A New Probe of Cosmic Birefringence Using Galaxy Polarization and Shapes

Weichen Winston Yin,[1,*] Liang Dai,[1] Junwu Huang,[2] Lingyuan Ji,[1] and Simone Ferraro[3,1]

[1] *Department of Physics, 366 Physics North MC 7300, University of California, Berkeley, CA 94720, USA*

[2] *Perimeter Institute for Theoretical Physics, 31 Caroline St. N., Waterloo, Ontario N2L 2Y5, Canada*

[3] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

(Dated: February 29, 2024)

We propose a new method to search for parity-violating new physics via measurements of cosmic birefringence and demonstrate its power in detecting the topological effect originating from an axion string network with an axion-photon coupling as a motivated source of cosmic birefringence. The method, using large galaxy samples, exploits an empirical correlation between the polarization direction of the integrated radio emission from a spiral galaxy and its apparent shape. We devise unbiased minimum-variance quadratic estimators for discrete samples of galaxies with both integrated radio polarization and shape measurements. Assuming a synergy with overlapping optical imaging surveys, we forecast the sensitivity to polarization rotation of the forthcoming SKA radio continuum surveys of spiral galaxies out to $z \sim 1.5$. The angular noise power spectrum of polarization rotation using our method can be lower than that expected from CMB Stage-IV experiments, when assuming a wide survey covering $\sim 1000\,\mathrm{deg}^2$ and reaching an RMS flux of $\sim 1\,\mu$Jy. Our method will be complementary to CMB-based methods as it will be subject to different systematics. It can be generalized to probe time-varying or redshift-varying birefringence signals.

arXiv:2402.18568v1 [astro-ph.CO] 28 Feb 2024

Introduction — The question of whether discrete transformations—charge conjugation (C), parity (P), and time reversal (T)—are fundamental symmetries of the Universe has led to important discoveries [1–3]. The axion was originally proposed to solve the strong CP problem in Quantum Chromodynamics [4–7]. More recently, generic axions are shown to arise abundantly in string theory [8, 9]. The Chern-Simons interaction between the axion and electromagnetism leads to the parity-violating cosmic birefringence effect [10]:

$$\mathcal{L} = -\frac{1}{4}\, g_{a\gamma\gamma}\, a\, F^{\mu\nu}\, \tilde{F}_{\mu\nu} = g_{a\gamma\gamma}\, a\, \mathbf{E}\cdot\mathbf{B}, \qquad (1)$$

where a is the axion field, $F^{\mu\nu}$ is the electromagnetic field strength tensor, and $\tilde{F}^{\mu\nu} \equiv (1/2)\,\epsilon_{\mu\nu\rho\sigma}\, F^{\rho\sigma}$ is its dual. As a photon travels through a background axion field, its plane of polarization rotates by an angle proportional to the net change in the axion field value along the photon's worldline, multiplied by the axion-photon coupling constant $g_{a\gamma\gamma}$. The rotation is independent of the photon's frequency due to the topological nature of the interaction [11–13]. Recent analyses of the cosmic microwave background (CMB) have shown hints of a rotation of its polarization anisotropies that is highly coherent across the sky, on the order of a fraction of a degree [14, 15]. Possible contamination from foregrounds is being investigated [16]. On the other hand, searches for anisotropic birefrigence have not so far uncovered a signal [17–20].

Measuring birefringence toward any single astrophysical source or along any single sightline is hindered by our lack of knowledge about the *intrinsic* polarization direction of the source. When the CMB is used as the probe, this fundamental problem is solved statistically. Rotation angles that are correlated across the sky are measurable owing to the state-of-the-art cosmological model that accurately predicts the Gaussian statistics of the primary CMB polarization anisotropies [20–23]. To confidently confirm a signal of cosmic birefringence, it is important to consider alternative methods of detection based on distinct cosmological or astrophysical probes whose systematics are independent of CMB measurements.

In this paper, we propose a new probe of cosmic birefringence that relies on a large sample of spiral galaxies with both radio polarization and apparent shape measurements. This is motivated by empirical findings that the *integrated* radio emission of nearby spiral galaxies, at least at 4.8 GHz, is significantly polarized and aligns on average with the apparent minor axis of the optical galactic disk [24]. A plausible explanation for this alignment is that the polarized radio emission is predominantly synchrotron radiation powered by star-formation feedback and has a polarization direction perpendicular to the local toroidal magnetic field that is ordered on the galactic scale. Integrating over the entire disk results in a nonzero net polarization if the disk is inclined. It seems reasonable to assume that a similar empirical correlation is true for a cosmological population of spiral galaxies. The apparent disk shape can be measured with either optical imaging or interferometric radio imaging, with the former available from existing weak lensing catalogs [25–27] or from upcoming imaging surveys [28, 29].

In reality, the polarization-shape alignment is imperfect either due to the poloidal component of the ordered magnetic field or due to randomly arising asymmetry in the disk structure. Nonetheless, one expects that the mean polarization-shape misalignment angle is zero among a sample of unrelated galaxies, as long as all the physics on the (sub-)galactic scale responsible for the

* winstonyin@berkeley.edu

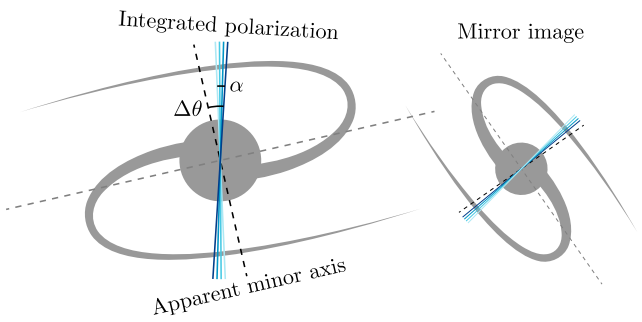


FIG. 1. An illustration of the parity-violating effect cosmic birefringence has on the (mis)alignment between the linear polarization of the integrated radio emission (solid blue) and the apparent minor axis (dashed black) of a spiral galaxy. For every spiral galaxy with a mislignment angle $\Delta\theta$ (left), another galaxy whose structure is its mirror image (right) would have the opposite misalignment angle, and so $\Delta\theta$ is expected to average to zero over many galaxies. In the presence of birefringence, the integrated polarization is rotated by an angle α (from light to dark blue) for both the galaxy and its mirror image, and $\Delta\theta$ would be statistically biased from zero.

polarization-shape correlation respects parity. If there is one galaxy whose integrated radio polarization is offset from the apparent minor axis to the clockwise side, it would be equally probable that another galaxy appears as its mirror reflection and has a misalignment to the counter-clockwise side (Fig. 1).

It has been proposed that such polarization-shape correlation can be exploited to calibrate the intrinsic shapes of galaxies, assuming that polarization is not externally modified, which reduces noise in weak lensing shear measurements [30–32]. By reversing this logic, one can calibrate the intrinsic polarization direction relative to source morphology [33, 34]. With cosmic birefringence, unrelated galaxies in a patch of the sky will have their polarization rotated by nearly the same angle, leading to a polarization-shape correlation that is biased from zero.

In this Letter, we construct a minimum-variance quadratic estimator for polarization rotation angles using a catalog of galaxies with both integrated radio polarization and shape measurements, and forecast the angular noise power spectrum for the estimator. Radio polarization measurements may come from the existing VLA Sky Survey [35] or the upcoming SKA continuum surveys [36], while the optical counterparts can come from an overlapping photometric galaxy catalog with shape measurements, such as from the DESI Legacy Surveys (DECaLS) [28] or the forthcoming LSST [29].

Galaxy Radio Polarization — When a galaxy is viewed at an inclination, the magnetic fields appear aligned with the apparent major axis. This leads to a net linear polarization of the integrated radio emission as found in [24]. The degree of polarization of the integrated emission depends on the galaxy's disk inclination and radio luminosity, as well as the presence of a randomly oriented component of the magnetic field throughout the galaxy. Ref. [24] found in the 4.8 GHz emission from nearby spiral galaxies that the net degree of polarization ranges from $\sim 1\%$ to $\sim 18\%$, with a mean of $\sim 4\%$. While the alignment between the net polarization and the minor axis is not exact in any individual galaxy, the sample of [24] shows a mean misalignment angle consistent with zero and an RMS value of $\sim 26^\circ$. As we will see later, the effective scatter in this misalignment angle, which limits the precision of cosmic birefringence measurements, is only as small as $\sim 5^\circ$, once a proper quadratic estimator is constructed. This rather remarkable result is uncovered empirically from the sample of [24], which indicates a strong correlation between the scatter in misalignment angle and disk inclination.

Estimating Cosmic Birefrigence — the linear polarization of a galaxy can be described by a pair of Stokes parameters (Q, U). A spin-2 variable $P_\pm$ can be defined

$$P_\pm \equiv Q \pm i\,U = I\,p\,e^{\pm 2\,i\,\theta_P}, \tag{2}$$

where I is the intensity [37], p is the degree of polarization $(0 < p < 1)$, and θ_P is the position angle of the linear polarization in the sky. The apparent elliptical shape of an inclined galaxy can be described by another spin-2 variable, the ellipticity $\epsilon_\pm$ [38, p. 11]

$$\epsilon_\pm \equiv q \pm i\,u = (a-b)/(a+b)\,e^{\pm 2\,i\,\theta_\epsilon}, \tag{3}$$

where a and b are the semi-minor and semi-major axes, respectively, and θ_ϵ is the position angle of the semi-major axis in the sky.

Since $\theta_P - \theta_\epsilon$ is invariant under rotations of the sky, we define the following (dimensionless) spin-0 quantities

$$X \pm i\,Y \equiv -P_\pm\,\epsilon_\mp / I = p\,|\epsilon|\,e^{\pm 2\,i\,\Delta\theta}, \tag{4}$$

where $\Delta\theta \equiv \theta_P - \theta_\epsilon - \pi/2$ is the polarization-shape misalignment angle and $|\epsilon| \equiv (a-b)/(a+b)$. The quantities X and Y have even and odd parity, respectively.

The quantities p, $|\epsilon|$, and $\Delta\theta$ are intrinsic to each galaxy. Since there is no known mechanism for *macroscopic* parity violation in the synchrotron radiation in the interstellar medium (ISM), we assume that $\Delta\theta$ has a distribution symmetric about $\Delta\theta = 0$, and that possible parity-violating effects occur externally as a result of cosmic birefringence. Cosmic birefringence can be quantified by a sightline-dependent polarization rotation field $\alpha(\hat{\mathbf{n}})$, coherent over some angular scale on the sky. Depending on the birefringence source, $\alpha(\hat{\mathbf{n}})$ may be uniform on the sky [39], or behave as a Gaussian random field [40, 41], or have a highly non-Gaussian spatial pattern [13]. Thus, cosmic birefringence causes the distribution of $\Delta\theta$ to be biased from zero by an amount $\alpha(\hat{\mathbf{n}})$ for galaxies in a local sky patch in the direction $\hat{\mathbf{n}}$. For a first analysis, we assume that $\alpha(\hat{\mathbf{n}})$ is independent of galaxy distance, and that $(I\,p, \Delta\theta)$ are independently drawn from the same global distribution for individual galaxies.

While the Faraday effect of the Milky Way (MW)'s ISM also rotates the polarization coherently on the sky,

the rotation angle scales as the wavelength squared. This can be distinguished from the wavelength-independent birefringence effect induced by an axion [11]. We therefore assume that MW Faraday rotation is subtracted as a foreground effect prior to the following analysis. This does require observation in multiple radio bands, the effectiveness of which is subject to further study. On the other hand, Faraday effect within the source galaxy causes incoherent polarization rotation across the galactic disk, which leads to the depolarization of the integrated radio emission that is stronger at lower radio frequencies [24, 42]. This is not a concern in the following analysis, as depolarization at 4.8 GHz is included in the empirical polarization-shape (mis)alignment data.

Parity-odd combinations of X and Y have a vanishing mean over the ensemble of galaxies when $\alpha(\hat{\mathbf{n}}) = 0$:

$$\langle Y \rangle = \langle X\, Y \rangle = 0. \tag{5}$$

Statistically significant deviation of $\langle Y \rangle$ or $\langle XY \rangle$ from zero would therefore be a smoking gun of parity violation.

In the presence of a nonzero $\alpha(\hat{\mathbf{n}})$, the disk shape is unaffected while the integrated polarization is rotated:

$$P_\pm \to P_\pm\, e^{\pm 2\, i\, \alpha(\hat{\mathbf{n}})}, \quad \epsilon_\pm \to \epsilon_\pm. \tag{6}$$

This implies transformation rules for X and Y

$$X \to X\, \cos 2\alpha(\hat{\mathbf{n}}) - Y\, \sin 2\alpha(\hat{\mathbf{n}}), \tag{7}$$

$$Y \to Y\, \cos 2\alpha(\hat{\mathbf{n}}) + X\, \sin 2\alpha(\hat{\mathbf{n}}). \tag{8}$$

Notably, the value of $\alpha(\hat{\mathbf{n}})$ does not change observables of a galaxy along a different direction $\hat{\mathbf{n}}' \neq \hat{\mathbf{n}}$.

We construct an estimator for $\alpha(\hat{\mathbf{n}})$ using X and Y, inspired by the minimum-variance quadratic estimators for weak lensing [43] and for polarization rotation [21–23, 44] in analyzing CMB anistropies. In App. B, this estimator is confirmed by a maximum likelihood approach [45].

For CMB anisotropies, quadratic estimators for cosmic shear and for polarization rotation are orthogonal to each other to first order in each of the effects [20]. Cosmic shear induces a white-noise $Y \approx p\,|\epsilon|\,\gamma = 0.002\,(p/0.1)\,(\gamma/0.02)\,|\epsilon|$, where $\sqrt{\langle\gamma^2\rangle} \approx 0.02$ at $z \sim 1$ [46], which however is smaller than the intrinsic $\sigma_Y = 0.004$ (Eq. (13)). With both weak lensing and galaxy intrinsic alignment, a correlated noise contribution to $\hat{\alpha}$ arises, but with a magnitude significantly smaller than the noise floor derived here. We will therefore disregard weak lensing in the following calculation.

To estimate the polarization rotation $\alpha(\hat{\mathbf{n}})$ using a single galaxy in the direction $\hat{\mathbf{n}}$, we write down an estimator $\hat{\alpha}(\hat{\mathbf{n}})$ as the general combination of the observed quantities $\tilde{X}$ and $\tilde{Y}$ (including measurement noise) for this galaxy up to quadratic terms:

$$\hat{\alpha} \equiv u\,\tilde{X} + v\,\tilde{Y} + u'\,\tilde{X}^2 + v'\,\tilde{Y}^2 + w\,\tilde{X}\,\tilde{Y}, \tag{9}$$

where the coefficients u, v, u', v', w are to be determined. We omit the dependence on $\hat{\mathbf{n}}$ for clarity of notation whenever only a single galaxy is concerned.

Following the idea underlying the minimum-variance quadratic estimator, we require that $\hat{\alpha}$ be an unbiased estimator of α. That is, $\langle\hat{\alpha}\rangle = \alpha$ when averaged over an ensemble of galaxies with independent *intrinsic* properties. Subject to this requirement, we then minimize the variance of $\hat{\alpha}$ under the null hypothesis $\alpha = 0$. To carry out the calculation analytically, we approximate X as a Gaussian random variable with mean $\overline{X}$ and variance σ_X^2, and Y a Gaussian random variable with *zero* mean and variance σ_Y^2. Since X and Y are only approximately Gaussian, this gives a nearly optimal quadratic estimator. The true optimal estimator could be obtained by numerically minimizing $\langle\hat{\alpha}^2\rangle$, given the empirically measured distributions for Ip, $|\epsilon|$, and $\Delta\theta$.

Observed quantities have measurement uncertainties. If the radio intensity measurement has an RMS uncertainty of σ_I, then the uncertainty on the polarization degree p is σ_I/I, which equals the inverse unpolarized signal-to-noise ratio (SNR) of the galaxy. We find that with realistic deep optical imaging surveys that include shape measurement, the radio polarization measurement error dominates over the error in ellipticity measurement. This leads to the overall errors for $\tilde{X}$ and $\tilde{Y}$, also assumed to be Gaussian:

$$\tilde{\sigma}^2_{X,Y} = \sigma^2_{X,Y} + (1/2)\,\left(\Delta\epsilon/\mathrm{SNR}\right)^2, \tag{10}$$

where $\Delta\epsilon \approx 0.3$ is the intrinsic spread in galaxy ellipticity, and $\sigma^2_{X,Y}$ are the intrinsic variances of X and Y.

Since large rotation angles have been ruled out by observation, it is reasonable to assume that polarization rotation is small $\alpha \ll 1$. We find the following unbiased minimum-variance quadratic estimator (detailed derivation can be found in App. A):

$$\hat{\alpha} = \frac{\tilde{\sigma}_Y^2\,\overline{X}\,\tilde{Y} + (\sigma_X^2 - \sigma_Y^2)\,\tilde{X}\,\tilde{Y}}{2\left[(\sigma_X^2 - \sigma_Y^2)^2 + \tilde{\sigma}_X^2\,\overline{X}^2\right]}, \tag{11}$$

whose variance under the null hypothesis is

$$\left\langle\hat{\alpha}^2\right\rangle = \frac{\tilde{\sigma}_X^2\,\tilde{\sigma}_Y^2}{4\left[(\sigma_X^2 - \sigma_Y^2)^2 + \tilde{\sigma}_X^2\,\overline{X}^2\right]}. \tag{12}$$

Conceptually, we have obtained the expected result that the estimator for the parity-violating $\alpha(\hat{\mathbf{n}})$ is constructed from the parity-odd observables Y and $X\,Y$ only. Since $\langle\hat{\alpha}(\hat{\mathbf{n}})\rangle = 0$ when $\alpha(\hat{\mathbf{n}}) = 0$, a statistically significant deviation of $\langle\hat{\alpha}(\hat{\mathbf{n}})\rangle$ from zero toward any direction $\hat{\mathbf{n}}$ over a number of galaxies would be evidence for cosmic birefringence in that direction.

Sensitivity — From the observables p, $|\epsilon|$, and $\Delta\theta$, X and Y for each galaxy can be computed. We obtain values for p and $\Delta\theta$ for 13 nearby spiral galaxies from Ref. [24], and compute values for $|\epsilon|$ from the angular major and minor axes given in the SIMBAD Astronomical Database [47]. We exclude M31 from that sample, as its close proximity and its particularly high degree of

polarization at 18% hint at possible systematics. This results in the following:

$$\overline{X} \pm \sigma_X = (1.3 \pm 2.1) \times 10^{-2},$$
$$\overline{Y} \pm \sigma_Y = (0.04 \pm 4.4) \times 10^{-3}. \quad (13)$$

Despite the small sample size used here, we find $\sqrt{\langle|\epsilon|^2\rangle} = 0.30$, in good agreement with the typical value for cosmological samples. It is more uncertain whether the properties of the integrated radio polarization of this sample are representative of a cosmological sample. Verification would require analyses of samples of distant galaxies (e.g. [35, 42, 48]). Using values in Eq. (13) for $\overline{X}$, σ_X and σ_Y while fixing $\overline{Y} = 0$, Eq. (12) gives the standard error in the rotation estimator for a single galaxy, $\sqrt{\langle\hat{\alpha}^2\rangle} = 5°$, for negligible measurement error.

Remarkably, $\sqrt{\langle\hat{\alpha}^2\rangle} = 5°$ is about five times smaller than the standard deviation of $\Delta\theta$ at 26° as estimated from nearby spiral galaxies. We attribute this startling difference to an anti-correlation between the scatter in the misalignment angle $\Delta\theta$ and the product $p\,|\epsilon|$. According to Eq. (4), when computing the statistics of X and Y in Eq. (13), galaxies with large $\Delta\theta$ values are weighted less due to smaller $p\,|\epsilon|$ values.

This anti-correlation has a geometrical origin (the effect is quantitatively modelled in [24]). The intrinsic misalignment angle has a large variance for galaxies showing a low integrated polarization degree and a low ellipticity, as these galaxies mostly have a low inclination and appear rather circular. Since p and $|\epsilon|$ both increase with inclination, the prefactor $p\,|\epsilon|$ effectively down-weights face-on galaxies and emphasizes edge-on galaxies on which birefringence would have a larger detectable effect. While we do not pursue it in this work, it seems possible to us that if this anti-correlation is accurately modelled as a function of the inclination, then an improved definition of X and Y could be derived to replace Eq. (4).

To make sure that the surprisingly small value of $\sqrt{\langle\hat{\alpha}^2\rangle} = 5°$ is not a fluke of low-number statistics from only 13 galaxies, we perform bootstrapping on the data extracted from Refs. [24, 47]. In each run, 4,000 pairs of $(p|\epsilon|, \Delta\theta)$ are generated by independent draws from the 13 galaxies with replacement. The value of $\langle\hat{\alpha}^2\rangle$ is then computed from the 4,000 artificial pairs using Eq. (12). We repeat this procedure 2,000 times, and obtain $\sqrt{\langle\hat{\alpha}^2\rangle} = (44 \pm 5)°$ (see Fig. 2), strongly suggesting that the anti-correlation between $\Delta\theta$ and $p\,|\epsilon|$ is real.

Galaxy Survey — Under the flat-sky approximation, the estimator is given by a finite sum over N discrete galaxies

$$\hat{\alpha}(\mathbf{L}) = (\Omega/N) \sum_{i=1}^{N} e^{-i\mathbf{L}\cdot\hat{\mathbf{n}}_i}\, \hat{\alpha}(\hat{\mathbf{n}}_i), \quad (14)$$

where Ω is survey area in steradians. The variance of the estimator is then

$$\langle\hat{\alpha}(\mathbf{L})\,\hat{\alpha}(\mathbf{L}')^*\rangle = (\Omega/N)^2\, \langle\hat{\alpha}^2\rangle \sum_{i=1}^{N} e^{-i\,(\mathbf{L}-\mathbf{L}')\cdot\hat{\mathbf{n}}_i}. \quad (15)$$

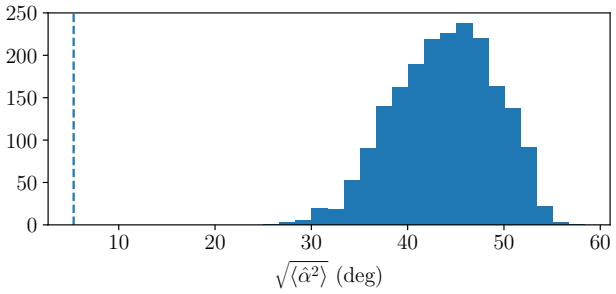


FIG. 2. Histogram of the standard error of $\hat{\alpha}$ when the anti-correlation between $p|\epsilon|$ and $\Delta\theta$ is erased by bootstrapping using a sample of 13 galaxies, repeated 2,000 times. The dashed line marks the standard error of $\hat{\alpha}$ measured from the 13 galaxies without boostrapping. Its significantly lower value compared to the bootstrapped distribution indicates that there is a genuine anti-correlation between $\Delta\theta$ and $p\,|\epsilon|$.

When galaxies are dense on the sky, the sum approximates a Dirac delta function:

$$\langle\hat{\alpha}(\mathbf{L})\,\hat{\alpha}(\mathbf{L}')^*\rangle = (\Omega/N)\, \langle\hat{\alpha}^2\rangle\, (2\pi)^2\, \delta^{(2)}(\mathbf{L}-\mathbf{L}'). \quad (16)$$

In the absence of birefringence, the estimation for the rotation power spectrum is therefore limited by a white-noise floor

$$C_{\hat{\alpha}\hat{\alpha}}(\mathbf{L}) = (\Omega/N)\, \langle\hat{\alpha}^2\rangle, \quad (17)$$

analogous to the shot noise in cosmic shear estimation from a discrete sample of galaxies.

We adapt the procedure used in Ref. [49] to count the number of available spiral galaxies. We first compute the redshift-dependent radio luminosity function of spiral galaxies. The radio luminosity function at 4.8 GHz at redshift z can be expressed in terms of the radio luminosity function at 1.4 GHz at $z = 0$, taking into account luminosity evolution [50] and the spectral index ($S_\nu \propto \nu^\alpha$ where $\alpha = -0.7$ [51, 52]). We choose 4.8 GHz because Ref. [24] established the empirical correlation at this frequency and because Faraday rotation in both the source and the MW is greatly reduced at higher frequencies. Next, the local radio luminosity function can be obtained from the far-infrared (FIR) luminosity function [53, 54] by exploiting an approximately linear relation between the radio and FIR luminosities of star-forming galaxies [55, 56].

The number of spiral galaxies with an observed spectral flux density S_ν above a threshold $S_\nu^{\rm min}$ is given by

$$N[S_\nu > S_\nu^{\rm min}] = \quad (18)$$
$$\Omega \int \mathrm{d}z \int_{\log \lambda(z)^{-1} L_{4.8}^{\rm min}(z)}^{\log L_{4.8}^{\rm max}} \mathrm{d}\log L_{4.8}\, \frac{c\, s(z)^2}{H(z)}\, \phi(\log L_{4.8}),$$

where $s(z) = \int_0^z c/H(z')\,\mathrm{d}z'$ is the comoving distance to redshift z; $H(z)$ is the expansion rate for $H_0 =$

$70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}, \Omega_\Lambda = 0.7, \Omega_m = 0.3$; Ω is the surveyed sky area in steradians; ϕ is the 4.8 GHz radio luminosity function (per unit comoving volume) of spiral galaxies at $z = 0$ (details can be found in App. D); $L_{4.8}^{\max} = 10^{25}\,\mathrm{W\,Hz^{-1}}$ is a generous upper bound on the radio luminosity of a spiral galaxy; $L_{4.8}^{\min}$ is the luminosity of a galaxy at redshift z with an observed spectral flux density of $S_\nu^{\min}$:

$$L_{4.8}^{\min}(z) = 4\pi\, s(z)^2\,(1+z)^{1-\alpha}\, S_\nu^{\min}; \tag{19}$$

and $\lambda(z)$ describes the luminosity evolution with z for spiral galaxies [50]. The upper limit in the z integral should not exceed $z = 1.5$; above that, nearly all massive halos host spheroidal galaxies [57]. The lower limit in the z integral ideally should exceed the redshift of detectable sources of birefringence. There is then a tradeoff between the galaxy number and the accumulation of birefringence along the line of sight. A more sophisticated treatment would consider birefringence sources interspersed amongst the galaxies throughout all redshift $0 < z < 1.5$. We leave such model-dependent calculations to a future study.

The highest sensitivity achievable with available data in the near future will likely come from correlating polarized SKA radio galaxies with LSST shape data (or using radio shapes from resolved SKA images). Similar to Ref. [58], we consider `Wide`, `Deep`, and `Ultradeep` reference surveys, which have $\sigma_I = 1\,\mu$Jy, $0.2\,\mu$Jy, and $0.05\,\mu$Jy, and cover $1000\,\mathrm{deg}^2$, $20\,\mathrm{deg}^2$, and $1\,\mathrm{deg}^2$ on the sky, respectively, but for SKA-MID Band 5a (4.6–8.5 GHz [59]). Including galaxies whose integrated radio intensities have SNR > 5 at 4.8 GHz, we estimate the number of measurable spiral galaxies at $0.2 < z < 1.5$ to be $N = 1.6\times10^6$, 2.1×10^5 and 2.8×10^4 for each survey, respectively. For $0.8 < z < 1.5$, the numbers are $N = 2.8\times10^5$, 9.5×10^4 and 1.7×10^4, respectively. App. D details the calculation behind these numbers.

While most detectable galaxies will have polarization measurement errors making an important contribution to the variance of the estimator $\hat\alpha$, they collectively provide a non-negligible amount of information to reduce the variance in the estimation of α. Consider N detectable spiral galaxies in a small sky patch $\mathrm{d}\Omega$, each with a rotation estimator $\hat\alpha_i$. From a general linear combination $\hat\alpha_{\rm best} \equiv \sum_{i=1}^N w_i\,\hat\alpha_i$, we can obtain the weights that lead to an unbiased minimum-variance estimator, namely $w_i = \langle\hat\alpha_i^2\rangle^{-1} / \sum_j \langle\hat\alpha_j^2\rangle^{-1}$. For different galaxies, $\langle\hat\alpha_i^2\rangle$ only depends on the unpolarized radio SNR, which in turn only depends on z and $L_{4.8}$. Let us denote it as $\langle\hat\alpha^2\rangle(x)$, where $x = (z, \log L_{4.8})$.

With Eq. (18) rewritten as $N = \mathrm{d}\Omega \int \Phi(x)\,\mathrm{d}x$, the weights take the form

$$w(x) = \mathrm{d}\Omega^{-1}\,\langle\hat\alpha^2\rangle(x)^{-1}\left[\int\langle\hat\alpha^2\rangle(x')^{-1}\Phi(x')\,\mathrm{d}x'\right]^{-1}, \tag{20}$$

where $\Phi(x)$ quantifies the galaxy distribution with redshift and luminosity as defined in Eq. (C1). This leads to

$$\langle\hat\alpha^2_{\rm best}\rangle = \mathrm{d}\Omega^{-1}\left[\int\langle\hat\alpha^2\rangle(x)^{-1}\Phi(x)\,\mathrm{d}x\right]^{-1}. \tag{21}$$

Accounting for the SNR distribution is therefore equivalent to having N identically measured galaxies, each with an effective estimator variance of

$$\langle\hat\alpha^2\rangle_{\rm eff} = N\,\langle\hat\alpha^2_{\rm best}\rangle = \frac{\int\Phi(x)\,\mathrm{d}x}{\int\langle\hat\alpha(x')^2\rangle^{-1}\Phi(x')\,\mathrm{d}x'}. \tag{22}$$

We replace $\langle\hat\alpha^2\rangle$ in Eq. (17) with $\langle\hat\alpha^2\rangle_{\rm eff}$ to estimate the effective noise power spectrum

$$C_{\hat\alpha\hat\alpha}(\mathbf{L}) = \left[\int\langle\hat\alpha(x)^2\rangle^{-1}\Phi(x)\,\mathrm{d}x\right]^{-1}. \tag{23}$$

While the x-integration formally depends on the choice of the threshold intensity $S_\nu^{\min}$ as in Eq. (18), the integral approaches a minimal value as $S_\nu^{\min}$ decreases, until it is practically saturated for $S_\nu^{\min}/\sigma_I \lesssim 5$.

In Fig. 3, the noise angular power spectrum of the rotation estimator is plotted, along with the corresponding sensitivity curves for current and projected CMB surveys [20]. SKA will enable a noise power spectrum lower than that of CMB-S4 for $L \gtrsim 5$, but probe a more limited z range for the birefringence sources. The noise spectra are compared to the maximal signal allowed by Planck [20] due to a network of ultralight axion strings [13] approximated by the loop crossing model [60].

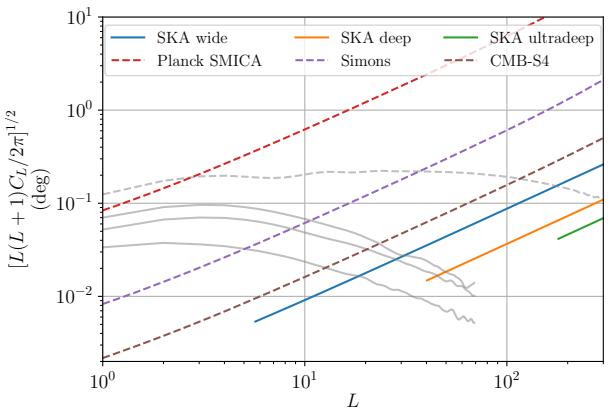


FIG. 3. Polarization rotation noise power spectra for SKA radio galaxy surveys combined with an overlapping optical imaging galaxy survey, compared to those for three CMB surveys. The low-L cutoffs of SKA curves are set by the proposed survey areas. We show signal power spectra from ultralight axion strings (see main text) up to $z = z_{\rm cmb}$ (gray dashed) and up to $z = 1.5$, 1, 0.5 (gray solid; from top to bottom).

Discussion — Using Eq. (11), noisy rotation estimations can be done at discrete galaxy locations $\{\hat\alpha_i\}_i$, which is suited for real-space approaches that exploit the non-Gaussian statistics of $\alpha(\hat{\mathbf{n}})$. This is applicable to birefringence from an axion string network [61, 62].

Measuring polarization rotation in the CMB has the advantage of being sensitive to all birefringence sources up to $z \approx 1100$. Spiral galaxies do not effectively probe birefringence sources at $z \gtrsim 1.5$. Nonetheless, galaxies present an opportunity for extracting tomographic information, *i.e.* the z-dependence of the birefringence sources. One could compare rotation estimated from the CMB with that from radio galaxies, or compare estimates from galaxies in different z bins. For example, if axion strings exist for much of the cosmic history, this would enable the measurement of their size and number evolution with z [60]. Furthermore, an axion field that is homogeneous but dynamic for $z \lesssim 1100$ is indistinguishable from a constant uniform axion field via a CMB measurement, but could show up as z-dependent in galaxy-based birefringence tomography. This also means that any detection of cosmic birefringence through the CMB [14] can then be validated through galaxy surveys. Foreground systematics that are unaccounted for, *e.g.* from the MW, would show up coherently for galaxies in every redshift bin and could thus be identified and subtracted.

Conclusion — We have proposed a new statistical technique to estimate anisotropic cosmic birefringence from correlation between radio polarization and shape in spiral galaxies. This offers an independent technique that complements CMB probes. This will be enabled by a synergy between upcoming large radio galaxy surveys such as SKA and optical imaging surveys such as LSST. The noise floor of the polarization rotation angular power spectrum of the galaxy-based method using SKA will be better than that of CMB-S4. We plan to test the empirical correlations of [24] on cosmological galaxy samples using data from existing smaller radio surveys [35, 42, 48].

ACKNOWLEDGMENTS

We thank Neal Dalal, Marc Kamionkowski, Dustin Lang, Jessie Muir and Kendrick Smith for useful discussions and comments. L.D. acknowledges research grant support from the Alfred P. Sloan Foundation (Award Number FG-2021-16495), and support of Frank and Karen Dabby STEM Fund in the Society of Hellman Fellows. S.F. is supported by Lawrence Berkeley National Laboratory and the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

[1] T. D. Lee and C.-N. Yang, Question of Parity Conservation in Weak Interactions, Phys. Rev. **104**, 254 (1956).
[2] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, Experimental Test of Parity Conservation in β Decay, Phys. Rev. **105**, 1413 (1957).
[3] M. Kobayashi and T. Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, Prog. Theor. Phys. **49**, 652 (1973).
[4] S. Weinberg, A New Light Boson?, Phys.Rev.Lett. **40**, 223 (1978).
[5] F. Wilczek, Problem of Strong p and t Invariance in the Presence of Instantons, Phys.Rev.Lett. **40**, 279 (1978).
[6] R. Peccei and H. R. Quinn, CP Conservation in the Presence of Instantons, Phys.Rev.Lett. **38**, 1440 (1977).
[7] R. Peccei and H. R. Quinn, Constraints Imposed by CP Conservation in the Presence of Instantons, Phys.Rev. **D16**, 1791 (1977).
[8] P. Svrcek and E. Witten, Axions In String Theory, JHEP **06**, 051, arXiv:hep-th/0605206.
[9] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, String Axiverse, Phys. Rev. D **81**, 123530 (2010), arXiv:0905.4720 [hep-th].
[10] S. M. Carroll, G. B. Field, and R. Jackiw, Limits on a Lorentz and Parity Violating Modification of Electrodynamics, Phys. Rev. D **41**, 1231 (1990).
[11] D. Harari and P. Sikivie, Effects of a Nambu-Goldstone boson on the polarization of radio galaxies and the cosmic microwave background, Phys. Lett. B **289**, 67 (1992).
[12] M. A. Fedderke, P. W. Graham, and S. Rajendran, Axion Dark Matter Detection with CMB Polarization, Phys. Rev. D **100**, 015040 (2019), arXiv:1903.02666 [astro-ph.CO].
[13] P. Agrawal, A. Hook, and J. Huang, A CMB Millikan experiment with cosmic axiverse strings, JHEP **07**, 138, arXiv:1912.02823 [astro-ph.CO].
[14] Y. Minami and E. Komatsu, New extraction of the cosmic birefringence from the planck 2018 polarization data, Phys. Rev. Lett. **125**, 221301 (2020).
[15] J. R. Eskilt and E. Komatsu, Improved constraints on cosmic birefringence from the WMAP and Planck cosmic microwave background polarization data, Phys. Rev. D **106**, 063503 (2022), arXiv:2205.13962 [astro-ph.CO].
[16] S. E. Clark, C.-G. Kim, J. C. Hill, and B. S. Hensley, The Origin of Parity Violation in Polarized Dust Emission and Implications for Cosmic Birefringence, Astrophys. J. **919**, 53 (2021), arXiv:2105.00120 [astro-ph.GA].
[17] V. Gluscevic, D. Hanson, M. Kamionkowski, and C. M. Hirata, First cmb constraints on direction-dependent cosmological birefringence from wmap-7, Phys. Rev. D **86**, 103529 (2012).
[18] S. Lee, G.-C. Liu, and K.-W. Ng, Cosmic Birefringence Fluctuations and Cosmic Microwave Background B-mode Polarization, Phys. Lett. B **746**, 406 (2015), arXiv:1403.5585 [astro-ph.CO].
[19] D. Contreras, P. Boubel, and D. Scott, Constraints on direction-dependent cosmic birefringence from Planck polarization data, JCAP **12**, 046, arXiv:1705.06387 [astro-ph.CO].
[20] W. W. Yin, L. Dai, and S. Ferraro, Probing cosmic strings by reconstructing polarization rotation of the cosmic microwave background, JCAP **06** (06), 033, arXiv:2111.12741 [astro-ph.CO].

[21] M. Kamionkowski, How to derotate the cosmic microwave background polarization, Phys. Rev. Lett. **102**, 111302 (2009).

[22] A. P. S. Yadav, R. Biswas, M. Su, and M. Zaldarriaga, Constraining a spatially dependent rotation of the cosmic microwave background polarization, Phys. Rev. D **79**, 123009 (2009).

[23] V. Gluscevic, M. Kamionkowski, and A. Cooray, Derotation of the cosmic microwave background polarization: Full-sky formalism, Phys. Rev. D **80**, 023510 (2009).

[24] J. M. Stil, M. Krause, R. Beck, and A. R. Taylor, The Integrated Polarization of Spiral Galaxy Disks, Astrophys. J. **693**, 1392 (2009), arXiv:0810.2303 [astro-ph].

[25] M. Gatti, E. Sheldon, A. Amon, M. Becker, M. Troxel, A. Choi, C. Doux, N. MacCrann, A. Navarro-Alsina, I. Harrison, D. Gruen, G. Bernstein, M. Jarvis, L. F. Secco, A. Ferté, T. Shin, J. McCullough, R. P. Rollins, R. Chen, C. Chang, S. Pandey, I. Tutusaus, J. Prat, J. Elvin-Poole, C. Sanchez, A. A. Plazas, A. Roodman, J. Zuntz, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Bacon, E. Bertin, S. Bhargava, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, C. Conselice, M. Costanzi, M. Crocce, L. N. da Costa, T. M. Davis, J. De Vicente, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, K. Eckert, S. Everett, I. Ferrero, J. Frieman, J. García-Bellido, D. W. Gerdes, T. Giannantonio, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, D. J. James, T. Jeltema, E. Krause, R. Kron, N. Kuropatkin, M. Lima, M. A. G. Maia, J. L. Marshall, R. Miquel, R. Morgan, J. Myles, A. Palmese, F. Paz-Chinchón, E. S. Rykoff, S. Samuroff, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, D. L. Tucker, T. N. Varga, R. H. Wechsler, J. Weller, W. Wester, and R. D. Wilkinson, Dark energy survey year 3 results: weak lensing shape catalogue, Monthly Notices of the Royal Astronomical Society **504**, 4312 (2021), arXiv:2011.03408 [astro-ph.CO].

[26] B. Giblin, C. Heymans, M. Asgari, H. Hildebrandt, H. Hoekstra, B. Joachimi, A. Kannawadi, K. Kuijken, C.-A. Lin, L. Miller, T. Tröster, J. L. van den Busch, A. H. Wright, M. Bilicki, C. Blake, J. de Jong, A. Dvornik, T. Erben, F. Getman, N. R. Napolitano, P. Schneider, H. Shan, and E. Valentijn, KiDS-1000 catalogue: Weak gravitational lensing shear measurements, Astronomy & Astrophysics **645**, A105 (2021), arXiv:2007.01845 [astro-ph.CO].

[27] X. Li, H. Miyatake, W. Luo, S. More, M. Oguri, T. Hamana, R. Mandelbaum, M. Shirasaki, M. Takada, R. Armstrong, A. Kannawadi, S. Takita, S. Miyazaki, A. J. Nishizawa, A. A. Plazas Malagon, M. A. Strauss, M. Tanaka, and N. Yoshida, The three-year shear catalog of the Subaru Hyper Suprime-Cam SSP Survey, Publications of the Astronomical Society of Japan **74**, 421 (2022), arXiv:2107.00136 [astro-ph.CO].

[28] A. Dey *et al.* (DESI), Overview of the DESI Legacy Imaging Surveys, Astron. J. **157**, 168 (2019), arXiv:1804.08657 [astro-ph.IM].

[29] C. Chang, M. Jarvis, B. Jain, S. M. Kahn, D. Kirkby, A. Connolly, S. Krughoff, E. Peng, and J. R. Peterson, The Effective Number Density of Galaxies for Weak Lensing Measurements in the LSST Project, Mon. Not. Roy. Astron. Soc. **434**, 2121 (2013), arXiv:1305.0793 [astro-ph.CO].

[30] M. L. Brown and R. A. Battye, Polarization as an indicator of intrinsic alignment in radio weak lensing, Monthly Notices of the Royal Astronomical Society **410**, 2057 (2011), arXiv:1005.1926 [astro-ph.CO].

[31] M. L. Brown and R. A. Battye, Mapping the Dark Matter with Polarized Radio Surveys, The Astrophysical Journal Letters **735**, L23 (2011), arXiv:1101.5157 [astro-ph.CO].

[32] L. Whittaker, M. L. Brown, and R. A. Battye, Separating weak lensing and intrinsic alignments using radio observations, Mon. Not. Roy. Astron. Soc. **451**, 383 (2015), arXiv:1503.00061 [astro-ph.CO].

[33] M. Kamionkowski, Non-Uniform Cosmological Birefringence and Active Galactic Nuclei, Phys. Rev. D **82**, 047302 (2010), arXiv:1004.3544 [astro-ph.CO].

[34] L. Whittaker, R. A. Battye, and M. L. Brown, Measuring cosmic shear and birefringence using resolved radio sources, Mon. Not. Roy. Astron. Soc. **474**, 460 (2018), arXiv:1702.01700 [astro-ph.CO].

[35] M. Lacy *et al.*, The Karl G. Jansky Very Large Array Sky Survey (VLASS). Science case and survey design, Publ. Astron. Soc. Pac. **132**, 035001 (2020), arXiv:1907.01981 [astro-ph.IM].

[36] M. Jarvis, D. Bacon, C. Blake, M. Brown, S. Lindsay, A. Raccanelli, M. Santos, and D. J. Schwarz, Cosmology with SKA Radio Continuum Surveys, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (2015) p. 18, arXiv:1501.03825 [astro-ph.CO].

[37] In radio astronomy, I is the spectral flux density typically measured in units of jansky (Jy).

[38] M. Kilbinger, Cosmology with cosmic shear observations: a review, Rept. Prog. Phys. **78**, 086901 (2015), arXiv:1411.0115 [astro-ph.CO].

[39] S. M. Carroll, Quintessence and the rest of the world, Phys. Rev. Lett. **81**, 3067 (1998), arXiv:astro-ph/9806099.

[40] M. Pospelov, A. Ritz, C. Skordis, A. Ritz, and C. Skordis, Pseudoscalar perturbations and polarization of the cosmic microwave background, Phys. Rev. Lett. **103**, 051302 (2009), arXiv:0808.0673 [astro-ph].

[41] M. Li and X. Zhang, Cosmological CPT violating effect on CMB polarization, Phys. Rev. D **78**, 103516 (2008), arXiv:0810.0403 [astro-ph].

[42] A. R. Taylor, S. Sekhar, L. Heino, A. M. M. Scaife, J. Stil, M. Bowles, M. Jarvis, I. Heywood, and J. D. Collier, MIGHTEE polarization early science fields: the deep polarized sky, Monthly Notices of the Royal Astronomical Society **528**, 2511 (2024), arXiv:2312.13230 [astro-ph.GA].

[43] W. Hu and T. Okamoto, Mass reconstruction with cosmic microwave background polarization, The Astrophysical Journal **574**, 566 (2002).

[44] T. Namikawa, Testing parity-violating physics from cosmic rotation power reconstruction, Phys. Rev. D **95**, 043523 (2017), arXiv:1612.07855 [astro-ph.CO].

[45] C. M. Hirata and U. Seljak, Analyzing weak lensing of the cosmic microwave background using the likelihood function, Phys. Rev. D **67**, 043001 (2003), arXiv:astro-ph/0209489.

[46] E. Krause and C. M. Hirata, Weak lensing power spectra for precision cosmology: Multiple-deflection, reduced

shear and lensing bias corrections, Astron. Astrophys. **523**, A28 (2010), arXiv:0910.3786 [astro-ph.CO].

[47] M. Wenger, F. Ochsenbein, D. Egret, P. Dubois, F. Bonnarel, S. Borde, F. Genova, G. Jasniewicz, S. Laloë, S. Lesteven, and R. Monier, The simbad astronomical database: The cds reference database for astronomical objects, Astronomy and Astrophysics Supplement Series **143**, 9–22 (2000).

[48] K. Knowles, W. D. Cotton, L. Rudnick, F. Camilo, S. Goedhart, R. Deane, M. Ramatsoku, M. F. Bietenholz, M. Brüggen, C. Button, H. Chen, J. O. Chibueze, T. E. Clarke, F. de Gasperin, R. Ianjamasimanana, G. I. G. Józsa, M. Hilton, K. C. Kesebonye, K. Kolokythas, R. C. Kraan-Korteweg, G. Lawrie, M. Lochner, S. I. Loubser, P. Marchegiani, N. Mhlahlo, K. Moodley, E. Murphy, B. Namumba, N. Oozeer, V. Parekh, D. S. Pillay, S. S. Passmoor, A. J. T. Ramaila, S. Ranchod, E. Retana-Montenegro, L. Sebokolodi, S. P. Sikhosana, O. Smirnov, K. Thorat, T. Venturi, T. D. Abbott, R. M. Adam, G. Adams, M. A. Aldera, E. F. Bauermeister, T. G. H. Bennett, W. A. Bode, D. H. Botha, A. G. Botha, L. R. S. Brederode, S. Buchner, J. P. Burger, T. Cheetham, D. I. L. de Villiers, M. A. Dikgale-Mahlakoana, L. J. du Toit, S. W. P. Esterhuyse, G. Fadana, B. L. Fanaroff, S. Fataar, A. R. Foley, D. J. Fourie, B. S. Frank, R. R. G. Gamatham, T. G. Gatsi, M. Geyer, M. Gouws, S. C. Gumede, I. Heywood, M. J. Hlakola, A. Hokwana, S. W. Hoosen, D. M. Horn, J. M. G. Horrell, B. V. Hugo, A. R. Isaacson, J. L. Jonas, J. D. B. Jordaan, A. F. Joubert, R. P. M. Julie, F. B. Kapp, V. A. Kasper, J. S. Kenyon, P. P. A. Kotzé, A. G. Kotze, N. Kriek, H. Kriel, V. K. Krishnan, T. W. Kusel, L. S. Legodi, R. Lehmensiek, D. Liebenberg, R. T. Lord, B. M. Lunsky, K. Madisa, L. G. Magnus, J. P. L. Main, A. Makhaba, S. Makhathini, J. A. Malan, J. R. Manley, S. J. Marais, M. D. J. Maree, A. Martens, T. Mauch, K. McAlpine, B. C. Merry, R. P. Millenaar, O. J. Mokone, T. E. Monama, M. C. Mphego, W. S. New, B. Ngcebetsha, K. J. Ngoasheng, M. T. Ockards, A. J. Otto, A. A. Patel, A. Peens-Hough, S. J. Perkins, N. M. Ramanujam, Z. R. Ramudzuli, S. M. Ratcliffe, R. Renil, A. Robyntjies, A. N. Rust, S. Salie, N. Sambu, C. T. G. Schollar, L. C. Schwardt, R. L. Schwartz, M. Serylak, R. Siebrits, S. K. Sirothia, M. Slabber, L. Sofeya, B. Taljaard, C. Tasse, A. J. Tiplady, O. Toruvanda, S. N. Twum, T. J. van Balla, A. van der Byl, C. van der Merwe, C. L. van Dyk, V. Van Tonder, R. Van Wyk, A. J. Venter, M. Venter, M. G. Welz, L. P. Williams, and B. Xaia, The MeerKAT Galaxy Cluster Legacy Survey. I. Survey Overview and Highlights, Astronomy & Astrophysics **657**, A56 (2022), arXiv:2111.05673 [astro-ph.GA].

[49] A. Bonaldi, I. Harrison, S. Camera, and M. L. Brown, Ska weak lensing– ii. simulated performance and survey design considerations, Monthly Notices of the Royal Astronomical Society **463**, 3686–3698 (2016).

[50] M. Negrello, F. Perrotta, J. G.-N. Gonzalez, L. Silva, G. De Zotti, G. L. Granato, C. Baccigalupi, and L. Danese, Astrophysical and cosmological information from large-scale submillimetre surveys of extragalactic sources, Monthly Notices of the Royal Astronomical Society **377**, 1557–1568 (2007).

[51] E. M. Sadler, C. A. Jackson, R. D. Cannon, V. J. McIntyre, T. Murphy, J. Bland-Hawthorn, T. Bridges, S. Cole, M. Colless, C. Collins, W. Couch, G. Dalton, R. de Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, P. Norberg, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor, Radio sources in the 2dF Galaxy Redshift Survey – II. Local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz, Monthly Notices of the Royal Astronomical Society **329**, 227 (2002), https://academic.oup.com/mnras/article-pdf/329/1/227/3879739/329-1-227.pdf.

[52] J. J. Condon, W. D. Cotton, and J. J. Broderick, Radio sources and star formation in the local universe, The Astronomical Journal **124**, 675 (2002).

[53] T. T. Takeuchi, K. Yoshikawa, and T. T. Ishii, The luminosity function of iras point source catalog redshift survey galaxies, The Astrophysical Journal **587**, L89 (2003).

[54] T. T. Takeuchi, K. Yoshikawa, and T. T. Ishii, Erratum: "the luminosity function of iras point source catalog redshift survey galaxies" (apj, 587, l89 [2003]), The Astrophysical Journal **606**, L171 (2004).

[55] M. S. Yun, N. A. Reddy, and J. J. Condon, Radio properties of infrared-selected galaxies in theiras2 jy sample, The Astrophysical Journal **554**, 803–822 (2001).

[56] G. De Zotti, M. Massardi, M. Negrello, and J. Wall, Radio and millimeter continuum surveys and their astrophysical implications, The Astronomy and Astrophysics Review **18**, 1–65 (2009).

[57] G. L. Granato, G. D. Zotti, L. Silva, A. Bressan, and L. Danese, A physical model for the coevolution of qsos and their spheroidal hosts, The Astrophysical Journal **600**, 580 (2004).

[58] R. T. Coogan, M. T. Sargent, A. Cibinel, I. Prandoni, A. Bonaldi, E. Daddi, and M. Franco, Looking ahead to the sky with the square kilometre array: simulating flux densities & resolved radio morphologies of $0 < z < 2.5$ star-forming galaxies (2023), arXiv:2307.06358 [astro-ph.GA].

[59] R. Braun, A. Bonaldi, T. Bourke, E. Keane, and J. Wagg, Anticipated Performance of the Square Kilometre Array – Phase 1 (SKA1), arXiv e-prints , arXiv:1912.12699 (2019), arXiv:1912.12699 [astro-ph.IM].

[60] M. Jain, A. J. Long, and M. A. Amin, Cmb birefringence from ultralight-axion string networks, Journal of Cosmology and Astroparticle Physics **2021** (05), 055.

[61] W. W. Yin, L. Dai, and S. Ferraro, Testing charge quantization with axion string-induced cosmic birefringence, JCAP **07**, 052, arXiv:2305.02318 [astro-ph.CO].

[62] R. Hagimoto and A. J. Long, Measures of non-gaussianity in axion-string-induced cmb birefringence (2023), arXiv:2306.07351 [astro-ph.CO].

[63] L. Silva, G. De Zotti, G. L. Granato, R. Maiolino, and L. Danese, Observational tests of the evolution of spheroidal galaxies, Mon. Not. Roy. Astron. Soc. **357**, 1295 (2005), arXiv:astro-ph/0412340.

[64] G. Helou, B. T. Soifer, and M. Rowan-Robinson, Thermal infrared and nonthermal radio - Remarkable correlation in disks of galaxies, Astrophys. J. Lett. **298**, L7 (1985).

[65] L. J. Kewley, M. J. Geller, R. A. Jansen, and M. A. Dopita, The hα and infrared star formation rates for the nearby field galaxy survey, The Astronomical Journal **124**, 3135 (2002).

[66] J. J. Condon, Radio emission from normal galaxies., Annual Rev. Astron. Astrophys. **30**, 575 (1992).

[67] Garrett, M. A., The fir/radio correlation of high redshift galaxies in the region of the hdf-n, A& A **384**, L19 (2002).

Appendix A: Quadratic estimator

Polarization and shape measurements have errors. For example, the polarized flux has a squared error

$$\sigma_P^2 = \frac{\sigma_I^2}{2}, \tag{A1}$$

where σ_I is the RMS flux of the detector in units of Jansky. We assume that the effect of measurement errors on the variance of the quadratic estimator $\hat{\alpha}$ is dominated by the measurement error for polarization, and the shape error is much smaller. In this case,

$$\tilde{\sigma}_{X,Y}^2 = \sigma_{X,Y}^2 + \frac{\sigma_I^2\,\sigma_\epsilon^2}{2\,I^2} = \sigma_{X,Y}^2 + \frac{\sigma_\epsilon^2}{2\,\mathrm{SNR}^2}, \tag{A2}$$

where the total flux SNR enters the expression.

We now calculate the propagation of this error into the variance of the quadratic estimator. Let $\hat{\alpha}$ for each galaxy be constructed from the observed quantities (with tilde) including measurement errors, up to quadratic combinations:

$$\hat{\alpha} \equiv u\,\tilde{X} + v\,\tilde{Y} + u'\,\tilde{X}^2 + v'\,\tilde{Y}^2 + w\,\tilde{X}\,\tilde{Y}, \tag{A3}$$

where the coefficients are to be determined. Under a polarization rotation α at the galaxy's location, the observed quantities are transformed:

$$\tilde{X} \to \tilde{X} - 2\alpha\,Y, \quad \tilde{Y} \to \tilde{Y} + 2\alpha\,X. \tag{A4}$$

Note that the transformation is proportional to quantities without measurement noises, since cosmic birefringence is applied before detection. By requiring that the estimator $\hat{\alpha}$ be unbiased, namely $\langle\hat{\alpha}\rangle = \alpha$, we have

$$0 = u\,\langle X\rangle + u'\,\langle X^2\rangle + v'\,\langle Y^2\rangle \tag{A5}$$

$$\frac{1}{2} = v\,\langle X\rangle + w\,\langle X^2\rangle - w\,\langle Y^2\rangle\,. \tag{A6}$$

Using these, u can be expressed in terms of u' and v', and v can be expressed in terms of w. It remains to minimize the variance of the estimator, $\langle\hat{\alpha}^2\rangle$, under the null hypothesis, with respect to u', v', and w. After some algebra, we obtain the claimed result in Eq. (11):

$$u = u' = v' = 0, \quad v = \frac{\tilde{\sigma}_Y^2\,\overline{X}}{2\left[(\sigma_X^2 - \sigma_Y^2)^2 + \tilde{\sigma}_X^2\,\overline{X}^2\right]}, \quad w = \frac{\sigma_X^2 - \sigma_Y^2}{2\left[(\sigma_X^2 - \sigma_Y^2)^2 + \tilde{\sigma}_X^2\,\overline{X}^2\right]}. \tag{A7}$$

Appendix B: Maximum likelihood approach

Inspired by the idea originally put forth in Ref. [45] and closely following the formalism presented in Ref. [20], we pursue an maximum-likelihood approach to estimating the polarization rotation field $\alpha(\hat{\mathbf{n}})$. For detailed justifications of some assumptions we are about to make, we refer to Sections 3 and 4.1 of Ref. [20].

Let $\mathbf{X}_i \equiv \begin{pmatrix}\tilde{X}_i & \tilde{Y}_i\end{pmatrix}^T$ denote the observed quantities including measurement error for the i-th galaxy, located in the direction $\hat{\mathbf{n}}_i$, among a total of N galaxies. We start by assuming the following log-likelihood function:

$$\ln\mathcal{L}[\alpha;\{\mathbf{X}_i,\mathbf{N}_i\}] = -\frac{1}{2}\sum_{i=1}^{N}\left\{\left[M(\alpha(\hat{\mathbf{n}}_i))^{-1}(\mathbf{X}_i-\mathbf{N}_i)-\overline{\mathbf{X}}_i\right]^T C_{\mathbf{X}_i}^{-1}\left[M(\alpha(\hat{\mathbf{n}}_i))^{-1}(\mathbf{X}_i-\mathbf{N}_i)-\overline{\mathbf{X}}_i\right] + \mathbf{N}_i^T C_{\mathbf{N}_i}^{-1}\mathbf{N}_i\right\}, \tag{B1}$$

where $\mathbf{N}_i$ is the measurement error of $\mathbf{X}_i$; $\overline{\mathbf{X}}_i$ is the mean of $\mathbf{X}$ over an ensemble of galaxies; $C_{\mathbf{X}_i}$ is the covariance matrix of the intrinsic, unrotated quantities $\mathbf{X}_i$ (which in general may depend on the type or other properties of the i-th galaxy); $C_{\mathbf{N}_i}$ is the covariance matrix of the measurement noise given by the second term in Eq. (10); and $M(\alpha(\hat{\mathbf{n}}_i))$ is the effect of polarization rotation on the intrinsic quantities. Eq. (B1) assumes that each $\mathbf{X}_i$ is a random

Gaussian variable, the same approximation we used in the main text, meaning $C_{\mathbf{X}_i} = \mathrm{diag}(\tilde{\sigma}^2_{X_i}, \tilde{\sigma}^2_{Y_i})$. Our goal is then to obtain an estimator $\hat{\alpha}_{\mathrm{ML}}(\hat{\mathbf{n}})$ which maximizes $\ln \mathcal{L}$ given a set of observations $\{\mathbf{X}_i\}$.

Since the sum in Eq. (B1) has completely independent terms, it suffices to individually maximize each term:

$$\ln \mathcal{L}_i[\alpha(\hat{\mathbf{n}}_i); \mathbf{X}_i, \mathbf{N}_i] = -\frac{1}{2}\left[M(\alpha(\hat{\mathbf{n}}_i))^{-1}(\mathbf{X}_i - \mathbf{N}_i) - \overline{\mathbf{X}}_i\right]^T C^{-1}_{\mathbf{X}_i}\left[M(\alpha(\hat{\mathbf{n}}_i))^{-1}(\mathbf{X}_i - \mathbf{N}_i) - \overline{\mathbf{X}}_i\right] - \frac{1}{2}\mathbf{N}_i^T C^{-1}_{\mathbf{N}_i}\mathbf{N}_i. \quad \text{(B2)}$$

From now on, we will omit the galaxy index i and suppress the $\hat{\mathbf{n}}_i$ dependence for clarity. For any fixed $\mathbf{X}$, the likelihood is maximized when the noise is

$$\mathbf{N}_{\max} = (\mathbf{M}^{-T} C^{-1}_{\mathbf{X}} \mathbf{M}^{-1} + C^{-1}_{\mathbf{N}})^{-1} \mathbf{M}^{-T} C^{-1}_{\mathbf{X}} (\mathbf{M}^{-1}\mathbf{X} - \overline{\mathbf{X}}). \quad \text{(B3)}$$

The value of the likelihood at this noise realisation is

$$\ln \mathcal{L}[\alpha; \mathbf{X}, \mathbf{N} = \mathbf{N}_{\max}] = -\frac{1}{2}(\mathbf{X} - \mathbf{M}^{-1}\overline{\mathbf{X}})^{-1}(C_{\mathbf{N}} + \mathbf{M}^T C_{\mathbf{X}} \mathbf{M})^{-1}(\mathbf{X} - \mathbf{M}^{-1}\overline{\mathbf{X}}). \quad \text{(B4)}$$

It would have been more justifiable to marginalize over the noise $\mathbf{N}$, which would produce an additional logarithmic term that is numerically negligible.

The matrix $M(\alpha)$ depends on the angle α non-linearly. Using Eqs. (7)–(8), we expand it up to $O(\alpha^2)$:

$$M(\alpha) = e^{2\,\alpha\, T} = 1 + 2\,\alpha\, T + 2\,\alpha^2\, T^2, \quad T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad \text{(B5)}$$

We insert this expansion into Eq. (B4) and take the derivative with respect to α, keeping terms up to linear order:

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{2}{\tilde{\sigma}^2_X\, \tilde{\sigma}^2_Y}\left[\tilde{\sigma}^2_Y\, \overline{X}\, \tilde{Y} + (\sigma^2_X - \sigma^2_Y)\, \tilde{X}\, \tilde{Y}\right] - \frac{4\,\alpha}{\tilde{\sigma}^4_X\, \tilde{\sigma}^4_Y}\left[(\cdots) + (\cdots)\tilde{X} + (\cdots)\tilde{X}^2 + (\cdots)\tilde{Y}^2\right], \quad \text{(B6)}$$

where $(\cdots)$ are combinations of statistical quantities whose exact forms are not enlightening. This results in the maximum-likelihood estimator

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\tilde{\sigma}^2_X\, \tilde{\sigma}^2_Y\left[\tilde{\sigma}^2_Y\, \overline{X}\, \tilde{Y} + (\sigma^2_X - \sigma^2_Y)\, \tilde{X}\, \tilde{Y}\right]}{2\left[(\cdots) + (\cdots)\tilde{X} + (\cdots)\tilde{X}^2 + (\cdots)\tilde{Y}^2\right]}. \quad \text{(B7)}$$

We then introduce the approximation by replacing the denominator of the estimator with its ensemble average (expectation value):

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\tilde{\sigma}^2_X\, \tilde{\sigma}^2_Y\left[\tilde{\sigma}^2_Y\, \overline{X}\tilde{Y} + (\sigma^2_X - \sigma^2_Y)\, \tilde{X}\, \tilde{Y}\right]}{2\left[(\cdots) + (\cdots)\overline{X} + (\cdots)(\tilde{\sigma}^2_X + \overline{X}^2) + (\cdots)\tilde{\sigma}^2_Y\right]}. \quad \text{(B8)}$$

After some simplification, the quadratic estimator in Eq. (11) is recovered:

$$\hat{\alpha}_{\mathrm{ML}} = \frac{\tilde{\sigma}^2_Y\, \overline{X}\, \tilde{Y} + (\sigma^2_X - \sigma^2_Y)\, \tilde{X}\, \tilde{Y}}{2\left[(\sigma^2_X - \sigma^2_Y)^2 + \tilde{\sigma}^2_X\, \overline{X}^2\right]}, \quad \text{(B9)}$$

which agrees with the quadratic estimator Eq. (11). When $\alpha = 0$, the variance $\langle \hat{\alpha}^2_{\mathrm{ML}} \rangle$ is therefore also given by Eq. (12).

Appendix C: Effective estimator variance

Eq. (12) and Eq. (10) tell us the dependence of the estimator error $\sqrt{\langle \hat{\alpha}^2 \rangle}$ on the RMS flux σ_I of a given radio survey. For a cosmological sample of galaxies, we can use the redshift-dependent radio luminosity function of galaxies to obtain a distribution of values for $\langle \hat{\alpha}^2 \rangle$ when measurement errors are accounted for.

We wish to include as many galaxies as possible in our estimation of α, even if many may have low SNR for the polarized flux. We now calculate the lowest estimator variance achievable by including all galaxies observable by an experiment with a given σ_I.

Suppose we are estimating the birefringence α within a small solid angle $\mathrm{d}\Omega$ around a direction $\hat{\mathbf{n}}$. The galaxies located within this patch of sky follow the redshift-dependent radio luminosity function. That is, in a redshift bin from z to $z + \mathrm{d}z$ and a luminosity bin from $\log L$ to $\log L + \mathrm{d}\log L$, the number of galaxies in a solid angle $\mathrm{d}\Omega$ is

$$\mathrm{d}N = \Phi(z, \log L)\, \mathrm{d}z\, \mathrm{d}\log L\, \mathrm{d}\Omega. \tag{C1}$$

Each of these galaxies corresponds to an independent estimator $\hat{\alpha}_i$ whose variance $\langle \hat{\alpha}(p)^2 \rangle$ only depends on $p = (z, \log L)$ of the galaxy. We define a new estimator as a linear combination of them:

$$\hat{\alpha}_{\mathrm{best}} \equiv \sum_i w_i\, \hat{\alpha}_i. \tag{C2}$$

The weights are required to sum to unity:

$$1 = \sum_i w_i. \tag{C3}$$

We assume that the weight $w_i = w(p)$ only depends on p. In the continuum limit, we have

$$1 = \mathrm{d}\Omega \int w(p)\, \Phi(p)\, \mathrm{d}p. \tag{C4}$$

Under this constraint, we choose weights that minimise the variance of $\hat{\alpha}_{\mathrm{best}}$:

$$\langle \hat{\alpha}^2_{\mathrm{best}} \rangle = \sum_{ij} w_i w_j \langle \hat{\alpha}_i \hat{\alpha}_j \rangle = \sum_i w_i^2 \langle \hat{\alpha}_i^2 \rangle. \tag{C5}$$

In the continuum limit, this can be written as

$$\langle \hat{\alpha}^2_{\mathrm{best}} \rangle = \mathrm{d}\Omega \int w(p)^2 \langle \hat{\alpha}(p)^2 \rangle\, \Phi(p)\, \mathrm{d}p. \tag{C6}$$

A standard application of the Lagrange multiplier method leads to

$$w(p) = \mathrm{d}\Omega^{-1} \left[\int \langle \hat{\alpha}(p')^2 \rangle^{-1}\, \Phi(p')\, \mathrm{d}p' \right]^{-1} \langle \hat{\alpha}(p)^2 \rangle^{-1}. \tag{C7}$$

The minimum variance is

$$\langle \hat{\alpha}^2_{\mathrm{best}} \rangle = \mathrm{d}\Omega^{-1} \left[\int \langle \hat{\alpha}(p')^2 \rangle^{-1}\, \Phi(p')\, \mathrm{d}p' \right]^{-1}. \tag{C8}$$

This is the best estimator variance achieved from a total of N galaxies following the luminosity function, where

$$N = \mathrm{d}\Omega \int \Phi(p)\, \mathrm{d}p. \tag{C9}$$

If there were instead N galaxies with the same estimator variance $\langle \hat{\alpha}^2 \rangle_{\mathrm{eff}}$, the variance of the arithmetic mean of their corresponding estimators would be $\langle \hat{\alpha}^2 \rangle_{\mathrm{eff}} / N$. What would $\langle \hat{\alpha}^2 \rangle_{\mathrm{eff}}$ have to be so that the arithmetic mean would produce the same variance as $\hat{\alpha}_{\mathrm{best}}$? The answer is

$$\langle \hat{\alpha}^2 \rangle_{\mathrm{eff}} = N\, \langle \hat{\alpha}^2_{\mathrm{best}} \rangle = \frac{\int \Phi(p)\, \mathrm{d}p}{\int \langle \hat{\alpha}(p)^2 \rangle^{-1}\, \Phi(p)\, \mathrm{d}p}, \tag{C10}$$

which is correctly independent of the solid angle $\mathrm{d}\Omega$.

We can use this $\langle \hat{\alpha}^2 \rangle_{\mathrm{eff}}$ to estimate the noise floor of the power spectrum:

$$C_{\hat{\alpha}\hat{\alpha}} = \frac{\Omega}{N_{\mathrm{tot}}}\, \langle \hat{\alpha}^2 \rangle_{\mathrm{eff}} = \frac{\Omega}{N_{\mathrm{tot}}} \frac{\int \Phi(p)\, \mathrm{d}p}{\int \langle \hat{\alpha}(p)^2 \rangle^{-1}\, \Phi(p)\, \mathrm{d}p} = \left[\int \langle \hat{\alpha}(p)^2 \rangle^{-1}\, \Phi(p)\, \mathrm{d}p \right]^{-1}, \tag{C11}$$

which is independent of the sky area surveyed.

Appendix D: Radio luminosity function

In this Appendix, we outline our procedure for estimating the source count of spiral galaxies for radio continuum surveys, thus providing the calculation details underlying Eq. (18).

The number of galaxies observable within a solid angle $\mathrm{d}\Omega$ of the sky, in a redshift interval between z and $z + \mathrm{d}z$ and in a logarithmic interval of intrinsic spectral luminosity between $\log L$ and $\log L + \mathrm{d}\log L$, can be written as

$$\mathrm{d}N = \frac{c\, s^2(z)}{H(z)}\, \phi(z, \log L)\, \mathrm{d}\Omega\, \mathrm{d}z\, \mathrm{d}\log L, \tag{D1}$$

where the luminosity function $\phi(z, \log L)$ gives the number density (per unit comoving volume) of galaxies per decade of luminosity at redshift z, and $c\, s(z)^2/H(z)\, \mathrm{d}\Omega\, \mathrm{d}z$ is the comoving volume corresponding to the solid angle $\mathrm{d}\Omega$ and the redshift slice from z to $z + \mathrm{d}z$. We need to know the functional form of ϕ for radio-emitting spiral galaxies.

The radio galaxy luminosity function ϕ at $z > 0$ can be related to that at $z = 0$:

$$\phi(z, \log L) = \phi(0, \log[\lambda(z) L]), \tag{D2}$$

where the factor $\lambda(z)$ accounts for the redshift evolution of spiral galaxies. For this, we adopt the simple formula from [63] that fits observed counts (also see [50]):

$$\lambda(z) = \begin{cases} (1+z)^{-1.5}, & 0 \le z < 1 \\ 2^{-1.5}, & 1 \le z < 1.5 \end{cases}. \tag{D3}$$

The number density of spiral galaxies appear to drop quickly above $z = 1.5$ [57], so we set $\phi(z, \log L) = 0$ for $z \ge 1.5$.

Suppose the radio luminosity function $\phi_\nu(z, \log L)$ at some frequency ν is known. The function at a different frequency ν' can be derived as:

$$\phi_{\nu'}(z, \log L) = \phi_\nu(z, \log[(\nu'/\nu)^{-\alpha} L]), \tag{D4}$$

by exploiting the power-law frequency dependence of the spectral flux density with respect to frequency: $S_\nu \propto \nu^\alpha$, where the spectral index has an appropriate value $\alpha = -0.7$ [51, 52]. While [24] reported the empirical polarization-shape correlation at 4.8 GHz, we would like to relate sources at 4.8 GHz to sources at $\nu = 1.4$ GHz in order to make contact with observed counts.

For local star-forming galaxies, it is known that the radio and far infrared (FIR) luminosities are nearly linearly related [64–67]. Consider a general mean relation between the FIR luminosity and the radio luminosity at $z = 0$:

$$L_{1.4} = f(L_{\mathrm{FIR}}). \tag{D5}$$

The corresponding luminosity functions at $z = 0$ are related by

$$\phi_{1.4}(0, \log L_{1.4}) = \frac{L_{1.4}}{f^{-1}(L_{1.4})\, f'(f^{-1}(L_{1.4}))}\, \phi_{\mathrm{FIR}}(f^{-1}(\log L_{1.4})), \tag{D6}$$

where f^{-1} is the inverse function to f. This reduces to a simple shift in the logarithmic luminosity when f is linear.

The FIR luminosity function is given by a phenomenological Schechter model [53]:

$$\phi_{\mathrm{FIR}}(\log L_{\mathrm{FIR}}) = \phi_* \left(\frac{L_{\mathrm{FIR}}}{L_*}\right)^{1-a} \exp\left\{-\frac{1}{2\,\sigma^2}\left[\log\left(1 + \frac{L_{\mathrm{FIR}}}{L_*}\right)\right]^2\right\}. \tag{D7}$$

The parameter values are taken from Eq. (6) in [53], which correspond to the "cool" galaxies, which are interpreted as star-forming spiral galaxies. Corrections from [54] are applied, assuming $h = 0.7$:

$$\phi_* = 0.635\,\mathrm{Mpc}^{-3}, \quad a = 1.25, \quad L_* = 1.95 \times 10^9\, L_\odot, \quad \sigma = 0.5. \tag{D8}$$

Following Eq. (16) in [56], we find that the following fitting function for Eq. (D5) in combination with the FIR luminosity function above reproduces the appropriate curve in Fig. 8 of [56]:

$$L_{1.4} = A \left[\left(\frac{L_{\mathrm{FIR}}}{L_b}\right)^{-3.1} + \left(\frac{L_{\mathrm{FIR}}}{L_b}\right)^{-1}\right]^{-1}, \tag{D9}$$

where the fitting parameters are given as

$$A = 7.66 \times 10^{20}\,\mathrm{W\,Hz^{-1}}, \quad L_b = 3.34 \times 10^{35}\,\mathrm{W}. \tag{D10}$$

Having found the redshift-dependent radio luminosity function $\phi_\nu(z, \log L)$ for any radio frequency ν, we can now compute the total source count within a sky patch by integrating Eq. (D1):

$$N = \Omega \int \mathrm{d}z \int_{L_{\min}}^{L_{\max}} \mathrm{d}\log L \, \frac{c\, s(z)^2}{H(z)} \, \phi_{\nu(z)}(z, \log L), \tag{D11}$$

where $\nu(z) = \nu_0\,(1+z)$ is the intrinsic frequency at the location of a galaxy at redshift z corresponding to an observed frequency ν_0. If the radio continuum survey has an RMS flux of $S_\nu^{\min}$, then the luminosity cutoff is given by

$$L_{\min} = 4\pi\, s(z)^2\,(1+z)\, S_\nu^{\min}. \tag{D12}$$

Writing $\phi(\log L) \equiv \phi_{\nu_0}(0, \log L)$ and performing a change of variables on $\log L$ from 1.4 GHz to 4.8 GHz, we derive Eq. (18) in the main text.

Prepared for submission to JHEP

Three-leg form factor on Coulomb branch

A.V. Belitsky,[a] **L.V. Bork,**[b] **J.M. Grumski-Flores,**[a] **V.A. Smirnov**[c]

[a] *Department of Physics, Arizona State University, Tempe, AZ 85287-1504, USA*

[b] *Institute for Theoretical and Experimental Physics, 117218 Moscow, Russia*
The Center for Fundamental and Applied Research, 127030 Moscow, Russia

[c] *Skobeltsyn Institute of Nuclear Physics, Moscow State University 119992 Moscow, Russia*
Moscow Center for Fundamental and Applied Mathematics 119992 Moscow, Russia

arXiv:2402.18475v1 [hep-th] 28 Feb 2024

Abstract: We study the form factor of the lowest component of the stress-tensor multiplet away from the origin of the moduli space in the spontaneously broken, aka Coulomb, phase of the maximally supersymmetric Yang-Mills theory for decay into three massive W-bosons. The calculations are done at two-loop order by deriving and solving canonical differential equations in the asymptotical limit of nearly vanishing W-masses. We confirm our previous findings that infrared physics of 'off-shell observables' is governed by the octagon anomalous dimension rather than the cusp. In addition, the form factor in question possesses a nontrivial remainder function, which was found to be identical to the massless case, upon a proper subtraction of infrared logarithms (and finite terms). However, the iterative structure of the object is more intricate and is not simply related to the previous orders in coupling as opposed to amplitudes/form factors at the origin of the moduli space.

Contents

1 Introduction **1**

2 Setting up conventions **3**

3 One loop **4**

4 Two loops **6**
- 4.1 Tri-pentagon 8
- 4.2 Tri-box 13
- 4.3 Double boxes and nonplanar tri-box 13

5 Adding things up **14**
- 5.1 Infrared exponentiation and general structure 15
- 5.2 Finite part 16
- 5.3 Iterative structure 17

6 Conclusion **18**

1 Introduction

The so-called Coulomb branch of the spontaneously broken maximally supersymmetric Yang-Mills (sYM) theory [1] is a natural laboratory to study off-shell amplitudes and form factors in four-dimensional gauge theories. Endowing some (or all) scalars of the model with vacuum expectation values, one can adjust their values in a way as to yield matrix elements which possess massive external states, i.e., W-bosons, Higgs-like scalars etc., but only massless excitations propagating in quantum loops.

The $\mathcal{N} = 4$ model away from the origin of the moduli space can naturally be obtained from a generalized form of the dimensional reduction [2–4] akin to the original one used to discover its Lagrangian in the first place from the ten-dimensional $\mathcal{N} = 1$ sYM [5, 6], or the six-dimensional $\mathcal{N} = (1, 1)$ sYM (see, e.g., [7]). Instead of setting the extra-dimensional, i.e., $D > 4$, components of momenta to zero, one can trade them in lieu of the scalars' moduli, i.e., vacuum averages of $D > 4$ components of the ten-dimensional gauge field. The 16 supercharges remain unbroken in this phase, but the supersymmetric algebra gets a central extension with BPS charges induced by nonvanishing masses, and thus the theory shares a gamut of properties of its conformal sibling. One-loop analyses demonstrated that the Coulomb branch scattering amplitudes obey a no-triangle rule[1], thus enjoying only boxes

[1]Bubbles and tadpoles are excluded form the get-go based on their poor ultraviolet properties.

in their integral expansion [3, 8] at a generic point of the moduli space, — a generalization of the background-field gauge proof from Refs. [9, 10] applicable to off-shell massless amplitudes. Further, there are no rational terms as well [3, 11] and, therefore, integrands on the Coulomb branch are cut-constructible [12]. Making use of this latter property, a proof of the dual conformal invariance of massive loop integrals from a six-dimensional viewpoint was elucidated in Refs. [7, 13]. Also correctness of the four-leg amplitudes at four-loop order (including nonplanar contributions) [7] was demonstrated to expressions built using solely four-dimensional momenta in the cuts [14] by lifting four-dimensional inner products of momenta up to six dimensions.

The above higher-dimensional perspective provides a natural bridge between the dimensionally regularized theory and its massive version to tame infrared divergences in scattering amplitudes and form factors in a gauge invariant manner. Their explicit structure for the four-gluon amplitude and the Sudakov form factor was deduced at up to three-loop level [1, 15–17] by promoting massless integral bases constructed in four dimensions to involve massive propagators only around graphs' periphery. Infrared structure was shown to be in accord with the well-known conformal phase of $\mathcal{N}=4$ sYM in $D=4-2\varepsilon$ [18] (see Refs. [19–21], for earlier QCD studies) for a minor difference in kinematically-independent contributions and in compliance with a common wisdom that infrared properties of gauge theories are encoded in the so-called cusp anomalous dimension [22, 23].

The situation drastically changes, however, when all internal propagators are left massless, but the external legs are kept massive, or off-shell, as we will refer to them hereafter. Four- [24] and five-leg [25] W-boson amplitudes as well as the two–W-boson Sudakov form factor [26, 27] enjoyed the same recurrent feature in variance to the naive expectation: the infrared logarithms are governed by an exponent different from the cusp anomalous dimension. Instead they exhibit dependence on the so-called octagon anomalous dimension which made its debut in completely different circumstances: the light-cone limit of correlation functions of infinitely heavy BPS operators [28, 29] and the near-origin asymptotics of the six-gluon remainder function [30].

In the current paper, we continue our exploration of the Coulomb branch by addressing a more involved quantity, the three–W-boson form factor $\mathcal{F}_3$ of the lowest component of the stress tensor multiplet

$$\int d^4x\, \mathrm{e}^{-iq\cdot x}\langle p_1,p_2,p_3|\,\mathrm{tr}\,\phi_{12}^2(x)|0\rangle = (2\pi)^4\delta^{(4)}(q-p_1-p_2-p_3)\mathcal{F}_3\,. \tag{1.1}$$

Here we explicitly extracted the energy-momentum conserving delta function. This is the simplest 'observable' which possesses nontrivial remainder function after factoring out infrared divergences [31]. In the conformal phase of the theory, it was bootstrapped to a staggering eight-loop order [32, 33] using techniques adopted from scattering amplitudes [34]. Our goal will be much more modest: we will calculate its off-shell version at two loops. The incentive for our analysis is multifold. First, we would like to confirm the octagon anomalous dimension as the Sudakov exponent of 'off-shell observables'. Second, we will

establish similarities/differences to the iterative structure of the form factor with increased perturbative order compared to its conformal analogue. Third, given that the infrared logarithms are different in the on- and off-shell cases, will the remainder functions differ as well?

Our subsequent presentation will be organized as follows. In the next section, we set up our notations. Then, in Sect. 3, we perform the one-loop calculation, which is then followed by two loops in Sect. 4. The only graph that was not touched upon in the existing literature corresponds to the tri-pentagon. So we perform its calculation from scratch in Sect. 4.1. It is then followed by all other contributing graphs. In Sect. 5, we add them up and use symbol analysis to simplify the sum and uncover the structure of the form factor at two-loop order. Finally, we conclude.

2 Setting up conventions

The form factor of three W-bosons contains an overall prefactor encoding polarization dependence of the external states. We will not be interested in it in what follows and thus introduce the ratio function

$$F_3 \equiv \mathcal{F}_3/\mathcal{F}_{3,\text{tree}}\,. \tag{2.1}$$

F_3 depends on three invariants s_{ij} and the off-shellnesses of the W-legs, which will be taken to have the same value μ,

$$s_{ij} \equiv (p_i+p_j)^2\,, \quad p_i^2 \equiv -\mu\,. \tag{2.2}$$

These are linearly dependent, however,

$$s_{12}+s_{23}+s_{31} = q^2-3\mu\,. \tag{2.3}$$

Since the form factor is a homogeneous function of these kinematical variables, one can set one scale to one, e.g., $q^2=-1$ below. Equivalently this can be done by introducing Mandelstam-like variables and the 'mass parameter' m

$$u = s_{12}/q^2\,, \quad v = s_{23}/q^2\,, \quad w = s_{31}/q^2\,, \quad m = -\mu/q^2\,, \tag{2.4}$$

and ignore the overall mass scale q^2: F_3 is dimensionless.

F_3 admits perturbative series expansion in the gauge coupling g^2_{YM} accompanied at each order by the number of colors N (in the planar limit), allowing us to introduce

$$g^2 \equiv \frac{g^2_{\text{YM}}N}{(4\pi)^2}(4\pi \mathrm{e}^{-\gamma_{\text{E}}})^{\varepsilon}\,, \tag{2.5}$$

which comes hand-in-hand with a measure of the dimensionally-regularized loop momentum integrals,

$$(\mu^2_{\text{DR}}\mathrm{e}^{\gamma_{\text{E}}})^{\varepsilon}\int\frac{d^D\ell}{i\pi^{D/2}}\,, \tag{2.6}$$

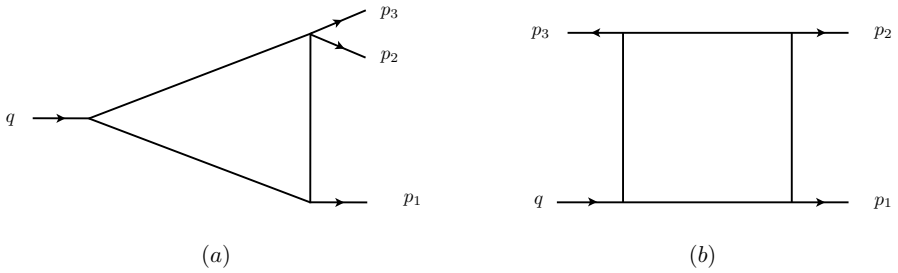


Figure 1. Graphs contributing to the one-loop form factor $F_3^{(1)}$.

in $D = 4 - 2\varepsilon$. We will dwell on the necessity to deal with the space-time away from $D = 4$, even though we already have an infrared regulator m, when it becomes indispensable at two-loop order.

Thus, we have to the lowest two orders

$$F_3 = 1 + g^2 F_3^{(1)} + g^4 F_3^{(2)} + \dots , \tag{2.7}$$

where $F_3^{(i)}$ are given by linear combinations of one- and two-loop integrals for $i = 1, 2$, respectively. Instead of using Feynman rules of the Coulomb phase of $\mathcal{N} = 4$ sYM in order to find the latter, we will employ, as was advocated in the introduction, the connection between the spontaneously broken phase in $D = 4$ and higher-dimensional theory with exact gauge symmetry to recycle generalized unitarity analyses from Refs. [35, 36] and [31], to ascertain integral families defining the one and two-loop integrands, respectively. All calculations will be done in the limit $m \to 0$, i.e., they will be valid up to power corrections in m.

3 One loop

Without further ado, let us start our analysis with the one-loop form factor. At this order, $F_3^{(1)}$ receives the following expansion in terms of the triangle Tri and box Box integrals [35, 36]

$$F_3^{(1)} = \sum_{n=0}^{2} \mathbb{P}^n \left[(s_{12} + s_{13}) \mathrm{Tri}(p_1, p_2 + p_3) + \frac{1}{2} s_{12} s_{23} \mathrm{Box}(p_1, p_2, p_3) \right] , \tag{3.1}$$

shown in Fig. 1. In this equation, we introduced an operator $\mathbb{P}$ that shifts momentum indices of any function to its right by one

$$\mathbb{P} f_{ij\dots} = f_{i+1(\mathrm{mod}\,3), j+1(\mathrm{mod}\,3)\dots} , \tag{3.2}$$

modulo 3, which imposes periodicity. The Mandelstam-like variables then transform as $\mathbb{P}(u, v, w) = (v, w, u)$. Both integrals in Eq. (3.1) can immediately be expressed in terms of

the Davydychev-Ussyukina function $\Phi_1(x,y)$ [37, 38],

$$\Phi_\ell(x,y) = -\sum_{j=\ell}^{2\ell} \frac{j!(-1)^j \log^{2\ell-j}\left(\frac{y}{x}\right)}{\ell!(j-\ell)!(2\ell-j)!} \frac{\mathrm{Li}_j\big(-(\rho x)^{-1}\big) - \mathrm{Li}_j\big(-(\rho y)^{+1}\big)}{\lambda}\,, \tag{3.3}$$

where ρ and λ are functions of x and y,

$$\lambda(x,y) = [(1-x-y)^2 - 4xy]^{1/2}\,, \qquad \rho(x,y) = 2[1-x-y-\lambda(x,y)]^{-1}\,, \tag{3.4}$$

via

$$\mathrm{Tri}(p_1, p_2+p_3) = \Phi_1(m,v)/(1-v) \tag{3.5}$$

$$\mathrm{Box}(p_1,p_2,p_3) = \Phi_1\left(m^2/(uv), m/(uv)\right)/(uv)\,. \tag{3.6}$$

Their small-mass expansion yields the following expressions for the triangle and the box

$$\mathrm{Tri}(p_1, p_2+p_3) = -\frac{\log m \log v + 2\mathrm{Li}_2(1-v)}{1-v}\,, \tag{3.7}$$

$$\mathrm{Box}(p_1,p_2,p_3) = -\frac{2\log^2 m - 2\log m \log(uv) + \log^2(uv) + 2\zeta_2}{uv}\,, \tag{3.8}$$

where we used the condition (2.3), up to terms vanishing as a power of m. Adding all contributions up, we find

$$\begin{aligned} F_3^{(1)} = &-\log^2\frac{m}{u} - \log^2\frac{m}{v} - \log^2\frac{m}{w} \\ &- \log u \log v - \log v \log w - \log w \log u \\ &- 2\mathrm{Li}_2(1-u) - 2\mathrm{Li}_2(1-v) - 2\mathrm{Li}_2(1-w) - 3\zeta_2\,. \end{aligned} \tag{3.9}$$

It is instructive to compare this results to the conformal case, calculated within dimensional regularization (or rather reduction), [35],

$$\begin{aligned} F_3^{(1)}(\varepsilon) = &-\frac{1}{\varepsilon^2}\left[\left(\frac{-\mu^2_{\mathrm{DR}}}{u}\right)^{\varepsilon} + \left(\frac{-\mu^2_{\mathrm{DR}}}{v}\right)^{\varepsilon} + \left(\frac{-\mu^2_{\mathrm{DR}}}{w}\right)^{\varepsilon}\right] \\ &- \log u \log v - \log v \log w - \log w \log u \\ &- 2\mathrm{Li}_2(1-u) - 2\mathrm{Li}_2(1-v) - 2\mathrm{Li}_2(1-w) + \frac{9}{2}\zeta_2\,. \end{aligned} \tag{3.10}$$

We observe that the finite parts are identical in the two cases, except for the coefficient of ζ_2. When Eq. (3.10) expanded in the Laurent series, the coefficient of the double logarithms of $\mu^2_{\mathrm{DR}}/(u,v,w)$ are half of the off-shell case, as anticipated. This is the well-known doubling phenomenon observed back in the early days of QED [39, 40] and well-understood by now as a result of an extra, the so-called ultra-soft, region [41–43] of loop momentum producing leading effects on par with other regimes present in both.

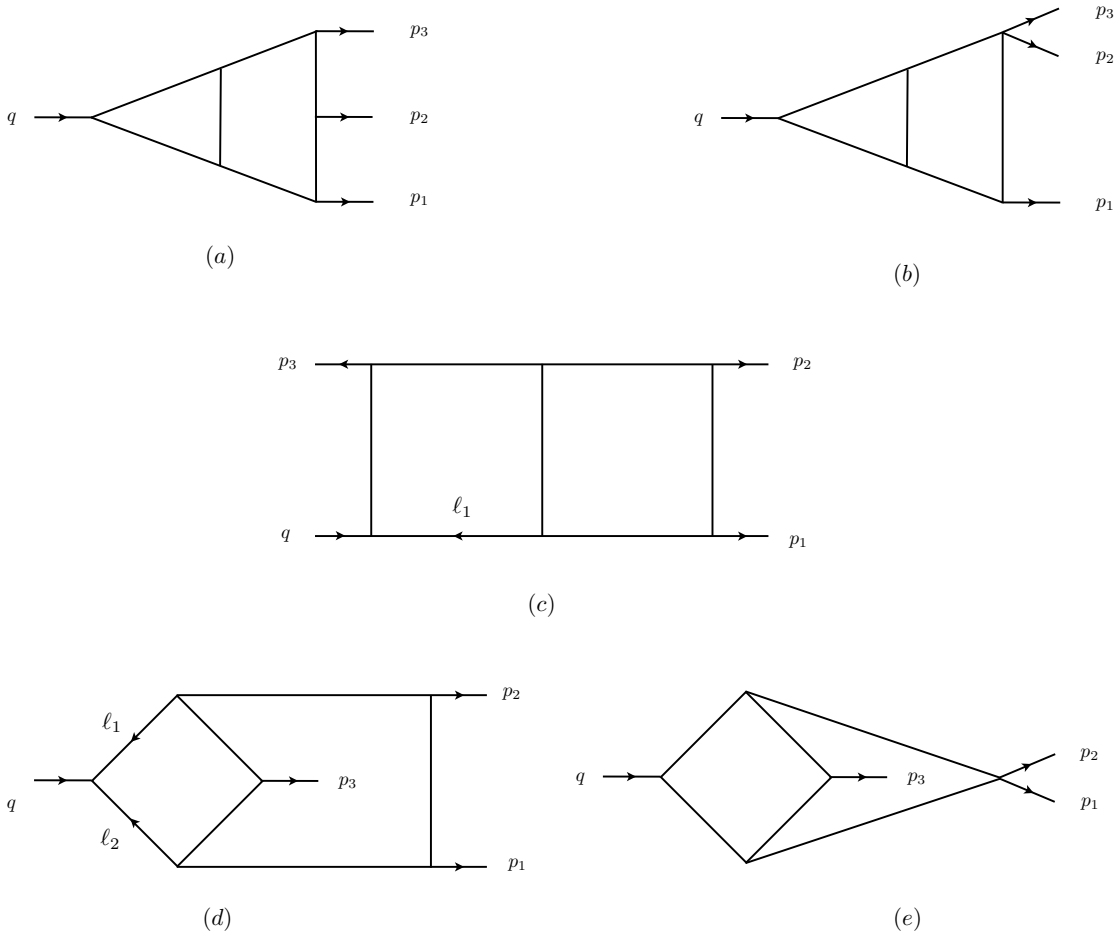


Figure 2. Graphs contributing to the two-loop form factor $F_3^{(2)}$. The integrands, built from product of propagators read off from these diagrams, are accompanied by numerators according to Eqs. (4.1)–(4.5).

4 Two loops

We now proceed to the two loop calculation. The integrands for the form factor $F_3^{(2)}$ were constructed in Ref. [31] using (generalized) unitarity cut technique. With a slight change of the nomenclature compared to [31], the relevant graphs shown in Fig. 2 generate the following integrals[2]

$$\text{TriPent}(p_1,p_2,p_3) = \mathrm{e}^{2\varepsilon\gamma_{\rm E}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \frac{q^2 s_{12} s_{23}}{\text{denom}_{(a)}}\,, \tag{4.1}$$

$$\text{TriBox}(p_1,p_2+p_3) = \mathrm{e}^{2\varepsilon\gamma_{\rm E}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \frac{q^2[s_{12}+s_{31}]}{\text{denom}_{(b)}}\,, \tag{4.2}$$

[2]Here and below, we set the mass scale of dimensional regularization to one, $\mu^2_{\rm DR} = 1$

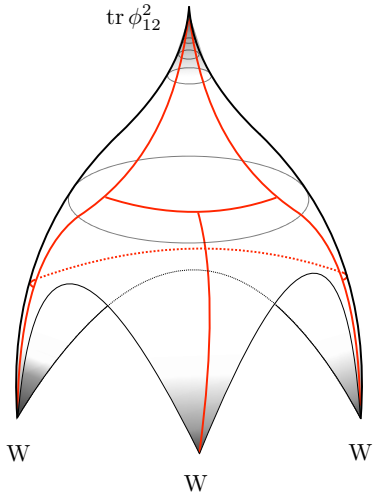


Figure 3. World-sheet perspective of the three-leg form factor and the non-planar graph from Fig. 2 (d) overlayed on it: it demonstrates why it produces contribution of leading order in color.

$$\text{DBox}(p_1,p_2,p_3) = \mathrm{e}^{2\varepsilon\gamma_{\text{E}}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \frac{s_{12}[s_{31}\,\ell_1\cdot p_1 - s_{23}\,\ell_1\cdot p_2]}{\text{denom}_{(c)}}\,, \tag{4.3}$$

$$\text{NBox}(p_1,p_2,p_3) = \mathrm{e}^{2\varepsilon\gamma_{\text{E}}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \frac{s_{12}[\frac{1}{2}s_{23}s_{31} - s_{23}\ell_1\cdot p_2 - s_{31}\,\ell_2\cdot p_1]}{\text{denom}_{(d)}}\,, \tag{4.4}$$

$$\text{NTriBox}(p_1+p_2,p_3) = \mathrm{e}^{2\varepsilon\gamma_{\text{E}}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \frac{\frac{1}{2}q^2[s_{23}+s_{31}]}{\text{denom}_{(e)}}\,, \tag{4.5}$$

where the denominator structure can readily be read off from the corresponding graphs. In terms of these integrals, the two-loop form factor is given by the expression

$$\begin{aligned} F_3^{(2)} = \sum_{n=0}^{2} \mathbb{P}^n \Big[&\text{TriBox}(p_1,p_2+p_3) + \text{TriBox}(p_3,p_1+p_2) + \text{TriPent}(p_1,p_2,p_3) \\ &+ \text{DBox}(p_1,p_2,p_3) + \text{DBox}(p_3,p_2,p_1) + \text{NBox}(p_1,p_2,p_3) \\ &+ \text{NTriBox}(p_1+p_2,p_3)\Big]\,. \end{aligned} \tag{4.6}$$

Notice that starting from this order, there are non-planar graphs which are leading order in color, i.e., Fig. 2 (d) and (e). The reason for this is that the operator $\text{tr}\,\phi_{12}^2$ is a singlet with respect to the SU(N) and thus does not 'participate' in color traces. It becomes quite obvious from the world-sheet perspective of the matrix element (1.1) demonstrated in Fig. 3 where the operator corresponds to the closed string state, while the W-bosons to the open ones.

Out of all the contributions in Eq. (4.6), a truly new integral, which was not addressed in existing literature, is the tri-pentagon, Fig. 2 (a). So we start with its analysis first in the next section.

4.1 Tri-pentagon

Let us begin with the construction of the canonical basis for the tri-pentagon family, see Fig. 2 (a), by routing the loop momenta ℓ_1 and ℓ_2 according to the following definitions of propagator denominators D_i ($i = 1, \ldots, 7$) and irreducible scalar products D_8 and D_9,

$$\begin{aligned} &D_1 = -\ell_1^2\,, \quad D_2 = -(\ell_1 + p_1)^2\,, \quad D_3 = -(\ell_1 + p_1 + p_2)^2\,, \qquad D_4 = -(\ell_1 + p_1 + p_2 + p_3)^2\,, \\ &D_5 = -\ell_2^2\,, \quad D_6 = -(\ell_2 - \ell_1)^2\,, \quad D_7 = -(\ell_2 + p_1 + p_2 + p_3)^2\,, \quad D_8 = -(\ell_2 + p_1)^2\,, \\ &\qquad\qquad\qquad\quad D_9 = -(\ell_2 + p_1 + p_2)^2\,, \end{aligned} \tag{4.7}$$

such that

$$G_{a_1a_2a_3a_4a_5a_6a_7a_8a_9} \equiv \mathrm{e}^{2\varepsilon\gamma_\mathrm{E}} \int \frac{d^D\ell_1}{i\pi^{D/2}} \int \frac{d^D\ell_2}{i\pi^{D/2}} \prod_{i=1}^{9} D_i^{-a_1}\,. \tag{4.8}$$

The use of dimensionally regularized integrals is required for proper use of the integration-by-part technique in order to work with vanishing surface loop integrals, which are at the heart of the formalism [44]. An IBP reduction with the `FIRE` code [45–47] immediately reveals 49 initial Master Integrals (MIs), which are further reduced to 46 by finding equivalences among them with the `LiteRed` software [48, 48], generating thus the primary basis

$$\begin{aligned} \boldsymbol{I} = \{&G_{000111000}, G_{001011000}, G_{010001100}, G_{010011000}, G_{001011100}, G_{001011200}, \\ &G_{001111000}, G_{001112000}, G_{010011100}, G_{010011200}, G_{010110100}, G_{010111000}, \\ &G_{010112000}, G_{011001100}, G_{011001200}, G_{011011000}, G_{011012000}, G_{100110100}, \\ &G_{101001100}, G_{101001200}, G_{101010100}, G_{110001100}, G_{110001200}, G_{110010100}, \\ &G_{001111100}, G_{010111100}, G_{011011100}, G_{011011200}, G_{011012100}, G_{011012200}, \\ &G_{011021100}, G_{012011100}, G_{021011100}, G_{011110100}, G_{011111000}, G_{011112000}, \\ &G_{101011100}, G_{101110100}, G_{110011100}, G_{110110100}, G_{111001100}, G_{111001200}, \\ &G_{111010100}, G_{011111100}, G_{111011100}, G_{111110100}\}\,. \end{aligned} \tag{4.9}$$

Next we turn to the derivation of differential equations for $\boldsymbol{I}$ making use of a `FIRE` interface to `LiteRed`,

$$\partial_i \boldsymbol{I} = \boldsymbol{M}_i \cdot \boldsymbol{I}\,, \tag{4.10}$$

in the kinematical invariants $i = u, v, w, m$. The goal is now to convert them to the canonical form [49]

$$\partial_i \boldsymbol{J} = \varepsilon \boldsymbol{A}_i \cdot \boldsymbol{J}\,, \qquad \varepsilon \boldsymbol{A}_i = \boldsymbol{T}^{-1} \cdot \boldsymbol{M} \cdot \boldsymbol{T} - \boldsymbol{T}^{-1} \cdot \partial_i \boldsymbol{T}\,, \tag{4.11}$$

with some transformation matrix $\boldsymbol{T}$. In fact, what we need is an *asymptotically* canonical basis, which captures all logarithmically enhanced and constant terms in m as m goes to

zero. This can easily be accomplished by keeping track of only singular power-like terms in the 'virtuality' matrix

$$\boldsymbol{A}_m = \frac{\varepsilon}{m}\boldsymbol{A}_m^0 + O(m^0)\,, \tag{4.12}$$

as was explained at length in Ref. [50].

Splitting the basis elements of $\boldsymbol{I}$ into sectors, we form their linear combinations accompanied by unknown functions of the Mandelstam-like variables (u, v, w) and fix the former by enforcing the ε-form of the differential equations (4.11). Having fixed the diagonal blocks in this manner, the off-diagonal ones can be constrained by using two available software packages `Canonica` [51, 52] and `Libra` [53, 54]. To achieve this, one first transforms the equations to the Fuchsian, i.e., dLog, form followed by factorization of the ε-dependence into an overall factor [53]. `Canonica` is solely based on built-in Mathematica commands and fails to successfully solve corresponding systems of linear equations. Therefore, we used two strategies in our analysis. One was based exclusively on `Libra`. However, having constructed canonical form of differential equations, we discovered that five of its elements did not possess uniform transcendentality[3] (UT), namely, J_i's with indices $i = 34, 43, 44, 45, 46$. So in our attempt to alleviate this problem, we deduced yet another form of the canonical differential equations by the combined use of `Canonica` (to bring equations to the Fuchsian form) and `Libra` (for the derivation of the ε-form). Though, the basis found was slightly different from the first one, nevertheless the very same five elements suffered from the very same problem. Obviously, this was not in any way an obstruction in our subsequent steps of solving theses 'canonical' equations rather it was merely a nuisance: instead of fixing a set of integration constant of uniform transcendentality at each ε^n-order, we had to use a sum of constants of increasing transcendental weight $w_i \leq n$. The asymptotically canonical basis, which we used in the explicit iterative solution of the differential equations, is

$$J_1 = \varepsilon^2(3m - u - v - w)G_{000122000}\,, \tag{4.13}$$
$$J_2 = \varepsilon^2 u G_{001022000}\,, \tag{4.14}$$
$$J_3 = \varepsilon^2 v G_{010002200}\,, \tag{4.15}$$
$$J_4 = \varepsilon^2 m G_{010022000}\,, \tag{4.16}$$
$$J_5 = \varepsilon^3(v + w)G_{002011100}\ ,\,, \tag{4.17}$$
$$J_6 = \varepsilon^2(u + v + w)((2\varepsilon - 1)G_{001011200} + \varepsilon G_{002011100})\,, \tag{4.18}$$
$$J_7 = \varepsilon^2(2\varepsilon - 1)m G_{002111000}\,, \tag{4.19}$$
$$J_8 = \varepsilon^3(v + w)G_{001112000}\,, \tag{4.20}$$
$$J_9 = \varepsilon^3(u + w)G_{020011100}\,, \tag{4.21}$$
$$J_{10} = \varepsilon^2(1 - 2\varepsilon)(u + v + w)G_{010011200}\,, \tag{4.22}$$
$$J_{11} = \frac{7}{25}\varepsilon^2(1 - 2\varepsilon)^2 G_{010110100}\,, \tag{4.23}$$

[3]We would like to thank Johannes Henn for instructive communications on this point.

$$J_{12} = \varepsilon^2(2\varepsilon-1)(3\varepsilon-1)G_{010111000}\,, \tag{4.24}$$
$$J_{13} = \varepsilon^3(u+w)G_{010112000}\,, \tag{4.25}$$
$$J_{14} = \varepsilon^2(2\varepsilon-1)(3\varepsilon-1)G_{011001100}\,, \tag{4.26}$$
$$J_{15} = \varepsilon^3 v G_{011001200}\,, \tag{4.27}$$
$$J_{16} = \varepsilon^2(2\varepsilon-1)(3\varepsilon-1)G_{011011000}\,, \tag{4.28}$$
$$J_{17} = \varepsilon^3 u G_{011012000}\,, \tag{4.29}$$
$$J_{18} = \frac{1}{25}\varepsilon^2(1-2\varepsilon)^2 G_{100110100}\,, \tag{4.30}$$
$$J_{19} = \varepsilon^2(2\varepsilon-1)(3\varepsilon-1)G_{101001100}\,, \tag{4.31}$$
$$J_{20} = \varepsilon^3(v+w)G_{101001200}\,, \tag{4.32}$$
$$J_{21} = \frac{7}{25}\varepsilon^2(1-2\varepsilon)^2 G_{101010100}\,, \tag{4.33}$$
$$J_{22} = \varepsilon^2(1-2\varepsilon)m G_{210001100}\,, \tag{4.34}$$
$$J_{23} = \varepsilon^3(u+w)G_{110001200}\,, \tag{4.35}$$
$$J_{24} = \frac{7}{25}\varepsilon^2(1-2\varepsilon)^2 G_{110010100}\,, \tag{4.36}$$
$$J_{25} = \varepsilon^4(v+w)G_{001111100}\,, \tag{4.37}$$
$$J_{26} = \varepsilon^4(u+w)G_{010111100}\,, \tag{4.38}$$
$$J_{27} = \varepsilon^4(u+v)G_{011011100}\,, \tag{4.39}$$
$$J_{28} = \frac{1}{2}\varepsilon^2\big[-vG_{010002200} + mG_{010022000} + 2\varepsilon v((u+v+w)G_{011011200} - 2G_{011001200})\big]\,, \tag{4.40}$$
$$J_{29} = \frac{1}{2}\varepsilon^2\big[-2vG_{010002200} - 11mG_{010022000} + 2\varepsilon uvG_{011012100}\big]\,, \tag{4.41}$$
$$J_{30} = \frac{\varepsilon^2}{v+w}\big[-uvG_{010002200} - m(u-2v-2w)G_{010022000} + u\left[2\left(6\varepsilon^2-5\varepsilon+1\right)G_{011001100} - 4\varepsilon vG_{011001200} + mv(u+v+w)G_{011012200}\right]\big]\,, \tag{4.42}$$
$$J_{31} = \frac{1}{2}\varepsilon^2\big[vG_{010002200} + 12mG_{010022000} + 2\varepsilon u(u+v+w)G_{011021100}\big]\,, \tag{4.43}$$
$$J_{32} = \frac{1}{4}\varepsilon^2\big[vG_{010002200} + 5mG_{010022000} + 4\big[\left(-6\varepsilon^2+5\varepsilon-1\right)G_{011001100} + \varepsilon vG_{011001200} + \varepsilon m(v+w)G_{012011100}\big]\big]\,, \tag{4.44}$$
$$J_{33} = -\frac{1}{4}\varepsilon^2\big[3vG_{010002200} + 17mG_{010022000} + 4\left(6\varepsilon^2-5\varepsilon+1\right)G_{011001100} - 4\varepsilon m(u+w)G_{021011100}\big]\,, \tag{4.45}$$
$$J_{34} = \frac{1}{5}\varepsilon^2\big[\left(14\varepsilon^2-9\varepsilon+1\right)(2G_{010110100} - 4G_{110010100}) + 5\varepsilon vG_{011110100}\big]\,, \tag{4.46}$$
$$J_{35} = (1-2\varepsilon)\varepsilon^3 vG_{011111000}\,, \tag{4.47}$$
$$J_{36} = \varepsilon^3 uvG_{011112000}\,, \tag{4.48}$$

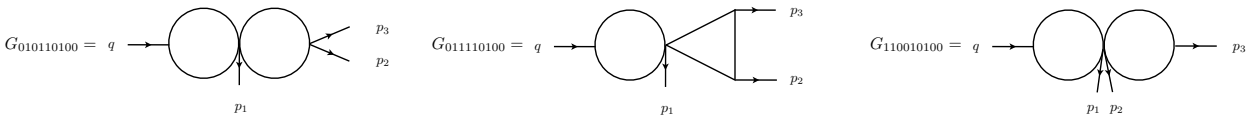


Figure 4. Diagrammatic form of the integrals forming the element J_{34} of the canonical basis (4.13).

$$J_{37} = \varepsilon^4(v+w)G_{101011100}\,, \tag{4.49}$$

$$J_{38} = -\frac{7}{5}\varepsilon^3(2\varepsilon-1)(v+w)G_{101110100}\,, \tag{4.50}$$

$$J_{39} = \varepsilon^4(u+w)G_{110011100}\,, \tag{4.51}$$

$$J_{40} = -\frac{7}{5}\varepsilon^3(2\varepsilon-1)(u+w)G_{110110100}\,, \tag{4.52}$$

$$J_{41} = \varepsilon^3(1-2\varepsilon)uG_{111001100}\,, \tag{4.53}$$

$$J_{42} = \varepsilon^3 uvG_{111001200}\,, \tag{4.54}$$

$$J_{43} = \frac{1}{5}\varepsilon^2\left[2\left(14\varepsilon^2-9\varepsilon+1\right)\left(2G_{101010100}-4G_{110010100}\right)+5\varepsilon uG_{111010100}\right], \tag{4.55}$$

$$\begin{aligned} J_{44} = \frac{1}{7}\varepsilon^2\big[&7\varepsilon^2 v(v+w)G_{011111100}+2\varepsilon(7\varepsilon-1)vG_{011110100} \\ &+2(2\varepsilon-1)(7\varepsilon-1)(2G_{010110100}-4G_{110010100})\big]\,, \end{aligned} \tag{4.56}$$

$$\begin{aligned} J_{45} = \frac{1}{7}\varepsilon^2\big[&\left(14\varepsilon^2-9\varepsilon+1\right)\left(2G_{101010100}-4G_{110010100}\right) \\ &+\varepsilon u\big[2(7\varepsilon-1)G_{111010100}+7\varepsilon(u+w)G_{111011100}\big]\big]\,, \end{aligned} \tag{4.57}$$

$$\begin{aligned} J_{46} = \frac{1}{5}\varepsilon^2\big[&(2\varepsilon-1)(7\varepsilon-1)\big[2G_{010110100}+2G_{101010100}-8G_{110010100}\big] \\ &+2\varepsilon(7\varepsilon-1)\big[vG_{011110100}+uG_{111010100}\big]-7\varepsilon(2\varepsilon-1)uvG_{111110100}\big]\,. \end{aligned} \tag{4.58}$$

First, we solved the 'virtuality' differential equation, in the small-virtuality limit

$$\boldsymbol{J} = m^{\varepsilon \boldsymbol{A}^0_m}\cdot \boldsymbol{J}_0 \tag{4.59}$$

related to the 'massless' MIs $\boldsymbol{J}_0$ via the matrix exponent $m^{\varepsilon \boldsymbol{A}^0_m}$. Next, we solved the $m=0$ limit of the differential equations in Mandelstam-like variables via the Chen iterated integrals on a piece-wise contour [55]

$$\boldsymbol{J}_0 = P_\gamma \exp\left(\varepsilon \int_{[0,u]\cup[0,v]\cup[0,w]} \boldsymbol{A}^0\right)\boldsymbol{J}_{00}\,, \tag{4.60}$$

with the differential of the A-matrices $\boldsymbol{A} = du\boldsymbol{A}^0_u + dv\boldsymbol{A}^0_v + dw\boldsymbol{A}^0_w$. At each order of the ε-expansion, we found solutions in terms of multiple polylogarithms [56].

Finally, we had to fix the vector of the integration constants $\boldsymbol{J}_{00}$ at each order of the ε-expansion

$$\boldsymbol{J}_{00} = \sum_{p\geq 0} \varepsilon^p \boldsymbol{c}^{(p)} \,. \tag{4.61}$$

To accomplish this, we used two criteria: (i) the absence of spurious poles in the right-hand sides of differential equations at the location of $u+v$, $v+w$ and $w+u$ poles and (ii) numerical integration with `FIESTA` [57] with subsequent use of the `PSLQ` algorithm [58]. However, these considerations alone did no allow us to fully analytically determine all of the integration constants. We needed further input. We found that all undetermined contributions are reduced a set of unknowns which can be determined in turn by evaluating one of the elements of the canonical basis explicitly. The element in question is J_{34}, which is given by a linear combination of factorized products of bubbles and triangles, Eq. (4.46), as demonstrated in Fig. 4. This can be easily calculated making use of the code `MBcreate.m` [59]. It yielded the following expressions

$$G_{010110100} = v^{-\varepsilon}(u+v+w)^{-\varepsilon}\frac{\mathrm{e}^{2\gamma\varepsilon}\,\Gamma(1-\varepsilon)^4\Gamma(\varepsilon)^2}{\Gamma(2-2\varepsilon)^2}\,, \tag{4.62}$$

$$\begin{aligned} G_{011110100} = {} & m^{-2\varepsilon}v^{-\varepsilon-1}(u+v+w)^{-\varepsilon}\frac{\mathrm{e}^{2\gamma\varepsilon}\,\Gamma(1-\varepsilon)^2\Gamma(\varepsilon)}{\Gamma(1-2\varepsilon)\Gamma(2-2\varepsilon)} \\ & \times\left[m^{2\varepsilon}\Gamma(-\varepsilon)^2\Gamma(\varepsilon+1)+v^{\varepsilon}\Gamma(1-\varepsilon)\Gamma(\varepsilon)\left(2m^{\varepsilon}\Gamma(-\varepsilon)+v^{\varepsilon}\Gamma(1-2\varepsilon)\Gamma(\varepsilon)\right)\right], \end{aligned} \tag{4.63}$$

$$G_{110010100} = m^{-\varepsilon}(u+v+w)^{-\varepsilon}\frac{\mathrm{e}^{2\gamma\varepsilon}\,\Gamma(1-\varepsilon)^4\Gamma(\varepsilon)^2}{\Gamma(2-2\varepsilon)^2}\,. \tag{4.64}$$

Matching their expansions to the iterative solution, we found our final result. The expressions are too lengthy to be displayed here in the body or appendices, so they are relegated to the accompanying Mathematica notebook `TriPentagonA2Z.nb`, where an interested reader could find as well all steps from-A-to-Z for the determination of their expressions starting with necessary initial IBP reductions. The tri-pentagon (4.1) is then given by the integral

$$\mathrm{TriPent} = uv(u+v+w)G_{111111100}\,, \tag{4.65}$$

which is not one of the elements of the above basis, but can be easily reduced to them by means of an IBP reduction. The latter gives

$$\begin{aligned} \mathrm{TriPent} = \frac{1}{\varepsilon^4}\Bigg[& -J_3-\frac{11}{2}J_4-\frac{2(1-7\varepsilon)(u+v+w)}{7(v+w)}[5J_{11}-J_{34}] \\ & -\frac{2(1-7\varepsilon)(u+v+w)}{7(u+w)}[5J_{21}-J_{43}] \\ & +\frac{20(1-7\varepsilon)(u^2+2uv+v^2+3uw+3vw+2w^2)}{7(u+w)(v+w)}J_{24} \\ & -J_{29}+J_{36}+J_{42}+\frac{u}{v+w}J_{44}+\frac{v}{u+w}J_{45}+\frac{5}{7}J_{46}\Bigg]\,. \end{aligned} \tag{4.66}$$

Notice that some of the MIs in this expression do not possess UT individually, however, in the sum TriPent is indeed UT. However, it is not a pure function: multiple polylogarithms are accompanied by rational prefactors of the (u, v, w) variables.

Before we move on to the next graph, let us point out that there is an alternative way to obtain the analytic result for the tri-pentagon (4.1), which bypasses the need for construction of the MI basis. It is based on the famous triangle relation established in Ref. [44]. Namely, by applying it to the left triangle subgraph in Fig. 2 (a), $G_{111111100}$ can be recast in terms of integrals corresponding to graphs with either the central propagator or one of the two (adjacent) left ones shrunk into a point. The first of these contributions yields a factorized diagram of a bubble with an attached triangle that can be calculated with, say, `MBcreate.m` package [59]. While the second one reduces to the double box integrals evaluated in [50] for the required kinematics.

4.2 Tri-box

The tri-box graph in Fig. 2 (b) is related to the Davydychev-Ussyukina function Φ_2 given in Eq. (3.3),

$$\mathrm{TriBox}(p_1, p_2 + p_3) = (u + w)\Phi_2(m, v)\,. \tag{4.67}$$

Its small-m expansion immediately produces the sought after expression

$$\begin{aligned}\mathrm{TriBox}(p_1, p_2 + p_3) = {}& \left[\frac{1}{2}\mathrm{Li}_2\left(\frac{u}{u-1}\right) + \frac{1}{4}\log^2\frac{1-u}{u} + \frac{\zeta_2}{2}\right]\log^2 m \\ &+ \left[3\mathrm{Li}_3\left(\frac{u}{u-1}\right) - \mathrm{Li}_2\left(\frac{u}{u-1}\right)\log u - \frac{1}{2}\log(1-u)\log^2\frac{1-u}{u} - \zeta_2\log\frac{(1-u)^3}{u^2}\right]\log m \\ &+ 6\mathrm{Li}_4\left(\frac{u}{u-1}\right) - 3\mathrm{Li}_3\left(\frac{u}{u-1}\right)\log(u) + \frac{1}{2}\mathrm{Li}_2\left(\frac{u}{u-1}\right)\log^2 u \\ &+ \frac{1}{4}\log^2(1-u)\log^2\frac{1-u}{u} + \frac{1}{2}\zeta_2\log^2 u + 3\zeta_2\log(1-u)\log\frac{1-u}{u} + \frac{21}{2}\zeta_4\,.\end{aligned} \tag{4.68}$$

4.3 Double boxes and nonplanar tri-box

The non- and planar double boxes in the near–off-shell kinematics were calculated recently in Ref. [50]. The asymptotically canonical bases for the families of graphs in Fig. 2 (c) and (d) consist of 62 and 97 elements, respectively. Thus, all we need to do in order to evaluate the integrals in Eqs. (4.3) and (4.4) is to perform their IBP reduction to the canonical elements constructed in [50]. This task is elementary making use of the `FIRE` code and we found

$$\mathrm{DBox} = -\frac{1}{2\varepsilon^4}\left[\frac{w}{v}J_{54} + \frac{v}{u+w}J_{55} - J_{59} - J_{61}\right]\,, \tag{4.69}$$

and

$$\mathrm{NBox} = -\frac{1}{4\varepsilon^4}\left[\frac{1585645}{22176}J_1 - \frac{949}{528}J_2 + \frac{72337}{1848}J_3 - \frac{4403017}{22176}J_4 + \frac{7045}{528}J_5 - \frac{1153}{88}J_8\right.$$

$$
\begin{aligned}
&- \frac{1715}{88} J_9 - \frac{589}{44} J_{11} + \frac{3406615}{5544} J_{12} - \frac{58561}{11088} J_{13} - \frac{84}{11} J_{14} - \frac{1/22}{J}_{15} + \frac{6794}{63} J_{17} \\
&+ \frac{8495}{126} J_{18} - \frac{688}{21} J_{19} - \frac{12673}{3696} J_{20} - \frac{42541}{308} J_{21} - \frac{64103}{3696} J_{22} - \frac{2}{3} J_{23} - \frac{793}{44} J_{24} \\
&- \frac{78637}{308} J_{25} + \frac{55897}{1848} J_{26} + \frac{5064085}{22176} J_{27} + \frac{799}{352} J_{28} - \frac{42905}{7392} J_{29} + J_{33} - 8 J_{34} \\
&- \frac{5}{2} J_{35} + \frac{7}{2} J_{36} + 6 J_{37} - 4 J_{38} + 3 J_{39} + 15 J_{40} - 2 J_{41} + \frac{1}{6} J_{43} - \frac{17}{6} J_{44} - \frac{71}{21} J_{45} \\
&- \frac{82}{7} J_{46} - J_{47} + \frac{788}{21} J_{48} + \frac{464}{21} J_{49} + 8 J_{50} - 2 J_{54} + \frac{5}{2} J_{55} - \frac{13}{2} J_{56} - \frac{176}{21} J_{57} \\
&- \frac{281}{21} J_{58} - \frac{667}{21} J_{59} + \frac{352}{21} J_{60} - 2 J_{61} - 2 J_{62} + 6 J_{63} + 11 J_{64} + J_{65} + 7 J_{66} \\
&- J_{67} + 6 J_{68} + 2 J_{69} + 4 J_{70} - \frac{122711}{924} J_{71} - \frac{60253}{462} J_{72} + \frac{271}{132} J_{73} - \frac{1}{3} J_{74} \\
&- \frac{82953}{308} J_{75} + \frac{118609}{462} J_{76} - \frac{17}{44} J_{77} - \frac{2w}{v} J_{78} + \frac{2v}{u+w} J_{79} + \frac{2w}{u+v} J_{80} \\
&- \frac{2(v+w)}{w} J_{81} - \frac{3}{11} J_{82} + \frac{3}{22} J_{83} - \frac{25}{22} J_{84} + \frac{25}{22} J_{85} - 4 J_{86} + \frac{25}{22} J_{87} \\
&+ \frac{8(u+v+w)}{v+w} J_{88} + \frac{21}{22} J_{89} + \frac{1}{44} J_{90} - \frac{25}{132} J_{91} + \frac{4}{3} J_{92} - \frac{25}{66} J_{93} + \frac{25}{132} J_{94} - J_{95} + J_{96} \Big] ,
\end{aligned}
\tag{4.70}
$$

for the planar and non-planar graphs in Figs. 2 (c) and (d), respectively. These are way too lengthy to be presented in the explicit form in the body of the paper. Therefore, for the reader's convenience, we spell them out in the `Mathematica` notebook `Integrals.nb` attached with this submission.

Finally, the nonplanar tri-box in Fig. (2) (e) is just one of the MIs in the nonplanar doublebox basis, namely,

$$
\text{NTriBox} = \frac{2(u+v+w)}{\varepsilon^4 (v+w)} J_{88} \,. \tag{4.71}
$$

Of course, this graph was calculated in Ref. [60], where it was found into factorize after a Fourier transform to the square of Φ_1:

$$
\text{NTriBox} = \frac{1-u}{2} [\Phi_1(m,u)]^2 \,. \tag{4.72}
$$

We indeed confirmed our agreement with it on the constraint (2.3), $u+v+w = 1 + O(m)$. This concludes our calculation of contributing two-loop graphs. All of the integrals reported in this section are UT, however, none are pure.

5 Adding things up

Finally, we are in a position to add up all of the calculated integrals.

5.1 Infrared exponentiation and general structure

As we alluded to in the introduction, we anticipate [25–27] that the infrared logarithms, i.e., $\log m$, exponentiate such that the form factor takes the form

$$\log F_3 = -\frac{\Gamma_{\rm oct}(g)}{4}\left[\log^2\left(\frac{m}{u}\right)+\log^2\left(\frac{m}{v}\right)+\log^2\left(\frac{m}{w}\right)\right]+\mathrm{Fin}_3\left(u,v,w;g\right)+O(m^2)\,, \quad (5.1)$$

with $\Gamma_{\rm oct}$ being the octagon anomalous dimension [28–30] and Fin_3 being a finite part: it depends only on scalar products of momenta of external states and the 't Hooft coupling constant g. It also depends on the type of the operator insertion in (1.1) as well as helicities of external states. Fin_3 develops a perturbative expansion

$$\mathrm{Fin}_3 = g^2 f_3^{(1)} + g^4 f_3^{(2)} + \dots\,. \quad (5.2)$$

The infrared exponent $\Gamma_{\rm oct}$ is known exactly to all orders in the coupling g and is given by [29]:

$$\Gamma_{\rm oct}(g) = -\frac{2}{\pi^2}\log\cosh\left(2\pi g\right) = 4g^2 - 16\zeta_2 g^4 + \dots\,. \quad (5.3)$$

Here, we expanded it to the first two orders, relevant for our current study. We would like to point out the absence in Eq. (5.1) of linear powers in $\log m$ in contrast to the kinematical regime considered in Refs. [1, 15–17], where all external particles' momenta were strictly massless and states propagating in loops' perimeters where taken massive[4]: there is no analogue of the collinear anomalous dimension in the off-shell regime!

The expansion of $\log F_3$ in powers of g is given by

$$\log F_3 = g^2 F_3^{(1)} + g^4\left(F_3^{(2)} - \frac{1}{2}[F_3^{(1)}]^2\right) + \dots\,, \quad (5.4)$$

and can be matched onto the expressions for $F_3^{(1)}$ and $F_3^{(2)}$ in terms of scalar integrals given by (3.1) and (4.6), respectively. Focusing on the infrared divergent part first, we combine the integrals computed above to find

$$\log F_3\Big|_{\rm div} = [-3g^2 + 12\zeta_2 g^4 + \dots]\log^2 m + [2g^2 - 8\zeta_2 g^2 + \dots]\log m \log(uvw)\,, \quad (5.5)$$

in full agreement with our expectation (5.1).

Several comments are in order. Individual two-loop integrals in (4.6) contain $\log^4 m$ as well $\log^3 m$ terms. They cancel, however, in the difference between $F_3^{(2)}$ and the square of one-loop form factor $F_3^{(1)}$ in the $O(g^4)$ coefficient in (5.4). Individual two-loop integrals, i.e., coefficients accompanying the powers of $\log m$, are, in general, expressed in terms of multiple polylogarithms [56]. As can be seen in attached `Mathematica` notebook `Integrals.nb`, the

[4]The relation (5.1) is, strictly speaking, a conjecture supported by an array of explicit computations [25–27] as well as a general intuition about IR properties of gauge theories [19–21].

coefficients of $\log^2 m$ and $\log m$ in (5.5) are determined solely by ordinary logarithms. To observe the cancellations of higher powers of the infrared logarithms as well as simplifications of $\log^2 m$ and $\log m$ terms in (5.5) we used a combination of the symbol map [61, 62] along with high-precision numerical computations offered by the `GiNaC` integrator [63] through the interactive `Ginsh` environment of the `PolyLogTools` package [64]. As we emphasized in earlier sections, individual two-loop integrals are not pure UT functions. They, however, do neatly combine into a pure UT expression when collected together in $F_3^{(2)}$.

5.2 Finite part

Let us now move on to the finite part Fin_3. From Eq. (3.1) it is easy to see that at one loop we have

$$\begin{aligned} f_3^{(1)}(u,v,w) = &- \log u \log v - \log v \log w - \log w \log u \\ &- 2\text{Li}_2(1-u) - 2\text{Li}_2(1-v) - 2\text{Li}_2(1-w) - 3\zeta_2 \,. \end{aligned} \tag{5.6}$$

The two-loop finite part $f_3^{(2)}$ is given by the $\log m$-free term of the $O(g^4)$ coefficient[5] in (5.4). It is a complicated combination of multiple polylogarithms of weight 4. On the route to simplify this expression, it is instructive to consider its symbol map first [62]. Using the `PolyLogTools`, we found out that the symbol of $f_3^{(3)}$ is given by

$$\begin{aligned} \mathcal{S}[f_3^{(3)}] = &- 2u \otimes (1-u) \otimes (1-u) \otimes \frac{1-u}{u} + u \otimes (1-u) \otimes u \otimes \frac{1-u}{u} \\ &- u \otimes (1-u) \otimes v \otimes \frac{1-v}{v} - u \otimes (1-u) \otimes w \otimes \frac{1-w}{w} \\ &- u \otimes v \otimes (1-u) \otimes \frac{1-v}{v} - u \otimes v \otimes (1-v) \otimes \frac{1-u}{u} \\ &+ u \otimes v \otimes w \otimes \frac{1-u}{u} + u \otimes v \otimes w \otimes \frac{1-v}{v} \\ &+ u \otimes v \otimes w \otimes \frac{1-w}{w} - u \otimes w \otimes (1-u) \otimes \frac{1-w}{w} \\ &+ u \otimes w \otimes v \otimes \frac{1-u}{u} + u \otimes w \otimes v \otimes \frac{1-v}{v} \\ &+ u \otimes w \otimes v \otimes \frac{1-w}{w} - u \otimes w \otimes (1-w) \otimes \frac{1-u}{u} \\ &+ \text{ cyclic permutations} \,. \end{aligned} \tag{5.7}$$

This symbol is identical to the symbol of a local function of the following combination of logarithms and classical polylogarithms:

$$R_3^{(2)}(u,v,w) = -2\left[\text{J}\left(-\frac{uv}{w}\right) + \text{J}\left(-\frac{vw}{u}\right) + \text{J}\left(-\frac{wu}{v}\right)\right] - 8\sum_{i=1}^{3}\left(\text{Li}_4\left(1-u_i^{-1}\right) + \frac{\log^4 u_i}{4!}\right)$$

[5] Less $4\zeta_2\left(\log^2(u) + \log^2(v) + \log^2(w)\right)$ due to our definition of the divergent part which includes a finite term as well.

$$- 2\left(\sum_{i=1}^{3}\mathrm{Li}_2(1-u_i^{-1})\right)^2 + \frac{1}{2}\left(\sum_{i=1}^{3}\log^2 u_i\right)^2 - \frac{\log^4(uvw)}{4!}\,, \tag{5.8}$$

with the $\mathrm{J}(z)$ function defined as

$$\mathrm{J}(z) = \mathrm{Li}_4(z) - \log(-z)\mathrm{Li}_3(z) + \frac{\log^2(-z)}{2!}\mathrm{Li}_2(z) - \frac{\log^3(-z)}{3!}\mathrm{Li}_1(z) - \frac{\log^4(-z)}{48}\,. \tag{5.9}$$

Here for brevity of the presentation, we employed the set of variables $u_1 \equiv u$, $u_2 \equiv v$ and $u_3 \equiv w$. The $R_3^{(2)}$ function was first uncovered in the computation of the finite part of the three-gluon form factor in the conformal regime [31], i.e., at the origin of the moduli space of $\mathcal{N}=4$ sYM. However, numerical evaluations of $f_3^{(2)}$ and $R_3^{(2)}$ in several kinematical points clearly indicate that they are different and the difference is not a constant. This is not surprising given that the symbol map is blind to terms such as $\pi^2 \times \text{function}(u,v,w)$. We have constructed an ansatz of all possible terms[6] of the form $\pi^2 \times \{\log(x_i)\log(x_j),\, \mathrm{Li}_2(x_i),\, \pi^2\}$ with rational coefficients plus $R_3^{(2)}$. The values of x_i were taken from the following list

$$\left\{u, v, w, 1-u, 1-v, 1-w, 1-\frac{1}{u}, 1-\frac{1}{v}, 1-\frac{1}{w}, -\frac{uv}{w}, -\frac{vw}{u}, -\frac{wu}{v}\right\}\,. \tag{5.10}$$

Evaluating numerically our ansatz and $f_3^{(2)}$ in several kinematical points using the `Ginsh` integrator allowed us to unambiguously fix these coefficients, and we arrived at

$$\begin{aligned} f_3^{(2)}(u,v,w) = R_3^{(2)}(u,v,w) &+ 3\zeta_2\left[\log(u)\log(v) + \log(v)\log(w) + \log(w)\log(u)\right] \\ &- 4\zeta_2\sum_{i=1}^{3}\mathrm{Li}_2\left(1-u_i^{-1}\right) + 9\zeta_2\sum_{i=1}^{3}\log^2 u_i + \frac{63\zeta_4}{4}\,. \end{aligned} \tag{5.11}$$

This concludes our calculation of the finite part at the two-loop order. We see that it is a pure function of uniform transcendentality just as in the conformal case.

5.3 Iterative structure

In the massless case of scattering amplitudes, it became customary to split results according to the so called BDS ansatz [18] and a finite remainder [65, 66]. The same decomposition was established for the case of form factors as well [31]. Such a decomposition admits the following generic from

$$F_3^{(2)} = \frac{1}{2}\left[F_3^{(1)}\right]^2 + 4\zeta_2\widetilde{F}_3^{(1)} + \mathcal{R}_3^{(2)}\,. \tag{5.12}$$

In the massless case, $\widetilde{F}_3^{(1)}$ was found to enjoy a very powerful feature, namely, it was determined at two loops to be merely given by the one-loop form factor [31]

$$\widetilde{F}_3^{(1)} = \frac{1}{4}F_3^{(1)}(2\varepsilon)\,, \tag{5.13}$$

[6] Taking into account cyclic symmetry as well as functional relations between Li_2 reduces the number of terms in the ansatz quite significantly.

where the factor of $\frac{1}{4}$ is introduced to accommodate the change from $\Gamma_{\rm oct}$ to $\Gamma_{\rm cusp}$ of the massless case. This is the well-known cross-order relation [18] encoding the iterative structure of massless amplitudes. It was also confirmed on the Coulomb branch where the external legs were kept massless [16].

Adopting the same nomenclature in the current 'off-shell' case, we find that $\widetilde{F}_3^{(1)}$ possesses all of the building blocks of the one-loop form factor $F_3^{(1)}$ but is not directly related to it except for the infrared-divergent terms. It has the form

$$\begin{aligned}\widetilde{F}_3^{(1)} &= 3\log^2 m - 2\log m \log uvw \\ &\quad + \frac{3}{4}\big[\log^2 u + \log^2 v + \log^2 w + \log u \log v + \log u \log w + \log v \log w\big] \\ &\quad + \mathrm{Li}_2(1-u) + \mathrm{Li}_2(1-v) + \mathrm{Li}_2(1-w)\,,\end{aligned} \tag{5.14}$$

cf. Eq. (5.5), such that $\widetilde{F}_3^{(1)}|_{\rm div} = F_3^{(1)}|_{\rm div}$. With this convention, the 'off-shell' remainder function $\mathcal{R}_3^{(2)}$ is related by a constant shift

$$\mathcal{R}_3^{(2)}(u,v,w) = R_3^{(2)}(u,v,w) + \frac{63\zeta_4}{4}\,. \tag{5.15}$$

to the one of the conformal case[7], $R_3^{(2)}$ [31]! Indeed, we could enforce the same iterative structure of the 'off-shell' form factor as in the conformal case at the expense of changing the remainder function $\mathcal{R}_3^{(2)}$.

6 Conclusion

With this paper, we continued our excursion into the land of the Coulomb branch away from the origin in its moduli space. The object under our study was form factor of the lowest component of the stress-tensor multiplet for three massive W-bosons. We were particularly interested in the asymptotic region of their vanishing masses, $m \to 0$. In this case, the emerging infrared divergences are encoded by the logarithms of m, which replace inverse powers of ε in dimensional regularization. However, this is not to be confused with another use of the Coulomb branch advocated in Ref. [1], as a means to make amplitudes and form factors finite by giving vacuum expectation values to scalars propagating around quantum loops perimeters. In the latter case, it was established that amplitudes and Sudakov form factors echo the well-known infrared behavior of massless scattering amplitudes and form factors with the infrared physics driven by the cusp anomalous dimension. In counter-distinction, we find instead, that like in the case of scattering amplitudes of four- [24] and five W-bosons [25] and the Sudakov form factor of two W-bosons [26, 27], the infrared logarithms are accompanied by a completely different function of the coupling, the octagon

[7]Note that the remainder function in Ref. [31] contains an additive constant $-\frac{23}{2}\zeta_4$, which we did not include in our definition (5.8).

anomalous dimension [28–30]. This reconfirms the role of the latter as the critical infrared exponent of the off-shell kinematics.

Further, the form factor of three W-bosons possesses a nontrivial remainder function. After a proper subtraction of infrared logarithms with judiciously-chosen finite parts, we found it to be identical to the one in the massless case (up to a constant), i.e., the origin of the moduli space. The structure of the collinear limit is however quite different in the two cases. While the massless case inherits its iterative structure in terms of one-loop amplitude/form factor, the case of massive W-bosons is trickier. In order to put it on a firmer foundation, analysis of the five-W amplitude at generic values of Mandelstam-like variables needs to be studied, as opposed to the symmetric point discussed in Ref. [25].

