alpha_y(\sqrt{t})}{4\pi}\frac{1}{t}dt\right). \tag{44}$$

For the above values of y_t, M_t, we find that the value of Q for which $\Delta(Q, M_t) = 0.5$ is $Q = 1.5\times10^8$ TeV. We thus take this scale as the upper limit of our SCET2 theory. This may seem like a shockingly high scale at which to still be tracking finite M_t. After all, we are nine orders of magnitude larger and $(M_t/Q)^2$ effects are beyond negligible. However, the lesson of the ultracollinear terms is that naive power suppression breaks down as kinematic enhancements can provide compensating factors. Said differently, in the limit that $\alpha_y \to 0$, even if $M_X \gg M_t$, the $t\bar{t}$ pair in the X decays into will necessarily be on-shell as they will have no interactions that allow them to split and shower into other particles. Furthermore, since our Yukawa theory has only collinear, rather than soft & collinear divergences like gauge theories, the Sudakov factor is only single-log enhanced. This causes a relative suppression in the exponential argument, causing an extension of the regime where we need to track M_t.

For the case of real QCD, if we use the double-log $q \to gq$ Sudakov in [3] and plug in the observed values for top and bottom quark masses (m_t, m_b) along with α_s, we find that one should keep bottom quark masses in calculations up to $Q \approx 235$ GeV and top quark masses up to $Q \approx 43$ TeV. Returning to our toy theory, we summarize the parameters used in our example in Table 2. The interesting mass effects occur in the SCET2 theory and its matching to the low-scale theory. Thus, to simplify our example we take the SCET1 regime to run over a negligible range of scales. In effect, we match directly from the full theory to the SCET2 theory, since the Wilson coefficients we obtain are independent of M_t.[11] We do not need to specify a numerical value for M_X, but we take it to be just a small factor greater than the $M_{\rm Intermediate} = 1.5 \times 10^8$ TeV, determined by our Sudakov condition.

As discussed in Section 4.2, matching to the full theory at NLO gives $\mathcal{O}(\alpha_y)$ contributions to $C_{h,b}$ since it is possible to produce h, b particles from a hard three-body decay of the X. By construction, these are the portions of the X decay rate that do not follow from producing a t, which subsequently fragments into the other fields following the SCET Feynman rules. We give the z-space boundary conditions in Eq. 40, but in practice we run & match in Mellin space. Those boundary conditions are:

$$\begin{aligned}
C_t(s, \mu = M_X) &= 1 \\
C_b(s, \mu = M_X) &= \alpha_y\left(\frac{s-2}{8\pi s(s+1)}\right) \\
C_h(s, \mu = M_X) &= \alpha_y\left(-\frac{H(s-2)}{4\pi(s-2)} + \frac{1-5s}{8\pi\,(s-s^3)}\right).
\end{aligned} \tag{45}$$

Presenting results is more physically transparent back in z or plus momentum-fraction space. Thus, after numerically solving our RG equation to run and/or match our Wilson coefficients at the scale $\mu = M_t$, we perform an inverse Mellin transform. We set our contour as $s(t) = 1 + t \cdot \exp(i\,\phi)$, where $\phi = 3\pi/4$, and integrate numerically over t,

$$C_i(z) = \mathrm{Im}\left\{\frac{1}{\pi}\int_0^\infty e^{i\phi} z^{-s(t)} C_i\left(s(t)\right) dt\right\}, \tag{46}$$

[11] The boundary conditions we set in Eq. 45 are thus taken as the initial boundary conditions for the SCET2 running, *i.e.* $C_i(M_X) = C_i(M_{\rm Intermediate})$.

We tested robustness of the results in shifting the definition of $s(t) = a + t \cdot \exp(i\,\phi)$, varying a and ϕ by $\mathcal{O}(1)$ amounts. Solving the RG equation necessitates running the coupling, α_y, as well,

$$\beta(\alpha_y) = \frac{7}{4\pi}\alpha_y^2; \qquad \alpha_y(\mu) = \frac{4\pi}{16\pi^2 - 7\log\left(\frac{\mu}{M_t}\right)} \tag{47}$$

where we use $\alpha_y(M_t) = 1/4\pi$ as a boundary condition for $\alpha_y(\mu)$.

Energy Scale	EFT	Boundary Conditions
$\gtrsim$ Few $\times 10^8$ TeV	Full Theory	
Few-1.5 $\times 10^8$ TeV	SCET1	At $M_X =$ Few $\times 10^8$ TeV; $C_t = \delta(1-z)$; $C_{b,h}$ =*cf.* Eq. 40
1.5 $\times 10^8$ TeV-$M_t (= 173$ GeV)	SCET2	At $M_{\text{Intermediate}} = 1.5 \times 10^8$ TeV; Trivial from matching SCET1 onto SCET2
Below M_t	Low-Scale Theory (No running)	At M_t; Matching Conditions that remove t

Table 2: Numerical inputs & operational steps for computing with our toy theory.

5.1 Effect of Mass on Running

In Figure 2 we show the numerical significance of the finite M_t contributions by comparing three different versions of running the Wilson Coefficients. Firstly, we have the massless case where no mass contributions are considered throughout our running. That is to say that we include neither finite-width (Γ_t) nor ultracollinear contributions. This is just LL conventional running with the usual splitting functions as obtainable from [17]. Said differently, this is what the running would be if we had a massless SCET1 theory from $M_{\text{Intermediate}}$ to M_t. In the second case, we include the finite width effect, which enters the LL anomalous dimension (*cf.* Eq. 34 and Table 3). This is the case spelled out in Table 2 with a SCET2 theory involving a finite M_t running over our scales of interest. We see that the effect is numerically large, as we expect since the finite width shifts the $\mathcal{O}(\alpha_y)$ anomalous dimension at this same level. This does not involve however, any of the ultracollinear contributions. To see their effect we plot a third curve, which includes a subset of the $\mathcal{O}(\alpha_y^2)$ anomalous dimension corrections (*cf.* Tables 4-6). These give our "NLL-UC" running, which is not a complete accounting of NLL, as discussed above. Nonetheless, we see the ultracollinear terms on their own give a non-negligible shift to the C_i.

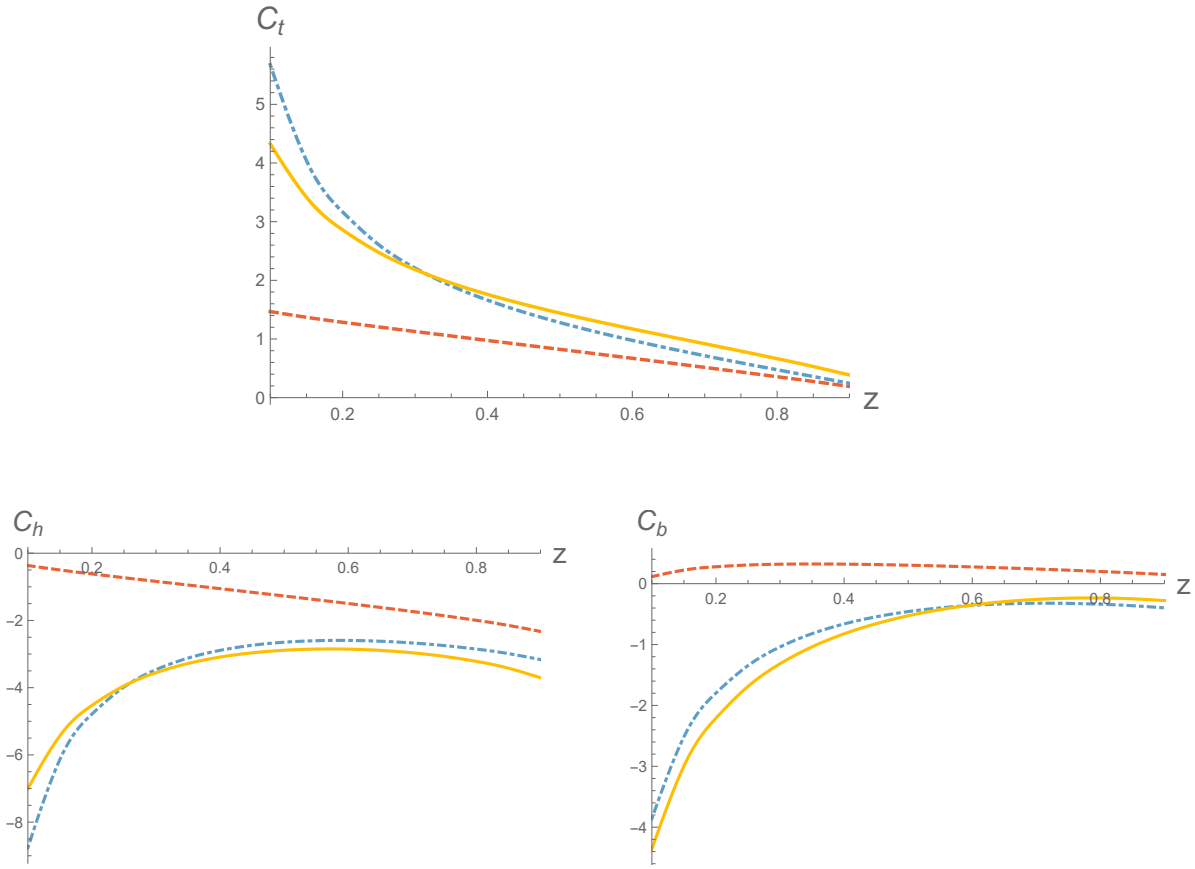


Figure 2: Running of $C_t(z), C_h(z), C_b(z)$ at massless LL *(dashed red)*, at LL including the finite width *(dotted dashed blue)*, and at NLL-UC *(solid yellow)*.

5.2 Effect of Mass on Matching

If we were generating particles in this t, b, h theory originating from some high-energy process (which we take to be the decay of some heavy X for concreteness), running down in scale would eventually drop below M_t. This is the final step of Table 2, matching to a theory without the t. In Fig. 3, we plot the result for the two remaining Wilson coefficients, C_b, C_h. Once again, we show three variations. In all cases, we assume that the UV theory in this matching is SCET2 with finite M_t. Thus, the LL running does account for finite Γ_t. By matching at LO, we again include the finite width, which is unavoidable even in a leading treatment with a massive particle. At NLO, we include the ultracollinear terms in the fragmentation functions (*cf.* Table 1). We see in Fig. 4, this just provides a percent-level shift. However, in both Figs. 3 and 4, we see that running at NLL-UC provides a qualitatively larger shift. It is an interesting question for future work if this survives in the complete NLL result or gets substantially cancelled by the other $\mathcal{O}(\alpha_y^2)$ contributions to γ_{ij}.

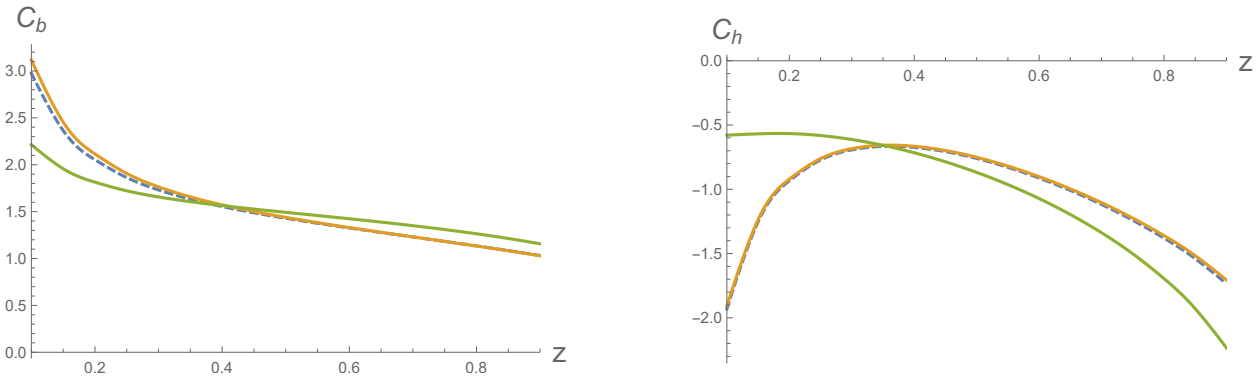


Figure 3: Results of $C_h(z), C_b(z)$ in Running + Matching at LL+LO *(dashed blue)*, LL+NLO *(solid orange)*, NLL-UC+NLO *(solid green)*.

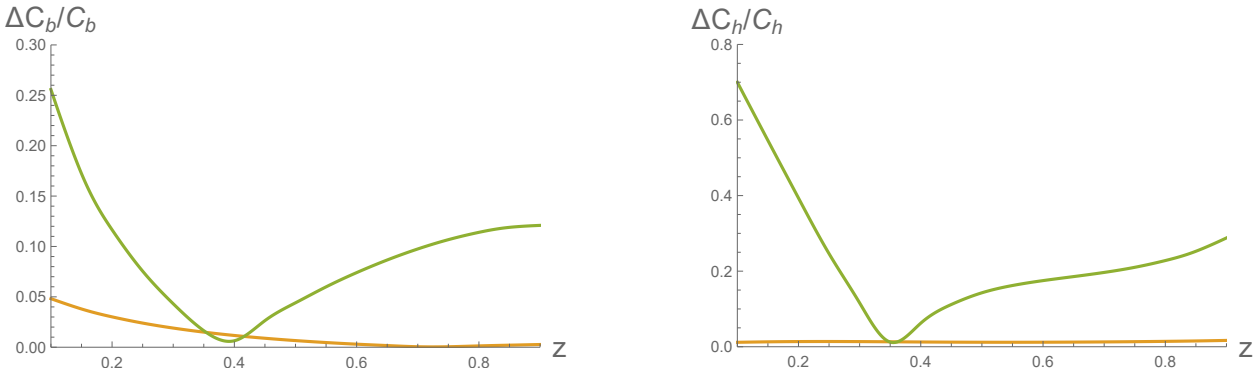


Figure 4: Ratio of $\frac{|C^{\rm LL}_{\rm NLO} - C^{\rm LL}_{\rm LO}|}{C^{\rm LL}_{\rm LO}}$ *(orange)* and $\frac{|C^{\rm NLL\text{-}UC}_{\rm NLO} - C^{\rm LL}_{\rm LO}|}{C^{\rm LL}_{\rm LO}}$ *(green)* for C_b, C_h.

5.3 Conclusions

From a rather ordinary mathematical problem arises a very interesting physical paradox that requires nontrivial conceptual labor to resolve. As Smilga noticed long ago [1], small fermion mass terms are not obeying commutation of the massless limit, and thus it is unsafe to simply toss them aside. The use of Effective Field Theory and its systematic formalism of operators and Wilson coefficients reveals when one must track a fermion mass. The answer boils down to when one expects to produce an on-shell particle, as quantified by the Sudakov factor. It is not hard to cook up an example, as we have done here, where finite mass effects endure long beyond the naive intution of $(m_{\rm light}/m_{\rm heavy})^2$ suppression.

We see furthermore that the numerical stakes of keeping finite particle masses (or not) can be significant. There are $\mathcal{O}$(few) shifts in the Wilson coefficients, C_i, depending on whether one tracks the finite width in running, with further possible $\mathcal{O}(10\text{s}\%)$ corrections from the ultracollinear contributions to the anomalous dimensions (*cf.* Fig. 2). Similar scale shifts from these terms persist in our final matching step, where we integrate out the massive fermion (*cf.* Figs. 3 & 4).

Now that we have studied and understood the subtleties of finite fermion mass effects in a controlled setting, it is a natural follow-up to study more realistic theories. Our quick exercise with QCD Sudakovs in Section 5 shows that even with double-log running, one may need to track b

and t quark masses about an order of magnitude higher than one might expect. For showers used for astroparticle physics, like HDMSpectra [9], the range of scales under consideration is enormous, stretching from weak to Planck scale. One could imagine all manner of BSM physics in this range. For each particle, we must answer the question of when to keep its mass. By extending this work, one can arrive at a systematic answer.

Acknowledgments: This work has benefitted from useful discussions with Benoit Assi, Sean Fleming, Stefan Hoeche, Aneesh Manohar, Thomas Mehen, Davison Soper, & Ubirajara van Kolck. We especially thank Varun Vaidya for clarifying subtle points in SCET. We were grateful to have Ozan Erdoğan's collaboration during the early stages of this project. MB and PC are supported by the DOE (HEP) Award DE-SC0019470.

A Fragmentation Formalism

We review here the fragmentation formalism and how our calculational quantities of interest arise. From SCET, we take the operator definitions in [24, 25], which follow from the classic treatment in perturbative QCD given in [26]. As originally formulated, the physical concern was observing hadrons in jets initiated by perturbative partons. Our situation here is qualitatively different, as the theory is fully perturbative. Nonetheless, we can still work with fragmentation functions, as in [9], which provide a means to compute the rate to observe a particle j that radiated from a possibly different particle i.

It is natural to conceive of the fragmentation function as a kind of probability distribution. In its original QCD framing, this would give the probability to observe a hadron with a particular energy fraction of its mother parton. In our perturbative theory, this becomes the probability of observing a "parton" with a particular plus-momentum fraction of its mother parton. The probabilistic interpretation breaks down beyond the Leading-Log level. One can get explicitly negative results, in fact. While inconvenient, this not a fundamental pathology. Only observable amplitudes need to respect the strict unitarity of quantum mechanics. Nonetheless, having even an approximately probabilistic quantity is useful for guiding its construction. Also, the need to conserve probability at LL provides a useful sum rule with which to check the calculation, as detailed below.

Following Collins, as reviewed in [27], constructing a function which gives the number density of particles with a particular fraction of their mother's E or p^+ starts us in a frame where the mother has zero transverse momentum, ($k_\perp = 0$), and we integrate over the daughter's $\vec{p}_\perp$. We also have that $p^+ = z\,k^+$ and $p^- = (\vec{p}_\perp^{\,2} + m_d^2)/p^+$, where in our case the daughter mass $m_d = M_t$ if it is a top and zero otherwise. In this frame, if the initiating particle is a fermion, the fragmentation function to observe particle j emitted from it is

$$D_{ij}(z) = \frac{1}{z}\int d^2p_\perp \int \frac{dx^- d^2x_\perp}{2(2\pi)^3} e^{ik^+x^-/2} \frac{1}{2N_{c,i}} \mathrm{Tr} \sum_X \langle 0|\frac{\bar{n}\!\!\!/}{2}\chi_i(0,x^-,x_\perp)|j(p)\,X\rangle\langle j(p)\,X|\chi_i^\dagger(0)|0\rangle \Big|_{p_\perp^{X_j}=0}. \tag{48}$$

We average over spin and color of the initiating particle (in our case $N_c = 2$ for top and higgs, $N_c = 1$ for the bottom). Naively, $D_{ij}(z)$ depends on the plus-momentum of both mother (k^+) and daughter (p^+). However, we have boost invariance in directions orthogonal to the $\perp$ direction. Thus, it can only depend on $z(\equiv p^+/k^+)$.

The difficulty with the form of Eq. 48 is that it fixes an unobservable quantity, the mother's perp momentum, $k_\perp = 0$. A better operational definition fixes the axis of the observed daughter's momentum. Thus, we rotate and boost to a $p_\perp = 0$ frame with p^+ unchanged. The mother perp

momentum becomes $\vec{k}_\perp = -\vec{p}_\perp/z$. Doing this gives

$$D_{ij}(z) = z\int \frac{dx^-}{4\pi} e^{ik^+x^-/2} \frac{1}{2N_{c,i}} \,\mathrm{Tr} \sum_X \langle 0|\frac{\bar{\not{n}}}{2}\chi_i(0,x^-,0_\perp)|j(p)\,X\rangle\langle j(p)\,X|\chi_i^\dagger(0)|0\rangle \Big|_{p_\perp^j=0}. \quad (49)$$

In practice, we compute purely in momentum space and thus calculate the following quantities,

$$D_{ij}^{\text{fermion}}(z) = \frac{z}{Q}\frac{1}{2N_{c,i}} \,\mathrm{Tr}\int \Pi_X\,\delta\left(1-\frac{p_X^+ + p^+}{Q}\right)\langle 0|\frac{\bar{\not{n}}}{2}\chi_i(k)|j(p)\,X\rangle\langle j(p)\,X|\chi_i^\dagger(k)|0\rangle\Big|_{p_\perp^j=0}$$
$$D_{ij}^{\text{scalar}}(z) = z\frac{1}{N_{c,i}} \,\mathrm{Tr}\int \Pi_X\,\delta\left(1-\frac{p_X^+ + p^+}{Q}\right)\langle 0|\phi_i(k)|j(p)\,X\rangle\langle j(p)\,X|\phi_i^\dagger(k)|0\rangle\Big|_{p_\perp^j=0}. \quad (50)$$

We get a very useful set of constraints from the set of sum rules that massless fragmentation functions satisfy up to order $\alpha_y \log(\mu^2/m_t^2)$, which follow from the conservations of probability and momentum:

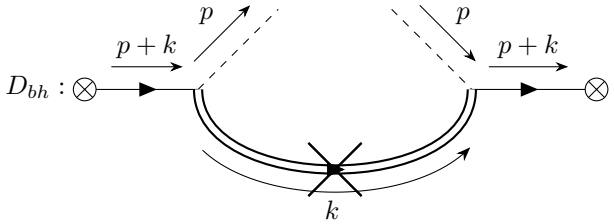
$$\frac{1}{2}\sum_j \int dz\, D_{ij}^R(z) + \int dz\, D_{ii}^V(z) = 1$$
$$\sum_j \int dz\, z\, D_{ij}^R(z) + \int dz\, z\, D_{ii}^V(z) = 1 \quad (51)$$

where R stands for fragmentation with real-emission while V indicates self-fragmentation. Including the effects of finite M_t, the probabilistic interpretation of D_{ij} breaks down. Contributions due to the finite width Γ_t give positive contributions to D_{tb}, D_{th}. However, in our $\overline{\text{MS}}$ scheme, there is no compensating negative term in D_{tt}. Furthermore, the $\mathcal{O}(\alpha_y)$ corrections (without log enhancement) arising from the ultracollinear contributions contain strictly negative terms (*cf.* Table 1) allowing D_{ij} to take negative values for some range of μ and z.

B Soft & Zero-Bin Contribution

Throughout the following calculations of this Appendix we note the different notation used to label the daughter and mother momenta with respect to the one used in the definition of the fragmentation functions in A.

The calculation of the soft contribution to a graph can come from the Full Theory calculation, followed by taking the appropriate soft and collinear limits for the graph's momenta. We recall that it is only in the SCET1 theory that a vertex with a soft field is leading-power. Thus, we can work in the simpler $M_t = 0$ case. As an example we look at the bare $b \to h$ fragmentation function,



(52)

$$D_{0,bh} = \frac{z}{2}y^2\int \frac{d^4k}{(2\pi)^4}\,\mathrm{Tr}\left\{\frac{\bar{\not{n}}}{2}\not{p}\not{k}\not{p}\right\}2\pi\delta\left(k^2\right)\frac{\delta\left(Q-p^+-k^+\right)}{\left((p+k)^2\right)^2}, \quad (53)$$

with $p^- = 0$, $p_\perp = 0$ and computing the trace:

$$D_{0,bh} = \frac{z}{2}y^2\int \frac{d^4k}{(2\pi)^4}(p^+k^-p^+)2\pi\delta\left(k^2\right)\frac{\delta\left(Q-p^+-k^+\right)}{\left(p^+k^- - k_\perp^2\right)^2}. \quad (54)$$

We now take k to have the scaling of a soft momentum k, where $(k^+, k^-, k_\perp) \sim (\lambda^2, \lambda^2, \lambda^2)$, compared to $(p^+, p^-, p_\perp) \sim (1, \lambda^2, \lambda)$, where λ is the SCET bookkeeping power counting parameter:

$$D_{0,bh}^{\text{soft}} = \frac{z}{2} y^2 \int \frac{d^4k}{(2\pi)^4} (p^+ k^- p^+) 2\pi\delta\left(k^2\right) \frac{\delta\left(Q - p^+\right)}{\left(p^+ k^-\right)^2}. \tag{55}$$

Using the on-shell condition of k, integrating over k^-:

$$D_{0,bh}^{\text{soft}} = \frac{z}{2} y^2 \int \frac{d^2k_\perp dk^+}{2(2\pi)^3} \frac{(p^+)^2}{(k^+)^2} k_\perp^2 \frac{(k^+)^2}{(p^+)^2} \frac{\delta\left(Q - p^+\right)}{\left(k_\perp^2\right)^2}. \tag{56}$$

Thus, our soft contribution will be:

$$D_{0,bh}^{\text{soft}} = \frac{z}{2Q} y^2 \int \frac{d^2k_\perp dk^+}{2(2\pi)^3} k_\perp^2 \frac{\delta\left(1 - z\right)}{\left(k_\perp^2\right)^2}. \tag{57}$$

For the zero-bin contribution of the same graph, we compute it in the collinear sector of SCET1,

$$D_{0,bh}^{\text{zero}} = \frac{z}{2} y^2 \int \frac{d^4k}{(2\pi)^4} \text{Tr} \left\{ \frac{\bar{\not{n}}}{2} \left(p^+ + k^+\right) \frac{\not{n}}{2} \left(\frac{\not{k}_\perp}{k^+} - \frac{\not{p}_\perp + \not{k}_\perp}{p^+ + k^+} \right) \frac{\bar{\not{n}}}{2} \frac{\not{n}}{2} k^+ \right.$$
$$\left. \left(\frac{\not{k}_\perp}{k^+} - \frac{\not{p}_\perp + \not{k}_\perp}{p^+ + k^+} \right) \frac{\bar{\not{n}}}{2} \left(p^+ + k^+\right) \frac{\not{n}}{2} \right\} 2\pi\delta\left(k^2\right) \frac{\delta\left(Q - p^+ - k^+\right)}{\left((p+k)^2\right)^2}, \tag{58}$$

and then take the soft limit k. We also impose on-shellness for p with $p_\perp = 0$, $p^- = 0$,

$$D_{0,bh}^{\text{zero}} = \frac{z}{2} y^2 \int \frac{d^4k}{(2\pi)^4} \text{Tr} \left\{ \frac{\bar{\not{n}}}{2} \left(p^+\right) \frac{\not{n}}{2} \left(\frac{\not{k}_\perp}{k^+} \right) \frac{\bar{\not{n}}}{2} \frac{\not{n}}{2} k^+ \left(\frac{\not{k}_\perp}{k^+} \right) \frac{\bar{\not{n}}}{2} \left(p^+\right) \frac{\not{n}}{2} \right\} 2\pi\delta\left(k^2\right) \frac{\delta\left(Q - p^+\right)}{\left(p^+ k^-\right)^2}. \tag{59}$$

Computing the trace:

$$D_{0,bh}^{\text{zero}} = \frac{z}{2} y^2 \int \frac{d^4k}{(2\pi)^4} \frac{(p^+)^2}{k^+} k_\perp^2 2\pi\delta\left(k^2\right) \frac{\delta\left(Q - p^+\right)}{\left(p^+ k^-\right)^2}. \tag{60}$$

Integrating over k^-:

$$D_{0,bh}^{\text{zero}} = \frac{z}{2} y^2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{\delta\left(Q - p^+\right)}{k_\perp^2}, \tag{61}$$

and defining $z = p^+/Q$:

$$D_{0,bh}^{\text{zero}} = \frac{z}{2Q} y^2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{\delta\left(1 - z\right)}{k_\perp^2}. \tag{62}$$

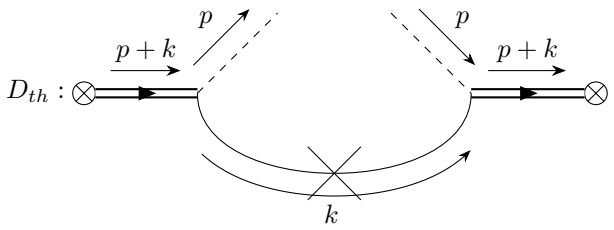
We thus see from equations 57 and 62, that $\boxed{D_{0,bh}^{\text{zero}} \equiv D_{0,bh}^{\text{soft}}}$. Thus the soft modes do not affect our results as they are cancelled by the zero-bin modes. The same result holds for the other anomalous dimensions, as well.

C Calculational Details

Throughout the following calculations of this Appendix we note the different notation used to label the daughter and mother momenta with respect to the one used in the definition of the fragmentation functions in A.

C.1 Calculation of D_{th} in Intermediate Theory (SCET2 with $M_t \neq 0$)

The diagram for D_{th} calculation is the following:



(63)

Using the operator definition of the fragmentation function as pointed out in A, along with the SCET2 Feynman rules, as presented in Section 2, in the complex mass scheme, we get:

$$D_{0,th} = \frac{z}{2}y^2 \int \frac{d^4k}{(2\pi)^4} \mathrm{Tr}\left\{ \frac{\bar{n}\!\!\!/}{2}\left(p^+ + k^+\right)\frac{n\!\!\!/}{2}\left(\frac{k\!\!\!/_\perp}{k^+} - \frac{p\!\!\!/_\perp + k\!\!\!/_\perp}{p^+ + k^+} - \frac{\sqrt{M_t^2 - iM_t\Gamma_t}}{p^+ + k^+}\right)\frac{\bar{n}\!\!\!/}{2}\frac{n\!\!\!/}{2}k^+ \right.$$
$$\left. \left(\frac{k\!\!\!/_\perp}{k^+} - \frac{p\!\!\!/_\perp + k\!\!\!/_\perp}{p^+ + k^+} + \frac{\sqrt{M_t^2 + iM_t\Gamma_t}}{p^+ + k^+}\right)\frac{\bar{n}\!\!\!/}{2}\left(p^+ + k^+\right)\frac{n\!\!\!/}{2}\right\} 2\pi\delta\left(k^2\right)\frac{\delta\left(Q - p^+ - k^+\right)}{\left((p+k)^2 - M_t^2\right)^2 + M_t^2\Gamma^2}, \tag{64}$$

where the "0" subscript denotes that this is the bare fragmentation function. We follow the steps outlined in 3.1, performing the calculation in the reference frame, discussed in A where daughter's perp momentum is zero ($p_\perp = 0$):

$$D_{0,th} = \frac{z}{2}y^2 \int \frac{d^4k}{(2\pi)^4}$$
$$\mathrm{Tr}\left\{ \frac{\bar{n}\!\!\!/}{2}\left(p^+ + k^+\right)\frac{n\!\!\!/}{2}\left(\frac{p^+ k\!\!\!/_\perp}{(p^+ + k^+)k^+} - \frac{\sqrt{M_t^2 - iM_t\Gamma_t}}{p^+ + k^+}\right)\frac{\bar{n}\!\!\!/}{2}\frac{n\!\!\!/}{2}k^+\left(\frac{p^+ k\!\!\!/_\perp}{(p^+ + k^+)k^+} + \frac{\sqrt{M_t^2 + iM_t\Gamma_t}}{p^+ + k^+}\right)\right.$$
$$\left.\frac{\bar{n}\!\!\!/}{2}\left(p^+ + k^+\right)\frac{n\!\!\!/}{2}\right\} 2\pi\delta\left(k^2\right)\frac{\delta\left(Q - p^+ - k^+\right)}{\left((p+k)^+(p+k)^- - k_\perp^2 - M_t^2\right)^2 + M_t^2\Gamma^2}. \tag{65}$$

The next step is to calculate the trace and simplify the Dirac structure:

$$D_{0,th} = \frac{z}{2}y^2 \int \frac{d^4k}{(2\pi)^3}\left(p^+ + k^+\right)^2 k^+ \left(\frac{k_\perp^2 p_+^2 + M_t^2 k_+^2 - M_t\Gamma k_+^2}{k_+^2(p_+ + k_+)^2}\right)\frac{\delta\left(k^2\right)\delta\left(Q - p^+ - k^+\right)}{\left((p+k)^+(p+k)^- - k_\perp^2 - M_t^2\right)^2 + M_t^2\Gamma^2}. \tag{66}$$

We integrating k^- with k on-shell and also $p^- = 0$ due to p on-shellness since we are in the daughter's reference frame:

$$D_{0,th} = \frac{z}{2}\frac{y^2}{2}\int \frac{d^2k_\perp dk^+}{(2\pi)^3}\left(k_\perp^2\frac{p_+^2}{k_+^2} + M_t^2\right)\frac{\delta\left(Q - p^+ - k^+\right)}{\left((p+k)^+\frac{k_\perp^2}{k^+} - k_\perp^2 - M_t^2\right)^2 + M_t^2\Gamma^2} + \mathcal{O}\left(\alpha_Y^2\right). \tag{67}$$

Here in this step we focused only on the terms that are leading order in α_y, any terms of $\mathcal{O}\left(\alpha_y^2\right)$ are subsequently dropped. Defining $z \equiv \frac{p^+}{Q}$ and integrating over k^+ with the remaining delta function,

$\delta\left(Q-k^{+}-p^{+}\right)$:

$$D_{0,th}=\frac{z}{2}\frac{y^2}{2}\int\frac{d^2k_\perp}{(2\pi)^3}\left(k_\perp^2\frac{z^2}{(1-z)^2}+M_t^2\right)\frac{\frac{(1-z)^2}{z^2}}{\left(k_\perp^2-M_t^2\frac{1-z}{z}\right)^2+M_t^2\Gamma^2\frac{(1-z)^2}{z^2}}. \tag{68}$$

Regularizing the UV divergence through Dimensional Regularization in $d=2-2\varepsilon$:

$$D_{0,th}=\frac{z}{2}\frac{y^2}{2}\mu^{2\varepsilon}\int\frac{d^dk_\perp}{(2\pi)^3}\frac{\left(k_\perp^2+M_t^2\frac{(1-z)^2}{z^2}\right)}{\left(k_\perp^2-M_t^2\frac{1-z}{z}\right)^2+M_t^2\Gamma^2\frac{(1-z)^2}{z^2}}, \tag{69}$$

Integrating $k_\perp$ from $[0,\infty)$ and expanding at $\varepsilon=0$:

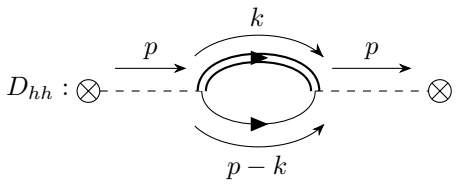
$$D_{0,th}=\frac{\alpha_y}{4\pi}\left(z\frac{1}{\varepsilon}+z\log\left(\frac{\mu^2}{M_t^2\frac{1-z}{z}\sqrt{1+\frac{\Gamma^2}{M_t^2}}}\right)-\frac{M_t}{\Gamma_t}\arctan\left(\frac{\Gamma_t}{M_t}\right)+\pi\frac{M_t}{\Gamma_t}\right). \tag{70}$$

We note that the M_t^2 term in the numerator of Eq. 69 contributes the $z\log[z/(1-z)]$ in $D_{0,th}$ that one can spot in Eq. 70. This is precisely the ultracollinear contribution. It relies on keeping finite M_t in the calculation, but it persists after the fact even if we take $M_t\to 0$. The final form of the fragmentation function is given after keeping the LO terms of $\dfrac{\Gamma_t}{M_t}\to 0$ (to $\mathcal{O}(\alpha_y)$, $\Gamma_t=\alpha_y M_t/4$) and subtracting the $\dfrac{1}{\varepsilon}$ term. Therefore, we have the renormalized fragmentation function, D_{th}:

$$\boxed{D_{th}=\frac{\alpha_y}{4\pi}\left(z\log\left(\frac{\mu^2}{M_t^2}\right)-z\log\left(\frac{1-z}{z}\right)-1+\pi\frac{M_t}{\Gamma_t}\right)} \tag{71}$$

C.2 Calculation of D_{hh} in Intermediate Theory with $M_t\neq 0$

The diagram for D_{hh} calculation is the following:



(72)

We follow the steps outlined in 3.1 and we calculate the residue of the expression represented by this graph:

$$\text{Res}\left[\Sigma\left(p\right)\right]=\lim_{p^2\to 0}\frac{d}{dp^2}\left\{\left(p^2\right)^2\cdot\frac{1}{p^2}\Sigma\left(p\right)\cdot\frac{1}{p^2}\right\}=\lim_{p^-\to 0}\frac{dp^-}{dp^2}\frac{d}{dp^-}\Sigma\left(p\right)=\lim_{p^-\to 0}\frac{1}{p^+}\frac{d}{dp^-}\Sigma\left(p\right), \tag{73}$$

where $\Sigma(p)$ is given by:

$$\Sigma(p) = (-1)y^2 \int \frac{d^4k}{(2\pi)^4} \mathrm{Tr}\left\{ \left(\frac{-\not{k}_\perp}{p^+ - k^+} - \frac{\not{k}_\perp}{k^+} - \frac{M_t}{k^+} \right) \frac{\not{n}}{2} \frac{\not{n}}{2} \left(\frac{-\not{k}_\perp}{p^+ - k^+} - \frac{\not{k}_\perp}{k^+} + \frac{M_t}{k^+} \right) \frac{\not{n}}{2} \frac{\not{n}}{2} \right\} \frac{k^+}{k^+k^- - k_\perp^2 - M_t^2} \left(\frac{(p-k)^+}{(p-k)^+(p-k)^- - k_\perp^2} \right), \tag{74}$$

and the (-1) comes from the fermion loop factor. The derivative gives:

$$\begin{aligned} \mathrm{Res}\left[\Sigma(p)\right] = \lim_{p^-\to 0} \frac{1}{p^+}\frac{d}{dp^-}\Sigma(p) &= -y^2 2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{p^+} \left(\frac{p_+^2 k_\perp^2}{k^+(p^+ - k^+)} - \frac{M_t^2(p^+ - k^+)}{k^+} \right) \\ &\quad \frac{(k-p)^+}{k^+k^- - k_\perp^2 - M_t^2} \left(\frac{1}{(p-k)^+(-k^-) - k_\perp^2} \right)^2 \\ &= -y^2 2 \int \frac{d^4k}{(2\pi)^4} \frac{1}{p^+} \left(\frac{p_+^2 k_\perp^2}{k^+(p^+ - k^+)} - \frac{M_t^2(p^+ - k^+)}{k^+} \right) \\ &\quad \frac{1}{k^+k^- - k_\perp^2 - M_t^2} \frac{1}{(k^+ - p^+)} \left(\frac{1}{k^- - \frac{k_\perp^2}{(k^+ - p^+)}} \right)^2 . \end{aligned} \tag{75}$$

Above, the trace is $\mathrm{Tr}\left\{ \frac{\not{n}}{2}\frac{\not{n}}{2}\frac{\not{n}}{2}\frac{\not{n}}{2} \right\} = 2$. Using the residue theorem, the integral over k^- can be calculated by expanding the integrand around the pole, $k^- = (k_\perp^2 + M_t^2)/k^+$. The integral is nonzero if $p^+ > k^+ > 0$. Thus,

$$\begin{aligned} \mathrm{Res}\left[\Sigma(p)\right] &= -\frac{y^2}{2} 2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{1}{p^+} \left(\frac{p_+^2 k_\perp^2}{k^+(p^+ - k^+)} - \frac{M_t^2(p^+ - k^+)}{k^+} \right) \frac{1}{(k^+ - p^+)} \frac{1}{k^+} \left(\frac{1}{\frac{k_\perp^2}{k^+} + \frac{M_t^2}{k^+} - \frac{k_\perp^2}{(k^+ - p^+)}} \right)^2 \\ &= -\frac{y^2}{2} 2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{1}{p^+} \left(\frac{p_+^2 k_\perp^2}{k^+(p^+ - k^+)} - \frac{M_t^2(p^+ - k^+)}{k^+} \right) \frac{1}{(k^+ - p^+)k^+} \left(\frac{1}{\frac{M_t^2}{k^+} - \frac{k_\perp^2 p^+}{k^+(k^+ - p^+)}} \right)^2 \\ &= -\frac{y^2}{2} 2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{1}{p^+} \left(-k_\perp^2 + M_t^2 \frac{(k^+ - p^+)^2}{(p^+)^2} \right) \left(\frac{1}{k_\perp^2 - \frac{M_t^2(k^+ - p^+)}{p^+}} \right)^2 . \end{aligned} \tag{76}$$

The $k_\perp$ in the numerator of expression above is not Euclidean. Thus we need to rewrite the expression as follows:

$$\mathrm{Res}\left[\Sigma(p)\right] = -\frac{y^2}{2} 2 \int \frac{d^2k_\perp dk^+}{(2\pi)^3} \frac{1}{p^+} \left(k_\perp^2 + M_t^2 \frac{(k^+ - p^+)^2}{(p^+)^2} \right) \left(\frac{1}{k_\perp^2 - \frac{M_t^2(k^+ - p^+)}{p^+}} \right)^2 . \tag{77}$$

Integrating first over $k_\perp$ on $[0, \infty)$ using Dimensional Regularization:

$$\mathrm{Res}\left[\Sigma(p)\right] = -\frac{y^2}{2} 2 \int_0^{p^+} \frac{dk^+}{8(\pi)^2} \frac{\pi(k^+ - p^+)\mu^{2\epsilon} \csc(\pi\epsilon)(k^+\epsilon - 2p^+) \left(\frac{p^+}{M_t^2(p^+ - k^+)} \right)^\epsilon}{(p^+)^2(\epsilon - 2)}. \tag{78}$$

Integrating over k^+ and expanding to $\mathcal{O}(\epsilon^0)$, we get the 1-loop contribution as:

$$\text{Res}\left[\Sigma\left(p\right)\right] = \frac{\alpha_y}{2\pi}\left[-\log\left(\frac{\mu^2}{M_t^2}\right) - \frac{1}{\epsilon} - \frac{1}{2}\right]. \tag{79}$$

Working toward the fragmentation function, we restore the z-dependence, through a $\delta(1-z)$ as we have in the self-fragmentation:

$$D_{0,hh} = \frac{\alpha_y}{2\pi}\left[-\log\left(\frac{\mu^2}{M_t^2}\right) - \frac{1}{\epsilon} - \frac{1}{2}\right]\delta(1-z). \tag{80}$$

The final form of the renormalized fragmentation function, after subtracting the $\frac{1}{\varepsilon}$, is

$$\boxed{D_{hh}(z) = \frac{\alpha_y}{2\pi}\left[-\log\left(\frac{\mu^2}{M_t^2}\right) - \frac{1}{2}\right]\delta\left(1-z\right).} \tag{81}$$

The -1/2 term is the ultracollinear contribution in this case.

D Mellin Space Anomalous Dimensions

As we saw in 4.2, performing the running requires us to calculate the anomalous dimension matrix in Mellin space. This can be done at LL and NLL-UC, as dictated in equations 34, 35, using the matrix elements for fragmentation functions in Mellin space from 17. We list the expressions for anomalous dimension matrix elements at LL and NLL-UC in the tables below:

	γ_{bi}^{LL} for $i \in \{h, b, t\}$
γ_{bh}^{LL}	$\frac{\alpha_y}{\pi}\frac{(s-1)}{s^2}$
γ_{bb}^{LL}	$-\frac{\alpha_y}{\pi}\frac{\left(s^3+s^2+2\right)}{2s^2(s+1)}$
γ_{bt}^{LL}	$\frac{\alpha_y}{\pi}\frac{1}{(s^2+s)}$

	γ_{hi}^{LL} for $i \in \{h, b, t\}$
γ_{hh}^{LL}	$-\frac{\alpha_y}{\pi}\frac{\left(s^2+1\right)}{s^2}$
γ_{hb}^{LL}	$\frac{\alpha_y}{\pi}\frac{(s-1)}{s^2}$
γ_{ht}^{LL}	$\frac{\alpha_y}{\pi}\frac{1}{s}$

	γ_{ti}^{LL} for $i \in \{h, b, t\}$
γ_{th}^{LL}	$\frac{\alpha_y}{\pi}\frac{1+3s}{4(s^2+s)}$
γ_{tb}^{LL}	$\frac{\alpha_y}{\pi}\frac{3+s}{4(s^2+s)}$
γ_{tt}^{LL}	$-\frac{\alpha_y}{\pi}\frac{1}{4}$

Table 3: Mellin Transforms of the Anomalous Dimension at LL.

	$\gamma_{bi}^{\text{NLL-UC}}$ for $i \in \{h, b, t\}$
$\gamma_{bh}^{\text{NLL-UC}}$	$\alpha_y^2\frac{\left(-2(s+1)s^4H(s+1)-\left(5s^4+10s^3+9s^2+12s+8\right)sH(s)+2s^7-5s^6-2s^5+35s^4+38s^3+26s^2+32s+16\right)}{8\pi^2(s+1)^3s^5}$
$\gamma_{bb}^{\text{NLL-UC}}$	$-\alpha_y^2\frac{\left(4\left(5s^4+12s^3+11s^2+12s+8\right)sH(s)+27s^8+81s^7+117s^6+51s^5-144s^4-148s^3-48s^2-112s-64\right)}{32\pi^2s^5(s+1)^3}$
$\gamma_{bt}^{\text{NLL-UC}}$	$\alpha_y^2\frac{\left(\left(5s^4+10s^3+9s^2+12s+8\right)sH(s)+\left(9s^6+13s^5-24s^4-38s^3-22s^2-32s-16\right)\right)}{8\pi^2s^4(s+1)^3}$

Table 4: Mellin Transforms of the Anomalous Dimension for $b \to \{h, b, t\}$ at NLL-UC.

	$\gamma_{hi}^{\text{NLL-UC}}$ for $i \in \{h, b, t\}$
$\gamma_{hh}^{\text{NLL-UC}}$	$-\alpha_y^2 \frac{\left(4(s+1)s^4 H(s+1) + 2\left(11s^4 + 22s^3 + 11s^2 + 8s + 8\right) sH(s) + 11s^7 + 22s^6 - 11s^5 - 78s^4 - 80s^3 - 44s^2 - 48s - 32\right)}{16\pi^2 s^5 (s+1)^2}$
$\gamma_{hb}^{\text{NLL-UC}}$	$\alpha_y^2 \frac{\left(-\left(11s^4 + 24s^3 + 13s^2 + 8s + 8\right) sH(s) + \left(3s^6 - 6s^5 + 7s^4 + 28s^3 + 8s^2 + 20s + 16\right)\right)}{8\pi^2 s^5 (s+1)^2}$
$\gamma_{ht}^{\text{NLL-UC}}$	$\alpha_y^2 \frac{\left(\left(11s^4 + 22s^3 + 11s^2 + 8s + 8\right) sH(s) - \left(11s^5 + 37s^4 + 44s^3 + 18s^2 + 24s + 16\right)\right)}{8\pi^2 s^4 (s+1)^2}$

Table 5: Mellin Transforms of the Anomalous Dimension for $h \to \{h, b, t\}$ at NLL-UC.

	$\gamma_{ti}^{\text{NLL-UC}}$ for $i \in \{h, b, t\}$
$\gamma_{th}^{\text{NLL-UC}}$	$\alpha_y^2 \frac{\left(2s\left(3s^2 + 5s + 4\right) H(s) + (s+1)\left(8s^4 H(s+1) - \left(5s^4 + 15s^3 - 4s^2 + 16s + 16\right)\right)\right)}{16\pi^2 s^4 (s+1)^2}$
$\gamma_{tb}^{\text{NLL-UC}}$	$\alpha_y^2 \frac{\left(4\left(4s^4 + 11s^3 + 12s^2 + 9s + 4\right) sH(s) - \left(13s^6 + 49s^5 + 59s^4 + 39s^3 + 68s^2 + 92s + 32\right)\right)}{32\pi^2 s^4 (s+1)^3}$
$\gamma_{tt}^{\text{NLL-UC}}$	$-\alpha_y^2 \frac{\left(s\left(3s^3 + 8s^2 + 9s + 4\right) H(s) - \left(2s^4 + 7s^3 + 19s^2 + 24s + 8\right)\right)}{8\pi^2 s^3 (s+1)^3}$

Table 6: Mellin Transforms of the Anomalous Dimension for $t \to \{h, b, t\}$ at NLL-UC.

References

[1] A. V. Smilga, *Quasiparadoxes of massless QED*, *Comments Nucl. Part. Phys.* **20** (1991), no. 1+2 69–84.

[2] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8, [arXiv:2203.11601].

[3] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[4] C. W. Bauer, S. Fleming, and M. E. Luke, *Summing Sudakov logarithms in $B \to X_s\gamma$ in effective field theory.*, *Phys. Rev. D* **63** (2000) 014006, [hep-ph/0005275].

[5] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, *Phys. Rev. D* **63** (2001) 114020, [hep-ph/0011336].

[6] C. W. Bauer and I. W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett. B* **516** (2001) 134–142, [hep-ph/0107001].

[7] C. W. Bauer, D. Pirjol, and I. W. Stewart, *Power counting in the soft collinear effective theory*, *Phys. Rev. D* **66** (2002) 054005, [hep-ph/0205289].

[8] J. Chen, T. Han, and B. Tweedie, *Electroweak Splitting Functions and High Energy Showering*, *JHEP* **11** (2017) 093, [arXiv:1611.00788].

[9] C. W. Bauer, N. L. Rodd, and B. R. Webber, *Dark matter spectra from the electroweak to the Planck scale*, *JHEP* **06** (2021) 121, [arXiv:2007.15001].

[10] T. Han, Y. Ma, and K. Xie, *Electroweak fragmentation at high energies: A Snowmass White Paper*, in *Snowmass 2021*, 3, 2022. arXiv:2203.11129.

[11] D. Pagani, T. Vitos, and M. Zaro, *Improving NLO QCD event generators with high-energy EW corrections*, *arXiv e-prints* (Sept., 2023) arXiv:2309.00452, [arXiv:2309.00452].

[12] T. Ježo, J. M. Lindert, and S. Pozzorini, *Resonance-aware NLOPS matching for off-shell $t\bar{t}$ + tW production with semileptonic decays*, *JHEP* **10** (2023) 008, [arXiv:2307.15653].

[13] B. Assi and S. Höche, *A new approach to QCD evolution in processes with massive partons*, arXiv:2307.00728.

[14] A. Gavardi, M. A. Lim, S. Alioli, and F. Tackmann, *NNLO+PS W^+W^- production using jet veto resummation at NNLL′*, arXiv:2308.11577.

[15] S. Ferrario Ravasio, K. Hamilton, A. Karlberg, G. P. Salam, L. Scyboz, and G. Soyez, *Parton Showering with Higher Logarithmic Accuracy for Soft Emissions*, *Phys. Rev. Lett.* **131** (2023), no. 16 161906, [arXiv:2307.11142].

[16] M. van Beekveld and S. Ferrario Ravasio, *Next-to-leading-logarithmic PanScales showers for Deep Inelastic Scattering and Vector Boson Fusion*, arXiv:2305.08645.

[17] P. Ciafaloni, D. Comelli, A. Riotto, F. Sala, A. Strumia, and A. Urbano, *Weak Corrections are Relevant for Dark Matter Indirect Detection*, *JCAP* **03** (2011) 019, [arXiv:1009.0224].

[18] M. Beneke, A. P. Chapovsky, A. Signer, and G. Zanderighi, *Effective theory approach to unstable particle production*, *Phys. Rev. Lett.* **93** (2004) 011602, [hep-ph/0312331].

[19] M. Beneke, A. P. Chapovsky, A. Signer, and G. Zanderighi, *Effective theory calculation of resonant high-energy scattering*, *Nucl. Phys. B* **686** (2004) 205–247, [hep-ph/0401002].

[20] M. Beneke, *Unstable-particle effective field theory*, *Nucl. Part. Phys. Proc.* **261-262** (2015) 218–231, [arXiv:1501.07370].

[21] A. Manohar, B. Shotwell, C. Bauer, and S. Turczyk, *Non-cancellation of electroweak logarithms in high-energy scattering*, *Phys. Lett. B* **740** (2015) 179–187, [arXiv:1409.1918].

[22] S. Davidson, P. Gambino, M. Laine, M. Neubert, and C. Salomon, eds., *Proceedings, Les Houches summer school: EFT in Particle Physics and Cosmology: Les Houches (Chamonix Valley), France, July 3-28, 2017*, vol. 108 of *Lecture Notes of the Les Houches Summer School*. Oxford University Press, Oxford, 4, 2020.

[23] A. V. Manohar and I. W. Stewart, *The Zero-Bin and Mode Factorization in Quantum Field Theory*, *Phys. Rev. D* **76** (2007) 074002, [hep-ph/0605001].

[24] M. Procura and I. W. Stewart, *Quark Fragmentation within an Identified Jet*, *Phys. Rev. D* **81** (2010) 074009, [arXiv:0911.4980]. [Erratum: Phys.Rev.D 83, 039902 (2011)].

[25] A. Jain, M. Procura, and W. J. Waalewijn, *Parton Fragmentation within an Identified Jet at NNLL*, *JHEP* **05** (2011) 035, [arXiv:1101.4953].

[26] J. C. Collins and D. E. Soper, *Parton Distribution and Decay Functions*, *Nucl. Phys. B* **194** (1982) 445–492.

[27] J. Collins, *Foundations of Perturbative QCD*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2011.

arXiv:2402.16958v1 [hep-ph] 26 Feb 2024

Bare mass effects on the reheating process after inflation

Simon Clery[a],* Marcos A. G. Garcia[b],† Yann Mambrini[a],‡ Keith A. Olive[c§]
[a] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France*
[b] *Departamento de Física Teórica, Instituto de Física, Universidad Nacional Autónoma de México, Ciudad de México C.P. 04510, Mexico*
[c] *William I. Fine Theoretical Physics Institute, School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA*

(Dated: February 28, 2024)

We consider the effects of a bare mass term for the inflaton, when the inflationary potential takes the form $V(\phi) = \lambda\phi^k$ about its minimum with $k \geq 4$. We concentrate on $k = 4$, but discuss general cases as well. Further, we assume $\lambda\phi_{\rm end}^2 \gg m_\phi^2$, where $\phi_{\rm end}$ is the inflaton field value when the inflationary expansion ends. We show that the presence of a mass term (which may be present due to radiative corrections or supersymmetry breaking) can significantly alter the reheating process, as the equation of state of the inflaton condensate changes from $w_\phi = \frac{1}{3}$ to $w_\phi = 0$ when $\lambda\phi^2$ drops below m_ϕ^2. We show that for a mass $m_\phi \gtrsim T_{\rm RH}/250$, the mass term will dominate at reheating. We compute the effects on the reheating temperature for cases where reheating is due to inflaton decay (to fermions, scalars, or vectors) or to inflaton scattering (to scalars or vectors). For scattering to scalars and in the absence of a decay, we derive a strong upper limit to the inflaton bare mass $m_\phi < 350$ MeV$(T_{\rm RH}/10^{10}$ GeV$)^{3/5}$, as there is always a residual inflaton background which acts as cold dark matter. We also consider the effect of the bare mass term on the fragmentation of the inflaton condensate.

I. INTRODUCTION

The hypothesis of a violent inflationary phase during the first moments of the Universe makes it possible to address several cosmological issues, ranging from the flatness of the Universe to the horizon or entropy problem [1]. However, a complete inflationary model requires above all a mechanism for a graceful exit. Indeed, the prolonged period of exponential expansion must end with a sufficiently efficient transfer of the oscillation modes of the inflaton condensate ϕ to a thermal bath [2, 3], i.e. reheating, that ensures a temperature $\gtrsim 2$ MeV to allow for standard big bang nucleosynthesis. Moreover, the density fluctuation spectrum produced during inflation should agree with observations of the CMB anisotropy spectrum [4], which in turn constrains the parameters of the inflaton potential $V(\phi)$.

The process of transferring the energy stored in inflaton oscillations to Standard Model particles is not instantaneous [5–8]. Rather, in many models, an oscillating inflaton condensate decays or scatters progressively producing a bath of relativistic particles. The efficiency of the reheating process depends on the rate of the energy transfer as well as on the shape of the inflaton potential, $V(\phi)$, about its minimum [9, 10]. Even if the exact shape of the potential at the end of inflation is unknown it can often be approximated about its minimum by a polynomial function of ϕ.

In many models of inflation, the inflaton potential can be approximated about its minimum by a quadratic term, $V(\phi) = \frac{1}{2}m_\phi^2\phi^2$. The Starobinsky model [11] is one example. In this case, only one Fourier mode of the inflaton oscillation contributes to the reheating process. The energy density in radiation, ρ_R, grows rapidly at first, and redshifts as $\rho_{\rm R} \propto a^{-\frac{3}{2}}$ where a is the cosmological scale factor, as decays continue to add to the radiation bath. Because $\rho_\phi \propto a^{-3}$, eventually, the radiation bath comes to dominate the total energy density, at which time we can define a reheating temperature. This occurs when the cosmological scale factor, $a_{\rm RH}$ satisfies $\rho_{\rm R}(a_{\rm RH}) = \rho_\phi(a_{\rm RH})$. This occurs (up to a numerical factor) when $H(a_{\rm RH}) \simeq \Gamma_\phi$, or $T_{\rm RH} \simeq \sqrt{\Gamma_\phi M_P}$, where H is the Hubble parameter, Γ_ϕ is the width of the inflaton condensate, and $M_P = 1/\sqrt{8\pi G_N} \simeq 2.4 \times 10^{18}$ GeV is the reduced Planck mass.