Last but certainly not least is the question of the dual description of scattering amplitudes and form factors on the Coulomb branch. A proposal for an off-shell Wilson loop was put forward in Ref. [67] starting from a higher-dimensional holonomy and dimensionally reducing it down to four-dimensions. However, while the one-loop expectation value for four sites was found to be in agreement with the amplitude of the W-bosons, starting from two loops the two 'observables' started to deviate. The reason for this fact remains obscure. The T-dual gauge theory was chosen to be the conformal $\mathcal{N} = 4$ sYM. Had it rather be something else or one had to use a different variant of dimensional reduction? This question will have to be readdressed in the future.

Acknowledgments

We would like to thank Johannes Henn, Roman Lee and Alexander Smirnov for useful communications and discussions. The work of A.B. was supported by the U.S. National Science Foundation under the grant No. PHY-2207138. The work of L.B. was supported by the Foundation for the Advancement of Theoretical Physics and Mathematics "BASIS". The work of V.S. was supported by the Russian Science Foundation under the agreement No. 21-71-30003 (solving differential equations for Feynman integrals) and by the Ministry of Education and Science of the Russian Federation as part of the program of the Moscow Center for Fundamental and Applied Mathematics under Agreement No. 075-15-2022-284 (applying `FIESTA` for checks and to obtain precision values for integration constants).

References

[1] L. F. Alday, J. M. Henn, J. Plefka and T. Schuster, *Scattering into the fifth dimension of N=4 super Yang-Mills*, *JHEP* **01** (2010) 077, [0908.0684].

[2] K. G. Selivanov, *An Infinite set of tree amplitudes in Higgs-Yang-Mills*, *Phys. Lett. B* **460** (1999) 116–118, [hep-th/9906001].

[3] R. H. Boels, *No triangles on the moduli space of maximally supersymmetric gauge theory*, *JHEP* **05** (2010) 046, [1003.2989].

[4] N. Craig, H. Elvang, M. Kiermaier and T. Slatyer, *Massive amplitudes on the Coulomb branch of N=4 SYM*, *JHEP* **12** (2011) 097, [1104.2050].

[5] L. Brink, J. H. Schwarz and J. Scherk, *Supersymmetric Yang-Mills Theories*, *Nucl. Phys. B* **121** (1977) 77–92.

[6] F. Gliozzi, J. Scherk and D. I. Olive, *Supersymmetry, Supergravity Theories and the Dual Spinor Model*, *Nucl. Phys. B* **122** (1977) 253–290.

[7] Z. Bern, J. J. Carrasco, T. Dennen, Y.-t. Huang and H. Ita, *Generalized Unitarity and Six-Dimensional Helicity*, *Phys. Rev. D* **83** (2011) 085022, [1010.0494].

[8] R. M. Schabinger, *Scattering on the Moduli Space of N=4 Super Yang-Mills*, 0801.1542.

[9] S. J. Gates, M. T. Grisaru, M. Rocek and W. Siegel, *Superspace Or One Thousand and One Lessons in Supersymmetry*, vol. 58 of *Frontiers in Physics*. 1983.

[10] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, *One loop n point gauge theory amplitudes, unitarity and collinear limits*, *Nucl. Phys. B* **425** (1994) 217–260, [hep-ph/9403226].

[11] S. D. Badger, *Direct Extraction Of One Loop Rational Terms*, *JHEP* **01** (2009) 049, [0806.4600].

[12] Z. Bern, L. J. Dixon, D. C. Dunbar and D. A. Kosower, *Fusing gauge theory tree amplitudes into loop amplitudes*, *Nucl. Phys. B* **435** (1995) 59–101, [hep-ph/9409265].

[13] T. Dennen and Y.-t. Huang, *Dual Conformal Properties of Six-Dimensional Maximal Super Yang-Mills Amplitudes*, *JHEP* **01** (2011) 140, [1010.5874].

[14] Z. Bern, J. J. M. Carrasco, L. J. Dixon, H. Johansson and R. Roiban, *The Complete Four-Loop Four-Point Amplitude in N=4 Super-Yang-Mills Theory*, *Phys. Rev. D* **82** (2010) 125040, [1008.3327].

[15] J. M. Henn, S. G. Naculich, H. J. Schnitzer and M. Spradlin, *Higgs-regularized three-loop four-gluon amplitude in N=4 SYM: exponentiation and Regge limits*, *JHEP* **04** (2010) 038, [1001.1358].

[16] J. M. Henn, S. G. Naculich, H. J. Schnitzer and M. Spradlin, *More loops and legs in Higgs-regulated N=4 SYM amplitudes*, *JHEP* **08** (2010) 002, [1004.5381].

[17] J. M. Henn, S. Moch and S. G. Naculich, *Form factors and scattering amplitudes in N=4 SYM in dimensional and massive regularizations*, *JHEP* **12** (2011) 024, [1109.5057].

[18] Z. Bern, L. J. Dixon and V. A. Smirnov, *Iteration of planar amplitudes in maximally supersymmetric Yang-Mills theory at three loops and beyond*, *Phys. Rev. D* **72** (2005) 085001, [hep-th/0505205].

[19] A. H. Mueller, *On the Asymptotic Behavior of the Sudakov Form-factor*, *Phys. Rev. D* **20** (1979) 2037.

[20] L. Magnea and G. F. Sterman, *Analytic continuation of the Sudakov form-factor in QCD*, *Phys. Rev. D* **42** (1990) 4222–4227.

[21] G. F. Sterman and M. E. Tejeda-Yeomans, *Multiloop amplitudes and resummation*, *Phys. Lett. B* **552** (2003) 48–56, [hep-ph/0210130].

[22] A. M. Polyakov, *Gauge Fields as Rings of Glue*, *Nucl. Phys. B* **164** (1980) 171–188.

[23] G. P. Korchemsky and A. V. Radyushkin, *Renormalization of the Wilson Loops Beyond the Leading Order*, *Nucl. Phys. B* **283** (1987) 342–364.

[24] S. Caron-Huot and F. Coronado, *Ten dimensional symmetry of $\mathcal{N} = 4$ SYM correlators*, *JHEP* **03** (2022) 151, [2106.03892].

[25] L. V. Bork, N. B. Muzhichkov and E. S. Sozinov, *Infrared properties of five-point massive amplitudes in $\mathcal{N} = 4$ SYM on the Coulomb branch*, *JHEP* **08** (2022) 173, [2201.08762].

[26] A. V. Belitsky, L. V. Bork, A. F. Pikelner and V. A. Smirnov, *Exact Off Shell Sudakov Form Factor in N=4 Supersymmetric Yang-Mills Theory*, *Phys. Rev. Lett.* **130** (2023) 091605, [2209.09263].

[27] A. V. Belitsky, L. V. Bork and V. A. Smirnov, *Off-shell form factor in $\mathcal{N}$=4 sYM at three loops*, *JHEP* **11** (2023) 111, [2306.16859].

[28] F. Coronado, *Bootstrapping the Simplest Correlator in Planar $\mathcal{N} = 4$ Supersymmetric Yang-Mills Theory to All Loops*, *Phys. Rev. Lett.* **124** (2020) 171601, [1811.03282].

[29] A. V. Belitsky and G. P. Korchemsky, *Exact null octagon*, *JHEP* **05** (2020) 070, [1907.13131].

[30] B. Basso, L. J. Dixon and G. Papathanasiou, *Origin of the Six-Gluon Amplitude in Planar $N = 4$ Supersymmetric Yang-Mills Theory*, *Phys. Rev. Lett.* **124** (2020) 161603, [2001.05460].

[31] A. Brandhuber, G. Travaglini and G. Yang, *Analytic two-loop form factors in N=4 SYM*, *JHEP* **05** (2012) 082, [1201.4170].

[32] L. J. Dixon, A. J. McLeod and M. Wilhelm, *A Three-Point Form Factor Through Five Loops*, *JHEP* **04** (2021) 147, [2012.12286].

[33] L. J. Dixon, O. Gurdogan, A. J. McLeod and M. Wilhelm, *Bootstrapping a stress-tensor form factor through eight loops*, *JHEP* **07** (2022) 153, [2204.11901].

[34] L. J. Dixon, J. M. Drummond and J. M. Henn, *Bootstrapping the three-loop hexagon*, *JHEP* **11** (2011) 023, [1108.4461].

[35] A. Brandhuber, B. Spence, G. Travaglini and G. Yang, *Form Factors in N=4 Super Yang-Mills and Periodic Wilson Loops*, *JHEP* **01** (2011) 134, [1011.1899].

[36] L. V. Bork, D. I. Kazakov and G. S. Vartanov, *On MHV Form Factors in Superspace for N=4 SYM Theory*, *JHEP* **10** (2011) 133, [1107.5551].

[37] N. I. Usyukina and A. I. Davydychev, *An Approach to the evaluation of three and four point ladder diagrams*, *Phys. Lett. B* **298** (1993) 363–370.

[38] N. I. Usyukina and A. I. Davydychev, *Exact results for three and four point ladder diagrams with an arbitrary number of rungs*, *Phys. Lett. B* **305** (1993) 136–143.

[39] V. V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, *Sov. Phys. JETP* **3** (1956) 65–71.

[40] R. Jackiw, *Dynamics at high momentum and the vertex function of spinor electrodynamics*, *Annals Phys.* **48** (1968) 292–321.

[41] P. M. Fishbane and J. D. Sullivan, *Asymptotic behavior of the vertex function in quantum electrodynamics*, *Phys. Rev. D* **4** (1971) 458–475.

[42] A. H. Mueller, *Perturbative QCD at High-Energies*, *Phys. Rept.* **73** (1981) 237.

[43] G. P. Korchemsky, *Sudakov Form-factor in QCD*, *Phys. Lett. B* **220** (1989) 629–634.

[44] K. G. Chetyrkin and F. V. Tkachov, *Integration by parts: The algorithm to calculate β-functions in 4 loops*, *Nucl. Phys. B* **192** (1981) 159–204.

[45] A. V. Smirnov, *Algorithm FIRE – Feynman Integral REduction*, *JHEP* **10** (2008) 107, [0807.3243].

[46] A. V. Smirnov and F. S. Chuharev, *FIRE6: Feynman Integral REduction with Modular Arithmetic*, *Comput. Phys. Commun.* **247** (2020) 106877, [1901.07808].

[47] A. V. Smirnov and M. Zeng, *FIRE 6.5: Feynman Integral Reduction with New Simplification Library*, 2311.02370.

[48] R. N. Lee, *LiteRed 1.4: a powerful tool for reduction of multiloop integrals*, *J. Phys. Conf. Ser.* **523** (2014) 012059, [1310.1145].

[49] J. M. Henn, *Multiloop integrals in dimensional regularization made simple*, *Phys. Rev. Lett.* **110** (2013) 251601, [1304.1806].

[50] A. V. Belitsky and V. A. Smirnov, *Near mass-shell double boxes*, 2312.00641.

[51] C. Meyer, *Transforming differential equations of multi-loop Feynman integrals into canonical form*, *JHEP* **04** (2017) 006, [1611.01087].

[52] C. Meyer, *Algorithmic transformation of multi-loop master integrals to a canonical basis with CANONICA*, *Comput. Phys. Commun.* **222** (2018) 295–312, [1705.06252].

[53] R. N. Lee, *Reducing differential equations for multiloop master integrals*, *JHEP* **04** (2015) 108, [1411.0911].

[54] R. N. Lee, *Libra: A package for transformation of differential systems for multiloop integrals*, *Comput. Phys. Commun.* **267** (2021) 108058, [2012.00279].

[55] K.-T. Chen, *Iterated path integrals*, *Bulletin of the American Mathematical Society* **83** (1977) 831–879.

[56] A. B. Goncharov, *Multiple polylogarithms and mixed Tate motives*, math/0103059.

[57] A. V. Smirnov, N. D. Shapurov and L. I. Vysotsky, *FIESTA5: Numerical high-performance Feynman integral evaluation*, *Comput. Phys. Commun.* **277** (2022) 108386, [2110.11660].

[58] H. Ferguson, D. Bailey and S. Arno, *Analysis of PSLQ, an integer relation finding algorithm*, *Mathematics of Computation* **68** (1999) 351–369.

[59] A. V. Belitsky, A. V. Smirnov and V. A. Smirnov, *MB tools reloaded*, *Nucl. Phys. B* **986** (2023) 116067, [2211.00009].

[60] N. I. Usyukina and A. I. Davydychev, *New results for two loop off-shell three point diagrams*, *Phys. Lett. B* **332** (1994) 159–167, [hep-ph/9402223].

[61] A. B. Goncharov, *A simple construction of Grassmannian polylogarithms*, 0908.2238.

[62] A. B. Goncharov, M. Spradlin, C. Vergu and A. Volovich, *Classical Polylogarithms for Amplitudes and Wilson Loops*, *Phys. Rev. Lett.* **105** (2010) 151605, [1006.5703].

[63] C. W. Bauer, A. Frink and R. Kreckel, *Introduction to the GiNaC framework for symbolic computation within the C++ programming language*, *J. Symb. Comput.* **33** (2002) 1–12, [cs/0004015].

[64] C. Duhr and F. Dulat, *PolyLogTools — polylogs for the masses*, *JHEP* **08** (2019) 135, [1904.07279].

[65] Z. Bern, L. J. Dixon, D. A. Kosower, R. Roiban, M. Spradlin, C. Vergu et al., *The Two-Loop Six-Gluon MHV Amplitude in Maximally Supersymmetric Yang-Mills Theory*, *Phys. Rev. D* **78** (2008) 045007, [0803.1465].

[66] J. M. Drummond, J. Henn, G. P. Korchemsky and E. Sokatchev, *Hexagon Wilson loop = six-gluon MHV amplitude*, *Nucl. Phys. B* **815** (2009) 142–173, [0803.1466].

[67] A. V. Belitsky and V. A. Smirnov, *An off-shell Wilson loop*, *JHEP* **04** (2023) 071, [2110.13206].

arXiv:2402.18429v1 [hep-lat] 28 Feb 2024

Real-time scattering in the lattice Schwinger model

Irene Papaefstathiou,[1,2] Johannes Knolle,[3,2,4] and Mari Carmen Bañuls[1,2]

[1] *Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany*
[2] *Munich Center for Quantum Science and Technology (MCQST), 80799 Munich, Germany*
[3] *Department of Physics TQM, Technische Universität München, James-Franck-Straße 1, D-85748 Garching, Germany*
[4] *Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom*

(Dated: February 29, 2024)

Tensor network methods have demonstrated their suitability for the study of equilibrium properties of lattice gauge theories, even close to the continuum limit. We use them in an out-of-equilibrium scenario, much less explored so far, by simulating the real-time collisions of composite mesons in the lattice Schwinger model. Constructing wave-packets of vector mesons at different incoming momenta, we observe the opening of the inelastic channel in which two heavier mesons are produced and identify the momentum threshold. To detect the products of the collision in the strong coupling regime we propose local quantitites that could be measured in current quantum simulation platforms.

I. INTRODUCTION

Scattering experiments are a well-established tool for probing fundamental physics. In particular, collision experiments allow the production of high energy and rare particles and thereby a study of their interactions. Precise theoretical predictions of such processes, necessary for their interpretation, often involve contributions that can not be extracted from diagrammatic perturbation theory. This is the case, for instance, for hadron collisions, where non-perturbative effects of Quantum Chromodynamics (QCD) may play a significant role [1]. The most powerful tool to address such non-perturbative regimes is lattice gauge theory (LGT), the discrete formulation of gauge field theories [2]. Using advanced numerical methods, like Quantum Monte Carlo [3, 4], LGT has allowed the successful exploration of strong coupling phenomena, such as the hadron spectra in QCD, but real time dynamics represents a challenge. Despite recent progress [5], a precise first-principles calculation of the scattering processes has not yet been possible, one of the reasons that motivates the search for alternative techniques [6].

In recent years quantum methods have unveiled potential alternative ways to explore fundamental physics (see [7–11] for reviews). Their central focus are LGTs, which also appear as effective low-energy descriptions of strongly correlated condensed matter systems [12]. The Hamiltonian version of LGTs then constitutes a natural object for quantum simulations. While a full simulation of QCD is still beyond reach, current efforts are focused on simplified and low-dimensional LGTs. An important development are the quantum-information based tensor network (TN) methods [13–18] which have been successfully applied to study spectral and thermal equilibrium properties of low dimensional LGTs [6, 19–23], in 1+1D cases achieving some of the most precise existing extrapolations to the continuum.

Real time evolution phenomena, such as scattering dynamics, are among the most promising problems for a potential quantum advantage [24], since Monte Carlo methods suffer in this case from a sign problem. Tensor network algorithms, on the other hand, can be used to simulate LGT dynamics, but only for a limited time, before the entanglement in the system becomes too large for an efficient TN description. Yet, out-of-equilibrium simulations of LGT with TN have been used to explore the dynamics of pair production [25] and string breaking dynamics in several LGT models [26–30]. Whereas far from the precision achieved by their equilibrium counterparts and not extrapolated to the continuum limit, such simulations overcome the limits of classical Monte Carlo methods and constitute a valuable tool to prepare and analyze the potential quantum simulations.

More recently, scattering experiments in LGT have attracted the attention of TN calculations and quantum simulation proposals [28, 31, 32]. The basic phenomenology of confinement [33, 34] and meson scattering has also been explored recently in quantum spin models which share properties of LGT especially in 1+1D [35–40].

In this paper we use TN techniques to study elastic and inelastic scattering of composite particles in the lattice Schwinger model, which is the discrete lattice version of Quantum Electrodynamics in one spatial dimension. Due to its similarities with more complex LGTs, including confinement and chiral symmetry breaking, the Schwinger model is a standard testbench for LGT methods, and has been focus of TN simulations, quantum simulation proposals and even experimental realizations (see [7, 41] for reviews). Studies of scattering processes in the Schwinger model have been more scarce.

In absence of a background field, the spectrum of the model contains two stable particles, referred to as *vector* and *scalar*. Elastic processes between two vector mesons, below the threshold of production of the scalar, have been simulated in the strong coupling [28] and in the weak and intermediate coupling [31] regimes. In this regime, the particles collide and can bounce back, without the creation of new particles post-collision, but generating entanglement between the mesons. A first approach to the inelastic regime was explored recently in [32], with a focus on the thermodynamic limit and the non-perturbative regime near the confinement-deconfinement critical point, and using a bosonized formulation of the model, aimed at proposing a quantum simulation scheme that could be

implemented in circuit-QED.

In this work, we focus instead on the strong coupling regime at vanishing background field, and explore the energy threshold required to open the inelastic channel and obtain particle production after the collision. Working in this regime allows us to propose simple initial state preparations and observables that could be amenable to realization on quantum simulators.

The paper is structured as follows: In Sec. II we briefly present the Schwinger model in the continuum limit and we show the lattice discretization we use in this work. In Sec. III we discuss the preparation of the initial state of two meson wavepackets, indicate the momentum threshold for particle production and present the method used for the time evolution of the system. In Sec. IV we show the results of meson-meson collisions. We specifically identify the energy threshold for particle production and we compare our results with the predicted threshold obtained with DMRG calculations.

II. MODEL

The Schwinger model [42] is Quantum Electrodynamics (QED) in one spatial dimension. It is the simplest gauge theory including dynamical matter, and exactly solvable for massless fermions, yet it exhibits common features with those of Quantum Chromodynamics, like confinement and chiral symmetry breaking [43]. Its lattice discretization is commonly used as a benchmark for numerical LGT methods [44, 45], and has in particular been widely adopted in studies of Tensor Network techniques for lattice gauge theories, which started [46] with the application to the model of the Density Matrix Renormalization Group (DMRG) algorithm.

A. The continuum limit

The continuum Hamiltonian of the Schwinger model in the temporal gauge can be written:

$$\begin{aligned}\mathcal{H} = &-\int dx\Big(i\bar{\Psi}(x)\gamma^1\left[\partial_1 + igA_1(x)\right]\Psi(x)\Big)\\ &+\int dx\Big(m\bar{\Psi}(x)\Psi(x) + \frac{1}{2}E(x)^2\Big),\end{aligned} \tag{1}$$

with γ^0 and γ^1 the Dirac matrices in the temporal and spatial dimension, whereas ∂_1 is the partial derivative with respect to x. The fermionic field $\Psi(x)$ has two components and its conjugate is $\bar{\Psi} = \Psi^\dagger\gamma^0$. m is the fermionic mass and g is the coupling constant. In an infinite volume, the only independent parameter of the model is the adimensional ratio m/g. Note that in the temporal gauge $A^0 = 0$. The electric field $E(x)$ is then related to the vector gauge field $A^1(x)$ as $E(x) = -\partial_0 A^1(x)$.

The continuum Schwinger model is exactly solvable for the case $m/g = 0$ and $m/g \to \infty$ [42, 47]. For $m/g = 0$, the Schwinger model can be solved using bosonization and corresponds to a free massive boson, the so-called Schwinger boson. For non-zero m/g, this boson becomes interacting.

We denote the ground state energy as E_{GS}. The first stable particle of the theory is the vector meson, a bound state (Schwinger boson) with mass M_{V}. The second stable particle is the scalar meson, with mass M_{S}, which can be considered a bound state of two Schwinger bosons [43, 48].

These two bound states are stable particles with parity and charge conjugation being well-defined quantum numbers (CP symmetry) [43]. The bound state of two Schwinger bosons belongs in the same sector as the ground state (scalar sector), whereas the Schwinger boson belongs to the vector sector.

In open boundary conditions, integrating Gauss law fixes the electric field up to a constant, which can be interpreted as an external background field $g\theta/2\pi$, with θ the so-called vacuum angle [43, 49]. In the strong coupling (small mass) regime we target in this study, it is possible to estimate the particle masses perturbatively. Specifically, the mass of the mesons M_{V}, in the strong coupling regime can be obtained from second-order perturbation theory as [50, 51]:

$$\begin{aligned}M_{\mathrm{V}}^2 &\equiv \mu_2^2\\ &= \mu_0^2\Big(1 + 3.5621\frac{m}{\mu_0}\cos(\theta)\\ &\quad + 5.4807\frac{m^2}{\mu_0^2} - 2.0933\frac{m^2}{\mu_0^2}\cos(2\theta)\Big)\end{aligned} \tag{2}$$

with $\mu_0^2 = \frac{g^2}{\pi}$ the result from zero-order perturbation theory. The mass of the second excitation of the theory, the bound state M_{S}, is given in second-order perturbation theory by [48]:

$$M_{\mathrm{S}}^2 = 4M_{\mathrm{V}}^2 - \Delta \tag{3}$$

with $\Delta = \frac{4\pi^4 m^2\Sigma^2\cos^2(\theta)}{M_{\mathrm{V}}^2}$. $\Sigma = \frac{g^\gamma}{2\pi}\mu_0$ is the fermion condensate with γ the Euler constant [50].

The form of the wavefuction of the bound states in the strong coupling regime of the continuum limit was previously studied first by Ref. [52] up to and including four-body states and then by Ref. [53] up to and including six-body states with the use of the six-body light-front Tamm-Dancoff approximation. In these works it was found that the composition of the bound states depends on the fermionic mass m/g. They specifically consist of a two-body component, a four-body component and a negligible six-body component. The contribution of the four-body component increases as the fermionic mass m/g is getting smaller.

B. The lattice

This section presents the lattice Schwinger model, the scattering processes of which will be studied in this paper. Here we are using the Kogut-Susskind staggered fermion

formulation for the model [54]:

$$W = \frac{g^2 a}{2}\sum_n L_n^2 + m\sum_n (-1)^n \Phi_n^\dagger \Phi_n, \\ -\frac{i}{2a}\sum_n \left(\Phi_n^\dagger e^{i\theta_n}\Phi_{n+1} - h.c.\right), \quad (4)$$

with the lattice spacing being ga in coupling units. Specifically, we focus on the strong coupling regime, which means $m/g \ll 1$.

The fermionic annihilation(creation) operators $\Phi_n(\Phi_n^\dagger)$ reside on the lattice sites n. As single-component fermionic operators, they satisfy canonical anticommutation relations $\{\Phi_n^\dagger, \Phi_m\} = \delta_{nm}$ and $\{\Phi_n, \Phi_m\} = 0$. The gauge fields θ_n and L_n are the lattice equivalent to the vector potential and the electric field in the continuum. They occupy the links between two nearest neighbouring sites (n and $n+1$). They are canonically conjugate operators, satisfying the commutation relation $[\theta_n, L_m] = i\delta_{nm}$. In addition, the lattice version of the Gauss law reads [45]:

$$L_n - L_{n-1} = \Phi_n^\dagger \Phi_n - \frac{1}{2}\left[1 - (-1)^n\right] \quad (5)$$

and needs to be satisfied by the physical states of the system.

By a Jordan-Wigner transformation, the Hamiltonian (4) can be mapped onto a spin model. The Gauss law, which now reads:

$$L_n - L_{n-1} = \frac{1}{2}\big((-1)^n + (\sigma_n^z)\big), \quad (6)$$

can be integrated out if we consider open boundary conditions, resulting in a spin Hamiltonian with long range interactions [45, 55]. After rescaling it with a global $2/(ag^2)$ factor, one is left with the adimensional Hamiltonian:

$$H = \frac{2}{g^2 a}W = H_x + H_\mu + H_l \\ = x\sum_{n=0}^{N-2}\left[\sigma_n^+\sigma_{n+1}^- + \sigma_n^-\sigma_{n+1}^+\right] \\ + \frac{\mu}{2}\sum_{n=0}^{N-1}\left[1 + (-1)^n\sigma_n^z\right] \\ + \sum_{n=0}^{N-2}\left[l + \frac{1}{2}\sum_{k=0}^{n}((-1)^k + \sigma_k^z)\right]^2, \quad (7)$$

with $x = \frac{1}{g^2a^2}$ and $\mu = \frac{2m}{g^2 a}$ being adimensional parameters and $\sigma^\pm = \frac{1}{2}(\sigma^x \pm i\sigma^y)$. l is the background electric field and is related to the vacuum angle as following: $l = \theta/2\pi$. It is set to zero in this work, $l = 0$.

The lattice model with open boundary conditions breaks most of the continuum symmetries. In particular, charge conjugation is no longer a good quantum number. However, the symmetry is restored in the thermodynamic limit and, if the system is large enough, it is possible to define an approximate charge conjugation operator whose expectation value can be used to distinguish the scalar and vector sectors [56].

1. The strong coupling regime

In this paper we study the spin formulation of the lattice Schwinger model solely focusing on the strong coupling regime. Since here we use the rescaled Hamiltonian of Eq. (7) with the adimensional parameters μ and x, being in the strong coupling regime for a finite lattice means $x \ll 1$. In this regime H_x can be considered the perturbation, with $H_\mu + H_l$ being the unperturbed part of the Hamiltonian. The strong coupling regime offers the advantage of having a simple Ansatz for the initial mesons of the collision, as well as a clear energy threshold for producing the scalar mesons of interest. We note that, a truncated Hilbert space approach, which has been employed for spin chains [39], would not work here.

In the infinite coupling limit, $x = 0$, the ground state of the unperturbed Hamiltonian $H_\mu + H_l$ is the state with all the sites empty:

$$|0\rangle = |\downarrow\uparrow\downarrow\uparrow\cdots\downarrow\uparrow\rangle \quad (8)$$

The sites are numbered from left to right $n = 0, 1, \ldots, N-1$. Due to the fact that we are using the staggered fermions formulation, the odd sites correspond to the fermions and the even sites to the antifermions. Specifically, the occupied sites are $|\downarrow\rangle_O$ for the odd sites and $|\uparrow\rangle_E$ for the even sites ($|\uparrow\rangle_O$ and $|\downarrow\rangle_E$ are the empty odd and even sites correspondingly). Therefore, $|\downarrow\rangle_O$ are the particles (electrons) and $|\uparrow\rangle_E$ are the anti-particles (positrons).

In the limit of infinite coupling, $x = 0$, the first excitation of the theory $|1_V\rangle$ (the vector meson), with energy E_V, can be obtained simply as:

$$|1_V\rangle = \frac{1}{\sqrt{N}}\sum_n \left(\sigma_n^+\sigma_{n+1}^- - \sigma_{n+1}^+\sigma_n^-\right)|0\rangle. \quad (9)$$

For small but non-vanishing values of x, the strong coupling expansion (SCE) [45], a perturbative expansion in x, can be used to obtain very precise estimates of the ground state energy and the mass spectrum, for small bare fermion mass [57].

Our goal is to create states of the mesons with well defined momenta. Although we do not have the explicit solution for the operator that creates these particles, in the strong coupling regime we can use an Ansatz that resembles the first order of SCE (10), but where the strong coupling vacuum has been substituted by the true ground state at finite x, which we denote $|\Omega\rangle$ and compute numerically following the method in [56]. Therefore, our Ansatz for the state of the mesons can be written as:

$$|1_V\rangle = \frac{1}{\sqrt{N}}\sum_n \left(\sigma_n^+\sigma_{n+1}^- - \sigma_{n+1}^+\sigma_n^-\right)|\Omega\rangle. \quad (10)$$

III. METHODS

A. Preparing the initial state

The first step for studying the elastic and inelastic meson-meson collisions is creating two meson wavepackets

in the initial state and giving them opposite momentum so that they eventually collide. We prepare two wavepackets for the mesons as Gaussian superpositions of the operator in Eq. (10) acting on different positions, namely:

$$|\phi_i\rangle = \mathcal{N}_i \sum_{n=\alpha_i^{\rm beg}}^{\alpha_i^{\rm end}} e^{-(n-c_i)^2/(2\sigma^2)} e^{-ink_i} \left(\sigma_n^+\sigma_{n+1}^- - \sigma_{n+1}^+\sigma_n^-\right)|\Omega\rangle = \mathcal{O}_i|\Omega\rangle \tag{11}$$

with indices $i = 1,\ 2$ corresponding to the two mesons. We choose the wavepackets to have centers c_i, and common width σ. Initially, the wavepacket of the first (second) meson is spread in the left (right) half of the system. We therefore have $\alpha_1^{\rm beg} = 0$, $\alpha_1^{\rm end} = N/2 - 1$ for the first meson and $\alpha_2^{\rm beg} = N/2$ and $\alpha_2^{\rm end} = N - 1$ for the second one. We consider the mesons to have opposite momenta with the same magnitude, $k_1 = -k_2 = k$. $\mathcal{N}_i$ are the necessary normalization factors. The total initial state is constructed as the product:

$$|\phi\rangle = (\mathcal{O}_1\mathcal{O}_2)|\Omega\rangle. \tag{12}$$

Notice that this approach is made possible by working in the strong coupling regime, as this provides us with the explicit approximate description of the vector meson state shown in Eq. (10), which can be directly used for the creation of the wavepackets.

B. Momentum threshold $k_{\rm thr}$

In order to open an inelastic channel (i.e. to produce a final state allowed by the symmetries of the problem), the momentum of the incoming wavepackets needs to be high enough to produce the new particles satisfying the kinematic constraints imposed by energy and momentum conservation.

We consider collisions in the center of mass of the incoming mesons, where the initial wavepackets have opposite spatial momenta $p_1 = -p_2 \equiv p$. Momentum conservation imposes that the outgoing particles, two identical scalar mesons, have the same momenta with opposite directions, q and $-q$. Conservation of energy then requires:

$$2\sqrt{M_{\rm V}^2 + p^2} = 2\sqrt{M_{\rm S}^2 + q^2} \tag{13}$$

The final state with the smallest possible energy would correspond to creating the pair of scalars at rest, $q = 0$. This sets the momentum threshold for the incoming vector mesons at:

$$p^2 = p_{\rm thr}^2 = M_{\rm S}^2 - M_{\rm V}^2. \tag{14}$$

Notice that the relations above are written in terms of physical (continuum) momenta. Our initial state (11) is however written in terms of adimensional lattice quantities. In particular, the lattice momentum k is related to the physical one as $k = p/(g\sqrt{x})$.

Because we work on a finite lattice with open boundary conditions, the scalar mesons cannot be produced at rest, but will carry at least the smallest lattice momentum $k_{\rm min} = O(1/N)$, which will correct the threshold.

In the continuum limit, one could calculate the momentum threshold with the use of perturbation theory and Eq. (2) and Eq. (3). Since in our parameter regime, $N = 100$, $x = 1$ and $\mu = 2 \times 10^{-5}$, we are far away from that limit, we instead calculate the masses using a variational optimization of MPS, with the algorithm in ref. [56] for which we obtain the masses $M_V/g = 0.785$, $M_S/g = 1.370$. With these values we estimate the lattice momentum threshold at:

$$k_{\rm thr}^{\rm MPS} \approx 1.123. \tag{15}$$

C. Time Evolution

In our work, we use TN methods to prepare the initial state and simulate the time evolution. Firstly, we find a matrix product state [14] approximation to the interacting vacuum $|\Omega\rangle$ by optimizing variationally the MPS ansatz that minimizes the energy, using the method in [56]. We truncate the bond dimension of the ground state to $D_{\rm GS} = 40$. The operators $\mathcal{O}_1$ and $\mathcal{O}_2$ of Eq.(11) can be written as exact matrix product operators (MPOs) [58] with bond dimension $D = 4$. They can be then applied to the ground state, $|\Omega\rangle$ to create the meson wavepackets of Eq. (11). We truncate the bond dimension of the wavepackets to $D_{\rm V} = 50$, which, when compared to the exact initial state of meson wavepackets (with bond dimension $4D_{\rm GS}$), has fidelity very close to 1 (with negligible corrections).

After creating the initial state with the two meson wavepackets, we perform time evolution with standard tensor network techniques [13, 14]. Specifically, we use a second order Suzuki-Trotter expansion of the evolution operator [59, 60] with time step δ, using the splitting of the Hamiltonian described in [61], namely:

$$e^{-i\tau H} \approx \left(e^{-i\frac{\delta}{2}H_x^e} e^{-i\frac{\delta}{2}H_z} e^{-i\delta H_x^o} e^{-i\frac{\delta}{2}H_z} e^{-i\frac{\delta}{2}H_x^e}\right)^{\tau/\delta}. \tag{16}$$

H_x^e and H_x^o are the hopping terms (part of H_x) acting on the two-body terms of even-odd and odd-even pairs of sites correspondingly. H_z is the mass term of the Hamiltonian together with the electric field term, $H_z = H_\mu + H_l$.

The exponentials $e^{-i\frac{\delta H_x^e}{2}}$ and $e^{i\frac{-\delta H_x^o}{2}}$ can be written as exact MPOs with bond dimension 4. The exponential of the term H_z can also be written as an exact MPO, diagonal in the z basis, with bond dimension that scales linearly with the system size as $\chi = N + 1$ [61].

Specifically, the elements of the MPO are:

$$(M_n^{ij})_{L_{n-1}L_n} = e^{-i\delta h_n(i)}\delta_{ij}\delta_{L_n - L_{n-1}, G_n(i)}, \tag{17}$$

where the virtual indices label the electric flux on the links, $L_n \in [-N/2, N/2]$ and the second delta ensures the fulfilment of Gauss law (6),

$$G_n(i) = \frac{1}{2}[(-1)^n + (-1)^i], \quad i = 0, 1, \tag{18}$$

and $h_n(i) = \mu(-1)^n G_n(i) + L_n^2$ for $n < N-1$, with $h_{N-1}(i) = \mu(-1)^{N-1} G_{N-1}(i)$.

To be able to deal with large system sizes (of the order of 100 or more sites) it is convenient to truncate the bond dimension of this MPO to a constant value $\chi_{\rm tr}$, which can be related to a cutoff in the absolute value of the electric flux $L_{\rm max}$, as $\chi_{\rm tr} = 2L_{\rm max}+1$. This truncation, which was successfully used for thermal equilibrium states in [61, 62], will be also valid in our case as far as the dynamics does not generate states with large occupations of the electric flux. The cutoff in the electric flux is restricted to $L_{\rm max} = 8$ in our study. Applying the evolution operator increases the bond dimension of the time-evolved state $|\Psi(t)\rangle$ at each time step. In order to control the computational cost, we truncate the bond dimension of the evolved state to a maximum value of $D = 50$. Even though we use relatively small bond dimensions for the ground state, the initial mesons and the evolved state, as well as a relatively small cutoff of the electric flux, we check that our qualitative observations are stable under changes of the truncation parameters, and this level of precision suffices to observe the effects of inelastic scattering in the cases under study. Specifically, knowing that the total energy has to be conserved in our system, we control the errors by measuring the (relative) change of the total energy in comparison to the initial total energy of the system. We find that for the times of interest (up to $gt_{\rm phys} = 140$) the relative difference is less than 6%.

IV. RESULTS

In this section we present the results of elastic and inelastic meson-meson scattering processes. Specifically, in subsection IV A we observe the first signs of particle production, when having energies above the momentum threshold of Eq. (15) (inelastic channel). In subsection IV B we calculate quantities that can be directly measured in experiments, indicating the signal of particle production for the inelastic channel. In subsection IV C we study the four-body projector.

The results shown in this section correspond to a fixed fermion mass $m/g = 10^{-5}$ and physical volume $N/\sqrt{x} = 100$ (notice that we express all magnitudes in units of the coupling g).

These magnitudes are related to their corresponding lattice analogues through the lattice spacing $ga = 1/\sqrt{x}$. In particular, the time parameter τ appearing in the evolution operator $e^{-iH\tau}$ (16) is the adimensional lattice time. It is related to the physical time $gt_{\rm phys}$ as:

$$\tau = \frac{gt_{\rm phys}}{2\sqrt{x}} \tag{19}$$

The lattice size was fixed to $N = 100$ sites, and the inverse lattice spacing to the value $x = 1$ in all the cases below.

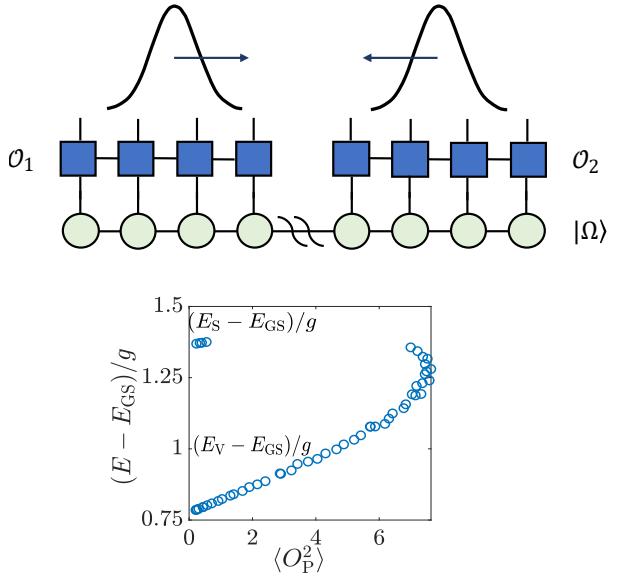


FIG. 1. The upper panel shows the initial state preparation following Eq. (12). The lower panel shows the dispersion relation of the adimensional Hamiltonian of Eq. (7) as $E = f(\langle O_{\rm P}^2 \rangle)$, with $O_P = -ix\sum_n(\sigma_n^- \sigma_{n+1}^z \sigma_{n+2}^+ - h.c.)$ the dimensionless momentum operator for the fermion field (see [56]). The lower branch corresponds to the vector meson and the upper one to the scalar, for which only the lowest momentum excitations are shown. The ground state energy, $E_{\rm GS}/g$, is subtracted from all the energies plotted. We consider a finite system with lattice size $N = 100$ and physical volume $N/\sqrt{x} = 100$. The mass is $m/g = 10^{-5}$. The energies were calculated with variational MPS as in [56] with small bond dimension, $D = 50$.

A. Entropy

We first calculate the entanglement entropy $S(n,t)$ corresponding to a bipartition where one subsystem includes two neighboring sites n and $n+1$, with $n = 0, \dots N-2$, and the other subsystem consists of the remaining sites:

$$S(n,t) = -\,{\rm Tr}[\rho_{n,n+1}(t)\log_2 \rho_{n,n+1}(t)], \tag{20}$$

with $\rho_{n,n+1}$ being the reduced density matrix of one of the subsystems. $S(n,t)$ is interpreted as the entanglement between the sites $\{n, n+1\}$ and the rest of the system. Since the two-site reduced density matrix can in principle be recovered from local measurements, as recently realized in optical lattices in [63], this quantity could be more easily accessible in an experiment than the half-chain entanglement that has been analyzed in other scattering simulations [28, 31, 32].

The results of the calculation of the entanglement entropy can be seen in Fig. 2 and in Fig. 3. We study three different cases depending on the magnitude of the momenta of the incoming mesons. The three cases consider momenta below, slightly above and above the momentum threshold $k_{\rm thr}^{\rm MPS}$. The first case is the case where the momenta are below the momentum threshold, as seen in the left plot of Fig. 2. We specifically have lattice momenta $k_1 = -k_2 = k = 1$. Here, the incoming mesons do not have enough energy to create new particles. Therefore,

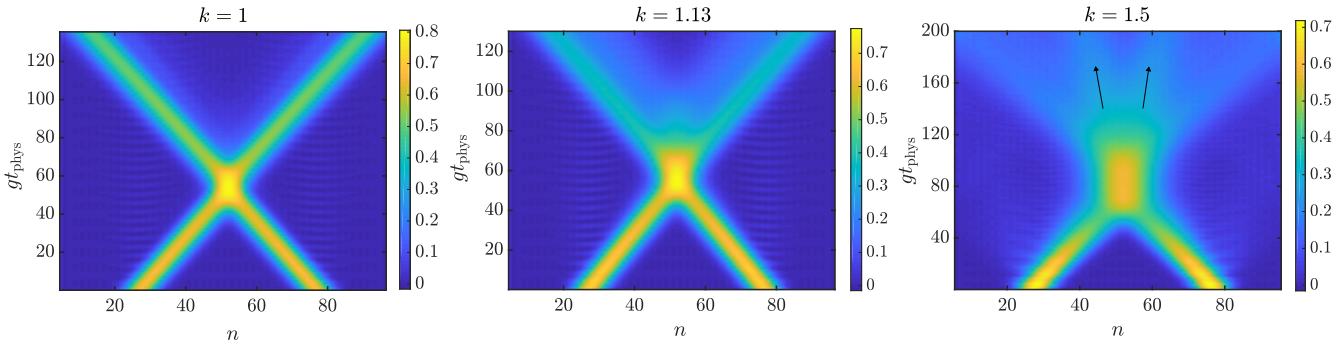


FIG. 2. The plots show the two-site entanglement entropy $S(n,t)$, defined in the main text (20), with the entanglement entropy of the vacuum subtracted from all the plots. The fermion mass is fixed to $m/g = 10^{-5}$. We consider a finite system with lattice size $N = 100$ and physical volume $N/\sqrt{x} = 100$. The standard deviation of the wavepackets is $\sigma = 4$. In the left plot, the momentum magnitude of the incoming mesons is $k = 1$ and only the elastic channel is present. The elastic channel corresponds to two mesons exiting the collision (yellow colour) with the same momentum as the momentum of the incoming mesons. In the middle plot, the momentum of the incoming mesons is $k = 1.13$, slightly above the threshold predicted by the variational MPS calculation $k_{\rm thr}^{\rm MPS}$ (15). At this point, the inelastic channel opens, with two bound states produced with velocities close to 0. The signal of the two bound states that were created after the collision is visible a bit below $gt_{\rm phys} \sim 80$, with the two outgoing mesons coming from the elastic channel also being visible. The momentum of the incoming mesons for the plot on the right is $k = 1.5$, also above the momentum threshold. Both the elastic and inelastic channels are open. The two bound states created from the collision (inelastic channel) have non-zero velocities with opposite signs and less magnitude than the momenta of the outgoing mesons of the elastic channel. The elastic channel is represented by the light blue cone-like shape, after the collision, whereas the velocities of the bound states are shown with the two arrows. The parameters used in the numerical simulation were bond dimension $D = 50$, Trotter step $\delta = 0.1$ and maximum electric flux $L_{trunc} = 8$.

only the elastic channel is present and the outgoing particles are two vector mesons with the same momenta as the incoming ones. Fig. 3 shows the excess entanglement entropy with respect to the vacuum after the collision, for physical time $gt_{\rm phys} = 90$ (blue line). The two peaks represent the wavepackets of the two outgoing mesons. Notice the small increase in the entanglement between the two wavepackets in comparison to the entanglement entropy of the vacuum. Because the local entropy we compute is not directly measuring entanglement between the outgoing mesons, we attribute this increment to the fluctuations imprinted onto the vacuum by the collision.

The second case is the middle plot of Fig. 2, where the incoming particles have enough energy to produce two scalar mesons of mass M_S with zero velocity. Specifically, as mentioned in Sec. III A in Eq. 15, the threshold for particle production is at $k_{\rm thr}^{\rm MPS} \approx 1.12$. As the initial momenta of the vector mesons is $k = 1.13$, we expect producing scalar mesons that are almost at rest after the collision. Notice here that it is not possible to create particles with momentum exactly 0, because as mentioned in Sec. III B, the scalar mesons have to have at least momentum $k_{\rm min} = O(1/N)$. One could observe that after the collision ($t_{\rm phys} \sim 80$) there is a sign of particle production in the middle of the system, where the entanglement entropy is greater than the entanglement entropy between the outgoing particles of the elastic channel. This signal can be more clearly observed in Fig. 3 (orange line). The peak in the middle of the system corresponds to the two bound states produced with almost zero velocities. The higher peaks on the left and right of this signal correspond to the outgoing mesons of the elastic channel.

Finally, the right plot of Fig. 2 represents the case where the momentum of the incoming particles is considerably

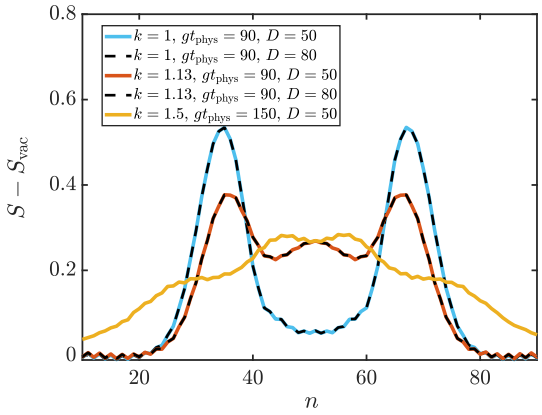


FIG. 3. The plot shows the two-site entanglement entropy $S(n,t)$ (20) minus the entanglement entropy of the vacuum, S_{vac}, on a lattice with size $N = 100$ and physical volume $N/\sqrt{x} = 100$, for the cases where the initial particles have momenta below, slightly above and above the momentum threshold, at $k = 1$, $k = 1.13$ and $k = 1.5$. The mass is $m/g = 10^{-5}$, and the standard deviation of the wavepackets $\sigma = 4$. As in Fig. 2, the parameters used in the numerical simulation were bond dimension $D = 50$, Trotter step $\delta = 0.1$ and maximum electric flux $L_{trunc} = 8$. To demonstrate the numerical convergence, for the cases $k = 1$ and $k = 1.13$ we show also the results for $D = 80$ (black dash lines) while keeping all the other parameters the same.

larger than the threshold, such that they have enough energy to produce two scalar particles of mass M_S, and the excess is transformed in kinetic energy of the outgoing mesons. In this case, we can notice both the elastic and the inelastic channels. Since the scalar mesons are produced with small velocities, their signal is found near the

middle of the system, as also seen by Fig 3. One could also observe the signal of the elastic channel, near the edges of the same plot.

We notice here that the momentum threshold for particle production calculated from the variational MPS spectrum calculations was $k_{\rm thr}^{\rm MPS} \approx 1.12$. The simulated collisions behave as expected, with the first particle production observed at $k = 1.13$, and no signs of particle production at $k = 1$. We could also compare with the momentum threshold predicted by perturbation theory in the continuum limit, $p_{\rm thr}^{\rm continuum}/g \approx 0.98$, which was done with the use of Eq. (14), Eq. (2) and Eq. (3).

B. Correlators

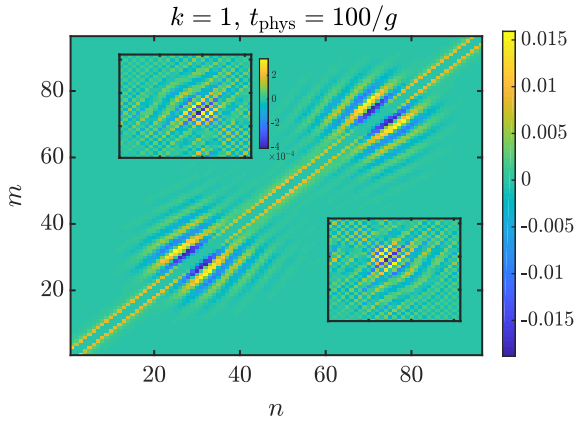


FIG. 4. This plot shows the electric flux correlator, C_l^{mn}, of a system size $N = 100$, corresponding to physical volume $N/\sqrt{x} = 100$ at lattice spacing $x = 1$, for times after the meson-meson collision, $gt_{\rm phys} = 100$. The initial momentum of the mesons, $k = 1$, is below the momentum threshold. The fermion mass is $m/g = 10^{-5}$, and the standard deviation of the initial wavepackets $\sigma = 4$. We set $C^{mn} = 0$ for $|m - n| \leq 1$, so that $L^{m,n}$ do not overlap. The insets show the enclosed area rescaled by a factor, to highlight the signal of the correlations between the two outgoing meson wavepackets of the elastic channel. The bond dimension used for the calculations is $D = 50$ and the maximum electric flux $L_{trunc} = 8$.

In this subsection we plot the electric flux correlator,

$$\mathcal{C}^{mn}(t) = \langle L^m L^n \rangle - \langle L^m \rangle \langle L^n \rangle \tag{21}$$

with $L^n = \frac{1}{2}\sum_{k=0}^{n}((-1)^k + \sigma_k^z)$. In principle, we can expect that this quantity can be measured in experiments, showing the signals of particle production.

The plots in this section show the electric flux correlators, $C^{mn}(t)$, specifically for times after the collision. Fig. 4 shows $C^{mn}(t)$ for $gt_{\rm phys} = 100$ for the case in which only the elastic channel appears, with the incoming mesons having momentum $k = 1$. As seen in the left plot of Fig. 2, the outgoing mesons appear to be occupying the lattice sites $20 - 40$ and $60 - 80$. This is verified by Fig. 4, where the two oval shapes around the diagonal correspond to the two outgoing mesons of the elastic channel and are concentrated around the same positions.

In Fig. 6 we plot the correlator for $gt_{\rm phys} = 180$ and initial meson momenta above the momentum threshold, at $k = 1.5$. In this case both the elastic and inelastic channel are present. The inelastic channel appears as the yellow signal around the main diagonal, with the two bound states visible around sites $40 - 50$ and $60 - 70$.

C. Four-body projector

As seen in the subsections above, two different particles are produced when the momenta of the initial mesons are above the momentum threshold $k_{\rm thr}^{\rm MPS} \approx 1.12$. These are scalar mesons, which can be interpreted, as discussed in section II A as bound states of two Schwinger bosons.

Ideally, to unambiguously identify the particles produced as scalar mesons, we would like to measure the corresponding scalar operator, which is not known, in general. However, in the strong coupling regime, $m/g \to 0$, the continuum theory results ensure that the composition of the scalar meson in that case is dominated by four fermion terms, whereas the vector meson is dominated by two-fermion terms [52, 53]. Even though we are far from that limit and both the ground states and excitations are strongly renormalised, we propose probing the appearance of the scalar mesons by calculating the time-dependent expectation value of the four-body projector:

$$P_n(t) = \langle \Pi_\uparrow^n \Pi_\downarrow^{n+1} \Pi_\uparrow^{n+2} \Pi_\downarrow^{n+3} \rangle|_{n=even} \tag{22}$$

with $\Pi_\uparrow^n$, $\Pi_\downarrow^n$ the projectors:

$$\begin{aligned} \Pi_\uparrow^n|\uparrow\rangle &= |\uparrow\rangle \\ \Pi_\downarrow^n|\downarrow\rangle &= |\downarrow\rangle \end{aligned} \tag{23}$$

and $\Pi_\uparrow^n|\downarrow\rangle = 0$, $\Pi_\downarrow^n|\uparrow\rangle = 0$. This will give a signal that the time-evolved state has a four-body component iff all the consecutive sites n, $n+1$, $n+2$ and $n+3$ are occupied. The underlying assumption is that the wave functions are continuously connected to the strong-coupling regime, such that we can identify its signal as the presence of the scalar.

The results shown in Fig. 5 give yet another indication of the presence of the bound states. Specifically, in the left plot of Fig. 5 ($k = 1.13$) one observes the first signal of bound state production (yellow colour). This corresponds to two scalar mesons produced almost at rest (approximately zero velocities). The right plot of Fig. 5 shows the case where the incoming mesons have momentum $k = 1.5$. After the collision, one observes signals of two bound states with opposite velocities (yellow colour).

V. DISCUSSION

In this work we have used TNS methods to simulate scattering processes in the lattice version of the Schwinger model. More concretely, we have studied the scattering in the center of mass of two vector mesons (composite particles) at different incident momenta, and have been able to observe the threshold at which the inelastic channel opens

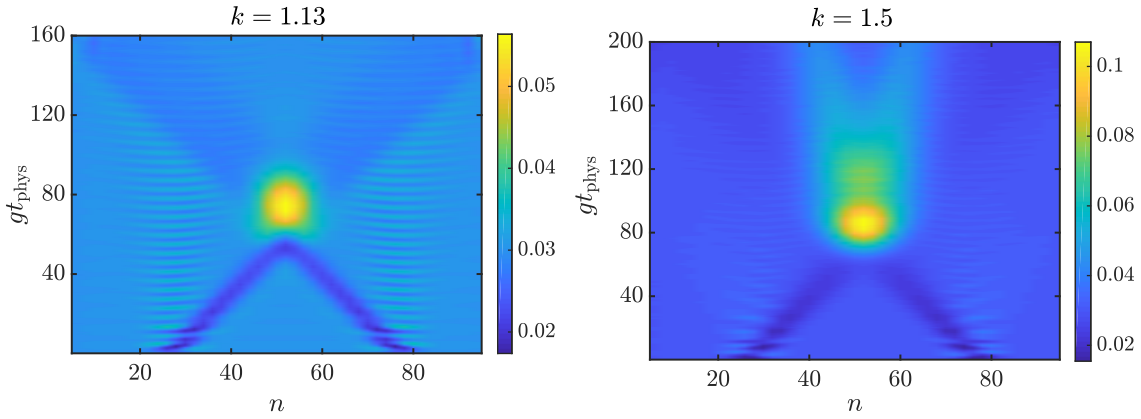


FIG. 5. This plot shows the expectation value of the four-body projector, $P_n(t)$, of a system with lattice size $N = 100$ and physical volume $N/\sqrt{x} = 100$. For the first plot the initial momenta of the mesons are $k = 1.13$ and for the second plot are $k = 1.5$. The mass is $m/g = 10^{-5}$, and the standard deviation of the wavepackets $\sigma = 4$. The bond dimension is $D = 50$ and the maximum electric flux $L_{trunc} = 8$.

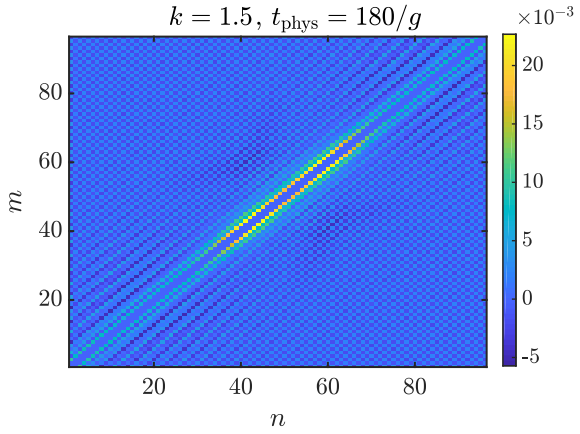


FIG. 6. This plot shows the electric flux correlator, C_l^{mn}, of a system size $N = 100$ and physical volume $N/\sqrt{x} = 100$, for times after the meson-meson collision, $gt_{\text{phys}} = 180$. The initial momenta of the mesons are above the momentum threshold, $k = 1.5$. The mass is $m/g = 10^{-5}$, and the standard deviation of the wavepackets $\sigma = 4$. The bond dimension is $D = 50$ and the maximum electric flux $L_{trunc} = 8$. We set $C^{mn} = 0$ for $m = n$, $m = n - 1$ and $m = n + 1$.

and other particles of higher mass are produced. We have proposed entanglement and local observables to identify the character of the collision and the nature of the products. More concretely, we show that the entanglement of two neighboring sites (which would be accessible from measurements on two sites) with the rest of the chain and the correlators of the electric flux provide clear signatures of the collision character. Moreover, a 4-site projector can be used as another measure for observing the presence of the collision products.

Our initial state preparation and detection observables are based on perturbative expansions valid in the strong coupling (small fermion mass and small x parameter) regime, and we performed our numerical simulations at a fixed value of the lattice spacing, far from the continuum limit. Yet, this parameter regime would be closer to the ones possible for current quantum simulation platforms, for example ultracold atoms as recently proposed for a related U(1) quantum link model [64].

Quantum inspired classical simulations with TNS, as well as proposed quantum simulation protocols are tackling progressively challenging regimes of LGT, and could in the future study problems that are out of the reach of traditional lattice methods. Out of equilibrium processes, such as collision experiments, are among the most interesting ones, but also come with their own challenges. In particular, performing a systematic extrapolation of lattice results to the continuum limit for a real-time problem, with a precision close to the one attained for spectral or thermal properties, requires performing real-time simulations on larger lattices far from the strong coupling limit as $x \to \infty$. It is thus a natural extension of this work to investigate how to prepare suitable initial states and to identify other observables that can detect the products of the collision away from the strong coupling perturbative regime.

Further extensions of our work could involve studying the effect of a background electric field, which is known to induce a phase transition, or considering collisions at much higher energies, with the possibility to open other inelastic channels. Overall, we hope that our results provide a stepping stone for exploring real-time scattering phenomena in regimes inaccessible by perturbative methods.

ACKNOWLEDGMENTS

We thank Joseph Vovrosh, Krzysztof Cichy, Manuel Schneider and David C.-J. Lin for very helpful discussions. This work was partly supported by the DFG (German Research Foundation) under Germany's Excellence Strategy – EXC-2111 – 390814868, Research Unit FOR 5522 (grant nr. 499180199), and TRR 360 - 492547816; and by the EU-QUANTERA project TNiSQ (BA 6059/1-1). Research was supported by the Munich Quantum

Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. J.K. acknowledges support from the Imperial-TUM flagship partnership.

[1] N. Brambilla, S. Eidelman, P. Foka, S. Gardner, A. S. Kronfeld, M. G. Alford, R. Alkofer, M. Butenschoen, T. D. Cohen, J. Erdmenger, L. Fabbietti, M. Faber, J. L. Goity, B. Ketzer, H. W. Lin, F. J. Llanes-Estrada, H. B. Meyer, P. Pakhlov, E. Pallante, M. I. Polikarpov, H. Sazdjian, A. Schmitt, W. M. Snow, A. Vairo, R. Vogt, A. Vuorinen, H. Wittig, P. Arnold, P. Christakoglou, P. Di Nezza, Z. Fodor, X. Garcia i Tormo, R. Höllwieser, M. A. Janik, A. Kalweit, D. Keane, E. Kiritsis, A. Mischke, R. Mizuk, G. Odyniec, K. Papadodimas, A. Pich, R. Pittau, J. W. Qiu, G. Ricciardi, C. A. Salgado, K. Schwenzer, N. G. Stefanis, G. M. von Hippel, and V. I. Zakharov, Qcd and strongly coupled gauge theories: challenges and perspectives, The European Physical Journal C **74**, 2981 (2014).

[2] K. G. Wilson, Confinement of quarks, Phys. Rev. D **10**, 2445 (1974).

[3] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum monte carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[4] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Quantum monte carlo methods for nuclear physics, Rev. Mod. Phys. **87**, 1067 (2015).

[5] K. Cichy and M. Constantinou, A guide to light-cone pdfs from lattice qcd: An overview of approaches, techniques, and results, Advances in High Energy Physics **2019**, 3036904 (2019).

[6] M. C. Bañuls and K. Cichy, Review on novel methods for lattice gauge theories, Reports on Progress in Physics **83**, 024401 (2020).

[7] Bañuls, Mari Carmen, Blatt, Rainer, Catani, Jacopo, Celi, Alessio, Cirac, Juan Ignacio, Dalmonte, Marcello, Fallani, Leonardo, Jansen, Karl, Lewenstein, Maciej, Montangero, Simone, Muschik, Christine A., Reznik, Benni, Rico, Enrique, Tagliacozzo, Luca, Van Acoleyen, Karel, Verstraete, Frank, Wiese, Uwe-Jens, Wingate, Matthew, Zakrzewski, Jakub, and Zoller, Peter, Simulating lattice gauge theories within quantum technologies, Eur. Phys. J. D **74**, 165 (2020).

[8] Z. Davoudi, M. Hafezi, C. Monroe, G. Pagano, A. Seif, and A. Shaw, Towards analog quantum simulations of lattice gauge theories with trapped ions, Phys. Rev. Res. **2**, 023015 (2020).

[9] S. D. Bass and E. Zohar, Quantum technologies in particle physics, Phil. Trans. R. Soc. A **380**, 20210072 (2022).

[10] L. Funcke, T. Hartung, K. Jansen, and S. Kühn, Review on Quantum Computing for Lattice Field Theory, PoS **LATTICE2022**, 228 (2023).

[11] C. W. Bauer, Z. Davoudi, A. B. Balantekin, T. Bhattacharya, M. Carena, W. A. de Jong, P. Draper, A. El-Khadra, N. Gemelke, M. Hanada, D. Kharzeev, H. Lamm, Y.-Y. Li, J. Liu, M. Lukin, Y. Meurice, C. Monroe, B. Nachman, G. Pagano, J. Preskill, E. Rinaldi, A. Roggero, D. I. Santiago, M. J. Savage, I. Siddiqi, G. Siopsis, D. Van Zanten, N. Wiebe, Y. Yamauchi, K. Yeter-Aydeniz, and S. Zorzetti, Quantum simulation for high-energy physics, PRX Quantum **4**, 027001 (2023).

[12] J. B. Kogut, An introduction to lattice gauge theory and spin systems, Reviews of Modern Physics **51**, 659 (1979).

[13] F. Verstraete, V. Murg, and J. Cirac, Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, Adv. Phys **57**, 143 (2008).

[14] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Annals of Physics **326**, 96 (2011), january 2011 Special Issue.

[15] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Annals Phys. **349**, 117 (2014).

[16] P. Silvi, F. Tschirsich, M. Gerster, J. Jünemann, D. Jaschke, M. Rizzi, and S. Montangero, The Tensor Networks Anthology: Simulation techniques for many-body quantum lattice systems, SciPost Phys. Lect. Notes , 8 (2019).

[17] K. Okunishi, T. Nishino, and H. Ueda, Developments in the tensor network — from statistical mechanics to quantum entanglement, J. Phys. Soc. Jpn **91**, 062001 (2022).

[18] M. C. Bañuls, Tensor network algorithms: A route map, Annu. Rev. Condens. Matter Phys **14** (2023).

[19] T. Felser, P. Silvi, M. Collura, and S. Montangero, Two-dimensional quantum-link lattice quantum electrodynamics at finite density, Phys. Rev. X **10**, 041040 (2020).

[20] P. Emonts, M. C. Bañuls, I. Cirac, and E. Zohar, Variational monte carlo simulation with tensor networks of a pure F_3 gauge theory in $(2+1)$D, Phys. Rev. D **102**, 074501 (2020).

[21] D. Robaina, M. C. Bañuls, and J. I. Cirac, Simulating $2+1$D Z_3 lattice gauge theory with an infinite projected entangled-pair state, Phys. Rev. Lett. **126**, 050401 (2021).

[22] G. Magnifico, T. Felser, P. Silvi, and S. Montangero, Lattice quantum electrodynamics in (3+1)-dimensions at finite density with tensor networks, Nature Communications **12**, 3600 (2021).

[23] P. Emonts, A. Kelman, U. Borla, S. Moroz, S. Gazit, and E. Zohar, Finding the ground state of a lattice gauge theory with fermionic tensor networks: A $2+1$D F_2 demonstration, Phys. Rev. D **107**, 014505 (2023).

[24] Y. Tong, V. V. Albert, J. R. McClean, J. Preskill, and Y. Su, Provably accurate simulation of gauge theories and bosonic systems, Quantum **6**, 816 (2022).

[25] B. Buyens, J. Haegeman, K. Van Acoleyen, H. Verschelde, and F. Verstraete, Matrix product states for gauge field theories, Phys. Rev. Lett. **113**, 091601 (2014).

[26] S. Kühn, E. Zohar, J. I. Cirac, and M. C. Bañuls, Non-abelian string breaking phenomena with matrix product states, Journal of High Energy Physics **2015**, 130 (2015).

[27] B. Buyens, J. Haegeman, H. Verschelde, F. Verstraete, and K. Van Acoleyen, Confinement and string breaking for qed_2 in the hamiltonian picture, Phys. Rev. X **6**, 041040 (2016).

[28] T. Pichler, M. Dalmonte, E. Rico, P. Zoller, and S. Montangero, Real-time dynamics in u(1) lattice gauge theories with tensor networks, Phys. Rev. X **6**, 011023 (2016).

[29] T. Chanda, J. Zakrzewski, M. Lewenstein, and L. Tagliacozzo, Confinement and lack of thermalization after quenches in the bosonic schwinger model, Phys. Rev. Lett. **124**, 180602 (2020).

[30] S. Notarnicola, M. Collura, and S. Montangero, Real-time-dynamics quantum simulation of $(1+1)$-dimensional lattice qed with rydberg atoms, Phys. Rev. Res. **2**, 013288

(2020).
[31] M. Rigobello, S. Notarnicola, G. Magnifico, and S. Montangero, Entanglement generation in (1+1)D qed scattering processes, Phys. Rev. D **104**, 114501 (2021).
[32] R. Belyansky, S. Whitsitt, N. Mueller, A. Fahimniya, E. R. Bennewitz, Z. Davoudi, and A. V. Gorshkov, High-energy collision of quarks and hadrons in the schwinger model: From tensor networks to circuit qed, arXiv preprint 10.48550/arXiv.2307.02522 (2023).
[33] M. Kormos, M. Collura, G. Takács, and P. Calabrese, Real-time confinement following a quantum quench to a non-integrable model, Nature Physics **13**, 246 (2017).
[34] J. Vovrosh, H. Zhao, J. Knolle, and A. Bastianello, Confinement-induced impurity states in spin chains, Phys. Rev. B **105**, L100301 (2022).
[35] M. Van Damme, L. Vanderstraeten, J. De Nardis, J. Haegeman, and F. Verstraete, Real-time scattering of interacting quasiparticles in quantum spin chains, Phys. Rev. Res. **3**, 013078 (2021).
[36] F. M. Surace and A. Lerose, Scattering of mesons in quantum simulators, New Journal of Physics **23**, 062001 (2021).
[37] P. Karpov, G.-Y. Zhu, M. P. Heller, and M. Heyl, Spatiotemporal dynamics of particle collisions in quantum spin chains, Phys. Rev. Res. **4**, L032001 (2022).
[38] A. Milsted, J. Liu, J. Preskill, and G. Vidal, Collisions of false-vacuum bubble walls in a quantum spin chain, PRX Quantum **3**, 020316 (2022).
[39] J. Vovrosh, R. Mukherjee, A. Bastianello, and J. Knolle, Dynamical hadron formation in long-range interacting quantum spin chains, PRX Quantum **3**, 040309 (2022).
[40] Y. Chai, A. Crippa, K. Jansen, S. Kühn, V. R. Pascuzzi, F. Tacchino, and I. Tavernelli, Entanglement production from scattering of fermionic wave packets: a quantum computing approach, arXiv:2312.02272 (2023), arXiv:2312.02272 [quant-ph].
[41] A. F. Shaw, P. Lougovski, J. R. Stryker, and N. Wiebe, Quantum Algorithms for Simulating the Lattice Schwinger Model, Quantum **4**, 306 (2020).
[42] J. Schwinger, Gauge invariance and mass. ii, Phys. Rev. **128**, 2425 (1962).
[43] S. R. Coleman, More About the Massive Schwinger Model, Annals Phys. **101**, 239 (1976).
[44] The massive schwinger model on a lattice: Background field, chiral symmetry and the string tension, Nuclear Physics B **208**, 413 (1982).
[45] C. J. Hamer, W.-h. Zheng, and J. Oitmaa, Series expansions for the massive Schwinger model in Hamiltonian lattice theory, Phys. Rev. D **56**, 55 (1997), arXiv:hep-lat/9701015.
[46] T. Byrnes, P. Sriganesh, R. J. Bursill, and C. J. Hamer, Density matrix renormalization group approach to the massive Schwinger model, Phys. Rev. D **66**, 013002 (2002), arXiv:hep-lat/0202014.
[47] J. Schwinger, Gauge invariance and mass, Phys. Rev. **125**, 397 (1962).
[48] C. Adam, Scattering processes in the massive schwinger model, Phys. Rev. D **55**, 6299 (1997).
[49] S. Coleman, R. Jackiw, and L. Susskind, Charge shielding and quark confinement in the massive schwinger model, Annals of Physics **93**, 267 (1975).
[50] C. Adam, The schwinger mass in the massive schwinger model, Physics Letters B **382**, 383 (1996).
[51] C. Adam, Improved vector and scalar masses in the massive schwinger model, Physics Letters B **555**, 132 (2003).
[52] Y. Mo and R. J. Perry, Basis function calculations for the massive schwinger model in the light-front tamm-dancoff approximation, Journal of Computational Physics **108**, 159 (1993).
[53] K. Harada, A. Okazaki, and M.-a. Taniguchi, Six-body light-front tamm-dancoff approximation and wave functions for the massive schwinger model, Phys. Rev. D **52**, 2429 (1995).
[54] J. Kogut and L. Susskind, Hamiltonian formulation of wilson's lattice gauge theories, Phys. Rev. D **11**, 395 (1975).
[55] T. Banks, L. Susskind, and J. Kogut, Strong-coupling calculations of lattice gauge theories: (1 + 1)-dimensional exercises, Phys. Rev. D **13**, 1043 (1976).
[56] M. C. Bañuls, K. Cichy, J. I. Cirac, and K. Jansen, The mass spectrum of the schwinger model with matrix product states, Journal of High Energy Physics **2013**, 1 (2013).
[57] K. Cichy, A. Kujawa-Cichy, and M. Szyniszewski, Lattice hamiltonian approach to the massless schwinger model: Precise extraction of the mass gap, Computer Physics Communications **184**, 1666 (2013).
[58] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, New Journal of Physics **12**, 025012 (2010).
[59] H. F. Trotter, On the product of semi-groups of operators (1959) pp. 545–551.
[60] M. Suzuki, Fractal decomposition of exponential operators with applications to many-body theories and monte carlo simulations, Physics Letters A **146**, 319 (1990).
[61] M. C. Bañuls, K. Cichy, J. I. Cirac, K. Jansen, and H. Saito, Thermal evolution of the schwinger model with matrix product operators, Phys. Rev. D **92**, 034519 (2015).
[62] M. C. Bañuls, K. Cichy, K. Jansen, and H. Saito, Chiral condensate in the schwinger model with matrix product operators, Phys. Rev. D **93**, 094512 (2016).
[63] A. Impertro, S. Karch, J. F. Wienand, S. Huh, C. Schweizer, I. Bloch, and M. Aidelsburger, Local readout and control of current and kinetic energy operators in optical lattices, arXiv preprint 10.48550/arXiv.2312.13268 (2023).
[64] G.-X. Su, J. Osborne, and J. C. Halimeh, A cold-atom particle collider, arXiv preprint 10.48550/arXiv.2401.05489 (2024).

arXiv:2402.18038v1 [hep-lat] 28 Feb 2024

Infrared fixed point in the massless twelve-flavor SU(3) gauge-fermion system

Anna Hasenfratz[1,*] and Curtis T. Peterson[1,†]

[1] *Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

We present strong numerical evidence for the existence of an infrared fixed point in the renormalization group flow of the SU(3) gauge-fermion system with twelve massless fermions in the fundamental representation. Our numerical simulations using nHYP-smeared staggered fermions with Pauli-Villars improvement do not exhibit any first-order bulk phase transition in the investigated parameter region. We utilize an infinite volume renormalization scheme based on the gradient flow transformation to determine the renormalization group β function. We identify an infrared fixed point at $g^2_{\mathrm{GF}\star} = 6.69(68)$ in the GF scheme and calculate the leading irrelevant critical exponent $\gamma_g^\star = 0.204(36)$. Our prediction for $\gamma_g^\star$ is consistent with available literature at the 1-2σ level.

I. INTRODUCTION

The infrared properties of the SU(3) gauge-fermion system with $N_f = 12$ massless fundamental flavors has been studied extensively using a variety of analytical and numerical techniques. Such techniques include perturbation theory [1–5], use of the gap equation [6, 7], functional renormalization group methods [8, 9], conformal expansion [10], conformal bootstrap [11], the background field method [12], perturbative non-relativistic quantum chromodynamics [13], large-N expansion [14], and nonperturbative lattice simulations [15–38]. Investigations based on lattice simulations have utilized finite-volume step-scaling [15–24], Monte Carlo renormalization group methods [25, 26], hadron mass and decay constant spectroscopy [28–36], and the Dirac eigenmode spectrum [37, 38]. Many investigations suggest that the $N_f = 12$ system is infrared conformal[1], though a minority of studies conclude that the system is confining with chiral symmetry breaking, or are inconclusive, as they find neither direct evidence of chiral symmetry breaking nor of an infrared fixed point[2].

Most lattice studies are affected by the presence of a bulk first-order phase transition [40–45]. Such unphysical phase transitions are triggered by strong ultraviolet fluctuations in the fermion sector that prevent lattice simulations from reaching deep into the infrared regime. Even when strong couplings are reached, lattice cutoff effects make it difficult to take the proper continuum limit, leading to inconsistent results between different lattice formulations. It is imperative to reduce the ultraviolet fluctuations that trigger first-order bulk phase transitions. Ref. [46] suggested to include unphysical heavy Pauli-Villars fields to achieve the necessary improvement. Lattice Pauli-Villars fields are similar to their continuum analogue - they have the same action as the fermions, but possess bosonic statistics. Their mass is at the level of the cutoff. Therefore, they decouple in the infrared limit, while in the ultraviolet they compensate for cutoff effects introduced by the fermions. This idea has been tested in simulations of the SU(3) gauge-fermion system with ten massless fundamental Dirac fermions (flavors) and the SU(4) gauge-fermion system four massless fundamental and four massless two-index Dirac fermions [47, 48]. Both studies extended the reach into the infrared regime significantly, and both present clear evidence for infrared conformality in those systems. See also Refs. [43, 44] for an alternative proposal to remove unphysical bulk phase transitions.

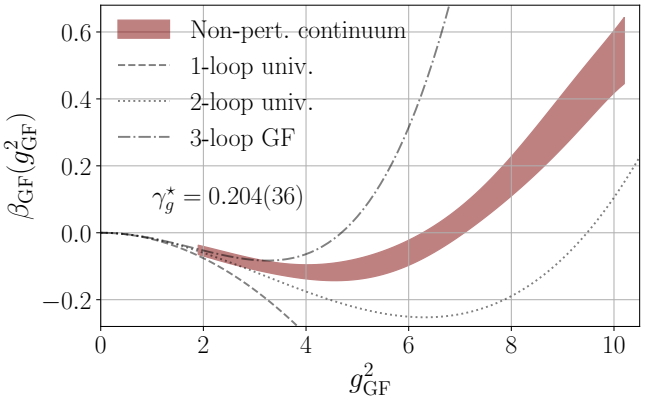


FIG. 1. Our continuum prediction for $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ as a function of g^2_{GF} (maroon band) juxtaposed against the 1- (dashed), 2- (dotted) and 3-loop (dashed-dotted) gradient flow β function from perturbation theory [49]. The width of the maroon band indicates the error. Also given is our prediction for the leading irrelevant critical exponent at the IRFP, $\gamma_g^\star = 0.204(36)$.

In this work, we utilize Pauli-Villars (PV) improvement to study the infrared properties of the massless SU(3) gauge-fermion system with $N_f = 12$ fundamental flavors. We calculate the renormalization group (RG) β function, defined as the logarithmic derivative of the renormalized running coupling $g^2(\mu)$ with respect to an energy scale μ as

$$\beta\left(g^2\right) = \mu^2 \frac{\mathrm{d}g^2(\mu)}{\mathrm{d}\mu^2}. \tag{1}$$

If the β function exhibits a fixed point at some coupling

* anna.hasenfratz@colorado.edu
† curtis.peterson@colorado.edu

[1] See Refs. [1–3, 5–7, 10, 11, 15–17, 20, 23–26, 28, 30, 31, 33–36, 38]
[2] See Refs. [12, 14, 18, 19, 29, 32, 39]

$g_\star^2$, the gauge coupling becomes irrelevant and the system is infrared conformal. The value of $g_\star^2$ depends on the renormalization scheme, but the existence of the infrared fixed point (IRFP) does not.

On the lattice, it is convenient to use gradient flow (GF) transformation [50–52] as a continuous smearing operation to define an infinite-volume renormalization scheme [27, 53–57]. In the GF scheme, the renormalized coupling $g^2_{\rm GF}(t)$ in infinite volume at flow time $t \propto 1/\mu^2$ is defined in terms of the Yang-Mills energy density $E(t)$ as

$$g^2_{\rm GF}(t) \equiv \mathcal{N}\langle t^2 E(t)\rangle, \tag{2}$$

where $\mathcal{N} = 128\pi^2/(3N_c^2 - 3)$ is a constant chosen to match $g^2_{\rm GF}$ to $g^2_{\overline{\rm MS}}$ at tree-level [51]. The corresponding RG β function is

$$\beta_{\rm GF}\left(g^2_{\rm GF}\right) \equiv -t\frac{{\rm d}g^2_{\rm GF}(t)}{{\rm d}t}. \tag{3}$$

To calculate the infinite volume gradient flow β function $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ from finite-volume simulations, we utilize the continuous β function method (CBFM) proposed in Refs. [27, 53, 56] and deployed extensively in Refs. [47, 48, 58–61] to a variety of strongly-coupled gauge-fermion systems. In this paper we follow the steps described in [59, 61] with additional extensions for improved error estimation. We discuss the continuous β function method and its implementation in further detail in Sec. III.

As a preview, we show our nonperturbative prediction for $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ as a function of $g^2_{\rm GF}$ in Fig. 1. The predicted β function converges to the universal 1-/2-loop and 3-loop gradient flow perturbative β functions at small $g^2_{\rm GF}$ [49]. Around $g^2_{{\rm GF}\star} = 6.69(68)$, the nonperturbative β function unambiguously exhibits an infrared fixed point. From the slope of $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ at $g^2_{{\rm GF}\star}$, we calculate the leading irrelevant critical exponent $\gamma_g^\star = 0.204(36)$ and find that it is consistent with the perturbative calculations of Refs. [4, 5] at the 1σ level and the lattice calculation of Ref. [20] at the 2σ level. We control for systematical errors in the infinite volume extrapolation step of the CBFM using Bayesian model averaging. Additionally, our Pauli-Villars improved simulations offer tight control over systematics in the continuum extrapolation step of the CBFM. Our data is publicly available at Ref. [62].

This paper is laid out as follows. In Sec. II, we summarize details of our numerical simulations. In Sec. III, we review the continuous β function method and discuss our analysis. We explain our calculation of the leading irrelevant critical exponent in Sec. IV, and wrap up in Sec. V with conclusions.

II. NUMERICAL DETAILS

We simulate the massless twelve-flavor SU(3) gauge-fermion system using an adjoint-plaquette gauge action with $\beta_F/\beta_A = -0.25$ ($\beta_F \equiv \beta_b$) and a massless ($am_f = 0.0$) nHYP-smeared staggered fermion action with four massive ($am_{\rm PV} = 0.5$) Pauli-Villars (PV) fields per staggered fermion [46, 63–65]. The "pions" of these PV fields have mass $am^{\rm (PV)}_{\rm PS} \gtrsim 1.0$[3] and generate gauge loops in the effective gauge action with a size that decays exponentially with $am^{\rm (PV)}_{\rm PS}$ [46]. As long as the volume is much larger than $1/m^{\rm (PV)}_{\rm PS}$ and $m_f \ll m^{\rm (PV)}_{\rm PS}$ the PV fields decouple in the infrared. Their only effect is a modified, but local, gauge action. One of the goals of the present work is to illustrate the validity of this expectation.

β_b	L/a				
	24	28	32	36	40
9.20	340	253	188	188	133
9.40	347	262	215	273	186
9.60	244	233	251	203	166
9.80	275	329	250	297	280
10.0	271	246	312	151	134
10.2	184	209	217	221	133
10.4	283	241	299	221	142
10.8	246	220	288	208	306
11.0	236	288	156	151	156
11.4	188	194	223	193	183
12.0	182	248	200	254	167
12.8	180	179	204	254	209
13.6	251	183	168	254	228
14.6	253	191	178	251	226

TABLE I. The number of thermalized configurations analyzed at each bare coupling β_b and volume L/a. The configurations are separated by 10 MDTUs.

We use antiperiodic boundary conditions in all four directions for both the staggered fermion fields and PV fields. Our numerical simulations are performed using the hybrid Monte Carlo algorithm [66] implemented in a modified version of the `MILC` library[4] and the `Quantum EXpressions` (`QEX`) library[5] [67]. We set the molecular dynamics trajectory length to $\tau = 1.0$. Our configurations are separated by ten trajectories (10 molecular dynamics time units). We perform our simulations at fourteen bare gauge couplings ($9.20 \le \beta_b \le 14.6$) and five symmetric volumes ($24 \le L/a \le 40$). In Table I, we list the total number of thermalized configurations on each ensemble.

Our gradient flow measurements are performed using either the modified `MILC` or `QEX` libraries [67]. We flow our configurations using Wilson flow [51, 52], integrating the gradient flow equations using the 4th-order Runge-Kutta algorithm discussed in Ref. [51] with time step ${\rm d}t/a^2 = 0.02$ for $t/a^2 \le 5.0$ and ${\rm d}t/a^2 = 0.1$ for

[3] PS = pseudoscalar

[4] The modified `MILC` library can be found at https://github.com/daschaich/KS_nHYP_FA

[5] Our fork of `QEX` can be found at https://github.com/ctpeterson/qex

$t/a^2 > 5.0$. At each integration step, we measure the Yang-Mills energy density $E(t)$ using the Wilson (W) and clover (C) discretizations. In the rest of this paper, we refer to results based on Wilson flow and Wilson operator as WW, while we refer to results based on Wilson flow and clover operator as WC. Our data for the Yang-Mills energy density $E(t)$ from the Wilson and clover operator is available at Ref. [62].

III. NONPERTURBATIVE β FUNCTION

We measure the gradient flow coupling in finite volume in terms of the Yang-Mills energy density $E(t)$ as

$$g^2_{\mathrm{GF}}\big(t; L, g_0^2\big) \equiv \frac{\mathcal{N}}{1+\delta(t,L)} \ \big\langle t^2 E(t)\big\rangle\,, \tag{4}$$

where $\delta(t,L)$ corrects for gauge zero modes [68]. From $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$, we calculate the gradient flow β function in finite volume as

$$\beta_{\mathrm{GF}}\big(t; L, g_0^2\big) \equiv -t\frac{\mathrm{d}}{\mathrm{d}t} g^2_{\mathrm{GF}}\big(t; L, g_0^2\big), \tag{5}$$

where we discretize $\mathrm{d}/\mathrm{d}t$ with a 5-point stencil. The autocorrelation time for $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$ and $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$ is typically between 20-80 molecular dynamics time units (MDTUs), with occasional jumps to 120-200 MDTUs.

To extract the continuum $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ as a function of g^2_{GF}, we follow the CBFM procedure outlined in Refs. [59, 61].

1. Take the infinite volume limit by independently extrapolating both $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$ and $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$ linearly in $(a/L)^4 \to 0$ at fixed t/a^2 and β_b.
2. Interpolate $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ in $g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ at fixed t/a^2.
3. Take the continuum limit by extrapolating $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ linearly in $a^2/t \to 0$ at fixed $g^2_{\mathrm{GF}}(t)$.

Correlated uncertainties are propagated throughout our analysis using the automatic error propagation tools provided by the `gvar` library [69]. Fits are performed using the `SwissFit` library, which integrates directly with `gvar` [70]. The steps of the CBFM are detailed in the rest of this section.

A. Infinite volume extrapolation

Because the Yang-Mills energy density is a dimension-4 operator, leading finite-volume corrections to $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$ and $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$ are expected to be $\mathcal{O}\big(t^2/L^4\big)$. Therefore, we extrapolate both $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$ and $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$ to $a/L \to 0$ by independently fitting them to the ansatz

$$\mathrm{FV}(L/a) = k_1(a/L)^4 + k_2 \tag{6}$$

at fixed t/a^2 and β_b. This analysis strategy was first outlined in Ref. [61] and subsequently applied in Refs. [59, 60]. Alternative methods are discussed in Refs. [27, 53, 56, 58].

We account for the systematic uncertainty that is associated with choosing a particular subset of volumes for the infinite volume extrapolation using Bayesian model averaging [71–73]. We do so by first fitting over all possible subsets of volumes $L/a \in \{24, 28, 32, 36, 40\}$ with at least three volumes in each subset η. We calculate the model weight w_η for a particular subset η as

$$w_\eta \propto \exp\left[-\frac{1}{2}\big(\chi^2_\eta + 2d_\eta\big)\right], \tag{7}$$

where χ^2_η is the χ^2 statistic of fit η and d_η is the number of data points *not* included in fit η from the full set of volumes $\{24, 28, 32, 36, 40\}$. Denoting the mean of k_i from fit η as $\overline{k}_i^{(\eta)}$, our model-averaged prediction for the mean $\overline{k}_i$ of k_i is

$$\overline{k}_i = \sum_\eta \overline{k}_i^{(\eta)} w_\eta, \tag{8}$$

where the weights w_η have been normalized such that $\sum_\eta w_\eta = 1$. The covariance C_{ij} of our model-averaged prediction for $\{k_i\}_{i=1,2}$ is

$$C_{ij} = \sum_\eta C_{ij}^{(\eta)} w_\eta + \sum_\eta \overline{k}_i^{(\eta)} \overline{k}_j^{(\eta)} w_\eta - \overline{k}_i \overline{k}_j, \tag{9}$$

where $C_{ij}^{(\eta)}$ is the covariance of $\{k_i^{(\eta)}\}_{i=1,2}$ from fit η.

In Figs. 2 and 3, we show the result of our model-averaged infinite volume extrapolation for the W and C discretization of $E(t)$, respectively, over a range of flow times $2.5 \le t/a^2 \le 6.0$ (different colors). The left panels of Figs. 2-3 show our infinite extrapolation of $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$, while the right panels show our infinite volume extrapolation of $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$. The bare gauge couplings that we chose for these plots are in the vicinity where the continuum β function predicts an IRFP. For all three bare gauge couplings shown in Figs. 2-3, the model average is dominated by subsets containing $L/a \in \{28, 32, 36, 40\}$, as $L/a = 24$ often deviates from the linear trend in a^4/L^4, particularly as the flow time increases. This is reflected in the model average, as fits including $L/a = 24$ possess small model weights w_η and contribute negligibly to the model average.

B. Intermediate interpolation

The continuum limit $a^2/t \to 0$ of $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ is taken at fixed g^2_{GF}. We predict pairs $\big(t/a^2, \beta_{\mathrm{GF}}\big(t; g_0^2\big)\big)$ at a set of fixed g^2_{GF} for the continuum extrapolation by interpolating $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ in $g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ at fixed t/a^2 using the ansatz

$$\mathcal{I}_N\big(g^2_{\mathrm{GF}}\big) = g^4_{\mathrm{GF}} \sum_{i=0}^{N-1} p_n g^{2n}_{\mathrm{GF}}. \tag{10}$$

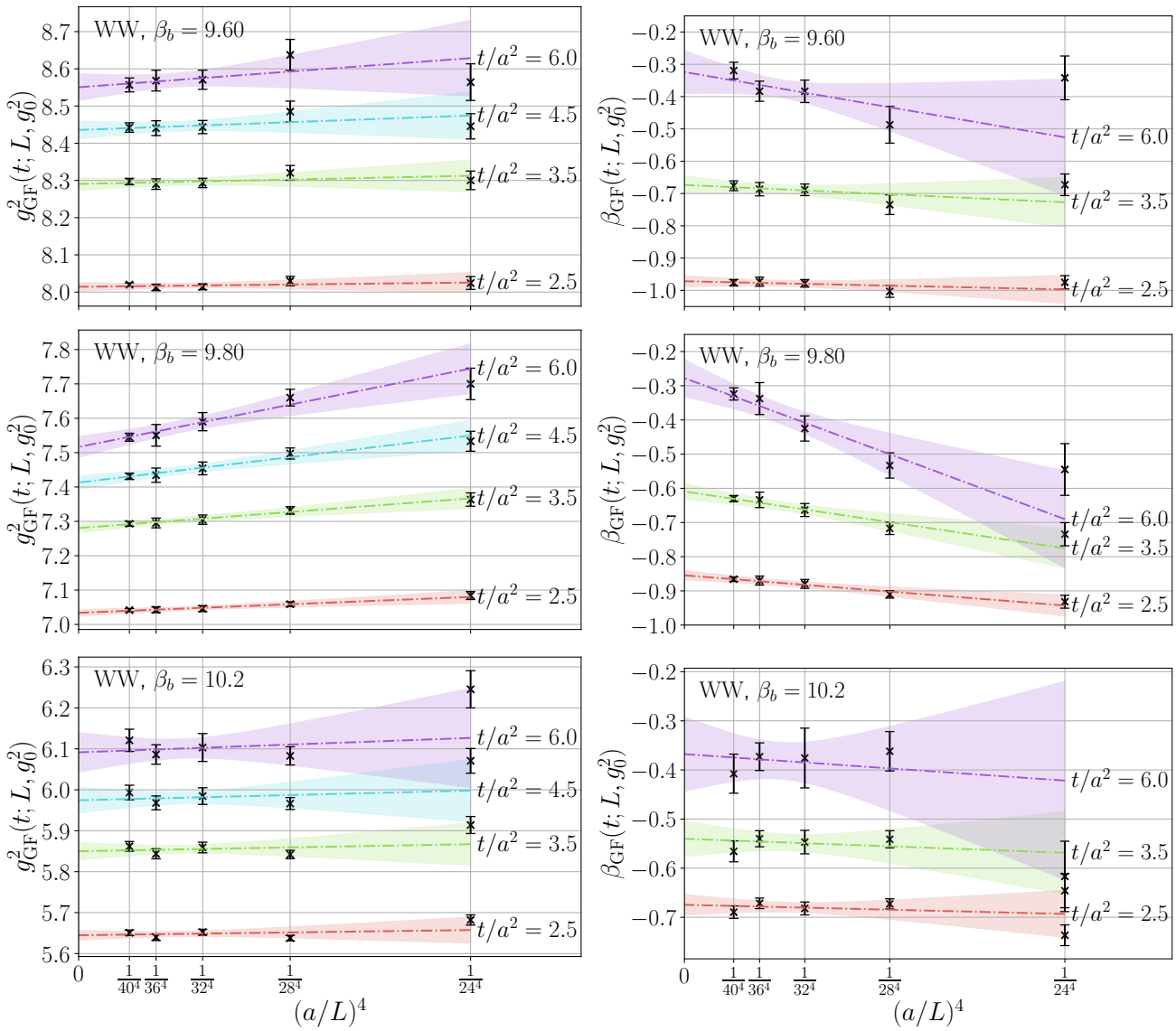


FIG. 2. Result of our infinite volume extrapolation of $g^2_{\rm GF}(t; L, g_0^2)$ (left panels) and $\beta_{\rm GF}(t; L, g_0^2)$ (right panels) for the Wilson (W) operator at $\beta_b = 9.60$ (top panels), 9.80 (middle panels) and 10.2 (bottom panels). Black ($\times$) markers with error bars are the data included in our extrapolation. Extrapolations with errors that are predicted from Bayesian model averaging are indicated by multi-colored bands at $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). We do not show the infinite volume extrapolation of $\beta_{\rm GF}(t; L, g_0^2)$ at $t/a^2 = 4.5$ for visualization purposes.

At each t/a^2, we account for the uncertainty in $g^2_{\rm GF}(t; g_0^2)$ by including the mean and covariance of $g^2_{\rm GF}(t; g_0^2)$ as a Gaussian prior. We also set a Gaussian prior on each coefficient p_n with zero mean and a width of 0.1, which helps stabilize the fit. We choose $N = 4$, as it is the lowest value of N that fits the data well. In Fig. 4, we show the result of our interpolation for the W operator (top panel) and the C operator (bottom panel) for several flow time values in the range $2.5 \leq t/a^2 \leq 6.0$ (different colors). Our fits have p-values in the $35\% - 100\%$ range, but they strongly skew toward the higher end. This could indicate that we are either overfitting or the errors in our data are overestimated. Reducing the order makes each interpolation significantly worse, as interpolations with $N \leq 3$ are unable to accommodate the varying curvature at weak/strong coupling. Therefore, use $N = 4$ for our central analysis. We will discuss the systematic effect that is associated with the order N in our estimate of $g^2_{\rm GF\star}$ and $\gamma_g^\star$ in Sec. IV.

C. Continuum extrapolation

The final step is the continuum ($a^2/t \to 0$) limit over a set of fixed $g^2_{\rm GF}$ that predicts $\beta_{\rm GF}(g^2_{\rm GF})$ as a function of $g^2_{\rm GF}$. The range of $[t_{\rm min}/a^2, t_{\rm max}/a^2]$ used in the contin-

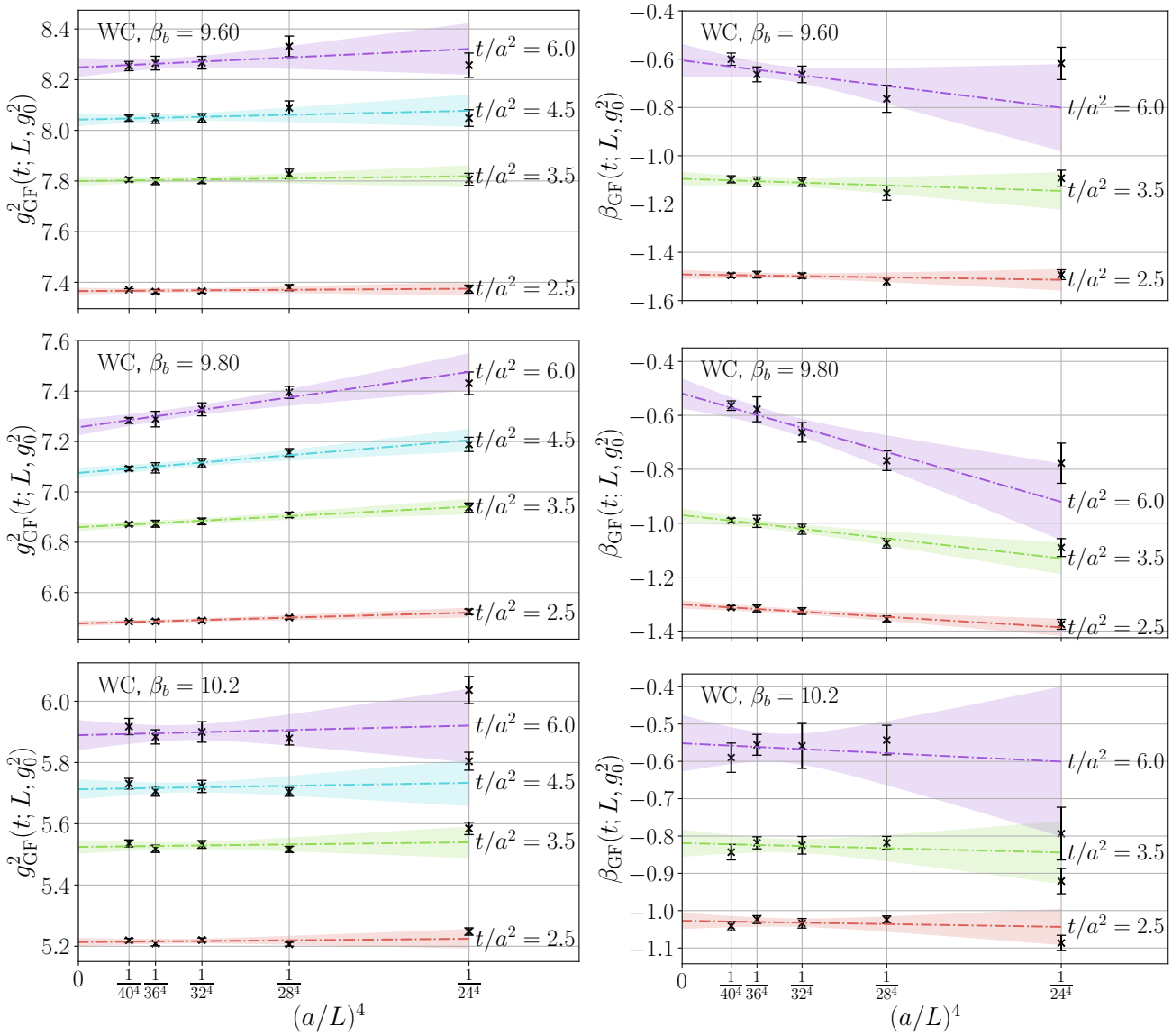


FIG. 3. Result of our infinite volume extrapolation of $g^2_{\rm GF}\big(t; L, g_0^2\big)$ (left panels) and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ (right panels) for the clover (C) operator at $\beta_b = 9.60$ (top panels), 9.80 (middle panels) and 10.2 (bottom panels). Black ($\times$) markers with error bars are the data included in our extrapolation. Extrapolations with errors that are predicted from Bayesian model averaging are indicated by multi-colored bands at $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). We do not show the infinite volume extrapolation of $\beta_{\rm GF}\big(t; L, g_0^2\big)$ at $t/a^2 = 4.5$ for visualization purposes.

uum extrapolation must be chosen with care. The value of $t_{\rm min}/a^2$ must be large enough for the RG flow to reach the renormalized trajectory. Once this is the case, finite-cutoff effects are $\mathcal{O}(a^2/t)$. In practice, one can identify when $t_{\rm min}/a^2$ is close enough to the renormalized trajectory by the overlap between the continuum prediction for $\beta_{\rm GF}(g^2_{\rm GF})$ from both operators. Because finite-volume effects are expected to be $\mathcal{O}\big(t^2/L^4\big)$, $t_{\rm max}/a^2$ must be chosen such $t/a^2 < t_{\rm max}/a^2$ has a reliable infinite volume extrapolation. Ideally, we would apply Bayesian model averaging to the continuum extrapolation to automatically account for systematic effect that is associated with making a particular choice in $[t_{\rm min}/a^2, t_{\rm max}/a^2]$. However, at this point in the analysis, we no longer have access to the full covariance matrix, which means that we no longer have access to a reliable estimate of the model weights. To estimate the error in our continuum extrapolation, we use the half-difference of the prediction from the continuum extrapolation performed at $\pm 1\sigma$. This approach was also taken in Ref. [59].

Fig. 5 shows examples of the continuum extrapolation performed in the range $2.0 \leq g^2_{\rm GF} \leq 6.0$ (different colors) using $[t_{\rm min}/a^2, t_{\rm max}/a^2] = [3.5, 6.0]$. For $t/a^2 \lesssim 3.5$, $\beta\big(t; g_0^2\big)$ has a slight curvature in a^2/t, indicating emerg-

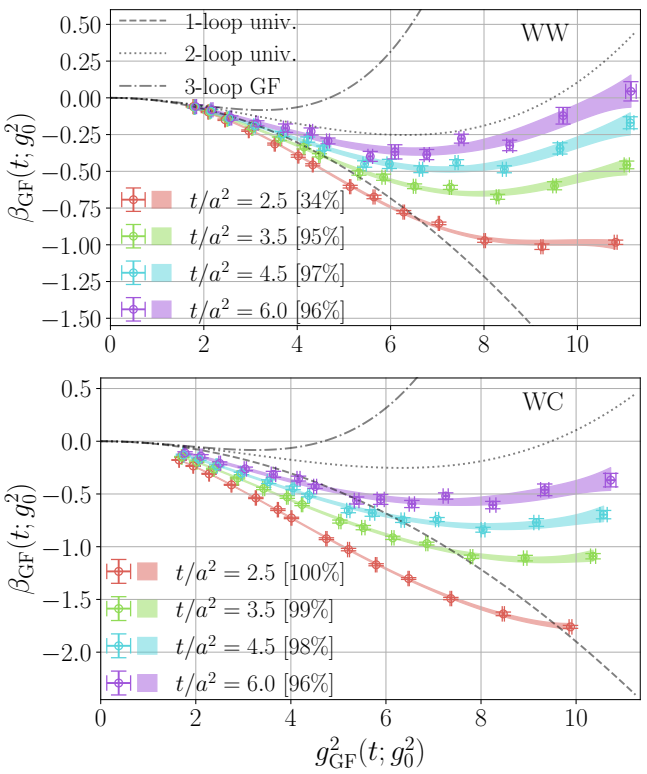


FIG. 4. Illustration of our interpolation of $\beta_{\rm GF}(t; g_0^2)$ in $\beta_{\rm GF}(t; g_0^2)$ for the Wilson operator (top panel) and clover operator (bottom panel). Interpolations at fixed t/a^2 are indicated by colored bands, with $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). The width of the band indicates the error. The data contributing to each interpolation is indicated by an open circular marker with both x- and y-errors. We compare our interpolation against the continuum 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) gradient flow β function from perturbation theory [49].

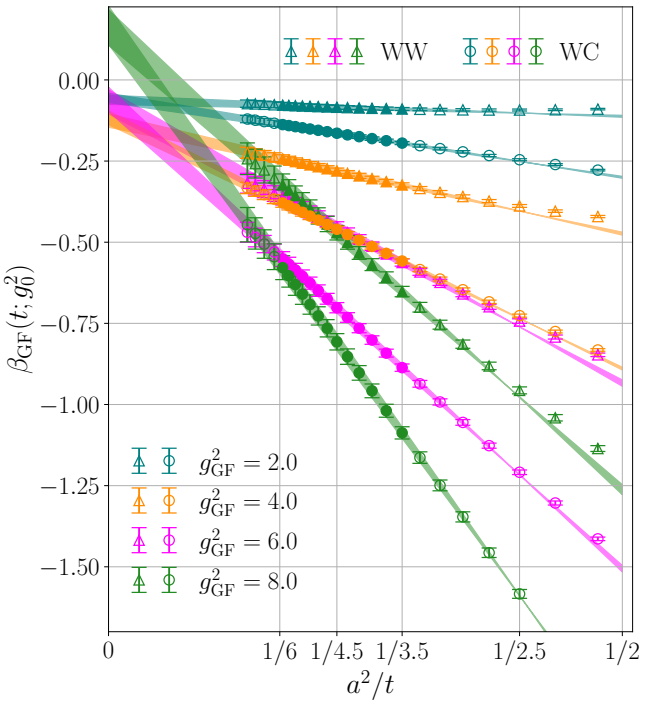


FIG. 5. Illustration of our continuum extrapolation of $\beta_{\rm GF}(t; g_0^2)$ at fixed $g^2_{\rm GF} = 2.0$ (teal), 4.0 (dark orange), 6.0 (magenta), and 8.0 (forest green). Data contributing to our extrapolation with the W operator are shown as error bars with triangular markers and the C operator are shown as error bars with circular markers. Our extrapolations are shown as colored bands, where the error is indicated by the width of the band.

ing higher-order cutoff effects for both the W and C operator. For $t/a^2 \gtrsim 6.0$, the data begins to deviate from a linear trend in a^2/t, indicating that the infinite volume extrapolation is getting unreliable. Our choice of $[t_{\rm min}/a^2, t_{\rm max}/a^2] = [3.5, 6.0]$ avoids both of these two regimes. In Sec. IV, we discuss the sensitivity of our prediction for $g^2_{\rm GF\star}$ and $\gamma_g^\star$ to our choice of $t_{\rm min}/a^2, t_{\rm max}/a^2$.

We show our prediction for the continuum $\beta_{\rm GF}(g^2_{\rm GF})$ in Fig. 6 from the WW (gold band) and WC (maroon band) combination. The continuum predictions from both operators are consistent with one another across the entire range of investigated renormalized couplings $g^2_{\rm GF}$. At small $g^2_{\rm GF}$, the continuum $\beta_{\rm GF}(g^2_{\rm GF})$ appears to converge to the 1-, 2-, and 3-loop perturbative gradient flow β function [49]. At $g^2_{\rm GF\star} \approx 6.69$, our continuum β function predicts an infrared fixed point. The location of the fixed point is slightly below the predicted IRFP from the step-scaling calculation of Ref. [20]. Note that, because the calculation in Ref. [20] was done in a different gradient-flow-based renormalization scheme, the predicted $g^2_{\rm GF\star}$

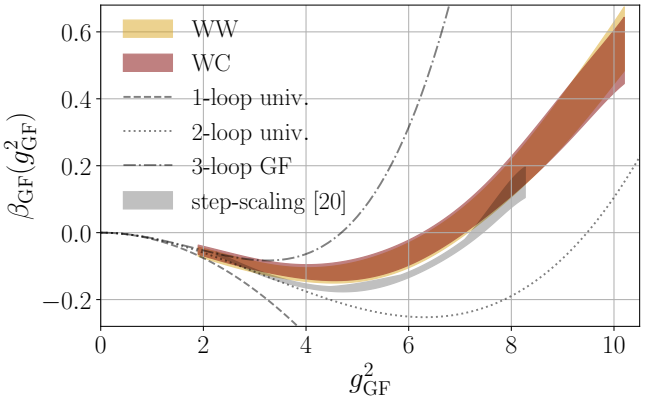


FIG. 6. Our continuum prediction for $\beta_{\rm GF}(g^2_{\rm GF})$ as a function of $g^2_{\rm GF}$ for the W operator (gold band) and C operator (maroon band). The width of the band indicates the error. The nonperturbative results are juxtaposed against the 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) gradient flow β function from perturbation theory [49]. Also shown is the step-scaling β function in the $c = 0.25$ scheme from Ref. [20] as a grey band.

values do not have to agree. Our final result for $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ from both operators is provided as an ASCII file.

IV. THE IRFP AND ITS LEADING IRRELEVANT CRITICAL EXPONENT

In the vicinity of the RG fixed point $g^2_{\rm GF\star}$

$$\beta\left(g^2_{\rm GF}\right) \approx \frac{\gamma^\star_g}{2}\left(g^2_{\rm GF} - g^2_{\rm GF\star}\right), \quad g^2_{\rm GF} \to g^2_{\rm GF\star} \tag{11}$$

where $\gamma^\star_g$ is the universal critical exponent of the irrelevant gauge coupling. The factor of $1/2$ is chosen to match the convention of Refs. [5, 20]. We estimate $g^2_{\rm GF\star}$ and $\gamma^\star_g$ via the following procedure.

1. Interpolate the central value of $\beta\left(g^2_{\rm GF}\right)$ and the central value of $\beta\left(g^2_{\rm GF}\right) \pm 1\sigma$ in $g^2_{\rm GF}$ using a monotonic spline.
2. Estimate the central value of $g^2_{\rm GF\star}$ from the root of the spline interpolation of $\beta\left(g^2_{\rm GF}\right)$ in $g^2_{\rm GF}$.
3. Estimate the central value of $\gamma^\star_g$ from the derivative of the spline at $g^2_{\rm GF\star}$.
4. Repeat Steps (2)-(3) with a spline interpolation of $\beta\left(g^2_{\rm GF}\right) \pm 1\sigma$ in $g^2_{\rm GF}$.
5. Estimate the error in $g^2_{\rm GF\star}$ and $\gamma^\star_g$ from the half difference of their predictions from the interpolations in Step (4).

Steps (1)-(5) yield $g^2_{\rm GF\star} = 6.77(39), 6.69(40)$ and $\gamma^\star_g = 0.210(22), 0.204(21)$ for WW and WC, respectively. In Fig. 7 we look at how the continuum limit predictions for $g^2_{\rm GF\star}$ (top panels) and $\gamma^\star_g$ (bottom panels) vary with our choice of $t_{\rm min}/a^2$ (x-axes) and $t_{\rm max.}/a^2$ (different colors) for the W (left panels) and C (right panels) operators. The central values for both quantities and both operators are stable; they vary well within error. The stability in $t_{\rm min}/a^2$ is attributed to the linearity of the continuum extrapolation over a wide range of a^2/t, while the stability in $t_{\rm max}/a^2$ is likely attributed to our control over the infinite volume extrapolation.

We take the result for $\gamma^\star_g$ from the WC combination with the value for $[t_{\rm min}/a^2, t_{\rm max}/a^2] = [3.5, 6.0]$ from Sec. V as our central result. We estimate additional systematic errors by varying our analysis as follows.

- Choosing a higher-order polynomial for the intermediate interpolation in Sec. III B. The highest-order interpolation in N that we can use before we lose control over our continuum extrapolation due to overfitting is $N = 7$. This shifts the value of $\gamma^\star_g$ by ≈ 0.003, and we take the latter difference as an estimate for the systematic error that is associated with our choice of N for the intermediate interpolation.
- Choosing a different $t_{\rm min}/a^2, t_{\rm max}/a^2$ in the continuum extrapolation. We estimate the systematic error that is associated with our choice of $t_{\rm min}/a^2, t_{\rm max}/a^2$ by the difference in the most extreme values of $\gamma^\star_g$ in our variations illustrated in Fig. 7. This yields a systematic error of ≈ 0.006.
- Choosing instead the prediction for $\gamma^\star_g = 0.210(22)$ from the WW combination as our central result. We take the difference in these predictions (≈ 0.006) as an estimate of the systematic error associated with making a particular choice of flow/operator combination.

To be conservative, we combine the error in our analysis of $\gamma^\star_g$ with the systematic error estimates above linearly. This yields the final of prediction $\gamma^\star_g = 0.204(36)$. Repeating the same exercise for $g^2_{\rm GF\star}$ yields a systematic error of ≈ 0.11 from the interpolation order, ≈ 0.09 from the continuum extrapolation, and ≈ 0.08 from the flow/operator combination. Including the systematic error in $g^2_{\rm GF\star}$ linearly yields a final prediction of $g^2_{\rm GF\star} = 6.69(68)$.

In Fig. 8 we compare our prediction for $\gamma^\star_g$ with those available in the literature [5, 20]. Our result is plotted as a maroon star with errors indicated by an error bar. The smaller error bar is our error estimate before accounting for systematic effects and the larger error bar includes systematic effects. The result for $\gamma^\star_g$ from the perturbative calculation of Ref. [5] (dark gold error bar) is within 1σ of our estimate for γ^*_g. The lattice calculation of $\gamma^\star_g$ from Ref. [20] (cyan error bar) is within 2σ of our result. However, it is important to note that the lattice calculation in Ref. [20] uses smaller volumes and very coarse lattices in comparison to the present work. It is also worth noting the "scheme-independent" prediction of $\gamma^\star_g \approx 0.228$ from Ref. [4] is also within 1σ of our predicted $\gamma^\star_g$; we don't show this result in Fig. 8 because no estimate of the systematic error in this result is available.

V. CONCLUSIONS

We have calculated the non-perturbative β function of the SU(3) gauge-fermion system with twelve massless fundamental fermions using a Pauli-Villars improved lattice action. We find strong evidence for an infrared fixed point at $g^2_{\rm GF\star} = 6.69(68)$ from our gradient-flow-based renormalization scheme. Our study utilizes a wide range of couplings and volumes. In particular, we can reach renormalized gauge coupling values well above the predicted IRFP without the interference of a bulk phase transition. We include systematic effects from the infinite volume extrapolation directly into our analysis using Bayesian model averaging. Our data exhibits cutoff effects that are consistent with the leading $\mathcal{O}(a^2/t)$ form over a wide range t/a^2. The consistency between the W and C operators further supports the leading order scaling behavior. We believe the improved scaling is due to

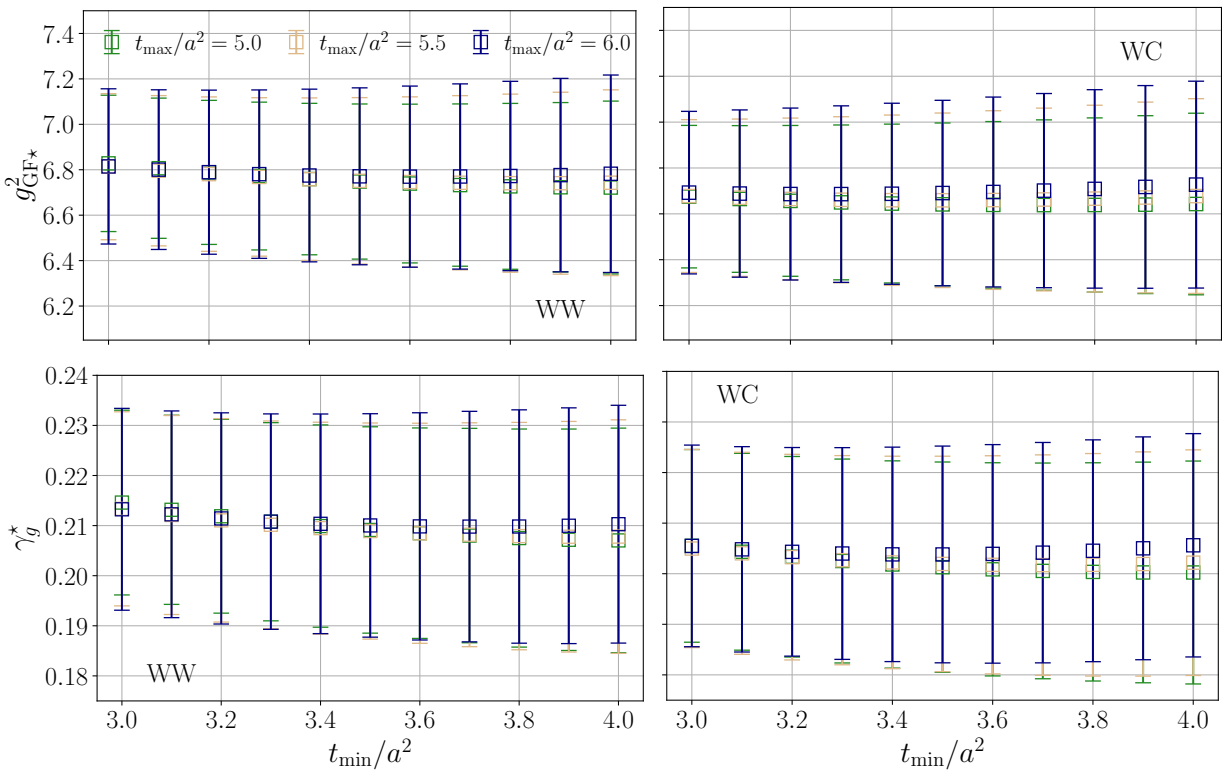


FIG. 7. Comparison of our estimated $g^2_{\mathrm{GF}\star}$ and $\gamma^\star_g$ for different $t_{\min}/a^2$ (x-axes) and $t_{\max}/a^2 = 5.0$ (green), 5.5 (gold), 6.0 (navy) from the continuum extrapolation.

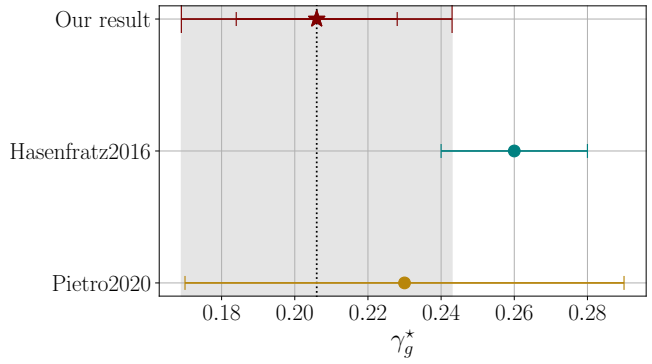


FIG. 8. Comparison of our value for $\gamma^\star_g = 0.204(36)$ (maroon errorbar) against Ref. [20] (teal error bar) and Ref. [5] (dark gold error bar). The smaller error bar on our result indicates the error without accounting for systematic effects; the larger error bar indicates our error after accounting for systematic effects. We indicate our total error with a grey band for visualization.

the additional PV bosons that reduce cutoff effects. In contrast, we found significantly larger cutoff effects when we re-analyzed the data that were generated without PV fields and used in Ref. [20]. Our data is publicly available at Ref. [62].

Overall, the systematics of our continuum extrapolation are well controlled. Based on the continuum prediction for $\beta\left(g^2_{\mathrm{GF}}\right)$ in g^2_{GF}, we estimate the leading irrelevant critical exponent $\gamma^\star_g$. This estimate includes systematic errors from various choices in our analysis. The value $\gamma^\star_g = 0.204(36)$ agrees with Refs. [4, 5] at the 1σ level and Ref. [20] at the 2σ level.

ACKNOWLEDGMENTS

Both authors acknowledge support by DOE Grant No. DE-SC0010005. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2040434. The research reported in this work made use of computing and long-term storage facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy. This work utilized the Alpine high-performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, and Colorado State University. We thank James Osborn and Xiaoyong Jin for writing `QEX` and helping us develop our `QEX`-based hybrid Monte Carlo and gradient flow codes.

[1] Thomas A. Ryttov and Robert Shrock, "Higher-loop corrections to the infrared evolution of a gauge theory with fermions," Phys. Rev. **D83**, 056011 (2011), arXiv:1011.4542.

[2] Claudio Pica and Francesco Sannino, "UV and IR Zeros of Gauge Theories at The Four Loop Order and Beyond," Phys. Rev. D **83**, 035013 (2011), arXiv:1011.5917 [hep-ph].

[3] Thomas A. Ryttov and Robert Shrock, "Infrared Zero of β and Value of γ_m for an SU(3) Gauge Theory at the Five-Loop Level," Phys. Rev. **D94**, 105015 (2016), arXiv:1607.06866 [hep-th].

[4] Thomas A. Ryttov and Robert Shrock, "Scheme-Independent Series Expansions at an Infrared Zero of the Beta Function in Asymptotically Free Gauge Theories," Phys. Rev. **D94**, 125005 (2016), arXiv:1610.00387 [hep-th].

[5] Lorenzo Di Pietro and Marco Serone, "Looking through the QCD Conformal Window with Perturbation Theory," JHEP **07**, 049 (2020), arXiv:2003.01742 [hep-th].

[6] Thomas Appelquist, Anuradha Ratnaweera, John Terning, and L. C. R. Wijewardhana, "The Phase structure of an SU(N) gauge theory with N(f) flavors," Phys. Rev. D **58**, 105017 (1998), arXiv:hep-ph/9806472.

[7] A. Bashir, A. Raya, and J. Rodriguez-Quintero, "QCD: Restoration of Chiral Symmetry and Deconfinement for Large N_f," Phys. Rev. D **88**, 054003 (2013), arXiv:1302.5829 [hep-ph].

[8] Jens Braun and Holger Gies, "Chiral phase boundary of QCD at finite temperature," JHEP **06**, 024 (2006), arXiv:hep-ph/0602226.

[9] Jens Braun, Christian S. Fischer, and Holger Gies, "Beyond Miransky Scaling," Phys. Rev. D **84**, 034045 (2011), arXiv:1012.4279 [hep-ph].

[10] Jong-Wan Lee, "Conformal window from conformal expansion," Phys. Rev. D **103**, 076006 (2021), arXiv:2008.12223 [hep-ph].

[11] Zhijin Li and David Poland, "Searching for gauge theories with the conformal bootstrap," (2020), arXiv:2005.01721 [hep-th].

[12] Seth Grable and Paul Romatschke, "Elements of Confinement for QCD with Twelve Massless Quarks," (2023), arXiv:2310.12203 [hep-th].

[13] Hee Sok Chung and Daniel Nogradi, "f$_\varrho$/m$_\varrho$ and f$_\pi$/m$_\varrho$ ratios and the conformal window," Phys. Rev. D **107**, 074039 (2023), arXiv:2302.06411 [hep-ph].

[14] Paul Romatschke, "An alternative to perturbative renormalization in 3+1 dimensional field theories," (2024), arXiv:2401.06847 [hep-th].

[15] Thomas Appelquist, George T. Fleming, and Ethan T. Neil, "Lattice study of the conformal window in QCD-like theories," Phys.Rev.Lett. **100**, 171607 (2008), arXiv:0712.0609 [hep-ph].

[16] Thomas Appelquist, George T. Fleming, and Ethan T. Neil, "Lattice Study of Conformal Behavior in SU(3) Yang-Mills Theories," Phys. Rev. D **79**, 076010 (2009), arXiv:0901.3766 [hep-ph].

[17] C. J. David Lin, Kenji Ogawa, Hiroshi Ohki, and Eigo Shintani, "Lattice study of infrared behaviour in SU(3) gauge theory with twelve massless flavours," JHEP **08**, 096 (2012), arXiv:1205.6076 [hep-lat].

[18] C. J. David Lin, Kenji Ogawa, and Alberto Ramos, "The Yang-Mills gradient flow and SU(3) gauge theory with 12 massless fundamental fermions in a colour-twisted box," JHEP **12**, 103 (2015), arXiv:1510.05755 [hep-lat].

[19] Zoltan Fodor, Kieran Holland, Julius Kuti, Santanu Mondal, Daniel Nogradi, and Chik Him Wong, "Fate of the conformal fixed point with twelve massless fermions and SU(3) gauge group," Phys. Rev. **D94**, 091501 (2016), arXiv:1607.06121 [hep-lat].

[20] Anna Hasenfratz and David Schaich, "Nonperturbative beta function of twelve-flavor SU(3) gauge theory," JHEP **02**, 132 (2018), arXiv:1610.10004 [hep-lat].

[21] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chik Him Wong, "The twelve-flavor β-function and dilaton tests of the sextet scalar," EPJ Web Conf. **175**, 08015 (2018), arXiv:1712.08594 [hep-lat].

[22] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chik Him Wong, "Extended investigation of the twelve-flavor β-function," Phys. Lett. **B779**, 230–236 (2018), arXiv:1710.09262 [hep-lat].

[23] A. Hasenfratz, C. Rebbi, and O. Witzel, "Nonperturbative determination of β functions for SU(3) gauge theories with 10 and 12 fundamental flavors using domain wall fermions," Phys. Lett. **B798**, 134937 (2019), arXiv:1710.11578 [hep-lat].

[24] Anna Hasenfratz, Claudio Rebbi, and Oliver Witzel, "Gradient flow step-scaling function for SU(3) with twelve flavors," Phys. Rev. **D100**, 114508 (2019), arXiv:1909.05842 [hep-lat].

[25] Anna Hasenfratz, "Conformal or Walking? Monte Carlo renormalization group studies of SU(3) gauge models with fundamental fermions," Phys. Rev. D **82**, 014506 (2010), arXiv:1004.1004 [hep-lat].

[26] Anna Hasenfratz, "Infrared fixed point of the 12-fermion SU(3) gauge model based on 2-lattice MCRG matching," Phys. Rev. Lett. **108**, 061601 (2012), arXiv:1106.5293 [hep-lat].

[27] Anna Hasenfratz and Oliver Witzel, "Continuous β function for the SU(3) gauge systems with two and twelve fundamental flavors," PoS **LATTICE2019**, 094 (2019), arXiv:1911.11531 [hep-lat].

[28] A. Deuzeman, M. P. Lombardo, and E. Pallante, "Evidence for a conformal phase in SU(N) gauge theories," Phys. Rev. D **82**, 074503 (2010), arXiv:0904.4662 [hep-ph].

[29] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chris Schroeder, "Twelve massless flavors and three colors below the conformal window," Phys. Lett. **B703**, 348–358 (2011), arXiv:1104.3124 [hep-lat].

[30] T. Appelquist, G.T. Fleming, M.F. Lin, E.T. Neil, and D.A. Schaich, "Lattice Simulations and Infrared Conformality," Phys.Rev. **D84**, 054501 (2011), arXiv:1106.2148 [hep-lat].

[31] Thomas DeGrand, "Finite-size scaling tests for spectra in SU(3) lattice gauge theory coupled to 12 fundamental flavor fermions," Phys.Rev. **D84**, 116901 (2011), arXiv:1109.1237 [hep-lat].

[32] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, Chris Schroeder, and Chik Him Wong, "Conformal finite size scaling of twelve fermion flavors," PoS **Lattice 2012**, 279 (2012), arXiv:1211.4238.

[33] Yasumichi Aoki, Tatsumi Aoyama, Masafumi Kurachi, Toshihide Maskawa, Kei-ichi Nagai, Hiroshi Ohki, Akihiro Shibata, Koichi Yamawaki, and Takeshi Yamazaki (LatKMI), "Lattice study of conformality in twelve-flavor QCD," Phys. Rev. **D86**, 054506 (2012), arXiv:1207.3060 [hep-lat].

[34] Yasumichi Aoki, Tatsumi Aoyama, Masafumi Kurachi, Toshihide Maskawa, Kei-ichi Nagai, Hiroshi Ohki, Enrico Rinaldi, Akihiro Shibata, Koichi Yamawaki, and Takeshi Yamazaki (LatKMI), "Light composite scalar in twelve-flavor QCD on the lattice," Phys. Rev. Lett. **111**, 162001 (2013), arXiv:1305.6006 [hep-lat].

[35] Anqi Cheng, Anna Hasenfratz, Yuzhi Liu, Gregory Petropoulos, and David Schaich, "Finite size scaling of conformal theories in the presence of a near-marginal operator," Phys.Rev. **D90**, 014509 (2014), arXiv:1401.0195 [hep-lat].

[36] M. P. Lombardo, K. Miura, T. J. Nunes da Silva, and E. Pallante, "On the particle spectrum and the conformal window," JHEP **12**, 183 (2014), arXiv:1410.0298 [hep-lat].

[37] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chris Schroeder, "Nearly conformal gauge theories in finite volume," Phys. Lett. B **681**, 353–361 (2009), arXiv:0907.4562 [hep-lat].

[38] Anqi Cheng, Anna Hasenfratz, Gregory Petropoulos, and David Schaich, "Scale-dependent mass anomalous dimension from Dirac eigenmodes," JHEP **1307**, 061 (2013), arXiv:1301.1355 [hep-lat].

[39] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chik Him Wong, "Case studies of near-conformal β-functions," PoS **LATTICE2019**, 121 (2019), arXiv:1912.07653 [hep-lat].

[40] Albert Deuzeman, Maria Paola Lombardo, Tiago Nunes Da Silva, and Elisabetta Pallante, "The bulk transition of QCD with twelve flavors and the role of improvement," Phys. Lett. B **720**, 358–365 (2013), arXiv:1209.5720 [hep-lat].

[41] Tiago Nunes da Silva and Elisabetta Pallante, "The strong coupling regime of twelve flavors QCD," PoS **LATTICE2012**, 052 (2012), arXiv:1211.3656 [hep-lat].

[42] David Schaich, Anqi Cheng, Anna Hasenfratz, and Gregory Petropoulos, "Bulk and finite-temperature transitions in SU(3) gauge theories with many light fermions," PoS **LATTICE2012**, 028 (2012), arXiv:1207.7164 [hep-lat].

[43] Tobias Rindlisbacher, Kari Rummukainen, and Ahmed Salami, "Bulk-preventing actions for SU(N) gauge theories," PoS **LATTICE2021**, 576 (2022), arXiv:2111.00860 [hep-lat].

[44] Tobias Rindlisbacher, Kari Rummukainen, and Ahmed Salami, "Bulk-transition-preventing actions for SU(N) gauge theories," Phys. Rev. D **108**, 114511 (2023), arXiv:2306.14319 [hep-lat].

[45] Felix Springer, David Schaich, and Enrico Rinaldi (Lattice Strong Dynamics (LSD)), "First-order bulk transitions in large-N lattice Yang–Mills theories using the density of states," (2023), arXiv:2311.10243 [hep-lat].

[46] Anna Hasenfratz, Yigal Shamir, and Benjamin Svetitsky, "Taming lattice artifacts with Pauli-Villars fields," Phys. Rev. D **104**, 074509 (2021), arXiv:2109.02790 [hep-lat].

[47] Anna Hasenfratz, Ethan T. Neil, Yigal Shamir, Benjamin Svetitsky, and Oliver Witzel, "Infrared fixed point and anomalous dimensions in a composite Higgs model," Phys. Rev. D **107**, 114504 (2023), arXiv:2304.11729 [hep-lat].

[48] Anna Hasenfratz, Ethan T. Neil, Yigal Shamir, Benjamin Svetitsky, and Oliver Witzel, "Infrared fixed point of the SU(3) gauge theory with Nf=10 flavors," Phys. Rev. D **108**, L071503 (2023), arXiv:2306.07236 [hep-lat].

[49] Robert V. Harlander and Tobias Neumann, "The perturbative QCD gradient flow to three loops," JHEP **06**, 161 (2016), arXiv:1606.03756 [hep-ph].

[50] R. Narayanan and H. Neuberger, "Infinite N phase transitions in continuum Wilson loop operators," JHEP **0603**, 064 (2006), arXiv:hep-th/0601210 [hep-th].

[51] Martin Lüscher, "Trivializing maps, the Wilson flow and the HMC algorithm," Commun.Math.Phys. **293**, 899–919 (2010), arXiv:0907.5491 [hep-lat].

[52] Martin Lüscher, "Properties and uses of the Wilson flow in lattice QCD," JHEP **1008**, 071 (2010), arXiv:1006.4518 [hep-lat].

[53] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chik Him Wong, "A new method for the beta function in the chiral symmetry broken phase," EPJ Web Conf. **175**, 08027 (2018), arXiv:1711.04833 [hep-lat].

[54] Andrea Carosso, Anna Hasenfratz, and Ethan T. Neil, "Nonperturbative Renormalization of Operators in Near-Conformal Systems Using Gradient Flows," Phys. Rev. Lett. **121**, 201601 (2018), arXiv:1806.01385 [hep-lat].

[55] Andrea Carosso, "Stochastic Renormalization Group and Gradient Flow," JHEP **01**, 172 (2020), arXiv:1904.13057 [hep-th].

[56] Anna Hasenfratz and Oliver Witzel, "Continuous renormalization group β function from lattice simulations," Phys. Rev. **D101**, 034514 (2020), arXiv:1910.06408 [hep-lat].

[57] Anna Hasenfratz, Christopher J. Monahan, Matthew David Rizik, Andrea Shindler, and Oliver Witzel, "A novel nonperturbative renormalization scheme for local operators," PoS **LATTICE2021**, 155 (2022), arXiv:2201.09740 [hep-lat].

[58] Julius Kuti, Zoltán Fodor, Kieran Holland, and Chik Him Wong, "From ten-flavor tests of the β-function to α_s at the Z-pole," PoS **LATTICE2021**, 321 (2022), arXiv:2203.15847 [hep-lat].

[59] Anna Hasenfratz, Curtis Taylor Peterson, Jake van Sickle, and Oliver Witzel, "Λ parameter of the SU(3) Yang-Mills theory from the continuous β function," Phys. Rev. D **108**, 014502 (2023), arXiv:2303.00704 [hep-lat].

[60] Chik Him Wong, Szabolcs Borsanyi, Zoltan Fodor, Kieran Holland, and Julius Kuti, "Toward a novel determination of the strong QCD coupling at the Z-pole," in *39th International Symposium on Lattice Field Theory* (2023) arXiv:2301.06611 [hep-lat].

[61] Curtis T. Peterson, Anna Hasenfratz, Jake van Sickle, and Oliver Witzel, "Determination of the continuous β function of SU(3) Yang-Mills theory," in *38th International Symposium on Lattice Field Theory* (2021) arXiv:2109.09720 [hep-lat].

[62] Curtis Peterson and Anna Hasenfratz, "Twelve flavor SU(3) gradient flow data for the continuous beta-function," (2024).

[63] Anna Hasenfratz and Francesco Knechtli, "Flavor symmetry and the static potential with hypercubic blocking," Phys. Rev. **D64**, 034504 (2001), arXiv:hep-lat/0103029.

[64] Anna Hasenfratz, Roland Hoffmann, and Stefan Schaefer, "Hypercubic smeared links for dynamical fermions,"

JHEP **0705**, 029 (2007), arXiv:hep-lat/0702028 [hep-lat].
[65] Anqi Cheng, Anna Hasenfratz, and David Schaich, "Novel phase in SU(3) lattice gauge theory with 12 light fermions," Phys.Rev. **D85**, 094509 (2012), arXiv:1111.2317 [hep-lat].
[66] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," Phys.Lett. **B195**, 216–222 (1987).
[67] J. Osborn and Xiao-Yong Jin, "Introduction to the Quantum EXpressions (QEX) framework," PoS **LATTICE2016**, 271 (2017).
[68] Zoltan Fodor, Kieran Holland, Julius Kuti, Daniel Nogradi, and Chik Him Wong, "The Yang-Mills gradient flow in finite volume," JHEP **1211**, 007 (2012), arXiv:1208.1051 [hep-lat].
[69] Peter Lepage, "gvar," (2015).
[70] Curtis Peterson, "SwissFit," `https://github.com/ctpeterson/SwissFit`.
[71] William I. Jay and Ethan T. Neil, "Bayesian model averaging for analysis of lattice field theory results," Phys. Rev. D **103**, 114502 (2021), arXiv:2008.01069 [stat.ME].
[72] Ethan T. Neil and Jacob W. Sitison, "Improved information criteria for Bayesian model averaging in lattice field theory," Phys. Rev. D **109**, 014510 (2024), arXiv:2208.14983 [stat.ME].
[73] Ethan T. Neil and Jacob W. Sitison, "Model averaging approaches to data subset selection," Phys. Rev. E **108**, 045308 (2023), arXiv:2305.19417 [stat.ME].

arXiv:2402.17864v1 [quant-ph] 27 Feb 2024

Casimir Physics beyond the Proximity Force Approximation: The Derivative Expansion

César D. Fosco,[1] Fernando C. Lombardo,[2,3] and Fransisc D. Mazzitelli[1]

[1] *Centro Atómico Bariloche, Instituto Balseiro, Comisión Nacional de Energía Atómica, 8400 Bariloche, Argentina*
[2] *Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina*
[3] *Instituto de Física de Buenos Aires (IFIBA), CONICET, Universidad de Buenos Aires, Argentina*

We review the derivative expansion (DE) method in Casimir physics, an approach which extends the proximity force approximation (PFA). After introducing and motivating the DE in contexts other than the Casimir effect, we present different examples which correspond to that realm. We focus on different particular geometries, boundary conditions, types of fields, and quantum and thermal fluctuations. Besides providing various examples where the method can be applied, we discuss a concrete example for which the DE cannot be applied; namely, the case of perfect Neumann conditions in 2 + 1 dimensions. By the same example, we show how a more realistic type of boundary condition circumvents the problem. We also comment on the application of the DE to the Casimir–Polder interaction which provides a broader perspective on particle–surface interactions.

I. INTRODUCTION

Casimir forces are one of the most intriguing macroscopic manifestations of quantum fluctuations in Nature. Their existence, first realized in the specific context of the interaction between the quantum electromagnetic (EM) field and the boundaries of two neutral bodies, manifests itself as an attractive force between them. That force depends, in an intricate manner, on the shape and EM properties of the objects. Since the discovery of this effect by Hendrik Casimir 75 years ago [1] this, and closely related phenomena, have been subjected to intense theoretical and experimental research [2–5]. The outcome of that work has not just revealed fundamental aspects of quantum field theory, but also subtle aspects of the models used to describe the EM properties of material bodies. Besides, it has become increasingly clear that this research has potential applications to nanotechnology.

Theoretical and experimental reasons have called for the calculation of the Casimir energies and forces for different geometries and materials [6], and with an ever increasing accuracy. The simplicity of the theoretical predictions when two parallel plates are involved, corresponds to a difficult experimental setup, due to alignment problems (in spite of this, the Casimir force for this geometry has been measured at the 10% accuracy level [7]). Equivalently, geometries which are more convenient from the experimental point of view, and allow for higher precision measurements, lead to more involved theoretical calculations. Such is the case of a cylinder facing a plane [8], or a sphere facing a plate, which is free from the alluded alignment problems [9–15].

From a theoretical standpoint, finding the dependence of the Casimir energies and forces on the geometry of the objects, poses an interesting challenge. Indeed, even when evaluating the self-energies which result from the coupling on an object to the vacuum field fluctuations, results may be rather non-intuitive; as in the case of a single spherical surface [16].

For a long time, calculations attempting to find analytical results for the Casimir and related interactions had been restricted to using the so called proximity force approximation (PFA). In this approach, the interaction energies and the resulting forces are computed approximating the geometry by a collection of parallel plates and then adding up the contributions obtained for this approximate geometry. This procedure was presumed to work well enough, at least for smooth surfaces when they are sufficiently close to each other; in more precise terms: when the curvature radii of the surfaces R_i are much larger than the distance d between them. Indeed, this is the main content of the Derjaguin approximation (DA), developed by Boris Derjaguin in the 1930s [17–19] , which is pivotal in the study of surface interactions, especially in the context of colloidal particles and biological cells. This approach has significant implications in understanding colloidal stability, adhesion, and thin film formation.

It is worth introducing some essentials of the DA, in particular, of the geometrical aspects involved. Assuming the interaction energy per unit area between two parallel planes at a distance h is known, and given by $E_{\shortparallel}(h)$, the DA yields an expression for the interaction energy between two curved surfaces, $U_{\rm DA}$ [2, 4, 17–20]. Indeed,

$$U_{\rm DA}(a) = 2\pi R_{\rm eff} \int_a^\infty E_{\shortparallel}(h) dh \,, \tag{1}$$

where a denotes the distance between the surfaces, R_1 and R_2 are their curvature radii (at the point of closest distance), while $R_{\rm eff} = R_1 R_2/(R_1 + R_2)$. It is rather straightforward to implement the approximation at the level of the force $f_{\rm DA}$ between surfaces:

$$f_{\rm DA}(a) = 2\pi R_{\rm eff} E_{\shortparallel}(a) \,. \tag{2}$$

This approximation is usually derived from a quite reasonable assumption, namely, that the interaction energy

can be approximated by means of the PFA expression:

$$U_{\rm PFA} = \int dS\, E_{_{\rm II}} \,. \tag{3}$$

Here, the surface integration may be performed over one of the participating surfaces, but it could also be over an imaginary, "interpolating surface", which lies between them. The DA is obtained from the expression above, by approximating the surfaces by (portions of) the osculating spheres (with radii R_1 and R_2) at the point of closest approach.

Based on this hypothesis, on dimensional grounds one can expect the corrections to the PFA to be of order $\mathcal{O}(a/R_i)$. Note, however, that since the PFA had not been obtained as the leading-order term in a well-defined expansion, the approximation itself did not provide any quantitative method to asses the validity of that assumption.

A need for reliable measure of the accuracy of the results obtained using different methods became increasingly crucial, specially since the development of the "precision era" in the measurement of the Casimir forces [9–15]. It was in this context that the Derivative Expansion (DE) approach, was first introduced by us in 2011 [21], as a tool to asses the validity of the PFA, by putting it in the framework of an expansion, and to calculate corrections to the PFA using that very same expansion. When one realizes that the PFA had previously been proposed in contexts which are rather different to Casimir physics, it becomes clear that the improvement on the PFA which represents the DE may and does have relevance on those realms, regardless of them having an origin in vacuum fluctuations or not. Indeed, when one strips off the DE of the particularities of Casimir physics, one can see the ingredients that allowed one to implement it are also found, for example, in electrostatics, nuclear physics, and colloidal surface interactions.

Here, we present the essential features of the DE, its derivation, and consider some examples of its applications. The review is organized as follows. In Section II , we recall some aspects of the DA which stem from its application to nuclear and colloidal physics. We start with the DA not just for historical reasons, but also because we believe that this sheds light on some geometrical aspects of the approximation, in a rather direct way (like the relevance of curvature radii and distances).

Then, in Section III, we introduce the DE in one of its simplest realizations, namely, in the context of electrostatics, for a system consisting of two conducting surfaces kept at different potentials [22]. We first evaluate the PFA in this example, and then introduce the DE as a method to improve on that approximation. In Section IV, we introduce a more abstract, and therefore more general, formulation of the DE [23]. By putting aside the particular features of an specific interaction, and keeping just the ones that are common to all of them, we are lead to formulate the problem as follows: the DE is a particular kind of expansion of a functional having as argument a surface (or surfaces). We mean "functional" here in its mathematical sense: a function that assigns a number to a function or functions. We elucidate and demonstrate some of the aspects of the DE in this general context; the purpose of presenting those aspects are not just a matter of consistency or justification, but they also provide a concrete way of applying and implementing the DE to any example where it is applicable.

Then, in Section V, we focus on the DE in the specific context of the Casimir interaction between surfaces, for perfect boundary conditions at zero temperature; i.e., vacuum fluctuations [21, 24]. Then in Section VI we review the extension of those results to the case of finite temperatures and real materials [25, 26]. As we shall see, the temperature introduces another scale, which affects the form one must adopt for the different terms in the DE. Then we comment on an aspect which first manifests itself here: as it happens with any expansion, it is to be expected to break down for some specific examples, when the hypothesis that justified it are not satisfied. We show this for the case of the Casimir effect with Neumann conditions at finite temperatures [26, 27]. We also show that the application of the DE to the EM field is free of this problem, if dissipative effects are included in the model describing the media [28].

The application of the DE to Casimir-Polder forces for atoms near smooth surfaces [29] is described in Section VII. Other alternatives to compute Casimir energies beyond PFA [30] are described in Section VIII. Section IX contains our conclusions.

II. PROXIMITY APPROXIMATIONS IN NUCLEAR AND COLLOIDAL PHYSICS

The introduction of the Derjaguin Approximation (DA) to nuclear physics dates back to the seminal paper [31]. In this paper, the DA was rediscovered and applied to calculate nuclear interactions, starting with a Derjaguin-like formula for the surface interaction energies. The approach was based on a crucial "universal function" - a term referring here to the interaction energy between flat surfaces, calculated using a Thomas-Fermi approximation. In spite of the rather different context, the analogy with the approach followed in the DA becomes clear when one introduces three surfaces, the physical ones, Σ_L and Σ_R, and the intermediate one Σ which one uses to parametrize the interacting ones. Then, if the physical surfaces are sufficiently smooth, the interaction energy should, to a reasonable approximation, be described by the PFA, in a similar fashion as in Equation (3). To render the assertion above more concrete, we yet again use the function $h : \Sigma \to \mathbb{R}$, measuring the distance between Σ_L and Σ_R at each point on Σ. Since h will have level sets which are, except for a zero measure set, one-dimensional (closed curves), and the interaction depends just on h, the PFA expression for the interaction

energy U may be rendered as a one-dimensional integral:

$$U_{\rm PFA} = \int dh\, J(h)\, E_{\shortparallel}(h)\ , \tag{4}$$

where $J(h)dh$ is the infinitesimal area between two level curves on Σ: the ones between h and $h+dh$, while $E_{\shortparallel}$ is the universal function.

We now assume that Σ is a plane, and that the physical surfaces may be both described by means of just one Monge patch based on Σ. This surface is then naturally thought of (in descriptive geometry terms) as the projection plane. Using Cartesian coordinates $(x_1, x_2) \equiv \mathbf{x}_{\shortparallel}$ on Σ, assuming (for smooth enough surfaces) that J may be regarded as constant, and using a second-order Taylor expansion of h around a (the distance of closest approach):

$$h(\mathbf{x}_{\shortparallel}) \simeq a + \frac{1}{2}\left(\frac{x_1^2}{R_1} + \frac{x_2^2}{R_2}\right) \tag{5}$$

produces, when evaluating the PFA interaction energy (4), the DA energy (1). Here, R_1 and R_2 are the radii of curvature of the surface by $x_3 = h(\mathbf{x}_{\shortparallel})$ at $x_3 = a$.

This result may be improved, even within the spirit of the PFA, by introducing some refinements. Indeed, in [32], a generalization of the PFA has been introduced such that the starting point was Equation (4), but now allowing for the surfaces to have larger curvatures, as long as they remained almost parallel locally. The main difference that follows from those weaker assumptions is that, now, the Jacobian J may become a non-trivial function of h. For instance, introducing a linear expansion:

$$J(h) \approx J_0 + J_1 h\ , \tag{6}$$

a straightforward calculation shows that the force f becomes:

$$f_{\rm PFA}(a) \;=\; J_0 E_{\shortparallel}(a) - J_1(a) \int_a^{\infty} dh\, E_{\shortparallel}(h)\ . \tag{7}$$

Note that the result is the sum of the DA term plus a second term proportional to the derivative of the Jacobian with respect to h. This is a correction to the DA obtained from the same starting point we used for the DA: $U_{\rm PFA}$. In other words, Equation (7) is still determined by the energy density for parallel plates. As we shall see, the DE will introduce corrections that go beyond $E_{\shortparallel}(a)$. The correction will depend on both the geometry and the nature of the interaction.

We wish to point out that the lack of knowledge of an exact expression for $E_{\shortparallel}$ is not specific to nuclear physics, but of course it may appear in other applications. The general PFA approach can nevertheless be introduced; the accuracy of its predictions will then be limited not just by the fulfillment or not of the geometrical assumptions, but also by the reliability of the expression for $E_{\shortparallel}$. Using different approximations for $E_{\shortparallel}$ gives as many results for the PFA. For a recent review in the case of nuclear physics, see Ref. [33, 34].

An apparently unrelated approximation, based on different physical assumptions, was introduced in the context of colloidal physics. Let us now see how it yields a result which agrees with the DA: it is the so called Surface Element Integration (SEI) [35], or Surface Integration Approach (SIA) [36]. This approach may be introduced as follows: let us consider a compact object facing the $x_3 = 0$ plane. x_3 is then the normal coordinate to the plane, pointing towards the compact object. With this conventions, the SEI approximation applied to the interaction energy amounts to the following:

$$U_{\rm SEI} = -\int_{\rm plane} dx_1 dx_2\, \frac{\hat{n}\cdot\hat{e}_3}{|\hat{n}\cdot\hat{e}_3|}\, E_{\shortparallel}\ . \tag{8}$$

Here, $\hat{n}$ denotes the outwards pointing unit normal to each surface element of the object. We see that, when the compact object may be thought of as delimited by just two surfaces, one of them facing the plane and the other away from it, the SEI consists of the difference between the PFA energies of those surfaces. This (possibly startling) fact is, as we shall see, related to the fact that the SEI becomes exact for almost transparent bodies, a situation characterized by the fact that the interaction is the result of adding all the (volumetric) pairwise contributions.

In the context of colloidal physics , the SEI method relies heavily upon the existence of a pressure on the compact object. The effect of that pressure should be integrated over the closed surface surrounding the compact object, in order to find the total force [35]. An alternative route to understand the SEI is to showthat Equation (8) becomes exact when the interaction between macroscopic bodies is the superposition of the interactions for the pair potentials of their constituents [36]. That may be interpreted by using a simple example. Consider two media, one of them, the left medium L, corresponding to the $x_3 \le 0$ half-space, while the right medium, R, is defined as the region:

$$R \;=\; \{(x_1, x_2, x_3) : \psi_1(\mathbf{x}_{\shortparallel}) \le x_3 \le \psi_2(\mathbf{x}_{\shortparallel})\}\ . \tag{9}$$

The interaction energy U is a functional of the two functions $\psi_{1,2}$. When the media are diluted, we expect the interaction energy to have the form

$$U[\psi_1, \psi_2] \;=\; \int d^2\mathbf{x}_{\shortparallel}\, (E_{\shortparallel}(\psi_1) - E_{\shortparallel}(\psi_2))\ , \tag{10}$$

where $E_{\shortparallel}(a)$ is the interaction energy per unit area, between two half-spaces at a distance a. This formula can be interpreted as follows: to obtain the interaction energy for the configuration described by ψ_1 and ψ_2, one must certainly subtract from $E_{\shortparallel}(\psi_1)$ the contributions from $x_3 > \psi_2$. This "linearity" is expected to be valid only for dilute media, and in that situation it coincides with the result obtained using the SEI. One expects then the SEI to give an exact result for almost-transparent media, for which the superposition principle holds true, and the total interaction energy is due to the sum of all the different

pairwise potentials [36]. It is worth noting, at this point, the important fact that the PFA also becomes exact in Casimir physics when the media constituting the objects are dilute. Indeed, this has been pointed out in [37, 38].

The examples just described illustrate the relevance of the DA, and of some of its variants, to different areas of physics. At the same time, the main drawback is made rather evident: in spite of being based on reasonable physical assumptions, it is difficult to assess its validity. The reason for this difficulty is that the approximation is uncontrolled, and therefore the estimation of the error incurred is difficult, within a self-contained approach.

The DE provides a systematic method to improve the PFA, and to compute its next-to-leading-order (NTLO) correction in a consistent set up.

III. INTRODUCING THE DERIVATIVE EXPANSION

A. The PFA in an Electrostatic Example

We introduce the PFA, and then the DE, in an example which neatly illustrates the DE main aspects, in the context of electrostatics. Here, contrary to what happens when dealing with more involved systems, like, say, Van der Waals, nuclear or Casimir forces, the physical assumptions and their implementations are more transparent. We follow closely Ref.[22]

The set-up we want to describe consists of two perfectly-conducting surfaces, one of them an infinite grounded plane and the other a smoothly curved surface kept at an electrostatic potential V_0. We use coordinates such that the plane corresponds to $z = 0$ while the smooth surface is such that it can be described by a single function, namely, by an equation of the form $z = \psi(x_{\shortparallel})$. The electrostatic energy contained between surfaces can then be written as follows:

$$U = \frac{\epsilon_0}{2} \int d^2\mathbf{x}_{\shortparallel} \int_0^{\psi(\mathbf{x}_{\shortparallel})} dz \, |\mathbf{E}|^2 \,, \tag{11}$$

where ϵ_0 denotes the permittivity of vacuum. In terms of U and V_0, the capacitance C of the system is then given by $C = 2U/V_0^2$.

Let us see how one implements the PFA in order to calculate U (from which one can extract, for instance, an approximate expression for C) expecting it to be accurate when the distance between the two surfaces is shorter than the curvature radius of the curved conductor. To that end, one first finds and approximation to the electric field between the conductors, by proceeding as follows: the smooth conductor is regarded as a set of parallel plates (Fig.1), in the sense that the electric field $\mathbf{E}$ points along the z direction and has a z-independent value. The electric field does, however, depend on $\mathbf{x}_{\shortparallel}$ since it is assumed to have, for every $\mathbf{x}_{\shortparallel}$, the same intensity as the electric field due to two (infinite) conducting planes at a distance $\psi(\mathbf{x}_{\shortparallel})$. Namely, $\mathbf{E}(\mathbf{x}) = -V_0/\psi(\mathbf{x}_{\shortparallel})\,\hat{\mathbf{z}}$. Therefore, the approximated expression for the electrostatic energy becomes:

$$U_{\rm PFA} = \frac{\epsilon_0 V_0^2}{2} \int d^2\mathbf{x}_{\shortparallel} \, \frac{1}{\psi(\mathbf{x}_{\shortparallel})} \,. \tag{12}$$

It is implicitly understood in the equation above, that the region to integrate is such that the assumption on the distance and curvature is satisfied. On the contrary, regions such that the assumption is not satisfied can be consistently ignored (see the example below).

It should be evident that Equation (12) provides a rather convenient tool to obtain estimates for the electrostatic energy in many relevant situations. Indeed, to illustrate this point we consider a cylinder of length L and radius R in front of a plane, and denote by a the minimum distance between the two surfaces. The cylinder is not a surface that can be described by a single patch; namely, one needs at least two functions. However, in the context of the PFA, it is reasonably to assume that only the half that is closer to the plane should be relevant. Assuming the axis of the cylinder to be along y, the function ψ reads:

$$\psi(x) = a + R\left(1 - \sqrt{1 - \frac{x^2}{R^2}}\right) \,, \tag{13}$$

with the variable x assumed to be in the range $-x_M < x < x_M < R$. Note that for $x_M/R = O(1) < 1$ the assumption on the distance and the curvature is satisfied. It is to be expected that, as long as $R \gg a$ (where the PFA gives an accurate value of the electrostatic energy), the final result will not depend on x_M. This can be readily checked by inserting Equation (13) into Equation (12), computing the integral, and expanding that result for $a \ll R$. Doing this we obtain:

$$U^{\rm cp}_{\rm PFA} \approx \frac{\epsilon_0 V_0^2 L \pi}{\sqrt{2}} \sqrt{\frac{R}{a}} \,, \tag{14}$$

which is independent of x_M. An immediate consequence of this is that, when the cylinder approaches the plane, the electrostatic force behaves as $a^{-3/2}$.

Let us check now the accuracy of $U^{\rm cp}_{\rm PFA}$. We take advantage of the knowledge of the exact expression for the electrostatic interaction energy:

$$U^{\rm cp} = \frac{\pi L \epsilon_0 V_0^2}{\operatorname{arccosh}\left(1 + \frac{a}{R}\right)} \,. \tag{15}$$

For $a/R \ll 1$, $U^{\rm cp}$ yields the PFA result $U^{\rm cp}_{\rm PFA}$ (14). The relevance of the corrections to the PFA can be estimated by expanding the exact result, but keeping also the next-to-leading order (NTLO) when $a << R$:

$$U^{\rm cp} \approx \frac{\epsilon_0 V_0^2 L \pi}{\sqrt{2}} \sqrt{\frac{R}{a}} \left(1 + \frac{1}{12}\frac{a}{R}\right) \,. \tag{16}$$

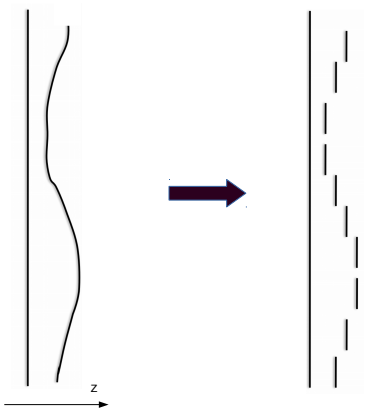


Figure 1. In the PFA, the interaction between a smoothly curved surface and a plane is approximated by that of a set of parallel plates. For each pair of parallel plates, border effects are ignored.

We will now introduce the DE. By construction, it should produce the NTLO result (for this an other surfaces), without resorting to the expansion of any exact expression (the knowledge of which, needless to say, is usually lacking).

B. Improvement of the PFA Using a Derivative Expansion

We begin by noting that the electrostatic energy is a functional of the function which defines the shape of the surface. A second observation is that, in principle, there is no reason to assume that the functional is local in ψ. Here, "local" means that it contains just one integral over $\mathbf{x}_{\shortparallel}$ of a sum of terms involving powers of $\psi(\mathbf{x}_{\shortparallel})$ and derivatives at $\psi(\mathbf{x}_{\shortparallel})$. On the contrary, the exact functional will generally involve terms where, for example, there are two or more integrals over $\mathbf{x}_{\shortparallel}$, and kernels depending of those variables, and products of ψ with different arguments. However, regardless of the non locality of the exact expression, it must become local when the surfaces are sufficiently smooth and close to each other. Indeed, if the PFA becomes valid asymptotically in that limit, then the energy must approach a result which is a local function of ψ. Not whatever local functional but just one without derivatives.

The way we found to depart slightly but significantly from the PFA, has been to add terms involving derivatives of ψ. Namely, we shall assume that the electrostatic energy can be expanded in local terms involving derivatives of ψ. One can think of the condition $|\nabla\psi| \ll 1$, as introducing a small, dimensionless expansion parameter. In physical terms, this means that the curved surface is almost parallel to the plane on the points where it is satisfied.

To introduce the first departure from the PFA, we include terms with up to two derivatives. Then the electrostatic energy has to be (up to this order) of the form:

$$U_{\rm DE} \simeq \int d^2\mathbf{x}_{\shortparallel} \left[V(\psi) + Z(\psi)|\nabla\psi|^2\right] , \qquad (17)$$

for some functions V and Z. The gradient is the two-dimensional one, and it can only appear in such a way that the energy is a scalar (ψ is a scalar under changes of coordinates on the plane). Besides, recalling the equations of electrostatics, and on dimensional grounds, the result must be proportional to $\epsilon_0 V^2$. On top of that it must reproduce $U_{\rm PFA}$ for constant ψ. Furthermore, as ψ is the only other dimensionful quantity, both functions V and Z have to be proportional to ψ^{-1}. Thus, we have restricted even further the functional to:

$$U_{\rm DE} \simeq \frac{\epsilon_0 V_0^2}{2} \int d^2\mathbf{x}_{\shortparallel} \frac{1}{\psi}(1 + \beta_{\rm E}|\nabla\psi|^2) , \qquad (18)$$

where $\beta_{\rm E}$ is a numerical coefficient to be determined (the subindex E stands for electrostatics). It is worth stressing that it is independent of the specific surface being considered, as long as it is smooth. Therefore, it can be obtained once and for all just from its evaluation for a particular case. A simple procedure to obtain the coefficient $\beta_{\rm E}$, when an exact analytic solution to the problem is known, would be to retrieve its value by expanding that solution. Let us do that for the configuration of a cylinder in front of a plane. Inserting Equation (13) into Equation (18), and performing the integrals, an expansion of the result in powers of a/R, allow us to fix $\beta_{\rm E}$. Indeed, in order to agree with the expansion of the exact result in Equation (16), this fixes its value to $\beta_{\rm E} = 1/3$. Of course, one will obtain the same value for any other particular example for which the exact solution was known.

It is worthy of noting that, since the DE is a perturbative approach, it should be desirable to have a perturbative method to calculate the coefficient $\beta_{\rm EM}$. In other words, to compute it from first principles, using the appropriate expansion. One can do that, for instance, by solving perturbatively the Laplace equation and then resorting to the method described in Section V. We have performed that calculation in Ref.[22], and refer the reader to that work for details, and also for the application of the DE to other electrostatic examples.

C. Two Smooth Surfaces

As a natural generalization of the previously discussed situation, let us now consider two surfaces described by the two functions $\psi_1(\mathbf{x}_{\shortparallel})$ and $\psi_2(\mathbf{x}_{\shortparallel})$, each one of them measuring the respective height of a surface with respect to a reference plane Σ. This geometry was first considered in the context of the DE for the Casimir effect in Ref.[24].

To construct the DE for the electrostatic energy in this case, we keep up to two derivatives of the functions. This

allows we to write the general expression:

$$U_{\rm DE}[\psi_1,\psi_2] \;=\; \int_\Sigma d^2\mathbf{x}_{\shortparallel}\, U_{\shortparallel}(\psi)\, \left[1+\beta_1|\nabla\psi_1|^2 + \beta_2|\nabla\psi_2|^2+\beta_\times\nabla\psi_1\cdot\nabla\psi_2 + \beta_-\,\hat{\mathbf{z}}\cdot\nabla\psi_1\times\nabla\psi_2)+\cdots\right], \qquad (19)$$

where $\psi = |\psi_2-\psi_1|$ is the height difference, $U_{\shortparallel}(\psi) = \epsilon_0 V_0^2/(2\psi)$ is the electrostatic energy between parallel plates, and the dots denote higher derivative terms. Equation (19) actually contains four numerical constants: β_1, β_2, $\beta_\times$, and β_-. However, symmetry considerations imply some constraints on them: the energy must be invariant under the interchange of ψ_1 and ψ_2, since that is just a relabeling: $\beta_1=\beta_2$ and $\beta_-=0$. Furthermore, in order to reproduce the result for a single smooth surface in front of a plane we must have $\beta_1=\beta_2=1/3$. The coefficient $\beta_\times$ can be determined taking into account that the energy should be invariant under a simultaneous rotation of both surfaces [24]. Indeed, for an infinitesimal rotation of each surface by an angle ϵ in the plane (x,z), the changes induced on the functions ψ_i are $\delta\psi_i = \epsilon(x+\psi_i\partial_x\psi_i)$, for $i=1,2$. To simplify the determination of $\beta_\times$ we can assume that, initially, $\psi_1=0$ and that ψ_2 is only a function of x. Computing explicitly the variation of $U_{\rm DE}$ to linear order in ϵ one can show that

$$\delta U_{\rm DE}=0 \Rightarrow \beta_\times = 1/3\,, \qquad (20)$$

and therefore

$$U_{\rm DE}[\psi_1,\psi_2] = \frac{\epsilon_0 V_0^2}{2}\int_\Sigma d^2\mathbf{x}_{\shortparallel}\,\frac{1}{\psi}\,\left[1+\frac{1}{3}\left(|\nabla\psi_1|^2+|\nabla\psi_2|^2+\nabla\psi_1\cdot\nabla\psi_2\right)\right]. \qquad (21)$$

Note that, by taking the variation of the electrostatic energy Equation (19) with respect to translations or rotations of one of the surfaces, one can obtain the vertical and lateral components of the force, as well as the torque, due to the remaining surface.

The identities $\beta_1=\beta_2$ and $\beta_-=0$ are universally valid, regardless of the interaction (as long as the surfaces are of an identical nature), but $\beta_\times=\beta_1$ holds true for the electrostatic interaction. This depends upon the fact that the leading term is proportional to ψ^{-1} (i.e. it is then not valid for the Casimir energy).

For later use, let us recall that, for a general function $E_{\shortparallel}(\psi)$, the relation between the different coefficients becomes [24]:

$$2(\beta_1+\beta_2)+2\beta_\times+\psi\frac{d\log E_{\shortparallel}}{d\psi}=1\,. \qquad (22)$$

The relation (22) shows that, for any interaction, the DE for the interaction energy between two curved surfaces can be reduced to the problem of a single surface in front of a plane. Indeed, in the later case one can determine β_1 and β_2, while Equation (22) determines the remaining coefficient $\beta_\times$.

To summarize: when computing the electrostatic energy associated with a configuration of two conductors at different potentials, with smoothly curved surfaces, one can go beyond the PFA by simply assuming that the energy admits an expansion in derivatives of the functions that define the shapes of the conductors. If the exact electrostatic energy for a single non trivial curved configuration is known, one can determine all the free parameters in the expansion.

Finally, the NTLO correction produces an appreciable improvement in the DA and, by the same token, also provides an assessment for its validity. An interesting alternative approach to compute electrostatic forces beyond the PFA can be found in Ref. [39].

IV. OBTAINING THE DE FROM A PERTURBATIVE EXPANSION

Regardless of the interaction considered, the DA and its improvement, the DE, can be obtained by performing the proper resummation of a perturbative expansion [23]. The required expansion is in powers of the departure of the surfaces, about a two flat parallel planes configuration. This connection yields a systematic and quite general approach to obtain the DE, even when an exact solution is not available.

To keep things general, we work with a general functional of the surface; that functional may correspond to an energy, free energy, force, etc. Besides, we do not make any assumption about the kind of interaction involved, not even about whether it satisfies a superposition principle or not.

To begin, let us we assume a geometry where there are two surfaces, one of which, L, is a plane, which with a proper choice of Cartesian coordinates (x_1, x_2, x_3), is described by $x_3=0$. The other one, R, is assumed to be describable by $x_3=\psi(\mathbf{x}_{\shortparallel})$.

The object for which we implement the approximation is denoted by $F[\psi]$, a functional of ψ. Then we note that the PFA for F, to be denoted here by F_0, is obtained as follows: add, for each $\mathbf{x}_{\shortparallel}$, the product of a local surface density $\mathcal{F}_0(\psi(\mathbf{x}_{\shortparallel}))$ depending only on the value of ψ at the point $\mathbf{x}_{\shortparallel}$, times the surface element area; namely,

$$F_0[\psi] \;=\; \int d^2\mathbf{x}_{\shortparallel}\,\mathcal{F}_0(\psi(\mathbf{x}_{\shortparallel}))\,. \qquad (23)$$

The surface density is, in turn, determined by the (assumed) knowledge of the exact form of F for the case of two parallel surfaces, as follows:

$$\mathcal{F}_0(a) \;=\; \lim_{\mathcal{S}\to\infty}\big[\frac{F[a]}{\mathcal{S}}\big]\;, \tag{24}$$

where $\mathcal{S}$ denotes the area of the L plate and a is a constant. Namely, to determine the density one needs to know the functional F just for constant functions $\psi \equiv a$. Note that, if the functional F is the interaction energy between the surfaces, $\mathcal{F}_0$ becomes the interaction energy per unit area $E_{\shortparallel}$, and F_0 becomes $U_{\rm PFA}$ (see Equation (3)).

Let us now show how to derive the PFA (and its corrections) by the resummation of a perturbative expansion. To that end, we evaluate F for a ψ having the form:

$$\psi(\mathbf{x}_{\shortparallel}) = a + \eta(\mathbf{x}_{\shortparallel})\;. \tag{25}$$

and write the resulting perturbative expansion in powers of η, which has the general form:

$$F[\psi] = \mathcal{S}\mathcal{F}_0(a) + \sum_{n\geq 1}\int \frac{d^2k_{\shortparallel}^{(1)}}{(2\pi)^2}...\frac{d^2k_{\shortparallel}^{(n)}}{(2\pi)^2}\,\delta(k_{\shortparallel}^{(1)}+...+k_{\shortparallel}^{(n)})\,h^{(n)}(k_{\shortparallel}^{(1)},...,k_{\shortparallel}^{(n)})\,\widetilde{\eta}(k_{\shortparallel}^{(1)})...\widetilde{\eta}(k_{\shortparallel}^{(n)})\;, \tag{26}$$

where $\delta(\cdot)$ is the Dirac delta function, and the form factors $h^{(n)}$ can be computed by using perturbative techniques. For the Dirichlet-Casimir effect, this can be done in a rather systematic way [40]. Although the approach to follow in order to obtain those form factors may depend strongly on the kind of system considered, the form of the expansion shall be the same. Note that the form factors may depend on a, although, in order to simplify the notation, we will not make that dependence explicit.

Up to now, we have not used the hypothesis of smoothness of the R surface. We do that now by assuming that the Fourier transform $\widetilde{\eta}$ is peaked at the zero momentum. What follows is to make use of this assumption for all terms in the expansion. In Equation (26), we set then: $h^{(n)}(k_{\shortparallel}^{(1)},...,k_{\shortparallel}^{(n)}) \simeq h^{(n)}(0,...,0)$, and, as a consequence:

$$F(\psi) \simeq \mathcal{S}\mathcal{F}_0(a) + \sum_{n\geq 1} h^{(n)}(0,..,0)\int d^2\mathbf{x}_{\shortparallel}[\eta(\mathbf{x}_{\shortparallel})]^n\;. \tag{27}$$

One could evaluate the form factors at the zero momentum straighforwardly. However, there is a shortcut here that allows one to obtain all of them immediately: consider a constant $\eta(\mathbf{x}_{\shortparallel}) = \eta_0$, so that the interaction energy is given by Equation (27) with the replacement $\int d^2\mathbf{x}_{\shortparallel}\,\eta(\mathbf{x}_{\shortparallel})^n \to \mathcal{S}\eta_0^n$. For this particular case, F becomes just the functional corresponding to parallel plates, which are separated by a distance $a+\eta_0$:

$$\mathcal{F}_0(a+\eta_0) = \mathcal{F}_0(a) + \sum_{n\geq 1} h^{(n)}(0,..,0)\eta_0^n\;. \tag{28}$$

We then conclude that, in this low-momentum approximation, the series can be summed up with the result:

$$F_0[\psi] \simeq \int d^2\mathbf{x}_{\shortparallel}\mathcal{F}_0\left(a+\eta(\mathbf{x}_{\shortparallel})\right) = \int d^2\mathbf{x}_{\shortparallel}\mathcal{F}_0(\psi)\;, \tag{29}$$

which is just the PFA.

The calculation just above shows that, for the class of geometries considered in this paper, the PFA can be justified from first principles as the result of a resummation of a perturbative calculation corresponding to almost flat surfaces. In order to be well defined, the PFA requires that the form factors $h^{(n)}(k_{\shortparallel}^{(1)},...,k_{\shortparallel}^{(n)})$ have a finite limit as $k_{\shortparallel}^{(i)} \to 0$.

This procedure also suggests how the PFA could be improved; one can include the NTLO terms in the low-momentum expansions of the form factors. We assume that they can be expanded in powers of the momenta up to the second order. We stress that this is by no means a trivial assumption. Indeed, depending on the the interaction considered, the form factors could include nonanalyticities (we will discuss some explicit examples below). In case of no nonanalyticities, one can introduce the expansions:

$$h^{(n)}(k_{\shortparallel}^{(1)},...,k_{\shortparallel}^{(n)}) = h^{(n)}(0,...,0) + \sum_{i,\alpha} A_{i\alpha}^{(n)}k_{\shortparallel\,\alpha}^{(i)} + \sum_{i,j,\alpha,\beta} B_{ij\alpha\beta}^{(n)}k_{\shortparallel\,\alpha}^{(i)}k_{\shortparallel\,\beta}^{(j)}\;\ldots, \tag{30}$$

for some $a-$dependent coefficients $A_{i\alpha}^{(n)}$ and $B_{ij\alpha\beta}^{(n)}$. Here $i,j=1,...,n$ label arguments while $\alpha,\beta=1,2$ label their components. Symmetry considerations are crucial, since they allow us to simplify the above expression (30), as fol-

lows: rotational invariance implies that the form factors depend only on the scalar products $k_{\shortparallel}^{(i)} \cdot k_{\shortparallel}^{(j)}$. Additionally, they have to be symmetric under the interchange of any two momenta. This thus leads to

$$h^{(n)}(k_{\shortparallel}^{(1)}, ..., k_{\shortparallel}^{(n)}) = h^{(n)}(0, ..., 0) + B^{(n)} \sum_i k_{\shortparallel}^{(i)\,2} + C^{(n)} \sum_{i \neq j} k_{\shortparallel}^{(i)} \cdot k_{\shortparallel}^{(j)} \,, \tag{31}$$

for some coefficients $B^{(n)}$ and $C^{(n)}$.

Inserting Equation (31) into Equation (26) and taking integrations by parts, one then finds the form of the first correction to the PFA:

$$F_2[\psi] = \int d^2\mathbf{x}_{\shortparallel} \left[\sum_{n \geq 2} D^{(n)} \, \eta^{n-2} \right] |\nabla \eta|^2 \,, \tag{32}$$

where the coefficients $D^{(n)}$ are linear combinations of $B^{(n)}$ and $C^{(n)}$. The subindex 2 in F indicates that this is the part of the functional containing two derivatives.

We complete the calculation by calculating the sum in Equation (32). To that end, we evaluate the correction F_2 for a particular case: $\eta(\mathbf{x}_{\shortparallel}) = \eta_0 + \epsilon(\mathbf{x}_{\shortparallel})$, with $\epsilon \ll \eta_0$, and expand up to the second order in ϵ. Thus,

$$F_2[a + \eta_0 + \epsilon] = \int d^2\mathbf{x}_{\shortparallel} \left[\sum_{n \geq 2} D^{(n)} \, \eta_0^{n-2} \right] |\nabla \epsilon|^2 \,. \tag{33}$$

The resummation can be obtained in this case, by considering the usual perturbative evaluation of the interaction energy up to second order in ϵ. This evaluation does, naturally, depend on the interaction considered, but, once one has that result one can obtain the sum of the series above. We we will denote by Z that sum, namely:

$$Z(a + \eta_0) \equiv \sum_{n \geq 2} D^{(n)} \, \eta_0^{n-2} \,. \tag{34}$$

Upon replacement $\eta_0 \to \eta$ in Equation (34), one obtains

$$F_2[\psi] = \int d^2\mathbf{x}_{\shortparallel} \, Z(\psi) |\nabla \psi|^2 \,. \tag{35}$$

This is the NTLO correction to the PFA. This concludes our systematic derivation of the PFA, including its first correction, a result which may be put as follows:

$$F_{\rm DE}[\psi] \;=\; \int d^2\mathbf{x}_{\shortparallel} \left[V(\psi) \,+\, Z(\psi) |\nabla \psi|^2 \right] , \tag{36}$$

where $V(\psi) = \mathcal{F}_0(\psi)$ is determined from the (known) expression for the interaction energy between parallel surfaces, while $Z(\psi)$ can be computed using a perturbative technique. In practice, $Z(\psi)$ can be evaluated setting $\eta_0 = 0$ in Equation (34).

The higher orders may be derived by an extension of the procedure described just above. It should be evident that, for the expansion to be well-defined, the analytic structure of the form factors is quite relevant. Indeed, the existence of nonanalytic zero-momentum contributions can render the DE non applicable. This should be expected on physical grounds, since the presence of nonanalytic terms implies that the functional cannot be approximated, in coordinate space, by the single integral of a local density. Physically, it is a signal that the nonlocal aspects of the interaction cannot be ignored. That should not come up as a surprise, when one recalls that the same kind of phenomenon does happen when evaluating the effective action in quantum field theory, and the quantum effects contain contributions due to virtual massless particles. In this case, the effective action may develop nonanalyticities at zero momentum.

The main messages of this Section are the following: irrespective of the nature of the interaction, the energy and forces between objects are functionals of their shapes. The PFA is recovered when the form factors of the functionals are evaluated at zero momentum. Enhancements to this approximation are achievable by expanding these form factors at low momenta. If the expansion is analytic, a resummation of the form factors produces the DE.

V. DE FOR THE ZERO-TEMPERATURE CASIMIR EFFECT

The application of the DE to the Casimir interaction energy between two objects was, actually, our original motivation to introduce the approximation, and it is useful briefly review some aspects of this application here. We consider first a real vacuum scalar field satisfying Dirichlet boundary conditions (Section V A) and then we move to the EM field with perfect-conductor boundary conditions (Section V B). We follow Ref. [21] for the derivation of the DE in the Dirichlet case.

A. Scalar Field with Dirichlet Boundary Conditions

We consider here a massless real scalar field φ in $3+1$ dimensions, coupled to two mirrors which impose Dirichlet boundary conditions. In our Euclidean conventions, we use x_0, x_1, x_2, x_3 to denote the spacetime coordinates, x_0 being the imaginary time. As before, the mirrors oc-

cupy two surfaces, denoted by L and R, defined by the equations $x_3 = 0$ and $x_3 = \psi(\mathbf{x}_{\shortparallel})$, respectively.

On only dimensional grounds, and using natural units ($\hbar \equiv c \equiv 1$), the DE approximation to the interaction energy to be of the form

$$E_{\rm DE} = -\frac{\pi^2}{1440}\int d^2\mathbf{x}_{\shortparallel}\,\frac{1}{\psi^3}\left[\alpha_{\rm D} + \beta_{\rm D}(\partial_\alpha\psi)^2\right], \qquad (37)$$

where $\alpha_{\rm D}$ and $\beta_{\rm D}$ are dimensionless coefficients that do not depend on the geometry. The subindex D stands for Dirichlet. An evaluation of the above expression for parallel plates fixes $\alpha_D \equiv 1$. As in the electrostatic case, the coefficient $\beta_{\rm D}$ could be computed from explicit examples where the interaction energy is known exactly.

Let us recall, from Section IV, that the interaction energy can also be computed from an expansion of the Casimir energy in powers of η for

$$\psi(\mathbf{x}_{\shortparallel}) \;=\; a \,+\, \eta(\mathbf{x}_{\shortparallel})\,, \qquad (38)$$

where a (assumed to be greater than zero) is the spatial average of ψ whereas η contains its varying piece. The expansion needed is of the second order in η, and with up to two spatial derivatives.

To obtain such an expansion, we start from a rather general yet formal expression for the energy (for earlier perturbative computations of the Casimir force see, for example, Ref.[41, 42]). That formal expression follows from the functional approach to the Casimir effect, where we deal with $\mathcal{Z}$, the zero-temperature limit of a partition function. That partition function, for a scalar field in the presence of two Dirichlet mirrors is given by

$$\mathcal{Z} \;=\; \int \mathcal{D}\varphi\;\delta_L(\varphi)\,\delta_R(\varphi)\;e^{-S_0(\varphi)}\,, \qquad (39)$$

with S denoting the real scalar field free (Euclidean) action

$$S_0(\varphi) \;=\; \frac{1}{2}\int d^4x\,(\partial\varphi)^2\,, \qquad (40)$$

while the δ_L and δ_R impose Dirichlet boundary conditions on the L and R surface, respectively.

The vacuum energy, E, is then obtained as follows:

$$E \;=\; -\lim_{T\to\infty}\log\mathcal{Z}/T\,, \qquad (41)$$

where T is the extent of the time dimension (or β^{-1}, in a thermal partition function setting). We discard from E the terms that do not contribute to the Casimir interaction energy between the two surfaces. These terms will appear as factors in $\mathcal{Z}$; among them the one describing the zero point energy of the field in the absence of the plates, and also the 'self-energy' contributions, due to the vacuum distortion produced by each mirror, even when the other is infinitely far apart.

Exponentiating the two Dirac delta functions by introducing two auxiliary fields, λ_L and λ_R, we obtain for $\mathcal{Z}$ an equivalent expression:

$$\mathcal{Z} \;=\; \int \mathcal{D}\varphi\mathcal{D}\lambda_L\,\mathcal{D}\lambda_R\;e^{-S(\varphi;\lambda_L,\lambda_R)}\,, \qquad (42)$$

with

$$S(\varphi;\lambda_L,\lambda_R) = S_0(\varphi) - i\int d^4x\varphi(x)\left[\lambda_L(x_{\shortparallel})\delta(x_3) + \lambda_R(x_{\shortparallel})\sqrt{g_R(\mathbf{x}_{\shortparallel})}\,\delta(x_3-\psi(\mathbf{x}_{\shortparallel}))\right]$$

where we have introduced $x_{\shortparallel} \equiv (x_0, x_1, x_2) = (x_0, \mathbf{x}_{\shortparallel})$. The factor depending on the determinant of the induced metric on the R, $g_R(\mathbf{x}_{\shortparallel}) \equiv 1 + |\nabla\psi(\mathbf{x}_{\shortparallel})|^2$ makes the expression above reparametrization invariant. However, by a redefinition of the auxiliary field λ_R one gets rid of that factor, at the expense of generating a Jacobian. That Jacobian does not depend on the distance between the two surfaces, since only derivatives of ψ are involved. Therefore it will not contribute the the Casimir interaction energy and thus we shall subsequently ignore such factor, as well as others that will appear in the course of the calculations.

Integrating out φ, we see that $\mathcal{Z}_0$, corresponding to the field φ in the absence of boundary conditions factors out, while the rest becomes an integral over the auxiliary fields:

$$\mathcal{Z} = \mathcal{Z}_0\int \mathcal{D}\lambda_L\mathcal{D}\lambda_R e^{-\frac{1}{2}\int d^3x_{\shortparallel}\int d^3y_{\shortparallel}\sum_{\alpha,\beta}\lambda_\alpha(x_{\shortparallel})\mathbb{T}_{\alpha\beta}\lambda_\beta(y_{\shortparallel})}, \qquad (43)$$

with:

$$\mathcal{Z}_0 \;=\; \int \mathcal{D}\varphi\;e^{-S_0(\varphi)}\,, \qquad (44)$$

and $\alpha,\,\beta = L, R$. We have introduced the objects:

$$\mathbb{T}_{LL}(x_{\shortparallel},y_{\shortparallel}) = \langle x_{\shortparallel},0|(-\partial^2)^{-1}|y_{\shortparallel},0\rangle \qquad (45)$$

$$\mathbb{T}_{LR}(x_{\shortparallel},y_{\shortparallel}) = \langle x_{\shortparallel},0|(-\partial^2)^{-1}|y_{\shortparallel},\psi(\mathbf{y}_{\shortparallel})\rangle \qquad (46)$$

$$\mathbb{T}_{RL}(x_{\shortparallel},y_{\shortparallel}) = \langle x_{\shortparallel},\psi(\mathbf{x}_{\shortparallel})|(-\partial^2)^{-1}|y_{\shortparallel},0\rangle \qquad (47)$$

$$\mathbb{T}_{RR}(x_{\shortparallel},y_{\shortparallel}) = \langle x_{\shortparallel},\psi(\mathbf{x}_{\shortparallel})|(-\partial^2)^{-1}|y_{\shortparallel},\psi(\mathbf{y}_{\shortparallel})\rangle \qquad (48)$$

where we use a "bra-ket" notation to denote matrix elements of operators, and ∂^2 is the four-dimensional Lapla-

cian. Thus, for example,

$$\langle x|(-\partial^2)^{-1}|y\rangle \;=\; \int \frac{d^4k}{(2\pi)^4}\,\frac{e^{ik\cdot(x-y)}}{k^2}\;. \qquad (49)$$

A subtraction of the zero point contribution contained in $\mathcal{Z}_0$ leads to:

$$E \;=\; \lim_{T\to\infty}\left(\frac{1}{2T}\mathrm{Tr}\log\mathbb{T}\right), \qquad (50)$$

which still contains self-energies. Up to now, we have obtained a formal expression for the vacuum energy; let us now proceed to evaluate its DE.

We need to expand E to the second order in η, keeping up to the second order term in an expansion in derivatives. It is convenient to do so first for $\Gamma \;\equiv\; \frac{1}{2}\mathrm{Tr}\log\mathbb{T}$. Namely,

$$\Gamma(a,\eta) \;=\; \Gamma^{(0)}(a) \;+\; \Gamma^{(1)}(a,\eta) \;+\; \Gamma^{(2)}(a,\eta) \;+\; \ldots \qquad (51)$$

where the upper index denotes the order in derivatives. Each term will be a certain coefficient times the spatial integral over $\mathbf{x}_{\shortparallel}$ of a local term, depending on a and on η-derivatives. Additionally, because the configuration is time-independent, they should be proportional to T (a factor that will cancel out). Expanding first the matrix $\mathbb{T}$ in powers of η

$$\mathbb{T} = \mathbb{T}^{(0)} + \mathbb{T}^{(1)} + \mathbb{T}^{(2)} + \ldots\,, \qquad (52)$$

we obtain: $\Gamma \;=\; \Gamma^{(0)} + \Gamma^{(1)} + \Gamma^{(2)} \;+\; \ldots$,

$$\begin{aligned}
\Gamma^{(0)} \;&=\; \frac{1}{2}\mathrm{Tr}\log\mathbb{T}^{(0)}\\
\Gamma^{(1)} \;&=\; \frac{1}{2}\mathrm{Tr}\log\left[(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(1)}\right]\\
\Gamma^{(2)} \;&=\; \frac{1}{2}\mathrm{Tr}\log\left[(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(2)}\right]\\
&\quad-\;\frac{1}{4}\mathrm{Tr}\log\left[(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(1)}(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(1)}\right]\;, \qquad (53)
\end{aligned}$$

where, in $\Gamma^{(l)}$, we need to keep up to l derivatives of η.

Then, the zeroth-order term is obtained as follows: replace ψ by a constant, a, and then subtract from the result its $a\to\infty$ limit (this gets rid of self-energies). This leads to:

$$\Gamma^{(0)}(a) \;=\; \frac{1}{2}\,\mathrm{Tr}\log\left[1-(T^{(0)}_{LL})^{-1}T^{(0)}_{LR}(T^{(0)}_{RR})^{-1}T^{(0)}_{RL}\right]. \qquad (54)$$

Here, the $T^{(0)}_{\alpha\beta}$ are identical to the ones for two flat parallel mirrors separated by a distance a .

Taking the trace, leads to:

$$\Gamma^{(0)} \;=\; \frac{T}{2}\int d^2\mathbf{x}_{\shortparallel}\int\frac{d^3k_{\shortparallel}}{(2\pi)^3}\log[1-e^{-2k_{\shortparallel}a}]\,. \qquad (55)$$

Then, we recall the general derivation to note that the replacement $a\to\psi$ leads to:

$$\begin{aligned}
E^{(0)} \;&=\; \frac{1}{2}\int d^2\mathbf{x}_{\shortparallel}\int\frac{d^3k_{\shortparallel}}{(2\pi)^3}\log[1-e^{-2k_{\shortparallel}\psi(\mathbf{x}_{\shortparallel})}]\\
&=\; -\frac{\pi^2}{1440}\int d^2\mathbf{x}_{\shortparallel}\,\frac{1}{\psi(\mathbf{x}_{\shortparallel})^3}, \qquad (56)
\end{aligned}$$

which is the PFA expression for the vacuum energy.

To improve on the previous result, we consider its first non trivial correction. There can be no first order term because of symmetry considerations. while to terms contribute to the second order

$$\Gamma^{(2)} \;=\; \Gamma^{(2,1)} \;+\; \Gamma^{(2,2)} \qquad (57)$$

where,

$$\Gamma^{(2,1)} \;=\; \frac{1}{2}\mathrm{Tr}\log\left[(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(2)}\right] \qquad (58)$$

and

$$\Gamma^{(2,2)} \;=\; -\frac{1}{4}\mathrm{Tr}\log\left[(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(1)}(\mathbb{T}^{(0)})^{-1}\mathbb{T}^{(1)}\right]\;. \qquad (59)$$

In the terms above, we have to keep just up to two derivatives of η. We see that, in Fourier space, and before implementing any expansion in momentum (derivatives), they have the structure:

$$\Gamma^{(2,j)} \;=\; \frac{T}{2}\int\frac{d^2\mathbf{k}_{\shortparallel}}{(2\pi)^2}f^{(2,j)}(\mathbf{k}_{\shortparallel})\,|\tilde{\eta}(\mathbf{k}_{\shortparallel})|^2 \qquad (60)$$

$(j=1,2)$, with $\tilde{\eta}$ denoting the Fourier transform of η, and with the $f^{(2,j)}$ kernels denoting the $k_0\to 0$ (i.e., static) limits of the more general expressions:

$$\begin{aligned}
f^{(2,1)}(k_{\shortparallel}) \;&=\; -\int\frac{d^3p_{\shortparallel}}{(2\pi)^3}\,\frac{|p_{\shortparallel}||p_{\shortparallel}+k_{\shortparallel}|}{1-e^{-2|p_{\shortparallel}+k_{\shortparallel}|a}}\\
f^{(2,2)}(k_{\shortparallel}) \;&=\; -\int\frac{d^3p_{\shortparallel}}{(2\pi)^3}\,\frac{|p_{\shortparallel}||p_{\shortparallel}+k_{\shortparallel}|e^{-2|p_{\shortparallel}+k_{\shortparallel}|a}(1+e^{-2|p_{\shortparallel}|a})}{(1-e^{-2|p_{\shortparallel}|a})(1-e^{-2|p_{\shortparallel}+k_{\shortparallel}|a})}.
\end{aligned}$$

By subtracting all the a-independent contributions, one finds:

$$\Gamma^{(2)} \;=\; \frac{T}{2}\int\frac{d^2\mathbf{k}_{\shortparallel}}{(2\pi)^2}f^{(2)}(\mathbf{k}_{\shortparallel})\,|\tilde{\eta}(\mathbf{k}_{\shortparallel})|^2 \qquad (61)$$

with:

$$f^{(2)}(k_{\shortparallel}) \;=\; -2\int\frac{d^3p_{\shortparallel}}{(2\pi)^3}\,\frac{|p_{\shortparallel}|\,|p_{\shortparallel}+k_{\shortparallel}|}{(1-e^{-2|p_{\shortparallel}|a})(e^{2|p_{\shortparallel}+k_{\shortparallel}|a}-1)}\,. \qquad (62)$$

The low-momentum behaviour of $f^{(2)}$ determines whether the DE can be applied or not. In this case, the function is analytic and therefore a local expansion of the vacuum energy exists. We need to extract its $\mathbf{k}^2$ order term in a Taylor expansion at zero momentum, namely $f^{(2)}(\mathbf{k}_{\shortparallel}) \;\simeq\; \chi\,{\mathbf{k}_{\shortparallel}}^2$. We find:

$$\chi \;=\; -\frac{\pi^2}{1080\,a^3}\;. \qquad (63)$$

Thus,

$$\Gamma^{(2)}(a,\eta) \;=\; -\frac{T}{2}\frac{\pi^2}{1080}\int\frac{d^2\mathbf{k}_{\shortparallel}}{(2\pi)^2}\frac{{\mathbf{k}_{\shortparallel}}^2}{a^3}\,|\tilde{\eta}(\mathbf{k}_{\shortparallel})|^2$$

$$= -\frac{T}{2}\frac{\pi^2}{1080}\int d^2\mathbf{x}_{\shortparallel}\,\frac{1}{a^3}\,(\partial_\alpha\eta)^2\,. \qquad (64)$$

Therefore, the NTLO term in the DE becomes:

$$E^{(2)} = \frac{\Gamma^{(2)}(\psi)}{T} = -\frac{1}{2}\frac{\pi^2}{1080}\int d^2\mathbf{x}_{\shortparallel}\,\frac{(\partial_\alpha\psi)^2}{\psi^3}\,, \qquad (65)$$

where the index α runs from 1 to 2.

Putting together the terms up to second order,

$$E_{\rm DE} \equiv E^{(0)} + E^{(2)} = -\frac{\pi^2}{1440}\int d^2\mathbf{x}_{\shortparallel}\,\frac{1}{\psi^3}\left[1+\frac{2}{3}(\partial_\alpha\psi)^2\right]. \qquad (66)$$

The leading-order term above is the Casimir energy according to the PFA , while the second order one represents the first significant deviation from it. We note that the structure of both terms had been anticipated by dimensional analysis and symmetry considerations. The overall normalization, on the other hand, had been fixed by our previous knowledge of the (well-established) result for parallel plates.

We would like to insist on the fact that the relative weight between the PFA and its correction term–the factor $\beta_{\rm D} = 2/3$–is independent of the surface geometry. This value of $\beta_{\rm D}$ has been independently corroborated in concrete examples by expanding the exact Casimir energy expressions. Interesting cases among them are, for example, either a sphere or a cylinder positioned in front of a plane.

We conclude this Section with an application of the DE to the particular geometry of a sphere in front of a plane. Let us express the function ψ of Equation (13) in polar coordinates ρ, with R the radius of the sphere and d the distance to the plane. The function $\psi(\rho)$ describes an hemisphere when $0 \le \rho \le R$. By inserting the expression of ψ into the DE for the Casimir energy, it becomes possible to explicitly calculate the integrals, to get a rather compact analytical expression:

$$E_{\rm DE} = E_{\rm PFA}\left(1+\frac{1}{3}\frac{a}{R}\right), \qquad (67)$$

where $E_{\rm PFA} = -\hbar c\pi^3 R/(1440a^2)$.

B. The EM Case

The results for the scalar field satisfying Dirichlet boundary conditions, described in Section (V A) above, have been generalized to different boundary conditions and fields. Results for the EM field case and two curved surfaces have been presented in Ref.[24]. Note that, as pointed out at the end of Section III C, symmetry considerations allow for the two-surface problem to be reduced to the one of a curved surface facing a plane, namely, the geometry we have just dealt with in the Dirichlet case above. Indeed, as shown in Ref. [24], the extension of Ref.[21] to two curved surfaces is restricted among other things by the tilt invariance of the reference plane, to which the two surfaces can be projected. This served as a rigorous test for the self-consistency of perturbative results.

Venturing beyond the scalar Dirichlet (D) case of Ref.[21], they calculated the DE for Neumann (N), mixed D/N, and electromagnetic (EM) (perfect metal) surfaces. Interestingly, they observed that the EM correction must align with the sum of D and N corrections. They also replicated previous findings for cylinders under D, N, and mixed D/N conditions, as well as for the sphere with D boundary conditions. However, their calculations did not confirm previous results for the sphere/plane geometry, either with N or EM boundary conditions. Indeed, the results for β were found to disagree with those obtained from Refs. [43–45]. This discrepancy was later resolved in Ref.[46] in favour of the results in [24].

Another interesting concrete example presented in [24] is the DE for two spheres of radii R_1 and R_2, both imposing the same boundary conditions. It was found there that

$$E = E_{\rm PFA}\left[1-\frac{a}{R_1+R_2}+(2\beta-1)\left(\frac{a}{R_1}+\frac{a}{R_2}\right)\right], \qquad (68)$$

where $E_{\rm PFA} = -(\alpha\pi^3 R_1 R_2))/[1440a^2(R_1+R_2)]$; a is chosen to be the distance of closest separation, and β is a number that depends on the type of boundary condition, as can be seen from Table I. $\alpha = \alpha_{\rm EM} = 2$ in the EM boundary conditions case. The corresponding formula for the sphere/plane case can be obtained by taking one of the two radii to infinity (in fact it coincides with the D case in Equation (67) when $\alpha = \alpha_{\rm D} = 1$ and $\beta = \beta_{\rm D} = 2/3$).

A rather different example corresponds to two circular cylinders (with identical boundary conditions) whose axes are inclined at a relative angle θ. Using the DE, the interaction Casimir energy reads:

$$E = -\frac{\alpha\pi^3\sqrt{R_1R_2}}{1440a^2\sin\theta}\left[1+\left(\beta-\frac{3}{8}\right)\frac{a}{R_1+R_2}\right]. \qquad (69)$$

For this particular geometry, the interaction energy has been computed numerically in [47]. The numerical results reproduce Equation (69) at short distances.

The results obtained for the β-coefficients in each case are summarized in Table I.

Having presented in this Section a derivation and some interesting results obtained by applying the DE to the Casimir effect at zero temperature and for perfect boundary conditions, we present in the rest of the review some generalizations and applications.

VI. FINITE TEMPERATURE, NONANALYTICITIES, AND DE

The DE can be extended to the finite temperature case [25, 26, 28], the free energy being the relevant functional

$\beta_{\rm D}$	$\boldsymbol{\beta_{\rm N}}$	$\boldsymbol{\beta_{\rm DN}}$	$\boldsymbol{\beta_{\rm ND}}$	$\boldsymbol{\beta_{\rm EM}}$
$2/3$	$2/3(1-30/\pi^2)$	$2/3$	$2/3-80/7\pi^2$	$2/3(1-15/\pi^2)$

Table I. β coefficient from (68) for the following five cases: a scalar field obeying Dirichlet (D) or Neumann (N) boundary conditions on both surfaces, or D boundary condition on one surface and N boundary condition on the other, or vice versa, and for the electromagnetic (EM) field with ideal metal boundary conditions [24]

to approximate. There are at least two reasons why this extension is not trivial: firstly, the temperature introduces a dimensionful magnitude, and this will reflect itself in the form of the DE (part of it was fixed by dimensional analysis). Second, a known phenomenon in quantum field theory at finite temperature is the so-called "dimensional reduction", by which a bosonic model which is defined in $d+1$ dimensions at zero temperature, becomes effectively d-dimensional at high temperatures. The DE should therefore manifest (and interpolate between) those two cases.

We first describe, in Section VI A, the results for a scalar field satisfying Dirichlet conditions [26] in $d+1$ dimensions. Then, Section VI B discusses the appearance of nonanaliticities for Neumann boundary conditions [26, 27]. Finally, we comment on the results for the EM field with imperfect boundary conditions [25, 28] (Section VI C) and on semianalytic formula for plane–sphere geometry (Section VI D).

A. Dirichlet Boundary Conditions

In the finite-temperature case, and for the same geometry that we have considered in the zero temperature case, the functions $V(\psi)$ and $Z(\psi)$ cannot be completely determined from dimensional analysis alone. Indeed, on general grounds, we can assert that the Casimir free energy in $d+1$ dimensions, if the DE is applicable, must have the form:

$$F_{\rm DE}[\psi] \;=\; \int d^{d-1}\mathbf{x}_{\shortparallel}\left\{b_0(\frac{\psi}{\beta},d)\frac{1}{[\psi(\mathbf{x}_{\shortparallel})]^d} \;+\; b_2(\frac{\psi}{\beta},d)\,\frac{(\nabla\psi)^2}{[\psi(\mathbf{x}_{\shortparallel})]^d}\right\} \tag{70}$$

where b_0 and b_2 are dimensionless and depends on the ratio of the local distance between surfaces ψ and the inverse temperature β. They can be obtained from the knowledge of the Casimir free energy for small departures around $\psi(\mathbf{x}_{\shortparallel})=a=$ constant. They are given by [26]

$$\begin{aligned} b_0(\xi) &= \frac{\xi}{2}\sum_{n=-\infty}^{\infty}\int\frac{d^{d-1}\mathbf{p}_{\shortparallel}}{(2\pi)^{d-1}}\log\left(1-e^{-2\sqrt{(2\pi n\xi)^2+(\mathbf{p}_{\shortparallel})^2}}\right)\\ b_2(\xi) &= \frac{1}{2}\left[\frac{\partial F^{(2)}(\xi,n,\mathbf{l}_{\shortparallel})}{\partial|\mathbf{l}_{\shortparallel}|^2}\right]_{n\to 0,|\mathbf{l}_{\shortparallel}|\to 0} \end{aligned} \tag{71}$$

where

$$F^{(2)}(\xi;n,|\mathbf{l}_{\shortparallel}|)=-2\xi\sum_{m=-\infty}^{+\infty}\int\frac{d^{d-1}\mathbf{p}_{\shortparallel}}{(2\pi)^{d-1}}\Big\{\frac{\sqrt{(2\pi m\,\xi)^2+\mathbf{p}_{\shortparallel}^2}}{1-\exp\big[-2\sqrt{(2\pi m\xi)^2+\mathbf{p}_{\shortparallel}^2}\big]}\frac{\sqrt{(2\pi(m+n)\,\xi)^2+(\mathbf{p}_{\shortparallel}+\mathbf{l}_{\shortparallel})^2}}{\exp\big\{2\sqrt{[2\pi(m+n)\xi]^2+(\mathbf{p}_{\shortparallel}+\mathbf{l}_{\shortparallel})^2}\big\}-1}\Big\}. \tag{72}$$

In the zero temperature limit, the Matsubara sum becomes an integral that can be analytically computed. The results are described in Table II. The ratio b_2/b_0 tends to 1 for large values of d.

In the high temperature limit, we find

$$\begin{aligned} \big[b_0(\xi,d)\big]_{\xi>>1} &\simeq \xi\,\big[b_0(\xi,d-1)\big]_{\xi\to 0}\equiv\xi\,b_0(d-1)\,,\\ \big[b_2(\xi,d)\big]_{\xi>>1} &\simeq \xi\big[b_2(\xi,d-1)\big]_{\xi\to 0}\equiv\xi b_2(d-1) \end{aligned} \tag{73}$$

where $\xi=\psi/\beta$. The coefficients $b_0(d-1)$ and $b_2(d-1)$ agree with those for perfect mirrors at zero temperature, but in $d-1$ dimensions, i.e., the ' dimensional reduction" effect.

An interesting result is found when this is applied to the (Dirichlet) Casimir interaction for a system consisting of a sphere in front of an infinite plane. Denoting by a the distance between the surfaces, and by R the radius of the sphere, we get for the free energy at high temperatures:

$$F_{\rm DE}[\psi]|_{\psi/\beta>>1,d=3}\;\sim\;-\frac{\zeta(3)R}{8\beta a}\left(1-\frac{1}{6\zeta(3)}\frac{a}{R}\log\left(\frac{a}{R}\right)\right). \tag{74}$$

Dimension	$b_2(d)/b_0(d)$	Approximate Value
$d=1$	$\frac{1}{\pi^2}\left(1+\frac{\pi^2}{3}\right)$	0.435
$d=2$	$\frac{1+6\zeta(3)}{12\zeta(3)}$	0.569
$d=3$	$2/3$	0.667
$d=4$	$\frac{-\zeta(3)+10\zeta(5)}{12\zeta(5)}$	0.737
$d=5$	$\frac{10\pi^2-21}{10\pi^2}$	0.787
$d=6$	$\frac{-2\zeta(5)+7\zeta(7)}{6\zeta(7)}$	0.824

Table II. Values of the ratio $b_2(d)/b_0(d)$ for different dimensions. The ratio tends to 1 for $d\to\infty$. See text for details.

We see that the R/a^2-behavior corresponding to the dominant contribution at zero temperature changes to $R/a\beta$ in the high temperature case. This could be expected on dimensional grounds, if one assumes that the free energy is linear in the temperature in this limit. Note that the same problem has been exactly solved in Ref.[48], and one can show that Equation (74) does agree with the small-distance expansion obtained from the exact solution.

It is worth to remark that the NTLO correction from the DE becomes nonanalytic, because of the integration, in the ratio a/R. This behavior has been observed in numerical calculations of the Casimir interaction energy for this geometry, in the infinite temperature limit, for the electromagnetic case (see Refs.[48, 49]). It is important to recognize that this nonanalyticity has nothing to do with the nonanalyticity in momenta of the form factors described in Section 4, and is a non trivial prediction of the DE.

B. Neumann Boundary Conditions

This case, discussed in Ref.[27], highlights a potential warning to the applicability of the DE, already mentioned previously: the appearance of nonanalyticities in the form factors. To begin with, we deal with the zero temperature case in $2+1$ dimensions, since the nonanalyticity appears because of the existence of a Matsubara mode which behaves as a massless field in $2+1$ dimensions, with Neumann boundary conditions.

The free Euclidean action for the vacuum (i.e., $T=0$) field φ is given by

$$\mathcal{S}_0 \;=\; \frac{1}{2}\int d^3x\,(\partial\varphi)^2\;, \tag{75}$$

and, instead of imposing perfect Neumann boundary conditions on the surfaces, we add the following action to describe the interaction between the vacuum field and the mirrors:

$$\mathcal{S}_{\rm I}[\varphi]=\frac{1}{2\bar{\mu}}\int d^3x\left[\delta(x_2)(\partial_2\varphi(x))^2+\sqrt{g(x_\parallel)}\delta(x_2-\psi(x_\parallel))(\partial_n\varphi(x))^2\right]\;. \tag{76}$$

The constant $\bar{\mu}$, which has the dimensions of a mass, is used to impose Neumann boundary conditions in the $\bar{\mu}\to 0$ limit. We use the same $\bar{\mu}$ on both L and R mirrors, since we will assume them to have identical properties, differing just in their position and geometry.

The DE approximation to the Casimir energy can be computed following standard steps. The result reads, in the limit $\bar{\mu}\psi\to 0$ [27],

$$E_{\rm DE}[\psi] \;=\; -\frac{1}{16\pi}\int_{-\infty}^{\infty}dx_1\,\frac{1}{\psi(x_1)^2}\left[\zeta(3)+\log[\bar{\mu}\psi(x_1)]\left(\frac{d\psi(x_1)}{dx_1}\right)^2\right]\;. \tag{77}$$

In the expression above, the first term is the PFA contribution while the second one is a non trivial correction to it, and depends on the shape of the boundary (defined by ψ). It is then clear that, as this equation shows, the DE is well posed when imposing imperfect Neumann boundary conditions in $2+1$ dimensions. On the contrary, it cannot be applied when the boundary conditions become perfect ($\bar{\mu}=0$). The reason is that the hypothesis of analyticity in momentum, used to derive the DE, is clearly violated. The non-existence of a local expansion is due to the existence of massless modes, allowed by Neumann boundary conditions.

Since, at finite temperatures, a $3+1$ dimensional theory may be decomposed into the sum of an infinite tower of decoupled $2+1$ dimensional Matsubara modes, each one satisfying N boundary conditions, and with a mass $\frac{2n\pi}{\beta}$, $n=0,1,2,\ldots$ The existence of the massless $n=0$ mode (the only one surviving in the high temperature limit) means that analyticity will be lost in $3+1$ dimensions, for any non zero temperature. That is indeed the case [26]. We summarize here some of the main features of that example: the free energy in the $d+1$ dimen-

sional Neumann case can again be written as before (see Eq.(70), but with coefficients c_0 and c_2 instead of b_0 and b_2. The zero order term coincides with the one for the Dirichlet case; namely: $c_0 = b_0$.

When $d = 3$, the NTLO term contains, besides a local term, a nonlocal contribution which is linear in T, and thus present for any $T > 0$. Hence, there is no local DE for perfect Neumann boundary conditions at $d = 2$ at zero temperature and for $d = 3$ at any finite temperature. Indeed, an expansion for small values of $|k_{\shortparallel}|$ of the form factor contains, in addition to a term proportional to $k_{\shortparallel}^2$, one proportional to $(Ta)k_{\shortparallel}^2 \log\big(k_{\shortparallel}^2 a^2\big)$.

C. The Electromagnetic Case for Imperfect Boundary Conditions

We have seen that, for a real scalar field in the presence of Neumann boundary conditions, the DE cannot be applied when in $2+1$ dimensions at zero temperature, or in $3+1$ dimensions at a non-zero temperature [26]. The reason is that, as we have shown, nonanalyticities in the form factors appear. We have shown that the nonanalyticity could be cured by introducing a small departure from perfect Neumann conditions [27]. It is natural to wonder whether the nonanalyticities could also be cured by a similar approach for the EM field in $3+1$ dimensions at finite temperatures. We know, based on the insight obtained from Ref.[27], that nonanalyticities are originated in contributions due to dimensionally reduced massless modes: zero Matsubara frequency terms. To obtain an answer to this question, in Ref.[28] we singled out in detail the zero-mode contributions to the free energy, for a media described by non trivial permittivity $\epsilon(\omega)$ and permeability $\mu(\omega)$ functions.

We start from the free energy F for the EM field, which can be written in terms of the partition function $\mathcal{Z}(\psi)$, as follows:

$$F(\psi) \;=\; -\frac{1}{\beta}\,\log\big[\frac{\mathcal{Z}(\psi)}{\mathcal{Z}_0}\big]\,, \tag{78}$$

where the denominator, $\mathcal{Z}_0$, denotes the partition function for the EM field in the absence of media and

$$\mathcal{Z}(\psi) \;=\; \int \big[\mathcal{D}A\big]\; e^{-\mathcal{S}_{\rm inv}(A)}\,. \tag{79}$$

The gauge invariant action $\mathcal{S}_{\rm inv}(A)$ reads

$$\mathcal{S}_{\rm inv}(A) \;=\; \int_0^\beta d\tau \int_0^\beta d\tau' \int d^3\mathbf{x}\,\big[\frac{1}{2}\,F_{0j}(\tau,\mathbf{x})\epsilon(\tau-\tau',\mathbf{x})F_{0j}(\tau',\mathbf{x}) + \frac{1}{4}F_{ij}(\tau,\mathbf{x})\mu^{-1}(\tau-\tau',\mathbf{x})F_{ij}(\tau,\mathbf{x})\big]\,. \tag{80}$$

Here, indices like i, j ... run over spatial indices, Einstein summation convention is assumed, and $\epsilon(\tau-\tau',\mathbf{x})$ and $\mu(\tau-\tau',\mathbf{x})$ denote the imaginary time versions of the permittivity and permeability, respectively (μ^{-1} is the inverse integral kernel of μ).

The geometry of the system is determined by same two surfaces L and R we have considered before, and defined by $x_3 = 0$ and $x_3 = \psi(\mathbf{x}_{\shortparallel})$, but now they correspond to the boundaries of the media, i.e.,

$$\begin{aligned}
\epsilon(\tau-\tau',\mathbf{x}) \;&=\; \theta(-x_3)\epsilon_L(\tau-\tau') + \theta(x_3)\theta(\psi(\mathbf{x}_{\shortparallel})-x_3) + \theta(x_3-\psi(\mathbf{x}_{\shortparallel}))\epsilon_R(\tau-\tau')\\
\mu(\tau-\tau',\mathbf{x}) \;&=\; \theta(-x_3)\mu_L(\tau-\tau') + \theta(x_3)\theta(\psi(\mathbf{x}_{\shortparallel})-x_3) + \theta(x_3-\psi(\mathbf{x}_{\shortparallel}))\,\mu_R(\tau-\tau'),
\end{aligned}$$

where $\epsilon_{L,R}(\tau-\tau')$ and $\mu_{L,R}(\tau-\tau')$ characterize the permittivity and permeability of the respective mirror.

We can expand the fields and the electromagnetic properties as

$$\begin{aligned}
A_\mu(\tau,\mathbf{x}) \;&=\; \frac{1}{\beta}\sum_{n=-\infty}^{+\infty} \widetilde{A}_\mu^{(n)}(\mathbf{x})\, e^{i\omega_n\tau}\\
\epsilon(\tau-\tau',\mathbf{x}) \;&=\; \frac{1}{\beta}\sum_{n=-\infty}^{+\infty} \widetilde{\epsilon}^{(n)}(\mathbf{x})\, e^{i\omega_n(\tau-\tau')}\\
\mu(\tau-\tau',\mathbf{x}) \;&=\; \frac{1}{\beta}\sum_{n=-\infty}^{+\infty} \widetilde{\mu}^{(n)}(\mathbf{x})\, e^{i\omega_n(\tau-\tau')}
\end{aligned} \tag{81}$$

where $\omega_n \equiv 2\pi n/\beta$ ($n \in \mathbb{Z}$) are the Matsubara frequencies.

Inserting these expansions into the partition function one can readily check the factorization

$$\mathcal{Z}(\psi) \;=\; \prod_{n=0}^{\infty} \mathcal{Z}^{(n)}(\psi)\,, \tag{82}$$

and therefore

$$F(\psi) \;=\; \sum_{n=0}^{\infty} F^{(n)}(\psi)\,. \tag{83}$$

As mentioned, we are particularly interested in the $n =$

0 contribution,

$$\mathcal{Z}^{(0)}(\psi) \;=\; \int [\mathcal{D}\widetilde{A}_0^{(0)}\mathcal{D}\widetilde{A}_j^{(0)}]\; e^{-\mathcal{S}^{(0)}(\widetilde{A}_0^{(0)},\widetilde{A}_j^{(0)})}\,, \qquad (84)$$

where

$$\mathcal{S}^{(0)}(\widetilde{A}_0^{(0)},\widetilde{A}_j^{(0)}) \;=\; \frac{1}{\beta}\int d^3\mathbf{x}\,\big[\frac{1}{2}\,\widetilde{\epsilon}^{(0)}(\mathbf{x})(\partial_j\widetilde{A}_0^{(0)})^2 \;+\; \frac{1}{4\,\widetilde{\mu}^{(0)}(\mathbf{x})}(\widetilde{F}_{jk}^{(0)})^2 \;+\; \frac{1}{2}\,\Omega_0^2(\mathbf{x})(\widetilde{A}_j^{(0)})^2\big]\,, \qquad (85)$$

and

$$\Omega_0^2(\mathbf{x}) \;\equiv\; \lim_{n\to 0}\,\left[\omega_n^2\,\widetilde{\epsilon}^{(n)}(\mathbf{x})\right]\,. \qquad (86)$$

Note that Ω_0 vanishes for a dielectric and also for a metal described by the Drude model. On the other hand, it equals the plasma frequency for a metal described by the plasma model.

The zero mode contribution to the free energy therefore splits into a scalar (s) and a vector (v) contribution, the former associated to the field $\widetilde{A}_0^{(0)}$ and the later to $\widetilde{A}_j^{(0)}$

$$F^{(0)} \;=\; F_s(\psi) \;+\; F_v(\psi)\,. \qquad (87)$$

To discuss the emergence of non-analyticities in the derivative expansion we computed F_s and F_v assuming $\psi(\mathbf{x}_{\shortparallel}) = a+\eta(\mathbf{x}_{\shortparallel})$ up to second order in η. The quadratic contributions can be written as

$$F_{s,v}^{(2)} \;=\; \frac{1}{2}\int \frac{d^2k_{\parallel}}{(2\pi)^2}\,f_{s,v}^{(2)}(k_{\shortparallel},a)\,|\tilde{\eta}(k_{\shortparallel})|^2\,, \qquad (88)$$

and the crucial point is whether the functions $f_{s,v}^{(2)}$are analytic or not in $k_{\parallel}$.

Omitting the details, we summarize the main results [28]: for finite values of μ and ϵ, the scalar contribution $f_s^{(2)}$ analytic, including the limit $\epsilon\to\infty$, in which it tends to the $2+1$ dimensional Dirichlet value. It develops a nonanalytic (logarithmic) contribution for $\mu=\infty$, since the kernel corresponds in this case to that of a scalar field in $2+1$ dimensions satisfying Neumann boundary conditions. In other words, magnetic materials regulate the non-analyticity of the TE zero mode.

On the other hand, the TM zero mode is nonanalytic whenever $\omega^2\epsilon(\omega)\to\Omega^2\neq 0$ as $\omega\to 0$ for both mirrors. In terms of the models usually considered in the Casimir literature to describe real materials, this condition corresponds to the plasma model.

In summary, the nonanalyticities we observed for perfect conductors in our previous work [26], survive only under the assumption of perfectly lossless materials. The NTLO corrections to PFA for metals (gold) at room temperature have been computed in Ref. [25].

D. A Semianalytic Formula for Plane-Sphere Geometry

As a final application of the DE to compute the Casimir free energy we mention the results of Ref.[50], where the author combined exact calculations for the zero mode and the DE to obtain a precise formula for the interaction between a sphere and a plane at a finite temperature which is valid at all separations. We briefly describe here these findings.

Formally, the free energy for this geometry can be written as

$$F = k_BT\,{\sum_{n\geq 0}}'\,Tr\log[1-\hat{M}(i\xi_n)]\,, \qquad (89)$$

where the sum is over the Matsubara frequencies $\xi_n = 2\pi n k_B T/\hbar$ and $\hat{M}$ denotes scattering matrix elements for this geometry. The prime on the sum indicates that the $n=0$ term has an additional $1/2$ factor.

The $n=0$ contribution can be computed exactly using the Drude model to describe the materials of the plane and the sphere, and plays a crucial role. Indeed, the proposed approximation for the Casimir force on the sphere of radius R at a distance a from the plane is

$$F_{\rm approx} = F_{n=0}^{(\rm exact)} + F_{n>0}^{(\rm PFA)}(1-\theta\frac{a}{R})\,, \qquad (90)$$

where θ can be computed using the DE. Notably, $F_{\rm approx}$ describes with high precision the Casimir force at all separations, as can be checked by comparison with high precision numerical simulations of the exact scattering formula.

These results have been generalized in subsequent studies to the case of the two spheres- geometry [51], also considering the differences that come from the use of the Drude vs plasma models, as well as for grounded vs isolated spheres [52]. The relevance of the use of grounded conductors in Casimir experiments has also been discussed in Ref. [53].

VII. CASIMIR-POLDER FORCES

The DE approach has also been applied to the calculation of the Casimir-Polder interaction between a po-

larizable particle and a gently curved surface [29]. We present in this Section a simplified version of the results contained in that reference.

When a small polarizable particle is at a distance a of a planar surface, the Casimir-Polder potential reads [4]

$$U(a) = -\frac{1}{a^4}\int \frac{d\xi}{2\pi}\alpha(i\omega_c\xi)\beta^{(0)}(\xi)\,, \tag{91}$$

where $\alpha(\omega)$ is the frequency dependent polarizability (which is assumed isotropic), $\omega_c = c/a$, and

$$\beta^{(0)} = \frac{e^{-2\xi}}{2}(1+2\xi+2\xi^2)\,. \tag{92}$$

For moderate distances such that $\alpha(\omega) \approx \alpha(0)$ one obtains the usual Casimir-Polder potential [54]

$$U(a) = -\frac{3}{8\pi}\frac{\alpha(0)}{a^4}\,. \tag{93}$$

Assume now that the particle is in front of a slightly curved surface. The particle is at the origin of coordinates, and the surface is described, as usual, by the height function $z = \psi(\mathbf{x}_{\shortparallel})$. The DE for the Casimir-Polder interaction $U_{\rm DE}$ assumes that the interaction depends on the derivatives of the height function ψ evaluated at $\mathbf{x}_{\shortparallel} = 0$, the point on the surface closest to the particle (a local minimum for ψ). If the surface is homogeneous and isotropic, then the interaction energy must be invariant under rotations of the $\mathbf{x}_{\shortparallel}$ coordinates. The more general expression compatible with this properties describes the Casimir -Polder interaction energy at $T = 0$ reads [29]:

$$U_{\rm DE} = -\frac{1}{\psi^4}\int \frac{d\xi}{2\pi}\alpha(i\omega_c\xi)\left(\beta^{(0)}(\xi) + \beta^{(1)}(\xi)\psi\nabla^2\psi\right)\,. \tag{94}$$

The dimensionless function $\beta^{(1)}$ can be read from the perturbative expansion of the potential U, carried to second order in the deformation, that is, for $\psi(\mathbf{x}_{\shortparallel}) = a + \eta(\mathbf{x}_{\shortparallel})$ with $\eta(\mathbf{x}_{\shortparallel}) \ll a$. We stress that here the Casimir-Polder energy is not a functional but a function of ψ and its derivatives evaluated at the origin of coordinates (recall that $\nabla\psi(\mathbf{0}) = 0$). The DE is expected to be valid when $a \ll R_1, R_2$, the radii of curvature of the surface at $\mathbf{x}_{\shortparallel} = 0$. Note that $\psi(\mathbf{0}) = d$ and

$$\nabla^2\psi(\mathbf{0}) = \frac{1}{R_1} + \frac{1}{R_2}\,. \tag{95}$$

Using again the static polarizability approximation, $\alpha(\omega) \approx \alpha(0)$, one obtains

$$U_{\rm DE} = -\frac{1}{\pi a^4}\alpha(0)\left[\frac{3}{8} - \frac{13}{60}a(\frac{1}{R_1}+\frac{1}{R_2})\right]\,. \tag{96}$$

The results presented in Ref. [29] are much more general than those described here: they include the Casimir-Polder potential for a general polarization tensor $\alpha_{\mu\nu}(\omega)$ and higher order corrections proportional to $(a/R_i)^2$, as well as the details of the computation of the corresponding functions $\beta^{(p)}$. Additional applications can be found in [55, 56].

VIII. OTHER TECHNIQUES BEYOND PFA

In Ref. [57] a detailed analysis of the Casimir effect's roughness correction in a setting involving parallel metallic plates is presented. The plates were defined through the plasma model. The approach used is perturbative, factoring in the roughness amplitude and allowing for the consideration of diverse values of the plasma wavelength, plate separation, and roughness correlation length. A notable finding was that the roughness correction *exceed* the predictions of the PFA. The authors have calculated the second-order response function, $G(k)$, across a spectrum of values encompassing the plasma wavelength (λ_P), distance (a), and roughness wave vector (k):

$$G(k) = -\frac{A}{8\pi^2}\frac{1}{a^5 q}\int_0^\infty \frac{dK e^{-2K}}{1-e^{-2K}}\int_{|K-q|}^{K+q} dK' \frac{(KK')^2 + \frac{1}{4}(K^2+K'^2-q^2)^2}{1-e^{-2K'}}\,, \tag{97}$$

applicable when $\lambda_p \to 0$. Here, A represents the plate surface area, K the dimensionless integration variable denoting the imaginary wave vector's z-component scaled by plate separation d, K' the longitudinal component of the imaginary wave vector for the diffracted wave, and $q = ka$.

The calculation in Ref. [57] helps to compute the second-order roughness correction as a function of the surface profiles, h_1 and h_2. Analytical solutions were determined for specific limiting cases, revealing a more complex relationship with the perfect reflectors model than previously recognized [58, 59], particularly in scenarios involving extended distances and small roughness wavelengths. While the asymptotic case of long roughness wavelengths aligns with PFA predictions, it was established that PFA generally underestimates the roughness correction, a critical aspect for exploring constraints on potentially new weak forces at sub-millimeter ranges.

As a further expansion to [57], in Ref. [60], the authors explored the Casimir interaction between a plane and a sphere of radius R at a finite temperature T, in terms of the distance of closest approach, a. Noting that, un-

der the usual experimental conditions, the thermal wavelength λ_T satisfies $a \ll \lambda_T \ll R$, they evaluated the leading correction to the PFA, applicable to such intermediate temperatures. They resorted to developing the scattering formula in the plane-wave basis. The result captures the combined effect of spherical geometry and temperature, and is expressed as a sum of temperature-dependent logarithmic terms. Remarkably, two of these logarithmic terms originated from the Matsubara zero-frequency contribution.

Defining the variables $x = a/R$ and $\tau = a/\lambda_T$, and the deviation $\delta F(T) = F(T) - F(0)$, in the intermediate temperature regime $x \ll \tau \ll 1$, it is found in Ref. [60] that

$$\Delta = \frac{\delta F(T) - \delta F_{\rm PFA}(T)}{F_{\rm PFA}(T)} \approx \frac{45}{\pi^3} x\tau \left[-\log^2(x) + 2[1-\log(2)]\log(x) + 2\log^2(\tau) + \mathcal{O}(\log(\tau)) \right]. \tag{98}$$

The leading neglected terms stem from non-zero Matsubara frequencies.

In Ref. [61], the leading-order correction to PFA in a plane-sphere geometry was derived. The momentum representation connected this with geometrical optics and semiclassical Mie scattering. The primary contributions are shown to come from diffraction, with TE polarization becoming more relevant than TM polarization. The diffraction contribution is calculated at leading order, using the saddle-point approximation, considering leading order curvature effects at the sphere tangent plane.

Additionally, the next-to-leading order (NTLO) term in the saddle-point expansion contributed to the PFA correction. This involved computing the round-trip operator within the WKB approximation, representing sequences of reflections between the plane and the sphere. A key aspect was the tilt in the scattering planes, allowing TE and TM polarizations to mix.

Comprehending the implications of polarization mixing channels on the geometric optical correction applied to PFA holds considerable importance. Indeed, these channels are recognized for inducing negative Casimir entropies with a geometric foundation. [63**?** –67]. In spite of the non-vanishing contribution of the polarization mixing matrix elements, the total correction associated with the tilt between the scattering and Fresnel planes is zero at NTLO. This implies that the primary correction to the PFA would remain unchanged even if the complexities arising from the differences between the Fresnel and scattering polarization bases were initially ignored. The latter points to the fact that a different approach, one that completely omits the effect of polarization mixing, could directly produce the leading order correction to PFA. Plane waves proved to be a well-suited basis for studying the Casimir effect, as has been evidenced in the more recent study [30]. The utility of that basis ranges from analytical to numerical applications, particularly when dealing with objects in close proximity, the most relevant situation in experiments. It has been also shown that the use of plane waves was notably effective in improving the interpretation of results in the realms of geometrical optics and diffractive corrections.

In the context of a setup involving two spheres with arbitrary radii in vacuum, it was shown in [30] that the PFA emerged as the leading term in an asymptotic expansion for large radii. Extending a prior calculation based on the saddle-point approximation, involving a trace over multiple round-trips of electromagnetic waves between the spheres, the study encompassed spheres made of bi-isotropic material, requiring the consideration of polarization mixing during reflection processes. The result was naturally elucidated within the framework of geometrical optics.

Then, by relying on a saddle-point approximation framework, the authors derived leading-order corrections, of geometrical and diffractive origins. Explicit results, at first obtained for perfect electromagnetic conductors (PEMC) spheres at zero temperature, indicated that for certain material parameters, the PFA contribution vanishes; should that be the case, the leading-order correction would be the dominant term in the Casimir energy.

In the lowest-order saddle-point approximation, but including diffractive corrections, one can show that the expression for the Casimir energy becomes:

$$E_{\rm LO-SPA} = -\frac{\pi^3 R_{\rm eff}}{720 a^2}\left[1 - \frac{15}{\pi^2}x + \frac{15(10+3\pi)}{4\pi^3}x^{3/2} + ..\right], \tag{99}$$

where $x = a/R_{\rm eff}$. As expected, this result reproduces the PFA result and its leading-order diffractive correction. The NTLO correction behaves as $x^{3/2}$. However, the prefactor obtained accounts for about 90% of the one coming from numerical results [61]. This discrepancy may be traced back to having neglected the NTLO-SPA and NNTLO-SPA contributions.

IX. CONCLUSIONS

In this review, we have discussed several properties and applications of the DE approach, mostly as a method to improve the predictions of the Proximity Force Approximation, of long standing use in many different fields.

We started the review by briefly discussing the precursor of the PFA: the Derjaguin (and related) approximations, since we have found them rather appropriate in order to display the essentially geometric nature of the kind of problem we discuss: two quite close smooth

surfaces, and an interaction energy between them. Depending on the kind of system being considered, that interaction between the two surfaces may or not be the result of the superposition of the interactions between pairs. An example of an interaction which is not the result of such a superposition is the Casimir effect. Note, however, that even when the fundamental interaction satisfies a superposition principle, like in electrostatics, the actual evaluation of the Coulomb integral to calculate the total interaction energy could be a rather involved problem because the actual charge density may not be known a priori. That is indeed the case when the surfaces involved are conductors, since that usually requires finding the electrostatic potential. We have used precisely this problem in order to present the idea of the DE in a concrete example: to calculate the electrostatic energy between two conducting surfaces held at different potentials.

After introducing and applying the DE in that example, we have discussed its more general proof of that expansion, by first putting the problem in a more general and abstract way: how to approximate, under certain smoothness assumptions, a functional of a pair of surfaces. At the same time, the proof provides a concrete way to determine the PFA and its NTLO correction, the DE: one just needs to perform an expansion in powers of the deformation of the surfaces about the situation of two flat and parallel surfaces.

The derivations and examples here have been presented for a geometrical setting were one surface is a plane, while the other may be described by a single Monge patch based on that plane. However, as shown by other authors, under quite reasonable and general assumptions, the results obtained for that situation may be generalized to the case of two curved surfaces parametrized by their respective patches, based on a common plane (which now does not coincide with one of the physical surfaces).

Then we reviewed different applications of the DE to the zero temperature Casimir effect, considering different fields and boundary conditions, staring from the cases of the scalar field with Dirichlet boundary conditions, then the EM field in the presence of perfectly conducting surfaces, and commented on the scalar field with Neumann conditions.

We afterwards presented a description and brief review of the extension of DE to finite temperature cases, and different numbers of spatial dimensions. The temperature is a dimensionful magnitude and the phenomenon of dimensional reduction presents a problem when there are Neumann boundary conditions or when an EM field is involved. Indeed, dimensional reduction implies the existence of a massless 2+1 dimensional field (with Neumann conditions), and this mode introduces a nonanalyticity in momentum space, which violated one of the hypothesis of the DE, and therefore it cannot be applied. Nevertheless, we have shown that the introduction of a small departure from ideal Neumann conditions solves this issue, namely, analyticity is recovered and the DE may be applied.

We also mentioned the application of DE to the Casimir-Polder interaction, particularly between a polarizable particle and a gently curved surface. This example highlights the broader implications of DE in understanding particle-surface interactions beyond the Casimir force itself.

To conclude, we have presented in this review the main features of the DE approach, with a focus in the Casimir effect, but pointing at the fact that its applicability can certainly go beyond that realm. We have shown that explicitly for electrostatics, but we expect it to be applicable to, for example, the same kind of systems where the DA, SEI and SIA were introduced.

This research was funded by Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad de Buenos Aires (UBA), and Universidad Nacional de Cuyo (UNCuyo), Argentina.

[1] Casimir, H.B.G. On the attraction between two perfectly conducting plates. *Proc. Kon. Ned. Akad. Wetensch. B* **1948**, *51*, 793 – 795. Available online: `https://dwc.knaw.nl/DL/publications/PU00018547.pdf` (accessed on 6 Januaary 2023).

[2] Milonni, P.W. *The Quantum Vacuum. An Introduction to Quantum Electrodynamics*; Academic Press, Inc.: San Diego, CA, USA, 1994. https://doi.org/10.1016/C2009-0-21295-5

[3] Milton, K.A. *The Casimir Effect. Physical Manifestations of Zero-Point Energy*; World Scientific: Singapore, 2001. https://doi.org/10.1142/4505

[4] Bordag, M.; Klimchitskaya, G.L.; Mohideen, U.; Mostepanenko, V.M. *Advances in the Casimir Effect*; Oxford University Press: Oxford, UK, 2009. https://doi.org/10.1093/acprof:oso/9780199238743.001.0001

[5] Dalvit, D.; Milonni, P.; Roberts, D.; da Rosa, F. (Eds.) *Casimir Physics*; Springer: Berlin/Heidelberg, Germany, 2011. https://doi.org/10.1007/978-3-642-20288-9

[6] Mostepanenko, V.M. Casimir puzzle and Casimir conundrum: Discovery and search for resolution. *Universe* **2021**, *7*, 84.

[7] Bressi, G.; Carugno, G.; Onofrio, R.; Ruoso, G. Measurement of the Casimir force between parallel metallic surfaces. *Phys. Rev. Lett.* **2002**, *88*, 041804.

[8] Brown-Hayes, M.; Dalvit, D.A.R.; Mazzitelli, F.D.; Kim, W.J.; Onofrio, R. Towards a precision measurement of the Casimir force in a cylinder-plane geometry. *Phys. Rev. A* **2005**, *72*, 052102.

[9] Lamoreaux, S.K. Demonstration of the Casimir force in the 0.6 to 6 μm range. *Phys. Rev. Lett.* **1997**, *78*, 5–8.

[10] Mohideen, U.; Roy, A. Precision measurement of the

Casimir force from 0.1 to 0.9 μm. *Phys. Rev. Lett.* **1998**, *81*, 4549–4552.
[11] Chan, H.B. Aksyuk, V.A. Kleiman, R.N. Bishop, D.J. Capasso, F. Quantum mechanical actuation of microelectromechanical system by the Casimir effect. *Science* **2001**, *291*, 1941–1944.
[12] Decca, R.S.; López, D.; Fischbach, E.; Klimchitskaya, G.L.; Krause, D.E.; Mostepanenko, V.M. Precise comparison of theory and new experiment for the Casimir force leads to stronger constraints on thermal quantum effects and long-range interactions. *Ann. Phys.* **2005**, *318*, 37–80.
[13] Chang, C.C.; Banishev, A.A.; Castillo-Garza, R.; Klimchitskaya, G.L.; Mostepanenko, V.M.; Mohideen, U. Gradient of the Casimir force between Au surfaces of a sphere and a plate measured using an atomic force microscope in a frequency-shift technique. *Phys. Rev. B* **2012**, *85*, 165443.
[14] Bimonte, G.; López, D.; Decca, R.S. Isoelectronic determination of the thermal Casimir force. *Phys. Rev. B* **2016**, *93*, 184434.
[15] Bimonte, G.; Spreng, B.; Maia Neto, P.A.; Ingold, G.-L.; Klimchitskaya, G.L.; Mostepanenko, V.M.; Decca, R.S. Measurement of the Casimir force between 0.2 and 8 μm: Experimental procedures and comparison with theory. *Universe* **2021**, *7*, 93.
[16] Boyer, T.H. Quantum electromagnetic zero point energy of a conducting spherical shell and the Casimir model for a charged particle. *Phys. Rev.* **1968**, *174*, 1764–1774.
[17] Derjaguin, B.V. Untersuchungen über die Reibung und Adhäsion, IV. *Koll.-Z.* **1934**, *69*, 155–164. https://doi.org/10.1007/BF01433225
[18] Deriagin, B.V.; Abrikosova, I.I. Direct measurement of the molecular attraction of solid bodies. I. Statement of the problem and method of measuring forces by using negative feedback. *Sov. Phys. JETP* **1957**, *4*, 819–829. Available online: `http://jetp.ras.ru/cgi-bin/e/index/e/3/6/p819?a=list` (Accessed on 6 January 2024).
[19] Derjaguin, B.V. The force between molecules. *Sci. Am.* **1960**, *203*, 47. https://doi.org/10.1038/scientificamerican0760-47
[20] Israelachvili, J.N. *Intermolecular and Surface Forces*; Academic Press Elsevier, Inc.: Oxford, UK, 2011. https://doi.org/10.1016/C2009-0-21560-1
[21] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. The proximity force approximation for the Casimir energy as a derivative expansion. *Phys. Rev. D* **2011**, *84*, 105031.
[22] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. An improved proximity force approximation for electrostatics. *Ann. Phys.* **2012**, *327*, 2050–2059.
[23] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. Derivative-expansion approach to the interaction between close surfaces. *Phys. Rev. A* **2014**, *89*, 062120.
[24] Bimonte, G.; Emig, T.; Jaffe, R.; Kardar, M. Casimir forces beyond the proximity approximation. *Europhys. Lett.* **2012**, *97*, 50001.
[25] Bimonte, G.; Emig, T.; Kardar, M. Material dependence of Casimir forces: Gradient expansion beyond proximity. *App. Phys. Lett.* **2012**, *100*, 074110.
[26] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. Derivative expansion for the Casimir effect at zero and finite temperature in $d+1$ dimensions. *Phys. Rev. D* **2012**, *86*, 045021.
[27] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. Derivative expansion for the electromagnetic and Neumann Casimir effects in $2+1$ dimensions with imperfect mirrors. *Phys. Rev. D* **2015**, *91*, 105019.
[28] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. On the derivative expansion for the electromagnetic Casimir free energy at high temperatures. *Phys. Rev. D* **2015**, *92*, 125007.
[29] Bimonte, G.; Emig, T.; Kardar, M. Casimir-Polder interaction for gently curved surfaces. *Phys. Rev. D* **2014**, *90*, 081702.
[30] Schoger, T.; Spreng, B.; Ingold, G.L.; Neto, P.A.M. Casimir effect between spherical objects: Proximity-force approximation and beyond using plane waves. *Int. J. Mod. Phys. A* **2022**, *37*, 2241009.
[31] Blocki, J.; Randrup, J.; Swiatecki, W.J.; Tsang, C.F. Proximity forces. *Ann. Phys.* **1977**, *105*, 427–462.
[32] Blocki, J.; Swiatecki, W.J. A generalization of the proximity force theorem. *Ann. Phys.* **1981**, *132*, 53–65.
[33] Myers, W.D.; Swiatecki, W.J. Nucleus-nucleus proximity potential and superheavy nuclei. *Phys. Rev. C* **2000**, *62*, 044610.
[34] Dutt, I.; Puri, R.K. Comparison of different proximity potentials for asymmetric colliding nuclei. *Phys. Rev. C* **2010**, *81*, 064609.
[35] Bhattacharjee, S.; Elimelech, M. Surface element integration: A novel technique for evaluation of DLVO interaction between a particle and a flat plate. *J. Colloid. Interface Sci.* **1997**, *193*, 273–285.
[36] Dantchev, D.; Valchev, G. Surface integration approach: A new technique for evaluating geometry dependent forces between objects of various geometry and a plate. *J. Colloid Interface Sci.* **2012**, *372*, 148–163.
[37] Decca, R.S.; Fischbach, E.; Klimchitskaya, G.L.; Krause, D.E.; Lopez, D.; Mostepanenko, V.M. Application of the proximity force approximation to gravitational and Yukawa-type forces. *Phys. Rev. D* **2009**, *79*, 124021.
[38] Milton, K.A.; Parashar, P.; Wagner, J.; Shajesh, K.V. Exact Casimir energies at nonzero temperature: Validity of proximity force approximation and interaction of semitransparent spheres. *arXiv* **2009**, arXiv:0909.0977. https://doi.org/10.48550/arXiv.0909.097
[39] Hudlet, S.; Saint Jean, M.; Guthmann, C.; Berger, J. Evaluation of the capacitive force between an atomic force microscopy tip and a metallic surface. *Eur. Phys. J. B.* **1998**, *2*, 5–10.
[40] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. Fourth order perturbative expansion for the Casimir energy with a slightly deformed plate. *Phys. Rev. D* **2012**, *86*, 125018.
[41] Emig, T.; Hanke, A.; Golestanian, R.; Kardar, M. Normal and lateral Casimir forces between deformed plates. *Phys. Rev. A* **2003**, *67*, 022114.
[42] Lambrecht, A.; Neto, P.A.M.; Reynaud, S. The Casimir effect within scattering theory. *New J. Phys.* **2006**, *8*, 243.
[43] Bordag, M.; Nikolaev, V. Casimir force for a sphere in front of a plane beyond proximity force approximation. *J. Phys. A* **2008**, *41*, 164001.
[44] Bordag, M.; Nikolaev, V. Analytic corrections to the electromagnetic Casimir interaction between a sphere and a plate at short distances. *Int. Mod. J. Phys. A* **2010**, *25*, 2171–2176.
[45] Bordag, M.; Nikolaev, V. First analytic correction beyond the proximity force approximation in the Casimir

effect for the electromagnetic field in sphere-plane geometry. *Phys. Rev. D* **2010**, *81*, 065011.
[46] Teo, L.P.; Bordag, M.; Nikolaev, V. Corrections beyond the proximity force approximation. *Phys. Rev. D* **2011**, *84*, 125037.
[47] Rodriguez-Lopez, P.; Emig, T. Casimir interaction between inclined metallic cylinders. *Phys. Rev. A* **2012**, *85*, 032510.
[48] Bimonte, G.; Emig, T. Exact results for classical Casimir interactions: Dirichlet and Drude model in the sphere-sphere and sphere-plane geometry. *Phys. Rev. Lett.* **2012**, *109*, 160403.
[49] Canaguier-Durant, A.; Ingold, G.L.; Jaekel, M.T.; Lambrecht, A.; Neto, P.A.M.; Reynaud, S. Classical Casimir interaction in the plane-sphere geometry. *Phys. Rev. A* **2012**, *85*, 052501.
[50] Bimonte, G. Going beyond PFA: A precise formula for the sphere-plate Casimir force. *EPL (Europhys. Lett.)* **2017**, *118*, 20002.
[51] Bimonte, G. Beyond-proximity-force-approximation Casimir force between two spheres at finite temperature. *Phys. Rev. D* **2018**, *97*, 085011.
[52] Bimonte, G. Beyond-proximity-force-approximation Casimir force between two spheres at finite temperature II: Plasma versus Drude modeling, grounded versus isolated spheres. *Phys. Rev. D* **2018**, *98*, 105004.
[53] Fosco, C.D.; Lombardo, F.C.; Mazzitelli, F.D. Difference between the vacuum Casimir energies for grounded and isolated conductors. *Phys. Rev. D* **2016**, *94*, 085024.
[54] Casimir, H.B.G.; Polder, D. The Influence of retardation on the London–van der Waals forces. *Phys. Rev.* **1948**, *73*, 360–372.
[55] Bimonte, G.; Emig, T.; Kardar, M. Casimir-Polder force between anisotropic nanoparticles and gently curved surfaces. *Phys. Rev. D* **2015**, *92*, 025028.
[56] Bimonte, G.; Emig, T. Interplay of curvature and temperature in the Casimir–Polder interaction. *J. Phys. Condens. Mater.* **2015**, *27*, 214018.
[57] Neto, P.A.M.; Lambrecht, A.; Reynaud, S. Roughness correction to the Casimir force: Beyond the proximity force approximation. *Europhys. Lett.* **2005**, *69*, 924–930.
[58] Genet, C.; Lambrecht, A.; Maia, Neto, P.; Reynaud, S. The Casimir force between rough metallic plates. *Europhys. Lett.* **2003**, *62*, 484–490.
[59] Emig, T.; Hanke, A.; Golestanian, R.; Kardar, M. Probing the strong boundary shape dependence of the Casimir force. *Phys. Rev. Lett.* **2001**, *87*, 260402.
[60] Henning, V.; Spreng, B.; Neto, P.A.M.; Ingold, G.L. Casimir interaction between a plane and a sphere: Correction to the proximity-force approximation at intermediate temperatures. *Universe* **2021**, *7*, 129.
[61] Henning, V.; Spreng, B.; Hartmann, M.; Ingold, G.L.; Neto, P.A.M. Role of diffraction in the Casimir effect beyond the proximity force approximation. *J. Opt. Soc. Am. B* **2019**, *36*, C77–C87.
[62] Canaguier-Durand, A.; Neto, P.A.M.; Lambrecht, A.; Reynaud, S. Thermal Casimir effect in the plane-sphere geometry. *Phys. Rev. Lett.* **2010**, *104*, 040403.
[63] Canaguier-Durand, A.; Neto, P.A.M.; Lambrecht, A.; Reynaud, S. Thermal Casimir effect for Drude metals in the plane-sphere geometry. *Phys. Rev. A* **2010**, *82*, 012511.
[64] Zandi, R.; Emig, T.; Mohideen, U. Quantum and thermal Casimir interaction between a sphere and a plate: Comparison of Drude and plasma models. *Phys. Rev. B* **2010**, *81*, 195423.
[65] Milton, K.A.; Guérout, R.; Ingold, G.-L.; Lambrecht, A.; Reynaud, S. Negative Casimir entropies in nanoparticle interactions. *J. Phys. Condens. Matter* **2015**, *27*, 214003.
[66] Ingold, G.-L.; Umrath, S.; Hartmann, M.; Guérout, R.; Lambrecht, A.; Reynaud, S.; Milton, K.A. Geometric origin of negative Casimir entropies: A scattering-channel analysis. *Phys. Rev. E* **2015**, *91*, 033203.
[67] Umrath, S.; Hartmann, M.; Ingold, G.-L.; Neto, P.A.M. Disentangling geometric and dissipative origins of negative Casimir entropies. *Phys. Rev. E* **2015**, *92*, 042125.

Breakdown of Hawking Evaporation opens new Mass Window for Primordial Black Holes as Dark Matter Candidate

Valentin Thoss,[1,2,3]★ Andreas Burkert[1,2,3] and Kazunori Kohri[4,5,6]
[1]Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 Munich, Germany
[2]Max-Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
[3]Excellence Cluster ORIGINS, Boltzmannstrasse 2, 85748 Garching, Germany
[4]Division of Science, National Astronomical Observatory of Japan (NAOJ), and SOKENDAI, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[5]Theory Center, IPNS, and QUP, KEK, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan
[6]Kavli IPMU (WPI), University of Tokyo, Kashiwa, Chiba 277-8568, Japan

arXiv:2402.17823v1 [astro-ph.CO] 27 Feb 2024

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The energy injection through Hawking evaporation has been used to put strong constraints on primordial black holes as a dark matter candidate at masses below 10^{17} g. However, Hawking's semiclassical approximation breaks down at latest after half-decay. Beyond this point, the evaporation could be significantly suppressed as was shown in recent work. In this study, we review existing cosmological and astrophysical bounds on primordial black holes taking this effect into account. We show that the constraints disappear completely for a reasonable range of parameters, which opens a new window below 10^{10} g for light primordial black holes as a dark matter candidate.

Key words: dark matter – black hole physics – gamma-rays: general

1 INTRODUCTION

The hypothesis of black holes forming in the early universe has been discussed for more than 50 years (Zel'dovich & Novikov 1967; Hawking 1971; Carr & Hawking 1974), with Chapline (1975) first to suggest that primordial black holes (PBHs) could constitute the entire dark matter of the universe. Since the 1970s, people have studied the consequences of PBHs as a dark matter candidate from the Planck mass $M_{\rm PBH} = M_{\rm pl}$ up to the 'incredulity limit'[1] beyond $M_{\rm PBH} \sim 10^{10}\,M_\odot$. This has led to strong bounds which exclude PBHs of a single mass from constituting the entirety of the dark matter with the exception of a mass window in the asteroid range $M_{\rm PBH} \in [10^{17}, 10^{22}]$ g (Carr et al. 2021, and references therein).

The lower limit is a result of constraints due to black hole evaporation at low masses. This process was first described by Hawking (1974) as he was studying the consequences of light PBHs. He showed that a black hole will emit a thermal spectrum of particles, with the temperature of the radiation scaling as $T \sim 1/M_{\rm PBH}$. The described evaporation process is self-similar and ends with a final burst as $M \to 0$.

It was soon realized that the energy injection from low-mass PBHs is in conflict with observations of γ rays, the cosmic microwave background (CMB) and the abundance of light elements produced during big bang nucleosynthesis (BBN) unless these black holes constitute only a tiny fraction of the dark matter (Chapline 1975; Hawking 1975; Novikov et al. 1979; Carr et al. 2010, for a historical overview). Furthermore, if the PBHs have a mass below $M \simeq 5 \times 10^{14}$ g, they would have completely evaporated by now (see Auffinger (2023) for a review on constraints of evaporating PBHs).

However, it is possible to avoid some of the constraints that are a result of black hole evaporation. Pacheco et al. (2023) have studied 'quasi-extremal' PBHs and found that they can be a viable dark matter candidate. Friedlander et al. (2022) and Anchordoqui et al. (2022) have investigated PBHs in the context of large extra dimensions (Arkani-Hamed et al. 1998) and showed that this opens up new mass windows for light PBHs as dark matter candidate.

In this work, we focus on 'ordinary' 4D black holes. Dvali et al. (2020) have shown that Hawking's semiclassical calculations break down at latest when the black hole has lost roughly half of its initial mass. Hawking's result is only exact in the limit $M \to \infty, G \to 0, r_s = \text{const.}$, where G is the gravitational constant and $r_s = 2GM/c^2$ the Schwarzschild radius. It entirely neglects the backreaction of the emission on the black hole itself. However, this effect can no longer be ignored when the energy of the released quanta becomes comparable to that of the black hole. Dvali et al. (2020) summarize this as: 'An old black hole that lost half of its mass is by no means equivalent to a young classical black hole of equal mass.' The crucial insight by Dvali et al. (2020) is that this backreaction leads to a universal effect of so-called 'memory burden', first introduced by Dvali (2018). This can significantly suppress further evaporation.

As Dvali et al. (2020) have discussed, there are two possibilities for the fate of an evaporating black hole beyond half-decay. Either it continues to emit quanta with a strongly suppressed rate due to the memory burden or a new classical instability sets in. In the latter case, light PBHs cannot constitute the dark matter. However, if the

★ E-mail: vthoss@usm.lmu.de

[1] This term was coined by B. Carr and refers to the limit that at least one black hole must exist in a given environment (e.g. galaxy, universe).

© 2023 The Authors

evaporation is strongly suppressed there is the possibility that PBHs can be a viable dark matter candidate as was already pointed out by Dvali et al. (2020).

In our work we want to investigate the constraints on PBHs that are subject to the effect of memory burden and compare them to the results obtained within the semiclassical Hawking picture. Following Dvali et al. (2020), we will describe the strength of the suppression by a single parameter k and study the bounds on the dark matter fraction of PBHs $f_{\rm PBH}(k, M_{\rm PBH})$.

We want to emphasise that there is no precise understanding of the evaporation process beyond the semiclassical regime yet. Therefore, our results should be understood as a rough guide on how the landscape of PBH constraints changes as one goes beyond the Hawking picture. In the following, we will not discuss details of the memory burden effect and instead refer to the original publications (Dvali 2018; Dvali et al. 2020). In addition to the memory burden effect, there are also other potential quantum properties of black holes such as vortices (Dvali et al. 2022) than can affect the evaporation that we do not investigate here.

We want to mention the work of Alexandre et al. (2024) who studied the effect of memory burden on PBH constraints from BBN and CMB distortions and her results are complimentary to our study.

This paper is organized as follows: In Section 2 we discuss our modified model of the evaporation process and how we compute the various constraints. Our results are presented and discussed in Section 3. We conclude with a summary in Section 4.

2 METHODS

In this Section, we briefly introduce the physics of black hole evaporation and how it is modified in our model of the memory burden effect. We then discuss the methodology to obtain important constraints on PBHs. We conclude by reviewing further constraints that we do not study in this work in more detail.

Henceforth, we will denote the mass of PBHs at formation by M_0, and their mass today by M_t. To keep our results as general as possible we use a monochromatic mass function. The quantity $\beta_{\rm PBH} = \rho_{\rm PBH,0}/\rho_0$ gives the fraction of the universe's density in PBHs at the time of their formation. $f_{\rm PBH}$ refers to the present dark matter fraction of the PBHs, while $f_{\rm PBH,0}$ denotes the dark matter fraction at formation time. They are related via

$$f_{\rm PBH} = \frac{1}{1 + \frac{1-f_{\rm PBH,0}}{f_{\rm PBH,0}} \frac{M_0}{M_t}}, \qquad f_{\rm PBH,0} = \frac{\beta_{\rm PBH}}{\Omega_{\rm DM}}, \tag{1}$$

where $\Omega_{\rm DM}$ is the dark matter density at the time of PBH formation. We will use standard values for the Hubble rate $h = 0.67$, the number of relativistic degrees of freedom at formation $g_* = 106.75$ and for the factor $\gamma = 1$ which gives the ratio of the final black hole mass M_0 to the mass within a Hubble volume M_H at the time of formation (see Equations 2-6 in Carr et al. (2021)).

2.1 Evaporation process

The semiclassical (SC) emission rate of a particle species i with energy E from a black hole with current mass M and temperature T is given by (Hawking 1975)

$$\frac{{\rm d}^2 N_{i,\rm SC}}{{\rm d}E{\rm d}t}(E, M, s_i) = \frac{g}{2\pi\hbar} \frac{\Gamma(E, M, s_i)}{e^{E/k_B T(M)} - (-1)^{s_i}}, \tag{2}$$

where g_i specifies the degrees of freedom of the particle emitted, s_i its spin and Γ are the greybody factors. The black hole temperature T is directly related to its mass via

$$k_B T = \frac{\hbar c^3}{8\pi G M}. \tag{3}$$

The emission of fundamental particles leads to a mass loss rate $\dot{M} = -\mathcal{F}(M)/M^2$ of the black hole, where $\mathcal{F}(M)$ encompasses the degrees of freedom of the particles which the black hole emits at a certain mass. Both $\mathcal{F}(M)$ and the greybody factors are computed using the publicly available code BlackHawk (Arbey & Auffinger 2019). The lifetime of a black hole with mass M in the Hawking picture is given by $t_{\rm SC}(M) \approx M^3/(3\mathcal{F}(M))$.

Our modified model of the evaporation process assumes the validity of Hawking's results until the black hole has reached a mass of $M = qM_0$. Unless otherwise stated, we use $q = 1/2$, i.e. the breakdown happens at half-decay. Beyond this point, the emission rate is suppressed (denoted by MB for memory burden), which we will parameterise following Dvali et al. (2020) as

$$\frac{{\rm d}^2 N_{i,\rm MB}}{{\rm d}E{\rm d}t}(E, M_0, s_i) = \frac{1}{S(qM_0)^k} \frac{{\rm d}^2 N_{i,\rm SC}}{{\rm d}E{\rm d}t}(E, qM_0, s_i), \tag{4}$$

where k is an exponent that controls the strength of the suppression and S is the black hole entropy in units of the Boltzmann constant k_B given by

$$S = \frac{4\pi M^2 G}{\hbar c} \approx 2.6 \times 10^{10} \left(\frac{M}{1\,{\rm g}}\right)^2. \tag{5}$$

Note that the large value of the entropy implies that the emission rate decreases by many orders of magnitude for $k \sim 1$. In this model the evaporation rate and the black hole temperature remain constant. This implies a linear decay of the black hole with lifetime $t_{\rm MB}(M) \approx \frac{(qM_0)^3 S_0^k}{\mathcal{F}(qM_0)}$, neglecting the time spent in the SC regime.

As an alternative scenario, we could let the black hole temperature increase in the usual way, leading to a black hole explosion and

$$\frac{{\rm d}^2 N_{i,\rm MB}}{{\rm d}E{\rm d}t}(E, M, s_i) = \frac{1}{S(M)^k} \frac{{\rm d}^2 N_{i,\rm SC}}{{\rm d}E{\rm d}t}(E, M, s_i) \tag{6}$$

This leads to a very similar lifetime $t_{\rm MB}(M) \approx \frac{(qM_0)^3 S_0^k}{(3+2k)\mathcal{F}(qM_0)}$ and only makes a meaningful difference to the other scenario when the lifetime of the black hole is comparable or less than the Hubble time. However, such scenarios are heavily constrained ($f_{\rm PBH,0} \ll 1$) as we will show in this work. For simplicity and since we do not know the late-time behaviour of the black hole in such detail, we will restrict our discussion to the former scenario. We want to emphasise that (in both cases) due to the dependence on the initial black hole mass M_0 the evaporation process is no longer self-similar.

2.2 Secondary emission

Many of the particles that a sufficiently light black hole emits are not stable and decay or annihilate which leads to additional secondary emission of stable particles such as photons. The secondary emission of particles of type i is given by

$$\frac{{\rm d}^2 N_{i,\rm sec}}{{\rm d}E{\rm d}t} = \sum_j \int {\rm d}E'\, {\rm Br}_{j\to i}(E, E') \frac{{\rm d}^2 N_j}{{\rm d}E'{\rm d}t}, \tag{7}$$

where ${\rm Br}_{j\to i}(E, E')$ denotes the branching ratios that are calculated from particle physics codes. In our case we use the spectra computed by BlackHawk and additionally make use of the code HDMSpectra

(Bauer et al. 2021) for energies beyond the GeV-scale[2]. BlackHawk makes use of the codes PYTHIA (Sjöstrand et al. 2015), Herwig (Bellm et al. 2016) and Hazma (Coogan et al. 2020). All of them are suited for different energy ranges. For our purposes we use Hazma for black holes with $k_B T < 0.1$ GeV, HDMSpectra for temperatures $k_B T > 100$ GeV and Herwig in the intermediate range. Note that for a given black hole temperature we only include secondary emission for $E > 10^{-6} k_B T$ as we do not trust the results at lower energies.

2.3 Galactic γ ray emission

If PBHs make up a sizeable fraction $f_{\rm PBH}$ of the dark matter halo of the Milky way, they would be detectable through their photon emission. Comparisons with observations have been used to put strong constraints on PBHs in the mass range $M_0 \in [5\times10^{14}, 10^{17}]$ g (e.g. Carr et al. (2016), see also Auffinger (2023) for a recent review).

The measured flux of photons would be

$$\Phi_{\rm PBH} = \frac{f_{\rm PBH}}{4\pi M_t \Delta\Omega} \frac{{\rm d}^2 N_\gamma}{{\rm d}E{\rm d}t} \int_{\Delta\Omega} {\rm d}\Omega \int {\rm d}r \rho_{\rm DM}(R(r,l,b)), \tag{8}$$

with $\Delta\Omega$ being the observed field of view on the sky and R, l, b the galactocentric distance, longitude and latitude, respectively. The constraint on $f_{\rm PBH}$ is obtained by requiring that

$$\int_{E_{\rm low}}^{E_{\rm up}} {\rm d}E\, \Phi_{\rm PBH} \leq \Phi_{\rm gc}(E_{\rm up} - E_{\rm low}), \tag{9}$$

where $\Phi_{\rm gc}$ is the measured γ ray flux in the energy bin $[E_{\rm low}, E_{\rm up}]$.

For $\rho_{\rm DM}(R)$ we choose an NFW profile with the 'convenient' set of parameters from McMillan (2011). This means our results will be identical to Auffinger (2022) in the semiclassical limit, since we also use the observational data from their Isatis code. This includes data from *INTEGRAL*, *COMPTEL*, *EGRET* and *Fermi-LAT* (Bouchet et al. 2011; Strong et al. 1994, 1999; Strong & Mattox 1996; Strong 2011). Since we are also interested in PBHs of lower masses (and thus higher spectral energies) than usually investigated, we extend this dataset by the recent observational data from *LHAASO* in the energy band from 10^4 to 10^6 GeV (Cao et al. 2023, 'inner region').

2.4 Extragalactic γ ray background

The evaporation of PBHs between the time of recombination $t_{\rm rec}$ and today t_0 will lead to a contribution to the extragalactic γ ray background. This has been used to constrain PBHs with masses $M_0 \in [3\times10^{13}, 10^{17}]$ g (Carr et al. 2010; Ballesteros et al. 2020; Arbey et al. 2020; Chen et al. 2022).

Assuming a homogenous distribution of dark matter, the observed flux is

$$\Phi_{\rm PBH} = \frac{c n_t}{4\pi} \int_{t_{\rm rec}}^{t_0} {\rm d}t\, (1+z) \frac{{\rm d}^2 N_\gamma}{{\rm d}E{\rm d}t}\Big((1+z)E, M(t)\Big), \tag{10}$$

where n_t is today's number density of the PBHs which is related to $f_{\rm PBH,0}$ via

$$n_t \approx 2.2\times10^{-30}\,{\rm cm}^{-3}\, f_{\rm PBH,0} \left(\frac{M_0}{1\,{\rm g}}\right)^{-1}. \tag{11}$$

The constraint on $f_{\rm PBH,0}$ (and $f_{\rm PBH}$) is obtained analogously to Equation 9. We again use the observational data available in the Isatis code from *HEAO*, *COMPTEL*, *EGRET* and *Fermi-LAT* (Gruber et al. 1999; Ruiz-Lapuente et al. 2016; Strong et al. 2004; Ackermann et al. 2015). In addition, we include the flux data from *LHAASO* (Cao et al. 2023, 'outer region').

2.5 CMB anisotropies

The injection of energy from PBH evaporation after recombination can ionize the otherwise neutral medium. This in turn leads to the rescattering of CMB photons which affects the angular power spectrum of temperature and polarization. Comparisons with measurements of the CMB have led to strong constraints on PBH in the mass range $M_0 \in [3\times10^{13}, 10^{17}]$ g (Poulin et al. 2017; Stöcker et al. 2018; Acharya & Khatri 2020; Cang et al. 2022). Stöcker et al. (2018) have developed the publicly available code ExoCLASS, a branch of the Boltzmann code CLASS (Blas et al. 2011). It is able to compute the CMB power spectra for any value of M_0 and $f_{\rm PBH,0}$. To achieve this, the code computes the rate of energy density deposition as a function of redshift,

$$\left.\frac{{\rm d}^2 E}{{\rm d}t{\rm d}V}\right|_{{\rm dep},\alpha}(z) = h_\alpha(z) \left.\frac{{\rm d}^2 E}{{\rm d}t{\rm d}V}\right|_{\rm inj}(z) = \frac{h_\alpha(z) f_{\rm PBH,0} \rho_{\rm DM,t} (1+z)^3 \dot{M}}{M_0}, \tag{12}$$

where α denotes heating, ionization or excitation+ionization as the possible channels. The energy deposition function $h_\alpha(z)$ is a convolution of the PBH spectrum with a transfer function $T^i_\alpha(z', z, E)$. This function gives the fraction of the energy E injected at redshift z' in a channel α that is deposited at redshift z for the particle type i.

To obtain constraints on $f_{\rm PBH,0}$, we use ExoCLASS and Montepython (Audren et al. 2013; Brinckmann & Lesgourgues 2019) to perform a Markov chain Monte Carlo (MCMC) analysis. The model uses the Planck TT,TE,EE+lowE+lensing likelihood (Planck Collaboration et al. 2020) with $f_{\rm PBH,0}$ as an additional cosmological parameter. To perform the calculations, we modified ExoCLASS and implemented our model of the evaporation process as described in Section 2.1. In addition, we also change the computation of the PBH spectra in the DarkAges module of the code. We replace the geometric optics approximation with the full results from BlackHawk including secondary emission as described in Section 2.2. Note that in the semiclassical limit we reproduce the results from Stöcker et al. (2018) within a factor of a few.

Unfortunately, the transfer function used in ExoCLASS is only tabulated up to $E \sim 1$ TeV and for higher energies the code simply uses the last tabulated value. Stöcker et al. (2018) mention that this is a good approximation up to at least several hundred TeV, which corresponds to black holes with a mass of $M \sim 10^9$ g. For lower masses our results rely on the assumption that further extrapolation of the transfer function is a reasonable approximation.

Since we have k as an additional free parameter, it would be computationally very expensive to perform a large number of MCMC runs over the entire parameter space of M_0 and k and given the simple nature of our model we do not aim for very precise results. In fact, as we are going to show in this work, the constraints from CMB anisotropies are weaker than those from galactic and extragalactic γ rays by several orders of magnitude for most of the parameter space and thus not as relevant, which justifies approximations.

If the initial black hole mass is larger than $M_0 \simeq 3\times10^{13}$ g, then the semiclassical phase of the evaporation ($M_0 \to q M_0$) is still ongoing after recombination. In this case, one can expect the constraints to change only by a factor of a few if $q \sim 1/2$, as the black hole still releases a sizable fraction $1-q$ of its total energy $M_0 c^2$. To

[2] By the time of publication HDMSpectra has now been implemented into BlackHawk

investigate this, we perform a MCMC run for $M_0 = 5 \times 10^{13}$ g were the black hole evaporation is stopped after half-decay. We compare it to a simulation with the full evaporation process. The resulting contraints on $f_{\mathrm{PBH},0}$ only differ by roughly 20 per cent. As this is significantly smaller than the modelling uncertainty of around two dex (see e.g. Figure 6 in Auffinger (2023)), we ignore the effect of memory burden in this mass range and just use the semiclassical results from Stöcker et al. (2018).

To obtain the constraints for PBHs with $M_0 < 3 \times 10^{13}$ g we run a full MCMC for $M_0 \in [10^3, 10^5, 10^7, 10^9, 10^{11}, 10^{13}]$ g and interpolate the results for other values of M_0 and k. To do this, we assume that f_{PBH} scales linearly with the total amount of energy injection, as is motivated by Equation (12). Effectively, we are rescaling the constraints on $f_{\mathrm{PBH},0}(M_0, k)$ computed using ExoCLASS to obtain $f_{\mathrm{PBH},0}(M_0', k')$ through

$$f_{\mathrm{PBH},0}(M_0', k') \frac{\Delta M'}{M_0'} = f_{\mathrm{PBH},0}(M_0, k) \frac{\Delta M}{M_0}, \tag{13}$$

where ΔM is the change in mass from recombination to the end of reionization which depends on k. This does not require computing an angular power spectrum and is treating $h_\alpha(z)$ as constant. The latter approximation is exact if we keep M_0 constant as $h_\alpha(z)$ only depends on the shape of the PBH spectrum which is unaffected by k. We discuss this approximation, its computation and its validity in more detail in Appendix B.

2.6 Big bang nucleosynthesis

The abundance of light elements in the universe is the earliest cosmological probe to study primordial black hole evaporation. In the Hawking picture, black holes with masses $M_0 \in [10^{10}, 10^{13}]$ g evaporate during or after the formation of light elements. The emitted radiation will alter the neutron-to-proton ratio and lead to photo- and hadrodissociation of elements. Since the standard BBN scenario predicts the abundance of light elements with great success, any modification will be heavily constrained. Carr et al. (2021) (updated from Carr et al. (2010) through results of Kawasaki et al. (2018) and Hasegawa et al. (2019)) find that the initial dark matter fraction is constrained to $f_{\mathrm{PBH},0} \lesssim 10^{-3}$ in the denoted mass range with other studies obtaining comparable results (Acharya & Khatri 2020; Keith et al. 2020).

To understand how these results will be affected by the memory burden, we need to distinguish two regimes as in Section 2.5. For $M_0 \gtrsim 10^{10}$ g the semiclassical evaporation phase is still ongoing during or after nucleosynthesis. Therefore, one expects the constraints to change only by a factor of order unity as the black hole still releases half of its total energy $M_0 c^2$. To understand this in more detail we use of the model described in Kawasaki et al. (2018) and modify it by halting the evaporation process at half-mass. This is a good approximation for $k \gtrsim 0.5$, since the lifetime of the black hole exceeds the age of the universe for $M_0 > 10^{10}$ g in this case.

On the other hand, if the initial mass of the PBHs is below $M_0 \sim 10^{10}$ g then the memory burden slows down the evaporation before they can affect the abundance of light elements. If the black holes are still present today then their evaporation rate in the early universe will be completely negligible (compare this to the semiclassical case where there are no BBN bounds on PBHs with a lifetime greater than $t \sim 10^6$ yr). Conversely, if they are light enough to evaporate significantly in the early universe and affect the abundance of light elements then they will not survive until today. For our purposes we can therefore ignore this regime. This is also justified by the results of Alexandre et al. (2024) who discuss the constraints from BBN in more detail.

2.7 Other constraints

In addition to the constraints already discussed, there are a large number of other bounds on evaporating PBHs (for a full overview see the review of Auffinger 2023).

Many of these are conceptually similar to the constraints from the galactic γ ray emission but instead look at the flux of electrons, positrons or neutrinos (Boudaud & Cirelli 2019; Dasgupta et al. 2020), the 511 keV annihilation signal (Laha 2019; DeRocco & Graham 2019), radio emission from synchrotron radiation Chan & Lee (2020), heating of the ISM (Kim 2021) or take into account other astrophysical contributions to get stronger bounds (Berteaud et al. 2022). They are all usually within one order of magnitude of the conservative approach that we use here.

Another important constraint arises from spectral distortions of the cosmic microwave background (Chluba et al. 2020; Acharya & Khatri 2020; Lucca et al. 2020). It affects black holes with masses in the range $M_0 \in [10^{11}, 10^{13}]$ g. One can follow the same arguments as in the previous section to conclude that the bounds will not change dramatically in the regime where the semiclassical evaporation is still (partly) ongoing. During the phase of slowed down evaporation, the constraints only affect black holes which would not survive until today. This is also discussed in the work of Alexandre et al. (2024). Since the constraints from CMB spectral distortions are (currently) less strong than those from big bang nucleosynthesis and cover a smaller range of masses we will ignore them for our purposes.

If PBH are light enough and make up a sizeable fraction of the dark matter, then they will frequently come close to or even transit through the earth. The expected collision rate is (Luo et al. 2012)

$$\Gamma = 10^8\ \mathrm{yr}^{-1} f_{\mathrm{PBH}} \left(\frac{M}{1\,\mathrm{g}}\right)^{-1} \left(\frac{\langle v \rangle}{200\,\mathrm{km/s}}\right) \tag{14}$$

where $\langle v \rangle$ is the mean dark matter velocity. This assumes that the PBHs are smoothly distributed throughout the solar system. While this would imply a sizeable number of collisions per year for $M \ll 10^8$ g, these are virtually unobservable. In the same work it was found that a black hole with $M_0 = 10^{15}$ g would produce a seismic event with a magnitude of $M_w = 4$. Since the amplitude of the seismic waves scales linearly with the black hole mass according to Luo et al. (2012), this would imply immeasurable magnitudes $M_w \ll 1$ for black holes with mass $M_0 \ll 10^{12}$ g. While close encounters with the Earth would be more numerous they are still too weak to be detectable in the mass range that we are interested in.

Finally, we want to mention that in the standard formation scenario of PBHs from spherical collapse of overdensities there is a lower mass limit of $M_0 \sim 0.1$ g that is set by the time of the end of inflation.

3 RESULTS AND DISCUSSION

In this Section, we discuss the results and possible uncertainties in our model. The combined map of all constraints is shown in Fig. 1. The results for galactic and extragalactic γ ray emission are also shown separately in the Appendix.

3.1 Galactic γ ray emission

Figure A1 shows the constraints obtained from galactic γ ray emission. PBHs with initial mass $M_0 \gtrsim 5 \times 10^{14}$ g have not yet lost half of

their mass and thus the usual constraints apply which rule out PBHs to be the entirety of the dark matter up to $M_0 \sim 10^{17}$ g independent of k. For $k > 0.2$ a new mass window emerges that extends down to $M_0 = 10^9$ g for $k = 1$ and to $M_0 = 4 \times 10^5$ g for $k = 2$. Note that when only primary photon emission is considered (dashed lines in Figure A1), then there are no bounds for $k > 1.3$ or $M < 6 \times 10^5$ g beyond the trivial constraint that they would have evaporated by now. However, when considering the emission of secondary photons, we obtain bounds on these PBHs far beyond these limits. We do not find any constraints from galactic γ ray emission for $M_0 < 3 \times 10^2$ g. The radiation of these light black holes goes beyond the energy range covered by any current γ ray observations. In this case one would need to consider indirect effects of these highly energetic photons (beyond PeV) to obtain limits. For each value of k, the constraints on $f_{\rm PBH,0}$ extend roughly two orders of magnitude beyond the mass at which the PBHs would completely evaporate by today.

3.2 Extragalactic γ ray background

In Figure A2, the constraints from the extragalactic γ ray background are shown. In the Hawking picture, the dark matter fraction of PBHs is limited to $f_{\rm PBH,0} < 1$ in the mass range from $M_0 \approx 3 \times 10^{13}$ g to $M \sim 10^{17}$ g. As explained before, for initial black hole masses $M \gtrsim 5 \times 10^{14}$ g there are no changes. In the range $M_0 \in [3 \times 10^{13}, 5 \times 10^{14}]$ g the constraints are softened at most by one order of magnitude when taking into account all observational data. This is because the semiclassical evaporation phase is still ongoing after recombination. The lower end of the constrained mass range (which is roughly set by the mass of a black hole that evaporates at recombination) changes by only 4 per cent. A new window for PBHs as dark matter candidates opens at lower masses for $k > 0.35$. It extends down to $M = 10^9$ g for $k = 1$ and to $M = 10^5$ g for $k = 2$. If only primary photon emission is considered, then the constraints become significantly weaker as can be seen from the dashed lines in Figure A2. We are not able to constrain black hole holes with $M_0 \lesssim 10^2$ g as the energy exceeds the current observational range, as already discussed for the galactic γ ray emission.

3.3 Other constraints

The constraints on PBHs from CMB anisotropies are weaker than those from galactic and extragalactic γ ray emission by 4-6 orders of magnitude for $M_0 < 10^{12}$ g and by around 2 dex for $M_0 \in [10^{12}, 3 \times 10^{13}]$ g. Note that for $M_0 < 10^9$ g our results rely on extrapolation of the transfer function and are thus subject to some uncertainty. For $M_0 > 3 \times 10^{13}$ g we just give the constraints from Stöcker et al. (2018) as explained in Section 2.5.

The bound from BBN in the mass range $M_0 \in [10^{10}, 10^{13}]$ g weakens by around 45 per cent if the evaporation is stopped at half-mass. As discussed in Section 2.6 there are no other constraints from BBN for black holes which survive to the present day.

3.4 Combined constraints

The combined map of constraints $f_{\rm PBH,0}(k, M_0)$ is shown in Figure 1. Figure 2 shows $f_{\rm PBH,0}(M_0)$ for $k = 2$, which is chosen for illustrative purposes. One can identify the constraints from the semiclassical evaporation phase as the vertical, almost k-independent band in the range $M_0 \in [10^{10}, 10^{17}]$ g. The slowed down evaporation phase leads to constraints that are dependent on k. For $k \gtrsim 1.0$ a new mass window emerges for which PBHs can make up the entirety of the

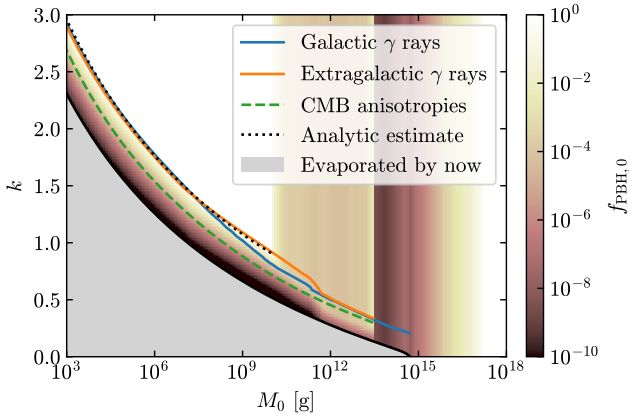


Figure 1. Combined constraints on $f_{\rm PBH,0}(k, M_0)$. The coloured lines show $f_{\rm PBH,0} = 1$ contours for each type of constraint. In the white region, PBHs can make up the entirety of the dark matter. The black, dotted line displays the analytic estimate given by Equation (15). The different linestyles are chosen for better readability.

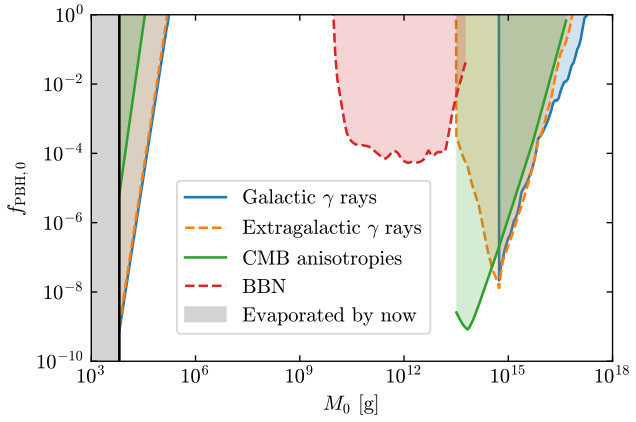


Figure 2. Constraints on $f_{\rm PBH,0}(M_0)$ for $k = 2$. The different linestyles are chosen for better readability.

dark matter. The upper mass limit of $M_0 \sim 10^{10}$ g is set by the time of nucleosynthesis as the PBHs must reach the stage of suppressed evaporation before the onset of BBN. On the other hand, the lower bound depends on the strength of the suppression k as it determines the lifetime of the black hole.