For a potential whose expansion about its minimum is $V(\phi) = \lambda\phi^k$, with $k \geq 4$, the exercise is more subtle, and requires a more involved analysis [9, 10]. The reheating process will in general depend on the spin of the final state particles in either inflaton decays or scatterings. In fact, in some cases reheating does not occur. For example, for $k = 4$, the evolution of $\rho_\phi \propto a^{-4}$ is the same as the evolution of $\rho_{\rm R} \propto a^{-4}$ for inflaton decays or scatterings to vector bosons [12], precluding the condition $\rho_\phi(a_{\rm RH}) = \rho_R(a_{\rm RH})$ to occur. However, we cannot exclude the presence of a bare mass term $\frac{1}{2}m_\phi^2\phi^2$, which may be subdominant at the end of inflation, and during

* simon.clery@ijclab.in2p3.fr
† marcos.garcia@fisica.unam.mx
‡ yann.mambrini@ijclab.in2p3.fr
§ olive@umn.edu

the early phases of the oscillations, but which becomes dominant when ϕ has redshifted down to a point $a_{\rm m}$ defined by $\lambda\phi^4(a_{\rm m}) = \frac{1}{2}m_\phi^2\phi^2(a_{\rm m})$. The presence of this term, even if it is small, would then modify the reheating mechanisms, making for example reheating by decays to vector bosons possible in the case $k = 4$.

Many models of inflation have potentials which, when expanded about their minimum, are described by a series of self interactions beyond their mass term. For example, the well studied Starobinsky potential [11], contains a full series of interaction terms. However, for $\phi < \phi_{\rm end}$, where $\phi_{\rm end} = \phi(a_{\rm end})$ is the inflaton field value when the inflationary expansion ends (when $\ddot{a} = 0$), terms which are higher order than the quadratic (mass term), become greatly suppressed and do not substantially affect the subsequent evolution of the inflaton condensate. In contrast, models such the so-called α-attractor T-models of inflation [13], described by a potential of the form,

$$V(\phi) \;=\; \lambda M_P^4 \left|\sqrt{6}\tanh\left(\frac{\phi}{\sqrt{6}M_P}\right)\right|^k , \qquad (1)$$

contain only even interaction terms starting with $\lambda M_P^{4-k}\phi^k$ yielding a massless inflaton for $k \geq 4$.

A bare mass term may be present at the tree level, may be produced as a result of supersymmetry breaking in a supersymmetric model, or may be produced radiatively. Though we will treat the mass as a free parameter, we note that there are 1) upper limits on the mass imposed by slow-roll parameters which determine the inflationary observables, n_s and r; 2) in the absence of fine-tuning, there is a lower bound on the mass derived from loop corrections to the potential à la Coleman-Weinberg. Both of these limits will be discussed below. In any case, the presence of a mass term seems unavoidable, at least at higher order, justifying a detailed analysis of its effect on the reheating process.

More specifically, the reheating phase in the T-models with $k \geq 4$, as an example, is altered when a mass term is added to the potential in Eq. 1. As a result, for $k = 4$, the evolution of the energy density transitions from a radiation-dominated Universe ($V(\phi) \propto \phi^4$, $\rho_\phi \propto a^{-4}$) to a matter-dominated Universe ($V(\phi) \propto \phi^2$, $\rho_\phi \propto a^{-3}$).[1] If the reheating process is sufficiently slow, the quadratic term can come to dominate the inflaton energy density and would result in higher reheating temperature than would have been achieved from the quartic term alone. The presence of a bare mass term generalizes previous results [9, 10].

Furthermore, it was recently shown in [14, 15] that the effects of the fragmentation of the inflaton condensate through its self interaction $\lambda\phi^k$, $k \geq 4$ could considerably affect the reheating process. It was noticed that in the case of reheating generated via fermion decay, fragmentation stopped the reheating process too early, leaving the Universe with a bath of massless (and thus stable) particles (inflatons). This would be in contradiction with CMB/BBN observations. However, if the quadratic term $\frac{1}{2}m_\phi^2\phi^2$ were to dominate *before* the end of the fragmentation of the inflaton ($a_{\rm m} < a_{\rm F}$ where $a_{\rm F}$ is the scale factor when fragmentation is complete), the latter would stop, allowing the condensate to continue the reheating process safely through its decay, $\Gamma_\phi \propto m_\phi$. For example, for $k = 4$, the conformal self-resonance responsible for the exponential growth of a narrow range of relativistic ϕ momentum modes is shut down as they become non-relativistic. The effective frequencies lose the oscillatory driving, and become incapable of fragmenting the inflaton condensate [14, 15]. This will be discussed in more detail below.

The paper is organized as follows: in Section II, we describe the effect of the transition from a $\phi^4 \to \phi^2$ potential on the evolution of the inflaton condensate and its impact on the reheating temperature. In Section III, we derive the upper limit to the inflaton mass from CMB observables and the bare mass expected from radiative corrections which in the absence of fine-tuning represents a lower limit to the mass. Then in Section IV, we derive the relations between the inflaton coupling to matter and the reheating temperature in view of the transition to a matter dominated expansion. These results are generalized to $k \neq 4$ in Section V and the consequences on the fragmentation of the inflaton condensate are discussed in Section VI. Our summary is found in Section VII.

II. THE TRANSITION, $\phi^4 \to \phi^2$

We begin by supposing that the dominant contribution in a series expansion of the inflaton potential about its minimum is the quartic term and that at the end of inflation, this dominates over a quadratic mass term, so that

$$\lambda\phi_{\rm end}^4 \gg \frac{1}{2}m_\phi^2\phi_{\rm end}^2 \,. \qquad (2)$$

For $a > a_{\rm end}$, the evolution of the energy density of ϕ is governed by the Friedmann equation for ρ_ϕ

$$\frac{d\rho_\phi}{dt} + 3(1+w)H\rho_\phi \simeq 0 \,. \qquad (3)$$

Where $\rho_\phi = \langle V(\phi)\rangle = V(\phi_0)$, the mean being taken over the oscillation of ϕ and ϕ_0 is the envelope of the oscillations. More precisely,

$$\phi(t) = \phi_0(t)\,\mathcal{P}(t) \,, \qquad (4)$$

[1] More generally, the Universe transitions from an expansion with an equation of state, $w = P_\phi/\rho_\phi = (k-2)/(k+2)$ ($V(\phi) \propto \phi^k$, $\rho_\phi \propto a^{-6k/(k+2)}$) to a matter-dominated Universe.

with $\mathcal{P}(t)$ a quasiperiodic function encoding the (an)harmonicity of short-timescale oscillations in the potential.

For $k = 4$, Eq. (3) gives

$$\rho_\phi = \rho_{\rm end}\left(\frac{a_{\rm end}}{a}\right)^4 , \tag{5}$$

where $\rho_{\rm end}$ is the value of the density of energy of the inflaton at the end of inflation, when $\ddot{a} = 0$. This condition is equivalent to $w = -1/3$ or $\dot{\phi}^2_{\rm end} = V(\phi_{\rm end})$. Hence,

$$\rho_{\rm end} = \frac{3}{2}V(\phi_{\rm end}) , \tag{6}$$

where for the T-models with potential given in Eq. (1) we have [10],

$$\phi_{\rm end} \simeq \sqrt{\frac{3}{8}}\, M_P \ln\left[\frac{1}{2} + \frac{k}{3}\left(k + \sqrt{k^2+3}\right)\right] . \tag{7}$$

The parameter λ in Eq. (1) is determined from the normalization of the CMB anisotropies [4]. The normalization of the potential for different values of k can be approximated by [10]

$$\lambda \simeq \frac{18\pi^2 A_{S*}}{6^{k/2} N_*^2} , \tag{8}$$

where N_* is the number of e-folds from horizon crossing to the end of inflation and $A_{S*} \simeq 2.1 \times 10^{-9}$ is the amplitude of the curvature power spectrum. For $N_* = 56$ e-folds we find $\lambda = 3.3 \times 10^{-12}$, and $\rho^{\frac{1}{4}}_{\rm end} = 4.8 \times 10^{15}$ GeV (when $k = 4$).

As ϕ_0 decreases, eventually the evolution of the condensate will be governed by the quadratic term. This occurs at $a = a_{\rm m}$ when

$$\frac{1}{2}m_\phi^2\phi_0^2(a_{\rm m}) = \lambda\phi_0^4(a_{\rm m}) . \tag{9}$$

Using $\phi_0^4(a) = (\rho_{\rm end}/\lambda)(\frac{a_{\rm end}}{a})^4$ for $a_{\rm end} < a < a_{\rm m}$ gives

$$\frac{a_{\rm m}}{a_{\rm end}} = \left(\frac{4\lambda\rho_{\rm end}}{m_\phi^4}\right)^{1/4} \simeq 9.1 \times 10^3 \left(\frac{10^9\ {\rm GeV}}{m_\phi}\right) . \tag{10}$$

In deriving (10), we note that the envelope function ϕ_0 is determined by the average energy density $\langle\rho_\phi\rangle = V(\phi_0)$. Thus unless reheating occurs rapidly, the quadratic term will dominate the reheating process even if the quartic dominates after when oscillations begin. This will have huge consequences on the reheating temperature, as well as on the physics of fragmentation as we will see.

Indeed, if reheating occurs at $a = a_{\rm RH} > a_{\rm m}$, the process is affected by the bare mass term. For $a > a_{\rm m}$, the equation of state changes from $w = 1/3$ (for $k = 4$) to $w = 0$ (for $k = 2$) and the solution for $a \ll a_m$ to the Friedmann equation becomes

$$\rho_\phi = \frac{1}{2}\rho_\phi(a_{\rm m})\left(\frac{a_{\rm m}}{a}\right)^3 = \rho_{\rm end}\left(\frac{a_{\rm end}}{a_{\rm m}}\right)^4\left(\frac{a_{\rm m}}{a}\right)^3 . \tag{11}$$

Furthermore,

$$\rho_{\rm m} \equiv \rho_\phi(a_{\rm m}) = 2\rho_{\rm end}\left(\frac{a_{\rm end}}{a_{\rm m}}\right)^4 = \frac{m_\phi^4}{2\lambda} . \tag{12}$$

Combining Eqs. (10) and (11) we obtain

$$\rho_\phi|_{a>a_{\rm m}} = \frac{m_\phi \rho_{\rm end}^{\frac{3}{4}}}{(4\lambda)^{\frac{1}{4}}}\left(\frac{a_{\rm end}}{a}\right)^3 . \tag{13}$$

This form for ρ_ϕ dominates the energy density until reheating defined by $\rho_\phi(a_{\rm RH}) = \rho_{\rm R}(a_{\rm RH})$. Here, $\rho_{\rm R}$ is the energy density transferred to the thermal bath via the Boltzmann equation

$$\frac{d\rho_R}{dt} + 4H\rho_R = (1+w)\Gamma_\phi\rho_\phi . \tag{14}$$

From the above, we can determine the reheating temperature for a given mass, m_ϕ for which the bare mass affects the reheating process, and therefore modifies the calculation of $T_{\rm RH}$. The condition $a_{\rm m} < a_{\rm RH}$ implies that $\rho_{\rm m} > \rho_\phi(a_{\rm RH})$ and thus the condition for the quadratic part to dominate the reheating process is given by

$$\rho_{\rm m} \gtrsim \rho_{\rm RH} \quad \Rightarrow \quad \rho_{\rm RH} \lesssim \frac{m_\phi^4}{2\lambda} , \tag{15}$$

from Eq.(12).

Defining $\rho_{\rm RH} = \alpha T_{\rm RH}^4$ with $\alpha = \frac{g_{\rm RH}\pi^2}{30}$ for $g_{\rm RH}$ relativistic degrees of freedom at $a_{\rm RH}$, we obtain

$$T_{\rm RH} \lesssim \frac{m_\phi}{(2\alpha\lambda)^{\frac{1}{4}}} \simeq 250\ m_\phi , \tag{16}$$

which means that if the energy transfer between the condensate and the thermal bath is slow and the reheating temperature $T_{\rm RH}$ lower than the limit obtained in the equation (16), we must take into account the quadratic term to determine $T_{\rm RH}$ when $a_{\rm RH} > a_{\rm m}$.

III. LIMITS ON THE INFLATON BARE MASS

As noted earlier, the CMB observables impose an upper limit to m_ϕ and in the absence of any fine-tuning, couplings of the inflaton to Standard Model fields (necessary for reheating), provide a lower bound to m_ϕ from radiative corrections to the potential.

Planck [4] has determined with relatively high precision, the value for the tilt of the CMB anisotropy spectrum, $n_s = 0.9649 \pm 0.0042$ (68% CL). In addition, the

tensor-to-scalar ratio, $r < 0.036$ is constrained by BICEP/Keck observations [16, 17]. To translate these limits to an upper limit on m_ϕ, we use the T-model in Eq. (1) as an example.

Recall that the conventional slow-roll parameters for a single-field inflationary model are given by

$$\epsilon \equiv \frac{1}{2} M_P^2 \left(\frac{V'}{V}\right)^2 , \qquad \eta \equiv M_P^2 \left(\frac{V''}{V}\right) , \tag{17}$$

where the prime denotes a derivative with respect to the inflaton field, ϕ. The number of e-folds can be computed using

$$N_* \simeq \frac{1}{M_P^2} \int_{\phi_{\rm end}}^{\phi_*} \frac{V(\phi)}{V'(\phi)} d\phi \simeq \int_{\phi_{\rm end}}^{\phi_*} \frac{1}{\sqrt{2\epsilon}} \frac{d\phi}{M_P} , \tag{18}$$

where ϕ_* corresponds to the horizon exit scale $k_* = 0.05\,{\rm Mpc}^{-1}$ used in the *Planck* analysis. The scalar tilt and tensor-to-scalar ratio can be expressed in terms of the slow roll parameters as

$$n_s \simeq 1 - 6\epsilon_* + 2\eta_* , \tag{19}$$

$$r \simeq 16\epsilon_* . \tag{20}$$

In a more precise model determination of N_*, and n_s, there is some dependence on the reheating temperature and equation of state [18, 19]. The computation is based on the self-consistent solution of the relation between N_* and its corresponding pivot scale k_*,

$$\begin{aligned} N_* = {} & \ln\left[\frac{1}{\sqrt{3}}\left(\frac{\pi^2}{30}\right)^{1/4}\left(\frac{43}{11}\right)^{1/3}\frac{T_0}{H_0}\right] - \ln\left(\frac{k_*}{a_0 H_0}\right) \\ & - \frac{1}{12}\ln g_{\rm RH} + \frac{1}{4}\ln\left(\frac{V(\phi_*)^2}{M_P^4 \rho_{\rm end}}\right) \\ & + \ln\left[\frac{a_{\rm end}}{a_{\rm RH}}\left(\frac{\rho_{\rm end}}{\rho_{\rm RH}}\right)^{1/4}\right] , \end{aligned} \tag{21}$$

where the present Hubble parameter and photon temperature are given by $H_0 = 67.36\,{\rm km\,s^{-1}\,Mpc^{-1}}$ [4] and $T_0 = 2.7255\,{\rm K}$ [20]. For the T-Models dominated by a quadratic term, agreement with Planck/BICEP/Keck data requires N_* between roughly 42 - 56 [21].

In the absence of a mass, $m_\phi = 0$, $N_* \simeq 56$ with $\phi_* = 6.96 M_P$ and $(n_s, r) = (0.964, 0.0034)$, independently of the efficiency of reheating [10, 15]. Therefore, to set limits on a possible mass term for $k = 4$, we set $N_* = 56$. For non-zero masses both n_s and r increase, but the limit on m_ϕ is determined mainly from n_s. Figs. 1 and 2 show the numerically computed CMB observables n_s and r for a variety of bare masses and inflaton-matter couplings. As is customary, the *Planck* ($k_* = 0.05\,{\rm Mpc}^{-1}$) and WMAP ($k_* = 0.002\,{\rm Mpc}^{-1}$) pivot scales are chosen for n_s and r, respectively. For $m_\phi \neq 0$, the effective equation-of-state parameter evolves as $w = -1/3 \to 1/3 \to 0 \to 1/3$ from the end of inflation to the end of reheating. The top panel of Fig. 1

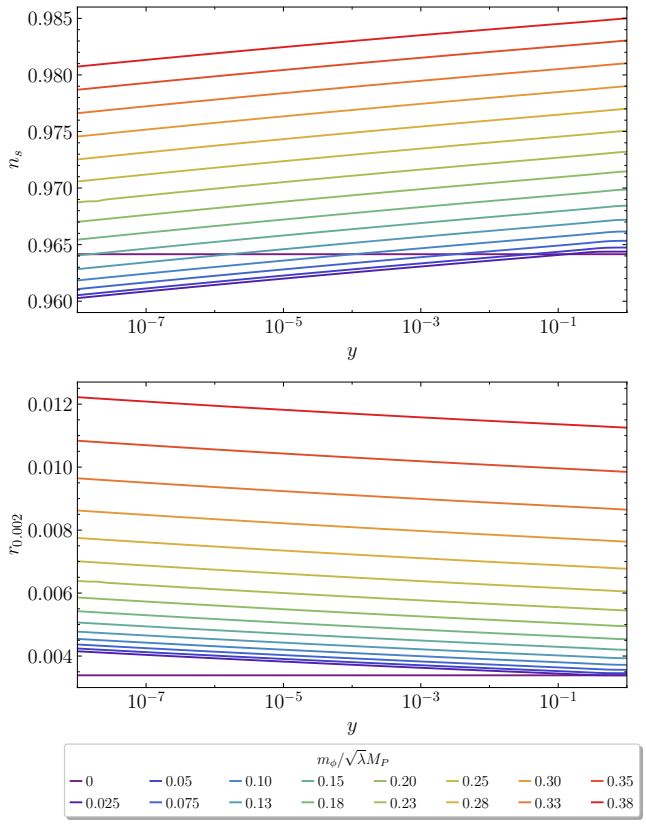


FIG. 1. *Scalar tilt n_s (top) and tensor-to-scalar ratio r (bottom) as functions of the Yukawa coupling y (22), for a selection of bare masses m_ϕ and $k = 4$.*

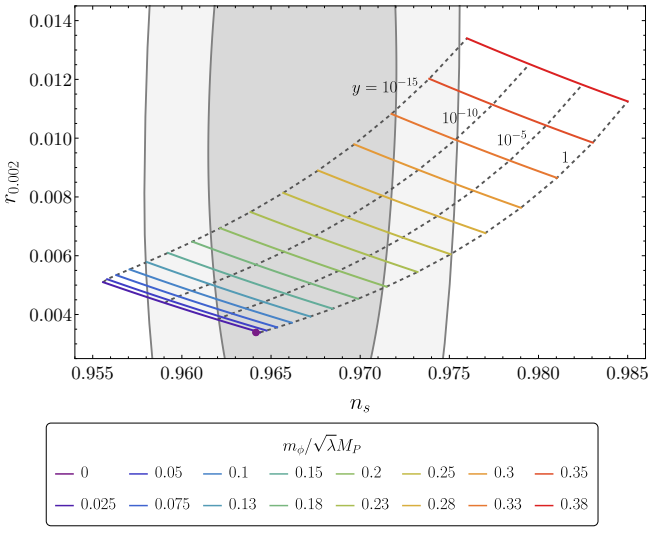


FIG. 2. *Same as Fig. 1 shown in the (n_s, r) plane. The gray (light gray) shaded regions correspond to the 68% (95%) C.L. Planck+BK18 regions [16].*

depicts the bare mass dependence of the scalar tilt, as a function of an inflaton-matter Yukawa coupling (see Eq. (22)). For $m_\phi = 0$, $N_* \simeq 56$ for any y, leading to

the purple horizontal line. For $m_\phi = 0.025\sqrt{\lambda}M_P$, the smallest non-zero mass in the Figure, the resulting curve presents two regimes. At $y \gtrsim 10^{-1}$, n_s is independent of y since reheating is completed before matter domination, $a_{\rm RH} < a_{\rm m}$. However, for $y \lesssim 10^{-1}$, reheating is completed by the dissipation of the quadratic, harmonic oscillations of ϕ. A dependence of n_s on y is induced, since now the last term of (21) is relevant for the determination of N_*. For smaller y reheating is delayed, resulting in a smaller N_* and as a consequence n_s. In the case of larger masses, the pure quartic regime is reduced, or outright lost, and the relation of n_s and y is determined by the duration of reheating in the matter dominated era, and the modification of the slow roll dynamics due to the presence of the large bare mass. Analogous conclusions can be drawn from the bottom panel of Fig. 1. In this case the addition of the bare mass increases the value of the tensor-to-scalar ratio, both from the modified inflation dynamics, and from the dependence on y of the number of e-folds N_*.

Fig. 2 compares the corresponding (n_s, r) curves against the *Planck*+BK18 constraints [16]. Here the range of couplings spans reheating temperatures from $T_{\rm RH} \sim 2 \times 10^{14}$ GeV for $y = 1$, to $T_{\rm RH} \sim \mathcal{O}(10)$ MeV for $y = 10^{-15}$. We note that for the smallest bare masses high reheating temperatures are favored by the CMB data. On the other hand, for the largest masses considered, lower $T_{\rm RH}$ are preferred. At the nominal $N_* = 56$, corresponding to $y \approx 1$ in the figure, we find that $m_\phi < 0.2\sqrt{\lambda}M_P \simeq 8.8 \times 10^{11}$ GeV at 68% CL with $(n_s, r) = (0.971, 0.0050)$ and $m_\phi < 0.25\sqrt{\lambda}M_P \simeq 1.1 \times 10^{12}$ GeV at 95% CL with $(n_s, r) = (0.975, 0.0061)$. Above these masses, the values of n_s and r rise very quickly and agreement with data is lost. Applying this limit on m_ϕ in Eq. (16) gives $T_{\rm RH} \lesssim 2.8 \times 10^{14}$ GeV. In other words, for larger reheating temperatures, the energy transfer is sufficiently efficient to avoid any interference of a possible quadratic interaction without violating the CMB data. Allowing for the full range in coupling y or equivalently $T_{\rm RH}$ and expanding the range in N_*, we see from Fig. 2, that the 68% CL upper limit is $m_\phi < 0.33\sqrt{\lambda}M_P = 1.4 \times 10^{12}$ GeV (for $y \geq 10^{-15}$ and a 95% CL upper limit of $m_\phi \lesssim 0.38\sqrt{\lambda}M_P = 1.6 \times 10^{12}$ GeV. For larger masses it becomes impossible to simultaneously satisfy the *Planck* constraints to 2σ and the BBN bound $T_{\rm RH} \gtrsim$ MeV.

In addition to an upper bound to m_ϕ, we expect that radiative corrections to the potential will provide finite mass which unless fine-tuned away, will determine a lower bound on the inflaton mass. We expect that through the coupling of the inflaton to either fermions or scalars, would lead to a mass term proportional to ym_f or μ (see Eqs. (22) and (35) for couplings to fermions and scalars respectively. While the former is probably no larger than the weak scale, the coupling to scalars could generate a significant contribution to m_ϕ. Furthermore, in a supersymmetric theory we would also expect contributions to the scalar mass of order the supersymmetry breaking scale. However, as noted, any lower limit to the inflaton mass would be subject to the degree of fine-tuning by canceling a bare mass term with any 1-loop corrections. Therefore unlike the upper limit discussed above, we do not apply a firm lower limit its mass, but recognize that it should not be surprising to generate weak scale masses, even in theories with the potential given in Eq. (1) for $k \geq 4$.

IV. CONSEQUENCES OF THE INFLATON COUPLING TO MATTER

Reheating to create a thermal bath of Standard Model particles requires some coupling of the inflaton to the Standard Model. The relation between this coupling and the reheating temperature is dependent not only on the shape of the inflaton potential about its minimum, but also on whether the reheating is produced by inflaton decay (in to either fermions, scalars or vectors) or scattering. As in [9, 10] we will study the three possible cases: fermion decay, scalar decay and scalar scattering, adding the vectorial final states (decay and scattering) analyzed in [12].

A. Inflaton decay to fermions

Given a Yukawa-like coupling of the inflaton to fermions,

$$\mathcal{L}_{\phi ff} = y\phi\bar{f}f\,, \tag{22}$$

the inflaton decay rate is

$$\Gamma_\phi = \frac{y_{\rm eff}^2}{8\pi} m_\phi\,. \tag{23}$$

Here, the effective Yukawa coupling $y_{\rm eff}(k) \neq y$ is defined by averaging over an oscillation. In general for $k \neq 2$, the effective coupling must be calculated numerically [10, 22, 23].

The general expressions for the reheating temperature, defined by $\rho_\phi(a_{\rm RH}) = \rho_{\rm R}(a_{\rm RH})$ and $\alpha T_{\rm RH}^4 = \rho_{\rm R}(a_{\rm RH})$, are given in the Appendix. $T_{\rm RH}$ depends strongly on the spin of the final state decay products, and for decays to fermions, Eq. (91) gives with $l = 1/2 - 1/k$ and $k < 7$

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}} \left[\frac{k\sqrt{3k(k-1)}}{7-k}\lambda^{\frac{1}{k}}\frac{y_{\rm eff}^2}{8\pi}\right]^{\frac{k}{4}} M_P\,, \tag{24}$$

or

$$T_{\rm RH} = \begin{cases} \left(\frac{\lambda}{\alpha}\right)^{\frac{1}{4}} \frac{y_{\rm eff}^2}{\pi} M_P \simeq 4.2 \times 10^{14} y_{\rm eff}^2 \text{ GeV} & k = 4\,, \\ \left(\frac{3}{\alpha}\right)^{\frac{1}{4}} \left(\frac{y_{\rm eff}^2 m_\phi M_P}{20\pi}\right)^{\frac{1}{2}} \\ \quad \simeq 3.3 \times 10^{12} y_{\rm eff}\sqrt{\frac{m_\phi}{10^9 \text{GeV}}} \text{ GeV} & k = 2\,. \end{cases} \tag{25}$$

Notable in Eq. (25) is that $T_{\rm RH}$ exhibits a different dependence on the coupling and mass of the inflaton. In particular, $T_{\rm RH} \propto y_{\rm eff}^2$ in the case $a_{\rm m} > a_{\rm RH}$, $T_{\rm RH} \propto y_{\rm eff}\sqrt{m_\phi}$ if $a_{\rm m} < a_{\rm RH}$. We will see that for sufficiently low coupling, the quadratic term can dominate the reheating process leading to a higher reheating temperature.

When the limit in Eq. (16) is satisfied, reheating is sufficiently late to be determined by the quadratic term ($k = 2$ in Eq. (25)) and that can be translated into a limit on the coupling $y_{\rm eff}$,

$$y_{\rm eff} \lesssim y_{\rm eff}^{\rm m} = 0.02\sqrt{\frac{m_\phi}{10^9\ {\rm GeV}}}\,. \tag{26}$$

We show in Fig. 3 the value of the reheating temperature as function of $y_{\rm eff}$ for different values of the inflaton bare mass $m_\phi = 10^3$, 10^9 and 10^{11} GeV, neglecting the effects of an effective final state mass (see below) and thus $y_{\rm eff} = y$. To obtain the figure, we solved numerically the complete set of Friedmann equations for ρ_R and ρ_ϕ, taking the full potential $V(\phi) = \frac{1}{2}m_\phi^2\phi^2 + \lambda\phi^4$. We also show for comparison with dashed lines, the analytical value of $T_{\rm RH}$ obtained in Eqs. (25). We clearly see the change of behavior $T_{\rm RH} = f(y_{\rm eff})$ below the limiting value in Eq. (26) where the bare mass term controls the final reheating temperature. For $y_{\rm eff} \lesssim y_{\rm eff}^{\rm m}$, $T_{\rm RH} \propto y_{\rm eff}$, whereas for larger values of $y_{\rm eff}$, when the reheating is dominated by the quartic part of the potential, the reheating temperature $\propto y_{\rm eff}^2$ and is independent of m_ϕ.

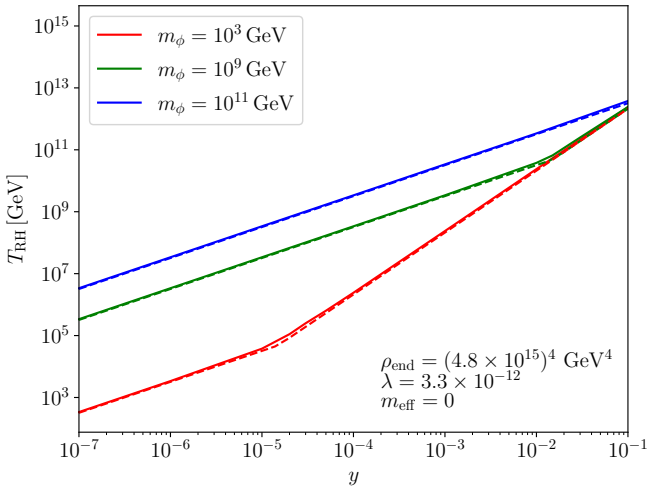


FIG. 3. *Reheating temperature as a function of the Yukawa coupling y when a bare mass term is added to a quartic potential ($k = 4$). Solid lines are obtained by solving numerically the Boltzmann equations for energy densities, while dashed lines are given by the analytical approximations in Eqs. (25) Here we neglect the effective mass of the final state fermion, $\mathcal{R} = 0$ and $y_{\rm eff} = y$.*

A background field value for ϕ, however, induces an effective mass for the fermion, f, $m_{\rm eff} = y\phi$, and the rates for producing the fermions are suppressed by $\mathcal{R}^{-1/2}$ where $\mathcal{R} \propto m_{\rm eff}^2/m_\phi^2 \propto y^2(\phi_0/M_P)^{4-k}/\lambda$ [10]. The mass of the inflaton is defined by

$$m_\phi^2(t) \;\equiv\; V''(\phi_0(t))\,. \tag{27}$$

When $\mathcal{R} \gg 1$, there is a significant suppression in the decay rate and $y_{\rm eff} \ll y$. Note that in the case of a quartic potential, $m_\phi \propto \phi$. As $m_{\rm eff} \propto \phi$ also, $\mathcal{R}$ is constant, $\mathcal{R} \simeq 1.4y^2/\lambda \simeq 4.2 \times 10^{11}y^2$. In other words, the effect of $\mathcal{R}$ results in a suppression of the reheating efficiency by a constant factor $\mathcal{R}^{-\frac{1}{2}} \simeq 1.5 \times 10^{-6}/y$ throughout the reheating process. This suppression begins to be efficient ($\mathcal{R} \gtrsim 1$) for $y \gtrsim 1.5 \times 10^{-6}$ [10]. On the other hand, for a quadratic potential, $\mathcal{R} = 4(\phi_0/m_\phi)^2y^2$ *decreases* with time, redshifting as $a^{-\frac{3}{2}}$. Which means that if there is no suppression during the quartic dominated era ($a < a_{\rm m}$), there is no suppression in the quadratic era ($a > a_{\rm m}$).

The kinematic suppression in the effective coupling $y_{\rm eff}$ for $\mathcal{R} \gg 1$ can be parametrized as [10]

$$y_{\rm eff}^2 \simeq c_k\mathcal{R}^{-1/2}(\omega/m_\phi)y^2 \tag{28}$$

where c_k is a k-dependent constant[2] and ω is the oscillation frequency. For $k = 4$, $c_4 \simeq 0.5$ and $\omega \simeq 0.49m_\phi$. This leads to

$$y_{\rm eff} \simeq \frac{1}{2} \times \frac{y}{\mathcal{R}^{\frac{1}{4}}} \simeq 6 \times 10^{-4}\sqrt{y} \qquad (k = 4)\,. \tag{29}$$

We note that only when $\mathcal{R} \sim 0.1$, do we recover $y_{\rm eff} = y$. Note also that unless $y_{\rm eff}$ is relatively small, $y_{\rm eff} \lesssim 2 \times 10^{-3}$, the Lagrangian coupling, y, is non-perturbative [10, 14], where this perturbativity limit on $y_{\rm eff}$ assumes $y \lesssim \sqrt{4\pi}$.

For $k = 2$, $c_2 \simeq 0.38$ and $\omega = m_\phi$. At the end of reheating, $\rho_{\rm RH} = \frac{1}{2}m_\phi^2\phi_0^2(a_{\rm RH}) = \alpha T_{\rm RH}^4$, so that

$$\phi_0(a_{\rm RH}) = \sqrt{2\alpha}\frac{T_{\rm RH}^2}{m_\phi}\,. \tag{30}$$

Then, for $\mathcal{R} \gg 1$, we can write

$$y_{\rm eff} \simeq 0.15(m_\phi/T_{\rm RH})\sqrt{y} \qquad (k = 2)\,, \tag{31}$$

and using Eq. (25) for $T_{\rm RH}$ in terms of $y_{\rm eff}$ we have

$$y_{\rm eff} = 6.7 \times 10^{-3}\left(\frac{m_\phi}{10^9\ {\rm GeV}}\right)^{\frac{1}{4}} y^{\frac{1}{4}} \qquad (k = 2)\,. \tag{32}$$

In this case, non- perturbativity sets in unless $y_{\rm eff} \lesssim 1.5(m_\phi/\phi_0)^{\frac{1}{2}}$, assuming that $y_{\rm eff} < y$. Note that for $k > 4$

[2] There is an additional dependence of $y_{\rm eff}$ on the sum of the Fourier modes associated with the inflaton oscillations in the potential $V(\phi) \sim \phi^k$, for each value of k. However, this additional dependence is $\mathcal{O}(1)$, as shown in [10].

the limit becomes more severe as $\mathcal{R}$ is larger and increases in time.

Because of the suppression in the decay rate, the relation between $T_{\rm RH}$ and the decay coupling y shown in Fig. 3 needs to be reassessed. Indeed, when $y \gtrsim 1.5\times10^{-6}$, $\mathcal{R} \gtrsim 1$ and the suppression effect should be taken into account. The relation between $T_{\rm RH}$ and y when the effects of kinematic suppression are included is shown in Fig. 4. At very low values of y, $\mathcal{R} \ll 1$ and the suppression effects can be ignored. In this case, the relation between $T_{\rm RH}$ and y is unaffected. However, when $y_{\rm eff} \leq y$ the relation is altered. From Eq. (32), this occurs when $y > 1.3\times10^{-6}(m_\phi/{\rm GeV})^{\frac{1}{3}}$, or when $y > 1.3\times10^{-5}(1.3\times10^{-3})(6\times10^{-3})$ when $m_\phi = 10^3$ (10^9) (10^{11}) GeV. These values are seen in Fig. 4 when the solid curves begin to deviate from the dashed curves. The dashed curves show the relation in Fig. 3 when suppression effects are ignored. The expression for $y_{\rm eff}$ in Eq. (32) can be inserted in Eq. (25) to obtain the relation between $T_{\rm RH}$ and y for when suppression effects are included and reheating is governed by the quadratic term,

$$T_{\rm RH} = 2.2\times10^{10}\ {\rm GeV}\left(\frac{m_\phi}{10^9{\rm GeV}}\right)^{\frac{3}{4}} y^{\frac{1}{4}} \qquad (k=2)\,. \tag{33}$$

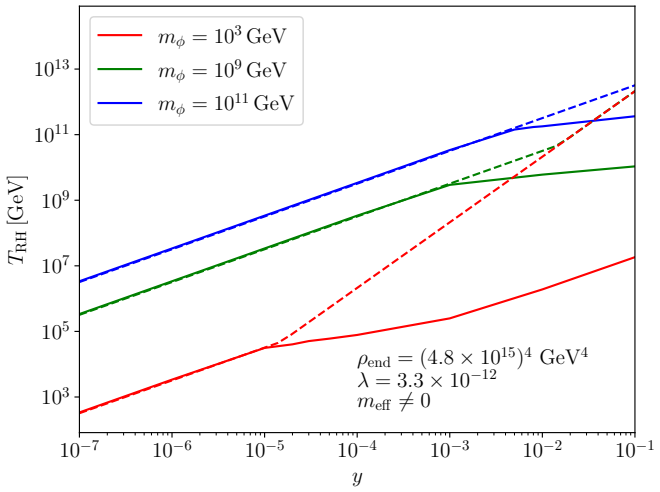


FIG. 4. *As in Fig. 3, the reheating temperature as function of the yukawa coupling y for different values of the inflaton bare mass $m_\phi = 10^3$ GeV (red-dotted), 10^9 GeV (green-dashed) and 10^{11} GeV (full-blue). Here we consider the effective mass of produced fermion, $\mathcal{R} = (2y\phi_0/\omega)^2$.*

We saw previously in Eq. (16) that the reheating temperature is determined by the quartic term only if $T_{\rm RH} \gtrsim 250 m_\phi$. When the kinematic suppression effects are ignored ($y = y_{\rm eff}$), this occurs when y does not satisfy Eq. (26). In this case, we can use Eq. (29) to determine the relation between $T_{\rm RH}$ and y,

$$T_{\rm RH} = 1.5\times10^8 y\ {\rm GeV} \qquad (k=4)\,, \tag{34}$$

and thus we expect that reheating is determined by the quartic term when $y > 1.7\times10^{-6} m_\phi/{\rm GeV}$. This occurs at $y = 1.7\times10^{-3}$ for $m_\phi = 10^3$ GeV as can be seen in Fig. 4. For the larger masses shown, we see that the transition would only occur in the non-perturbative regime (with $y \gg 1$) and so for the two higher masses, the reheating temperature is always determined by the quadratic mass term.

B. Decay to scalars

Another possibility is that reheating occurs predominantly through inflaton decay to scalars, through the coupling

$$\mathcal{L}_{\phi b^2} = \mu\phi b^2 \tag{35}$$

where b is a real scalar field. As was the case for the fermion decay, there is also an effect from the effective mass of the scalar field, and we parameterize it by considering an effective coupling $\mu_{\rm eff}$. We note that $\mu_{\rm eff}$ is now a dimensionful parameter and is *enhanced* (and not reduced) by $\mathcal{R}^{1/2}$ [10]. The associated decay rate is given by

$$\Gamma_{\phi b^2} = \frac{\mu_{\rm eff}^2}{8\pi m_\phi}\,. \tag{36}$$

For $k = 2$ this effective coupling reduces to the Lagrangian coupling μ but is different for $k > 2$. It is important to note that in this case, as m_ϕ decreases with time, the decay rate *increases* with time.

For decays to scalars, $l = 1/k - 1/2$, and using the appropriate expression found in the Appendix for γ_ϕ, we have

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{2k\sqrt{3}}{(4k+2)\sqrt{k(k-1)}}\lambda^{-\frac{1}{k}}\frac{\mu_{\rm eff}^2}{8\pi M_P^2}\right]^{\frac{k}{4(k-1)}} M_P\,, \tag{37}$$

or

$$T_{\rm RH} = \begin{cases} \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left(\frac{\mu_{\rm eff}^2}{36\pi M_P^2}\right)^{\frac{1}{3}}\lambda^{-\frac{1}{12}} M_P \\ \qquad \simeq 1.8\times10^{18}\left(\frac{\mu_{\rm eff}}{M_P}\right)^{\frac{2}{3}}{\rm GeV} \qquad k=4\,, \\ \left(\frac{3}{\alpha}\right)^{\frac{1}{4}}\left(\frac{M_P}{20\pi m_\phi}\right)^{\frac{1}{2}}\mu_{\rm eff} \\ \qquad \simeq 3.3\times10^3\mu_{\rm eff}\sqrt{\frac{10^9{\rm GeV}}{m_\phi}} \qquad k=2\,, \end{cases} \tag{38}$$

We show in Fig. 5 the evolution of $T_{\rm RH}$ as function of μ for the same set of masses $m_\phi = 10^3$, 10^9 and 10^{11} GeV, in the simplified case with $m_{\rm eff} = 0$. We clearly recognize the dependence $T_{\rm RH} \propto \mu$ for the smaller values of μ and $T_{\rm RH} \propto \mu^{2/3}$ for the larger values, when reheating is dominated by the quartic part of the potential. The value of μ for which reheating is dominated by the quadratic

term obtained from Eq. (38) with $k = 4$ is

$$\mu \lesssim 1.3 \times 10^8 \left(\frac{m_\phi}{10^9\ \text{GeV}}\right)^{\frac{3}{2}}\ \text{GeV}\,, \tag{39}$$

which is effectively what is observed in Fig. 5. From Eq. (39), we see that the reheating temperature for $m_\phi = 10^3$ GeV (red curve) is always due to the quartic term, as the transition from quadratic to quartic occurs at a low value of μ beyond the range shown. For the larger values of m_ϕ, Eq. (39) indicates when when the slopes of $T_{\rm RH}$ vs. μ begins to change.

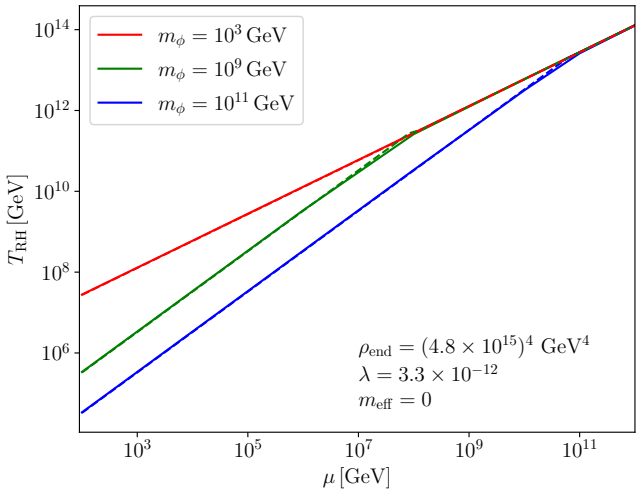


FIG. 5. *Reheating temperature as function of the bosonic coupling μ, for different values of the inflaton bare mass $m_\phi = 10^3$ GeV (red), 10^9 GeV (green) and 10^{11} GeV (blue). Solid lines are obtained by solving numerically the Boltzmann equations for energy densities, while dashed lines are given by the analytical approximations in Eqs. (38). Here we neglect the effective mass of produced bosons, $\mathcal{R} = 0$.*

In order to account for the effective mass $m_{\rm eff}^2 = 2\mu\phi_0$, we need to include an *enhancement* of the production rate $\propto \mathcal{R}^{\frac{1}{2}}$, with $\mathcal{R} = 8\mu\phi_0/m_\phi^2$ for $k = 2$ and $\mathcal{R} \simeq 2.8\mu/(\lambda\phi_0)$ for $k = 4$. The effective dimensionful coupling[3] when $\mathcal{R} \gg 1$ is [10]

$$\mu_{\rm eff}^2 \simeq \frac{c'_k}{4}(k+2)(k-1)\frac{\omega}{m_\phi}\mathcal{R}^{\frac{1}{2}}\mu^2\,, \tag{40}$$

with $c'_k \simeq \{0.38, 0.37, 0.36\}$ for $k = \{2, 4, 6\}$, so that $\mu_{\rm eff} \simeq 0.62(8\phi_0/m_\phi^2)^{\frac{1}{4}}\mu^{\frac{5}{4}}$ for $k = 2$. Then using Eq. (30) for ϕ_0 and Eq. (38) for $k = 2$ to replace $T_{\rm RH}$, we have

$$\mu_{\rm eff} \simeq 3.3 \times 10^{-10}\ \text{GeV}\left(\frac{10^9\ \text{GeV}}{m_\phi}\right)^2\left(\frac{\mu}{\text{GeV}}\right)^{\frac{5}{2}}. \tag{41}$$

[3] Again, an additional $\mathcal{O}(1)$ dependence of $\mu_{\rm eff}$ on the sum of the Fourier modes associated with the inflaton oscillations for each value of k is neglected here [10]. Note also that the values of c'_k were omitted in [10].

Then, the effects of the kinematic enhancement will occur when

$$\mu \gtrsim 2.1 \times 10^6 \left(\frac{m_\phi}{10^9\ \text{GeV}}\right)^{\frac{4}{3}}. \tag{42}$$

This can be see seen in Fig. 6 for $m_\phi = 10^9$ (10^{11} GeV as the point when the solid curves break away from the dashed curves at $\mu \simeq 2.1 \times 10^6$ (9.8×10^8) GeV respectively. At lower values of μ, the effects of the kinematic suppression can be ignored. For $m_\phi = 10^3$ GeV, this occurs at a value of μ below the range shown.

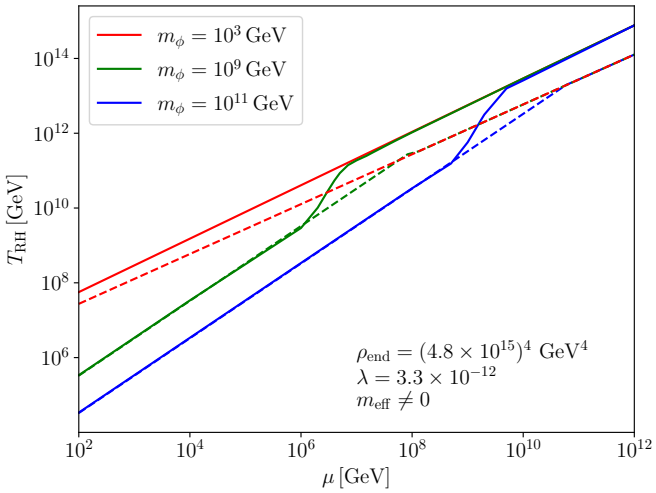


FIG. 6. *Reheating temperature as function of the bosonic coupling μ, for different values of the inflaton bare mass $m_\phi = 10^3$ GeV (red), 10^9 GeV (green) and 10^{11} GeV (blue). Solid lines are obtained by solving numerically the Boltzmann equations for energy densities, while dashed lines are given by the analytical approximations in Eqs. (38). Here we consider the effective mass of produced bosons, $\mathcal{R} = 8\mu\phi_0/\omega^2$.*

In the region when $\mu_{\rm eff} > \mu$ and quadratic reheating dominates, we can insert Eq. (41) into Eq. (38) to obtain

$$T_{\rm RH} = 1.1 \times 10^{-6}\ \text{GeV}\left(\frac{10^9\ \text{GeV}}{m_\phi}\right)^{\frac{5}{2}}\mu^{\frac{5}{2}} \qquad (k = 2)\,. \tag{43}$$

At higher values of μ the transition to quartic reheating occurs and using Eq. (40) with the expression for $\mathcal{R}$ for $k = 4$, we find that

$$\mu_{\rm eff} \simeq 2.5\ \text{GeV}\left(\frac{\mu}{\text{GeV}}\right)^{\frac{15}{14}} \tag{44}$$

which when inserted in Eq. (38) gives

$$T_{\rm RH} \simeq 2.5 \times 10^{19}\ \text{GeV}\left(\frac{\mu}{M_P}\right)^{\frac{5}{7}} \qquad (k = 4)\,. \tag{45}$$

C. Decay to Vectors

Recently, we have considered the possibility of inflaton decays to vectors [12] motivated by inflationary models in the context of no-scale supergravity [25] (which easily lend construction of the T-models considered here [9]). Often in such models, the inflaton couplings to matter fermions and scalars are highly suppressed [26–28] and reheating is only possible if the gauge kinetic functions contain inflaton couplings. The inflaton to vector couplings can be parameterized by

$$\mathcal{L} \supset -\frac{g}{4M_P}\phi F_{\mu\nu}F^{\mu\nu} - \frac{\tilde{g}}{4M_P}\phi F_{\mu\nu}\tilde{F}^{\mu\nu}\,, \tag{46}$$

From these Lagrangian couplings, we can derive the inflaton decay rate

$$\Gamma_{\phi\to A_\mu A_\mu} = \frac{\alpha_{\rm eff}^2 m_\phi^3}{M_P^2}\,, \tag{47}$$

where $\alpha_{\rm eff}^2 = (g_{\rm eff}^2 + \tilde{g}_{\rm eff}^2)/(64\pi)$. Note the dependence of the width on m_ϕ^3, which is very different from the decay into fermions ($\propto m_\phi$) and to scalars ($\propto 1/m_\phi$). $\Gamma_{\phi\to A_\mu A_\mu}$ decreases much more rapidly than $\Gamma_{\phi\to ff}$, rendering the reheating much less efficient, even impossible as long as the reheating is dominated by the quartic term.

Indeed, for decay to vectors, $l = 3/2 - 3/k$, and using the appropriate expression for γ_ϕ, we have from Eq. (91)

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{\sqrt{3}k^{\frac{5}{2}}(k-1)^{\frac{3}{2}}\lambda^{\frac{3}{k}}}{13-4k}\alpha_{\rm eff}^2\right]^{\frac{k}{4(3-k)}} M_P\,. \tag{48}$$

This expression is valid so long as $k+8-6kl > 0$, which is the case for $k = 2$, but not for $k \geq 4$. For $k+8-6kl < 0$, the reheating temperature is given by Eq. (92) for $k > 4$. For $k = 4$, the radiation density in Eq. (88) scales as a^{-4} as does the inflaton energy density in Eq. (84) and we never achieve the condition that $\rho_\phi(a_{\rm RH}) = \rho_R(a_{\rm RH})$ and reheating never occurs. Thus we have

$$T_{\rm RH} = \begin{cases} \text{no reheating} & k = 4\,, \\ \left(\frac{3}{\alpha}\right)^{\frac{1}{4}}\left(\frac{2m_\phi^3}{5M_P^3}\right)^{\frac{1}{2}}\alpha_{\rm eff}M_P & \\ \quad \simeq 7.0\times 10^3\alpha_{\rm eff}\left(\frac{m_\phi}{10^9\,\text{GeV}}\right)^{\frac{3}{2}}\ \text{GeV} & k = 2\,, \end{cases} \tag{49}$$

Thus for a $k = 4$ inflationary potential, reheating via the decays to vector bosons does not occur in the absence of a bare mass term. The bare mass term is then *necessary* to ensure a successful reheating. However, the bare mass term should ensure $T_{\rm RH} \gtrsim 2$ MeV, which means

$$m_\phi \gtrsim 40\alpha_{\rm eff}^{-\frac{2}{3}}\ \text{TeV}\,. \tag{50}$$

This value is the minimal bare mass necessary to have reheating through decay to vectors for $k = 4$.

Finally we note that there are no kinematic enhancement/suppression effects in this case. Since the inflaton is coupled to F^2 (as opposed to A^2), no mass term is generated. Then $g_{\rm eff} = g$ (and $\tilde{g}_{\rm eff} = \tilde{g}$) for $k = 2$, and for $k = 4$ only differs by a Fourrier coefficient in an expansion of $V(\phi)$ [12].

D. Scattering to scalars

We can also consider the case where the inflaton transfers its energy through the coupling

$$\mathcal{L}_{\phi^2 b^2} = \sigma\phi^2 b^2 \tag{51}$$

where b is a real scalar field. The associated decay rate is given by [10]

$$\Gamma_{\phi^2 b^2} = \frac{\sigma_{\rm eff}^2}{8\pi}\frac{\rho_\phi}{m_\phi^3} \tag{52}$$

where we have introduced the effective coupling $\sigma_{\rm eff}$ obtained, as for $y_{\rm eff}$ and $\mu_{\rm eff}$, after averaging over oscillations of the background inflaton condensate [10]. This effective coupling is equal to the Lagrangian coupling σ for $k = 2$ but is different for $k > 2$ and as in the case of decays to fermions there is a kinematic suppression.

For scattering to scalars, $l = 3/k - 1/2$, and using the appropriate expression found in the Appendix for γ_ϕ, we have from Eq. (91) valid when $k \geq 4$,

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{\sqrt{3}}{(2k-5)\sqrt{k}(k-1)^{\frac{3}{2}}}\lambda^{-\frac{3}{k}}\frac{\sigma_{\rm eff}^2}{8\pi}\right]^{\frac{k}{4(k-3)}} M_P\,. \tag{53}$$

For $k = 2$, $8+k-6kl < 0$ and ρ_R redshifts as a^{-4} which is faster than $\rho_\phi \propto a^{-3}$. Thus, in this case, reheating is not possible if the quadratic term becomes dominant before reheating is complete. The reheating temperature can then be written as,

$$T_{\rm RH} = \begin{cases} \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left(\frac{\sigma_{\rm eff}^2}{144\pi}\right)\lambda^{-\frac{3}{4}}M_P & \\ \quad \simeq 8.9\times 10^{23}\sigma_{\rm eff}^2\text{GeV} & k = 4\,, \\ \text{no reheating} & k = 2\,. \end{cases} \tag{54}$$

As one can see, the possibility of reheating through scattering to scalars is opposite the case of decays to vectors. Reheating is *not possible* when the quadratic part of the potential dominates the reheating process. Naively, when we neglect the kinematic suppression effects in $\mathcal{R}$, reheating is therefore only possible if the limit in Eq. (16) is violated, namely

$$\sigma_{\rm eff} \gtrsim 5.3\times 10^{-7}\sqrt{\frac{m_\phi}{10^9\ \text{GeV}}}\,. \tag{55}$$

For smaller couplings, the quadratic term will dominate before reheating is complete, and as a result never completes. We note in the expression for $T_{\rm RH}$ in Eq. (54), the

maximum value for $\sigma_{\rm eff}$ that can be used is determined from $a_{\rm RH} > a_{\rm end}$, which gives

$$\sigma_{\rm eff}^2 < 2.2 \times 10^{-9} \,. \tag{56}$$

Furthermore for self couplings this large, we expect that non-perturbative effects become non-negligible [24]. For larger values, we have a maximum reheating temperature of 2×10^{15} GeV, which is basically determined from $\rho_{\rm end}$.

As previously noted, for inflaton scattering to scalars, there is a kinematic suppression when $\mathcal{R} > 1$. In this case, for $k = 4$, $\mathcal{R} \simeq 2.8\sigma/\lambda$ is a constant and [4]

$$\begin{aligned} \sigma_{\rm eff}^2 &\simeq \frac{c_k''}{8} k(k+2)(k-1)^2 \mathcal{R}^{-1/2} (\omega/m_\phi) \sigma^2 \\ &\simeq 16 \mathcal{R}^{-1/2} \sigma^2 \simeq 9.6 \sqrt{\lambda} \sigma^{\frac{3}{2}} \,, \end{aligned} \tag{57}$$

using $c_4'' = 1.22$. Then the reheating temperature in terms of σ becomes

$$T_{\rm RH} = 1.6 \times 10^{19} \text{ GeV} \sigma^{\frac{3}{2}} \qquad (k = 4) \,. \tag{58}$$

In Fig. 8, we compare the reheating temperature as a function of σ when kinematic effects are ignored to the case where they are included. From Eq. (57), these effects become important when $\sigma > 3.1 \times 10^{-10}$. The dashed lines correspond to the solution when kinematic effects are ignored. The abrupt increase in $T_{\rm RH}$ occurs when Eq. (55) is satisfied (and $\sigma_{\rm eff} = 4\sigma$). In contrast, the solid lines include the kinematic suppression and reheating is possible when Eq. (55) is used with Eq. (57) or when

$$\sigma \gtrsim 6.4 \times 10^{-6} \left(\frac{m_\phi}{10^9 \text{ GeV}} \right)^{\frac{2}{3}} \,. \tag{59}$$

This limit accounts for the abrupt rise in $T_{\rm RH}$ for the solid lines in Fig. 8. At higher coupling, the reheating temperature follows Eq. (58) and scales as $\sigma^{\frac{3}{2}}$ as opposed to σ^2 when the suppression effects are ignored. In the latter case, we see the curves flatten at large coupling since $a_{\rm RH}$ is approaching $a_{\rm end}$ and the approximation used in (54) breaks down. These curves end when $a_{\rm RH} = a_{\rm end}$, indicated by the vertical gray dotted line. The solid curves would end when $\sigma \simeq 0.002$.