We can understand our results in simple terms by considering the fraction α of the PBH mass that evaporates during memory burden until today ($qM_0 \to (1-\alpha)qM_0$). From Equation (4) it follows that

$$M_0 \approx \left(\frac{\mathcal{F}(qM_0)t_0}{\alpha}\right)^{\frac{1}{3+2k}} \left(\frac{\hbar c}{4\pi G}\right)^{\frac{k}{3+2k}}, \tag{15}$$

where t_0 is the age of the universe. This is is plotted in Fig. 1 for $\alpha = 10^{-13}$ and $\mathcal{F}(qM_0) \approx \mathcal{F}(0) \approx 8.2 \times 10^{26}$ g^3 s^{-1} and gives a good approximation for the lower bound on the black hole mass in the parameter range that we study. Put differently, this implies that the memory burden effect must increase the lifetime of the black hole to at least $\sim 10^{13}$ times the age of the universe in order to avoid all existing constraints.

3.5 Model uncertainties

Due to the lack of a detailed understanding of the evaporation process beyond the semiclassical limit, we keep our model of the evaporation process as simple as possible. Motivated by Dvali et al. (2020) we model the strength of the suppression by a factor $1/S_0^k$ with S_0 being the black hole entropy. This means k only depends logarithmically on other parameters in our models. Therefore any uncertainties in our models of the constraints will only weakly affect the bound on $k(M_0)$ where $f_{\mathrm{PBH},0} = 1$ (i.e. the coloured lines in Figure 1). More specifically, if e.g. the rate of galactic γ ray emission decreases by a factor of 10 this will shift the bound $k(M_0)$ by less than 0.05 for the range of mass that we study.

Another uncertainty in our model is the question on how to set the black hole temperature and its spectrum. When the semiclassical approximation breaks down we can no longer expect the emission to be thermal. In our model we keep the shape of the spectrum fixed as a zeroth-order approximation. Should the spectrum change very drastically (but keeping $\mathrm{d}N/\mathrm{d}t$ constant) then this will most strongly affect the constraints from galactic and extragalactic emission. If the emission becomes more soft then the bounds would become weaker and vice versa. This is due to the larger observed γ ray flux at lower energies. However, even a drastic change in f_{PBH} will not shift the bound on $k(M_0)$ by much as explained. The constraints from CMB anisotropies are much more robust to uncertainties in the black hole spectrum since they depend mostly on the total rate of emission.

Similar arguments can be made about the uncertainties regarding the secondary emission of the PBHs. During the phase of memory burden it is in principle possible that the emission of particles is no longer 'democratic', i.e. the emission of certain particle species are preferred (see also the discussion in Alexandre et al. (2024)). However, as long as the emission rate only changes within an order of magnitude our bounds should not change dramatically.

Regarding the black hole temperature we discussed two different approaches in Section 2.1. For the results discussed so far we keep it fixed once the black hole has lost half of its mass. If we instead assume $T \sim 1/M$ throughout the memory burden phase (i.e. as in Equation (6)) then we find that it changes our results only when the lifetime is comparable to or less than a Hubble time. Thus it only affects the constraints in a regime where PBHs are already excluded from being a sizeable fraction of the dark matter ($f_{\mathrm{PBH},0} \sim 10^{-10}$).

Finally, for our analysis we defined the onset of the memory burden effect when the black hole loses half of its mass by setting $q = 1/2$. As Dvali et al. (2020) have shown this is the latest time at which a quantum backreaction is unavoidable. Dvali & Panchenko (2015) and Michel & Zell (2023) have shown that in certain prototype systems of black holes the breakdown of Hawking's calculations could happen already after a fraction $1/\sqrt{S}$ of the semiclassical lifetime. While this requires more microscopic justification it is still worthwhile to consider the effect on the constraints of PBH. Such an early transition into the phase of memory burden would correspond to

$$1 - q \approx 2 \times 10^{-6} (M_0/\mathrm{g})^{-1}, \tag{16}$$

which would imply that black holes with $M_0 > 10^{10}$ g reach the memory burden phase after losing a fraction of less than 10^{-15} of their initial mass. This implies that any constraints that arise from the semiclassical evaporation phase (the 'vertical band' in Figure 1) will be significantly weakened. When we recompute these constraints under this assumption then PBHs can make up the entirety of the dark matter for $M_0 > 10^{10}$ g. One is still left with the constraints from the slowed down phase of the evaporation. This would open a significantly larger window for PBHs as a viable dark matter candidate.

4 SUMMARY AND CONCLUSIONS

In this work we investigate how the constraints on evaporating primordial black holes change when one goes beyond the semiclassical calculations from Hawking (1974). When the quantum backreaction of the emitted particles is taken into account the evaporation can slow down drastically due to the effect of memory burden described by Dvali et al. (2020). Here we compute the bounds on the dark matter fraction of PBHs as a function of the currently unknown parameter k that quantifies the effect memory burden. It gives the suppression of the evaporation rate in powers of the black hole entropy $1/S^k$.

We find that for $k > 1.0$ a new mass window emerges where PBHs can be a viable dark matter candidate. It extends up to $M_0 \sim 10^{10}$ g with higher mass being ruled out due to constraints from big bang nucleosynthesis. The lower end of the mass window depends on the parameter k, extending to $M_0 \sim 10^5$ g for $k = 2$.

The bounds that we obtain with our simple model of the evaporation process are quite robust and should hold unless the breakdown of the semiclassical evaporation phase happens much earlier or the spectra of the PBH evaporation changes dramatically. It should be noted that all the constraints that we obtain in this work are computed for a monochromatic mass function and zero spin of the black hole to keep our results as general as possible. In reality, PBHs will form with an extended distribution of initial mass. Studying bounds on the dark matter fraction of PBHs for more realistic mass functions and a non-zero spin distribution is left for future work.

Our results are supposed to provide a first insight on the landscape of PBH constraints beyond the Hawking picture of black hole evaporation. A more detailed understanding of the actual evaporation process remains a significant theoretical challenge but will undoubtedly help to study the possibility that dark matter exists in the form of light primordial black holes.

In addition to open questions about the fundamental nature of black holes it would also be of great interest to study the potential observational signatures and possibilities to detect PBHs in this new mass window where they are a viable dark matter candidate.

ACKNOWLEDGEMENTS

We thank Gia Dvali for giving us the idea for this project, as well as for his supervision. We also would like to thank Florian Kühnel, Sebastian Zell and Manuel Behrendt for helpful comments. We are grateful for the help of Patrick Stöcker with regards to the ExoCLASS code. This research was supported by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 - 390783311.

DATA AVAILABILITY

The numerical results presented in this article are publicly accessible under Thoss (2024)

SOFTWARE

BlackHawk 2.1 (Arbey & Auffinger 2019), Exoclass (Stöcker et al. 2018) and Class (Blas et al. 2011), Montepython 3.6.0 (Audren et al. 2013; Brinckmann & Lesgourgues 2019), Julia 1.9.3 (Bezanson et al. 2017), Matplotlib 3.7.1 (Hunter 2007)

REFERENCES

Acharya S. K., Khatri R., 2020, J. Cosmology Astropart. Phys., 2020, 010
Ackermann M., et al., 2015, ApJ, 799, 86
Alexandre A., Dvali G., Koutsangelas E., 2024, arXiv e-prints, p. arXiv:2402.14069
Anchordoqui L. A., Antoniadis I., Lüst D., 2022, Phys. Rev. D, 106, 086001
Arbey A., Auffinger J., 2019, European Physical Journal C, 79, 693
Arbey A., Auffinger J., Silk J., 2020, Phys. Rev. D, 101, 023010
Arkani-Hamed N., Dimopoulos S., Dvali G., 1998, Physics Letters B, 429, 263
Audren B., Lesgourgues J., Benabed K., Prunet S., 2013, J. Cosmology Astropart. Phys., 2013, 001
Auffinger J., 2022, European Physical Journal C, 82, 384
Auffinger J., 2023, Progress in Particle and Nuclear Physics, 131, 104040
Ballesteros G., Coronado-Blázquez J., Gaggero D., 2020, Physics Letters B, 808, 135624
Bauer C. W., Rodd N. L., Webber B. R., 2021, Journal of High Energy Physics, 2021, 121
Bellm J., et al., 2016, European Physical Journal C, 76, 196
Berteaud J., Calore F., Iguaz J., Serpico P. D., Siegert T., 2022, Phys. Rev. D, 106, 023030
Bezanson J., Edelman A., Karpinski S., Shah V. B., 2017, SIAM Review, 59, 65
Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 2011, 034
Bouchet L., Strong A. W., Porter T. A., Moskalenko I. V., Jourdain E., Roques J.-P., 2011, ApJ, 739, 29
Boudaud M., Cirelli M., 2019, Phys. Rev. Lett., 122, 041104
Brinckmann T., Lesgourgues J., 2019, Physics of the Dark Universe, 24, 100260
Cang J., Gao Y., Ma Y.-Z., 2022, J. Cosmology Astropart. Phys., 2022, 012
Cao Z., et al., 2023, Phys. Rev. Lett., 131, 151001
Carr B. J., Hawking S. W., 1974, MNRAS, 168, 399
Carr B. J., Kohri K., Sendouda Y., Yokoyama J., 2010, Phys. Rev. D, 81, 104019
Carr B. J., Kohri K., Sendouda Y., Yokoyama J., 2016, Phys. Rev. D, 94, 044029
Carr B., Kohri K., Sendouda Y., Yokoyama J., 2021, Reports on Progress in Physics, 84, 116902
Chan M. H., Lee C. M., 2020, MNRAS, 497, 1212
Chapline G. F., 1975, Nature, 253, 251
Chen S., Zhang H.-H., Long G., 2022, Phys. Rev. D, 105, 063008
Chluba J., Ravenni A., Acharya S. K., 2020, MNRAS, 498, 959
Coogan A., Morrison L., Profumo S., 2020, J. Cosmology Astropart. Phys., 2020, 056
Dasgupta B., Laha R., Ray A., 2020, Phys. Rev. Lett., 125, 101101
DeRocco W., Graham P. W., 2019, Phys. Rev. Lett., 123, 251102
Dvali G., 2018, arXiv e-prints, p. arXiv:1810.02336
Dvali G., Panchenko M., 2015, arXiv e-prints, p. arXiv:1507.08952
Dvali G., Eisemann L., Michel M., Zell S., 2020, Phys. Rev. D, 102, 103523
Dvali G., Kühnel F., Zantedeschi M., 2022, Phys. Rev. Lett., 129, 061302
Friedlander A., Mack K. J., Schon S., Song N., Vincent A. C., 2022, Phys. Rev. D, 105, 103508
Gruber D. E., Matteson J. L., Peterson L. E., Jung G. V., 1999, ApJ, 520, 124
Hasegawa T., Hiroshima N., Kohri K., Hansen R. S. L., Tram T., Hannestad S., 2019, J. Cosmology Astropart. Phys., 2019, 012
Hawking S., 1971, MNRAS, 152, 75
Hawking S. W., 1974, Nature, 248, 30
Hawking S. W., 1975, Communications in Mathematical Physics, 43, 199
Hunter J. D., 2007, Computing in Science & Engineering, 9, 90
Kawasaki M., Kohri K., Moroi T., Takaesu Y., 2018, Phys. Rev. D, 97, 023502
Keith C., Hooper D., Blinov N., McDermott S. D., 2020, Phys. Rev. D, 102, 103512
Kim H., 2021, MNRAS, 504, 5475
Laha R., 2019, Phys. Rev. Lett., 123, 251101
Lucca M., Schöneberg N., Hooper D. C., Lesgourgues J., Chluba J., 2020, J. Cosmology Astropart. Phys., 2020, 026

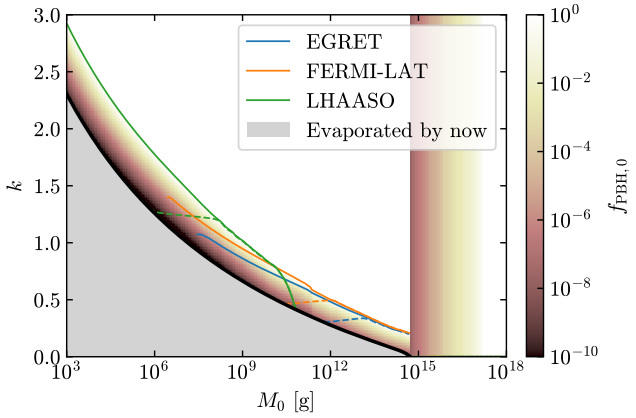


Figure A1. Combined constraints on $f_{\mathrm{PBH},0}(k, M_0)$ from galactic γ ray emission. The coloured lines show $f_{\mathrm{PBH}} = 1$ contours separately for selected observational data. The dashed lines are computed using only the primary photon emission.

Luo Y., Hanasoge S., Tromp J., Pretorius F., 2012, ApJ, 751, 16
McMillan P. J., 2011, MNRAS, 414, 2446
Michel M., Zell S., 2023, Fortschritte der Physik, 71, 2300163
Novikov I. D., Polnarev A. G., Starobinskii A. A., Zeldovich I. B., 1979, A&A, 80, 104
Pacheco J. A. d. F., Kiritsis E., Lucca M., Silk J., 2023, Phys. Rev. D, 107, 123525
Planck Collaboration et al., 2020, A&A, 641, A5
Poulin V., Lesgourgues J., Serpico P. D., 2017, J. Cosmology Astropart. Phys., 2017, 043
Ruiz-Lapuente P., The L.-S., Hartmann D. H., Ajello M., Canal R., Röpke F. K., Ohlmann S. T., Hillebrandt W., 2016, ApJ, 820, 142
Sjöstrand T., et al., 2015, Computer Physics Communications, 191, 159
Stöcker P., Krämer M., Lesgourgues J., Poulin V., 2018, J. Cosmology Astropart. Phys., 2018, 018
Strong A. W., 2011, in Giani S., Leroy C., Rancoita P. G., eds, Cosmic Rays for Particle and Astroparticle Physics. pp 473–481 (arXiv:1101.1381), doi:10.1142/9789814329033_0059
Strong A. W., Mattox J. R., 1996, A&A, 308, L21
Strong A. W., et al., 1994, A&A, 292, 82
Strong A. W., Bloemen H., Diehl R., Hermsen W., Schönfelder V., 1999, Astrophysical Letters and Communications, 39, 209
Strong A. W., Moskalenko I. V., Reimer O., 2004, ApJ, 613, 956
Thoss V. A., 2024, Data for the paper "Light Primordial Black Holes as Dark Matter due to the breakdown of Hawking Evaporation" (Valentin Thoss et. al. 2024), doi:10.5281/zenodo.10718218, https://doi.org/10.5281/zenodo.10718218
Zel'dovich Y. B., Novikov I. D., 1967, Soviet Ast., 10, 602

APPENDIX A: ADDITIONAL FIGURES

Here we provide additional plots which show the constraints on $f_{\mathrm{PBH},0}(k, M_0)$ for galactic γ ray emission in Figure A1 and for extragalactic γ ray emission in Figure A2.

APPENDIX B: CMB ANISOTROPIES

As explained in Section 2.5, to compute the constraints from CMB anisotropies we rescale the results from the full MCMC runs using Equation (13). In Fig. B1 we show that this method is able to reproduce the semiclassical constraints within one order of magnitude by

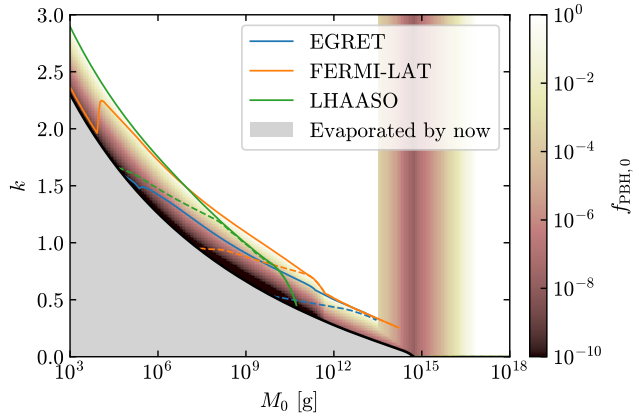


Figure A2. Combined constraints on $f_{\mathrm{PBH},0}(k, M_0)$ from the extragalactic γ ray background. The coloured lines show $f_{\mathrm{PBH}} = 1$ contours separately for selected observational data. Dashed lines are computed using only primary photon emission.

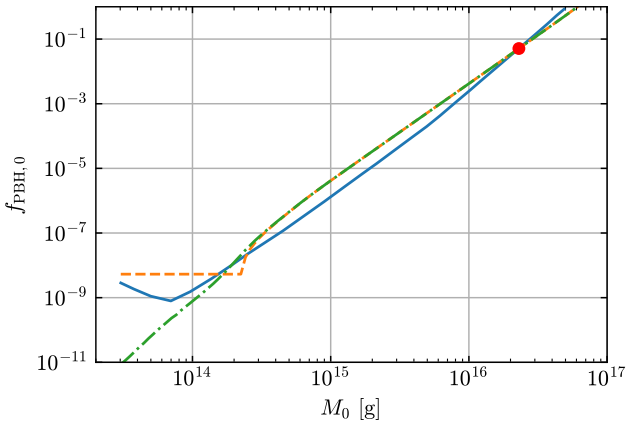


Figure B1. CMB constraints on the initial dark matter fraction of PBHs $f_{\mathrm{PBH},0}(M_0)$ in the semiclassical limit ($k = 0$). The solid blue line shows the result from Stöcker et al. (2018). The red dot shows the value $f(M_0 = 2.3\times10^{16}\mathrm{g}) = 5.1\times10^{-2}$ from which the rescaled constraints were obtained. The dashed orange line shows the constraints obtained through Equation (13), whereas the green dash-dotted line also takes into account the redshift-dependence of the integral.

rescaling a single value $f(M_0 = 2.3\times10^{16}\mathrm{g}) = 5.1\times10^{-2}$ obtained by Stöcker et al. (2018) over three orders of magnitude in mass. Note that in Fig. B1 we also show the resulting constraints when properly taking into acount the $(1+z)^3$ factor in Equation (12). There is no noticeable difference down to $M_0 \sim 3\times10^{14}\mathrm{g}$. The divergence at low masses is not relevant for us since it will in any case only affect those PBHs which will not survive until today - both in the SC limit as well as for $k > 0$.

To obtain constraints for $k > 0$ we compute 6 full MCMC runs for $M_0 \in [10^3, 10^5, 10^7, 10^9, 10^{11}, 10^{13}]$ g where k is chosen for each mass such that $f(k, M_0) = 10^{-2}$ when rescaled from $f(M_0 = 2.3\times10^{16}\mathrm{g}) = 5.1\times10^{-2}$ to the respective masses according to Equation (13) and taking into account the effect of the memory burden. The values of $f(k, M_0)$ obtained from the MCMC runs lie between 3.8×10^{-2} and 9.7×10^{-2} and thus up to one order of magnitude higher then the 10^{-2} expected from rescaling. These results are then used to obtain constraints for arbitrary values of k' and M_0' by rescaling from the two nearest values of M_0 and interpolating between the results. To test the validity of this approach we run a second set of MCMC runs with $M_0 \in [10^4, 10^6, 10^8, 10^{10}, 10^{12}]$ g and k such that our method would give $f(k, M_0) = 10^{-1}$. The resulting constraints on $f_{\mathrm{PBH},0}$ from the full computation are within a factor of two from our approximation which is more than sufficient. This translates to an error in the bound on $k(M_0, f = 1)$ of less than 0.01.

This paper has been typeset from a TEX/LATEX file prepared by the author.

arXiv:2402.17817v1 [hep-th] 27 Feb 2024

Gradient Properties of Perturbative Multiscalar RG Flows to Six Loops

William H. Pannell* and Andreas Stergiou†
Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom
(Dated: February 2024)

The gradient property of the renormalisation group (RG) flow of multiscalar theories is examined perturbatively in $d = 4$ and $d = 4 - \varepsilon$ dimensions. Such theories undergo RG flows in the space of quartic couplings λ^I. Starting at five loops, the relevant vector field that determines the physical RG flow is not the beta function traditionally computed in a minimal subtraction scheme in dimensional regularisation, but a suitable modification of it, the B function. It is found that up to five loops the B vector field is gradient, i.e. $B^I = G^{IJ}\partial A/\partial\lambda^J$ with A a scalar and G_{IJ} a rank-two symmetric tensor of the couplings. Up to five loops the beta function is also gradient, but it fails to be so at six loops. The conditions under which the B function (and hence the RG flow) is gradient at six loops are specified, but their verification rests on a separate six-loop computation that remains to be performed.

Introduction.—Dynamical systems, governed by differential equations of the form

$$\frac{d\vec{x}}{dt} = \vec{F}(\vec{x})\,, \tag{1}$$

appear in every corner of the natural world and are studied in depth in many branches of science and engineering. While general equations like (1) are often intractable with analytic methods, in some favourable cases features of such systems can be drawn out of the equations themselves, without having to resort to finding explicit solutions. One such case is that of gradient systems, where the function $\vec{F}(\vec{x})$ may be written in the form

$$\vec{F}(\vec{x}) = \vec{\nabla} f(\vec{x})\,. \tag{2}$$

Systems undergoing gradient flow have remarkable properties, and it is known from a theorem by Łojasiewicz [1] that at late times all solutions must flow to either a fixed point, where $\vec{F}(\vec{x})$ vanishes, or off to infinity.

In quantum field theory (QFT), the flow of field theories under a change of energy scale, governed by what is called the renormalisation group (RG), is one such dynamical system. As recognised by Wilson [2], the qualitative behaviour of such flows is of immense consequence to the physical system being modelled. In particular, a potential gradient property for the RG flow would have remarkable consequences, e.g. it would preclude periodic flow lines and severely limit the nature of potential endpoints of the RG flow.

In this Letter we will be concerned with the gradient property of RG flow in multiscalar QFTs with N real scalar fields in $d = 4$ and $d = 4 - \varepsilon$ dimensions. These QFTs are defined by the Lagrangian

$$\mathscr{L} = \tfrac{1}{2}\partial^\mu\phi_i\partial_\mu\phi_i + \tfrac{1}{4!}\lambda_{ijkl}\phi_i\phi_j\phi_k\phi_l\,, \tag{3}$$

where indices from the Latin alphabet take values from 1 to N [3]. Such theories are renormalisable in perturbation theory. This requires the bare symmetric tensor λ_{ijkl} to be substituted with a renormalised one, which we will also call λ_{ijkl}, which depends on the renormalisation scale μ:

$$\mu\frac{d\lambda_{ijkl}}{d\mu} = \beta_{ijkl}\,. \tag{4}$$

The quantity β_{ijkl}, called the beta function, is a function of λ and can be computed order by order in a weak-coupling expansion. Its first quantum contribution is of order λ^2 and takes the form

$$\beta_{ijkl} \supset \lambda_{ijmn}\lambda_{mnkl} + \lambda_{ikmn}\lambda_{mnjl} + \lambda_{ilmn}\lambda_{mnjk}\,, \tag{5}$$

which follows from the computation of a one-loop Feynman diagram in momentum space and we have rescaled λ to eliminate the factor of $1/16\pi^2$ that one encounters. This result can be extended to higher orders, and recently [4] used results of [5] to compute β_{ijkl} up to six loops, i.e. order λ^7. Due to the indices carried by λ there is a proliferation of contributions to β_{ijkl} at higher orders, with 2, 7, 23, 110, 571 distinct terms contributing at two, three, four, five and six loops, respectively.

One may think of β_{ijkl} as a vector field β^I in the space of couplings, where I denotes the collection of indices $(ijkl)$. A natural question then is if this vector field can be derived as a gradient of a scalar function. In that case the associated flow would be gradient, and the following equation would hold:

$$\beta^I = G^{IJ}\partial_J A\,, \quad \partial_I = \partial/\partial\lambda^I\,, \tag{6}$$

with G^{IJ} is a rank-two tensor and A a scalar. Both G^{IJ} and A are functions of λ. The metric G_{IJ} should be Riemannian, i.e.

(a) symmetric, $G_{IJ} = G_{JI}$, and

(b) positive-definite, $G \succ 0$.

In the present context of multiscalar theories, the validity of eq. (6) with a Riemannian metric was verified up to three loops some time ago in $d = 4$ and $d = 4-\varepsilon$ [6, 7], and this was more recently extended to four loops in $d = 4$ [8].

An important remark is that [9–11] have already established eq. (6) to any order in perturbation theory in $d = 4$, but the metric G_{IJ} in their treatment is substituted with a T_{IJ} which is not symmetric in general. However, the symmetry property of the metric has important consequences for potential fixed points of the flow [12]. In particular, it is necessary so that around any fixed point local coordinates can be chosen in which the metric is flat to leading order. In that case, $\beta_I = \partial_I A$ at leading order and the matrix $\partial_I \beta_J = \partial_I \partial_J A$ is symmetric at the fixed point. Consequently, its eigenvalues, which determine the scaling dimensions of operators quartic in ϕ, are real at real fixed points and unitarity is maintained.

Beyond its potential role for unitarity at fixed points of the RG flow, an equation like (6) encapsulates important features of the RG flow itself and has consequences for the existence of possible exotic endpoints one might seek, e.g. limit cycles and ergodic behaviour. In short, if eq. (6) is valid, then such exotic behaviour is not possible and the quantity A serves as a monotonically decreasing quantity along any RG flow to the infrared, since if G is positive-definite then

$$\mu \frac{d}{d\mu} A = \beta^I \partial_I A = G_{IJ} \beta^I \beta^J > 0 \tag{7}$$

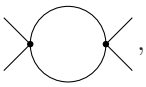
for non-zero β^I. Of course a possible antisymmetric contribution to G_{IJ} does not invalidate (7), so the results of [9–11] are enough to establish that limit cycles and ergodic behaviour are not possible endpoints of RG flows in perturbation theory in $d = 4$.

In this work we will consider the gradient property of the RG flow in $d = 4$ and we will also describe the extension to $d = 4 - \varepsilon$. We will analyse the RG flow to six loops, to test the validity of (6) in the most complicated setting available. As we will see, hundreds of non-trivial constraints among beta function coefficients need to be satisfied for the RG flow to be gradient at six loops. The beta function vector field satisfies the required constraints up to five loops, but fails to do so at six loops. However, a suitable modification of it, which gives the proper vector field that describes physical RG flows, may satisfy them but this rests on some hitherto undetermined five- and six-loop results. These can be obtained within the standard paradigm of renormalised perturbation theory, but we do not attempt the associated computations in this work.

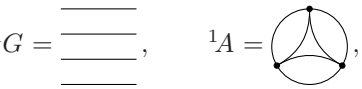
Summary of results up to four loops in $d = 4$.— Starting with (5) at one loop, it is trivial to verify (6) with

$${}^1G_{ijkl;mnpq} = \delta_{ijkl;mnpq}\,, \quad {}^1A = \lambda_{ijmn}\lambda_{mnpq}\lambda_{pqkl}\,. \tag{8}$$

Our notation ${}^L X$ captures the loop order L that the quantity X is first relevant in eq. (6). In a diagrammatic notation, we may write



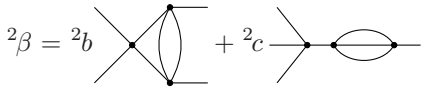
(9)



(10)

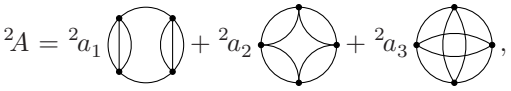
where proper index symmetrisations where necessary are understood.

At two loops



(11)

and



(12)



(13)

Demanding $\beta = G\partial A$ now gives

$${}^2a_1 - {}^2c = 0\,, \tag{14}$$

$${}^2a_2 + \tfrac{1}{4}\,{}^1b\,{}^2g = 0\,, \tag{15}$$

$${}^1b\,{}^2g + 2\,{}^2a_3 - 3\,{}^2b = 0\,, \tag{16}$$

where ${}^1b$ is the coefficient in the one-loop beta function (${}^1b = 1$ from (9)). These can be solved for ${}^1a_3, {}^2a_3, {}^3a_3$ with ${}^2g = 0$ for any ${}^1b, {}^2b, {}^2c$, and so there are no constraints among beta function coefficients and the metric can remain globally flat up to this order (although in general ${}^2g$ is simply an undetermined coefficient at this order).

At three loops the beta function again has the gradient property regardless of the specific values of coefficients that appear in it, but the metric cannot be chosen to be globally flat any more [7]. That the metric can be globally flat at two loops but not at three loops can be seen by a cursory examination of the diagrams in (12) and Fig. 1. Each of the diagrams in (12) has identical vertices, so that there will be a one-to-one correspondence between coefficients in ${}^2\beta_I$ and in $\partial_I\,{}^2A$. A cursory examination of Fig. 1 shows that at three-loops the complexity of the vacuum diagrams increases, erasing this property as a general feature. Further, at four loops the gradient property of the beta function vector field is not totally general, and rests on four constraints among beta function coefficients [8]. These are indeed satisfied, so the RG flow remains gradient in $d = 4$ to four loops.

Diagram Generation.—To consider higher orders in the loop expansion, we need to find efficient ways to generate

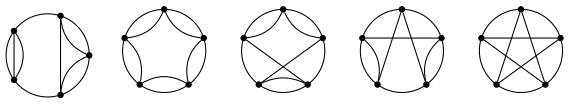


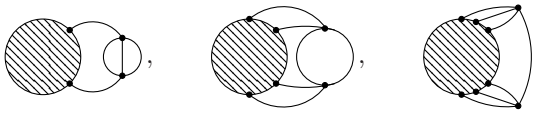
FIG. 1. The five vacuum diagrams contributing to ${}^3A$.

the terms contributing to A and G_{IJ}. To that end, we found it useful to map strings of λ_{ijkl} tensors to graphs analogous to Feynman diagrams. Determining whether or not two tensor structures are equivalent up to a permutation of their free indices then becomes a simple application of known algorithms for distinguishing between non-isomorphic graphs. These graph algorithms were handled using the `IGraph/M` package [13] for *Mathematica*, while the tensors themselves were implemented using `xAct` with `xTras` [14].

As A is a scalar function, all of the indices on its constituent λ's must be contracted, so that the resulting graphs will be the vacuum diagrams in the QFT. As the n-th loop order terms in the beta function will be order λ^{n+1}, the corresponding n-loop terms in A will look like

$${}^nA \supset \sum_{\text{Bubbles}} {}^na_i$$  , (17)

where the sum is over connected bubble diagrams containing $n+2$ vertices and the ${}^na_i$ are arbitrary constants. The beta function will only include contributions from connected graphs, so that only connected, one-vertex-irreducible (1VI) graphs can contribute to A with non-zero coefficients. We can then decompose the diagrams in A as

 . (18)

The essential information in these diagrams will be contained in an order λ^n tensor with two, four or six free indices, which we contract with either $\lambda_{iklm}\lambda_{jklm}$, $\lambda_{ijmn}\lambda_{klmn}$, or $\lambda_{ijkp}\lambda_{lmnp}$ respectively.

Similarly, to construct terms in the metric at a given order in λ one can notice that they will contain at most three factors of δ_{ij}, so that the λ-part of the term will have two, four, six or eight external indices. One could also consider constructing the terms in the metric by removing two vertices from a vacuum diagram, so that the connectedness of terms in the beta function requires that we only consider metric terms such that $G^{IJ}\lambda_I\lambda_J$ is a connected, 1VI graph.

We see that constructing A and G_{IJ} simply requires enumerating all possible combinations of n λ_{ijkl} tensors with at most eight free indices. This enumeration is done iteratively, beginning with only λ_{ijkl} at $\mathrm{O}(\lambda^1)$ and successively attaching new factors of λ_{ijkl} to build up the higher order structures. Depending on the number of contracted indices, attaching a new vertex will change the number of free indices by an integer $m \in \{-4, -2, 0, 2\}$, leading to the straightforward recursion relations. In order to construct the gradient flow equation up to six loops, we will need the vacuum diagrams to order λ^8 and the terms in the metric up to order λ^6. The derivative $\partial A/\partial\lambda_{ijkl}$ corresponds to summing over all of the different ways of removing a vertex, so that the number of equations in (6) will be equal to the number of these distinct structures. In Table I we list the numbers of the various objects appearing in this paper through six loops.

	1-loop	2-loop	3-loop	4-loop	5-loop	6-loop
A	1	3	5	17	42	177
G_{IJ}	1	1	7	18	97	453
Equations	1	3	10	36	164	819

TABLE I. Whenever there is overlap, these numbers agree with those listed in [15, 16].

Results at five and six loops.—At five loops we find 164 equations for 145 unknowns, including those remaining unfixed in A and G_{IJ} at lower orders. One immediately sees that the number of equations has grown faster than the number of unknowns, so that this order even by naive counting we expect that there must exist constraint equations for the coefficients of the various contributions to the beta function. The unknowns $\{{}^na_m, {}^ng_m\}$ do not all appear independently, and we can only solve for 127 of the coefficients, with 18 remaining as free parameters. This leaves 37 of the five-loop equations as constraints on the beta function. We find that these constraints are indeed satisfied by the $\overline{\text{MS}}$ values of the beta function coefficients $\{{}^nb_m, {}^nc_m\}$ (where $n \leq 5$).

Many of the constraints are complicated and entirely mysterious. Some, however, may be derived readily from simple arguments. Consider, for instance, the constraint

$${}^5b_{67} = {}^5b_{69} \left(= 36\zeta_3^2\right) , \tag{19}$$

involving the diagrams [17]

$${}^5\beta \supset {}^5b_{67}$$  $+ {}^5b_{69}$  . (20)

As these diagrams are primitive, gradient flow can only hold if they are absorbed by the same term in ${}^5A$, namely

$${}^5A \supset {}^5a_{42}$$  . (21)

The constraint (19) then just expresses the fact that this part of ${}^5\beta$ must be derivative.

At six loops we find 819 equations for 648 unknowns. Solving these equations for A and G_{IJ}, we find a total of 234 constraints at this order [18]. Of these, five are not

satisfied by the coefficients in $\overline{\text{MS}}$, indicating that β_{ijkl} is *not* gradient for a generic scalar field theory at six loops.

However, beginning at five loops there exists an essential ambiguity in the definition of β_{ijkl}, which prevents it from being the physical object associated with RG flow [9, 19]. This ambiguity arises from a mismatch between the object used to eliminate poles in propagator diagrams, Z_{ij}, and the quantity used to renormalise the bare field operator, $Z_{ij}^{1/2}$. While Z_{ij} is used for propagator diagrams, it is $Z_{ij}^{1/2}$ which is used to renormalise the two-point function. This may be calculated from Z_{ij} using the matrix equation

$$Z = (Z^{1/2})^T Z^{1/2} \,. \tag{22}$$

The crucial observation is that this does not determine $Z^{1/2}$ uniquely, due to the possible $O(N)$ rotation

$$Z_{ij}^{1/2} \to R_{ik} Z_{kj}^{1/2} \,, \qquad R \in O(N) \,, \tag{23}$$

which does not alter the pole structure of the propagator. If R contains $1/\varepsilon$ poles itself, the beta function can be affected, however. To see this, we will work with an infinitesimal rotation, and expand the Lie algebra element as a series in $1/\varepsilon$,

$$R_{ij} = \delta_{ij} + \sum_n \frac{O_{ij}^{(n)}}{\varepsilon^n} \,, \tag{24}$$

where we have used the notation of [19] and denoted the Lie algebra element by O_{ij}. One can then see that this will alter the poles in $Z_{ij}^{1/2}$, specifically sending the $1/\varepsilon$ pole to

$$Z_{ij}^{1/2,\,(1)} \to Z_{ij}^{1/2,\,(1)} + O_{ij}^{(1)} \,. \tag{25}$$

If $O_{ij}^{(1)}$ has dependence on λ_{ijkl} this extra term will contribute to both the anomalous dimension matrix γ_{ij} and thus also to the beta function, shifting them to become

$$\gamma_{ij} \to \gamma_{ij} - \omega_{ij} \qquad \beta_{ijkl} \to \beta_{ijkl} + (\omega_{im}\lambda_{mjkl} + \text{perms.}) \,, \tag{26}$$

where $\omega_{ij} = -\lambda_I \partial_I O_{ij}^{(1)}$. This ambiguity prevents the beta function from being physically meaningful as long as we can write down such a correction $O^{(1)} = O^{(1)}(\lambda)$. As $O_{ij}^{(1)} = O_{[ij]}^{(1)}$, these come from asymmetric propagator diagrams which appear first at five loops with the four shown in Fig. 2. Thus, beginning at five loops there is an essential ambiguity in β_{ijkl} which prevents it from being the physical object associated with RG flow.

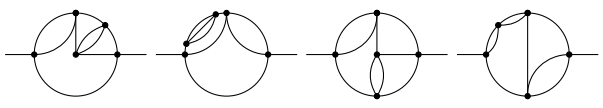


FIG. 2. The four asymmetrical propagator corrections contributing to S_{ij} at five-loops.

To resolve this, we must turn to the trace of the stress-tensor. As shown by Jack and Osborn in [9], $T^\mu{}_\mu$ will not be proportional $\beta_{ijkl}\hat{\mathcal{O}}_{ijkl}$, but rather $B_{ijkl}\hat{\mathcal{O}}_{ijkl}$, where the B-function is a modification of the beta function:

$$B_{ijkl} = \beta_{ijkl} - (S_{im}\lambda_{mjkl} + \text{perms.}) \,, \qquad S_{ij} = S_{[ij]} \,. \tag{27}$$

The only features of the additional term S_{ij}, which can be calculated using counterterms arising from coordinate-dependent couplings, which are needed for this work are that under (23) it will transform as

$$S_{ij} \to S_{ij} + \omega_{ij} \,, \tag{28}$$

and that it will have some expansion

$$S_{ij} = \sum_{n\geq 5} \sum_m {}^n s_m \, {}^n\mathcal{S}_{m,\,ij} \tag{29}$$

where ${}^n\mathcal{S}_{m,\,ij}$ are anti-symmetrised propagator diagrams, e.g. those contained in Fig. 2 for five loops. While the leading coefficients ${}^n s_m$ have been calculated in the context of scalar-fermion theories in $d = 4$ [19], where the first contributions to S appear at three loops, no such computation has been performed for multiscalar theories in $d = 4$, and we will leave them unfixed throughout our analysis. From the first property, one immediately sees that, unlike the beta function, B_{ijkl} is invariant under (23) and is thus unambiguous. Beginning at five loops we are then forced to identify the B-function as the proper vector field associated with RG flow.

Practically, the distinction between B_{ijkl} and β_{ijkl} is an antisymmetric shift of coefficients associated with asymmetric propagator diagrams. There are four such diagrams at five loops and 19 at six loops, leading to a total of 23 unknowns which may appear in the constraint equations. For five loops, we define the ${}^5 s_m$ by using the ordering shown in Fig. 2 for m, with the left leg getting the index i. Implementing this shift into the existing constraints for ${}^n b_m$ and ${}^n c_m$, we find that at five loops the ${}^5 c_m$ coefficients associated with Fig. 2 appear only symmetrically, so that the factors of ${}^5 s_m$ cancel. The gradient property of β_{ijkl} thus implies the gradient property of B_{ijkl} at five loops. At six loops we find that the five unsatisfied constraints become two constraints on ${}^5 s_m$. However, four of the previously satisfied constraints also receive contributions from ${}^n s_m$, leading to a total of four constraint equations:

$$\begin{aligned}
&{}^5 s_2 = \frac{259}{4608} \,, \qquad 192\,{}^5 s_3 + 384\,{}^5 s_4 = 31 - 36\zeta_3 \,, \\
&{}^5 s_1 - 12\,{}^5 s_2 + {}^5 s_4 = 12\,{}^6 s_8 - 6\,{}^6 s_9 - 6\,{}^6 s_{16} \,, \\
&4\,{}^5 s_1 - 2\,{}^5 s_3 = -2\,{}^6 s_3 + {}^6 s_6 + 4\,{}^6 s_{10} - 2\,{}^6 s_{15} + {}^6 s_{17} \,.
\end{aligned} \tag{30}$$

As these coefficients have not been calculated, these equations amount to a prediction of their values, an explicit

one in the case of ${}^5s_2$, if the B-function is to be gradient. Given that B_{ijkl} already satisfies 266 constraints through six loops, evidence suggests that gradient flow is an intrinsic property of multiscalar RG flows, and thus that an explicit calculation of S_{ij} must uphold (30).

Scheme Independence of Results.—Up to this point the discussion has been entirely confined to the $\overline{\text{MS}}$ renormalisation scheme. To ensure that our results are physically meaningful statements, we must demonstrate that the constraints are satisfied in all renormalisation schemes. While we considered arbitrary coefficients in deriving the constraints on the beta function, we had implicitly demanded the scheme-dependent statement that the coefficients of tensor structures not appearing in $\overline{\text{MS}}$ were zero. While it is possible to prove scheme-independence by recasting the constraints in terms of explicitly scheme-independent quantities as was done for the four-loop constraints in [8], we will take a more brute-force approach.

A change in renormalisation scheme can be realised as a generic shift in the interaction tensor,

$$\lambda_I \to \lambda'_I = \lambda_I + \sum_{n=1}^{\infty} \sum_m {}^n r_m \, {}^n T_{m,\,I}(\lambda)\,, \tag{31}$$

where the coefficients ${}^n r_m$ are arbitrary, and m indexes a sum over distinct tensor structures ${}^n T_m$ of order λ^n. The number of these structures will be identical with the number of equations listed in Table I. In order to determine how this shift will alter the coefficients appearing in the beta function, we must notice that

$$\beta'_I(\lambda') = \mu \frac{d\lambda'_I}{d\mu} = \beta_I(\lambda) + \sum_{n=1}^{\infty} \sum_m \sum_{\hat{T}\subset T} {}^n r_m \, {}^n \hat{T}_{m,\,IJ}\, \beta_J(\lambda)\,, \tag{32}$$

where $\hat{T}$ has a vertex removed compared to T. We also notice that $\beta'_I(\lambda')$ has an expansion in its own right as

$$\beta'_I(\lambda') = \sum_{n=1}^{\infty} \sum_m {}^n \mathfrak{b}_m \, {}^n T_{m,\,I}(\lambda')\,. \tag{33}$$

We then notice that we can isolate the new coefficients ${}^n\mathfrak{b}_m$ by expanding the λ'_I tensors in (33) using (31) and then matching with the right-hand side of (32). We perform this matching with the same computational method used to find the solution to the gradient flow equations, and find the resulting generic scheme change up to six loops.

Solving the gradient flow equations with $\beta'_I(\lambda')$, we find again the 275 constraints through six loops, now in terms of the ${}^n\mathfrak{b}_m$. At four and five-loops, substituting in the explicit expressions ${}^n\mathfrak{b}_m = {}^n\mathfrak{b}_m(r,b,c)$ yields constraints that are independent of the scheme-change parameters ${}^n r_m$, and reduce to those found in $\overline{\text{MS}}$. At six loops the constraints contain explicit factors of ${}^n r_m$, which naively seem to spoil scheme-invariance. However, these terms appear multiplied with combinations of ${}^n b$ and ${}^n c$ beta-function coefficients that vanish upon taking their $\overline{\text{MS}}$ values. These factors will then not spoil the conclusion that satisfying the $\overline{\text{MS}}$ constraints implies that the constraints are satisfied in a general scheme. The gradient property of the beta function in the $\overline{\text{MS}}$ scheme thus ensures the gradient property of the beta function in all schemes.

Extension to $d = 4 - \varepsilon$.—Working slightly away from four dimensions in $d = 4 - \varepsilon$, the coupling λ_{ijkl} becomes weakly relevant and β_{ijkl} receives a contribution from its classical scaling dimension in the form of a linear term,

$$\beta_{ijkl} \supset -\varepsilon \lambda_{ijkl}\,. \tag{34}$$

This term can only be absorbed by introducing a new term in A,

$${}^0A = -\frac{\varepsilon}{2}$$  . (35)

At n loops this contribution to A will bring in terms in the metric which in four dimensions first appeared at $n{+}1$ loops, namely

$$-\varepsilon \sum_m {}^{n+1} g_m \, {}^{n+1} G_m{}^{IJ} \lambda_J\,. \tag{36}$$

Extending our four-dimensional solution to $d = 4 - \varepsilon$ without the introduction of non-perturbative $1/\varepsilon$ poles [20] requires that we absorb this new term into ε corrections to a and g:

$${}^n a_m \to {}^n a_m + \varepsilon \, {}^n d_m(\varepsilon)\,, \qquad {}^n g_m \to {}^n g_m + \varepsilon \, {}^n h_m(\varepsilon)\,. \tag{37}$$

Here ${}^n d_m(\varepsilon)$ and ${}^n h_m(\varepsilon)$ are arbitrary polynomials in ε. These ε corrections will alter the gradient flow equation to read, schematically,

$$\varepsilon\left(\lambda + H(\varepsilon)\partial A + G\partial D(\varepsilon)\right) + \varepsilon^2 H(\varepsilon)\partial D(\varepsilon) = 0\,, \tag{38}$$

where we have discarded the $\mathrm{O}(\varepsilon^0)$ terms already solved for in terms of a and g. At $\mathrm{O}(\lambda^6)$ and below the values of G and A on the will largely be fixed by the four-dimensional solution, but the $\mathrm{O}(\lambda^7)$ equation will contain the novel metric term ${}^7G$. As we are truncating the expansion, and have already implicitly discarded all terms in (6) of $\mathrm{O}(\lambda^8)$ and above, we are free to include the coefficients ${}^7 g_m$ in the solution at this order. While treating these as free parameters here may not allow for a consistent solution at seven loops, it will be sufficient to draw conclusions through six loops. In general, solutions to (38) will include $1/\varepsilon$ poles, which both obstruct the smooth $\varepsilon \to 0$ limit and invalidate the existing $\mathrm{O}(\varepsilon^0)$ solution. Demanding that these poles vanish provides additional equations that $H(\varepsilon)$ and $D(\varepsilon)$ must satisfy.

We solve for $\{{}^n d_m,\, {}^n h_m\}$, and the freedom in G and A unfixed by the four dimensional solution using the same

techniques as before. When the dust settles, we find four constraints on the coefficients in the beta function at five loops, and 32 constraints at six loops. The additional freedom in ${}^7G$ means that at six loops the constraints only appear from the $\mathrm{O}(\lambda^6)$ equations. We note that these constraints are independent of those found in $d=4$, and crucially are all satisfied in $\overline{\mathrm{MS}}$. Importantly, the coefficients ${}^n s_m$ appearing in S_{ij} identically cancel in all of these constraints, so that there is no obstruction to B_{ijkl} satisfying them, still subject to (30).

To establish the physical relevance of the constraints in $d=4-\varepsilon$, we once again examine them under a generic scheme change, replacing the ${}^n\mathfrak{b}_m$ in (33) with

$$ {}^n\mathfrak{b}_m \to {}^n\mathfrak{b}_m + \varepsilon\, {}^n\mathfrak{e}_m \,. \tag{39}$$

Dividing (32) into terms of $\mathrm{O}(\varepsilon^0)$ and terms of $\mathrm{O}(\varepsilon)$, we see that the ${}^n\mathfrak{b}_m$ will be the same as in four dimensions. The $\mathrm{O}(\varepsilon)$ terms will be

$$ \sum_{n=1}^{\infty}\sum_m {}^n\mathfrak{e}_m\, {}^nT_{m,I}(\lambda') = -\sum_{n=1}^{\infty}\sum_m (n-1)\, {}^nr_m\, {}^nT_{m,I}(\lambda)\,, \tag{40}$$

where we have used the fact that ${}^0\mathfrak{e}=-1$ and (31). We once again solve these by expanding the factors of λ' on the left-hand side and matching coefficients to obtain ${}^n\mathfrak{e}_m = {}^n\mathfrak{e}_m(r)$. We then re-solve the equations, using now the four dimensional solution in terms of the coefficients $\mathfrak{b}$ in a general scheme change, noting now that contributions from $\mathfrak{e}$ will produce ε terms in (38) to all orders in λ. We again find 36 constraints through six loops, which reduce to the 36 we had found previously. Thus, the vanishing of the constraints in $\overline{\mathrm{MS}}$ implies the vanishing of the constraints in all schemes, so that the extension of the four dimensional solution to $d=4-\varepsilon$ is physically meaningful.

Conclusion.—While the conclusion that the RG flow is gradient to six loops in $d=4$ and $d=4-\varepsilon$ if the conditions in (30) are satisfied is scheme-independent and thus physical, the specific A and G_{IJ} that we determine along the way are scheme-dependent. This means that they cannot be given a direct physical meaning along an RG flow. It would be interesting to understand the extent to which A and G_{IJ} may contain physical information, along and at a fixed point of the RG flow.

There are other cases where the gradient property of the RG flow has been examined in perturbation theory, e.g. [21–24]. In all these cases the RG flow is gradient. In cases with defects, the next-to-leading order result for the line-defect beta function in scalar-fermion theories in $d=4-\varepsilon$ [25] is a gradient vector field. We expect that a direct examination of the diagrammatic expansion for the gradient flow equation, analogous to the one presented here, will provide a perturbative proof of the gradient property in these line defect systems, given the relatively simple structure of the terms involved.

In summary, the evidence for the gradient property of RG flow from perturbation theory is overwhelming. It would be interesting to look beyond perturbation theory and seek connections with existing literature, e.g. [26], in specific models. Let us finally emphasise that the gradient property of RG flow cannot be viewed as a fundamental physical requirement at this stage, so it is important to explore how it may arise from deeper physical principles.

We would like to thank Ian Jack and especially Hugh Osborn for useful discussions and communications. We would also like to thank Slava Rychkov for discussions that inspired this work during "Bootstrap 2023" at ICTP-SAIFR in July 2023, and acknowledge support received through FAPESP grant 2021/14335-0. AS is supported by the Royal Society under grant URF\R1\211417 and by STFC under grant ST/X000753/1.

* william.pannell@kcl.ac.uk
† andreas.stergiou@kcl.ac.uk

[1] S. Lojasiewicz, Sur les trajectoires du gradient d'une fonction analytique, Seminari di geometria **1983**, 115 (1982).
[2] K. G. Wilson, Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture, Phys. Rev. B **4**, 3174 (1971).
[3] We follow the Einstein summation convention for repeated indices. We have assumed a $\mathbb{Z}_2$ symmetry generated by $\phi \to -\phi$ and neglected the allowed term $\frac{1}{2}\kappa_{ij}\phi_i\phi_j$, which plays no role in our discussion.
[4] A. Bednyakov and A. Pikelner, Six-loop beta functions in general scalar theory, JHEP **04**, 233, arXiv:2102.12832 [hep-ph].
[5] M. V. Kompaniets and E. Panzer, Minimally subtracted six loop renormalization of $O(n)$-symmetric ϕ^4 theory and critical exponents, Phys. Rev. D **96**, 036016 (2017), arXiv:1705.06483 [hep-th].
[6] D. J. Wallace and R. K. P. Zia, Gradient Flow and the Renormalization Group, Phys. Lett. A **48**, 325 (1974).
[7] D. J. Wallace and R. K. P. Zia, Gradient Properties of the Renormalization Group Equations in Multicomponent Systems, Annals Phys. **92**, 142 (1975).
[8] I. Jack and C. Poole, Scheme invariants in ϕ^4 theory in four dimensions, Phys. Rev. D **98**, 065011 (2018), arXiv:1806.08598 [hep-th].
[9] I. Jack and H. Osborn, Analogs for the c Theorem for Four-dimensional Renormalizable Field Theories, Nucl. Phys. B **343**, 647 (1990).
[10] H. Osborn, Weyl consistency conditions and a local renormalization group equation for general renormalizable field theories, Nucl. Phys. B **363**, 486 (1991).
[11] I. Jack and H. Osborn, Constraints on RG Flow for Four Dimensional Quantum Field Theories, Nucl. Phys. B **883**, 425 (2014), arXiv:1312.0428 [hep-th].
[12] Such fixed points lie beyond the computational ability of perturbation theory in $d=4$, but become available in $d=4-\varepsilon$ with ε small.
[13] S. Horvát, J. Podkalicki, G. Csárdi, T. Nepusz,

V. Traag, F. Zanini, and D. Noom, IGraph/M: graph theory and network analysis for Mathematica, Journal of Open Source Software **8**, 4899 (2023).

[14] T. Nutma, xTras: A field-theory inspired xAct package for Mathematica, Computer Physics Communications **185**, 1719 (2014).

[15] I. R. Klebanov and G. Tarnopolsky, On Large N Limit of Symmetric Traceless Tensor Models, JHEP **10**, 037, arXiv:1706.00839 [hep-th].

[16] H. Kleinert and V. Schulte-Frohlinde, *Critical Poperties of ϕ^4 Theories* (World Scientific, Singapore, 2001).

[17] Note that both diagrams in (20) are primitive (i.e. have no subdivergences) and thus the value of the two coefficients given in (19) is scheme-independent.

[18] When considering the constraints in a general scheme we find 237. However, three of these are identically zero in $\overline{\text{MS}}$ as they involve only graphs which are either one-vertex reducible or have non-overlapping divergences.

[19] J.-F. Fortin, B. Grinstein, and A. Stergiou, Limit Cycles and Conformal Invariance, JHEP **01**, 184, arXiv:1208.3674 [hep-th].

[20] One notices that a solution always exists at these orders if we permit $1/\varepsilon$ poles to appear. In fact this is fully general, and one can prove holds to all loop orders as $G^{IJ}\lambda_J$ will be a full-rank in the coefficients g. While the coupling λ_{ijkl} will be $\mathrm{O}(\varepsilon)$ at the fixed points, there is nothing preventing a generic flow from moving beyond this region.

[21] I. Jack and C. Poole, The a-function for gauge theories, JHEP **01**, 138, arXiv:1411.1301 [hep-th].

[22] I. Jack, D. R. T. Jones, and C. Poole, Gradient flows in three dimensions, JHEP **09**, 061, arXiv:1505.05400 [hep-th].

[23] J. A. Gracey, I. Jack, and C. Poole, The a-function in six dimensions, JHEP **01**, 174, arXiv:1507.02174 [hep-th].

[24] I. Jack and C. Poole, α-function in three dimensions: Beyond the leading order, Phys. Rev. D **95**, 025010 (2017), arXiv:1607.00236 [hep-th].

[25] W. H. Pannell and A. Stergiou, Line defect RG flows in the ε expansion, JHEP **06**, 186, arXiv:2302.14069 [hep-th].

[26] J. Cotler and S. Rezchikov, Renormalization group flow as optimal transport, Phys. Rev. D **108**, 025003 (2023), arXiv:2202.11737 [hep-th].

arXiv:2402.17789v1 [physics.class-ph] 26 Feb 2024

Reference frame dependence of the periodically oscillating Coulomb field in the Proca theory

Bogdan Damski
Jagiellonian University, Faculty of Physics, Astronomy and Applied Computer Science, Łojasiewicza 11, 30-348 Kraków, Poland

The Proca theory of the real massive vector field admits non-equilibrium solutions, where the asymptotic dynamics of the electric field is dominated by the periodically oscillating Coulomb component. We discuss how such field configurations are seen in different reference frames, where we find an intriguing spatial pattern of the vector field and the electromagnetic field associated with it. Our studies are carried out in the framework of the classical Proca theory.

I. INTRODUCTION

We consider the simplest version of the Proca theory, which is defined by the following Lagrangian density

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2 + \frac{m^2}{2}(V_\mu)^2, \tag{1}$$

where V is the vector field and $m > 0$ is the mass of the vector boson (see Appendix A for our conventions). This theory is often used to illustrate various field theoretic issues and it describes some properties of massive vector bosons such as ρ and ω mesons [1–3]. Moreover, the Proca theory is considered as an extension of Maxwell electrodynamics [4–6], the one taking into account the possibility that the photon might be a massive particle.

We recently studied the quantum version of the Proca theory in [7], where the periodically oscillating Coulomb field was introduced for the first time. In this work, we focus entirely on the classical version of the Proca theory, where the reference frame dependence of such an unusual field can be most transparently discussed.

To proceed, we introduce the primed reference frame, where the vector field is denoted as V' and the spacetime coordinates are $(t', \boldsymbol{r}')$, where $\boldsymbol{r}' = (x', y', z')$. Then, we consider the following field configuration

$$V'^0(t', \boldsymbol{r}') = -\frac{q}{m^2}\int \frac{d^3k}{(2\pi)^3} f(\omega_k) \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}')\cos(\varepsilon_k t'), \tag{2a}$$

$$\boldsymbol{V}'(t', \boldsymbol{r}') = \frac{\mathrm{i}q}{m^2}\int \frac{d^3k}{(2\pi)^3} f(\omega_k)\frac{\varepsilon_k}{\omega_k^2}\boldsymbol{k} \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}')\sin(\varepsilon_k t'), \tag{2b}$$

where q is a constant non-zero parameter, $\omega_k = |\boldsymbol{k}|$, $\varepsilon_k = \sqrt{m^2+\omega_k^2}$, and real-valued $f(\omega_k)$ weights the Fourier modes in the studied field configuration ($f(0) \neq 0$ is assumed because low frequency modes are of interest in this work). We assume that V' represents the finite-energy field configuration and that its first and second order derivatives over space-time coordinates can be obtained by taking such derivatives under the integral symbol, which facilitates the computation of the electromagnetic field and the verification of Proca field equations. These properties can be ensured by various functions $f(\omega_k)$, e.g. such as those given by (9).

Taking into account the above remarks, one may easily verify that (2) satisfies the Proca field equations in the primed reference frame

$$\Box' V' = -m^2 V', \ \partial'\cdot V' = 0, \tag{3}$$

where $\Box' = \partial'\cdot\partial' = \partial_{t'}^2 - \partial_{x'}^2 - \partial_{y'}^2 - \partial_{z'}^2$. Moreover, via $\boldsymbol{E}' = -\partial_{t'}\boldsymbol{V}' - \boldsymbol{\nabla}' V'^0$ and $\boldsymbol{B}' = \boldsymbol{\nabla}'\times\boldsymbol{V}'$, where $\boldsymbol{\nabla}' = (\partial_{x'}, \partial_{y'}, \partial_{z'})$, one may also verify the following expressions for the electromagnetic field

$$\boldsymbol{E}'(t', \boldsymbol{r}') = -\mathrm{i}q\int \frac{d^3k}{(2\pi)^3} f(\omega_k)\frac{\boldsymbol{k}}{\omega_k^2}\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}')\cos(\varepsilon_k t'), \tag{4}$$

$$\boldsymbol{B}'(t', \boldsymbol{r}') = \boldsymbol{0}. \tag{5}$$

The discussed field configuration is described by spherically symmetric $V'^0(t', \boldsymbol{r}')$ as well as $\boldsymbol{V}'(t', \boldsymbol{r}') \propto \boldsymbol{r}'$ and $\boldsymbol{E}'(t', \boldsymbol{r}') \propto \boldsymbol{r}'$, where the proportionality factors are a function of t' and $r' = |\boldsymbol{r}'|$ [8]. Thereby, our field configuration is centered around $\boldsymbol{r}' = \boldsymbol{0}$ for all times and as such it is at rest in the primed reference frame. As we deal here with the non-equilibrium solution of (3), the term "at rest" refers to the fact that the analyzed field configuration does not move as a whole in any particular direction. In addition to that, we note that the energy of the studied field configuration is given by

$$\frac{q^2}{4\pi^2 m^2}\int_0^\infty d\omega_k f^2(\omega_k)\varepsilon_k^2. \tag{6}$$

This result has been obtained from the standard expression [1]

$$\frac{1}{2}\int d^3r' \left(|\boldsymbol{E}'(t', \boldsymbol{r}')|^2 + |\boldsymbol{B}'(t', \boldsymbol{r}')|^2 + m^2|\boldsymbol{V}'(t', \boldsymbol{r}')|^2 + m^2[V'^0(t', \boldsymbol{r}')]^2\right), \tag{7}$$

where the presence of the vector field does not cause any problems because such a quantity is observable in the Proca theory (e.g. [9] reports an attempt of the experimental measurement of the vector field of the Proca theory; see the discussion by the end of this section).

Our interest in (2) comes from the fact that the following quantity exhibits intriguing dynamics

$$Q'(t', \mathbb{R}^3) = \int d^3r' \boldsymbol{\nabla}'\cdot\boldsymbol{E}'(t', \boldsymbol{r}') = q\cos(mt'), \tag{8}$$

which has been obtained by employing (4) and the fact that (without loss of generality) we assume $f(0) = 1$ in this work. Adopting the nomenclature introduced in [4, 6], one may call (8) a spurious charge or a pseudocharge. We will call it simply the charge following the nomenclature employed in [7]. One should, however, keep in mind that we talk here about the charge associated with the internal Proca 4-current $-m^2V' = (\boldsymbol{\nabla}' \cdot \boldsymbol{E}', -m^2\boldsymbol{V}')$ (not the charge associated with some external 4-current describing electrically charged particles). We mention in passing that (8) illustrates the phenomenon of periodic charge oscillations in the classical Proca theory. Such a phenomenon was recently discussed in the quantum Proca theory [7], where it follows from the harmonic oscillator equation satisfied by the charge operator [10, 11].

The importance of (8) lies in the fact that such a result, via the Gauss' law argument, shows that vector field (2) describes some sort of the Coulomb field in the limit of $r' \to \infty$: the one exhibiting both the inverse-square law decay and oscillations in the time domain with the period $\tau' = 2\pi/m$. The reason why it is so is the following: the Fourier transform $\int d^3r' \boldsymbol{E}'(t', \boldsymbol{r}') \exp(-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}')$ for $\omega_k \to 0$ is the same as the Fourier transform of the Coulomb field multiplied by $\cos(mt')$.

To understand (8) better, we assume that [12]

$$f(\omega_k) = \left(\frac{m}{\varepsilon_k}\right)^{\beta}, \ \beta = 6, 8, 10, \cdots . \tag{9}$$

In such a case the energy is

$$m\frac{q^2}{8\pi^{3/2}}\frac{\Gamma(\beta - 3/2)}{\Gamma(\beta - 1)} \tag{10}$$

and we can use the results of [7] to verify that for $r' > |t'|$

$$V'(t', \boldsymbol{r}') \simeq \left(0, -\frac{q\hat{\boldsymbol{r}}'}{4\pi r'^2}\frac{\sin(mt')}{m}\right), \tag{11}$$

$$\boldsymbol{E}'(t', \boldsymbol{r}') \simeq \frac{q\hat{\boldsymbol{r}}'}{4\pi r'^2}\cos(mt'), \tag{12}$$

where the symbol $\simeq$ indicates that we have omitted the β-dependent terms because they are $\propto \exp(-mr')$ (such terms are discussed in Appendix B). Two remarks are in order now.

First, (12) shows that the periodically oscillating Coulomb component dominates the electric field not only for $r' \to \infty$ but for $r' > |t'|$ as long as $r' \gg 1/m$. Such a property should not be taken for granted because for $r' < |t'|$ the dynamics of the discussed field configuration is much more complicated and non-universal in the sense that it is β-dependent, which can be inferred from [7]. We focus on the universal features in the main body of this work.

Second, we would like to clarify one point that was not touched upon in [7]. Namely, given the fact that (12) describes the periodically oscillating Coulomb field, one may wonder why the vector field

$$\tilde{V}'(t', \boldsymbol{r}') \simeq \left(\frac{q}{4\pi r'}\cos(mt'), \boldsymbol{0}\right) \tag{13}$$

is not describing our field configuration for $r' > |t'|$ and $r' \gg 1/m$. Indeed, such a vector field also yields electromagnetic field (5) and (12) as well as it presents the straightforward generalization of the standard textbook expression to the periodically oscillating system. The problem with (13) is that it breaks one of Proca field equations in the $r' > |t'|$ region, where it leads to $\Box'\tilde{V}' = -m^2\tilde{V}'$ but

$$\partial' \cdot \tilde{V}' = -\frac{q}{4\pi r'}m\sin(mt') \neq 0. \tag{14}$$

This brings us to the point that the vector field is gauge fixed in the Proca theory [4], which one should keep in mind while reading this work. Such an observation is justified as follows. Suppose that both V' and $\tilde{V}' = V' + (\partial'^{\mu}\chi')$ satisfy Proca field equations, where χ' is a real-valued function. Then from $\partial' \cdot \tilde{V}' = 0$ one obtains $\Box'\chi' = 0$. By combining this result with $\Box'\tilde{V}' = -m^2\tilde{V}'$, one finds that $\partial'^{\mu}\chi'$ must vanish as long as $m \neq 0$. We mention in passing that there also exists the gauge invariant version of the Proca theory (see e.g. [11, 13, 14]).

II. ANOTHER REFERENCE FRAME: GENERAL CONSIDERATIONS

We will discuss now how the periodically oscillating Coulomb field is seen in the unprimed reference frame, which is moving with velocity $-\boldsymbol{v}$ with respect to the primed frame considered in Sec. I. The spacetime coordinates in the two systems are linked via the Lorentz transformation [15]

$$t' = \gamma(t - \boldsymbol{r}\cdot\boldsymbol{v}), \tag{15a}$$

$$\boldsymbol{r}' = \boldsymbol{r} - \gamma t\boldsymbol{v} + (\gamma - 1)(\boldsymbol{r}\cdot\hat{\boldsymbol{v}})\hat{\boldsymbol{v}}, \tag{15b}$$

where $\boldsymbol{r} = (x, y, z)$, $\gamma = 1/\sqrt{1 - v^2}$, and $v = |\boldsymbol{v}|$.

Moreover, the inverse transformation yields the vector field in the unprimed reference frame [15, 16]

$$V^0(t, \boldsymbol{r}) = \gamma\left[V'^0(t', \boldsymbol{r}') + \boldsymbol{V}'(t', \boldsymbol{r}')\cdot\boldsymbol{v}\right], \tag{16a}$$

$$\boldsymbol{V}(t, \boldsymbol{r}) = \boldsymbol{V}'(t', \boldsymbol{r}') + \gamma V'^0(t', \boldsymbol{r}')\boldsymbol{v} + (\gamma - 1)\left[\boldsymbol{V}'(t', \boldsymbol{r}')\cdot\hat{\boldsymbol{v}}\right]\hat{\boldsymbol{v}}, \tag{16b}$$

where the right-hand sides of these equations are supposed to be evaluated via (2) combined with (15). Such obtained vector field V satisfies the unprimed version of (3). Thereby, it presents the valid solution of the Proca field equations.

The electromagnetic field can be now computed from (16). The resulting expressions, however, are not insightful in their general $f(\omega_k)$-dependent form. Thereby, we

do not list them below. Still, some results can be obtained without assuming a specific form of $f(\omega_k)$.

First, the electric and magnetic fields satisfy $\boldsymbol{B}(t,\boldsymbol{r}) = \boldsymbol{v} \times \boldsymbol{E}(t,\boldsymbol{r})$ and $|\boldsymbol{E}(t,\boldsymbol{r})|^2/\gamma^2 + [\boldsymbol{v}\cdot\boldsymbol{E}(t,\boldsymbol{r})]^2 = |\boldsymbol{E}'(t',\boldsymbol{r}')|^2$. These expressions can be obtained by combining (5) with the following transformation law for the electromagnetic field [15, 16]

$$\boldsymbol{E}(t,\boldsymbol{r}) = \gamma\left[\boldsymbol{E}'(t',\boldsymbol{r}') - \boldsymbol{v}\times\boldsymbol{B}'(t',\boldsymbol{r}')\right] - (\gamma-1)\left[\boldsymbol{E}'(t',\boldsymbol{r}')\cdot\hat{\boldsymbol{v}}\right]\hat{\boldsymbol{v}}, \quad (17)$$

$$\boldsymbol{B}(t,\boldsymbol{r}) = \gamma\left[\boldsymbol{B}'(t',\boldsymbol{r}') + \boldsymbol{v}\times\boldsymbol{E}'(t',\boldsymbol{r}')\right] - (\gamma-1)\left[\boldsymbol{B}'(t',\boldsymbol{r}')\cdot\hat{\boldsymbol{v}}\right]\hat{\boldsymbol{v}}. \quad (18)$$

Second, the charge $Q(t,\mathbb{R}^3)$, given by the unprimed version of (8), can be discussed. As its value can be straightforwardly obtained via the integration of $-m^2V^0 = \boldsymbol{\nabla}\cdot\boldsymbol{E}$ (31), we present it as a function of

$$Q_1(t) = -m^2\gamma\int d^3r V'^0(t',\boldsymbol{r}') = q\gamma^2 f(\gamma m v)\cos(mt), \quad (19)$$

$$Q_2(t) = -m^2\gamma\int d^3r \boldsymbol{V}'(t',\boldsymbol{r}')\cdot\boldsymbol{v} = -q\gamma^2 f(\gamma m v)\cos(mt), \quad (20)$$

where the above expressions have been computed via the following formula

$$\int\frac{d^3r}{(2\pi)^3}\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}' \pm \mathrm{i}\varepsilon_k t') = \gamma\delta(\boldsymbol{k}\mp\gamma m\boldsymbol{v})\exp(\pm\mathrm{i}mt) \quad (21)$$

after reversing the order of spatial (d^3r) and momentum (d^3k) integrations; t' and $\boldsymbol{r}'$ have been replaced by (15) during the evaluation of (19)–(21). In terms of $Q_{1,2}(t)$, we find that

$$Q(t,\mathbb{R}^3) = Q_1(t) + Q_2(t) = 0 \text{ for } \boldsymbol{v}\neq\boldsymbol{0}, \quad (22)$$

$$Q(t,\mathbb{R}^3) = Q_1(t)|_{\boldsymbol{v}=\boldsymbol{0}} = q\cos(mt) \text{ for } \boldsymbol{v}=\boldsymbol{0}. \quad (23)$$

These results are rather surprising because they show that $Q(t,\mathbb{R}^3)$ is discontinuous at $\boldsymbol{v}=\boldsymbol{0}$, which is not evident from (16) or (17). The technical reason for the discontinuity is the following. For $\boldsymbol{v}=\boldsymbol{0}$, there is only one contribution to $Q(t,\mathbb{R}^3)$, which coincides with $Q_1(t)|_{\boldsymbol{v}=\boldsymbol{0}}$. For $\boldsymbol{v}\neq\boldsymbol{0}$, there are two contributions to $Q(t,\mathbb{R}^3)$: $Q_1(t)$ and $Q_2(t)$. The latter one does not disappear for $\boldsymbol{v}\to\boldsymbol{0}$ despite the appearances. Namely, the $\boldsymbol{v}$ "factor" under the integral in (20), which suggests vanishing of $Q_2(t)$ in such a limit, gets canceled during the calculations because it appears in the following setting

$$\frac{\boldsymbol{k}\cdot\boldsymbol{v}}{\omega_k^2}\delta(\boldsymbol{k}\mp\gamma m\boldsymbol{v}) = \pm\frac{\delta(\boldsymbol{k}\mp\gamma m\boldsymbol{v})}{\gamma m}, \quad (24)$$

which can be verified by combining (2b), (20), and (21). Nonvanishing of $Q_2(t)$ for $\boldsymbol{v}\to\boldsymbol{0}$ makes $Q(t,\mathbb{R}^3)$ discontinuous. The physical reason behind such a feature will be identified in Sec. III B.

III. ANOTHER REFERENCE FRAME: PARTICULAR SOLUTION

We turn our attention here towards the particular solution specified by (9). The following results can be compactly presented after the introduction of $\boldsymbol{R} = \boldsymbol{r} - \boldsymbol{v}t$, which leads to

$$\boldsymbol{r}' = \boldsymbol{R} + (\gamma-1)(\boldsymbol{R}\cdot\hat{\boldsymbol{v}})\hat{\boldsymbol{v}},\ r' = \sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}. \quad (25)$$

For $r > |t|$, the discussed Proca field configuration is characterized by

$$V^0(t,\boldsymbol{r}) \simeq V^0(\boldsymbol{R})\frac{\sin\left(mt/\gamma - m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right)}{m}, \quad (26a)$$

$$V^0(\boldsymbol{R}) = -q\frac{\gamma^2\boldsymbol{R}\cdot\boldsymbol{v}}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}, \quad (26b)$$

$$\boldsymbol{V}(t,\boldsymbol{r}) \simeq \boldsymbol{V}(\boldsymbol{R})\frac{\sin\left(mt/\gamma - m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right)}{m}, \quad (27a)$$

$$\boldsymbol{V}(\boldsymbol{R}) = -q\frac{\boldsymbol{R}+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})\boldsymbol{v}}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}, \quad (27b)$$

$$\boldsymbol{E}(t,\boldsymbol{r}) \simeq \boldsymbol{E}(\boldsymbol{R})\cos\left(mt/\gamma - m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right), \quad (28a)$$

$$\boldsymbol{E}(\boldsymbol{R}) = q\frac{\gamma\boldsymbol{R}}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}, \quad (28b)$$

$$\boldsymbol{B}(t,\boldsymbol{r}) \simeq \boldsymbol{B}(\boldsymbol{R})\cos\left(mt/\gamma - m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right), \quad (29a)$$

$$\boldsymbol{B}(\boldsymbol{R}) = \boldsymbol{v}\times\boldsymbol{E}(\boldsymbol{R}). \quad (29b)$$

Some technical remarks are in order now.

First, we note that (26)–(29) have been obtained by combining (5), (11), and (12) with (16)–(18). We also note that the subleading contributions to (26)–(29) are discussed in Appendix B.

Second, we observe that the range of validity of (26)–(29) follows from the fact that $r' > |t'|$ maps onto $r > |t|$ under the Lorentz transformation. This is easily seen if one notes that these two inequalities are equivalent to $X'\cdot X' < 0$ and $X\cdot X < 0$, where $X' = (t',\boldsymbol{r}')$ and $X = (t,\boldsymbol{r})$.

Third, we note that (26)–(29) satisfy the following equations in the $r > |t|$ region

$$\partial_t V^0(t,\boldsymbol{r}) + \boldsymbol{\nabla}\cdot\boldsymbol{V}(t,\boldsymbol{r}) = 0, \quad (30)$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{E}(t,\boldsymbol{r}) = -m^2V^0(t,\boldsymbol{r}), \quad (31)$$

$$\boldsymbol{\nabla}\times\boldsymbol{E}(t,\boldsymbol{r}) = -\partial_t\boldsymbol{B}(t,\boldsymbol{r}), \quad (32)$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{B}(t,\boldsymbol{r}) = 0, \quad (33)$$

$$\boldsymbol{\nabla}\times\boldsymbol{B}(t,\boldsymbol{r}) = -m^2\boldsymbol{V}(t,\boldsymbol{r}) + \partial_t\boldsymbol{E}(t,\boldsymbol{r}), \quad (34)$$

where $\boldsymbol{\nabla} = (\partial_x, \partial_y, \partial_z)$. These equations are equivalent to the unprimed version of (3). We also note that the primed version of (30)–(34) is satisfied by (5), (11), and (12) in the $r' > |t'|$ region. It is presumably worth to stress that all this happens despite the fact that (26)–(29) as well as (11) and (12) provide approximate expressions, which the symbol $\simeq$ indicates.

Fourth, we observe that the $\boldsymbol{R}$-dependence of (26)–(29) indicates that the studied field configuration moves as a whole with velocity $\boldsymbol{v}$, which is in accordance with our expectations.

In the following, we take a close look at the above expressions (Sec. III A) and the charge that can be computed out of them (Sec. III B).

A. Comparision to Maxwell theory

In the $m \to 0$ limit, the electric and magnetic fields, far away from the center of the studied field configuration, are described by $\boldsymbol{E}(\boldsymbol{R})$ and $\boldsymbol{B}(\boldsymbol{R})$: the electromagnetic field of the charge q moving with the constant velocity $\boldsymbol{v}$ [17]. Accordingly, vector field (26) and (27) for $m \to 0$ must be related to the expressions known in the context of the Maxwell theory. This is not immediately evident because

$$\left(\frac{q}{4\pi} \frac{\gamma}{\sqrt{R^2 + \gamma^2 (\boldsymbol{R}\cdot\boldsymbol{v})^2}}, \frac{q\boldsymbol{v}}{4\pi} \frac{\gamma}{\sqrt{R^2 + \gamma^2 (\boldsymbol{R}\cdot\boldsymbol{v})^2}} \right) \tag{35}$$

is traditionally discussed as the vector field of the charge moving with the velocity $\boldsymbol{v}$ [17]. However, $\lim_{m\to 0}(V^\mu)$ and (35) are related via the gauge transformation. Indeed, $\lim_{m\to 0}(V^\mu)$ is obtained by subtracting $(\partial^\mu \chi)$ from (35), where

$$\chi = \frac{q}{4\pi} \frac{t/\gamma - \gamma \boldsymbol{R}\cdot\boldsymbol{v}}{\sqrt{R^2 + \gamma^2 (\boldsymbol{R}\cdot\boldsymbol{v})^2}}. \tag{36}$$

The key differences between (26)–(29) and the Maxwell theory of a charge moving with a constant velocity come from the oscillatory terms, which are both time- and position-dependent.

The time dependence seen in (26)–(29) implies the oscillation period

$$\tau = \frac{2\pi}{m\sqrt{1-v^2}} = \gamma \tau', \tag{37}$$

which is dilated with respect to the oscillation period in the primed reference frame, where the center of the studied field configuration is at rest. While looking at this result, it is instructive to realize that the periods τ and τ' appear in slightly different contexts. The period τ' describes the harmonic oscillations of $\boldsymbol{V}'$ and $\boldsymbol{E}'$ taking place at any fixed point in the primed reference frame that is far-enough from the center of the discussed field configuration. In such a context, the periodically oscillating Coulomb field can be seen as a timekeeping element operating in the unit of τ'. Switching to the unprimed reference frame, the period τ describes the harmonic oscillations of V, $\boldsymbol{E}$, and $\boldsymbol{B}$ taking place at the set of spatially-separated space-time points: $(t, \boldsymbol{R} + \boldsymbol{v}t)$, where $\boldsymbol{R}$ is a large-enough constant displacement.

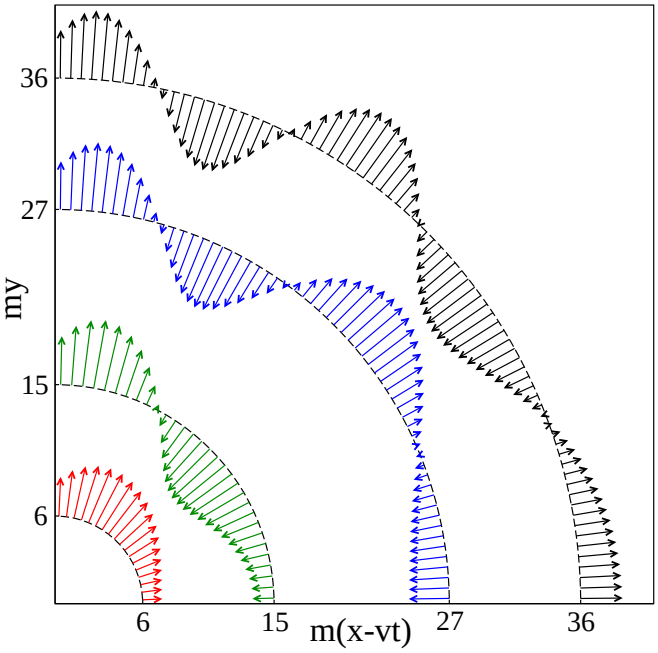


FIG. 1: The rescaled electric field in the unprimed reference frame. Namely, $\boldsymbol{E}(t,\boldsymbol{r}) \times |\boldsymbol{r} - \boldsymbol{v}t|^2$ in arbitrary units on the plane $z = 0$ for $mt = 1$, $\boldsymbol{v} = (1/3, 0, 0)$, and $mR = m|\boldsymbol{r} - \boldsymbol{v}t| = 6, 15, 27, 36$. The dashed lines depict quarter-circular segments. The plotted electric field has two components: leading contribution (28) and subleading contribution (B12) computed for $\beta = 6$. The latter contribution rapidly vanishes as mR grows. For the data displayed on the plot, mean $|\delta\boldsymbol{E}|/|\boldsymbol{E}|$ is approximately given by $2 \cdot 10^{-1}$ ($mR = 6$), $2 \cdot 10^{-4}$ ($mR = 15$), $8 \cdot 10^{-9}$ ($mR = 27$), and $2 \cdot 10^{-12}$ ($mR = 36$) [18].

The position dependence seen in the trigonometric functions in (26)–(29) adds a non-trivial local structure to the studied field configuration, which we illustrate in Figs. 1 and 2 in the electric field context. This can be analyzed by noting that the trigonometric functions in (26)–(29) are unchanged under the transformation

$$\boldsymbol{R} \to \boldsymbol{R} + \xi(\theta)\hat{\boldsymbol{R}}, \;\; \theta = \measuredangle(\boldsymbol{R}, \boldsymbol{v}), \tag{38a}$$

where

$$\xi(\theta) = \frac{2\pi}{mv\cos(\theta)} \sqrt{1-v^2}. \tag{38b}$$

The quasi-periodic behavior under (38) is exhibited by (26)–(29) when $R \gg \xi(\theta)$ because such a condition guarantees that $V(\boldsymbol{R})$, $\boldsymbol{E}(\boldsymbol{R})$, and $\boldsymbol{B}(\boldsymbol{R})$ negligibly change when (38) is performed (see Fig. 3 and comments therein). All in all, we note that the intriguing pattern is encoded in (26)–(29).

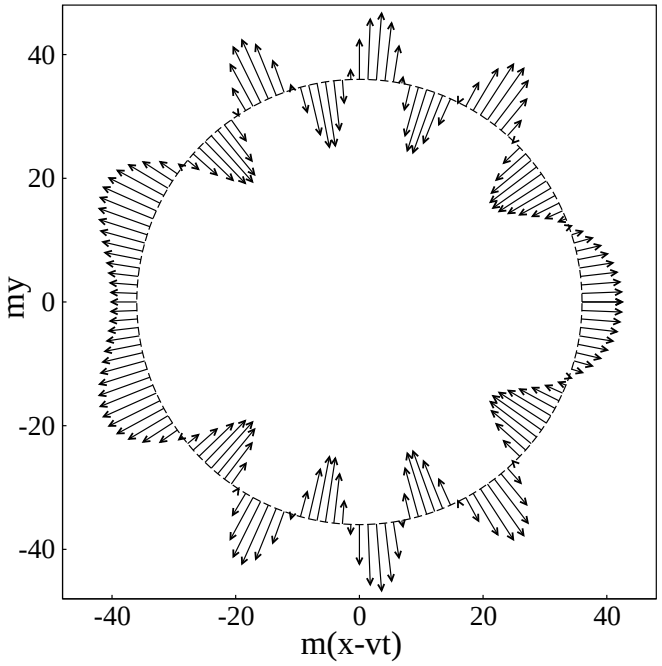


FIG. 2: The same as in Fig. 1 but on the whole x-y plane and for $mR = 36$ only. As $\boldsymbol{E}(t,\boldsymbol{r}) \propto \boldsymbol{R}$, where the spatial dependence of the proportionality coefficient is governed by R and $\boldsymbol{R}\cdot\boldsymbol{v} \propto \cos[\measuredangle(\boldsymbol{R},\boldsymbol{v})]$, one can easily visualize the three dimensional pattern of the electric field associated with this plot. Such a pattern is obtained by rotating the plotted vector field about the $y = 0$ line.

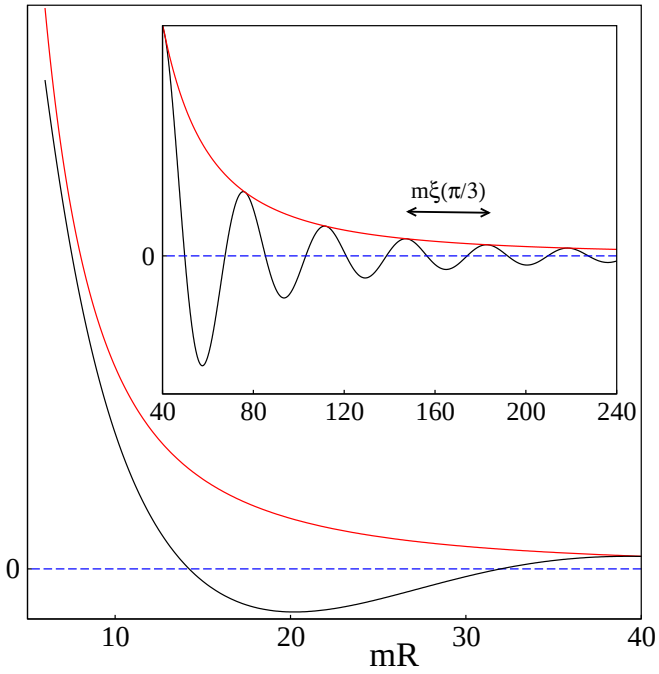


FIG. 3: The black line depicts $\boldsymbol{E}(t,\boldsymbol{r})\cdot\hat{\boldsymbol{R}}$ along the line set by the constraint $\boldsymbol{R}\cdot\boldsymbol{v} = Rv\cos(\pi/3)$ (leading (28) and subleading (B12) contributions are taken into account). The red line shows the upper envelope of the leading contribution to the electric field, i.e. $\boldsymbol{E}(\boldsymbol{R})\cdot\hat{\boldsymbol{R}}$ computed via (28b). The blue dashed line facilitates identification of the points, where the electric field changes direction. The inset shows the same as the main plot but for larger distances. The plot is prepared for $mt = 1$, $\boldsymbol{v} = (1/3, 0, 0)$, and $\beta = 6$. There are no quasi-periodic oscillations in the main plot, where mR is at most of the order of $m\xi(\pi/3) \approx 36$. The situation is different in the inset, where mR becomes substantially larger than 36.

B. Charge

We begin by defining

$$Q(t,\mathcal{V}) = \int_{\mathcal{V}} d^3r \boldsymbol{\nabla}\cdot\boldsymbol{E}(t,\boldsymbol{r}), \tag{39}$$

which conforms to previously introduced expression (8). To get quantitative insights into (22), we choose the integration volume in (39) to be the ball of radius R centered at $\boldsymbol{v}t$, where $R > (1+v)t$. We denote such a ball as $\mathcal{B}(R)$. Its surface lies inside the region, where (28) holds. Thereby, we proceed by replacing the volume integral by the surface one, which leads to

$$Q(t,\mathcal{B}(R)) \simeq q\gamma\cos(mt/\gamma)\int_0^1 ds \frac{\cos(pRs)}{[1+(ps/m)^2]^{3/2}}, \tag{40}$$

where $p = \gamma m v$ and $\simeq$ reminds us that expression (28), which becomes exact in the limit of $mR \to \infty$, has been used in order to derive (40). For $\boldsymbol{v} = \boldsymbol{0}$, (40) instantly yields the expected result. For $\boldsymbol{v} \neq \boldsymbol{0}$,

$$Q(t,\mathcal{B}(R)) \simeq \frac{q}{\gamma^2}\cos(mt/\gamma)\frac{\sin(pR)}{pR} + O\left(\frac{1}{(mR)^2}\right), \tag{41}$$

which follows from the computations presented in Appendix C. Taking the limit of $mR \to \infty$, (22) is obtained.

The disappearance of the charge $Q(t,\mathbb{R}^3)$ for $\boldsymbol{v} \neq \boldsymbol{0}$ can be seen as the result of self-averaging to zero of the Coulomb-like component of the electric field over large distances. This is caused by the spatial dependence of the cosine function in (28), which is parametrized by the length scale

$$1/p = \frac{\sqrt{1-v^2}}{mv}. \tag{42}$$

Such an observation is consistent with the fact that the damping factor in (41),

$$\frac{\sin(pR)}{pR}, \tag{43}$$

approaches zero on length scales set by (42). To conclude, we note that the oscillatory character of the electric field provides the physical reason for the discontinuity of $Q(t,\mathbb{R}^3)$ at $\boldsymbol{v} = \boldsymbol{0}$.

IV. SUMMARY

There exist field configurations in the Proca theory, where the electric field has the periodically oscillating Coulomb component. This unusual possibility, having no counterpart in the traditional Maxwell theory, was recently discussed in [7]. In the current work, we have described how such field configurations are seen in different reference frames. This has lead to the following new observations.

First, the periodically oscillating Coulomb field, when observed in a reference frame in which it moves as a whole, exhibits quasi-periodic spatial oscillations. Quite interestingly, the resulting pattern of the electric field, which is depicted in Figs. 1–3, does not resemble the well-known pattern of the electric field of the moving charge in the Maxwell theory (similar remarks apply to the magnetic field and the vector field obtained in our studies).

Second, the charge associated with the periodically oscillating Coulomb field–defined via the surface integral of the electric field at spatial infinity–is surprisingly sensitive to the choice of the reference frame. Namely, (i) it periodically oscillates in the reference frame in which the "center" of the studied field configuration is at rest and (ii) it vanishes in all other reference frames. These observations trigger two remarks. First, as only the (i) result was described in [7], the present work fills an important gap. Second, the (ii) result follows from the above-mentioned spatial oscillations of the electric field.

Finally, we mention that the dipole-charged field configurations, where the periodic oscillations of either magnetic or electric dipole moment are taking place, were recently studied in the quantum version of the Proca theory [19]. The reference frame dependence of their classical counterparts can be analyzed akin to what we have presented in this work. In fact, one may show that the same trigonometric factors as in (26)–(29) are responsible for the spatial oscillations of such Lorentz-transformed dipole-charged fields in Proca theory.

ACKNOWLEDGMENTS

I am indebted to Bogdan S. Damski for stimulating comments about this work. These studies have been supported by the Polish National Science Centre (NCN) Grant No. 2019/35/B/ST2/00034. The research for this publication has been also supported by a grant from the Priority Research Area DigiWorld under the Strategic Programme Excellence Initiative at Jagiellonian University.

APPENDIX A: Conventions

We adopt the Heaviside-Lorentz system of units and set $\hbar = c = 1$. Greek and Latin indices of tensors take values $0, 1, 2, 3$ and $1, 2, 3$, respectively. The metric signature is $(+ - --)$. 3-vectors are written in bold, e.g. $V = (V^\mu) = (V^0, \boldsymbol{V})$. We use the Einstein summation convention, $(X_{\mu\cdots})^2 = X_{\mu\cdots}X^{\mu\cdots}$. $\Gamma(x)$ is the gamma function, $\hat{\boldsymbol{x}} = \boldsymbol{x}/|\boldsymbol{x}|$, $\measuredangle(\boldsymbol{x}, \boldsymbol{y})$ is the angle between 3-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\partial_X = \partial/\partial X$.

APPENDIX B: Subleading contributions

We provide here subleading contributions to the vector field and the electromagnetic field in both reference frames discussed in the main body of this work. We denote them with the symbol δ (e.g. $\delta\boldsymbol{E}'$ ($\delta\boldsymbol{E}$) is the subleading contribution to the electric field in the primed (unprimed) reference frame).

To write the discussed expressions in the compact form, we introduce

$$F_\beta(a,b) = (1 + a - a\partial_a)P_\beta(a,b), \tag{B1}$$

$$\tilde{F}_\beta(a,b) = (1 + a - a\partial_a)\partial_b P_\beta(a,b), \tag{B2}$$

$$G_\beta(a,b) = (1 - 2\partial_a + \partial_a^2)P_\beta(a,b), \tag{B3}$$

where $P_\beta(a,b)$ are the bivariate polynomials whose properties are discussed in [7, 19]. In particular,

$$P_6(a,b) = 1 + \frac{5a}{8} + \frac{a^2}{8} - \frac{b^2}{2} - \frac{ab^2}{4} + \frac{b^4}{24}, \tag{B4}$$

$$P_8(a,b) = 1 + \frac{11a}{16} + \frac{3a^2}{16} + \frac{a^3}{48} - \frac{b^2}{2} - \frac{5ab^2}{16} - \frac{a^2b^2}{16} + \frac{b^4}{24} + \frac{ab^4}{48} - \frac{b^6}{720}, \tag{B5}$$

$$P_{10}(a,b) = 1 + \frac{93a}{128} + \frac{29a^2}{128} + \frac{7a^3}{192} + \frac{a^4}{384} - \frac{b^2}{2} - \frac{11ab^2}{32} - \frac{3a^2b^2}{32} - \frac{a^3b^2}{96} + \frac{b^4}{24} + \frac{5ab^4}{192} + \frac{a^2b^4}{192} - \frac{b^6}{720} - \frac{ab^6}{1440} + \frac{b^8}{40320}. \tag{B6}$$

The subleading contributions to (11) and (12) read

$$\delta V'^0(t', \boldsymbol{r}') = -\frac{q\exp(-mr')}{4\pi r'} G_\beta(mr', mt'), \tag{B7}$$

$$\delta\boldsymbol{V}'(t', \boldsymbol{r}') = -\frac{q\hat{\boldsymbol{r}}'\exp(-mr')}{4\pi r'^2}\frac{\tilde{F}_\beta(mr', mt')}{m}, \tag{B8}$$

$$\delta\boldsymbol{E}'(t', \boldsymbol{r}') = -\frac{q\hat{\boldsymbol{r}}'\exp(-mr')}{4\pi r'^2} F_\beta(mr', mt'). \tag{B9}$$

(B7)–(B9) should be understood as follows. If we add $(\delta V'^0, \delta\boldsymbol{V}')$ to the right-hand side of (11), we will get, in the $r' > |t'|$ region, an exact result for V'. Similarly, the addition of $\delta\boldsymbol{E}'$ to the right-hand side of (12) yields an

exact result for $\boldsymbol{E}'$ in the above-mentioned region. Finally, we note that $\delta\boldsymbol{B}'(t',\boldsymbol{r}')=\boldsymbol{0}$ because the magnetic field identically vanishes in the primed reference frame.

The subleading contributions to (26)–(29) read

$$\delta V^0(t,\boldsymbol{r})=-q\frac{\gamma^2\exp\left(-m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}\right)\boldsymbol{R}\cdot\boldsymbol{v}}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}\cdot\frac{\tilde{F}_\beta\left(m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2},mt/\gamma-m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right)}{m}-q\frac{\gamma\exp\left(-m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}\right)}{4\pi\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}}\cdot G_\beta\left(m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2},mt/\gamma-m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right),\quad\text{(B10)}$$

$$\delta\boldsymbol{V}(t,\boldsymbol{r})=-q\frac{\exp\left(-m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}\right)\left[\boldsymbol{R}+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})\boldsymbol{v}\right]}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}\cdot\frac{\tilde{F}_\beta\left(m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2},mt/\gamma-m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right)}{m}-q\frac{\gamma\exp\left(-m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}\right)\boldsymbol{v}}{4\pi\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}}\cdot G_\beta\left(m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2},mt/\gamma-m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right),\quad\text{(B11)}$$

$$\delta\boldsymbol{E}(t,\boldsymbol{r})=-q\frac{\gamma\exp\left(-m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2}\right)\boldsymbol{R}}{4\pi\left[R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2\right]^{3/2}}\cdot F_\beta\left(m\sqrt{R^2+\gamma^2(\boldsymbol{R}\cdot\boldsymbol{v})^2},mt/\gamma-m\gamma\boldsymbol{R}\cdot\boldsymbol{v}\right),\quad\text{(B12)}$$

$$\delta\boldsymbol{B}(t,\boldsymbol{r})=\boldsymbol{v}\times\delta\boldsymbol{E}(t,\boldsymbol{r}).\quad\text{(B13)}$$

Several remarks are in order now.

First, results (B7)–(B9) can be found in [7]. They are gathered here for the sake of completeness of our discussion. Results (B10)–(B13) have been obtained from (B7)–(B9) via (16)–(18). If we add (B10), (B11), (B12), and (B13) to (26), (27), (28), and (29), respectively, we will get exact expressions for V^0, $\boldsymbol{V}$, $\boldsymbol{E}$, and $\boldsymbol{B}$ in the $r>|t|$ region.

Second, the above expressions show that subleading contributions exponentially vanish with the distance from the center of the studied field configuration. This happens in both reference frames. The distance at which the subleading contributions become negligible depends on the parameter β.

Third, we note that (28) and $\delta\boldsymbol{E}(t,\boldsymbol{r})$ are proportional to $\boldsymbol{R}$, where the spatial dependence of the proportionality coefficient is determined by R and $\boldsymbol{R}\cdot\boldsymbol{v}$. This should be kept in mind when one looks at Figs. 1–3.

Fourth, we note that in the $r>|t|$ region (B10)–(B13) satisfy (30)–(34), where V^0, $\boldsymbol{V}$, $\boldsymbol{E}$, and $\boldsymbol{B}$ are replaced by δV^0, $\delta\boldsymbol{V}$, $\delta\boldsymbol{E}$, and $\delta\boldsymbol{B}$. In the same sense, the subleading contributions in the primed reference frame satisfy the primed version of (30)–(34) in the $r'>|t'|$ region.

APPENDIX C: Integral from expression (40)

The integral of interest here is

$$I=\int_0^1 ds\,g(s)\cos(pRs),\quad\text{(C1a)}$$

$$g(s)=\frac{1}{\left[1+(ps/m)^2\right]^{3/2}},\quad\text{(C1b)}$$

where $p=\gamma mv$. Integrating twice by parts, we find

$$I=\frac{1}{\gamma^3}\frac{\sin(pR)}{pR}+\Delta,\quad\text{(C2a)}$$

$$\Delta=\frac{1}{(pR)^2}\int_0^1 ds\,g''(s)\left[\cos(pR)-\cos(pRs)\right],\quad\text{(C2b)}$$

where $d^2g/ds^2=g''$. To bound Δ, we compute

$$\begin{aligned}|\Delta|&\le\frac{2}{(pR)^2}\int_0^1 ds\,|g''(s)|\\&=\frac{6}{(mR)^2}\int_0^1 ds\frac{\left|1-4(ps/m)^2\right|}{\left[1+(ps/m)^2\right]^{7/2}}\\&\le\frac{6}{(mR)^2}\int_0^1 ds\frac{1+4(ps/m)^2}{\left[1+(ps/m)^2\right]^{7/2}}\\&=\left(6+4v^2-\frac{18}{5}v^4\right)\frac{1}{\gamma(mR)^2}\\&\le\frac{64}{9}\frac{1}{\gamma(mR)^2},\end{aligned}\quad\text{(C3)}$$

where knowledge of the following easy-to-verify indefinite integral has been employed

$$\int ds\frac{1+4(ps/m)^2}{\left[1+(ps/m)^2\right]^{7/2}}=\frac{s+\frac{8}{3}(p/m)^2s^3+\frac{16}{15}(p/m)^4s^5}{\left[1+(ps/m)^2\right]^{5/2}}.\quad\text{(C4)}$$

[1] R. Greiner and J. Reinhardt, *Field Quantization* (Springer-Verlag, 1996).

[2] B. G.-g. Chen, D. Derbes, D. Griffiths, B. Hill, R. Sohn, and Y.-S. Ting, *Lectures of Sidney Coleman on Quantum*

Field Theory (World Scientific, 2018).
[3] S. Weinberg, *The Quantum Theory of Fields*, vol. I: Foundations (Cambridge University Press, 2010).
[4] A. S. Goldhaber and M. M. Nieto, Rev. Mod. Phys. **43**, 277 (1971).
[5] L.-C. Tu, J. Luo, and G. T. Gillies, Rep. Prog. Phys. **68**, 77 (2005).
[6] A. S. Goldhaber and M. M. Nieto, Rev. Mod. Phys. **82**, 939 (2010).
[7] B. Damski, Nucl. Phys. B **994**, 116300 (2023).
[8] These statements can be justified by doing the angular integration in (2) and (4). Writing d^3k as $d\omega_k\omega_k^2 d\Omega(\hat{\boldsymbol{k}})$, this amounts to the evaluation of $\int d\Omega(\hat{\boldsymbol{k}})$ in the abovementioned expressions.
[9] R. Lakes, Phys. Rev. Lett. **80**, 1826 (1998).
[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).
[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Broken symmetries and the Goldstone theorem, in *Advances in Particle Physics*, edited by R. L. Cool and R. E. Marshak (Interscience Publishers, New York, 1968), Vol. 2, pp. 567–708.
[12] Positive even β's are preferable because such a choice facilitates analytical evaluation of the discussed fields. The chosen values of β guarantee differentiability of the vector field in the sense specified below (2). These technical remarks follow from the discussion presented in [7].
[13] B. M. Pimentel and G. E. R. Zambrano, Nucl. Part. Phys. Proc. **267-269**, 183 (2015).
[14] B. Damski, Phys. Rev. D **107**, 045016 (2023).
[15] A. Dragan, *Unusually Special Relativity* (World Scientific, 2021).
[16] M. Maggiore, *A Modern Introduction to Classical Electrodynamics* (Oxford University Press, 2023).
[17] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Elsevier Butterworth-Heinemann, 2004).
[18] The mean is defined as $\sum_{i=1}^{n} |\delta\boldsymbol{E}_i|/|\boldsymbol{E}_i|$ divided by n, where $\boldsymbol{E}_i$ is the i-th vector displayed in Fig. 1 for a given mR and $\delta\boldsymbol{E}_i$ is the subleading contribution to such $\boldsymbol{E}_i$ (e.g. $n = 18$ for $mR = 6$).
[19] B. Damski, Nucl. Phys. B **998**, 116398 (2024).

arXiv:2402.18557v1 [hep-ph] 28 Feb 2024

Dark matter from mediator decay in early matter domination

Rouzbeh Allahverdi [*1], Ngo Phuc Duc Loc [†1], and Jacek K. Osiński [‡2]

[1]Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
[2]AstroCeNT, Nicolaus Copernicus Astronomical Center of the Polish Academy of Sciences, ul. Rektorska 4, 00-614 Warsaw, Poland

Abstract

We study dark matter production from mediator decays in scenarios with an epoch of early matter domination. Particles that mediate interactions between dark matter and the standard model particles are kinematically accessible to the thermal bath as long as their mass is below the reheating temperature of the Universe after inflation. Decay of on-shell mediators can then lead to copious production of dark matter during early matter domination or a preceding radiation-dominated phase. In particular, for mediators that are charged under the standard model, it can exceed the standard freeze-in channel due to inverse annihilations at much lower temperatures (often by many orders of magnitude). The requirement to obtain the correct relic abundance severely constrains the parameter space for dark matter masses above a few TeV.

1 Introduction

While there are many lines of evidence for the existence of dark matter (DM) in the Universe [1], the identity of DM is an outstanding problem at the interface of particle physics and cosmology. Explaining the observed DM abundance is another challenge that in addition depends on the details of the thermal history of the early Universe. Thermal freeze-out in a radiation-dominated (RD) Universe is a simple and attractive mechanism that can yield the correct relic abundance if the (thermally averaged) DM annihilation rate takes the specific value $\langle\sigma_{\rm ann}v\rangle = 3\times10^{-26}{\rm cm}^3\ {\rm s}^{-1}$. However, in nonstandard cosmological histories, the correct DM abundance can be obtained for much larger or smaller values of $\langle\sigma_{\rm ann}v\rangle$ [2, 3].

It is known that well-motivated classes of models arising from string theory generically lead to nonstandard histories that involve one or more epochs of early matter domination (EMD) [4]. In general, an EMD phase arises when a matter-like component dominates the energy density of the Universe and eventually decays to establish a RD Universe prior to big bang nucleosynthesis (BBN). The matter equation of state can be due to coherent oscillations

[*]rouzbeh@unm.edu
[†]locngo148@gmail.com
[‡]jaksaosinski@gmail.com

of a scalar field, like a string modulus, that is displaced from the minimum of its potential during inflation [5, 6, 7, 8]. It can also arise from long-lived nonrelativistic quanta produced in the postinflationary Universe [9, 10, 11, 12]. In fact, it has been recently shown that a visible-sector long-lived particle (LLP) with a weak-scale mass may achieve this [13, 14].

The indirect DM searches have set stringent constraints on the DM annihilation rate, most notably the recent Fermi-LAT results from dwarf spheroidal galaxies [15] and Milky Way satellites [16]. An analysis of these results has pushed the upper limit on $\langle\sigma_{\rm ann}v\rangle$ below the nominal value of 3×10^{-26} cm^3 s^{-1} for DM masses below 20 GeV in a model-independent way [17] (larger masses can be excluded for specific annihilation channels). The increasingly tighter experimental bounds therefore warrant studying cases with a small DM annihilation rate. If $\langle\sigma_{\rm ann}v\rangle$ is very small, inverse annihilation will never be in thermal equilibrium and DM production occurs in the freeze-in regime.

Freeze-in production during EMD has been studied in the literature [18, 19, 20, 21, 22] (for a detailed review, see [23]). These studies have mostly focused on the latter part of an EMD epoch where evolution of radiation is nonadiabatic. It has been noted that earlier stages of the thermal history can significantly contribute to freeze-in production and thereby dominate the DM relic abundance (for example, see [24]). This is an important consideration because an epoch of EMD is typically preceded by other phases in the postinflationary history (notably the RD phase established at the end of inflationary reheating), especially for the last epoch of EMD that can end just before BBN.

A period of EMD starting with a non-negligible initial amount of radiation goes through an adiabatic phase during which $T\propto H^{2/3}$, as opposed to $T\propto H^{1/4}$ in the nonadiabatic phase. In fact, as we will see, the Universe is generally in this adiabatic phase for most of the EMD period, in terms of the temperature evolution. The faster redshift of temperature in this phase then implies that particles that mediate DM interactions with the standard model (SM) particles can be kinematically accessible to the thermal bath during EMD (or in a preceding phase) even if they are very heavy. The mediator particles can indeed reach equilibrium if their couplings to the SM particles are not too small, and hence their decay can be an important source of DM production.

In this work, we present a detailed study of the contribution from on-shell mediator decays to the DM relic abundance in the freeze-in regime. By paying special attention to the adiabatic phase of EMD, we will show that these decays can easily dominate over the standard freeze-in production from inverse annihilations at much lower temperatures by many orders of magnitude. In fact, mediator decays alone can easily lead to DM overproduction in large parts of the parameter space. As we will see, the parameter space is tightly constrained for DM masses above a few TeV. The resulting constraints are milder and the allowed parameter space will extend to smaller DM masses if DM and the mediator belong to a hidden sector with its own gauge symmetry.

The rest of this paper is organized as follows. In Section 2, we introduce the general picture for mediator coupling to DM and SM particles. In Section 3, we discuss some important details of EMD epochs. In Section 4, we present our main results and identify the allowed region of the parameter space in order for mediator decays not to overproduce DM. We also comment on how our results are modified in the case of hidden sector DM models. We conclude the paper in Section 5. Some details of our calculations are discussed in the Appendix.

2 The set up

The main particles that are of interest in this work are the DM candidate χ and the mediator of its interactions with the SM particles X. In general, three types of renormalizable interactions that are linear in X and involve χ and the SM particles (collectively denoted by ψ) are possible:

$$h_1 X\chi\psi \quad , \quad h_2 X\chi\chi \quad , \quad h_3 X\psi\psi, \tag{1}$$

where h is a dimensionless coupling. We assume that X is charged under the SM gauge group, but remain agnostic about the bosonic or fermionic nature of it and χ, as well as the Lorentz structure of the interaction terms. All that is required is that the terms in Eq. (1) are invariant under the SM gauge symmetry and other symmetries that a given model may possess (for example, the symmetry that is responsible for stability of DM).

To put the terms in Eq. (1) in perspective, let us consider some general classes of models. In supersyymetric extensions of the SM with R-parity conservation, the lightest supersymmetric particle (LSP) is stable and hence a DM candidate (for a review, see [25]). Then interactions of the type $X\chi\psi$ arise with X being any superpartner that is coupled to DM. Interactions of the form $X\chi\chi$ and $X\psi\psi$ can also arise if X is a sufficiently heavy R-parity even particle. In models where the SM gauge symmetry is extended, interactions of the $X\chi\chi$ and $X\psi\psi$ types can arise with X being a gauge boson of the new symmetries. For example, in the $U(1)_{B-L}$ extension of the SM [26], Z'_{B-L} is coupled to all SM fermions as well as the RH neutrinos the lightest of which can be a DM candidate [27]. Finally, in minimal extensions of the SM that only include new fields, interactions of the type $X\chi\psi$ and $X\psi\psi$ can arise. For example, in the model with GeV-scale DM proposed in [28], X is a color-triplet and iso-singlet scalar that is coupled to the up-type and down-type RH quarks and DM.

In cases that $m_\chi \ll m_X$, the first two interaction terms in Eq. (1) result in the following partial decay width:

$$\Gamma_{X\to\chi} \simeq C_1 \times \frac{h^2}{8\pi} m_X, \tag{2}$$

where h is the larger of h_1 and h_2, and C_1 is a factor whose exact value depends on the nature of X and χ and details of their couplings. In the following, we will take $C_1 \sim 1$. Given that X is in equilibrium with the thermal bath at $T \geq m_X$, because of carrying SM charges, its decay can lead to copious production of DM even if h is very small.

The interactions in Eq. (1) also lead to annihilation/inverse annihilation $\chi\chi \leftrightarrow \psi\bar{\psi}$ and scattering $\chi\psi \leftrightarrow \chi\psi$ processes mediated by X. The former contributes to χ production from the thermal bath. However, X decay typically dominates χ production as long as the Universe starts at a temperature $T \gg m_X$ so that on-shell X particles exist in sufficient abundance. This is because $\Gamma_{\rm ann} \ll \Gamma_{X\to\chi}$, where $\Gamma_{\rm ann} = \langle\sigma_{\rm ann} v\rangle n_\chi$.

We note that if $\Gamma_{\rm ann} \ll H(T \sim m_\chi)$, the comoving number density of DM particles produced from X decay will remain frozen. The annihilation rate at energies $E \sim m_\chi$ depends on the dimension of the effective operator $\chi\chi\psi\psi$ and its Lorentz structure. For

example, assuming S-wave dominance, at $E \ll m_X$ we have:

$$\langle \sigma_{\rm ann} v \rangle_{\rm dim-5} \simeq C_2 \times \frac{h^4}{16\pi m_X^2}\,, \tag{3}$$

$$\langle \sigma_{\rm ann} v \rangle_{\rm dim-6} \simeq C_2 \times \frac{h^4 E^2}{16\pi m_X^4}\,, \tag{4}$$

where C_2 is a multiplicity factor.

3 Early matter domination: generalities

In this work, we consider EMD that starts at some point after the completion of inflationary reheating. The matter component ϕ carries an energy density ρ_ϕ, and its decay at the rate Γ_ϕ feeds the radiation energy density $\rho_{\rm r}$. The evolution of $\rho_{\rm r}$ and ρ_ϕ is governed by the following system of Boltzmann equations:

$$\begin{aligned} \frac{d\rho_{\rm r}}{dt} + 4H\rho_{\rm r} &= \Gamma_\phi \rho_\phi\,, \\ \frac{d\rho_\phi}{dt} + 3H\rho_\phi &= -\Gamma_\phi \rho_\phi\,, \end{aligned} \tag{5}$$

where the Hubble rate H is set by the total energy density $\rho_{\rm tot} = \rho_\phi + \rho_{\rm r}$.

We take the moment when $\rho_\phi = \rho_{\rm tot}/2$ to be the onset of the EMD epoch. We denote the Hubble rate and the temperature at this time by $H_{\rm O}$ and $T_{\rm O}$ respectively. Given that $\rho_{\rm r} = \rho_{\rm tot}/2$, we have:

$$H_{\rm O} = \left(\frac{\pi^2 g_{*\rm O}}{45}\right)^{1/2} \left(\frac{T_{\rm O}^4}{M_{\rm P}^2}\right)^{1/2}\,, \tag{6}$$

where $M_{\rm P}$ is the reduced Planck mass and g_* counts the relativistic degrees of freedom at the time indicated by the subscript. Note that in the most extreme case we have $T_{\rm O} \simeq T_{\rm reh}$ (where $T_{\rm reh}$ is the inflationary reheat temperature). The EMD epoch that starts at $H \simeq H_{\rm O}$ consists of two phases:

Adiabatic phase - In this phase, the initial radiation dominates over that produced from ϕ decay. As a result, $T \propto a^{-1}$, with a being the scale factor, similar to RD even though the Universe is in an EMD epoch where $a \propto t^{2/3}$. During this stage, the relation between H and T follows:

$$H \simeq \left(\frac{\pi^2 g_*^{3/4} g_{*\rm O}^{1/4}}{90}\right)^{1/2} \frac{T^{3/2} T_{\rm O}^{1/2}}{M_{\rm P}}. \tag{7}$$

This phase eventually ends when the temperature drops to the following value:

$$T_{\rm tr} \simeq \left(\frac{g_{*\rm R}^4 g_{*\rm O}}{g_*^5}\right)^{1/20} (T_{\rm R}^4 T_{\rm O})^{1/5}. \tag{8}$$

Nonadiabatic phase - There is a transition to the nonadiabatic phase at $T_{\rm tr}$ where radiation produced by ϕ decay becomes dominant over the preexisting amount. During this stage, we

have:

$$H \simeq \frac{5g_*}{6g_{*\rm R}^{1/2}} \left(\frac{\pi^2}{30}\right)^{1/2} \left(\frac{T^4}{T_{\rm R}^2 M_{\rm P}}\right), \tag{9}$$

which implies that $T \propto a^{-3/8}$. This stage extends all the way from $T \simeq T_{\rm tr}$ to the end of the EMD epoch when $H \sim \Gamma_\phi$ and $T \simeq T_{\rm R}$.

This discussion holds even if some non-radiation component with an equation-of-state parameter $0 < w \leq 1$ makes the dominant contribution to $\rho_{\rm tot}/2$ at the onset of EMD, provided that it is not converted into radiation. Note that a component with $w > 0$ is redshifted faster than matter, and hence will become increasingly subdominant during EMD without any need to decay to radiation. If it feeds radiation, on the other hand, entropy will increase at a faster rate during EMD thereby leading to a shorter adiabatic phase and a higher $T_{\rm tr}$ than what is mentioned above.

To illustrate the relative duration of the two stages of EMD, we consider two generic possibilities regarding the origin of such an epoch: coherent oscillations of a scalar field that is displaced from the minimum of its potential, and nonrelativistic particles produced from the thermal bath.

(1) Coherent oscillations - Consider a scalar field ϕ with mass m_ϕ (a notable example is string moduli [4]). If $m_\phi \ll H_{\rm inf}$, where $H_{\rm inf}$ is the Hubble rate during inflation, then ϕ can be significantly displaced from the minimum of its low-energy potential during inflation. It typically starts to oscillate about this minimum when $H \simeq m_\phi$ and its coherent oscillations behave like non-relativistic particles of mass m_ϕ. If the Universe is RD at this time, and the initial fractional energy density of ϕ is α, coherent oscillations start to dominate when the Hubble expansion rate is:

$$H_{\rm O} \simeq \alpha^2 m_\phi. \tag{10}$$

This signals the start of an EMD epoch when the temperature reaches the following value:

$$T_{\rm O} \simeq \alpha \left(\frac{45}{\pi^2 g_{*\rm O}}\right)^{1/4} (m_\phi M_{\rm P})^{1/2}. \tag{11}$$

On the other hand, if the Universe is still dominated by inflaton oscillations when $H \sim m_\phi$, we will instead have:

$$H_{\rm O} \simeq \alpha^2 H_{\rm reh}, \tag{12}$$

where $H_{\rm reh} < m_\phi$ is the Hubble rate when a RD Universe is established after inflationary reheating. Therefore, given that $\alpha \lesssim 1$ [1], the temperature at the onset of the EMD epoch follows:

$$T_{\rm O} \lesssim \left(\frac{45}{\pi^2 g_{*\rm O}}\right)^{1/4} (m_\phi M_{\rm P})^{1/2}. \tag{13}$$

(2) Nonrelativistic quanta - Consider a (bosonic or fermionic) particle ϕ from decay or scattering processes in the thermal bath. The produced quanta become nonrelativistic at

[1] In realistic situations, we typically have $\alpha \gtrsim \mathcal{O}(10^{-2})$ [8].

$T \sim m_\phi$. If the fractional energy density of ϕ at this time is α, its quanta dominate the Universe when the Hubble expansion rate is:

$$H_{\rm O} \sim \alpha^2 \left(\frac{\pi^2 g_{*m_\phi}}{45}\right)^{1/2} \frac{m_\phi^2}{M_{\rm P}}. \tag{14}$$

We note that $\alpha \lesssim g_{*m_\phi}^{-1}$, with the maximum occurring when ϕ reaches thermal equilibrium. Thus, in this case, we have:

$$T_{\rm O} \lesssim \left(\frac{g_{*m_\phi}^3}{g_{*\rm O}}\right)^{1/2} m_\phi. \tag{15}$$

To get a feeling of the relative duration of the adiabatic and nonadiabatic phases, let us consider the $m_\phi \simeq 10^2 - 10^6$ GeV mass range. The upper and lower limits correspond to the cases when ϕ is a string modulus [6, 7] and a weak-scale visible sector particle [13, 14] respectively. Eqs. (13,15) then give us:

$$\begin{aligned} T_{\rm O} &\sim 10^{10} - 10^{12}\ \text{GeV} \qquad \text{(Case 1)}\,, \\ T_{\rm O} &\sim 1 - 10^{4}\ \text{GeV} \qquad\ \ \text{(Case 2)}\,, \end{aligned} \tag{16}$$

where we have used $g_* \sim \mathcal{O}(100)$ as in the case of the SM. We therefore see that $T_{\rm O}$ can be within a vast range spanning many orders of magnitude.

The important point is that $T_{\rm tr}$ is in general not extremely larger than $T_{\rm R}$. It follows from Eq. (8) that $T_{\rm tr}/T_{\rm R}$ scales as $(T_{\rm O}/T_{\rm R})^{1/5}$. For example, for $T_{\rm R} \sim 10$ MeV, we find:

$$\begin{aligned} T_{\rm tr} &\sim 100 - 300\ \text{GeV} \qquad \text{(Case 1)}\,, \\ T_{\rm tr} &\sim 0.03 - 10\ \text{GeV} \qquad \text{(Case 2)}\,, \end{aligned} \tag{17}$$

We see that while $T_{\rm O}$ is typically much higher in case 1, see Eq. (16), in both cases the majority of the EMD epoch is spent in the adiabatic phase $T_{\rm tr} \lesssim T \lesssim T_{\rm O}$.

4 Results

We consider the evolution of the χ number density produced from the decay of X in the bath, as described by the following Boltzmann equation [18, 21, 29, 30]:

$$\frac{dn_\chi}{dt} + 3Hn_\chi = \Gamma_{X\to\chi} \frac{n_X^{\rm eq}}{n_\chi^{\rm eq}} \frac{K_1(m_X/T)}{K_2(m_X/T)} (n_\chi^{\rm eq} - n_\chi)\,, \tag{18}$$

where K_i are the modified Bessel functions of the second kind. The superscript 'eq' denotes the corresponding equilibrium number density. Note that X reaches equilibrium with the thermal bath at $T \gg m_X$ due to its SM charges. n_X will follow its equilibrium value at $T < m_X$ if annihilation to the SM particles is efficient or in the presence of an efficient decay mode via the $hX\psi\psi$ term in Eq. (1).

In general, the χ number density will become frozen at some time in the cosmological history. Depending on the values of the parameters, this can occur in any of the phases

described in the previous section. Furthermore, the freezing process may proceed as a freeze-in, where χ never achieves equilibrium with the bath, or a freeze-out, where the production rate is sufficient to reach equilibrium, with χ subsequently decoupling.

We can analytically estimate the abundance of χ produced from X decay, expressed as the number density normalized by the entropy density n_χ/s, in the following way (see Appendix A for details). We first define the temperature of the background bath at the time when the χ number density freezes as T_f. Beyond this point the number density is only redshifted with expansion as $n_\chi \propto a^{-3}$. We express the number density n_χ at this time as:

$$n_\chi(T_\mathrm{f}) = F(\gamma_\chi) n_\chi^\mathrm{eq}(T_\mathrm{f}), \tag{19}$$

where the function $F(\gamma_\chi)$ accounts for the possibility of not reaching equilibrium, with $\gamma_\chi \equiv \Gamma_{X\to\chi}/H(T = m_X)$. We will typically consider $T_\mathrm{f} \gg m_\chi$ such that the equilibrium number density above is given by the relativistic expression proportional to T_f^3.

In general, the χ number density produced from X decay can become frozen at different stages of the cosmological history described in the previous section:

(1) $T_\mathrm{R} < m_X \lesssim T_\mathrm{tr}$ - In this case, n_χ freezes during the nonadiabatic phase of EMD, and the final DM abundance is given by:

$$\left(\frac{n_\chi}{s}\right)^\mathrm{NA} \simeq 0.53 \frac{g_\chi g_{*\mathrm{R}}}{g_{*\mathrm{f}}^2} F(\gamma_\chi) \left(\frac{T_\mathrm{R}}{T_\mathrm{f}}\right)^5, \tag{20}$$

where g_χ is the number of degrees of freedom in χ.

(2) $T_\mathrm{tr} < m_X \lesssim T_\mathrm{O}$ **-** In this case, n_χ freezes during the adiabatic phase of EMD and the DM relic abundance follows:

$$\left(\frac{n_\chi}{s}\right)^\mathrm{A} \simeq 0.28 \frac{g_\chi}{g_{*\mathrm{f}}} F(\gamma_\chi) \left(\frac{T_\mathrm{R}}{T_\mathrm{O}}\right). \tag{21}$$

(3) $T_\mathrm{O} < m_X \lesssim T_\mathrm{reh}$ **-** In this case, n_χ freezes during the RD phase preceding the EMD epoch. Given that the frozen n_χ and s are both redshifted $\propto a^{-3}$ during this RD phase and the adiabatic phase of EMD alike, we have:

$$\left(\frac{n_\chi}{s}\right)^\mathrm{RD} \simeq 0.28 \frac{g_\chi}{g_{*\mathrm{f}}} F(\gamma_\chi) \left(\frac{T_\mathrm{R}}{T_\mathrm{O}}\right). \tag{22}$$

If $\gamma_\chi \ll 1$, production of χ will not be strong enough to reach equilibrium and the number density of χ saturates to $n_\chi \approx 7\gamma_\chi n_\chi^\mathrm{eq}$. The temperature at the time of χ freeze-in in this case falls roughly within the range $m_X/10 \lesssim T_\mathrm{f} \lesssim m_X/5$. On the other hand, if equilibrium is reached, we have $\gamma_\chi \gtrsim 1$ and $F(\gamma_\chi) = 1$, with the temperature at χ freeze-out not being limited to the previous range.

In the case that χ reaches equilibrium in the nonadiabatic phase (or maintains it through the transition from the adiabatic phase), we can estimate the decoupling temperature using the RHS of the Boltzmann equation Eq. (18):

$$\Gamma_{X\to\chi} \frac{g_X m_X^2}{g_\chi m_\chi^2} \frac{K_1(m_X/T_\mathrm{f})}{K_2(m_\chi/T_\mathrm{f})} \approx 3H(T_\mathrm{f}), \tag{23}$$

where we have made use of $n_\chi^{\rm eq}(T) = (g_\chi/2\pi^2)m_\chi^2 T K_2(m_\chi/T)$, and similarly for $n_X^{\rm eq}(T)$ [2]. Using this to find $T_{\rm f}$, which we identify as the decoupling temperature in this case, we can obtain the abundance at the end of EMD $(n_\chi/s)_{\rm R}$ for cases where χ remains relativistic and in equilibrium in the nonadiabatic phase at temperatures below m_X. We note that if equilibrium is reached during the adiabatic phase, but equilibrium is not maintained through the transition (i.e., decoupling occurs before $T_{\rm tr}$), then there is no dependence on the decoupling temperature, as seen in Eq. (21) above.

With $T_{\rm f}$ determined, we can now compare the frozen χ abundance to the observed DM relic abundance in order to constrain such scenarios:

$$\frac{n_\chi}{s} \lesssim \left(\frac{n_\chi}{s}\right)_{\rm obs} \simeq 4\times 10^{-10} \times \frac{1\ {\rm GeV}}{m_\chi}\,. \tag{24}$$

Here, we only require that DM is not overproduced from X decays. Underproduction may be compensated for by other sources that create χ (for example, direct decay of ϕ or inverse annihilations).

We numerically solve the Boltzmann equations for the background energy densities, given by Eq. (5) together with the equation for the DM number density produced from X decay, Eq. (18). In Figs. 1, 2, and 3 we show contours in the $T_{\rm O}$-m_X plane that achieve the observed DM relic abundance for the values of the parameters shown in each figure. The region above and to the right of a given contour corresponds to underproduction of DM. In Fig. 1 we vary $T_{\rm R}$ for fixed h and m_χ. In Fig. 2 we vary the DM mass m_χ while keeping h and $T_{\rm R}$ fixed. In Fig. 3 we vary h for fixed values of $T_{\rm R}$ and m_χ.

The vertical dashed line denotes the current LHC limit on new particles charged under the SM, which we loosely take to be $m_X \gtrsim \mathcal{O}({\rm TeV})$ [3]. One could also constrain $T_{\rm O}$ by considering the absolute upper bound $T_{\rm O} < T_{\rm reh}$ where $T_{\rm reh} \lesssim 10^{-1}(H_{\rm inf}M_{\rm P})^{1/2}$ [4]. The current Planck bounds [33] on the tensor-to-scalar ratio correspond to $H_{\rm inf} \lesssim 10^{13}$ GeV, which results in a very weak limit $T_{\rm reh} \lesssim 10^{15}$ GeV. Other considerations can significantly tighten this bound. For example, the recently proposed trans-Planckian censorship conjecture of the swampland program sets an upper limit of $(H_{\rm inf}M_{\rm P})^{1/2} \lesssim 10^9$ GeV in order for modes with sub-Planckian wavelength to not exit the horizon during inflation. This would result in $T_{\rm O} < 10^9$ GeV thereby removing the very top of the $T_{\rm O} - m_X$ plane. Model-dependent bounds on $H_{\rm inf}$ lead to similar restrictions.

The diagonal dashed line shows where $T_{\rm O} = m_X$. Extending the contours to the region below this line assumes that the Universe is in a RD phase from $T = T_{\rm O}$ all the way up to $T \simeq m_X$. This will not be the case if m_X exceeds the reheating temperature after inflation $T_{\rm reh}$, or if there is another epoch of EMD at $H > H_{\rm O}$. Thus, the dependence on the details of the thermal history above $T_{\rm O}$ requires the $m_X > T_{\rm O}$ half-plane to be treated with some care.

[2] Given that $g_\chi = 1$ (3/4) for a single bosonic (fermionic) degree of freedom, we take $g_\chi = 1$ in numerical calculations.

[3] The exact bound depends on the SM charge assignment of X (for example, colored versus colorless particles).

[4] This upper bound is saturated in the case of instant reheating and thermalization. The exact value of $T_{\rm reh}$ depends on the details of reheating (for reviews, see [31, 32]) and can be much lower than this upper bound.

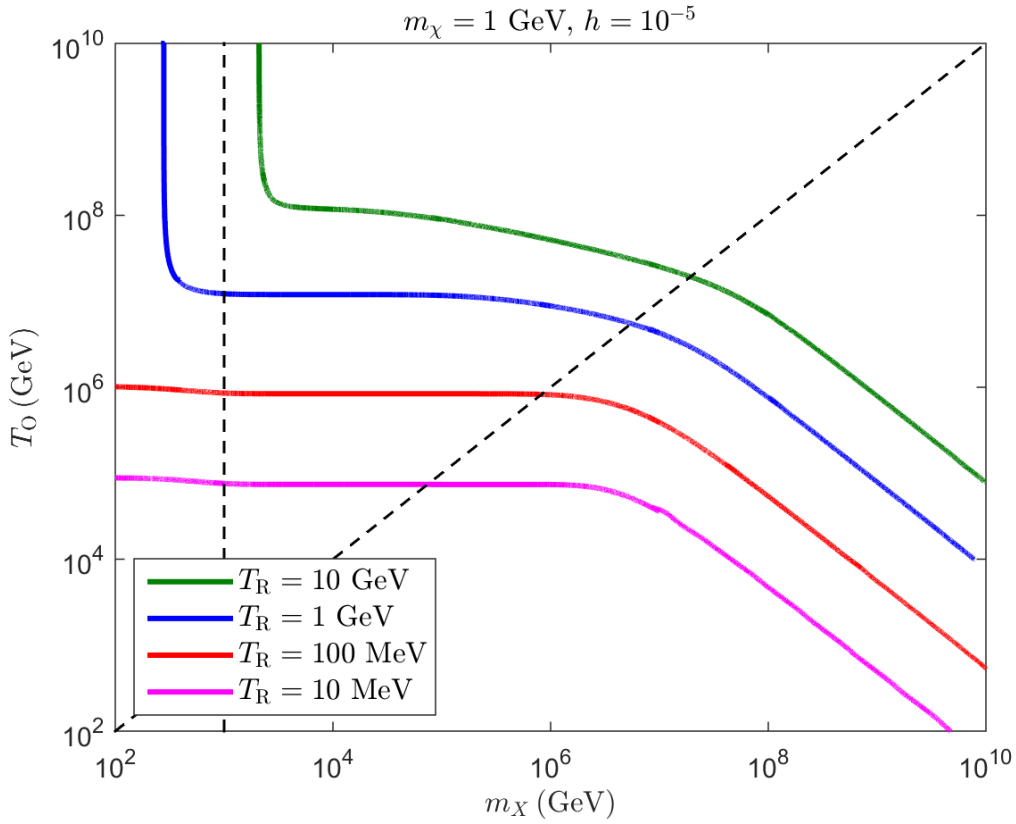


Figure 1: The $T_{\rm O}$-m_X plane for varied EMD end temperature $T_{\rm R}$ with other parameters held fixed. Solid contours show parameter combinations that achieve the observed DM relic abundance. The region above (below) a given contour corresponds to underproduction (overproduction) and is therefore allowed (constrained). The diagonal dashed line shows $T_{\rm O} = m_X$ for reference, while the vertical dashed line sits at $m_X = 1$ TeV.

A general feature seen in all figures is that each contour starts as a vertical line at sufficiently small m_X. Along the vertical segments the bulk of DM production takes place in the nonadiabatic phase corresponding to $m_X \ll T_{\rm tr}$ where, see Eq. (20), the DM relic abundance is independent from $T_{\rm O}$. The vertical segment starts to turn at $m_X \sim T_{\rm tr}$. We note that in Fig. 2 the turning points are stacked on top of each other around values of m_X within the same order of magnitude (unlike Fig. 1). This is because $T_{\rm tr}$ is much more sensitive to $T_{\rm R}$ than $T_{\rm O}$ as can be seen in Eq. (8). Since $T_{\rm R}$ is fixed in Fig. 2, the change in $T_{\rm tr}$ for different contours is rather minimal, while this is not the case in Fig. 1. The more significant dependence of $T_{\rm tr}$ on $T_{\rm R}$ also explains the apparently missing vertical segment in the contours for $T_{\rm R} = 10$ MeV and 100 MeV in Fig. 1. In this case, $T_{\rm tr}$ is smaller than the range of m_X shown in that figure.

Another visible feature in all figures is the presence of a plateau beyond $m_X \sim T_{\rm tr}$ where the contours continue as a horizontal line for some range of m_X. Along this segment, DM production from X decay occurs in the adiabatic phase. According to Eq. (21), the relic abundance has no dependence on m_X as long as $F(\gamma_\chi) = 1$ (happening for $\gamma_\chi \gtrsim 1$). Note

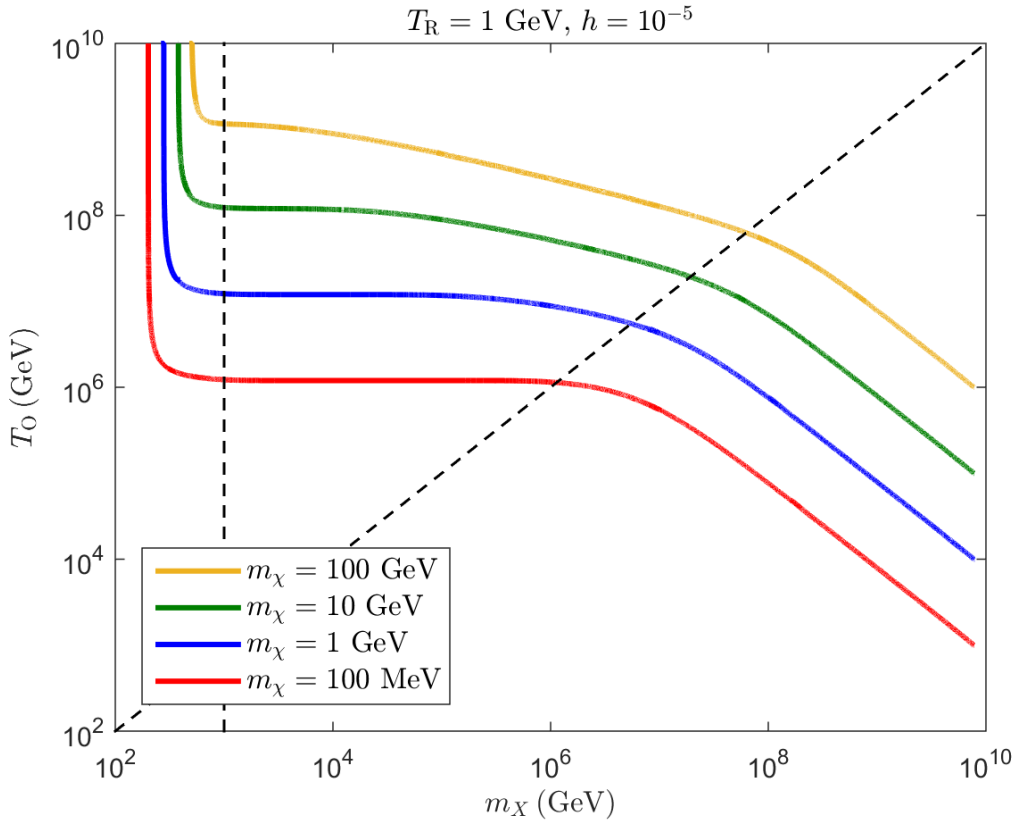


Figure 2: The $T_{\rm O}$-m_X plane for varied DM mass m_χ with other parameters held fixed. As in Fig. 1, solid contours correspond to the observed DM relic abundance, with the region above (below) a given contour being allowed (constrained).

that in the adiabatic phase, see Eq. (7), $\gamma_\chi \propto (m_X T_{\rm O})^{-1/2}$. This implies that for sufficiently small values of $T_{\rm O}$, we have $\gamma_\chi \gtrsim 1$ all the way up to $m_X = T_{\rm O}$. Assuming a RD phase prior to EMD, $\gamma_\chi \propto m_X^{-1}$ for $m_X > T_{\rm O}$. It will therefore drop below one at some value of m_X, beyond which DM production is not in equilibrium and the contour has a negative slope. This behavior is seen for contours with smaller values of $T_{\rm R}$ (m_χ) in Fig. 1 (Fig. 2). Increasing $T_{\rm O}$ results in a smaller value for γ_χ during the EMD epoch, in which case DM production can drop out of equilibrium during the adiabatic phase of EMD. The contour then starts to fall at $m_X < T_{\rm O}$ and keeps doing so for $m_X > T_{\rm O}$ (but with a different slope due to the different dependence of γ_χ on m_X). This is the behavior seen for contours with larger values of $T_{\rm R}$ (m_χ) in Fig. 1 (Fig. 2).

An additional feature seen in Fig. 3 is the asymptotic behavior of the contours as h increases. In fact, the flat horizontal segment of contours with sufficiently large h starts at the same $T_{\rm O}$ and m_X. This is because once $\gamma_\chi \sim \mathcal{O}(1)$ at $T_{\rm tr}$, χ reaches thermal equilibrium during the adiabatic phase of EMD. Further increasing h only extends the plateau toward larger values of m_X when γ_χ falls below one again.