In the absence of a decay term for the inflaton, a bare mass term will eventually lead to a non-zero relic density of inflatons after annihilations freeze out. Indeed, even if $\sigma_{\rm eff}$ is sufficiently large and respects the condition (55), the presence of a quadratic term may dominate the energy budget of the Universe. Thus we can derive a limit

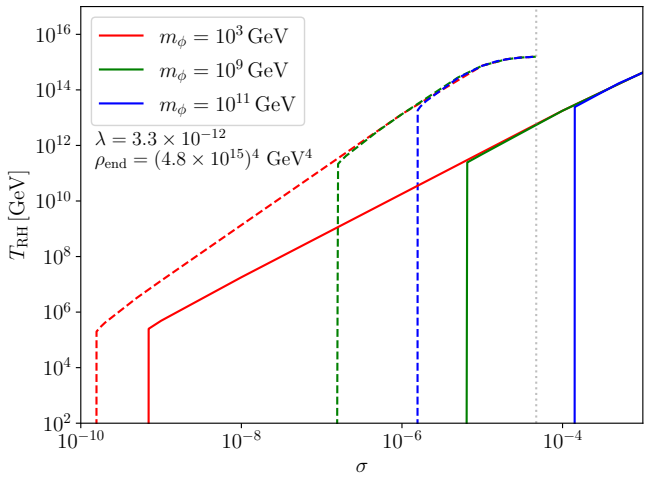


FIG. 7. *Reheating temperature as function of the scattering coupling σ, for different values of the inflaton bare mass $m_\phi = 10^3$ GeV (red), 10^9 GeV (green) and 10^{11} GeV (blue). Solid lines are obtained by solving numerically the Boltzmann equations for energy densities including the effect of $\mathcal{R}$, while dashed lines neglect the effect of effective masses. The vertical gray dotted line corresponds to the limit Eq.(56), when neglecting the effect of $\mathcal{R}$.*

on a combination of the inflaton mass, $T_{\rm RH}$ and the coupling σ. Saturating the limit leaves us with the inflaton as a cold dark matter candidate! [5]

Indeed, for $\sigma_{\rm eff}$ sufficiently large to ensure reheating with $k = 4$, for $a > a_{\rm RH}$, the evolution of ρ_ϕ is determined from the Boltzmann equation including dissipative effects [10]

$$\frac{d}{da} \left(\rho_\phi a^{\frac{6k}{k+2}} \right) = - \frac{\gamma_\phi}{aH} \frac{2k}{k+2} \frac{\rho_\phi^{l+1}}{M_P^{4l}} a^{\frac{6k}{k+2}} \,. \tag{60}$$

and for $k = 4$, $l = \frac{1}{4}$ and γ_ϕ as given in the Appendix, we find that ρ_ϕ scales as

$$\rho_\phi = 256 \rho_{\rm RH} \left(\frac{a_{\rm RH}}{a} \right)^8 \,. \tag{61}$$

Here, we used $H = \sqrt{\rho_R / 3 M_P^2}$. In the absence of a mass term, since $\Gamma_\phi \propto \gamma_\phi \rho_\phi^{\frac{1}{4}} \propto a^{-2}$ and after reheating, $H \propto \rho_R^{\frac{1}{2}} \propto a^{-2}$, the ratio Γ/H remains constant and the scaling in Eq. (61) remains true indefinitely and the density of inflatons becomes negligibly small.

However, when $m_\phi \neq 0$, eventually the mass term dominates over the quartic term (at $a = a_{\rm m}$) and we can

[4] We neglect the dependence of $\sigma_{\rm eff}$ on the sum of the Fourier modes associated with the inflaton oscillations for each value of k [10].

[5] The possibility of inflaton dark matter in a similar context was considered in [29] where the conditions for freeze-out of a thermal inflaton given. See also [30–34].

determine $a_{\rm m}$, when $\rho_\phi(a_{\rm m}) = \frac{1}{2}m_\phi^2\phi^2(a_{\rm m})$,

$$\frac{a_{\rm m}}{a_{\rm RH}} = \frac{\rho_{\rm RH}^{\frac{1}{8}}\lambda^{\frac{1}{8}}2^{\frac{1}{8}}}{\sqrt{m_\phi}}\,, \tag{62}$$

where the inflaton density is given by

$$\rho_\phi^{\rm m} = \rho_\phi(a_{\rm m}) = \frac{m_\phi^4}{2\lambda}\,. \tag{63}$$

as was previously found in Eq. (12).

For $a > a_{\rm m}$, Eq. (60) can be solved, now with $k = 2$ and $l = 1$. In the limit that $a \gg a_{\rm m}$, the residual inflaton density is given by

$$\rho_\phi(a) \simeq \rho_\phi^{\rm m}\left(\frac{a_{\rm m}}{a}\right)^3\,, \tag{64}$$

so long as $(m_\phi/M_P) \ll (2\lambda)^{\frac{1}{4}}/3^{\frac{1}{3}} \approx .001$, which is always true given the upper limits on m_ϕ discussed in Section III. Thus the presence of a mass term in the case where reheating is determined by a quartic coupling of the inflaton to scalars (which requires $k > 2$), leads automatically to cold dark matter candidate.

Given the inflaton density in Eq. (64), it is straightforward to compute the relic density today and in effect set a limit on the inflaton bare mass. Today,

$$\rho_\phi = \frac{8m_\phi^{\frac{5}{2}}\alpha^{\frac{3}{8}}T_0^3}{(2\lambda)^{\frac{5}{8}}T_{\rm RH}^3 2}\xi\,, \tag{65}$$

where $\xi = (43/427)(4/11) \simeq 0.036$ and relative to the critical density we have

$$\Omega_\phi h^2 = 1.6\left(\frac{m_\phi}{1\ {\rm GeV}}\right)^{\frac{5}{2}}\left(\frac{10^{10}\ {\rm GeV}}{T_{\rm RH}}\right)^{\frac{3}{2}} \tag{66}$$

and thus

$$m_\phi < 0.35\left(\frac{T_{\rm RH}}{10^{10}\ {\rm GeV}}\right)^{\frac{3}{5}}\ {\rm GeV}\,, \tag{67}$$

using $\Omega_\phi h^2 < 0.12$. This is a remarkably strong limit on a bare mass term for the inflaton if it remains stable.

E. Scattering to Vectors

If the gauge kinetic function is quadratic in the inflaton, then scattering rather decay to vectors occurs. In this case, the inflaton to vector couplings can be parameterized by

$$\mathcal{L} \supset -\frac{\kappa}{4M_P^2}\phi^2 F_{\mu\nu}F^{\mu\nu} - \frac{\tilde\kappa}{4M_P^2}\phi^2 F_{\mu\nu}\tilde F^{\mu\nu}\,, \tag{68}$$

From these Lagrangian couplings, we can derive the inflaton decay rate

$$\Gamma_{\phi\phi\to A_\mu A_\mu} = \frac{\beta^2\rho_\phi}{M_P^4}m_\phi\,, \tag{69}$$

where $\beta^2 = (\kappa_{\rm eff}^2 + \tilde\kappa_{\rm eff}^2)/(4\pi)$.

For scattering to vectors, $l = 3/2 - 1/k$, and using the appropriate expression for γ_ϕ, we have from Eq. (92)

$$\begin{aligned} T_{\rm RH} &= \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{\sqrt{3}k}{4k-7}\beta^2(k(k-1))^{\frac{1}{2}}\lambda^{\frac{1}{k}}\right]^{\frac{3k}{4k-16}} \\ &\times\left(\frac{\rho_{\rm end}}{M_P^4}\right)^{\frac{4k-7}{4k-16}}M_P\,. \end{aligned} \tag{70}$$

since $8 + k - 6kl < 0$ for $k \geq 2$. However, Eq. (70) is only valid for $k > 4$. For $k = 2(4)$, $\rho_\phi \propto a^{-3}(a^{-4})$ while $\rho_R \propto a^{-4}$ for all k and reheating is not possible for $k < 6$. For these specific cases, we then have

$$T_{\rm RH} = \begin{cases} \text{no reheating} & k = 4\,, \\ \text{no reheating} & k = 2\,, \end{cases} \tag{71}$$

In this case, the presence of a bare mass will not change the lack of reheating through the scattering to vectors.

As a conclusion, whereas in the case of decays to fermions or bosons, the presence of a quadratic term only acts on the *value* of $T_{\rm RH}$, decreasing the reheating temperature in the former case, increasing it in the latter case, the quadratic term when dominant removes the possibility of reheating through scattering to scalars but reopens the possibility of reheating through decay to vectors, but does not allow reheating through the scattering to vectors.

V. GENERALIZED POTENTIALS

The inflationary potential may be dominated by higher order terms if $k > 4$. In this section, we generalize some of the arguments made above in the event that the inflationary potential is approximated by

$$\frac{1}{2}m_\phi^2\phi_0^2 + \lambda\phi_0^k M_P^{4-k} \tag{72}$$

about its minimum. In this case, the general expression for the scale factor when the mass term dominates is given by

$$\frac{a_{\rm m}}{a_{\rm end}} = \left(\frac{2\lambda^{\frac{2}{k}}M_P^{\frac{2(4-k)}{k}}\rho_{\rm end}^{\frac{k-2}{k}}}{m_\phi^2}\right)^{\frac{k+2}{6k-12}}\,, \tag{73}$$

with $\rho_{\rm end}$ given by Eq. (6) and λ by Eq. (8). Then

$$\rho_\phi(a_m) = 2\left(\frac{m_\phi^2}{2}\right)^{\frac{k}{k-2}}\lambda^{\frac{-2}{k-2}}M_P^{\frac{2(k-4)}{k-2}}\,, \tag{74}$$

which clearly reduces to Eq. (12) for $k = 4$. A parallel derivation leading to Eq. (16) implies that

$$\rho_{\rm RH} \lesssim 2 \left(\frac{m_\phi^2}{2}\right)^{\frac{k}{k-2}} \lambda^{\frac{-2}{k-2}} M_P^{\frac{2(k-4)}{k-2}} \tag{75}$$

for the mass term to dominate at reheating. In terms of the reheating temperature, this amounts to

$$T_{\rm RH} \lesssim \left(\frac{1}{\alpha}\right)^{\frac{1}{4}} \left(\frac{m_\phi M_P^{\frac{k-4}{k}}}{(2\lambda)^{\frac{1}{k}}}\right)^{\frac{k}{2(k-2)}} . \tag{76}$$

For comparison with Eq. (16), we have

$$T_{\rm RH} \lesssim \begin{cases} 5.0\times 10^5 \text{ GeV} \left(\frac{m_\phi}{\rm GeV}\right)^{\frac{3}{4}} & k = 6\,, \\ 6.3\times 10^6 \text{ GeV} \left(\frac{m_\phi}{\rm GeV}\right)^{\frac{2}{3}} & k = 8\,, \end{cases} \tag{77}$$

using $\lambda = 5.7\times 10^{-13}$ and 9.5×10^{-14} for $k = 6$ and 8, respectively.

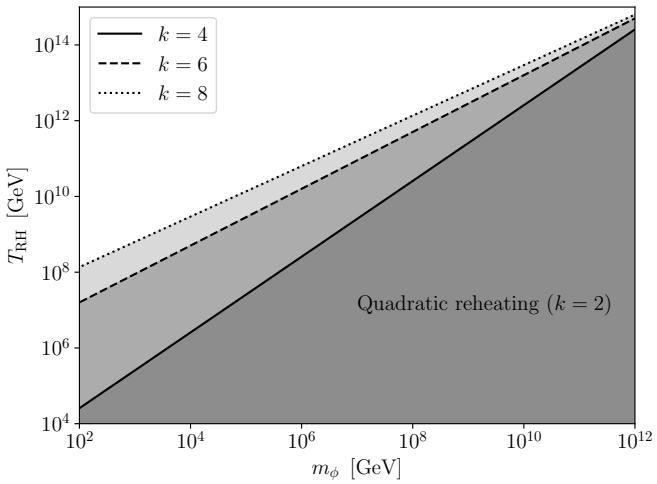


FIG. 8. *Minimal reheating temperature from Eq.(76), below which the inflaton mass term drives the process, as a function of the bare mass m_ϕ and for different values of k, $k = 4$ (solid line), $k = 6$ (dashed), $k = 8$ (dotted). In the different shaded regions, reheating occurs while the inflaton oscillates in a quadratic potential ($k = 2$), given by its bare mass m_ϕ. Above the lines, for different k, reheating occurs while the inflaton oscillates in the potential $V(\phi) \sim \phi^k$.*

VI. CONSEQUENCE ON THE INFLATON FRAGMENTATION

Recently, the authors of [14] and [15] have shown that fragmentation can significantly alter the reheating process. Indeed, the fragmentation of the inflaton condensate results in the population of an inflaton-particle bath, whose very low mass, proportional to the density of the condensate which remains unfragmented, may not allow reheating temperatures above the BBN bounds for inflaton decays to fermions. This fragmentation is due to the presence of a self-scattering term of type $\lambda\phi^k$, with $k \geq 4$. The inflaton condensate does not fragment in the absence of self-interactions allowing for reheating to occur as discussed above.

However, the study [14] was carried out in the context of a monomial potential of the type $V(\phi) = \lambda\phi^k$. It is then easy to see that the presence of a bare mass term of the type $\frac{1}{2}m_\phi^2\phi^2$ can change the conclusions of this study, in particular if the quadratic term begins to dominate *before* the fragmentation halts. If we define $a_{\rm F}$ as the value of the scale factor at the end of fragmentation, then $a_{\rm F}/a_{\rm end} = 180, 4.5\times 10^4, 6\times 10^6$ and 7×10^8, for $k = 4, 6, 8$ and 10 respectively [14]. In order for a quadratic term to affect the fragmentation process, we must have $a_{\rm m} \lesssim a_{\rm F}$ and using Eq.(10) it becomes easy to compute, for each value of k, the minimal value of m_ϕ necessary to ensure that the quadratic term dominates the potential *before* the end of fragmentation. The problem of a leftover bath of massless inflatons can then be avoided by stopping the fragmentation process.

More precisely, when reheating begins, self interactions can source the growth of the inflaton fluctuations $\delta\phi(t,\boldsymbol{x}) = \phi(t,\boldsymbol{x}) - \bar\phi(t)$, where $\bar\phi$ denotes the homogeneous condensate. At early times, this growth can be captured by the linear equation of motion

$$\ddot{\delta\phi} + 3H\dot{\delta\phi} - \frac{\nabla^2\delta\phi}{a^2} + V''(\bar\phi)\,\delta\phi \;=\; 0\,, \tag{78}$$

where

$$V''(\bar\phi) \;\simeq\; k(k-1)\lambda\bar\phi^{k-2}M_P^{4-k} + m_\phi^2\,. \tag{79}$$

For $m_\phi = 0$, the oscillating nature of this resulting effective mass term drives the resonant growth of $\delta\phi$ and the eventual fragmentation, $\delta\phi \gg \bar\phi$ [14, 15, 35–43]. However, if m_ϕ dominates before fragmentation, $V'' \sim$ const., strongly suppressing the oscillatory driving force.[6]

Fig. 9 shows the evolution of the total inflaton energy density ρ_ϕ, compared to the energy density in its fluctuations $\rho_{\delta\phi}$, as computed numerically for a T-model of inflation [13] with $k = 4$ and three choices of the bare mass (see [15] for details). The top panel depicts the zero bare mass scenario. In it, the rapid growth of inflaton fluctuations driven by parametric resonance can be appreciated. This growth only stops when $\rho_\phi \simeq \rho_{\delta\phi}$ ($a/a_{\rm end} \simeq 180$), corresponding to the near-complete fragmentation of the inflaton condensate in favor of free ϕ-particles.[7] For the

[6] For a purely quadratic inflaton potential the growth of fluctuations is still present, albeit not exponentially enhanced, due to the coupling of $\delta\phi$ with the fluctuations of the metric [44–46].

[7] The fragmentation of the inflaton condensate is not total even for $m_\phi = 0$. A small but nonvanishing homogeneous component $\bar\phi$ remains, and its presence can induce the decay of the free inflaton quanta $\delta\phi$ [14, 15].

bottom two panels we take $m_\phi > 0$. In both cases, the quartic → quadratic transition time has been chosen to be posterior to the complete fragmentation of the inflaton, $a_{\rm m} > a_{\rm F}$. A naive estimate from Eq. (79) would indicate that the resonant growth of $\delta\phi$ would not stop until

$$\frac{a}{a_{\rm end}} = \frac{\sqrt{12\lambda}\phi_{\rm end}}{m_\phi} \simeq 2.6\,\frac{a_{\rm m}}{a_{\rm end}}\,, \tag{80}$$

that is, the field would be fully fragmented before matter domination. However, the full numerical solution of the equation of motion (78) shows that the growth of fluctuations is in reality suppressed from $a \lesssim a_{\rm m}/2$, as both panels of Fig. 9 demonstrate. Therefore, reaching quadratic dominance is a sufficient condition to avert full fragmentation. Note that for smaller masses than those used in Fig. 9, fragmentation would nearly completely destroy the condensate and potentially disrupt the reheating process entirely. On the other hand, for larger masses, the fragmentation process would not be operative at all.

A qualitative depiction of this result for potentials with $k \geq 4$ is shown in Fig. 10, where we plot the limit on the mass m_ϕ above which the bare mass term dominates over $\lambda\phi^k$ in the potential as a function of k. We see that for larger value of k, where the fragmentation is less efficient due to the increasing difficulty for the self scattering to occurs for higher modes, even a small bare mass term can be sufficient to stop the fragmentation process and ensure a successful reheating.

VII. CONCLUSION

Reheating in most models of inflation is accomplished through either inflaton decay or scattering to Standard Model particles. This typically after inflationary expansion ends and a period of inflaton oscillations begins. When the potential is dominated by a quadratic term about its minimum, decays are necessary, as scatterings will not in general lead to a radiation dominated universe. However, potentials dominated by higher order interactions, $k > 2$, have anharmonic oscillations and scattering may lead to reheating, though these models may be subject to additional constraints arising from the fragmentation of the inflaton condensate. In addition, the details of the reheating process and the final reheating temperature depend on the spin of the final state particles produced in the decay or scattering.

In models of inflation for which the potential can be expanded about its minimum as $V(\phi) \propto \phi^k$, typically the lowest power, k appearing in the expansion dominates the reheating process. For $k > 2$, it is quite possible, as we have argued that in addition to the inflationary potential, a bare mass term in the full scalar potential is also present. This may arise from radiative corrections or supersymmetry breaking. In section III, we derived upper

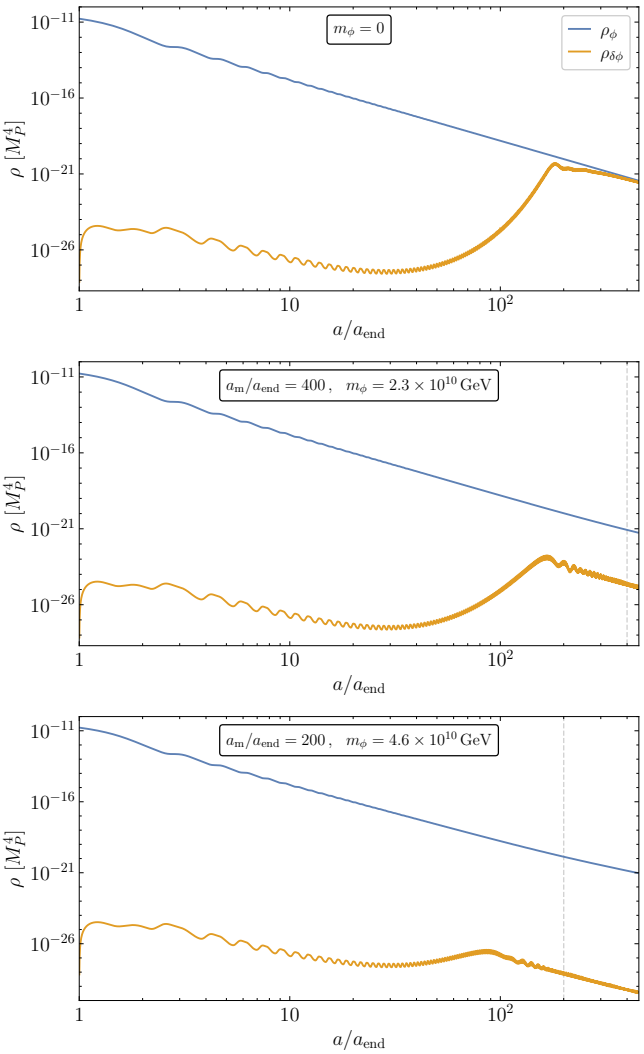


FIG. 9. *Energy density of the inflaton fluctuations $\rho_{\delta\phi}$ compared to the total energy density ρ_ϕ, for three values of the bare mass, for $k = 4$. The vertical dashed line corresponds to the value of $a_{\rm m}/a_{\rm end}$ when $m_\phi \neq 0$. In both of these cases, although $a_{\rm m} > a_{\rm F}$, the exponential growth of $\delta\phi$ is stopped by the transition to matter-domination.*

limits to this mass from CMB observables. These limits are sufficiently weak so that the presence of the mass term may affect the reheating process. Indeed, quite generally, if Eqs. (16) for $k = 4$ or (76) more generally are satified, the final reheating temperature will be determined by the quadratic rather than a higher order term.

The qualitative effect of the mass term also depends on the reheating mechanism (decays or scattering) as well as the spin of the final states. For decays to fermions, the reheating temperature is increased by the presence of mass term, while for scalars, it is decreased. For decays to vectors, reheating does not occur for $k = 4$ in the absence of a mass term and its presence allows for the possibility of reheating in this case. In contrast, if the mass term becomes important before the end of re-

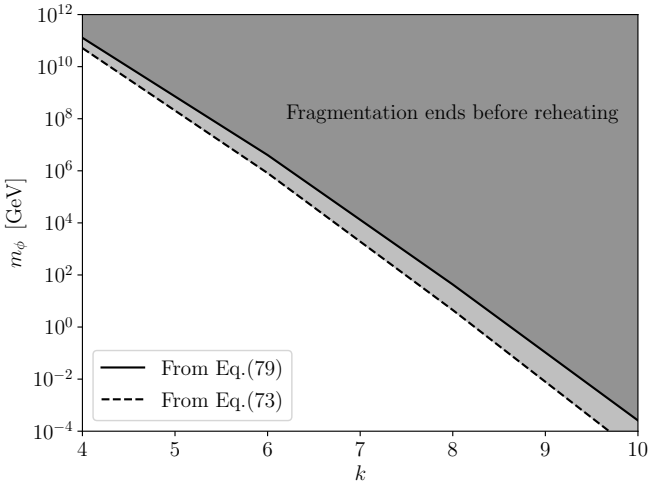


FIG. 10. *Region in the parameter space where the fragmentation happens after the domination by the bare mass term $\frac{1}{2}m_\phi^2\phi^2$ over $\lambda\phi^k$, allowing for a quadratic reheating. The dashed line is obtained from Eq.(73), while the solid line is obtained from Eq.(79).*

heating for scattering to scalars, the reheating process is halted. Furthermore, when reheating is accomplished through scattering to scalars with $k \geq 4$, the density of inflatons is quickly redshifts (as a^{-8}) until the mass term comes to dominate. In this case, the residual inflaton matter density acts as cold dark matter and a strong limit on the inflaton mass has been derived in Eq. (67). Finally we have seen that for scattering to vectors, reheating with $k = 4$ is not possible ($k \geq 6$ is required) and the mass term does not come to the rescue in this case.

Understanding the reheating process after inflation is of great importance as it is not only responsible for providing an early period of radiation domination necessary for big bang nucleosynthesis, but may be the source of dark matter. Thermal production of dark matter in equilibrium remains an important mechanism, however, it is well established that non-equilibrium process just as freeze-in [47] may also be the ultimate source of dark matter in the Universe. For these cases, a detailed understanding of reheating is essential and here we examined the role of a bare mass term for the inflaton in models where the inflationary dynamics are governed by higher order interactions.

ACKNOWLEDGMENTS

The authors thank Mathieu Gross and Jong-Hyun Yoon for extremely valuable discussions during the completion of our work. This project has received support from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 860881-HIDDeN, and the IN2P3 Master Projet UCMN. The work of M.A.G.G. was supported by the DGAPA-PAPIIT grant IA103123 at UNAM, the CONAHCYT "Ciencia de Frontera" grant CF-2023-I-17, and the Programa de Investigación del Instituto de Física 2023 (PIIF23). The work of K.A.O. was supported in part by DOE grant DE-SC0011842 at the University of Minnesota. The authors acknowledge the support of the Institut Pascal at Université Paris-Saclay during the Paris-Saclay Astroparticle Symposium 2023, with the support of the P2IO Laboratory of Excellence (program "Investissements d'avenir" ANR-11-IDEX-0003-01 Paris-Saclay and ANR-10-LABX-0038) and the IN2P3 master project UCMN.

APPENDIX

The decay (or scattering) rate of the inflaton, averaged over several oscillations, can be neatly expressed as [10, 12]

$$\Gamma_\phi(t) \;=\; \gamma_\phi \left(\frac{\rho_\phi}{M_P^4}\right)^l , \tag{81}$$

where

$$\gamma_\phi \;=\; \begin{cases} \sqrt{k(k-1)}\lambda^{1/k} M_P \dfrac{y_{\rm eff}^2}{8\pi} , & \phi \to \bar{f} f , \\ \dfrac{\mu_{\rm eff}^2}{8\pi\sqrt{k(k-1)}\lambda^{1/k} M_P} , & \phi \to bb , \\ \alpha^2 \left[k(k-1)\right]^{\frac{3}{2}} \lambda^{\frac{3}{k}} M_P , & \phi \to AA , \\ \dfrac{\sigma_{\rm eff}^2 M_P}{8\pi[k(k-1)]^{3/2}\lambda^{3/k}} , & \phi\phi \to bb , \\ \beta^2 \left[k(k-1)\right]^{\frac{1}{2}} \lambda^{\frac{1}{k}} M_P , & \phi\phi \to AA , \end{cases} \tag{82}$$

and

$$l \;=\; \begin{cases} \frac{1}{2} - \frac{1}{k} , & \phi \to \bar{f} f , \\ \frac{1}{k} - \frac{1}{2} , & \phi \to bb , \\ \frac{3}{2} - \frac{3}{k} , & \phi \to AA , \\ \frac{3}{k} - \frac{1}{2} , & \phi\phi \to bb , \\ \frac{3}{2} - \frac{1}{k} , & \phi\phi \to AA . \end{cases} \tag{83}$$

So long as $\gamma_\phi \ll H$, the Friedmann equation for ρ_ϕ (3) can be integrated to give

$$\rho_\phi(a) = \rho_{\rm end} \left(\frac{a}{a_{\rm end}}\right)^{-\frac{6k}{k+2}} , \tag{84}$$

which sources the Boltzmann equation (14). This can be rewritten as

$$\frac{1}{a^4}\frac{d}{da}\left(\rho_R a^4\right) \;=\; \frac{2k}{k+2}\frac{\gamma_\phi}{aH}\frac{\rho_\phi^{l+1}}{M_P^{4l}} , \tag{85}$$

and can be integrated to give

$$\rho_R = \frac{2k}{k+8-6kl}\frac{\gamma_\phi}{H_{\rm end}}\frac{\rho_{\rm end}^{l+1}}{M_P^{4l}}\left(\frac{a_{\rm end}}{a}\right)^4\left[\left(\frac{a}{a_{\rm end}}\right)^{\frac{k+8-6kl}{k+2}}-1\right], \tag{86}$$

where $H_{\rm end}^2 = \rho_{\rm end}/3M_P^2$. At later times when $a \gg a_{\rm end}$ and $8+k-6kl>0$, we can approximate ρ_R as

$$\rho_R^{a\gg a_{\rm end}} = \frac{2k}{k+8-6kl}\frac{\gamma_\phi}{H_{\rm end}}\frac{\rho_{\rm end}^{l+1}}{M_P^{4l}}\left(\frac{a_{\rm end}}{a}\right)^{\frac{3k+6kl}{k+2}}. \tag{87}$$

If $8+k-6kl<0$,

$$\rho_R^{a\gg a_{\rm end}} = \frac{2k}{6kl-k-8}\frac{\gamma_\phi}{H_{\rm end}}\frac{\rho_{\rm end}^{l+1}}{M_P^{4l}}\left(\frac{a_{\rm end}}{a}\right)^4, \tag{88}$$

which implies that the temperature would simply redshift as $T\propto a^{-1}$.

Finally, when $\rho_R(T_{\rm RH}) = \rho_\phi(T_{\rm RH})$, we obtain [10]

$$\frac{a_{\rm RH}}{a_{\rm end}} = \left[\frac{k+8-6kl}{2k}\frac{M_P^{4l-1}\rho_{\rm end}^{\frac{1}{2}-l}}{\sqrt{3}\gamma_\phi}\right]^{\frac{k+2}{3k-6kl}}, \tag{89}$$

for $8+k-6kl>0$ and for $8+k-6kl<0$,

$$\frac{a_{\rm RH}}{a_{\rm end}} = \left[\frac{6kl-k-8}{2k}\frac{M_P^{4l-1}\rho_{\rm end}^{\frac{1}{2}-l}}{\sqrt{3}\gamma_\phi}\right]^{\frac{k+2}{2k-8}}. \tag{90}$$

Note that Eq. (90) is only true for $k>4$. When $k\leq 4$ and $8+k-6kl<0$, reheating never occurs.

Evaluating ρ_R at $a=a_{\rm RH}$ gives

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{2k}{k+8-6kl}\frac{\sqrt{3}\gamma_\phi}{M_P^{4l-1}}\right]^{\frac{1}{2-4l}}, \tag{91}$$

for $8+k-6kl>0$, and

$$T_{\rm RH} = \left(\frac{1}{\alpha}\right)^{\frac{1}{4}}\left[\frac{2k}{6kl-k-8}\frac{\sqrt{3}\gamma_\phi}{M_P^{4l-1}}\rho_{\rm end}^{\frac{6kl-k-8}{6k}}\right]^{\frac{3k}{4k-16}}. \tag{92}$$

for $8+k-6kl<0$ and $k>4$

[1] K. A. Olive, Phys. Rept. **190** (1990) 307; A. D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, Switzerland, 1990); D. H. Lyth and A. Riotto, *Phys. Rep.* **314** (1999) 1 [arXiv:hep-ph/9807278]; J. Martin, C. Ringeval and V. Vennin, Phys. Dark Univ. **5-6**, 75-235 (2014) [arXiv:1303.3787 [astro-ph.CO]]; J. Martin, C. Ringeval, R. Trotta and V. Vennin, JCAP **1403** (2014) 039 [arXiv:1312.3529 [astro-ph.CO]]; J. Martin, Astrophys. Space Sci. Proc. **45**, 41 (2016) [arXiv:1502.05733 [astro-ph.CO]].

[2] A. D. Dolgov and A. D. Linde, Phys. Lett. **116B**, 329 (1982); L. F. Abbott, E. Farhi and M. B. Wise, Phys. Lett. **117B**, 29 (1982).

[3] D. V. Nanopoulos, K. A. Olive and M. Srednicki, Phys. Lett. B **127**, 30 (1983).

[4] N. Aghanim *et al.* [Planck], Astron. Astrophys. **641**, A6 (2020) [arXiv:1807.06209 [astro-ph.CO]].

[5] G. F. Giudice, E. W. Kolb and A. Riotto, Phys. Rev. D **64** (2001) 023508 [hep-ph/0005123]; D. J. H. Chung, E. W. Kolb and A. Riotto, Phys. Rev. D **60** (1999) 063504 [hep-ph/9809453].

[6] K. Mukaida and M. Yamada, JCAP **02**, 003 (2016) [arXiv:1506.07661 [hep-ph]].

[7] M. A. G. Garcia, Y. Mambrini, K. A. Olive and M. Peloso, Phys. Rev. D **96**, no.10, 103510 (2017) [arXiv:1709.01549 [hep-ph]].

[8] N. Bernal, JCAP **10**, 006 (2020) [arXiv:2005.08988 [hep-ph]]; A. Di Marco and G. Pradisi, Int. J. Mod. Phys. A **36**, no.15, 2150095 (2021) [arXiv:2102.00326 [gr-qc]].

[9] M. A. G. Garcia, K. Kaneta, Y. Mambrini and K. A. Olive, Phys. Rev. D **101** (2020) no.12, 123507 [arXiv:2004.08404 [hep-ph].

[10] M. A. G. Garcia, K. Kaneta, Y. Mambrini and K. A. Olive, JCAP **04**, 012 (2021) [arXiv:2012.10756 [hep-ph]].

[11] A. A. Starobinsky, Phys. Lett. B **91**, 99 (1980).

[12] M. A. G. Garcia, K. Kaneta, W. Ke, Y. Mambrini, K. A. Olive and S. Verner, [arXiv:2311.14794 [hep-ph]].

[13] R. Kallosh and A. Linde, JCAP **07**, 002 (2013) [arXiv:1306.5220 [hep-th]].

[14] M. A. G. Garcia, M. Gross, Y. Mambrini, K. A. Olive, M. Pierre and J. H. Yoon, JCAP **12**, 028 (2023) [arXiv:2308.16231 [hep-ph]].

[15] M. A. G. Garcia and M. Pierre, JCAP **11** (2023), 004 [arXiv:2306.08038 [hep-ph]].

[16] P. A. R. Ade *et al.* [BICEP and Keck Collaborations], Phys. Rev. Lett. **127**, no.15, 151301 (2021) [arXiv:2110.00483 [astro-ph.CO]].

[17] M. Tristram, A. J. Banday, K. M. Górski, R. Keskitalo, C. R. Lawrence, K. J. Andersen, R. B. Barreiro, J. Borrill, L. P. L. Colombo and H. K. Eriksen, *et al.* Phys. Rev. D **105**, no.8, 083524 (2022) [arXiv:2112.07961 [astro-ph.CO]].

[18] A. R. Liddle and S. M. Leach, Phys. Rev. D **68**, 103503 (2003) [astro-ph/0305263];

[19] J. Martin and C. Ringeval, Phys. Rev. D **82**, 023511 (2010) [arXiv:1004.5525 [astro-ph.CO]].

[20] D. J. Fixsen, Astrophys. J. **707** (2009), 916-920

[arXiv:0911.1955 [astro-ph.CO]].
[21] J. Ellis, M. A. G. Garcia, D. V. Nanopoulos, K. A. Olive and S. Verner, Phys. Rev. D **105** (2022) no.4, 043504 [arXiv:2112.04466 [hep-ph]].
[22] Y. Shtanov, J. H. Traschen and R. H. Brandenberger, Phys. Rev. D **51** (1995), 5438-5455 [arXiv:hep-ph/9407247 [hep-ph]].
[23] K. Ichikawa, T. Suyama, T. Takahashi and M. Yamaguchi, Phys. Rev. D **78** (2008), 063545 [arXiv:0807.3988 [astro-ph]].
[24] M. A. G. Garcia, K. Kaneta, Y. Mambrini, K. A. Olive and S. Verner, JCAP **03**, no.03, 016 (2022) doi:10.1088/1475-7516/2022/03/016 [arXiv:2109.13280 [hep-ph]].
[25] J. Ellis, M. A. G. Garcia, N. Nagata, N. D. V., K. A. Olive and S. Verner, Int. J. Mod. Phys. D **29**, no.16, 2030011 (2020) [arXiv:2009.01709 [hep-ph]].
[26] J. Ellis, M. A. G. Garcia, D. V. Nanopoulos and K. A. Olive, JCAP **10**, 003 (2015) [arXiv:1503.08867 [hep-ph]].
[27] J. Ellis, D. V. Nanopoulos, K. A. Olive and S. Verner, JCAP **08**, 037 (2020) [arXiv:2004.00643 [hep-ph]].
[28] M. Endo, K. Kadota, K. A. Olive, F. Takahashi and T. T. Yanagida, JCAP **02**, 018 (2007) [arXiv:hep-ph/0612263 [hep-ph]].
[29] M. A. G. Garcia, Y. Mambrini, K. A. Olive and S. Verner, JCAP **10**, 061 (2021) [arXiv:2107.07472 [hep-ph]].
[30] A. R. Liddle and L. A. Ureña-López, Phys. Rev. Lett. **97**, 161301 (2006) [arXiv:astro-ph/0605205 [astro-ph]].
[31] A. de la Macorra, Astropart. Phys. **35**, 478-484 (2012) [arXiv:1201.6302 [astro-ph.CO]].
[32] K. Mukaida and K. Nakayama, JCAP **08**, 062 (2014) [arXiv:1404.1880 [hep-ph]].
[33] L. Ji and M. Kamionkowski, Phys. Rev. D **100**, no.8, 083519 (2019) [arXiv:1905.05770 [astro-ph.CO]].
[34] O. Lebedev and J. H. Yoon, Phys. Lett. B **821**, 136614 (2021) [arXiv:2105.05860 [hep-ph]]; O. Lebedev, T. Nerdi, T. Solomko and J. H. Yoon, Phys. Rev. D **106**, no.4, 043537 (2022) [arXiv:2110.05488 [hep-ph]].
[35] A. V. Frolov, Class. Quant. Grav. **27** (2010), 124006 [arXiv:1004.3559 [gr-qc]].
[36] P. B. Greene, L. Kofman, A. D. Linde and A. A. Starobinsky, Phys. Rev. D **56** (1997), 6175-6192 [arXiv:hep-ph/9705347 [hep-ph]].
[37] D. I. Kaiser, Phys. Rev. D **56** (1997), 706-716 [arXiv:hep-ph/9702244 [hep-ph]].
[38] J. Garcia-Bellido, D. G. Figueroa and J. Rubio, Phys. Rev. D **79** (2009), 063531 [arXiv:0812.4624 [hep-ph]].
[39] M. A. Amin, R. Easther, H. Finkel, R. Flauger and M. P. Hertzberg, Phys. Rev. Lett. **108** (2012), 241302 [arXiv:1106.3335 [astro-ph.CO]].
[40] M. P. Hertzberg, J. Karouby, W. G. Spitzer, J. C. Becerra and L. Li, Phys. Rev. D **90** (2014), 123528 [arXiv:1408.1396 [hep-th]].
[41] K. D. Lozanov and M. A. Amin, Phys. Rev. Lett. **119** (2017) no.6, 061301 [arXiv:1608.01213 [astro-ph.CO]].
[42] K. D. Lozanov and M. A. Amin, Phys. Rev. D **97** (2018) no.2, 023533 [arXiv:1710.06851 [astro-ph.CO]].
[43] D. G. Figueroa and F. Torrenti, JCAP **02** (2017), 001 [arXiv:1609.05197 [astro-ph.CO]].
[44] F. Finelli and R. H. Brandenberger, Phys. Rev. Lett. **82** (1999), 1362-1365 [arXiv:hep-ph/9809490 [hep-ph]].
[45] K. Jedamzik, M. Lemoine and J. Martin, JCAP **09** (2010), 034 [arXiv:1002.3039 [astro-ph.CO]].
[46] J. Martin, T. Papanikolaou and V. Vennin, JCAP **01** (2020), 024 [arXiv:1907.04236 [astro-ph.CO]].
[47] L. J. Hall, K. Jedamzik, J. March-Russell and S. M. West, JHEP **1003** (2010) 080 [arXiv:0911.1120 [hep-ph]]; X. Chu, T. Hambye and M. H. G. Tytgat, JCAP **1205** (2012) 034 [arXiv:1112.0493 [hep-ph]]; A. Biswas, D. Borah and A. Dasgupta, Phys. Rev. D **99**, no.1, 015033 (2019) [arXiv:1805.06903 [hep-ph]].

Exploring the sensitivity to non-standard and generalized neutrino interactions through coherent elastic neutrino-nucleus scattering with a NaI detector

Sabya Sachi Chatterjee,[1,2,*] Stéphane Lavignac,[2,†] O. G. Miranda,[3,‡] and G. Sanchez Garcia[4,§]

[1] *Institut für Astroteilchenphysik, Karlsruher Institut für Technologie (KIT), 76131 Karlsruhe, Germany*

[2] *Institut de Physique Théorique, Université Paris Saclay, CNRS, CEA, F-91191 Gif-sur-Yvette, France*

[3] *Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740 07000, Ciudad de Mexico, Mexico*

[4] *Instituto de Física Corpuscular (CSIC-Universitat de València), Parc Científic UV, C/ Catedràtico José Beltrán, 2, 46980 Paterna, Spain*

arXiv:2402.16953v1 [hep-ph] 26 Feb 2024

After the first observation of coherent elastic neutrino-nucleus scattering (CEνNS) by the COHERENT collaboration, many efforts are being made to improve the measurement of this process, making it possible to constrain new physics in the neutrino sector. In this paper, we study the sensitivity to non-standard interactions (NSIs) and generalized neutrino interactions (GNIs) of a NaI detector with characteristics similar to the one that is currently being deployed at the Spallation Neutron Source at Oak Ridge National Laboratory. We show that such a detector, whose target nuclei have significantly different proton to neutron ratios (at variance with the current CsI detector), could help to partially break the parameter degeneracies arising from the interference between the Standard Model and NSI contributions to the CEνNS cross section, as well as between different NSI parameters. By contrast, only a slight improvement over the current CsI constraints is expected for parameters that do not interfere with the SM contribution. We find that a significant reduction of the background level would make the NaI detector considered in this paper very efficient at breaking degeneracies among NSI parameters.

I. INTRODUCTION

Coherent elastic neutrino-nucleus scattering (CEνNS) [1] is a privileged process to probe new physics in the neutrino sector. So far, the only measurements of CEνNS have been done at the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory, USA, by the COHERENT collaboration. The first observation of this process was performed with a CsI detector [2, 3]. Another detector, with liquid Argon as a target, was subsequently used by the COHERENT collaboration [4].

* sabya.chatterjee@kit.edu
† stephane.lavignac@ipht.fr
‡ omar.miranda@cinvestav.mx
§ gsanchez@fis.cinvestav.mx

These measurements allowed to test the Standard Model at low energy [5], to constrain the neutron root mean square *(rms)* radius of Cesium, Iodine [6] and Argon [7], and to probe various possible manifestations of physics beyond the Standard Model, such as non-standard neutrino interactions (NSIs) [8–11], generalized neutrino interactions (GNIs) [10, 12–14], neutrino electromagnetic properties [15–17], new light gauge bosons [13, 18–20] and sterile neutrinos [10, 21, 22], among others. Future CEνNS experiments at spallation neutron sources [23–25] or nuclear reactors [26–29] will provide more data using different target materials. Particularly interesting is the possibility to disentangle new physics from Standard Model parameters using the interplay between experiments with different neutrino sources [30, 31]. In addition, using nuclear targets with different proportions of protons and neutrons can help reduce the parameter degeneracies arising from the interference between standard and non-standard contributions to CEνNS, e.g., in the presence of NSIs [8].

A broad class of new physics models predict low-energy effects in the lepton sector that can effectively be described in terms of NSIs or GNIs [32–35]. CEνNS is an excellent process to test such new physics effects and to resolve the degeneracies that appear in the interpretation of neutrino oscillation data in the presence of NSIs [36, 37]. However, the possibility of constraining NSI parameters with CEνNS is limited by the possible cancellations between standard and non-standard contributions to the cross section, as well as between different NSI couplings. It is, therefore, crucial to perform CEνNS measurements with different neutrino sources and targets in order to break these degeneracies [8, 9, 24, 37–39].

In this work, we study the expected sensitivity to non-standard and generalized neutrino interactions of a NaI detector with characteristics similar to the one that is currently being deployed at the SNS. NaI, as opposed to CsI, has the important feature of consisting of two target nuclei with significantly different proton-to-neutron ratios ($p/n \simeq 0.72$ for Iodine and $p/n \simeq 0.92$ for Sodium). We show that this property of the NaI detector makes it particularly efficient at reducing the degeneracies between NSI parameters, provided that the signal-over-background ratio is sufficiently large. For generalized neutrino interactions and off-diagonal NSIs, no significant improvement over the constraints extracted from the current CsI data is expected, unless the signal systematic uncertainty can be reduced below 10%.

The paper is structured as follows. In Section II, we discuss the NSI and GNI contributions to the CEνNS cross section. In Section III, we present the experimental setup at the SNS and the characteristics of the current CsI and future NaI detectors. Section IV details our analysis procedure. The expected sensivities of the future NaI detector to NSI and GNI parameters are presented in Section V, and compared with the current CsI constraints. The impact of the statistics, signal

systematic uncertainty and background level are also studied in this section. Finally, we present our conclusions in Section VI.

II. THEORETICAL FRAMEWORK

In this section, we discuss the tools needed to probe GNIs through CEνNS measurements. We begin by introducing the SM prediction for the CEνNS cross section, then we discuss how this cross section is modified when GNI parameters are introduced.

A. The standard CEνNS cross section

In the Standard Model, the cross section for the coherent elastic scattering of a neutrino or antineutrino of flavor α ($\alpha = e, \mu, \tau$) and energy E_ν off a nucleus with Z protons, N neutrons and mass M is given (up to subleading terms of order T/E_ν, $(T/E_\nu)^2$ [8] and small radiative corrections [40]) by [1]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}T}(E_\nu, T) \;=\; \frac{G_F^2 M}{\pi}\left(1 - \frac{MT}{2E_\nu^2}\right) F^2(|\vec{q}|^2)\, Q_{W,\alpha}^2 \;, \tag{1}$$

where G_F is the Fermi constant, T the nuclear recoil energy, $F(|\vec{q}|^2)$ the form factor of the nucleon distribution in the nucleus [41] (assumed to be the same for protons and neutrons) evaluated at the transferred three-momentum $|\vec{q}| \simeq \sqrt{2MT}$, and

$$Q_{W,\alpha} \;=\; Z g_V^p + N g_V^n \,, \tag{2}$$

with $g_V^p = \frac{1}{2} - 2\sin^2\theta_W$ and $g_V^n = -\frac{1}{2}$, is the weak nuclear charge. Since $|\vec{q}|$ is typically very small in CEνNS, the value of the weak mixing angle is taken at zero momentum transfer ($\sin^2\theta_W = 0.23867$ [42, 43]).

The form factor $F(|\vec{q}|^2)$ describe the loss of coherence of the scattering when $|\vec{q}| \gtrsim R^{-1}$, where R is the nucleon radius. Following the COHERENT collaboration, we use the Klein-Nystrand parametrization for the nuclear form factor [44],

$$F(|\vec{q}|^2) \;=\; 3\,\frac{j_1(|\vec{q}|R_A)}{|\vec{q}|R_A}\left(\frac{1}{1 + |\vec{q}|^2 a_k^2}\right) \,, \tag{3}$$

where $j_1(x) = (\sin x - x\cos x)/x^2$ is the spherical Bessel function of order one, $a_k = 0.9\,\mathrm{fm}$ and $R_A = 1.23A^{1/3}\,\mathrm{fm}$ is the nuclear radius, with A the total number of nucleons of the target nucleus.

B. Non-standard and generalized neutrino interactions

The CEνNS cross section is potentially affected by any interaction between neutrinos and quarks that may be generated by new physics beyond the Standard Model. The simplest example of such interactions are the so-called (neutral-current) non-standard neutrino interactions (NC-NSIs, or NSIs for short), whose phenomenology has been widely studied [33–35, 45]. They are usually parametrized as

$$\mathcal{L}_{\text{NC-NSI}} \;=\; -\sqrt{2}\,G_F \left\{ \varepsilon^{fV}_{\alpha\beta} \left(\overline{\nu}_\alpha \gamma^\mu P_L \nu_\beta\right) \left(\overline{f} \gamma_\mu f\right) + \varepsilon^{fA}_{\alpha\beta} \left(\overline{\nu}_\alpha \gamma^\mu P_L \nu_\beta\right) \left(\overline{f} \gamma_\mu \gamma^5 f\right) \right\}, \tag{4}$$

where $\alpha, \beta = e, \mu, \tau$ denote the neutrino flavors, $f = u, d$ label the up and down quarks (we omit the leptonic NSIs, $f = e$, as they do not contribute to CEνNS), summation over α, β and f is implicit, and $\varepsilon^{fV}_{\alpha\beta}$, $\varepsilon^{fA}_{\alpha\beta}$ are the vector and axial-vector NSI couplings, respectively. In this paper, we do not consider the axial-vector NSIs, because their contribution to the CEνNS cross section depends on the nuclear spin and is therefore negligible for heavy nuclei [12], similarly to the contribution of the neutral current axial couplings in the SM [8, 46].

Non-standard neutrino-quark interactions can have a more general Lorentz structure than the one of Eq. (4). Namely, one may also consider scalar, pseudo-scalar and tensor couplings[1], usually dubbed "generalized neutrino interactions" (GNIs) in the literature [12, 32, 48]

$$\begin{aligned} \mathcal{L}_{\text{GNI}} \;=\; -\sqrt{2}\,G_F \quad & \left\{ \varepsilon^{qS}_{\alpha\beta} \left(\overline{\nu}_\alpha P_L \nu_\beta\right) \left(\overline{q} q\right) + \varepsilon^{qP}_{\alpha\beta} \left(\overline{\nu}_\alpha P_L \nu_\beta\right) \left(\overline{q} \gamma^5 q\right) \right. \\ & \left. + \varepsilon^{qT}_{\alpha\beta} \left(\overline{\nu}_\alpha \sigma^{\mu\nu} P_L \nu_\beta\right) \left(\overline{q} \sigma_{\mu\nu} P_L q\right) \right\} + \text{h.c.}\,, \end{aligned} \tag{5}$$

where $\sigma^{\mu\nu} = \frac{i}{2}\left[\gamma^\mu, \gamma^\nu\right]$ and $q = u, d$. Generally, NSIs are also included in the GNI Lagrangian. In the following, we will refer to $\varepsilon^{qS}_{\alpha\beta}$, $\varepsilon^{qV}_{\alpha\beta}$ and $\varepsilon^{qT}_{\alpha\beta}$ as scalar, vector and tensor couplings, respectively (we will not consider pseudo-scalar interactions, which give a negligible contribution to CEνNS [12, 32]). Note that hermiticity of the Lagrangian implies $\varepsilon^{qV}_{\beta\alpha} = \left(\varepsilon^{qV}_{\alpha\beta}\right)^*$, while the matrices of scalar and tensor couplings do not have any particular symmetry property if neutrinos are Dirac fermions. If they are Majorana fermions, instead, one has $\varepsilon^{qS}_{\beta\alpha} = \varepsilon^{qS}_{\alpha\beta}$ and $\varepsilon^{qT}_{\beta\alpha} = -\varepsilon^{qT}_{\alpha\beta}$ (in particular, flavor-diagonal tensor couplings vanish for Majorana neutrinos).

In the presence of scalar, vector and tensor non-standard neutrino interactions, the differential CEνNS cross section (1) for an incident neutrino of flavor α is modified to [32]

$$\begin{aligned} \frac{\mathrm{d}\sigma}{\mathrm{d}T}(E_\nu, T) \;=\; \frac{G_F^2 M}{\pi}\, F^2(|\vec{q}|^2) \sum_{\beta\,=\,e,\,\mu,\,\tau} & \left[\left|C^S_{\beta\alpha}\right|^2 \frac{MT}{8E_\nu^2} + \left|C^V_{\beta\alpha} + Q_{W,\alpha}\,\delta_{\alpha\beta})\right|^2 \left(1 - \frac{MT}{2E_\nu^2}\right) \right. \\ & \left. + 2\left|C^T_{\beta\alpha}\right|^2 \left(1 - \frac{MT}{4E_\nu^2}\right) \right], \end{aligned} \tag{6}$$

[1] The tensor operators $\left(\overline{\nu}_\alpha \sigma^{\mu\nu} P_L \nu_\beta\right)\left(\overline{q} \sigma_{\mu\nu} P_R q\right)$ vanish by Lorentz symmetry [47], hence do not appear in Eq. (5).

where subleading terms of order $T/E_\nu \ll MT/E_\nu^2$ have been omitted[2]. For an incident antineutrino of flavor α, one should replace $C^S_{\beta\alpha} \to C^S_{\alpha\beta}$, $C^V_{\beta\alpha} \to C^V_{\alpha\beta}$ and $C^T_{\beta\alpha} \to C^T_{\alpha\beta}$ in Eq. (6). The SM contribution is contained in the weak nuclear charge $Q_{W,\alpha}$, given by Eq. (2), while the non-standar contributions are encapsulated in the coefficients $C^S_{\alpha\beta}$, $C^V_{\alpha\beta}$ and $C^T_{\alpha\beta}$, which are given by [12]

$$C^S_{\alpha\beta} = \sum_{q=u,d} \left(Z \frac{m_p}{m_q} f_q^p + N \frac{m_n}{m_q} f_q^n \right) \varepsilon^{qS}_{\alpha\beta} \,, \tag{7}$$

$$C^V_{\alpha\beta} = (2Z + N)\, \varepsilon^{uV}_{\alpha\beta} + (Z + 2N)\, \varepsilon^{dV}_{\alpha\beta} \,, \tag{8}$$

$$C^T_{\alpha\beta} = \sum_{q=u,d} \left(Z\delta_q^p + N\delta_q^n \right) \varepsilon^{qT}_{\alpha\beta} \,, \tag{9}$$

For the numerical coefficients appearing in Eqs. (7) and (9), we adopt the values given in Refs. [49] and [50], respectively:

$$f_u^p = 0.0208 \,, \quad f_d^p = 0.0411 \,, \quad f_u^n = 0.0189 \,, \quad f_d^n = 0.0451 \,, \tag{10}$$

$$\delta_u^p = 0.792 \,, \quad \delta_d^p = -0.194 \,, \quad \delta_u^n = -0.194 \,, \quad \delta_d^n = \ 0.792 \,. \tag{11}$$

As for quark masses, we use the central values given by the Particle Data Group [43], $m_u = 2.2\,\text{MeV}$ and $m_d = 4.7\,\text{MeV}$. Given the large uncertainties associated with the light quark masses and with f_q^N and δ_q^N, many authors provide constraints on the coefficients $C^S_{\alpha\beta}$ and $C^T_{\alpha\beta}$ rather than on the Lagragian parameters $\varepsilon^{qS}_{\alpha\beta}$ and $\varepsilon^{qT}_{\alpha\beta}$. In this paper, we choose instead to constrain the Lagragian parameters for fixed values of m_q, f_q^N and δ_q^N. It is straightforward to translate the current bounds and future sensitivity estimates on $\varepsilon^{qS}_{\alpha\beta}$ and $\varepsilon^{qT}_{\alpha\beta}$ presented in the next sections into constraints on $C^S_{\alpha\beta}$ and $C^T_{\alpha\beta}$.

Other authors, e.g. in Ref. [10, 12, 32, 51, 52], use the symbols C_S^q, C_V^q, C_T^q instead of ε^{qS}, ε^{qV}, ε^{qT} in Eqs. (7) to (9), with depending on the normalization convention either $C_V^q = \varepsilon^{qV}$ [52] (with the CEνNS cross section given by Eq. (6)) or $C_V^q = 2\varepsilon^{qV}$ [10, 12, 32, 51] (in which case one should replace $C^V_{\alpha\beta}$ by $\frac{1}{2}\, C^V_{\alpha\beta}$ in Eq. (6)).

III. EXPERIMENTAL SETUP AND DETECTORS

In this section, we describe the SNS neutrino beam and the main characteristics of the two detectors considered in this paper, namely the current CsI detector used by the COHERENT collaboration and a future NaI detector similar to the one under deployment by the same collaboration [53].

[2] In particular, we omit a scalar-tensor interference term in Eq. (6), which is proportional to T/E_ν and whose sign is different for neutrinos and antineutrinos [32].

A. Neutrino beam and predicted number of CEνNS events

At the SNS, the neutrinos used for CEνNS measurements arise from the decays at rest of the π^+ produced, together with neutrons and π^-'s, from the collision of high-energy protons on a mercury target (in which the π^- are absorbed before decaying). The SNS neutrino beam therefore consists of three components: a prompt, mono-energetic ν_μ component produced in the two-body decays of stopped positively charged pions ($\pi^+ \to \mu^+ + \nu_\mu$), and two delayed $\bar{\nu}_\mu$ and ν_e components arising from the subsequent decays at rest of the antimuons ($\mu^+ \to e^+ + \bar{\nu}_\mu + \nu_e$). These contributions can be analytically computed and are given by

$$\frac{dN_{\nu_\mu}}{dE_\nu} = \xi\,\delta\left(E_\nu - \frac{m_\pi^2 - m_\mu^2}{2m_\pi}\right) , \tag{12}$$

$$\frac{dN_{\bar{\nu}_\mu}}{dE_\nu} = \xi\,\frac{64E_\nu^2}{m_\mu^3}\left(\frac{3}{4} - \frac{E_\nu}{m_\mu}\right) , \tag{13}$$

$$\frac{dN_{\nu_e}}{dE_\nu} = \xi\,\frac{192E_\nu^2}{m_\mu^3}\left(\frac{1}{2} - \frac{E_\nu}{m_\mu}\right) , \tag{14}$$

where E_ν is the neutrino energy and $\xi = rN_{\rm POT}/4\pi L^2$ is a normalization factor that depends on several experimental features: $N_{\rm POT}$, the number of protons on target (POT) throughout the operation time; r, the number of neutrinos per flavor produced by each POT; and L, the distance between the source and the detector.

Given the neutrino flux, the predicted number of CEνNS events in the nuclear recoil energy bin $[T_i, T_{i+1}]$ is given by, for a target material consisting of a single nucleus with mass M,

$$N_i^{th} = \mathcal{N}\int_{T_i}^{T_{i+1}} A(T)\,dT \int_0^{T'_{\rm max}} \mathcal{R}(T,T')\,dT' \sum_{\nu=\nu_e,\nu_\mu,\bar{\nu}_\mu} \int_{E_{\rm min}(T')}^{E_{\rm max}} dE_\nu\ \frac{{\rm d}N_\nu}{{\rm d}E_\nu}(E_\nu)\,\frac{{\rm d}\sigma}{{\rm d}T'}(E_\nu,T') , \tag{15}$$

where ${\rm d}\sigma/{\rm d}T$ is the relevant CEνNS cross section, given by Eq. (1) or (6), and $\mathcal{N} = N_A M_{\rm det}/M_{\rm mol}$ is the number of nuclei in the detector, which is computed from the detector mass $M_{\rm det}$ and the molar mass of the nucleus $M_{\rm mol}$, with N_A the Avogadro constant. The lower limit of the integral over E_ν is the minimal neutrino energy that can induce a nuclear recoil of energy T', $E_{\rm min}(T') = \sqrt{MT'/2}$, while the upper limit $E_{\rm max}$ is the maximum neutrino energy, which for the SNS beam is 52.8 MeV. Notice that Eq. (15) depends on both the real (T') and the reconstructed (T) nuclear recoil energies, with $\mathcal{R}(T,T')$ the smearing function, which may be different for each detector, and the upper limit of the integral over T' is the maximal nuclear recoil energy that can be induced by a neutrino from the beam, $T'_{\rm max} = 2E_{\rm max}^2/M$. Finally, $A(T)$ is an acceptance function, which also depends

Target nucleus	Z	N	Z/N	M (a.m.u)
Cs	55	78	0.71	132.91
I	53	74	0.72	126.90
Na	11	12	0.92	22.99

TABLE I. Main properties of the target nuclei used in the two detectors considered in this work: number of protons (Z) and neutrons (N), proton to neutron ratio (Z/N) and mass in atomic units (M).

on the detector. If the target material consists of two nuclei, as is the case for the CsI and NaI detectors, the total number of CEνNS events is the sum of the numbers of neutrino scatterings on each nucleus, computed separately from Eq. (15) with the appropriate values of M and $M_{\rm mol}$.

When the timing information of the experiment is available, one can also bin the data in time intervals. In this case, the predicted number of events in the nuclear recoil energy bin i and time bin j is given by

$$N_{ij}^{th} = \sum_{\nu=\nu_e,\nu_\mu,\overline{\nu}_\mu} N_{\nu,i}^{th} \int_{t_j}^{t_{j+1}} f_\nu(t)\, \varepsilon_t(t)\, dt\,, \tag{16}$$

where $N_{\nu,i}^{th}$ is the predicted number of CEνNS events induced by the component ν ($\nu = \nu_e, \nu_\mu, \overline{\nu}_\mu$) of the neutrino flux in the nuclear recoil energy bin $[T_i, T_{i+1}]$, $f_\nu(t)$ is the time distribution of the neutrino flux component ν taken from Ref. [54], and $\varepsilon(t)$ is the timing efficiency provided in Ref. [3]. To compute $N_{\nu,i}^{th}$, we use Eq. (15) without performing the sum over the contributions of the three components of the neutrino flux.

In our analysis, we will use the data of the current CsI detector of the COHERENT collaboration, as well as the expected future data from a NaI detector at the SNS. The main characteristics of these detectors are described in the following subsections.

B. Current CsI detector

The COHERENT collaboration performed CEνNS measurements with a 14.6 kg CsI detector located at 19.3 m from the neutrino source [3]. The accumulated data corresponds to $N_{POT} = 3.198 \times 10^{23}$, with a number of neutrinos per flavor of $r = 0.0848$ [3]. These numbers are used to compute the neutrino fluxes though Eqs. (12), (13) and (14). Following the COHERENT collaboration, we take the smearing function $\mathcal{R}(T, T')$ in Eq. (15) to be the Gamma function given in Ref. [3] (we refer the reader to Ref. [10] for details about the smearing procedure). Finally, we take into account the timing information provided by the COHERENT collaboration in our analysis and use Eq. (16) to compute the predicted number of events in each recoil energy and time bin.

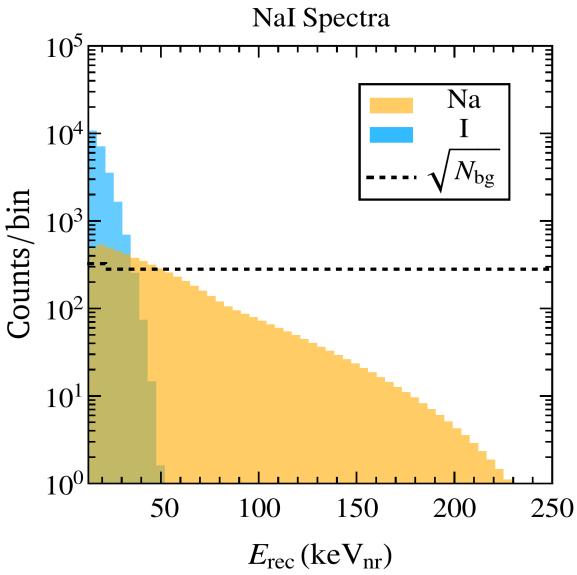


FIG. 1. Projected numbers of signal and background events in the SM as a function of the reconstructed nuclear recoil energy for the future NaI detector considered in this paper, assuming 3 years of data taking. The signal event spectra are displayed separately for the Sodium (yellow) and Iodine (light blue) target nuclei and include the contributions of the three components of the SNS neutrino beam (ν_e, ν_μ and $\overline{\nu}_\mu$). The expected steady-state background is represented by the black dashed curve, which for the sake of presentation shows the square root of the actual number of events.

C. Future NaI detector

The complete program of the COHERENT collaboration includes a NaI-based detector comprising several modules, each consisting of individual 7.7 kg NaI crystals [23]. In its final design, this detector is expected to have up to seven modules, each containing 63 crystals, giving a total mass $M_{\rm det} = 3395.7\,{\rm kg}$, which we shall consider in this work. The detector is expected to be located at a distance $L = 22\,{\rm m}$ from the neutrino source [23]. The SNS provides around $1.1{\times}10^{23}$ POT per calendar year, corresponding to $\simeq 5000$ hours (208.3 days) of operation per year. In this work, we consider three years of data taking with the SNS operating at current beam power, giving a total number of protons on target $N_{\rm POT} = 3.3 \times 10^{23}$. We also assume the current value of neutrinos per flavor, $r = 0.0848$. To compute the expected number of events, we assume a conservative constant acceptance of 80% and a Gaussian smearing function $\mathcal{R}(T, T')$ with an energy-dependent resolution $\sigma(T) = \eta\sqrt{TT_{th}}$, where[3] $\eta = 0.14$ and $T_{th} = 13\,{\rm keV_{nr}}$ is the nuclear recoil energy threshold of the detector [53].

[3] This choice reproduces the energy resolution given in [55] to a good approximation.

Under these assumptions and using the properties of the Sodium and Iodine nuclei given in Table I, we obtain the event spectra shown in Fig. 1, where the width of the recoil energy bins corresponds to 1 $\mathrm{keV_{ee}}$ [53]. To convert an electron recoil energy of 1 keV into the equivalent nuclear recoil energy (i.e., the nuclear recoil energy giving the same number of photoelectrons as an electron recoil energy of 1 keV), we use a constant quenching factor of 0.23 (estimated from Fig. 4.12 of Ref. [53]), resulting in a bin width of 4.33 $\mathrm{keV_{nr}}$. Fig. 1 displays the individual contributions of the Sodium (yellow) and Iodine (light blue) target nuclei to the signal event spectra. As can be seen, the CEνNS signal is dominated by the Iodine contribution below $T \approx 35\,\mathrm{keV_{nr}}$, while for larger nuclear recoil energies the Sodium contribution is dominant. This is due to the fact that the Iodine nucleus is heavier than Sodium, resulting in a larger cross section at low recoil energies, but in a lower endpoint (namely, $T_{\mathrm{max}}(\mathrm{I}) \approx 47\,\mathrm{keV_{nr}}$ while $T_{\mathrm{max}}(\mathrm{Na}) \approx 259\,\mathrm{keV_{nr}}$). As we will see later, the different features of the Sodium and Iodine contributions to the CEνNS signal can help reduce the degeneracies arising in the presence of vector-type GNIs. Fig. 1 also shows the expected steady-state background, which we assumed to be flat and equal to its average value of 300 ckkd (counts per keV_{ee} per kg per day) [53], except in the first two bins where the background is larger. There, we took a conservative value of 400 ckkd. In practice, the number of background events is reduced by a factor of 8000 [53] taking into account the fact that the SNS neutrino beam is pulsed. This reduction factor has been applied in Fig. 1. It is important to mention that the background model adopted here results in a smaller signal-over-background ratio than for the current CsI detector. For comparison, the signal-over-background ratio ranges from 4.2 $\times 10^{-3}$ to 0.15 in the first eight recoil energy bins of Fig. 1 (the ones that contribute the most to the statistics), while it varies between 0.04 and 2.2, depending on the energy bin, for the CsI measurement. This relatively large background level will affect the expected sensitivity of the NaI detector to generalized neutrino interactions. We will see in particular that, in spite of a larger statistics, the NaI detector will not be able to significantly improve the current CsI constraints on scalar and tensor GNI parameters.

IV. ANALYSIS PROCEDURE

For the analysis of the current CsI detector data, we follow the same procedure as Ref. [10] and use the Poissonian χ^2 function

$$\chi^2(\kappa) \,=\, \min_{\xi} \left\{ 2 \sum_{i,j} \left[N_{ij}(\kappa,\xi) - \tilde{N}_{ij} + \tilde{N}_{ij} \ln \left(\frac{\tilde{N}_{ij}}{N_{ij}(\kappa,\xi)} \right) \right] + \sum_{m=1}^{2} \frac{\xi_{sig,m}^2}{\sigma_{\xi_{sig,m}}^2} + \sum_{k=1}^{3} \frac{\xi_{bg,k}^2}{\sigma_{\xi_{bg,k}}^2} \right\}, \quad (17)$$

where i, j run over the nuclear recoil energy and time bins, respectively, $\tilde{N}_{ij}$ is the experimental number of events in bins (i, j), $N_{ij}(\kappa, \xi)$ the predicted number of events in the same bins in the presence of GNIs, κ stands for the set of GNI parameters under test, and ξ is the set of nuisance parameters over which we minimize the expression within braces. $N_{ij}(\kappa, \xi)$ is computed as

$$\begin{aligned} N_{ij}(\kappa,\xi) = \quad & (1+\xi_{sig,1})N_{ij}^{sig}(\kappa,\xi_{sig,2},\xi_1,\xi_2) + (1+\xi_{bg,1})N_{ij}^{BRN}(\xi_1) \\ & + (1+\xi_{bg,2})N_{ij}^{NIN}(\xi_1) + (1+\xi_{bg,3})N_{ij}^{SSB}, \end{aligned} \tag{18}$$

where N_{ij}^{sig} is the number of signal events, and N_{ij}^{BRN}, N_{ij}^{NIN} and N_{ij}^{SSB} are the numbers of beam-related neutron (BRN), neutrino-induced neutron (NIN) and steady-state (SSB) background events, respectively. The signal nuisance parameter $\xi_{sig,1}$ accounts for the detector efficiency, neutrino flux, and quenching factor (QF) normalizations, while $\xi_{sig,2}$ is related to the nuclear radius. The corresponding uncertainties are $\sigma_{\xi_{sig,1}} = 11.45\%$ and $\sigma_{\xi_{sig,2}} = 5\%$ [10]. As for the nuisance parameters $\xi_{bg,k}$, they are associated with the different sources of background: BRN ($\xi_{bg,1}$), NIN ($\xi_{bg,2}$) and SSB ($\xi_{bg,3}$), with corresponding uncertainties $\sigma_{\xi_{bg,1}} = 25\%$, $\sigma_{\xi_{bg,2}} = 35\%$ and $\sigma_{\xi_{bg,3}} = 2.1\%$ [3]. Following Ref. [10], we include in Eq. (17) two additional nuisance parameters ξ_1 and ξ_2, with no penalization term, which account for deviations in the uncertainty on the CEνNS detection efficiency and on beam timing, respectively. By minimizing over all nuisance parameters, we obtain allowed regions at a given confidence level for the GNI parameters under test.

To assess the sensitivity of the future NaI detector to GNI parameters, we consider the following χ^2 function

$$\chi^2(\kappa) = \min_{\xi} \left[\sum_i 2 \left\{ N_i(\kappa,\xi) - \tilde{N}_i + \tilde{N}_i \ln\left(\frac{\tilde{N}_i}{N_i(\kappa,\xi)} \right) \right\} + \left(\frac{\xi_{sig}}{\sigma_{sig}} \right)^2 + \left(\frac{\xi_{bg}}{\sigma_{bg}} \right)^2 \right], \tag{19}$$

where we have divided the data in recoil energy bins labelled by i. Since we are dealing with simulated data, we compare the theoretical predictions in the absence and in the presence of GNIs. Hence, we take

$$\tilde{N}_i = N_i^{sig}(SM) + N_i^{bg}\,, \tag{20}$$

where $N_i^{sig}(SM)$ represents the number of signal events in bin i predicted by the SM and N_i^{bg} is the simulated steady-state background described in Section III, while

$$N_i(\kappa,\xi) = N_i^{sig}(\kappa)(1+\xi_{sig}) + N_i^{bg}(1+\xi_{bg})\,, \tag{21}$$

where $N_i^{sig}(\kappa)$ is the predicted number of signal events in bin i in the presence of GNIs, and the nuisance parameters ξ_{sig}, ξ_{bg} account for the signal and background normalizations, respectively.

For the associated systematic uncertainties, we conservatively assume $\sigma_{sig} = 10\%$ and $\sigma_{bg} = 5\%$, motivated by the known values of the uncertainties for the current CsI detector. Notice that, due to the absence of detailed information on the signal and background systematic uncertainties, for the NaI detector currently under deployment at the SNS, we only consider two nuisance parameters.

V. EXPECTED SENSITIVITIES TO GNI PARAMETERS

In this section, we study the sensitivities to GNI parameters[4] of a future NaI detector with characteristics described in Subsection III C, and compare them with the constraints set by the COHERENT CsI detector. As discussed in Subsection II B, we consider only scalar, vector and tensor interactions and denote the corresponding couplings by $\varepsilon^{qS}_{\alpha\beta}$, $\varepsilon^{qV}_{\alpha\beta}$ and $\varepsilon^{qT}_{\alpha\beta}$, respectively, with $\alpha, \beta = e, \mu, \tau$ and $q = u, d$.

Since the SNS neutrino beam consists only of electron neutrinos, muon neutrinos and muon antineutrinos, not all of these parameters can be constrained by CEνNS measurements. Vector GNIs, also known as NSIs, are subject to the hermiticity condition $\varepsilon^{qV}_{\beta\alpha} = \left(\varepsilon^{qV}_{\alpha\beta}\right)^*$. As a result, there are 6 independent parameters for each quark flavor, but only 5 of them can be constrained at the SNS: two real flavor-diagonal couplings, ε^{qV}_{ee} and $\varepsilon^{qV}_{\mu\mu}$, and three complex flavor off-diagonal couplings, $\varepsilon^{qV}_{e\mu}$, $\varepsilon^{qV}_{e\tau}$ and $\varepsilon^{qV}_{\mu\tau}$. In this paper, we study the sensitivities of the CsI and NaI detectors to ε^{qV}_{ee}, $\varepsilon^{qV}_{\mu\mu}$ and $\varepsilon^{qV}_{e\mu}$ $(q = u, d)$, leaving aside the other off-diagonal couplings, which are less constrained[5].

For scalar and tensor GNIs, the number of independent parameters depends on whether neutrinos are Dirac of Majorana fermions. In this paper, we assume that they are Majorana fermions, both for simplicity and because this possibility is better motivated from a theoretical point of view. Scalar couplings are therefore symmetric, and for each quark flavor, 5 out of the 6 independent ones can be constrained by CEνNS measurements at the SNS. However, due to the absence of interference between the scalar and SM contributions to the CEνNS cross section, the same constraints apply to ε^{qS}_{ee} and $\varepsilon^{qS}_{e\tau}$, as well as to $\varepsilon^{qS}_{\mu\mu}$ and $\varepsilon^{qS}_{\mu\tau}$. It is therefore sufficient to study the sensitivities of the CsI and NaI detectors to the scalar couplings ε^{qS}_{ee}, $\varepsilon^{qS}_{\mu\mu}$ and $\varepsilon^{qS}_{e\mu}$ $(q = u, d)$. Finally, tensor couplings are antisymmetric for Majorana neutrinos, leaving only 3 indepedent parameters $\varepsilon^{qT}_{e\mu}$, $\varepsilon^{qT}_{e\tau}$ and $\varepsilon^{qT}_{\mu\tau}$ for each value of q, all of which can induce scatterings of SNS neutrinos off nuclei.

Throughout this paper, we assume all GNI couplings to be real. The upper bounds on these

[4] From now on, we will collectively refer to vector, scalar and tensor couplings as GNI parameters, and we will more specifically use NSI parameters for vector couplings.

[5] This is due to the fact that $\varepsilon^{qV}_{e\mu}$ can induce coherent scatterings for all three components of the SNS neutrino flux, while only ν_e (resp. ν_μ and $\nu_{\bar{\mu}}$) can scatter off a nucleus via $\varepsilon^{qV}_{e\tau}$ (resp. $\varepsilon^{qV}_{\mu\tau}$).

parameters obtained from the currently available CsI data and the expected sensitivity of the future NaI detector are determined following the numerical procedure described in Section IV.

A. Expected sensitivities to individual GNI parameters

We first consider the expected sensitivities of the NaI detector to individual GNI parameters, i.e. we consider a single nonvanishing coupling at a time. Note that in this case, the assumption of real couplings does not mean any loss of generality, as the CEνNS cross section only depends on the moduli of the GNI parameters (except for diagonal vector couplings, which are real and can interfere with the SM contribution).

The results for vector couplings are shown in Fig. 2, where the $\Delta\chi^2$ functions defined in Section IV are plotted against the parameter assumed to be nonvanishing (from left to right: ε_{ee}^{qV}, $\varepsilon_{e\mu}^{qV}$ and $\varepsilon_{\mu\mu}^{qV}$, with $q = u$ in the upper panels and $q = d$ in the lower panels). The black solid curves are obtained using the currently available CsI data, and are consistent with the results obtained in Ref. [10]. The red dashed curves represent the expected sensitivities of the future NaI detector after 3 years of data taking, assuming conservatively $\sigma_{sig} = 10\%$ and $\sigma_{bg} = 5\%$. For the flavor-diagonal parameters ε_{ee}^{qV} and $\varepsilon_{\mu\mu}^{qV}$ $(q = u, d)$, two minima can be observed in the $\Delta\chi^2$ profiles[6], as a result of the interference between the SM and NSI contributions in the CEνNS cross section, Eq. (6). The sharp rise of the $\Delta\chi^2$ for $\varepsilon_{\mu\mu}^{qV}$ between the two minima is due to the fact that when $\varepsilon_{\mu\mu}^{qV} \approx 0.2$, the NSI and SM contributions to the CEνNS cross section almost completely cancel out for ν_μ's and $\bar{\nu}_\mu$'s, resulting in a reduction of the predicted number of CEνNS events by roughly 2/3, in strong tension with data. The same effect is present for ε_{ee}^{qV}, but since this coupling only affects ν_e-induced CEνNS events, i.e. roughly 1/3 of the signal, the $\Delta\chi^2$ reaches smaller values between the two minima (especially for CsI, which suffers from a smaller statistics than the NaI detector). In Table II, we give the 90% C.L. and 2σ expected sensitivities of the future NaI detector to the vector couplings ε_{ee}^{qV}, $\varepsilon_{e\mu}^{qV}$ and $\varepsilon_{\mu\mu}^{qV}$ $(q = u, d)$, as well as the constraints on these parameters from the COHERENT CsI data.

As can be seen from Fig. 2 and Table II, a clear improvement of the constraints on these diagonal vector couplings is expected from the future NaI detector, already with 3 years of data taking. This is partly due to the larger statistics of the NaI detector, as already mentioned. However, the most noticeable feature, namely the fact that the second minimum of the χ^2 function tends to

[6] Note that for the CsI detector, the first minimum of the χ^2 function is not located at $\varepsilon_{ee}^{qV} = 0$ (resp. $\varepsilon_{\mu\mu}^{qV} = 0$). This is due to the fact that we are analysing real data, which does not exactly match the SM prediction.

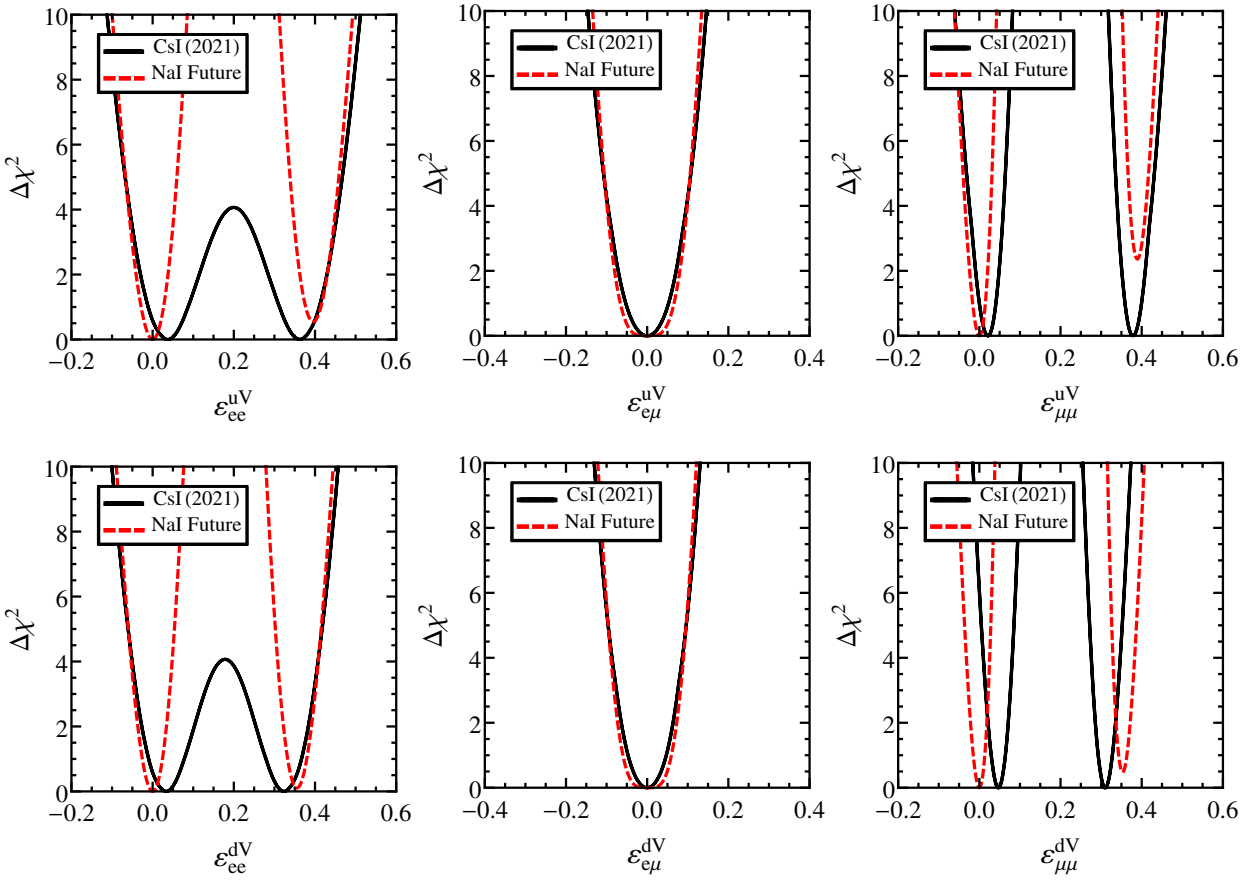


FIG. 2. One-dimensional $\Delta\chi^2$ profiles (17) and (19) for the vector coupling $\varepsilon_{e\ell}^{qV}$ (left panels), $\varepsilon_{e\mu}^{qV}$ (middle panels) and $\varepsilon_{\mu\mu}^{qV}$ (right panels), with $q = u$ in the upper panels and $q = d$ in the lower panels. The black solid curves correspond to Eq. (17) and are obtained using the currently available data of the COHERENT CsI detector. The red dashed curves correspond to Eq. (19) and represent the expected sensitivities of the NaI detector described in Subsection III C, assuming 3 years of data taking at the SNS. The off-diagonal NSI parameter $\varepsilon_{e\mu}^{qV}$ is assumed to be real.

be disfavoured (mostly for the parameter $\varepsilon_{\mu\mu}^{uV}$, and to a lesser extent for ε_{ee}^{uV} and $\varepsilon_{\mu\mu}^{dV}$), cannot be explained by statistics alone. This can be understood instead by noting that the SM–NSI interference term in the CEνNs cross section is controlled by the quantity (where we have neglected a term suppressed by $g_V^p \ll 1$)

$$C_{\alpha\alpha}^V + Q_{W,\alpha} \simeq N\left[\left(2\,\frac{Z}{N}+1\right)\varepsilon_{\alpha\alpha}^{uV} + \left(\frac{Z}{N}+2\right)\varepsilon_{\alpha\alpha}^{dV} - \frac{1}{2}\right], \tag{22}$$

and is therefore sensitive to the proton to neutron ratio of the target nucleus. In particular, the second $\Delta\chi^2$ minimum for $\varepsilon_{\alpha\alpha}^{uV}$ (resp. $\varepsilon_{\alpha\alpha}^{dV}$) is located at $\varepsilon_{\alpha\alpha}^{uV} \approx 1/(2Z/N+1)$ (resp. $\varepsilon_{\alpha\alpha}^{dV} \approx 1/(Z/N+2)$). While the CsI detector involves two nuclei with approximately the same proton

Detector	q	ε_{ee}^{qV}		$\varepsilon_{e\mu}^{qV}$		$\varepsilon_{\mu\mu}^{qV}$	
		90% C.L.	2σ C.L.	90% C.L.	2σ C.L.	90% C.L.	2σ C.L.
NaI	u	$[-0.047, 0.044]$ $\cup [0.356, 0.438]$	$[-0.059, 0.053]$ $\cup [0.345, 0.451]$	$[-0.088, 0.088]$	$[-0.099, 0.099]$	$[-0.027, 0.024]$ $\cup [0.380, 0.399]$	$[-0.036, 0.028]$ $\cup [0.371, 0.411]$
	d	$[-0.043, 0.039]$ $\cup [0.317, 0.398]$	$[-0.053, 0.048]$ $\cup [0.308, 0.408]$	$[-0.079, 0.079]$	$[-0.09, 0.09]$	$[-0.026, 0.021]$ $\cup [0.333, 0.376]$	$[-0.032, 0.026]$ $\cup [0.329, 0.382]$
CsI	u	$[-0.04, 0.134]$ $\cup [0.266, 0.440]$	$[-0.056, 0.186]$ $\cup [0.212, 0.456]$	$[-0.081, 0.081]$	$[-0.097, 0.097]$	$[-0.015, 0.053]$ $\cup [0.346, 0.414]$	$[-0.024, 0.060]$ $\cup [0.34, 0.424]$
	d	$[-0.035, 0.120]$ $\cup [0.237, 0.393]$	$[-0.05, 0.168]$ $\cup [0.190, 0.408]$	$[-0.073, 0.073]$	$[-0.087, 0.087]$	$[-0.016, 0.075]$ $\cup [0.282, 0.340]$	$[-0.010, 0.081]$ $\cup [0.276, 0.347]$

TABLE II. Sensitivities to the vector couplings ε_{ee}^{qV}, $\varepsilon_{e\mu}^{qV}$ and $\varepsilon_{\mu\mu}^{qV}$ ($q = u, d$) at the 90% and 2σ confidence levels with 1 degree of freedom (i.e. $\Delta\chi^2 \leq 2.71$ and $\Delta\chi^2 \leq 4$, respectively). For the CsI detector, the intervals actually correspond to the ranges of parameter values allowed by the currently available data, while for the NaI detector they represent the expected sensitivities assuming 3 years of data taking at the SNS. The off-diagonal NSI parameter $\varepsilon_{e\mu}^{qV}$ is assumed to be real.

to neutron ratio, this is not the case for the NaI detector (see Table I), making it more difficult to fit SM-like experimental data with a nonzero value of $\varepsilon_{\alpha\alpha}^{uV}$ or $\varepsilon_{\alpha\alpha}^{dV}$. The same effect is observed when combining the data collected by two detectors with target nuclei characterized by different proton to neutron ratios [24, 39]; in the case of the NaI detector, the two nuclei are present in the same target. However, the NaI detector considered in this paper is not as efficient at constraining flavor-diagonal vector couplings as expected from suitably chosen combinations of the detectors proposed for CEνNS measurements at the European Spallation Source [24], such as CsI + Si and Xe + Si [39]. The latter indeed benefit from a large statistics and from a lower background level than the NaI detector. Also, the advantage of combining two target nuclei with different proton to neutron ratios in the same detector is counterbalanced, in the case of NaI, by the fact that the recoil energy bins in which the Na-induced scatterings contribute significantly are the ones in which the signal over background ratio is the least favorable, as can be seen from Fig. 1.

For off-diagonal vector couplings, whose contributions do not interfere with the SM one, no significant improvement on the current constraints is expected from the future NaI detector, in spite of the larger statistics (the results for $\varepsilon_{e\tau}^{qV}$ and $\varepsilon_{\mu\tau}^{qV}$ are qualitatively similar to the ones for $\varepsilon_{e\mu}^{qV}$ shown in Fig. 2). This is due to the fact that the signal-over-background ratio is significantly smaller than for the CsI detector (see the discussion at the end of Subsection III C).

Let us now move on to the results for scalar GNIs. As explained before, CEνNS measurements

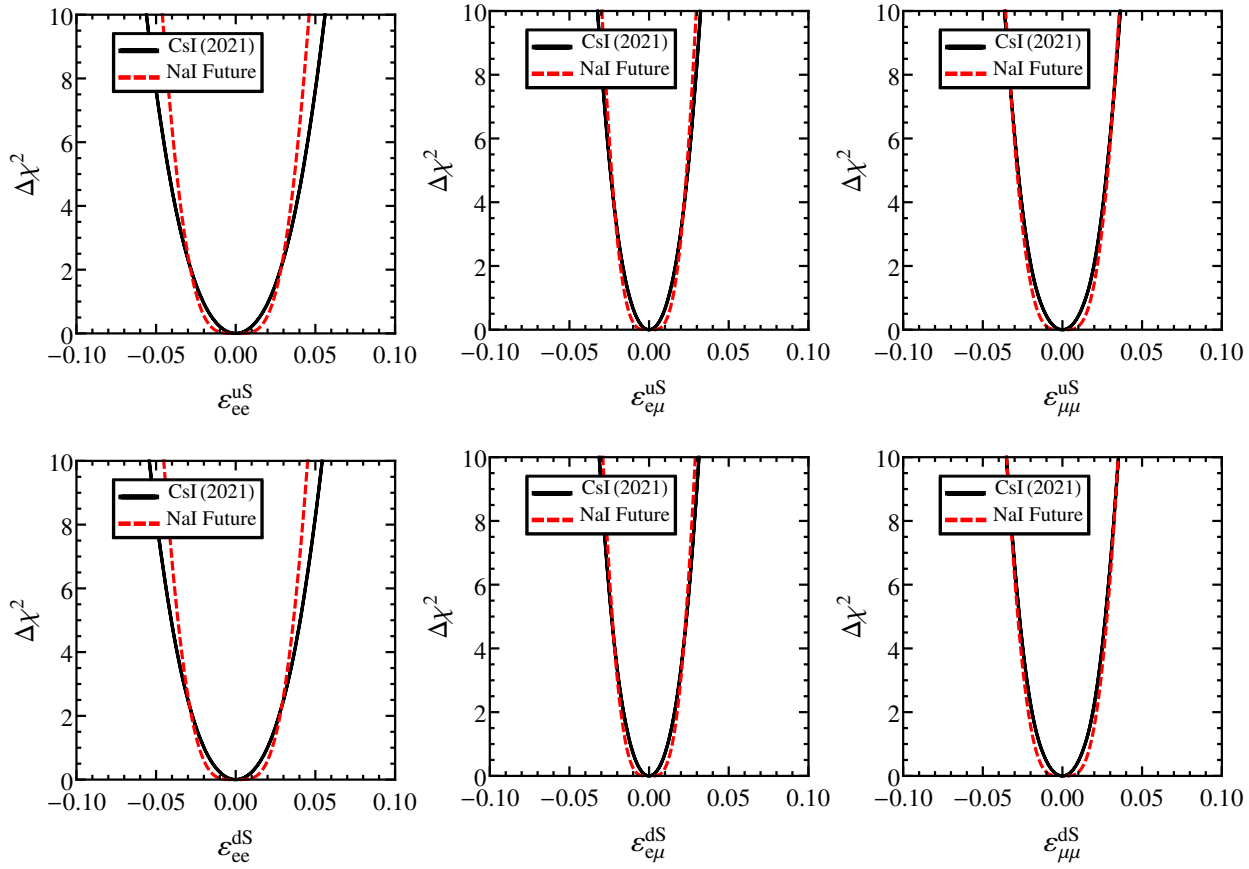


FIG. 3. Same as Fig. 2, but for the scalar couplings ε^{qS}_{ee} (left panels), $\varepsilon^{qS}_{e\mu}$ (middle panels) and $\varepsilon^{qS}_{\mu\mu}$ (right panels), with $q = u$ in the upper panels and $q = d$ in the lower panels. All couplings are assumed to be real.

Detector	q	ε^{qS}_{ee}		$\varepsilon^{qS}_{e\mu}$		$\varepsilon^{qS}_{\mu\mu}$	
		90% C.L.	2σ C.L.	90% C.L.	2σ C.L.	90% C.L.	2σ C.L.
NaI	u	$[-0.031, 0.031]$	$[-0.034, 0.034]$	$[-0.02, 0.02]$	$[-0.022, 0.022]$	$[-0.024, 0.024]$	$[-0.027, 0.027]$
	d	$[-0.031, 0.031]$	$[-0.033, 0.033]$	$[-0.019, 0.019]$	$[-0.021, 0.021]$	$[-0.023, 0.023]$	$[-0.026, 0.026]$
CsI	u	$[-0.032, 0.032]$	$[-0.038, 0.038]$	$[-0.018, 0.018]$	$[-0.022, 0.022]$	$[-0.022, 0.022]$	$[-0.026, 0.026]$
	d	$[-0.031, 0.031]$	$[-0.037, 0.037]$	$[-0.018, 0.018]$	$[-0.022, 0.022]$	$[-0.022, 0.022]$	$[-0.025, 0.025]$

TABLE III. Same as Table II, but for the scalar couplings ε^{qS}_{ee}, $\varepsilon^{qS}_{e\mu}$ and $\varepsilon^{qS}_{\mu\mu}$ ($q = u$, d).

at the SNS only give six independent contraints on scalar couplings, since $\varepsilon^{qS}_{e\tau}$ and $\varepsilon^{qS}_{\mu\tau}$ contribute the same way to the CEνNS cross section as ε^{qS}_{ee} and $\varepsilon^{qS}_{\mu\mu}$, respectively[7]. We therefore only present results for the 6 scalar couplings ε^{qS}_{ee}, $\varepsilon^{qS}_{\mu\mu}$ and $\varepsilon^{qS}_{e\mu}$ ($q = u, d$). The $\Delta\chi^2$ profiles are displayed in Fig. 3, and the 90% C.L. and 2σ expected sensitivities of the future NaI detector, together

[7] This statement holds for symmetric scalar couplings, hence for Majorana neutrinos (as assumed in this paper), but not for Dirac neutrinos.

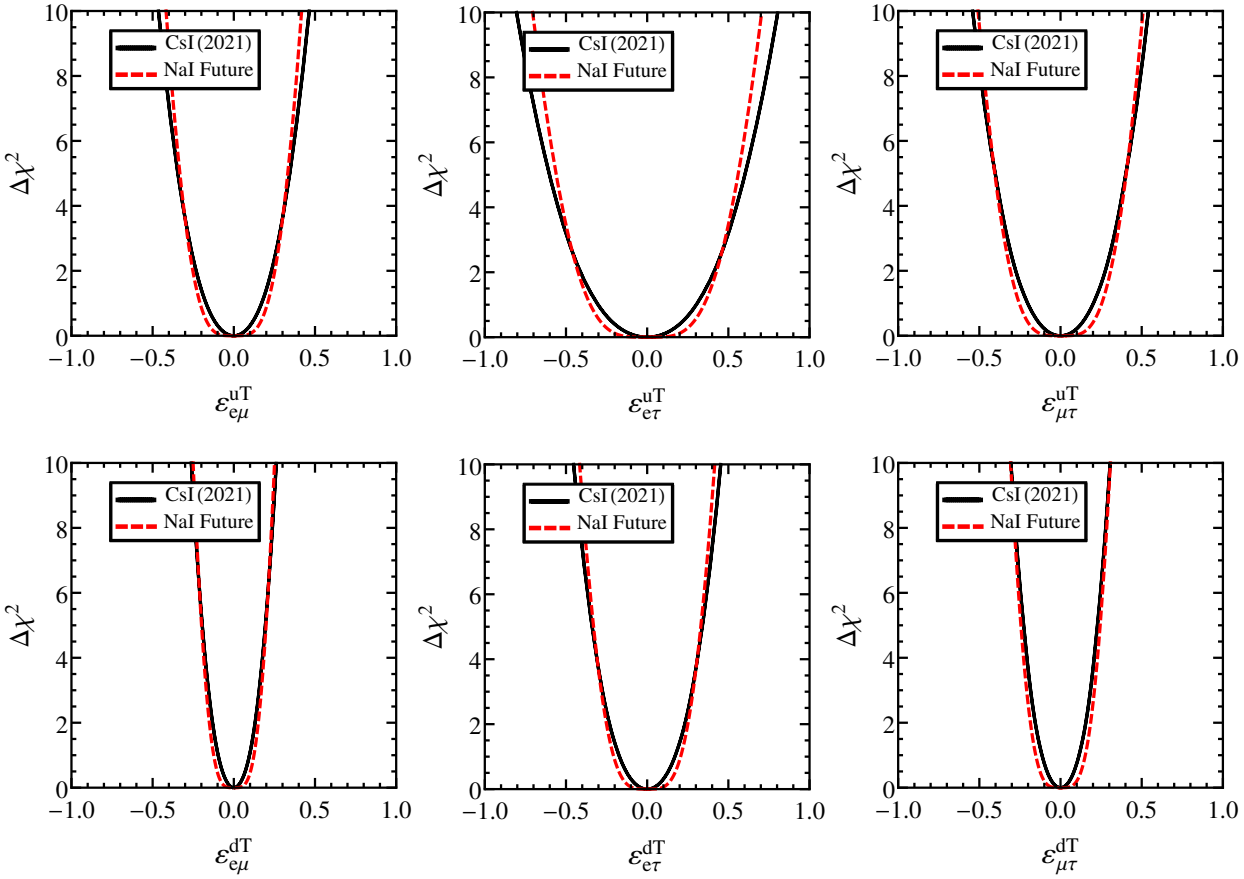


FIG. 4. Same as Fig. 2, but for the tensor couplings $\varepsilon^{qT}_{e\mu}$ (left panels), $\varepsilon^{qT}_{e\tau}$ (middle panels) and $\varepsilon^{qT}_{\mu\tau}$ (right panels), with $q = u$ in the upper panels and $q = d$ in the lower panels. All couplings are assumed to be real.

with the constraints from the COHERENT CsI data, are given in Table III. As can be seen by comparing Figs. 2 and 3, the results for scalar couplings are similar to the ones for off-diagonal vector couplings, which is not a surprise since both types of interactions share the property of not interfering with the SM contribution in the CEνNS cross section. In particular, we do not expect the future NaI detector to significantly improve the current CsI constraints on scalar couplings within the assumptions made in our analysis (3 years of running time, $\sigma_{sig} = 10\%$ and $\sigma_{bg} = 5\%$). Note however that, at the $\gtrsim 3\sigma$ confidence level, there is a slight improvement for ε^{uS}_{ee} and ε^{dS}_{ee}.

Finally, the results for tensor GNIs are shown in Fig. 4 and Table IV. Recall that there are only 6 independent tensor couplings if, as assumed in this paper, neutrinos are Majorana fermions, namely $\varepsilon^{qT}_{e\mu}$, $\varepsilon^{qT}_{e\tau}$ and $\varepsilon^{qT}_{\mu\tau}$ $(q = u, d)$. Comparing the expected sensitivities of the future NaI detector with the constraints from the COHERENT CsI data (see Table IV), we only expect a marginal improvement of the constraints on tensor GNI paremeters. However, at the $\gtrsim 3\sigma$ confidence level,

Detector	q	$\varepsilon^{qT}_{e\mu}$		$\varepsilon^{qT}_{e\tau}$		$\varepsilon^{qT}_{\mu\tau}$	
		90% C.L.	2σ C.L.	90% C.L.	2σ C.L.	90% C.L.	2σ C.L.
NaI	u	$[-0.275, 0.275]$	$[-0.310, 0.310]$	$[-0.465, 0.465]$	$[-0.525, 0.525]$	$[-0.337, 0.337]$	$[-0.380, 0.380]$
	d	$[-0.165, 0.165]$	$[-0.186, 0.186]$	$[-0.274, 0.274]$	$[-0.310, 0.310]$	$[-0.203, 0.203]$	$[-0.228, 0.228]$
CsI	u	$[-0.265, 0.265]$	$[-0.314, 0.314]$	$[-0.467, 0.467]$	$[-0.55, 0.55]$	$[-0.314, 0.314]$	$[-0.369, 0.369]$
	d	$[-0.15, 0.15]$	$[-0.177, 0.177]$	$[-0.264, 0.264]$	$[-0.310, 0.310]$	$[-0.177, 0.177]$	$[-0.208, 0.208]$

TABLE IV. Same as Table II, but for the tensor couplings $\varepsilon^{qT}_{e\mu}$, $\varepsilon^{qT}_{e\tau}$ and $\varepsilon^{qT}_{\mu\tau}$ ($q = u, d$).

a slight improvement can be observed for $\varepsilon^{uT}_{e\tau}$.

B. Expected sensitivities to pairs of GNI parameters

In this section, we study the expected sensitivity of the NaI detector to pairs of GNI parameters, i.e. we consider two nonvanishing couplings at a time. As it is well known, this situation leads to degeneracies in the form of extended allowed regions in the two-dimensional GNI parameter space. Combining data on target nuclei characterized by different proton to neutron ratios can help reduce these degeneracies, as shown e.g. in Refs. [8, 24, 39]. In the case of the NaI detector studied in this paper, the two nuclei have significantly different proton to neutron ratios 0.92 (for Na) and 0.72 (for I), while the Cesium and Iodine targets of the CsI detector have approximately the same ratio.

We begin our analysis by considering the case where the two nonvanishing GNI parameters are of vector type (i.e., NSIs). Many different combinations of two NSI parameters are possible; we only present results for a few representative examples, as the other combinations are characterized by similar allowed regions. Fig. 5 shows the allowed regions at the 90% confidence level for four different pairs of vector couplings. The upper panels correspond to combinations of diagonal couplings involving the same lepton flavor but different quarks, namely $\left(\varepsilon^{uV}_{ee}, \varepsilon^{dV}_{ee}\right)$ in the left panel and $\left(\varepsilon^{uV}_{\mu\mu}, \varepsilon^{dV}_{\mu\mu}\right)$ in the right panel. By contrast, the lower panels correspond to pairs of vector couplings involving the same quark but different lepton flavors. The combination in the left panel, $\left(\varepsilon^{uV}_{ee}, \varepsilon^{uV}_{\mu\mu}\right)$, involves two diagonal couplings, while the one in the right panel, $\left(\varepsilon^{uV}_{e\mu}, \varepsilon^{uV}_{\mu\mu}\right)$, involves a diagonal and an off-diagonal coupling. In all cases, the green regions are the ones allowed by the current CsI data (these allowed regions were first presenteded in Ref. [10], and are shown here for reference), while the regions delimited by red lines correspond to the expected sensitivity of the NaI detector described in Subsection III C, assuming 3 years of data taking at the SNS.

In the upper panels of Fig. 5, the allowed regions consist of two parallel bands [9]. This is a

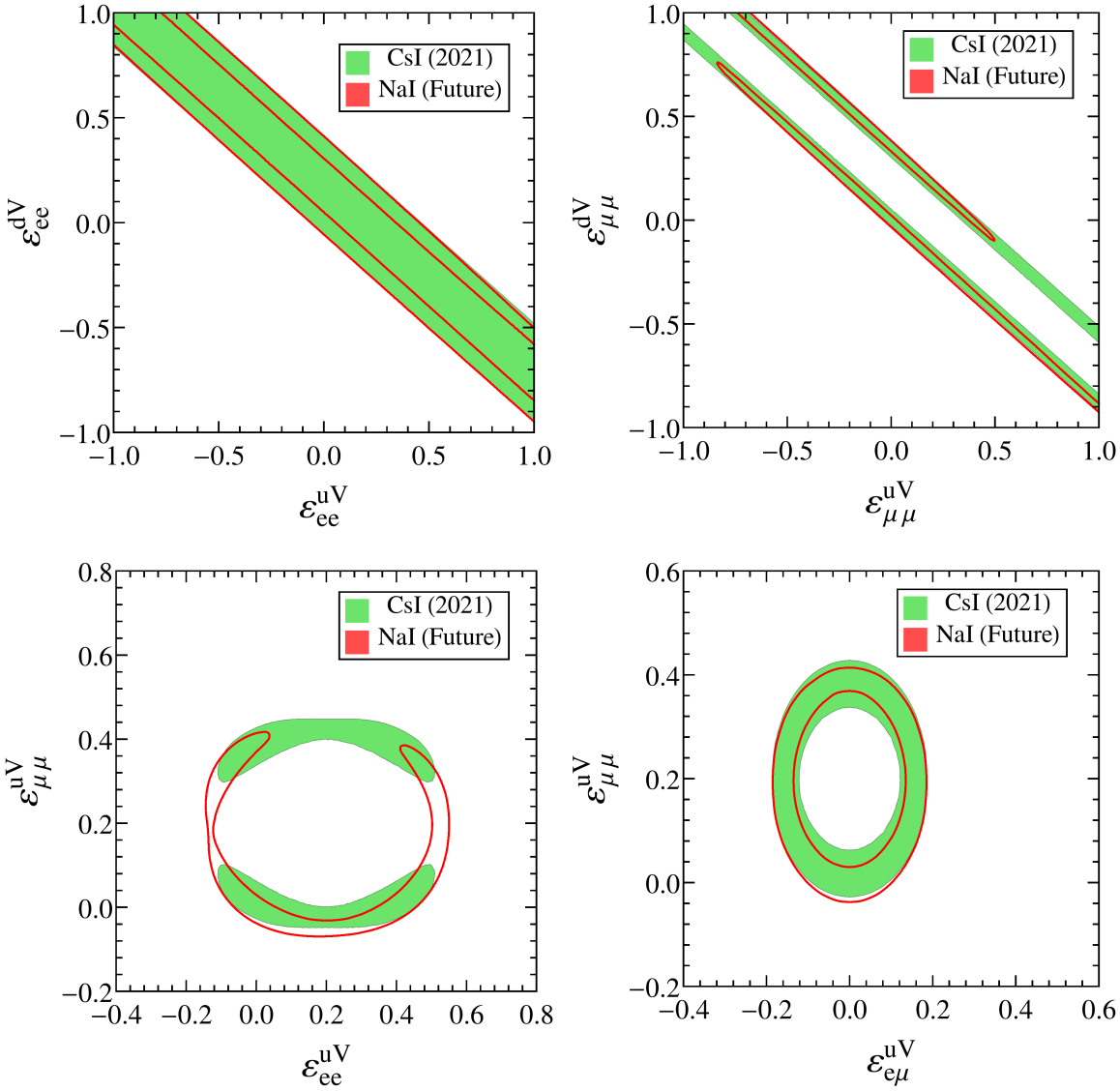


FIG. 5. 90% C.L. allowed regions (with two degrees of freedom, i.e. $\Delta\chi^2 \leq 4.61$) for different combinations of two NSI parameters: $(\varepsilon^{uV}_{ee}, \varepsilon^{dV}_{ee})$ (upper left panel), $(\varepsilon^{uV}_{\mu\mu}, \varepsilon^{dV}_{\mu\mu})$ (upper right panel), $(\varepsilon^{uV}_{ee}, \varepsilon^{uV}_{\mu\mu})$ (lower left panel) and $(\varepsilon^{uV}_{e\mu}, \varepsilon^{uV}_{\mu\mu})$ (lower right panel). The green regions correspond to the constraints from the current CsI data, while the red contours represent the expected sensitivity of the NaI detector described in Subsection III C, assuming 3 years of data taking at the SNS. The off-diagonal parameter $\varepsilon^{qV}_{e\mu}$ is assumed to be real.

consequence of the fact that, for a single-nucleus target (or a target composed of two nuclei with approximately the same proton to neutron ratio, like CsI), the CEνNS cross section depends on NSI parameters only through the modulus of the combination (22), which can be more conveniently

rewritten as

$$C_{\alpha\alpha}^{V} + Q_{W,\alpha} \simeq (Z+2N)\left(\varepsilon_{\alpha\alpha}^{dV} - m\,\varepsilon_{\alpha\alpha}^{uV}\right) - \frac{1}{2}\,, \qquad m = -\frac{2Z+N}{Z+2N}\,. \tag{23}$$

Thus, values of $\varepsilon_{\alpha\alpha}^{uV}$ and $\varepsilon_{\alpha\alpha}^{dV}$ lying on a straight ligne with slope m [8] produce the same number of CEνNS events in each recoil energy bin. This is an example of the degeneracies mentioned above. It follows that the parameter space region allowed by SM-like experimental data at a given confidence interval (here 90% C.L.) consists of two parallel bands around the straight lines $\varepsilon_{\alpha\alpha}^{dV} - m\,\varepsilon_{\alpha\alpha}^{uV} = 0$ and $\varepsilon_{\alpha\alpha}^{dV} - m\,\varepsilon_{\alpha\alpha}^{uV} = 1/(Z+2N)$. This is what can be observed in Fig. 5 for CsI, where the two bands overlap for $\alpha = e$ due to the limited statistics (there are twice as many ν_μ's and $\overline{\nu}_\mu$'s as ν_e's in the SNS neutrino beam). For NaI, the bands are thinner due to the larger statistics, and they reduce to closed or semi-closed areas in the $(\varepsilon_{\mu\mu}^{uV},\, \varepsilon_{\mu\mu}^{dV})$ plane[8]. This partial breaking of the degeneracy between $\varepsilon_{\mu\mu}^{uV}$ and $\varepsilon_{\mu\mu}^{dV}$ is due to the fact that Sodium and Iodine have different proton to neutron ratios, hence different slopes m_I and m_{Na}.

The elliptic-like shape of the allowed regions in the lower panels of Fig. 5 can be understood from the expression for the predicted CEνNS events in each reconstructed nuclear recoil energy bin [39]. Considering for instance the couplings $\left(\varepsilon_{ee}^{uV},\, \varepsilon_{\mu\mu}^{uV}\right)$, the predicted number of events in the recoil energy bin $[T_i, T_{i+1}]$, for a single-nucleus detector, can be written as

$$N_i^{th} = C_i^e\,(2Z+N)^2\left(\varepsilon_{ee}^{uV} + \frac{Zg_V^p + Ng_V^n}{2Z+N}\right)^2 + C_i^\mu\,(2Z+N)^2\left(\varepsilon_{\mu\mu}^{uV} + \frac{Zg_V^p + Ng_V^n}{2Z+N}\right)^2\,, \tag{24}$$

where C_i^e (resp. C_i^μ) is the number obtained by performing the integral over E_{ν_e} (resp. E_{ν_μ} and $E_{\overline{\nu}_\mu}$) and the true and reconstructed nuclear recoil energies in Eq. (15), after having factorized out $|C_{\alpha\alpha}^V + Q_{W,\alpha}|^2$. Eq. (24) is the equation of an ellipse with semi-major and semi-minor axes (given respectively by $\sqrt{N_i^{th}/C_i^e\,(2Z+N)^2}$ and $\sqrt{N_i^{th}/C_i^\mu\,(2Z+N)^2}$) that depend on both the nucleus and the energy bin (the center of the ellipse also depends on the nucleus). All sets of $(\varepsilon_{ee}^{uV},\, \varepsilon_{\mu\mu}^{uV})$ values lying on this ellipse predict the same number of CEνNS events in the ith energy bin, but not necessarily in the other bins, resulting in an approximate degeneracy between these two parameters. This explains why the 90% C.L. allowed regions in the lower left panel of Fig. 5 have the form of approximate, partly broken ellipses. The degeneracy is broken by the energy binning and, in the case of the NaI detector, by the fact that the nuclei composing the target have different numbers of protons and neutrons (while they are approximately the same for Cesium and Iodine).

From the lower left panel of Fig. 5, one can see that the CsI and NaI detectors play a complementary role in reducing the degeneracy between the vector couplings ε_{ee}^{uV} and $\varepsilon_{\mu\mu}^{uV}$. For CsI, the

[8] The same effect should occur for $(\varepsilon_{ee}^{uV},\, \varepsilon_{ee}^{dV})$ as well, but is not seen in the left panel of Fig. 5 due to the relatively smaller statistics of the ν_e beam component with respect to the $\nu_\mu + \overline{\nu}_\mu$ components.

allowed region has a symmetric shape, reflecting the fact that Cesium and Iodine have almost the same numbers of protons and neutrons. There are no allowed points around $\varepsilon_{\mu\mu}^{uV} \approx 0.2$, where the NSI and SM contributions to ν_μ- and $\overline{\nu}_\mu$-induced CEνNS almost completely cancel out. While an increase of the number of ν_e-induced CEνNS events could in principle compensate for this effect, the required increase (hence the required value of ε_{ee}^{uV}) is not the same in all energy bins, so that no value of ε_{ee}^{uV} can mimic the experimental data within 90% C.L. A similar effect occurs around $\varepsilon_{ee}^{uV} \approx 0.2$, but only leads to a small shriking of the allowed region. By contrast, the expected allowed region for NaI is not symmetric. Around $\varepsilon_{ee}^{uV} \approx 0.2$, negative values of $\varepsilon_{\mu\mu}^{uV}$ can compensate for the suppression of ν_e-induced CEνNS events, but this is not possible for positive values of $\varepsilon_{\mu\mu}^{uV}$, as a consequence of the different proton to neutron ratios of the Sodium and Iodine nuclei[9]. Around $\varepsilon_{\mu\mu}^{uV} \approx 0.2$, the allowed region shrinks for negative values of ε_{ee}^{uV}, but the degeneracy is not completely broken.

Finally, we also observe an "ellipse-shaped" degeneracy between the vector couplings $\varepsilon_{e\mu}^{uV}$ and $\varepsilon_{\mu\mu}^{uV}$ in the lower right panel of Fig. 5, but neither the CsI nor the NaI detector are able to reduce this degeneracy. This can be understood by noting that $\varepsilon_{e\mu}^{uV}$ induces both ν_e and $\nu_\mu/\overline{\nu}_\mu$ coherent scatterings on the target nuclei, without interfering with the SM contribution. Nonvanishing values of $\varepsilon_{e\mu}^{uV}$ can therefore compensate for the destructive interference between the SM and $\varepsilon_{\mu\mu}^{uV}$ contributions to the CEνNS cross section, which would otherwise reduce the number of $\nu_\mu/\overline{\nu}_\mu$-induced signal events. As a result, the ellipse is not broken around $\varepsilon_{\mu\mu}^{uV} \approx 0.2$ in the case of the CsI detector, at variance with what one can see in the lower left panel. The expected sensitivity of the future NaI detector only represents a slight improvement on the CsI constraints. In particular, we do not observe a partial breaking of the degeneracy around $\varepsilon_{\mu\mu}^{uV} \approx 0.4$, unlike in the lower left panel.