Given that the region below and to the left of each contour is ruled out as it gives rise to DM overproduction, we can draw some important conclusions about the allowed parts of

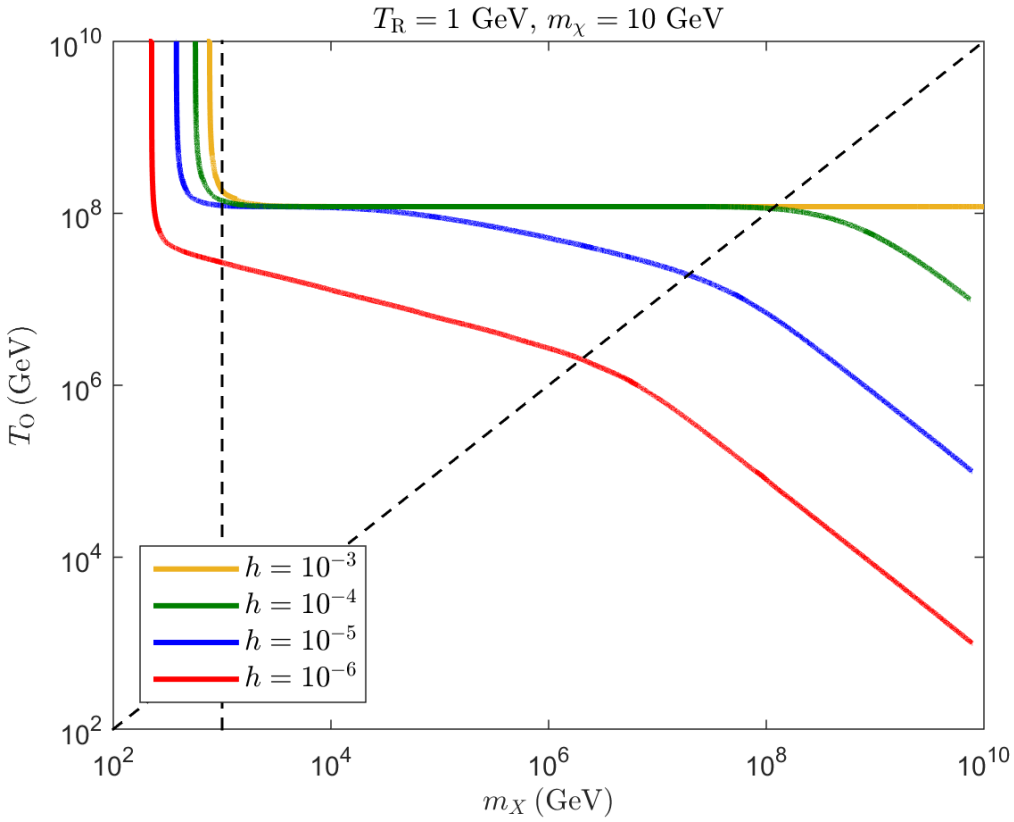


Figure 3: The $T_{\rm O}$-m_X plane for varied coupling h with other parameters held fixed. As in Fig. 1, solid contours correspond to the observed DM relic abundance, with the region above (below) a given contour being allowed (constrained).

the $T_{\rm O} - m_X$ plane:

• The constraint $m_X \gtrsim \mathcal{O}(\text{TeV})$ does not severely restrict regions with DM underproduction unless $T_{\rm R} \ll 1$ GeV and/or $h \ll 10^{-5}$. It mainly affects regions where DM is overproduced, which are already ruled out.

• The parameter space significantly opens up in the lower half-plane $T_{\rm O} < m_X$. However, in this part the main contribution to the DM relic abundance comes from the pre-EMD phase that requires detailed knowledge of the postinflationary thermal history.

• Cases with $m_\chi > \mathcal{O}(\text{TeV})$ are essentially ruled out unless $T_{\rm O}$ and m_X are (well) above 10^{10} GeV (see Fig. 2). This is only marginally possible in models with large $T_{\rm O}$ (like modulus-driven EMD), and not viable in models with intermediate values of $T_{\rm O}$, see Eq. (16).

We would like to note that throughout the allowed parameter space shown in Figs. 1,2. the expressions given in Eqs. (3,4) result in $\langle\sigma_{\rm ann} v\rangle \lesssim 10^{-43}$ cm^3 s^{-1}. The most extreme case in Fig. 3, corresponding to $h = 10^{-3}$ and $m_X = 1$ TeV, results in $\langle\sigma_{\rm ann} v\rangle \simeq 10^{-37}$ cm^3 s^{-1}. This is still small enough to be in the freeze-in regime [20]. It also results in an annihilation rate for χ that satisfies $\Gamma_{\rm ann} \ll H$ at temperatures $T \lesssim m_\chi$, which ensures that the comoving number density of χ remains frozen upon production from X decay. In

Appendix B, we discuss DM production from inverse annihilations and find that in large regions of the parameter space it can be completely neglected. In cases with $m_\chi < T_{\rm R}$ and $m_X \gg T_{\rm tr}$, this contribution becomes significant and may even give rise to the correct abundance while X decays lead to DM underproduction.

4.1 Scenarios with hidden sector DM

In this section, we consider the case where X and χ are part of a hidden sector with its own gauge symmetry and temperature that can be distinct from those of the visible sector. The Boltzmann equations in this case are slightly modified to include the radiation energy density of the hidden sector:

$$\begin{aligned} \frac{d\rho_{\rm r}^{\rm VS}}{dt} + 4H\rho_{\rm r}^{\rm VS} &= {\rm Br}_{\phi\to{\rm VS}}\Gamma_\phi\rho_\phi\,, \\ \frac{d\rho_{\rm r}^{\rm HS}}{dt} + 4H\rho_{\rm r}^{\rm HS} &= (1-{\rm Br}_{\phi\to{\rm VS}})\,\Gamma_\phi\rho_\phi\,, \end{aligned} \tag{25}$$

where superscripts denote the relevant sector and ${\rm Br}_{\phi\to{\rm VS}}$ is the branching fraction for ϕ decays to the visible sector. In order to guarantee a RD Universe that is compatible with observations, we need to have $\rho_{\rm r}^{\rm VS} \gg \rho_{\rm r}^{\rm HS}$ at the end of EMD. This implies that ϕ must decay predominantly to the visible sector, and hence ${\rm Br}_{\phi\to{\rm VS}} \simeq 1$.

There are two major differences between this case and that considered in the previous sections:

(1) m_X does not need to be restricted to being larger than $\mathcal{O}$(TeV). This is because X is a hidden-sector particle, and hence it can easily evade the LHC bounds. All that is required for our analysis to be valid in this case is that $m_X \gg m_\chi$.

(2) Given that ϕ decay must mainly reheat the visible sector, the amount of radiation injected to the hidden sector does not necessarily dominate over the initial radiation therein. As a result, the nonadiabatic phase of EMD may be shortened or even non-existent as far as the hidden sector is concerned.

In order to solve the Boltzmann equations in Eq. (25), we need to set the initial values of $\rho_{\rm r}^{\rm HS}$ and $\rho_{\rm r}^{\rm VS}$ as well as the value of ${\rm Br}_{\phi\to{\rm VS}}$. Here, we consider a case where radiation is almost entirely in the hidden sector prior to EMD. This, for example, is the case in the model of high scale SUSY discussed in [34, 35]. We also use the current observational limits from PLANCK 2018 on dark radiation in the form of $\Delta N_{\rm eff}$ using the full TT,TE,EE+lowE+lensing+BAO data [36] to set a lower bound on ${\rm Br}_{\phi\to{\rm VS}}$. These conditions yield a maximally constraining case for the configuration of the HS in that it is initially dominant and reheated to the maximum amount allowed by observations. An initially subdominant HS, as well as a shorter or even absent nonadiabatic phase, results in weaker constrains in the $T_{\rm O}$-m_X plane.

In Fig. 4 we show contours for different values of m_χ that yield the correct DM relic abundance for the benchmark with $T_{\rm R} = 1$ GeV and $h = 10^{-5}$. We note that the vertical segments of the contours have shifted to the left, compared with Fig. 1, or completely disappeared. This is a consequence of having ${\rm Br}_{\phi\to{\rm HS}} \equiv 1 - {\rm Br}_{\phi\to{\rm VS}} \ll 1$. Combined with the removal of the $m_X \gtrsim \mathcal{O}$(TeV) constraint, this results in a significant opening up of the allowed parameter space for small values of m_X in comparison with Fig. 1.

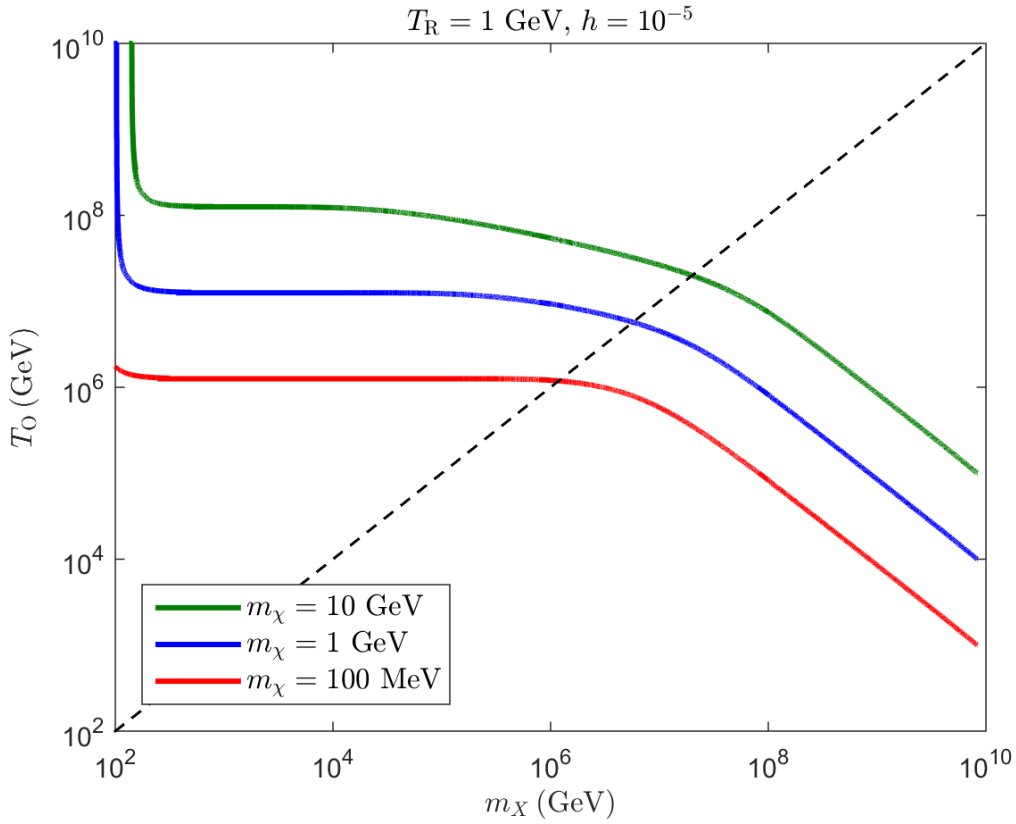


Figure 4: The $T_{\rm O}$-m_X plane for varied DM mass m_χ, with other parameters held fixed, for the case that X and χ are in a hidden sector. The hidden sector is taken to be dominant in the RD period before EMD, and is partially reheated during EMD to the amount allowed by PLANCK 2018 constraints on dark radiation from $\Delta N_{\rm eff}$.

5 Conclusion

We presented a detailed study of mediator production and decay and their contribution to the DM relic abundance in scenarios with an epoch of EMD. Special attention was paid to mediators that are charged under the SM gauge group and decay during early stages of EMD (or before its onset).

We showed that decay of on-shell mediators can totally dominate over the standard freeze-in contribution from inverse annihilations that happen at much lower temperatures. The requirement of not overproducing DM then leads to stringent constraints on the parameter space, demonstrated in Figs. 1-3, for DM masses above a few TeV. Additional information on the scale of inflation $H_{\rm inf}$, or an upper bound on it, will further restrict the allowed parameter space in favor of small DM masses.

The situation will be more relaxed if the DM candidate and mediators belong to a hidden sector with its own gauge symmetries. The absence of collider bounds on the mediator mass combined with a subdominant reheating of the hidden sector (as required by observational bounds) opens up new regions of the parameter space in this case as seen in Fig. 4.

While mentioning some specific examples of cosmological histories with EMD and DM

candidates, we remained largely agnostic about the particle physics origin of DM and the underlying model governing the cosmological history in this work. Our results can therefore be used by both phenomenologists and early Universe model builders to constrain DM models for a given thermal history and vice versa. They can also be used to pinpoint the parameters of a top-down high energy physics model that describes evolution of the early Universe from inflation to BBN and explains DM.

Acknowledgments

The work of R.A. and N.P.D.L. is supported in part by NSF Grant No. PHY-2210367. N.P.D.L. also acknowledges financial support from the UNM Department of Physics and Astronomy through the Origins of the Universe Award. J.K.O. is supported by the project AstroCeNT: Particle Astrophysics Science and Technology Centre, carried out within the International Research Agendas programme of the Foundation for Polish Science financed by the European Union under the European Regional Development Fund. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452.

A DM abundance from X decay

To estimate the final abundance of χ particles produced by decay of X during EMD, we consider the redshifted number density of χ from the time that its comoving value becomes frozen, marked by $T = T_{\rm f}$, to the end of EMD when $T = T_{\rm R}$, after which point entropy injection ceases:

$$n_\chi(T_{\rm R}) = n_\chi(T_{\rm f}) \left(\frac{a_{\rm f}}{a_{\rm R}}\right)^3 = n_\chi(T_{\rm f}) \left(\frac{H_{\rm R}}{H_{\rm f}}\right)^2 , \tag{26}$$

with

$$n_\chi(T_{\rm f}) = F(\gamma_\chi) n_\chi^{\rm eq}(T_{\rm f}) = F(\gamma_\chi) \frac{\zeta(3)}{\pi^2} g_\chi T_{\rm f}^3 , \tag{27}$$

assuming $T \gg m_\chi$. Taking $T_{\rm R}$ to be when $\rho_\phi = \rho_{\rm r}$ at the end of EMD, to better match numerical results, we have

$$H_{\rm R} = \left(\frac{\pi^2 g_{*\rm R}}{45}\right)^{1/2} \frac{T_{\rm R}^2}{M_{\rm P}} . \tag{28}$$

(1) $T_{\rm tr} < T_{\rm f} < T_{\rm O}$ - In this case the comoving χ number density freezes during the adiabatic phase and the Hubble rate at that time is therefore given by:

$$H_{\rm f} = \left(\frac{\pi^2}{45} g_{*\rm f}\right)^{1/2} \frac{T_{\rm f}^{3/2} T_{\rm O}^{1/2}}{M_{\rm P}} , \tag{29}$$

where we have used a factor of 2 for $\rho_\phi(T_{\rm O}) = \rho_{\rm r}(T_{\rm O})$, and $g_* T^3 \propto a^3$. After normalizing by the entropy density at $T_{\rm R}$, this gives:

$$\left(\frac{n_\chi}{s}\right)_{\rm dec}^{\rm A} \approx 0.28 F(\gamma_\chi) \frac{g_\chi}{g_{*\rm f}} \left(\frac{T_{\rm R}}{T_{\rm O}}\right) . \tag{30}$$

(2) $T_{\rm R} < T_{\rm f} < T_{\rm tr}$ - The number density freezes during the nonadiabatic phase and the corresponding Hubble rate is given by:

$$H_{\rm f} = \frac{1}{\sqrt{6}} \left(\frac{5\pi g_{*\rm f}}{6\sqrt{10} g_{*\rm R}^{1/2}} \right) \frac{T_{\rm f}^4}{T_{\rm R}^2 M_{\rm P}} \,, \tag{31}$$

where we have taken $\Gamma_\phi = \sqrt{3} H_{\rm R}$ to match our numerical results. After normalization, this gives:

$$\left(\frac{n_\chi}{s} \right)_{\rm dec}^{\rm NA} \approx 0.53 F(\gamma_\chi) \frac{g_\chi g_{*\rm R}}{g_{*\rm f}^2} \left(\frac{T_{\rm R}}{T_{\rm f}} \right)^5 . \tag{32}$$

B Production from inverse annihilations

Freeze-in production of DM from inverse annihilations can happen in three different regimes:

(1) $m_\chi > T_{\rm tr}$. In this regime, inverse annihilations during the adiabatic phase of the EMD epoch (or whatever phase precedes it) will be the dominant source. In fact, as shown in [24], the main contribution comes from the highest temperature at which the expressions in Eqs. (3,4) remain valid, namely $T \sim m_X$. The relic abundance from inverse annihilations with $m_X < T_{\rm O}$ then follows [14, 24]

$$\left(\frac{n_\chi}{s} \right)_{\rm ann}^{\rm A} \sim 10^{-1} \frac{g_\chi}{g_{*m_X}^{3/2}} \langle \sigma_{\rm ann} v \rangle (E \sim m_X) M_{\rm P} T_{\rm R} \left(\frac{m_X}{T_{\rm O}} \right)^{3/2} . \tag{33}$$

If instead $T_{\rm O} < m_X$, one can obtain a similar expression based on the type of dominant energy density before EMD.

(2) $T_{\rm R} < m_\chi \lesssim T_{\rm tr}$. In this case, the main contribution from inverse annihilations arises in the nonadiabatic phase of EMD (at $T \sim m_\chi/4$) giving rise to the following (for example, see [20]):

$$\left(\frac{n_\chi}{s} \right)_{\rm ann}^{\rm NA} \sim \frac{g_{*\rm R}^{3/2} g_\chi}{g_{*\rm m_\chi/4}^3} \langle \sigma_{\rm ann} v \rangle (E \sim m_\chi/4) M_{\rm P} T_{\rm R} \left(\frac{T_{\rm R}}{m_\chi} \right)^6 . \tag{34}$$

(3) $m_\chi \lesssim T_{\rm R}$. In this case, production from inverse annihilations is most efficient in the RD phase after EMD. The abundance is set at $T \sim T_{\rm R}$. and can be estimated as:

$$\left(\frac{n_\chi}{s} \right)_{\rm ann}^{\rm RD} \sim 10^{-1} \frac{g_\chi}{g_{*\rm R}^{3/2}} \langle \sigma_{\rm ann} v \rangle (E \sim T_{\rm R}) M_{\rm P} T_{\rm R} \,. \tag{35}$$

Note that the value of $\langle \sigma_{\rm ann} v \rangle$ at the relevant temperature in each case should be used, which is important when it is not a constant, see Eq. (4).

The values of m_X, m_χ, $T_{\rm R}$, and $T_{\rm O}$ determine the relevant regimes for DM production from inverse annihilations and X decay (discussed in Appendix A). To compare the contribution of inverse annihilations with that from X decay, we express $\langle \sigma_{\rm ann} v \rangle$ in terms of the particle-physics parameters using Eqs. (3,4) as an example. Inverse annihilations become most

pronounced in the regime where $m_\chi < T_{\rm R}$ and $m_X > T_{\rm tr}$. In this case, the number density of DM particles produced from X decay has undergone the highest redshift by the time inverse annihilations become efficient. For the parameter combinations shown in Figs. 1-3, we find that inverse annihilations contribute more than the observed DM abundance only in the regions where decay production also overproduces DM, with one exception. For $h = 10^{-3}$ in Fig. 3, inverse annihilations in the nonadiabatic phase lead to $\Omega_\chi h^2 \sim 0.1$ for $m_X \approx 1$ TeV and large values of $T_{\rm O}$, essentially coinciding with the contour from decay production. For larger values of m_X the abundance from inverse annihilation decreases.

References

[1] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279 [hep-ph/0404175].

[2] M. Kamionkowski and M.S. Turner, *THERMAL RELICS: DO WE KNOW THEIR ABUNDANCES?*, *Phys. Rev. D* **42** (1990) 3310.

[3] R. Allahverdi et al., *The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe*, 2006.16182.

[4] G. Kane, K. Sinha and S. Watson, *Cosmological Moduli and the Post-Inflationary Universe: A Critical Review*, *Int. J. Mod. Phys. D* **24** (2015) 1530022 [1502.07746].

[5] B.S. Acharya, P. Kumar, K. Bobkov, G. Kane, J. Shao and S. Watson, *Non-thermal Dark Matter and the Moduli Problem in String Frameworks*, *JHEP* **06** (2008) 064 [0804.0863].

[6] R. Allahverdi, M. Cicoli, B. Dutta and K. Sinha, *Nonthermal dark matter in string compactifications*, *Phys. Rev. D* **88** (2013) 095015 [1307.5086].

[7] R. Allahverdi, M. Cicoli, B. Dutta and K. Sinha, *Correlation between Dark Matter and Dark Radiation in String Compactifications*, *JCAP* **10** (2014) 002 [1401.4364].

[8] M. Cicoli, K. Dutta, A. Maharana and F. Quevedo, *Moduli Vacuum Misalignment and Precise Predictions in String Inflation*, *JCAP* **08** (2016) 006 [1604.08512].

[9] J.A. Dror, E. Kuflik and W.H. Ng, *Codecaying Dark Matter*, *Phys. Rev. Lett.* **117** (2016) 211801 [1607.03110].

[10] A. Berlin, D. Hooper and G. Krnjaic, *Thermal Dark Matter From A Highly Decoupled Sector*, *Phys. Rev. D* **94** (2016) 095019 [1609.02555].

[11] J.A. Dror, E. Kuflik, B. Melcher and S. Watson, *Concentrated dark matter: Enhanced small-scale structure from codecaying dark matter*, *Phys. Rev. D* **97** (2018) 063524 [1711.04773].

[12] M. Cirelli, Y. Gouttenoire, K. Petraki and F. Sala, *Homeopathic Dark Matter, or how diluted heavy substances produce high energy cosmic rays*, *JCAP* **02** (2019) 014 [1811.03608].

[13] R. Allahverdi and J.K. Osiński, *Early matter domination from long-lived particles in the visible sector*, *Phys. Rev. D* **105** (2022) 023502 [2108.13136].

[14] R. Allahverdi, N.P.D. Loc and J.K. Osiński, *Dark matter and baryogenesis from visible-sector long-lived particles*, *Phys. Rev. D* **107** (2023) 123510 [2212.11303].

[15] Fermi-LAT collaboration, *Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi Large Area Telescope Data*, *Phys. Rev. Lett.* **115** (2015) 231301 [1503.02641].

[16] Fermi-LAT, DES collaboration, *Searching for Dark Matter Annihilation in Recently Discovered Milky Way Satellites with Fermi-LAT*, *Astrophys. J.* **834** (2017) 110 [1611.03184].

[17] R.K. Leane, T.R. Slatyer, J.F. Beacom and K.C.Y. Ng, *GeV-scale thermal WIMPs: Not even slightly ruled out*, *Phys. Rev. D* **98** (2018) 023016 [1805.10305].

[18] L.J. Hall, K. Jedamzik, J. March-Russell and S.M. West, *Freeze-In Production of FIMP Dark Matter*, *JHEP* **03** (2010) 080 [0911.1120].

[19] G.F. Giudice, E.W. Kolb and A. Riotto, *Largest temperature of the radiation era and its cosmological implications*, *Phys. Rev. D* **64** (2001) 023508 [hep-ph/0005123].

[20] A.L. Erickcek, *The Dark Matter Annihilation Boost from Low-Temperature Reheating*, *Phys. Rev. D* **92** (2015) 103505 [1504.03335].

[21] L. Calibbi, F. D'Eramo, S. Junius, L. Lopez-Honorez and A. Mariotti, *Displaced new physics at colliders and the early universe before its first second*, *JHEP* **05** (2021) 234 [2102.06221].

[22] A. Ahmed, B. Grzadkowski and A. Socha, *Higgs boson induced reheating and ultraviolet frozen-in dark matter*, *JHEP* **02** (2023) 196 [2207.11218].

[23] N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen and V. Vaskonen, *The Dawn of FIMP Dark Matter: A Review of Models and Constraints*, *Int. J. Mod. Phys. A* **32** (2017) 1730023 [1706.07442].

[24] R. Allahverdi and J.K. Osiński, *Freeze-in Production of Dark Matter Prior to Early Matter Domination*, *Phys. Rev. D* **101** (2020) 063503 [1909.01457].

[25] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195 [hep-ph/9506380].

[26] R.N. Mohapatra and R.E. Marshak, *Local B-L Symmetry of Electroweak Interactions, Majorana Neutrinos and Neutron Oscillations*, *Phys. Rev. Lett.* **44** (1980) 1316.

[27] Z.M. Burell and N. Okada, *Supersymmetric minimal B-L model at the TeV scale with right-handed Majorana neutrino dark matter*, *Phys. Rev. D* **85** (2012) 055011 [1111.1789].

[28] R. Allahverdi and B. Dutta, *Natural GeV Dark Matter and the Baryon-Dark Matter Coincidence Puzzle*, *Phys. Rev. D* **88** (2013) 023525 [1304.0711].

[29] F. D'Eramo, R.Z. Ferreira, A. Notari and J.L. Bernal, *Hot Axions and the H_0 tension*, *JCAP* **11** (2018) 014 [1808.07430].

[30] Y. Du, F. Huang, H.-L. Li, Y.-Z. Li and J.-H. Yu, *Revisiting dark matter freeze-in and freeze-out through phase-space distribution*, *JCAP* **04** (2022) 012 [2111.01267].

[31] R. Allahverdi, R. Brandenberger, F.-Y. Cyr-Racine and A. Mazumdar, *Reheating in Inflationary Cosmology: Theory and Applications*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 27 [1001.2600].

[32] M.A. Amin, M.P. Hertzberg, D.I. Kaiser and J. Karouby, *Nonperturbative Dynamics Of Reheating After Inflation: A Review*, *Int. J. Mod. Phys. D* **24** (2014) 1530003 [1410.3808].

[33] Planck collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [1807.06211].

[34] R. Allahverdi, I. Broeckel, M. Cicoli and J.K. Osiński, *Superheavy dark matter from string theory*, *JHEP* **02** (2021) 026 [2010.03573].

[35] R. Allahverdi, C. Arina, M. Chianese, M. Cicoli, F. Maltoni, D. Massaro et al., *Phenomenology of superheavy decaying dark matter from string theory*, 2312.00136.

[36] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

arXiv:2402.18547v1 [hep-ph] 28 Feb 2024

Minimal seesaw and leptogenesis with the smallest modular finite group

Simone Marciano,[†] Davide Meloni,[‡] Matteo Parriciatu,[§]

Dipartimento di Matematica e Fisica, Università di Roma Tre
INFN Sezione di Roma Tre, Via della Vasca Navale 84, 00146, Roma, Italy

Abstract

We propose a model for leptons based on the smallest modular finite group $\Gamma_2 \cong S_3$ that, for the first time, accounts for both the hints of large low-energy CP-violation in the lepton sector and the matter-antimatter asymmetry of the Universe, generated by only two heavy right-handed neutrinos. These same states are also employed in a Minimal seesaw mechanism to generate light neutrino masses. Besides the heavy neutrinos, our particle content is the same as the Standard Model (SM), with the addition of one single modulus τ, whose vacuum expectation value is responsible for both the modular and CP-symmetry breakings. We show that this minimalistic SM extension is enough to get an excellent fit to low energy neutrino observables and to the required baryon asymmetry η_B. Predictions for the neutrino mass ordering, effective masses in neutrinoless double beta decay and tritium decay as well as for the Majorana phases are also provided.

Contents

1 Introduction 2

2 Modular flavor symmetry at level 2 3

3 The model 6

4 Numerical analysis and results 9

5 Leptogenesis 14

6 Conclusions 21

[†]e-mail address: simone.marciano@uniroma3.it
[‡]e-mail address: davide.meloni@uniroma3.it
[§]e-mail address: matteo.parriciatu@uniroma3.it

A Beyond the N_1−dominated scenario **22**

B Modular forms and S_3 representations **24**

C Mapping to $\mathcal{D}$ and rescaling **25**

1 Introduction

The flavor structure of the Standard Model (SM) is still an outstanding puzzle in theoretical particle physics. The quite pronounced mass hierarchies between the three generations of charged fermions in both quark and lepton sectors are not addressed by the SM gauge symmetry, while the smallness of the neutrino mass scale calls for an entirely different mechanism beyond the SM. The mixing of the generations and CP-violation are also completely unconstrained by gauge symmetry. As we enter the precision era of neutrino experiments, we face the theoretical challenge to derive at least a major portion of the free parameters from an organizing principle. In particular, the peculiar mixing pattern of leptons (characterized by two large and one small angles), together with the recent hints of CP violation in neutrino oscillations could be understood with discrete symmetries through non-Abelian groups [1]. However, this approach proved to be quite unsatisfactory due to the unconstrained number of free parameters, and to the complicated scalar sector equipped with a substantial number of flavons (spurion fields) which are needed to reproduce the observed mixing [2]. A step forward was made by Feruglio in 2017 [3] where modular symmetry takes the role of a flavor discrete symmetry in a bottom-up approach. The Yukawa couplings of the SM become modular forms: pre-determined functions of a complex field τ called "modulus", which describes the geometry of compactified extra-dimensions in superstring theory [4, 5]. As a result, models based on modular symmetry are in general more predictive as they require only a limited number of free parameters, and the flavon vacuum alignment problem is replaced by the moduli stabilization, at least in top-down approaches [6–9]. Both the usual matter superfields and Yukawa couplings transform in irreducible representations of the modular finite group Γ_N of a given level $N \in \mathbb{N}$. The appealing feature is that for $N \leq 5$ these groups are isomorphic to the more familiar non-Abelian discrete groups S_3, A_4, S_4 and A_5. As opposed to traditional flavor models based on those groups, in modular symmetry a generic vacuum expectation value (VEV) of the modulus breaks the modular group completely, and the observables of the lepton sector are completely determined by the value of this VEV, up to a limited number of free parameters in the superpotential. Currently, in the literature there is a substantial number of phenomenologically viable modular models for the lepton sector, some examples are given by: $\Gamma_2 \cong S_3$ [10,11], $\Gamma_3 \cong A_4$ [3,10,12–31], $\Gamma_4 \cong S_4$ [20, 32–40], $\Gamma_5 \cong A_5$ [37, 41, 42]. Models based on the double-cover Γ'_N of the modular finite groups have also been put forward, some examples are [43–50]. For a more exhaustive list of references, see the recent reviews [51, 52]. In addition to the flavor puzzle, the source of the asymmetry between matter and antimatter in the early universe remains an open issue. In this regard, once a seesaw mechanism is introduced to give mass to neutrinos, leptogenesis [53–58]

becomes almost inevitable. Indeed, Sakharov's conditions [59] can be naturally satisfied in this context; in fact, the seesaw mechanism predicts a violation of lepton number through processes involving Majorana neutrinos, as well as complex Yukawa couplings that induce CP violation. In addition, for a wide range of parameter values, these processes can occur out of thermal equilibrium.

In the supersymmetric modular symmetry business, a bunch of papers have already discussed the issues related to leptogenesis; in particular, models based on $\Gamma_3 \cong A_4$ have been presented in [19,27,60–66], those on $\Gamma_4 \cong S_4$ in [38,40] while the double covering group $\Gamma'_5 \cong A'_5$ has been discussed in [67].

In this paper, we contribute to the subject presenting a modular model based on S_3 capable of explaining the low-energy lepton observables and CP-violation as well as to account for the correct amount of the baryon asymmetry of the Universe (BAU), η_B. Our construction is "minimal" in the sense that only two right-handed neutrinos (RN) are added to the particle content of the SM and that the unique modulus τ generates both the low energy CP-violating phase and the high energy CP violation needed for the leptogenesis mechanism to account for the BAU, without the addition of any other complex phases in the Yukawa matrices. Our results for η_B are obtained solving the appropriate Boltzmann equations in the N_1-dominated scenario, that we discuss in detail. A number of interesting predictions can be drawn from our model, and are related to the effective masses in the neutrinoless double beta decay and tritium decay (all below the current experimental values), as well as to the values of the Majorana phases. The remaining sections of this paper are organized as follows: in section 2 we recall the main ingredients of the modular symmetry and its use in the supersymmetric (SUSY) context; section 3 is devoted to the analytic description of our model, which is subsequently analyzed numerically in section 4; the relevant part of the present manuscript related to leptogenesis is carefully described in section 5; in 6 we draw our conclusions. Three appendices are supplied: in appendix A we discuss the solution of the Boltzmann equations in presence of two sterile neutrinos and justify the adopted N_1-dominated scenario; in appendix B we report our choice for the arbitrary normalization of the level $N = 2$ modular forms; finally, in the last appendix C we show how to map the points outside the fundamental domain into points belonging to it.

2 Modular flavor symmetry at level 2

The modular group $\overline{\Gamma} = \mathrm{SL}(2,\mathbb{Z})/\{\pm\mathbb{1}\}$ acts on the modulus τ, restricted to the upper-half complex plane, through the transformation $\gamma : \tau \to \gamma(\tau)$,

$$\gamma(\tau) = \frac{a\tau + b}{c\tau + d} \quad , \quad a, b, c, d \in \mathbb{Z} \quad , \quad ad - bc = 1 \,, \tag{2.1}$$

which is generated by S and T defined as:

$$\begin{cases} S : \tau \to -\dfrac{1}{\tau} \\ \\ T : \tau \to \tau + 1 \end{cases} \quad , \qquad S^2 = (ST)^3 = \mathbb{1} \,. \tag{2.2}$$

The generators (2.2) are represented in $\mathrm{SL}(2,\mathbb{Z})$ by the 2×2 matrices $S=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$, and $T=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$. A modular form is a holomorphic function of τ which, under $\overline{\Gamma}$, transforms as:

$$f(\gamma(\tau))=(c\tau+d)^k f(\tau)\,, \tag{2.3}$$

where k is a positive integer called "*weight*".
While the group $\overline{\Gamma}$ is infinite and non-compact, compact and finite groups can be constructed from its infinite normal subgroups $\Gamma(N)$ for $N=1,2,3...$ defined as:

$$\Gamma(N)=\left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in \mathrm{SL}(2,\mathbb{Z})\,\middle|\,\begin{pmatrix}a & b\\ c & d\end{pmatrix}\equiv\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\,(\mathrm{mod}\ N)\right\}\,, \tag{2.4}$$

where the natural number N is called the *level*. In [3] Feruglio has shown that it is always possible to find a basis where modular forms of a given level transform in unitary representations of the finite groups $\Gamma_N\equiv\overline{\Gamma}/\overline{\Gamma}(N)$ which, for $N\leq 5$, are isomorphic to the non-Abelian discrete groups S_3, A_4, S_4, A_5. Here we focus on the lowest level $N=2$, which labels the smallest finite modular group $\Gamma_2\cong S_3$. As reviewed in [11], modular forms of level 2 and (even) weight k span a linear space of finite dimension $k/2+1$, and a basis of lowest weight forms for all modular forms of level 2 is provided by the S_3 doublet $Y_1(\tau), Y_2(\tau)$ given in appendix B. Those can be employed to construct modular forms of higher weights using the S_3 composition rules.

In what follows, we will use the convenient notation[1]:

$$Y^{(1)}_{\mathbf{2}}(\tau)\equiv\begin{pmatrix}Y_1(\tau)\\ Y_2(\tau)\end{pmatrix}_{\mathbf{2}}\,. \tag{2.5}$$

Making use of the composition rules listed in appendix B, the 3 linearly independent modular forms of weight 4 are constructed in the following way:

$$Y^{(1)}_{\mathbf{2}}(\tau)\otimes Y^{(1)}_{\mathbf{2}}(\tau)=(Y_1^2(\tau)+Y_2^2(\tau))_{\mathbf{1}}\oplus\begin{pmatrix}Y_2^2(\tau)-Y_1^2(\tau)\\ 2Y_1(\tau)Y_2(\tau)\end{pmatrix}_{\mathbf{2}}\,, \tag{2.6}$$

which, according to our notation, will be expressed as $Y^{(2)}_{\mathbf{1}}(\tau)$ and $Y^{(2)}_{\mathbf{2}}(\tau)$, respectively. Similarly, there are 4 linearly independent modular forms of weight 6:

$$\begin{aligned}Y^{(2)}_{\mathbf{2}}(\tau)\otimes Y^{(1)}_{\mathbf{2}}(\tau)=[3Y_1(\tau)Y_2^2(\tau)-Y_1^3(\tau)]_{\mathbf{1}}\oplus[Y_2^3(\tau)-3Y_2(\tau)Y_1^2(\tau)]_{\mathbf{1'}}\oplus\\ \oplus\begin{pmatrix}Y_1(\tau)(Y_1^2(\tau)+Y_2^2(\tau))\\ Y_2(\tau)(Y_1^2(\tau)+Y_2^2(\tau))\end{pmatrix}_{\mathbf{2}}\,,\end{aligned} \tag{2.7}$$

labelled in our notation as $Y^{(3)}_{\mathbf{1}}$, $Y^{(3)}_{\mathbf{1'}}$ and $Y^{(3)}_{\mathbf{2}}$, respectively. Since the number of independent forms of weight k grows with $k/2+1$, we stop at weight 6 to aim at a minimalistic number of

[1] In $Y^{(a)}_{\boldsymbol{\rho}}(\tau)$, the symbol $\boldsymbol{\rho}$ denotes the irrep, the upper index (a) signals how many times the basis $Y_1(\tau), Y_2(\tau)$ was contracted to form $Y^{(a)}_{\boldsymbol{\rho}}(\tau)$, thus having weight $2a$. Individual components of the multiplets are simply given by $Y^p_j(\tau)$, where $j=\{1,2\}$ and p is an exponent.

free parameters in the superpotential. One of the appealing features of modular invariance is that the value of the modulus can be restricted to a *fundamental* domain defined as:

$$\mathcal{D} = \left\{ \tau \in \mathbb{C} \,:\, \mathrm{Im}\,\tau > 0\,,\;\; |\mathrm{Re}\,\tau| \leq \frac{1}{2}\,,\;\; |\tau| \geq 1 \right\}. \tag{2.8}$$

This corresponds to the gray region $\mathcal{D}$ in figure C.1, appendix C. Values outside $\mathcal{D}$ are redundant and can be mapped inside (2.8) through the appropriate modular transformation (2.1). It is interesting to observe that, in particular for $\Gamma_2 \cong S_3$, the modular forms of lowest weight $Y_1(\tau), Y_2(\tau)$ are intrinsically hierarchical for $\tau \in \mathcal{D}$, see eq.(B.2); this allows us to define the appropriate ratio:

$$\frac{Y_2(\tau)}{Y_1(\tau)} \equiv \zeta = |\zeta|\, e^{i\,g}\,, \tag{2.9}$$

where $g = g(\mathrm{Re}\,\tau)$ is a real function depending on $\mathrm{Re}\,\tau$. The absolute value $|\zeta|$ satisfies $|\zeta| \lesssim 1$ for $\tau \in \mathcal{D}$ and is suppressed by $e^{-\pi \mathrm{Im}\,\tau}$, as shown in figure B.1. Thus, we can use $Y_2(\tau)/Y_1(\tau)$ as an expansion parameter, independently of the chosen normalization of the modular forms (more about that in appendix C).

2.1 The SUSY framework

The supersymmetric modular-invariant action, turning off gauge interactions for simplicity, is written as:

$$\mathcal{S} = \int d^4x \int d^2\theta d^2\bar{\theta}\, K(\Phi, \bar{\Phi}) + \left[\int d^4x \int d^2\theta\, \mathcal{W}(\Phi) + \mathrm{h.c.} \right], \tag{2.10}$$

where $\theta, \bar{\theta}$ are the superspace coordinates; $\Phi = (\tau, \varphi)$ are chiral superfields, and φ labels the usual matter supermultiplets; $K(\Phi, \bar{\Phi})$ is the Kähler potential and $\mathcal{W}(\Phi)$ is the superpotential. We assume that chiral superfields transform under $\overline{\Gamma}$ as:

$$\tau \to \gamma(\tau) = \frac{a\tau + b}{c\tau + d} \qquad , \qquad \varphi^{(I)} \to (c\tau + d)^{-k_I} \rho^{(I)}(\gamma) \varphi^{(I)}\,, \tag{2.11}$$

where $\gamma \in \overline{\Gamma}$ and $\rho^{(I)}(\gamma)$ is the matrix for the irreducible representation of Γ_N. Here the k_I of the superfields are called modular charges. The action (2.10) is invariant under (2.11) provided that the superpotential $\mathcal{W}(\Phi)$ and the Kähler potential transform as:

$$\begin{cases} \mathcal{W}(\Phi) \to \mathcal{W}(\Phi) \\ \\ K(\Phi, \bar{\Phi}) \to K(\Phi, \bar{\Phi}) + f(\Phi) + \bar{f}(\bar{\Phi}) \end{cases} \tag{2.12}$$

with $f(\Phi)$ being holomorphic. The superpotential $\mathcal{W}(\Phi)$ will generally be a combination of Yukawa couplings $Y_{I_1 \ldots I_n}(\tau)$ and chiral superfields

$$\mathcal{W}(\Phi) = \sum (Y_{I_1 \ldots I_n}(\tau)\, \varphi^{(I_1)} \ldots \varphi^{(I_n)})_{\mathbf{1}}\,. \tag{2.13}$$

The Yukawa couplings are assumed to be modular forms which transform as:

$$Y_{I_1\dots I_n}(\gamma(\tau)) = (c\tau + d)^{k_Y}\rho(\gamma)Y_{I_1\dots I_n}(\tau)\,, \tag{2.14}$$

with $k_Y \geq 2$ being an even integer in the case of $N = 2$. The novelty compared to classic non-Abelian discrete groups model building is that the invariance of (2.13) is achieved by satisfying not only the existence of a singlet contraction $\rho \otimes \rho_{I_1} \otimes \rho_{I_2} \dots \otimes \rho_{I_n} \supset \mathbf{1}$ between all the irreps involved, but also that every operator in (2.13) must be *weightless*, i.e. $k_Y = k_{I_1} + k_{I_2} + \dots + k_{I_n}$, the total weight of the matter fields must be counterbalanced by appropriate Yukawa modular forms of weight k_Y.
Finally, in most economical realizations of modular flavor models, the invariant Kähler potential is given by:

$$K(\Phi, \bar{\Phi}) = -h\Lambda_\tau^2 \log(-i\tau + i\bar{\tau}) + \sum_I (-i\tau + i\bar{\tau})^{-k_I} |\varphi^{(I)}|^2\,, \tag{2.15}$$

where $h > 0$ and Λ_τ has mass-dimensions of one. Note that we used the same symbol to denote the superfield and its scalar component.

In what follows, the vacuum expectation value (VEV) of the modulus τ is the only source of flavor symmetry breaking, thus no additional flavons are needed.[2] The scale of modular symmetry breaking can be located at the compactification scale, which is supposed to be near the Planck scale. The modular group is completely broken for every value of τ except for the so-called *symmetric* points $\tau_{\text{sym}} = \{\omega, i, i\infty\}$ where $\omega \equiv e^{2\pi i/3}$, which are left invariant under ST, S, and T respectively [33].

2.2 gCP symmetry

In order to reduce the number of model free parameters (thus possibly making the model itself more predictive), we assume a generalized CP symmetry (gCP). This was made consistent with modular symmetry in [68] where it was shown that under certain requirements for the group basis and the normalization of the Yukawa forms, imposing a gCP symmetry consists in the reality condition of the superpotential couplings.

Given this assumption, the only source of CP violation in gCP models is the VEV of the complex modulus. Since τ appears in the expansions (B.2) of appendix B through powers of $e^{2\pi i\tau}$, the CP violation is controlled entirely by $\mathrm{Re}\,\tau$ for $\tau \in \mathcal{D}$. This means that the only complex phase in the mass matrices is represented by powers of $e^{2\pi i\,\mathrm{Re}\,\tau}$. Apart from the obvious $\mathrm{Re}\,\tau = 0$, the other CP-conserving points consist in the boundary of $\mathcal{D}$. Note that all the previously mentioned symmetric points are CP-conserving.

3 The model

Given that our main concern here is to find a viable seesaw model for generating neutrino masses, for the charged-leptons sector we simply employ the transformation properties under

[2] Here we do not deal with the dynamical justification for this VEV, which is a subject of ongoing studies. See for examples [18], and [6–9].

Γ_2 and a charge assignment compatible with *naturally* hierarchical charged leptons, adopted from the construction illustrated in [11]. In this respect, the superfields E_i^c correspond to the three flavors of right-handed charged leptons (respectively $\{i = 1, 2, 3\} \equiv \{e, \mu, \tau\}$) with modular charges $4, 0$ and -2, respectively. The first two lightest left-handed $SU(2)_{\rm L}$ doublets are grouped into a doublet of S_3 while the heaviest family belongs to a pseudo-singlet. Finally, the Higgs doublets transform as invariant singlets with modular charges set to 0. In table 1 we summarize the chiral supermultiplets of our model, the transformation properties under $\Gamma_2 \cong S_3$ and the related modular charges. As a consequence, the allowed terms in the superpotential are the following:

$$\mathcal{W}_e^H = \alpha E_1^c H_d (D_\ell Y_{\mathbf{2}}^{(3)})_{\mathbf{1}} + \beta E_2^c H_d (D_\ell Y_{\mathbf{2}})_{\mathbf{1'}} + \gamma E_3^c H_d \ell_3 + \alpha_D E_1^c H_d \ell_3 Y_{\mathbf{1'}}^{(3)} \,, \tag{3.1}$$

and the resulting mass matrix (in the right-left basis) after electroweak symmetry breaking reads[3]:

$$M_\ell = \begin{pmatrix} \alpha (Y_{\mathbf{2}}^{(3)})_1 & \alpha (Y_{\mathbf{2}}^{(3)})_2 & \alpha_D Y_{\mathbf{1'}}^{(3)} \\ \beta Y_2 & -\beta Y_1 & 0 \\ 0 & 0 & \gamma \end{pmatrix}_{\rm RL} v_d \,, \tag{3.2}$$

where v_d is the VEV of H_d and the modular forms of weight 6 are given by (2.7). One can recast the matrix (3.2) in a form suitable for getting perturbative expressions of the eigenvalues; in order to do that, we define $A \equiv \alpha_D/\alpha$, $B \equiv (\beta/\alpha) Y_1^{-2}$, $C \equiv (\gamma/\alpha) Y_1^{-3}$ and ζ is the expansion parameter defined in (2.9). Thus, we get:

$$M_\ell = v_d \alpha Y_1^3 \begin{pmatrix} (1+\zeta^2) & \zeta(1+\zeta^2) & -A\zeta(3-\zeta^2) \\ B\zeta & -B & 0 \\ 0 & 0 & C \end{pmatrix}_{\rm RL} . \tag{3.3}$$

Assuming $\frac{\beta}{\alpha} \sim \frac{\gamma}{\alpha} \approx 1$ and using the q-expansions (B.2) we estimate $|B|, |C| \gg A$ for $\tau \in \mathcal{D}$. It is now easy to obtain the approximated eigenvalues in powers of ζ as follows:

$$m_e = v_d \alpha \left(|Y_1^3| + \frac{3}{2} |Y_1^3| |\zeta|^2 + \mathcal{O}(\zeta^3) \right) \tag{3.4}$$

$$m_\mu = v_d \alpha \left(|Y_1| + \frac{1}{2} |Y_1| |\zeta|^2 + \mathcal{O}(\zeta^3) \right) \tag{3.5}$$

$$m_\tau = v_d \alpha \left(1 + \frac{9A^2}{2} |Y_1^6| |\zeta|^2 + \mathcal{O}(\zeta^3) \right) . \tag{3.6}$$

The hierarchy $(m_\tau, m_\mu, m_e) \sim m_\tau (1, |Y_1|, |Y_1|^3)$ naturally arises considering that $|Y_1| \approx 7/100$. Given the non-diagonal structure of the charged-leptons mass matrix (3.3), a non-trivial contribution to the PMNS matrix arises; considering that $A/|C|^2 \ll |\zeta|$, the unitary matrix U_ℓ that diagonalises $M_\ell^\dagger M_\ell$ can be expressed as follows:

$$U_\ell \sim \begin{pmatrix} 1 - \frac{|\zeta|^2}{2} & -\overline{\zeta} & 0 \\ \zeta & 1 - \frac{|\zeta|^2}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} + \mathcal{O}(\zeta^3) \,, \tag{3.7}$$

[3] In the parentheses notation $(...)_{1,2}$ we denote the two components of the corresponding doublet $Y_{\mathbf{2}}^{(a)}$.

	E_1^c	E_2^c	E_3^c	D_ℓ	ℓ_3	$H_{d,u}$	N^c
$SU(2)_L \times U(1)_Y$	$(\mathbf{1}, +1)$	$(\mathbf{1}, +1)$	$(\mathbf{1}, +1)$	$(\mathbf{2}, -1/2)$	$(\mathbf{2}, -1/2)$	$(\mathbf{2}, \mp 1/2)$	$(\mathbf{1}, 0)$
$\Gamma_2 \cong S_3$	$\mathbf{1}$	$\mathbf{1'}$	$\mathbf{1'}$	$\mathbf{2}$	$\mathbf{1'}$	$\mathbf{1}$	$\mathbf{2}$
k_I	4	0	-2	2	2	0	2

Table 1: *Chiral supermultiplets, transformation properties under $\Gamma_2 \cong S_3$ and modular charges. For the sign convention of the modular charges, we refer to our choice in eq. (2.11).*

up to $\mathcal{O}(1)$ coefficients. In particular, we see that U_ℓ would provide a non-negligible correction to the solar angle, for which we expect only a small $\mathcal{O}(\zeta)$ contribution for $\mathrm{Im}\,\tau$ sufficiently large. All in all, the hierarchical texture of the charged-leptons is provided by the modular invariance through our construction. The mass spectrum is completely determined by five dimensionless real parameters, $\mathrm{Re}\,\tau$, $\mathrm{Im}\,\tau$, α_D/α, β/α and γ/α and one global scale $v_d\alpha$.

3.1 Neutrino sector

To keep the number of free parameters at minimum, we adopt the Minimal seesaw scenario [69, 70] consisting in the introduction of two SM singlets right-handed Majorana fields N_1^c, N_2^c. As listed in table 1, these are assigned to an S_3 doublet $N^c \sim \mathbf{2}$. The invariant superpotential of the neutrino sector is given by:

$$\begin{aligned}\mathcal{W}_\nu = g H_u N^c D_\ell Y_{\mathbf{2}}^{(2)} + g' H_u (N^c Y_{\mathbf{2}}^{(2)})_{\mathbf{1'}} \ell_3 + g'' H_u (N^c D_\ell)_{\mathbf{1}} Y_{\mathbf{1}}^{(2)} + \\ + \Lambda [(N^c N^c)_{\mathbf{2}} Y_{\mathbf{2}}^{(2)} + \lambda (N^c N^c)_{\mathbf{1}} Y_{\mathbf{1}}^{(2)}]\,,\end{aligned} \tag{3.8}$$

where g, g', g'', λ are dimensionless free parameters and Λ is the Majorana right-handed mass scale. As mentioned in section 2.2, all these couplings (as well as the ones from the charged-leptons sector) are real due to the imposed gCP symmetry. The only complex parameter is the modulus τ.

Exploiting the composition rules of S_3 given in appendix B, the Dirac mass matrix reads:

$$M_D = g v_u \begin{pmatrix} -(Y_2^2 - Y_1^2) + \dfrac{g''}{g}(Y_1^2 + Y_2^2) & 2Y_1 Y_2 & \dfrac{g'}{g}(2Y_1 Y_2) \\ 2Y_1 Y_2 & (Y_2^2 - Y_1^2) + \dfrac{g''}{g}(Y_1^2 + Y_2^2) & -\dfrac{g'}{g}(Y_2^2 - Y_1^2) \end{pmatrix}_{\mathrm{RL}}\,, \tag{3.9}$$

where v_u denotes the VEV of the neutral scalar component of H_u. The Majorana mass matrix of right-handed neutrinos is given by:

$$\mathcal{M}_R = \Lambda \begin{pmatrix} -(Y_2^2 - Y_1^2) + \lambda(Y_1^2 + Y_2^2) & 2Y_1 Y_2 \\ 2Y_1 Y_2 & (Y_2^2 - Y_1^2) + \lambda(Y_1^2 + Y_2^2) \end{pmatrix}_{\mathrm{RR}}\,. \tag{3.10}$$

The mass matrix of the light Majorana neutrinos is then obtained from the well known formula of type-I seesaw:

$$m_\nu = -M_D^T \mathcal{M}_R^{-1} M_D\,. \tag{3.11}$$

Parameter	Best-fit value and 1σ range	
	NO	**IO**
$r \equiv \Delta m^2_{\rm sol}/\lvert\Delta m^2_{\rm atm}\rvert$	0.0295 ± 0.0008	0.0298 ± 0.0008
$\sin^2\theta_{12}$	$0.303^{+0.012}_{-0.012}$	$0.303^{+0.012}_{-0.011}$
$\sin^2\theta_{13}$	$0.02225^{+0.00056}_{-0.00059}$	$0.0223^{+0.00058}_{-0.00058}$
$\sin^2\theta_{23}$	$0.451^{+0.019}_{-0.016}$	$0.569^{+0.016}_{-0.021}$
$J_{\rm CP}$	$-0.027^{+0.010}_{-0.010}$	$-0.032^{+0.007}_{-0.007}$
m_e/m_μ	0.0048 ± 0.0002	
m_μ/m_τ	0.0565 ± 0.0045	

Table 2: *Neutrino observables and their 1σ ranges, from NuFIT 5.2 [72] using the dataset with SK atmospheric data [73]. Here "Normal ordering" and "Inverted ordering" for the mass spectrum are indicated with **NO** and **IO**, respectively. The mass ratios of charged-leptons are taken from [3]. Here we defined $\Delta m^2_{sol} \equiv m_2^2 - m_1^2$, $|\Delta m^2_{atm}| \equiv |m_3^2 - (m_1^2+m_2^2)/2|$. The Jarlskog invariant is defined in (4.2) and can be expressed as $J_{CP} = c_{12}s_{12}c_{23}s_{23}c_{13}^2 s_{13}\sin\delta_{CP}$. To extract its best-fit value and 1σ uncertainty, we referred to its one-dimensional χ^2 projection from NuFIT 5.2. The charged-leptons mass ratios are taken from [3]. The running of low-energy neutrino observables can be neglected since in our fit we do not obtain a quasi-degenerate spectrum [74].*

Note that, as a consequence of the Minimal seesaw, the matrix m_ν has rank 2, i.e. in the limit of exact SUSY our model predicts that the lightest neutrino is massless.[4] From (3.11) it follows that the mass-scale of light neutrinos is completely determined by the global parameter $g^2v_u^2/\Lambda$ and, from the discussion below eq. (3.7), we conclude that the low-energy neutrino mixing and mass-splittings will be mainly dictated by five parameters: $\mathrm{Re}\,\tau, \mathrm{Im}\,\tau, g'/g, g''/g, \lambda$. Note that, among these, the real and imaginary part of τ already appear in the charged-leptons mass matrix.

4 Numerical analysis and results

In total, the parameter space of our model consists of only eight dimensionless parameters, as discussed above.[5] To verify that our model is able to reproduce the seven experimental dimensionless observables reported in table 2, we performed a χ^2 analysis for which we used

[4] Regarding the stability of $m_{lightest} = 0$ against the quantum corrections, it has been shown that the condition still holds at one-loop for a running between the seesaw and electroweak scales. The two-loop effect is vanishingly small [71].

[5] The overall mass scales αv_d and $g^2v_u^2/\Lambda$ are easily recovered through the values of m_τ and $|\Delta m^2_{\rm atm}|$ and thus are not introduced into the fit. In doing so, we used for m_τ the value reported in [26] and $|\Delta m^2_{\rm atm}|/(10^{-3}\ \mathrm{eV}^2) = 2.507^{+0.026}_{-0.027}$ **(NO)**, $2.486^{+0.025}_{-0.028}$ **(IO)** from NuFIT 5.2.

the Gaussian approximation

$$\chi^2(p_i) = \sum_{j=1}^{7} \left(\frac{q_j(p_i) - q_j^{\text{b-f}}}{\sigma_j} \right)^2 , \tag{4.1}$$

for the mixing $\{\sin^2\theta_{12}, \sin^2\theta_{13}, \sin^2\theta_{23}\}$, the mass ratios $\{m_e/m_\mu, m_\mu/m_\tau, r\}$ and for the invariant J_{CP}. Here $q_j(p_i)$ are the observables obtained from our model with the set of dimensionless parameters $p_i = \{\beta/\alpha, ..., g'/g, ..., \lambda\}$ taken as input. On the other hand, $q_j^{\text{b-f}}$ and σ_j are the best-fit values and the 1σ symmetrized uncertainties taken from table 2. The Jarlskog invariant J_{CP} parameterises the amount of CP violation in the lepton sector and is defined as [72]:

$$J_{\text{CP}} = \text{Im}\left[U_{11}U_{12}^*U_{21}^*U_{22}\right], \tag{4.2}$$

where $U \equiv U_\ell^\dagger U_\nu$ is the PMNS matrix [75, 76] for the lepton mixing, which in the case of Majorana neutrinos can be written in the standard convention:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\text{CP}}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{\text{CP}}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{\text{CP}}} & c_{23}c_{13} \end{pmatrix} \text{diag}(e^{-i\alpha_1}, e^{-i\alpha_2}, 1) , \tag{4.3}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$ and α_1, α_2 are the two Majorana phases. Note that, if the lightest neutrino is massless (as in our case), only one Majorana phase is physical.

It must be noted that, given the CP symmetry of the model which consists in the invariance under $\tau \to -\tau^*$ (a reflection across the imaginary axis), we obtain two sets of points distinguished only by $\pm\text{Re}\,\tau$ with the CP violating (CPV) phases given by $\{\pm\delta_{\text{CP}}, \pm\alpha_1, \pm\alpha_2\}$. For thoroughness, we reported both sets in figure 1. The results of our fit[6] are given in table 3, where the 1σ ranges correspond to the increase of one unit of the one dimensional χ^2 projections from χ^2_{min}.

[6] We explored the parameter space with an algorithm inspired by ref. [77] which mimics the Brownian motion of a particle in a potential that depends on the figure of merit $l(p_i) = \sqrt{\chi^2(p_i)}$, where p_i is the array of free dimensionless parameters mentioned in the text.

	Best-fit and 1σ range
$\mathrm{Re}\,\tau$	$\pm 0.244^{+0.012}_{-0.067}$
$\mathrm{Im}\,\tau$	$1.132^{+0.027}_{-0.297}$
β/α	$0.92^{+0.85}_{-0.03}$
γ/α	$-1.20^{+0.06}_{-2.14}$
$\log_{10}(\alpha_D/\alpha)$	$-13.4^{+13.2}_{-76.3}$
g'/g	$2.76^{+0.21}_{-0.23}$
g''/g	$-2.53^{+0.13}_{-0.03}$
$\log_{10}(\lvert\lambda\rvert)$	$-12.2^{+10.9}_{-59.2}$
$v_d\,\alpha$, [GeV]	$1.08^{+0.06}_{-0.69}$
$v_u^2\,g^2/\Lambda$ [eV]	$3.46^{+0.55}_{-1.65}$
$\sin^2\theta_{12}$	$0.305^{+0.011}_{-0.011}$
$\sin^2\theta_{13}$	$0.0221^{+0.0006}_{-0.0005}$
$\sin^2\theta_{23}$	$0.448^{+0.014}_{-0.016}$
r	$0.0296^{+0.0006}_{-0.0008}$
m_e/m_μ	$0.0048^{+0.0001}_{-0.0002}$
m_μ/m_τ	$0.0574^{+0.0032}_{-0.0050}$
Ordering	**NO**
J_{CP}	$-0.018^{+0.002}_{-0.002}$
α_1/π	0
α_2/π	$\pm 0.112^{+0.792}_{-0.014}$
m_1 [meV]	0
m_2 [meV]	$8.620^{+0.095}_{-0.123}$
m_3 [meV]	$50.806^{+0.016}_{-0.021}$
$\sum_i m_i$ [eV]	$0.0594^{+0.0001}_{-0.0001}$
$\lvert m_{\beta\beta}\rvert$ [meV]	$3.61^{+0.09}_{-0.09}$
m_β^{eff} [meV]	$8.90^{+0.10}_{-0.09}$
d_{FT}	3.03
χ^2_{min}	0.98

Table 3: *Best fit values of the free parameters and observables with their* 1σ *ranges, obtained as described in the text.*

The fit is excellent, with $\chi^2_{\mathrm{min}} = 0.98$ and most dimensionless free parameters being $\mathcal{O}(1)$

with only two exceptions, related to the parameters α_D/α (which can take values arbitrarily close to zero with a spread reaching $\mathcal{O}(1)$ values, all inside its 1σ range) and λ, which takes values arbitrarily close to zero as selected by the fit procedure. The degree of "fine-tuning" for the free parameters of the model, has been evaluated by means of a normalised Altarelli-Blankenburg measure $d_{\rm FT}$ [78]:

$$d_{\rm FT} = \frac{\displaystyle\sum_i \left|\frac{{\rm par}_i}{\delta{\rm par}_i}\right|}{\displaystyle\sum_i \left|\frac{{\rm obs}_i}{\sigma_i}\right|} \,. \tag{4.4}$$

Here $|{\rm par}_i|$ is the best-fit value of the free parameter and $|\delta{\rm par}_i|$ is the shift of the parameter from the best-fit value which causes $\chi^2 - \chi^2_{\rm min} = 1$, while keeping all the other parameters fixed to their best-fit values. Intuitively, a model can be considered fine-tuned if it takes even a very small shift from the best-fit value of some parameters to produce a large variation in the χ^2. The measure (4.4) is normalized by the absolute sum of the ratios between the observables best-fit values and their uncertainties. For the fit shown in table 3 we obtained $d_{\rm FT} \sim \mathcal{O}(1)$ which suggests that our model is not affected by a fine-tuning much larger than the one which already affects the input data.

Let us now turn our attention to the predictions of our model. First of all, we found that the Inverted Ordering of the neutrino mass eigenstates is strongly disfavoured compared to the Normal Ordering (**NO**); equally relevant is the fact that, as anticipated, the lightest neutrino mass is exactly vanishing. This allows us to easily understand the numerical results on the Majorana effective mass of neutrinoless double-beta decay and the effective neutrino mass of tritium decay. In the first case, the usual definition (inserting $\alpha_1 = 0$) gives:

$$|m_{\beta\beta}| = \left|\sum_i m_i U_{ei}^2\right| = \sqrt{\Delta m^2_{\rm atm}}\left|\sqrt{r}\, s^2_{12} c^2_{13} + \sqrt{1+r}\, s^2_{13} e^{-i(\delta_{\rm CP}-\alpha_2)}\right| \sim \mathcal{O}(1)\,{\rm meV}\,, \tag{4.5}$$

while in the latter case:

$$m^{\rm eff}_\beta = \sqrt{\sum_i m_i^2 |U_{ei}|^2} = \sqrt{\Delta m^2_{\rm atm}}\,\sqrt{s^2_{13} + r(1 - c^2_{13} c^2_{12})} \sim \mathcal{O}(10)\,{\rm meV}\,, \tag{4.6}$$

in perfect agreement with the results shown in table 3. Finally, the non-vanishing neutrino mass eigenstates lie in the region of tens of meV, implying $\sum_i m_i \sim 0.06$ eV, perfectly compatible with the most recent upper bound of 0.115 eV (95 % C.L.) from [79]. This, again, is mainly dictated by the **NO** scenario with vanishing lightest neutrino mass:

$$\sum_{i=1}^{3} m_i = \sqrt{\Delta m^2_{\rm atm}}(\sqrt{r} + \sqrt{1+r}) \sim 0.06\,{\rm eV}\,. \tag{4.7}$$

In figure 1 some correlations between observables (both fitted and predicted ones) and free parameters are displayed. All points in the plots have been selected by our algorithm and satisfy the relation $\sqrt{\chi^2} \leq 5$. The yellow point corresponds to the minimum χ^2 reported in table 3 while the black one is the CP-transformed value of the modulus, which produces the

inversion of all the CPV phases in the mass matrices. The orange lines and bands correspond to the experimental best fit and 1 σ ranges, respectively, taken from table 2. For illustrative purposes, we also report points located outside the fundamental domain.

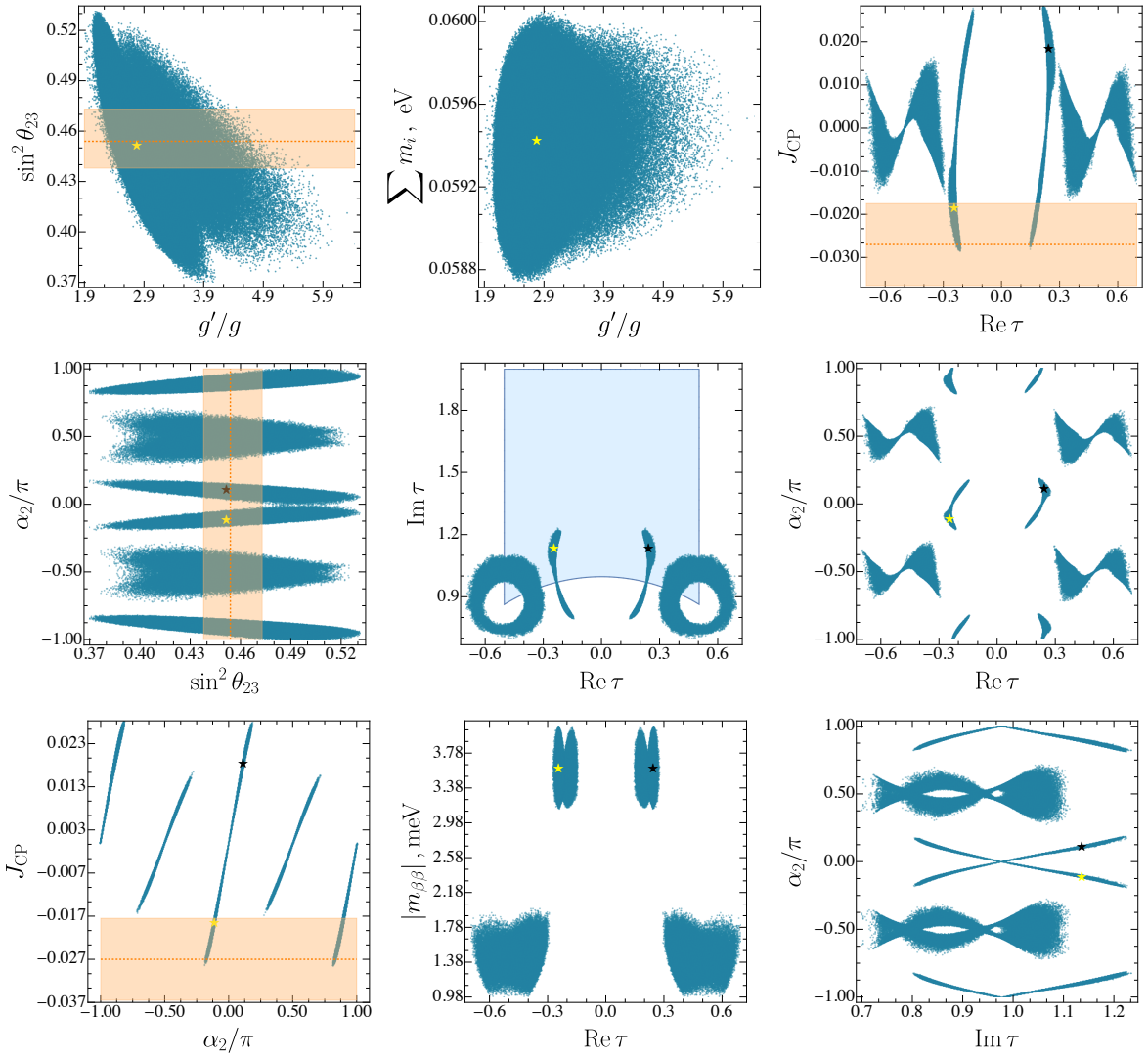


Figure 1: *Correlations between observables and free parameters. All the plotted points were accepted by the algorithm used for the fit and satisfy* $\sqrt{\chi^2} \leq 5$. *The yellow point corresponds to the minimum* χ^2. *The black point is the CP-transformed value of the modulus, which produces the inversion of all the CPV phases in the mass matrices. Orange lines and bands correspond to the experimental best fit and 1*σ *ranges taken from table 2. Points located outside the fundamental domain are displayed for the purpose of illustration.*

5 Leptogenesis

We now discuss the mechanism of leptogenesis. It provides an elegant explanation for the matter-antimatter asymmetry of the Universe, attributing it to a lepton number asymmetry generated by the decays of heavy right-handed neutrinos. This asymmetry is then converted into a baryon asymmetry through non-perturbative processes referred to as *sphalerons* [56, 80, 81].

As commonly done, we express the baryon asymmetry by the parameter η_B, as

$$\eta_B \equiv \left. \frac{n_B - n_{\bar{B}}}{n_\gamma} \right|_0 , \tag{5.1}$$

where $n_B, n_{\bar{B}}$ and n_γ refer to the number densities of the baryons, antibaryons and photons, while the subscript "0" means "at present time".

We have worked so far in the limit of exact supersymmetry. We will assume that the SUSY breaking scale m_{SUSY} is above the mass scale of the right-handed neutrinos of our model, so that we can avoid considering the superparticles contribution to the final baryon asymmetry, as well as all the complications related to the gravitino densities [58, 82–84]. Nevertheless, it is imperative to exercise caution regarding the magnitude of $m_{\rm SUSY}$, as arbitrarily large values may introduce complications. This is discussed in depth in Ref. [85]. Notably, the lepton masses undergo a correction which is proportional to the ratio $m_{\rm SUSY}/\mathcal{F}$, where $\mathcal{F}$ represents the characteristic scale at which the supersymmetry-breaking sector communicates with the visible sector. This scale can be chosen in proximity to the Planck scale M_{pl}. As we will discuss in detail in the following, our model provides right-handed neutrinos with masses of order $\mathcal{M} \sim \mathcal{O}(10^{12})$ GeV. Thus, we could choose $m_{\rm SUSY} \simeq 10^{14}$ GeV and $\mathcal{F} \simeq 10^{18}$ GeV so that the corrections to the low energy observables do not affect our results, and we can proceed in discussing the leptogenesis without considering any massive superparticle.
In this section, we will focus on the so-called N_1 dominated scenario (N1DS) [86–88], even though our model involves one additional sterile state N_2. Then, we will solve numerically the semi-classical Boltzmann Equations [55, 89, 90] for the baryon asymmetry evolution, and we will present a realization that provides the value of the baryon-to-photon ratio η_B inside the 3σ experimental allowed region [91].

5.1 Thermal leptogenesis: basic physics

The lepton number violating decays of the sterile Majorana neutrinos N_i are the main ingredients for the thermal leptogenesis recipe [83, 92]. They depend, in general, on the mass of the heavy neutrinos, as well as their Yukawa couplings with the charged leptons and the Higgs doublet. In order to generate a lepton asymmetry, we first need a thermal production of a distribution of N_i through scattering in the early Universe at temperatures around $T \sim M_{N_i}$, where we indicate with M_{N_i} the mass of the N_i sterile neutrino. As the temperature drops below the mass M_{N_i}, these sterile neutrinos undergo decays via processes that violate the lepton number by one unit ($\Delta L = 1$), shown in figure 2 and, if these decays are slow enough compared to the expansion rate of the Universe, the N_i abundance does not decrease according to the

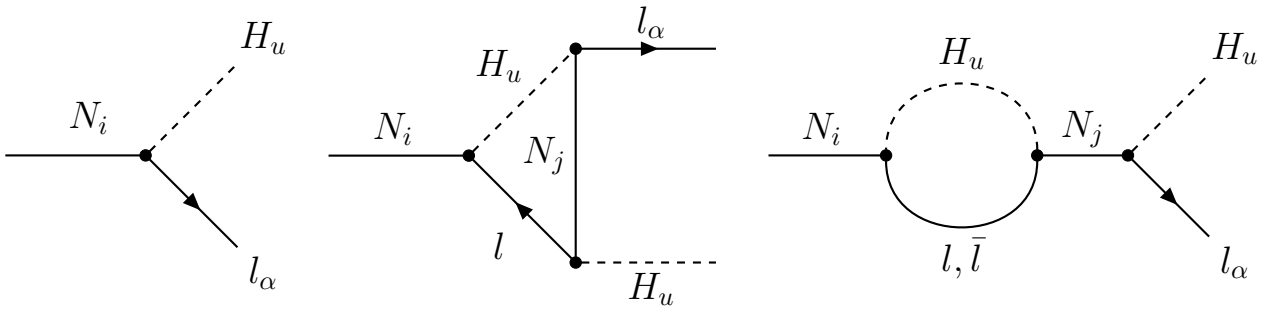


Figure 2: *CP-violating N_i decay.*

Boltzmann equilibrium statistics $\propto e^{-M_{N_i}/T}$, as the equilibrium demanded. This is how a net lepton asymmetry is generated. Thus, the crucial ingredient is to estimate the mass of the heavy neutrinos. To this aim, it is useful to recast the Majorana and Dirac mass matrices of eqs. (3.9) and (3.10) in the following way:

$$M_D = v_u\, g Y_1^2 \begin{pmatrix} (1+g''/g)+(g''/g-1)\,\zeta^2 & 2\,\zeta & 2\,\zeta\, g'/g \\ 2\,\zeta & (1+g''/g)\,\zeta+(g''/g-1) & (1-\zeta^2)\, g'/g \end{pmatrix}_{\rm RL} , \tag{5.2}$$

$$\mathcal{M}_R = \Lambda\, Y_1^2 \begin{pmatrix} \lambda(1+\zeta^2)-(\zeta^2-1) & 2\,\zeta \\ 2\,\zeta & \lambda(\zeta^2+1)+(\zeta^2-1) \end{pmatrix}_{\rm RR} , \tag{5.3}$$

where $\zeta = Y_2/Y_1$. With the modulus τ in the fundamental domain, and for small values of $\mathrm{Re}\,\tau$, it reads:

$$\zeta \simeq 8\sqrt{3}\, t^2\ (1 + i\,\pi\,\mathrm{Re}\tau)\ , \quad t \equiv e^{-(\pi\,\mathrm{Im}\tau)/2}\,, \tag{5.4}$$

where only terms up to $\mathcal{O}(t^2)$ are considered.
Therefore, with $|\zeta| \ll 1$ and all the free parameters close to their best fit values quoted in table 3, it is evident that the mass scale of the right-handed neutrinos is basically $\mathcal{M}_N \sim \Lambda Y_1^2$ and the Yukawas $\mathcal{Y} \sim g\, Y_1^2$. In other words, both the mass scale of the Majorana neutrinos and the Yukawas get suppressed by the modular form Y_1, which can be expanded as:

$$Y_1(\tau) \simeq \frac{7}{100} + \frac{42}{25}\, t^2\ (1 + 2\, i\,\pi\,\mathrm{Re}\tau) + \mathcal{O}(t^3)\,. \tag{5.5}$$

Since the mass-scale of light neutrinos is completely determined by the global parameter $g^2\, v_u^2/\Lambda$, we can estimate Λ. By imposing that the dimensionless parameter g assumes a value of order $\mathcal{O}(1)$, we get $\Lambda \sim \mathcal{O}(10^{14})$ GeV. This, along with the suppression provided by the modular form Y_1, implies a mass scale $\mathcal{M}_N$ for the Majorana neutrinos of approximately 10^{12} GeV.

In principle, the interference between the tree-level and the loop contributions in figure 2 produces CP violating decays that can generate asymmetries in all flavors α. Flavor effects [93–95] become relevant when the rate of interaction Γ_β of the processes involving a lepton with flavor β ($\beta = e, \mu, \tau$) are rapid enough to discern distinct lepton flavors. Under these circumstances, the coherence of leptons and antileptons as flavor superpositions is lost during

leptogenesis. Consequently, it becomes necessary to adopt a flavor basis to calculate the baryon asymmetry. Typically, the temperature below which the flavor effect cannot be neglected is such that the interaction rate Γ_β for a charged lepton with flavor β becomes faster than the expansion rate $H(T)$ of the Universe. The interaction rate can be estimated [96, 97] as:

$$\Gamma_\beta \simeq 10^{-2}\, \alpha_\beta^2\, T \,, \tag{5.6}$$

where α_β is the charged-lepton Yukawa coupling. For a certain flavor-dependent $T \lesssim T_\beta$, the condition $\Gamma_\beta(T) > H(T)$ is satisfied, so that the typical temperatures below which the leptogenesis becomes sensitive to lepton flavors are $T_e \simeq 4 \cdot 10^4$ GeV, $T_\mu \simeq 2 \cdot 10^9$ GeV and $T_\tau \simeq 5 \cdot 10^{11}$ GeV [98–100]. The amount of CP asymmetry generated by the interference between the tree and one loop decay diagrams of figure 2 can be described by the ϵ_i parameters, defined as:

$$\begin{aligned} \epsilon_i \, &= \sum_\alpha \frac{\Gamma(N_i \to L_\alpha\, H_u) - \Gamma(N_i \to \overline{L_\alpha}\, H_u^*)}{\Gamma(N_i \to L_\alpha\, H_u) + \Gamma(N_i \to \overline{L_\alpha}\, H_u^*)} = \\ &= \sum_{j \neq i} \frac{\mathrm{Im}\left[\left(\mathcal{Y}^\dagger \mathcal{Y}\right)^2_{ji}\right]}{8\pi (\mathcal{Y}^\dagger \mathcal{Y})_{ii}} \left[f(x_{ij}) + g(x_{ij})\right] \quad \text{with } i = 1, 2 \end{aligned} \tag{5.7}$$

with $x_{ij} = M_j^2/M_i^2$ and the loop functions [101, 102]:

$$\begin{aligned} f(x_{ij}) \, &= \sqrt{x_{ij}} \left[1 - (1 + x_{ij}) \log\left(1 + \frac{1}{x_{ij}}\right) + \frac{1}{1 - x_{ij}}\right], \\ g(x_{ij}) \, &= \frac{\sqrt{x_{ij}}(1 - x_{ij})}{(1 - x_{ij})^2 + x_{ij} (\mathcal{Y}^\dagger \mathcal{Y})^2_{jj}/(16\pi^2)} \,, \end{aligned} \tag{5.8}$$

which are related to the vertex correction and the self energy correction, respectively.
The lepton number violation produced by the sterile neutrino decays can be depleted by scatterings and inverse processes, which are often referred to as *wash-out processes*. If they occur out-of-equilibrium, *i.e.* they are not very fast compared to the Hubble expansion rate of the Universe $H(T)$ at temperature $T = M_{N_i}$, a net asymmetry can survive.
Typically, we can quantify the departure from thermal equilibrium by the *decay parameter* which, in terms of the decay rates Γ_i of the N_i sterile neutrino, reads as follows:

$$K_i \equiv \frac{\Gamma_i}{H(T = M_i)} = \frac{\widetilde{m}_i}{m^*} \,, \tag{5.9}$$

where the *effective neutrino mass* [103] is defined as:

$$\widetilde{m}_i = \frac{\left(\mathcal{Y}^\dagger \mathcal{Y}\right)_{ii} v_u^2}{M_i} \,, \tag{5.10}$$

and $m^* \simeq 1.1 \cdot 10^{-3}$ eV. We can distinguish three different regimes: $K \ll 1$, $K \simeq 1$ and $K \gg 1$, which we identify as the *weak*, *intermediate* and *strong* wash-out regime, respectively. Using our best fit values of the free parameters of the model, we deduce we are always in a strong

wash-out regime; thus, the contributions to the leptogenesis from the $\Delta L = 1$ and $\Delta L = 2$ processes[7] can be safely neglected, as discussed in [55,104,105]. Therefore, from now on, we will only consider the decays and the inverse decays. Let us now discuss the relevant Boltzmann Equation (BE) for the leptogenesis in our model, with two heavy sterile neutrinos $N_{1,2}$ with masses $M_1 \lesssim M_2$. Assuming the reheating temperature $T_{\rm rh} > M_{1,2}$, both the sterile neutrinos $N_{1,2}$ are thermally produced [80,106–108] in the Early Universe, and the set of classical kinetic equations [89] could be written as:[8]

$$\begin{aligned} \frac{\mathrm{d}N_{N_i}}{\mathrm{d}z} &= -\, D_i \left(N_{N_i} - N_{N_i}^{eq}\right) , \quad \text{with } i = 1, 2 \\ \frac{\mathrm{d}N_{B-L}}{\mathrm{d}z} &= -\sum_{i=1}^{2} \epsilon_i\, D_i \left(N_{N_i} - N_{N_i}^{eq}\right) - \sum_{i=1}^{2} W_i\, N_{B-L} \end{aligned} \tag{5.11}$$

with $z = M_1/T$. Here, D_i is the decay term while W_i is the washout factor; futhermore, the N_i's are the number densities of the RH sterile neutrinos, while N_{B-L} is the amount of $B - L$ asymmetry, both normalized by the comoving volume [56]. With this choice, we automatically take into account the expansion of the Universe, which was not possible if we were to employ the number density n_X of the particle species. The equilibrium abundances of N_i's are given by $N_i^{eq} = \frac{3}{8} z_i^2 \mathcal{K}_2(z_i)$, with $\mathcal{K}_2$ the Bessel function of the second kind. In the literature, the comoving volume $R_*^3(t)$ is usually chosen as to contain one photon at the time t_*, before the onset of the leptogenesis [105]:

$$N_X(t) = n_X(t) R_*^3(t) , \tag{5.12}$$

with

$$R_*(t_*) = \left(n_\gamma^{eq}(t_*)\right)^{-1/3} , \tag{5.13}$$

so that $N_\gamma(t_*) = 1$. Using $z_i = z\sqrt{x_{1i}}$, we can write the decay term D_i and the washout factor W_i as:

$$\begin{aligned} D_i &= K_i\, x_{1i}\, z\, \langle 1/\gamma_i \rangle , \\ W_i \simeq W_i^{ID} &= \frac{1}{4}\, K_i \sqrt{x_{1i}} \mathcal{K}_1(z_i) z_i^3 , \end{aligned} \tag{5.14}$$

where W_i^{ID} is the inverse decay term which contributes to the washout of the lepton asymmetry, and $\langle 1/\gamma_i \rangle$ is the thermally averaged dilation factor, given as the ratio of the modified Bessel functions of the second kind:

$$\left\langle \frac{1}{\gamma_i} \right\rangle = \frac{\mathcal{K}_1(z_i)}{\mathcal{K}_2(z_i)} . \tag{5.15}$$

The final $B - L$ asymmetry N_{B-L}^f is given by:

$$N_{B-L}^f = N_{B-L}(z = \infty) = N_{B-L}^{in} e^{-\sum_i \int dz' W_i(z')} + N_{B-L}^{lep} , \tag{5.16}$$

[7] We refer to the relevant lepton number violating processes in the thermal plasma: the $\Delta L = 1$ scatterings between N_i and leptons mediated by the Higgs boson, involving the top quark or electroweak gauge bosons, and the $\Delta L = 2$ scatterings between the leptons and the Higgs boson with an intermediate N_i heavy neutrino.

[8] It is convenient to write the BE for the $B - L$ asymmetry evolution, because all the electroweak processes conserve it, including those mediated by the sphalerons.

where N^{lep}_{B-L} is the $B-L$ produced via thermal leptogenesis and N^{in}_{B-L} indicates the possible pre-existing asymmetry [109, 110], at an initial temperature T_{in} before the leptogenesis onset. The potential pre-existing $B-L$ asymmetry could introduce further constraints on the parameter space of the model [107, 111–113]. In order to consider a scenario of pure leptogenesis from RH neutrino decays, we will fix $N^{in}_{B-L}=0$ throughout the rest of the paper.

During the expansion of the Universe, the $B-L$ asymmetry is reprocessed in a baryon asymmetry through sphaleron processes. If sphalerons decouple before the electroweak phase transition, the amount of B asymmetry produced is [114]:

$$N^f_B = \frac{(8\,N_F + 4\,N_H)}{(22\,N_F + 13\,N_H)} N^f_{B-L}\,, \tag{5.17}$$

where $N_F = 3$ and $N_H = 2$ stand for the number of fermion generations and the number of Higgs doublets, respectively. The predicted baryon-to-photon ratio has to be compared with the value η_B measured at recombination [91]. The relation is:

$$\eta_B = \frac{N^f_B}{f}\,, \tag{5.18}$$

where $f = 2387/86$ is the dilution factor calculated assuming standard photon production from the onset of leptogenesis till recombination [89].

5.2 Validity of the N_1-dominated scenario

In presence of a hierarchical mass spectrum $M_1 \lesssim M_2$, only the contribution from the lightest sterile neutrino to the leptogenesis is commonly considered, the so-called N_1-dominated scenario. However, as widely discussed in literature [86–88], there exists a parameter space for which such an assumption is no longer valid, and the heaviest sterile neutrino contribution to the final baryon asymmetry cannot be neglected. An important parameter that can be used to test whether the N_1-dominated scenario sets in is the parameter $\delta_i = (M_i - M_1)/M_1$. This is related to x_{1i} by $\sqrt{x_{1i}} = 1+\delta_i$, and it will tell us whether the CP-asymmetry can undergo resonant enhancement [55, 57, 115–117]. In fact, it can be noted that, when dealing with an almost degenerate mass spectrum, the self-energy loop function $g(x_{ij})$ in eq.(5.8) can be reformulated in terms of δ_j in the following manner:

$$g(\delta_j) \simeq \frac{\delta_j}{2\left[(\delta_j)^2 + (\mathcal{Y}^\dagger\mathcal{Y})^2_{jj}/(64\pi^2)\right]}\,. \tag{5.19}$$

Therefore, if the Yukawa matrix and the mass splittings are independent quantities, the resonance enhancement occurs if the condition [101]

$$\delta^2_j \simeq (\mathcal{Y}^\dagger\mathcal{Y})^2_{jj}/(64\pi^2) \tag{5.20}$$

holds. As already noted, our model provides Yukawa couplings suppressed by the modular form Y_1 while the mass splitting parameter δ_2 belongs to the range $[0.7, 2.3]$, as our evaluation

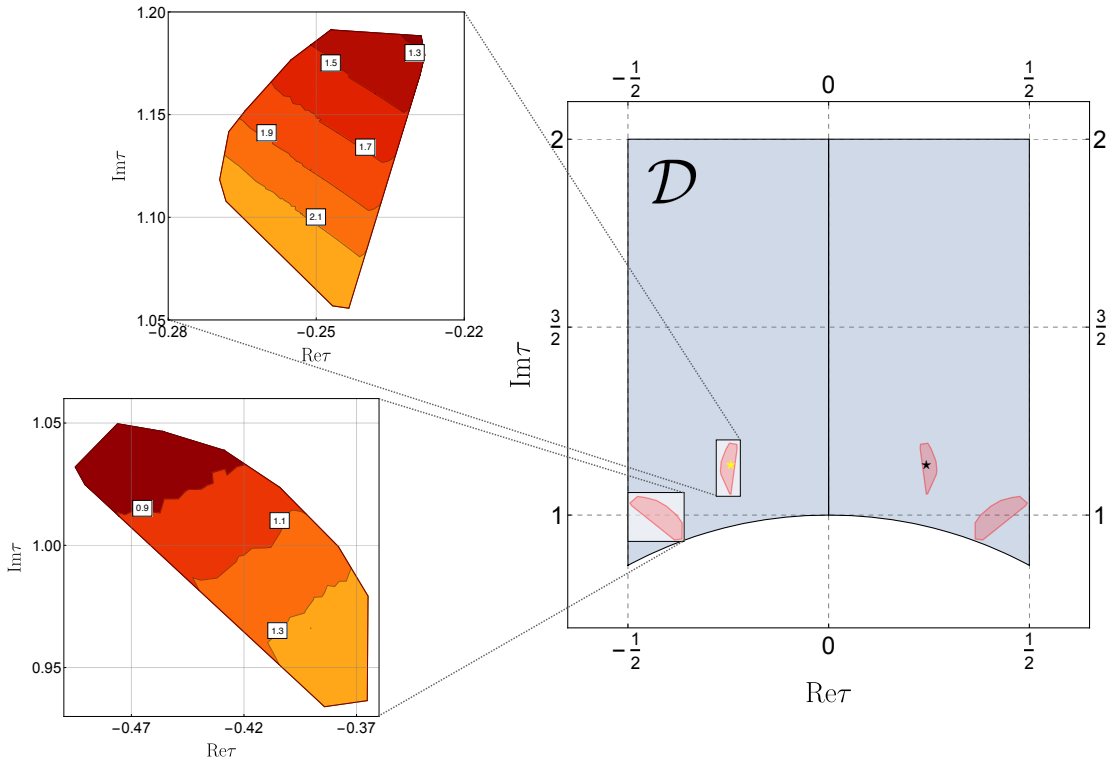


Figure 3: *Contour plots showing the values assumed by δ_2 in the best fit regions. For a clearer and more concise data visualization, all the fitted points from section 4 were mapped into $\mathcal{D}$, see appendix C for details.*

shows in figure 3. Given that the right-hand side of eq.(5.20) is approximately on the order of $\mathcal{O}(10^{-5})$, we safely deduce that the resonance condition is never satisfied in the best fit regions.[9] We can now discuss the validity of the N_1-dominated scenario in presence of another heavy neutrino N_2. In general, one should solve eqs.(5.11) taking into account the wash-out effects due to both N_1 and N_2 sterile neutrinos. The strength of the N_2 wash-out effects is encoded into W_2^{ID}, see eq.(5.14), which depends on the δ_2 parameter. There exists a specific value δ_2^* above which the influence of W_2^{ID} becomes irrelevant compared to W_1^{ID}; in that case, leptogenesis would be predominantly driven by the presence of the lightest neutrino, from which the name of N_1-dominated scenario. Conversely, for $\delta_2 < \delta_2^*$, the effect of W_2^{ID} becomes relevant. Usually, in a strong wash-out regime, a value $\delta_2^* \gtrsim 1.5 \div 5$ is found [86,88]. In our model, $\delta_2^* \simeq 1$, so that the N_1-dominated scenario is valid for $\text{Re}\tau \in [-0.44, -0.22]$ (which includes our best fit point) where $\delta_2 \gtrsim \delta_2^*$, while in the other regions of the parameter space we cannot a priori neglect the wash-out effects of the N_2 sterile neutrinos. In appendix A we illustrate that in the absence of specific correlations among the parameters, the contribution of the heavier sterile neutrino to

[9]Notice that the resonance would have probably ruined the prediction of the model regarding the baryon-to-photon ratio η_B; indeed, with heavy sterile neutrinos with masses around 10^{12} GeV, the resonance would imply an enhancement of the CP-violating parameter such that the final baryon asymmetry exceeds the observed value by many order of magnitudes (see *e.g.* [118]).

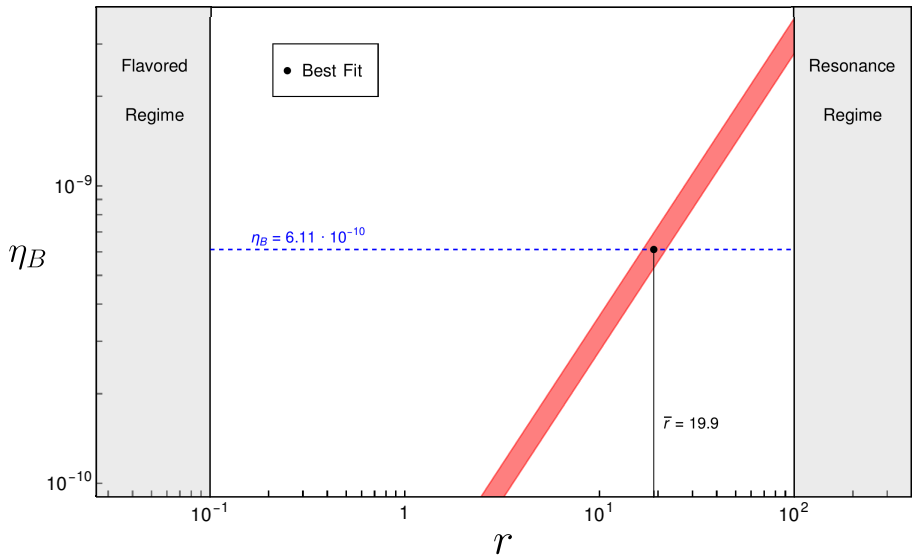


Figure 4: *Pink band: dependence of the final baryon-to-photon asymmetry η_B on the rescaling parameter "r" ($\Lambda \to r\,\Lambda$) for a fixed set of the theory's free parameters, satisfying $\sqrt{\chi^2} \leq 2$. The observed value of η_B [91] is shown with the blue line. The unflavored leptogenesis is provided for $10^{-1} < r < 10^2$, and it turns out to be successful for $16.8 \leq r \leq 22.1$. In particular, at the best fit, η_B in agreement with the observation for $r = \bar{r} = 19.9$.*

leptogenesis is negligible for $\delta_2 \gtrsim 1.5$.

5.3 Numerical analysis

We solve the full Boltzmann Equations presented in eqs.(5.11) for the best fit points that satisfy $\sqrt{\chi^2} \leq 2$. With the heavy sterile neutrino masses at $\mathcal{O}(10^{12})$ GeV, the model provides $\eta_B \lesssim 10^{-10}$, slightly below the observed value. However, since only the ratio $g^2\, v_u^2/\Lambda$ is constrained by the masses of the light neutrinos, the mass scale of the Majorana neutrinos can be regarded as a free parameter of the model. We can safely rescale $\Lambda \to r\,\Lambda$ and $g \to \sqrt{r}\, g$ without altering the results of the fit in table 3.[10] Notice that varying the mass-scale of the right-handed neutrinos have relevant consequences. For $r \lesssim 10^{-1}$ the mass-scale of the sterile neutrinos is lowered down to $\mathcal{O}(10^{11})$, below which the *flavored regime* sets in; on the other hand, if $r \gtrsim 10^2$ the *resonant regime* occurs. We confine the analysis to the range $r \in [10^{-1}, 10^2]$, focusing on the unflavored scenario that involves solely the lightest sterile neutrino [53, 54, 56, 83, 104, 105, 120]. The results are shown in figure 4, where the pink band represents the dependence of the final baryon-to-photon asymmetry η_B on the parameter r for a fixed set of the theory's free parameters, satisfying $\sqrt{\chi^2} \leq 2$; in *blue*, instead, we report the observed value of η_B (due to the small uncertainty on this observation, the 1σ allowed range is basically indistinguishable

[10] A similar work has been done in [119], in a context of a minimal extended seesaw, with an additional singlet field.

from the shown dotted line). In particular, the best fit realization of our model provides a successful leptogenesis if $\Lambda \to \bar{r}\Lambda$, where $\bar{r} = 19.9$. It is important to observe that this selection does not compromise the quality of our fit; in fact, it modifies g by a multiplicative factor of approximately $\sqrt{\bar{r}} \sim 4$, thereby maintaining g as an order $\mathcal{O}(1)$ parameter, as for most of the other parameters of the model.

6 Conclusions

In the present work, we have addressed the problem of finding a modular invariant minimalistic construction based on the smallest modular finite group $\Gamma_2 \cong S_3$ that was compatible with low energy neutrino oscillation data and, with the same model parameters, able to accommodate the measured value of the baryon asymmetry of the Universe (BAU). Our attempts have been guided by the requirement to minimally extend the Standard Model particle content; with the aim to naturally generate small neutrino masses, we relied on a seesaw mechanism built with the help of only two heavy right-handed neutrinos. A straightforward prediction of our model is that the lightest neutrino mass is zero at tree level. Having imposed a generalized CP-symmetry, not only we were able to restrict the number of free parameters to ten (eight dimensionless, of which two come from $\tau = \mathrm{Re}\,\tau + i\,\mathrm{Im}\,\tau$, and two global scales) but we have also secured that the modulus τ is the only source of CP-violation. We showed that, even within this limitation, this was enough to reproduce the current experimental value of the leptonic Jarlskog invariant within 1σ. Furthermore, the model strongly favours the Normal Ordering of neutrino masses, a prediction that can be tested in upcoming experiments [121–124]. The predictions for the Majorana effective mass of the neutrinoless double-beta decay and the effective neutrino mass of tritium decay are both substantially within the current upper bounds from [125, 126], and are both analitically understood from the constraint on the lightest neutrino mass, which is in turn a consequence of the Minimal seesaw scenario.

Since the only source of CP violation in our model is dictated by $\mathrm{Re}\,\tau$, it was highly non-trivial to generate a sufficient amount of BAU through the leptonic asymmetry η_B; we have demonstrated, indeed, that this is the case in our model. Given that in the presence of a mild hierarchical heavy mass spectrum with $\delta_2 \gtrsim 1$ only the contribution from the lightest sterile neutrino to the leptogenesis is commonly considered, we solved the appropriate Boltzmann Equations in the N_1-dominated scenario and shown that a mild rescaling of the Majorana right-handed mass scale Λ (not fixed by the fit) is enough to guarantee the required η_B value. Given the minimality of our $\Gamma_2 \cong S_3$ construction, we consider this feature a relevant result in the model building based on modular invariant theories.

Acknowledgements

The authors would like to thank J. Penedo for useful discussions.

Appendix A Beyond the N_1-dominated scenario

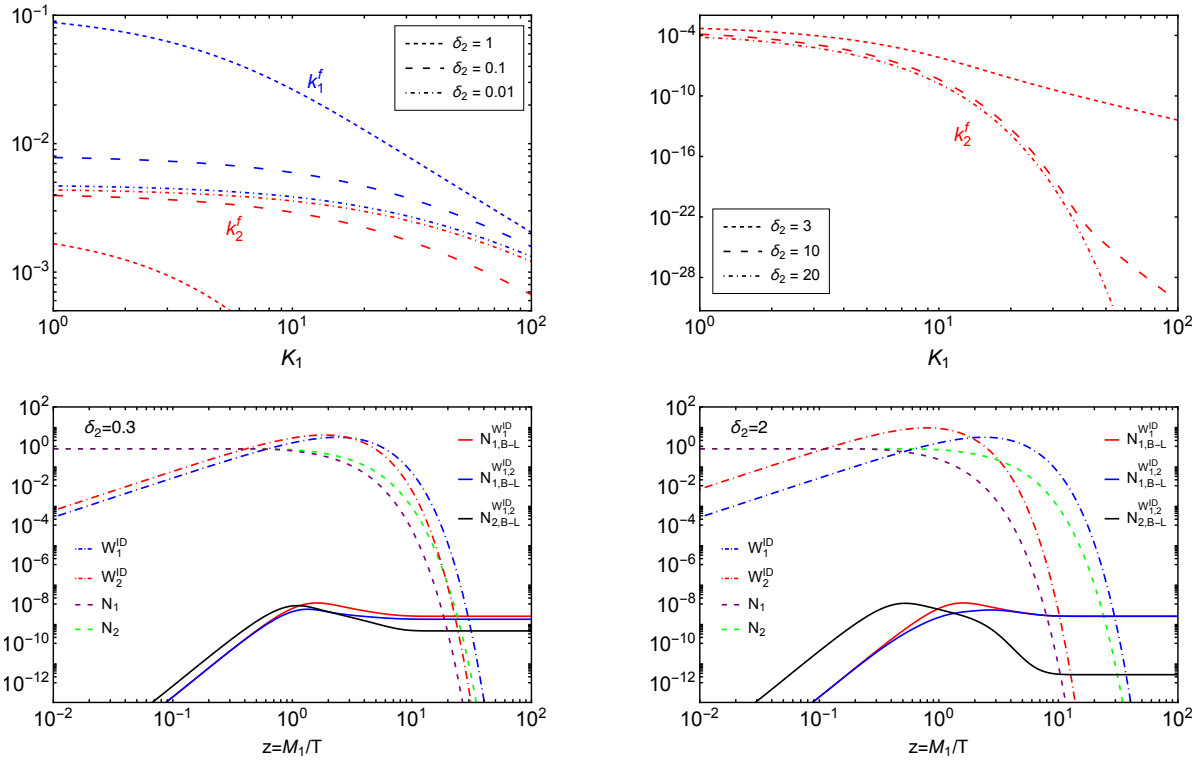


Figure A.1: *Top left (right) panel: evolutions of the efficiency factors k_i^f in terms of K_1 with $\delta_2 = 1,\ 0.1,\ 0.01$ ($\delta_2 = 3,\ 10,\ 20$), with $K_2 \simeq 25$. The efficiency factors of N_1 and N_2 are shown in blue and red dashed lines, respectively. Bottom left (right) panel: $|N_{B-L}|$ as a function of $z = M_1/T$ and other relevant quantities such as N_i and W_i^{ID}, with $\delta_2 = 0.3$ ($\delta_2 = 2$), $N_i^{in} = 3/4$ and $K_1 = K_2 = 10$. The colour code for N_i, W_i^{ID} is highlighted in the legend. Solid red line shows $|N_{B-L}|$ for a pure N_1-dominated scenario. The solid blue line shows the asymmetry generated by N_1, partially depleted by the wash-out effects of N_1 and N_2, given by W_1^{ID} and W_2^{ID}, respectively. The solid black line indicates $|N_{B-L}|$ produced by N_2, undergoing $W_{1,2}^{ID}$ washout.*

In this appendix we firstly illustrate that, in the absence of specific correlations among the decay parameters, the feasibility of the N_1-dominated scenario hinges crucially on the value of the δ_2 parameter, defined in section 5.2. This is shown in figure A.1. Then, we discuss the concrete example related to the best fit realization of our model.

In presence of two right-handed sterile neutrinos, the solution of the BEs in eq.(5.11) can be written as [127]:

$$N_{B-L}^f = \sum_i \epsilon_i k_i^f \ , \quad i = 1, 2 \tag{A.1}$$

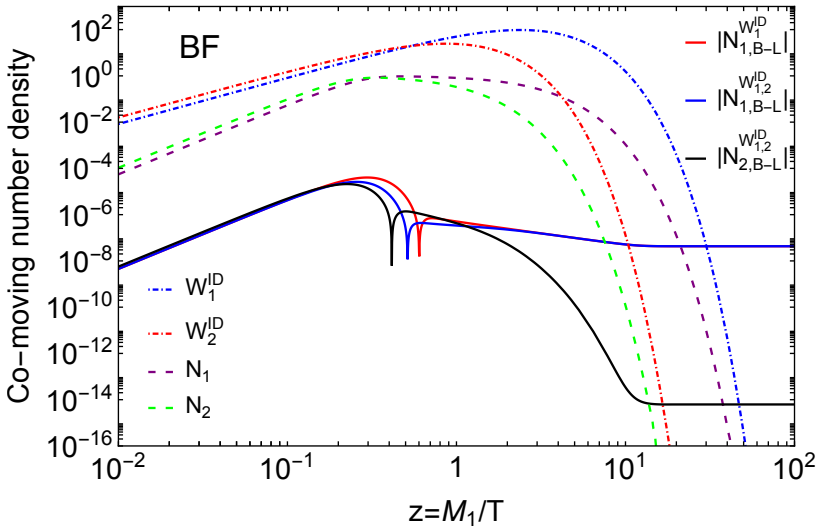


Figure A.2: $|N_{B-L}|$ *as a function of* $z = M_1/T$ *for the best fit realization with* $r = \bar{r} = 19.9$ *(see section 5.3 for further details). The colour code is the same as the bottom panels of figure A.1. A zero initial abundance* N_i^{in} *is assumed.*

with ϵ_i the usual CP-violating parameters (see eqs.(5.7) and (5.8)) and k_i^f the *efficiency factors*. Without strongly hierarchical Yukawa couplings, for quasi-degenerate heavy neutrinos we can assume $\epsilon_1 \simeq \epsilon_2$. Thus, the contribution of each N_i to the final baryon asymmetry is governed by the efficiency factor k_i^f, which can be written as:

$$k_i^f \equiv k_i(z=\infty) = -\int_{z_{in}\to 0}^{z_{fin}\to\infty} \frac{\mathrm{d}N_i}{\mathrm{d}z'} \exp\left[-\sum_i \int_{z'}^{z} W_i(z'')\mathrm{d}z''\right]\mathrm{d}z' \,. \tag{A.2}$$

We focus on the *strong washout regime*, for which the Yukawa couplings are strong enough to let any species reach the equilibrium even if starting from a vanishing initial density. Therefore, the efficiency factors can be computed by inserting into the general formula (A.2) the equilibrium values $\frac{\mathrm{d}N_i^{\mathrm{eq}}}{\mathrm{d}z'}$ for both rate abundances. Also, the k_i^f factors depend on the washout factors K_1, K_2, as well as on the $\delta_2 = (M_2 - M_1)/M_1$ factor which accounts for the mass difference between the RH neutrinos. In the left (right) upper panel of figure A.1 we plotted the k_i^f factors for $\delta_2 = 1,\, 0.1,\, 0.001$ ($\delta_2 = 3,\, 10,\, 20$) values. In order to highlight the effects of the washout, we assumed $K_2 \simeq 25$. As it is evident, as the wash-out parameter increases, the hierarchy between the two efficiency factor becomes stronger, implying that the contribution to the finally BAU from the heavier neutrino can be neglected, and the N_1-dominated scenario is recovered. This is confirmed by the numerical solutions of the BE in case of $\delta_2 = 0.3$ and $\delta_2 = 2$, shown in figure A.1 in the left lower panel and in the right lower panel, respectively. Usually, a good rule of thumb to check whether the N_1-dominated scenario is a correct approximation consists in computing the δ_2 parameter: for $\delta_2 \gtrsim 1.5 \div 5$, the contribution of the heaviest sterile neutrino N_2 to the leptogenesis can be safely neglected and the hierarchical N1DS is realized. Nevertheless, in a realistic model one should take into account possible correlations between

the decay parameters K_1 and K_2 as well as possible hierarchies in the Yukawa couplings. Such correlations could either establish a strong hierarchy between the two efficiency factors k_i^f, even for relatively small values of the δ_2 parameter, or limit the accessibility of the N1DS to cases where $\delta_2 > 5$.
In the case of our best fit realization analyzed in section 5.3, we have $r = \bar{r} = 19.9$ and a value of $\delta_2 \simeq 1.7$; these, along with a particular correlation between the two decay parameters $K_1 \sim 10\, K_2$, lead to a significant depletion of the asymmetry produced by N_2. Such a situation is well depicted in figure A.2, where the $|B-L|$ evolution with z, as well as the other relevant quantities such as N_i and W_i^{ID} is shown. In the figure, the red and blue solid lines correspond to the asymmetry produced by N_1, partially depleted by the wash-out effect due to N_1 and $N_1 + N_2$, respectively. The black line indicates the asymmetry produced by N_2, undergoing wash-out from $N_1 + N_2$. The particular correlation between the two decay parameters together with $\delta_2 \simeq 1.7$, provides a huge gap between the value at the plateau reached by the black line and the red line. When the strength of the inverse decays of N_1 reaches its maximum value, the asymmetry produced by N_2 is practically switched off, so it is efficiently washed-out. In other words, the final dynamics of $|B-L|$ is governed only by the processes involving N_1. This is clear by analysing the evolution of the red line and the blue line: if one takes into account the asymmetry produced by N_1 subjected to the N_2 wash-out along with the N_1 wash-out (blue line), the final asymmetry coincides with the pure N1DS (red line).

Appendix B Modular forms and S_3 representations

The group S_3 is characterized by three irreducible representations: the singlet $\mathbf{1}$, the pseudo-singlet $\mathbf{1}'$ and the doublet $\mathbf{2}$. We work in a basis where the S_3 doublet generators are given by:

$$\rho(S) = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix} \quad , \quad \rho(T) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{B.1}$$

$$(\rho(S))^2 = (\rho(T))^2 = (\rho(S)\rho(T))^3 = \mathbb{1} \quad .$$

The level $N = 2$ modular forms of lowest weight transform as a doublet and can be expanded as:

$$\begin{pmatrix} Y_1(\tau) \\ Y_2(\tau) \end{pmatrix}_{\mathbf{2}} = \begin{pmatrix} \dfrac{7}{100} + \dfrac{42}{25}q + \dfrac{42}{25}q^2 + \dfrac{168}{25}q^3 + ... \\ \\ \dfrac{14\sqrt{3}}{25}q^{1/2}(1 + 4q + 6q^2 + ...) \end{pmatrix} , \tag{B.2}$$

where $q \equiv e^{2\pi i \tau}$, and $\tau = \mathrm{Re}\,\tau + i\,\mathrm{Im}\,\tau$. The normalization of our modular forms differs from the one chosen in [10], and it is justified in [11]. We stress that to our knowledge there is no physical prescription on the normalization of modular forms: it is an arbitrary choice. Recent discussions on this issue can be found in refs. [128, 129]. Apart from that, as discussed in the

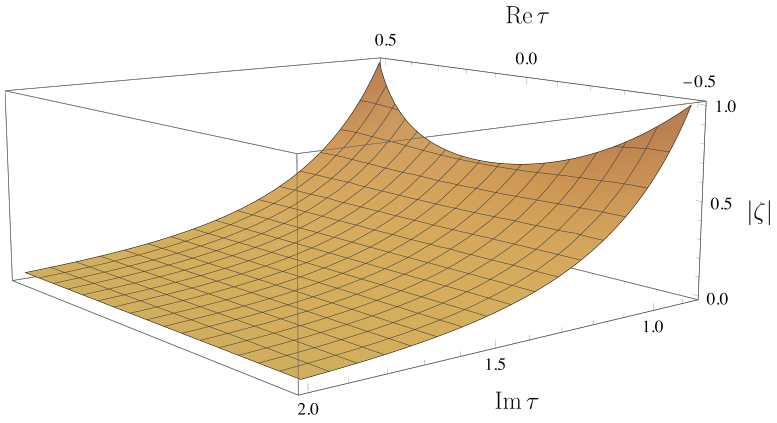


Figure B.1: *Visualization of* $|\zeta| = |Y_2/Y_1|$ *for* $\tau \in \mathcal{D}$.

text, the ratio $|Y_2(\tau)/Y_1(\tau)|$ can be used as an expansion parameter for $\tau \in \mathcal{D}$. This is shown in figure B.1 and the behaviour is obviously independent of the chosen normalization.

Labelling some S_3 pseudo-singlets with y_i for $i = 1, 2, 3...$, and with $\psi_{1,2}, \varphi_{1,2}$ two doublet components, the tensor decomposition rules are given by:

$$\begin{aligned} \mathbf{1}' \otimes \mathbf{1}' = \mathbf{1} \quad &\sim \quad y_1 y_2 \,, \\ \mathbf{1}' \otimes \mathbf{2} = \mathbf{2} \quad &\sim \quad \begin{pmatrix} -y_1 \psi_2 \\ y_1 \psi_1 \end{pmatrix} , \end{aligned} \tag{B.3}$$

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{1}' \oplus \mathbf{2} \begin{cases} \mathbf{1} \quad \sim \quad \psi_1 \varphi_1 + \psi_2 \varphi_2 \\ \mathbf{1}' \quad \sim \quad \psi_1 \varphi_2 - \psi_2 \varphi_1 \\ \mathbf{2} \quad \sim \quad \begin{pmatrix} \psi_2 \varphi_2 - \psi_1 \varphi_1 \\ \psi_1 \varphi_2 + \psi_2 \varphi_1 \end{pmatrix} . \end{cases} \tag{B.4}$$

Appendix C Mapping to $\mathcal{D}$ and rescaling

Points outside the fundamental domain (2.8) are redundant: some examples of modular mappings to $\mathcal{D}$ are given in figure C.1. The act of mapping into $\mathcal{D}$ implies that some free parameters of the superpotential may have to be rescaled, as argued in [33] . The Kähler potential for the matter fields is given by:

$$\sum_I (-i(\tau - \bar{\tau}))^{-k_I} |\varphi^{(I)}|^2 = \sum_I (2\mathrm{Im}\,\tau)^{-k_I} |\varphi^{(I)}|^2 \,, \tag{C.1}$$

and it is modular-invariant since under a modular transformation the relevant quantities transform as:

$$\mathrm{Im}\,\tau \quad \to \quad |c\tau + d|^{-2} \mathrm{Im}\,\tau \qquad , \quad \varphi^{(I)} \quad \to \quad (c\tau + d)^{-k_I} \varphi^{(I)} \,. \tag{C.2}$$

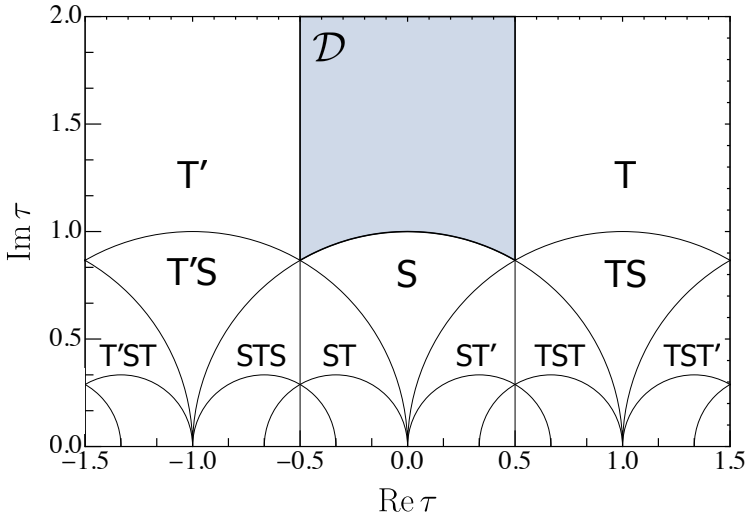


Figure C.1: *Examples of regions outside the fundamental domain $\mathcal{D}$. The mapping is provided by combinations of the generators S, T defined in (2.2). Here $T' \equiv T^{-1}$.*

In order to read the canonical kinetic terms from the potential, the superfields can be rescaled as:

$$\varphi^{(I)} = \tilde{\varphi}^{(I)} (2\mathrm{Im}\,\tau)^{k_I/2} \,, \tag{C.3}$$

and the modular-invariant operators of the superpotential change as:

$$\alpha_i \varphi^{(1)}\varphi^{(2)}\varphi^{(3)} Y_{k_Y} \quad \to \quad \alpha_i (2\mathrm{Im}\,\tau)^{\frac{k_1+k_2+k_3}{2}} \tilde{\varphi}^{(1)}\tilde{\varphi}^{(2)}\tilde{\varphi}^{(3)} Y_{k_Y} \,, \tag{C.4}$$

where α_i is the associated coupling and Y_{k_Y} is the required modular form of weight k_Y. Recall that from the modular constraints on the superpotential we know that $k_1 + k_2 + k_3 = k_Y$. If one performed a numerical scan and extracted best-fit values for α_i, namely $\hat{\alpha}_i$, the kinetic rescaling reads:

$$\hat{\alpha}_i = \alpha_i (2\mathrm{Im}\,\tau)^{k_Y/2} \,. \tag{C.5}$$

Acting with a modular transformation, $\mathrm{Im}\,\tau$ transforms as in (C.2). As a consequence:

$$\alpha_i (2\mathrm{Im}\,\tau)^{k_Y/2} \quad \to \quad \underbrace{\alpha_i (2\mathrm{Im}\,\tau)^{k_Y/2}}_{\hat{\alpha}_i} |c\tau + d|^{-k_Y} \,. \tag{C.6}$$

Thus, the coupling extracted from the numerical scan is mapped into $\hat{\alpha}'_i \equiv |c\tau + d|^{-k_Y} \hat{\alpha}_i$. By modular invariance, the parameter sets $\{\tau, \hat{\alpha}_i\}$ and $\{\gamma(\tau), \hat{\alpha}'_i\}$ lead to the same observables, as can be verified. As a common practice, the fitted parameters are the dimensionless ratios $\beta/\alpha, \gamma/\alpha, g'/g, g''/g$... etc. These are mapped to:

$$\frac{g_i}{g_j} \quad \to \quad \frac{g_i}{g_j} |c\tau + d|^{-k_Y^i + k_Y^j} \,, \tag{C.7}$$

where k_Y^i, k_Y^j are the weights of the modular forms associated with the i-th and j-th operators of the superpotential. In our model, all the operators from the neutrino sector have the same $k_Y^{i,j}$, thus the dimensionless ratios (C.7) are invariant under the modular mapping. The only non-trivial rescaling of dimensionless parameters happens in the charged-leptons sector where, in general, $k_Y^i \neq k_Y^j$ for $i \neq j$. The neutrino global mass-scale $g^2 v_u^2/\Lambda$ is mapped into $|c\tau + d|^{-k_Y} g^2 v_u^2/\Lambda$ where $k_Y = 4$ in our case of eq. (3.8).

References

[1] G. Altarelli and F. Feruglio, "Discrete Flavor Symmetries and Models of Neutrino Mixing," *Rev. Mod. Phys.* **82** (2010) 2701–2729, 1002.0211 [hep-ph].

[2] F. Feruglio and A. Romanino, "Lepton flavor symmetries," *Rev. Mod. Phys.* **93** no. 1, (2021) 015007, 1912.06028 [hep-ph].

[3] F. Feruglio, *Are neutrino masses modular forms?*, pp. 227–266. 2019. 1706.08749 [hep-ph].

[4] J. Lauer, J. Mas, and H. P. Nilles, "Duality and the Role of Nonperturbative Effects on the World Sheet," *Phys. Lett. B* **226** (1989) 251–256.

[5] J. Lauer, J. Mas, and H. P. Nilles, "Twisted sector representations of discrete background symmetries for two-dimensional orbifolds," *Nucl. Phys. B* **351** (1991) 353–424.

[6] K. Ishiguro, T. Kobayashi, and H. Otsuka, "Landscape of Modular Symmetric Flavor Models," *JHEP* **03** (2021) 161, 2011.09154 [hep-ph].

[7] P. P. Novichkov, J. T. Penedo, and S. T. Petcov, "Modular flavour symmetries and modulus stabilisation," *JHEP* **03** (2022) 149, 2201.02020 [hep-ph].

[8] K. Ishiguro, H. Okada, and H. Otsuka, "Residual flavor symmetry breaking in the landscape of modular flavor models," *JHEP* **09** (2022) 072, 2206.04313 [hep-ph].

[9] V. Knapp-Perez, X.-G. Liu, H. P. Nilles, S. Ramos-Sanchez, and M. Ratz, "Matter matters in moduli fixing and modular flavor symmetries," 2304.14437 [hep-th].

[10] T. Kobayashi, K. Tanaka, and T. H. Tatsuishi, "Neutrino mixing from finite modular groups," *Phys. Rev. D* **98** no. 1, (2018) 016004, 1803.10391 [hep-ph].

[11] D. Meloni and M. Parriciatu, "A simplest modular S_3 model for leptons," *JHEP* **09** (2023) 043, 2306.09028 [hep-ph].

[12] T. Kobayashi, Y. Shimizu, K. Takagi, M. Tanimoto, T. H. Tatsuishi, and H. Uchida, "Finite modular subgroups for fermion mass matrices and baryon/lepton number violation," *Phys. Lett. B* **794** (2019) 114–121, 1812.11072 [hep-ph].

[13] J. C. Criado and F. Feruglio, "Modular Invariance Faces Precision Neutrino Data," *SciPost Phys.* **5** no. 5, (2018) 042, 1807.01125 [hep-ph].

[14] T. Kobayashi, N. Omoto, Y. Shimizu, K. Takagi, M. Tanimoto, and T. H. Tatsuishi, "Modular A_4 invariance and neutrino mixing," *JHEP* **11** (2018) 196, 1808.03012 [hep-ph].

[15] H. Okada and M. Tanimoto, "CP violation of quarks in A_4 modular invariance," *Phys. Lett. B* **791** (2019) 54–61, 1812.09677 [hep-ph].

[16] H. Okada and M. Tanimoto, "Towards unification of quark and lepton flavors in A_4 modular invariance," *Eur. Phys. J. C* **81** no. 1, (2021) 52, 1905.13421 [hep-ph].

[17] G.-J. Ding, S. F. King, and X.-G. Liu, "Modular A_4 symmetry models of neutrinos and charged leptons," *JHEP* **09** (2019) 074, 1907.11714 [hep-ph].

[18] T. Kobayashi, Y. Shimizu, K. Takagi, M. Tanimoto, and T. H. Tatsuishi, "A_4 lepton flavor model and modulus stabilization from S_4 modular symmetry," *Phys. Rev. D* **100** no. 11, (2019) 115045, 1909.05139 [hep-ph]. [Erratum: Phys.Rev.D 101, 039904 (2020)].

[19] T. Asaka, Y. Heo, T. H. Tatsuishi, and T. Yoshida, "Modular A_4 invariance and leptogenesis," *JHEP* **01** (2020) 144, 1909.06520 [hep-ph].

[20] G.-J. Ding, S. F. King, X.-G. Liu, and J.-N. Lu, "Modular S_4 and A_4 symmetries and their fixed points: new predictive examples of lepton mixing," *JHEP* **12** (2019) 030, 1910.03460 [hep-ph].

[21] D. Zhang, "A modular A_4 symmetry realization of two-zero textures of the Majorana neutrino mass matrix," *Nucl. Phys. B* **952** (2020) 114935, 1910.07869 [hep-ph].

[22] S. J. D. King and S. F. King, "Fermion mass hierarchies from modular symmetry," *JHEP* **09** (2020) 043, 2002.00969 [hep-ph].

[23] G.-J. Ding and F. Feruglio, "Testing Moduli and Flavon Dynamics with Neutrino Oscillations," *JHEP* **06** (2020) 134, 2003.13448 [hep-ph].

[24] T. Asaka, Y. Heo, and T. Yoshida, "Lepton flavor model with modular A_4 symmetry in large volume limit," *Phys. Lett. B* **811** (2020) 135956, 2009.12120 [hep-ph].

[25] H. Okada and M. Tanimoto, "Spontaneous CP violation by modulus τ in A_4 model of lepton flavors," *JHEP* **03** (2021) 010, 2012.01688 [hep-ph].

[26] C.-Y. Yao, J.-N. Lu, and G.-J. Ding, "Modular Invariant A_4 Models for Quarks and Leptons with Generalized CP Symmetry," *JHEP* **05** (2021) 102, 2012.13390 [hep-ph].

[27] H. Okada, Y. Shimizu, M. Tanimoto, and T. Yoshida, "Modulus τ linking leptonic CP violation to baryon asymmetry in A_4 modular invariant flavor model," *JHEP* **07** (2021) 184, 2105.14292 [hep-ph].

[28] T. Nomura, H. Okada, and Y. Orikasa, "Quark and lepton flavor model with leptoquarks in a modular A_4 symmetry," *Eur. Phys. J. C* **81** no. 10, (2021) 947, 2106.12375 [hep-ph].

[29] M.-C. Chen, V. Knapp-Perez, M. Ramos-Hamud, S. Ramos-Sanchez, M. Ratz, and S. Shukla, "Quasi–eclectic modular flavor symmetries," *Phys. Lett. B* **824** (2022) 136843, 2108.02240 [hep-ph].

[30] T. Nomura, H. Okada, and Y. Shoji, "$SU(4)_C \times SU(2)_L \times U(1)_R$ models with modular A_4 symmetry," 2206.04466 [hep-ph].

[31] Y. Gunji, K. Ishiwata, and T. Yoshida, "Subcritical regime of hybrid inflation with modular A_4 symmetry," *JHEP* **11** (2022) 002, 2208.10086 [hep-ph].

[32] J. T. Penedo and S. T. Petcov, "Lepton Masses and Mixing from Modular S_4 Symmetry," *Nucl. Phys. B* **939** (2019) 292–307, 1806.11040 [hep-ph].

[33] P. P. Novichkov, J. T. Penedo, S. T. Petcov, and A. V. Titov, "Modular S_4 models of lepton masses and mixing," *JHEP* **04** (2019) 005, 1811.04933 [hep-ph].

[34] I. de Medeiros Varzielas, S. F. King, and Y.-L. Zhou, "Multiple modular symmetries as the origin of flavor," *Phys. Rev. D* **101** no. 5, (2020) 055033, 1906.02208 [hep-ph].

[35] T. Kobayashi, Y. Shimizu, K. Takagi, M. Tanimoto, and T. H. Tatsuishi, "New A_4 lepton flavor model from S_4 modular symmetry," *JHEP* **02** (2020) 097, 1907.09141 [hep-ph].

[36] S. F. King and Y.-L. Zhou, "Trimaximal TM_1 mixing with two modular S_4 groups," *Phys. Rev. D* **101** no. 1, (2020) 015001, 1908.02770 [hep-ph].

[37] J. C. Criado, F. Feruglio, and S. J. D. King, "Modular Invariant Models of Lepton Masses at Levels 4 and 5," *JHEP* **02** (2020) 001, 1908.11867 [hep-ph].

[38] X. Wang and S. Zhou, "The minimal seesaw model with a modular S_4 symmetry," *JHEP* **05** (2020) 017, 1910.09473 [hep-ph].

[39] X. Wang, "Dirac neutrino mass models with a modular S_4 symmetry," *Nucl. Phys. B* **962** (2021) 115247, 2007.05913 [hep-ph].

[40] B.-Y. Qu, X.-G. Liu, P.-T. Chen, and G.-J. Ding, "Flavor mixing and CP violation from the interplay of an S_4 modular group and a generalized CP symmetry," *Phys. Rev. D* **104** no. 7, (2021) 076001, 2106.11659 [hep-ph].

[41] P. P. Novichkov, J. T. Penedo, S. T. Petcov, and A. V. Titov, "Modular A_5 symmetry for flavour model building," *JHEP* **04** (2019) 174, 1812.02158 [hep-ph].

[42] G.-J. Ding, S. F. King, and X.-G. Liu, "Neutrino mass and mixing with A_5 modular symmetry," *Phys. Rev. D* **100** no. 11, (2019) 115005, 1903.12588 [hep-ph].

[43] X.-G. Liu and G.-J. Ding, "Neutrino Masses and Mixing from Double Covering of Finite Modular Groups," *JHEP* **08** (2019) 134, 1907.01488 [hep-ph].

[44] X.-G. Liu, C.-Y. Yao, and G.-J. Ding, "Modular invariant quark and lepton models in double covering of S_4 modular group," *Phys. Rev. D* **103** no. 5, (2021) 056013, 2006.10722 [hep-ph].

[45] X. Wang, B. Yu, and S. Zhou, "Double covering of the modular A_5 group and lepton flavor mixing in the minimal seesaw model," *Phys. Rev. D* **103** no. 7, (2021) 076005, 2010.10159 [hep-ph].

[46] C.-Y. Yao, X.-G. Liu, and G.-J. Ding, "Fermion masses and mixing from the double cover and metaplectic cover of the A_5 modular group," *Phys. Rev. D* **103** no. 9, (2021) 095013, 2011.03501 [hep-ph].

[47] P. P. Novichkov, J. T. Penedo, and S. T. Petcov, "Double cover of modular S_4 for flavour model building," *Nucl. Phys. B* **963** (2021) 115301, 2006.03058 [hep-ph].

[48] P. P. Novichkov, J. T. Penedo, and S. T. Petcov, "Fermion mass hierarchies, large lepton mixing and residual modular symmetries," *JHEP* **04** (2021) 206, 2102.07488 [hep-ph].

[49] I. de Medeiros Varzielas, S. F. King, and M. Levy, "Littlest modular seesaw," *JHEP* **02** (2023) 143, 2211.00654 [hep-ph].

[50] G.-J. Ding, X.-G. Liu, and C.-Y. Yao, "A minimal modular invariant neutrino model," *JHEP* **01** (2023) 125, 2211.04546 [hep-ph].

[51] T. Kobayashi and M. Tanimoto, "Modular flavor symmetric models," 7, 2023. 2307.03384 [hep-ph].

[52] G.-J. Ding and S. F. King, "Neutrino Mass and Mixing with Modular Symmetry," 2311.09282 [hep-ph].

[53] M. Fukugita and T. Yanagida, "Baryogenesis Without Grand Unification," *Phys. Lett. B* **174** (1986) 45–47.

[54] R. Barbieri, P. Creminelli, A. Strumia, and N. Tetradis, "Baryogenesis through leptogenesis," *Nucl. Phys. B* **575** (2000) 61–77, arXiv:hep-ph/9911315.

[55] S. Davidson, E. Nardi, and Y. Nir, "Leptogenesis," *Physics Reports* **466** no. 4–5, (Sept., 2008) 105–177. http://dx.doi.org/10.1016/j.physrep.2008.06.002.

[56] W. Buchmuller, P. Di Bari, and M. Plumacher, "Leptogenesis for pedestrians," *Annals Phys.* **315** (2005) 305–351, arXiv:hep-ph/0401240.

[57] A. Pilaftsis and T. E. J. Underwood, "Resonant leptogenesis," *Nucl. Phys. B* **692** (2004) 303–345, arXiv:hep-ph/0309342.

[58] W. Buchmuller, R. D. Peccei, and T. Yanagida, "Leptogenesis as the origin of matter," *Ann. Rev. Nucl. Part. Sci.* **55** (2005) 311–355, arXiv:hep-ph/0502169.

[59] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe," *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35.

[60] J. Gogoi, L. Sarma, and M. K. Das, "Leptogenesis and dark matter in minimal inverse seesaw using A_4 modular symmetry," arXiv:2311.09883 [hep-ph].

[61] G.-J. Ding, S. F. King, J.-N. Lu, and B.-Y. Qu, "Leptogenesis in SO(10) models with A_4 modular symmetry," *JHEP* **10** (2022) 071, 2206.14675 [hep-ph].

[62] D. W. Kang, J. Kim, T. Nomura, and H. Okada, "Natural mass hierarchy among three heavy Majorana neutrinos for resonant leptogenesis under modular A_4 symmetry," *JHEP* **07** (2022) 050, 2205.08269 [hep-ph].

[63] P. Mishra, M. K. Behera, P. Panda, and R. Mohanta, "Type III seesaw under A_4 modular symmetry with leptogenesis," *Eur. Phys. J. C* **82** no. 12, (2022) 1115, 2204.08338 [hep-ph].

[64] A. Dasgupta, T. Nomura, H. Okada, O. Popov, and M. Tanimoto, "Dirac Radiative Neutrino Mass with Modular Symmetry and Leptogenesis," 2111.06898 [hep-ph].

[65] M. Kashav and S. Verma, "Broken scaling neutrino mass matrix and leptogenesis based on A_4 modular invariance," *JHEP* **09** (2021) 100, 2103.07207 [hep-ph].

[66] M. K. Behera, S. Mishra, S. Singirala, and R. Mohanta, "Implications of A4 modular symmetry on neutrino mass, mixing and leptogenesis with linear seesaw," *Phys. Dark Univ.* **36** (2022) 101027, 2007.00545 [hep-ph].

[67] M. K. Behera and R. Mohanta, "Linear Seesaw in A_5' Modular Symmetry With Leptogenesis," *Front. in Phys.* **10** (2022) 854595, 2201.10429 [hep-ph].

[68] P. P. Novichkov, J. T. Penedo, S. T. Petcov, and A. V. Titov, "Generalised CP Symmetry in Modular-Invariant Models of Flavour," *JHEP* **07** (2019) 165, 1905.11970 [hep-ph].

[69] P. H. Frampton, S. L. Glashow, and T. Yanagida, "Cosmological sign of neutrino CP violation," *Phys. Lett. B* **548** (2002) 119–121, hep-ph/0208157.

[70] M. Raidal and A. Strumia, "Predictions of the most minimal seesaw model," *Phys. Lett. B* **553** (2003) 72–78, hep-ph/0210021.

[71] Z.-z. Xing and Z.-h. Zhao, "The minimal seesaw and leptogenesis models," *Rept. Prog. Phys.* **84** no. 6, (2021) 066201, 2008.12090 [hep-ph].

[72] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, "The fate of hints: updated global analysis of three-flavor neutrino oscillations," *JHEP* **09** (2020) 178, 2007.14792 [hep-ph].

[73] **Super-Kamiokande** Collaboration, K. Abe *et al.*, "Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV," *Phys. Rev. D* **97** no. 7, (2018) 072001, 1710.09126 [hep-ex].

[74] S. Antusch, J. Kersten, M. Lindner, and M. Ratz, "Running neutrino masses, mixings and CP phases: Analytical results and phenomenological consequences," *Nucl. Phys. B* **674** (2003) 401–433, hep-ph/0305273.

[75] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles," *Prog. Theor. Phys.* **28** (1962) 870–880.

[76] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge," *Zh. Eksp. Teor. Fiz.* **34** (1957) 247.

[77] P. Novichkov, *Aspects of the Modular Symmetry Approach to Lepton Flavour*. PhD thesis, SISSA, Trieste, 2021.

[78] G. Altarelli and G. Blankenburg, "Different $SO(10)$ Paths to Fermion Masses and Mixings," *JHEP* **03** (2011) 133, 1012.2697 [hep-ph].

[79] **eBOSS** Collaboration, S. Alam *et al.*, "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory," *Phys. Rev. D* **103** no. 8, (2021) 083533, 2007.08991 [astro-ph.CO].

[80] G. Giudice, A. Notari, M. Raidal, A. Riotto, and A. Strumia, "Towards a complete theory of thermal leptogenesis in the sm and mssm," *Nuclear Physics B* **685** no. 1, (2004) 89–149. https://www.sciencedirect.com/science/article/pii/S0550321304001270.

[81] G. 't Hooft, "Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle," *Phys. Rev. D* **14** (1976) 3432–3450. [Erratum: Phys.Rev.D 18, 2199 (1978)].

[82] S. Davidson and A. Ibarra, "A Lower bound on the right-handed neutrino mass from leptogenesis," *Phys. Lett. B* **535** (2002) 25–32, arXiv:hep-ph/0202239.

[83] G. Giudice, A. Notari, M. Raidal, A. Riotto, and A. Strumia, "Towards a complete theory of thermal leptogenesis in the sm and mssm," *Nuclear Physics B* **685** no. 1–3, (May, 2004) 89–149. http://dx.doi.org/10.1016/j.nuclphysb.2004.02.019.

[84] J. Pradler and F. D. Steffen, "Thermal gravitino production and collider tests of leptogenesis," *Phys. Rev. D* **75** (2007) 023509, arXiv:hep-ph/0608344.

[85] J. C. Criado and F. Feruglio, "Modular invariance faces precision neutrino data," *SciPost Physics* **5** no. 5, (Nov., 2018) . http://dx.doi.org/10.21468/SciPostPhys.5.5.042.

[86] R. Samanta and M. Sen, "Flavoured leptogenesis and cp $\mu - \tau$ symmetry," *Journal of High Energy Physics* **2020** no. 1, (Jan., 2020) . http://dx.doi.org/10.1007/JHEP01(2020)193.

[87] S. Blanchet and P. D. Bari, "The minimal scenario of leptogenesis," *New Journal of Physics* **14** no. 12, (Dec, 2012) 125012. https://dx.doi.org/10.1088/1367-2630/14/12/125012.

[88] S. Blanchet and P. D. Bari, "Leptogenesis beyond the limit of hierarchical heavy neutrino masses," *Journal of Cosmology and Astroparticle Physics* **2006** no. 06, (Jun, 2006) 023. https://dx.doi.org/10.1088/1475-7516/2006/06/023.

[89] W. Buchmüller, P. Di Bari, and M. Plümacher, "Leptogenesis for pedestrians," *Annals of Physics* **315** no. 2, (Feb., 2005) 305–351. http://dx.doi.org/10.1016/j.aop.2004.02.003.

[90] C. S. Fong, E. Nardi, and A. Riotto, "Leptogenesis in the Universe," *Adv. High Energy Phys.* **2012** (2012) 158303, arXiv:1301.3062 [hep-ph].

[91] **Planck** Collaboration, N. Aghanim *et al.*, "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641** (2020) A6, arXiv:1807.06209 [astro-ph.CO]. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[92] W. Buchmüller, P. D. Bari, and M. Plümacher, "Some aspects of thermal leptogenesis," *New Journal of Physics* **6** (Aug., 2004) 105–105. http://dx.doi.org/10.1088/1367-2630/6/1/105.

[93] T. Endoh, T. Morozumi, and Z. Xiong, "Primordial lepton family asymmetries in seesaw model," *Progress of Theoretical Physics* **111** no. 1, (Jan., 2004) 123–149. http://dx.doi.org/10.1143/PTP.111.123.