Let us now consider scalar and tensor GNIs. Fig. 6 shows the 90% C.L. allowed regions for two different pairs of scalar couplings with the same lepton flavor indices, $\left(\varepsilon_{e\mu}^{uS}, \varepsilon_{e\mu}^{dS}\right)$ in the left panel and $\left(\varepsilon_{\mu\mu}^{uS}, \varepsilon_{\mu\mu}^{dS}\right)$ in the right panel. As in Fig. 5, the regions allowed by the current CsI data are in green, while the ones delimited by red lines correspond to the expected sensitivity of the future NaI detector, assuming 3 years of data taking at the SNS. Since the scalar couplings $\varepsilon_{\alpha\beta}^{uS}$ and $\varepsilon_{\alpha\beta}^{dS}$ contribute to the CEνNS cross section through the linear combination $C_{\alpha\beta}^{S}$, given in Eq. (7), the allowed regions are straight bands around a line passing through the origin (as a consequence of the fact that scalar couplings do not interfere with the SM contribution). While the slope of this line depends on the proton to neutron ratio of the target nucleus, it is in practice very close to -1

[9] The same asymmetric behaviour can be seen in the $\Delta\chi^2$ for $\varepsilon_{\mu\mu}^{uV}$, in the upper right panel of Fig. 2, where the second minimum is "uplifted" with respect to the first one, a feature that is not observed in the case of the CsI detector.

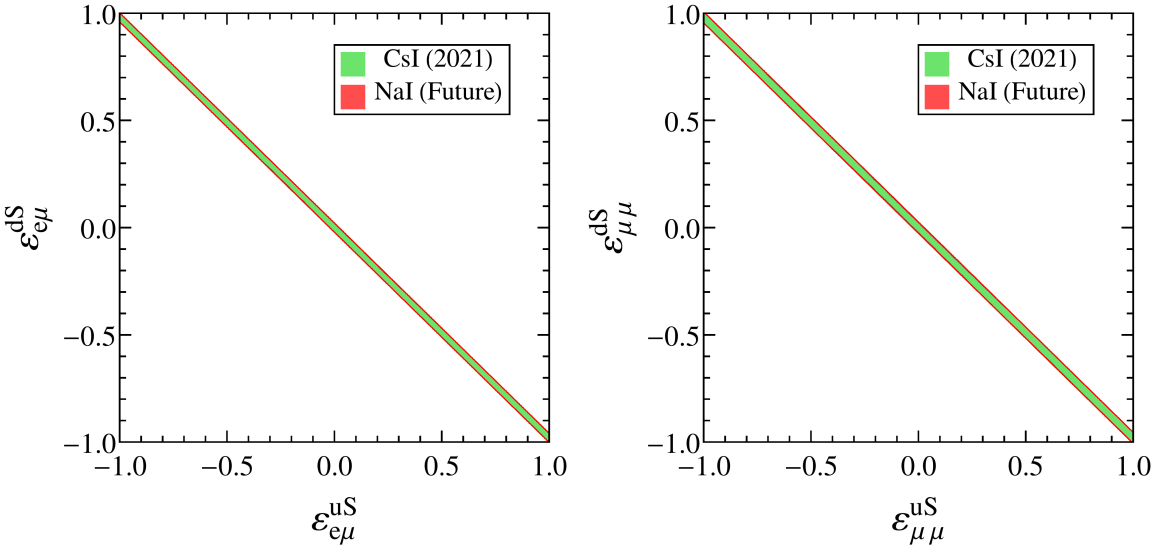


FIG. 6. 90% C.L. allowed regions for the pairs of scalar GNI parameters $\left(\varepsilon^{uS}_{e\mu}, \varepsilon^{dS}_{e\mu}\right)$ (left panel) and $\left(\varepsilon^{uS}_{\mu\mu}, \varepsilon^{dS}_{\mu\mu}\right)$ (right panel). The green regions correspond to the constraints from the current CsI data, while the red contours represent the expected sensitivity of the NaI detector described in Subsection III C, assuming 3 years of data taking at the SNS. All GNI parameters are assumed to be real.

for Cs, I and Na, as can be seen in Fig. 6. A consequence of this is that even though Sodium and Iodine have different proton to neutron ratios, the NaI detector is unable to reduce the degeneracy between $\varepsilon^{uS}_{\alpha\beta}$ and $\varepsilon^{dS}_{\alpha\beta}$.

Finally, Fig. 7 displays the 90% C.L. allowed regions for the pairs of tensor couplings $\left(\varepsilon^{uT}_{e\mu}, \varepsilon^{dT}_{e\mu}\right)$ (left panel) and $\left(\varepsilon^{uT}_{\mu\tau}, \varepsilon^{dT}_{\mu\tau}\right)$ (right panel). Qualitatively similar features to the scalar case can be observed, except that CEνNS is much less sensitive to tensor couplings, and that the line slope has a different dependence on the proton to neutron ratio of the nucleus. This explains the shape of the NaI region, but the poor sensitivity of CEνNS to tensor GNI parameters makes it difficult to reduce the degeneracy between $\varepsilon^{uT}_{\alpha\beta}$ and $\varepsilon^{dT}_{\alpha\beta}$ with a NaI detector.

C. Impact of statistics, systematic uncertainties and background level

In the previous subsections, we argued that the ability of the future NaI detector to set better constraints on generalized neutrino interactions than the current CsI detector was limited by its low signal-over-background ratio. In order to support this statement, and to identify other limiting factors, we now study the impact of statistics, systematic uncertainties and background level on the expected sensitivity of the future NaI detector to GNI parameters. To this end, we consider three

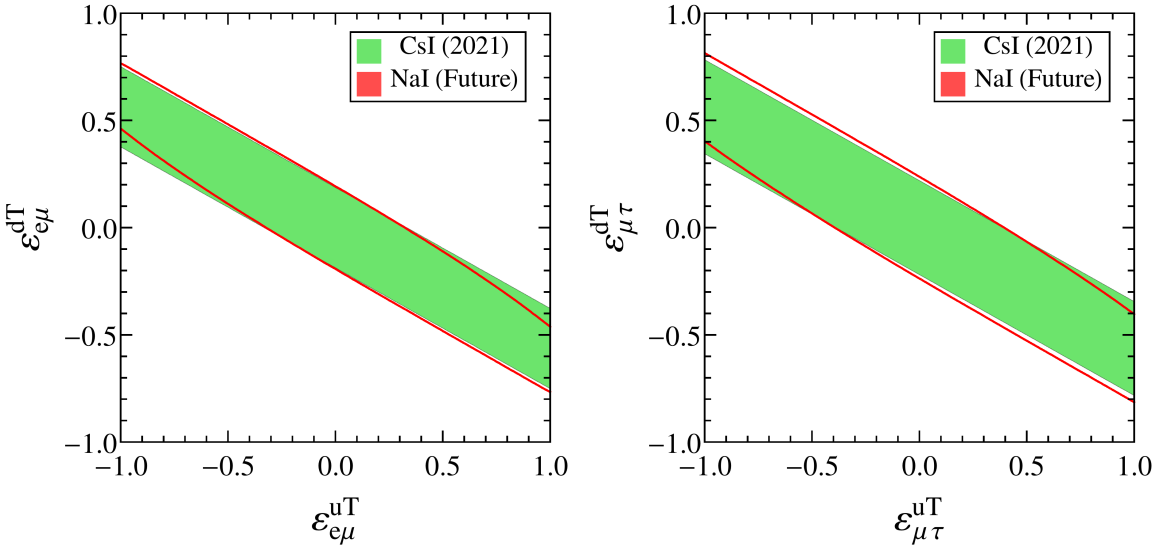


FIG. 7. Same as Fig. 6, but for the pairs of tensor GNI parameters $\left(\varepsilon^{uT}_{e\mu}, \varepsilon^{dT}_{e\mu}\right)$ (left panel) and $\left(\varepsilon^{uT}_{\mu\tau}, \varepsilon^{dT}_{\mu\tau}\right)$ (right panel).

scenarios characterized by more optimistic assumptions about the exposure time, signal systematic uncertainty or background control than the ones made in the previous subsections. In Scenario A, we assume five years of data taking with the same systematic errors as before (namely, $\sigma_{sig} = 10\%$ and $\sigma_{bg} = 5\%$). In Scenario B, we consider a running time of three years as in the previous subsections, but reduce the signal systematic uncertainty to $\sigma_{sig} = 5\%$, while keeping $\sigma_{bg} = 5\%$. Finally, in Scenario C, we assume a reduced background of 100 ckkd (to be compared with 400 ckkd in the first two recoil energy bins and 300 ckkd in the other bins in the previous subsections), with the same running time and systematic uncertainties as before.

Let us first discuss Scenario A. In Fig. 8, we show the expected sensitivity of the future NaI detector to some representative GNI parameters. In each panel, we compare the results of the previous subsections (red solid lines/red region) with the ones obtained with the larger statistics of Scenario A (blue dashed lines/blue contour). In the upper left panel, one can see that increasing the statistics makes it possible for the NaI detector to exclude the second minimum of the $\Delta\chi^2$ profile for the vector coupling $\varepsilon^{uV}_{\mu\mu}$ at the 2σ confidence level. Similarly, in the upper right panel, the larger statistics of Scenario A allows to further reduce the degeneracy between the vector couplings ε^{uV}_{ee} and $\varepsilon^{uV}_{\mu\mu}$, especially around $\left(\varepsilon^{uV}_{ee}, \varepsilon^{uV}_{\mu\mu}\right) \approx (-0.12, 0.2)$. This result illustrates the capability of a detector made of two nuclei with different proton to neutron ratios to lift degeneracies among NSI parameters, almost as efficiently as combinations of two different detectors, as discussed in Ref. [39] in the context of proposed experiments at the European Spallation Source. However, statistics

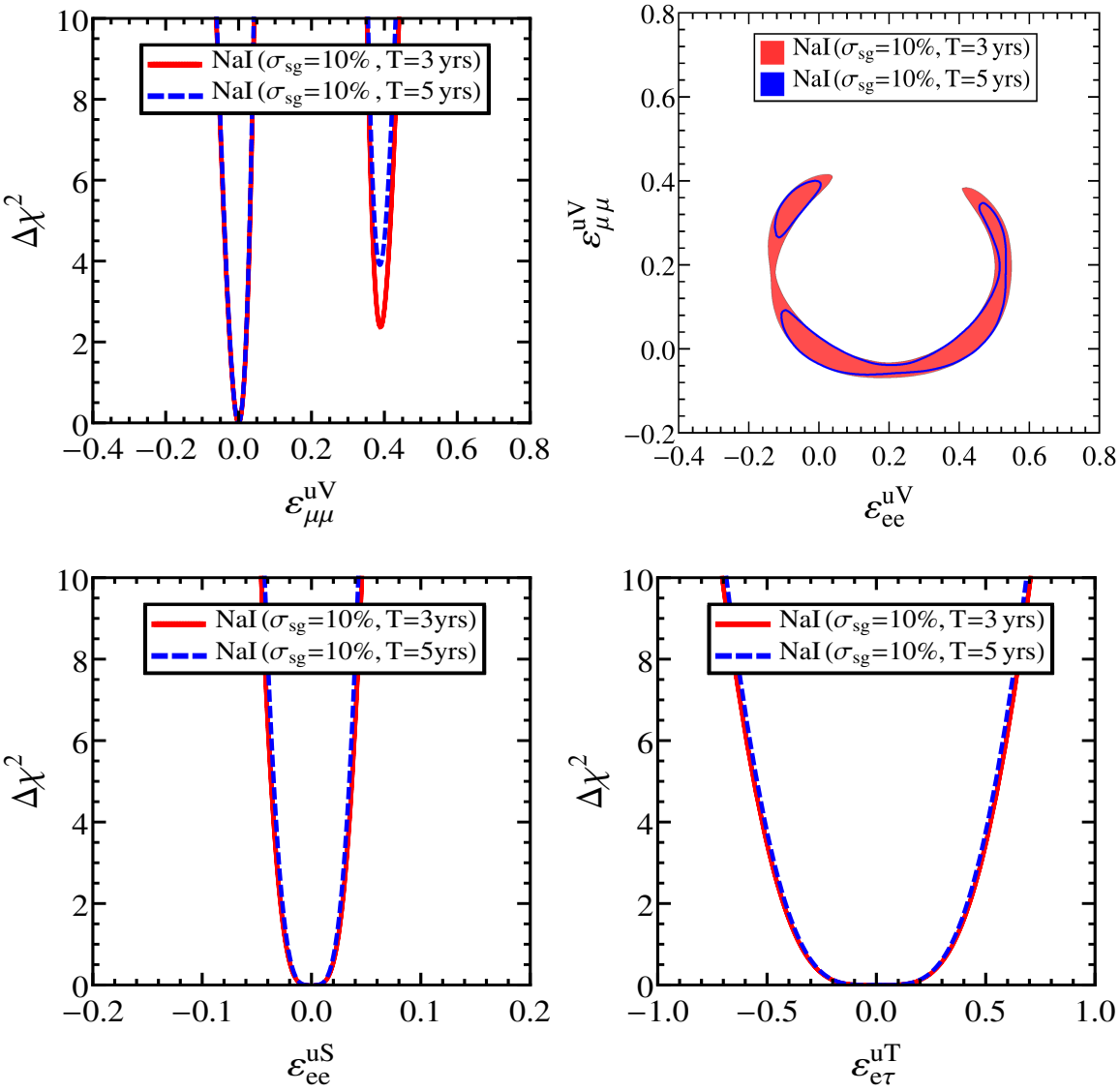


FIG. 8. Expected sensitivities of the future NaI detector to different GNI parameters, assuming 3 years (assumption of Subsections V A and V B, red solid lines/red region) or 5 years (Scenario B, blue dashed lines/blue contour) of data taking at the SNS. Upper left panel: one-dimensional $\Delta\chi^2$ profile for the vector coupling $\varepsilon^{uV}_{\mu\mu}$. Upper right panel: 90% C.L. allowed region for the pair of vector couplings (ε^{uV}_{ee}, $\varepsilon^{uV}_{\mu\mu}$). Lower panels: one-dimensional $\Delta\chi^2$ profiles for the scalar coupling ε^{uS}_{ee} (left) and tensor coupling $\varepsilon^{uT}_{e\tau}$ (right), both assumed to be real.

and (as we will see later) a good signal-over-background ratio are crucial to fully benefit from the presence of nuclei with different proton to neutron ratios in the same target material. Finally, the bottom panels of Fig. 8 show the $\Delta\chi^2$ profiles for the scalar coupling ε^{uS}_{ee} (left panel) and tensor coupling $\varepsilon^{uT}_{e\tau}$ (right panel). One can see that increasing the exposure time from 3 to 5 years does

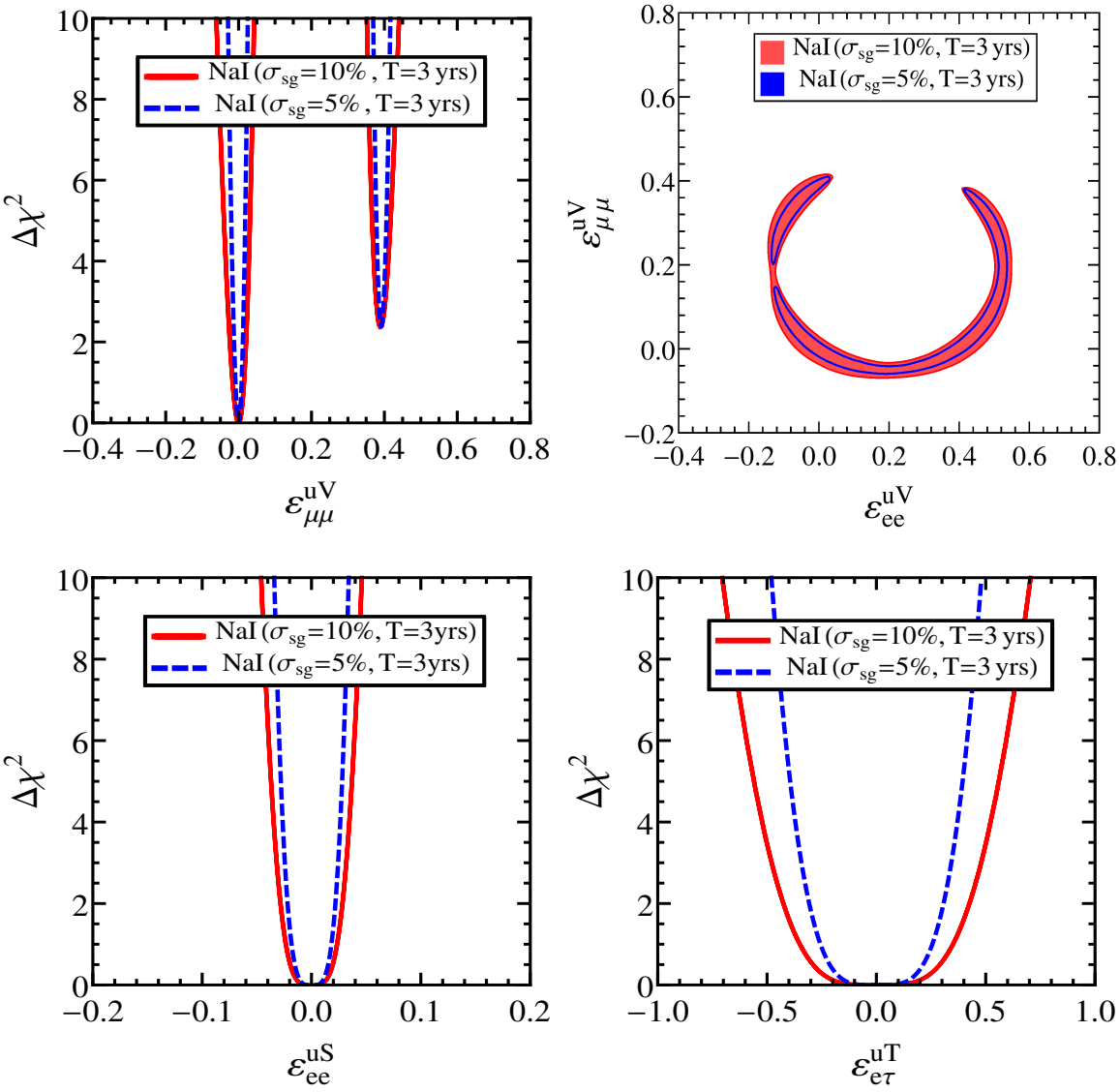


FIG. 9. Expected sensitivities of the future NaI detector to different GNI parameters, assuming a signal systematic uncertainty of 10% (assumption of Subsections V A and V B, red solid lines/red region) or 5% (Scenario B, blue dashed lines/blue contour). Upper left panel: one-dimensional $\Delta\chi^2$ profile for the vector coupling $\varepsilon^{uV}_{\mu\mu}$. Upper right panel: 90% C.L. allowed region for the pair of vector couplings (ε^{uV}_{ee}, $\varepsilon^{uV}_{\mu\mu}$). Lower panels: one-dimensional $\Delta\chi^2$ profiles for the scalar coupling ε^{uS}_{ee} (left) and tensor coupling $\varepsilon^{uT}_{e\tau}$ (right), both assumed to be real.

not result in a significant improvement of the expected sensitivity to those GNI parameters.

Let us now study the impact of the signal systematic uncertainty. In Fig. 9, we compare the results of the previous subsections (red solid lines/red region) with the ones obtained with the smaller σ_{sig} of Scenario B (blue dashed lines/blue contour), for the same GNI parameters as in Fig. 8. As can be seen, the main effect of reducing the signal systematic error is to improve the

sensitivity to GNI parameters, resulting in smaller allowed intervals or regions at a given confidence level. The impact on degeneracies is less important. In particular, the allowed region in the upper right panel is thinner for $\sigma_{sig} = 5\%$ than for $\sigma_{sig} = 10\%$ (which results in a partial breaking of the degeneray around $\left(\varepsilon_{ee}^{uV}, \varepsilon_{\mu\mu}^{uV}\right) \approx (-0.12, 0.2)$), but the region keeps roughly the same extent. Also, the second minimum of the $\Delta\chi^2$ profile for $\varepsilon_{\mu\mu}^{uV}$ (upper left panel) is excluded at the same confidence level for $\sigma_{sig} = 5\%$ and $\sigma_{sig} = 10\%$.

Finally, we consider Scenario C, which assumes a reduced background level of 100 ckkd in all recoil energy bins, as opposed to 400 ckkd in the first two bins and 300 ckkd in the other bins in Subsections V A and V B. In Fig. 10, we present the expected sensitivity of the future NaI detector to different GNI parameters, both in Scenario C and under the assumptions of Subsections V A and V B. The running time, signal and background systematic uncertainties are the same in both cases. As can be seen from the upper right panel (see also the lower left panel for a different type of degeneracy), a reduction of the background level by 66% (75% in the first two recoil energy bins) is more efficient at breaking degeneracies between NSI parameters than an increase of the running time from 3 to 5 years (Fig. 8), or a decrease of the signal systematic uncertainty by 50% (Fig. 9). The same statement holds for the second minimum of the $\Delta\chi^2$ profile for $\varepsilon_{\mu\mu}^{uV}$ (upper left panel), which is excluded at the 99% confidence level in Scenario C. In fact, such a level of background reduction would make the future NaI detector competitive with some of the best combinations of future advanced detectors at the European Spallation Source [39]. By contrast, the sensivity to tensor and scalar couplings of the NaI detector is only slightly ameliorated by a better background control (as illustrated in the right panel in the case of the parameter $\varepsilon_{e\tau}^{uT}$), while a reduction of the signal systematic error has a much stronger impact (see the lower panels of Fig. 9).

VI. CONCLUSIONS

In this work, we studied the potential of a NaI detector with characteristics similar to the one that is currently being deployed by the COHERENT collaboration to constrain scalar, vector, and tensor GNIs via CEνNS measurements at the Spallation Neutron Source. Compared with the current CsI detector, this new detector has the interesting property that its target material is composed of two nuclei with significantly different proton to neutron ratios, a feature that is known to help break degeneracies among NSI parameters. Furthermore, the future NaI detector benefits from a larger statistics than the current CsI detector, but suffers from a much less favorable signal-over-background ratio. As a consequence, the NaI detector is not expected to significantly improve

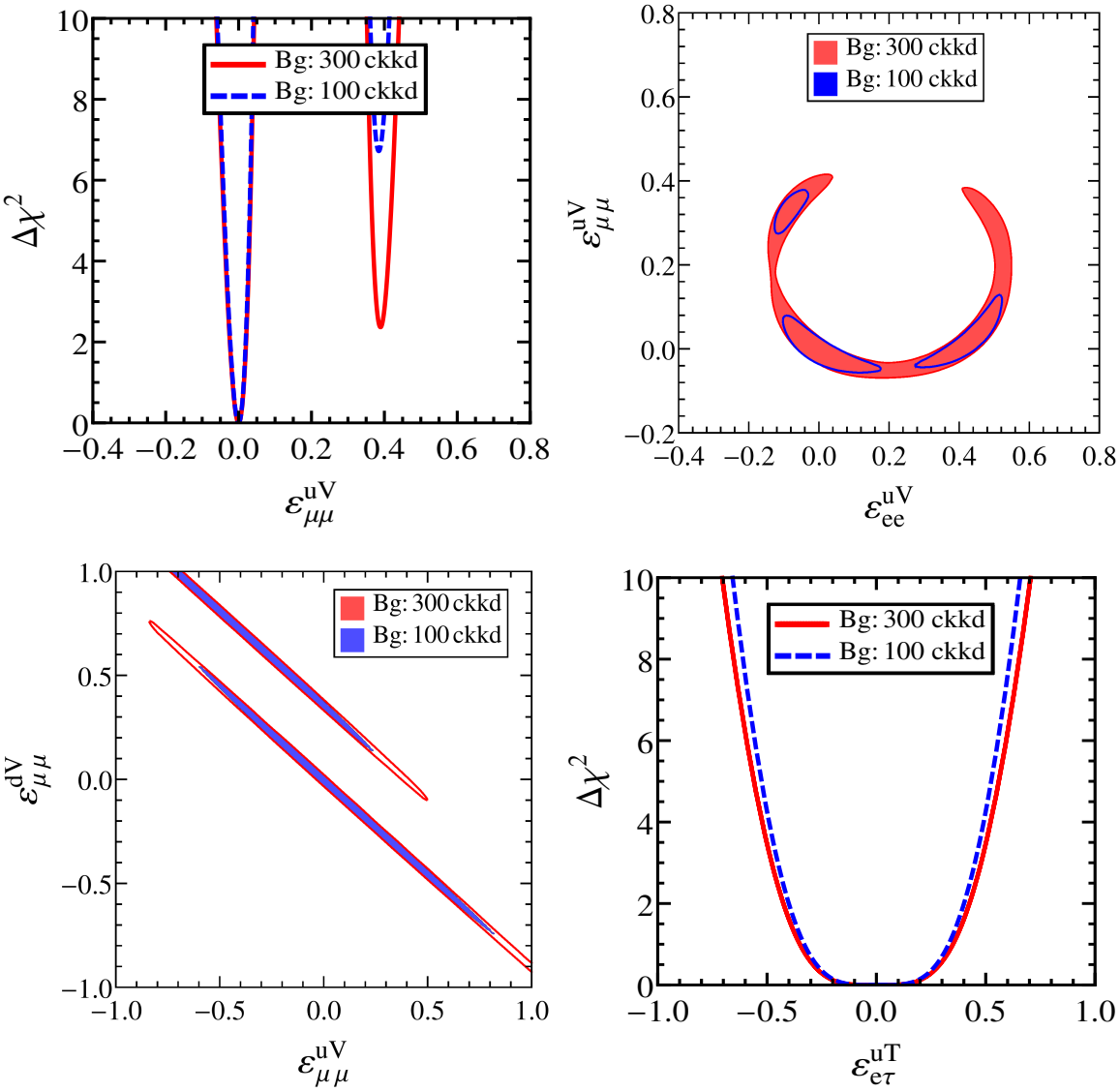


FIG. 10. Expected sensitivities of the future NaI detector to different GNI parameters, assuming the background level of Subsections V A and V B (red solid lines/red region) or 100 ckkd (Scenario C, blue dashed lines/blue contour). Upper left panel: one-dimensional $\Delta\chi^2$ profile for the vector coupling $\varepsilon^{uV}_{\mu\mu}$. Upper right panel: 90% C.L. allowed region for the pair of vector couplings (ε^{uV}_{ee}, $\varepsilon^{uV}_{\mu\mu}$). Lower left panel: 90% C.L. allowed region for the pair of vector couplings ($\varepsilon^{uV}_{\mu\mu}$, $\varepsilon^{dV}_{\mu\mu}$). Lower right panel: one-dimensional $\Delta\chi^2$ profile for the tensor coupling $\varepsilon^{uT}_{e\tau}$, assumed to be real.

the current CsI bounds on individual GNI parameters, except for diagonal vector couplings, whose contribution interfere with the Standard Model one in the CEνNS cross section. The NaI detector is also able to partially break degeneracies between two diagonal vector couplings, in a way that is complementary to the current CsI detector. These results are based on the technical details about

the NaI detector given in Refs. [53, 55] and on the background studies of Ref. [53], as well as on our assumptions about signal and background systematic uncertainties.

We also studied the impact of statistics, signal systematic uncertainty and background level on the expected sensitivity of the future NaI detector to GNIs. We found that reducing the signal systematic uncertainty from 10% to 5% significantly improves the sensitivity to all GNI parameters, while increasing the running time from three to five years makes it possible to further break degeneracies among diagonal NSIs, but has little impact on the sensitivity to GNI parameters whose contribution to CEνNS does not interfere with the SM contribution. Our results also show that a good background control is crucial to fully benefit from the different proton to neutron ratios of the Sodium and Iodine targets. More specifically, a background level of 100 counts per keV$_{ee}$ per kg per day (ckkd), if achievable, would make the NaI detector considered in this work very efficient at breaking degeneracies among NSI parameters, already after three years of data taking.

ACKNOWLEDGMENTS

The work of S.S.C. is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 510963981. S. S. C. also acknowledges financial support from the LabEx P2IO (ANR-10-LABX-0038 - Project "BSMNu") in the framework of the "Investissements d'Avenir" (ANR-11-IDEX-0003-01) managed by the Agence Nationale de la Recherche (ANR), France. This work has been supported by the Spanish grants PID2020-113775GB-I00 (MCIN/AEI/10.13039/501100011033) and CIPROM/2021/054 (Generalitat Valenciana). G.S.G. acknowledges financial support by the CIAPOS/2022/254 grant funded by Generalitat Valenciana. The work of SL is supported in part by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 860881-HIDDeN. O. G. M. was supported by the CONAHCyT grant 23238 and by SNII-Mexico.

[1] D. Z. Freedman, "Coherent effects of a weak neutral current," Phys. Rev. D **9** (Mar, 1974) 1389–1392. https://link.aps.org/doi/10.1103/PhysRevD.9.1389.

[2] **COHERENT** Collaboration, D. Akimov et al., "Observation of Coherent Elastic Neutrino-Nucleus Scattering," Science **357** no. 6356, (2017) 1123–1126, arXiv:1708.01294 [nucl-ex].

[3] D. Akimov et al., "Measurement of the Coherent Elastic Neutrino-Nucleus Scattering Cross Section on CsI by COHERENT," arXiv:2110.07730 [hep-ex].

[4] **COHERENT** Collaboration, D. Akimov et al., "First Measurement of Coherent Elastic Neutrino-Nucleus Scattering on Argon," Phys. Rev. Lett. **126** no. 1, (2021) 012002, arXiv:2003.10630 [nucl-ex].

[5] M. Cadeddu, N. Cargioli, F. Dordei, C. Giunti, Y. F. Li, E. Picciau, C. A. Ternes, and Y. Y. Zhang, "New insights into nuclear physics and weak mixing angle using electroweak probes," Phys. Rev. C **104** no. 6, (2021) 065502, arXiv:2102.06153 [hep-ph].

[6] M. Cadeddu, C. Giunti, Y. Li, and Y. Zhang, "Average CsI neutron density distribution from COHERENT data," Phys. Rev. Lett. **120** no. 7, (2018) 072501, arXiv:1710.02730 [hep-ph].

[7] O. Miranda, D. Papoulias, G. Sanchez Garcia, O. Sanders, M. Tórtola, and J. Valle, "Implications of the first detection of coherent elastic neutrino-nucleus scattering (CEvNS) with Liquid Argon," JHEP **05** (2020) 130, arXiv:2003.12050 [hep-ph].

[8] J. Barranco, O. G. Miranda, and T. I. Rashba, "Probing new physics with coherent neutrino scattering off nuclei," JHEP **12** (2005) 021, arXiv:hep-ph/0508299.

[9] K. Scholberg, "Prospects for measuring coherent neutrino-nucleus elastic scattering at a stopped-pion neutrino source," Phys. Rev. D **73** (2006) 033005, arXiv:hep-ex/0511042.

[10] V. De Romeri, O. G. Miranda, D. K. Papoulias, G. Sanchez Garcia, M. Tórtola, and J. W. F. Valle, "Physics implications of a combined analysis of COHERENT CsI and LAr data," JHEP **04** (2023) 035, arXiv:2211.11905 [hep-ph].

[11] V. Bresó-Pla, A. Falkowski, M. González-Alonso, and K. Monsálvez-Pozo, "EFT analysis of New Physics at COHERENT," JHEP **05** (2023) 074, arXiv:2301.07036 [hep-ph].

[12] D. Aristizabal Sierra, V. De Romeri, and N. Rojas, "COHERENT analysis of neutrino generalized interactions," Phys. Rev. D **98** (2018) 075018, arXiv:1806.07424 [hep-ph].

[13] L. J. Flores, N. Nath, and E. Peinado, "Non-standard neutrino interactions in U(1)' model after COHERENT data," JHEP **06** (2020) 045, arXiv:2002.12342 [hep-ph].

[14] T. Han, J. Liao, H. Liu, and D. Marfatia, "Scalar and tensor neutrino interactions," JHEP **07** (2020) 207, arXiv:2004.13869 [hep-ph].

[15] M. Cadeddu, F. Dordei, C. Giunti, Y. F. Li, and Y. Y. Zhang, "Neutrino, electroweak, and nuclear physics from COHERENT elastic neutrino-nucleus scattering with refined quenching factor," Phys. Rev. D **101** no. 3, (2020) 033004, arXiv:1908.06045 [hep-ph].

[16] O. G. Miranda, D. K. Papoulias, M. Tórtola, and J. W. F. Valle, "Probing neutrino transition magnetic moments with coherent elastic neutrino-nucleus scattering," JHEP **07** (2019) 103, arXiv:1905.03750 [hep-ph].

[17] M. Cadeddu, F. Dordei, C. Giunti, Y. F. Li, E. Picciau, and Y. Y. Zhang, "Physics results from the first COHERENT observation of coherent elastic neutrino-nucleus scattering in argon and their combination with cesium-iodide data," Phys. Rev. D **102** no. 1, (2020) 015030, arXiv:2005.01645 [hep-ph].

[18] Y. Farzan, M. Lindner, W. Rodejohann, and X.-J. Xu, "Probing neutrino coupling to a light scalar with coherent neutrino scattering," JHEP **05** (2018) 066, arXiv:1802.05171 [hep-ph].

[19] L. M. G. de la Vega, L. J. Flores, N. Nath, and E. Peinado, "Complementarity between dark matter direct searches and CEνNS experiments in U(1)' models," JHEP **09** (2021) 146, arXiv:2107.04037 [hep-ph].

[20] E. Bertuzzo, G. Grilli di Cortona, and L. M. D. Ramos, "Probing light vector mediators with coherent scattering at future facilities," JHEP **06** (2022) 075, arXiv:2112.04020 [hep-ph].

[21] B. Dutta, Y. Gao, R. Mahapatra, N. Mirabolfathi, L. E. Strigari, and J. W. Walker, "Sensitivity to oscillation with a sterile fourth generation neutrino from ultra-low threshold neutrino-nucleus coherent scattering," Phys. Rev. D **94** no. 9, (2016) 093002, arXiv:1511.02834 [hep-ph].

[22] T. S. Kosmas, D. K. Papoulias, M. Tortola, and J. W. F. Valle, "Probing light sterile neutrino signatures at reactor and Spallation Neutron Source neutrino experiments," Phys. Rev. D **96** no. 6, (2017) 063013, arXiv:1703.00054 [hep-ph].

[23] D. Akimov et al., "The COHERENT Experimental Program," in Snowmass 2021. 4, 2022. arXiv:2204.04575 [hep-ex].

[24] D. Baxter et al., "Coherent Elastic Neutrino-Nucleus Scattering at the European Spallation Source," JHEP **02** (2020) 123, arXiv:1911.00762 [physics.ins-det].

[25] **CCM** Collaboration, A. A. Aguilar-Arevalo et al., "First Dark Matter Search Results From Coherent CAPTAIN-Mills," arXiv:2105.14020 [hep-ex].

[26] **CONUS** Collaboration, H. Bonet et al., "Novel constraints on neutrino physics beyond the standard model from the CONUS experiment," JHEP **05** (2022) 085, arXiv:2110.02174 [hep-ph].

[27] **CONNIE** Collaboration, A. Aguilar-Arevalo et al., "Results of the Engineering Run of the Coherent Neutrino Nucleus Interaction Experiment (CONNIE)," JINST **11** no. 07, (2016) P07024, arXiv:1604.01343 [physics.ins-det].

[28] **νGeN** Collaboration, I. Alekseev et al., "First results of the νGeN experiment on coherent elastic neutrino-nucleus scattering," Phys. Rev. D **106** no. 5, (2022) L051101, arXiv:2205.04305 [nucl-ex].

[29] D. Y. Akimov, I. Alexandrov, R. Alyev, V. Belov, A. Bolozdynya, A. Etenko, A. Galavanov, E. Glagovsky, Y. Gusakov, A. Khromov, et al., "The red-100 experiment," Journal of Instrumentation **17** no. 11, (2022) T11011.

[30] B. C. Canas, E. A. Garces, O. G. Miranda, A. Parada, and G. Sanchez Garcia, "Interplay between nonstandard and nuclear constraints in coherent elastic neutrino-nucleus scattering experiments," Phys. Rev. D **101** no. 3, (2020) 035012, arXiv:1911.09831 [hep-ph].

[31] R. R. Rossi, G. Sanchez Garcia, and M. Tórtola, "Probing nuclear properties and neutrino physics with current and future CEνNS experiments," arXiv:2311.17168 [hep-ph].

[32] M. Lindner, W. Rodejohann, and X.-J. Xu, "Coherent Neutrino-Nucleus Scattering and new Neutrino Interactions," JHEP **03** (2017) 097, arXiv:1612.04150 [hep-ph].

[33] T. Ohlsson, "Status of non-standard neutrino interactions," Rept. Prog. Phys. **76** (2013) 044201, arXiv:1209.2710 [hep-ph].

[34] O. G. Miranda and H. Nunokawa, "Non standard neutrino interactions: current status and future prospects," New J. Phys. **17** no. 9, (2015) 095002, arXiv:1505.06254 [hep-ph].

[35] Y. Farzan and M. Tortola, "Neutrino oscillations and Non-Standard Interactions," Front. in Phys. **6** (2018) 10, arXiv:1710.09360 [hep-ph].

[36] O. G. Miranda, M. A. Tortola, and J. W. F. Valle, "Are solar neutrino oscillations robust?," JHEP **10** (2006) 008, arXiv:hep-ph/0406280.

[37] P. Coloma and T. Schwetz, "Generalized mass ordering degeneracy in neutrino oscillation experiments," Phys. Rev. D **94** no. 5, (2016) 055005, arXiv:1604.05772 [hep-ph]. [Erratum: Phys.Rev.D 95, 079903 (2017)].

[38] A. Galindo-Uribarri, O. G. Miranda, and G. S. Garcia, "Novel approach for the study of coherent elastic neutrino-nucleus scattering," Phys. Rev. D **105** no. 3, (2022) 033001, arXiv:2011.10230 [hep-ph].

[39] S. S. Chatterjee, S. Lavignac, O. G. Miranda, and G. Sanchez Garcia, "Constraining nonstandard interactions with coherent elastic neutrino-nucleus scattering at the European Spallation Source," Phys. Rev. D **107** no. 5, (2023) 055019, arXiv:2208.11771 [hep-ph].

[40] O. Tomalak, P. Machado, V. Pandey, and R. Plestid, "Flavor-dependent radiative corrections in coherent elastic neutrino-nucleus scattering," JHEP **02** (2021) 097, arXiv:2011.05960 [hep-ph].

[41] J. D. Lewin and P. F. Smith, "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil," Astropart. Phys. **6** (1996) 87–112.

[42] J. Erler and R. Ferro-Hernández, "Weak Mixing Angle in the Thomson Limit," JHEP **03** (2018) 196, arXiv:1712.09146 [hep-ph].

[43] **Particle Data Group** Collaboration, R. L. Workman et al., "Review of Particle Physics," PTEP **2022** (2022) 083C01.

[44] S. R. Klein and J. Nystrand, "Exclusive vector meson production in relativistic heavy ion collisions," Phys. Rev. C **60** (Jun, 1999) 014903. https://link.aps.org/doi/10.1103/PhysRevC.60.014903.

[45] C. Giunti, "General COHERENT constraints on neutrino nonstandard interactions," Phys. Rev. D **101** no. 3, (2020) 035039, arXiv:1909.00466 [hep-ph].

[46] O. G. Miranda, G. Sanchez Garcia, and O. Sanders, "Coherent elastic neutrino-nucleus scattering as a precision test for the Standard Model and beyond: the COHERENT proposal case," Adv. High Energy Phys. **2019** (2019) 3902819, arXiv:1902.09036 [hep-ph].

[47] E. E. Jenkins, A. V. Manohar, and P. Stoffer, "Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching," JHEP **03** (2018) 016, arXiv:1709.04486 [hep-ph].

[48] I. Bischer and W. Rodejohann, "General Neutrino Interactions at the DUNE Near Detector," Phys. Rev. D **99** no. 3, (2019) 036006, arXiv:1810.02220 [hep-ph].

[49] M. Hoferichter, J. Ruiz de Elvira, B. Kubis, and U.-G. Meißner, "High-Precision Determination of the Pion-Nucleon σ Term from Roy-Steiner Equations," Phys. Rev. Lett. **115** (2015) 092301, arXiv:1506.04142 [hep-ph].

[50] T. Bhattacharya, V. Cirigliano, S. Cohen, R. Gupta, H.-W. Lin, and B. Yoon, "Axial, Scalar and Tensor Charges of the Nucleon from 2+1+1-flavor Lattice QCD," Phys. Rev. D **94** no. 5, (2016) 054508, arXiv:1606.07049 [hep-lat].

[51] L. J. Flores, N. Nath, and E. Peinado, "CEνNS as a probe of flavored generalized neutrino interactions," Phys. Rev. D **105** no. 5, (2022) 055010, arXiv:2112.05103 [hep-ph].

[52] A. Majumdar, D. K. Papoulias, R. Srivastava, and J. W. F. Valle, "Physics implications of recent Dresden-II reactor data," Phys. Rev. D **106** no. 9, (2022) 093010, arXiv:2208.13262 [hep-ph].

[53] S. C. Hedges, Low Energy Neutrino-Nucleus Interactions at the Spallation Neutron Source. PhD thesis, Duke U., 2021 (https://hdl.handle.net/10161/24413).

[54] E. Picciau, Low-energy signatures in DarkSide-50 experiment and neutrino scattering processes. PhD thesis, Cagliari U., 2022.

[55] S. C. Hedges, "The NaI[Tl] COHERENT Detector,". Talk at the Magnificent CEνNS Workshop 2018 (doi:10.5281/zenodo.3462345).

LA-UR-24-21703

Froggatt-Nielsen Meets the SMEFT

Eetu Loisa[a] and Jim Talbert[a,b]

[a] *DAMTP, University of Cambridge, Wilberforce Rd., Cambridge, CB3 0WA, United Kingdom*
[b] *Theoretical Division, Group T-2, MS B283, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545, USA*

E-mail: eal47@cam.ac.uk, rjt89@cantab.ac.uk

arXiv:2402.16940v1 [hep-ph] 26 Feb 2024

Abstract

We study the matching of Froggatt-Nielsen theories of flavour onto the Standard Model Effective Field Theory (SMEFT), upon integrating out a heavy Beyond-the-Standard-Model (BSM) scalar 'flavon' whose vacuum expectation value breaks an Abelian flavour symmetry at energies $\Lambda_{\rm FN}$ well above the electroweak scale, $\Lambda_{\rm FN} > \Lambda_{\rm SM}$. We include matching contributions to the infrared $d_{\rm SM} = 6$ (Warsaw basis) SMEFT sourced from ultraviolet contact terms suppressed up to order $1/\Lambda_{\rm UV}^2$ in the Froggatt-Nielsen Lagrangian, where $\Lambda_{\rm UV} > \Lambda_{\rm FN}$ is an arbitrary deep-ultraviolet scale where further unspecified BSM particles are dynamical. This includes tree-level (one-loop) ultraviolet diagrams with $d_{\rm FN} = 6$ (5) effective vertices. We first do so with a toy model, but then generalize our findings to arbitrary Frogatt-Nielsen charges. Our results indicate a rich and non-trivial signature of Froggatt-Nielsen theories on the (otherwise) model-independent operators of the SMEFT, and we briefly speculate on extending our analysis to broader classes of BSM flavour models, e.g. non-Abelian and/or gauged theories. We thus take an important step towards determining how to use rapidly developing theoretical and experimental SMEFT technologies to gain unambiguous insight into the SM's longstanding fermion flavour puzzle.

Contents

1 Introduction **1**

2 Matching the Froggatt-Nielsen Mechanism **4**
2.1 A Toy Model for the Down Quarks 7
2.1.1 Dimension Four ($d_{\rm FN} = 4$) 8
2.1.2 Dimension Five ($d_{\rm FN} = 5$) 13
2.1.3 Dimension Six ($d_{\rm FN} = 6$) 17
2.1.4 Matching Summary and Phenomenological Outlook 24

3 Generalizations and Extensions Beyond the Toy Model **26**
3.1 Charge Generalizations and Extensions to Different Families 27
3.2 Non-Abelian and/or Continuous Flavour Models 35

4 Summary and Outlook **40**

1 Introduction

If not the oldest, the Froggatt-Nielsen (FN) mechanism [1] is arguably the most famous attempt to dynamically resolve the Standard Model's (SM) *flavour puzzle*, i.e. to explain the otherwise arbitrary (and often dramatic) hierarchies of fermionic mass, mixing, and CP-violation inferred from decades of experiment. Allowing for simple generalizations of the original FN idea, the ingredients necessary for solving such a longstanding puzzle of the SM are strikingly simple: a new symmetry (gauged or global, Abelian or non-Abelian, continuous or discrete), a new scalar sector to break it, and a sufficiently high mass scale to suppress entries in the SM's Yukawa couplings.

Indeed, the FN mechanism[1] is most transparent at intermediate energy scales, where we can discuss the Beyond-the-SM (BSM) dynamics most relevant to the flavour puzzle without reference to fully ultraviolet (UV)-complete Lagrangians. Namely, when the Higgs boson is non-trivially charged under an Abelian (continuous or discrete) BSM flavour symmetry $\mathcal{S}_{\rm FN}$, the SM's renormalizable Yukawa couplings are, a priori, disallowed for arbitrary flavours. Non-renormalizable corrections to the Yukawa sector can appear, however, given successive insertions of a novel scalar flavon θ, itself non-trivially charged under $\mathcal{S}_{\rm FN}$:

$$\mathcal{L} \supset y_{ij}\,\overline{\psi}_i\,H\,\psi_j \longrightarrow \mathcal{L} \supset c_{ij}\,\overline{\psi}_i\,H\,\psi_j\left(\frac{\theta}{\Lambda_{\rm UV}}\right)^{x_{ij}} \quad \text{with} \quad x_{ij} = -\frac{(q_H + q_{\overline{\psi}_i} + q_{\psi_j})}{q_\theta}\,, \tag{1}$$

where ψ are SM fermions, q is the respective $\mathcal{S}_{\rm FN}$ charge of a given field, x_{ij} represents the power suppression necessary to achieve $\mathcal{S}_{\rm FN}$ invariance, c_{ij} are generally considered to be $\mathcal{O}(1)$ parameters, and $\Lambda_{\rm UV}$ represents an arbitrary UV mass scale where heavier degrees of freedom propagate. For example, heavy vector-like fermions can 'complete' the effective interaction on the right-hand-side (RHS) of (1), yielding a fully renormalizable Lagrangian, although

[1] ...or, rather, what we actually mean when we refer to said mechanism!

we wish to remain agnostic about such (likely unfalsiable) UV-completions. Regardless, well below $\Lambda_{\rm UV}$, the scalar flavon θ can obtain a vacuum expectation value (vev) which breaks the $\mathcal{S}_{\rm FN}$ symmetry of the effective Lagrangian, such that $\langle\theta\rangle/\Lambda_{UV} \equiv \lambda \lesssim 0.1$. Then, upon the Higgs itself acquiring a vev in the electroweak phase transition, $\langle H^\dagger H\rangle \equiv v_H^2/2$, the SM fermions obtain masses,

$$M^\psi_{ij} \simeq c_{ij} \cdot \frac{v_H}{\sqrt{2}} \cdot \lambda^{x_{ij}} \,, \tag{2}$$

where we use a capitalized M to indicate that the mass matrix is in a interaction basis, and not the physical mass-eigenstate basis, and where $c_{ij} \to y_{ij}$ for matrix elements allowed at the renormalizable level ($x_{ij} = 0$). In this way, hierarchical patterns of mass and mixing can be built up via hierarchies in powers of λ, in accord with observation, as opposed to unnatural hierarchies in Lagrangian parameters.

Unfortunately, however, FN-type models of flavour have not convincingly predicted the complete structure of the deeply hierarchical SM quark mass spectrum, including the well-constrained [2] Cabibbo-Kobayashi-Maskawa (CKM) matrix, much less the mass scales and mixing patterns that analogously appear in the lepton sector given non-zero neutrino masses (see e.g. [3]). Indeed, given that only the leptonic Dirac CP-violating phase, further phases associated to the potential Majorana nature of neutrinos, and the exact values of the neutrino-mass eigenvalues (and hence also their ordering) remain largely unresolved in the SM's pure mass and mixing sector, most predictions coming from new FN-type models should actually be considered *retrodictions*. Furthermore, the mass scale associated with θ is often assumed to be very heavy for simplicity, such that any additional phenomenology associated to θ-dependent vertices at (e.g.) high-energy colliders and/or precision flavour factories is often limited (purposely overlooked) in the literature — for recent analyses of FN-type phenomenology that extend beyond mass and mixing predictions, see e.g. [4–6].

In short, the falsifiability of FN-type models (as typically studied) is becoming increasingly challenged. Given this somewhat uninspiring context, a natural question arises:

> *Is it possible to probe the UV dynamics of FN-type theories in a largely model-independent manner, thereby exposing any universal infrared (IR) predictions that can be efficiently falsified by experiment?* ($\triangle$)

One is immediately led to effective field theory (EFT) considerations, and in particular to the SMEFT [7, 8] (see [9, 10] for reviews), for an answer. The SMEFT is the theory including all renormalizable and non-renormalizable interactions of the SM field content, invariant under spacetime and linearly-realized SM gauge symmetries $\mathcal{G}_{\rm SM} = {\rm SU(3)_c} \times {\rm SU(2)_L} \times {\rm U(1)_Y}$, assuming new BSM physics is parametrically heavier than characteristic SM mass scales:[2]

$$\mathcal{L}_{\rm SMEFT} \equiv \mathcal{L}_{\rm SM} + \sum_i \frac{C_i^{(d)}}{\Lambda^{d-4}} \mathcal{Q}_i^{(d)} \,. \tag{3}$$

Here C_i are arbitrary Wilson coefficients that encode the impact of UV dynamics on local, IR contact interactions $\mathcal{Q}_i$ of mass-dimension d, suppressed by appropriate powers of a BSM mass scale $\Lambda > \Lambda_{\rm SM}$, above which the unspecified UV dynamics can propagate. The counting of operators at arbitrary mass dimensions d is by now greatly simplified via Hilbert Series

[2] We have suppressed flavour labels in (3).

methods [11], while complete and non-redundant bases up to $d \leq 6$ are well-utilized in the literature — the lone operator appearing at $d = 5$ is the Weinberg operator [12] giving a non-zero Majorana neutrino mass term, while a plethora of operators already appear at $d = 6$ [8]. Critically, (3) parameterizes model-dependent BSM physics into a single, largely model-independent IR Lagrangian; constraining C_i experimentally amounts to constraining every BSM model that 'matches' to $\mathcal{L}_{\text{SMEFT}}$ (with specific predictions for C_i). Whilst technically challenging, global fits to $C_i^{(6)}$ are now becoming available in the combined top, Higgs, and electroweak sector [13–16], while fits addressing embedded flavour assumptions in particular observables are also appearing [17–27].

In an attempt to address ($\triangle$), we aim to 'match' the FN mechanism distilled in (1) to the SMEFT in (3). That is, we will determine which operators $\mathcal{Q}_i$ have generically non-zero Wilson coefficients C_i upon integrating the flavon field θ out of the UV Lagrangian. We can therefore identify $\Lambda \sim \Lambda_{\text{FN}}$ in (3), and hence our job is to calculate C_i explicitly. We will do so up to $d \equiv d_{\text{SM}} = 6$ in (3) (assuming the Warsaw basis of [8]), including contributions from operators up to $d_{\text{FN}} \equiv 4 + x_{ij} = 6$ in the intermediate-energy FN EFT $\mathcal{L}_{\text{FN}}$ introduced in (4) below. Pursuing matching at $d_{\text{FN}} = 6$ has also required enumerating a minimal, non-redundant basis for the FN EFT at this order, including SM and flavon degrees of freedom.

We will use analytic functional techniques (cf. the recent presentations in (e.g.) [28, 29]) when matching tree-level FN interactions to the SMEFT, and the automated functional techniques embedded in `Matchete` [30] when considering one-loop UV contributions (many of which we will also check/intuit diagramatically by hand). In general, we find that a host of SMEFT operators appear already at $d_{\text{SM}} = 6$, including Higgs-scalar (H^6), Higgs-derivative (H^4D^2), Higgs-gauge (X^2H^2), four-fermion (ψ^4), Higgs-kinetic (ψ^2H^2D), and Higgs-enhanced Yukawa ($H^3\psi^2$) operators. These effects are driven both by the $\mathcal{S}_{\text{FN}}$ charge structure of the FN theory, giving distinct IR SMEFT flavour phenomenology, but also by flavour-*independent* effects, including Higgs-flavon mixing. Our results are given explicitly in Tables 2–8.

The IR signatures of FN-type theories are therefore quite distinct, and hence we address how one might use the SMEFT to unambiguously probe theories aiming to dynamically resolve the flavour puzzle. Simply put, phenomenological analyses that constrain the SMEFT operators we have identified simultaneously constrain the FN-type theories that ubiquitously turn them on. Our work therefore adds to a growing list of theoretical studies addressing what one can learn about UV BSM flavour physics from the flavour structure of the IR SMEFT [31–47].

The paper develops as follows: in Section 2 we introduce a toy model of (1), review the functional matching techniques/tools we will use, and then perform a tree- and loop-level matching to the SMEFT. As mentioned above, we also present a complete and minimal $d_{\text{FN}} = 6$ EFT basis for the FN theory in this Section, before also generalizing our matching to arbitrary toy models of this type when presenting final results in Tables 2–8. Then, in Section 3, we briefly discuss a number of generalizations to the naive toy setup of Section 2, including considerations of non-Abelian flavour models, as well as continuous (global or gauged) constructions. Finally, we provide a summary and outlook in Section 4.

2 Matching the Froggatt-Nielsen Mechanism

We begin with a generic FN effective Lagrangian valid at scales $\Lambda_{\rm FN} > \Lambda_{\rm SM}$, which is composed of both renormalizable and non-renormalizable interactions furnished by the SM field content and the novel flavon field θ, which we take to be a complex scalar charged under $\mathcal{S}_{\rm FN}$ but transforming as a singlet of the SM gauge group. We can decompose the resulting $\mathcal{G}_{\rm SM} \times \mathcal{S}_{\rm FN}$ symmetric Lagrangian as:

$$\mathcal{L}_{\rm FN} \equiv \mathcal{L}_{\rm SM'} + \mathcal{L}^4_\theta + \sum_{d>4}^{\infty} \mathcal{L}^d_{\rm FN}\,, \tag{4}$$

where $\mathcal{L}^4_\theta$ is the set of θ-dependent renormalizable operators,[3]

$$\mathcal{L}^4_\theta = -\theta^*\Box\theta + \mu^2_\theta\theta^*\theta - \lambda_{02}(\theta^*\theta)^2 - \lambda_{11}\big(H^\dagger H\big)(\theta^*\theta), \tag{5}$$

$\mathcal{L}_{\rm SM'}$ is the $\mathcal{S}_{\rm FN}$-symmetric SM, and $\mathcal{L}^d_{\rm FN}$ is the set of effective, non-renormalizable operators that parameterize further BSM effects appearing in the deep UV. At this stage, we wish to remain agnostic about any specific UV completions of the effective FN Lagrangian, and thus allow all operators invariant under $\mathcal{G}_{\rm SM} \times \mathcal{S}_{\rm FN}$ to be turned on in (4). Hence the non-renormalizable terms in the equation can be further decomposed into $\mathcal{S}_{\rm FN}$-symmetric SMEFT operators (that is, terms independent of θ) and operators with θ-dependence,

$$\mathcal{L}^{d>4}_{\rm FN} \equiv \mathcal{L}^{d>4}_{\rm SMEFT'} + \mathcal{L}^{d>4}_\theta\,. \tag{6}$$

Hence one critical task for our study is to enumerate the minimal non-redundant operator basis for $\mathcal{L}^{d>4}_{\rm UV}$ order-by-order, i.e. in a way that is consistent with the output of a Hilbert series calculation. Thankfully this task is somewhat simplified given our assumptions. For example, we assume that both H and θ are non-trivially charged under $\mathcal{S}_{\rm FN}$ and that θ is a $\mathcal{G}_{\rm SM}$-singlet. In this case it is easy to deduce that there are no odd-mass-dimension terms in the effective scalar potential, for instance. We will return to the issue of operator bases in Section 2.1.3 below.

Given (4), we must adopt a prescription for integrating θ out of the dynamical action. We will do so with functional matching techniques, where the one-light-particle-irreducible (1LPI) effective action of the UV theory is matched with that of the IR EFT. At tree-level:

$$\mathcal{L}^{\rm tree}_{\rm EFT}[\phi] = \mathcal{L}^{\rm tree}_{\rm UV}[\Phi,\phi]|_{\Phi=\Phi_c[\phi]}\,, \tag{7}$$

where Φ collectively labels the heavy fields, ϕ are the light SM fields and $\Phi_c[\phi]$ means the solutions to the equations of motion (EOM) of the heavy fields in terms of the light fields (see [29] for a review). At tree-level, functional matching thus reduces to a simple application of the Euler-Lagrange equation,

$$\frac{\delta\mathcal{L}}{\delta\theta^{(\star)}} = \partial_\mu \frac{\delta\mathcal{L}}{\delta(\partial_\mu\theta^{(\star)})}\,, \tag{8}$$

yielding the classical EOM for the heavy dynamical fields (the flavon θ in our case). To obtain an IR EFT Lagrangian up to a fixed operator dimension, one solves the EOM (8) iteratively

[3]Here we have anticipated working in the broken $\mathcal{S}_{\rm FN}$ phase.

by expanding the solution $\theta_c[\phi]$ as a series organized in inverse powers of m_θ^2, with each term written in terms of the SM fields ϕ assumed to be light compared to θ. Substituting the classical power series solution back into the UV Lagrangian and truncating at the desired operator order yields the tree-level-matched EFT.

We apply (8) to (4) in the broken $\mathcal{S}_{\rm FN}$ phase, in order to match to the SMEFT where $\mathcal{S}_{\rm FN}$ is no longer a linearly-realized symmetry. That is, at scales far above the electroweak scale, but below the $\mathcal{S}_{\rm FN}$ breaking scale, we may expand the theory about its classical vacuum,

$$\langle H^\dagger H\rangle = 0, \quad \langle \theta^* \theta\rangle = \frac{v_\theta^2}{2} = \frac{\mu_\theta^2}{2\lambda_{02}}\,, \tag{9}$$

by writing

$$H \to H, \quad \theta \to \frac{v_\theta + \vartheta + i\pi_\theta}{\sqrt{2}}\,, \tag{10}$$

where ϑ is the new, massive CP-even scalar state and π_θ is a potential CP-odd Nambu-Goldstone boson associated with the $\mathcal{S}_{\rm FN}$ symmetry breaking.

At this point we must further discuss the group properties of $\mathcal{S}_{\rm FN}$. The classic scenario is to take $\mathcal{S}_{\rm FN} \sim {\rm U}(1)_{\rm FN}$, i.e. a continuous Abelian symmetry which, a priori, can either be global or gauged. While the former choice implies that π_θ is classically massless, one might expect it to acquire a small mass through radiative effects, e.g. mixed ${\rm U}(1)_{\rm FN}$-gauge-gauge anomalies.[4] As such, matching to the SMEFT may be inappropriate when considering a global symmetry-breaking scenario, given the expectation that π_θ will be light (cf. Section 3.2). In the latter case, where ${\rm U}(1)_{\rm FN}$ is gauged, π_θ can be absorbed into the longitudinal component of the family non-universal ${\rm U}(1)_{\rm FN}$ gauge boson, whose mass will be of the order of the symmetry breaking scale v_θ. This scenario necessitates the fermionic charges to be fixed such that all gauge anomalies cancel (see e.g. [4,50]), unless anomaly cancellation is assumed to occur through the Green-Schwartz mechanism [51], and also warrants further study regarding the SMEFT matching effects sourced from integrating out the presumably heavy BSM gauge boson alongside the flavon.

On the other hand, the dominant flavour mechanism of interest in (1) (ultimately responsible for solving the flavour puzzle) can just as easily be realized in a *discrete* Abelian scenario.[5] In this case we can neglect the subtleties surrounding π_θ while preserving the SMEFT matching results associated with the physics of the vev v_θ and the real component ϑ of the flavon field, thereby exposing the most pertinent IR signatures of the FN mechanism. Hence, for the remainder of the section, we will effectively take

$$\mathcal{S}_{\rm FN} \sim \mathbb{Z}_{\rm FN} \tag{11}$$

and thus content ourselves with the assumption that either **(1)** a discrete symmetry may truly be the ultimate UV source of flavour (as can easily be motivated in stringy scenarios — see e.g. [63–65] for recent progress), **(2)** that the $\mathbb{Z}_{\rm FN}$ represents an IR residual symmetry of a (yet further) UV continuous symmetry, perhaps realized at or above $\Lambda_{\rm UV}$, or that **(3)** the IR physics

[4] We refer the reader to e.g. [48, 49], where π_θ is identified as the QCD axion, for related studies.

[5] Anomaly constraints can also be important in models employing (Abelian or non-Abelian) discrete symmetries [52–61], which arguably must be gauged in the deep UV to avoid quantum gravity (wormhole) constraints [62].

arising from the flavon Goldstone or the flavourful gauge boson plays a phenomenological 'second fiddle' to that of ϑ. Regardless, in Section 3.2 we further discuss the generalization of $\mathbb{Z}_{\rm FN}$ to both non-Abelian and/or continuous symmetry constructions.

No matter what the symmetry structure is, the matching procedure becomes more complex at the one-loop level despite still remaining methodical. Matching the 1LPI effective actions of the IR EFT and the full UV theory at one-loop, one finds, using the method of regions, that only the hard momentum region of the UV effective action contributes to the Wilson coefficients of the EFT

$$\int d^d \mathcal{L}_{\rm EFT}^{\text{1-loop}}[\phi] = \Gamma_{\rm L,UV}^{\text{1-loop}}[\phi]\big|_{\rm hard}. \tag{12}$$

The background field method (see [66] for an introduction) is used to write this in a more practical form,

$$\int d^d x\, \mathcal{L}_{\rm EFT}^{\text{1-loop}} = \frac{i}{2} \log \text{Sdet}\left(-\frac{\partial^2 S_{\rm UV}}{\partial \varphi^2}\bigg|_{\Phi=\Phi_c[\phi]} \right)\Bigg|_{\rm hard}, \tag{13}$$

where $\mathcal{S}_{\rm UV}$ is the action of the full theory, φ stands for both the heavy and light fields and Sdet is a generalization of the functional determinant capturing integration over both bosonic and fermionic fields. Following on from [67], the `Matchete` package [30] version 0.1.5 is able to match a UV Lagrangian to the SMEFT at the one-loop level in an automated way using this functional method. The only manual task is to manipulate the output Lagrangian into the desired physical basis. Note of course that diagrammatic techniques for one-loop matching (e.g. those embedded in `Matchmakereft` [68]) are also applicable, and in what follows we will often intuit the loop-level functional output `Matchete` with UV diagrams that we also use to check many of `Matchete`'s results, particularly those unique to FN models.

Furthermore, besides these normal complications regarding functional matching, there are also additional subtleties in our calculation stemming from the fact that we are matching an EFT to an EFT, as opposed to a renormalizable theory to an EFT. For example, our principle expansion parameter is phenomenologically bound by fermionic mass and mixing textures to $\lambda \sim \mathcal{O}(10^{-1})$, which necessitates considering the IR matching effects of higher-order operators in the UV (see the discussion below). Additionally, care has to be taken in the treatment of evanescent operators when matching UV models to EFTs at the one-loop-level or higher, and when transforming the results into a physical basis (see [69] for details). The subtleties arise because the complete operator basis in $d = 4 - 2\varepsilon$ dimensions is infinite dimensional whereas the $d = 4$ basis of dimension-6 operators is finite dimensional. Whether this will be a problem at one-loop hinges on which EFT operators are generated when the heavy new physics is integrated out at the tree-level. For the FN model in question, the flavon couplings to the SM are such that no evanescent operators are generated up to the operator dimensions reached in this paper. The absence of such operators means that we can safely relate the redundant set of operators in the `Matchete` output into the physical Warsaw basis using the Fierz identities. Indeed, all of our upcoming matching results at mass-dimension 6 are presented in the non-redundant, physical Warsaw basis [8].

$\mathcal{S}_{\rm FN}$ Charge	$\overline{Q}_1$	$\overline{Q}_2$	$\overline{Q}_3$	d_1	d_2	d_3	H	θ
q	6	4	0	5	3	3	-3	-2

Table 1: The charge assignments for a toy three-generation FN model in the down-quark sector. The subscripts denote the fermion generation.

2.1 A Toy Model for the Down Quarks

We begin by setting up a simple benchmark model of flavour, concentrating only on the down-type quarks and aiming to achieve somewhat realistic Yukawa textures. This model, whose charge assignments are displayed in Table 1, allows us to illustrate the key features of our results in a concrete way and paves the way for the more abstract discussion of Section 3.[6] All SM fields not listed in Table 1 are assumed to carry unspecified $\mathcal{S}_{\rm FN}$ charges such that some renormalizable Yukawa elements are allowed for the charged leptons and up quarks. In other words, while we allow FN suppression patterns to arise in these family sectors, we do not assume that FN dynamics *successfully* resolve/predict their flavour structure.

Imposing this set of $\mathcal{S}_{\rm FN}$ charges, the down-quark mass matrix and the corresponding mass dimension of the operators required to populate its individual matrix elements are given by

$$M^d \simeq \frac{v_H}{\sqrt{2}} \begin{pmatrix} c_{11}^{d_8}\,\lambda^4 & c_{12}^{d_7}\,\lambda^3 & c_{13}^{d_7}\,\lambda^3 \\ c_{21}^{d_7}\,\lambda^3 & c_{22}^{d_6}\,\lambda^2 & c_{23}^{d_6}\,\lambda^2 \\ c_{31}^{d_5}\,\lambda & y_{32}^{d}\,\lambda^0 & y_{33}^{d}\,\lambda^0 \end{pmatrix} \implies \left[\mathcal{O}^d\right] = \begin{pmatrix} 8 & 7 & 7 \\ 7 & 6 & 6 \\ 5 & 4 & 4 \end{pmatrix}, \tag{14}$$

where it is clear that the FN charges allow for a reduced renormalizable Yukawa sector given by

$$\mathcal{L}_{\rm SM'} \supset y_{32}^d\,\overline{Q}_3 H d_2 + y_{33}^d\,\overline{Q}_3 H d_3 + \text{h.c.}\,, \tag{15}$$

in addition to the aforementioned unspecified flavour(s) $\{i,j\}$ of up quark and charged lepton Yukawas, $y_{ij}^u \overline{Q}_i \tilde{H} u_j$ and $y_{ij}^e \overline{L}_i H e_j$, that turn on at mass-dimension 4. The superscript d_k in the coefficients $c_{ij}^{d_k}$ above denotes a down-type Yukawa Wilson coefficient which enters at operator dimension k — the Yukawa element is thus suppressed by λ^{k-4}. The exception to this notation is the dimension-5 coefficient $c_{ij}^{d_5}$, which will appear so frequently in our analysis that we leave the dimension label implicit; the coefficient c_{ij}^d will everywhere stand for $c_{ij}^{d_5}$ in what follows.

The matrix (14) indicates that, in order to be fully consistent, one would need to consider the effective FN Lagrangian corresponding to this toy model up to $d_{\rm FN} = 8$. However, the phenomenological implications of the FN theory beyond (14) are of course most relevant at *lower* orders in $1/\Lambda_{\rm UV}$. Indeed, we will find in upcoming sections that the IR SMEFT effects sourced from FN operators as low as $d_{\rm FN} \leq 6$ are already extremely rich, and sufficiently challenging to enumerate on a technical level. This comment is true upon considering both tree-level *and* one-loop UV contributions with FN vertices. The latter loop contributions are

[6] We are aware of this toy model thanks to a publicly available talk given by C. Luhn at FLASY 2014. To the best of our knowledge, the earliest use of these textures traces back to [70, 71], where one can also find more details about the implied mass and mixing phenomenology they generate.

especially important in the context of our study due to the fact that, unlike most matching analyses, we have a relatively fixed hierarchy of new physics scales built-in, thanks to the constraint $\lambda \lesssim \mathcal{O}(10^{-1})$ driven by mass and mixing considerations. Since $(4\pi)^{-1} \sim 10^{-(1-2)}$, naive power counting tells us that $\lambda/(4\pi) \sim \lambda^2$. Assuming that dimensionless Lagrangian parameters are $\mathcal{O}(1)$, this indicates that a tree-level UV contribution sourced from a dimension-$d_{\rm FN}$ vertex can become competitively large with a one-loop UV contribution sourced from a $d_{\rm FN} - 1$ dimension vertex.

In what follows we will therefore systematically match our toy model implied in (14) to the $d_{\rm SM} = 6$ SMEFT, considering up to $d_{\rm FN} \leq 6$ UV operators, and including loop contributions from $d_{\rm FN} \leq 5$ operators. When presenting our results, we adopt the notational conventions of [8] unless otherwise stated.

2.1.1 Dimension Four ($d_{\rm FN} = 4$)

We first consider our toy model's renormalizable FN Lagrangian, which is given by (5) and the $\mathcal{S}_{\rm FN}$-invariant SM. Working in the broken $\mathcal{S}_{\rm FN}$ phase, we expand the flavon field about its vacuum and, ignoring operators without ϑ or v_θ dependence, obtain[7]

$$\mathcal{L}^4_{\not{\rm FN}} \supset -\frac{1}{2}\vartheta\Box\vartheta - \mu_\theta^2\vartheta^2 - \frac{\mu_\theta^2}{v_\theta}\vartheta^3 - \frac{\mu_\theta^2}{4v_\theta^2}\vartheta^4 - \frac{\lambda_{11}}{2}v_\theta^2\big(H^\dagger H\big) - \lambda_{11}v_\theta\big(H^\dagger H\big)\vartheta - \frac{\lambda_{11}}{2}\big(H^\dagger H\big)\vartheta^2\,, \quad (16)$$

which of course includes 'static' terms proportional to the flavon vev v_θ and terms dynamical in ϑ, which we will integrate out with (8). In doing so, we will find familiar results, given that $\mathcal{L}^4_{\rm FN}$ effectively reduces to well-studied case of the SM extended by a singlet (complex) scalar, with the only difference being that the Yukawa matrices controlled by FN-type dynamics are mostly populated by zeros at the renormalizable level thanks to $\mathcal{S}_{\rm FN}$. We can and will therefore compare our findings with those in (e.g.) [72, 73].

Tree-Level Matching

From (16) we can immediately derive the following EOM for ϑ:

$$\big(\Box + 2\mu_\theta^2\big)\vartheta = -\lambda_{11}v_\theta\big(H^\dagger H\big) - \lambda_{11}\big(H^\dagger H\big)\vartheta - \frac{3\mu_\theta^2}{v_\theta}\vartheta^2 - \frac{\mu_\theta^2}{v_\theta^2}\vartheta^3. \quad (17)$$

Rearranging and identifying $2\mu_\theta^2 \equiv m_\theta^2$, one obtains the (recursive) expression for ϑ:

$$\vartheta = \frac{-\lambda_{11}v_\theta}{\Box + m_\theta^2}\big(H^\dagger H\big) - \frac{\lambda_{11}}{\Box + m_\theta^2}\big(H^\dagger H\big)\vartheta - \frac{3m_\theta^2}{2v_\theta}\frac{1}{\Box + m_\theta^2}\vartheta^2 - \frac{m_\theta^2}{2v_\theta^2}\frac{1}{\Box + m_\theta^2}\vartheta^3. \quad (18)$$

Following (e.g.) [74], we will solve this expression iteratively, and proceed by seeking a power series solution with $1/m_\theta^2$ the power-counting parameter. We thus write

$$\vartheta = \vartheta^{(2)} + \vartheta^{(4)} + \vartheta^{(6)} + \dots \quad (19)$$

[7] The slashed notation here and below, F$\not{\rm N}$, simply indicates interactions in the broken $\mathcal{S}_{\rm FN}$ phase.

which, upon collecting terms and solving, yields

$$\vartheta^{(2)} = -\frac{\lambda_{11} v_\theta \big(H^\dagger H\big)}{m_\theta^2}\,,$$

$$\vartheta^{(4)} = -\frac{\lambda_{11}^2 v_\theta}{2m_\theta^4}\big(H^\dagger H\big)^2 + \frac{\lambda_{11} v_\theta}{m_\theta^4}\Box\big(H^\dagger H\big)\,,$$

$$\vartheta^{(6)} = -\frac{\lambda_{11}^3 v_\theta}{2m_\theta^6}\big(H^\dagger H\big)^3 + \frac{\lambda_{11}^2 v_\theta}{2m_\theta^6}\Box\big(H^\dagger H\big)^2 + \frac{2\lambda_{11}^2 v_\theta}{m_\theta^6}\big(H^\dagger H\big)\Box\big(H^\dagger H\big) - \frac{\lambda_{11} v_\theta}{m_\theta^6}\Box^2\big(H^\dagger H\big)\,. \quad (20)$$

Substituting the flavon EOM back into (16), keeping terms up to $d_{\rm SM} = 6$, and appreciating a number of nice cancellations, one finally obtains

$$\mathcal{L}_{\rm SMEFT} \supset -\frac{\lambda_{11} v_\theta^2}{2}\big(H^\dagger H\big) + \frac{\lambda_{11}^2 v_\theta^2}{2m_\theta^2}\big(H^\dagger H\big)^2 - \frac{\lambda_{11}^2 v_\theta^2}{2m_\theta^4}\big(H^\dagger H\big)\Box\big(H^\dagger H\big). \quad (21)$$

We can already make a few important observations from this simple result. To begin with, two renormalizable terms, $H^\dagger H$ and $(H^\dagger H)^2$, receive contributions from integrating out the flavon. This changes the quadratic and quartic Higgs potential terms from

$$V(H) \supset -\mu_H^2\big(H^\dagger H\big) + \lambda_{20}\big(H^\dagger H\big)^2 \quad (22)$$

to

$$V(H) \supset -\underbrace{\left(\mu_H^2 - \frac{\lambda_{11} v_\theta^2}{2}\right)}_{\mu_H'^2}\big(H^\dagger H\big) + \underbrace{\left(\lambda_{20} - \frac{\lambda_{11}^2 v_\theta^2}{2m_\theta^2}\right)}_{\lambda_{20}'}\big(H^\dagger H\big)^2\,, \quad (23)$$

where the shifts may be absorbed into redefinitions of the quadratic and quartic Higgs couplings as shown.[8]

In addition, one dimension-6 operator in the Warsaw basis, the bosonic $\mathcal{Q}_{H\Box}$, is turned on. This operator universally modifies the couplings between the physical Higgs field h and the other SM fields in the broken EW phase and can be interpreted as capturing the mixing between the Higgs and the flavon field.[9] Another "Higgs-like operator" which one may have

[8] As it stands, the flavon contributes to the Higgs hierarchy problem by shifting the Higgs mass term as expected for a heavy BSM scalar. We do not address the hierarchy problem in this work.

[9] As shown and explained in [75], after renormalizing the Higgs kinetic term, all *renormalizable* operators containing the Higgs field will see their couplings $\{g_i\}$ modified by a shift $\delta g_i \propto C_{H,\rm kin}$ where

$$C_{H,\rm kin} \equiv \left(C_{H\Box} - \frac{1}{4}C_{HD}\right)v_H^2 = -\frac{\lambda_{11}^2 v_\theta^2 v_H^2}{2m_\theta^4}.$$

For instance, the trilinear Higgs self-coupling will shift as

$$-\lambda_{20} h^3 \to -\lambda_{20}(1 + 3C_{H,\rm kin})h^3.$$

Had we not integrated out the flavon ϑ, this phenomenon would be attributed to Higgs-flavon mixing which makes the scalar mass matrix non-diagonal. Diagonalizing the mass matrix requires a redefinition of the physical scalar fields h and ϑ which modifies the couplings between h and the SM fields by an amount shown above. When accounting for the modified couplings in the case of dimension-6 SMEFT operators, the shift arising from the rescaling of the kinetic term is formally of dimension-8 and can thus be neglected when working at dimension-6 level in the EFT.

expected, $\mathcal{Q}_H$, vanishes when we expand about the minimum of the scalar potential in the broken phase.

Solving for ϑ up to mass-dimension 4 terms and substituting back into the UV Lagrangian of (16), we find that dependence on $\vartheta^{(4)}$ cancels out when the series is truncated at dimension-6 terms. The higher dimension term $\vartheta^{(4)}$ first appears at dimension-8 in the SMEFT.

Hence, before even considering the FN-type operators driving the flavour mechanism of core interest, we see that the presence of the additional $\mathcal{S}_{\rm FN}$ symmetry (and its breaking via a scalar flavon) generates non-trivial consequences in the SMEFT. Indeed, the Wilson coefficient of $Q_{H\Box}$ is already experimentally bound by available global SMEFT fits [13] which, for $\lambda_{11} = \mathcal{O}(1)$ and $m_\theta \sim v_\theta$, imply a lower bound $m_\theta \gtrsim 1$ TeV on the flavon mass.

As a final comment, we note that we have checked our analytic result in (21) against the output of `Mathchete`, which gives

$$\begin{aligned}\mathcal{L}_{\rm SMEFT}\big|_{\texttt{Matchete}} \supset &-\frac{\lambda_{11}v_\theta^2}{2}H^\dagger H + \frac{\lambda_{11}^2 v_\theta^2}{2m_\theta^2}\big(H^\dagger H\big)^2 - \frac{\lambda_{11}^2 v_\theta^2}{m_\theta^4}\big(H^\dagger H\big)(D_\mu H)^\dagger(D^\mu H) \\ &- \frac{\lambda_{11}^2 v_\theta^2}{2m_\theta^4}\big(H^\dagger H\big)\big(H^\dagger D_\mu D^\mu H + \text{h.c.}\big) \end{aligned} \tag{24}$$

upon asking that (16) be matched to the $d_{\rm SM} = 6$ SMEFT. The reason this looks different from our analytic result calculated in (21) is that both of the two-derivative bosonic operators listed here are not in the Warsaw basis, whereas our result from above is. However, using the standard IBP identity

$$\big(H^\dagger H\big)(D_\mu H)^\dagger\big(D^\mu H\big) = \frac{1}{2}\big(H^\dagger H\big)\Big(\Box\big(H^\dagger H\big) - H^\dagger D_\mu D^\mu H - (D_\mu D^\mu H)^\dagger H\Big)\,, \tag{25}$$

we may exchange the bosonic two-derivative operators in (24) for $(H^\dagger H)\Box(H^\dagger H)$ to arrive at the same result, thus completing our crosscheck and giving us further confidence in `Matchete`'s functionality, which we will now use for loop-level matching.[10]

One-Loop Matching

Following the functional algorithm briefly reviewed around (12)–(13), we have allowed `Matchete` to calculate the one-loop contributions sourced by matching (16) to the $d_{\rm SM} \leq 6$ SMEFT. In so doing we have uncovered a host of further non-trivial operator structures in the Yukawa, Higgs-scalar, Higgs-gauge, Higgs-derivative, Higgs-kinetic, and four-fermion sectors. For the remainder of this work, we use $\hbar$ as shorthand for the loop suppression factor $(4\pi)^{-2}$, a choice that follows the convention in `Matchete`.

We first consider contributions to the (Higgs-enhanced) *Yukawa* sector ($\psi^2 H^{1+d_{\rm SM}-4}$), including both modifications to the renormalizable SM terms and higher-dimensional SMEFT operators:

$$\mathcal{L}_{\rm SMEFT} \supset \frac{\hbar\lambda_{11}^2 v_\theta^2 y^e_{ij}}{12m_\theta^4}\big[3m_\theta^2 - m_H^2(17+6\mathbb{L})\big]\overline{L}_i H e_j + \frac{\hbar\lambda_{11}^2 v_\theta^2 y^d_{ij}}{12m_\theta^4}\big[3m_\theta^2 - m_H^2(17+6\mathbb{L})\big]\overline{Q}_i H d_j$$

[10] In what follows we will perform this identity and others (e.g. Fierz identities) by hand, in order to translate `Matchete` output to the Warsaw basis for all operators that appear.

$$+\frac{\hbar y^u_{ij}\lambda_{11}^2 v_\theta^2}{12m_\theta^4}\big[3m_\theta^2-m_H^2(17+6\mathbb{L})\big]\overline{Q}_i\tilde{H}u_j+\frac{\hbar\lambda_{11}^2 v_\theta^2}{36m_\theta^4}\Big[9(y^e y^{e\dagger}y^e)_{ij}(5+6\mathbb{L})+6y^e_{ij}\lambda'_{20}(29+6\mathbb{L})$$

$$-y^e_{ij}g^2(31+30\mathbb{L})\Big)\big(H^\dagger H\big)\overline{L}_i H e_j+\frac{\hbar\lambda_{11}^2 v_\theta^2 y^d_{ij}}{36m_\theta^4}\big[9\big(|y^d_{32}|^2+|y^d_{33}|^2\big)(5+6\mathbb{L})$$

$$-g^2(31+30\mathbb{L})+6\lambda'_{20}(29+6\mathbb{L})\big]\big(H^\dagger H\big)\overline{Q}_i H d_j+\frac{\hbar\lambda_{11}^2 v_\theta^2}{36m_\theta^4}\big[9(y^u y^{u\dagger}y^u)_{ij}(5+6\mathbb{L})$$

$$+6y^u_{ij}\lambda'_{20}(29+6\mathbb{L})-y^u_{ij}g^2(31+30\mathbb{L})\big]\big(H^\dagger H\big)\overline{Q}_i\tilde{H}u_j\,, \tag{26}$$

where m_H is the Higgs mass parameter in the UV Lagrangian and $\mathbb{L}\equiv\ln(\mu^2/m_\theta^2)$ with μ the $\overline{\rm MS}$ renormalization/matching scale). We also see in (26) the first manifestation of our $\mathcal{S}_{\rm FN}$ charges from Table 1, in that generic flavour indices appear on up-quark and charged-lepton coefficients, whereas specific indices ((32) and (33)) appear on the down quarks.

There are also contributions to the purely bosonic sector of the $d_{\rm SM}\leq 6$ SMEFT, including augmented *Higgs-scalar potential* and *Higgs-derivative* (H^4D^2) terms given by

$$\mathcal{L}_{\rm SMEFT}\supset\frac{\hbar\lambda_{11}^2}{9m_\theta^8}\Big[6\lambda_{11}m_\theta^6-6\lambda_{11}^4v_\theta^6+18\lambda_{11}^2v_\theta^4m_\theta^2(\lambda_{11}-6\lambda'_{20})$$

$$+v_\theta^2m_\theta^4\Big(-18\lambda_{11}^2+108\lambda'_{20}\lambda_{11}+\lambda'_{20}\big(12\lambda'_{20}(41+30\mathbb{L})-g^2(31+30\mathbb{L})\big)\Big)\Big]\big(H^\dagger H\big)^3+$$

$$\Big\{-\frac{\lambda_{11}^2v_\theta^2}{2m_\theta^4}+\frac{\hbar\lambda_{11}^2}{12m_\theta^2}\Big[16-18\mathbb{L}-\frac{2\lambda_{11}v_\theta^2}{m_\theta^2}(7-12\mathbb{L})+\frac{\lambda_{11}^2v_\theta^4}{m_\theta^4}(19+12\mathbb{L})-\frac{g_Y^2v_\theta^2}{6m_\theta^2}(31+30\mathbb{L})$$

$$-\frac{g^2v_\theta^2}{2m_\theta^2}(31+30\mathbb{L})\Big]\Big\}\big(H^\dagger H\big)\Box\big(H^\dagger H\big)-\frac{\hbar g_Y^2v_\theta^2\lambda_{11}^2}{18m_\theta^4}[31+30\mathbb{L}](H^\dagger D^\mu H)^\dagger(H^\dagger D_\mu H)\,, \tag{27}$$

as well as mixed *Higgs-gauge* (H^2X^2) effective operators turning on

$$\mathcal{L}_{\rm SMEFT}\supset\frac{\hbar\lambda_{11}^2v_\theta^2}{12m_\theta^4}\big[\big(H^\dagger H\big)\big(g_Y^2B^{\mu\nu}B_{\mu\nu}+g^2W^{I\mu\nu}W^I_{\mu\nu}\big)+2gg_Y\big(H^\dagger\tau^IH\big)B^{\mu\nu}W^I_{\mu\nu}\big] \tag{28}$$

with hypercharge (B), weak isospin (W), and mixed (BW) field-strength tensors appearing.

Additional *Higgs-kinetic terms* with non-trivial fermion structure (ψ^2H^2D) also appear,

$$\mathcal{L}_{\rm SMEFT}\supset\frac{\hbar\lambda_{11}^2v_\theta^2}{144m_\theta^4}\Big[9\big(|y^d_{32}|^2+|y^d_{33}|^2\big)\delta_{3i}\delta_{3j}(5+2\mathbb{L})-9\big(y^uy^{u\dagger}\big)_{ij}(5+2\mathbb{L})$$

$$-\frac{g_Y^2}{3}(17+6\mathbb{L})\delta_{ij}\Big]\Big(H^\dagger i\overleftrightarrow{D}_\mu H\Big)\overline{Q}_i\gamma^\mu Q_j+\frac{\hbar v_\theta^2\lambda_{11}^2}{144m_\theta^4}\Big[9\big(|y^d_{32}|^2+|y^d_{33}|^2\big)\delta_{3i}\delta_{3j}(5+2\mathbb{L})$$

$$+9\big(y^uy^{u\dagger}\big)_{ij}(5+2\mathbb{L})-g^2\delta_{ij}(17+6\mathbb{L})\Big]\Big(H^\dagger i\overleftrightarrow{D}^I_\mu H\Big)\overline{Q}_i\tau^I\gamma^\mu Q_j$$

$$+\frac{\hbar\lambda_{11}^2v_\theta^2}{216m_\theta^4}\Big[g_Y^2\delta_{ij}(17+6\mathbb{L})-27\big(y^{d\dagger}y^d\big)_{ij}(5+2\mathbb{L})\Big]\Big(H^\dagger i\overleftrightarrow{D}_\mu H\Big)\overline{d}_i\gamma^\mu d_j$$

$$
\begin{aligned}
&- \left\{ \frac{\hbar\lambda_{11}^2 v_\theta^2}{4m_\theta^4} \Big[\big(y^{u\dagger} y^d\big)_{ij} (5 + 2\mathbb{L}) \Big] \Big(\tilde{H}^\dagger i D_\mu H \Big) \overline{u}_i \gamma^\mu d_j + \text{h.c.} \right\} \\
&- \frac{\hbar\lambda_{11}^2 v_\theta^2}{216m_\theta^4} \big[2g_Y^2 \delta_{ij} (17 + 6\mathbb{L}) - 27 (y^{u\dagger} y^u)_{ij} (5 + 2\mathbb{L}) \big] (H^\dagger i \overleftrightarrow{D}_\mu H) (\overline{u}_i \gamma^\mu u_j) \\
&- \frac{\hbar\lambda_{11}^2 v_\theta^2}{72m_\theta^4} \big[9 (y^{e\dagger} y^e)_{ij} (5 + 2\mathbb{L}) - g_Y^2 \delta_{ij} (17 + 6\mathbb{L}) \big] (H^\dagger i \overleftrightarrow{D}_\mu H) (\overline{e}_i \gamma^\mu e_j) \\
&+ \frac{\hbar\lambda_{11}^2 v_\theta^2}{144m_\theta^4} \big[9 (y^e y^{e\dagger})_{ij} (5 + 2\mathbb{L}) + g_Y^2 \delta_{ij} (17 + 6\mathbb{L}) \big] (H^\dagger i \overleftrightarrow{D}_\mu H) (\overline{L}_i \gamma^\mu L_j) \\
&+ \frac{\hbar\lambda_{11}^2 v_\theta^2}{144m_\theta^4} \big[9 (y^e y^{e\dagger})_{ij} (5 + 2\mathbb{L}) - g^2 \delta_{ij} (17 + 6\mathbb{L}) \big] (H^\dagger i \overleftrightarrow{D}_\mu^I H) (\overline{L}_i \tau^I \gamma^\mu L_j) \,, \qquad (29)
\end{aligned}
$$

where the product of down-type Yukawa matrices can be written in matrix form as

$$
\big(y^{d\dagger} y^d\big)_{ij} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & |y_{32}^d|^2 & y_{32}^{d*} y_{33}^d \\ 0 & y_{32}^d y_{33}^{d*} & |y_{33}^d|^2 \end{pmatrix} \qquad (30)
$$

for the toy model in question. To help with intuitive understanding, Fig. 1[11] presents some Feynman diagrams that contribute to the Wilson coefficient of the operator $(H^\dagger i \overleftrightarrow{D}_\mu H)\big(\overline{d}_i \gamma^\mu d_j\big)$ upon integrating out the flavon.

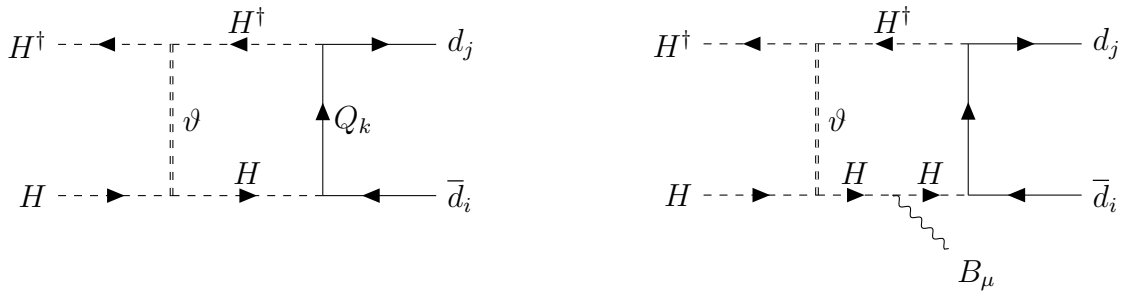
One also finds a number of *four-fermion* (ψ^4) operator classes:

$$
\begin{aligned}
\mathcal{L}_{\text{SMEFT}} \supset &\left\{ \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{ij}^e y_{kl}^{d\dagger}}{6m_\theta^4} \big(\overline{L}_i^\alpha e_j\big) \big(\overline{d}_k Q_l^\alpha\big) + \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{ij}^e y_{kl}^u}{6m_\theta^4} \epsilon_{\alpha\beta} \big(\overline{L}_i^\alpha e_j\big) \Big(\overline{Q}_k^\beta u_l\Big) \right. \\
&\left. + \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{ij}^u y_{kl}^d}{6m_\theta^4} \epsilon_{\alpha\beta} \big(\overline{Q}_i^\alpha u_j\big) \Big(\overline{Q}_k^\beta d_l\Big) + \text{h.c.} \right\} - \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{il}^e y_{kj}^{e\dagger}}{12m_\theta^4} (\overline{L}_i \gamma_\mu L_j) (\overline{e}_k \gamma^\mu e_l) \\
&- \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{il}^d y_{kj}^{d\dagger}}{12m_\theta^4} (\overline{Q}_i \gamma_\mu Q_j) (\overline{d}_k \gamma^\mu d_l) - \frac{\hbar\lambda_{11}^2 v_\theta^2 y_{il}^u y_{kj}^{u\dagger}}{12m_\theta^4} (\overline{Q}_i \gamma_\mu Q_j) (\overline{u}_k \gamma^\mu u_l), \qquad (31)
\end{aligned}
$$

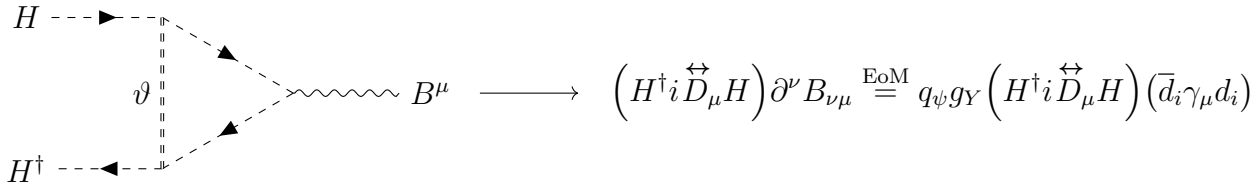
where $\epsilon_{\alpha\beta}$ is the antisymmetric $\text{SU}(2)_\text{L}$ tensor.

The one-loop matching coefficients presented in this section already exist in the literature in the context of the Standard Model singlet extension, obtained using functional methods [72,77] and diagrammatic techniques [73]. Our results from `Matchete` are in full agreement with those in the references once we account for the fact that we have restricted the form of the down-type Yukawa matrices in a flavour non-universal way. The restriction of these Yukawas to only the (33) and (32) elements leads to patterns in family space best illustrated by equations (26) and (29)–(31). We see for instance that the Wilson coefficient $(C_{Hd})_{ij}$ of the Higgs-kinetic operator with down-type quarks, understood as a matrix in generation space, contains a fully populated 2×2 block shown partly in (30) and a (11) element proportional to g_Y^2. The other

[11] All the Feynman diagrams in this paper have been drawn using `TikZ-Feynman` [76].



(a) These two diagrams give a contribution to the Wilson coefficient C_{Hd} proportional to $\lambda_{11}^2 v_\theta^2 (y^{d\dagger} y^d)_{ij}/m_\theta^4$.



(b) A diagram contributing to the term proportional to g_Y^2 in the Wilson coefficient C_{Hd}.

Figure 1: Some diagrams contributing to the Wilson coefficient C_{Hd} of the Higgs-kinetic operator $(H^\dagger i\overleftrightarrow{D}_\mu H)\left(\overline{d}_i\gamma^\mu d_j\right)$ coming from the dimension-4 UV Lagrangian at one-loop.

elements in the first row and columnn are zero. This flavour pattern can be recognised as a prediction of the toy model.

In conclusion, we find that, before even considering the θ-dependent operators of principal interest to the FN mechanism (those populating the Yukawa matrix in (14)), a wide array of SMEFT operator classes appear already from $d_{\rm FN} = 4$ UV contributions. Many of the Wilson coefficients that appear contain patterns in generation space dictated by the renormalizable Yukawa elements. In fact, we will find that tree- and loop-level corrections sourced from $d_{\rm FN} = 5$ operators will only introduce a few more operator classes beyond those found in this subsection, although the associated Wilson coefficients of most of the operators appearing between (26)–(31) will also be augmented by higher-order corrections.

2.1.2 Dimension Five ($d_{\rm FN} = 5$)

At $d_{\rm FN} = 5$ one only has the Weinberg operator [12] and the effective Yukawa operator implied in (14),

$$\mathcal{L}^5_{\rm FN} = \frac{c^W_{ij}}{\Lambda_{\rm UV}}\left(H\overline{L}^c_i\right)(L_j H) + \frac{c^d_{31}}{\Lambda_{\rm UV}}\,\overline{Q}_3 H d_1\theta + \text{h.c.}, \tag{32}$$

assuming that the (unspecified) $\mathcal{S}_{\rm FN}$-charge of the $\mathrm{SU(2)_L}$ lepton doublet L respects $q_{L_i} + q_{L_j} \overset{!}{=} -2q_H$ and that the $\mathcal{S}_{\rm FN}$ charges of the up quarks and/or charged leptons do not allow for a leading $d_{\rm FN} = 5$ contribution to the Yukawa matrix.[12] Substituting the expanded flavon field

[12]We will discuss FN charge generalizations to the lepton sector in Section 3 below.

into this Lagrangian and ignoring operators with no ϑ dependence results in

$$\mathcal{L}^{5}_{\text{F}\not{\text{N}}} \supset c^{d}_{31}\,\lambda\,\overline{Q}_3 H d_1 + \frac{c^{d}_{31}}{\sqrt{2}\Lambda_{\text{UV}}}\overline{Q}_3 H d_1 \vartheta + \text{h.c.} \tag{33}$$

The first term is of course the operator driving the core FN mechanism, in that it populates the Yukawa couplings with a non-zero (31) matrix element, suppressed by one power of $\lambda = v_\theta/(\sqrt{2}\Lambda_{\text{UV}})$, while the second term allows for a propagating ϑ that again needs to be integrated out in order to make contact with the SMEFT.

Tree-Level Matching

Following exactly the same procedure as between (17)–(20) above, one quickly arrives at the tree-level $d_{\text{FN}} = 5$ contribution to the $d_{\text{SM}} = 6$ SMEFT,

$$\mathcal{L}_{\text{SMEFT}} \supset -\frac{\lambda_{11}\lambda\, c^{d}_{31}}{m^2_\theta}\big(H^\dagger H\big)\overline{Q}_3 H d_1 + \text{h.c.}, \tag{34}$$

which is of course understood to be in *addition* to those operators found in Section 2.1.1. We see that the enhanced Yukawa-sector operator Q_{dH} of the Warsaw basis is turned on, with a distinctive Wilson coefficient suppressed both by λ and two powers of the flavon mass m^2_θ. Additionally, the presence of λ_{11} and c^d_{31} indicates that this term is sourced by an interplay between scalar potential (Higgs-flavon mixing) and Yukawa potential EFT operators. As Fig. 2 shows, this is easy to understand in terms of diagrammatic matching. The contribution arises from connecting a $H^\dagger H\vartheta$ vertex of (16) to the second vertex of (34) via an internal flavon propagator, yielding the local $d_{\text{SM}} = 5$ term in (34). We have again checked (34) with `Matchete`, finding perfect agreement.

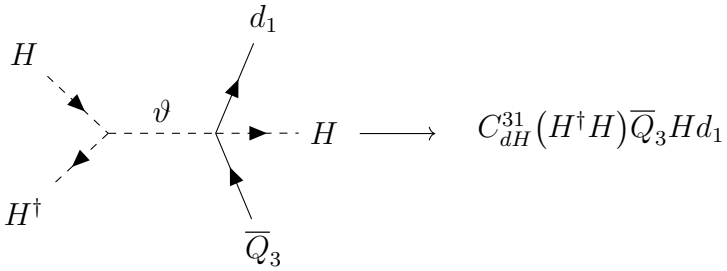


Figure 2: A tree-level flavon exchange, leading to an enhanced Yukawa operator in the Warsaw basis. Note that above the electroweak scale, the Higgs field (H) carries hypercharge $Y = 1/2$, which is denoted by the arrows in the external Higgs legs.

As a final comment for intuitive readers, one also obtains Higgs-enhanced four-fermions ($H^2\psi^4$) from $d_{\text{FN}} = 5$ vertices at tree-level (via a single internal ϑ propagator), but these represent $d_{\text{SM}} = 8$ SMEFT operators (which we briefly discuss in our Summary and Outlook). At low energies, upon the Higgs acquiring a vev, these will turn into Low-Energy EFT (LEFT) [78, 79] operators $\propto v^2_H\psi^4$.

One-Loop Matching

We now move to our final loop-level analysis by enumerating the contributions to IR SMEFT operators coming, in the diagrammatic interpretation, from loops that can be drawn with both renormalizable and $d_{\rm FN} = 5$ vertices. We again use `Matchete` to perform this analysis using the functional matching method, and later do analytic computations of individual diagrams to crosscheck the results.

Unsurprisingly, we find that (32) does not source any novel contributions to the purely bosonic sector of the SMEFT beyond those already presented in Section 2.1.1. We do of course find additional contributions to the enhanced Yukawa sector given by

$$\begin{aligned}\mathcal{L}_{\rm SMEFT} \supset -\frac{\hbar c_{31}^d \lambda\lambda_{11}}{4m_\theta^2}\Bigg[&6\big(|y_{32}^d|^2+|y_{33}^d|^2\big)(1+\mathbb{L}) - 6\frac{m_\theta^2}{v_\theta^2}(1-\mathbb{L}) - \frac{\lambda_{11}^2 v_\theta^2}{m_\theta^2}(15+4\mathbb{L}) \\ &+ 2\lambda_{11}(4-6\mathbb{L}) + \lambda'_{20}(7+6\mathbb{L})\Bigg]\big(H^\dagger H\big)\overline{Q}_3 H d_1 + \frac{\hbar y_{kl}^d}{144m_\theta^2}\Bigg[18\frac{\lambda^2 m_\theta^2 |c_{31}^d|^2}{v_\theta^2}(1+2\mathbb{L}) \\ &+ \frac{4\lambda_{11}^2 v_\theta^2}{m_\theta^2}\big(9\big(|y_{32}^d|^2+|y_{33}^d|^2\big)(5+6\mathbb{L}) - g^2(31+30\mathbb{L}) + 6\lambda'_{20}(29+6\mathbb{L})\big)\Bigg]\big(H^\dagger H\big)\overline{Q}_k H d_l + \text{h.c.},\end{aligned} \tag{35}$$

where $(k,l) \neq (3,1)$, as well as contributions to Higgs-kinetic operators

$$\begin{aligned}\mathcal{L}_{\rm SMEFT} \supset\ & \frac{\hbar}{8m_\theta^2}\Bigg[4\lambda\lambda_{11}\Big(c^{d\dagger}y^d + y^{d\dagger}c^d\Big)_{ij} + \frac{2\lambda^2 m_\theta^2 |c_{31}^d|^2}{v_\theta^2}\delta_{i1}\delta_{j1}(1+2\mathbb{L})\Bigg]\Big(H^\dagger i\overleftrightarrow{D}_\mu H\Big)\big(\overline{d}_i\gamma^\mu d_j\big) \\ &+ \Bigg\{\frac{\hbar\lambda\lambda_{11}}{2m_\theta^2}\big(y^{u\dagger}c^d\big)_{ij}(\tilde{H}^\dagger i D_\mu H)(\overline{u}_i\gamma^\mu d_j) + \text{h.c.}\Bigg\} - \frac{\hbar\lambda^2|c_{31}^d|^2}{16v_\theta^2}\delta_{i3}\delta_{j3}(1+2\mathbb{L})\Big(H^\dagger i\overleftrightarrow{D}_\mu H\Big)\big(\overline{Q}_i\gamma^\mu Q_j\big) \\ &- \frac{\hbar\lambda^2|c_{31}^d|^2}{16v_\theta^2}\delta_{i3}\delta_{j3}(1+2\mathbb{L})\Big(H^\dagger i\overleftrightarrow{D}_\mu^I H\Big)\big(\overline{Q}_i\tau^I\gamma^\mu Q_j\big).\end{aligned} \tag{36}$$

It is again instructive to show the emergent flavour structure for the toy model defined in Table 1. We may write the product of flavour space matrices on the first line of (36) as

$$\Big(c^{d\dagger}y^d + y^{d\dagger}c^d\Big)_{ij} = \begin{pmatrix} 0 & c_{31}^{d*}y_{32}^d & c_{31}^{d*}y_{33}^d \\ c_{31}^d y_{32}^{d*} & 0 & 0 \\ c_{31}^d y_{33}^{d*} & 0 & 0 \end{pmatrix}, \tag{37}$$

whereas $y^{u\dagger}c^d$ from the line that follows is represented by

$$\big(y^{u\dagger}c^d\big)_{ij} = \begin{pmatrix} c_{31}^d y_{31}^{u*} & 0 & 0 \\ c_{31}^d y_{32}^{u*} & 0 & 0 \\ c_{31}^d y_{33}^{u*} & 0 & 0 \end{pmatrix}. \tag{38}$$

Note again that we have not made any assumptions about the up-type Yukawa matrix elements y_{ij}^u of the renormalizable Lagrangian. Finally there are also four-fermion operators,

$$\mathcal{L}_{\rm SMEFT} \supset \Bigg\{-\frac{\hbar\lambda\lambda_{11}y_{ij}^e c_{13}^{d*}}{2m_\theta^2}\big(\overline{L}_i^\alpha e_j\big)\big(\overline{d}_1 Q_3^\alpha\big) - \frac{\hbar\lambda\lambda_{11}y_{ij}^u c_{31}^d}{2m_\theta^2}\epsilon_{\alpha\beta}\big(\overline{Q}_i^\alpha u_j\big)\Big(\overline{Q}_3^\beta d_1\Big) + \text{h.c.}\Bigg\}$$

$$- \frac{\hbar\lambda^2|c_{31}^d|^2}{2v_\theta^2}(1+\mathbb{L})\big(\overline{Q}_3\gamma_\mu Q_3\big)\big(\overline{d}_1\gamma^\mu d_1\big)\,. \tag{39}$$

Hence (35)–(39) complete the modifications to SMEFT Wilson coefficients sourced from $d_{\rm FN}=5$ loop corrections. As in Section 2.1.1, in Fig. 3 we have sketched a number of diagrams that allow us to physically intuit the `Matchete` output in (35)–(39), and we have explicitly calculated a subset of those diagrams, confirming the associated Wilson coefficients that appear above.

It is worth pausing at this stage to study the results. Firstly, we see that the inclusion of a single FN term in the dimension-5 Lagrangian generates substantial flavour structure in the Higgs-enhanced Yukawa, Higgs-kinetic and four-fermion operators of the SMEFT output. For the Wilson coefficient $(C_{Hd})_{ij}$ associated with the Higgs-kinetic operator with down-type quarks $(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{d}_i\gamma^\mu d_j)$ in (36), a 2×2 block of zero elements in flavour space is predicted, whilst the first column and row are populated. These findings are hardly a surprise, given that the renormalizable terms of (15) and the dimension-5 term in (32) single out flavoured pairs of quarks which, after the flavon is integrated out, are mapped to SMEFT operators with flavour-non-universal Wilson coefficients. Nonetheless, the results demonstrate that at a fixed operator dimension in the SMEFT, the terms associated with FN dynamics are capable of generating distinguishable flavour patterns at the one-loop level.

To gain a better understanding of the patterns in the Wilson coefficients, we combine the results obtained above with those coming out of the $d_{\rm FN}=4$ part of the Lagrangian in Section 2.1.1. We may illustrate this by again considering the Wilson coefficients $(C_{Hd})_{ij}$ of the Higgs-kinetic operator with a down-current. Adding the Wilson coefficients from the two stages of matching gives

$$(C_{Hd}) = \frac{\hbar}{2m_\theta^2}\times$$

$$\begin{pmatrix} \frac{54\lambda^2(1+2\mathbb{L})m_\theta^4|c_{31}^d|^2+g_Y^2\lambda_{11}^2v_\theta^4(17+6\mathbb{L})}{108m_\theta^2v_\theta^2} & \lambda\lambda_{11}c_{31}^{d*}y_{32}^d & \lambda\lambda_{11}c_{31}^{d*}y_{33}^d \\ \lambda\lambda_{11}c_{31}^d y_{32}^{d*} & \frac{\lambda_{11}^2v_\theta^2\left(g_Y^2(17+6\mathbb{L})-27(5+2\mathbb{L})|y_{32}^d|^2\right)}{108m_\theta^2} & -\frac{\lambda_{11}^2v_\theta^2(5+2\mathbb{L})y_{32}^{d*}y_{33}^d}{4m_\theta^4} \\ \lambda\lambda_{11}c_{31}^d y_{33}^{d*} & -\frac{\lambda_{11}^2v_\theta^2(5+2\mathbb{L})y_{32}^d y_{33}^{d*}}{4m_\theta^4} & \frac{\lambda_{11}^2v_\theta^2\left(g_Y^2(17+6\mathbb{L})-27(5+2\mathbb{L})|y_{33}^d|^2\right)}{108m_\theta^2} \end{pmatrix} \tag{40}$$

We may assume $c_{31}\sim y_{ij}^d\sim 1$, as required by our toy model, and $v_\theta\sim m_\theta$ which stems from the expectation that the flavon vev and mass originate from the same physical mechanism. The matrix $(C_{Hd})_{ij}$ then becomes

$$C_{Hd}\sim\frac{\hbar}{2m_\theta^2}\begin{pmatrix} \frac{54\lambda^2(1+2\mathbb{L})+\lambda_{11}^2(17+6\mathbb{L})}{108} & \lambda\lambda_{11} & \lambda\lambda_{11} \\ \lambda\lambda_{11} & -\frac{\lambda_{11}^2(59+24\mathbb{L})}{54} & -\frac{\lambda_{11}^2(5+2\mathbb{L})}{4} \\ \lambda\lambda_{11} & \frac{\lambda_{11}^2(5+2\mathbb{L})}{4} & -\frac{\lambda_{11}^2(59+24\mathbb{L})}{54} \end{pmatrix}. \tag{41}$$

The elements of the matrix come with three different parametric dependencies. The bottom 2×2 block is proportional to λ_{11}^2, the off-diagonals on the first row and column scale as $\lambda\lambda_{11}$

whereas the (11) element has two competing terms, controlled by the squares of each dimensionless parameter. If one makes the further assumption that $\lambda_{11} \sim 1$ as befits an otherwise undetermined dimensionless parameter, recalls that $\lambda \sim 0.1$ and evaluates the Wilson coefficient at the matching scale, one arrives at the prediction $(C_{Hd})_{1i} \sim (C_{Hd})_{i1} \sim 0.1(C_{Hd})_{jk}$, where $i \in \{1,2,3\}$ and $j,k \in \{2,3\}$. Despite its simplicity, this analysis shows that it is possible to predict the relative sizes of flavourful Wilson coefficients in our model whilst making minimal assumptions about the UV details of the FN model in question. Given two different UV complete models, where the hitherto unknown model parameters are fixed, the precise form of the predicted flavour structure could be used to discriminate between the two.

In the same vein, there are also hierarchies between different classes of SMEFT operators. Unsurprisingly, the relative sizes are dictated by the powers of λ at which the operators appear, and whether they appear at tree- or loop-level. There is for instance a hierarchy between the Wilson coefficients of the operator $(H^\dagger H)\overline{Q}_3 H d_1$ and that of $\left(H^\dagger i \overleftrightarrow{D}_\mu H\right)\left(\overline{d}_i \gamma^\mu d_j\right)$. In this case, the distinction between the two is caused by the former receiving contributions from tree-level matching and the latter only appearing at the one-loop level, leading to the $16\pi^2$ suppression factor.

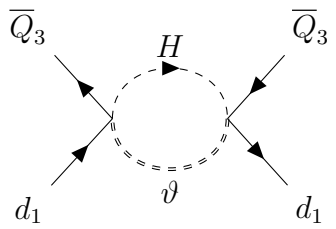
2.1.3 Dimension Six ($d_{\rm FN} = 6$)

Many more operators appear at mass-dimension 6. For example, the θ-dependent scalar potential and $\mathcal{S}_{\rm FN}$-type operators driving our UV flavour mechanism give 11 additional interactions on their own,

$$\begin{aligned}\mathcal{L}^6_{\rm FN} \supset\ & \frac{1}{\Lambda^2_{\rm UV}}\Big[\lambda_{03}(\theta^\star\theta)^3 + \lambda_{12}\big(H^\dagger H\big)(\theta^\star\theta)^2 + \lambda_{21}\big(H^\dagger H\big)^2(\theta^\star\theta)\Big] \\ & + \frac{1}{\Lambda^2_{\rm UV}}\left[c^{d_6}_{22}\,\overline{Q}_2 H d_2\,\theta^2 + c^{d_6}_{23}\,\overline{Q}_2 H d_3\,\theta^2 + c^{d_6}_{32}\,\overline{Q}_3 H d_2\,(\theta^\star\theta) + c^{d_6}_{33}\,\overline{Q}_3 H d_3\,(\theta^\star\theta) + \text{h.c.}\right],\end{aligned} \tag{42}$$

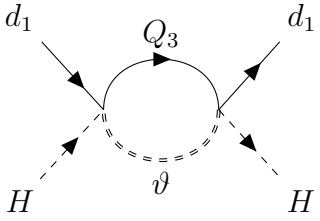
although this is but a subset of the full $d_{\rm FN} = 6$ Lagrangian, and would be yet further augmented upon considering $(\theta^\star\theta)$-enhanced renormalizable Yukawa terms for up quarks and leptons.

One option is to simplify the $d_{\rm FN} = 6$ analysis by only considering (e.g.) these operators in our matching since they are expected to be the most relevant for the flavour dynamics. However, doing so potentially misses important physics effects that might be sourced from energies around $\sim \Lambda_{\rm UV}$, where we remain purposefully agnostic about hypothetical UV completions to the FN class of EFTs. This might of course occur directly, upon integrating out some new vector-like fermion or novel (gauge) boson, or it could occur via Renormalization-Group-Evolution (RGE) effects sourced from running between the scales $\Lambda_{\rm UV}$ and $\Lambda_{\rm FN}$, where we recall that $\Lambda_{\rm FN}$ sits at or around the flavon mass. As is well known [75, 80, 81] in the case of (e.g.) the SMEFT, RGE mixes EFT operators via anomalous dimensions such that Wilson coefficients in the IR are functions of *multiple* Wilson coefficients (associated to different EFT operators) that may have existed in the UV. In other words, in the absence of an explicit UV completion, it is likely important to incorporate the complete basis of EFT operators in the IR (here meaning physics propagating around $\sim \Lambda_{\rm FN}$), such that said effects can be parameterized in the full set of otherwise unconstrained Wilson coefficients.



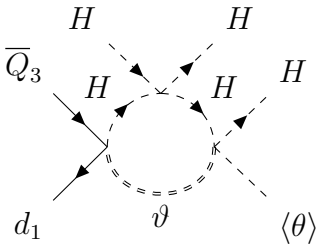
$$\longrightarrow \; -2C_{qd}^{3311}\big(\overline{Q}_3 d_1\big)\big(\overline{d}_1 Q_3\big) \overset{\text{Fierz}}{=} C_{qd}^{3311}\big(\overline{Q}_3\gamma^\mu Q_3\big)\big(\overline{d}_1\gamma_\mu d_1\big)$$

(a) A four-fermion operator at dimension-6 sourced by the FN model vertices of (33). The diagram illustrates why the resulting Wilson coefficient, shown in the last term of (39), is proportional to $|c_{31}^d|^2$.



$$\longrightarrow \quad (H^\dagger i\overset{\leftrightarrow}{D}_\mu H)\big(\overline{d}_1\gamma^\mu d_1\big)$$

(b) A contribution to the Higgs-derivative operator $(H^\dagger i\overset{\leftrightarrow}{D}_\mu H)\big(\overline{d}_1\gamma^\mu d_1\big)$ built out of dimension-5 FN vertices and thus proportional to $|c_{31}^d|^2$.



$$\longrightarrow \quad (H^\dagger H)\overline{Q}_3 H d_1$$

(c) A contribution to the Yukawa-like operator $(H^\dagger H)(\overline{Q}_3 H d_1)$ proportional to c_{31}^d.