[94] A. Abada, S. Davidson, F.-X. Josse-Michaux, M. Losada, and A. Riotto, "Flavour issues in leptogenesis," *Journal of Cosmology and Astroparticle Physics* **2006** no. 04, (Apr., 2006) 004–004. http://dx.doi.org/10.1088/1475-7516/2006/04/004.

[95] E. Nardi, Y. Nir, E. Roulet, and J. Racker, "The importance of flavor in leptogenesis," *Journal of High Energy Physics* **2006** no. 01, (Jan., 2006) 164–164. http://dx.doi.org/10.1088/1126-6708/2006/01/164.

[96] B. A. Campbell, S. Davidson, J. Ellis, and K. A. Olive, "On the baryon, lepton-flavour and right-handed electron asymmetries of the universe," *Physics Letters B* **297** no. 1–2, (Dec., 1992) 118–124. http://dx.doi.org/10.1016/0370-2693(92)91079-O.

[97] J. M. Cline, K. Kainulainen, and K. A. Olive, "Protecting the primordial baryon asymmetry from erasure by sphalerons," *Physical Review D* **49** no. 12, (June, 1994) 6394–6409. http://dx.doi.org/10.1103/PhysRevD.49.6394.

[98] S. Davidson and A. Ibarra, "A lower bound on the right-handed neutrino mass from leptogenesis," *Physics Letters B* **535** no. 1–4, (May, 2002) 25–32. http://dx.doi.org/10.1016/S0370-2693(02)01735-5.

[99] L. Covi, E. Roulet, and F. Vissani, "Cp violating decays in leptogenesis scenarios," *Physics Letters B* **384** no. 1–4, (Sept., 1996) 169–174. http://dx.doi.org/10.1016/0370-2693(96)00817-9.

[100] K. Hamaguchi, H. Murayama, and T. Yanagida, "Leptogenesis from a n1-dominated early universe," *Physical Review D* **65** no. 4, (Jan., 2002) . http://dx.doi.org/10.1103/PhysRevD.65.043512.

[101] S. Pascoli, S. Petcov, and A. Riotto, "Leptogenesis and low energy cp-violation in neutrino physics," *Nuclear Physics B* **774** no. 1–3, (July, 2007) 1–52. http://dx.doi.org/10.1016/j.nuclphysb.2007.02.019.

[102] G. C. Branco, R. Gonzalez Felipe, M. N. Rebelo, and H. Serodio, "Resonant leptogenesis and tribimaximal leptonic mixing with A(4) symmetry," *Phys. Rev. D* **79** (2009) 093008, arXiv:0904.3076 [hep-ph].

[103] W. Buchmüller and M. Plümacher, "Baryon asymmetry and neutrino mixing," *Physics Letters B* **389** no. 1, (Dec., 1996) 73–77. http://dx.doi.org/10.1016/S0370-2693(96)01232-4.

[104] W. Buchmüller, P. Di Bari, and M. Plümacher, "The neutrino mass window for baryogenesis," *Nuclear Physics B* **665** (Aug., 2003) 445–468. http://dx.doi.org/10.1016/S0550-3213(03)00449-8.

[105] W. Buchmuller, P. Di Bari, and M. Plumacher, "Cosmic microwave background, matter - antimatter asymmetry and neutrino masses," *Nucl. Phys. B* **643** (2002) 367–390, arXiv:hep-ph/0205349. [Erratum: Nucl.Phys.B 793, 362 (2008)].

[106] D. Croon, N. Fernandez, D. McKeen, and G. White, "Stability, reheating and leptogenesis." 2019.

[107] G. F. Giudice, A. Riotto, I. Tkachev, and M. Peloso, "Production of massive fermions at preheating and leptogenesis," *Journal of High Energy Physics* **1999** no. 08, (Aug., 1999) 014–014. http://dx.doi.org/10.1088/1126-6708/1999/08/014.

[108] M. Y. Khlopov and A. D. Linde, "Is It Easy to Save the Gravitino?," *Phys. Lett. B* **138** (1984) 265–268.

[109] S. H.-S. Alexander, M. E. Peskin, and M. M. Sheikh-Jabbari, "Gravi-Leptogenesis: Leptogenesis from Gravity Waves in Pseudo-scalar Driven Inflation Models," *eConf* **C0605151** (2006) 0022, arXiv:hep-ph/0701139.

[110] P. Adshead, A. J. Long, and E. I. Sfakianakis, "Gravitational Leptogenesis, Reheating, and Models of Neutrino Mass," *Phys. Rev. D* **97** no. 4, (2018) 043511, arXiv:1711.04800 [hep-ph].

[111] E. Bertuzzo, P. Di Bari, and L. Marzola *Nuclear Physics B* **849** no. 3, (Aug., 2011) 521–548. http://dx.doi.org/10.1016/j.nuclphysb.2011.03.027.

[112] P. Di Bari and L. Marzola *Nuclear Physics B* **877** no. 3, (Dec., 2013) 719–751. http://dx.doi.org/10.1016/j.nuclphysb.2013.10.027.

[113] M. Chianese and P. Di Bari, "Strong thermal so(10)-inspired leptogenesis in the light of recent results from long-baseline neutrino experiments," *Journal of High Energy Physics* **2018** no. 5, (May, 2018) . http://dx.doi.org/10.1007/JHEP05(2018)073.

[114] J. A. Harvey and M. S. Turner, "Cosmological baryon and lepton number in the presence of electroweak fermion-number violation," *Phys. Rev. D* **42** (Nov, 1990) 3344–3349. https://link.aps.org/doi/10.1103/PhysRevD.42.3344.

[115] B. Dev, M. Garny, J. Klaric, P. Millington, and D. Teresi, "Resonant enhancement in leptogenesis," *Int. J. Mod. Phys. A* **33** (2018) 1842003, arXiv:1711.02863 [hep-ph].

[116] A. Pilaftsis and T. E. J. Underwood, "Electroweak-scale resonant leptogenesis," *Physical Review D* **72** no. 11, (Dec., 2005) . http://dx.doi.org/10.1103/PhysRevD.72.113001.

[117] Z.-z. Xing and S. Zhou, "Tri-bimaximal neutrino mixing and flavor-dependent resonant leptogenesis," *Physics Letters B* **653** no. 2–4, (Sept., 2007) 278–287. http://dx.doi.org/10.1016/j.physletb.2007.08.009.

[118] G. Arcadi, S. Marciano, and D. Meloni, "Neutrino mixing and leptogenesis in a $l_e - l_\mu - l_\tau$ model," *The European Physical Journal C* **83** no. 2, (Feb., 2023) . http://dx.doi.org/10.1140/epjc/s10052-023-11268-6.

[119] K.-Y. Jung and K. Siyeon, "Light sterile neutrino and leptogenesis," *Journal of the Korean Physical Society* **81** no. 12, (Dec., 2022) 1211–1224. http://dx.doi.org/10.1007/s40042-022-00674-w.

[120] S. Blanchet and P. Di Bari, "New aspects of leptogenesis bounds," *Nucl. Phys. B* **807** (2009) 155–187, arXiv:0807.0743 [hep-ph].

[121] **DUNE** Collaboration, B. Abi *et al.*, "Long-baseline neutrino oscillation physics potential of the DUNE experiment," *Eur. Phys. J. C* **80** no. 10, (2020) 978, 2006.16043 [hep-ex].

[122] **Hyper-Kamiokande Proto-** Collaboration, 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, 1502.05199 [hep-ex].

[123] **JUNO** Collaboration, F. An *et al.*, "Neutrino Physics with JUNO," *J. Phys. G* **43** no. 3, (2016) 030401, 1507.05613 [physics.ins-det].

[124] **IceCube-Gen2** Collaboration, M. G. Aartsen *et al.*, "Combined sensitivity to the neutrino mass ordering with JUNO, the IceCube Upgrade, and PINGU," *Phys. Rev. D* **101** no. 3, (2020) 032006, 1911.06745 [hep-ex].

[125] **KamLAND-Zen** Collaboration, S. Abe *et al.*, "Search for the Majorana Nature of Neutrinos in the Inverted Mass Ordering Region with KamLAND-Zen," *Phys. Rev. Lett.* **130** no. 5, (2023) 051801, 2203.02139 [hep-ex].

[126] **KATRIN** Collaboration, M. Aker *et al.*, "Direct neutrino-mass measurement with sub-electronvolt sensitivity," *Nature Phys.* **18** no. 2, (2022) 160–166, 2105.08533 [hep-ex].

[127] E. W. Kolb and M. S. Turner, *The Early Universe*, vol. 69. 1990.

[128] I. de Medeiros Varzielas, M. Levy, J. T. Penedo, and S. T. Petcov, "Quarks at the modular S_4 cusp," 2307.14410 [hep-ph].

[129] S. T. Petcov, "On the Normalisation of the Modular Forms in Modular Invariant Theories of Flavour," 2311.04185 [hep-ph].

ULB-TH/24-01

Dark matter bound-state formation in the Sun

Xiaoyong Chu,[1,*] Raghuveer Garani,[2,†] Camilo García-Cely,[3,‡] and Thomas Hambye[4,5,§]

[1] *Institute of High Energy Physics, Austrian Academy of Sciences, Nikolsdorfer Gasse 18, 1050 Vienna, Austria*

[2] *INFN Sezione di Firenze, Via G. Sansone 1, I-50019 Sesto Fiorentino, Italy*

[3] *Instituto de Física Corpuscular (IFIC), Universitat de València-CSIC, Parc Científic UV, C/ Catedrático José Beltrán 2, E-46980 Paterna, Spain*

[4] *Service de Physique Théorique Université Libre de Bruxelles, Boulevard du Triomphe, CP225, 1050 Brussels, Belgium*

[5] *CERN, Theoretical Physics Department, Geneva, Switzerland*

arXiv:2402.18535v1 [hep-ph] 28 Feb 2024

The Sun may capture asymmetric dark matter (DM), which can subsequently form bound-states through the radiative emission of a sub-GeV scalar. This process enables generation of scalars without requiring DM annihilation. In addition to DM capture on nucleons, the DM-scalar coupling responsible for bound-state formation also induces capture from self-scatterings of ambient DM particles with DM particles already captured, as well as with DM bound-states formed in-situ within the Sun. This scenario is studied in detail by solving Boltzmann equations numerically and analytically. In particular, we take into consideration that the DM self-capture rates require a treatment beyond the conventional Born approximation. We show that, thanks to DM scatterings on bound-states, the number of DM particles captured increases exponentially, leading to enhanced emission of relativistic scalars through bound-state formation, whose final decay products could be observable. We explore phenomenological signatures with the example that the scalar mediator decays to neutrinos. We find that the neutrino flux emitted can be comparable to atmospheric neutrino fluxes within the range of energies below one hundred MeV. Future facilities like Hyper-K, and direct DM detection experiments can further test such scenario.

* xiaoyong.chu@oeaw.ac.at

† garani@fi.infn.it

‡ camilo.garcia@ific.uv.es

§ thomas.hambye@ulb.be

CONTENTS

I. Introduction 2

II. Number evolution of DM particles 3
A. Determination of the various rates 5
B. Integrating the Boltzmann equations 10

III. Neutrino flux and Terrestrial detection 13

IV. Results and constraints for an explicit model 14
A. Model 14
B. Constraints and results 15

V. Summary 19

A. Approximate parametric solution to number evolution of DM particles 20
1. Case with no self-capture on DM bound-states 21
2. Case with self-capture on DM bound-states 22

B. (Non-)perturbative treatment of elastic DM scattering 24
1. Integrated DM self-scattering in DM halos 25
2. Differential DM-DM(BS) scattering for DM capture 26

C. DM thermalization 28

References 29

I. INTRODUCTION

Wherever dark matter (DM) particles are numerous, it is possible that two or more of them form bound-states [1–20]. This could occur if DM undergoes attractive self-interactions mediated by a scalar or vector boson. In this case bound-state formation (BSF) can occur radiatively via the emission of this mediator, leading to an observable flux of particles. Indirect signals of BSF in the center of the Milky Way has been studied for DM with no particle-antiparticle asymmetry, as well as for asymmetric DM [20–23]. In this work we consider the possibility that the BSF process occurs in the Sun and study the corresponding DM indirect detection signals.

Symmetric DM accumulating in the Sun from its capture on nucleons and their corresponding indirect detection signals due to emitted meta-stable mediators have been studied in numerous works, see e.g. [24–41]. In these scenarios, the accumulation of DM particles in the Sun reaches a saturation point when the annihilation rate matches the capture rate. This is not the case of asymmetric DM scenarios as DM cannot annihilate. Interestingly, this absence of annihilation permits a greater accumulation of DM particles. However, it comes with the drawback of not generating any indirect detection signal.

The BSF of asymmetric DM particles in the Sun from radiative emission of a light scalar has the interesting property of allowing both DM indirect detection and large accumulation of DM particles in the Sun. A crucial aspect of this scenario is that the DM-scalar interaction that is needed for BSF inherently implies that DM capture results not only from interactions with nucleons but also from interactions with previously accumulated DM particles and DM bound-states (DMBS). This means that the capture rate is larger than that of the usual symmetric or asymmetric scenarios in which the capture only arises from DM-nucleon scatterings. This allows for a larger accumulation and an enhanced flux of emitted particles. To quantify this effect, it will be necessary to compute the rates for DM-DM and DM-DMBS scatterings, which –as we will show– receive non-perturbative contributions which can be calculated in the semi-classical approximation. To our knowledge this possibility that asymmetric DM particles form bound-states in the Sun and that DM particles are captured by scattering off DM bound-states has not been considered before.[1]

For concreteness, in this work we assume that the associated light scalars, once emitted when the bound-states form, decay into SM particles, in particular to neutrinos. As is well known, unlike other SM particles, low energy neutrinos can escape the Sun leading to observable signatures even if the decay takes place inside the Sun. After solving the set of Boltzmann equations that describe the dynamics of DM accumulation in the Sun in section II, we show in section III that an observable neutrino flux could arise from the decays of the emitted mediator via the BSF of those accumulated DM. Finally, in section IV we discuss relevant constraints from both astrophysical observations and terrestrial experiments, such as DM self-interactions, BBN, CMB, and direct/indirect searches.

II. NUMBER EVOLUTION OF DM PARTICLES

Before considering quantitatively any concrete model, this section introduces the basic relevant processes and related Boltzmann equations determining the number of DM particles captured in

[1] Particle-antiparticle BSF inside the Sun for symmetric DM was considered in [14].

the Sun. The framework of our interest is based on an interaction between the DM particle and a lighter particle. For definiteness in the following we consider that DM is made of Dirac fermions χ, with a Yukawa coupling g_s to a lighter scalar particle ϕ

$$\mathcal{L} \supset -g_s \phi \bar{\chi} \chi \,. \tag{1}$$

To assume a scalar mediator is convenient because it induces an attractive interaction.[2] We also assume that DM is asymmetric, thus it does not annihilate in the present epoch. Nevertheless, in addition to the usual capture due to scattering on target nuclei ($C_\star$) [44, 45], three additional terms come into play in the Boltzmann equation determining the DM particle number N_χ, due to DM-ϕ interactions. First, depending on the strength of self-interaction, DM particles could efficiently form χ-χ bound-states by emitting the particle mediating DM self-interactions. The rate of the latter is denoted by $A_{\rm bsf}$. Since the number of free particles is reduced by two per process, the term $-2 \cdot \frac{1}{2} A_{\rm bsf} N_\chi^2$ is introduced in the equation that describes the evolution of N_χ, where the second factor, $\frac{1}{2}$, counts for double-counting of identical initial states. In contrast to the case of annihilation, in this scenario DM particles are not lost by the system. In effect, there is a build-up of DM bound-states within the celestial body, whose number $N_{2\chi}$ is determined by a second Boltzmann equation that simply involves a $+\frac{1}{2} A_{\rm bsf} N_\chi^2$ term. Second, there is a term coming from capture of Galactic DM on already-captured DM particles, whose rate is denoted by C_χ. Finally, there is a term from capture of Galactic DM on the formed bound-states, denoted by $C_{2\chi}$. The presence of both C_χ and $C_{2\chi}$ terms increase DM accretion. In summary, both populations evolve according to the following set of differential equations,

$$\frac{{\rm d}N_\chi}{{\rm d}t} = C_\star - A_{\rm bsf} N_\chi^2 + C_\chi N_\chi + C_{2\chi} N_{2\chi} \,, \tag{2}$$

$$\frac{{\rm d}N_{2\chi}}{{\rm d}t} = \frac{1}{2} A_{\rm bsf} N_\chi^2 \,. \tag{3}$$

The various rates $C_\star$, C_χ and $C_{2\chi}$ take into account the fact that the three kinds of DM capture occur within different spheres around the solar center. Here we have neglected self-capture via direct BSF between a galactic DM and a captured DM particle, as such inelastic scattering is typically much weaker than the elastic scattering with two DM initial states. Also, in these Boltzmann equations we do not write down explicitly additional terms coming from possible evaporation of the captured DM particles. We have checked that the DM mass above which evaporation is negligible is not significantly different than in the standard scenario (no-self interaction effects).

[2] Such interactions do not result in the formation of mini black hole in the Sun, i.e. the Chandrasekhar limit for fermions is not modified [42, 43].

The value is found to be $m_{\rm evap} \approx 5$ GeV, which is about 50% larger than that in the standard scenario [34]. Therefore, in the following phenomenological discussion we consider only DM masses above 5 GeV. Once DM particles are captured they thermalize with the solar material, leading to efficient formation of bound-states. We always assume that the thermalization happens quickly, and refer to Appendix C for detailed calculation of the thermalization process.

The number of DM particles captured in any of the three ways cannot be larger than the corresponding geometric rates, as the latter assume that all DM particles crossing the corresponding thermal spheres are captured. This is taken into account through the following conservative matching for the regimes of small optical depth (thin) to large (thick)[3]

$$C_\chi N_\chi \;\to\; C_\chi N_\chi \Theta(C^g_\chi - C_\chi N_\chi) + C^g_\chi \Theta(C_\chi N_\chi - C^g_\chi)\;, \tag{4}$$

$$C_{2\chi} N_{2\chi} \;\to\; C_{2\chi} N_{2\chi} \Theta(C^g_{2\chi} - C_{2\chi} N_{2\chi}) + C^g_{2\chi} \Theta(C_{2\chi} N_{2\chi} - C^g_{2\chi})\;. \tag{5}$$

where the respective geometric rates on the DM and DMBS thermal spheres are denoted by C^g_χ and $C^g_{2\chi}$. These geometric rates determine the maximal possible capture rates, independently of the underlying particle model.

Note that we do not consider bound-states containing more than two χ particles, under the assumption that there exist bottlenecks to form heavier bound-states, such as fast decays of $(3\chi) \to (2\chi) + \chi$ and $(4\chi) \to (2\chi) + (2\chi)$. See e.g. [46–49] for further discussions. In practice, if there is formation of many-body bound-states, each captured DM particle may cause the emission of a few more mediator particles while thermal radii for heavier bound-states shrink by a factor of few. The total effect at most modifies our results mildly.[4]

We begin in the next subsection by determining the various rates. The reader seeking immediate understanding of their interplay in the Boltzmann equations can directly refer to section II B.

A. Determination of the various rates

a. Thermal radius and geometric rates $C^g_\star$, C^g_χ and $C^g_{2\chi}$: Once a DM particle has been captured in the Sun, or a DMBS has formed, these particles will thermalize with the SM material of the Sun and lie within different spheres of thermal radius $r_{\rm th}$. Noting that the mass of a bound-state is approximately twice that of a DM particle, these radii are obtained [44, 50] by equating the average

[3] These equations are accurate if only one geometric rate can be saturated. This will be the case studied here, for which the geometric capture rates on free DM (C^g_χ) and on nucleons ($C^g_\star$) are never reached.

[4] The presence of stable many-body bound-states, as well as self-capture via inelastic scattering, may reduce DM evaporation efficiency by quickly capturing free χ particles into heavier bound-states. This can have qualitative consequences for DM candidates below GeV or those captured by Earth, which is left for a future study.

thermal energy, $3T_\odot/2$, to the gravitational potential energy per particle, $2\pi G r_{\rm th}^2 \rho_\odot n m_\chi/3$, as

$$r_{\rm th} = \left(\frac{9T_\odot}{4\pi G\rho_\odot n m_\chi}\right)^{\frac{1}{2}} = 0.03 R_\odot \left(\frac{T_\odot}{2.2\,{\rm keV}}\frac{150\,\frac{\rm g}{\rm cm^3}}{\rho_\odot}\frac{10\,{\rm GeV}}{n\,m_\chi}\right)^{\frac{1}{2}}, \quad \text{with } n=1(2) \text{ for DM (DMBS).} \quad (6)$$

That is, the DM bound-state thermal radius is smaller than the DM thermal radius by a factor of $\sqrt{2}$. We take the solar core temperature, $T_\odot$, to be 2.2 keV, and the core mass density to be $\rho_\odot \sim 150\,{\rm g/cm^3}$. For example, the thermal radius of 3 GeV DM particles is about one-tenth of the Solar radius, $R_\odot \simeq 6.9\times10^5\,{\rm km}$ [51]. In addition, we will assume throughout that the DM radial distribution is isothermal, with its temperature $T \approx T_\odot$ [44, 52].

Upon taking into account the relative motion of the Sun with respect to the galactic DM halo, the geometric capture rate on nucleons is [53]

$$C_\star^g = \frac{\rho_\chi}{m_\chi}\pi R_\odot^2\,\bar{v} \simeq 6\times10^{28}\,{\rm s^{-1}}\left(\frac{\rho_\chi}{0.3\,{\rm GeV/cm^3}}\cdot\frac{10\,{\rm GeV}}{m_\chi}\right). \quad (7)$$

Here $\bar{v}$ is a factor with units of velocity, accounting for the relative motion of the Sun and the velocity distribution of DM, see e.g. [52]. Throughout this work we take the local DM density $\rho_\chi = 0.3\,{\rm GeV/cm^3}$. Apparently, the geometric capture rate is proportional to the corresponding cross-sectional area. So this expression can be rescaled to obtain the geometric rate for DM self-capture on DM and on DMBS (i.e. when the mean free path of DM is smaller than the corresponding thermal sphere) by adding the corresponding factor $(r_{\rm th}/R_\odot)^2$. This results in geometric self-capture rates which scale as m_χ^{-2},

$$C_\chi^g \;=\; 5\times10^{25}\,{\rm s^{-1}}\left(\frac{10\,{\rm GeV}}{m_\chi}\right)^2, \qquad C_{2\chi}^g \approx 2.6\times10^{25}\,{\rm s^{-1}}\left(\frac{10\,{\rm GeV}}{m_\chi}\right)^2. \quad (8)$$

b. *Capture on nucleons rate $C_\star$:* The capture rate of DM particles in the Sun has been well studied in the literature. It can be written as [25]

$$C_\star \;\approx\; \sum_i \int_0^{R_\odot} 4\pi r^2 n_i(r){\rm d}r \int_0^\infty {\rm d}u_\chi \left(\frac{\rho_\chi}{m_\chi}\right)\times\, u_\chi\,\omega^2(r) f_\odot(u_\chi)\int_{E_R^{min}}^{E_R^{max}}\frac{{\rm d}\sigma_i}{{\rm d}E_R}\,{\rm d}E_R\,, \quad (9)$$

where $f_\odot(u_\chi)$ is the normalized asymptotic DM velocity distribution far from the Sun in the solar frame. The values of radial number density distribution of each element i, denoted as $n_i(r)$ above, are adopted from the AGSS09 Solar model [51]. Together with the escape velocity at a distance r away from the centre of the Sun, $v_e(r)$, it provides the relative velocity of a DM particle when it scatters with the nucleus, $\omega(r) = \sqrt{u_\chi^2 + v_e(r)^2}$. The differential cross section, ${\rm d}\sigma_i/{\rm d}E_R$, encodes the energy dependence of elastic scattering between DM and nucleus i in the non-relativistic limit. The recoil energy in the solar frame is given by $E_R = \mu_{\chi N}^2/m_N \times w^2(1-\cos\theta_{CM})\,,$

where $\mu_{\chi N} = m_\chi m_N/(m_\chi + m_N)$ is the reduced mass of the DM-nucleus system. For the Solar capture to actually occur, we need to make sure that the recoil energy lies within the range of $E_R^{min} < E_R < E_R^{max}$, with

$$E_R^{min} = \frac{1}{2} m_\chi u_\chi^2, \quad E_R^{max} = \frac{2\mu_{\chi N}^2 \omega^2}{m_N}\,. \tag{10}$$

For definiteness we assume that the dark sector communicates with the SM particles through the Higgs portal, and that the light mediator ϕ mixes with the Higgs boson with a mixing angle θ_ϕ. The differential cross section in the Born approximation is then

$$\frac{\mathrm{d}\sigma}{\mathrm{d}E_R} = \frac{g_s^2 \cos^2\theta_\phi \sin^2\theta_\phi}{2\pi} \frac{m_N^2 f_N^2}{v_H^2} \frac{m_N}{\omega^2 (2 m_N E_R + m_\phi^2)^2} F^2\left(\frac{E_R}{Q_0}\right)\,, \tag{11}$$

where $F(E_R/Q_0)$ is the Helm form factor and $f_N = f_u^N + f_d^N + f_s^N + 6/27 f_G$, which is an $\mathcal{O}(1)$ number (see e.g. [54, 55]).

The DM capture rates on nucleons are shown in Fig. 1, corresponding to a Higgs mixing angle $\sin\theta_\phi = 10^{-10}$ and $g_s = 1$, for several mediator masses (labeled in red). The geometric capture and self-capture rates, Eq. (7) and (8), are shown as black dashed and dotted curves. For mediator mass below MeV the differential cross section is independent of the mediator mass since $m_N E_R^{max} \sim \mu_{\chi N} v_e^2$, which is above MeV for $m_\chi \geq m_{\rm evap}$. This leads to a t-channel enhancement, which in the $m_\chi \gg m_N$ limit gives a capture cross section scaling as m_χ^{-1}, leading to a capture rate scaling as m_χ^{-2}. While, for mediator masses larger than ~ 100 MeV and $m_\chi \lesssim m_N$, the capture rate in turn scales as m_χ. Note that in Eq. (11) we have neglected the effect of DM-electron scattering, which is subleading due to lower energy loss and tiny electron Yukawa coupling.

c. Bound-state formation rate A_{bsf} As stated above, we consider DM to be asymmetric and made of Dirac fermions (χ). The interaction between two (identical) DM particles is attractive when the mediator is a scalar (ϕ), with a classical Yukawa potential $V(r) = -\frac{\alpha}{r} e^{-m_\phi r}$. The binding energy of the bound-state is $E_{\rm bind} \simeq \frac{m_\chi \alpha^2}{4} - \alpha m_\phi$, with the dark coupling $\alpha \equiv g_s^2/4\pi$. From the merging of two identical fermions, the radiative BSF rate has been estimated in [56], yielding

$$\sigma_{\rm BSF} v \simeq \frac{256\pi^2 \alpha^5}{5 e^4 m_\chi^2 v_{\rm rel}}\,, \tag{12}$$

in the limit of $\alpha/v_{\rm rel} \gg 1$ and $m_\phi \to 0$, where the DM relative velocity is $v_{\rm rel} \simeq 2 v_{\chi,\odot}$. The recoil velocity of the DMBS is $\alpha^2/8$ in this limit, which must be below the escape velocity from the Solar center, resulting in the requirement $\alpha \lesssim 0.18$.[5] Finally, the total annihilation rate that enters the

[5] Due to velocity enhancement, the BSF process in the early Universe could be efficient such that very few free DM particles could exist in the current epoch. Then qualitative changes in some regions of parameter space of interest are expected. However, this depends on the assumed cosmological history, and it is not explored in this work. See e.g. [49, 56] for related discussions.

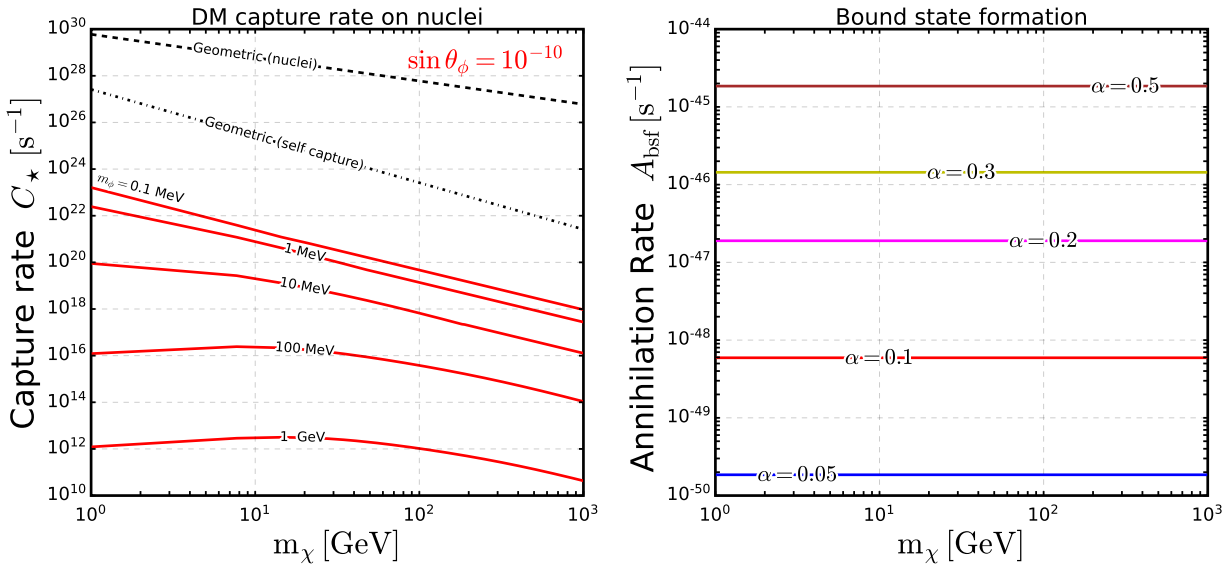


FIG. 1. *Left*: DM capture rate on nucleons for several mediator masses for a Higgs mixing angle $\sin\theta_\phi = 10^{-10}$. The geometric capture on nucleons $C_\star^g$ and C_χ^g self-capture rates are shown by the dashed and dashed-dotted black curves. The $C_{2\chi}^g$ self-capture rate line (not shown) is a factor 2 below the C_χ^g line. *Right*: The DM BSF rate inside the Sun, for several values of α.

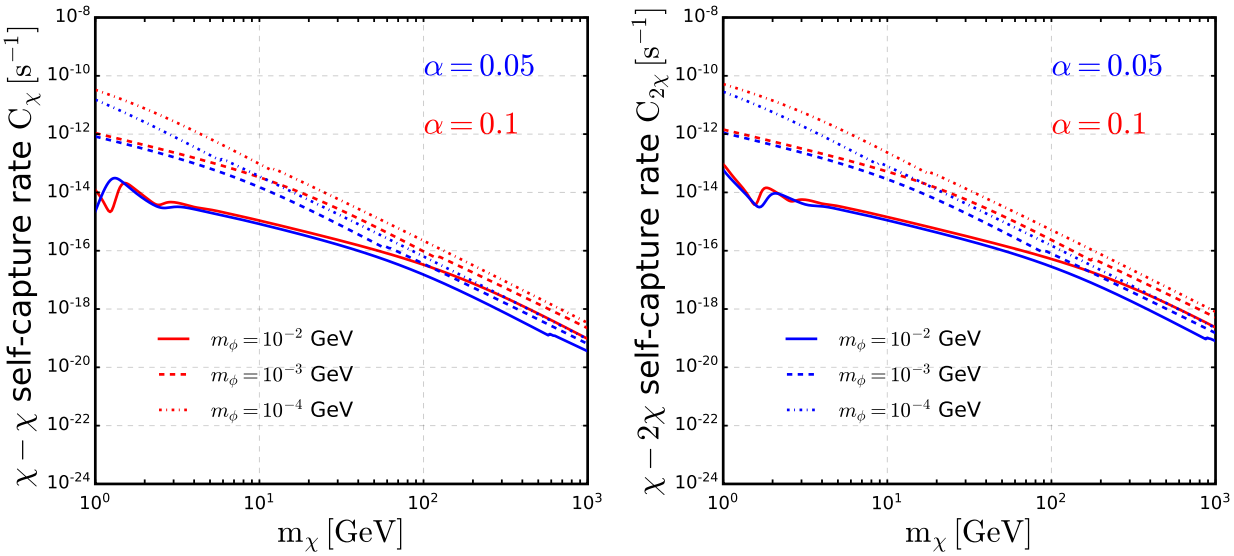


FIG. 2. *Left* (*right*): DM self-capture rates due to DM-DM (DM-DMBS) scattering. See text for details.

number evolution equation is given by

$$A_{\rm bsf} = \frac{\int n_\chi^2\, \sigma_{\rm BSF} v\, dV}{\left(\int n_\chi dV\right)^2} \,. \tag{13}$$

The normalized DM radial distribution is

$$n_\chi(r,t) = N_\chi(t)\, \frac{e^{-m_\chi \phi(r)/T}}{\int_0^{R_\odot} e^{-m_\chi \phi(r)/T}\, 4\pi r^2\, {\rm d}r} \,, \tag{14}$$

which corresponds to an isothermal sphere, with a radial dependence set by the gravitational potential $\phi(r) = \int_0^r GM_\odot(r')/r'^2\,\mathrm{d}r'$, with G the gravitational constant and $M_\odot(r')$ the mass inside a sphere of radius r', and $N_\chi(t)$ is the total population of DM particles at a given time t.

Once DM particles are thermalized its average velocity is $v_\chi \simeq \sqrt{2T/m_\chi}$. The smallness of this value considerably boosts the BSF in Eq. (12): $v_{rel} \simeq 4.2\times10^{-3}\sqrt{1\,\mathrm{GeV}/m_\chi}$. Consequently, $A_{\rm bsf}$ scales as $\alpha^5(m_\chi^2 v_{\rm rel})^{-1}(r_{\rm th}^\chi)^{-3}$, be approximately independent of the DM mass in practice, as shown in right panel of Fig. 1.

d. DM self capture rates C_χ and $C_{2\chi}$: Analogous to DM-nucleon capture rate, we estimate the self-capture rate per target DM particle in the Sun as follows

$$C_\chi = \int_0^{r_{th}^\chi} \mathrm{d}r 4\pi r^2 n_\chi(r) \int_0^\infty \mathrm{d}u_\chi \left(\frac{\rho_\chi}{m_\chi}\right) u_\chi\, f_\odot(u_\chi)\, w^2 \int_{\cos\theta_{\rm min}}^{\cos\theta_{\rm max}} \frac{\mathrm{d}\sigma_{\chi-\chi}}{\mathrm{d}\cos\theta}\,\mathrm{d}\cos\theta\ . \tag{15}$$

The minimum scattering angle is set by the requirement of a minimum energy that has to be lost in a single scattering event to be captured, i.e. DM kinetic energy at infinity must be smaller than $1/2\,m_\chi u_\chi^2$. The maximum scattering angle for scattering of identical particles is $\pi/2$. Therefore $\cos\theta_{\rm min} = 1 - 2\frac{u_\chi^2}{\omega^2}$ and $\cos\theta_{\rm max} = 0$. Similar to the DM-DM self-capture, the capture rate due to DM-DMBS scattering, per target DMBS in the Sun, is given by

$$C_{2\chi} = \int_0^{r_{th}^{2\chi}} \mathrm{d}r 4\pi r^2 n_{2\chi}(r) \int_0^\infty \mathrm{d}u_\chi \left(\frac{\rho_\chi}{m_\chi}\right) u_\chi\, f_\odot(u_\chi)\, \omega^2 \int_{\cos\theta_{\rm min}}^{\cos\theta_{\rm max}} \frac{\mathrm{d}\sigma_{2\chi-\chi}}{\mathrm{d}\cos\theta}\, F_\chi^2\left(\frac{E_r}{Q_\chi}\right)\mathrm{d}\cos\theta\ , \tag{16}$$

with $\cos\theta_{\rm min} = 1 - \frac{9}{8}\frac{u_\chi^2}{\omega^2}$ and $\cos\theta_{\rm max} = -1$. We have used $m_{2\chi} \approx 2m_\chi$ and set that the coupling to bound-state is 2α (as the vertex is of the scalar type). We assume that the form factor has the form of $\exp(E_r/Q_\chi)$. The typical size of the bound-state is set by the Bohr radius, hence $Q_\chi = m_\chi\alpha$. To a good approximation $F_\chi \to 1$ for a typical value of $\alpha = 0.1$. The radial number density of DMBS ($n_{2\chi}$) is assumed to be isothermal, analogous to n_χ.

Bound-state formation suggests that non-perturbative effects are non-negligible for the calculation of the DM self-scattering cross section (see e.g. [57, 58]). This is indeed our case. For the parameter values of interest $\alpha \in [0.02, 0.2]$, $m_\chi > 5$ GeV, MeV $\lesssim m_\phi \ll m_\chi$, and $v \sim v_{esc} \sim 4\times10^{-3}$, we are never in the Born regime of self-scattering, for which $m_\chi < m_\phi/\alpha$ or $\alpha \lesssim v$ [59]. As a result, a perturbative expansion of the scattering cross section is not justified here, see left panel of Fig. 7 in Appendix B 1. However, for these parameters above, it is not necessary to solve the Schrödinger equation associated with the non-relativistic DM collisions because the scattering process is typically semi-classical, for which the range of the potential is larger than the DM de Broglie

wavelength, that is, $m_\chi v/m_\phi \gtrsim 1$. In this case classical mechanics can be employed to estimate the scattering cross section.[6]

More importantly, unlike phenomenological studies of DM self-scattering in galaxies and galaxy clusters, where the transfer and viscosity cross sections are relevant, here the differential cross section is needed to calculate the integrated self-capture rates introduced above. This is because DM capture needs that enough initial kinetic energy is lost in a single scattering. This condition is imposed by integrating the differential cross section from minimum possible scattering angle ($\theta_{\rm min}$, set by required energy loss) to the maximum one where the scattered particle still does not gain enough energy to escape. These limits are explicitly indicated in Eqs. (15) and (16). To calculate the differential cross sections, we follow Refs. [60, 61] and solve the elastic scattering with a Yukawa potential classically. Further discussions and full expressions are presented in Appendix B.

We present the results for DM self capture rates through scattering on free DM particles (DM bound-states) in the left (right) panel of Fig. 2. The rates scale proportionally as m_χ^{-2} for $m_\chi \gg m_\phi$. For moderate values of DM masses the scaling is less steep. Note that the capture rate on bound-states $C_{2\chi}$ is approximately larger by a factor of two with respect to C_χ, due to larger Yukawa coupling induced by bound-states and larger maximal allowed scattering angle that keep both particles gravitationally captured.

B. Integrating the Boltzmann equations

The set of coupled Boltzmann equations of Eqs. (2) and (3) has no closed form analytical solution. In Appendix A we describe at length how they can be solved approximately. We summarize in this subsection the main outcome of this discussion. To this end, in Fig. 3 we present the evolution of both populations for two parameter sets. The evolution of N_χ ($N_{2\chi}$) from the numerical integration of the full Boltzmann Eqs. (2) and (3) are given by the solid (dashed) red curves. Curves with other colors correspond to what is obtained switching off both the C_χ and $C_{2\chi}$ terms (black), or only the C_χ term (blue) or only the $C_{2\chi}$ term (green).

At early times the number of free particles N_χ grows as $C_\star\, t$ (i.e. red and black solid lines coincide). As N_χ increases, the BSF starts to occur efficiently and the corresponding term (quadratic in N_χ) quickly catches up with the constant term associated with capture on nucleons, so that in absence of the $C_{2\chi}$ term, a quasi-static equilibrium between both terms is reached (see black and green solid curves in Fig. 3). Solving the Boltzmann equation for N_χ one gets that the associated

[6] A precise treatment would require to solve the phase shift for a large number of partial waves, see e.g. [24, 57, 58].

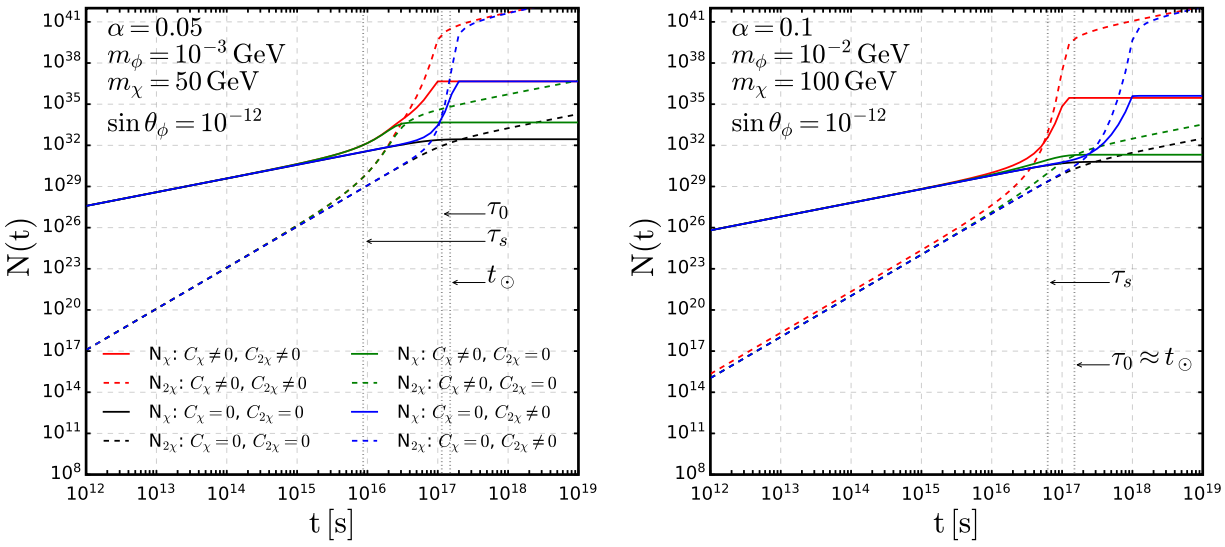


FIG. 3. Number evolution of DM and DM bound-states for two parameter sets, as a function of time. Here τ_s and τ_0 are characteristic times explained in the main text (and exactly defined in Appendix A), $t_\odot$ is the age of the Sun. The color code of the curves in the left panel are the same as that in the right panel.

time scale is $\tau_0 \equiv (C_\star A_{\rm bsf})^{-1/2}$ if one drops the C_χ term –which has a subleading effect (black solid curve)– or $\tau_s \equiv (C_\star A_{\rm bsf} + C_\chi^2/4)^{-1/2}$ if one includes it (green solid curve). For more details, see the end of this section and Appendix A.

Another consequence of efficient BSF is that the number of bound-states $N_{2\chi}$ quickly catches up to the number of DM particles N_χ, and never stops growing thereafter, as it is not counterbalanced by any other term in Eq. (3). From this point, the term associated with the capture on bound-states, $C_{2\chi}N_{2\chi}$, takes over and leads to an exponential growth of N_χ (red and blue solid curves in Fig. 3) and of $N_{2\chi}$ (dashed curves with same colors). The exponential growth become significant at $t \gtrsim \tau_s + C_{2\chi}^{-1}$, or at $\tau_0 + C_{2\chi}^{-1}$, depending on whether one includes the C_χ term. This growth lasts until the capture rate saturates the geometric rate within the bound-state thermal sphere, i.e. when the term $C_{2\chi}N_{2\chi}$ reaches $C_{2\chi}^g$. This happens when the mean free path of DM particles becomes much smaller than the bound-state thermal radius. At this moment, we can safely neglect the $C_\chi N_\chi$ term as $N_\chi \ll N_{2\chi}$, thus the Boltzmann equation takes the form, $dN_\chi/dt \simeq C_\star + C_{2\chi}^g - A_{\rm bsf}N_\chi^2$. This means that, quickly after the geometric capture rate is saturated, the BSF term compensates the constant capture rate from both nucleon and bound-state terms. Setting $dN_\chi/dt \simeq 0$ gives the maximum final value

$$N_{\chi,\rm eq} \simeq \left(\frac{C_\star + C_{2\chi}^g}{A_{\rm bsf}}\right)^{1/2} . \tag{17}$$

We denote as τ_g the time when N_χ freezes in such a way. An analytic approximation is given in Appendix A.

Note that, contrary to the $C_{2\chi}N_{2\chi}$ term –which saturates the geometric rate $C^g_{2\chi}$– the average DM density within the thermal radius $r^\chi_{\rm th}$ is smaller than the DMBS density within $r^{2\chi}_{\rm th}$, so the self-capture term $C_\chi N_\chi$ does not reach C^g_χ at $t=\tau_g$, and will never reach it after, as N_χ stops increasing. As for $N_{2\chi}$, it keeps increasing in time forever. For $t>\tau_g$, it increases linearly in time and the rate of BSF is half the capture rate,

$$\Gamma(t)=\frac{dN_{2\chi}}{dt}=\frac{1}{2}A_{\rm bsf}N^2_{\chi,{\rm eq}}\simeq\frac{C_\star+C^g_{2\chi}}{2}\,, \tag{18}$$

inducing a flux of mediators given by the same rate.

The reason why the C_χ term has a subleading effect with respect to the effect of the $C_{2\chi}$ term stems from the fact that these two terms are very different in nature. If we switch off the $C_{2\chi}$ term (green curves), the C_χ term also leads to an exponential grow, starting slightly before $t=\tau_s$, but it is much less important than the exponential growth from the $C_{2\chi}$ term. This is analogous to what happens for the symmetric DM case with BSF playing the role of annihilation if $C_{2\chi}=0$.[7] In Eq. (3) the C_χ term (linear in N_χ) is quickly counter-balanced by formation of bound-states from the $A_{\rm bsf}$ term (quadratic in N_χ): after a short period of exponential growth, $N_\chi\sim C_\star(e^{C_\chi t}-1)/C_\chi$, the green curve ($C_{2\chi}=0$), goes to a plateau from an equilibration of capture and BSF. In contrast, if $C_{2\chi}\neq 0$, the capture rate on bound-state grows as $N_{2\chi}$. Thus, due to capture on bound-state term $C_{2\chi}$, if it were not for the geometric rate upper floor, both N_χ and $N_{2\chi}$ would grow forever. Moreover, the argument of the exponential from the $C_{2\chi}$ term is larger than the one from the C_χ term, because $C_{2\chi}$ is about a factor of two times C_χ, see Fig. 2.

In presence of both C_χ and $C_{2\chi}$ terms, the C_χ term has only a moderate impact on the evolution of N_χ and $N_{2\chi}$. As a comparison of the full evolution (red curves in Fig. 3) and the $C_\chi=0$ evolution (blue curves) shows, including C_χ reduces the timescale when the exponential growth starts, from $t\sim\tau_0$ to $t\sim\tau_s$. Thus, due to the C_χ term, the exponential effect of the $C_{2\chi}$ term starts somewhat earlier, and the geometric rate within the bound-state thermal sphere is also reached somewhat earlier. The values of N_χ and $N_{2\chi}$ are insensitive to the C_χ term at $t\gg\tau_s$.

[7] In the limit of $C_{2\chi}\to 0$ and $A_{\rm bsf}=0$ (and with an additional DM annihilation term), the solution has an analytical form, first presented in [62].

III. NEUTRINO FLUX AND TERRESTRIAL DETECTION

As explained above, thanks to the capture on DMBS, the capture rate will saturate the geometric rate within the DMBS thermal sphere, leading to an equilibrium between the capture and the BSF processes. The time τ_g, at which this happens, can be easily shorter than the age of the Sun, as solved in Eqs. (A10) and (A11). At this point, the flux of mediators emitted from BSF becomes constant and is approximately given by Eq. (18) above.

For concreteness (and other reasons explained below) we will assume in the following that the mediator decays dominantly into a pair of neutrinos. Under the assumption that the mediator mass is much smaller than the binding energy, $E_{\rm bind}$, the differential flux at a terrestrial detector is

$$\frac{d^2\Phi_\nu}{dE_\nu d\Omega} = \frac{A_{\rm bsf} N_\chi^2(t_\odot)}{2\Delta\Omega}\frac{dN_\nu}{dE_\nu} = \frac{A_{\rm bsf} N_\chi^2(t_\odot)}{8\pi d_\odot^2\,(1-\cos\vartheta)}\frac{dN_\nu}{dE_\nu}\,, \tag{19}$$

where $d_\odot$ is the Earth-Sun distance (AU), with ϑ being the angular sensitivity of detector (as the apparent angular diameter of the Sun $\vartheta_\odot \approx 0.5^\circ$). Since the light mediator is boosted, the neutrino energy spectrum this decay leads to is not monochromatic but has a characteristic box shape [63]

$$\frac{dN_\nu}{dE_\nu} \simeq \frac{2}{E_{\rm bind}} \quad \text{for } E_- \le E_\nu \le E_+\,, \qquad \text{with} \qquad E_\pm = \frac{E_{\rm bind}}{2}\left(1 \pm \sqrt{1-\frac{4m_\phi^2}{E_{bind}^2}}\right). \tag{20}$$

In particular, $E_- \simeq 0$ and $E_+ \simeq E_{\rm bind}$ if $m_\phi \ll E_{\rm bind}$. The differential flux of neutrinos at the detector (ignoring oscillations) is

$$\frac{d\Phi_\nu}{dE_\nu} \;=\; \frac{1}{4\pi d_\odot^2}\Gamma(m_\chi, m_\phi, \alpha_\chi, s_\theta)\frac{2}{E_{\rm bind}}\Theta(E_+ - E_\nu)\Theta(E_\nu - E_-)\;. \tag{21}$$

The resulting neutrino fluxes are shown in Fig. 4 for a fixed Higgs mixing angle $\sin\theta_\phi = 10^{-12}$, for $\alpha = 0.05$, 0.1 and 0.15 (in blue, red and maroon colors), respectively. For comparison, the current limits (90% C.L.) on diffuse supernova ν_e neutrino background (DSNB) flux from Super-K runs III, IV and KamLAND are shown in blue, red and black points [64–66]. Also shown are predicted atmospheric fluxes of ν_e down to 10 MeV in thin green line, assuming 30° angular resolution in the sky, adapted from [67–69]. Current measurements of atmospheric neutrino fluxes are shown in green points [70]. As this figure suggests, the neutrino flux induced by BSF could be observable. When the geometric limit for DMBS sphere is reached, the emitted flux of mediators is proportional to $C_{2\chi}^g \propto m_\chi^{-2}$. This scaling is seen in Fig. 4. If the geometric limit is not reached, heavier mediators would lead to smaller self-capture rates, as shown by the falling tails of the neutrino flux.

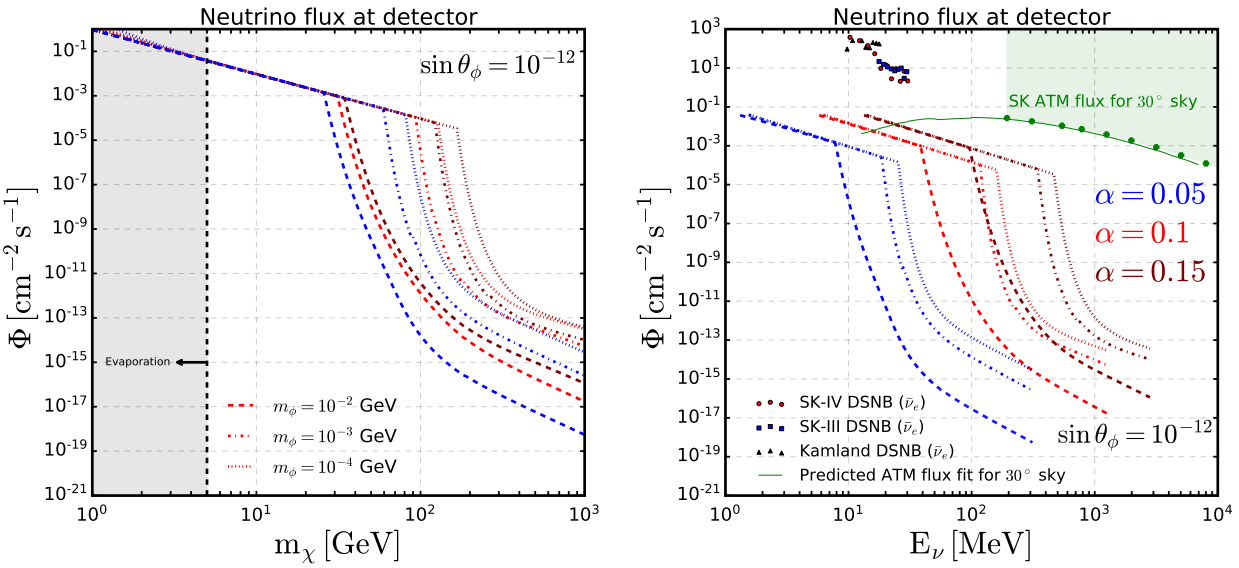


FIG. 4. *Left*: Flux of light mediators emitted by the BSF process for various parameter choices. *Right*: Neutrino flux as a function of neutrino energy, coming from the Sun for a detector placed on the surface of the Earth. Shown in red, blue and black scattered points are the current limits on diffuse supernova neutrinos adapted from [65]. Atmospheric low energy neutrino measurements from Super-K are shown by the green dots [64]. Thin green line denotes the predicted atmospheric neutrino flux for 30° sky [69].

Finally, note that our results can be easily generalized, e.g. if the mediator decays electromagnetically. Nevertheless, the experimental constraints, as studied below, also become more stringent for electromagnetic decays, further narrowing down allowed regions of parameter space.

IV. RESULTS AND CONSTRAINTS FOR AN EXPLICIT MODEL

A. Model

As mentioned above, we consider DM in the form of a (vector-like) Dirac fermion self-interacting through the exchange of a light scalar. Concretely the Lagrangian is (see e.g. [71–73])

$$\mathcal{L} \supset \mathcal{L}_{\rm SM} + \mathcal{L}_{N_R} + \bar{\chi}\left(i\partial\!\!\!/ - m_\chi\right)\chi + \frac{1}{2}\left(\partial\phi\right)^2 - g_s\phi\bar{\chi}\chi - V(\phi, H)\,. \tag{22}$$

Here we do not specify the scalar potential $V(\phi, H)$, and simply assume that it induces a ϕ-H mixing angle θ_ϕ (which can be achieved in various ways), so that, upon rotation to the mass basis, the following interactions are obtained:

$$\mathcal{L}_{\rm int} \;=\; g_s\cos\theta_\phi\phi'\bar{\chi}\chi + g_s\sin\theta_\phi h'\bar{\chi}\chi + \left(\cos\theta_\phi h' + \sin\theta_\phi\phi'\right)\frac{m_f}{v_H}\bar{f}_L f_R\,. \tag{23}$$

If we assume that neutrino masses are generated through the usual type-I seesaw mechanism, nothing prevents the right-handed neutrinos to couple in pairs to the light mediator ϕ. Thus one has the extra interactions

$$-\mathcal{L}_{N_R} \;=\; Y_\nu \bar{N} L \cdot H + \frac{m_N}{2}\overline{N^c} N + Y_\phi \phi \overline{N^c} N + \text{h.c.}\;. \tag{24}$$

In the physical ν', N' and ϕ', h' mass eigenstate basis, this leads to the following relevant Yukawa interactions for the light scalar eigenstate ϕ', up to second order in the neutrino mixing angle,

$$\begin{aligned} -\mathcal{L}_{\rm N_R} \;\supset\; & Y_\phi \phi' \cos\theta_\phi (\overline{N'^c} N' \cos^2\theta_\nu + \overline{N'^c}\nu'^c \sin\theta_\nu + \overline{\nu'} N' \sin\theta_\nu + \overline{\nu'}\nu'^c \sin^2\theta_\nu) + h.c. \\ - \;& Y_\nu \phi' \sin\theta_\phi (\bar{N}'\nu' \cos^2\theta_\nu + \overline{\nu'^c}\nu' \sin\theta_\nu - \bar{N}' N'^c \sin\theta_\nu - \overline{\nu'^c} N'^c \sin^2\theta_\nu) + h.c.\,. \end{aligned} \tag{25}$$

Thus, both the Y_ϕ and Y_ν interactions induce a decay of the light mediator into a pair of light neutrinos. In practice the decay induced by Y_ϕ will be dominant because various constraints require $Y_\nu \sin\theta_\phi \ll Y_\phi \sin\theta_\nu$. In the following, for simplicity, we will consider only one left-handed neutrino and one right-handed Majorana neutrino, with the neutrino mixing angle $\sin\theta_\nu \approx Y_\nu v_h/\sqrt{2}M_N$. Likewise for the associated Yukawa coupling, the typical seesaw gives

$$\sin^2\theta_\nu \sim \frac{m_\nu}{m_N} = 10^{-13} \cdot \frac{m_\nu}{0.1\text{eV}} \cdot \frac{1\,\text{TeV}}{m_N} \tag{26}$$

with $Y_\nu \sim \sqrt{2 m_\nu m_N / v_h^2}$.

If $m_N > m_\phi$ the light mediator cannot decay into NN or $N\nu$ and the $\nu\nu$ is the only possible channel induced by the seesaw interactions. The decay width is

$$\Gamma(\phi \to \nu\nu) = \frac{1}{16\pi} |Y_\phi|^2 \, m_\phi \, \sin^4\theta_\nu \,. \tag{27}$$

In addition to this neutrino-mixing channel, ϕ can also decay, through the ϕ-H scalar mixing, into pairs of charged leptons or quarks (or SM bosons for large ϕ mass). As will be discussed below, this needs to be suppressed.

B. Constraints and results

An observable neutrino flux requires sufficiently large capture and BSF rates, eventually leading to an equilibrium between both processes. This scenario must also fulfill additional phenomenological constraints:

Self-interactions: On the one hand, the corresponding self-scattering cross section is bounded from above by the observation of galaxy cluster collisions, e.g., [74–78]

$$\frac{\sigma_{\rm SI}}{m_\chi} \lesssim 0.5\,\text{cm}^2/\text{g}\;, \tag{28}$$

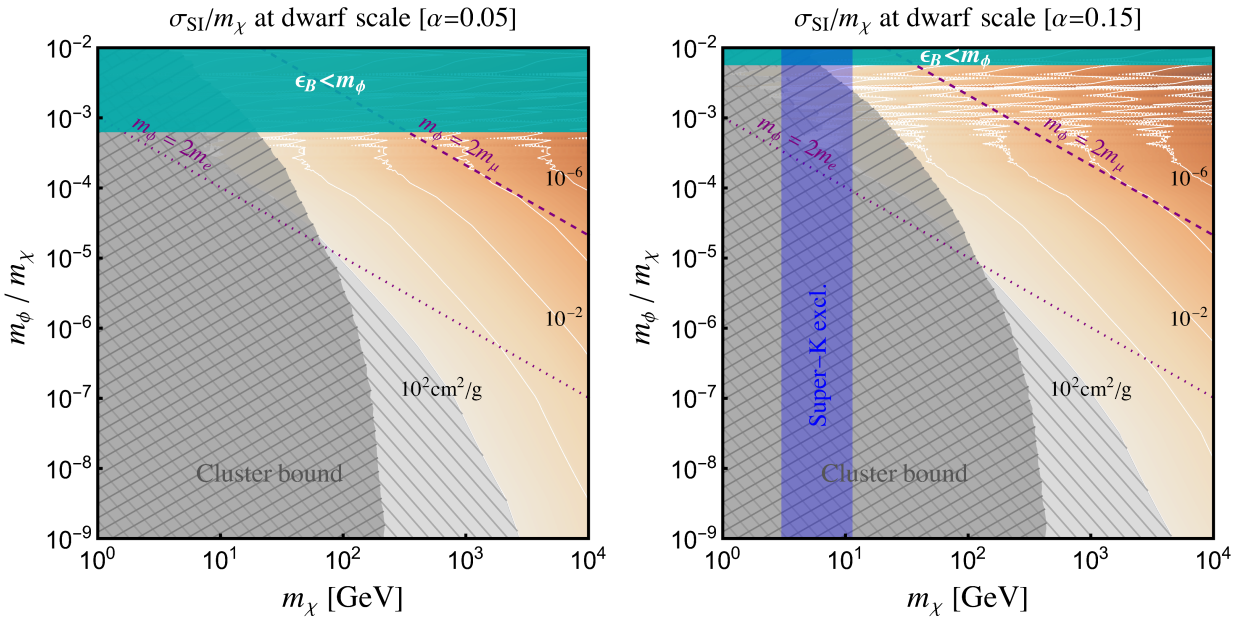


FIG. 5. Constraints for two example values of the dark coupling α. Contour lines give the values of $\sigma_{\rm SI}/m_\chi$ in dwarf-sized halos. Black (gray) hatched regions are excluded because $\sigma_{\rm SI}/m_\chi \geq 0.5\,{\rm cm}^2/{\rm g}$ at cluster scales (or $100\,{\rm cm}^2/{\rm g}$ at dwarfs scales, see contour lines). The resonant regime (with peaks) will become smaller in the Milky Way halo, where the average DM velocity is much larger than that at dwarfs scale.

for DM velocities around $v \sim 1000\,{\rm km/s}$. On the other hand, the non-observation of gravothermal collapse in dwarf-sized halos sets an upper bound of about $100\,{\rm cm}^2/{\rm g}$ at small scales, for which we take $v \sim 25\,{\rm km/s}$ [79, 80]. Concretely, we derive the bounds taking $\sigma_{\rm SI}$ equal to the viscosity cross section of DM particles. Adopting the modified transfer cross section barely changes the results, see Appendix B 1 for detailed discussions of these effective cross sections.

Fig. 5 shows the resulting values for this cross section as a function of m_χ and m_ϕ/m_χ for two values of α. The cluster bound excludes χ masses below $\sim$ 10-1000 GeV depending on m_ϕ/m_χ and α. This figure also shows the region where m_ϕ is larger than the binding energy where BSF is not possible. Values of the cross section that could address the small scale anomalies are allowed, i.e. $\sigma_{\rm SI}/m_\chi \in [1,\,100]\,{\rm cm}^2/{\rm g}$ at typical DM velocities of the order of 25 km/s in dwarf galaxies [81–84]. In the m_χ-α plane, Fig. 6 (right) shows the corresponding upper bound on α.

BBN: In order to observe the neutrino flux in the detector, the in-flight lifetime of the mediator must be smaller than eight minutes. This constraint is typically less stringent than that resulting from upper bounds on extra radiation after neutrino decoupling, which requires its lifetime to be shorter than one second. Using the typical seesaw expectation of Eq. (26), one gets

$$|Y_\phi|^2 \sim 0.03\,\frac{100\,{\rm MeV}}{m_\phi}\frac{1\,{\rm s}}{\tau_\phi}\left(\frac{0.1{\rm eV}}{m_\nu}\frac{m_N}{1{\rm GeV}}\right)^2. \tag{29}$$

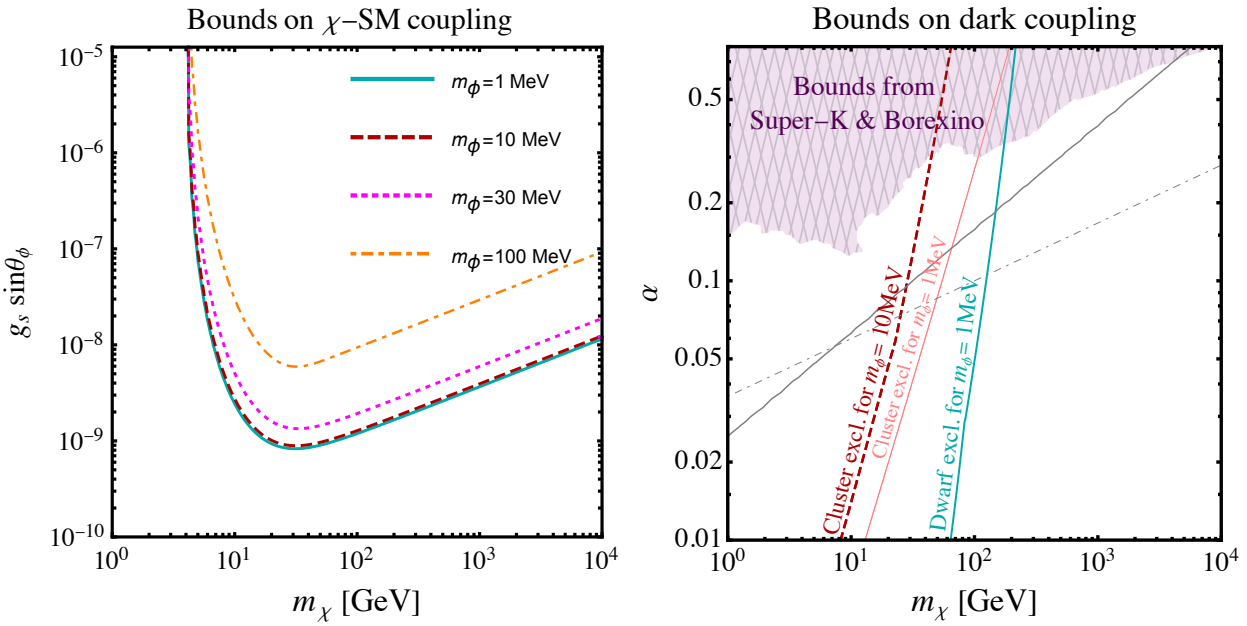


FIG. 6. *Left*: Upper bounds on the DM-SM interaction strength from XENON1T [85], depending on m_ϕ. *Right*: Upper bounds on DM coupling α from indirect search for neutrino flux (purple shaded area) and DM self-scattering at dwarfs scales. The bounds from the Bullet cluster, Eq. (28), are given for $m_\phi = 1\,\text{MeV}$ and $10\,\text{MeV}$. The conservative dwarf scale upper bound, $\sigma_{\rm SI}/m_\chi \lesssim 100\,\text{cm}^2/\text{g}$ is given for $m_\phi = 1$ MeV. This case is disfavored by BBN observables, given its sizeable coupling to neutrinos. For $m_\phi = 100\,\text{MeV}$, the dwarf galaxy scale bound is weaker and basically irrelevant for BSF. For α above the gray solid line, one expects a galactic flux from BSF in the galactic halo larger than the flux from BSF in the Sun. Efficient BSF in the early Universe is possible for parameters above the dot-dashed gray line [49].

Thus a fast enough decay is obtained provided that Y_ϕ is not too small and m_N is not too large. BBN and perturbative couplings, $Y_\phi^2 \lesssim 4\pi$, require $m_N \lesssim 20\,\text{GeV}\cdot(m_\nu/0.1\,\text{eV})\cdot(m_\phi/100\,\text{MeV})^{1/2}$. Thus low scale seesaw is favored along this scenario. This BBN bound can be relaxed if the ϕ number density gets suppressed before neutrino decoupling.

Direct detection: As already mentioned above, due to evaporation, DM masses below a few GeV are not relevant for our purpose. Thus, non-observation of spin-independent nucleon recoil signals provides the best direct-detection bounds. For purely scalar-mediated interactions, the differential elastic scattering cross section with nuclei has been given by Eq. (11). If $m_\phi \lesssim \sqrt{2m_N E_R} \lesssim 1\,\text{MeV}$ for typical keV-scale recoil energies, direct detection rates, albeit in the Born regime, are boosted due to t-channel exchange of the light mediator, leading to more stringent bounds than for heavier mediators. The corresponding measurements by the Xenon1T experiment [85, 86] set an upper bound on $g_s \sin\theta_\phi$, which we show in Fig. 6 (left) for different masses of mediator and DM. This

obviously translates into an upper bound on the capture rate on nucleons, as both processes involve the same DM-nucleon cross section. Recent LZ data can improve the limit on $g_s \sin\theta_\phi$ by a factor of $2-3$, depending on the mediator mass [87].

Indirect detection from galactic center emission: If DM BSF occurs in the Sun, it is reasonable to anticipate its occurrence in the galactic center of the Milky Way as well. Consequently, we would expect corresponding emission of energetic neutrinos originating from the galactic center, also with a box-shaped energy spectrum. Here we estimate the indirect search limit by re-scaling the current bounds on symmetric DM from neutrino telescope observations. Regarding indirect signals, a BSF process that generates one mediator and eventually two neutrinos with $E_\nu \approx E_{\rm bind}/2$ is equivalent to symmetric DM annihilation with a mass of $E_{\rm bind}/2$, if the latter only makes up a fraction of the observed DM abundance. Therefore, existing bounds can be rescaled as follows,

$$(\sigma_{\rm symm.} v)|_{m_\chi \to E_{\rm bind}/2} \to \left(\frac{E_{\rm bind}/2}{m_\chi}\right)^2 (\sigma_{\rm BSF} v)\,, \tag{30}$$

where $(\sigma_{\rm symm.} v)$ denotes the known bounds on symmetric DM as a function of the DM mass.

Quantitatively, the neutrino flux generated from the BSF process of halo DM particles is estimated to be

$$\Phi_\nu^{\rm D} \sim 1\,{\rm cm}^{-2}{\rm s}^{-1} \left(\frac{\rm GeV}{m_\chi}\right)^2 \left(\frac{\sigma_{\rm BSF} v}{3\times 10^{-22}\,{\rm cm}^3/{\rm s}}\right)\,. \tag{31}$$

This is similar to boosted DM case [88], which can vary mildly due to the uncertainty of the Galactic J-factor [89]. The corresponding upper bound on α is shown in the right panel of Fig. 6, obtained from re-scaling the indirect bounds holding in the symmetric DM case [90]. As also illustrated in Fig. 5, this excludes values of m_χ between 3.0 and 11.4 GeV for $\alpha = 0.15$, and gives irrelevant bound for $\alpha = 0.05$.

We can further speculate when this galactic neutrino flux is larger than the one induced by BSF inside the Sun. The latter is approximately $\Phi_\nu^\odot/({\rm GeV}/m_\chi)^2 \sim \mathcal{O}(1)\,{\rm cm}^{-2}{\rm s}^{-1}$ when the DM capture on DMBS is saturated by its geometric rate. Using Eq. (12) results in $\alpha^{5/2}/m_\chi \lesssim 10^{-4}/{\rm GeV}$, which in turn requires $m_\chi \gtrsim 5.6\,{\rm GeV}$ (87.1 GeV) for $\alpha = 0.05$ (0.15), see also the right panel of Fig. 6. In practice, the non-vanishing mass of the mediator suppresses the BSF cross section, reducing the neutrino flux, but barely affects the BSF-induced flux from the Sun.

CMB: Since the CMB observables are mostly sensitive to the total electromagnetic energy injected by extra processes in the high-redshift Universe, the associated constraint on BSF can be obtained by re-scaling the CMB bound on the annihilation of a symmetric DM candidate [91],

$$f_{\rm eff} \frac{\sigma_{\rm BSF} v}{m_\chi} \frac{{\rm Br}_{\phi\to {\rm EM}} E_{\rm bind}}{2m_\chi} \lesssim 4.1\times 10^{-28}\,\frac{{\rm cm}^3/{\rm s}}{\rm GeV}\,. \tag{32}$$

Here, $f_{\rm eff}$ is an efficiency factor that depends on the spectrum of injected electrons and photons, and we have taken into account that the energy injected per process is not about $2m_\chi$ as in the case of DM annihilation, but is given by the binding energy $\approx \alpha^2 m_\chi/4$, times the electromagnetic branching ratio of ϕ decay, $\mathrm{Br}_{\phi\to\mathrm{EM}}$. For $m_\phi \ll E_{\rm bind}$, the mass of the light mediator does not play any role in the CMB bound. For decay into SM particles other than neutrinos, the fact that BSF is strongly enhanced when the DM relative velocity is small gives a CMB bound which is much stronger than other bounds derived from local cosmic-ray observations such as Fermi-Lat and AMS experiments. For instance, taking the DM velocity at CMB to be a few km/s leads to $\mathrm{Br}_{\phi\to\mathrm{EM}}^{1/3}\alpha^{7/3} \lesssim 10^{-5} m_\chi/\mathrm{GeV}$. Combining this condition with Eq. (27) gives the constraint $Y_\phi \theta_\nu^2/\theta_\phi \gtrsim 10^{2.2}\alpha^{7/2}(\mathrm{GeV}/m_\chi)^{3/2}$ for m_ϕ below twice the muon mass. For typical seesaw values of the neutrino mixing angle, Eq. (26), this translates to the condition $m_N \lesssim 22\,\mathrm{TeV}\cdot Y_\phi\,(m_\chi/5\,\mathrm{GeV})^{3/2}(10^{-12}/\theta_\phi)(0.1/\alpha)^{7/2}(m_\nu/0.1\,\mathrm{eV})$.

Mediator decay into right-handed neutrinos: If $2m_N < m_\phi$ the dominant decay is not anymore into a pair of light neutrinos but into a pair of right-handed neutrinos (or into νN for $m_\phi/2 < m_N < m_\phi$), with their decay widths given by,

$$\Gamma(\phi\to NN) = \frac{1}{16\pi}|Y_\phi|^2\, m_\phi\,, \qquad\qquad \Gamma(\phi\to N\nu) = \frac{1}{16\pi}|Y_\phi|^2\, m_\phi \sin^2\theta_\nu\,. \tag{33}$$

Using the typical seesaw expectation of Eq. (26), one gets

$$|Y_\phi|^2 \sim 10^{-22}\,\frac{1\,\mathrm{GeV}}{m_\phi}\frac{1\,\mathrm{s}}{\tau_\phi}\,, \qquad\qquad |Y_\phi|^2 \sim 10^{-12}\,\frac{1\,\mathrm{GeV}}{m_\phi}\frac{1\,\mathrm{s}}{\tau_\phi}\frac{0.1\mathrm{eV}}{m_\nu}\frac{m_N}{1\mathrm{GeV}} \tag{34}$$

for $2m_N < m_\phi$ and $m_\phi/2 < m_N < m_\phi$, respectively. This can be compatible with the extra radiation constraint $\tau_\phi \lesssim 1\,\mathrm{s}$ but is basically excluded by the CMB bound because the right-handed neutrino decay product contains a non-negligible amount of electromagnetic material.

V. SUMMARY

We have considered the possibility that asymmetric DM forms bound-states in the Sun, and showed that this leads to novel phenomenology. BSF in the Sun can proceed via emission of light scalar particles that carry energy roughly equal to the binding energy. Their decays to neutrinos lead to potentially testable low energy signals at neutrino detectors.

Unlike for annihilating DM, BSF produces a flux of particles without reducing the number of DM particles in the Sun. We point out that on top of the DM particles captured in the Sun, the bound-states piling up in this way become additional scattering targets through which DM from

the galactic halo could be captured. We have determined the associated DM accretion rates on DM and DM bound-states by evaluating the differential cross section, taking into account that for typical parameters, $v \sim 10^{-3}$, $\alpha \sim 0.1$, $m_\phi <$ GeV and $m_\chi \sim 100$ GeV, the DM-DM and DM-DMBS scattering processes proceed in the semi-classical regime. As soon as these self capture rates are larger than $t_\odot^{-1}$, they can become phenomenologically relevant. In particular, we have shown that, thanks to the self-capture on bound-states, the number of DM particles in the Sun can exponentially increase, so much that the capture rate can reach the geometric rate, i.e. all the DM particles intercepting the DM bound-state thermal sphere are captured as the mean free path becomes smaller than the sphere. As a result, this exponential effect also considerably boosts the BSF and thus the associated flux of light mediators. In an example model, where DM is a Dirac fermion which self-interacts through exchange of light scalar that mixes with the Higgs boson, with the scalar decaying into two neutrinos through seesaw interactions, this leads to a neutrino flux which reach the predicted atmospheric neutrino fluxes at energies below hundred MeV. Near future experiments such as Hyper-K, as well as direct detection experiments, will be able to probe further this scenario.

Acknowledgments: We thank Sergio Palomares-Ruiz and Hai-bo Yu for discussions, and Sebastian Wild for help with the implementation of direct detection rates in DDCalc [86]. X.C. is supported by the Research Network Quantum Aspects of Spacetime (TURIS) and co-funded by the European Union (ERC, NLO-DM, 101044443). R.G. is supported by MIUR grants PRIN 2017FMJFMW and 2017L5W2PT, and thanks Galileo Galilei Institute for hospitality. C.G.C. is supported by a Ramón y Cajal contract with Ref. RYC2020-029248-I, the Spanish National Grant PID2022-137268NA-C55 and Generalitat Valenciana through the grant CIPROM/22/69. The work of T.H. is supported by the Excellence of Science (EoS) project No. 30820817 - be.h "The H boson gateway to physics beyond the Standard Model", by the IISN convention 4.4503.15, and by CERN where part of this work has been done. The authors acknowledge the workshop on Self-Interacting Dark Matter: Models, Simulations and Signals 2023, Pollica, for hospitality.

Appendix A: Approximate parametric solution to number evolution of DM particles

The set of coupled Boltzmann equations for N_χ and $N_{2\chi}$, Eqs. (2)-(3), has no closed-form analytical solution. However, we can construct an approximate solution and better understand the underlying physics by proceeding step by step as follows.

1. Case with no self-capture on DM bound-states

Without the $C_{2\chi}$ term, the equation for N_χ does not depend on $N_{2\chi}$

$$\frac{\mathrm{d}N_\chi}{\mathrm{d}t} \;=\; C_\star - A_{\rm bsf} N_\chi^2 + C_\chi N_\chi \,, \tag{A1}$$

which can be solved easily as follows.

BSF with capture only on nucleons: $C_\chi = C_{2\chi} = 0$: Without any capture due to DM self-interactions, the above equation has the solution

$$N_\chi^0(t) = C_\star \tau_0 \tanh\left(\frac{t}{\tau_0}\right) \qquad , \qquad N_{2\chi}^0(t) = \frac{1}{2} C_\star \left(t - \tau_0 \tanh\left(\frac{t}{\tau_0}\right)\right) \,, \tag{A2}$$

with the time constant given by $\tau_0 = (C_\star A_{\rm bsf})^{-1/2}$. Without BSF, i.e. when $\tau_0 \to \infty$, this solution gives a N_χ linearly increasing in time as expected, $N_\chi^0(t \lesssim \tau_0) = C_\star t$. With BSF, such an increase occurs until there are enough DM particles for BSF to proceed. For $t \gtrsim \tau_0$, an equilibrium between BSF and capture on nucleons is reached so that N_χ saturates becoming constant: $N_\chi = C_\star \tau_0 = (C_\star / A_{\rm bsf})^{1/2}$. This value can be obtained also directly from the Boltzmann equation, Eq. (2), whose right-hand side vanishes when $C_\star - A_{\rm bsf} N_\chi^2 = 0$. As expected, the larger $A_{\rm bsf}$, the sooner the saturation is reached and the smaller the asymptotic value of N_χ is. Conversely, the larger $C_\star$, the larger N_χ must be for the term associated with BSF to compensate that of $C_\star$. Note, nevertheless, that the larger $C_\star$ and the larger $A_{\rm bsf}$, the smaller the saturation time scale τ_0. This stems from the fact that the number of particles captured increases more rapidly, and BSF starts to become important earlier. Consequently, BSF catches up with the capture process sooner.

BSF with capture on nucleons and on free DM particles: $C_\chi \neq 0$, $C_{2\chi} = 0$: Early in the evolution the contribution of $C_{2\chi} N_{2\chi}$ is negligible. The analytical solution when $C_{2\chi} = 0$ and $C_\chi \neq 0$ is

$$N_\chi^s(t) \;=\; \frac{C_\chi}{2A_{\rm bsf}} + \frac{1}{A_{\rm bsf}\tau_s} \tanh\left(\frac{t}{\tau_s} - \operatorname{arctanh}\left(\frac{C_\chi \tau_s}{2}\right)\right) \,, \tag{A3}$$

$$2N_{2\chi}^s(t) + N_\chi^s(t) \;=\; \left(C_\star + \frac{C_\chi^2}{2A_{\rm bsf}} + \frac{C_\chi}{A_{\rm bsf}\tau_s}\right) t + \frac{C_\chi}{A_{\rm bsf}} \log\left(\frac{2 - C_\chi\tau_s + e^{-\frac{2t}{\tau_s}}\,(2 + C_\chi \tau_s)}{4}\right) \tag{A4}$$

with $\tau_s = (C_\star A_{\rm bsf} + C_\chi^2/4)^{-1/2}$. As it must be, for $C_\chi = 0$ these equations reduce to Eq. (A2) and τ_s reduces to τ_0.

Here again, the BSF term equilibrates with capture in Eq. (2), as it is negative and quadratic in N_χ, whereas the $C_\star$ (C_χ) term is constant (linear) in it. Thus, N_χ saturates when the right-hand side of Eq. (2) vanishes, $C_\star - A_{\rm bsf} N_\chi^2 + C_\chi N_\chi = 0$, which gives $N_\chi^{s,\rm eq} = C_\chi/(2A_{\rm bsf}) + 1/(\tau_s A_{\rm bsf}) =$

$C_\star/(\tau_s^{-1} - C_\chi/2)$. The equilibrium time scale τ_s is smaller than that without the C_χ term, τ_0, because the C_χ term increases the capture, so that BSF becomes important earlier. Note that before equilibrium is reached the solution of Eq. (A3) is exponential. However the effect of this exponential is very limited as the equilibrium is reached soon, as can be seen in Fig. (3), for various examples of parameter sets.[8] Note also that the total number of particles grows linearly, Eq. (A4), except for a logarithmic correction, particularly for $t \gg \tau_s$. More importantly, for realistic values of $C_\star$ and C_χ, the rate associated with the saturation value always lies well below the geometric rate C_χ^g of thermalized DM particles in the Sun.

2. Case with self-capture on DM bound-states

The appearance of a non-vanishing $C_{2\chi}$ term drastically changes the physics in several ways. First of all, the $C_{2\chi}$ term implies that the number of free DM particles in the Sun depends on the number of DM bound-states so that the Boltzmann equations for N_χ and $N_{2\chi}$ are coupled. Differentiating Eq. (2) with respect to time and using the equation for the bound-state, Eq. (3), results in the following second-order differential equation,

$$\frac{\mathrm{d}^2 N_\chi}{\mathrm{d}^2 t} = (-2A_{\rm bsf} N_\chi + C_\chi)\frac{\mathrm{d}N_\chi}{\mathrm{d}t} + \frac{1}{2}C_{2\chi}A_{\rm bsf}N_\chi^2\,, \tag{A5}$$

with boundary conditions: $\frac{\mathrm{d}N_\chi(t=0)}{\mathrm{d}t} = C_\star$ and $N_\chi(t=0) = 0$. Formally, Eq. (A5) is of the form $N'' = f(N)N' + g(N)$, known as a Liénard equation, with no known general analytical solutions.

Second, the self-capture associated with $C_{2\chi}$ is larger than that induced by C_χ. This and the fact that $N_{2\chi}$ never saturates –as Eq. (3) shows– imply that the term $C_{2\chi}N_{2\chi}$ in Eq. (2) becomes much larger than $C_\chi N_\chi$. In fact, N_χ saturates to values much larger than in the case of $C_\chi \neq 0, C_{2\chi} = 0$, because the right-hand side of Eq. (2) vanishes for $A_{\rm bsf}N_\chi^2 \simeq C_{2\chi}N_{2\chi} \gg C_\chi N_\chi$. Thus, unlike in the case of $C_\chi \neq 0$, $C_{2\chi} = 0$, where the exponential growth induced by the C_χ term does not last long, the $C_{2\chi}$ term induces an exponential growth for N_χ which is both faster (since the argument of the exponential will be proportional to $C_{2\chi}$ rather than to C_χ) and much larger. Therefore, N_χ can reach in this way much higher values, large enough for the capture rate to reach the geometric limit within the DMBS thermal sphere. Approximate solutions can be obtained as follows.

[8] Actually the solution of Eq. (A3) can be rewritten in exponential form as $2C_\star\tau_s(e^{2t/\tau_s} - 1)/(A + Be^{2t/\tau_s})$ with $A = -2 - C_\chi\tau_s$ and $B = -2 + C_\chi\tau_S$. In practice if $C_\star A_{\rm bsf} < C_\chi^2/4$ (as will be the case for our scenario), then $\tau_s \simeq 2/C_\chi$ which gives $B < A$ so that $N_\chi \simeq 2C_\star\tau_s(e^{2t/\tau_s} - 1)/A$, which is exponentially growing until it reaches the equilibrium plateau when Be^{2t/τ_s} becomes larger than A. The amount of exponential grows is limited because in practice B is not much smaller than A.

BSF with capture on nucleons and on DM bound-states: $C_\chi = 0$, $C_{2\chi} \neq 0$: For the parameters of interest in this work, $C_{2\chi} \ll 100/\tau_0$, we numerically find that the following exponential ansatz provides a good approximation from $t = \tau_0$ to the point when $C_{2\chi}N_{2\chi}$ saturates the geometric capture rate. That is, in practice

$$N_\chi(\tau_0 \lesssim t \lesssim \tau_g) = N_\chi^0(\tau_0) \exp\left[\frac{C_{2\chi}}{4}(t-\tau_0)\right] . \tag{A6}$$

This is also suggested by the quasi-static solution of $dN_\chi/dt \sim 0$ in Eqs. (2) and (3), where

$$\frac{dN_{2\chi}}{dt} \simeq \frac{C_\star}{2} + \frac{C_{2\chi}}{2}N_{2\chi} \quad \text{and} \quad N_\chi \simeq \left(\frac{C_\star + C_{2\chi}N_{2\chi}}{A_{\rm bsf}}\right)^{1/2} . \tag{A7}$$

For the opposite case, $C_{2\chi} \geq 100/\tau_0$, the divergence happens even faster.

Therefore, capture on bound-states exponentially increases the number of DM particles within the Solar lifetime if the condition, $t_\odot - \tau_0 \gg C_{2\chi}^{-1}$ is satisfied. The fact that the exponential growth starts when $N_{2\chi}$ becomes of order N_χ can also be seen in the numerical examples shown in Fig. 3.

BSF in the full general case: $C_\chi \neq 0$, $C_{2\chi} \neq 0$: For the reasons explained above, switching on the C_χ term –in addition to the $C_{2\chi}$ term– does not drastically change the result. It induces an additional moderate exponential growth that makes the contribution of $C_{2\chi}$ important slightly earlier (i.e. around $t = \tau_s$ rather than at τ_0), see Fig. 3. An approximate solution to this general case is obtained in the same way as Eq. (A6), by matching Eq. (A3) at τ_s rather than at τ_0. Interestingly, since $\tau_s < 2/C_\chi \sim 4/C_{2\chi}$, we have $\tau_s C_{2\chi} \lesssim 4$. That is, the condition for the validity of Eq. (A6), $C_{2\chi} \ll 100/\tau_0$, in which τ_0 is now replaced by τ_s, is automatically satisfied. This suggests

$$N_\chi(\tau_s \lesssim t \lesssim \tau_g) = N_\chi^s(\tau_s) \exp\left[\frac{C_{2\chi}}{4}(t-\tau_s)\right] , \tag{A8}$$

where $N_\chi^s(\tau_s)$ is well approximated by Eq. (A3).

This exponential growth lasts until the geometric capture rate within the DMBS thermal sphere is saturated. Quickly after $t = \tau_g$, N_χ stops increasing when the BSF term (quadratic in N_χ) compensates the constant capture rate.[9] The quasi-static equilibrium solution, obtained from $dN_\chi/dt \simeq C_\star + C_{2\chi}^g - A_{bsf}N_\chi^2 \simeq 0$, yields the final particle number[10]

$$N_\chi(t > \tau_g) = \left(\frac{C_{2\chi}^g + C_\star}{A_{\rm bsf}}\right)^{1/2} . \tag{A9}$$

[9] Once the capture rate reaches the DMBS sphere geometric rate, the time it takes for N_χ to become constant is small, $\Delta t \simeq 1/\sqrt{A_{\rm bsf}C_{2\chi}^g}$.

[10] Here we neglect the subleading capture contribution, from the C_χ term, within the DM thermal sphere outside the DMBS thermal sphere. To be accurate the $C_\star$ term in Eq. (A9) has to be taken into account only outside the DMBS thermal sphere as the $C_{2\chi}^g$ term already accounts for the maximal total capture rate within this sphere, but this concerns a negligible effect.

The associated time τ_g, where such an equilibrium is reached, is determined as

$$\tau_g = \tau_s + \frac{2}{C_{2\chi}} \log\left(\frac{C^g_{2\chi} + C_\star}{A_{\rm bsf}(N^s_\chi(\tau_s))^2}\right) . \tag{A10}$$

The condition that $\tau_g \leq t_\odot$, so that the DM number can be maximized at present, is satisfied if

$$\frac{C_{2\chi}}{2}\,(t_\odot - \tau_s) \geq \log\left(\frac{C^g_{2\chi} + C_\star}{A_{\rm bsf}(N^s_\chi(\tau_s))^2}\right) = \begin{cases} \log\left(\frac{C^g_{2\chi}/C_\star+1}{\tanh[1]^2}\right) & \text{if } C^2_\chi \leq A_{\rm bsf}C_\star \\ \log\left(\frac{C^g_{2\chi}/C_\star+1}{(e^2-1)^2}\,\frac{C^2_\chi}{A_{\rm bsf}C_\star}\right) & \text{if } C^2_\chi \gg A_{\rm bsf}C_\star \end{cases} . \tag{A11}$$

The latter case has $\tau_0 \gg \tau_s \simeq 2/C_\chi \sim 4/C_{2\chi}$, so the saturation should happen within a few τ_s.

We can now analytically match the above two solutions at times $\sim \tau_s$ and τ_g. The final approximate solution for species i has the parametric form

$$N_i(t) \;=\; N_i^{t<\tau_s}\Theta(\tau_s - t) + N_i^{\tau_s<t<\tau_g}\Theta(t-\tau_s)\Theta(\tau_g - t) + N_i^{t>\tau_g}\Theta(t-\tau_g)\,, \tag{A12}$$

where $N_i^{t<\tau_s}$, $N_i^{\tau_s<t<\tau_g}$ and $N_i^{t>\tau_g}$ are given by Eqs. (A3), (A8) and (A9), respectively. We have checked that these analytical expressions agree well with our numerical results presented in the various figures from solving the Boltzmann equations.

Appendix B: (Non-)perturbative treatment of elastic DM scattering

Suppose that DM scatters off a target with a mass m_T with a negligible form factor, the cross section differential in the recoil energy is

$$\frac{{\rm d}\sigma}{{\rm d}E_R} = \frac{2\pi\, m_T}{k^2}\left|\mathcal{M}(k,\theta)\right|^2, \qquad \text{with} \qquad \cos\theta = 1 - \frac{m_T E_R}{k^2}\,. \tag{B1}$$

Here θ and $k = \mu v$ are respectively the scattering angle and the incoming momentum in the center of mass frame, while the reduced mass of the DM-target system $\mu = m_\chi m_T/(m_\chi + m_T)$. We aim to calculate this when the scattering is triggered by the exchange of a scalar mediator, particularly when non-perturbative effects cannot be disregarded. In the non-relativistic limit, this is described by a free DM particle scattered by the corresponding Yukawa potential $V(r) = -\alpha' e^{-m_\phi r}/r$, with $\alpha' = g_s y_N \sin\theta_\phi/4\pi, \alpha$ or 2α when target is a nucleon, a DM particle or a DMBS, respectively.

The amplitude of such quantum scattering can be obtained from a partial-wave expansion

$$\mathcal{M}(k,\theta) = \frac{1}{2ik}\sum_{\ell=0}^{\infty}(2l+1)P_\ell\left(\cos\theta\right)\left(e^{2i\delta_\ell} - 1\right) , \tag{B2}$$

where ℓ is the orbital angular momentum, $P_\ell(\cos\theta)$ is the corresponding Legendre polynomial and δ_ℓ is the phase-shift. In general, the phase-shift must be obtained by solving the radial part of the

Schrödinger equation describing the collision, which is equivalent to solving [92]

$$\delta'_{\ell,k}(r) = -2k\mu\, r^2 V(r)\ \mathrm{Re}\left[e^{i\delta_{\ell,k}(r)} h_\ell^{(1)}(kr)\right]^2 \quad \text{with } \delta_{\ell,k}(0) = 0 \ \text{ and } \ \delta_{\ell,k}(r)|_{r\to\infty} \to \delta_\ell\,. \tag{B3}$$

Here, $h_\ell^{(1)} = j_\ell + i\, n_\ell$ is the spherical Hankel function of the first kind. Nevertheless, solving Eq. (B3) for all values of ℓ is impractical, and thus simplifications are necessary for different parameter regimes, as shown in left panel of Fig. 7. This will be discussed below for integrated and differential DM-DM(BS) scattering, separately. The Born approximation is always justified for the very weak scattering between DM and nucleon.

1. Integrated DM self-scattering in DM halos

Sufficiently large DM self-scattering can alter the evolution of halos, which gives rise to constraints on the corresponding scattering rates. As argued in Ref. [93], the gravothermal evolution of such halos is best characterized by the so-called viscosity cross section, $\sigma_V \equiv \frac{3}{2}\int_{-1}^{1} d\cos\theta\, \sin^2\theta \frac{d\sigma}{d\cos\theta}$.[11] The parameter regimes where the Born and semi-classical approximations apply are given in left panel of Fig. 7.

For $\alpha \lesssim v$ or $m_\chi \lesssim m_\phi/\alpha$, the first-order Born approximation is justified [59]. Note that $\mu = m_\chi/2$ and $\alpha' = \alpha$ in this case. Moreover, accounting for the indistinguishability of the DM particles in $\chi - \chi$ scattering, we find

$$\sigma_V = \frac{6\pi\beta^2}{m_\phi^2}\left(\frac{\left(\frac{m_\chi v}{m_\phi}\right)^4 + 5\left(\frac{m_\chi v}{m_\phi}\right)^2 + 5}{\left(\frac{m_\chi v}{m_\phi}\right)^4 + 2\left(\frac{m_\chi v}{m_\phi}\right)^2}\log\left[\left(\frac{m_\chi v}{m_\phi}\right)^2 + 1\right] - \frac{5}{2}\right)\,. \tag{B4}$$

Where we have introduced $\beta = 2m_\phi\alpha/(m_\chi v^2)$ for later convenience. Note that if one adopts the modified transfer cross section, defined as $\sigma_{|T|} \equiv \int_{-1}^{1} d\cos\theta\,(1 - |\cos\theta|)\frac{d\sigma}{d\cos\theta}$, approximately $\sigma_V \simeq 2\sigma_{|T|}$ if $m_\chi v/m_\phi \ll 1$ as well as $\sigma_V \simeq 3\sigma_{|T|}$ if $m_\chi v/m_\phi \gg 1$.

Outside the Born regime, non-perturbative effects are non-negligible, especially in the resonant regime ($\alpha m_\chi/m_\phi \gtrsim 1$ and $m_\chi v/m_\phi \lesssim 1$). In spite of this, in this parameter region the scattering is generally dominated by the s-wave, $\ell = 0$ [57], for which we can adopt the analytical results obtained by solving for the phase shift, δ_0 of $\ell = 0$, under the assumption that the Yukawa potential can be approximated by a Hulthén potential [94]. For identical fermions, after adding the symmetric factor 1/2 for final states, this leads to

$$\sigma_V = \frac{4\pi}{2}\frac{d\sigma}{d\Omega}\bigg|_{\ell=0} = \frac{2\pi}{k^2}\left|e^{i\delta_0}\Big(P_0(\cos\theta) + P_0(-\cos\theta)\Big)\sin\delta_0\right|^2 = \frac{8\pi}{k^2}\left|e^{i\delta_0}\sin\delta_0\right|^2\,, \tag{B5}$$

[11] Following Ref. [93], we introduce a prefactor of 3/2 to make $\sigma_V = \sigma$ for isotropic scatterings.

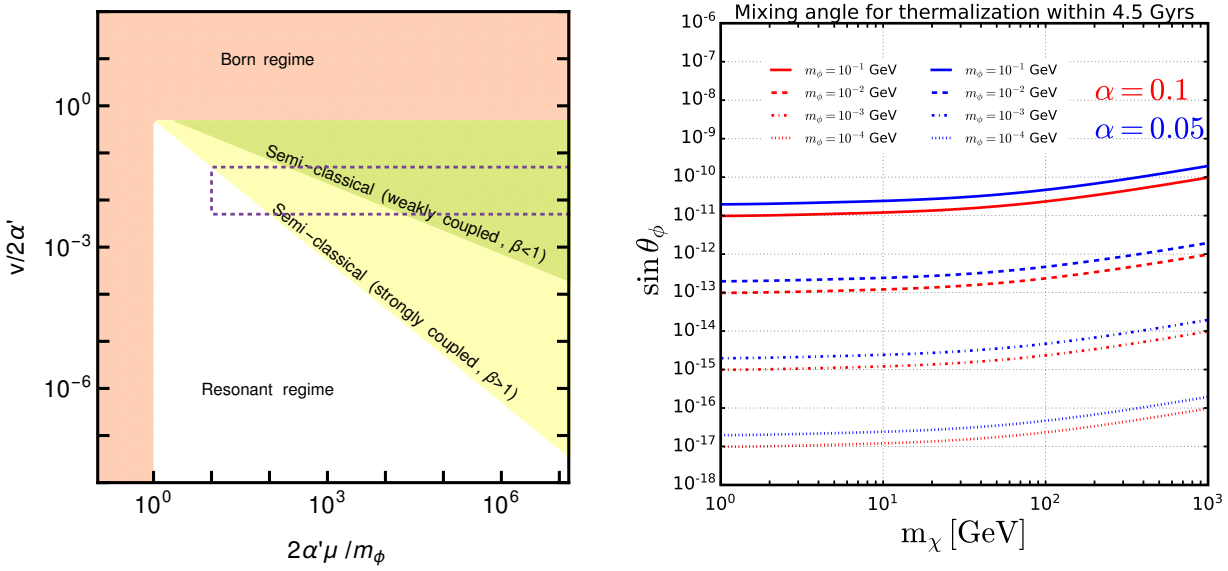


FIG. 7. *Left*: Dark matter self-scattering regimes. For self-capture of dark matter by the Sun, the relevant parameters roughly lie within the dotted line, and correspond to $v \sim 4 \cdot 10^{-3}$, $m_\chi \gtrsim 5$ GeV, 10^{-4} GeV $\lesssim m_\phi \lesssim 10^{-2}$ GeV, and $0.02 \lesssim \alpha < 0.2$. See text for details. *Right*: Curves above which the captured DM particles thermalize with the nucleons in the Sun within $t_\odot$.

where the scattering is isotropic, $k = m_\chi v/2$, and P_0 is the 0th-order Legendre polynomial. Note that this cross section only includes an even spatial wave-function and an anti-symmetric spin wave-function. We take the probability of such anti-parallel spin alignment to be 1/4. That is, its corresponding cross section for identical fermion is reduced by half with respect to that of non-identical fermions. The characteristic peaks of this regime are visible in Figs. 2 and 5.

Finally, with the resulting expressions of the viscosity cross section for all regimes, we impose bounds on DM self-interaction at both cluster and dwarf scales, obtaining the observationally allowed parameter region shown in Fig. 5, see also section IV.

2. Differential DM-DM(BS) scattering for DM capture

As explained in the main text and above, while the impact of DM self-interactions in galaxies and galaxy clusters can be characterized by the integrated transfer or viscosity total cross sections, for DM self-capture in the Sun, it is crucial to obtain the differential scattering cross section. In the latter case, left panel of Fig. 7 (dotted contour) illustrates the parameter region of interest for the DM capture via scatterings off free and bound DM particles already captured in the Sun.

The relevant parameters largely lie in the semi-classical regime, where $k \gg m_\phi^{-1}$. That is, the

de Broglie wavelength is much larger than the range of the Yukawa potential and the scattering amplitude can be estimated semi-classically. More precisely, the resulting sum in Eq. (B2) can be performed using the stationary phase approximation [59], which, up to an inconsequential global phase, gives

$$\mathcal{M} = \frac{1}{k}\left(\frac{\ell_0(\theta)}{\sin\theta}\left|\frac{\mathrm{d}\ell_0(\theta)}{\mathrm{d}\theta}\right|\right)^{\frac{1}{2}}, \quad \text{where} \quad \int_{r_0}^{\infty} \frac{\ell_0(\theta)+\frac{1}{2}}{r^2\sqrt{k^2-\left(\frac{\ell_0(\theta)+\frac{1}{2}}{r}\right)^2-2\mu V(r)}}\mathrm{d}r = \frac{\pi+\theta}{2}\,. \tag{B6}$$

Here r_0 is the closest distance from the potential centre, and $\ell_0(\theta)$ is the angular-momentum minimizing the phase in Eq. (B2) for a given scattering angle θ. All this applies as long as $\ell_0(\theta) \gg 1$, which justifies neglecting 1/2 in Eq. (B6). To see how Eq. (B6) resembles the classical equations of motion, we define the impact parameter as $\rho \equiv \ell_0(\theta)/k$, which together with Eq. (B1), leads to

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta} = 2\pi\rho\frac{\mathrm{d}\rho}{\mathrm{d}\cos\theta}\,, \qquad \text{where} \qquad \int_{r_0}^{\infty} \frac{\rho\,\mathrm{d}r}{r^2\sqrt{1-\frac{\rho^2}{r^2}+2\beta\left(\frac{e^{-m_\phi r}}{m_\phi r}\right)}} = \frac{\pi+\theta}{2}\,, \tag{B7}$$

with $\beta \equiv \mu m_\phi \alpha'/k^2$. The integral in Eq. (B7) gives the scattering angle according to classical mechanics. In this work, we follow the numerical approach presented in Ref. [60, 61] to solve Eq. (B7) and obtain the scattering rates.

As shown by Ref. [61], the calculation can be further simplified for $\beta > 13.2$, where the angular dependence of $\mathrm{d}\sigma/\mathrm{d}\cos\theta$ becomes insensitive to β. Performing a numerical fit to the cross sections of non-identical scattering and defining $\mathrm{d}\sigma/\mathrm{d}\cos\theta \equiv \pi\rho_*^2(\beta, m_\phi)\mathcal{N}(\cos\theta)$, we find

$$\mathcal{N}(c_\theta) \simeq 0.35 + 0.05c_\theta + 2.09c_\theta^3 + 3.25c_\theta^4 - 5.51c_\theta^5 - 8.53c_\theta^6 + 5.06c_\theta^7 + 7.22c_\theta^8\,. \tag{B8}$$

Note that the normalization, $\int_{-1}^{1} dc_\theta(1-c_\theta)\mathcal{N}(c_\theta) \simeq 1$, is chosen in a way we can directly adopt the overall prefactor from [60] as

$$\rho_*^2(\beta, m_\phi) = \begin{cases} \frac{1}{m_\phi^2}\left(\log\beta + 1 - \frac{1}{2\log\beta}\right)^2 & \text{for} \quad \beta > 1000\,, \\ \frac{8}{m_\phi^2}\left(\frac{\beta^2}{1+1.5\,\beta^{1.65}}\right) & \text{for} \quad 1000 > \beta > 13.2\,. \end{cases} \tag{B9}$$

The amplitudes at the vicinity of $\cos\theta = 1$, which may approach infinity, are not included in the numerical fit above, given that small-angle scatterings barely contribute to the self-capture rate as a result of $\theta_{\min}$ in Eqs. (15) and (16).

Besides, under the assumption that the interference is subleading in the (semi-)classical regime [95], the previous approach can also be applied to the scattering of identical particles

by symmetrizing the differential scattering amplitude $\theta \to \pi - \theta$, as

$$\left.\frac{d\sigma}{d\cos\theta}\right|_{\chi\chi} = \frac{d\sigma}{d\cos\theta} + \left.\frac{d\sigma}{d\cos\theta}\right|_{\cos\theta\to-\cos\theta}, \tag{B10}$$

where the range of the scattering angle θ reduces to $[0,\, \pi/2]$.

Appendix C: DM thermalization

We summarize below the general formalism for DM thermalization with the Solar medium, and present the results applicable to the scenario presented in the main text. After being captured, DM particle orbits become smaller than the size of the celestial body. For contact interactions, the time taken for the orbits to shrink to its thermal radius is given by [96]

$$t_{\rm therm} = \frac{m_\chi}{(\sum_N \rho_N \sigma_{\chi N})}\sqrt{\frac{m_\chi}{2\,T}}\ . \tag{C1}$$

For the Sun, the time scale is estimated via

$$t_2^\odot \ = \ 1.5\,{\rm yrs}\left(\frac{m_\chi}{10\,{\rm GeV}}\right)^{3/2}\left(\frac{2.2\,{\rm KeV}}{T_c}\right)^{1/2}\left(\frac{150\,{\rm g/cm^3}}{\rho_c}\right)\left(\frac{10^{-40}\,{\rm cm^2}}{\sigma_{\chi N}}\right)\ . \tag{C2}$$

Thermalization time is dominated by the last stages of the process with the typical energy and momentum transfer of the order of temperature [97]. Formally, thermalization time is computed through the energy loss rate. Consider the elastic scattering of a DM particle χ off a distribution of target particles T: $\chi(k)+T(p) \to \chi(k')+T(p')$. The interaction rate per DM particle reads [98, 99]:

$$\begin{aligned} d\Gamma &= 2\frac{{\rm d}^3k'}{(2\pi)^3}S(q_0,q)\ , \\ S(q_0,q) &= \iint \frac{{\rm d}^3p'{\rm d}^3p}{(2\pi)^6 2E_{p'}2E_{k'}2E_p2E_k}(2\pi)^4\delta^4\left(k+p-k'-p'\right)|\mathcal{M}|^2 f(E_p)\left(1-f(E_{p'})\right)\ , \end{aligned} \tag{C3}$$

where the second line above is the response function, and $f(E_{p,p'})$ being the Maxwell-Boltzmann distributions w.r.t. the core temperature T_c. For non-degenerate medium, there exists little Pauli-blocking from the final states, so $1 - f(E_{p'}) \simeq 1$. In turn, the rate of energy loss is given by

$$\Phi = 2\int \frac{{\rm d}^3k'}{(2\pi)^3}S(q_0,q)\times(E_i - E_f)\,, \tag{C4}$$

where k' is integrated from 0 to k.

Using this we can write down the time taken to thermalize with the celestial body, i.e. to reach a final energy $E_f = 3/2\,T_{\rm c}$ starting with an initial kinetic energy $E_i = m_\chi v_{\rm esc}^2/2$:

$$\tau_{\rm therm} = -\int_{E_i}^{E_f}\frac{{\rm d}E_i}{\Phi}. \tag{C5}$$

Integrating Eq. (C3) over p', the response function becomes

$$S(q_0,q) \;=\; \int \frac{\mathrm{d}^3p}{(2\pi)^2 2}\frac{|\mathcal{M}|^2}{16m_N^2m_\chi^2}\delta^3\delta\left(q_0-E_p+E_{p'}\right)f(E_p), \tag{C6}$$

where $E_k = E_{k'} = m_\chi$ and $E_p = E_{p'} = m_N$ have been taken in the denominator, while for the numerator we use $E_p = m_N + p^2/2m_N$ and $E'_p = m_N + p'^2/2m_N$. Moreover, the δ^3 enforces momentum conservation, $\mathbf{q} = \mathbf{p}' - \mathbf{p}$, and the energy delta-function is recast in terms of angle between the incoming nuclei and the momentum transfer as shown in Ref. [100], yielding

$$S(q_0,q) \approx \int \frac{\mathrm{d}p\,\mathrm{d}\cos\theta}{(2\pi)^2 2}2\pi p^2\frac{|\mathcal{M}|^2}{16m_N^2m_\chi^2}e^{\bar{\mu}/T_c}e^{-p^2/2m_NT_c}\frac{m_N}{|p||q|}\delta\left(\cos\theta-\cos\theta_0\right)\Theta(p^2-p_-^2)\;, \tag{C7}$$

along with

$$\cos\theta_0 = \frac{m_N}{|p||q|}\left(q_0-\frac{q^2}{2m_N}\right) \qquad \text{and} \qquad p_-^2 = \frac{m_N^2}{q^2}\left(q_0-\frac{q^2}{2m_N}\right)^2\;. \tag{C8}$$

Integrating the above equation we get

$$S(q_0,q) \approx \frac{|\mathcal{M}|^2}{64\pi m_\chi^2}\frac{T_c}{|q|}e^{\bar{\mu}/T_c}e^{-p_-^2/2m_NT_c}\,. \tag{C9}$$

Taking the Higgs-mixing model introduced in the main text, the squared amplitude for scattering on nucleon in the non-relativistic limit reads

$$|\mathcal{M}|^2 = g_s^2\cos^2_{\theta_\phi}\sin^2_{\theta_\phi}\frac{m_N^2f_N^2}{v_H^2}\frac{m_\chi^2m_N^2}{\left(q_0^2-q^2-m_\phi^2\right)^2}\;, \tag{C10}$$

which, in the limit of small energy transfer, corresponds to DM-nuclei scattering cross section of

$$\sigma_{\chi N} = g_s^2\cos^2_{\theta_\phi}\sin^2_{\theta_\phi}\frac{1}{16\pi(m_\chi+m_N)^2}\frac{m_N^2f_N^2}{v_H^2}\frac{m_\chi^2m_N^2}{m_\phi^4}\,. \tag{C11}$$

The k' integral is recast in terms of momentum transfer and energy transfer. The limits of integration are $0 < q_0 < E_k$ and $q_0 < q < 2E_k - q_0$. Therefore the energy loss rate for a particle with initial momenta k is

$$\frac{d\Gamma}{dq_0}\Delta E = \int\frac{dq}{(2\pi)^2}q\frac{E_{k'}}{E_k}S(q_0,q)(E_k-E_{k'}) = \int\frac{dq}{(2\pi)^2}q\frac{E_k-q_0}{E_k}S(q_0,q)q_0\;. \tag{C12}$$

In the non-degenerate limit the above equation factorizes. Hence the total rate is proportional to cross section on one target particle times their number density [100]. The corresponding numerical results are shown in the right panel of Fig. 7, showing that the thermalization can be reached within the lifetime of the Sun, even with very tiny mixing angles. Including the DM self-interactions would further reduce the time scale of thermalization.

---

[1] J. D. March-Russell and S. M. West, Phys. Lett. B **676**, 133 (2009), arXiv:0812.0559 [astro-ph].

[2] M. Pospelov and A. Ritz, Phys. Lett. B **671**, 391 (2009), arXiv:0810.1502 [hep-ph].

[3] W. Shepherd, T. M. P. Tait, and G. Zaharijas, Phys. Rev. D **79**, 055022 (2009), arXiv:0901.2125 [hep-ph].

[4] R. Laha and E. Braaten, Phys. Rev. D **89**, 103510 (2014), arXiv:1311.6386 [hep-ph].

[5] B. von Harling and K. Petraki, JCAP **12**, 033 (2014), arXiv:1407.7874 [hep-ph].

[6] K. Petraki, L. Pearce, and A. Kusenko, JCAP **07**, 039 (2014), arXiv:1403.1077 [hep-ph].

[7] K. Petraki, M. Postma, and M. Wiechers, JHEP **06**, 128 (2015), arXiv:1505.00109 [hep-ph].

[8] L. Pearce, K. Petraki, and A. Kusenko, Phys. Rev. D **91**, 083532 (2015), arXiv:1502.01755 [hep-ph].

[9] R. Laha, Phys. Rev. D **92**, 083509 (2015), arXiv:1505.02772 [hep-ph].

[10] H. An, B. Echenard, M. Pospelov, and Y. Zhang, Phys. Rev. Lett. **116**, 151801 (2016), arXiv:1510.05020 [hep-ph].

[11] H. An, M. B. Wise, and Y. Zhang, Phys. Rev. D **93**, 115020 (2016), arXiv:1604.01776 [hep-ph].

[12] H. An, M. B. Wise, and Y. Zhang, Phys. Lett. B **773**, 121 (2017), arXiv:1606.02305 [hep-ph].

[13] X.-J. Bi, Z. Kang, P. Ko, J. Li, and T. Li, Phys. Rev. D **95**, 043540 (2017), arXiv:1602.08816 [hep-ph].

[14] C. Kouvaris, K. Langæble, and N. G. Nielsen, JCAP **10**, 012 (2016), arXiv:1607.00374 [hep-ph].

[15] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, JCAP **05**, 036 (2017), arXiv:1612.07295 [hep-ph].

[16] P. Asadi, M. Baumgart, P. J. Fitzpatrick, E. Krupczak, and T. R. Slatyer, JCAP **02**, 005 (2017), arXiv:1610.07617 [hep-ph].

[17] K. Petraki, M. Postma, and J. de Vries, JHEP **04**, 077 (2017), arXiv:1611.01394 [hep-ph].

[18] E. Johnson, E. Braaten, and H. Zhang, PoS **ICHEP2016**, 210 (2016), arXiv:1611.06212 [hep-ph].

[19] A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, JCAP **05**, 006 (2017), arXiv:1702.01141 [hep-ph].

[20] I. Baldes, M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, SciPost Phys. **4**, 041 (2018), arXiv:1712.07489 [hep-ph].

[21] L. Pearce and A. Kusenko, Phys. Rev. D **87**, 123531 (2013), arXiv:1303.7294 [hep-ph].

[22] R. Mahbubani, M. Redi, and A. Tesi, Phys. Rev. D **101**, 103037 (2020), arXiv:1908.00538 [hep-ph].

[23] R. Mahbubani, M. Redi, and A. Tesi, JCAP **02**, 039 (2021), arXiv:2007.07231 [hep-ph].

[24] J. Chen, Z.-L. Liang, Y.-L. Wu, and Y.-F. Zhou, JCAP **12**, 021 (2015), arXiv:1505.04031 [hep-ph].

[25] J. L. Feng, J. Smolinsky, and P. Tanedo, Phys. Rev. D **93**, 115036 (2016), [Erratum: Phys.Rev.D 96, 099903 (2017)], arXiv:1602.01465 [hep-ph].

[26] C. Arina, M. Backović, J. Heisig, and M. Lucente, Phys. Rev. D **96**, 063010 (2017), arXiv:1703.08087 [astro-ph.HE].

[27] R. K. Leane, K. C. Y. Ng, and J. F. Beacom, Phys. Rev. D **95**, 123016 (2017), arXiv:1703.04629 [astro-ph.HE].

[28] M. Lucente, C. Arina, M. Backović, and J. Heisig, PoS **EPS-HEP2017**, 628 (2017), arXiv:1710.03947 [hep-ph].

[29] C. Niblaeus, A. Beniwal, and J. Edsjo, JCAP **11**, 011 (2019), arXiv:1903.11363 [astro-ph.HE].

[30] A. Cuoco, P. De La Torre Luque, F. Gargano, M. Gustafsson, F. Loparco, M. N. Mazziotta, and D. Serini, Phys. Rev. D **101**, 022002 (2020), arXiv:1912.09373 [astro-ph.HE].

[31] M. N. Mazziotta, F. Loparco, D. Serini, A. Cuoco, P. De La Torre Luque, F. Gargano, and M. Gustafsson, Phys. Rev. D **102**, 022003 (2020), arXiv:2006.04114 [astro-ph.HE].

[32] B. Dasgupta, A. Gupta, and A. Ray, JCAP **10**, 023 (2020), arXiv:2006.10773 [hep-ph].

[33] N. F. Bell, J. B. Dent, and I. W. Sanderson, Phys. Rev. D **104**, 023024 (2021), arXiv:2103.16794 [hep-ph].

[34] R. Garani and S. Palomares-Ruiz, JCAP **05**, 042 (2022), arXiv:2104.12757 [hep-ph].

[35] A. A. Abud *et al.* (DUNE), JCAP **10**, 065 (2021), arXiv:2107.09109 [hep-ex].

[36] R. K. Leane, T. Linden, P. Mukhopadhyay, and N. Toro, Phys. Rev. D **103**, 075030 (2021), arXiv:2101.12213 [astro-ph.HE].

[37] R. Abbasi *et al.* (IceCube), Phys. Rev. D **105**, 062004 (2022), arXiv:2111.09970 [astro-ph.HE].

[38] N. F. Bell, M. J. Dolan, and S. Robles, JCAP **11**, 004 (2021), arXiv:2107.04216 [hep-ph].

[39] M. Zakeri and Y.-F. Zhou, JCAP **04**, 026 (2022), arXiv:2109.11662 [hep-ph].

[40] D. Bose, T. N. Maity, and T. S. Ray, Phys. Rev. D **105**, 123013 (2022), arXiv:2112.08286 [hep-ph].

[41] T. N. Maity, A. K. Saha, S. Mondal, and R. Laha, (2023), arXiv:2308.12336 [hep-ph].

[42] M. I. Gresham and K. M. Zurek, Phys. Rev. D **99**, 083008 (2019), arXiv:1809.08254 [astro-ph.CO].

[43] R. Garani, M. H. G. Tytgat, and J. Vandecasteele, Phys. Rev. D **106**, 116003 (2022), arXiv:2207.06928 [hep-ph].

[44] W. H. Press and D. N. Spergel, Astrophys. J. **296**, 679 (1985).

[45] K. Griest and D. Seckel, Nucl. Phys. B **283**, 681 (1987), [Erratum: Nucl.Phys.B 296, 1034–1036 (1988)].

[46] M. B. Wise and Y. Zhang, JHEP **02**, 023 (2015), [Erratum: JHEP 10, 165 (2015)], arXiv:1411.1772 [hep-ph].

[47] E. Hardy, R. Lasenby, J. March-Russell, and S. M. West, JHEP **06**, 011 (2015), arXiv:1411.3739 [hep-ph].

[48] M. I. Gresham, H. K. Lou, and K. M. Zurek, Phys. Rev. D **97**, 036003 (2018), arXiv:1707.02316 [hep-ph].

[49] M. I. Gresham, H. K. Lou, and K. M. Zurek, Phys. Rev. D **96**, 096012 (2017), arXiv:1707.02313 [hep-ph].

[50] D. N. Spergel and W. H. Press, Astrophys. J. **294**, 663 (1985).

[51] M. Asplund, N. Grevesse, A. Sauval, and P. Scott, Ann. Rev. Astron. Astrophys. **47**, 481 (2009), arXiv:0909.0948 [astro-ph.SR].

[52] R. Garani and S. Palomares-Ruiz, JCAP **05**, 007 (2017), arXiv:1702.02768 [hep-ph].

[53] A. Bottino, G. Fiorentini, N. Fornengo, B. Ricci, S. Scopel, and F. L. Villante, Phys. Rev. D **66**, 053005 (2002), arXiv:hep-ph/0206211.

[54] J. D. Lewin and P. F. Smith, Astropart. Phys. **6**, 87 (1996).

[55] G. Duda, A. Kemper, and P. Gondolo, JCAP **04**, 012 (2007), arXiv:hep-ph/0608035.

[56] M. B. Wise and Y. Zhang, Phys. Rev. D **90**, 055030 (2014), [Erratum: Phys.Rev.D 91, 039907 (2015)], arXiv:1407.4121 [hep-ph].

[57] S. Tulin, H.-B. Yu, and K. M. Zurek, Phys. Rev. D **87**, 115007 (2013), arXiv:1302.3898 [hep-ph].

[58] B. Colquhoun, S. Heeba, F. Kahlhoefer, L. Sagunski, and S. Tulin, Phys. Rev. D **103**, 035006 (2021), arXiv:2011.04679 [hep-ph].

[59] L. D. Landau and E. M. Lifshits, *Quantum Mechanics*, Course of Theoretical Physics, Vol. v.3 (Butterworth-Heinemann, Oxford, 1991).

[60] S. A. Khrapak, A. V. Ivlev, G. E. Morfill, and S. K. Zhdanov, Phys. Rev. Lett. **90**, 225002 (2003).

[61] S. Khrapak, A. Ivlev, G. Morfill, S. Zhdanov, and H. Thomas, IEEE Transactions on Plasma Science **32**, 555 (2004).

[62] A. R. Zentner, Phys. Rev. D **80**, 063501 (2009), arXiv:0907.3448 [astro-ph.HE].

[63] A. Ibarra, S. Lopez Gehler, and M. Pato, JCAP **07**, 043 (2012), arXiv:1205.0007 [hep-ph].

[64] E. Richard *et al.* (Super-Kamiokande), Phys. Rev. D **94**, 052001 (2016), arXiv:1510.08127 [hep-ex].

[65] K. Abe *et al.* (Super-Kamiokande), Phys. Rev. D **104**, 122002 (2021), arXiv:2109.11174 [astro-ph.HE].

[66] S. Abe *et al.* (KamLAND), Astrophys. J. **925**, 14 (2022), arXiv:2108.08527 [astro-ph.HE].

[67] G. Battistoni, A. Ferrari, T. Montaruli, and P. R. Sala, Astropart. Phys. **23**, 526 (2005).

[68] B. Zhou and J. F. Beacom, (2023), arXiv:2311.05675 [hep-ph].

[69] A. M. Suliga and J. F. Beacom, Phys. Rev. D **108**, 043035 (2023), arXiv:2306.11090 [hep-ph].

[70] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara, and S. Midorikawa, Phys. Rev. D **92**, 023004 (2015), arXiv:1502.03916 [astro-ph.HE].

[71] G. Krnjaic, Phys. Rev. D **94**, 073009 (2016), arXiv:1512.04119 [hep-ph].

[72] M. W. Winkler, Phys. Rev. **D99**, 015018 (2019), arXiv:1809.01876 [hep-ph].

[73] S. Matsumoto, Y.-L. S. Tsai, and P.-Y. Tseng, JHEP **07**, 050 (2019), arXiv:1811.03292 [hep-ph].

[74] D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tittley, Science **347**, 1462 (2015), arXiv:1503.07675 [astro-ph.CO].

[75] K. Bondarenko, A. Boyarsky, T. Bringmann, and A. Sokolenko, JCAP **04**, 049 (2018), arXiv:1712.06602 [astro-ph.CO].

[76] D. Harvey, A. Robertson, R. Massey, and I. G. McCarthy, Mon. Not. Roy. Astron. Soc. **488**, 1572 (2019), arXiv:1812.06981 [astro-ph.CO].

[77] L. Sagunski, S. Gad-Nasr, B. Colquhoun, A. Robertson, and S. Tulin, JCAP **01**, 024 (2021), arXiv:2006.12515 [astro-ph.CO].

[78] D. Cross *et al.* (DES), (2023), arXiv:2304.10128 [astro-ph.CO].

[79] S. Tulin and H.-B. Yu, Phys. Rept. **730**, 1 (2018), arXiv:1705.02358 [hep-ph].

[80] S. Adhikari *et al.*, (2022), arXiv:2207.10638 [astro-ph.CO].

[81] M. Vogelsberger, J. Zavala, and A. Loeb, Mon. Not. Roy. Astron. Soc. **423**, 3740 (2012), arXiv:1201.5892 [astro-ph.CO].

[82] J. Zavala, M. Vogelsberger, and M. G. Walker, Mon. Not. Roy. Astron. Soc. **431**, L20 (2013), arXiv:1211.6426 [astro-ph.CO].

[83] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, Mon. Not. Roy. Astron. Soc. **453**, 29 (2015), arXiv:1412.1477 [astro-ph.GA].

[84] T. Ren, A. Kwa, M. Kaplinghat, and H.-B. Yu, Phys. Rev. X **9**, 031020 (2019), arXiv:1808.05695 [astro-ph.GA].

[85] E. Aprile *et al.* (XENON), Phys. Rev. Lett. **121**, 111302 (2018), arXiv:1805.12562 [astro-ph.CO].

[86] P. Athron *et al.* (GAMBIT), Eur. Phys. J. C **79**, 38 (2019), arXiv:1808.10465 [hep-ph].

[87] J. Aalbers *et al.* (LZ), Phys. Rev. Lett. **131**, 041002 (2023), arXiv:2207.03764 [hep-ex].

[88] K. Agashe, Y. Cui, L. Necib, and J. Thaler, JCAP **10**, 062 (2014), arXiv:1405.7370 [hep-ph].

[89] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, JCAP **03**, 051 (2011), [Erratum: JCAP 10, E01 (2012)], arXiv:1012.4515 [hep-ph].

[90] C. A. Argüelles, A. Diaz, A. Kheirandish, A. Olivares-Del-Campo, I. Safa, and A. C. Vincent, Rev. Mod. Phys. **93**, 035007 (2021), arXiv:1912.09486 [hep-ph].

[91] G. Elor, N. L. Rodd, T. R. Slatyer, and W. Xue, JCAP **06**, 024 (2016), arXiv:1511.08787 [hep-ph].

[92] X. Chu, C. Garcia-Cely, and H. Murayama, JCAP **06**, 043 (2020), arXiv:1908.06067 [hep-ph].

[93] D. Yang and H.-B. Yu, JCAP **09**, 077 (2022), arXiv:2205.03392 [astro-ph.CO].

[94] S. Cassel, J. Phys. G **37**, 105009 (2010), arXiv:0903.5307 [hep-ph].

[95] L. Landau and E. Lifshitz, *Mechanics: Volume 1*, Course of theoretical physics (Elsevier Science, 1976).

[96] C. Kouvaris and P. Tinyakov, Phys. Rev. D **83**, 083512 (2011), arXiv:1012.2039 [astro-ph.HE].

[97] P. Tinyakov, M. Pshirkov, and S. Popov, Universe **7**, 401 (2021), arXiv:2110.12298 [astro-ph.HE].

[98] B. Bertoni, A. E. Nelson, and S. Reddy, Phys. Rev. D **88**, 123505 (2013), arXiv:1309.1721 [hep-ph].

[99] R. Garani, A. Gupta, and N. Raj, Phys. Rev. D **103**, 043019 (2021), arXiv:2009.10728 [hep-ph].

[100] S. Reddy, M. Prakash, and J. M. Lattimer, Phys. Rev. D **58**, 013009 (1998), arXiv:astro-ph/9710115.

Polarized and unpolarized off-shell $H^* \to ZZ \to 4\ell$ decay above the $2m_Z$ threshold

A. I. Hernández-Juárez*,[1] R. Gaitán,[1] and G. Tavares-Velasco[2]

[1] *Departamento de Física, FES-Cuautitlán, Universidad Nacional Autónoma de México, C.P. 54770, Estado de México, México.**

[2] *Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Apartado Postal 1152, Puebla, Puebla, México*

(Dated: February 29, 2024)

An analysis of the off-shell $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width is presented for both unpolarized and polarized Z gauge bosons in the scenario with the most general H^*ZZ vertex function, which is given in terms of two CP-even ($\hat{b}_Z$ and $\hat{c}_Z$) and one CP-odd ($\tilde{b}_Z$) anomalous couplings. The SM contributions to the H^*ZZ coupling up to the one-loop level are also included. Explicit analytic results for the unpolarized and polarized $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ square amplitudes and the four-body phase space are presented, out of which several observable quantities can be obtained straightforwardly. As far as the numerical analysis is concerned, a cross-check is performed via Madgraph, where our model was implemented with the aid of FeynRules. We then consider the most stringent bounds on anomalous complex H^*ZZ couplings and analyze the effects of the polarizations of the Z gauge bosons through the polarized $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width as well as left-right and forward-backward asymmetries, which are found to be sensitive to new-physics effects. Particular focus is put on the effects of the absorptive parts of the anomalous H^*ZZ couplings, which have been largely overlooked up to now in LHC analyses. It is found that the studied observable quantities, particularly the left-right asymmetries, can be helpful to look for effects of CP-violation in the H^*ZZ coupling and set bounds on the absorptive parts at the LHC and future colliders. For completeness we also analyze the case of unpolarized Z gauge bosons.

arXiv:2402.18497v1 [hep-ph] 28 Feb 2024

I. INTRODUCTION

All the experimental data collected until now indicate that the properties of the scalar particle discovered in 2012 at the LHC by the ATLAS and CMS collaborations [1, 2] are compatible with those of the long-sought-for Higgs boson, which is the remanent of the mechanism of spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry in the Glashow-Salam-Weinberg Standard Model (SM). Nonetheless, there are still several couplings of such a particle that remain to be measured with enough accuracy, such as for the $H\bar{b}b$ and $H\mu^-\mu^+$ interactions, whereas other ones are still not yet at the reach of experimental measurement, such as the Higgs boson couplings to light fermions, let alone the Higgs boson self couplings. Therefore, the study of the Higgs boson phenomenology is expected to play an important role to probe the SM and search for any new physics effects in present and future colliders. Along these lines, the Higgs boson couplings to the weak gauge bosons HZZ and HWW, which in the SM arise at the tree level, have long drawn considerable attention both theoretically and experimentally. As far as the HZZ coupling is concerned, it has played a crucial part in the experimental study of the Higgs boson at the LHC via the $H \to ZZ^* \to 4\ell$ decay, which provides a clear signal and allows for a precise resolution of the Higgs boson mass despite that its branching ratio is considerably smaller than those of other decay channels such as $H \to WW^*$, $H \to \tau^+\tau^-$, and $H \to \bar{b}b$.

Although one-loop corrections to the HZZ vertex were calculated long ago in the framework of the SM [3, 4], more recently a new evaluation was presented [5] that could be more suitable for the recent progress on the theoretical and experimental study of this coupling. The most general HZZ vertex can be parametrized by an effective Lagrangian given in terms of four coupling constants:

$$\mathcal{L}^{HZZ} = \frac{g}{c_W} m_Z \left[\frac{(1-a_Z)}{2} H Z_\mu Z^\mu + \frac{1}{2m_Z^2} \left\{ \hat{b}_Z H Z_{\mu\nu} Z^{\mu\nu} + \hat{c}_Z H Z_\mu \partial_\nu Z^{\mu\nu} + \widetilde{b}_Z H Z_{\mu\nu} \widetilde{Z}^{\mu\nu} \right\} \right], \tag{1}$$

where a_Z stands for the corrections to the SM tree-level coupling via loop contributions or new physics effects, whereas $\hat{b}_Z$, $\hat{c}_Z$ and $\widetilde{b}_Z$ are absent at the tree level in the SM and thus are dubbed anomalous couplings. While the anomalous CP-conserving coupling $\hat{b}_Z$ is induced at the one-loop level [3, 4] and can reach values of the order of 10^{-3} [5], the CP-violating one $\widetilde{b}_Z$ would arise up to the three-loop level and has been estimated to be of the order of 10^{-11} [6].

* alaban7_3@hotmail.com

The phenomenology of the HZZ anomalous couplings has been studied both in lepton [7–9] and hadron colliders [10]: it was found that non-SM contributions to the HZZ coupling could be strongly constrained in the future, which opens up the window for the detection of new physics effects via HZZ-mediated processes. In particular, the anomalous couplings have been studied through polarization observables in the $Z^* \to HZ$ process at the LHC, e^-e^-, ep and γe colliders [8, 11–21]. Very recently, the CMS collaboration used effective ratios to obtain indirect bounds on such anomalous couplings [22]. Even more, a milestone was achieved in 2022 when the CMS and ATLAS collaborations reported the measurement of the off shell H^*ZZ coupling via the $pp \to H^* \to ZZ$ process [22, 23], which has motivated a plethora of phenomenological studies. For instance, bounds on the anomalous HZZ couplings were obtained by combining LHC data and theoretical results [5], which can be as tight as $10^{-2} - 10^{-4}$ depending of the energy region. Other off-shell couplings have been also studied recently, such as trilinear neutral gauge boson couplings [24–27] and the coupling of the gluon with a quark-antiquar pair [28, 29].

In this work we are interested in the study of the potential detection of new physics effects on the HZZ anomalous couplings via polarized Z gauge bosons at the LHC through the off-shell decay $H^* \to ZZ \to 4\ell$. We will consider the most general scenario with both Z gauge bosons on shell and complex anomalous HZZ couplings. We will analyze the behavior of some observables that can be sensitive to new physics contributions. In particular, we are interested in the effects of the absorptive parts of the anomalous HZZ couplings, which have been overlooked in the past but can lead to interesting results that could be explored at the LHC and future colliders. For the sake of completeness we also discuss the scenario with unpolarized Z gauge bosons.

The rest of this work is as follows. Section II is devoted to discuss the theoretical framework appropriate for the study of the $H^* \to ZZ \to 4\ell$ decay width, with explicit analytical expressions for the polarized and unpolarized square amplitudes, out of which the respective $H^* \to ZZ \to 4\ell$ differential decay widths are obtained. The numerical analysis is presented in Sec. II, where we cross-check our calculation method with an alternative evaluation performed via MadGraph. The behavior of the differential decay, angular distributions along with left-right and a forward-backward asymmetries are analyzed in some realistic scenarios for the values of the real and absorptive parts of the anomalous HZZ couplings. In Section IV we present our concluding remarks and outlook. Finally, the kinematics of the $H^* \to ZZ \to 4\ell$ decay and the four-body phase space are discussed in Appendix A.

II. POLARIZED AND UNPOLARIZED $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ DECAY WIDTHS

We now present all the analytical formulas required for the calculation of the unpolarized and polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width. We will use the most general form of the HZZ vertex functions with complex anomalous couplings and also include the contributions of the SM up to the one-loop level [5].

A. HZZ vertex function

The S-matrix element obtained from Lagrangian (1) can be written as

$$i\mathcal{M} = i\frac{g}{c_W} m_Z \Gamma^{ZZH}_{\mu\nu}(q^2) Z^\mu(p_1) Z^\nu(p_2) H(q), \tag{2}$$

where the vertex function $\Gamma^{ZZH}_{\mu\nu}(q^2)$ can be written as follows for on-shell Z gauge bosons and an off-shell Higgs boson

$$\Gamma^{ZZH}_{\mu\nu}(q^2) = h_1^H(q^2) g_{\mu\nu} + \frac{h_2^H(q^2)}{m_Z^2} p_{1\nu} p_{2\mu} + \frac{h_3^H(q^2)}{m_Z^2} \epsilon_{\mu\nu\alpha\beta} p_1^\alpha p_2^\beta, \tag{3}$$

where h_i^H (i =1, 2 ,3) are form factors that are given in terms of the anomalous couplings of Lagrangian (1) as follows

$$h_1^H(q^2) = 1 + a_Z - \hat{b}_Z \frac{q^2 - 2m_Z^2}{m_Z^2} + \hat{c}_Z, \tag{4}$$

$$h_2^H(q^2) = \pm 2\hat{b}_Z, \tag{5}$$

$$h_3^H(q^2) = \pm 2\widetilde{b}_Z. \tag{6}$$

These form factors are both explicit and implicit functions of q^2 since the anomalous couplings also depend on the off-shell Higgs boson four-momentum.

Since the main interest of this work is the study of the anomalous contributions, we will consider a vanishing a_Z. For a study of the SM one-loop contributions to the a_Z coupling along with the renormalization procedure, we refer the reader to Ref. [3].

Since the anomalous couplings can be taken as complex in general [5, 30], the form factors of (4)-(6) will be written as follows

$$h_i^H = \mathrm{Re}\big[h_i^H\big] + i\mathrm{Im}\big[h_i^H\big]. \tag{7}$$

Although the absorptive parts have been neglected in experimental analyses at the LHC, their effects have been discussed in some theoretical works [7, 12, 13] and more recently via a left-right asymmetry [5], which may be reconstructed in the final state of the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay. Polarization effects of the Z gauge bosons on the Higgs boson decay $H \to ZZ$ have already been studied [31, 32], though only non-anomalous HZZ couplings were considered and only one Z gauge boson was taken on-shell. Similar calculations were also reported in [6, 33–35]. On the other hand, the scenario where both Z gauge bosons are on-shell and the HZZ coupling is purely real was studied in Refs. [4, 36, 37]. Finally, a restrictive scenario where not all the HZZ couplings were taken as complex was considered in Ref. [12].

In this work, we are interested in the most general scenario where all the anomalous HZZ couplings are complex and calculate the unpolarized and polarized off-shell $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width. We will then analyze the implications of the absorptive parts of the HZZ anomalous couplings on some observables.