Figure 3: Some Feynman diagrams contributing to the one-loop matching results coming out of a dimension-5 UV Lagrangian. We have verified the Wilson coefficients arising from the above diagrams manually and found agreement with `Matchete`.

A Complete and Minimal Basis

The construction of a complete, minimal, on-shell operator basis in EFTs is, a priori, tricky. However, assuming the charge assignments of Table 1 and that all other SM multiplets have trivial $\mathcal{S}_{\text{FN}}$ charges, one can readily calculate the Hilbert series $\mathbb{H}$ associated to the UV Lagrangian of our theory at successively higher orders. This then serves as a guide for the construction of a sensible basis before pursuing matching to the SMEFT, either at tree- or loop-level.

We have calculated the relevant Hilbert Series using `ECO` [82,83], an automated `form`-based package with a number of functionalities that builds on prior Hilbert Series results [84–86]. `ECO` includes the ability to add BSM field content as well as U(1) (global or gauged) BSM symmetries to the computation, and is furthermore able to apply integration-by-parts and EOM identities to reach a minimal on-shell basis. This is all we need in the simplified FN class of models we are currently considering. We collect our results for the Hilbert Series in

terms of the number of insertions of $\theta^{(\star)}$:[13]

$$\mathbb{H}(\{\Phi\}_{\rm SM},\theta)|_{d=5} = \theta\big(H\overline{Q}_3 d_1\big) + \theta^\star\big(H^\dagger Q_3 \overline{d}_1\big)\,, \tag{43}$$

$$\begin{aligned}\mathbb{H}(\{\Phi\}_{\rm SM},\theta)|_{d=6} &= \theta^2 H\overline{Q}_2(d_3+d_2) + \theta^{\star 2} H^\dagger Q_2\big(\overline{d}_3+\overline{d}_2\big)\\ &+ \theta^\star\theta\big(2G^2+2W^2+2B^2+H^2H^{\dagger 2}+H\overline{Q}_3(d_2+d_3)+H^\dagger Q_3(\overline{d}_2+\overline{d}_3)\big)\end{aligned} \tag{44}$$

$$+\,\theta^\star\theta\,p\left(\sum_{e,u,d,L,Q}\psi_i\overline{\psi_i}+\sum_{e,u,L}\psi_i\overline{\psi}_j+d_3\overline{d}_2+d_2\overline{d}_3+2pHH^\dagger\right)+(\theta^\star\theta)^2\big(p^2+HH^\dagger\big)+(\theta^\star\theta)^3\,,$$

where a summation over latin (flavour) indices and fermions $\psi\in\{e,u,d,...\}$ is implied in the third row of (44), where $i\neq j\in\{1,2,3\}$, where p stands for a (covariant) derivative, and where there are not yet terms (e.g.) $\propto\theta^\star\theta HLe$ or similarly for the up quarks (we have effectively set up and lepton $\mathcal{S}_{\rm FN}$ charges to zero for the above sample HS calculation, which also explains the off-diagonal flavour summation in the first bracketed term in the last line of (44)). Note that we have neglected terms with no $\theta^{(\star)}$-dependence.

We can now look at (43) and readily confirm that no other terms besides the FN operator in (32) and its Hermitian conjugate appear at $d_{\rm FN}=5$, while from (44) we see the three θ-dependent terms expected in the $d_{\rm FN}=6$ scalar potential,

$$\mathcal{O}_\theta=(\theta^\star\theta)^3,\qquad \mathcal{O}_{H\theta}=\big(H^\dagger H\big)^2(\theta^\star\theta)\,,\qquad \mathcal{O}_{\theta H}=\big(H^\dagger H\big)(\theta^\star\theta)^2, \tag{45}$$

as well as three θ-dependent kinetic operators

$$\mathcal{O}_{\theta\Box\theta}=(\theta^\star\theta)\Box(\theta^\star\theta)\,,\qquad \mathcal{O}_{H\Box\theta}=\big(H^\dagger H\big)\Box(\theta^\star\theta)\,,\qquad \mathcal{O}_{\theta\Box H}=(D^\mu H^\dagger D_\mu H)\theta^*\theta\,. \tag{46}$$

Additionally, we can draw one-to-one correspondences between many other terms in (44) and $d_{\rm SM}=6$ operators of the SMEFT Warsaw basis [8] (also see [87]), albeit with $H^{(\dagger)}\to\theta^{(\star)}$. Critically, the θ-enhanced Yukawa operators driving the FN mechanism of principal interest to our study are obvious in (44),

$$\mathcal{O}_{\overline{Q}_3 d_i\theta}=(\theta^\star\theta)\big(\overline{Q}_3 H d_2+\overline{Q}_3 H d_3\big)+\text{h.c.}\,,\qquad \mathcal{O}_{\overline{Q}_2 d_i\theta}=\theta^2\big(\overline{Q}_2 H d_3+\overline{Q}_2 H d_2\big)+\text{h.c.}\,, \tag{47}$$

and we also acknowledge that, for charged lepton and up quark flavours whose Yukawa elements are realized at the renormalizable level, additional operators will appear:

$$\mathcal{O}_{\overline{L}_i e_j\theta}=(\theta^\star\theta)\,\overline{L}_i H e_j+\text{h.c.}\,,\qquad \mathcal{O}_{\overline{Q}_i u_j\theta}=(\theta^\star\theta)\,\overline{Q}_i\tilde{H}u_j+\text{h.c.} \tag{48}$$

More care must of course be taken if the leading Yukawa structure in these family sectors appears at $d_{\rm FN}=6$, as in the (22) and (23) matrix elements of the toy model in (14), or if the Weinberg operator in (32) can only appear at $d_{\rm FN}=6$ via a $\theta^{(\star)}$ insertion (cf. Section 3 for more commentary on this point).

Six scalar-field strength interaction operators, the last three of which are CP violating, are also obvious in (44):

$$\mathcal{O}_{\theta G}=(\theta^\star\theta)\,G^A_{\mu\nu}G^{A\mu\nu},\qquad \mathcal{O}_{\theta W}=(\theta^\star\theta)\,W^I_{\mu\nu}W^{I\mu\nu},\qquad \mathcal{O}_{\theta B}=(\theta^\star\theta)\,B_{\mu\nu}B^{\mu\nu},$$

[13] Recall that the RHS of these equations are not to be interpreted as operators.

$$\mathcal{O}_{\theta\tilde{G}} = (\theta^\star\theta)\,\widetilde{G}^A_{\mu\nu}G^{A\mu\nu}, \qquad \mathcal{O}_{\theta\tilde{W}} = (\theta^\star\theta)\,\widetilde{W}^I_{\mu\nu}W^{I\mu\nu}, \qquad \mathcal{O}_{\theta\tilde{B}} = (\theta^\star\theta)\,\widetilde{B}_{\mu\nu}B^{\mu\nu}. \tag{49}$$

Note that, contrary to field strength operators constructed using the SM Higgs doublet, no cross term of the type $\theta^*\theta WB$ appears because such an operator would not be $\mathrm{SU}(2)_\mathrm{L}$ invariant.

We also have contact interactions between a flavon current and a fermion current,

$$\mathcal{O}_{\partial\overline{\psi}_i\psi_j} \in \theta^\star i\overset{\leftrightarrow}{\partial_\mu}\theta\left(\sum_{e,u,d,L,Q}\overline{\psi}_i\gamma^\mu\psi_i + \sum_{e,u,L}\overline{\psi}_i\gamma^\mu\psi_j + \overline{d}_2\gamma^\mu d_3 + \overline{d}_3\gamma^\mu d_2\right), \tag{50}$$

where a summation over latin (flavour) indices and fermions is again implied, and where the number and flavour structure of off-diagonal elements in the latter three terms is subject to modification depending on the complete set of unspecified $\mathcal{S}_\mathrm{FN}$ charges for the charged leptons and up-type quarks. In total, (45)–(50) represent the θ-dependent operators of the non-redundant, flavour-dependent $d_\mathrm{FN} = 6$ UV FN Lagrangian.[14]

There are of course additional contributions from all $\mathcal{S}_\mathrm{FN}$-symmetric SMEFT operators at dimension six [8], $\mathcal{L}^6_{\mathrm{SMEFT}'}$, which have no θ-dependence. While in the absence of a complete set of $\mathcal{S}_\mathrm{FN}$ charges for all SM multiplets it is again impossible to fully enumerate $\mathcal{L}^6_{\mathrm{SMEFT}'}$ (or analogous bases at higher mass dimensions for that matter), they will generically include a host of flavour-symmetric four-fermion (ψ^4), Higgs-enhanced Yukawa (ψ^2H^3),[15] dipole-type (ψ^2XH), and Higgs-kinetic (ψ^2H^2D) operators, in addition to purely bosonic structures. However, we will only attempt tree-level matching at $d_\mathrm{FN} = 6$, and hence we only need (45)–(50) for our calculations.

Furthermore, given that we are matching in the *broken* $\mathcal{S}_\mathrm{FN}$ phase, it turns out that we will only need a subset of the operators enumerated between (45)–(50). Indeed, when θ can effectively be treated as a real scalar, the operators of (50) go to zero due to their left-right derivative structure, as do the first (θ-dependence only) and third (mixed θ-H contribution) kinetic terms in (46). We have performed an additional Hilbert Series calculation confirming the results in [87], which found a minimal EFT basis of 18 operators at $d = 6$ for a real, gauge-singlet scalar enhancement of the SM, since our results in (45)–(50) map to [87] up to flavour-symmetry assumptions. The $\mathcal{S}_\mathrm{FN}$-breaking expansion of θ about the true vacuum transforms dimension-6 operators of the type $(\theta^*\theta)\times\mathcal{O}^{d=4}$, which constitute the majority of the basis in (45)–(50), into dimension-5 operators of the type $v_\theta\vartheta\times\mathcal{O}^{d=4}$. It is this set of dimension-5 operators that is responsible for generating EFT operators when matching at tree-level to the SMEFT: diagrammatically, one connects the flavon ϑ of the dimension-5 vertex to the three-point vertex $-\lambda_{11}v_\theta H^\dagger H\vartheta$ in a fashion illustrated in Fig. 4 where we have drawn a matching contribution to the SMEFT operator $(H^\dagger H)\overline{Q}_2Hd_2$.

In summary, we find that there is a subset of 18 operators in (45)–(50) that are required when matching the $d_\mathrm{FN} = 6$ toy FN model to the SMEFT, not including the unspecified number of $(\theta^\star\theta)\left[\overline{Q}_i\tilde{H}u_j + \overline{L}_iHe_j\right]$+h.c. operators allowed. We will now discuss the IR SMEFT operators that appear as each class of UV operators in said subset turns on.

[14] See Section 3 for a discussion of other operators that can appear when we generalize the $\mathcal{S}_\mathrm{FN}$ charges.

[15] After $\mathcal{S}_\mathrm{FN}$ symmetry breaking, these types of operators lead to $\mathcal{O}(\lambda^2)$ corrections to the (23) and (33) entries of M^d, for instance. Higher-order corrections of this type appear generically in FN-type models and have implications for flavoured phenomenology, despite often being neglected.

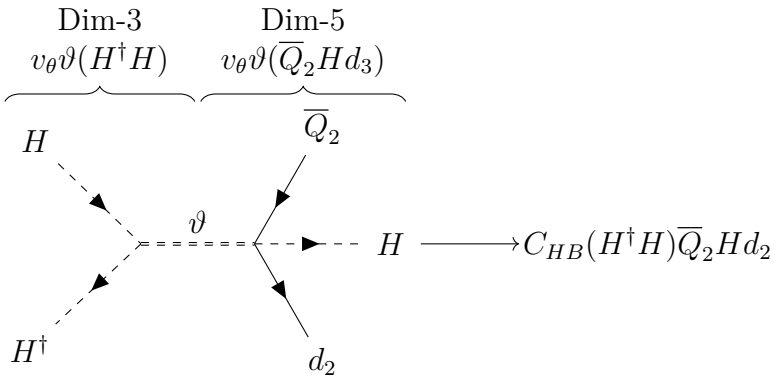


Figure 4: A diagrammatic illustration of how tree-level matching of the FN model yields the SMEFT operator $(H^\dagger H)\overline{Q}_2 H d_2$. A flavon line is used to glue together a dimension-3 vertex and a dimension-5 vertex both arising from the $\mathcal{S}_{\rm FN}$ symmetry breaking; the dimension-5 vertex in the figure is sourced from a $d_{\rm FN} = 6$ operator $(\overline{Q}_2 H d_2)\theta^2$ in the unbroken phase. The full Wilson coefficient C_{HB}, which obtains a second contribution from a diagram not drawn here, is written explicitly in (56).

Matching Scalar Potential and Derivative Operators

As is well known for the SMEFT in the broken EW phase, the presence of higher-order scalar potential operators shifts the minimum of the classical vacuum. This is also true for the FN EFT we are considering here, with $\mathcal{O}_\theta$ responsible for the shift in the flavon vev. Following the analogous discussion in [75] concerning the Higgs, the θ vev from (9) is modified to

$$\langle \theta^\star \theta \rangle = \frac{v_\theta^2}{2}\left(1 + \frac{3\,\lambda_{03}\,v_\theta^4}{4\,\mu_\theta^2\,\Lambda_{\rm UV}^2}\right) \equiv \frac{\overline{v}_\theta^2}{2} \tag{51}$$

at leading order in the implied expansion parameters on the right-hand side of the equation. We can then re-expand the complete scalar potential from (5), now including all of the operators in (45), about its new vacuum,

$$H \to H, \qquad \theta \to (\overline{v}_\theta + \vartheta)/\sqrt{2}\,, \tag{52}$$

to obtain the potential in the broken $\mathcal{S}_{\rm FN}$ phase analogous to (16),

$$\begin{aligned}
\mathcal{L}^6_{\not{\rm FN}} \supset\; & -\frac{\overline{m}_\theta^2}{8\overline{v}_\theta^2}\vartheta^4 - \frac{\overline{m}_\theta^2}{2\,\overline{v}_\theta}\vartheta^3 - \left(\frac{\lambda_{11}}{2}H^\dagger H + \frac{\overline{m}_\theta^2}{2}\right)\vartheta^2 - \overline{v}_\theta \lambda_{11} H^\dagger H\,\vartheta - \frac{\overline{v}_\theta^2 \lambda_{11}}{2}H^\dagger H + \frac{\overline{v}_\theta^2 \lambda_{21}}{2\Lambda_{\rm UV}^2}\big(H^\dagger H\big)^2 \\
& + \frac{1}{\Lambda_{\rm UV}^2}\left\{\frac{\lambda_{03}}{8}\vartheta^6 + \frac{3\overline{v}_\theta \lambda_{03}}{4}\vartheta^5 + \left(\frac{\lambda_{12}}{4}H^\dagger H + \frac{3\,\overline{v}_\theta^2\,\lambda_{03}}{2}\right)\vartheta^4 + \big(\overline{v}_\theta\,\lambda_{12}\,H^\dagger H + \overline{v}_\theta^3\,\lambda_{03}\big)\vartheta^3 \right. \\
& \left. + \left[\frac{3\overline{v}_\theta^2 \lambda_{12}}{2}H^\dagger H + \frac{\lambda_{21}}{2}(H^\dagger H)^2\right]\vartheta^2 + \big[\overline{v}_\theta^3 \lambda_{12} H^\dagger H + \overline{v}_\theta\,\lambda_{21}(H^\dagger H)^2\big]\vartheta + \frac{1}{4}v_\theta^4 \lambda_{12} H^\dagger H\right\},
\end{aligned} \tag{53}$$

where we have ignored static terms, have only kept contributions up to linear order in the dimension-6 Wilson coefficients of (45), and have defined the modified flavon mass as

$$\overline{m}_\theta^2 \equiv 2\,\mu_\theta^2 - \frac{3\,v_\theta^4\,\lambda_{03}}{2\,\Lambda_{\rm UV}^2}. \tag{54}$$

Upon implementing (53) and $\mathcal{O}_{H\Box\theta}$ from (46) in `Matchete`, we then match to $d_{\rm SM} = 6$ at tree-level, obtaining

$$\begin{aligned}
\mathcal{L}_{\rm SMEFT} \supset\, & +\frac{\overline{\lambda}^2\lambda_{12}\overline{v}_\theta^2}{2}H^\dagger H + \frac{1}{\overline{m}_\theta^2}\Big[\overline{\lambda}^2\big(\lambda_{21}m_\theta^2 - 2\lambda_{11}\lambda_{12}\overline{v}_\theta^2\big) + 2\overline{\lambda}^4\lambda_{12}^2\overline{v}_\theta^2\Big]\big(H^\dagger H\big)^2 \\
& + \left[\frac{2c_{31}^d\overline{\lambda}^3\lambda_{12}}{\overline{m}_\theta^2}\big(H^\dagger H\big)\overline{Q}_3 H d_1 + \text{h.c.}\right] - \frac{2}{m_\theta^6}\Big[\lambda_{11}\overline{\lambda}^2\big(\lambda_{03}\lambda_{11}^2\overline{v}_\theta^4 - \lambda_{11}\lambda_{12}m_\theta^2\overline{v}_\theta^2 + \lambda_{21}m_\theta^4\big) \\
& -2\overline{\lambda}^4\big(3\lambda_{03}\lambda_{11}^2\lambda_{12}\overline{v}_\theta^4 - 2\lambda_{11}\lambda_{12}^2\overline{m}_\theta^2\overline{v}_\theta^2 + \lambda_{12}\lambda_{21}\overline{m}_\theta^4\big) + \mathcal{O}\Big(\overline{\lambda}^6\Big)\Big]\big(H^\dagger H\big)^3 \\
& + \frac{2}{\overline{m}_\theta^4}\Big[\lambda_{11}\overline{\lambda}^2\big(\lambda_{12}\overline{v}_\theta^2 - c_{H\Box\theta}\overline{m}_\theta^2\big) - \lambda_{12}\overline{\lambda}^4\big(\lambda_{12}\overline{v}_\theta^2 - 2c_{H\Box\theta}m_\theta^2\big)\Big]\big(H^\dagger H\big)\Box\big(H^\dagger H\big)\,,
\end{aligned} \tag{55}$$

where we have defined the shifted FN expansion parameter $\overline{\lambda} \equiv \overline{v}_\theta/(\sqrt{2}\Lambda_{\rm UV})$ and expanded the results in powers of this order $\mathcal{O}(0.1)$ parameter. The expansions are shown up to and including order $\overline{\lambda}^4$ terms, but we note that contributions from dimension-8 operators in the scalar potential will also enter at this order and are expected to augment the results. Note in (55) the renormalizations of the Higgs potential and kinetic terms, in addition to the novel Higgs-flavon mixing term impacting the (31) Yukawa coupling that first gets turned on at $d_{\rm FN} = 5$. We emphasize that the shifted vacuum in (51) will also impact the lower-order matching presented in above sections, but do not show these results explicitly, for brevity. Furthermore, the shifts in $\overline{m}_\theta^2$ and $\overline{v}_\theta$ are formally of order $(v_\theta/\Lambda_{\rm UV})^2$ and so enter at the same order as dimension-8 SMEFT operators.

Matching Yukawa-Like FN Operators

Continuing with the Yukawa-like operators in (47), our tree-level `Matchete` matching analysis gives

$$\begin{aligned}
\mathcal{L}_{\rm SMEFT} \supset & \Big[\overline{Q}_2 H\big(c_{22}^{d_6}\,d_2 + c_{23}^{d_6}\,d_3\big) + \overline{Q}_3 H\big(c_{32}^{d_6}\,d_2 + c_{33}^{d_6}\,d_3\big)\Big] \\
& \times \left[\overline{\lambda}^2 + \frac{2\overline{\lambda}^2\Big(2\overline{\lambda}^2\,\lambda_{12} - \lambda_{11}\Big)}{\overline{m}_\theta^2}\big(H^\dagger H\big)\right] + \text{h.c.}
\end{aligned} \tag{56}$$

As in (33), in the first bracketed term we have kept the explicit FN contributions to the down-type Yukawas, responsible for the texturing of mass and mixing in the family. The second bracketed term represents the novel contributions to the Higgs-enhanced Yukawa operators of

the $d_{\rm SM} = 6$ SMEFT sourced by the $d_{\rm FN} = 6$ FN theory. Comparing the dimension-6 Higgs-enhanced Yukawa terms for the (31) and (22) flavour index pairs in (34) and (56), respectively, we see that the latter Wilson coefficient is suppressed by one more power of λ. This finding generalizes to matching to the SMEFT at higher $d_{\rm FN}$; as higher-dimension FN terms, of type $\overline{\psi}_i H \psi_j \theta^n$, are added to the UV Lagrangian and matching is done at tree-level, the Wilson coefficients of the resulting Higgs-enhanced Yukawa operators show the same hierarchies in $\overline{\lambda}$ as the Yukawa matrices themselves. We note that this is consistent with the Minimally-Flavour-Violating expectation from [88], where Yukawa matrices (and their flavour patterns) source all flavour violation BSM. The toy model thus predicts the following hierarchies in the down-type Higgs-enhanced Yukawas:

$$(C_{dH})_{ij} \sim \begin{pmatrix} \overline{\lambda}^4 & \overline{\lambda}^3 & \overline{\lambda}^3 \\ \overline{\lambda}^3 & \overline{\lambda}^2 & \overline{\lambda}^2 \\ \overline{\lambda} & \overline{\lambda}^0 & \overline{\lambda}^0 \end{pmatrix}, \tag{57}$$

where the $\overline{\lambda}^0$ terms come from $\mathcal{L}_{\rm SMEFT'}$, the part of the $\mathcal{S}_{\rm FN}$-symmetric UV Lagrangian without factors of θ, which contains $(C_{dH})_{33}$ and $(C_{dH})_{32}$ with order one coefficients.

Finally, for the flavour scenario where charged-lepton Yukawa couplings appear at the renormalizable level, the tree-level matching effects sourced from the $d_{\rm FN} = 6$ operators in (48) are given by

$$\mathcal{L}_{\rm SMEFT} \supset c^{e_6}_{ij} \overline{L}_i \, H \, e_j \left[\overline{\lambda}^2 + \frac{2\overline{\lambda}^2 \left(2\overline{\lambda}^2 \, \lambda_{12} - \lambda_{11} \right)}{\overline{m}^2_\theta} \left(H^\dagger H \right) \right] + \text{h.c.} \tag{58}$$

where we see the exact same structure as above for the (flavour-specific) down-quark sector. We expect that the same structure appears for the up-quarks.

Matching Field-Strength Operators

Finally, upon including (49) into the matching analysis, one finds the following contributions to the $d_{\rm SM} = 6$ SMEFT:

$$\mathcal{L}_{\rm SMEFT} \supset \frac{2\overline{\lambda}^2 \left(2\overline{\lambda}^2 \, \lambda_{12} - \lambda_{11} \right)}{\overline{m}^2_\theta} \left(H^\dagger H \right) \left(\overset{(\sim)}{C}_{\theta B} \overset{(\sim)}{B}_{\mu\nu} B^{\mu\nu} + \overset{(\sim)}{C}_{\theta W} \overset{(\sim)}{W}_{\mu\nu} W^{\mu\nu} + \overset{(\sim)}{C}_{\theta G} \overset{(\sim)}{G}_{\mu\nu} G^{\mu\nu} \right), \tag{59}$$

with the $(\sim)$ notation implying that both the H^2X^2 and $H^2\tilde{X}X$ field-strength operators are sourced with the same Wilson coefficients. The terms in (59) are again implied in addition to the operators presented above. We see contributions to the Higgs-enhanced gluonic field strength tensors of the Warsaw basis for the first time, while the contributions to the electroweak analogues are in addition to those already found in Section 2.1.2, sourced at one-loop from $d_{\rm FN} = 4$ operators. For those choices of parameters where the loop-matched Higgs-gauge Wilson coefficients of (28) are larger than those shown above, the model then predicts

a relative size difference in the Higgs-gauge operators of

$$\frac{C_{HW}}{C_{HG}} = \frac{g_Y^2 \lambda_{11} \overline{v}_\theta^2}{384\pi^2 \overline{m}_\theta^2 \overline{\lambda}^2} \sim \frac{g_Y^2 \lambda_{11}}{40}\,, \quad \frac{C_{HB}}{C_{HG}} = \frac{g^2 \lambda_{11} \overline{v}_\theta^2}{384\pi^2 \overline{m}_\theta^2 \overline{\lambda}^2} \sim \frac{g^2 \lambda_{11}}{40}\,, \tag{60}$$

where we have, once again, assumed $\overline{m}_\theta \sim \overline{v}_\theta$ and $\overline{\lambda} \sim 0.1$. However, given that one typically takes $\lambda_{11} \lesssim 1$, it becomes clear that it is more sensible to assume that the tree-level result of (59) dominates over the loop-level contribution and one therefore predicts $C_{HW} \sim C_{HB} \sim C_{HG}$.

2.1.4 Matching Summary and Phenomenological Outlook

The results above conclude our matching of the toy FN model implied in (14) to the SMEFT up to $d_{\rm SM} = 6$. In summary, we have considered the one-loop (tree-level) matching contributions to the SMEFT sourced by operators up to $d_{\rm FN} = 5$ (6) in the FN theory. The results in (21), (26)–(31), (34)–(36), (39), (55)–(56) and (59) represent the core findings of our study, and reveal that Higgs-scalar (H^6), Higgs-derivative (H^4D^2), Higgs-kinetic ($H^2D\psi^2$), Higgs-enhanced Yukawa ($H^3\psi^2$), Higgs-gauge (H^2X^2) and four-fermion (ψ^4) operators are all sourced with non-zero Wilson coefficients in the IR SMEFT. Their specific forms are also summarized in Tables 2–8.

While a dedicated phenomenological study of these results is beyond our scope in this paper, we have already attempted to highlight some of the most relevant signatures in the text above, as results have appeared. Here we simply summarize possible directions to falsify the FN mechanism with the SMEFT, recalling that falsification includes not only the FN mechanism itself, where $\mathcal{S}_{\rm FN}$-symmetry breaking occurs via a BSM flavon, but also the ability to distinguish one FN-type flavour model from another.

Concerning the former, the first and perhaps most obvious point to make regards which operators do *not* appear in Sections 2.1.1–2.1.3. Indeed, at least up to the orders we have considered (in the UV and IR), we do not see dipole-type (ψ^2XH) nor triple-field-strength (X^3) operators appearing in our matching analysis, although we do of course recall that such operators can appear in $\mathcal{L}_{\rm SMEFT'}$ and will filter down to the IR theory if so. This will be ubiquitous for the triple-field-strength operators which have no flavour content, but only flavour-specific combinations of the dipole SMEFT′ operators will appear (the same flavour combinations as allowed at the renormalizable level in (14)). Hence the presence of sizable non-zero Wilson coefficients in dipole operators outside of these limited combinations would signal a deviation and/or extension from the FN-type dynamics we have studied here.

Similarly, the generic FN mechanism relies on the presence of a BSM flavon and, as we have seen in Sections 2.1.1–2.1.3, its mixing with the SM Higgs boson, either via the renormalizable Higgs portal $\propto \lambda_{11}$ in (15) or its $d_{\rm FN} = 6$ portal analogues $\propto \lambda_{12,21}$ in (45), plays a major role in our matching analysis. In fact, we already pointed out that a non-zero $C_{H\Box}$ ($\propto \lambda_{11}^2$) is generated from renormalizable operators at tree level in (21), and leads to a straightforward flavon mass bound of $m_\theta \gtrsim 1$ TeV (also see [13]). Better bounds on $C_{H\Box}$ from (e.g.) Higgs coupling analyses at the LHC will immediately translate to stricter bounds on the two-dimensional λ_{11}-m_θ parameter space. We also notice that at $d_{\rm FN} = 6$ all of the newly obtained Wilson coefficients contain terms proportional to (powers of) $\overline{\lambda}^2\lambda_{12}$ whose origin is the $d_{\rm FN} = 6$

Higgs-flavon scalar potential portal term $(H^\dagger H)(\theta^*\theta)^2$ which, after $\mathcal{S}_{\rm FN}$ symmetry breaking, takes the form

$$\frac{\lambda_{12}}{\Lambda_{\rm UV}^2}(H^\dagger H)(\theta^*\theta)^2 \to 2\lambda_{12}\overline{\lambda}^2 v_\theta \vartheta H^\dagger H + \dots \tag{61}$$

This λ^2-suppressed three-point vertex can be used to build Feynman diagrams just like the familiar $\lambda_{11} v_\theta \vartheta H^\dagger H$ term from (16), explaining its frequent appearance in the Wilson coefficients of this section. These terms thus help generate non-trivial dimension-6 SMEFT operators even when the renormalizable Higgs-flavon mixing parameter λ_{11} is small. In general though, it is clear that in the absence of any sizable Higgs-flavon mixing term in the IR, regardless of its UV operator source, the number of phenomenologically relevant operators appearing in Sections 2.1.1–2.1.3 is greatly reduced, as expected.

On the other hand, phenomenologically relevant BSM effects can also be sourced from four-fermion interactions, and we observe from (31) and (39) that these are turned on at one-loop level in our FN setup by both $d_{\rm FN} = 4$ and $d_{\rm FN} = 5$ operators. Such four-fermion terms can contribute to flavour-changing neutral-current (FCNC) observables, such as neutral meson mixing, and often produce the most stringent constraints on flavourful BSM models. For the FN model in question the four-fermions may thus be expected to provide the leading lower bound on m_θ. However, as evident in (31) and (39), the four-fermion operators are suppressed not only by powers of m_θ but also $16\pi^2$ and (powers of) λ. Furthermore, by construction, the FN charge assignments ensure that four-fermion operators involving the lightest quarks, which contribute to increased neutral meson mixing, will be more suppressed than those containing the heavy quarks. One caveat of this analysis is that to achieve a realistic comparison with the FCNC bounds, the fermions would have to be rotated from the interaction eigenbasis to the mass eigenbasis. Setting the $c_{ij}^{d_i}$ coefficients to unity in the down-quark Yukawa matrix of (14) yields a rank-1 matrix with a degenerate eigenspace, and so the quark rotation matrices cannot be uniquely determined without making more elaborate assumptions about the Wilson coefficients. We do, of course, expect nearly diagonal quark rotation matrices, with the off-diagonal elements controlled by increasing powers of λ. All in all, the suppression structures discussed in this section suggest that the commonly-held assumption that the presence of four-fermion operators will push m_θ to PeV scales may require further thought in FN-type setups.

This possibility highlights the importance of studying SMEFT operators besides the four-fermion ones, given that other Wilson coefficients may exhibit significantly less suppression after the FN EFT is matched to the SMEFT at a given order. For example, measuring Higgs observables that acquire corrections from the Higgs-enhanced Yukawa terms found above (or the Higgs-kinetic operators $\sim \psi^2 H^2 D$), could yield constraints on m_θ and other FN parameters that are competitive with FCNC bounds. Furthermore, the form of the Higgs-enhanced Yukawas following from our matching shows that their Wilson coefficients exhibit the same flavour hierarchies, organised in powers of λ, as the Yukawa elements themselves. This indicates a distinct phenomenological SMEFT signature in the FN setup. More broadly, our results also indicate the importance of matching UV theories beyond tree level, and also potentially motivate studying the SMEFT output to higher orders than $d_{\rm SM} = 6$ — one may find that less parametrically suppressed higher-dimension operators contribute as much to the IR physics as more parametrically suppressed lower-dimension ones.

With respect to distinguishing different FN models in the Abelian $\mathbb{Z}_{\rm FN}$ -type constructions presented above, it is (unsurprisingly) clear that precision flavour constraints are required. After all, $\mathcal{S}_{\rm FN}$ is a UV flavour symmetry! As already discussed in Section 2.1.2, the differences in the parametric dependence of the flavourful Wilson coefficients appearing in Sections 2.1.1–2.1.3 suggest that their relative sizes may be used to distinguish independent FN models — one UV charge scenario will generate different hierarchies to another in a self-evident manner. Consider (e.g.) the final four-fermion operator $\propto (\overline{Q}_3\gamma_\mu Q_3)(\overline{d}_1\gamma^\mu d_1)$ in (39). That this particular flavour ensemble appears with no λ_{11}-dependence is evident from Table 1 and the loop diagram in Fig. 3(a). Were the charges of a particular UV model aligned such that different quark flavours $\{i,j\}$ leave the combination $\overline{Q}_i H\theta d_j$ $\mathcal{S}_{\rm FN}$-invariant, then those flavours would be the ones apparent in the SMEFT. Similar suppression patterns can be found and/or anticipated throughout all of the flavoured SMEFT operators that appear.

We also find it important to recall the obvious; the main goal of FN-type constructions as introduced in Section 2.1 is to dynamically explain observed patterns of fermionic mass, mixing, and CP violation. Hence data on extracted quark mass eigenvalues and CKM mixing angles and phases from (e.g.) the PDG [2] serves as the first point of phenomenological comparison between different FN models. However, care must be taken in the context of the SMEFT, given that the Higgs-enhanced Yukawa operators that appear throughout our matching (cf. e.g. (26) or (34)) will, after EWSB, effectively renormalize the Yukawa couplings and therefore mass and mixing outputs. Indeed, in [38] the definitions of mass eigenvalues and CKM parameters are given *exactly* in terms of Yukawa couplings, and these relations hold regardless of how many $(H^\dagger H)$ insertions serve to enhance these parameters' formal theory definitions. Furthermore, [31] correctly notes that, upon extracting a mass or mixing parameter by comparing theory (the SMEFT, at a given mass dimension) to data, there are indirect SMEFT constraints sourced in all processes that then depend on those extracted parameters, e.g. heavy flavour or W-boson decays.

We are of course aware that these speculations do not amount to a robust phenomenological study of the matching analysis in Sections 2.1.1–2.1.3. Such a study would ideally incorporate both a likelihood analysis with respect to existing global $d_{\rm SM} = 6$ SMEFT fits (see e.g. [13–16]), as well as more tailored studies with automated optimization tools like `smelli` [89], which further matches SMEFT Wilson coefficients / operators to the LEFT and subsequently compares them against low-energy flavour data. We will briefly reemphasize this point in our Summary and Outlook in Section 4.

3 Generalizations and Extensions Beyond the Toy Model

In the analysis of Section 2 we have assumed FN dynamics only in the down-quark sector and in a particular charge setup (cf. (15)). We will now discuss generalizations and extensions to this toy model analysis that more completely characterize the matching of generic FN models to the SMEFT. Note that our goal here is not to explore every scenario in full generality as above (for example, loop-level matching will be largely unconsidered in what follows) but rather to highlight the most relevant and/or obvious points regarding realistic FN theories beyond Table 1, many of which will evolve into future study.

3.1 Charge Generalizations and Extensions to Different Families

The model encapsulated in (14) is but one of many possible choices in an effort to resolve down quark mass hierarchies. Furthermore, we have left the $\mathbb{Z}_{\rm FN}$ charges for the up quarks and charged leptons unspecified in our matching procedures above, having simply allowed for the presence of renormalizable couplings in these sectors without the assumption that FN dynamics successfully structure their Yukawas. In this section we discuss the generalization of **(1)** the assignment of $\mathbb{Z}_{\rm FN}$ charges within our toy model and the extension of structured $\mathbb{Z}_{\rm FN}$ charges to **(2)** the up quark and **(3)** lepton families.

Generalization of Down Quark Charges

The simplest generalization to Section 2 is to still consider only down-type FN dynamics, but to allow for charge assignments beyond those in Table 1. It is clear that, in such a scenario, the results of Section 2.1 will hold, up to the explicit flavour labels that appear, which will of course be altered to reflect the new model's charges.

For example, everywhere one sees terms $\propto y^d_{32,33}$ in (e.g.) (26) or (27), one would instead expect to see terms $\propto y^d_{ij}$ in the new model, where y^d_{ij} are any renormalizable Yukawa matrix elements $\propto \lambda^0$ appearing in the model — cf. (14). One would clearly also expect straightforward modifications to the Wilson coefficients c^d_{ij} that appear at different mass dimensions in (14), and therefore contribute with different relative magnitudes to the Wilson coefficients of the SMEFT. We have been careful when presenting our results in Section 2 to point out those features which are readily generalizable to different charge configurations, and in fact Tables 4–8 are already presented in a fully generic flavour notation.

One might suspect that an interesting extension of the results in Section 2 could appear if the charges populating the $d_{\rm FN} = 5$ Yukawa sector were extended such that two or more Wilson coefficients akin to c^d_{31} appear in the UV Lagrangian. However, we have checked this possibility by imagining the (23) matrix element is λ (as opposed to λ^2) suppressed. We find that the only additional SMEFT operators that appear are given by

$$\mathcal{L}_{\rm SMEFT} \supset -\frac{\lambda_{11}\lambda\, c^d_{23}}{m^2_\theta}\big(H^\dagger H\big)\overline{Q}_2 H d_3 + \text{h.c.}, \tag{62}$$

when matching to the $d_{\rm SM} = 6$ SMEFT at tree level. These conclusions are subject to modification when matching at $d_{\rm SM} \geq 6$ and/or when considering loop-level matching effects.

Extension to Up Quarks

A complete description of quark flavour physics requires $\mathcal{S}_{\rm FN}$ dynamics in both the up and down quark sector. On the one hand, the up quark family exhibits the largest mass hierarchy, given a top mass on the order of the Higgs vev. On the other hand, the CKM matrix is constructed out of a product of two unitary matrices which transform the up- and down-type left-handed fermions from the gauge eigenbasis to the mass eigenbasis, $V_{\rm CKM} = V^\dagger_{u_L} V_{d_L}$. Each element of the matrix is measurable in charged-current interactions and ties together the up- and down-quark Yukawa matrices.

The simplest way to study the up quarks is to assign them a set of charges in complete analogy to Table 1, accounting, of course, for the fact that the Higgs boson enters the Yukawa

Tree-level matching at $d_{\rm FN}=4$

Operator	Wilson coefficient
$H^\dagger H$	$-\dfrac{\lambda_{11}v_\theta^2}{2}$
$(H^\dagger H)^2$	$\dfrac{\lambda_{11}^2 v_\theta^2}{2m_\theta^2}$
$\big(H^\dagger H\big)\Box\big(H^\dagger H\big)$	$-\dfrac{\lambda_{11}^2 v_\theta^2}{2m_\theta^4}$

Tree-level matching at $d_{\rm FN}=5$

Operator	Wilson coefficient
$\overline{Q}_i H d_j$	$c_{ij}^d\,\lambda$
$(H^\dagger H)\overline{Q}_i H d_j$	$-\dfrac{\lambda_{11}\,\lambda\,c_{ij}^d}{m_\theta^2}$

Table 2: The (SM)EFT operators and associated Wilson coefficients appearing from a tree-level matching of the $d_{\rm FN}=4(5)$ FN Lagrangian in (15) and (32), respectively.

Tree-level matching at $d_{\rm FN}=6$

Operator	Wilson coefficient
$H^\dagger H$	$\dfrac{\overline{\lambda}^2\lambda_{12}\overline{v}_\theta^2}{2}$
$\big(H^\dagger H\big)^2$	$\dfrac{1}{\overline{m}_\theta^2}\Big[\overline{\lambda}^2\big(\lambda_{21}m_\theta^2-2\lambda_{11}\lambda_{12}\overline{v}_\theta^2\big)+2\overline{\lambda}^4\lambda_{12}^2\overline{v}_\theta^2\Big]$
$\big(H^\dagger H\big)^3$	$-\dfrac{2}{m_\theta^6}\Big[\lambda_{11}\overline{\lambda}^2\big(\lambda_{03}\lambda_{11}^2\overline{v}_\theta^4-\lambda_{11}\lambda_{12}m_\theta^2\overline{v}_\theta^2+\lambda_{21}m_\theta^4\big)$
	$-2\overline{\lambda}^4\big(3\lambda_{03}\lambda_{11}^2\lambda_{12}\overline{v}_\theta^4-2\lambda_{11}\lambda_{12}^2\overline{m}_\theta^2\overline{v}_\theta^2+\lambda_{12}\lambda_{21}\overline{m}_\theta^4\big)+\mathcal{O}\Big(\overline{\lambda}^6\Big)\Big]$
$\big(H^\dagger H\big)\Box\big(H^\dagger H\big)$	$\dfrac{2}{\overline{m}_\theta^4}\Big[\lambda_{11}\overline{\lambda}^2\big(\lambda_{12}\overline{v}_\theta^2-c_{H\Box\theta}\overline{m}_\theta^2\big)-\lambda_{12}\overline{\lambda}^4\big(\lambda_{12}\overline{v}_\theta^2-2c_{H\Box\theta}m_\theta^2\big)\Big]$
$\overline{Q}_i H d_j$	$c_{ij}^{d_6}\,\overline{\lambda}^2$
$\big(H^\dagger H\big)\overline{Q}_i H d_j$	$\dfrac{2\,c_{ij}^d\,\overline{\lambda}^3\,\lambda_{12}}{\overline{m}_\theta^2}+\dfrac{2c_{ij}^{d_6}\overline{\lambda}^2\big(2\overline{\lambda}^2\,\lambda_{12}-\lambda_{11}\big)}{\overline{m}_\theta^2}$
$\big(H^\dagger H\big)\overset{(\sim)}{X}_{\mu\nu}X^{\mu\nu}$	$\overset{(\sim)}{C}_{\theta X}\,\dfrac{2\overline{\lambda}^2\Big(2\overline{\lambda}^2\,\lambda_{12}-\lambda_{11}\Big)}{\overline{m}_\theta^2}$

Table 3: The SMEFT operators appearing from a tree-level matching of the $d_{\rm FN}=6$ FN Lagrangian. Recall the barred notation accounts for corrections due to expanding about the new minimum of the associated scalar potential. In the sixth and seventh lines, the notation in the RH column indicates that the index pattern on c_{ij}^d and $c_{ij}^{d_6}$ follows that of the indices on $\overline{Q}_i$ and d_j in the left-hand column. Finally, $X\in\{B,W,G\}$ in the final line, with tilded notation following that in the text around (59).

One-loop matching at $d_{\rm FN}=5$ — $\psi^2 H$ (Yukawas)

Operator	Wilson coefficient
$\overline{L}_i H e_j$	$\dfrac{\hbar y^e_{ij}\lambda_{11}^2 v_\theta^2}{12 m_\theta^4}\big[3m_\theta^2 - m_H^2(17+6\mathbb{L})\big]$
$\overline{Q}_i H d_j$	$\dfrac{\hbar}{12 m_\theta^4 v_\theta^2}\Big[6c^d_{ij}\lambda m_\theta^2\big(3m_\theta^4(1+\mathbb{L}) - 2\lambda_{11}v_\theta^2 m_\theta^2(1+\mathbb{L}) + \lambda_{11}m_H^2 v_\theta^2(3+2\mathbb{L})\big)$ $+ y^d_{ij}\lambda_{11}^2 v_\theta^4\big(3m_\theta^2 - m_H^2(17+6\mathbb{L})\big)\Big]$
$\overline{Q}_i \tilde{H} u_j$	$\dfrac{\hbar y^u_{ij}\lambda_{11}^2 v_\theta^2}{12 m_\theta^4}\big[3m_\theta^2 - m_H^2(17+6\mathbb{L})\big]$

Table 4: The modified renormalizable Yukawa terms at one-loop. Whilst not SMEFT operators, we have presented them here to illustrate the one-loop corrections to the Yukawa matrices. Though not shown here for brevity, one also finds corrections to the coefficients of the renormalizable H^2 and H^4 operators.

elements in question as $\tilde{H}$, which has charge $+3$ under $\mathcal{S}_{\rm FN}$. For instance, in order to achieve the same hierarchies in the up and down quark Yukawa matrices, the right-handed up-type quarks could be assigned charges $q_{u_1}=-1$, $q_{u_2}=-3$ and $q_{u_3}=-3$ under the Abelian symmetry. We stay rather agnostic about the up-sector charges in this section and simply assume, in analogy to the down-sector, that some up-type Yukawa elements are allowed at the tree-level, denoted by y^u_{ij}, that others enter at dimension-5 level, with coefficients $c^u_{ij}/\Lambda_{\rm UV}$ and so on.

We first approach the matching to the SMEFT by turning off the down-type FN sector of the model and keeping only those FN terms which are responsible for up-type quark masses. In this instance we find all the same SMEFT operator structures as presented in Tables 4–8, but with straightforward "$d\to u$" replacements. For instance, when matching the $d_{\rm FN}=5$ UV Lagrangian to the SMEFT at tree-level up to dimension-6, the Higgs-enhanced Yukawa term changes as

$$-\frac{\lambda_{11}\lambda\, c^d_{ij}}{m_\theta^2}\big(H^\dagger H\big)\big(\overline{Q}_i H d_j + \text{h.c.}\big) \longrightarrow -\frac{\lambda_{11}\lambda\, c^u_{ij}}{m_\theta^2}\big(H^\dagger H\big)\Big(\overline{Q}_i \tilde{H} u_j + \text{h.c.}\Big) \tag{63}$$

as intuitively expected. At one-loop, the story remains largely the same. In Table 4, the Wilson coefficients of $\overline{Q}_i H d_j$ and $\overline{Q}\tilde{H}u_j$ are exchanged together with the swap $y^d_{ij}\leftrightarrow y^u_{ij}$ and replacement $c^d_{ij}\to c^u_{ij}$. The same applies to the Higgs-enhanced Yukawa terms of Table 1, whereas the purely bosonic operators from Tables 5 (H^6) and 6 (X^2H^2) remain unchanged. When it comes to the Higgs-kinetic terms in Table 7, we can once again make the obvious substitutions and exchanged as explained above. The coefficients of $(H^\dagger i\overset{\leftrightarrow}{D}_\mu H)(\overline{d}_i\gamma^\mu d_j)$ and $(H^\dagger i\overset{\leftrightarrow}{D}_\mu H)(\overline{u}_i\gamma^\mu u_j)$ are traded, and those of the operators $(H^\dagger i\overset{\leftrightarrow}{D}{}^I_\mu H)(\overline{Q}_i\tau^I\gamma^\mu Q_j)$ and $(H^\dagger i\overset{\leftrightarrow}{D}_\mu H)(\overline{Q}_i\gamma^\mu Q_j)$ receive a "$d\to u$" relabelling. In the same vein, the four-fermion operators of Table 8 change in the obvious way.

One-loop matching at $d_{\rm FN}=5$ — H^6 and H^4D^2

Operator	Wilson coefficient
$\big(H^\dagger H\big)^3$	$\frac{\hbar\lambda_{11}^2}{9m_\theta^8}\Big[6\lambda_{11}m_\theta^6-6\lambda_{11}^4v_\theta^6+18\lambda_{11}^2v_\theta^4m_\theta^2(\lambda_{11}-6\lambda'_{20})$
	$+v_\theta^2m_\theta^4\Big(-18\lambda_{11}^2+108\lambda'_{20}\lambda_{11}+\lambda'_{20}\big(12\lambda'_{20}(41+30\mathbb{L})-g^2(31+30\mathbb{L})\big)\Big)\Big]$
$\big(H^\dagger H\big)\Box\big(H^\dagger H\big)$	$-\frac{\lambda_{11}^2v_\theta^2}{2m_\theta^4}+\frac{\hbar\lambda_{11}^2}{12m_\theta^2}\Big[16-18\mathbb{L}-\frac{2\lambda_{11}v_\theta^2}{m_\theta^2}(7-12\mathbb{L})+\frac{\lambda_{11}^2v_\theta^4}{m_\theta^4}(19+12\mathbb{L})$
	$-\frac{g_Y^2v_\theta^2}{6m_\theta^2}(31+30\mathbb{L})-\frac{g^2v_\theta^2}{2m_\theta^2}(31+30\mathbb{L})\Big]$
$(H^\dagger D^\mu H)^\dagger(H^\dagger D_\mu H)$	$-\frac{\hbar g_Y^2\lambda_{11}^2v_\theta^2}{18m_\theta^4}[31+30\mathbb{L}]$

One-loop matching at $d_{\rm FN}=5$ — ψ^2H^3

Operator	Wilson coefficient
$(H^\dagger H)\overline{L}_iHe_j$	$\frac{\hbar\lambda_{11}^2v_\theta^2}{36m_\theta^4}\Big[9(y^ey^{e\dagger}y^e)_{ij}(5+6\mathbb{L})+6y^e_{ij}\lambda'_{20}(29+6\mathbb{L})-y^e_{ij}g^2(31+30\mathbb{L})\Big]$
$(H^\dagger H)\overline{Q}_iHd_j$	$\frac{-\lambda_{11}\lambda c^d_{ij}}{m_\theta^2}+\frac{\hbar}{144v_\theta^2m_\theta^6}\Big\{18v_\theta^2m_\theta^4\Big[\frac{m_\theta^2}{v_\theta^2}\lambda^2(y^dc^{d\dagger}c^d+c^dc^{d\dagger}y^d)_{ij}(1+2\mathbb{L})$
	$+4\lambda_{11}\lambda(y^dc^{d\dagger}y^d)_{ij}\Big]+36v_\theta^2m_\theta^2\Big[\Big(\frac{4\lambda^2m_\theta^4}{v_\theta^2}(c^dy^{d\dagger}c^d)_{ij}$
	$-6\lambda_{11}\lambda m_\theta^2(c^dy^{d\dagger}y^d)_{ij}\Big)(1+\mathbb{L})+\lambda_{11}\Big(\lambda_{11}v_\theta^2(y^dy^{d\dagger}y^d)_{ij}(5+6\mathbb{L})$
	$-6\lambda m_\theta^2(y^dy^{d\dagger}c^d)_{ij}(1+\mathbb{L})\Big)\Big]-2\lambda_{11}\Big[2\lambda_{11}m_\theta^2v_\theta^4y^d_{ij}\Big(-6\lambda'_{20}(29+6\mathbb{L})$
	$+g^2(31+30\mathbb{L})\Big)+18\lambda m_\theta^2c^d_{ij}\Big(6m_\theta^4(-1+\mathbb{L})-\lambda_{11}^2v_\theta^4(15+4\mathbb{L})$
	$+2v_\theta^2m_\theta^2\big(\lambda_{11}(4-6\mathbb{L})+2\lambda'_{20}(7+6\mathbb{L})\big)\Big)\Big]\Big\}$
$(H^\dagger H)\overline{Q}_i\tilde{H}u_j$	$\frac{\hbar\lambda_{11}^2v_\theta^2}{36m_\theta^4}\big[9(y^uy^{u\dagger}y^u)_{ij}(5+6\mathbb{L})+6y^u_{ij}\lambda'_{20}(29+6\mathbb{L})-y^u_{ij}g^2(31+30\mathbb{L})\big]$

Table 5: One-loop matching contributions from the dimension-5 FN EFT to the Higgs and Higgs-enhanced Yukawa operators in the Warsaw basis. The down-quark charges under $\mathcal{S}_{\rm FN}$ are taken to be arbitrary.

One-loop matching at $d_{\rm FN}=5$ — X^2H^2	
Operator	Wilson coefficient
$(H^\dagger H)B_{\mu\nu}B^{\mu\nu}$	$\dfrac{\hbar g_Y^2\lambda_{11}^2 v_\theta^2}{12m_\theta^4}$
$(H^\dagger\tau^I H)W^I_{\mu\nu}B^{\mu\nu}$	$\dfrac{\hbar g g_Y\lambda_{11}^2 v_\theta^2}{6m_\theta^4}$
$(H^\dagger H)W^I_{\mu\nu}W^{I,\mu\nu}$	$\dfrac{\hbar g^2\lambda_{11}^2 v_\theta^2}{12m_\theta^4}$

Table 6: One-loop matching contributions from the dimension-5 FN EFT to the Higgs-field strength operators in the Warsaw basis. These operators receive contributions only from Higgs-flavon mixing.

We next consider the matching analysis for the combined toy model of (14) and its up quark analogue, with both sectors turned on at once. There is hardly any mixing between the two sectors of the model. For nearly all operators one may collect the Wilson coefficients arising from the down-type FN sector, shown in the tables above, and add to these coefficients the terms which appear for the first time when turning on only the up-type sector in the UV.

The exception to this rule is the contributions to the Higgs kinetic operator $i(\tilde{H}^\dagger D_\mu H)(\overline{u}_i\gamma^\mu d_j)$. When the up-sector is turned on in addition to the down-sector, the Wilson coefficient of this operator contains terms which cross the dimension-5 coefficients c^d_{ij} and c^u_{ij}:

$$\begin{aligned}&\frac{\hbar}{4m_\theta^4}\Big[2\lambda_{11}\lambda m_\theta^2(y^{u\dagger}c^d)_{ij}-\lambda_{11}^2v_\theta^2(y^{u\dagger}y^d)_{ij}(5+2\mathbb{L})\Big]i(\tilde{H}^\dagger D_\mu H)(\overline{u}_i\gamma^\mu d_j)+\text{h.c.}\\ &\longrightarrow\frac{\hbar}{4m_\theta^4}\Big[2\lambda_{11}\lambda m_\theta^2(y^{u\dagger}c^d)_{ij}-\lambda_{11}^2v_\theta^2(y^{u\dagger}y^d)_{ij}(5+2\mathbb{L})+2\lambda_{11}\lambda m_\theta^2(c^{u\dagger}y^d)_{ij}\\ &\qquad\qquad+\lambda^2(c^{u\dagger}c^d)_{ij}\frac{m_\theta^4}{v_\theta^2}(1+2\mathbb{L})\Big]i(\tilde{H}^\dagger D_\mu H)(\overline{u}_i\gamma^\mu d_j)+\text{h.c.}\end{aligned}\tag{64}$$

The first two terms carry over from the down-quark-only scenario and the third term follows from the "$d\to u$" recipe outlined above, but the final term, proportional to $(c^{u\dagger}c^d)_{ij}$, is new.

Extension to Leptons

As with the up quarks, in the toy model analysis presented above we have not specified the $\mathcal{S}_{\rm FN}$ charges of L_i and/or e_i which, assuming a non-trivially charged Higgs, must respect

$$q_{\overline{L}_i}+q_{e_j}\stackrel{!}{=}-(q_H+x_{ij}\,q_\theta)\tag{65}$$

to allow for generically flavoured charged-lepton mass terms a la (1). Given the parity between the charged lepton and quark sectors of the (SM)EFT, we do not expect any substantive

One-loop matching at $d_{\rm FN}=5$ — ψ^2H^2D

Operator	Wilson coefficient
$(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{d}_i\gamma^\mu d_j)$	$\frac{\hbar}{432m_\theta^4}\Big[108\lambda_{11}\lambda m_\theta^2(c^{d\dagger}y^d)_{ij}+\frac{54\lambda^2m_\theta^4}{v_\theta^2}(c^{d\dagger}c^d)_{ij}(1+2\mathbb{L})$ $+2\lambda_{11}\big(\lambda_{11}g_Y^2v_\theta^2\delta_{ij}(17+6\mathbb{L})+54\lambda m_\theta^2(y^{d\dagger}c^d)_{ij}$ $-27\lambda_{11}v_\theta^2(y^{d\dagger}y^d)_{ij}(5+2\mathbb{L})\big)\Big]$
$i(\tilde{H}^\dagger D_\mu H)(\overline{u}_i\gamma^\mu d_j)$	$\frac{\hbar\lambda_{11}}{4m_\theta^4}\Big[2\lambda m_\theta^2(y^{u\dagger}c^d)_{ij}-\lambda_{11}v_\theta^2(y^{u\dagger}y^d)_{ij}(5+2\mathbb{L})\Big]$
$(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{e}_i\gamma^\mu e_j)$	$\frac{-\hbar\lambda_{11}^2v_\theta^2}{72m_\theta^4}\big[9(y^{e\dagger}y^e)_{ij}(5+2\mathbb{L})-g_Y^2\delta_{ij}(17+6\mathbb{L})\big]$
$(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{L}_i\gamma^\mu L_j)$	$\frac{\hbar\lambda_{11}^2v_\theta^2}{144m_\theta^4}\big[9(y^ey^{e\dagger})_{ij}(5+2\mathbb{L})+g_Y^2\delta_{ij}(17+6\mathbb{L})\big]$
$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\overline{L}_i\tau^I\gamma^\mu L_j)$	$\frac{\hbar\lambda_{11}^2v_\theta^2}{144m_\theta^4}\big[9(y^ey^{e\dagger})_{ij}(5+2\mathbb{L})-g^2\delta_{ij}(17+6\mathbb{L})\big]$
$(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{Q}_i\gamma^\mu Q_j)$	$\frac{-\hbar}{288m_\theta^4}\Big[36\lambda_{11}\lambda m_\theta^2(y^dc^{d\dagger})_{ij}+\frac{18m_\theta^4\lambda^2}{v_\theta^2}(c^dc^{d\dagger})_{ij}(1+2\mathbb{L})$ $+2\lambda_{11}\Big(18\lambda m_\theta^2(c^dy^{d\dagger})_{ij}-9\lambda_{11}v_\theta^2(y^dy^{d\dagger}-y^uy^{u\dagger})_{ij}(5+2\mathbb{L})$ $+\frac{\lambda_{11}g_Y^2v_\theta^2}{3}\delta_{ij}(17+6\mathbb{L})\Big)\Big]$
$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\overline{Q}_i\tau^I\gamma^\mu Q_j)$	$\frac{-\hbar}{288m_\theta^4}\Big[36\lambda_{11}\lambda m_\theta^2(y^dc^{d\dagger})_{ij}+\frac{18\lambda^2m_\theta^4}{v_\theta^2}(c^dc^{d\dagger})_{ij}(1+2\mathbb{L})$ $+2\lambda_{11}\Big(18\lambda m_\theta^2(c^dy^{d\dagger})_{ij}-9\lambda_{11}v_\theta^2(y^dy^{d\dagger}+y^uy^{u\dagger})_{ij}(5+2\mathbb{L})$ $+\lambda_{11}g^2v_\theta^2\delta_{ij}(17+6\mathbb{L})\Big)\Big]$
$(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{u}_i\gamma^\mu u_j)$	$\frac{\hbar\lambda_{11}^2v_\theta^2}{216m_\theta^4}\big[27(y^{u\dagger}y^u)_{ij}(5+2\mathbb{L})-2g_Y^2\delta_{ij}(17+6\mathbb{L})\big]$

Table 7: One-loop matching contributions from the dimension-5 FN EFT to the Higgs kinetic operators in the Warsaw basis. The down-quark charges charges under $\mathcal{S}_{\rm FN}$ are taken to be arbitrary.