B. Polarized Amplitudes

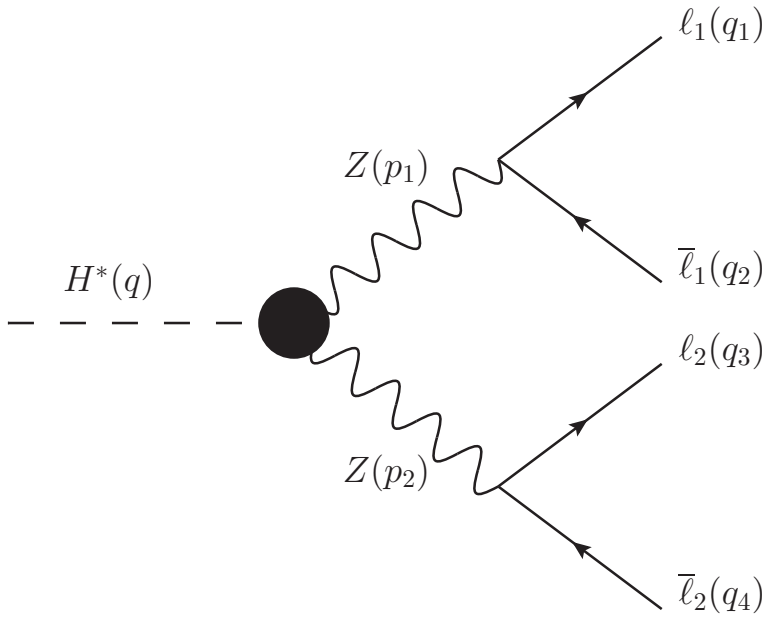


FIG. 1. Nomenclature for the four-momenta of the particles involved in the Feynman diagram for the contribution of the HZZ anomalous couplings to the Higgs boson decay $H^* \to Z(p_1)Z(p_2) \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$.

Using the nomenclature of Fig. 1, the amplitude for the $H^* \to Z(p_1)Z(p_2) \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ process can be written as

$$\mathcal{M}_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2} = \mathcal{M}^{\mu\nu}_{H^*\to Z(p_1)Z(p_2)} \frac{i\sum \epsilon_\eta(p_1,\lambda_1)\epsilon^*_\mu(p_1,\lambda_1)}{p_1^2 - m_Z^2 + i\Gamma_Z m_Z} \mathcal{M}^{\eta}_{Z(p_1)\to\bar{\ell}_1\ell_1} \frac{i\sum \epsilon_\kappa(p_2,\lambda_2)\epsilon^*_\nu(p_2,\lambda_2)}{p_2^2 - m_Z^2 + i\Gamma_Z m_Z} \mathcal{M}^{\kappa}_{Z(p_2)\to\bar{\ell}_2\ell_2}, \tag{8}$$

where Γ_Z is the Z gauge boson decay width, and, according to Lagrangian (1), the amplitude for the process $H^* \to ZZ$ is given in terms of the complex h_i^H form factors as follows

$$\begin{aligned}\mathcal{M}^{\mu\nu}_{H^*\to Z(p_1)Z(p_2)} =& i\Big(\frac{g}{c_W}\Big) m_Z \Big\{ g^{\mu\nu}\Big(\mathrm{Re}\big[h_1^H\big] + i\mathrm{Im}\big[h_1^H\big]\Big) + \frac{p_2^\mu {p_1}^\nu}{m_Z^2}\Big(\mathrm{Re}\big[h_2^H\big] + i\mathrm{Im}\big[h_2^H\big]\Big) \\ &+ \frac{\epsilon^{\mu\nu\alpha\beta} p_{1\alpha} p_{2\beta}}{m_Z^2}\Big(\mathrm{Re}\big[h_3^H\big] + i\mathrm{Im}\big[h_3^H\big]\Big)\Big\},\end{aligned} \tag{9}$$

whereas the amplitudes for the $Z \to \bar{\ell}_1 \ell_1$ and $Z \to \bar{\ell}_2 \ell_2$ processes are

$$\mathcal{M}^{\eta}_{Z(p_1)\to\bar{\ell}_1\ell_1} = i\frac{g}{c_W}\overline{u}(q_2)\gamma^{\eta}\big(g_V - g_A\gamma^5\big)u(q_1), \tag{10}$$

$$\mathcal{M}^{\kappa}_{Z(p_2)\to\bar{\ell}_2\ell_2} = i\frac{g}{c_W}\overline{u}(q_4)\gamma^{\kappa}\big(g_V - g_A\gamma^5\big)u(q_3). \tag{11}$$

For the Z gauge boson propagators of Eq. (8) we have used the Breit-Wigner approximation and the completeness relation [38]

$$-g^{\mu\nu} + \frac{k^{\mu}k^{\nu}}{m_Z^2} = \sum_{\lambda=1}^{3}\epsilon^{\mu}_{\lambda}(k)\epsilon^{*\nu}_{\lambda}(k), \tag{12}$$

where we only consider 3 polarizations since both Z gauge bosons are on-shell.

In order to study the polarization effects in the final state, it is convenient to rewrite Eq. (8) as a sum of the polarized amplitudes

$$\mathcal{M}_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}(\lambda_1,\lambda_2) = \sum_{\lambda_1}\sum_{\lambda_2}\frac{\mathcal{M}_{H^*\to Z(p_1)Z(p_2)}(\lambda_1,\lambda_2)\mathcal{M}_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda_1)\mathcal{M}_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda_2)}{\big(p_1^2 - m_Z^2 + i\Gamma m_Z\big)\big(p_2^2 - m_Z^2 + i\Gamma m_Z\big)} \tag{13}$$

where

$$\mathcal{M}_{H^*\to Z(p_1)Z(p_2)}(\lambda_1,\lambda_2) = \mathcal{M}^{\mu\nu}_{H^*\to Z(p_1)Z(p_2)}\epsilon^{*}_{\mu}(p_1,\lambda_1)\epsilon^{*}_{\nu}(p_2,\lambda_2), \tag{14}$$

$$\mathcal{M}_{Z(p_i)\to\bar{\ell}_i\ell_i}(\lambda_i) = \mathcal{M}^{\mu}_{Z(p_i)\to\bar{\ell}_i\ell_i}\epsilon_{\mu}(p_i,\lambda_i). \tag{15}$$

Since the $\mathcal{M}_{H^*\to Z(p_1)Z(p_2)}(\lambda_1,\lambda_2)$ polarized amplitude is non-vanishing for $\lambda_1 = \lambda_2$ only [5, 37], the square amplitude can be written as

$$\mathcal{M}^2_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2} = \frac{\sum_{\lambda}\mathcal{M}^2_{H^*\to Z(p_1)Z(p_2)}(\lambda,\lambda)\mathcal{M}^2_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda)\mathcal{M}^2_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda) + \mathcal{M}^2_{\rm int}}{\left((p_1^2 - m_Z^2)^2 + \Gamma^2 m_Z^2\right)\left((p_2^2 - m_Z^2)^2 + \Gamma^2 m_Z^2\right)}. \tag{16}$$

where the interference term $\mathcal{M}^2_{\rm int}$ is given by

$$\begin{aligned}\mathcal{M}^2_{\rm int} =& \sum_{\lambda_1}\sum_{\lambda_2\neq\lambda_1}\mathcal{M}_{H^*\to Z(p_1)Z(p_2)}(\lambda_1,\lambda_1)\mathcal{M}^{\dagger}_{H^*\to Z(p_1)Z(p_2)}(\lambda_2,\lambda_2)\mathcal{M}_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda_1)\mathcal{M}^{\dagger}_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda_2)\\ &\times\mathcal{M}_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda_1)\mathcal{M}^{\dagger}_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda_2).\end{aligned} \tag{17}$$

From the first term of the right-hand side of Eq. (16) we can extract the contributions of polarized Z gauge bosons with polarizations $\lambda = L$, R, and 0, which allows one to study the polarized decay width and other observables of interest. As described below, to obtain the $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width we will follow the approach of Ref. [33] for the calculation of the $H \to ZZ^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay. A similar approach was followed in Ref. [39, 40] for the decay $K^{\pm} \to \pi^{\pm}\pi^0 e^+e^-$, whereas an alternative approach was implemented in Ref. [34] for the $H \to ZZ^* \to \ell^+\ell^-\tau^+\tau^-$ and $H \to WW^* \to \ell^-\overline{\nu}_{\ell}\tau^+\nu_{\tau}$ decays using transformation properties of the helicity amplitudes under the rotation group.

The phase space for the $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ process, along with the kinematics, is presented in Appendix A. The relevant angular variables on which the square $\mathcal{M}_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}$ amplitude depends are θ_1, θ_2 and ϕ, which are described in Fig 2. We note that the square polarized amplitudes $\mathcal{M}^2_{H^*\to ZZ}(\lambda_1,\lambda_2)$, $\mathcal{M}^2_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda_1)$ and $\mathcal{M}^2_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda_2)$ can be computed in the corresponding H^*, $Z(p_1)$ and $Z(p_2)$ rest frames since each one depends only on scalar products of four-momenta given in the same inertial frame. Even more, these square amplitudes are Lorentz invariant and it is not necessary to apply a boost to bring them into a common reference system. On the other hand, the interference term $\mathcal{M}^2_{H^*Z(p_1)Z(p_2)}(\lambda_1,\lambda_2)$ is given in terms of partial amplitudes evaluated in distinct reference frames, so a boost from the $Z(p_i) \to \bar{\ell}_i\ell_i$ reference frames into the Higgs boson rest frame will be required to perform the sum of Eq. (17). The same is true for the calculation of the unpolarized interference term via standard techniques as scalar products of four-momenta given in distinct reference frames are also involved.

We present below the analytical expressions for the polarized square amplitudes and the interference term of Eq. (16), which were calculated using the FeynCalc package [41–43]. The relation $p_i^2 = m_Z^2$ ($i = 1, 2$) will be considered as the energy region where both Z gauge bosons are on-shell is of interest for our analysis.

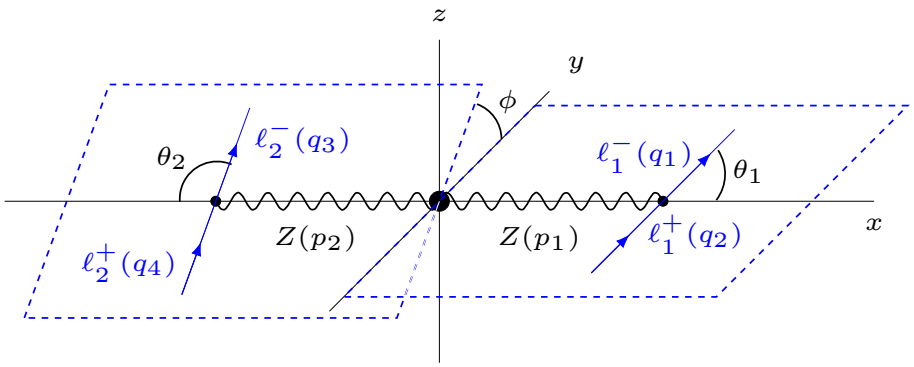


FIG. 2. Nomenclature for the reference systems and angles used in the calculation of the $H \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay, as described in Appendix A.

1. *Polarized $H^* \to Z(p_1)Z(p_2)$ square amplitudes*

The polarized square amplitudes $\mathcal{M}^2_{H^*}(\lambda,\lambda)$ given in terms of the real and absorptive parts of the complex anomalous couplings h_i^H of Lagrangian (1) were presented by some of us in a previous work [5]. We consider a reference frame in which the motion of the $Z(p_1)$ gauge boson is along the positive direction of the x axis and the off-shell Higgs boson is at rest. For further details see Appendix A. In the energy region where both Z gauge bosons are on-shell, the transversally polarized $H^* \to ZZ$ amplitudes read as follows

$$\mathcal{M}^2_{H^*\to Z(p_1)Z(p_2)}(L,L) = \left(\frac{g}{2c_W m_Z}\right)^2 \left(4m_Z^2 Q\sqrt{Q^2-4m_Z^2}\,\mathrm{Im}\left(h_3^H h_1^{H\dagger}\right) + Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2 + 4m_Z^4\left|h_1^H\right|^2\right), \tag{18}$$

and

$$\mathcal{M}^2_{H^*\to Z(p_1)Z(p_2)}(R,R) = \left(\frac{g}{2c_W m_Z}\right)^2 \left(-4m_Z^2 Q\sqrt{Q^2-4m_Z^2}\,\mathrm{Im}\left(h_3^H h_1^{H\dagger}\right) + Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2 + 4m_Z^4\left|h_1^H\right|^2\right), \tag{19}$$

whereas for the longitudinal polarization we obtain

$$\begin{aligned}\mathcal{M}^2_{H^*\to Z(p_1)Z(p_2)}(0,0) = \left(\frac{g}{4c_W m_Z^3}\right)^2 \Bigg(&4Q^6 m_Z^2\left(\mathrm{Re}\left(h_1^H h_2^{H\dagger}\right) - 2\left|h_2^H\right|^2\right) + 4Q^4 m_Z^4\left(\left|h_1^H\right|^2 + 4\left|h_2^H\right|^2 - 6\mathrm{Re}\left(h_1^H h_2^{H\dagger}\right)\right)\\ &- 16Q^2 m_Z^6\left(\left|h_1^H\right|^2 - 2\mathrm{Re}\left(h_1^H h_2^{H\dagger}\right)\right) + 16m_Z^8\left|h_1^H\right|^2 + Q^8\left|h_2^H\right|^2\Bigg),\end{aligned} \tag{20}$$

where we have introduced the four-lepton invariant mass $Q^2 = q^2$, instead of the four-momentum of the off-shell Higgs boson. Our results reproduce those reported in Ref. [37] in the scenario with real anomalous couplings.

Note that there is no dependence on the angular variables. Also, a non-vanishing h_3^H form factor may give rise to a left-right asymmetry in the four-lepton final state, which vanishes up to the one-loop level in the SM.

2. *$Z(p_i) \to 2\ell_i$ square amplitudes*

As already mentioned, the square amplitudes $\mathcal{M}^2_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda)$ and $\mathcal{M}^2_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda)$ can be worked out in the rest frame of the $Z(p_i)$ gauge boson. According to the kinematics presented in Appendix A, the polarization vectors of the $Z(p_1)$ ($Z(p_2)$) gauge boson are oriented along the positive (negative) direction of the x axis, thus the transversally

and longitudinally polarized amplitudes are given as

$$\mathcal{M}^2_{Z(p_i)\to\bar{\ell}_i\ell_i}(R/L) = \left(\frac{g}{c_W}\right)^2 \left(s^2_{\theta_i}\left(4m_i^2 - m_Z^2\right)\left(g_A^2+g_V^2\right) \pm 4g_A g_V c_{\theta_i} m_Z\sqrt{m_Z^2-4m_i^2} + 2m_Z^2\left(g_A^2+g_V^2\right) - 8m_i^2 g_A^2\right)$$
$$\simeq \left(\frac{g}{c_W}\right)^2 m_Z^2\left(\left(g_A^2+g_V^2\right)\left(1+c^2_{\theta_i}\right) \pm 4g_A g_V c_{\theta_i}\right), \tag{21}$$

and

$$\mathcal{M}^2_{Z(p_i)\to\bar{\ell}_i\ell_i}(0) = 2\left(\frac{g}{c_W}\right)^2 \left(c^2_{\theta_i}\left(4m_i^2 - m_Z^2\right)\left(g_A^2+g_V^2\right) + m_Z^2\left(g_A^2+g_V^2\right) - 4m_i^2 g_A^2\right)$$
$$\simeq 2\left(\frac{g}{c_W}\right)^2 s^2_{\theta_i} m_Z^2\left(g_A^2+g_V^2\right), \tag{22}$$

where $i = 1, 2$ and we have introduced the short-hand notation $c_\eta \equiv \cos\eta$ and $s_\eta \equiv \sin\eta$. The approximate results were obtained in the massless lepton limit, which is used in the phase space defined in Appendix A.

To cross-check our calculation, we have boosted Eqs. (21) and (22) into the Higgs boson rest frame via the Lorentz transformation given in Appendix A, which leaves them invariant as expected. We also note that the above square amplitudes are independent of the ϕ angle and the four-lepton invariant mass Q. Also, although the $g_A g_V$ term in Eq. (21) seems to give rise to a left-right asymmetry, it vanishes after integration over the θ_i angles. Therefore, the only observable effects of the left- and right-handed polarized $Z(p_i) \to \bar{\ell}_i\ell_i$ amplitudes would appear via either angular distributions or a forward-backward asymmetry.

3. Interference term

For the purpose of our work, it is enough to consider the massless lepton approximation, which yields

$$\mathcal{M}^2_{\rm int} \equiv \sum_{\lambda_1}\sum_{\lambda_2\neq\lambda_1} \mathcal{M}^2_{H^* Z(p_1) Z(p_2)}(\lambda_1,\lambda_2)$$
$$= \frac{s_{\theta_1}s_{\theta_2}}{2m_Z^2}\Big(s_{\theta_1}s_{\theta_2}m_Z^2\left(g_A^2+g_V^2\right)^2 f_1\left(Q^2\right) + \left(2g_A g_V + c_{\theta_1}\left(g_A^2+g_V^2\right)\right)\left(2g_A g_V + c_{\theta_2}\left(g_A^2+g_V^2\right)\right) f_2\left(Q^2\right)$$
$$+ \left(2g_A g_V - c_{\theta_1}\left(g_A^2+g_V^2\right)\right)\left(2g_A g_V - c_{\theta_2}\left(g_A^2+g_V^2\right)\right) f_3\left(Q^2\right)\Big), \tag{23}$$

where

$$f_1\left(Q^2\right) = 4\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\, m_Z^2 \mathrm{Re}\left(h_1^H h_3^{H^\dagger}\right) s_{2\phi} + \left(4m_Z^4\left|h_1^H\right|^2 - Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2\right) c_{2\phi}, \tag{24}$$

$$f_2\left(Q^2\right) = c_\phi\Bigg(\left(Q^2\left(Q^2-4m_Z^2\right)\right)^{3/2}\mathrm{Im}\left(h_3^H h_2^{H^\dagger}\right)$$
$$+ 2m_Z^2\left(\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\left(Q^2-2m_Z^2\right)\mathrm{Im}\left(h_3^H h_1^{H^\dagger}\right) + \left(4m_Z^2-Q^2\right)Q^2\mathrm{Re}\left(h_1^H h_2^{H^\dagger}\right)\right)\Bigg)$$
$$- s_\phi\Bigg(\left(Q^2\left(Q^2-4m_Z^2\right)\right)^{3/2}\mathrm{Re}\left(h_2^H h_3^{H^\dagger}\right)$$
$$+ 2m_Z^2\left(\sqrt{Q^2-4m_Z^2}\left(Q^2-2m_Z^2\right)\mathrm{Re}\left(h_1^H h_3^{H^\dagger}\right) + \left(Q^2-4m_Z^2\right)Q^2\mathrm{Im}\left(h_1^H h_2^{H^\dagger}\right)\right)\Bigg)$$
$$+ 4c_\phi m_Z^4\left(2m_Z^2-Q^2\right)\left|h_1^H\right|^2, \tag{25}$$

and

$$
\begin{aligned}
f_3\left(Q^2\right) = c_\phi \Bigg(& -\left(Q^2\left(Q^2-4m_Z^2\right)\right)^{3/2}\mathrm{Im}\left(h_3^H h_2^{H^\dagger}\right) \\
& + 2m_Z^2\left(\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\left(2m_Z^2-Q^2\right)\mathrm{Im}\left(h_3^H h_1^{H^\dagger}\right) + \left(4m_Z^2-Q^2\right)Q^2\mathrm{Re}\left(h_1^H h_2^{H^\dagger}\right)\right)\Bigg) \\
& - s_\phi\Bigg(\left(Q^2\left(Q^2-4m_Z^2\right)\right)^{3/2}\mathrm{Re}\left(h_2^H h_3^{H^\dagger}\right) \\
& + 2m_Z^2\left(\sqrt{Q^2-4m_Z^2}\left(Q^2-2m_Z^2\right)\mathrm{Re}\left(h_1^H h_3^{H^\dagger}\right) + \left(4m_Z^2-Q^2\right)Q^2\mathrm{Im}\left(h_1^H h_2^{H^\dagger}\right)\right)\Bigg) \\
& + 4c_\phi m_Z^4\left(2m_Z^2-4Q^2\right)\left|h_1^H\right|^2. \qquad (26)
\end{aligned}
$$

Although the terms dependent on the ϕ angle vanish after integration in the phase space, the study of observables related to ϕ can lead to interesting results, which will be addressed in a future work.

C. $H^* \to Z(p_1)Z(p_2) \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width

Following the discussion of Appendix A, the differential $H^* \to Z(p_1)Z(p_2) \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width can be written as

$$
\frac{d\Gamma_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}}{dp_1 dp_2 d\cos\theta_1 d\cos\theta_2 d\phi} = \frac{\sqrt{Q^2-4m_Z^2}}{512(2\pi)^6 Q^2}\mathcal{M}^2_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}. \qquad (27)
$$

Since we consider the scenario where both Z gauge bosons are on-shell, we use the narrow-width approximation for their Breit-Wigner propagators

$$
\lim_{m_Z\Gamma_Z\to 0}\frac{1}{(p_i^2-m_Z^2)+(m_Z\Gamma_Z)^2} = \delta(p_i^2-m_Z^2)\frac{\pi}{m_Z\Gamma_Z}, \quad (i=1,2), \qquad (28)
$$

which allows one to integrate over p_1 and p_2. Introducing the polarized and interference amplitudes, the differential decay width can be written as

$$
\begin{aligned}
\frac{d\Gamma_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}}{d\cos\theta_1 d\cos\theta_2 d\phi} = & \frac{\sqrt{Q^2-4m_Z^2}}{(32\sqrt{2})^2(2\pi)^4 Q^2(m_Z\Gamma_Z)^2}\Big(\sum_\lambda \mathcal{M}^2_{H^*\to Z(p_1)Z(p_2)}(\lambda,\lambda)\mathcal{M}^2_{Z(p_1)\to\bar{\ell}_1\ell_1}(\lambda)\mathcal{M}^2_{Z(p_2)\to\bar{\ell}_2\ell_2}(\lambda) \\
& + \mathcal{M}^2_{\mathrm{int}}\Big). \qquad (29)
\end{aligned}
$$

All our results presented below will be obtained from the above equation along with Eqs. (18)–(22), whereas the interference term (23) will be integrated out as we are mainly interested in the polarization effects.

Below we present the $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width in the scenario with polarized Z gauge bosons, though for completeness we also consider the case of unpolarized Z gauge bosons, which can be obtained via standard calculation techniques and serve to cross-check our results for polarized Z gauge bosons obtained from Eq. (29).

1. *Polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width*

To obtain the polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width we integrate Eq. (29) over the angular variables in the region given in Appendix A, which after some rearrangement yields

$$
\Gamma^\lambda_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2} = \frac{g^6\left(g_A^2+g_V^2\right)^2\sqrt{Q^2-4m_Z^2}}{(48\sqrt{2})^2(2\pi)^3 Q^2 c_W^6 m_Z^4\Gamma_Z^2}F^\lambda\left(Q^2\right), \qquad (30)
$$

with

$$F^{\lambda}\left(Q^2\right) = \left(\left(Q^2-2m_Z^2\right)^2\left(4m_Z^4\left|h_1^H\right|^2+Q^4\left|h_2^H\right|^2\right)+4Q^2m_Z^2\left(Q^4-6Q^2m_Z^2+8m_Z^4\right)\mathrm{Re}\left(h_1^H h_2^{H^\dagger}\right)\right)f_0$$
$$+4m_Z^4\left(4m_Z^2\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\mathrm{Im}\left(h_3^H h_1^{H^\dagger}\right)+Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2+4m_Z^4\left|h_1^H\right|^2\right)f_L$$
$$+4m_Z^2\left(-4m_Z^2\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\mathrm{Im}\left(h_3^H h_1^{H^\dagger}\right)+Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2+4m_Z^4\left|h_1^H\right|^2\right)f_R, \tag{31}$$

where the f_L, f_R and f_0 coefficients must be set to 0 or 1 according to the polarization λ of the Z gauge bosons. For instance, $f_L = 1$ and $f_R = f_0 = 0$ correspond to $\lambda = L$. As already pointed out, the interference term in Eq. (29) vanishes after ϕ integration, thus it is possible to discriminate between the distinct contributions of polarized Z gauge bosons, which are functions of the four-lepton invariant mass only.

The polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width can also be written in terms of the $Z(p_i) \to \bar{\ell}_i\ell_i$ branching fractions as follows

$$\Gamma^{\lambda}_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2} = 2\,\Gamma^{\lambda}_{H^*\to ZZ}\left(Q^2\right)\mathrm{Br}(Z\to\ell_1\bar{\ell}_1)\mathrm{Br}(Z\to\ell_2\bar{\ell}_2) = 0.00226418\times\Gamma_{H^*\to Z_\lambda Z_\lambda}\left(Q^2\right), \tag{32}$$

where an extra factor of 2 was included in Eq. (32), which stems from the fact that the Z bosons are not in the final state, and hence we have not considered the statical factor of 1/2 that corrects the double-counting of identical particles in the $H^* \to ZZ$ amplitude. This factor has been overlooked in the past [4]. In. Eq. (32), the polarized $H^* \to Z_\lambda Z_\lambda$ decay width is given by

$$\Gamma^{\lambda}_{H^*\to Z_\lambda Z_\lambda} = \frac{g^2\sqrt{Q^2-4m_Z^2}}{128\pi Q^2 c_W^2 m_Z^6}F^{\lambda}\left(Q^2\right). \tag{33}$$

and we have used

$$\mathrm{Br}(Z\to\ell_i\bar{\ell}_i) \simeq \frac{g^2 m_Z}{12\pi c_W^2\Gamma_Z}\left(g_A^2+g_V^2\right) \simeq 0.0336466, \tag{34}$$

in the massless lepton approximation [44].

From (32) it is clear that the four-lepton final state is sensitive to the polarization effects on the HZZ vertex: the polarization of the Z gauge bosons determine the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width. Nevertheless, for the four-lepton invariant mass distributions is only necessary to know the behavior of the $H^* \to ZZ$ process to describe the physics of the leptons in the final state. The left-right asymmetry in Z gauge boson pair production [5] has also effects on the decay of an off-shell Higgs boson into four-leptons. Such an asymmetry is a consequence of CP violation due to complex anomalous couplings. Hence, new physics effects may be detected via transversally polarized Z gauge bosons [45]. Moreover, they would provide an excellent probe of quantum field theory, which predict that the h_i^V (i=1, 2, 3) form factors are complex. On the other hand, longitudinally polarized Z gauge bosons are not useful to detect CP-violating effects, though it is still possible to detect new-physics contributions through the $\hat{c}_Z$ anomalous coupling, which enters into the h_1^H form factor.

2. *Unpolarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width*

For the sake of completeness we also consider unpolarized Z gauge bosons. From Eq. (30), we can obtain the unpolarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width by setting $f_L = f_R = f_0 = 1$, which yields

$$\Gamma_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2}\left(Q^2\right) = \frac{g^6\left(g_A^2+g_V^2\right)^2\sqrt{Q^2-4m_Z^2}}{(48\sqrt{2})^2(2\pi)^3 Q^2 c_W^6 m_Z^4\Gamma_Z^2}G\left(Q^2\right), \tag{35}$$

where

$$G^{\lambda}\left(Q^2\right) = Q^2\left(Q^2-4m_Z^2\right)\left(Q^2\left(Q^2-4m_Z^2\right)\left|h_2^H\right|^2+8m_Z^4\left|h_3^H\right|^2\right)+4m_Z^4\left(Q^4-4Q^2m_Z^2+12m_Z^4\right)\left|h_1^H\right|^2$$
$$+4Q^2m_Z^2\left(Q^4-6Q^2m_Z^2+8m_Z^4\right)\mathrm{Re}\left(h_1^H h_2^{H^\dagger}\right). \tag{36}$$

To cross-check this result we have inserted the left-hand side of Eq. (12) into Eq. (8) and used standard calculation techniques. In this case, the boost defined in Appendix A is necessary as the square amplitude depends on scalar products of four-momentum given in their respective $Z(p_i)$ gauge boson rest frame.

From Eq. (35) we can observe that the contribution of the CP-violating form factor h_3^H to the four-lepton final state is suppressed as compared to the CP-conserving contributions, which are proportional to higher powers of the four-lepton invariant mass Q. Thus the $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ unpolarized decay width is not really sensitive to CP-violating effects. We also note that one can extract from Eq. (35) the tree-level SM contribution, which is obtained for $h_2^H = h_3^H = 0$ and $h_1^H = 1$. As far as the one-loop level SM contribution is concerned, it is obtained by setting $h_3^H = 0$, whereas the remaining form factors are given in terms of $\hat{b}_Z$ only since $\hat{c}_Z$ vanishes at this order. The one-loop level contribution to $\hat{b}_Z$ was reported in [5].

Eq. (35) can be written in the following short form

$$\Gamma_{H^*\to\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2} = 2\,\Gamma_{H^*\to ZZ}\left(Q^2\right)\mathrm{Br}(Z\to\ell_1\bar{\ell}_1)\mathrm{Br}(Z\to\ell_2\bar{\ell}_2) = 0.00226418\times\Gamma_{H^*\to ZZ}\left(Q^2\right), \tag{37}$$

where the unpolarized $H^* \to ZZ$ decay width for complex anomalous couplings is given by

$$\Gamma_{H^*\to ZZ}\left(Q^2\right) = \frac{g^2\sqrt{Q^2-4m_Z^2}}{512\pi Q^2 c_W^2 m_Z^6} G\left(Q^2\right). \tag{38}$$

The inclusion of a factor of 2 in Eq. 37 has already been explained when obtaining the polarized decay width.

III. NUMERICAL ANALYSIS

We now turn to the study of the polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width and a few observables sensitive to new physics effects. We first present a cross-check of our numerical evaluation method, which relies on the analytical results presented in Sec. II along with the kinematics and four-body phase space discussed in Appendix A.

A. Consistency of our numerical evaluation method

Using our analytical results, we have implemented a Mathematica code to perform a numerical evaluation, which is available in our GitLab repository [46]. For an alternative numerical evaluation, we have implemented the Lagrangian of Eq. (1) into Madgraph5 [47] via the UFO format [48] using the FeynRules package [49]. We then compare the numerical results obtained via these two independent methods. We consider the $e^-e^+\mu^-\mu^+$ final state in our analysis, which, however, is indistinguishable from other $\bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ final states as the massless lepton approximation is used.

Since the anomalous HZZ couplings have negligible effects on the unpolarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width [5], we set $h_1^H = 1$ and $h_2^H = h_3^H = 0$ and calculate the tree-level SM contribution to the unpolarized $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ and $H^* \to ZZ$ decays. The results obtained by our own numerical evaluation method and Madgraph are shown in Fig. 3 as functions of the four-lepton invariant mass. We observe an excellent agreement between both results, with a slight variation of the order of 2 % of the decay widths. We have also verified that the $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ decay width obtained via Eq. (37) agree with that obtained from Eq. (35) and shown in the left plot of Fig. 3, which in turn shows the consistency of the extra factor of 2 inserted into Eq. (37).

We now turn to cross-check the numerical results for the polarized $H^* \to ZZ$ decay width, which is sensitive to the anomalous HZZ couplings and yields the unpolarized $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ decay via Eq. (32). For our numerical evaluation method we use the analytical result of Eq. (33) for the polarized partial decay widths $\Gamma_{H^*\to Z_\lambda Z_\lambda}$. We show in the left (right) plots of Fig. 4 the behavior of the $H^* \to Z_L Z_L$ ($H^* \to Z_R Z_R$) decay width as a function of the four-lepton invariant mass for two sets of values of the anomalous HZZ couplings consistent with the most stringent bounds [5]. For comparison purposes we also show the tree-level SM contributions.

Again, we can conclude that the results obtained by our numerical evaluation method and Madgraph agree nicely. We observe that the anomalous contributions yield a significant deviation from the tree-level SM contribution for large Q, and also note that the $H^* \to Z_\lambda Z_\lambda$ decay width shows a distinctive behavior for each polarizations in both scenarios of new physics. As discussed below, this stems from the CP-violating effects due to the complex $\hat{b}_Z$ form factor, which have already been studied through a left-right asymmetry in the decay $H^* \to Z_\lambda Z_\lambda$ [5]. A more detailed analysis of the polarized $H^* \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay widths, including the SM contribution up to one-loop level, together with the analysis of other interesting observables sensitive to new physics will be presented below.

In summary, there is an excellent agreement between the results obtained by our evaluation method and Madgraph. We also corroborate the consistency of Eqs. (37) and (32), which allows one to obtain the four-lepton decay widths by means of the $H^* \to ZZ$ decay width only. Note that in our evaluation method we use a constant Γ_Z decay width in the Breit-Wigner propagator, whereas in Madgraph the complex mass scheme is used. According to Ref. [50], both methods yield the same results. In the following, we will use our evaluation method as it is more suited for our analysis.

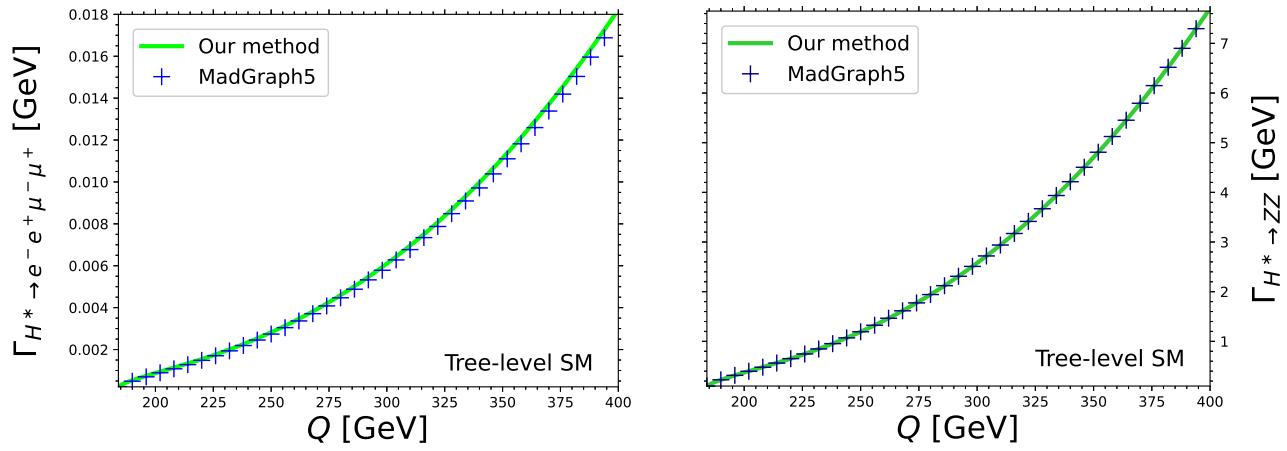


FIG. 3. Tree-level SM contribution to the unpolarized $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ and $H^* \to ZZ$ decay widths as functions of the four-lepton invariant mass Q obtained by both our own evaluation method and Madgraph.

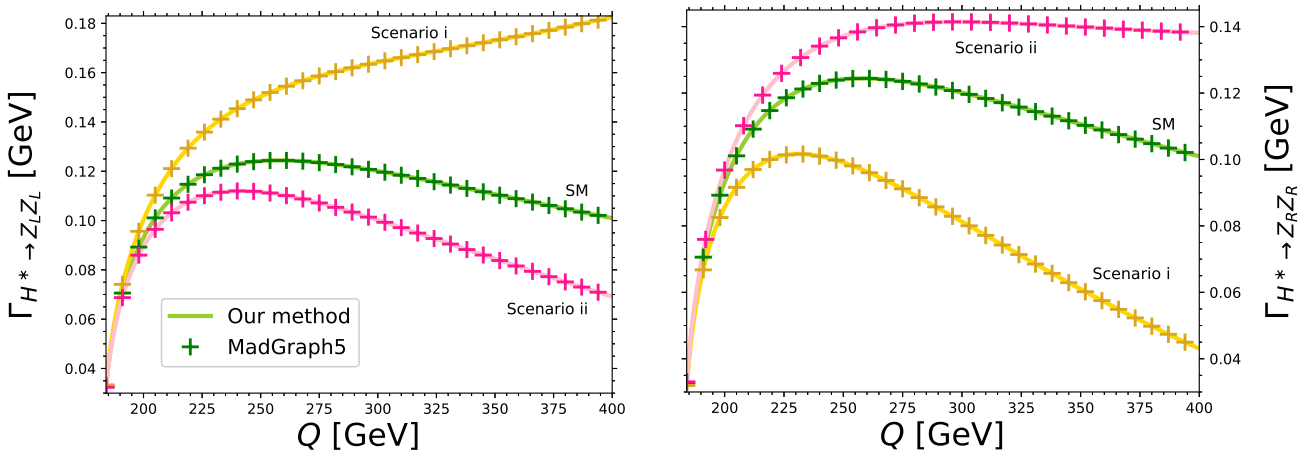


FIG. 4. Behavior of the polarized $H^* \to Z_L Z_L$ (left plot) and $H^* \to Z_R Z_R$ (right plot) decay widths as functions of the four-lepton invariant mass for two sets of values of the anomalous couplings: scenario i)$\hat{b}_Z = 0.001 + i0.003$, $\hat{c}_Z = 0.001 + i0.002$, and $\widetilde{b}_Z = 0.01 + i0.02$; scenario ii)$\hat{b}_Z = 0.0001 + i0.001$, $\hat{c}_Z = 0.0001 + i0.001$, and $\widetilde{b}_Z = 0.001 - i0.01$. We also show the SM tree-level contributions for comparison purposes. For the numerical evaluation we use our own evaluation method and Madgraph.

B. Polarized decay widths

We now consider the most general case with complex HZZ anomalous couplings and consider three realistic scenarios for their numerical values consistent with the most stringent current constraints, which are shown in Table I. Note that although we consider a non-vanishing $\hat{c}_Z$, it yields a negligible contribution to the h_1^H coupling, and hence we will consider the same values for the three scenarios. For the $\hat{b}_Z$ anomalous coupling we will include the SM contribution up to one-loop level, which was reported in Ref. [5] in terms of the Passarino-Veltman scalar functions and is function of Q the four-lepton invariant mass. The numerical evaluation of $\hat{b}_Z$ is obtained through LoopTools package [51]. The one-loop SM contribution is also studied to compare with the new physics scenarios.

We first analyze the effects of the Z gauge bosons polarizations on the $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ decay, for which we can use either Eq. (30) or Eq. (32) as they yield the same results. We show in Fig. 5, the results for the left- and right-handed polarized $H^* \to e^-e^+\mu^-\mu^+$ decay widths as functions of Q. We do not consider longitudinally

TABLE I. Scenarios for the HZZ anomalous couplings used in our analysis of the $H^* \to e^-e^+\mu^-\mu^+$ decay. For the $\hat{b}_Z$ coupling we considere SM contribution up to one-loop level [5].

Scenario	$\widetilde{b}_Z$	$\hat{c}_Z$
1	$0.01 + i0.02$	$0.0001 + 0.0003i$
2	$0.001 - i0.01$	$0.0001 + 0.0003i$
3	$0.0001 + i0.001$	$0.0001 + 0.0003i$

polarized Z gauge bosons as in this case the $H^* \to ZZ \to e^-e^+\mu^-\mu^+$ decay width is not sensitive to the CP-violating anomalous coupling $\tilde{b}_Z$.

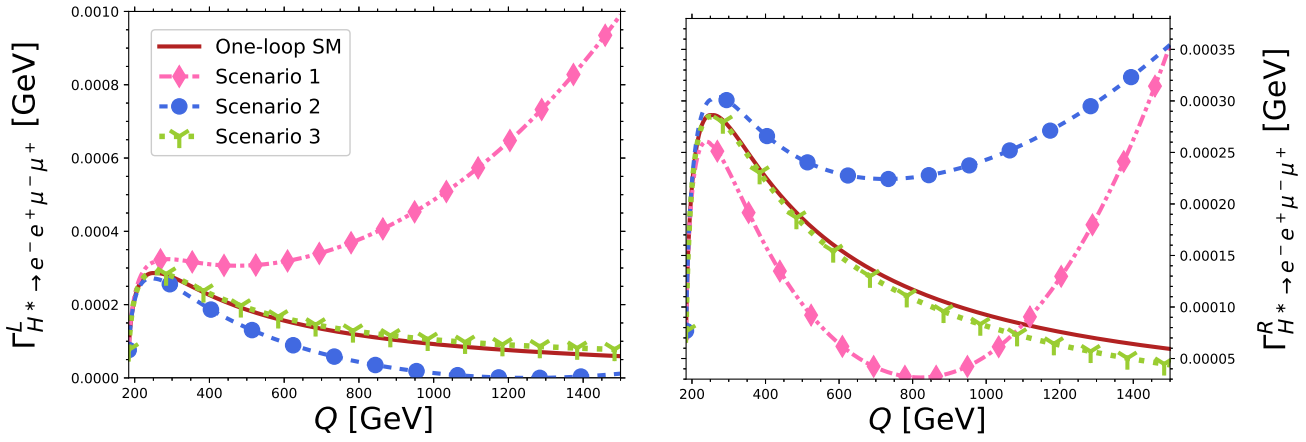


FIG. 5. Behavior of the transversally polarized $\Gamma^{\lambda}_{H^*\to e^-e^+\mu^-\mu^+}$ decay widths ($\lambda = L,\, R$) for the SM contribution and the three scenarios of Table I for the anomalous HZZ couplings.

We observe in Fig. 5 that for small values of the real and absorptive parts of the CP-violating anomalous coupling $\tilde{b}_Z$ (scenario 3), the polarized $\Gamma^{\lambda}_{H^*\to e^-e^+\mu^-\mu^+}$ decay widths shows little variation from the pure SM contribution, though such a deviation increases slightly as Q increases, being more appreciable in the case of right-handed polarization. On the other hand, in scenarios 1 and 2, in which $\tilde{b}_Z$ is assumed to be relatively large, a considerable deviation from the SM contribution develops around $Q = 250$ GeV, and becomes larger as Q increases. Therefore, the effects of the complex anomalous HZZ couplings in the four-lepton final state would be larger at high energies.

C. Left-right asymmetry $\mathcal{A}_{LR}$

The behavior of the $H^* \to e^-e^+\mu^-\mu^+$ polarized amplitudes give rise to a left-right asymmetry $\mathcal{A}_{LR}$

$$\mathcal{A}_{LR} = \frac{\Gamma^L_{H^*\to e^-e^+\mu^-\mu^+} - \Gamma^R_{H^*\to e^-e^+\mu^-\mu^+}}{\Gamma^L_{H^*\to e^-e^+\mu^-\mu^+} + \Gamma^R_{H^*\to e^-e^+\mu^-\mu^+}}, \tag{39}$$

which is non-vanishing as long as complex anomalous couplings are present and can be written as

$$\mathcal{A}_{LR} = \frac{4m_Z^2\sqrt{Q^2\left(Q^2-4m_Z^2\right)}\mathrm{Im}\left(h_3^H h_1^{H\dagger}\right)}{Q^2\left(Q^2-4m_Z^2\right)\left|h_3^H\right|^2 + 4m_Z^4\left|h_1^H\right|^2}, \tag{40}$$

thereby being suitable to detect CP-violating effects.

In Fig. 6 we show the behavior of $\mathcal{A}_{LR}$ as a function of the four-lepton invariant mass in the three scenarios of Table I for the anomalous HZZ couplings. The SM contribution is not shown since it vanishes up to the one-loop

level. We observe that the $\mathcal{A}_{LR}$ magnitude can be larger in scenario 2, in which the real and absorptive parts of h_3^H are of opposite sign. Also, in the three scenarios, $\mathcal{A}_{LR}$ reaches its largest magnitude at a relatively large Q and then tends to decrease, which stems from the fact that it behaves as $1/Q^2$ at large Q. Thus, CP-violating effects would be more significant beyond the $2m_Z$ threshold.

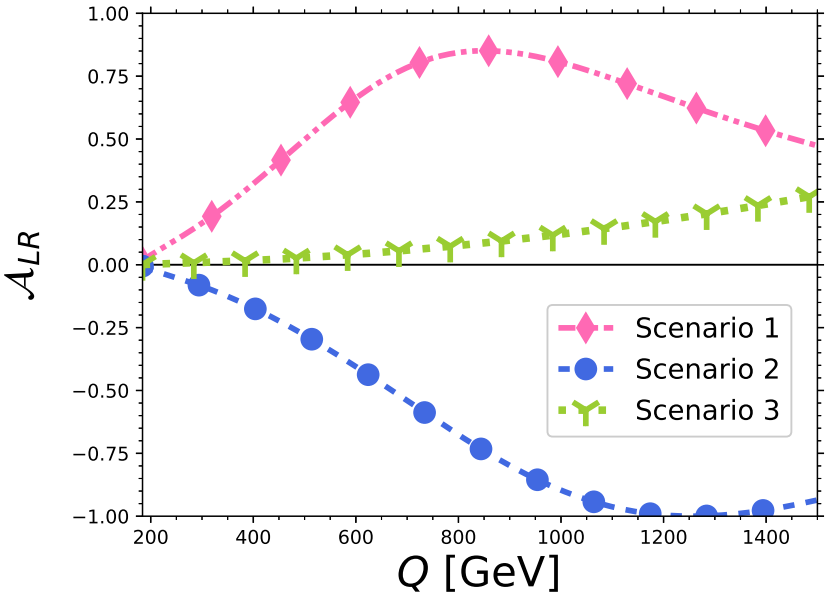


FIG. 6. Left-right asymmetry $\mathcal{A}_{LR}$ as a function of the four-lepton invariant mass in the three scenarios of Table I for the anomalous HZZ couplings.

The results of Figs. 5 and 6, together with Eq. (40), are in agreement with those reported in [5] for the $H^* \to Z_\lambda Z_\lambda$ process, which is a consequence of the four-leptons final state being a function of Q and independent of the $Z \to \bar{\ell}\ell$ processes (see Eq. 32). On the other hand, the angular variables only appear in the process of production of lepton pairs and could also serve to detect any new physics effects contributing to the HZZ anomalous couplings.

D. Angular distributions

We now turn to analyze the role of the angular variables in the polarized $H^* \to e^-e^+\mu^-\mu^+$ decay widths. Since the angle ϕ only enters into the interference terms of the full square amplitude, its effects are unobservable through polarized Z gauge bosons. After integrating Eq. (29) over ϕ, the differential $H^* \to e^-e^+\mu^-\mu^+$ decay width is given in terms of the angles $\theta_{1,2}$, as shown in Eq. (30). Therefore, it is still possible to study the polarization effects in the four lepton-final state through the angular variables. Furthermore, from Eq. (21) we observe that the left- and right-handed square amplitudes are symmetric in both θ_1 and θ_2 angles, which is evident in Fig. 7, where we show the contours of the SM contribution to the transversally polarized $H^* \to e^-e^+\mu^-\mu^+$ differential decay widths in the c_{θ_1} vs c_{θ_2} plane for $Q = 500$ GeV. Although both the left- and right-handed polarized $H^* \to e^-e^+\mu^-\mu^+$ differential decay widths are in general of the order of 10^{-5}, the former increases almost one order of magnitude around $\cos\theta_i = -1$, whereas the latter is larger for $\cos\theta_i = 1$. Furthermore, it is found that the transversally polarized differential $H^* \to e^-e^+\mu^-\mu^+$ decay widths behave similarly for other values of Q and if non-vanishing anomalous couplings are considered.

We now consider the three scenarios of Table I for the anomalous HZZ couplings and study their effects on the polarized $H^* \to e^-e^+\mu^-\mu^+$ angular distributions after integrating over ϕ and one of the angles $\theta_{1,2}$. The results are presented in Figs. 8 and 9 as functions of the cosine of the remaining angle c_{θ_i}. We observe that the angular distributions show a significant deviation from the SM contribution in scenarios 1 and 2, which becomes more pronounced at large Q. However, in scenario 3, where small values for the real and absorptive parts of the anomalous h_3^H coupling are considered, the deviation is insignificant for small Q, though it becomes noticeable as Q becomes

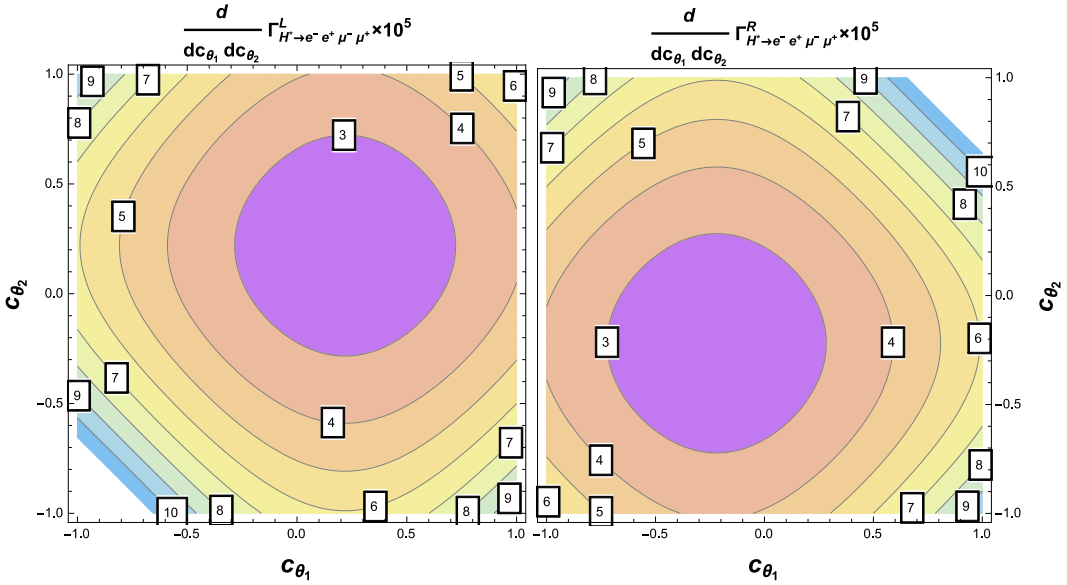


FIG. 7. Contours of the transversally polarized $H^* \to e^-e^+\mu^-\mu^+$ differential decay widths in the c_{θ_1} vs c_{θ_2} plane for $Q = 500$ GeV. We only include the SM one-loop contribution.

large, particularly in the case of right-handed polarization. We also note that, for left-handed polarization, the differential decay width is always above (below) the SM contribution in scenario 1 (scenario 2), whereas the opposite is true for right-handed polarization. Also, the left-handed (right-handed) polarized angular distributions reaches its larger magnitude as $c_{\theta_i} \to -1$ ($c_{\theta_i} \to 1$). This distinctive behavior of the angular distributions for each transverse polarization of the Z gauge bosons stems from the fact that the terms that give rise to the left-right asymmetry $\mathcal{A}_{LR}$ of Eq. (40) and the terms proportional to c_{θ_i} in Eq. (21), which have opposite signs, remain unchanged after ϕ integration. This hints to new asymmetries associated with the angular variables, which we will analyze below. We also would like to point out that the results shown in Figs. 8 and 9 for the SM contributions agree with those reported in Ref. [31].

1. *Polarized asymmetries*

Motivated by the behavior of the $H^* \to e^-e^+\mu^-\mu^+$ angular distributions for transverse polarizations, we define the following angular left-right asymmetry

$$\mathcal{A}_{LR\theta} = \frac{\frac{d}{dc_{\theta_i}}\Gamma^{L}_{H^*\to e^-e^+\mu^-\mu^+} - \frac{d}{dc_{\theta_i}}\Gamma^{R}_{H^*\to e^-e^+\mu^-\mu^+}}{\frac{d}{dc_{\theta_i}}\Gamma^{L}_{H^*\to e^-e^+\mu^-\mu^+} + \frac{d}{dc_{\theta_i}}\Gamma^{R}_{H^*\to e^-e^+\mu^-\mu^+}}, \tag{41}$$

where $i = 1$ (2) and the angle θ_2 (θ_1) was integrated over. Since $\Gamma_{H^*\to e^-e^+\mu^-\mu^+}$ is symmetric under θ_i ($i = 1, 2$), the same expression holds for both angles. The analytical form, obtained through Eq. (29), can be written as

$$\mathcal{A}_{LR\theta} = -\frac{4}{m_Z^2}\frac{f(Q^2, c_{\theta_i})}{h(Q^2, c_{\theta_i})}, \tag{42}$$

with the $f(Q^2, c_{\theta_i})$ and $h(Q^2, c_{\theta_i})$ functions being given as

$$\begin{aligned} f(Q^2, c_{\theta_i}) = {} & g_A g_V c_{\theta_i} m_Z^2 \left(Q^2\left(Q^2 - 4m_Z^2\right)\left|h_3^H\right|^2 + 4m_Z^4\left|h_1^H\right|^2\right) \\ & - \left(g_A^2 + g_V^2\right)\left(1 + c_{\theta_i}^2\right)\left(m_Z^4\sqrt{Q^2\left(Q^2 - 4m_Z^2\right)}\mathrm{Im}\left(h_3^H h_1^{H\dagger}\right)\right), \end{aligned} \tag{43}$$

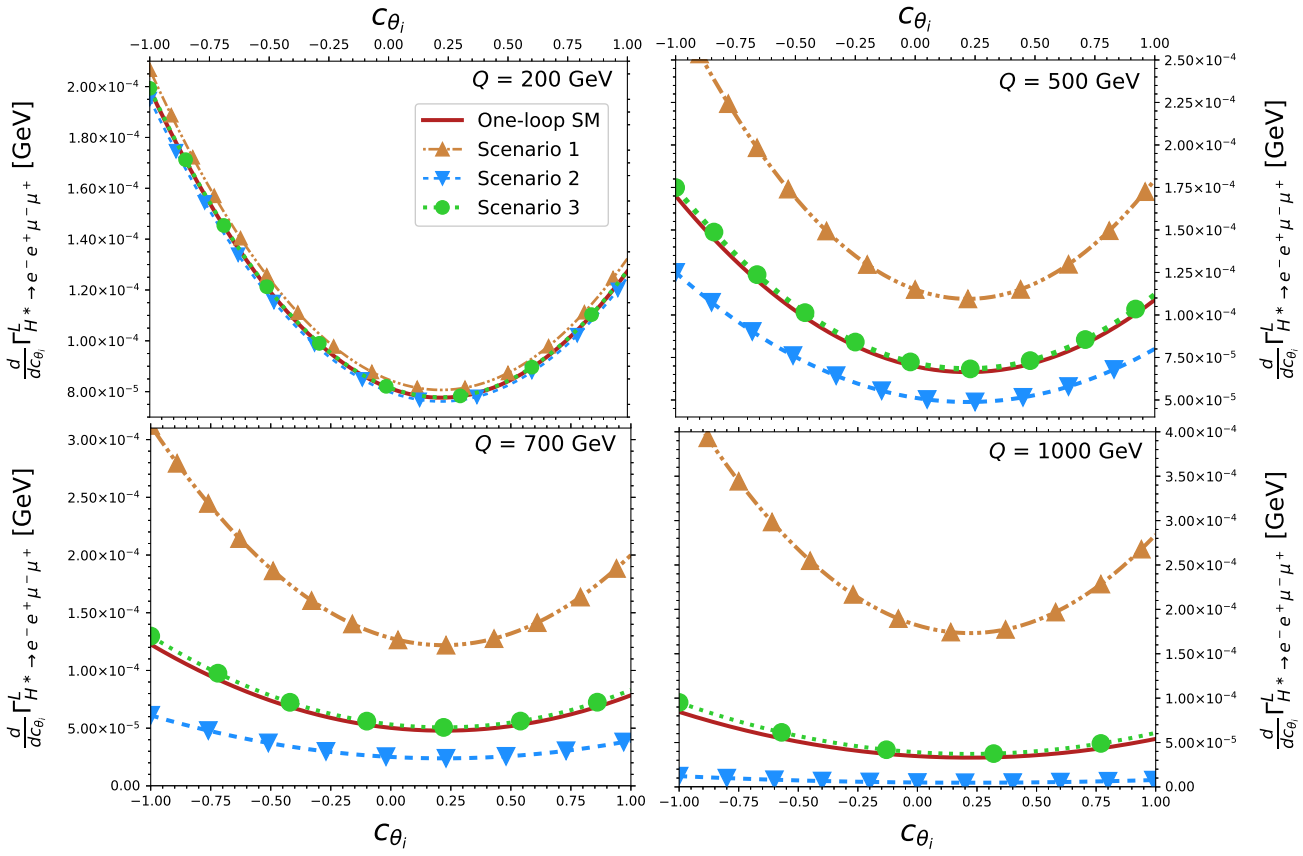


FIG. 8. Differential $H^* \to e^- e^+ \mu^- \mu^+$ decay width for left-handed polarization as a function of $\cos\theta_i$ $(i = 1, 2)$ for a few values of the four-lepton invariant mass Q in the three scenarios of Table I for the anomalous HZZ couplings. The SM contribution up to the one-loop level was included. We already integrated over one of the two angles.

$$h(Q^2, c_{\theta_i}) = -16 g_A g_V c_{\theta_i} m_Z^2 \sqrt{Q^2\left(Q^2 - 4m_Z^2\right)} \mathrm{Im}\left(h_3^H h_1^{H\dagger}\right) + \left(g_A^2 + g_V^2\right)\left(1 + c_{\theta_i}^2\right)\left(Q^2\left(Q^2 - 4m_Z^2\right)\left|h_3^H\right|^2 + 4m_Z^4\left|h_1^H\right|^2\right), \tag{44}$$

where the terms proportional to $g_{V,A}^2$ in $f\left(Q^2, c_{\theta_i}\right)$ are identical to those in the $\mathcal{A}_{LR}$ asymmetry (40), whereas the remaining terms arise from those proportional to c_{θ_i} in the $H^* \to e^- e^+ \mu^- \mu^+$ amplitude, which have been not integrated out yet. Furthermore, if the numerator and denominator Eq. (42) are integrated over c_{θ_i} we obtain the non-angular $\mathcal{A}_{LR}$ asymmetry. Although the $\mathcal{A}_{LR}$ and $\mathcal{A}_{LR\theta}$ asymmetries are related, they have a distinct origin: the former is induced via the HZZ anomalous couplings and the latter is a result of polarized Z gauge bosons decaying into leptons, where the angular variables do appear. It is also worth noting that even for a vanishing CP-violating form factor h_3^H, the $\mathcal{A}_{LR\theta}$ asymmetry is non-vanishing, thereby being non-vanishing in the SM at the one-loop level:

$$\mathcal{A}_{LR\theta}^{\mathrm{SM}} = -\frac{4 g_A g_V c_{\theta_i}}{\left(g_A^2 + g_V^2\right)\left(1 + c_{\theta_i}^2\right)}. \tag{45}$$

Again we consider the three scenarios of Table I for the anomalous HZZ couplings and evaluate the effects on the $\mathcal{A}_{LR\theta}$ asymmetry. The results are shown in Fig. 10, where we also include the SM contribution. We observe that at low Q, $\mathcal{A}_{LR\theta}$ shows a slight deviation from the SM in scenarios 1 and 2, which becomes more significant as Q increases, whereas such a deviation is negligible in scenario 3. It is interesting to note that the deviation of $\mathcal{A}_{LR\theta}$ from the SM value is of the order of $10^{-1} - 10^{-2}$ at $Q = 1000$ GeV, whereas the corresponding deviation of the polarized $H^* \to e^- e^+ \mu^- \mu^+$ decay widths from the SM contribution is of the order of $10^{-4} - 10^{-5}$. Thus, $\mathcal{A}_{LR\theta}$ is more sensitive to CP violation than the polarized differential widths.

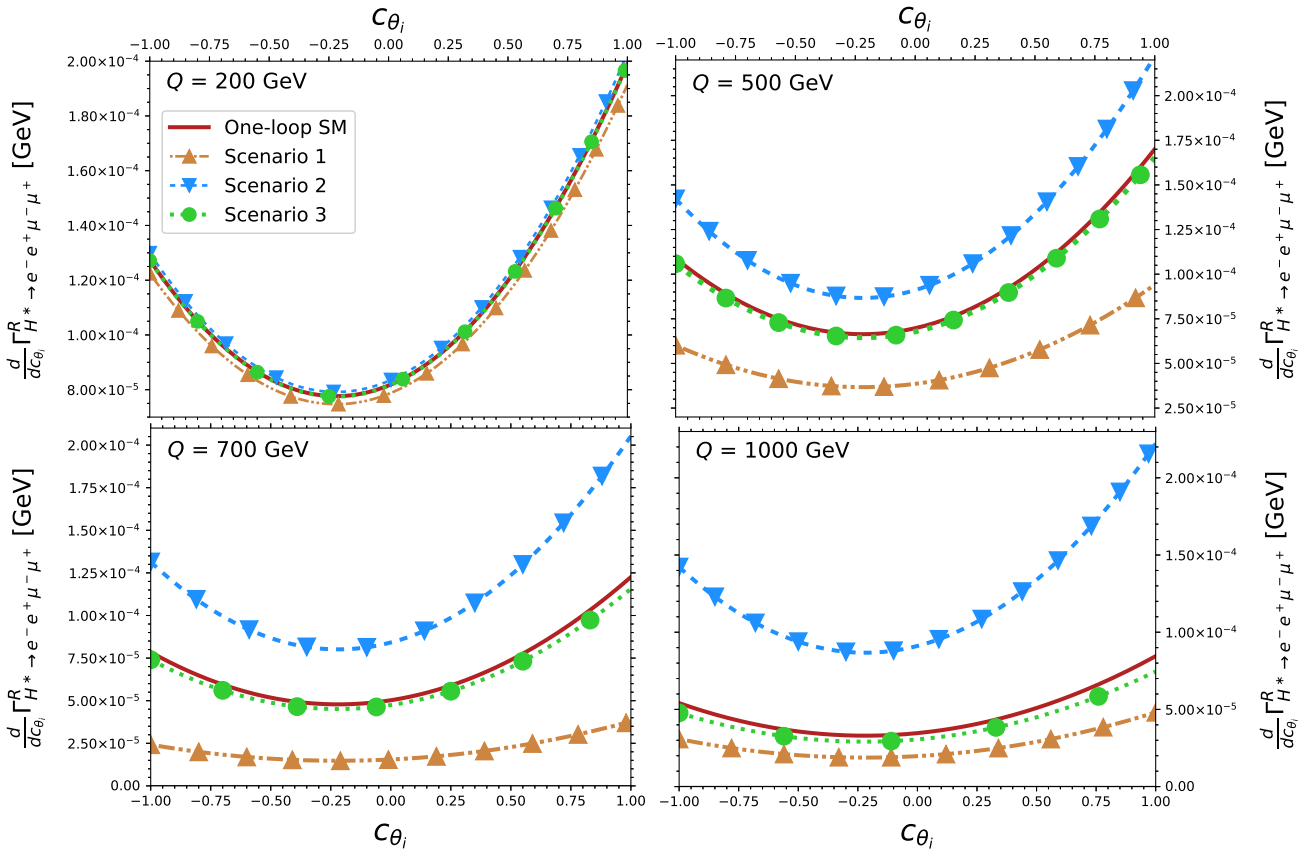


FIG. 9. The same as in Fig. 8, but for right-handed polarizations.

Given the behavior observed in the polarized differential $H^* \to e^-e^+\mu^-\mu^+$ decay width, we also expect the presence of a polarized forward-backward asymmetry, which in general can be defined as follows for the polarized $H^* \to \overline{f}_i f_i \overline{f}_j f_j$ decay

$$\mathcal{A}^{\lambda}_{FB} = \frac{\displaystyle\int_0^1 dc_{\theta_i} \frac{d}{dc_{\theta_i}}\Gamma^{\lambda}_{H^*\to\overline{f}_i f_i \overline{f}_j f_j} - \int_{-1}^0 dc_{\theta_i} \frac{d}{dc_{\theta_i}}\Gamma^{\lambda}_{H^*\to\overline{f}_i f_i \overline{f}_j f_j}}{\displaystyle\int_0^1 dc_{\theta_i} \frac{d}{dc_{\theta_i}}\Gamma^{\lambda}_{H^*\to\overline{f}_i f_i \overline{f}_j f_j} + \int_{-1}^0 dc_{\theta_i} \frac{d}{dc_{\theta_i}}\Gamma^{\lambda}_{H^*\to\overline{f}_i f_i \overline{f}_j f_j}}. \tag{46}$$

A straightforward calculation yields

$$\mathcal{A}^{L/R}_{FB} = \mp \frac{3 g_A g_V}{2\left(g_A^2 + g_V^2\right)}. \tag{47}$$

Such a simple expression stems from the fact that $\mathcal{A}^{L/R}_{FB}$ is due to the $Z \to \overline{\ell}\ell$ decay, thereby being independent the anomalous HZZ couplings and the four-lepton invariant mass: it is constant even in new-physics scenarios but a difference of sign arises for left-handed and right-handed polarizations.

We present in Table II the values of the $\mathcal{A}^{\lambda}_{FB}$ asymmetry for all the combinations of light fermion flavors in the $H^* \to \overline{f}_i f_i \overline{f}_j f_j$ final state. We consider massless fermions and distinct fermion-antifermion pairs. We also note that the result of Eq. (47) is similar to the left-right Z gauge boson asymmetry in the SM [52, 53].

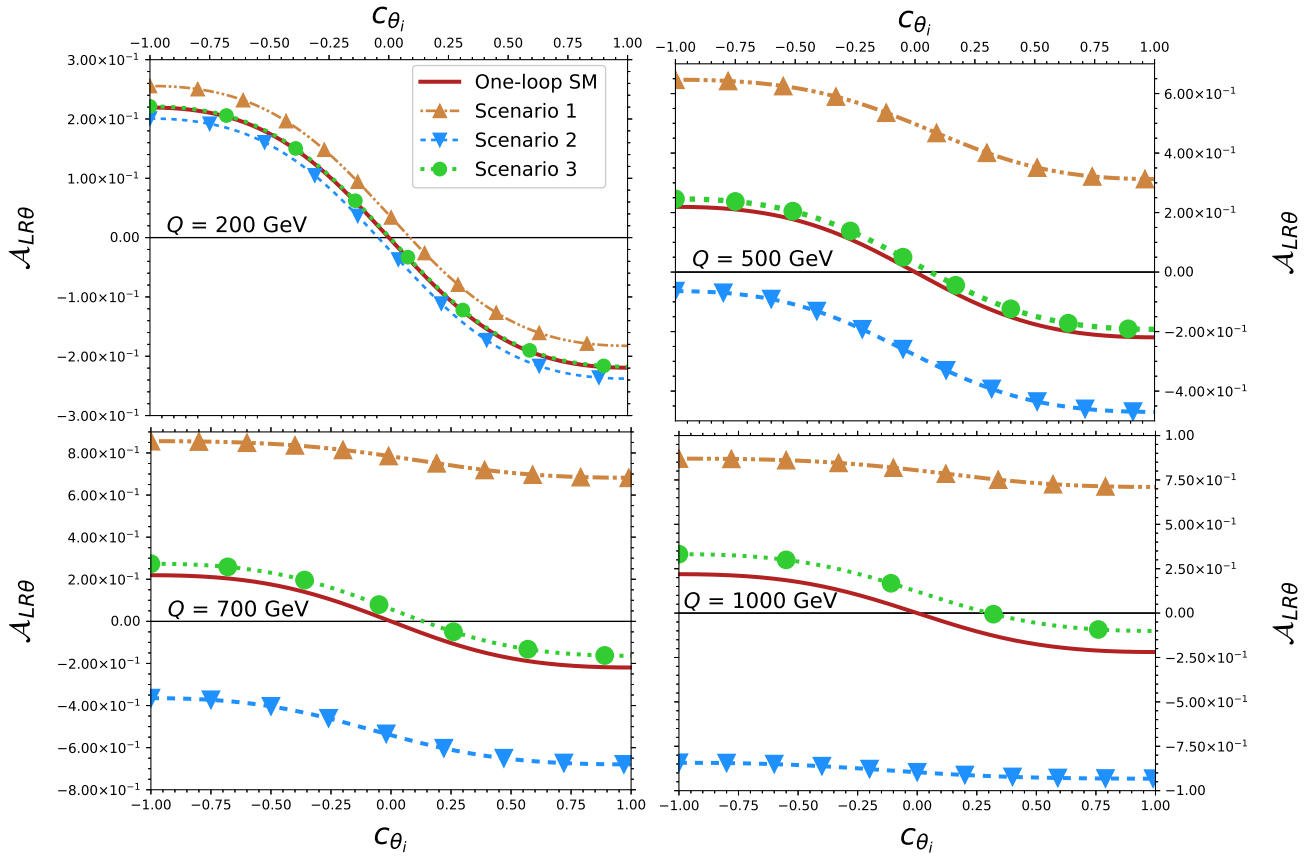


FIG. 10. Behavior of the left-right asymmetry $\mathcal{A}_{LR\theta}$ as a function $\cos\theta_i$ for a few values of the four-lepton invariant mass Q in the three scenarios of Table I for the anomalous HZZ couplings. We also include the SM contribution up to the one-loop level.

TABLE II. $\mathcal{A}_{FB}^{L/R}$ asymmetry for all combinations of distinct pairs of light fermion-antifermion flavors in the $H^* \to \overline{f}_i f_i \overline{f}_j f_j$ decay. The fermion masses have been neglected.

f_i	f_j	$\mathcal{A}_{FB}^{L/R}$
e, μ, τ	e, μ, τ	∓ 0.164
ν_e, ν_μ, ν_τ	ν_e, ν_μ, ν_τ	∓ 0.75
d, s, b	d, s, b	∓ 0.705
u, c	u, c	∓ 0.524

2. *Unpolarized asymmetries*

Up to now we have focused only on the effects of the Z gauge boson polarizations in the $H^* \to ZZ \to \overline{\ell}_1\ell_1\overline{\ell}_2\ell_2$ decay. Nevertheless, the change of sign in Eq. (47) hints the presence of a forward-backward asymmetry $\mathcal{A}_{FB}$ in the

unpolarized case, which is defined as follows

$$\mathcal{A}_{FB} = \frac{\displaystyle\int_0^1 dc_{\theta_i} \frac{d}{dc_{\theta_i}} \Gamma_{H^* \to ZZ \to \bar{\ell}_1 \ell_1 \bar{\ell}_2 \ell_2} - \int_{-1}^0 dc_{\theta_i} \frac{d}{dc_{\theta_i}} \Gamma_{H^* \to ZZ \to \bar{\ell}_1 \ell_1 \bar{\ell}_2 \ell_2}}{\displaystyle\int_0^1 dc_{\theta_i} \frac{d}{dc_{\theta_i}} \Gamma_{H^* \to ZZ \to \bar{\ell}_1 \ell_1 \bar{\ell}_2 \ell_2} + \int_{-1}^0 dc_{\theta_i} \frac{d}{dc_{\theta_i}} \Gamma_{H^* \to ZZ \to \bar{\ell}_1 \ell_1 \bar{\ell}_2 \ell_2}}. \tag{48}$$

After inserting Eq. (29) we obtain

$$\mathcal{A}_{FB} = -\frac{48 g_A g_V m_Z^6 Q \sqrt{Q^2 - 4m_Z^2}}{(g_A^2 + g_V^2)} \frac{\mathrm{Im}\left(h_3^H h_1^{H^\dagger}\right)}{G\left(Q^2\right)}, \tag{49}$$

where the $G\left(Q^2\right)$ function was given in (36). We note that $\mathcal{A}_{FB}$ vanishes in the SM up to the one-loop level since $h_3^H = 0$. Our result reproduces the one reported in Ref. [12], where the absorptive part of h_1^H was dismissed. However, such a term is induced at the one-loop level in the SM and its contribution, provided that a non-vanishing new-physics contribution to h_3^H is present, could be larger than that arising from the corresponding real part [5]. Thus, our result is more complete than those previously reported in the literature.

In contrast with the polarized forward-backward asymmetry, which is due only to the $Z \to \bar{\ell}_i \ell_i$ decay, the unpolarized one depends on the properties of both the $H^* \to ZZ$ and the $Z \to \bar{\ell}_i \ell_i$ decays: the dependence on the complex anomalous couplings is a reflect of the former and the presence of the vector g_V and axial g_A couplings arises from the latter.

We now show in Fig. 11 the unpolarized forward-backward asymmetry $\mathcal{A}_{FB}$ as a function of the four-lepton invariant mass in the three scenarios of Table I for the anomalous HZZ couplings. In addition to the anomalous contributions, we have included the one-loop SM contribution to the real and absorptive parts of the h_1^H form factor. Similar to the above results for other observables, $\mathcal{A}_{FB}$ can reach the largest values in scenarios 1 and 2, of the order of 10^{-3} at most, whereas it is one order of magnitude below in scenario 3. While $\mathcal{A}_{FB}$ can be large at small energies, it tends to vanish at very large Q as opposed to the behavior of the $\mathcal{A}_{LR}$ and $\mathcal{A}_{LR\theta}$ asymmetries, which in addition can reach values two orders of magnitude larger than that of $\mathcal{A}_{FB}$. To our knowledge, the $\mathcal{A}_{FB}$ asymmetry has not been analyzed in the literature.

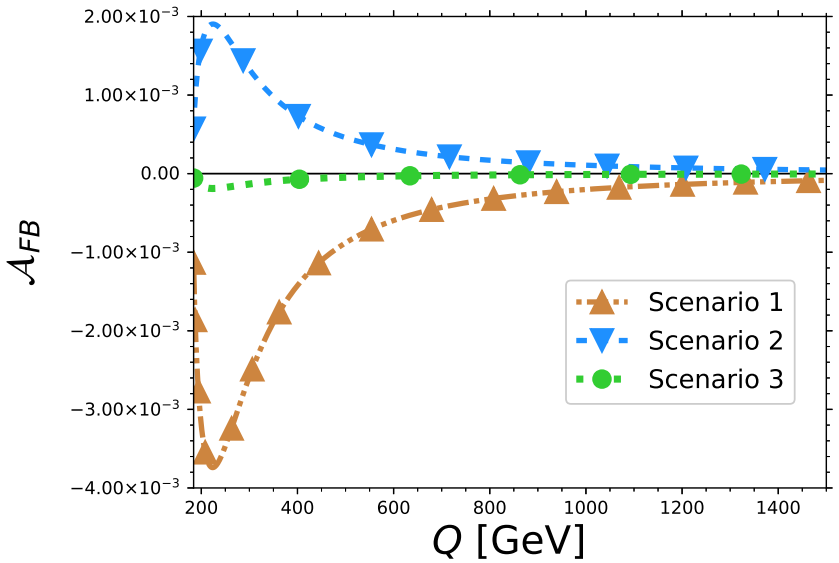


FIG. 11. Behavior of the unpolarized forward-backward asymmetry $\mathcal{A}_{FB}$ as a function of the four-lepton invariant mass Q in the three scenarios of Table I for the anomalous HZZ couplings. The one-loop SM contribution to the h_1^H form factor has been added to the new-physics contribution.

IV. CONCLUSIONS AND OUTLOOK

We have presented a novel analysis of the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width for both unpolarized and polarized Z gauge bosons considering the most general H^*ZZ vertex function, which is parametrized by three form factors h_i^H (i=1, 2, 3), for which we include the SM contributions up to the one-loop level and anomalous contributions arising from new-physics. We then consider the scenario where all the h_i^H couplings are complex, which to our knowledge has never been studied in the literature, and obtain analytical results for both the unpolarized and polarized $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ square amplitudes, out of which the decay widths, angular distributions, and left-right (forward-backward) asymmetries can be straightforwardly obtained. Our results reproduce previous ones obtained in more restrictive scenarios. It is found that the polarized $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay is mainly determined by the $H^* \to ZZ$ one for the four-lepton invariant mass distributions.

To cross-check the consistency of our numerical evaluation method we use Madgraph, with the corresponding Feynman rules for our model obtained via the FeynRules package. For the numerical analysis we consider some realistic scenarios for the complex HZZ anomalous couplings, consistent with the current experimental and indirect constraints, and analyze the role of the transversal (left-handed and right-handed) polarizations of the Z gauge bosons on the behavior of the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width as a function of the four-lepton invariant mass Q, which turns out to be very distinctive for each type of longitudinal polarization. We found that for some specific values of the complex HZZ couplings, the polarized $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay width can deviate considerably from the SM contribution, but such a deviation is more pronounced at large Q. We also analyze the angular distribution of the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay, which exhibits a similar behavior to that observed for the decay width and leads to a left-right asymmetry $\mathcal{A}_{LR}$, which vanishes in the SM as requires the presence of CP-violating anomalous couplings. Hence, we introduce an angular left-right asymmetry $\mathcal{A}_{LR\theta}$, which has a non-vanishing SM one-loop level contribution, but shows considerable deviations due to complex anomalous HZZ couplings. Finally, a polarized forward-backward asymmetry $\mathcal{A}_{FB}^{\lambda}$ is studied, which is found to be constant but differs by a sign for left-handed and right-handed polarizations of the Z gauge bosons. It is also worth noting that the $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay is insensitive to new-physics effects in the case of longitudinally polarized Z gauge bosons. For completeness, we also study the scenario with unpolarized Z gauge bosons, where all the observables discussed in the polarized case could be sensitive to new physics effects.

In summary, the study of the off-shell $H^* \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay, via the transverse polarizations of the Z gauge bosons, could be a useful tool to search for effects of new physics due to the real and absorptive parts of the HZZ anomalous couplings, which could lead to significant deviations from the SM contributions in the decay widths as well as other observables such as angular distributions and left-right (forward-backward) asymmetries. In particular, we have put special emphasis on the study of the effects of the absorptive parts of the HZZ anomalous couplings as they have been largely overlooked in the past but can serve as a probe of the SM at the LHC and future colliders. Such absorptive parts are a unique prediction of quantum field theory.

ACKNOWLEDGMENTS

This work was supported by UNAM Posdoctoral Program (POSDOC). We also acknowledge support from Sistema Nacional de Investigadores (Mexico).

Appendix A: Kinematics and phase space for the $H \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay

The phase space of an n-body decay was studied by Cabibbo and Maksymowicz [54], Pais and Treiman [55], and Byckling and Kajanate [56–60]. In particular, the $1 \to 4$ decay can be decomposed through a recursion relation into 3 partial $1 \to 2$ decays as follows [54, 58, 61]: $1 \to 2 \longrightarrow 1 \to 2+1 \to 2$. Following this approach we obtain below the differential phase space for the $H \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ decay.

For the four-momenta we use the nomenclature defined in Fig. 1, namely, $H(q) \to Z(p_1) + Z(p_2)$, followed by $Z(p_1) \to \ell_1(q_1) + \bar{\ell}_1(q_2)$ and $Z(p_2) \to \ell_2(q_3) + \bar{\ell}_2(q_4)$. For our calculation we consider the reference systems shown in Fig. 2. The differential decay width is thus given in terms of the following three angles θ_1, θ_2 and ϕ angles:

- θ_1 is the angle between $\vec{p}_1$ and $\vec{q}_1$ in the rest frame of the $Z(p_1)$ gauge boson.

- θ_2 is the angle between $\vec{p}_2$ and $\vec{q}_3$ in the rest frame of the $Z(p_2)$ gauge boson.

- ϕ is the relative angle between the $Z(p_1) \to 2\ell_1$ and $Z(p_2) \to 2\ell_2$ decay planes, taken as positive from the $Z(p_1)$ plane to the $Z(p_2)$ plane, with $\phi = 0$ in the case in which both planes coincide and both $\vec{q_1}$ and $\vec{q_3}$ are in the same direction.

We now discuss the kinematics for the distinct reference frames introduced above.

1. Off-shell Higgs boson rest frame

The kinematics of the $H^* \to ZZ$ decay in the Higgs boson rest frame can be described by the following relations

$$q^\mu = \big(Q,\ 0\big), \tag{A1}$$
$$p^\mu_{1,2} = \big(Q/2, \pm\vec{p}\big), \tag{A2}$$

where the magnitude of the three-momentum $\vec{p}$ is

$$\|\vec{p}\| = \frac{\sqrt{Q^2 - 4m_Z^2}}{2}. \tag{A3}$$

We consider that in this frame the Z gauge bosons move along the x axis, so their polarization vectors can be written as

$$\epsilon^\mu_{1,2}(0) = \frac{1}{m_Z}\Big(\|\vec{p}\|, \pm\frac{\sqrt{Q^2}}{2}, 0, 0\Big), \tag{A4}$$
$$\epsilon^\mu_1(R/L) = \frac{1}{\sqrt{2}}\Big(0, 0, -i, \pm 1\Big). \tag{A5}$$

2. $Z(p_i)$ gauge boson rest frame

In our calculation, we also use the rest frame of a $Z(p_i)$ gauge boson decaying into a lepton-antilepton pair, which obeys the following kinematics

$$p^\mu_i = \big(m_Z,\ 0\big), \tag{A6}$$
$$q^\mu_1 = \big(m_Z/2,\ \vec{k}_1\big), \quad q^\mu_2 = \big(m_Z/2,\ -\vec{k}_1\big), \tag{A7}$$
$$q^\mu_3 = \big(m_Z/2,\ \vec{k}_2\big), \quad q^\mu_4 = \big(m_Z/2,\ -\vec{k}_2\big), \tag{A8}$$

where, according to Fig. 2, the three-momenta $\vec{k}_i$ are given as

$$\vec{k}_1 = \frac{\sqrt{m_Z^2 - 4m_{\ell_i}^2}}{2}\big(\cos\theta_1, \sin\theta_1, 0\big), \tag{A9}$$
$$\vec{k}_2 = \frac{\sqrt{m_Z^2 - 4m_{\ell_i}^2}}{2}\big(-\cos\theta_2, \sin\theta_2\cos\phi, \sin\theta_2\sin\phi\big), \tag{A10}$$

whereas the form of the transverse polarization vectors of Eq. (A5) still holds, whereas for longitudinal polarizations we have

$$\epsilon^\mu_{1,2}(0) = \Big(0, \pm 1, 0, 0\Big). \tag{A11}$$

Since the interference term of the $H \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ square amplitude has to be calculated in the Higgs boson frame, we need to boost the four-momenta and polarization vectors defined in the $Z(p_i)$ rest frames into the Higgs boson rest frame, which is achieved via the following Lorentz matrix

$$\Lambda^\mu_\nu = \begin{pmatrix} \gamma & \gamma v & 0 & 0 \\ \gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A12}$$

with

$$v = \frac{\vec{p}_i}{E_i} = \pm\frac{\sqrt{Q^2 - 4m_Z^2}}{Q}. \tag{A13}$$

3. Phase space

The $H \to ZZ \to \bar{\ell}_1\ell_1\bar{\ell}_2\ell_2$ phase space $d\,(R_4)$ can be written as

$$d\,(R_4) = \frac{d^3q_1}{(2\pi)^3 2E_{q_1}}\frac{d^3q_2}{(2\pi)^3 2E_{q_2}}\frac{d^3q_3}{(2\pi)^3 2E_{q_3}}\frac{d^3q_4}{(2\pi)^3 2E_{q_4}}(2\pi)^4\delta^4\big(q - q_1 - q_2 - q_3 - q_4\big), \tag{A14}$$

where the lepton energies are $E_{q_i} = \sqrt{m_i^2 - \vec{p}_i^{\,2}}$ $(i = 1, 2, 3, 4)$. We will follow the approach of Ref. [61], which is analogous to that used in [54]. We thus introduce the relations

$$\int \frac{d^3p_1}{2E_{p_i}} dS_1 \delta^4\big(p_1 - q_1 - q_2\big) = 1, \tag{A15}$$

and

$$\int \frac{d^3p_2}{2E_{p_2}} dS_2 \delta^4\big(p_2 - q_3 - q_4\big) = 1, \tag{A16}$$

where $E_{p_i} = \sqrt{\vec{p}_i^{\,2} + S_i}$, with $S_i = p_i^2$ $(i = 1, 2)$.

The $d\,(R_4)$ phase space can thus be written as

$$d\,(R_4) = \frac{dS_1 dS_2}{(2\pi)^8} I_{p_1} I_{p_2} I_q, \tag{A17}$$

where, in the rest frames of the Z gauge bosons and the Higgs boson, the I_{p_i} and I_q integrals are given for massless leptons as follows

$$I_{p_1} = \int \frac{d^3q_1}{2E_{q_1}}\frac{d^3q_2}{2E_{q_2}}\delta^4\big(p_1 - q_1 - q_2\big) \tag{A18}$$

$$= \frac{\pi}{4} d\cos\theta_1, \tag{A19}$$

$$I_{p_2} = \int \frac{d^3q_3}{2E_{q_3}}\frac{d^3q_4}{2E_{q_4}}\delta^4\big(p_2 - q_3 - q_4\big) \tag{A20}$$

$$= \frac{1}{8} d\cos\theta_2 d\phi, \tag{A21}$$

and

$$I_q = \int \frac{d^3p_1}{2E_{p_1}}\frac{d^3p_2}{2E_{p_2}}\delta^4\big(q - p_1 - p_2\big) \tag{A22}$$

$$= \frac{\pi}{2Q}\sqrt{Q^2 - 4m_Z^2}. \tag{A23}$$

Therefore the $d\,(R_4)$ phase space reads

$$d\,(R_4) = \frac{\sqrt{Q^2 - 4m_Z^2}}{256Q(2\pi)^6} dS_1 dS_2 d\cos\theta_1 d\cos\theta_2 d\phi, \tag{A24}$$

For massless leptons, the integration region reads

$$0 < S_2 < \left(Q - \sqrt{S_1}\right)^2, \tag{A25}$$

$$0 < S_1 < Q^2, \tag{A26}$$

$$0 < \theta_1,\ \theta_2 < \pi, \tag{A27}$$

$$0 < \phi < 2\pi, \tag{A28}$$

where $Q^2 = q^2$ is the invariant mass of the four final leptons,which is the same in all frames of reference and is usually used in LHC analyses. For the purpose of our calculation we denote $Q = \|q\|$.

[1] S. Chatrchyan *et al.* (CMS), Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC, Phys. Lett. B **716**, 30 (2012), arXiv:1207.7235 [hep-ex].

[2] G. Aad *et al.* (ATLAS), 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**, 1 (2012), arXiv:1207.7214 [hep-ex].

[3] B. A. Kniehl, Radiative corrections for $H \to ZZ$ in the standard model, Nucl. Phys. B **352**, 1 (1991).

[4] K. H. Phan, D. T. Tran, and A. T. Nguyen, One-loop off-shell decay $H^* \to ZZ$ at future colliders, (2022), arXiv:2209.12410 [hep-ph].

[5] A. I. Hernández-Juárez, G. Tavares-Velasco, and A. Fernández-Téllez, New evaluation of the HZZ coupling: Direct bounds on anomalous contributions and CP-violating effects via a new asymmetry, Phys. Rev. D **107**, 115031 (2023), arXiv:2301.13127 [hep-ph].

[6] A. Soni and R. M. Xu, Probing CP violation via Higgs decays to four leptons, Phys. Rev. D **48**, 5259 (1993), arXiv:hep-ph/9301225.

[7] K. Hagiwara, S. Ishihara, J. Kamoshita, and B. A. Kniehl, Prospects of measuring general Higgs couplings at e+ e- linear colliders, Eur. Phys. J. C **14**, 457 (2000), arXiv:hep-ph/0002043.

[8] S. S. Biswal, R. M. Godbole, R. K. Singh, and D. Choudhury, Signatures of anomalous VVH interactions at a linear collider, Phys. Rev. D **73**, 035001 (2006), [Erratum: Phys.Rev.D 74, 039904 (2006)], arXiv:hep-ph/0509070.

[9] S. Dutta, K. Hagiwara, and Y. Matsumoto, Measuring the Higgs-Vector boson Couplings at Linear e^+e^- Collider, Phys. Rev. D **78**, 115016 (2008), arXiv:0808.0477 [hep-ph].

[10] M. Javurkova, R. Ruiz, R. C. L. de Sá, and J. Sandesara, Polarized ZZ pairs in gluon fusion and vector boson fusion at the LHC, (2024), arXiv:2401.17365 [hep-ph].

[11] B. A. Kniehl, Theoretical aspects of standard model Higgs boson physics at a future e+ e- linear collider, Int. J. Mod. Phys. A **17**, 1457 (2002), arXiv:hep-ph/0112023.

[12] R. M. Godbole, D. J. Miller, and M. M. Muhlleitner, Aspects of CP violation in the H ZZ coupling at the LHC, JHEP **12**, 031, arXiv:0708.0458 [hep-ph].

[13] K. Rao, S. D. Rindani, and P. Sarmah, Study of anomalous gauge-Higgs couplings using Z boson polarization at LHC, Nucl. Phys. B **964**, 115317 (2021), arXiv:2009.00980 [hep-ph].

[14] K. Rao, S. D. Rindani, P. Sarmah, and B. Singh, Polarized Z cross sections in Higgsstrahlung for the determination of anomalous ZZH couplings, Phys. Lett. B **840**, 137847 (2023), arXiv:2304.11573 [hep-ph].

[15] M. Fabbrichesi, R. Floreanini, E. Gabrielli, and L. Marzola, Stringent bounds on HWW and HZZ anomalous couplings with quantum tomography at the LHC, JHEP **09**, 195, arXiv:2304.02403 [hep-ph].

[16] I. T. Cakir, O. Cakir, A. Senol, and A. T. Tasci, Probing Anomalous HZZ Couplings at the LHeC, Mod. Phys. Lett. A **28**, 1350142 (2013), arXiv:1304.3616 [hep-ph].

[17] B. Şahin, Search for the anomalous ZZH couplings at the CLIC, Mod. Phys. Lett. A **34**, 1950299 (2019).

[18] R. Gauld, U. Haisch, and L. Schnell, SMEFT at NNLO+PS: Vh production, JHEP **01**, 192, arXiv:2311.06107 [hep-ph].

[19] S. Y. Klein, SusHi 2.0 - Higgs production cross sections in BSM models, PoS **EPS-HEP2023**, 416 (2024), arXiv:2311.05939 [hep-ph].

[20] K. Rao, S. D. Rindani, P. Sarmah, and B. Singh, Z polarization at an e^+e^- collider and properties of decay-lepton angular asymmetries, (2023), arXiv:2311.13473 [hep-ph].

[21] H. Xiong, H. Hou, Z. Qian, Q. Xu, and B. Wang, Electroweak corrections to Higgs boson production via Z Z fusion at the future LHeC, (2023), arXiv:2311.14912 [hep-ph].

[22] A. Tumasyan *et al.* (CMS), Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production, Nature Phys. **18**, 1329 (2022), arXiv:2202.06923 [hep-ex].

[23] G. Aad *et al.* (ATLAS), Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector, Phys. Lett. B **846**, 138223 (2023), arXiv:2304.01532 [hep-ex].

[24] G. J. Gounaris, J. Layssac, and F. M. Renard, New and standard physics contributions to anomalous Z and gamma selfcouplings, Phys. Rev. D **62**, 073013 (2000), arXiv:hep-ph/0003143.

[25] D. Choudhury, S. Dutta, S. Rakshit, and S. Rindani, Trilinear neutral gauge boson couplings, Int. J. Mod. Phys. A **16**, 4891 (2001), arXiv:hep-ph/0011205.

[26] A. I. Hernández-Juárez, A. Moyotl, and G. Tavares-Velasco, Contributions to ZZV^* ($V = \gamma, Z, Z'$) couplings from CP violating flavor changing couplings, Eur. Phys. J. C **81**, 304 (2021), arXiv:2102.02197 [hep-ph].

[27] A. I. Hernández-Juárez and G. Tavares-Velasco, Non-diagonal contributions to $Z\gamma V^*$ vertex and bounds on $Z\bar{t}q$ couplings, (2022), arXiv:2203.16819 [hep-ph].

[28] A. I. Hernández-Juárez, A. Moyotl, and G. Tavares-Velasco, New estimate of the chromomagnetic dipole moment of quarks in the standard model, Eur. Phys. J. Plus **136**, 262 (2021), arXiv:2009.11955 [hep-ph].

[29] A. I. Hernández-Juárez, G. Tavares-Velasco, and A. Moyotl, Chromomagnetic and chromoelectric dipole moments of quarks in the reduced 331 model, Chin. Phys. C **45**, 113101 (2021), arXiv:2012.09883 [hep-ph].

[30] A. I. Hernández-Juárez, A. Moyotl, and G. Tavares-Velasco, Bounds on the absorptive parts of the chromomagnetic and

chromoelectric dipole moments of the top quark from LHC data, Eur. Phys. J. Plus **137**, 925 (2022), arXiv:2109.09978 [hep-ph].
[31] E. Maina, Vector boson polarizations in the decay of the Standard Model Higgs, Phys. Lett. B **818**, 136360 (2021), arXiv:2007.12080 [hep-ph].
[32] E. Maina and G. Pelliccioli, Polarized Z bosons from the decay of a Higgs boson produced in association with two jets at the LHC, Eur. Phys. J. C **81**, 989 (2021), arXiv:2105.07972 [hep-ph].
[33] G. Buchalla, O. Cata, and G. D'Ambrosio, Nonstandard Higgs couplings from angular distributions in $h \to Z\ell^+\ell^-$, Eur. Phys. J. C **74**, 2798 (2014), arXiv:1310.2574 [hep-ph].
[34] S. Berge, S. Groote, J. G. Körner, and L. Kaldamäe, Lepton-mass effects in the decays $H \to ZZ^* \to \ell^+\ell^-\tau^+\tau^-$ and $H \to WW^* \to \ell\nu\tau\nu_\tau$, Phys. Rev. D **92**, 033001 (2015), arXiv:1505.06568 [hep-ph].
[35] H.-R. He, X. Wan, and Y.-K. Wang, Anomalous $H \to ZZ \to 4\ell$ decay and its interference effects on gluon–gluon contribution at the LHC, Chin. Phys. C **44**, 123101 (2020), arXiv:1902.04756 [hep-ph].
[36] Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, and N. V. Tran, Spin Determination of Single-Produced Resonances at Hadron Colliders, Phys. Rev. D **81**, 075022 (2010), arXiv:1001.3396 [hep-ph].
[37] S. Bolognesi, Y. Gao, A. V. Gritsan, K. Melnikov, M. Schulze, N. V. Tran, and A. Whitbeck, On the spin and parity of a single-produced resonance at the LHC, Phys. Rev. D **86**, 095031 (2012), arXiv:1208.4018 [hep-ph].
[38] M. Hoppe, M. Schönherr, and F. Siegert, Polarised cross sections for vector boson production with SHERPA, (2023), arXiv:2310.14803 [hep-ph].
[39] L. Cappiello, O. Cata, G. D'Ambrosio, and D.-N. Gao, $K^+ \to pi^+pi^0e^+e^-$: a novel short-distance probe, Eur. Phys. J. C **72**, 1872 (2012), [Erratum: Eur.Phys.J.C 72, 2208 (2012)], arXiv:1112.5184 [hep-ph].
[40] S. R. Gevorkyan and M. H. Misheva, Different approaches to calculate the $K^\pm \to \pi^\pm\pi^0e^+e^-$ decay width, Eur. Phys. J. C **74**, 2860 (2014), arXiv:1403.1053 [hep-ph].
[41] R. Mertig, M. Bohm, and A. Denner, FEYN CALC: Computer algebraic calculation of Feynman amplitudes, Comput. Phys. Commun. **64**, 345 (1991).
[42] V. Shtabovenko, R. Mertig, and F. Orellana, New Developments in FeynCalc 9.0, Comput. Phys. Commun. **207**, 432 (2016), arXiv:1601.01167 [hep-ph].
[43] V. Shtabovenko, R. Mertig, and F. Orellana, FeynCalc 9.3: New features and improvements, Comput. Phys. Commun. **256**, 107478 (2020), arXiv:2001.04407 [hep-ph].
[44] R. L. Workman *et al.* (Particle Data Group), Review of Particle Physics, PTEP **2022**, 083C01 (2022).
[45] S. J. Lee, M. Park, and Z. Qian, Probing unitarity violation in the tail of the off-shell Higgs boson in V_LV_L mode, Phys. Rev. D **100**, 011702 (2019), arXiv:1812.02679 [hep-ph].
[46] A. I. Hernández-Juárez, Mathematica code for *hzz* analysis, `https://gitlab.com/fcfm-buap-rc-group/zzh-anomalous-couplings`, accessed: 2022-11-22.
[47] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP **07**, 079, arXiv:1405.0301 [hep-ph].
[48] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, UFO - The Universal FeynRules Output, Comput. Phys. Commun. **183**, 1201 (2012), arXiv:1108.2040 [hep-ph].
[49] N. D. Christensen and C. Duhr, FeynRules - Feynman rules made easy, Comput. Phys. Commun. **180**, 1614 (2009), arXiv:0806.4194 [hep-ph].
[50] A. Denner, S. Dittmaier, M. Roth, and L. H. Wieders, Electroweak corrections to charged-current e+ e- —> 4 fermion processes: Technical details and further results, Nucl. Phys. B **724**, 247 (2005), [Erratum: Nucl.Phys.B 854, 504–507 (2012)], arXiv:hep-ph/0505042.
[51] T. Hahn and M. Perez-Victoria, Automatized one loop calculations in four-dimensions and D-dimensions, Comput. Phys. Commun. **118**, 153 (1999), arXiv:hep-ph/9807565.
[52] K. Abe *et al.* (SLD), Precise measurement of the left-right cross-section asymmetry in Z boson production by e+ e- collisions, Phys. Rev. Lett. **73**, 25 (1994), arXiv:hep-ex/9404001.
[53] S. Narita, *Measurement of the polarized forward - backward asymmetry of s quarks at SLD*, Ph.D. thesis, Tohoku U. (1998).
[54] N. Cabibbo and A. Maksymowicz, Angular Correlations in Ke-4 Decays and Determination of Low-Energy pi-pi Phase Shifts, Phys. Rev. **137**, B438 (1965), [Erratum: Phys.Rev. 168, 1926 (1968)].
[55] A. Pais and S. B. Treiman, Pion Phase-Shift Information from $K_{\ell 4}$ Decays, Phys. Rev. **168**, 1858 (1968).
[56] E. Byckling and K. Kajantie, Kinematic separation of three-particle channels in counter experiments, Nucl. Phys. B **14**, 355 (1969).
[57] E. Byckling and K. Kajantie, Reductions of the phase-space integral in terms of simpler processes, Phys. Rev. **187**, 2008 (1969).
[58] E. Byckling and K. Kajantie, N-particle phase space in terms of invariant momentum transfers, Nucl. Phys. B **9**, 568 (1969).
[59] E. Byckling and K. Kajantie, COUNTER EXPERIMENTS FOR THREE-PARTICLE FINAL STATES, (1969).
[60] E. Byckling and K. Kajantie, *Particle Kinematics: (Chapters I-VI, X)* (University of Jyvaskyla, Jyvaskyla, Finland, 1971).
[61] H.-Y. Cheng, C.-Y. Cheung, W. Dimm, G.-L. Lin, Y. C. Lin, T.-M. Yan, and H.-L. Yu, Heavy quark and chiral symmetry predictions for semileptonic decays anti-B —> D (D*) pi lepton anti-neutrino, Phys. Rev. D **48**, 3204 (1993), arXiv:hep-ph/9305340.

CPPC-2024-01, MS-TP-24-06

Prepared for submission to JCAP

arXiv:2402.18481v1 [hep-ph] 28 Feb 2024

Towards a precision calculation of $N_{\rm eff}$ in the Standard Model III: Improved estimate of NLO corrections to the collision integral

Marco Drewes,[a] **Yannis Georis,**[a] **Michael Klasen,**[b] **Luca Paolo Wiggering**[b] **and Yvonne Y. Y. Wong**[c]

[a]Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain, Louvain-la-Neuve B-1348, Belgium

[b]Institut für Theoretische Physik, Universität Münster, Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany

[c]Sydney Consortium for Particle Physics and Cosmology, School of Physics, The University of New South Wales, Sydney NSW 2052, Australia

E-mail: marco.drewes@uclouvain.be, yannis.georis@uclouvain.be, michael.klasen@uni-muenster.de, luca.wiggering@uni-muenster.de, yvonne.y.wong@unsw.edu.au

Abstract. We compute the dominant QED correction to the neutrino-electron interaction rate in the vicinity of neutrino decoupling in the early universe, and estimate its impact on the effective number of neutrino species $N_{\rm eff}$ in cosmic microwave background anisotropy observations. We find that the correction to the interaction rate is at the sub-percent level, consistent with a recent estimate by Jackson and Laine. Relative to that work we include the electron mass in our computations, but restrict our analysis to the enhanced t-channel contributions. The fractional change in $N_{\rm eff}^{\rm SM}$ due to the rate correction is of order 10^{-5} or below, i.e., about a factor of 30 smaller than that recently claimed by Cielo *et al.*, and below the nominal computational uncertainties of the current benchmark value of $N_{\rm eff}^{\rm SM} = 3.0440 \pm 0.0002$. We therefore conclude that aforementioned number remains to be the state-of-the-art benchmark for $N_{\rm eff}^{\rm SM}$ in the standard model of particle physics.

Contents

1 Introduction **1**

2 Preliminaries **4**
2.1 Damping approximation 6

3 Calculation and NLO results **6**
3.1 Evaluation of the closed fermion loop 7
3.2 Numerical results for NLO neutrino interaction rate 11
3.3 NLO decoupling temperatures 13

4 NLO effects on $N_{\rm eff}^{\rm SM}$ **15**
4.1 Entropy conservation 15
4.2 Solving the neutrino Boltzmann equations and the continuity equation 18

5 Conclusions **19**

A Bosonic and fermionic propagators at finite temperatures **21**

B Thermal integrals **21**

C 1PI-resummed photon propagator **22**
C.1 1-loop photon self-energy 23
C.2 The resummed photon propagator 28

D Parameterisation of the collision integrals **29**

E Functions in the continuity equation **30**

1 Introduction

The effective number of neutrinos, $N_{\rm eff}$, is an important parameter in standard hot big bang cosmology [1]. Defined as the energy density residing in free-streaming, ultra-relativistic particle species relative to the photon energy density in the post-neutrino decoupling early universe (i.e., at temperature $T \lesssim 1$ MeV), the primary role of $N_{\rm eff}$ is to fix the universal expansion rate at $T \lesssim 1$ MeV up to the end of the radiation-domination epoch. Its observable consequences are many—from setting the primordial light element abundances, to influencing the correlation statistics of the cosmic microwave background (CMB) anisotropies and the large-scale matter distribution. Coupled with the fact that many beyond-the-standard-model scenarios predict $N_{\rm eff}$-like effects (e.g., light sterile neutrinos [2, 3], axions [4, 5], gravitational waves [6], hidden sectors [7, 8], etc.), pinning down its value both theoretically and observationally has enjoyed an unwavering interest for over four decades [1, 9].

From the theoretical perspective, the expected value of $N_{\rm eff}$ in the context of the standard model (SM) of particle physics is 3 (for three generations), plus percent-level corrections due to residual energy transfer between the quantum electrodynamic (QED) plasma and the neutrino sector during neutrino decoupling [10–15] as well as deviations of the QED plasma itself

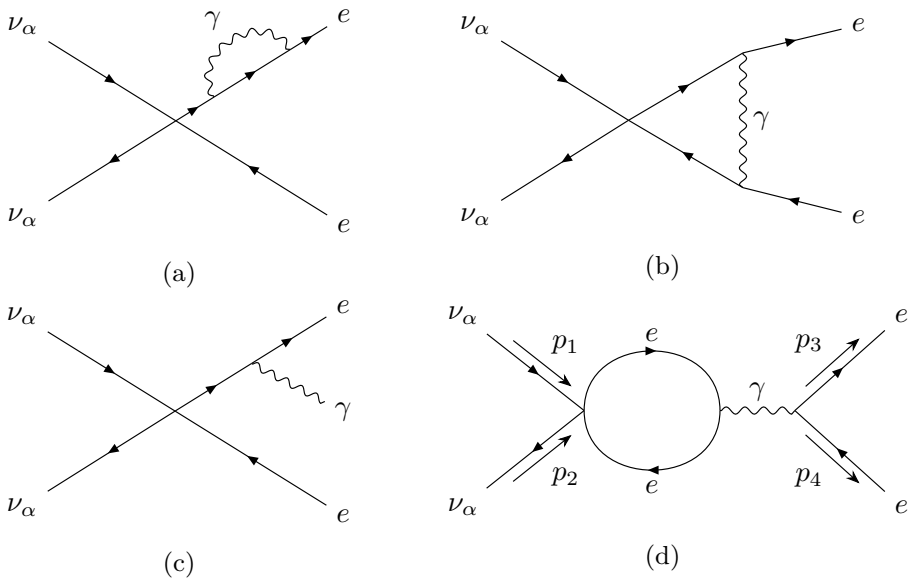


Figure 1: The four qualitatively different QED corrections to the neutrino interaction rate corresponding to (a) modification of the electron dispersion relation, (b) virtual photon exchange, (c) thermal photon emission and absorption, and (d) corrections with a closed fermion loop. All diagrams are schematic in that all time directions are possible.

from an ideal gas [16–21]. Historical estimates of these corrections have ranged from 0.011 to 0.052 [12, 22–25]. Detailed modelling in recent years [26–29], however, have drastically reduced the spread. Notably, two of us (Drewes and Wong) and collaborators reported in [29] a prediction of $N_{\rm eff}^{\rm SM} = 3.0440 \pm 0.0002$ from a fully momentum-dependent precision transport calculation that accounted for (i) neutrino oscillations, (ii) finite-temperature corrections to the QED equation of state (EoS) to $\mathcal{O}(e^3)$, and (iii) a first estimate of finite-temperature corrections of type (a) to the weak interaction rates depicted in figure 1 (see also table 1); the error bars are due mainly to numerical resolution and experimental uncertainties in the input neutrino mixing angles. More remarkably still, this result is in perfect agreement with the independent calculation of [28] modelling the same physics—to five significant digits in the central value and with comparable error estimates. The precision computation of $N_{\rm eff}^{\rm SM}$ appears therefore to have reached convergence, at least in a limited sense.

There are nonetheless reasons to be cautious. For one, while the general expectation is that the physics summarised in table 1 dominate corrections to $N_{\rm eff}^{\rm SM}$, as yet missing is a systematic study of possible higher-order effects that may be at least as important. Indeed, in estimating only next-to-leading-order (NLO) effects due to diagram (a) to the weak rates rather than the full range of corrections displayed in figure 1, one could argue that the computations of [28, 29] are, at least conceptually, incomplete. Two recent works have taken a first step towards filling this gap:

- Cielo *et al.* [30] took the finite-temperature rate corrections for $e^+e^- \to \nu_\alpha \bar{\nu}_\alpha$ from [31], computed originally in the context of energy loss in a stellar plasma, and appealed to detailed balance to estimate the corrections to collision integrals. The claimed effect of this correction on $N_{\rm eff}^{\rm SM}$ is quite substantial—at the ~ 0.001 level. We have reservations about this result: Aside from mapping rate corrections from an incompatible energy

Standard-model corrections to $N_{\rm eff}^{\rm SM}$	Leading-digit contribution
m_e/T_d correction	$+0.04$
$\mathcal{O}(e^2)$ FTQED correction to the QED EoS	$+0.01$
Non-instantaneous decoupling+spectral distortion	-0.006
$\mathcal{O}(e^3)$ FTQED correction to the QED EoS	-0.001
Flavour oscillations	$+0.0005$
Type (a) FTQED corrections to the weak rates	$\lesssim 10^{-4}$
Sources of uncertainty	
Numerical solution by `FortEPiaNO`	± 0.0001
Input solar neutrino mixing angle θ_{12}	± 0.0001

Table 1: Leading-digit contributions from various SM corrections, in order of importance, thus far accounted that make up the final $N_{\rm eff}^{\rm SM} - 3 = 0.0440 \pm 0.0002$ [28, 29]. Table adapted from [29].

regime ($\mathcal{O}(1)$ keV $\ll m_e$ in a stellar plasma versus $\mathcal{O}(1)$ MeV $> m_e$ in the early universe), the analysis of [30] also neglected (i) corrections due to diagram (d) in figure 1 (the "closed fermion loop"), (ii) corrections to elastic scattering reactions like $\nu_\alpha e \to \nu_\alpha e$, as well as (iii) Pauli blocking effects of the final-state neutrinos. Of particular note is that the neglected closed fermion loop diagram (d) contains a t-channel enhancement, which should, at least naïvely, dominate the weak rate corrections.

- The more recent work of Jackson and Laine [32] considered all diagrams in figure 1 in a first-principles calculation using the *imaginary-time* formalism of thermal field theory which accounts for both vacuum and finite-temperature corrections,[1] along with an estimate of hadronic corrections to diagram (d) following [34]. Because the imaginary-time formalism applies only to systems in thermal equilibrium, the computation of [32] effectively neglected non-equilibrium neutrino phase space distributions. Additionally, the authors assumed a negligible electron mass by setting $m_e = 0$, which is not necessitated by the formalism. The final result, presented as a set of corrections to the leading-order (LO) neutrino damping rate, confirms the expected t-channel enhancement in diagram (d). Nevertheless, these corrections are minute—of order 0.2 to 0.3%. While the authors of [32] did not report the corresponding change in $N_{\rm eff}^{\rm SM}$, it is clear that corrections of this magnitude cannot effect a shift in $N_{\rm eff}$ as sizeable as that claimed in reference [30].

The purpose of the present work is to clarify whether or not QED corrections to the neutrino interaction rate can alter the standard-model $N_{\rm eff}^{\rm SM}$ at the ~ 0.001 level as claimed in [30]. While this correction is small relative to the anticipated sensitivity of the next-generation CMB-S4 program to $N_{\rm eff}$, $\sigma(N_{\rm eff}) \simeq 0.02 - 0.03$ [35], an accurate theoretical prediction for $N_{\rm eff}^{\rm SM}$ at per-mille level accuracy is nonetheless desirable to justify neglecting the theoretical uncertainty in cosmological parameter inference. In this regard, our work shares a common motivation with Jackson and Laine [32]. However, our work also differs from [32] in three important ways:

[1] We note that computations of the vacuum corrections to the neutrino-electron interaction due to diagrams (a)–(d) already appeared previously in reference [33].

1. Our first-principles computation of the QED corrections uses the so-called *real-time* formalism and focuses on corrections of type (d) which contain a t-channel enhancement. Computation of the non-enhanced (and hence sub-dominant) diagrams (a)–(c) is postponed to a follow-up work. Like the imaginary-time formalism, the real-time formalism also automatically takes care of both vacuum and finite-temperature corrections. While in equilibrium situations the two formalisms are exactly equivalent [36], the latter has the advantage that it can easily be generalised to non-equilibrium settings [37, 38]. Thus, although we shall restrict the present analysis to the same equilibrium conditions as in [32] and hence provide an independent partial cross-check of their results, our calculation will also pave the way for incorporating NLO effects in neutrino decoupling codes such as `FortEPianNO` [26] to deliver a final verdict on $N_{\rm eff}^{\rm SM}$.

2. We retain a finite electron mass m_e in our calculation, which represents a departure from the $m_e = 0$ approximation made in [32]. Neutrino decoupling occurs at temperatures $T \sim 1$ MeV, in whose vicinity the setting of $m_e = 0$ may not be well justified. We shall examine the validity of the approximation and its impact on $N_{\rm eff}^{\rm SM}$.

3. Finally, we use our NLO results to estimate the corresponding change in $N_{\rm eff}^{\rm SM}$. This estimate was missing in [32].

The rest of the article is organised as follows. In section 2 we describe the physical system and sketch how $N_{\rm eff}^{\rm SM}$ is computed. Section 3 outlines the computation of QED corrections to the neutrino damping rate due to diagram (d) and presents our numerical estimates of these rate corrections. The shift in $N_{\rm eff}^{\rm SM}$ due to these corrections is presented in section 4, and section 5 contains our conclusions. Four appendices provide details on the bosonic and fermionic propagators at finite temperatures, explicit expressions for certain thermal integrals, derivation of the 1PI-resummed photon propagator, and parameterisation of the collision integrals.

2 Preliminaries

The SM effective number of neutrinos $N_{\rm eff}^{\rm SM}$ is defined via the ratio of the energy density contained in three generations of neutrinos and antineutrinos, ρ_ν, to the photon energy density ρ_γ in the limit $T/m_e \to 0$,

$$\left.\frac{\rho_\nu}{\rho_\gamma}\right|_{T/m_e\to 0} \equiv \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}^{\rm SM}. \tag{2.1}$$

Here, T is the photon temperature, m_e the electron mass, and the limit $T/m_e \to 0$ is understood to apply only to $T \gg m_i$, where m_i is the largest neutrino mass.

To compute the precise value of $N_{\rm eff}^{\rm SM}$ requires that we track the evolution of ρ_ν and ρ_γ simultaneously across the the neutrino decoupling epoch, i.e., $T \sim \mathcal{O}(10) \to \mathcal{O}(0.01)$ MeV. Assuming that at these temperatures the photons are held in a state of thermodynamic equilibrium together with the electrons/positrons— collectively referred to as the "QED plasma"— this is typically achieved by solving two sets of evolution equations: (i) a continuity equation that follows the total energy density of the universe, and (ii) some form of generalised Boltzmann equations—often referred to as the quantum kinetic equations (QKEs)—which describe the non-equilibrium dynamics in the neutrino sector across its decoupling.

Continuity equation In a Friedmann-Lemaître-Robertson-Walker (FLRW) universe, the continuity equation is given by

$$\frac{\mathrm{d}\rho_{\rm tot}}{\mathrm{d}t} + 3H(\rho_{\rm tot} + P_{\rm tot}) = 0, \tag{2.2}$$

where $\rho_{\rm tot}$ and $P_{\rm tot}$ are, respectively, the total energy density and pressure, $H \equiv (1/a)(\mathrm{d}a/\mathrm{d}t)$ the Hubble expansion rate, and a is the scale factor. For the physical system at hand, $\rho_{\rm tot} \equiv \rho_{\rm QED} + \rho_\nu$ and $P_{\rm tot} \equiv P_{\rm QED} + P_\nu$, where $\rho_{\rm QED} \equiv \rho_\gamma + \rho_e$ subsumes the photon and the electron/positron energy densities, and similarly for $P_{\rm QED}$. The assumption of thermodynamic equilibrium in the QED sector in the time frame of interest means that the standard thermodynamic relation $\rho_{\rm QED} = -P_{\rm QED} + T\,(\partial/\partial T)P_{\rm QED}$ applies. Then, the finite-temperature corrections to the QED equation of state summarised in table 1 simply refer to corrections to the QED partition function $Z_{\rm QED}$ and hence $P_{\rm QED} = (T/V)\ln Z_{\rm QED}$ that alter $\rho_{\rm QED} + P_{\rm QED} = T\,(\partial/\partial T)P_{\rm QED}$ from its ideal-gas expectation. Corrections to $Z_{\rm QED}$ are known to $\mathcal{O}(e^3)$ for arbitrary m_e and chemical potential μ [39] and to $\mathcal{O}(e^5)$ for $m_e = \mu = 0$ [40, 41]. Their effects on $N_{\rm eff}^{\rm SM}$ have been estimated in references [17, 20, 21] to $\mathcal{O}(e^4)$.

Quantum kinetic equations State-of-the-art neutrino decoupling calculations employ the QKEs to track simultaneously the effects of in-medium neutrino flavour oscillations and particle scatterings on the one-particle reduced density matrix of the neutrino ensemble, $\varrho = \varrho(p,t)$. Schematically, the QKEs take the form [42]

$$\partial_t \varrho - pH\partial_p \varrho = -\mathrm{i}[\mathbb{H}, \varrho] + \mathcal{I}[\varrho], \tag{2.3}$$

where $\mathbb{H} = \mathbb{H}(p,t) = \mathbb{H}_{\rm vac} + \mathbb{V}$ is the effective Hamiltonian incorporating vacuum flavour oscillations $\mathbb{H}_{\rm vac}$ and in-medium corrections from forward scattering $\mathbb{V}$,[2] and

$$\mathcal{I}[\varrho] = \frac{1}{2}\Big((\mathbb{1} - \varrho)\mathbb{\Gamma}^{<} - \varrho\mathbb{\Gamma}^{>}\Big) + \mathrm{h.c.} \tag{2.4}$$

is the collision integral encapsulating the non-unitary gains ($\mathbb{\Gamma}^{<} = \mathbb{\Gamma}^{<}(p,t)$) and losses ($\mathbb{\Gamma}^{>} = \mathbb{\Gamma}^{>}(p,t)$) of ϱ. In the context of precision $N_{\rm eff}^{\rm SM}$, the quantities ϱ, $\mathbb{H}$, and $\mathbb{\Gamma}^{\gtrless}$ are understood to be 3×3 hermitian matrices in flavour space, with the diagonal entries of ϱ corresponding to the occupation numbers $f_\alpha(p,t) \equiv \{\varrho(p,t)\}_{\alpha\alpha}$, for $\alpha = e, \mu, \tau$.[3] We assume a CP-symmetric universe, which is well justified if any asymmetry in the lepton sector mirrors the baryon asymmetry in the observable universe [43]. In practice this means it suffices to follow only one set of QKEs for the neutrinos; the antineutrinos evolve in the same way.

For the problem at hand, the gain and loss terms $\mathbb{\Gamma}^{\gtrless}$ incorporate in principle all weak scattering processes wherein at least one neutrino appears in either the initial or final state. In the temperature window of interest, however, it suffices to consider only $2 \to 2$ processes involving (i) two neutrinos and two electrons any way distributed in the initial and final states (labelled "νe"), and (ii) neutrino-neutrino scattering ("$\nu\nu$"). The leading-order $\mathbb{\Gamma}^{\gtrless}$ for these

[2] Depending on context, these modifications to the in-medium quasiparticle dispersion relations are variously known as thermal masses, matter potentials, or refractive indices.

[3] Strictly speaking, an occupation number refers to a diagonal element of ϱ in the basis in which $\mathbb{H}$ is diagonal, i.e., in the mass basis, rather than the interaction basis. This distinction is however unimportant for the present study, and we follow the common practice of calling, e.g., $\{\varrho(p,t)\}_{ee}$ the electron neutrino occupation number, etc.

processes are well known, and take the form of two-dimensional momentum integrals,

$$\begin{aligned}\mathbb{\Gamma}^{\gtrless}_{\nu e}(p,t) &\propto G_F^2 \int \mathrm{d}p_2\,\mathrm{d}p_3\,\mathbb{\Pi}_{\nu e}(p,p_2,p_3)\,\mathcal{F}^{\gtrless}_{\nu e}(\varrho(p_2,t),f_e(p_3,t),f_e(p_4,t)),\\ \mathbb{\Gamma}^{\gtrless}_{\nu\nu}(p,t) &\propto G_F^2 \int \mathrm{d}p_2\,\mathrm{d}p_3\,\mathbb{\Pi}_{\nu\nu}(p,p_2,p_3)\,\mathcal{F}^{\gtrless}_{\nu\nu}(\varrho(p_2,t),\varrho(p_3,t),\varrho(p_4,t)).\end{aligned} \tag{2.5}$$

Here, G_F is the Fermi constant; p_4 is fixed by energy-momentum conservation; $\mathbb{\Pi}_{\nu e}$ and $\mathbb{\Pi}_{\nu\nu}$ are scattering kernels, which are diagonal in the neutrino interaction basis; and $\mathcal{F}^{\gtrless}_{\nu e}$ and $\mathcal{F}^{\gtrless}_{\nu\nu}$ are real-time matrix products of $\varrho(p_2,t)$ with $f_e(p_3,t)$, $f_e(p_4,t)$ and $\varrho(p_3,t)$, $\varrho(p_4,t)$, respectively. Because ϱ can develop non-zero off-diagonal components from neutrino oscillations, $\mathcal{F}^{\gtrless}$ and hence $\mathbb{\Gamma}^{\gtrless}$ are generally not diagonal in the interaction basis.

The momentum integrals (2.5) are hard-coded in the purpose-built neutrino decoupling code `FortEPianNO` [26, 29], which solves numerically the continuity equation (2.2) and the three-flavour QKEs (2.3) in their entirety in precision $N_{\rm eff}^{\rm SM}$ computations.

2.1 Damping approximation

We would like to compute QED corrections to $\mathbb{\Gamma}^{\gtrless}_{\nu e}$, and estimate their impact on $N_{\rm eff}^{\rm SM}$. Ideally these corrections would take the form of corrections to the scattering kernel $\mathbb{\Pi}_{\nu e}$, to be incorporated into a neutrino decoupling code such as `FortEPianNO`. As a first pass, however, we may work within the damping approximation, which makes the simplifying assumption that all particle species—besides that at the momentum mode p represented by $\varrho(p)$—are in thermal equilibrium specified by a common temperature T equal to the photon temperature. Then, defining $\delta\varrho = \varrho(p) - \varrho_{\rm eq}(p)$, where $\varrho_{\rm eq}(p) = \mathrm{diag}(f_F(p), f_F(p), f_F(p))$ and $f_F(p) = [\exp(p/T)+1]^{-1}$ is the Fermi-Dirac distribution, the collision integral (2.4) can be expanded to linear order in $\delta\varrho$ to give

$$\{\mathcal{I}[\varrho(p)]\}_{\alpha\beta} \simeq -\{D(p,T)\}_{\alpha\beta}\Big[\{\varrho(p)\}_{\alpha\beta} - \delta_{\alpha\beta} f_F(p)\Big], \tag{2.6}$$

where $\delta_{\alpha\beta}$ is a Kronecker-δ, and

$$\{D(p,T)\}_{\alpha\beta} \equiv \frac{1}{2}\Big[\{\mathbb{\Gamma}(p,T)\}_{\alpha\alpha} + \{\mathbb{\Gamma}(p,T)\}_{\beta\beta}\Big] \tag{2.7}$$

are the damping coefficients composed of components of the damping rate matrix

$$\mathbb{\Gamma}(p,T) \equiv \mathbb{\Gamma}^{>}(p,T) + \mathbb{\Gamma}^{<}(p,T). \tag{2.8}$$

Linearisation in $\delta\varrho$ ensures that $\mathbb{\Gamma}(p,T)$ is diagonal in the neutrino interaction basis, and that $\mathbb{\Gamma}^{>}(p,T) = e^{E_p/T}\mathbb{\Gamma}^{<}(p,T)$ holds by detailed balance, where $E_p = p$ is the neutrino energy at the momentum mode p of interest. This is the approximation under which we shall compute $\mathbb{\Gamma}$ in section 3. Details of the derivation of (2.6) can be found in, e.g., appendix B of [29].

In the following, we shall devote section 3 to evaluating QED corrections to the damping rate $\{\mathbb{\Gamma}_{\nu e}(p,T)\}_{\alpha\alpha}$ due to the closed fermion loop (diagram (d) in figure 1). Our estimates of its impact on $N_{\rm eff}^{\rm SM}$ will be presented in section 4.

3 Calculation and NLO results

We compute QED corrections to the interaction rates of the weak processes $\nu_\alpha e \to \nu_\alpha e$ and $\nu_\alpha\bar{\nu}_\alpha \leftrightarrow e^+e^-$ within the framework of Fermi theory, whose effective Lagrangian can be

expressed as

$$\mathcal{L}_{4\mathrm{F}} = \frac{4G_F}{\sqrt{2}} \left[\bar{\psi}_\alpha \gamma_\mu P_L \psi_\alpha\right] \left[g_L^\alpha J_L^\mu + g_R J_R^\mu\right]. \tag{3.1}$$

Here, the spinors ψ_α represent neutrino fields of flavour α; $J^\mu_{L/R} = \bar{\psi}_e \gamma^\mu P_{L/R} \psi_e$ are the left- and right-handed electron current operators, with the chiral projectors $P_{L/R} = \frac{1}{2}(1 \mp \gamma^5)$; the right-handed coupling reads $g_R = \sin^2\theta_W$, while the left-handed coupling $g_L^\alpha = -\frac{1}{2} + \sin^2\theta_W + \delta^{\alpha e}$ depends on the neutrino flavor $\alpha = e, \mu, \tau$. Strictly speaking, the closed fermion loop in figure 1 receives in principle contributions also from quarks. This interaction is also well described by a Lagrangian of the form (3.1), with the couplings g_L^α and g_R updated for the quarks of interest.[4] We shall however omit the contributions from quarks in this work: at finite temperatures these contributions are Boltzmann-suppressed at the $\mathcal{O}(1)$ MeV temperatures of interest, while hadronic corrections in vacuum have been studied in [34]. As QED is a vector-like theory, we also introduce the vector-axial couplings $g^\alpha_{V,A} = \frac{1}{2}(g_L^\alpha \pm g_R)$ as an alternative notation.

3.1 Evaluation of the closed fermion loop

In order to compute the damping coefficients (2.7), we must evaluate the flavour-diagonal entries of the neutrino damping rate matrix in the equilibrium approximation for ϱ, i.e., $\Gamma_\alpha(p,T) \equiv \{\mathbb{\Gamma}_{\nu e}(p,T)\}_{\alpha\alpha}$, which is given by the sum

$$\Gamma_\alpha = \Gamma^>_\alpha(p) + \Gamma^<_\alpha(p), \tag{3.2}$$

where $\Gamma^{\gtrless}_\alpha$ are the α-flavoured neutrino production ($<$) and destruction ($>$) rates in the equilibrium limit. In the framework of the real-time formalism of finite-temperature QED, these rates can be extracted from the Wightman self-energies $\Sigma^{\gtrless}$,[5]

$$\Gamma^{\gtrless}_\alpha(p) = \mp \frac{1}{2p^0} \mathrm{Tr}\left[\not{p}\Sigma^{\gtrless}_\alpha\right]\big|_{p^0=E_p}, \tag{3.3}$$

where $\Sigma^{\gtrless}$ are identified with the self-energies with opposite (Keldysh) contour indices:[6] $\Sigma^< \equiv \Sigma^{12}$ and $\Sigma^> \equiv \Sigma^{21}$. In thermal equilibrium, detailed balance is established by the fermionic Kubo-Martin-Schwinger (KMS) relation $\Sigma^> = -e^{p^0/T}\Sigma^<$. Then, the mode-dependent damping rate can also be written as

$$\Gamma_\alpha(p) = \frac{1}{2p^0} \mathrm{Tr}\left[\not{p}(\Sigma^<_\alpha - \Sigma^>_\alpha)\right]\big|_{p^0=E_p} = \frac{1}{2p^0 f_F(p^0)} \mathrm{Tr}\left[\not{p}\Sigma^<_\alpha\right]\big|_{p^0=E_p}, \tag{3.4}$$

where the first equality follows from the fact that Γ_α can be related to the discontinuity of the retarded neutrino self-energy evaluated at the quasiparticle pole, $-i\mathrm{disc}\Sigma^R = 2\mathrm{Im}\Sigma^R = i(\Sigma^> - \Sigma^<)$, in accordance with the optical theorem at finite temperature. While the KMS relation makes explicit use of the fact that $\Sigma^{\gtrless}$ are computed in thermal equilibrium, the real-time formalism used here can be generalised in a straightforward manner to non-equilibrium situations [37].[7] Where there is no confusion, we shall drop the flavour index α in the following.

[4] The reason is that a Fierz identity has been used to bring the charged and neutral current contributions into the common form (3.1).

[5] There is no term involving the plasma four-velocity U because we consider the sum over helicities. See, e.g., appendix C of [44] in the context of heavy neutrinos.

[6] The idea to evaluate fields on a closed time path was already implicit in Schwinger's seminal work [45]. Here, we follow common convention and refer to the indices marking the forward and backward contours in

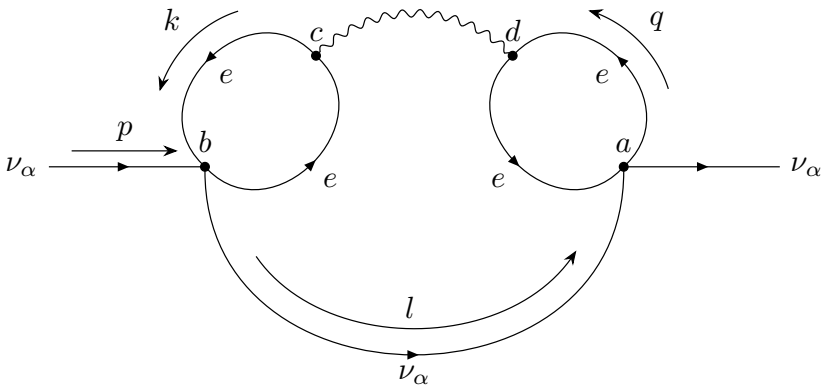


Figure 2: The three-loop diagram containing the closed fermion loop. Explicitly labelled are the external (a and b) and summed (c and d) real-time contour indices, the momenta (k, l, p and q) and particles (e and ν_α).

On general grounds the dominant QED correction to Γ_α can be expected from diagram (d) in figure 1 in the regime where the photon propagator is on-shell. This expectation has been confirmed in [32]. In the real-time formalism this process is contained in the contribution to the Wightman self-energy shown in figure 2, where, in terms of finite-temperature cutting rules, diagram (d) can be recovered from a cut through one closed electron loop and the internal neutrino line.[8] We shall compute only this contribution and denote the corresponding Wightman self-energy with Σ^{ba}_{NLO}. For notational convenience, we further split Σ^{ba}_{NLO} into a sum $\Sigma^{ba}_{\text{NLO}} = \sum_{c,d=1,2} \Sigma^{ba,cd}$ over the internal real-time indices c and d, with the partial self-energies $\Sigma^{ba,cd}$ given by

$$\begin{aligned}\text{Tr}\left[\not{p}\Sigma^{ba,cd}\right] &= (-1)^{a+b+c+d+1}(ie)^2\left(\frac{i4G_F}{\sqrt{2}}\right)^2 \\ &\times \int_{l,q,k} \text{Tr}\left[iS_e^{ad}(q)\gamma_\rho(P_L g_L^\alpha + P_R g_R) iS_e^{da}(q+p-l)\gamma^\mu\right] iD^{cd}_{\mu\nu}(p-l) \\ &\times \text{Tr}\left[\not{p}\gamma^\rho P_L iS^{ba}_{\nu_\alpha}(l)\gamma^\sigma P_L\right] \text{Tr}\left[iS_e^{cb}(k)\gamma^\nu iS_e^{bc}(k+p-l)\gamma_\sigma(P_L g_L^\alpha + P_R g_R)\right]. \quad (3.5)\end{aligned}$$

Here, we have introduced the abbreviation $\int_p \equiv \int \mathrm{d}^4p\,/(2\pi)^4$ for loop integrals; the traces are over the Clifford algebra; the subscripts "e" and "ν_α" on the fermionic propagators refer to the associated particle; and the definitions of the propagators can be found in appendix A.

---

the real-time formalism as Keldysh indices [46].

[7] Specifically, this entails replacing the equilibrium distribution in the relation (A.4) between the statistical and the spectral propagators with a dynamical function, and modifying all other propagators accordingly. See, e.g. [47, 48] for a discussion.

[8] Resumming the photon propagator (see discussion above equation (3.13) and appendix C) opens up an additional cut through the photon and the internal neutrino lines, corresponding to the so-called plasmon process $\gamma \to \nu\bar{\nu}$, even if the imaginary part of the photon self-energy is neglected. This can be seen in equation (C.32), which exhibits an on-shell δ-function in the limit of vanishing width (C.34). However, the plasmon process contributes at $\mathcal{O}(\alpha_{\text{em}}^2)$ (see appendix C in [49]), which can be understood as a phase space suppression from $m_\gamma \sim \alpha_{\text{em}} T$. We therefore do not consider it in this work.

As we would like to compute $\Sigma^< = \Sigma^{12}_{\rm NLO}$, we set the external contour indices to $a = 2$ and $b = 1$. Then, the contributions $\mathrm{Tr}\left[\not{p}\Sigma^{12,21}\right]$ and $\mathrm{Tr}\left[\not{p}\Sigma^{12,12}\right]$, which correspond to cutting both the photon and neutrino lines, vanish by momentum conservation (see also footnote 8). Of the remaining "11" and "22" contributions, the transformation behaviour of the thermal propagators under hermitian conjugation dictates that $\mathrm{Tr}\left[\not{p}\Sigma^{12,11}\right] = \mathrm{Tr}\left[\not{p}\Sigma^{12,22}\right]^*$. It then follows that

$$\mathrm{Tr}\left[\not{p}\Sigma^{12}_{\rm NLO}\right] = 2\mathrm{Re}\,\mathrm{Tr}\left[\not{p}\Sigma^{12,11}\right], \tag{3.6}$$

i.e., the Wightman self-energy represented by figure 2 can be determined entirely through the diagonal "11" contribution.

It is instructive to recast the expressions for the self-energy (3.5)–(3.6) into the form of a standard Boltzmann collision integral commonly found in textbooks (e.g., [50]). To do so, we first identify the 4-momenta p, l, q in figure 2 with the momenta p_1, p_2, p_3, p_4 of the external neutrinos and electrons of the underlying $2 \to 2$ process represented by diagram (d) of figure 1 via

$$p_1 = p, \qquad p_2 = -l, \qquad p_3 = q + p - l, \qquad p_4 = -q. \tag{3.7}$$

Then, writing out the propagators $iS_e^{12/21}$ and $iS_\nu^{12/21}$ explicitly, the self-energy (3.5)–(3.6) can be brought into the form

$$\mathrm{Tr}\left[\not{p}\Sigma^{12}_{\rm NLO}\right] = -(2\pi)^3 \int_{q,l} \delta(p_2^2)\delta(p_3^2 - m_e^2)\delta(p_4^2 - m_e^2)\mathcal{F}(p_2^0, p_3^0, p_4^0)\mathcal{T}_{\rm NLO}, \tag{3.8}$$

where the population factor

$$\mathcal{F}(p_2^0, p_3^0, p_4^0) = \left[f_F(|p_2^0|) - \Theta(p_2^0)\right]\left[f_F(|p_3^0|) - \Theta(-p_3^0)\right]\left[f_F(|p_4^0|) - \Theta(-p_4^0)\right] \tag{3.9}$$

contains the equilibrium phase space distributions of the three integrated external particles and is analogous to $\mathcal{F}^{\gtrless}$ in equation (2.5). The quantity $\mathcal{T}_{\rm NLO}$ plays the role of QED corrections to the LO squared matrix element,[9] and can be written in terms of the one-loop photon self-energy $\Pi^{\mu\nu}_{ab}$ as

$$\begin{aligned}\mathcal{T}_{\rm NLO} = -16G_F^2 g_V^\alpha \mathrm{Re}\Big\{&\mathrm{Tr}\left[(\not{p}_3 + m_e)\gamma^\mu\,(m_e - \not{p}_4)\gamma^\rho\,(g_L^\alpha P_L + g_R P_R)\right] \\ &\times \mathrm{Tr}\left[\not{p}_1\gamma_\rho P_L \not{p}_2 \gamma_\sigma P_L\right] D^{11}_{\mu\nu}(P)\Pi^{\nu\sigma}_{11}(P)\Big\},\end{aligned} \tag{3.10}$$

where we have introduced the photon momentum $P = p_1 + p_2$.

Since $\mathcal{T}_{\rm NLO}$ has the interpretation of a squared matrix element, we can split it into a vacuum and a thermal part according to temperature dependence,

$$\mathcal{T}_{\rm NLO} = \mathcal{T}_{\rm vac} + \mathcal{T}_{\rm th}. \tag{3.11}$$

The vacuum correction $\mathcal{T}_{\rm vac}$ has no intrinsic temperature dependence in the sense that it makes no explicit reference to the temperature or to any phase space distribution. It is simply the correction to the weak matrix elements arising from the interference of the closed

[9] For completeness, $\mathcal{T}_{\rm LO} = 2^7 G_F^2\left[(g_L^\alpha)^2\,(p_1\cdot p_4)^2 + g_R^2\,(p_2\cdot p_4)^2 + g_L^\alpha g_R m_e^2\,(p_1\cdot p_2)\right]$. If we were to replace $\mathcal{T}_{\rm NLO}$ with $\mathcal{T}_{\rm LO}$ in equation (3.8), we would recover the leading-order equilibrium neutrino production rates.

fermion loop diagram (d) with the LO graph in standard $T = 0$ quantum field theory, and can be expressed in terms of the vacuum photon self-energy as

$$\mathcal{T}_{\rm vac} = -2^{10} G_F^2 \, \alpha_{\rm em} \left[(g_V^\alpha)^2/(4\pi)\right] \left[m_e^2 (p_1 \cdot p_2) + 2(p_1 \cdot p_4)^2\right] \mathrm{Re}\, \Pi_2(P^2) \tag{3.12}$$

where $\alpha_{\rm em}$ is the electromagnetic fine-structure constant, and the form factor Π_2 is defined in appendix C; see equation (C.7). The simplicity of the expression follows from the fact that the integration domain is symmetric under the interchange $p_3 \leftrightarrow p_4$. This symmetry, along with momentum conservation, also ensures the absence of all antisymmetric terms containing Levi-Civita symbols arising from traces of γ^5 with four or more γ-matrices. Vacuum QED corrections to the neutrino-electron interaction matrix element were previously computed in reference [33].