One-loop matching at $d_{\rm FN}=5$ — $\psi^4:\ (\overline{L}R)(\overline{R}L)$ and $(\overline{L}R)(\overline{L}R)$

Operator	Wilson coefficient
$(\overline{L}_i^\alpha e_j)(\overline{d}_k Q_l^\alpha)$	$\dfrac{\hbar\lambda_{11}y^e_{ij}}{12m_\theta^4}\left[2\lambda_{11}(y^{d\dagger})_{kl}v_\theta^2-6\lambda(c^{d\dagger})_{kl}m_\theta^2\right]$
$(\overline{L}_i^\alpha e_j)\epsilon_{\alpha\beta}(\overline{Q}_k^\beta u_l)$	$\dfrac{\hbar\lambda_{11}^2v_\theta^2y^e_{ij}y^u_{kl}}{6m_\theta^4}$
$(\overline{Q}_i^\alpha u_j)\epsilon_{\alpha\beta}(\overline{Q}_k^\beta d_l)$	$\dfrac{\hbar\lambda_{11}y^u_{ij}}{12m_\theta^4}\left[2\lambda_{11}v_\theta^2y^d_{kl}-6c^d_{kl}\lambda m_\theta^2\right]$

One-loop matching at $d_{\rm FN}=5$ — $\psi^4:\ (\overline{L}L)(\overline{R}R)$

Operator	Wilson coefficient
$(\overline{Q}_i\gamma_\mu Q_j)(\overline{d}_k\gamma^\mu d_l)$	$-\dfrac{\hbar}{12}\left[\dfrac{\lambda_{11}^2v_\theta^2y^d_{il}(y^{d\dagger})_{kj}}{m_\theta^4}+\dfrac{6\lambda^2c^d_{il}(c^{d\dagger})_{kj}}{v_\theta^2}(1+\mathbb{L})\right]$
$(\overline{L}_i\gamma_\mu L_j)(\overline{e}_k\gamma^\mu e_l)$	$-\dfrac{\hbar\lambda_{11}^2v_\theta^2y^e_{il}(y^{e\dagger})_{kj}}{12m_\theta^4}$
$(\overline{Q}_i\gamma_\mu Q_j)(\overline{u}_k\gamma^\mu u_l)$	$-\dfrac{\hbar\lambda_{11}^2v_\theta^2y^u_{il}(y^{u\dagger})_{kj}}{12m_\theta^4}$

Table 8: One-loop matching contributions from the dimension-5 FN EFT to the four-fermion operators in the Warsaw basis. The down-quark charges charges under $\mathcal{S}_{\rm FN}$ are taken to be arbitrary.

deviations in the matching of FN models where (65) is respected and those already presented in Sections 2 and generalized in 3.1. In other words, if/when FN dynamics controls y^e_{ij} along the lines of (14), we expect the flavour structure of all derived Wilson coefficients dependent on L_i, e_i to be analogous to that found for the down quarks in Section 2. As a final note, we again caution that, when constructing gauged FN models with the aim of explaining all fermion masses and mixings, gauge anomalies provide an additional constraint on the charge assignment, and this is especially true when incorporating $\mathcal{S}_{\rm FN}$-charged leptons into the spectrum.

On the other hand, non-zero neutrino masses require a generation mechanism that is fundamentally BSM. One approach is to introduce SM-gauge-singlet, right-handed neutrinos analogous to $\{u, d, e\}$ into the dynamical particle spectrum, which immediately leads to renormalizable Dirac neutrino masses in analogy to the charged leptons and quarks (not to mention a lepton-number-violating Majorana mass for the gauge-singlets). Doing so is beyond our scope in this paper and, in any event, if said gauge singlets are sufficiently heavy, integrating them out of the Lagrangian leads to Weinberg operators [12] via the see-saw mechanism [90], and these already generically appear in our construction (cf. (32)). Indeed, it is well known that a Weinberg operator generating non-zero Majorana neutrino masses appears at $d_{\rm SM} = 5$, and this is also possible in a FN construction which generalizes the Weinberg operator class and its charge constraints as follows:

$$\mathcal{L}^{5+x_{ij}+y_{ij}}_{\rm FN} \supset \frac{c^{W_{xy}}_{ij}}{\Lambda^{1+x_{ij}+y_{ij}}_{\rm UV}} \big(H\overline{L}^c_i\big)(L_j H)\theta^{x_{ij}}\theta^{\star\, y_{ij}} + \text{h.c.} \quad \text{iff} \quad (x_{ij} - y_{ij})\, q_\theta = -(2q_H + q_{L_i} + q_{L_j})\,. \tag{66}$$

In this scenario FN dynamics also control hierarchies of neutrino masses, which is generally appealing for a complete flavour model.

However, implementing a flavour-generic, dimension-6 version of (66) into `Matchete`, and subsequently asking the program to match to the $d_{\rm SM} = 6$ SMEFT, only generates a Weinberg operator with the characteristic Wilson coefficient:

$$\mathcal{L}_{\rm SMEFT} \supset c^{W_{10}}_{ij} \frac{\overline{\lambda}}{\Lambda_{\rm UV}} \big(H\overline{L}^c_i\big)(L_j H) + \text{h.c.} \tag{67}$$

It is expected that, if matching to higher-orders in the SMEFT, more elaborate results will appear. Regardless, it is fun to observe that a Majorana neutrino mass term suppressed by $\overline{\lambda}/\Lambda_{\rm UV}$ is naturally generated in this naive FN setup. The suppression of the Weinberg operator by some power of $\overline{\lambda}$ will modify the typical lower bounds on the see-saw scale derived from neutrinoless double-beta decay searches.

A final comment regards the overlap of neutrino and charged lepton masses, whose diagonalization to their respective mass-eigenstate bases results in the (unitary contribution to the) Pontecorvo-Maki-Nakagawa-Sakata (PMNS) mixing matrix $U^{\rm PMNS}$ in the SMEFT. PMNS mixings are now quite well-constrained experimentally (cf. [3]) and, unlike the CKM quark mixing matrix, are generically large and non-hierarchical ($|U^{\rm PMNS}_{ij}| \sim \mathcal{O}(10^{-1}) - \mathcal{O}(1)$). Hence a successful FN model of lepton flavour must simultaneously explain strong charged-lepton mass hierarchies alongside of weak PMNS hierarchies, both as a function of λ. While not impossible, this challenge (along with the desire to limit the number of free UV model

parameters) has often led model-builders to consider non-Abelian theories of flavour — cf. Section 3.2 below.

3.2 Non-Abelian and/or Continuous Flavour Models

We now consider how our matching analysis might change in flavour models that are based on broken non-Abelian symmetries, and/or models based on continuous flavour symmetries, as opposed to the global, discrete, Abelian considerations of Section 2.

Non-Abelian Flavour Models

Non-Abelian flavour models typically assign multiple generations of a SM fermion family to a single BSM flavour-symmetry multiplet, which is not possible in Abelian frameworks. These symmetries can be continuous (e.g. U(2), SO(3) or SU(3)) or discrete (e.g. A4, S4, or $\Delta(27)$), and the literature concerning both is vast. A common feature of these constructions is the reduction of UV parameters compared to Abelian models, which can be understood thanks to their multi-dimensional irreducible representations — contracting N-plets of a given flavour symmetry populates *multiple* effective Yukawa couplings with a single free parameter (Wilson coefficient), and hence typically leads to increased predictivity. A trade-off is that multiple new flavon fields are often introduced to allow for distinct contractions, which typically have vacua that are especially aligned in flavour space to realize characteristic Yukawa textures.

As an example of a relatively recent and successful discrete model, we consider the Universal Texture Zero (UTZ) [91, 92] introduced by one of the authors and collaborators, based on a $\Delta(27) \times \mathbb{Z}_N$ flavour symmetry and consistent with an underlying $\mathrm{SO}(10) \to \mathrm{SU}(4) \times \mathrm{SU}(2)_\mathrm{R} \times \mathrm{SU}(2)_\mathrm{L}$ Grand Unified Theory (GUT). The UTZ is capable of controlling the Dirac masses[16] of both the quark and lepton sectors in terms of three flavons $\{\theta_{3,23,123}\}$ and two additional scalars $\{S, \Sigma\}$, with Σ related to the GUT symmetry breaking and S designed, along with the additional $\mathbb{Z}_N$ symmetry, to 'shape' the allowed interactions between θ_i:

$$\mathcal{L}_\mathrm{UTZ} \supset \psi_i \left[\frac{c_3^{(6)}}{\Lambda_{3,f}^2} \theta_3^i \theta_3^j + \frac{c_{23}^{(7)}}{\Lambda_{23,f}^3} \theta_{23}^i \theta_{23}^j \Sigma + \frac{c_{123}^{(7)}}{\Lambda_{123,f}^3} (\theta_{123}^i \theta_{23}^j + \theta_{23}^i \theta_{123}^j) S \right] \psi_j^c H + \mathcal{O}(1/\Lambda_i^4) + ... \quad (68)$$

Here ψ is a generic SM fermion with flavour $\{i, j\}$ of family f, $f \in \{u, d, e, \nu\}$, assigned to a $\Delta(27)$ triplet. As indicated, these flavour multiplets contract with the (anti-triplet) flavon fields, while $\{H, S, \Sigma\}$ are all trivial flavour singlets. Thanks to its elaborate $\Delta(27)$-invariant scalar potential — cf. [91] for details — the UTZ flavons develop the following vevs in flavour space (α and β are free complex phases):

$$\langle \theta_3 \rangle = \mathrm{v}_{\theta 3} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \langle \theta_{123} \rangle = \frac{\mathrm{v}_{123}}{\sqrt{3}} \begin{pmatrix} e^{i\beta} \\ e^{i\alpha} \\ -1 \end{pmatrix}, \quad \langle \theta_{23} \rangle = \frac{\mathrm{v}_{23}}{\sqrt{2}} \begin{pmatrix} 0 \\ e^{i\alpha} \\ 1 \end{pmatrix}, \quad (69)$$

[16] Majorana neutrino masses (and associated PMNS mixings) are also controlled with a similar equation. Note that the UTZ also assumes supersymmetry breaking, whose complications we will ignore in this discussion.

from which it is easy to intuit from (68) that the UTZ predicts very specific Yukawa/mass matrix textures:

$$M^f_{\text{Dirac}} \propto \begin{pmatrix} 0 & a\,e^{i(\alpha+\beta+\gamma)} & a\,e^{i(\beta+\gamma)} \\ a\,e^{i(\alpha+\beta+\gamma)} & (b\,e^{-i\gamma}+2a\,e^{-i\delta})\,e^{i(2\alpha+\gamma+\delta)} & b\,e^{i(\alpha+\delta)} \\ a\,e^{i(\beta+\gamma)} & b\,e^{i(\alpha+\delta)} & 1-2a\,e^{i\gamma}+b\,e^{i\delta} \end{pmatrix}, \tag{70}$$

which we have written after EW symmetry breaking, and where $\{a, b, \gamma, \delta\}$ are all rescaled functions of the (complex) coefficients and parameters given in (68)–(69). One immediately sees in (70) a concrete manifestation of the non-Abelian mechanism mentioned above in that, for a given family sector, only six *real* parameters structure Dirac mass and mixing up to EFT order implied in (68). This is to be compared to a *complex* parameter for each matrix element in the simple $\mathbb{Z}_N$ construction of Section 2.

From (68)–(70) we can already anticipate a number of complications, but also results, when matching the UTZ to the SMEFT, in comparison to the FN setup of Section 2. First and foremost, in *each* family sector f there are potentially three independent UV scales $\Lambda_{3,23,123}$ identified in (68), associated with the first appearance of each of the three flavons $\theta_{3,23,123}$. This allows for hierarchies amongst the flavon and/or UV mediator masses, which further implies a potentially complex sequential matching to intermediate EFTs. This would generate multiple expansion parameters that ultimately filter down to characterize the SMEFT Wilson coefficients (analogous to λ above). A naive option is to assume no hierarchy amongst the UV scales, $\Lambda_3 \sim \Lambda_{23} \sim \Lambda_{123}$, although as discussed in [91, 92] phenomenological considerations and the desire to avoid fine-tuned Wilson coefficients quickly demands $\Lambda_{3,u}/\Lambda_{23,u} < \Lambda_{3,d}/\Lambda_{23,d}$, while GUT considerations lead to an implied equivalence in the Yukawa expansion parameters between down quark and charged lepton sectors. Finally, it is straightforward to deduce from (68) that at least $\langle\theta_3\rangle/\Lambda_{3,u}$ should be relatively large to account for the top mass (cf. [93]). The takeaway is that the presence of multiple UV scales, associated to multiple flavons, might rapidly complicate the matching we have performed in Section 2 and that, in the absence of a fully generic / sequential matching to intermediate EFTs, one risks a heavily model-dependent analysis.

At the same time, (70) reveals that $M^d_{12,13,23} = M^d_{21,31,32}$, since these matrix elements are sourced by the same UTZ operators. This indicates that, upon integrating out the relevant flavon(s), each term in (68) will induce multiple SMEFT operators (corresponding to distinct IR flavour structures), each with the same Wilson coefficient. As an obvious example, we would expect an equivalent contribution to Higgs-enhanced Yukawa operators akin to (47) to be induced at the first tree-level matching order relevant to the UTZ, for each matrix element associated to the UTZ operator. This represents a distinct signature with respect to the Abelian FN construction, and it also holds up to the most important next-to-leading-order EFT operators in the UTZ (see [91, 92] for higher-order contributions to (70)). Additionally, we see that there are no renormalizable Yukawa couplings present in the UTZ, and so the analogous matching to Section 2.1.1 will be sourced fully from the scalar potential. Hence we would naively expect all one-loop matching terms proportional to $y^d_{32,33}$ to go to zero. It is also worth noting that while fewer terms in (68) populate M^d than in (14), and that the maximum mass dimension of the contributing operators is also less than in (14), the first order at which interesting Yukawa structure appears is *higher* than in (14) (mass-dimension 6, not 4).

As a final preliminary consideration, we recall that a rather particular extended scalar potential is required to dynamically drive the minimization of the vevs in (69). This is true already at the renormalizable level, while the model-building impacts of effective scalar potential operators akin to (45) is to our knowledge somewhat lacking in the literature (see [94] for a related counterexample, albeit in a different class of model), not to mention the role of effective operators in the UTZ theory outside of the scalar or Yukawa sectors (akin to the $d_{\rm FN} = 6$ basis we built in Section 2.1.3). As seen at virtually all stages of the matching in Section 2, the impact of Higgs-flavon mixing is extremely important in understanding the structure of the IR SMEFT Wilson coefficients. This will become even more intricate in non-Abelian, UTZ-type models. Besides the presence of more scalar operators, the special vev alignment in (69) requires hierarchies amongst their coefficients, which will have immediate implications for the SMEFT Wilson coefficients they populate as well. These must be understood in order to fully appreciate the IR imprint of the UTZ.

While the UTZ of (68) represents but one of many non-Abelian flavour models on the market, it shares a number of salient features with other constructions. For example, the renowned A4 models of Altarelli-Feruglio [95, 96] or Babu-Ma-Valle [97], the SU(3) models of Ross-King [98, 99], or even the more recent gauged SO(3) model of Reig-Valle-Wilczek [100] all depend on flavon(s) with particular vev alignments in flavour space, which, as mentioned above, will likely generate the dominant source of complexity when matching non-Abelian flavour models to the SMEFT. Distilling their most generic features beyond Points (1)–(3) above represents future work.

Goldstone Modes and Gauged Flavour Models

As mentioned below (10), the toy FN model we consider in Section 2 is effectively controlled by a discrete, Abelian FN symmetry ($\mathbb{Z}_N$), since we do not consider the implications of the (pseudo-)Goldstone boson associated to the breaking of a continuous $\mathrm{U}(1)_{\rm FN}$, regardless of whether it is global or gauged.

In the instance where we generalize to a *global* $\mathrm{U}(1)_{\rm FN}$, the π_θ in (10) are massless Goldstone particles in the strict symmetry limit, or light (with respect to the flavon ϑ) pseudo-Goldstone particles when $\mathrm{U}(1)_{\rm FN}$ is only approximate. In the absence of hypothetical model building that forces π_θ to be very heavy, we are left with an additional, dynamical BSM degree of freedom in either case, and so formally one should not match directly to the SMEFT as we have above, but instead to the IR EFT that also incorporates π_θ (which we will refer to as the πSMEFT). Furthermore, even if π_θ is heavy with respect to the SM, one would have to purposely integrate it out alongside of or sequential to ϑ, which would ultimately amount to matching a (flavour-specific) UV model with two real SM gauge singlets, since we work in the broken $\mathrm{U}(1)_{\rm FN}$ phase.

While beyond our scope to explore these situations in detail, we do note that, up to $d \leq 7$, a non-redundant basis for the proposed πSMEFT is effectively given in [87] up to flavour-symmetry assumptions. One can anticipate that, since the πSMEFT $\supset$ SMEFT, the contributions outlined in Section 2 will still hold. Also, up to $d_{\rm FN} = 6$, the only Lagrangian interactions allowed between ϑ and π_θ in the UV $\mathrm{U}(1)_{\rm FN}$ theory are sourced by the θ-current-fermion-current interaction in (50) and $\mathcal{O}_{\overline{Q}_2 d_i \theta}$ in (47) (cross terms proportional to $\pi_\theta\,\vartheta$ cancel in $\theta^\star\theta$ operator pairs). Hence some contributions to matched IR Wilson coefficients in the

πSMEFT would simply follow trivially from ϑ-independent UV interactions (e.g. the $d_{\rm FN} = 5$ operator $\propto \bar{\psi}_L H \psi_R \pi_\theta$). On the other hand, flavon-Goldstone mixing will complicate the results when matching up to higher operator dimensions in πSMEFT and especially in a loop-level analysis. Furthermore, even upon successfully matching to πSMEFT, utilizing the hypothetical πSMEFT-matched FN theory for precision phenomenology would require knowledge of the complete set of RGE equations for generic $d = 6$ πSMEFT Wilson coefficients, analogous to [75, 80, 81] for the SMEFT, to run down to scales relevant for experimental comparison or subsequent EFT matching(s). We note that some relatively recent work has been done on SMEFT interference effects when an additional, light, axion-like particle (akin to π_θ) is present [101].

If instead the $\mathrm{U}(1)_{\rm FN}$ is *gauged*, we must incorporate an additional Z' vector in the UV spectrum. In the unitary gauge, it will eat the Goldstone to acquire a mass proportional to the flavon vev and will therefore be integrated out of the spectrum alongside the scalar flavon as above. There are a host of papers discussing the matching of a Z' model to the SMEFT, but generic tree-level results are available from (e.g.) [102]. There it is clear that a number of dimension-6 SMEFT operators are turned on as a result of integrating out the Z'. Owing to its nature as a gauge boson, the Z' may only couple to the SM fermions via the covariant derivative $\overline{\psi}_i i \not{D} \psi_i$ at the renormalizable level. If $g_{\rm F}$ denotes the gauge coupling of $\mathcal{S}_{\rm FN}$ and q_{ψ_i} the charge of the fermion under the Abelian symmetry, we arrive at the interaction vertex $-g_{\rm F}\, q_{\psi_i} Z'_\mu \overline{\psi}_i \gamma^\mu \psi_i$ where one may assume without loss of generality that the interaction is diagonal in flavour space. For the down quark toy model of Table 1, the interaction may be expressed as

$$\mathcal{L}_{\rm FN} \supset -g_{\rm F} Z'_\mu \big(\Omega_{ij} \overline{d}_i \gamma^\mu d_j + \xi_{ij} \overline{Q}_i \gamma^\mu Q_j\big)\,, \tag{71}$$

where the two flavour matrices are

$$\Omega = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad \xi = -\begin{pmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{72}$$

When integrating out the Z' diagrammatically, (71) leads to four-fermion interactions via a Z' exchange:

$$\begin{aligned} \mathcal{L}^{d=6}_{\rm SMEFT} \supset &-\frac{g_{\rm F}^2\, \Omega_{ij}\Omega_{kl}}{2M_{Z'}^2} \big(\overline{d}_i \gamma^\mu d_j\big)\big(\overline{d}_k \gamma_\mu d_l\big) - \frac{g_{\rm F}^2\, \xi_{ij}\xi_{kl}}{2M_{Z'}^2} \big(\overline{Q}_i \gamma^\mu Q_j\big)\big(\overline{Q}_k \gamma_\mu Q_l\big) \\ &- \frac{g_{\rm F}^2\, \Omega_{ij}\xi_{kl}}{M_{Z'}^2} \big(\overline{d}_i \gamma^\mu d_j\big)\big(\overline{Q}_k \gamma_\mu Q_l\big). \end{aligned} \tag{73}$$

This of course generalizes to the remaining fermion species which would generate similar four-fermion terms $\mathcal{O}_{\psi\psi}$ for $(\psi\psi) \in \{LL, LQ, ee, uu, ed, eu, ud, Le, Ld, Lu, Qe, Qu\}$. Here, too, the results are presented in the interaction eigenbasis and a rotation to the mass basis is required in order to get to serious phenomenological predictions.

Provided that the Higgs is charged under $\mathcal{S}_{\rm FN}$, the Higgs kinetic term $D_\mu H^\dagger D^\mu H$ gives more renormalizable interactions between the Z' and the SM fields, one linear and one quadratic in the Z'. When it comes to dimension-6 results from tree-level matching, it is only the term

linear in the novel gauge boson, $\mathcal{L}_{\rm FN} \supset -ig_{\rm F}q_H Z'_\mu H^\dagger D^\mu H + \text{h.c.}$, that contributes to the Wilson coefficients. In our toy model, where $q_H = -3$, connecting two such vertices together via a Z' propagator gives rise to

$$\mathcal{L}^{d=6}_{\rm SMEFT} \supset -\frac{9g_{\rm F}^2}{2M_{Z'}^2}\big(H^\dagger H\big)\Box\big(H^\dagger H\big) - \frac{18g_{\rm F}^2}{M_{Z'}^2}\big(H^\dagger D^\mu H\big)^*\big(H^\dagger D_\mu H\big)\,, \tag{74}$$

whereas connecting a fermion current from (71) with a $Z'_\mu H^\dagger D^\mu H$ vertex contributes to Higgs kinetic operators

$$\mathcal{L}^{d=6}_{\rm SMEFT} \supset -\frac{3g_{\rm F}^2\Omega_{ij}}{M_{Z'}^2}(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{d}_i\gamma^\mu d_j) - \frac{3g_{\rm F}^2\xi_{ij}}{M_{Z'}^2}(H^\dagger i\overleftrightarrow{D}_\mu H)(\overline{Q}_i\gamma^\mu Q_j). \tag{75}$$

Again, a similar set of terms is obtained for the other fermion species, too.

One may also expect matching contributions that hinge on vertices that depend on both the flavon ϑ and the Z', originating from the flavon kinetic term $D_\mu\theta^* D^\mu\theta$. These vertices, however, will only yield SMEFT operators with $d_{\rm SM} > 6$ when matching at tree-level, and hence will not be addressed here. Similarly, when going beyond the renormalizable level in the FN EFT, it is possible to write down more complex interactions between the Z', the flavon, and the SM fields, but these vertices will not play a role in generating dimension-6 SMEFT operators at tree-level, and can introduce further complications regarding flavour-violating basis changes as well.

From the above analysis we conclude that, at tree-level, integrating out a Z' does not induce $d_{\rm SM} = 6$ dipole-type ($\psi^2 XH$) or triple-field-strength (X^3) operators in the SMEFT. As mentioned above, the X^3 operators will still be allowed in the IR with arbitrary Wilson coefficients due to their presence in $\mathcal{L}_{\rm SMEFT'}$, but IR SMEFT dipole operators will only appear with limited flavour combinations from $\mathcal{L}_{\rm SMEFT'}$. Hence the presence of other dipole signals (via sizable non-zero Wilson coefficients) in global SMEFT fits might serve as a powerful falsification tool for FN-type models, regardless of whether or not they are gauged or global. Simultaneously, signals for sizable four-fermion or Higgs-kinetic ($H^2 D\psi^2$) Wilson coefficient(s) could, in principle, be used to distinguish the gauged $\mathrm{U}(1)_{\rm FN}$ vs. global scenario, given that we found these operators only at loop-level in Section 2. What is more, the hierarchies of such operators in flavour space are different for the contributions driven by the flavon and controlled by powers of λ, versus those coming from the Z' and controlled by the fermion charges under $\mathcal{S}_{\rm FN}$. For the toy model we find, for instance, that the Wilson coefficients of the four-fermion operators grow as we move to lower generations, a feature that is capable of producing stringent lower bounds on the $M_{Z'}$ mass and thus the $\mathcal{S}_{\rm FN}$ breaking scale in the gauged scenario.

Finally, all of the above discussion can be further generalized to the case of continuous, *non-Abelian* flavour symmetries, as have been explored in the models of (e.g.) [98, 99] and, more recently, [100]. Here the presence of multiple (pseudo-)Goldstone bosons will source the obvious complication(s) beyond those already mentioned in the Subsection above. As with a $\mathrm{U}(1)_{\rm FN}$, breaking a global non-Abelian flavour symmetry will necessitate matching to a larger EFT than the SMEFT or even πSMEFT (incorporating *all* of the light Goldstones), while breaking a gauged non-Abelian flavour symmetry will require understanding the IR impacts coming from subsequently (or simultaneously) integrating out multiple heavy BSM vectors.

4 Summary and Outlook

We have systematically matched the Froggatt-Nielsen (FN) [1] flavour generation mechanism to the mass-dimension 6 ($d_{\rm SM} = 6$) SMEFT, considering tree- and one-loop level contributions sourced from UV operators in a FN effective theory up to mass-dimension 6 and 5 ($d_{\rm FN} = 6, 5$), respectively. Beginning from a simplified $\mathbb{Z}_{\rm FN}$ -symmetric construction, in Section 2 we studied an Abelian toy model where FN dynamics resolve fermionic mass and mixing relations derived from the down-quark Yukawa matrix in (14). This included enumerating a minimal, complete basis of EFT operators at $d_{\rm FN} = 6$ in the FN theory, to be fully generic in our analysis. Upon executing (and checking!) the functional matching procedure embedded in `Matchete` [30] we saw that, besides renormalizations of SM terms, a host of SMEFT operators also appear with characteristic Wilson coefficients. These include Higgs-scalar (H^6) and -derivative (H^4D^2), Higgs-gauge (H^2X^2), Higgs-kinetic ($H^2D\psi^2$), Higgs-enhanced Yukawa ($H^3\psi^2$), and four-fermion (ψ^4) operators. Many of these IR signatures are effectively flavour independent, appearing simply as a result of integrating out a heavy BSM scalar (the FN flavon, in this instance), while others are flavour-specific, reflecting the particular FN charge assignment (and therefore patterns of flavour suppression) of our UV toy model. Our final results are collected in Tables 2–8.

Our theoretical analysis can be further improved in a number of ways. In Section 3 we already discussed scenarios that obviously generalize our analysis, including the straightforward extension(s) of the FN dynamics we studied in the down quark sector to different fermion families. We also anticipated the implications of considering continuous and/or non-Abelian flavour models, either gauged or global, where we saw that a number of complications can arise due to the presence of additional IR (e.g. Goldstone boson) or UV (e.g. flavon and/or vector gauge boson) physics scales, depending on the particular flavour symmetry realization. We pointed out that, in the presence of additional light degrees of freedom, a better understanding of the matching to an intermediate EFT, with BSM propagating degrees of freedom, was necessary. We also commented on the possibility of distinguishing gauged vs. global UV flavour scenarios via the relative magnitudes of SMEFT Wilson coefficients that appear in each respective scenario. Indeed, all of the discussions in Section 3 stand to be refined with more dedicated theory analyses in the future, perhaps even by matching UV-complete (fully renormalizable) model(s) of FN flavour to the SMEFT, although this would be much less model-independent than the EFT approach taken in this paper.

On the other hand, in Section 3 we did not consider operators with $d > 6$, either in the IR or UV. From the UV perspective, the mass matrix in (14) has matrix elements that first appear at $d_{\rm FN} \geq 7$. While highly suppressed and very much model dependent, understanding such higher-order contributions in the context of the SMEFT may be interesting, if technically challenging. From the IR perspective, matching the SMEFT at $d_{\rm SM} > 6$ would also yield interesting results. Preliminary studies indicate a plethora of new, dimension-8 operators can appear, some of which may be phenomenologically relevant. One may for instance join together two dimension-5 FN vertices of the toy model from (33) in a way illustrated in Fig. 5 to obtain the Higgs-enhanced four-fermion terms

$$\mathcal{L}^{d=8}_{\rm SMEFT} \supset \left\{ \frac{(c^d_{31})^2 \tilde{\lambda}^2}{2\tilde{m}^2_\theta \tilde{v}^2_\theta} (\overline{Q}_3 H d_1)(\overline{Q}_3 H d_1) + \text{h.c.} \right\} + \frac{|c^d_{31}|^2 \tilde{\lambda}^2}{4\tilde{m}^2_\theta \tilde{v}^2_\theta} \left(H^\dagger H\right) \left(\overline{Q}_3 \gamma^\mu Q_3\right) \left(\overline{d}_1 \gamma_\mu d_1\right)$$

$$+\frac{|c_{31}^d|^2\tilde{\lambda}^2}{4\tilde{m}_\theta^2\tilde{v}_\theta^2}\big(H^\dagger\tau^I H\big)\big(\overline{Q}_3\gamma^\mu\tau^I Q_3\big)\big(\overline{d}_1\gamma_\mu d_1\big). \tag{76}$$

The tildes above the parameters remind us that they will be modified by contributions coming from the dimension-8 scalar potential in the broken $\mathcal{S}_{\rm FN}$ phase. To obtain a full set of $d_{\rm SM}=8$ results, it would of course be important to adopt a non-redundant, complete operator basis, for instance that of [103] or [104].

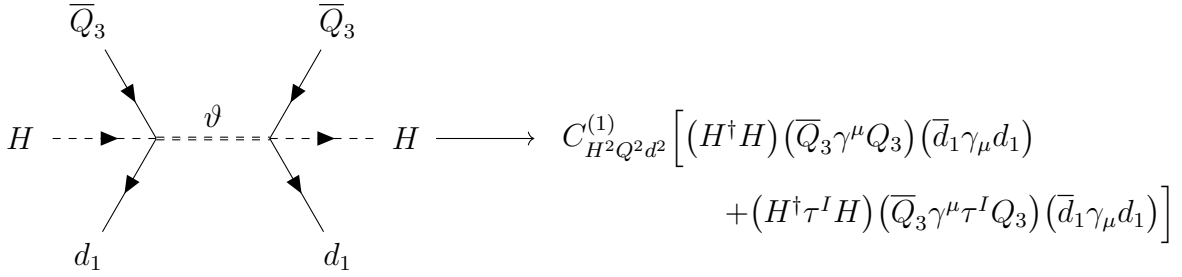


Figure 5: A diagrammatic illustration of how tree-level matching of the toy FN model yields the dimension-8 SMEFT operators $\big(H^\dagger H\big)\big(\overline{Q}_3\gamma^\mu Q_3\big)\big(\overline{d}_1\gamma_\mu d_1\big)$ and $\big(H^\dagger\tau^I H\big)\big(\overline{Q}_3\gamma^\mu\tau^I Q_3\big)\big(\overline{d}_1\gamma_\mu d_1\big)$. To get to the final operators, we use a Fierz identity and the Pauli matrix completeness relation. The Wilson coefficient follows the notation of [103].

Finally, we have not considered matching FN theories in the *unbroken* flavour-symmetry phase, nor have we considered matching to so-called geometric EFTs, where all-v_H/Λ-orders generalizations of (flavoured) Wilson coefficients, with implied geometries in BSM field space, appear (see e.g. [38, 105–109] for recent work incorporating fermions in these formalisms). All of these theoretical avenues can and should be explored.

Besides theory improvements, a major long-term goal of the SMEFT community is the robust, global fit of $d_{\rm SM}=6$ Warsaw basis operators to experimental data across a range of energy scales, and the subsequent projection of those operator bounds to concrete BSM physics conclusions. As mentioned in the Introduction, a number of studies have already made significant progress on the global fitting front, including in the flavour sector [17–27]. Hence an obvious next step in our study of FN-type effects on the SMEFT is to pursue precision phenomenology given the IR matching signatures we have outlined in Section 2. There we briefly mentioned bounds on flavon mass scales ($m_\theta \gtrsim 1$ TeV) sourced from $\mathcal{Q}_{H\Box}$ in (21), as well as the relative expectation of bounds from four-fermion vs. other operators appearing in Section 2. However, a more dedicated phenomenological study could be approached with the automated matching, running, and optimization tools embedded in (e.g.) `smelli` [89], which matches SMEFT Wilson coefficients to the lower-energy LEFT [78, 79], where many of the most relevant precision flavour constraints derived from (e.g.) meson mixing, charged lepton flavour violation, B decays, etc., can be applied. Indeed, data procured at energy scales spanning many orders of magnitude will likely be required to falsify FN-type models in general, and/or to distinguish one FN-type model from another.

In summary, by matching the FN mechanism to the SMEFT, we have presented an important first step in the effort to utilize the SMEFT and all of its associated experimental and theoretical technologies to efficiently gain concrete, falsifiable insight on one of the SM's

oldest and most perplexing problems: the flavour puzzle. We are hopeful that our work will help motivate a more EFT-centered approach to BSM flavour physics, and therefore both experimental and theoretical advances to that end.

Acknowledgements

We thank Tyler Corbett for his help understanding certain SMEFT Feynman Rules in the unbroken electroweak phase, Rodrigo Alonso for reminding us about how to treat Goldstone bosons in non-renormalizable operators, the Cambridge Pheno Working Group for helpful discussions while completing this work, and Ben Allanach for useful comments in the finishing stages of the paper. The work of EL is supported by STFC consolidated grants ST/T000694/1 and ST/X000664/1. JT gratefully acknowledges prior funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101022203, and guest support from the T-2 Group at LANL. JT also thanks the participants of HEFT2023 for valuable insight, Steve King for inspiring early discussions that slowly evolved into this work, and Robert Szafron for clarifying remarks regarding matching formalisms.

References

[1] C. D. Froggatt and Holger Bech Nielsen. Hierarchy of Quark Masses, Cabibbo Angles and CP Violation. *Nucl. Phys. B*, 147:277–298, 1979.

[2] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[3] Ivan Esteban, M. C. Gonzalez-Garcia, Michele Maltoni, Thomas Schwetz, and Albert Zhou. The fate of hints: updated global analysis of three-flavor neutrino oscillations. *JHEP*, 09:178, 2020.

[4] Q. Bonnefoy, E. Dudas, and S. Pokorski. Chiral Froggatt-Nielsen models, gauge anomalies and flavourful axions. *JHEP*, 01:191, 2020.

[5] Ben Allanach and Eetu Loisa. Flavonstrahlung in the B_3-L_2 Z' model at current and future colliders. *JHEP*, 03:253, 2023.

[6] Claudia Cornella, David Curtin, Ethan T. Neil, and Jedidiah O. Thompson. Mapping and Probing Froggatt-Nielsen Solutions to the Quark Flavor Puzzle. 6 2023.

[7] W. Buchmuller and D. Wyler. Effective Lagrangian Analysis of New Interactions and Flavor Conservation. *Nucl. Phys. B*, 268:621–653, 1986.

[8] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP*, 10:085, 2010.

[9] Ilaria Brivio and Michael Trott. The Standard Model as an Effective Field Theory. *Phys. Rept.*, 793:1–98, 2019.

[10] Gino Isidori, Felix Wilsch, and Daniel Wyler. The Standard Model effective field theory at work. 3 2023.

[11] Landon Lehman and Adam Martin. Hilbert Series for Constructing Lagrangians: expanding the phenomenologist's toolbox. *Phys. Rev. D*, 91:105014, 2015.

[12] Steven Weinberg. Baryon and Lepton Nonconserving Processes. *Phys. Rev. Lett.*, 43:1566–1570, 1979.

[13] John Ellis, Maeve Madigan, Ken Mimasu, Veronica Sanz, and Tevong You. Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory. *JHEP*, 04:279, 2021.

[14] Jacob J. Ethier, Giacomo Magni, Fabio Maltoni, Luca Mantani, Emanuele R. Nocera, Juan Rojo, Emma Slade, Eleni Vryonidou, and Cen Zhang. Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC. *JHEP*, 11:089, 2021.

[15] Shayan Iranipour and Maria Ubiali. A new generation of simultaneous fits to LHC data using deep learning. *JHEP*, 05:032, 2022.

[16] Ilaria Brivio, Sebastian Bruggisser, Nina Elmer, Emma Geoffray, Michel Luchmann, and Tilman Plehn. To Profile or To Marginalize – A SMEFT Case Study. 8 2022.

[17] Adam Falkowski, Martín González-Alonso, and Kin Mimouni. Compilation of low-energy constraints on 4-fermion operators in the SMEFT. *JHEP*, 08:123, 2017.

[18] Adam Falkowski and David Straub. Flavourful SMEFT likelihood for Higgs and electroweak data. *JHEP*, 04:066, 2020.

[19] Vincenzo Cirigliano, Andreas Crivellin, Wouter Dekens, Jordy de Vries, Martin Hoferichter, and Emanuele Mereghetti. CP Violation in Higgs-Gauge Interactions: From Tabletop Experiments to the LHC. *Phys. Rev. Lett.*, 123(5):051801, 2019.

[20] Rafael Aoude, Tobias Hurth, Sophie Renner, and William Shepherd. The impact of flavour data on global fits of the MFV SMEFT. *JHEP*, 12:113, 2020.

[21] Stefan Bißmann, Cornelius Grunwald, Gudrun Hiller, and Kevin Kröninger. Top and Beauty synergies in SMEFT-fits at present and future colliders. *JHEP*, 06:010, 2021.

[22] Vincenzo Cirigliano, Kaori Fuyuto, Christopher Lee, Emanuele Mereghetti, and Bin Yan. Charged Lepton Flavor Violation at the EIC. *JHEP*, 03:256, 2021.

[23] Luigi Bellafronte, Sally Dawson, and Pier Paolo Giardino. The importance of flavor in SMEFT Electroweak Precision Fits. *JHEP*, 05:208, 2023.

[24] Cornelius Grunwald, Gudrun Hiller, Kevin Kröninger, and Lara Nollen. More Synergies from Beauty, Top, Z and Drell-Yan Measurements in SMEFT. 4 2023.

[25] Lukas Allwicher, Darius A. Faroughy, Florentin Jaffredo, Olcyr Sumensari, and Felix Wilsch. Drell-Yan tails beyond the Standard Model. *JHEP*, 03:064, 2023.

[26] Svjetlana Fajfer, Jernej F. Kamenik, Nejc Košnik, Aleks Smolkovič, and Michele Tammaro. New Physics in CP Violating and Flavour Changing Quark Dipole Transitions. 6 2023.

[27] Lukas Allwicher, Claudia Cornella, Ben A. Stefanek, and Gino Isidori. New Physics in the Third Generation: A Comprehensive SMEFT Analysis and Future Prospects. 10 2023.

[28] Javier Fuentes-Martin, Jorge Portoles, and Pedro Ruiz-Femenia. Integrating out heavy particles with functional methods: a simplified framework. *JHEP*, 09:156, 2016.

[29] Timothy Cohen, Xiaochuan Lu, and Zhengkang Zhang. Functional Prescription for EFT Matching. *JHEP*, 02:228, 2021.

[30] Javier Fuentes-Martín, Matthias König, Julie Pagès, Anders Eller Thomsen, and Felix Wilsch. A proof of concept for matchete: an automated tool for matching effective theories. *Eur. Phys. J. C*, 83(7):662, 2023.

[31] Sébastien Descotes-Genon, Adam Falkowski, Marco Fedele, Martín González-Alonso, and Javier Virto. The CKM parameters in the SMEFT. *JHEP*, 05:172, 2019.

[32] Tobias Hurth, Sophie Renner, and William Shepherd. Matching for FCNC effects in the flavour-symmetric SMEFT. *JHEP*, 06:029, 2019.

[33] Andreas Helset and Andrew Kobach. Baryon Number, Lepton Number, and Operator Dimension in the SMEFT with Flavor Symmetries. *Phys. Lett. B*, 800:135132, 2020.

[34] Jason Aebischer and Jacky Kumar. Flavour violating effects of Yukawa running in SMEFT. *JHEP*, 09:187, 2020.

[35] Darius A. Faroughy, Gino Isidori, Felix Wilsch, and Kei Yamamoto. Flavour symmetries in the SMEFT. *JHEP*, 08:166, 2020.

[36] Sebastian Bruggisser, Ruth Schäfer, Danny van Dyk, and Susanne Westhoff. The Flavor of UV Physics. *JHEP*, 05:257, 2021.

[37] Tatsuo Kobayashi, Hajime Otsuka, Morimitsu Tanimoto, and Kei Yamamoto. Modular symmetry in the SMEFT. *Phys. Rev. D*, 105(5):055022, 2022.

[38] Jim Talbert and Michael Trott. Dirac masses and mixings in the (geo)SM(EFT) and beyond. *JHEP*, 11:009, 2021.

[39] Quentin Bonnefoy, Emanuele Gendy, Christophe Grojean, and Joshua T. Ruderman. Beyond Jarlskog: 699 invariants for CP violation in SMEFT. *JHEP*, 08:032, 2022.

[40] Gino Isidori, Julie Pagès, and Felix Wilsch. Flavour alignment of New Physics in light of the $(g-2)_\mu$ anomaly. *JHEP*, 03:011, 2022.

[41] Admir Greljo, Ajdin Palavrić, and Anders Eller Thomsen. Adding Flavor to the SMEFT. *JHEP*, 10:010, 2022.

[42] Sally Dawson and Pier Paolo Giardino. Flavorful electroweak precision observables in the Standard Model effective field theory. *Phys. Rev. D*, 105(7):073006, 2022.

[43] Sebastian Bruggisser, Danny van Dyk, and Susanne Westhoff. Resolving the flavor structure in the MFV-SMEFT. *JHEP*, 02:225, 2023.

[44] Camila S. Machado, Sophie Renner, and Dave Sutherland. Building blocks of the flavourful SMEFT RG. *JHEP*, 03:226, 2023.

[45] Admir Greljo and Ajdin Palavrić. Leading directions in the SMEFT. *JHEP*, 09:009, 2023.

[46] Quentin Bonnefoy, Emanuele Gendy, Christophe Grojean, and Joshua T. Ruderman. Opportunistic CP violation. *JHEP*, 06:141, 2023.

[47] Stefan Antusch, Admir Greljo, Ben A. Stefanek, and Anders Eller Thomsen. U(2) is Right for Leptons and Left for Quarks. 11 2023.

[48] Lorenzo Calibbi, Florian Goertz, Diego Redigolo, Robert Ziegler, and Jure Zupan. Minimal axion model from flavor. *Phys. Rev. D*, 95(9):095009, 2017.

[49] Yohei Ema, Koichi Hamaguchi, Takeo Moroi, and Kazunori Nakayama. Flaxion: a minimal extension to solve puzzles in the standard model. *JHEP*, 01:096, 2017.

[50] Aleks Smolkovič, Michele Tammaro, and Jure Zupan. Anomaly free Froggatt-Nielsen models of flavor. *JHEP*, 10:188, 2019. [Erratum: JHEP 02, 033 (2022)].

[51] Michael B. Green and John H. Schwarz. Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory. *Phys. Lett. B*, 149:117–122, 1984.

[52] Luis E. Ibanez and Graham G. Ross. Discrete gauge symmetry anomalies. *Phys. Lett. B*, 260:291–295, 1991.

[53] Luis E. Ibanez and Graham G. Ross. Should discrete symmetries be anomaly free? 1 1991.

[54] Tom Banks and Michael Dine. Note on discrete gauge anomalies. *Phys. Rev. D*, 45:1424–1427, 1992.

[55] Takeshi Araki. Anomaly of Discrete Symmetries and Gauge Coupling Unification. *Prog. Theor. Phys.*, 117:1119–1138, 2007.

[56] Takeshi Araki, Tatsuo Kobayashi, Jisuke Kubo, Saul Ramos-Sanchez, Michael Ratz, and Patrick K. S. Vaudrevange. (Non-)Abelian discrete anomalies. *Nucl. Phys. B*, 805:124–147, 2008.

[57] Hajime Ishimori, Tatsuo Kobayashi, Hiroshi Ohki, Yusuke Shimizu, Hiroshi Okada, and Morimitsu Tanimoto. Non-Abelian Discrete Symmetries in Particle Physics. *Prog. Theor. Phys. Suppl.*, 183:1–163, 2010.

[58] Mu-Chun Chen, Maximilian Fallbacher, Michael Ratz, Andreas Trautner, and Patrick K. S. Vaudrevange. Anomaly-safe discrete groups. *Phys. Lett. B*, 747:22–26, 2015.

[59] Jim Talbert. Pocket Formulae for Non-Abelian Discrete Anomaly Freedom. *Phys. Lett. B*, 786:426–431, 2018.

[60] Ben Gripaios. Gauge anomalies of finite groups. *Phys. Rev. D*, 105(10):105008, 2022.

[61] Joe Davighi, Ben Gripaios, and Nakarin Lohitsiri. Anomalies of non-Abelian finite groups via cobordism. *JHEP*, 09:147, 2022.

[62] Lawrence M. Krauss and Frank Wilczek. Discrete Gauge Symmetry in Continuum Theories. *Phys. Rev. Lett.*, 62:1221, 1989.

[63] Alexander Baur, Hans Peter Nilles, Andreas Trautner, and Patrick K. S. Vaudrevange. Unification of Flavor, CP, and Modular Symmetries. *Phys. Lett. B*, 795:7–14, 2019.

[64] Alexander Baur, Hans Peter Nilles, Andreas Trautner, and Patrick K. S. Vaudrevange. A String Theory of Flavor and $\mathscr{CP}$. *Nucl. Phys. B*, 947:114737, 2019.

[65] Alexander Baur, Hans Peter Nilles, Saul Ramos-Sanchez, Andreas Trautner, and Patrick K. S. Vaudrevange. Top-down anatomy of flavor symmetry breakdown. *Phys. Rev. D*, 105(5):055018, 2022.

[66] L. F. Abbott. Introduction to the Background Field Method. *Acta Phys. Polon. B*, 13:33, 1982.

[67] Javier Fuentes-Martin, Matthias König, Julie Pagès, Anders Eller Thomsen, and Felix Wilsch. SuperTracer: A Calculator of Functional Supertraces for One-Loop EFT Matching. *JHEP*, 04:281, 2021.

[68] Adrian Carmona, Achilleas Lazopoulos, Pablo Olgoso, and Jose Santiago. Matchmakereft: automated tree-level and one-loop matching. *SciPost Phys.*, 12(6):198, 2022.

[69] Javier Fuentes-Martín, Matthias König, Julie Pagès, Anders Eller Thomsen, and Felix Wilsch. Evanescent operators in one-loop matching computations. *JHEP*, 02:031, 2023.

[70] Pierre Binetruy, Stephane Lavignac, and Pierre Ramond. Yukawa textures with an anomalous horizontal Abelian symmetry. *Nucl. Phys. B*, 477:353–377, 1996.

[71] Eung Jin Chun and Andre Lukas. Quark and lepton mass matrices from horizontal U(1) symmetry. *Phys. Lett. B*, 387:99–106, 1996.

[72] Minyuan Jiang, Nathaniel Craig, Ying-Ying Li, and Dave Sutherland. Complete one-loop matching for a singlet scalar in the Standard Model EFT. *JHEP*, 02:031, 2019. [Erratum: JHEP 01, 135 (2021)].

[73] Ulrich Haisch, Maximilian Ruhdorfer, Ennio Salvioni, Elena Venturini, and Andreas Weiler. Singlet night in Feynman-ville: one-loop matching of a real scalar. *JHEP*, 04:164, 2020. [Erratum: JHEP 07, 066 (2020)].

[74] Sebastian A. R. Ellis, Jérémie Quevillon, Tevong You, and Zhengkang Zhang. Extending the Universal One-Loop Effective Action: Heavy-Light Coefficients. *JHEP*, 08:054, 2017.

[75] Rodrigo Alonso, Elizabeth E. Jenkins, Aneesh V. Manohar, and Michael Trott. Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology. *JHEP*, 04:159, 2014.

[76] Joshua Ellis. TikZ-Feynman: Feynman diagrams with TikZ. *Comput. Phys. Commun.*, 210:103–123, 2017.

[77] Minyuan Jiang, Nathaniel Craig, Ying-Ying Li, and Dave Sutherland. Erratum to: Complete one-loop matching for a singlet scalar in the standard model eft. *Journal of High Energy Physics*, 2021(1):135, 2021.

[78] Elizabeth E. Jenkins, Aneesh V. Manohar, and Peter Stoffer. Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching. *JHEP*, 03:016, 2018. [Erratum: JHEP 12, 043 (2023)].

[79] Wouter Dekens and Peter Stoffer. Low-energy effective field theory below the electroweak scale: matching at one loop. *JHEP*, 10:197, 2019. [Erratum: JHEP 11, 148 (2022)].

[80] Elizabeth E. Jenkins, Aneesh V. Manohar, and Michael Trott. Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence. *JHEP*, 10:087, 2013.

[81] Elizabeth E. Jenkins, Aneesh V. Manohar, and Michael Trott. Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence. *JHEP*, 01:035, 2014.

[82] Coenraad B. Marinissen, Rudi Rahn, and Wouter J. Waalewijn. ..., 83106786, 114382724, 1509048322, 2343463290, 27410087742, ... efficient Hilbert series for effective theories. *Phys. Lett. B*, 808:135632, 2020.

[83] Simon Calò, Coenraad Marinissen, and Rudi Rahn. Discrete symmetries and efficient counting of operators. *JHEP*, 05:215, 2023.

[84] Brian Henning, Xiaochuan Lu, Tom Melia, and Hitoshi Murayama. Hilbert series and operator bases with derivatives in effective field theories. *Commun. Math. Phys.*, 347(2):363–388, 2016.

[85] Brian Henning, Xiaochuan Lu, Tom Melia, and Hitoshi Murayama. 2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT. *JHEP*, 08:016, 2017. [Erratum: JHEP 09, 019 (2019)].

[86] Brian Henning, Xiaochuan Lu, Tom Melia, and Hitoshi Murayama. Operator bases, S-matrices, and their partition functions. *JHEP*, 10:199, 2017.

[87] Ben Gripaios and Dave Sutherland. An operator basis for the Standard Model with an added scalar singlet. *JHEP*, 08:103, 2016.

[88] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia. Minimal flavor violation: An Effective field theory approach. *Nucl. Phys. B*, 645:155–187, 2002.

[89] Jason Aebischer, Jacky Kumar, Peter Stangl, and David M. Straub. A Global Likelihood for Precision Constraints and Flavour Anomalies. *Eur. Phys. J. C*, 79(6):509, 2019.

[90] Peter Minkowski. $\mu \to e\gamma$ at a Rate of One Out of 10^9 Muon Decays? *Phys. Lett. B*, 67:421–428, 1977.

[91] Ivo de Medeiros Varzielas, Graham G. Ross, and Jim Talbert. A Unified Model of Quarks and Leptons with a Universal Texture Zero. *JHEP*, 03:007, 2018.

[92] Jordan Bernigaud, Ivo de Medeiros Varzielas, Miguel Levy, and Jim Talbert. Revisiting the universal texture zero of flavour: a Markov chain Monte Carlo analysis. *Eur. Phys. J. C*, 83(6):479, 2023.

[93] Ivo de Medeiros Varzielas and Graham G. Ross. SU(3) family symmetry and neutrino bi-tri-maximal mixing. *Nucl. Phys. B*, 733:31–47, 2006.

[94] Ivo de Medeiros Varzielas, David Emmanuel-Costa, and Philipp Leser. Geometrical CP Violation from Non-Renormalisable Scalar Potentials. *Phys. Lett. B*, 716:193–196, 2012.

[95] Guido Altarelli and Ferruccio Feruglio. Tri-bimaximal neutrino mixing from discrete symmetry in extra dimensions. *Nucl. Phys. B*, 720:64–88, 2005.

[96] Guido Altarelli and Ferruccio Feruglio. Tri-bimaximal neutrino mixing, A(4) and the modular symmetry. *Nucl. Phys. B*, 741:215–235, 2006.

[97] K. S. Babu, Ernest Ma, and J. W. F. Valle. Underlying A(4) symmetry for the neutrino mass matrix and the quark mixing matrix. *Phys. Lett. B*, 552:207–213, 2003.

[98] S. F. King and Graham G. Ross. Fermion masses and mixing angles from SU(3) family symmetry. *Phys. Lett. B*, 520:243–253, 2001.

[99] S. F. King and Graham G. Ross. Fermion masses and mixing angles from SU (3) family symmetry and unification. *Phys. Lett. B*, 574:239–252, 2003.

[100] Mario Reig, José W. F. Valle, and Frank Wilczek. SO(3) family symmetry and axions. *Phys. Rev. D*, 98(9):095008, 2018.

[101] Anne Mareike Galda, Matthias Neubert, and Sophie Renner. ALP — SMEFT interference. *JHEP*, 06:135, 2021.

[102] J. de Blas, J. C. Criado, M. Perez-Victoria, and J. Santiago. Effective description of general extensions of the Standard Model: the complete tree-level dictionary. *JHEP*, 03:109, 2018.

[103] Christopher W. Murphy. Dimension-8 operators in the Standard Model Eective Field Theory. *JHEP*, 10:174, 2020.

[104] Hao-Lin Li, Zhe Ren, Jing Shu, Ming-Lei Xiao, Jiang-Hao Yu, and Yu-Hui Zheng. Complete set of dimension-eight operators in the standard model effective field theory. *Phys. Rev. D*, 104(1):015026, 2021.

[105] Andreas Helset, Adam Martin, and Michael Trott. The Geometric Standard Model Effective Field Theory. *JHEP*, 03:163, 2020.

[106] Kieran Finn, Sotirios Karamitsos, and Apostolos Pilaftsis. Frame covariant formalism for fermionic theories. *Eur. Phys. J. C*, 81(7):572, 2021.

[107] Jim Talbert. The geometric νSMEFT: operators and connections. *JHEP*, 01:069, 2023.

[108] Viola Gattus and Apostolos Pilaftsis. Minimal supergeometric quantum field theories. *Phys. Lett. B*, 846:138234, 2023.

[109] Benoît Assi, Andreas Helset, Aneesh V. Manohar, Julie Pagès, and Chia-Hsien Shen. Fermion geometry and the renormalization of the Standard Model Effective Field Theory. *JHEP*, 11:201, 2023.