The thermal correction $\mathcal{T}_{\rm th}$, on the other hand, depends explicitly on equilibrium phase space distributions, f_F or f_B, of the internal particles ("F" for Fermi-Dirac; "B" for Bose-Einstein), and can be thought of as a temperature-dependent correction to the squared matrix element. This T-dependence originates in the thermal part of the "11" propagator, which is proportional to $\delta(p^2 - m^2) f_{F/B}(|p^0|)$ (see appendix A) and, where it is applied, effectively puts an internal line of the closed fermion loop diagram (d) in figure 1 on-shell. Purely from counting, there are altogether seven possible ways to put one, two, or all three internal lines of diagram (d) on-shell. Not all combinations contribute to $\mathcal{T}_{\rm NLO}$: Terms proportional to f_B correspond to putting the photon line on-shell, which are forbidden for diagram (d) for kinematic reasons [49, 51]. Similarly, putting both internal electrons on-shell leads to a purely imaginary contribution that is irrelevant to the real part of the self-energy we wish to compute. The only surviving two combinations correspond to putting either internal electron line on-shell, and are proportional to $f_F(|k^0|)$ and $f_F(|k^0 + P^0|)$ respectively.

Observe that $\mathcal{T}_{\rm NLO}$ contains a t-channel contribution from elastic $\nu_\alpha e$ scattering (i.e., $p_2^0 < 0$) that is logarithmically divergent for soft photon momenta. This divergence comes from the fact that the finite-temperature photon self-energy scales not as P^2 like in vacuum, but as T^2 in the hard-thermal loop limit which do not compensate anymore for the $1/P^2$ behaviour of the photon propagator. In addition, soft photons are Bose-enhanced. To remedy the problem, we replace the tree-level photon propagator $D_{\mu\nu}^{cd}$ in equation (3.10) with the fully-resummed photon propagator $\bar{D}_{\mu\nu}^{ab}$.[10] Furthermore, because both $\bar{D}_{\mu\nu}^{ab}$ and $\Pi_{11}^{\nu\sigma}$ split into a longitudinal ("L") and a transverse ("T") part, the same decomposition applies also to $\mathcal{T}_{\rm th}$, i.e., $\mathcal{T}_{\rm th} = \mathcal{T}_{\rm th}^L + \mathcal{T}_{\rm th}^T$, where $\mathcal{T}_{\rm th}^{L,T}$ can be brought into the form

$$\begin{aligned}\mathcal{T}_{\rm th}^{L/T} = -2^8 G_F^2 (g_V^\alpha)^2 \mathrm{Re}\, \bar{D}_R^{L/T}(P) \mathrm{Re}\, \Pi_{R,T\neq 0}^{L/T}(P) \Big[& 2(p_1 \cdot p_4) P_{L/T}^{1,4} \\ & + (p_1 \cdot p_2)(a_{L/T}(p_1 \cdot p_2) + P_{L,T}^{2,2} + P_{L/T}^{4,4}) \Big] . \end{aligned} \tag{3.13}$$

Here, $\bar{D}_R$ denotes the retarded resummed photon propagator; $\Pi_{R,T\neq 0}$ is the retarded thermal photon self-energy comprising the $f_F(|k^0|)$ and $f_F(|k^0 + P^0|)$ terms described above; we have employed the shorthand notation $P_{L/T}^{i,j} = P_{L/T}^{\mu\nu} p_{i,\mu} p_{j,\nu}$; and $a_{L,T} = (3 \mp 1)/4$ further differentiates between the longitudinal and the transverse contribution.

Note that the imaginary part of the photon propagator, $\mathrm{Im}\, \bar{D}_{11}^{L,T}$, does not appear in equation (3.13) because it is formally of higher-order in $\alpha_{\rm em}$ and we are only interested in the

[10] In principle, it is also necessary to replace the tree-level photon propagator in equation (3.12) with the fully resummed one. However, since the vacuum correction is not dominated by small photon momenta, the choice of photon propagator has in this case a negligible impact on the numerical outcome.

$\mathcal{O}(\alpha_{\rm em})$ corrections. We also only use the resummed photon propagator in the IR divergent t-channel contribution; where the divergence is absent, i.e., in the s-channel, we set $\bar{D}_R \to D_R$, where D_R is the un-resummed counterpart of $\bar{D}_R$. The final expressions for $\mathrm{Re}\Pi^{L/T}_{11,T\neq 0}$ and $\mathrm{Re}\,\bar{D}^{L/T}_{11}$ are given in equations (C.10), (C.12) and (C.32), which can be easily mapped to $\mathrm{Re}\,\Pi^{L/T}_{R,T\neq 0}$ and $\mathrm{Re}\,\bar{D}_R$ via $\mathrm{Re}\,\bar{D}^{L/T}_{R} = \mathrm{Re}\,\bar{D}^{L/T}_{11}$ and $\mathrm{Re}\,\Pi^{L/T}_{R} = \mathrm{Re}\,\Pi^{L/T}_{11}$. Details of their derivation, along with the relevant thermal integrals, can be found in appendices B and C.

3.2 Numerical results for NLO neutrino interaction rate

We evaluate the self-energy correction (3.8) and hence the correction to the neutrino damping rate (3.4) by numerically integrating over the two remaining momenta l and q in (3.8) using the parameterisation shown in appendix D. We adopt the experimentally-determined values given by the Particle Data Group [52] for the following input parameters:

- Fermi's constant: $G_F = 1.166\,378\,8(6)\,\mathrm{MeV}^{-2}$,
- Electron mass: $m_e = 0.510\,998\,950\,00(15)\,\mathrm{MeV}$,
- Electromagnetic fine-structure constant: $\alpha^{-1}_{\rm em}(0) = 137.035\,999\,180(10)$, and
- Weinberg angle: $\sin^2\theta_W(0)_{\overline{\rm MS}} = 0.238\,63(5)$.

The renormalisation scale μ_R appearing in the photon self-energy of the vacuum contribution is identified with the electron mass, $\mu_R = m_e$.

Figure 3 shows the closed fermion loop corrections to the damping rates $\Gamma_e(p)$ and $\Gamma_{\mu,\tau}(p)$ at the mean momentum $p = 3.15T$. Relative to their respective LO rates, we find the corrections at temperatures $T \sim 1 \to 3$ MeV to fall in the range $-0.2 \to +0.1\%$ and $-0.0005 \to +0.0002\%$, respectively, for ν_e and $\nu_{\mu,\tau}$. We further note that:

1. At $T \sim 2\,\mathrm{MeV}$, the vacuum and the thermal contributions to the correction are roughly equal in magnitude (green vs blue lines in figure 3), in contrast to the findings of [30], where finite-temperature corrections were determined to be subdominant. We note however that a direct comparison is not possible because a different set of diagrams was investigated in [30]—(a), (b), and (c) in figure 1—as opposed to the type (d) corrections considered in this work.

2. Reference [30] also found no significant flavour dependence in the rate corrections: their Γ_e and $\Gamma_{\mu,\tau}$ corrections differ by less than 1% in the temperature regime around neutrino decoupling. Our corrections are on the other hand strongly flavour-dependent—by more than two orders of magnitude—and trace their origin to the fact that electron neutrinos experience charge current interactions while the muon- and tau-flavoured neutrinos interact only via the neutral current with the $e^\pm$ thermal bath. This difference renders the corresponding vector couplings, $g^e_V \sim 0.49$ and $g^{\mu,\tau}_V \sim -0.012$, very roughly two orders of magnitude apart from one another. This strong flavour dependence in the rate corrections was also seen in reference [32].

3. We have computed the NLO contributions to the interaction rates in two different approximations: (i) retaining the full dependence on the electron mass (solid lines in figure 3), and (ii) in the limit $m_e \to 0$ (dashed lines). The massless calculation aims to quantify the effect of the $m_e = 0$ approximation used in reference [32], along with the Hard Thermal Loop (HTL) approximation of the photon propagator.

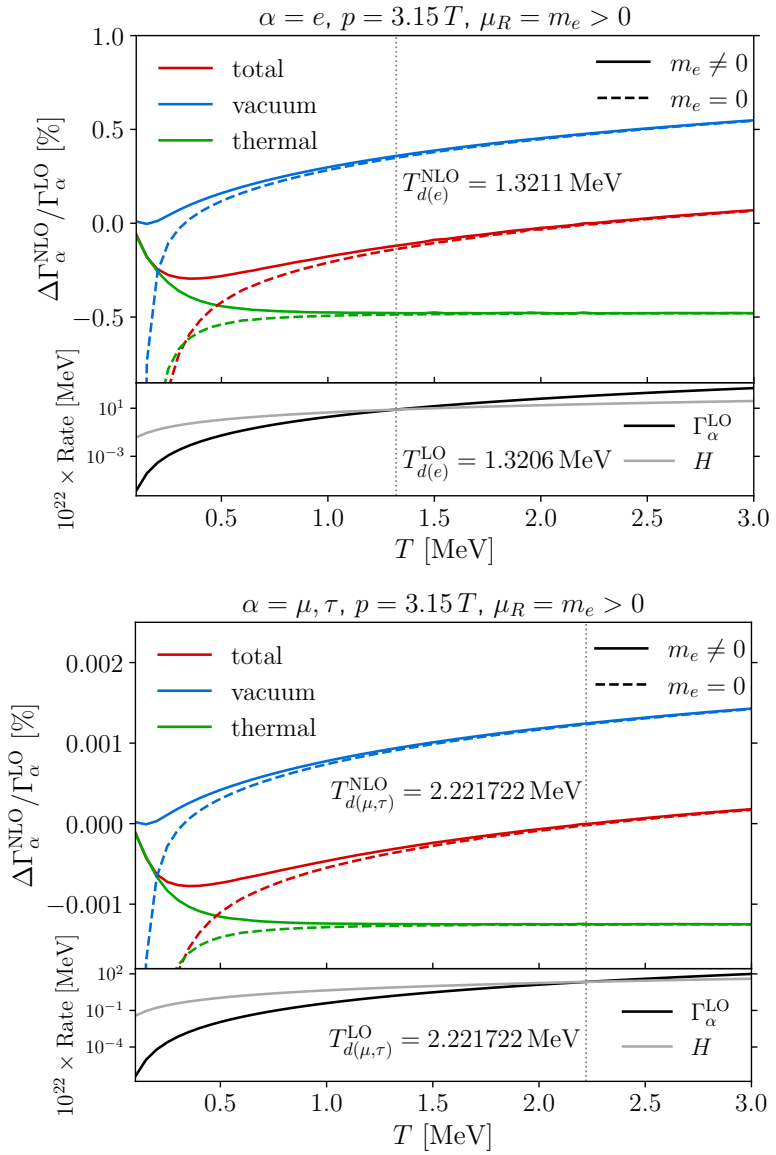


Figure 3: *Top*: NLO contributions to the ν_e interaction rate from the closed fermion loop with a finite electron mass (solid) or $m_e = 0$ (dashed) at different temperatures for the mean neutrino momentum $p = 3.15T$. For comparison we normalise all curves to the LO rate (without QED corrections), which we always evaluate with $m_e \neq 0$ to ensure a common normalisation. The total rate correction (red) is further split into the vacuum (blue) and the thermal contribution (green). At $T \gg m_e$, the green curve flattens out as the thermal correction contains no scale besides the temperature in this region, while the vacuum correction retains a mild dependence on the renormalisation scale μ_R. The lower panel shows the LO neutrino interaction rate compared to the Hubble rate, and T_d indicates the decoupling temperature, defined via $\Gamma_\alpha(T_{d(e)}) = H(T_{d(e)})$. *Bottom*: Same as top, but for $\nu_{\mu,\tau}$.

We observe that the error from neglecting m_e is relatively minor for $T \gtrsim 3m_e$, but becomes sizeable at low temperatures. In particular, in the limit $T \to 0$ the ra-

tio $\Gamma_\alpha^{\rm NLO}/\Gamma_\alpha^{\rm LO}$ vanishes for $m_e > 0$, but diverges for $m_e = 0$ because of Boltzmann suppression of the LO rates. Precision computations of $N_{\rm eff}^{\rm SM}$ track the evolution of neutrinos down to temperatures much below m_e [28, 29]. Thus, although it is commonly understood that (electron) neutrino decoupling occurs at relativistic temperatures $T \sim (2 \to 3) \times m_e$, what we assume for m_e in the NLO rate computations may yet have some impact on $N_{\rm eff}$ (see section 4.2).

Figure 4 focuses on the t-channel contribution to the interaction rate, where the enhancement near the photon mass-shell occurs. We show four sets of results, based on the resummed photon propagator detailed in appendix C: (i) the complete one-loop result including a finite m_e everywhere, (ii) the complete one-loop result in the limit $m_e \to 0$, (iii) using the HTL photon propagator (which does not depend on the electron mass), but with m_e everywhere else,[11] and (iv) using the HTL photon propagator and setting $m_e = 0$ everywhere. As expected, we observe that (ii) and (iv) match to a very good approximation. Indeed, since the scattering rates are dominated by the kinematic region around the t-channel singularity where photons are soft, the 1PI-resummed propagator is well-approximated by the HTL one when, in addition, we set $m_e = 0$. On the other hand, visible differences can be discerned between (i) and (iii) at $T \lesssim 1$ MeV, which can be explained by the fact that the HTL approximation only holds for $T \gg m_e$. In the lower panel of figure 4, we highlight the impact of m_e by displaying the ratio of (i) to (iv).

3.3 NLO decoupling temperatures

The lower panels of the two plots in figure 3 show the LO electron neutrino and muon/tau neutrino interaction rates juxtaposed with the Hubble expansion rate. The latter is given by $H^2(T) = (\rho_{\rm QED} + \rho_\nu)/(3M_{\rm Pl}^2)$, where $M_{\rm Pl} = 2.435\,36 \times 10^{21}\,\text{MeV}$ is the reduced the Planck mass, $\rho_\nu = 3g_\nu(7/8)(\pi^2/30)T^4$ is the energy density in three families of neutrinos with $g_\nu = 2$, and

$$\rho_{\rm QED} = g_\gamma \frac{\pi^2}{30} T^4 + \frac{g_e}{2\pi^2} \int_0^\infty \mathrm{d}p\, p^2 E_e f_F(E_e) \tag{3.14}$$

is the energy density in the QED plasma assuming an ideal gas configuration, with $g_\gamma = 2$, $g_e = 4$, and $E_e = \sqrt{p^2 + m_e^2}$. By solving the equation $\Gamma_\alpha^{\rm LO}(T_{d(\alpha)}) = H(T_{d(\alpha)})$ for the flavour-dependent decoupling temperatures $T_{d(\alpha)}$, we determine the LO decoupling temperatures to be

$$\begin{aligned} T_{d(e)}^{\rm LO} &\simeq 1.320\,58\,\text{MeV}, \\ T_{d(\mu,\tau)}^{\rm LO} &\simeq 2.221\,722\,09\,\text{MeV}. \end{aligned} \tag{3.15}$$

The first number differs from the earlier estimate of $T_{d(e)}^{\rm LO} \simeq 1.3453$ MeV from reference [21], which can be traced to a different choice of the Weinberg angle. Incorporating QED corrections to the damping rates, the decoupling temperatures shift to

$$\begin{aligned} T_{d(e)}^{\rm NLO} &\simeq 1.321\,10\,\text{MeV}, \\ T_{d(\mu,\tau)}^{\rm NLO} &\simeq 2.221\,722\,11\,\text{MeV} \end{aligned} \tag{3.16}$$

[11] Even though it is physically inconsistent to keep m_e in the electron propagators and not in the photon one, (iii) is nonetheless helpful as it isolates the dependence of the rate corrections to the 1PI-resummed vs the HTL propagator.

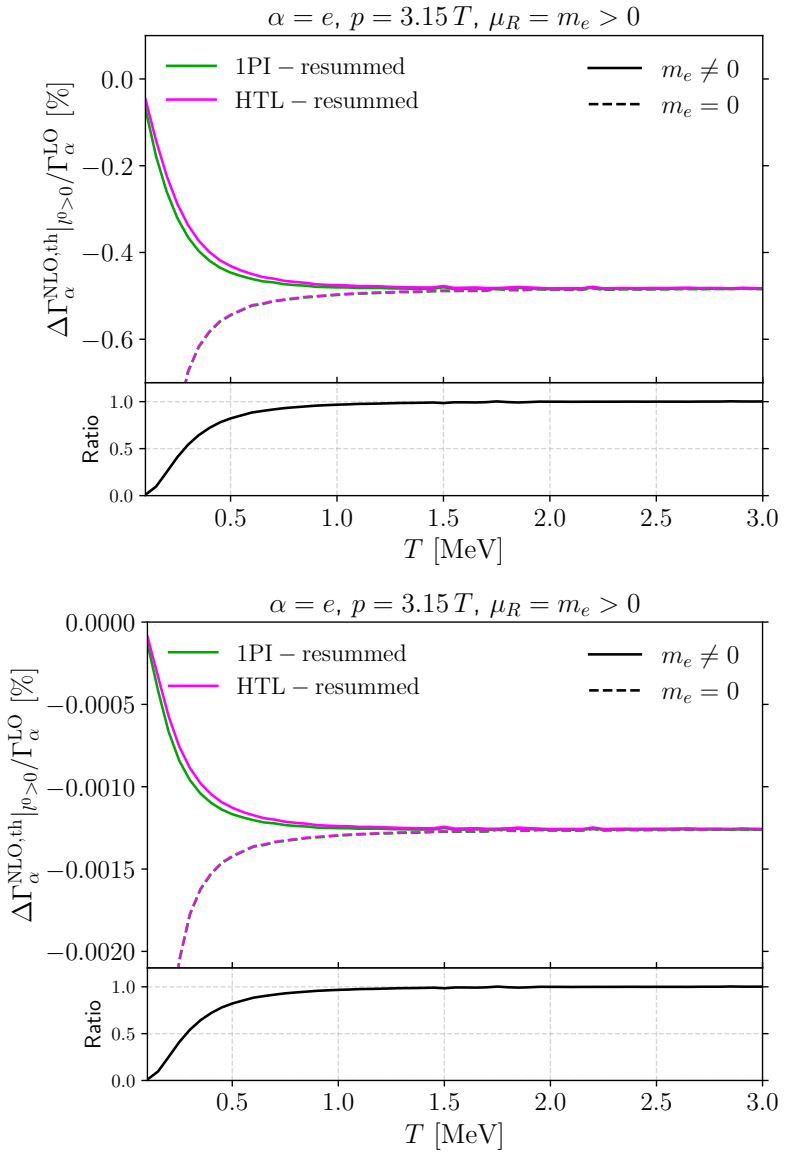


Figure 4: *Top:* The t-channel contribution ($l^0 > 0$) to the NLO electron neutrino interaction rate at the mean momentum $p = 3.15T$ in different approximations, namely, using 1PI-resummed (green) and HTL (magenta) photon propagators, in each case with $m_e \neq 0$ (solid lines) or $m_e = 0$ (dashed lines) in both electron loops of the self-energy diagram of figure 2. As in figure 3, we normalise all curves to the LO rate, which we always evaluate with $m_e \neq 0$. The lower panel shows the ratio of the 1PI result for $m_e \neq 0$ to the HTL result for $m_e = 0$ as a function of the temperature. *Bottom*: Same as top, but for $\alpha \neq e$.

corresponding to an increase of $\sim 0.04\,\%$ for ν_e and of $\sim 8 \times 10^{-7}\,\%$ for $\nu_{\mu,\tau}$. Around the muon neutrino decoupling temperature, the vacuum and thermal contributions to the NLO rates approximately cancel, explaining the smallness of the correction to $T_{d(\mu,\tau)}^{\mathrm{NLO}}$.

Given that [32] computed the NLO weak rates assuming $m_e = 0$, it is also of interest to study how such an assumption modifies the decoupling temperature shifts. Using our

Method	$\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$	$\delta N_{\rm eff}^{m_e=0}/N_{\rm eff}^{\rm LO}$
Entropy conservation (common decoupling)	-1.1×10^{-5}	-1.2×10^{-5}
Entropy conservation (flavour-dependent decoupling)	-5.4×10^{-6}	-6.1×10^{-6}
Boltzmann damping approximation (mean momentum)	-7.8×10^{-6}	-1.6×10^{-5}
Boltzmann damping approximation (full momentum)	-7.9×10^{-6}	-2.6×10^{-5}

Table 2: Estimates of the relative correction to $N_{\rm eff}^{\rm SM}$ due to NLO weak rate corrections, with and without the electron mass, using different methods.

computations of the $m_e = 0$ rate corrections (but with $m_e \neq 0$ LO rates), we find the corresponding NLO decoupling temperatures to be

$$\begin{aligned} T_{d(e)}^{\rm NLO,m_e=0} &\simeq 1.321\,18\,\text{MeV}, \\ T_{d(\mu,\tau)}^{\rm NLO,m_e=0} &\simeq 2.221\,722\,23\,\text{MeV}, \end{aligned} \tag{3.17}$$

i.e., a 0.05 % and 6×10^{-6} % shift for ν_e and $\nu_{\mu,\tau}$, respectively, relative to their corresponding LO decoupling temperature (3.15).

4 NLO effects on $N_{\rm eff}^{\rm SM}$

Having computed the closed fermion loop correction to the damping rate Γ_α, we are now in a position to estimate its effect on $N_{\rm eff}^{\rm SM}$. Within the damping approximation, two avenues are available to us:

- Given Γ_α at a representative momentum, we have already estimated in equation (3.16) the corresponding correction to the neutrino decoupling temperature T_d, defined via $\Gamma_\alpha(T_d) = H(T_d)$. Then, under the assumption of instantaneous neutrino decoupling, an estimate of the change to $N_{\rm eff}^{\rm SM}$, $\delta N_{\rm eff} \equiv N_{\rm eff}^{\rm NLO} - N_{\rm eff}^{\rm LO}$, can be obtained via entropy conservation arguments.

- We may also compute $\delta N_{\rm eff}$ by solving directly the continuity equation (2.2) and the QKEs (2.3) in the damping approximation (2.6) and neglecting neutrino oscillations.

We consider both approaches in the following. The corresponding estimates for $\delta N_{\rm eff}$ are summarised in table 2.

4.1 Entropy conservation

The entropy conservation argument posits that entropies in a comoving volume in two decoupled sectors are separately conserved, i.e.,

$$\begin{aligned} s(a_1)a_1^3 &= s(a_2)a_2^3, \\ s_{\nu_\alpha}(a_1)a_1^3 &= s_{\nu_\alpha}(a_2)a_2^3, \end{aligned} \tag{4.1}$$

where s_{ν_α} and $s \equiv s_\gamma + s_e + \sum_{\beta\neq\alpha} s_{\nu_\beta} + \cdots$ denote, respectively the entropy density of a decoupled neutrino species ν_α and of the QED plasma plus any neutrino species ν_β that may still be be in equilibrium with it, and the scale factors $a_{1,2}$ represent two different times

after decoupling. We take a_1 to correspond to the time immediately after ν_α decouples instantaneously from the QED plasma, i.e., $a_1 = a_d^+$, where the sector temperatures satisfy $T(a_1) = T_{\nu_\alpha}(a_1) \equiv T_d$, while $a_2 > a_1$ is some later time to be specified below.

The assumption of instantaneous decoupling allows the neutrinos to maintain to an excellent approximation an ideal-gas equilibrium phase space distribution at all times, so that $s_{\nu_\alpha}(a) \propto T^3_{\nu_\alpha}(a)$. It then follows simply from (4.1) that $T_{\nu_\alpha}(a_2) = (a_1/a_2)T_{\nu_\alpha}(a_1) = (a_1/a_2)T_d$. In the case of the QED+ν_β plasma, we parameterise its entropy density as

$$s(a) = \frac{2\pi^2}{45}\tilde{g}_s(a)T^3(a), \tag{4.2}$$

where the entropy degree of freedom parameter $\tilde{g}_s$ is given in the ideal gas limit by

$$\tilde{g}_s(a) \equiv g_\gamma + \frac{45}{4\pi^4}\frac{g_e}{T^4(a)}\int_0^\infty \mathrm{d}p\, p^2 \left(E_e + \frac{p^2}{3E_e}\right) f_F(E_e, T(a)) + \frac{7}{8}\sum_{\beta\neq\alpha} g_{\nu_\beta}, \tag{4.3}$$

with $g_\gamma = 2$, $g_e = 4$, and $g_{\nu_\beta} = 2$. For our current purpose of estimating δN_{eff} due to NLO contributions to the weak rates, it suffices to use the ideal-gas $\tilde{g}_s$. We do note however that QED entropy density at $T_d \sim \mathcal{O}(1)$ MeV is subject in principle to a sizeable finite-temperature correction to the QED equation of state, which needs to be included in precision $N_{\mathrm{eff}}^{\mathrm{SM}}$ calculations. See, e.g., reference [21] for details.

Then, combining the above, we find an estimate of the neutrino-to-photon temperature ratio of

$$\frac{T_{\nu_\alpha}(a_2)}{T(a_2)} = \left[\frac{\tilde{g}_s(a_2)}{\tilde{g}_s(a_d^+)}\right]^{1/3} \tag{4.4}$$

at the later time a_2. We use this temperature ratio in the following to provide two estimates of δN_{eff} due to the rate corrections.

Common decoupling temperature In the first estimate, we assume all neutrino flavours to decouple effectively at the same time, an assumption that may to some extent be justified by the observed large mixing in the neutrino sector. We choose $a_1 = a_d^+$ to correspond to the time immediately after ν_e decoupling, and set a_2 at a time significantly after $e^\pm$ annihilation, where $T(a_2) \ll m_e$. The latter leads immediately to an entropy degree of freedom of $\tilde{g}_s(a_2) = g_\gamma = 2$, while for the former we have

$$\tilde{g}_s(a_d^+) = 2 + \frac{45}{\pi^4 T^4_{d(e)}}\int_0^\infty \mathrm{d}p\, p^2 \left(E_e + \frac{p^2}{3E_e}\right) f_F(E_e, T_{d(e)}). \tag{4.5}$$

Then, using the temperature ratio (4.4) and the ideal-gas relations $\rho_\gamma \propto g_\gamma T^4$ and $\rho_{\nu_\alpha} \propto (7/8)g_{\nu_\alpha}T^4_{\nu_\alpha}$, we find

$$\left.\frac{\rho_\nu}{\rho_\gamma}\right|_{T/m_e\to 0} = \sum_\alpha \frac{\rho_{\nu_\alpha}(a_2)}{\rho_\gamma(a_2)} = 3\times\frac{7}{8}\left[\frac{2}{\tilde{g}_s(a_d^+)}\right]^{4/3}, \tag{4.6}$$

or, equivalently in terms of N_{eff},

$$N_{\mathrm{eff}} = 3\times\left[\frac{11}{4}\frac{2}{\tilde{g}_s(a_d^+)}\right]^{4/3}, \tag{4.7}$$

where the factor 4/11 corresponds to $2/\tilde{g}_s(T_d)$ evaluated in the limit $T_d \gg m_e$.

Using the LO and NLO electron neutrino decoupling temperatures in (3.15) and (3.16) respectively, we find a fractional shift of $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO} \simeq -1.1\times10^{-5}$ due to the rate corrections. Had we instead used the $m_e = 0$ NLO decoupling temperatures (3.17), the correspond shift in $N_{\rm eff}$ would have been $\delta N_{\rm eff}^{m_e=0} \simeq -3.8\times10^{-5}$, or $\delta N_{\rm eff}^{m_e=0}/N_{\rm eff}^{\rm LO} \simeq -1.2\times10^{-5}$. Thus, while setting $m_e = 0$ leads to a $\sim 10\%$ change in the estimate of $\delta N_{\rm eff}$, its ultimate impact on $N_{\rm eff}^{\rm SM}$ appears to be small, at least within the entropy conservation picture. These results are summarised in table 2.

Flavour-dependent decoupling In the absence of neutrino oscillations, ν_e and $\nu_{\mu,\tau}$ decouple at different temperatures, $T_{d(e)}$ and $T_{d(\mu,\tau)} > T_{d(e)}$. Then, to estimate $\delta N_{\rm eff}$ requires that we consider entropy conservation across four epochs: the time immediately after $\nu_{\mu,\tau}$ decoupling $a_1 = a^+_{d(\mu,\tau)}$; immediately *before* ν_e decoupling $a_2 = a^-_{d(e)}$; immediately *after* ν_e decoupling $a_3 = a^+_{d(e)}$; and a_4 is a time significantly after $e^\pm$ annihilation. The corresponding entropy degrees of freedom are

$$\begin{aligned}
\tilde{g}_s(a^+_{d(\mu,\tau)}) &= 2 + \frac{45}{\pi^4 T^4_{d(\mu,\tau)}}\int_0^\infty {\rm d}p\, p^2\left(E_e + \frac{p^2}{3E_e}\right) f_F(E_e, T_{d(\mu,\tau)}) + \frac{7}{8}g_{\nu_e},\\
\tilde{g}_s(a^-_{d(e)}) &= 2 + \frac{45}{\pi^4 T^4_{d(e)}}\int_0^\infty {\rm d}p\, p^2\left(E_e + \frac{p^2}{3E_e}\right) f_F(E_e, T_{d(e)}) + \frac{7}{8}g_{\nu_e},\\
\tilde{g}_s(a^+_{d(e)}) &= 2 + \frac{45}{\pi^4 T^4_{d(e)}}\int_0^\infty {\rm d}p\, p^2\left(E_e + \frac{p^2}{3E_e}\right) f_F(E_e, T_{d(e)}),
\end{aligned} \tag{4.8}$$

and $\tilde{g}_s(a_4) = 2$. An estimate of the ν_e-to-photon energy density ratio at a_4 follows straightforwardly from the temperature ratio (4.4) and ideal-gas temperature-energy relations:

$$\frac{\rho_{\nu_e}(a_4)}{\rho_\gamma(a_4)} = \frac{7}{8}\left[\frac{\tilde{g}_s(a_4)}{\tilde{g}_s(a^+_{d(e)})}\right]^{4/3} = \frac{7}{8}\left[\frac{2}{\tilde{g}_s(a^+_{d(e)})}\right]^{4/3}. \tag{4.9}$$

For $\rho_{\nu_{\mu,\tau}}(a_4)/\rho_\gamma(a_4)$, we note that ν_e decoupling at $a_{d(e)}$ introduces a discontinuity in $\tilde{g}_s$, thereby leading to a more complicated energy density ratio at a_4,

$$\frac{\rho_{\nu_{\mu,\tau}}(a_4)}{\rho_\gamma(a_4)} = \frac{7}{8}\left[\frac{2}{\tilde{g}_s(a^+_{d(e)})}\frac{\tilde{g}_s(a^-_{d(e)})}{\tilde{g}_s(a^+_{d(\mu,\tau)})}\right]^{4/3}. \tag{4.10}$$

Then, combining equations (4.9) and (4.10) to form $\rho_\nu = \sum_\alpha \rho_{\nu_\alpha}$, we find

$$N_{\rm eff} = \left(\frac{11}{4}\frac{2}{\tilde{g}_s(a^+_{d(e)})}\right)^{4/3}\left[1 + 2\times\left(\frac{\tilde{g}_s(a^-_{d(e)})}{\tilde{g}_s(a^+_{d(\mu,\tau)})}\right)^{4/3}\right] \tag{4.11}$$

for our estimate.

Evaluating (4.11) for the LO decoupling temperatures (3.15), we find $N_{\rm eff}^{\rm LO} \simeq 3.027867$. Including QED corrections to the interaction rates leads to a shift of $\delta N_{\rm eff} \simeq -1.6\times10^{-5}$ (or $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO} \simeq -5.4\times10^{-6}$) if we use the $m_e \neq 0$ corrections, or $\delta N_{\rm eff} \simeq -1.9\times10^{-5}$ ($\delta N_{\rm eff}^{m_e=0}/N_{\rm eff}^{\rm LO} \simeq -6.1\times10^{-6}$ using the $m_e = 0$ corrections. Compared with the common-decoupling estimates displayed in table 2, we note that the flavour-dependent decoupling

estimates of $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$ are generally about a factor of two smaller, in both the $m_e = 0$ and $m_e \neq 0$ cases. This difference is to be expected, as the QED corrections to the $\nu_{\mu,\tau}$ interaction rates are negligible compared with the corrections to the ν_e rates. We emphasise, however, both estimates are very crude approximations: the true shift in $\delta N_{\rm eff}$ will probably fall somewhere in-between.

4.2 Solving the neutrino Boltzmann equations and the continuity equation

Following reference [20], we introduce the comoving quantities $x \equiv m_e\, a$, $y \equiv a\, p$, and $z \equiv T\, a$, and rewrite the continuity equation (2.2) as a differential equation for the quantity z (corresponding to the photon temperature):

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\frac{x}{z}J(x/z) - \frac{1}{2z^3}\frac{\mathrm{d}\bar{\rho}_\nu}{\mathrm{d}x} + G_1(x/z)}{\frac{x^2}{z^2}J(x/z) + Y(x/z) + \frac{2\pi^2}{15} + G_2(x/z)}\,, \tag{4.12}$$

where $J(x/z)$ and $Y(x/z)$ describe the ideal-gas behaviour of the QED plasma, while the $G_{1,2}(\tau)$ account for deviations of the QED plasma from an idea gas. Explicit forms for these expression to $\mathcal{O}(e^2)$ can be found in appendix E.

Equation (4.12) also requires as input the total time derivative of the comoving neutrino density $\bar{\rho}_\nu$, which can generally be constructed from the neutrino density matrices via

$$\frac{\mathrm{d}\bar{\rho}_\nu}{\mathrm{d}x} = \frac{1}{\pi^2}\int \mathrm{d}y\; y^3 \sum_\alpha \frac{\mathrm{d}\{\varrho(x,y)\}_{\alpha\alpha}}{\mathrm{d}x}, \tag{4.13}$$

where $\mathrm{d}\{\varrho(x,y)\}_{\alpha\alpha}/\mathrm{d}x$ corresponds in general to the QKEs (2.3). Neglecting neutrino oscillations, the density matrices ϱ are diagonal in the flavour basis, in which case the QKEs (2.3) simplify to a set of classical Boltzmann equations. Then, together with the damping approximation (2.6) and in terms of the new variables, we find

$$\frac{\mathrm{d}\{\varrho(x,y)\}_{\alpha\alpha}}{\mathrm{d}x} \simeq -\frac{\Gamma_\alpha(x,y)}{xH(x)}\Big[\{\varrho(x,y)\}_{\alpha\alpha} - f_F((y/z)T)\Big]. \tag{4.14}$$

Equation (4.14) can be solved together with the continuity equation (4.12) for a range of momenta y covering the bulk of the neutrino population (a typical range would be $y \in [0.01, 30]$). We call this the "full-momentum" approach. Alternatively, we can simplify the equation (4.14) further by adopting the *ansatz*

$$\{\varrho(y)\}_{\alpha\alpha} = \{\varrho(\langle y\rangle)\}_{\alpha\alpha}\frac{f_F(yT)}{f_F(\langle y\rangle T)}, \tag{4.15}$$

and $\Gamma_\alpha(y) = \Gamma_\alpha(\langle y\rangle)$. Identifying $\langle y\rangle$ with the mean momentum mode $y_0 = 3.15z(T_0)$, where T_0 is the photon temperature at initialisation (which is equal to the neutrino temperature), we can rewrite equation (4.13) in this approximation as

$$\frac{\mathrm{d}\bar{\rho}_\nu}{\mathrm{d}x} = -\frac{7\pi^2}{120xH(x)}\left[\sum_\alpha \Gamma_\alpha(\langle y\rangle)\left(\frac{\{\varrho(x,\langle y\rangle)\}_{\alpha\alpha}}{f_F(\langle y\rangle T)} - z^3(x)\right)\right]. \tag{4.16}$$

We call this alternative the "mean-momentum" approach.

Irrespective of whether we use the full-momentum or the mean-momentum approach. The final $N_{\rm eff}$ can be estimated from the solutions to ρ_α and z at $x \to \infty$ using the definition of $N_{\rm eff}^{\rm SM}$ in equation (2.1), reproduced here in terms of the rescaled variables:

$$N_{\rm eff} = \frac{8}{7}\left(\frac{11}{4}\right)^{4/3} \frac{30}{2\pi^2} \left[\frac{z(x=0)}{z(x)}\right]^4 \bar{\rho}_\nu(x)\bigg|_{x\to\infty} . \tag{4.17}$$

Table 2 shows our estimates of $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$ due to QED corrections to the neutrino interaction rates using both the full-momentum and the mean-momentum approaches, with or without the electron mass in the correction.

Evidently from table 2, both the full-momentum and the mean-momentum estimates for $m_e \neq 0$, $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO} \simeq (-7.8 \to -7.9)\times 10^{-6}$, are broadly consistent with their counterpart obtained in section 4.1 from entropy arguments: in fact they sit between the common-decoupling and flavour-dependent decoupling estimates from entropy considerations. In contrast, the estimates assuming $m_e = 0$ differ by $\sim 50\%$ between the full-momentum ($\delta N_{\rm eff}^{m_e=0}/N_{\rm eff}^{\rm LO} \simeq -2.6\times 10^{-5}$) and the mean-momentum ($\delta N_{\rm eff}^{m_e=0}/N_{\rm eff}^{\rm LO} \simeq -1.6\times 10^{-5}$) approaches, and are furthermore $\sim 30\%$ to a factor of four larger than those from entropy arguments. This result is consistent with expectations. As we have seen previously in figures 3 and 4, the rate correction under the $m_e = 0$ assumption diverges as $T \to 0$ relative the LO rate, whereas its $m_e \neq 0$ counterpart tends to zero. Since neutrino decoupling in the early universe is not instantaneous but extends into the $e^\pm$-annihilation epoch at $T \sim m_e$, any estimate of $\delta N_{\rm eff}$ that accounts for non-instantaneous decoupling will to some extent be sensitive to exactly what we assume for m_e in the $T \lesssim 3m_e$ region. Indeed, the low-T effect of m_e on $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$ is not captured by entropy conservation arguments, which are based on one point estimate (the decoupling temperature) at $T > m_e$, where the $m_e \neq 0$ and $m_e = 0$ rate corrections differ by less than 10%. In contrast, the full-momentum approach, which is sensitive to the largest range of temperatures, yields the strongest dependence of $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$ on what we assume for the electron mass. Thus, while the overall effect of the rate corrections on $N_{\rm eff}^{\rm SM}$ is small, we conclude that neglecting m_e in their computations is strictly not a good approximation in high-precision calculations of $N_{\rm eff}^{\rm SM}$.

5 Conclusions

In this work we have computed the QED corrections to the neutrino-electron interaction rate in the vicinity of neutrino decoupling and evaluated its impact on the standard-model value of the effective number of neutrinos $N_{\rm eff}^{\rm SM}$. We have focused on diagram (d) in figure 1, because of the expectation of a t-channel enhancement and hence dominance over the other three diagrams. Similar corrections have also been recently investigated in Cielo *et al.* [30] and Jackson and Laine [32], which the former analysis [30] found to lead to a significant shift in $N_{\rm eff}^{\rm SM}$: from the benchmark value of 3.044 [28, 29] to 3.043.

Contrary [30], our first-principles calculations show that QED corrections to the neutrino interaction rates are modest. In the temperature range $T \sim 1 \to 3$ MeV, the corrections to electron neutrino interaction rate fall in the range $-0.2 \to +0.1\%$ relative to the LO rate, while for the muon and tau neutrinos the effect is even more minute: $-0.0005 \to +0.0002\%$. These results are consistent with those reported by Jackson and Laine [32], despite differences between the formalisms (imaginary vs real-time) and some approximations (zero vs finite m_e) used. The more than two orders of magnitude difference between the relative corrections for

the ν_e and the $\nu_{\mu,\tau}$ rates also confirms the strong flavour dependence found in [32], which was not observed in Cielo *et al.* [30].

Using our QED-corrected neutrino interaction rates, we proceed to estimate the corresponding shift in $N_{\rm eff}^{\rm SM}$ under a variety of approximations and methods: via entropy conservation arguments which assume instantaneous decoupling, and by solving the Boltzmann equation in the damping approximation. Depending on the exact method/approximation used, we find the relative change in $N_{\rm eff}^{\rm SM}$ to fall in the range $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO} \simeq (-0.5 \to -1.1)\times 10^{-5}$. That is, relative to the current SM benchmark of $N_{\rm eff}^{\rm SM} = 3.0440 \pm 0.0002$ [28, 29], QED corrections to the neutrino interaction rates can at best shift the number in the negative direction in the fifth decimal place, and are hence completely within the quoted uncertainties. Thus, while we can confirm the sign of $\delta N_{\rm eff}$ computed by Cielo *et al.* [30], even our most "optimistic" estimate is a factor of 30 smaller than their claimed correction.

It is also interesting to observe that setting $m_e = 0$ in the rate corrections can have an $\mathcal{O}(1)$ impact on $\delta N_{\rm eff}/N_{\rm eff}^{\rm LO}$, even though the rate corrections themselves at $T \sim 1 \to 3$ MeV differ by less than 10%. This is because corrections assuming $m_e = 0$ deviate from their $m_e \neq 0$ counterparts at $T \lesssim 3m_e$, and diverge relative to LO rates just as the $m_e \neq 0$ corrections vanish in the $T \to 0$ limit. Since neutrino decoupling in the early universe is not instantaneous, these $T \lesssim 3m_e$ effects will imprint on $N_{\rm eff}^{\rm SM}$ despite the common understanding that neutrino decoupling occurs at $T \sim 1$ MeV. Thus, while the overall contribution from QED weak rate corrections to $N_{\rm eff}^{\rm SM}$ is ultimately small, we argue that neglecting m_e their computations is strictly speaking not a good approximation in high-precision $N_{\rm eff}^{\rm SM}$ calculations.

In conclusion, our results strongly suggest that the SM benchmark value $N_{\rm eff}^{\rm SM} = 3.0440 \pm 0.0002$ [28, 29] is correct within the quoted uncertainties. Naturally, a full numerical solution of the QKEs by a dedicated neutrino decoupling code such as `FortEPiaNO` [26]—with all NLO contributions from diagrams (a) to (d) incorporated in the collision integral—would be highly desirable. However, short of a new and previously unaccounted-for effect, we believe it is unlikely that a more detailed investigation of NLO effects on the neutrino interaction rate will yield a deviation from the existing SM benchmark $N_{\rm eff}^{\rm SM}$ that is large enough to be of any relevance for cosmological observations in the foreseeable future.

Acknowledgments

MaD and Y^{3}W would like to thank Isabel Oldengott and Jamie McDonald for discussions. YG acknowledges the support of the French Community of Belgium through the FRIA grant No. 1.E.063.22F. The research of MK and LPW was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Research Training Group 2149 "Strong and Weak Interactions - from Hadrons to Dark matter". Y^{3}W is supported in part by the Australian Research Council's Future Fellowship (project FT180100031) and Discovery Project (project DP240103130). Computational resources have been provided by the University of Münster through the HPC cluster PALMA II funded by the DFG (INST 211/667-1), and the supercomputing facilities of the Université Catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11 and by the Walloon Region.

A Bosonic and fermionic propagators at finite temperatures

For vanishing chemical potentials the free thermal propagators for fermions are given by

$$\begin{aligned} iS_\psi^{11} &= \frac{i(\not{p}+m_\psi)}{p^2-m_\psi^2+i\epsilon} - 2\pi(\not{p}+m_\psi)\delta(p^2-m_\psi^2)f_F(|p^0|) = \gamma^0(iS_\psi^{22})^\dagger\gamma^0, \\ iS_\psi^{12} &= -2\pi(\not{p}+m_\psi)\delta(p^2-m_\psi^2)\left[f_F(|p^0|)-\Theta(-p^0)\right], \\ iS_\psi^{21} &= -2\pi(\not{p}+m_\psi)\delta(p^2-m_\psi^2)\left[f_F(|p^0|)-\Theta(p^0)\right], \end{aligned} \tag{A.1}$$

while the ones for bosons are

$$\begin{aligned} i\Delta^{11} &= \frac{i}{p^2-m^2+i\epsilon} + 2\pi\delta(p^2-m^2)f_B(|p^0|) = (i\Delta^{22})^*, \\ i\Delta^{12} &= 2\pi\delta(p^2-m^2)\left[f_B(|p^0|)+\Theta(-p^0)\right], \\ i\Delta^{21} &= 2\pi\delta(p^2-m^2)\left[f_B(|p^0|)+\Theta(p^0)\right], \end{aligned} \tag{A.2}$$

where $f_B(E) = 1/(e^{E/T}-1)$ is the Bose-Einstein distribution.

For the computation of the resummed photon propagator, it is useful to write down the advanced and retarded bosonic propagators,

$$i\Delta^{R/A} = \frac{i}{p^2-m^2\pm i\mathrm{sgn}(p^0)\epsilon}, \tag{A.3}$$

as well as the statistical propagator

$$i\Delta^+ \equiv \Delta^{12}+\Delta^{21} = -2\pi i\left[1+2f_B(|p^0|)\right]\delta(p^2-m^2) = i\left[1+2f_B(p^0)\right]\Delta^-, \tag{A.4}$$

where the relation to the spectral function $\Delta^- = i(\Delta^{21}-\Delta^{12})$ is another way of stating the KMS relation. In terms of the scalar propagators, the tree-level photon propagator is given by

$$iD_{\mu\nu}^{ab}(p) = id_{\mu\nu}\Delta^{ab}(p)|_{m=0}, \tag{A.5}$$

where we have defined the polarisation tensor,

$$d^{\mu\nu}(p) = -g^{\mu\nu}+(1-\xi)\frac{p^\mu p^\nu}{p^2} = -P_L^{\mu\nu}-P_T^{\mu\nu}-\xi\frac{p^\mu p^\nu}{p^2}, \tag{A.6}$$

for a general R_ξ-gauge.

B Thermal integrals

Throughout the calculations we encounter thermal integrals of the following forms

$$\begin{aligned} K^\mu(P) &= 2e^2\int\frac{\mathrm{d}^4k}{(2\pi)^3}k^\mu\frac{\delta(k^2-m_e^2)f_F(|k^0|)}{(P+k)^2-m_e^2+i\epsilon}, \\ K^{\mu\nu}(P) &= 2e^2\int\frac{\mathrm{d}^4k}{(2\pi)^3}k^\mu k^\nu\frac{\delta(k^2-m_e^2)f_F(|k^0|)}{(P+k)^2-m_e^2+i\epsilon}. \end{aligned} \tag{B.1}$$

To evaluate them, we choose $\mathbf{P}$ as the reference direction with the corresponding unit vector $\hat{\mathbf{P}}$, and parameterise the integration variable via

$$\mathbf{k} = |\mathbf{k}|(\cos\vartheta\hat{\mathbf{P}}+\cos\varphi\sin\vartheta\mathbf{e}_1+\sin\varphi\sin\vartheta\mathbf{e}_2), \tag{B.2}$$

where, together, $\hat{\mathbf{P}}$ and the two unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ form a complete orthogonal basis.

The contributions from $\mathbf{e}_1$ and $\mathbf{e}_2$ to the spatial components of K^μ vanish after the azimuthal integration, giving

$$\mathbf{K} = \frac{\mathbf{P}}{|\mathbf{P}|^3}\frac{\alpha_{\rm em}}{4\pi}\int_{m_e}^{\infty} \mathrm{d}\omega \left[P^2\ell_1(\omega,P) + 2\omega P_0 \ell_2(\omega,P) - 8|\mathbf{k}||\mathbf{P}|\right] f_F(\omega), \tag{B.3}$$

where $\omega = |k^0| = \sqrt{|\mathbf{k}|^2 + m_e^2}$, and

$$\begin{aligned} \ell_1(\omega,P) &= \ln\left|\frac{(P^2+2|\mathbf{k}||\mathbf{P}|)^2 - 4P_0^2\omega^2}{(P^2-2|\mathbf{k}||\mathbf{P}|)^2 - 4P_0^2\omega^2}\right|, \\ \ell_2(\omega,P) &= \ln\left|\frac{P^4 - 4(P_0\omega + |\mathbf{P}||\mathbf{k}|)^2}{P^4 - 4(P_0\omega - |\mathbf{P}||\mathbf{k}|)^2}\right| \end{aligned} \tag{B.4}$$

arise from integration over $\cos\vartheta$. Correspondingly,

$$K^0 = \frac{\alpha_{\rm em}}{2\pi|\mathbf{P}|}\int_{m_e}^{\infty} \mathrm{d}\omega\, \omega\, \ell_2(\omega,P) f_F(\omega) \tag{B.5}$$

is the remaining, temporal component of K^μ.

For $K^{\mu\nu}$, after application of the completeness relation $\delta^{ij} = \hat{\mathbf{P}}^i\hat{\mathbf{P}}^j + \sum_{k=1,2} e_k^i e_k^j$, the spatial components become

$$K^{ij} = \hat{\mathbf{P}}^i\hat{\mathbf{P}}^j G_1 + (\delta^{ij} - \hat{\mathbf{P}}^i\hat{\mathbf{P}}^j)(G_2 - G_1/2), \tag{B.6}$$

where we have introduced the two auxiliary functions

$$\begin{aligned} G_1 &= \frac{\alpha_{\rm em}}{8\pi|\mathbf{P}|^3}\int_{m_e}^{\infty} \mathrm{d}\omega \left[(P^4 + 4P_0^2\omega^2)\ell_1(\omega,P) + 4P^2 P_0\,\omega\,\ell_2(\omega,P) - 8|\mathbf{P}||\mathbf{k}|P^2\right] f_F(\omega), \\ G_2 &= \frac{\alpha_{\rm em}}{4\pi|\mathbf{P}|}\int_{m_e}^{\infty} \mathrm{d}\omega\, |\mathbf{k}|^2 \ell_1(\omega,P) f_F(\omega). \end{aligned} \tag{B.7}$$

The remaining components with at least one index being zero are

$$\begin{aligned} K^{00}(P) &= \frac{\alpha_{\rm em}}{2\pi|\mathbf{P}|}\int_{m_e}^{\infty} \mathrm{d}\omega\, \omega^2\, \ell_1(\omega,P) f_F(\omega), \\ K^{0i} &= \frac{\alpha_{\rm em}\hat{\mathbf{P}}^i}{4\pi|\mathbf{P}|^2}\int_{m_e}^{\infty} \mathrm{d}\omega\, \omega \left[2P_0\omega\ell_1(\omega,P) + P^2\ell_2(\omega,P)\right] f_F(\omega). \end{aligned} \tag{B.8}$$

In all cases, the remaining single integral over ω can, for our purposes, be evaluated numerically efficiently.

C 1PI-resummed photon propagator

A crucial ingredient in our calculation is the resummed photon propagator used in our NLO thermal matrix element (3.13). In this appendix, we provide an exact derivation of the photon self-energy at the one-loop level within the real-time formalism, assuming electrons in thermal equilibrium with zero chemical potentials. From there, we extract the resummed Feynman propagator. While exact expressions for this resummed photon propagator have been long known in the simplified scenarios wherein $m_e = 0$ [53] or in the HTL approximation (see, e.g., [54, 55]), here we remove these assumptions and derive a propagator valid for all values of the electron mass m_e and photon 4-momentum P. We note that, while the reference [49] also studied the photon propagator without HTL approximations (and accounting for non-zero electron chemical potentials), no explicit formulae were provided.

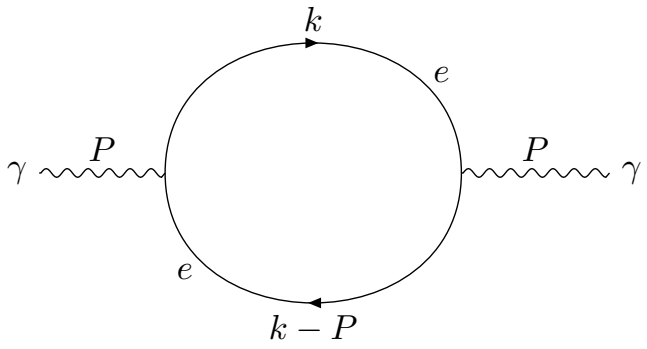


Figure 5: One-loop contribution to the photon self-energy.

C.1 1-loop photon self-energy

At the one-loop level, the thermal photon self-energy, displayed in figure 5, reads[12]

$$\Pi^{\mu\nu}_{ab}(P) = (-1)^{a+b} ie^2 \int \frac{\mathrm{d}^4k}{(2\pi)^4} \mathrm{Tr}\left[iS_e^{ab}(k)\gamma^\mu iS_e^{ba}(k+P)\gamma^\nu\right] \tag{C.1}$$

for general Keldysh indices $a, b \in \{1, 2\}$. It can always be split into a longitudinal and a transverse part,

$$\Pi^{\mu\nu}_{ab} = \Pi^L_{ab} P_L^{\mu\nu} + \Pi^T_{ab} P_T^{\mu\nu}, \tag{C.2}$$

with the corresponding projectors [36, 56]

$$\begin{aligned} &P_T^{00} = P_T^{0i} = P_T^{i0} = 0, \quad P_T^{ij} = -\delta^{ij} + \hat{\mathbf{P}}^i \hat{\mathbf{P}}^j, \\ &P_L^{\mu\nu} = g^{\mu\nu} - \frac{P^\mu P^\nu}{P^2} - P_T^{\mu\nu}. \end{aligned} \tag{C.3}$$

Writing the self-energy in this form makes also manifest that $\Pi^{\mu\nu}_{ab}$ fulfills the the Ward-Takahashi identity $P_\mu \Pi^{\mu\nu}_{ab} = 0$. The longitudinal projector can alternatively be expressed through the heat bath four-velocity $U^\mu = \delta^\mu_0$ (in its rest frame) as $P_L^{\mu\nu} = \tilde{U}^\mu \tilde{U}^\nu / \tilde{U}^2$ with $\tilde{U}^\mu = P^2 U^\mu - (U \cdot P) P^\mu$, so that the transverse projector becomes $P_T^{\mu\nu} = g^{\mu\nu} - \frac{P^\mu P^\nu}{P^2} - P_L^{\mu\nu}$.

In the following, we only highlight the derivation of Π_{11} and Π_{12}, as the remaining self-energies can be related to the first two via the KMS relation

$$\Pi_{21}(P) = e^{P_0/T} \Pi_{12}(P) \tag{C.4}$$

and

$$\Pi_{22}(P) = -\Pi_{11}(P)^* \tag{C.5}$$

for a general four-vector P.

Starting with $\Pi^{\mu\nu}_{11}$ ($\Pi^{\mu\nu}_{22}$), its diagonal component splits into the vacuum self-energy

$$\Pi_2^{\mu\nu} = ie^2 \int_k \frac{i}{k^2 - m_e^2 + i\epsilon} \frac{i}{(k+P)^2 - m_e^2 + i\epsilon} \mathrm{Tr}\left[(\not{k} + m_e)\gamma^\mu (\not{k} + \not{P} + m_e)\gamma^\nu\right] \tag{C.6}$$

[12] We only compute here the electron contribution to the photon self-energy, which is by far the dominant contribution at the relevant temperatures for N_{eff}. It is nonetheless straightforward to extend our results to higher temperatures where different charged particles can also contribute.

and the finite-temperature part $\Pi^{\mu\nu}_{11,T\neq0}$ ($\Pi^{\mu\nu}_{22,T\neq0}$). The renormalised vacuum part $\Pi^{\mu\nu}_2(P) = (\alpha_{\rm em}/\pi)(P^2 g^{\mu\nu} - P^\nu P^\mu)\Pi_2(P^2)$ is textbook material and, in the $\overline{\rm MS}$-scheme, reads

$$\begin{aligned}\Pi_2(P^2) &= \frac{2}{3}\ln\left(\frac{\mu}{m_e}\right) + \frac{5}{9} + \frac{\tau}{3} + \frac{1}{3}\left(1+\frac{\tau}{2}\right)\sqrt{1-\tau}\ln\left(\frac{\sqrt{1-\tau}-1}{\sqrt{1-\tau}+1}\right)\\ &\longrightarrow \frac{1}{3}\ln\left(\frac{\mu^2}{-P^2}\right) + \frac{5}{9} \quad \text{for } m_e \to 0,\end{aligned} \tag{C.7}$$

with $\tau = 4m_e^2/P^2$ (see, e.g., appendix A of [57] for the derivation). In practice, we note that the vacuum contribution to the photon self-energy is numerically irrelevant when resumming the photon propagator, as it vanishes in the soft photon exchange limit for which $P^2 = 0$. We can therefore safely neglect it in the resummed propagator.

In terms of the integrals K^μ and $K^{\mu\nu}$ computed in appendix B, the corresponding real parts of the two diagonal components read

$$\mathrm{Re}\,\Pi^{\mu\nu}_{11,T\neq0} = -\mathrm{Re}\,\Pi^{\mu\nu}_{22,T\neq0} = 4(-g^{\mu\nu}K\cdot P + 2K^{\mu\nu} + K^\mu P^\nu + K^\nu P^\nu). \tag{C.8}$$

After collecting all the terms according to the projectors defined in equation (C.3), we obtain for the longitudinal part

$$\mathrm{Re}\,\Pi^L_{11,T\neq0} = \frac{\alpha_{\rm em}P^2}{\pi|\mathbf{P}|^3}\int_{m_e}^\infty \mathrm{d}\omega\left[8k|\mathbf{P}| - \ell_1(\omega,P)(P^2+4\omega^2) - 4\ell_2(\omega,P)P_0\omega\right] f_F(\omega) \tag{C.9}$$

$$\overset{\text{HTL}}{\approx} -3m_\gamma^2\left(1-\frac{P_0^2}{|\mathbf{P}|^2}\right)\left[1 - \frac{P_0}{2|\mathbf{P}|}\ln\left|\frac{P_0+|\mathbf{P}|}{P_0-|\mathbf{P}|}\right|\right] \longrightarrow -3m_\gamma^2 \quad \text{for } P_0 = 0, \tag{C.10}$$

and for the transverse one

$$\begin{aligned}\mathrm{Re}\,\Pi^T_{11,T\neq0} = \frac{\alpha_{\rm em}}{\pi|\mathbf{P}|^3}\int_{m_e}^\infty \mathrm{d}\omega\,\big[&2\ell_2(\omega,P)P_0P^2\omega - 4|\mathbf{k}||\mathbf{P}|(P_0^2+|\mathbf{P}|^2)\\ &+\ell_1(\omega,P)(|\mathbf{P}|^2P^2 + 2P_0^2\omega^2 - 2|\mathbf{k}|^2|\mathbf{P}|^2 + \frac{1}{2}P^4)\Big] f_F(\omega)\end{aligned} \tag{C.11}$$

$$\overset{\text{HTL}}{\approx} -\frac{3}{2}m_\gamma^2\frac{P_0^2}{|\mathbf{P}|^2}\left[1 - \left(1-\frac{|\mathbf{P}|^2}{P_0^2}\right)\frac{P_0}{2|\mathbf{P}|}\ln\left|\frac{P_0+|\mathbf{P}|}{P_0-|\mathbf{P}|}\right|\right] \longrightarrow 0, \quad \text{for } P_0 = 0 \tag{C.12}$$

with the thermal photon mass $m_\gamma = eT/3$. Here, the symbol $\overset{\text{HTL}}{\approx}$ refers to the result obtained in the HTL approximation [58], defined through the assumption that the quantities in the integral can be separated into hard $\mathcal{O}(\pi T)$ and soft scales $\mathcal{O}(eT)$, and assuming that hard momenta k dominate the loop integral. The external momentum P on the other hand is defined to be soft and therefore also negligible compared to the loop momentum. More precisely, to go from the fully momentum- and mass-dependent expression (C.9) (or (C.11)) to the HTL result (C.10) (or (C.12)), we have expanded in $P_0/|\mathbf{k}|$ at the first step and then in $P_0/|\mathbf{P}|$ in the second step, using also

$$\ell_1(\omega,P) \overset{\text{HTL}}{\approx} -2\frac{|\mathbf{P}|}{|\mathbf{k}|}, \quad \ell_2(\omega,P) \overset{\text{HTL}}{\approx} 2\ln\left|\frac{P_0+|\mathbf{P}|}{P_0-|\mathbf{P}|}\right| \tag{C.13}$$

in the same expansion scheme. We further assume that $m_e \ll T$, such that the electron can be treated as effectively massless. This leaves us with an integral over ω that can be

performed analytically using the relation $\int_0^\infty \mathrm{d}k\, k\, f_F(k) = \pi^2 T^2/12$, thereby yielding the HTL result (C.10) (or (C.12)).

The (purely imaginary) off-diagonal component $\Pi_{12}^{\mu\nu}$, on the other hand, reads

$$\begin{aligned}\Pi_{12}^{\mu\nu} = i(ie)^2 \int \frac{\mathrm{d}^4k}{(2\pi)^4} 16\pi^2 \delta\left(k^2 - m_e^2\right) \left(f_F(|k^0|) - \Theta(k^0)\right) \delta\left((k+P)^2 - m_e^2\right) \\ \times \left(f_F(|k^0+P^0|) - \Theta(-k^0-P^0)\right) T^{\mu\nu},\end{aligned} \tag{C.14}$$

where the minus sign from the fermionic loop and the one from the fact that we have a type 1 and a type 2 vertex cancel out, and

$$T^{\mu\nu} = -g^{\mu\nu} k\cdot P + 2k^\mu k^\nu + k^\nu P^\mu + k^\mu P^\nu. \tag{C.15}$$

For the evaluation of the integrals, we make use again of the decomposition of $\mathbf{k}$ shown in equation (B.2). The additional δ-function here (compared to the diagonal case) fixes then the polar angle θ to the value

$$\cos\theta_\pm^* = \frac{P^2 \pm 2\omega P_0}{2|\mathbf{k}||\mathbf{P}|}. \tag{C.16}$$

With that, $T^{\mu\nu}$ separates into a transverse and a longitudinal part according to

$$T_\pm^{\mu\nu} = \frac{P^2}{2}(P_T^{\mu\nu} + P_L^{\mu\nu}) - |\mathbf{k}|^2 \sin^2\theta_\pm^* P_T^{\mu\nu} - \frac{P^2}{2|\mathbf{P}|^2}(P_0 \pm 2\omega)^2 P_L^{\mu\nu}. \tag{C.17}$$

The final result then reads

$$\Pi_{12}^{\mu\nu} = -i\frac{2\alpha_{\mathrm{em}}}{|\mathbf{P}|}\int_{m_e}^\infty \mathrm{d}\omega \sum_\pm \Theta(1-|\cos\theta_\pm^*|)(f_F(\omega) - \Theta(k^0))(f_F(|k^0+P^0|) - \Theta(-k^0-P^0))T_\pm^{\mu\nu}, \tag{C.18}$$

where the sum runs over positive and negative energies $k^0 = \pm\omega$. In the HTL limit we find

$$\Pi_{12}^L \overset{\mathrm{HTL}}{\approx} -i\frac{\pi e^2 P^2 T^3}{3|\mathbf{P}|^3}\Theta(|\mathbf{P}| - |P^0|), \tag{C.19}$$

$$\Pi_{12}^T \overset{\mathrm{HTL}}{\approx} i\frac{\pi e^2 P^2 T^3}{6|\mathbf{P}|^3}\Theta(|\mathbf{P}| - |P^0|), \tag{C.20}$$

where we have used the integral $\int_0^\infty \mathrm{d}k\, f_F(k)\,[1 - f_F(k)] = \pi^2 T^3/6$.

As will be explained in the subsequent section, it is convenient to compute the retarded and advanced photon self-energy $\Pi_{R/A}$ in order to perform the resummation of the photon propagator. Using the relation $\mathrm{Im}\,\Pi_{11} = \frac{i}{2}(\Pi_{12} + \Pi_{21})$, we can write the advanced and retarded self-energies as

$$\begin{aligned}\Pi_{R/A}^{T/L} = \Pi_{11}^{T/L} + \Pi_{12/21}^{T/L} &= \mathrm{Re}\,\Pi_{11}^{T/L} \pm \frac{1}{2}\left(\Pi_{12}^{T/L} - \Pi_{21}^{T/L}\right) \\ &= \mathrm{Re}\,\Pi_{11}^{T/L} \pm \frac{1}{2}\left(1 - e^{P_0/T}\right)\Pi_{12}^{T/L},\end{aligned} \tag{C.21}$$

where we have used the KMS relation (C.4) at the second equality. In particular, equation (C.21) implies that $\mathrm{Im}\Pi^R = -\mathrm{Im}\Pi^A$. With this in mind, we can directly compare our results for the advanced and the retarded transverse photon self-energies in the HTL limit to the results of Carrington *et al.* [54]. Given that $\Pi_{R,A}^{L/T} \overset{\mathrm{HTL}}{\approx} \mathrm{Re}\,\Pi_{11}^T \mp \frac{P^0}{2T}\Pi_{12}^T$ in the HTL limit, our

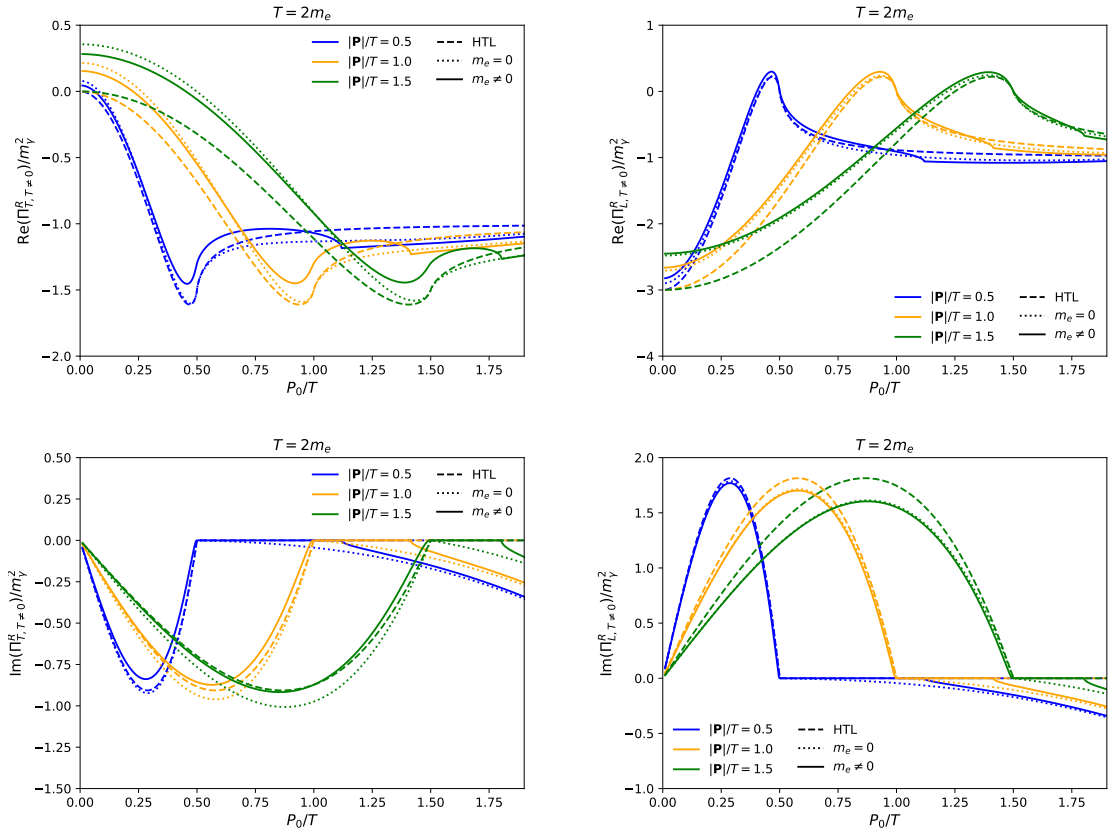


Figure 6: Comparison between the finite-temperature part of the photon self-energy at the one-loop level with (continuous lines) and without (dashed lines) the HTL approximation at $T = 2m_e$ and for various choices of $|\mathbf{P}|/T$.

results are in agreement with [54] after carefully sending the time-ordering parameter ϵ to zero. In the longitudinal part we differ by a term $P_0^2/|\mathbf{P}|^2$ with respect to [54],[13] but agree with [55]. We note in passing that the self-energy (C.18) leads to the unphysical process of photon decay $\gamma \to e^+e^-$ at high enough temperatures, where m_γ exceeds $2m_e$ [49]. In practice, this could be resolved by resumming the electron propagator. In our case, however, the relevant dynamics happen at temperatures much below that threshold, such that this resummation is not necessary.

In figure 6, we display the finite-temperature contribution to the real and the imaginary parts of the retarded transverse and longitudinal propagators, for $T = 2m_e$ and various choices of $|\mathbf{P}|/T$. We compare the exact one-loop photon self-energy to (i) the equivalent self-energy in the limit $m_e \to 0$, and (ii) the equivalent self-energy in the HTL limit. Similarly, in figure 7, we examine the impact of the temperature on the photon self-energy. We observe that the effect of the finite electron mass m_e, while small for large temperatures $T/m_e \sim 5$, remains sizeable for temperatures around the electron neutrino decoupling temperature where $T/m_e \sim 2$.

Beyond the photon self-energy, other quantities of interest are the residue of the trans-

[13] This difference can be seen by comparing figure 7 to figure 4 in [53].

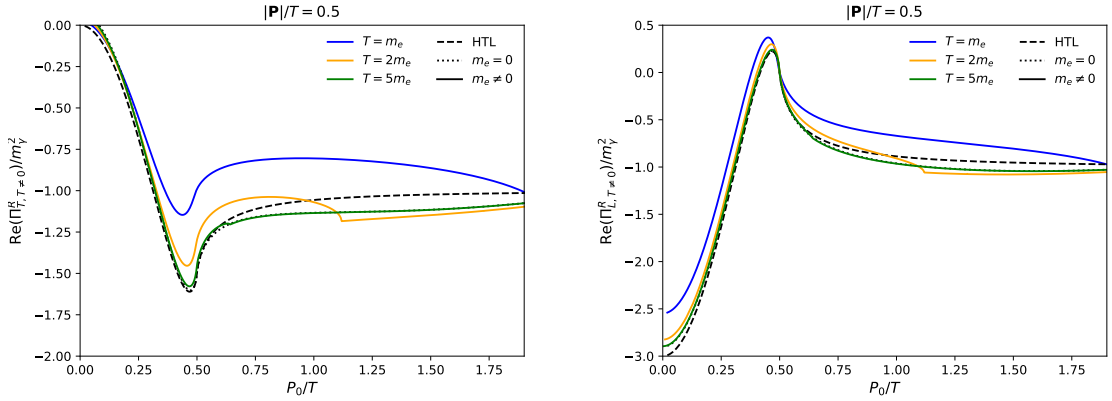


Figure 7: Comparison between the finite-temperature part of the photon self-energy at the one-loop level with (continuous lines) and without (dashed lines) the HTL approximation for different temperatures $T \in \{m_e, 2m_e, 5m_e\}$ at a fixed $|\mathbf{P}|/T = 0.5$.

verse and longitudinal photon[14] propagators which are quantified by

$$Z^{T/L} = \lim_{P_0 \to w(|\mathbf{P}|)} \left[1 + \frac{\partial \mathrm{Re}\,\Pi_{11}^{T/L}}{\partial P_0^2}\right]^{-1}, \tag{C.22}$$

where $P_0 = w(|\mathbf{P}|)$ defines the photon dispersion relation, i.e., solutions of the equation $P^2 + \mathrm{Re}\,\Pi_{11}^{T/L}(P) = 0$, which we show for the benchmark $T = 2m_e$ case in figure 8. The two residues take the form

$$\begin{aligned} Z^L = \Bigg[1 - \frac{\alpha_{\mathrm{em}}}{\pi|\mathbf{P}|^3}\int_{m_e}^{\infty} \mathrm{d}\omega f_F(\omega)\Big[\ell_2'(\omega,P)\left(4\omega P_0 P^2\right) + \ell_2(\omega,P)\left(2\omega(3P_0 - |\mathbf{P}|^2/P_0)\right) \\ + \ell_1'(\omega,P)\left(P^4 + 4\omega^2 P^2\right) + \ell_1(\omega,P)\left(2P^2 + 4\omega^2\right) - 8|\mathbf{k}||\mathbf{P}|\Big]_{|P_0 = w(|\mathbf{P}|)}\Bigg]^{-1}, \end{aligned} \tag{C.23}$$

and

$$\begin{aligned} Z^T = \Bigg[1 + \frac{\alpha_{\mathrm{em}}}{2\pi|\mathbf{P}|^3}\int_{m_e}^{\infty} \mathrm{d}\omega f_F(\omega)\Big[\ell_2'(\omega,P)\left(4\omega P_0 P^2\right) + \ell_2(\omega,P)\left(2\omega(3P_0 - |\mathbf{P}|^2/P_0)\right) \\ + \ell_1'(\omega,P)\left(P^4 + 4\omega^2 P_0^2 - 4|\mathbf{k}|^2|\mathbf{P}|^2 + 2P^2|\mathbf{P}|^2\right) \\ + \ell_1(\omega,P)\left(4\omega^2 + 2P_0^2\right) - 8|\mathbf{k}||\mathbf{P}|\Big]_{|P_0 = w(|\mathbf{P}|)}\Bigg]^{-1}, \end{aligned} \tag{C.24}$$

where we have defined the following derivatives

$$\begin{aligned} \ell_1'(\omega,P) &\equiv \frac{\partial \ell_1}{\partial P_0^2} = \frac{2(P^2 + 2|\mathbf{k}||\mathbf{P}|) - 4\omega^2}{(P^2 + 2|\mathbf{k}||\mathbf{P}|)^2 - 4P_0^2\omega^2} - \frac{2(P^2 - 2|\mathbf{k}||\mathbf{P}|) - 4\omega^2}{(P^2 - 2|\mathbf{k}||\mathbf{P}|)^2 - 4P_0^2\omega^2}, \\ \ell_2'(\omega,P) &\equiv \frac{\partial \ell_2}{\partial P_0^2} = \frac{2P^2 - 8\omega(P_0\omega + |\mathbf{P}|\mathbf{k}|)}{P^4 - 4(P_0\omega + |\mathbf{P}||\mathbf{k}|)^2} - \frac{2P^2 - 8\omega(P_0\omega - |\mathbf{P}||\mathbf{k}|)}{P^4 - 4(P_0\omega - |\mathbf{P}||\mathbf{k}|)^2}, \end{aligned} \tag{C.25}$$

of the logarithmic functions $\ell_1(\omega, P)$ and $\ell_2(\omega, P)$ of equation (B.4).

[14] Longitudinal photons are also sometimes referred to as "plasmons" in the literature.

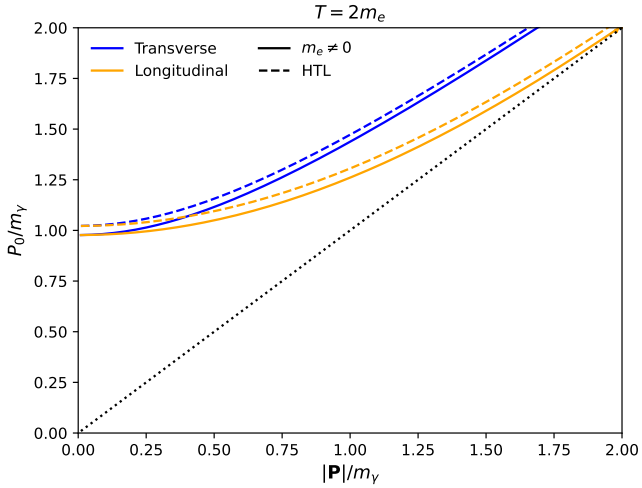


Figure 8: Dispersion relation for transverse (blue lines) and longitudinal (orange lines) photons at the one-loop level with (continuous lines) and without (dashed lines) the HTL approximation at $T = 2m_e$. The black dotted line indicates the dispersion relation at zero temperature, i.e., $P_0^2 = |\mathbf{P}|^2$.

C.2 The resummed photon propagator

The resummed photon propagator $\bar{D}^{ab}_{\mu\nu}$ is given by the Dyson-Schwinger equation

$$\bar{D}^{ab}_{\mu\nu} = D^{ab}_{\mu\nu} + \sum_{c,d} D^{ac}_{\mu\alpha}(\Pi^{\alpha\beta})^{cd}\bar{D}^{db}_{\beta\nu}. \tag{C.26}$$

The sum over the real-time labels c and d can be avoided [59] by re-expressing the equation in terms of the retarded and advanced propagators

$$\begin{aligned} G^R &= G_{11} - G_{12} = G_{21} - G_{22}, \\ G^A &= G_{11} - G_{21} = G_{12} - G_{22}, \end{aligned} \tag{C.27}$$

defined in a similar fashion to the retarded/advanced self-energies (C.21). This leads to the compact result

$$\bar{D}^{R/A}_{\mu\nu} = D^{R/A}_{\mu\nu} + D^{R/A}_{\mu\alpha}(\Pi^{\alpha\beta})^{R/A}\bar{D}^{R/A}_{\beta\nu}. \tag{C.28}$$

Decomposing the resummed photon propagator using the projectors from equations (C.3),

$$\bar{D}^{ab}_{\mu\nu} = d_{\mu\nu}\bar{D}^{ab} = -\bar{D}^{ab}_L P^L_{\mu\nu} - \bar{D}^{ab}_T P^T_{\mu\nu} - \xi\Delta^{ab}\frac{P_\mu P_\nu}{P^2}, \tag{C.29}$$

and using equation (C.28), the corresponding longitudinal and transverse components of the resummed retarded and advanced propagator take the form

$$\bar{D}^{R/A}_{L/T} = \frac{1}{(\Delta^{R/A})^{-1} + \Pi^{R/A}_{L/T}}. \tag{C.30}$$

With $D_{11} = \frac{1}{2}(D^A + D^R + D^+)$ and using the fact that *in equilibrium*, the thermal resummed statistical photon propagator $\bar{D}^+_{\mu\nu}$ is related to the retarded and advanced propagators via [54]

$$\bar{D}^+_{\mu\nu}(P) = [1 + 2f_B(|P^0|)]\mathrm{sgn}(P^0)\left[\bar{D}^R_{\mu\nu} - \bar{D}^A_{\mu\nu}\right], \tag{C.31}$$

we finally arrive at

$$\bar{D}_{11}^{L/T} = \frac{\Omega_{L,T}}{\Omega_{L,T}^2 + \Gamma_{L,T}^2} - i[1 + 2f_B(|p^0|)]\mathrm{sgn}(p^0)\frac{\Gamma_{L,T}}{\Omega_{L,T}^2 + \Gamma_{L,T}^2}, \tag{C.32}$$

with

$$\Omega^{L,T} = p^2 + \mathrm{Re}\,\Pi_R^{L/T}, \tag{C.33}$$

$$\Gamma^{L,T} = \mathrm{Im}\,\Pi_R^{L,T} = -\frac{1}{2f_B(P^0)}\mathrm{Im}\Pi_{12}^{L,T}. \tag{C.34}$$

It is understood that (C.34) should be replaced by the causality-respecting ϵ-prescription if the imaginary part of the photon self-energy vanishes for kinematic reasons at the given order in perturbation theory.

D Parameterisation of the collision integrals

We parameterise the momenta in the self-energy expression (3.8) as in reference [21],

$$\begin{aligned}
\mathbf{p} &= |\mathbf{p}|(0,0,1)^T, \\
\mathbf{l} &= |\mathbf{l}|(0,\sin\alpha,\cos\alpha)^T, \\
\mathbf{q} &= |\mathbf{q}|(\sin\theta\sin\beta,\sin\theta\cos\beta,\cos\theta)^T,
\end{aligned} \tag{D.1}$$

resulting in

$$\begin{aligned}
\mathrm{Tr}\left[\not{p}\Sigma_{\mathrm{NLO}}^{12}\right] &= \int \mathrm{d}^4q\,\mathrm{d}^4l\,\delta(l^2)\delta(q^2-m_e^2)\delta((q+p-l)^2-m_e^2)f(q,l,p) \\
&= \int \frac{\mathrm{d}^3q\,\mathrm{d}^3l}{4|\mathbf{l}|E_q}\sum_{\epsilon,\tau=\pm1} f(q,l,p)\delta((q+p-l)^2-m_e^2)\Big|_{\substack{q^0=\epsilon E_q\\ l^0=\tau|l|}} \\
&= \frac{2\pi}{4}\int \mathrm{d}|\mathbf{l}|\,\mathrm{d}|\mathbf{q}|\,\frac{|\mathbf{q}|^2|\mathbf{l}|}{E_q}\int_{-1}^{1}\mathrm{d}\cos\alpha\sum_{\epsilon,\tau=\pm1}\frac{\theta(b^2-4ac)}{\sqrt{|a|}}\int_{z_-}^{z^+}\mathrm{d}z\,\frac{f(q,l,p)}{\sqrt{(z-z_-)(z_+-z)}}\Big|_{\substack{q^0=\epsilon E_q\\ l^0=\tau|\mathbf{l}|}},
\end{aligned} \tag{D.2}$$

where $E_q = \sqrt{|\mathbf{q}|^2+m_e^2}$, and the function $f(q,l,p)$ acts as a placeholder for the integrand in equation (3.8). The last δ-function fixes the angle β to

$$\cos\beta_i = \frac{|\mathbf{p}||\mathbf{q}|\cos\theta + \epsilon\tau E_q|\mathbf{l}| + \tau|\mathbf{l}||\mathbf{p}| - |\mathbf{p}||\mathbf{l}|\cos\alpha - \epsilon E_q|\mathbf{p}| - |\mathbf{l}||\mathbf{q}|\cos\alpha\cos\theta}{|\mathbf{l}||\mathbf{q}|\sin\alpha\sin\theta}. \tag{D.3}$$

The integration limits on $z=\cos\theta$,

$$z_\pm = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}}, \tag{D.4}$$

are the roots of the quadratic polynomial az^2+bz+c, with

$$\begin{aligned}
a &= -q^2|\mathbf{p}-\mathbf{l}|^2, \\
b &= -2(\cos\alpha|\mathbf{l}|-|\mathbf{p}|)|\mathbf{q}|(|\mathbf{l}||\mathbf{p}|(\cos\alpha-\tau)+\epsilon E_q(|\mathbf{p}|-|\mathbf{l}|\tau)), \\
c &= (E_q^2(|\mathbf{p}|-|\mathbf{l}|\tau)^2) + 2\epsilon E_q|\mathbf{l}||\mathbf{p}|(\cos\alpha-\tau)(-|\mathbf{p}|+|\mathbf{l}|\tau) \\
&\quad + |\mathbf{l}|^2(-((\cos^2\alpha)|\mathbf{p}|^2) + |\mathbf{q}|^2(1-\cos^2\alpha) + 2\cos\alpha|\mathbf{p}|^2\tau - |\mathbf{p}|^2)
\end{aligned} \tag{D.5}$$

as the coefficients.

The z-integration of the vacuum and the thermal corrections can be performed analytically following

$$I^z = \int_{z_-}^{z^+} \mathrm{d}z \, \frac{\mathcal{T}(q,l,p)}{\sqrt{(z-z_-)(z_+-z)}} = \mathcal{A}\,\pi \left(\mathcal{G}^0 - \frac{b}{2a}\mathcal{G}^1 + \frac{3b^2-4ca}{8a^2}\mathcal{G}^2 \right), \tag{D.6}$$

where we have expanded the correction matrix $\mathcal{T}$ in powers of z according to

$$\mathcal{T} = \mathcal{A}\left(\mathcal{G}^0 + \mathcal{G}^1 z + \mathcal{G}^2 z^2\right). \tag{D.7}$$

Then, for the transverse thermal part, the coefficients read

$$\begin{aligned}
\mathcal{G}^0_{\mathrm{th},T} &= \frac{P^2}{4|\mathbf{P}|^2}\Big[2|\mathbf{l}|^2P^2 + 4l_0|\mathbf{p}|P_0q_0 + 8l_0|\mathbf{p}|q_0^2 + 2l_0P_0P^2 + 2m_e^2|\mathbf{P}|^2 + 8|\mathbf{p}|^2P_0^2q_0^2/P^2 \\
&\quad - 8|\mathbf{p}|^2q_0^2 + 4|\mathbf{p}|P_0^2q_0 + 4|\mathbf{p}|P_0q_0^2 - 2|\mathbf{p}|P^2q_0 + P_0^2P^2 + 2P_0P^2q_0 + 2P^2q_0^2\Big], \\
\mathcal{G}^1_{\mathrm{th},T} &= -\frac{|\mathbf{p}||\mathbf{q}|P^2}{2|\mathbf{P}|^2}\left[2l_0(P_0+2q_0) + 2P_0^2(1+4|\mathbf{p}|q_0/P^2) - 8|\mathbf{p}|q_0 - P^2 + 2P_0q_0\right], \\
\mathcal{G}^2_{\mathrm{th},T} &= 2|\mathbf{p}|^2|\mathbf{q}|^2,
\end{aligned} \tag{D.8}$$

while for the longitudinal thermal part we have

$$\begin{aligned}
\mathcal{G}^0_{\mathrm{th},L} &= -\frac{P^2}{8|\mathbf{P}|^2}\Big[4P_0q_0(P^2+2|\mathbf{p}|q_0) + 4P^2l^2 + (P^2+4|\mathbf{p}|q_0)P_0^2 + P^2(P^2+4q_0^2) \\
&\quad + 4l_0(P_0(P^2+2|\mathbf{p}|q_0) + 4|\mathbf{p}|q_0^2)\Big], \\
\mathcal{G}^1_{\mathrm{th},L} &= \frac{P^2|\mathbf{q}||\mathbf{p}|}{2|\mathbf{P}|^2}(l_0+|\mathbf{p}|)(P_0+2q_0), \\
\mathcal{G}^2_{\mathrm{th},L} &= 0,
\end{aligned} \tag{D.9}$$

with the prefactor $\mathcal{A}^{L/T}_{\mathrm{th}} = -2^8G_F^2(g_V^\alpha)^2\mathrm{Re}\,\bar{D}^{L,T}_{11}\mathrm{Re}\,\Pi^{L,T}_{11,T\neq0}$. For the vacuum part, the corresponding coefficients are

$$\begin{aligned}
\mathcal{G}^0_{\mathrm{vac}} &= -m_e^2(l\cdot p) + 2|\mathbf{p}|^2q_0^2, \\
\mathcal{G}^1_{\mathrm{vac}} &= -4|\mathbf{q}|q_0|\mathbf{p}|^2, \\
\mathcal{G}^2_{\mathrm{vac}} &= 2|\mathbf{p}|^2|\mathbf{q}|^2,
\end{aligned} \tag{D.10}$$

with $\mathcal{A}_{\mathrm{vac}} = -2^{10}G_F^2\,\alpha_{\mathrm{em}}\,[(g_V^\alpha)^2/(4\pi)]\,\Pi_2$.

E Functions in the continuity equation

The continuity (4.12), written in terms of the rescaled variables x, y, z, contains the functions $J(x/z)$, $Y(x/z)$, and $G_{1,2}(x/z)$. We give the explicit forms of these functions here to $\mathcal{O}(e^2)$ [20],

$$\begin{aligned}
J(\tau) &\equiv \frac{1}{\pi^2}\int_0^\infty \mathrm{d}\omega\,\omega^2 \frac{\exp\left(\sqrt{\omega^2+\tau^2}\right)}{[\exp\left(\sqrt{\omega^2+\tau^2}\right)+1]^2}, \\
Y(\tau) &\equiv \frac{1}{\pi^2}\int_0^\infty \mathrm{d}\omega\,\omega^4 \frac{\exp\left(\sqrt{\omega^2+\tau^2}\right)}{[\exp\left(\sqrt{\omega^2+\tau^2}\right)+1]^2},
\end{aligned} \tag{E.1}$$

$$
\begin{aligned}
G_1(\tau) =& 2\pi\alpha_{\rm em}\left[\frac{1}{\tau}\left(\frac{K(\tau)}{3}+2K^2(\tau)-\frac{J(\tau)}{6}-K(\tau)J(\tau)\right)\right. \\
&\left.+\left(\frac{K'(\tau)}{6}-K(\tau)K'(\tau)+\frac{J'(\tau)}{6}+J'(\tau)K(\tau)+J(\tau)K'(\tau)\right)\right],
\end{aligned}
\tag{E.2}
$$

$$
\begin{aligned}
G_2(\tau) =& -8\pi\alpha_{\rm em}\left(\frac{K(\tau)}{6}+\frac{J(\tau)}{6}-\frac{1}{2}K^2(\tau)+K(\tau)J(\tau)\right) \\
&+2\pi\alpha_{\rm em}\left(\frac{K'(\tau)}{6}-K(\tau)K'(\tau)+\frac{J'(\tau)}{6}+J'(\tau)K(\tau)+J(\tau)K'(\tau)\right),
\end{aligned}
\tag{E.3}
$$

where $'$ denotes a derivative with respect to τ. Expressions for $G_{1,2}(\tau)$ at $\mathcal{O}(e^3)$ can be found in reference [21].

References

[1] G. Steigman, D.N. Schramm and J.E. Gunn, *Cosmological Limits to the Number of Massive Leptons*, *Phys. Lett. B* **66** (1977) 202.

[2] R. Barbieri and A. Dolgov, *Neutrino oscillations in the early universe*, *Nucl. Phys. B* **349** (1991) 743.

[3] A.M. Abdullahi et al., *The present and future status of heavy neutral leptons*, *J. Phys. G* **50** (2023) 020501 [2203.08039].

[4] L. Di Luzio, J. Martin Camalich, G. Martinelli, J.A. Oller and G. Piazza, *Axion-pion thermalization rate in unitarized NLO chiral perturbation theory*, *Phys. Rev. D* **108** (2023) 035025 [2211.05073].

[5] F. D'Eramo, F. Hajkarim and S. Yun, *Thermal QCD Axions across Thresholds*, *JHEP* **10** (2021) 224 [2108.05371].

[6] C. Caprini and D.G. Figueroa, *Cosmological Backgrounds of Gravitational Waves*, *Class. Quant. Grav.* **35** (2018) 163001 [1801.04268].

[7] A. Aboubrahim, M. Klasen and P. Nath, *Analyzing the Hubble tension through hidden sector dynamics in the early universe*, *JCAP* **04** (2022) 042 [2202.04453].

[8] P. Agrawal et al., *Feebly-interacting particles: FIPs 2020 workshop report*, *Eur. Phys. J. C* **81** (2021) 1015 [2102.12143].

[9] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[10] S. Dodelson and M.S. Turner, *Nonequilibrium neutrino statistical mechanics in the expanding universe*, *Phys. Rev. D* **46** (1992) 3372.

[11] S. Hannestad and J. Madsen, *Neutrino decoupling in the early universe*, *Phys. Rev. D* **52** (1995) 1764 [astro-ph/9506015].

[12] A.D. Dolgov, S.H. Hansen and D.V. Semikoz, *Nonequilibrium corrections to the spectra of massless neutrinos in the early universe*, *Nucl. Phys. B* **503** (1997) 426.

[13] A.D. Dolgov, S.H. Hansen and D.V. Semikoz, *Nonequilibrium corrections to the spectra of massless neutrinos in the early universe: Addendum*, *Nucl. Phys. B* **543** (1999) 269.

[14] S. Esposito, G. Miele, S. Pastor, M. Peloso and O. Pisanti, *Nonequilibrium spectra of degenerate relic neutrinos*, *Nucl. Phys. B* **590** (2000) 539.

[15] J. Froustey and C. Pitrou, *Incomplete neutrino decoupling effect on big bang nucleosynthesis*, *Phys. Rev. D* **101** (2020) 043524 [1912.09378].

[16] D.A. Dicus, E.W. Kolb, A.M. Gleeson, E.C.G. Sudarshan, V.L. Teplitz and M.S. Turner, *Primordial Nucleosynthesis Including Radiative, Coulomb, and Finite Temperature Corrections to Weak Rates*, *Phys. Rev. D* **26** (1982) 2694.

[17] A. Heckler, *Astrophysical applications of quantum corrections to the equation of state of a plasma*, *Phys. Rev. D* **49** (1994) 611.

[18] N. Fornengo, C. Kim and J. Song, *Finite temperature effects on the neutrino decoupling in the early universe*, *Phys. Rev. D* **56** (1997) 5123 [hep-ph/9702324].

[19] R.E. Lopez and M.S. Turner, *An Accurate Calculation of the Big Bang Prediction for the Abundance of Primordial Helium*, *Phys. Rev. D* **59** (1999) 103502 [astro-ph/9807279].

[20] G. Mangano, G. Miele, S. Pastor and M. Peloso, *A Precision calculation of the effective number of cosmological neutrinos*, *Phys. Lett. B* **534** (2002) 8 [astro-ph/0111408].

[21] J.J. Bennett, G. Buldgen, M. Drewes and Y.Y.Y. Wong, *Towards a precision calculation of the effective number of neutrinos $N_{\rm eff}$ in the Standard Model I: the QED equation of state*, *JCAP* **03** (2020) 003 [1911.04504].

[22] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P.D. Serpico, *Relic neutrino decoupling including flavor oscillations*, *Nucl. Phys. B* **729** (2005) 221 [hep-ph/0506164].

[23] J. Birrell, C.-T. Yang and J. Rafelski, *Relic Neutrino Freeze-out: Dependence on Natural Constants*, *Nucl. Phys. B* **890** (2014) 481 [1406.1759].

[24] E. Grohs, G.M. Fuller, C.T. Kishimoto, M.W. Paris and A. Vlasenko, *Neutrino energy transport in weak decoupling and big bang nucleosynthesis*, *Phys. Rev. D* **93** (2016) 083522 [1512.02205].

[25] P.F. de Salas and S. Pastor, *Relic neutrino decoupling with flavour oscillations revisited*, *JCAP* **07** (2016) 051 [1606.06986].

[26] S. Gariazzo, P.F. de Salas and S. Pastor, *Thermalisation of sterile neutrinos in the early Universe in the 3+1 scheme with full mixing matrix*, *JCAP* **07** (2019) 014 [1905.11290].

[27] K. Akita and M. Yamaguchi, *A precision calculation of relic neutrino decoupling*, *JCAP* **08** (2020) 012 [2005.07047].

[28] J. Froustey, C. Pitrou and M.C. Volpe, *Neutrino decoupling including flavour oscillations and primordial nucleosynthesis*, *JCAP* **12** (2020) 015 [2008.01074].

[29] J.J. Bennett, G. Buldgen, P.F. De Salas, M. Drewes, S. Gariazzo, S. Pastor et al., *Towards a precision calculation of $N_{\rm eff}$ in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED*, *JCAP* **04** (2021) 073 [2012.02726].

[30] M. Cielo, M. Escudero, G. Mangano and O. Pisanti, *Neff in the Standard Model at NLO is 3.043*, *Phys. Rev. D* **108** (2023) L121301 [2306.05460].

[31] S. Esposito, G. Mangano, G. Miele, I. Picardi and O. Pisanti, *Neutrino energy loss rate in a stellar plasma*, *Nucl. Phys. B* **658** (2003) 217 [astro-ph/0301438].

[32] G. Jackson and M. Laine, *QED corrections to the thermal neutrino interaction rate*, 2312.07015.

[33] O. Tomalak and R.J. Hill, *Theory of elastic neutrino-electron scattering*, *Phys. Rev. D* **101** (2020) 033006 [1907.03379].

[34] R.J. Hill and O. Tomalak, *On the effective theory of neutrino-electron and neutrino-quark interactions*, *Phys. Lett. B* **805** (2020) 135466 [1911.01493].

[35] CMB-S4 collaboration, *CMB-S4 Science Book, First Edition*, 1610.02743.

[36] N.P. Landsman and C.G. van Weert, *Real and Imaginary Time Field Theory at Finite Temperature and Density*, *Phys. Rept.* **145** (1987) 141.

[37] K.-c. Chou, Z.-b. Su, B.-l. Hao and L. Yu, *Equilibrium and Nonequilibrium Formalisms Made Unified*, *Phys. Rept.* **118** (1985) 1.

[38] J. Berges, *Introduction to nonequilibrium quantum field theory*, *AIP Conf. Proc.* **739** (2004) 3 [hep-ph/0409233].

[39] J.I. Kapusta and C. Gale, *Finite-temperature field theory: Principles and applications*, Cambridge Monographs on Mathematical Physics, Cambridge University Press (2011), 10.1017/CBO9780511535130.

[40] C. Coriano and R.R. Parwani, *The Three loop equation of state of QED at high temperature*, *Phys. Rev. Lett.* **73** (1994) 2398 [hep-ph/9405343].

[41] R.R. Parwani and C. Coriano, *Higher order corrections to the equation of state of QED at high temperature*, *Nucl. Phys. B* **434** (1995) 56 [hep-ph/9409269].

[42] G. Sigl and G. Raffelt, *General kinetic description of relativistic mixed neutrinos*, *Nucl. Phys. B* **406** (1993) 423.

[43] L. Canetti, M. Drewes and M. Shaposhnikov, *Matter and Antimatter in the Universe*, *New J. Phys.* **14** (2012) 095012 [1204.4186].

[44] S. Antusch, E. Cazzato, M. Drewes, O. Fischer, B. Garbrecht, D. Gueter et al., *Probing Leptogenesis at Future Colliders*, *JHEP* **09** (2018) 124 [1710.03744].

[45] J.S. Schwinger, *Brownian motion of a quantum oscillator*, *J. Math. Phys.* **2** (1961) 407.

[46] L.V. Keldysh, *Diagram technique for nonequilibrium processes*, *Zh. Eksp. Teor. Fiz.* **47** (1964) 1515.

[47] B. Garbrecht and M. Garny, *Finite Width in out-of-Equilibrium Propagators and Kinetic Theory*, *Annals Phys.* **327** (2012) 914 [1108.3688].

[48] M. Drewes, S. Mendizabal and C. Weniger, *The Boltzmann Equation from Quantum Field Theory*, *Phys. Lett. B* **718** (2013) 1119 [1202.1301].

[49] E. Braaten and D. Segel, *Neutrino energy loss from the plasma process at all temperatures and densities*, *Phys. Rev. D* **48** (1993) 1478 [hep-ph/9302213].

[50] E.W. Kolb and M.S. Turner, *The Early Universe*, vol. 69 (1990), 10.1201/9780429492860.

[51] E. Braaten, *Neutrino emissivity of an ultrarelativistic plasma from positron and plasmino annihilation*, *Astrophys. J.* **392** (1992) 70.

[52] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[53] A. Peshier, K. Schertler and M.H. Thoma, *One loop selfenergies at finite temperature*, *Annals Phys.* **266** (1998) 162 [hep-ph/9708434].

[54] M.E. Carrington, D.-f. Hou and M.H. Thoma, *Equilibrium and nonequilibrium hard thermal loop resummation in the real time formalism*, *Eur. Phys. J. C* **7** (1999) 347 [hep-ph/9708363].

[55] M.L. Bellac, *Thermal Field Theory*, Cambridge Monographs on Mathematical Physics, Cambridge University Press (3, 2011), 10.1017/CBO9780511721700.

[56] A. Rebhan, *Hard thermal loops and QCD thermodynamics*, in *Cargese Summer School on QCD Perspectives on Hot and Dense Matter*, pp. 327–351, 11, 2001 [hep-ph/0111341].

[57] M. Vanderhaeghen, J.M. Friedrich, D. Lhuillier, D. Marchand, L. Van Hoorebeke and J. Van de Wiele, *QED radiative corrections to virtual Compton scattering*, *Phys. Rev. C* **62** (2000) 025501 [hep-ph/0001100].

[58] E. Braaten and R.D. Pisarski, *Soft Amplitudes in Hot Gauge Theories: A General Analysis*, *Nucl. Phys. B* **337** (1990) 569.

[59] J. Ghiglieri, A. Kurkela, M. Strickland and A. Vuorinen, *Perturbative Thermal QCD: Formalism and Applications*, *Phys. Rept.* **880** (2020) 1 [2002.10188].

Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

Z' boson mass reach and model discrimination at muon colliders

Kateryna Korshynska[a,1,2,3], **Maximilian Löschner**[b,1], **Mariia Marinichenko**[c,1,2,4], **Krzysztof Mękała**[d,1,5], **Jürgen Reuter**[e,1]

[1]Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany
[2]Department of Physics, Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Street, 01601 Kyiv, Ukraine
[3]Fundamentale Physik für Metrologie FPM, Physikalisch-Technische Bundesanstalt PTB, Bundesallee 100, 38116 Braunschweig, Germany
[4]Instituut-Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[5]Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warszawa, Poland

arXiv:2402.18460v1 [hep-ph] 28 Feb 2024

Received: date / Accepted: date

Abstract We study the discrimination power of future multi-TeV muon colliders for a large set of models with extended gauge symmetries and additional neutral gauge bosons ("Z'-models"). Our study is carried out using a χ^2-analysis of leptonic observables of s-channel scattering in effective Z'-models. We make use of angular and chiral asymmetries induced in such models to find the discovery reach of a given muon collider setup in terms of the Z' mass and to discriminate between the different scenarios. In this context, we discuss how polarized beams – should they become available at muon colliders – or polarization measurements can help in the discrimination. Our results show that typical muon collider setups which are currently under consideration can give a significantly higher reach compared to existing bounds and projections for high-luminosity LHC.

1 Motivation

The Standard Model (SM) of particle physics is among the most well-tested theories to date and can explain experimental observations to unprecedented precision despite thorough searches for deviations at modern collider experiments. Nonetheless, several well-known phenomena such as the nature of dark matter, the baryon-antibaryon asymmetry of the universe or even the universal law of gravity lack an explanation within this theory. To accommodate for effects Beyond the Standard Model (BSM), multiple extensions of the current formulation have been proposed. If such extensions are based on new gauge symmetries, the existence of additional neutral gauge bosons (which we denote as Z') is the natural consequence. In the absence of a unique well-motivated ultra-violet (UV) complete BSM theory, simplified models or effective theories can be useful tools to study potential deviations from SM predictions in a generic way. An example of this is the effective Z'-model with new gauge boson couplings to the SM fermions, as discussed e.g. in [1–3]. The idea is that an experimental determination of such couplings could eventually give guidance toward a specific UV-complete model which in turn might help explain other observational phenomena.

Recent measurements at the LHC have excluded the existence of a Z' boson with a mass up to about 5 TeV [4–11], which will be raised to close to 8 TeV [12]. Several ways to search for neutral gauge bosons have also been considered for future e^+e^- colliders which, mainly due to a cleaner collision environment, could probe masses up to about 20 TeV, depending on the collider energy [2, 13–22]. Given recent progress in the development of the muon collider accelerator design, the possibility of searching for heavy particles at such a multi-TeV machine is an appealing idea and has already been considered in [23] where several resonance production channels have been studied.

A multi-TeV muon collider has been proposed as the next energy-frontier collider in high-energy physics, combining the clean environment of e^+e^- machines with the energy reach of hadron colliders. Recent technological progress, especially in muon cooling [24], has led to a high priority in the US Snowmass Community Summer Study [25–28] and the Particle Physics Project Prioritization Panel (P5) report [29]. In Europe, the International Muon Collider Collaboration (IMCC) is leading the European effort, suggested in the last European Particle Physics Strategy Update (EPPSU 2020) [30, 31].

A comprehensive, systematic study of the indirect discovery reach of future muon colliders for heavy neutral gauge bosons originating from different unification scenar-

[a]*kateryna.korshynska@desy.de*
[b]*maximilian.loeschner@desy.de*
[c]*mariia.marinichenko@desy.de*
[d]*krzysztof.mekala@desy.de*
[e]*juergen.reuter@desy.de*

ios has been missing in the literature. We address the issue in this paper by proposing a framework based on leptonic observables suitable not only for searches for a Z' boson but also for specifying its nature. The work is structured as follows: in Section 2, we review models introducing additional neutral gauge bosons and in Section 3, we explain why future muon colliders may be used to search for them. In Section 4, we present our analysis procedure whose results are discussed in Section 5. The most important findings of the work are summarized in Section 6.

2 Theoretical background

In this section, we present the models considered in our study which contain additional neutral gauge bosons beyond the SM and review the relevant aspects of their phenomenology. Such a collection can of course never be complete, but we give a well-defined selection of different kinds of weakly and strongly-coupled models to cover theory space as comprehensively as possible.

2.1 Gauge bosons beyond the Standard Model

The Standard Model can be formulated as the most general renormalizable quantum field theory invariant under the gauge group $\mathcal{G}_{SM} = SU(3)_c \times SU(2)_L \times U(1)_Y$ with the matter content discovered as of today. The symmetry under this group gives rise to the existence of the well-known gauge bosons: $g, W^{\pm}, Z, \gamma$. Leaving gravity aside, it remains an open question whether $\mathcal{G}_{SM}$ represents the complete symmetry group responsible for what we can observe in Nature. It is possible that it merely represents the broken version of an enlarged symmetry group or *Grand Unified Theory* (GUT) [32–35] which in turn could lead to the existence of additional gauge bosons among other new particles. In general, any extension of the SM involving the introduction of additional gauge symmetries might extend the gauge boson sector [2]. The rich phenomenology of such models, reviewed for example in [3], could be connected to flavor non-universality [36, 37], neutrino masses [38–43] or the dark-matter problem [44–48].

Since no direct hint towards a single specific model with enlarged gauge symmetry has been experimentally found to date, we study a variety of models in a general way, namely in terms of the phenomenology of additional heavy neutral gauge bosons they would ensue. Therefore, we use an effective Z'-model where we introduce a Z' with couplings to the SM fermions that are the only trace of New Physics. In this model, the interactions of the neutral gauge bosons with the SM fermions are given by the generic neutral-current Lagrangian: [2, 3, 22, 49]:

$$\begin{aligned} -\mathcal{L}_{NC} &= eA_\mu J^\mu_A + g_Z Z_\mu J^\mu_Z + g_{Z'} Z'_\mu J^\mu_{Z'}, \\ J^\mu_A &= \sum_f \bar{f}\gamma^\mu q_f f, \\ J^\mu_Z &= \sum_f \bar{f}\gamma^\mu (v^{SM}_f - \gamma_5 a^{SM}_f) f, \\ J^\mu_{Z'} &= \sum_f \bar{f}\gamma^\mu (v_f - \gamma_5 a_f) f, \end{aligned} \tag{1}$$

where f denotes the SM fermions, A is the photon, Z is the SM neutral gauge boson, Z' is the new heavy neutral boson, e is the positron charge, g_Z and $g_{Z'}$ are the gauge couplings of the corresponding bosons, q_f is the fermion electric charge in units of e, v^{SM}_f and a^{SM}_f are the vector and axial-vector coupling of the fermion to the SM neutral boson and v_f and a_f are the couplings to the new neutral boson. Note that the splitting of the couplings into a prefactor $g_{Z'}$ and vector-/axial-vector components v_f, a_f, while customary for the SM, is rather arbitrary for a generic Z' model. Our convention will become clear below.

2.2 The set of Z'-models considered

In our analysis, we consider a vast set of BSM models, including the Sequential Standard Model (denoted as SSM) [50–52], the E6 Model (E6) [53, 54], the Left-Right Symmetric Model (LR) [33, 43, 55, 56], the Alternative Left-Right Model (ALR) [57, 58], the Littlest Higgs Model (LH) [59–61], the Universal Simplest Little Higgs Model (USLH) [62–64], and the $U(1)_X$ Model [65, 66]. While the SSM cannot be embedded into a renormalizable GUT-like model, it has become a standard candle for Z' searches with over-optimistically large couplings; for this reason, we keep it in our selection of models. Many early GUT models that were inspired by an embedding into a low-energy superstring action exhibit E_6 as a GUT gauge group. Several breaking scenarios via intermediate groups like $SU(5)$ [32], $SO(10)$ [67], Pati-Salam $SU(4) \times SU(2) \times SU(2)$ [33], or trinification $SU(3)^3$ [68] have been studied, and the rank reduction from six to four for the SM gives rise to two potential Z' candidates. The simplest class of models is left-right (LR)-symmetric models, which appear as low-scale effective theories of Pati-Salam or trinification likewise. Multi-step breaking with several GUT scales has been considered [69, 70], and so have the effects of mixing between different Z' bosons, both on model discrimination and reconstruction of the UV-scale model [71, 72]. Besides weakly-coupled, GUT-inspired models, there is the large class of models of compositeness or partial compositeness where some of the additional symmetries in the strongly-coupled sector have been gauged to avoid too light or massless Nambu-Goldstone models. One class of models is Little Higgs models, while another class of strongly-

Model	$g_{Z'}$	$2v_l$	$2a_l$
SSM	$\frac{e}{s_W c_W}$	$2s_W^2 - \frac{1}{2}$	$-\frac{1}{2}$
E_6	$\frac{e}{c_W}$	$\frac{2\cos\beta}{\sqrt{6}}$	$\frac{\cos\beta}{\sqrt{6}} + \frac{\sqrt{10}\sin\beta}{6}$
LR	$\frac{e}{c_W}$	$\frac{1}{\alpha} - \frac{\alpha}{2}$	$\frac{\alpha}{2}$
ALR	$\frac{e}{s_W c_W \sqrt{1-2s_W^2}}$	$\frac{5}{2}s_W^2 - 1$	$-\frac{1}{2}s_W^2$
LH	$\frac{e}{s_W}$	$-\frac{c}{4s}$	$-\frac{c}{4s}$
USLH	$\frac{e}{c_W \sqrt{3-4s_W^2}}$	$\frac{1}{2} - 2s_W^2$	$\frac{1}{2}$
$U(1)_X$	$\frac{e}{4c_W}$	-8	2

Table 1: The Z' couplings to leptons in the models considered in this paper. The sine of the Weinberg angle is denoted by s_W. For the description of the models and the explanation of the model-specific parameters see the main text. The table and notation are adapted from [22].

coupled models via their dual description to warped extra dimensions [73] directly interpolates into models with additional space-time dimensions like e.g. universal extra dimensions [74, 75].

For simplicity, we assume that all the considered models are flavor-universal, *i.e.* the couplings do not differ between the three fermion generations. The scenarios considered are listed in Table 1 where we present the respective axial and vector couplings of the Z' to the fermions together with the absolute normalization of the couplings, $g_{Z'}$. For the E_6 models, the specific values of the mixing angle $\beta = 0,\ \pi/2,\ \arctan(-\sqrt{5/3})$ correspond to the so-called χ, ψ and η models. We use $\beta = 0$ in our analysis as a representative case. For the LR model, the mixing angle between the two $SU(2)$ groups has to be in the range $\sqrt{2/3} \leq \alpha \leq \sqrt{c_W^2/s_W^2 - 1}$, and we use the upper bound for α. For the LH model, the mixing angles between the two $SU(2)$ groups obey $1/10 \leq c/s \leq 2$, and we use $c/s \equiv 1$.

2.3 Observables

The models under consideration specify different points in the parameter space of axial and vector couplings to fermions. Therefore, we study observables which encode angular and chiral properties in order to distinguish between the different scenarios. The analysis is particularly simple in leptonic scattering channels because all of the models considered are flavor universal so that only two coupling factors enter the scattering amplitudes (as opposed to the production of hadrons where leptonic and hadronic axial and vector couplings would play a role, and couplings of up- and down-type quarks would be overlaid). This makes it possible to study the models in the two-dimensional a_l-v_l plane with the only other unknown being the mass of the Z' (in principle, another additional parameter would be the Z' width, but we will only consider Z' states heavier than the collider energy such that the width does not have a significant impact on any observable). The experimental observables we consider are:

1. the total cross-section for the process $\mu^+\mu^- \to f\bar{f}$, denoted as σ^f, for $f \in \{e, \tau\}$ (we do not consider μ final states due to the contamination from t-channel exchange);
2. the forward-backward asymmetry, defined as:

$$A_{FB}^f = \frac{\sigma_F^f - \sigma_B^f}{\sigma^f}, \tag{2}$$

   where, for $f \in \{e, \tau\}$:
   - σ_F^f – the partial cross section for the fermion f going in the forward direction,

$$\sigma_F^f = \int_0^1 d\cos\theta \frac{d\sigma}{d\cos\theta}(\mu^-\mu^+ \to f\bar{f}) \tag{3}$$

   - σ_B^f – the partial cross section for the fermion f going in the backward direction;

$$\sigma_B^f = \int_{-1}^0 d\cos\theta \frac{d\sigma}{d\cos\theta}(\mu^-\mu^+ \to f\bar{f}) \tag{4}$$

3. the left-right asymmetry, defined as:

$$A_{LR}^f = \frac{\sigma_{LR}^f - \sigma_{RL}^f}{\sigma^f}, \tag{5}$$

   where, for $f \in \{e, \tau\}$:
   - $\sigma_{LR}^f = \sigma(\mu_L^-\mu_R^+ \to f\bar{f})$ – the partial cross section for the fully left-polarized muon beam and right-polarized antimuon beam,
   - $\sigma_{RL}^f = \sigma(\mu_R^-\mu_L^+ \to f\bar{f})$ – the partial cross section for the fully right-polarized muon beam and left-polarized antimuon beam;

   We will comment on the possibility of polarization at the muon collider below.
4. the polarization asymmetry, defined as:

$$A_{pol}^f = \frac{\sigma_{lh}^f - \sigma_{rh}^f}{\sigma^f}, \tag{6}$$

   where, for $f = \tau$:
   - $\sigma_{lh}^f = \sigma(\mu^-\mu^+ \to \tau_L^-\tau_R^+)$ – the partial cross section for the left-handed fermions in the final state,
   - $\sigma_{rh}^f = \sigma(\mu^-\mu^+ \to \tau_R^-\tau_L^+)$ – the partial cross section for the right-handed fermions in the final state.

Note that for massless fermions, $A_{pol}^f = A_{LR}^f$ holds [2] but we keep them as two independent quantities due to their vastly different experimental measurement. This will manifest itself in our statistical analysis via different systematic uncertainties.

In accelerator physics, it is not possible to fully polarize lepton beams and Eq. (5) should be corrected by a factor of the effective polarization

$$P_{\text{eff}} = \frac{P^+ + P^-}{P^+P^- + 1}, \tag{7}$$

where P^- (P^+) is the polarization fraction for the (anti-) muon beam and we assume opposite polarizations for the two beams. Analogously, fully perfect flavor tagging and τ polarization measurements are not possible and their efficiencies contribute to the systematic uncertainty of the study. Since the Muon Collider project is still in a preliminary phase, we tackle the issue by postulating an overall systematic uncertainty to 1% which roughly matches the order of magnitude of the statistical uncertainty. The polarization asymmetry is an exception though, because its uncertainty within the LEP measurements is much larger than other systematic uncertainty [76, 77]. Therefore, we conservatively assume a systematic uncertainty of 5% here. The impact of assuming different values is briefly discussed in Appendix A.

In the current scope of the analysis, we ignore additional information from studying hadronic observables. Their usage provides valuable input for the model discrimination; however, in view of the leptonic production channel at the muon collider, disentangling different products of couplings entering the expressions of our observables, including tagging and separating light up- and down-quark flavors, requires additional assumptions in our statistical analysis. This is beyond the scope of the present study. There is recent progress in tagging light-quark flavors using final-state QED radiation (see e.g. [78]), which would enable to partially disentangle these light-quark flavor couplings. We will extend the study towards hadronic observables in the future.

3 High-energy muon collider setup

For a sustainable future in high-energy collider physics, long-term planning is unavoidable. The high-luminosity upgrade of the Large Hadron Collider (HL-LHC) is now approved [79] which, after a decade of running, could run contemporarily with the start-up phase of or will be followed by an e^+e^- Higgs factory [29, 31, 80–85]. On a 20-year time scale, ambitious projects for parton collisions at energies of ~ 10 TeV could be realized either with a 100-TeV hadron machine [86–88] or a multi-TeV muon collider [30]. Since the latter has recently regained attention due to the successful demonstration of the muon-cooling principles [89], we decided to consider this proposal in our study, as already alluded to in the introduction.

Muon colliders would offer a versatile environment for both precision studies of SM and BSM phenomenology and high-energy searches, including the possible occurrence of new physics in electroweak interactions. Muons can be efficiently accelerated in circular machines, as they are more than 200 times more massive than electrons which significantly reduces bremsstrahlung. On the other hand, contrary to protons, muons are elementary, point-like particles offering a much cleaner collision environment[1]. Nevertheless, their finite lifetime poses a challenge for the design of both the accelerator complex and the detector. For brevity, we will simplify the discussion of the experimental effects in the following, assuming global systematic uncertainties only and leaving the meticulous study of the experimental conditions to the time when the final detector designs are available.

The muon collider community currently aims at achieving a collision energy of 10 TeV with a future machine [30]. An initial stage of 3 TeV is foreseen on the path towards the targeted energy and possible extensions of the project are not excluded if technology permits. As for the current design of the accelerator complex, the integrated luminosity, $\mathscr{L}_{\text{int}}$, scales with the square of the collision energy, E_{CM}. The results presented below scale mostly trivially with the integrated luminosity, and different running scenarios can be easily deduced. In our analysis, we assume that a 10-TeV muon collider would deliver 10 ab^{-1} of data which can be extrapolated to other energies by taking:

$$\mathscr{L}_{\text{int}}(E_{\text{CM}}) = 10\,\text{ab}^{-1}\left(\frac{E_{\text{CM}}}{10\,\text{TeV}}\right)^2. \tag{8}$$

The default setup of the muon collider does not assume polarization of the muon beams. However, due to the production of muons from pion decays there is a certain level of polarization inherent in the beams, and circular lepton colliders automatically build up transverse polarization. This could be converted into longitudinal polarization using spin rotators which has been discussed in the technical accelerator reports. Therefore, we will also consider the possibility of 30% polarization of both beams.

4 Analysis procedure

In this section, we present our analysis procedure. We demonstrate a statistical framework that we use to set limits on the Z' masses which could be probed at a future muon collider and show how one can distinguish between different models of New Physics. We carry out the analysis in the Born approximation which was shown to give reliable results in off-peak regions of the relevant observables, as long as appropriate kinematic cuts on photon radiation are applied [2].

[1] Note, however, the electroweak partonic content inside muons when applying a collinear factorization picture to high-energy collisions [90–92]

4.1 Mass reach

Our statistical analysis is based on the χ^2-test statistic:

$$\chi^2(a,v,M_{Z'}) = \sum_{i=1}^{n_{ob}} \left[\frac{O_i^{\rm SM} - O_i(a,v,M_{Z'})}{\Delta O_i^{\rm SM}} \right]^2 + n_{ob}, \tag{9}$$

where n_{ob} is the number of observables used, $O_i^{\rm SM}$ is the value of the i-th observable predicted by the SM, $O_i(a,v,M_{Z'})$ is the value of the observable in a given model (defined uniquely by a pair of axial and vector couplings, (a,v), and the Z' mass, $M_{Z'}$)[2] and $\Delta O_i^{\rm SM}$ is the measurement uncertainty, $\Delta O_i^{\rm SM} = \sqrt{\Delta O_{i,\rm stat}^2 + \Delta O_{i,\rm sys}^2}$ for $\Delta O_{i,\rm stat}$ and $\Delta O_{i,\rm sys}$ being the statistical and systematic uncertainties, respectively. The term "$+n_{ob}$" in Eq. (9) comes about via fluctuations of experimental values around the theory expectations of the $O_i^{\rm SM}$. We show in Appendix B that this simple approach perfectly coincides with the procedure of performing hundreds of pseudo-experiments to mimic real measurements by extracting observables from normal distributions. That this is indeed a meaningful framework can also be seen by assuming that a given model shows no significant discrepancy from the SM (*i.e.* the difference between all the observables is zero by construction) which also means the first term gives no contribution to the total value and the true expected value of the χ^2 distribution can only be restored by the second term.

In the following, we will assume that the statistical uncertainties are given by [14]:

$$\Delta\sigma^f = \frac{\sigma^f}{\sqrt{N_f}}, \tag{10a}$$

$$\Delta A_{FB}^f = \sqrt{\frac{1-\left(A_{FB}^f\right)^2}{N_f}}, \tag{10b}$$

$$\Delta A_{LR}^f = \sqrt{\frac{1-\left(P_{\rm eff}A_{LR}^f\right)^2}{N_f P_{\rm eff}^2}}, \tag{10c}$$

$$\Delta A_{pol}^f = \sqrt{\frac{1-\left(A_{pol}^f\right)^2}{N_f}}, \tag{10d}$$

where $P_{\rm eff}$ is the polarization fraction of the initial state lepton, as defined in Eq. (7), and $N_f = \mathscr{L}_{\rm int}\cdot\sigma^f$ is the number of expected events.

Given the design status of a future muon collider, a complete assessment of the systematic uncertainties is currently not possible. Therefore, for sake of simplicity of the analysis, we will assume that the measurement is statistically limited and the corresponding systematic uncertainties do not exceed significantly the statistical error. Thus, our analysis may be perceived as a hint of what the desired detector performance should be. In Appendix A, we show the impact of varying the systematic errors.

We assume that a model gives predictions distinct from the SM if the value of the χ^2 test exceeds the critical value at the confidence level of 95% for the given number of degrees of freedom, $n_{\rm d.o.f.}$, $\chi^2(a,v,M_{Z'}) > \chi^2_{\rm crit}(n_{\rm d.o.f.})$. In our analysis, we combine seven observables, namely:

1-2. the total cross section, σ^f, for $f \in \{e,\tau\}$,
3-4. the forward-backward asymmetry, A_{FB}^f, for $f \in \{e,\tau\}$,
5-6. the left-right asymmetry, A_{LR}^f, for $f \in \{e,\tau\}$,
7. the polarization asymmetry, A_{pol}^f, for $f = \tau$,

as defined in Sec. 2.3. For the mass reach, we have $\chi^2_{\rm crit}(n_{\rm d.o.f.}=7) = 12.02$ which corresponds to the 90%-quantile because it is a one-sided test, while for the resolution power discussed in the next section, we have $\chi^2_{\rm crit}(n_{\rm d.o.f.}=7) = 14.07$.

4.2 Resolution power

The resolution power measures the compatibility of a fictitious measurement of a parameter pair (a,v) for a fixed $M_{Z'}$ with a given reference Z'-model with couplings $(a_{\rm model}, v_{\rm model})$. Should the measurement of the couplings fall outside the region where $\chi^2_{\rm model} < \chi^2_{\rm crit}(n_{\rm d.o.f.})$, we can distinguish it from the theoretical prediction and thus, exclude the given model at the 95% confidence level. As in this case, our measurement is compared to a particular model, the test statistic given in Eq. (9) should be replaced by:

$$\chi^2_{\rm model}(a,v,M_{Z'}) = \sum_{i=1}^{n_{ob}} \left[\frac{O_i^{\rm model} - O_i(a,v,M_{Z'})}{\Delta O_i^{\rm model}} \right]^2 + n_{ob}, \tag{11}$$

where the only difference comes from the fact that the SM value of an observable is replaced by the observable predicted within the given model, $O_i^{\rm model}$.

5 Results

The results presented in the following two subsections are obtained by determining values for each observable analytically using the full $(\mu^+\mu^- \to \ell^+\ell^-)$-scattering amplitudes for $\ell \in \{e,\tau\}$ in Born approximation shown in Figure 1. We also include the width of the Z' using the Born approximations defined in [2]. The values for $\Gamma(Z' \to \text{any})$ range from ~ 40 GeV (USLH) to ~ 300 GeV (SSM) for $M_{Z'} = 30$ TeV. The exact width of the Z' does not play a major role in the analysis though, because the bounds we find are driven by the off-peak regions of the observables.

[2] For brevity, we will drop the parameters whenever there is no danger of confusion.

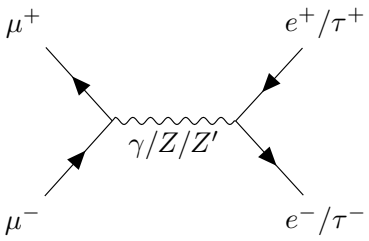


Fig. 1: The set of Feynman diagrams entering our χ^2-analysis.

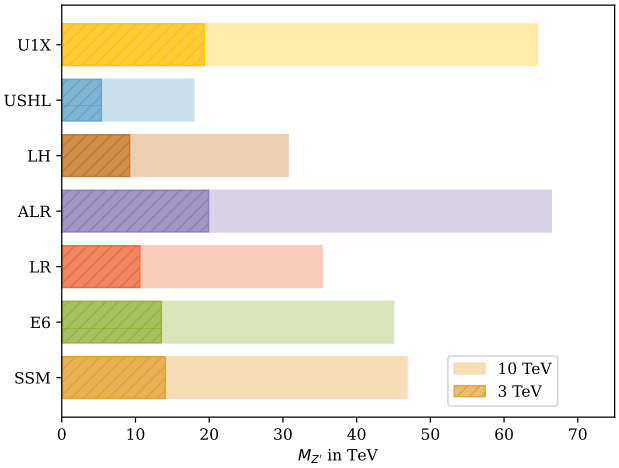


Fig. 2: The reach in $M_{Z'}$ for a 3 TeV and 10 TeV muon collider with luminosities of $0.9\,\text{ab}^{-1}$ and $10\,\text{ab}^{-1}$, respectively for $P_{\text{eff}} = 0$, $\Delta_{i,\text{sys}} = 1\%$ for σ_f, A_{FB}, A_{LR} and $\Delta_{A_{pol},\text{sys}} = 5\%$. The bars correspond to the exclusion limit of the given Z'-model at 95% confidence level.

We use Eq. (8) to set the integrated luminosity. Then, the statistical errors of Eq. (10) are typically of $\mathcal{O}(1\%)$ or lower (for $M_{Z'} = 3\sqrt{s}$). This could serve as a target value for the systematic errors of the prospective collider measurements in order for the precision to be driven by statistical fluctuations rather than systematics.

Cross-checks of the results obtained in this way were carried out in the Monte Carlo event generator framework Whizard 3.1 [93, 94], using the included generic Z'-model implementation as well as its UFO interface [95–97]. Note that the latter can in principle be used for more sophisticated event-level analyses, potentially also beyond the Born approximation [98], or involving detector simulation, but at the cost of higher computational effort. This is beyond the scope of this current study.

5.1 Mass reach

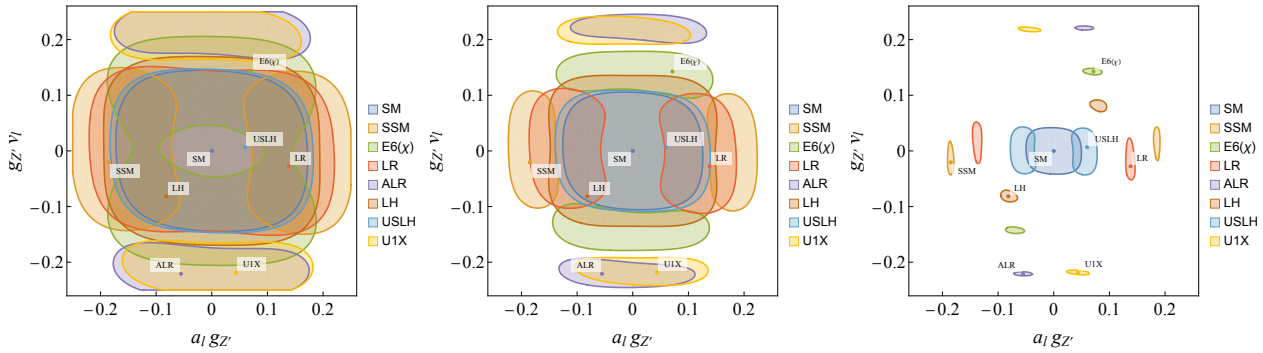
In Figure 2, we show the mass reach we find using the analysis procedure explained in Sec. 4. It shows that using leptonic observables alone, the exclusion limits at 95% C.L. for a 10 TeV muon collider extend up to $\sim$70 TeV, depending on the model. We find relatively low exclusion limits of $\sim$17 TeV for the Little/Littlest Higgs models due to the small magnitude of the leptonic axial and vector couplings[3]. The results shown are without beam polarization as the default setup of the muon collider, and because we find that even for polarization fractions close to 100%, the limits do not change significantly, *i.e.* only by up to 3%. This is due to the limited statistical significance of A_{LR}, *i.e.* a higher statistical error compared to A_{FB} (see Eq. (10)). The influence of the polarization fraction becomes important for the model discrimination though, as explained in the following section.

The discussion of [14] for e^+e^- colliders shows that the inclusion of hadronic observables can increase the reach by up to $\sim$ 50%, depending on the model. An example is the SSM, where one has axial and vector couplings to up- and down-type quarks of relatively large magnitude. The effect is much less drastic for models with smaller quark couplings though, such as the $U(1)_X$. For muon colliders, we expect a similar increase in mass reach when including hadronic observables.

5.2 Resolution power

Here, we determine the resolution power of a 10 TeV muon collider in the off-peak region of the observables considered, *i.e.* for masses above the collider energy. A study in the peak region would require a different type of analysis and for even smaller masses, a study for precision e^+e^- machine would be more sensible. In Figure 3, we show the resolution power for Z'-axial and -vector couplings, given that a signal for a Z' has been found at a mass $M_{Z'}$ of either 40 TeV, 30 TeV or 15 TeV. We find that one would be able to discriminate to a good extent between the models considered at around $M_{Z'} \sim 30\,\text{TeV} = 3\sqrt{s}$ or lower. Already for masses slightly higher than that, the discrimination power starts to decrease rather rapidly. On the other hand, it grows very fast for lower masses because the closer the Z'-mass is to the collider energy, the more one approaches the Z'-pole where deviations from the SM become substantial. Then, models with axial or vector couplings of comparatively large magnitude can be discriminated even without the use of polarized beams or a very accurate measurement of the τ polarization. In Appendix A, we discuss the influence of the latter measurements in more detail.

[3] Note that in such a case also the current LHC bounds are weaker as the signal strengths in Drell-Yan searches is reduced.



(a) $M_{Z'} = 40\,\text{TeV}$ (b) $M_{Z'} = 30\,\text{TeV}$ (c) $M_{Z'} = 15\,\text{TeV}$

Fig. 3: Resolution power for different masses of the Z'. Inputs: $\mathscr{L}_{\text{int}} = 10\,\text{ab}^{-1}$, $E_{\text{CM}} = 10\,\text{TeV}$, $P_{\text{eff}} = 0$.

6 Conclusions

A Z' boson occurs in a variety of BSM theories that are based on extensions of the SM gauge group. We showed that at a muon collider, using indirect search methods for leptonic observables alone, one would be able to probe Z' masses of up to $\sim 70\,\text{TeV}$. For masses up to $\sim 30\,\text{TeV}$, the same framework can be applied to discriminate its nature in terms of its axial and vector couplings to fermions, pointing towards a specific model of New Physics. The presented results give the most stringent limits up to date. Employing hadronic observables has been left for future explorations. They are expected to enhance the discovery reach towards higher Z' masses but are not expected to improve the model discrimination by a lot, as they suffer from combinatorics between light up- and down-type quarks. Final states consisting of charm and bottom will add discrimination power but are more complicated to quantify due to the charm- and bottom-tagging efficiencies. Z' decays into top quarks are highly interesting, especially for models where the top coupling is special (e.g. the Littlest Higgs or top-color models), but clearly these final states go beyond the simple two-fermion signatures considered here. Note that another complication is the non-factorization of leptonic production current and hadronic decay current.

A spectacular level of model discrimination is already possible using the default setup of the muon collider, using left-right and τ polarization asymmetries. If the systematic error on the τ polarization measurement dominates over its statistical uncertainty significantly, beam polarization of a degree of up to 30% via spin rotators could enhance the discriminative power as it allows to use left-right asymmetries as well.

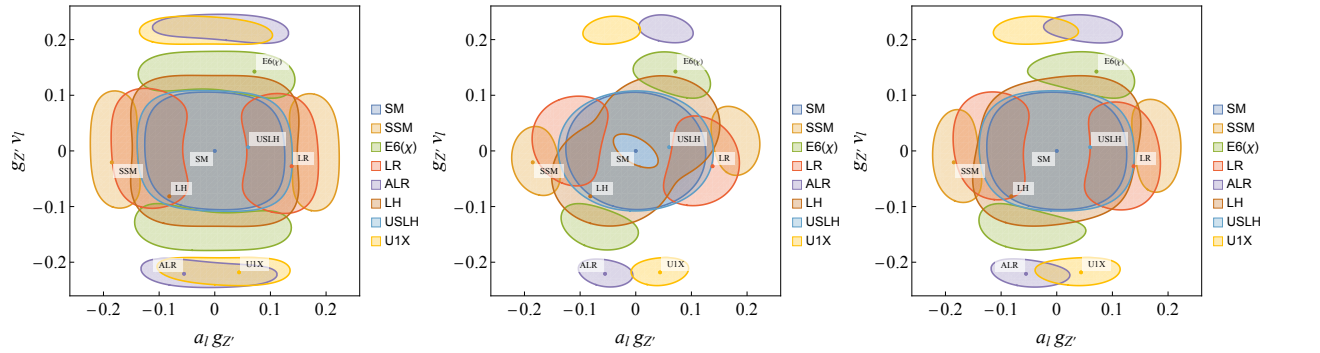
As statistical uncertainties can be expected to be of the order of 1%, we show that it is crucial to have the systematic uncertainties under control in order to guarantee a good model discrimination. This means that systematic uncertainties from lepton charge determination, angular resolution, τ polarization measurements etc. should not exceed a percent by much. This provides a ballpark what future muon collider detector development should aim for.

Acknowledgments

The work of KM and JRR was partially supported by the National Science Centre (Poland) under the OPUS research project no. 2021/43/B/ST2/01778. ML, KM and JRR acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Association) under Germany's Excellence Strategy-EXC 2121 "Quantum Universe"-390833306. KK and MM acknowledge Deutsches Elektronen-Synchrotron (DESY) for a hospitable environment during the DESY Ukraine Winter School. This work has also been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491245950. Furthermore, we acknowledge support from the COMETA COST Action CA22130.

Appendix A: Parametric variations of resolution power

In this Appendix, we present a list of plots for the discrimination power in the a_l-v_l plane of Z' models for variations of different input values. In Figure 4, we show that adding a polarization measurement or polarized beams to the analysis enables the distinction of models with different relative signs between the axial and vector couplings, e.g. between the ALR and $U(1)_X$ models. A simultaneous sign change of the axial and vector couplings would remain undetectable, because the left-right asymmetry A_{LR} only adds sensitivity to the product $a_l \cdot v_l$ [2]. Comparing Figure 4b to Figure 4c, we see that one can achieve similar resolutions by either bringing the systematic uncertainty on the τ polarization down to the 1%-level or by using polarized muon beams with $P_{\text{eff}} = 30\%$. Moreover, we find that a further increase



(a) $P_{\text{eff}} = 0\%$, $\Delta_{\text{sys}}(A_{pol}) = 5\%$ (b) $P_{\text{eff}} = 0\%$, $\Delta_{\text{sys}}(A_{pol}) = 1\%$ (c) $P_{\text{eff}} = 30\%$, $\Delta_{\text{sys}}(A_{pol}) = 5\%$

Fig. 4: Resolution power for different combinations of effective polarizations and systematic errors for the τ polarization asymmetry. Inputs: $\mathscr{L}_{\text{int}} = 10\,\text{ab}^{-1}$, $E_{\text{CM}} = 10\,\text{TeV}$, $M_{Z'} = 30\,\text{TeV}$.

in the effective polarization does not yield a significant increase in the resolution power. In Appendix C, we show how the individual observables contribute to the resolution power and present their analytic form in Appendix D.

In Figure 5 we show the variations for three different values of the systematic errors. We find that at $\sim 1\%$, systematic errors become competitive both with the statistical ones by explicit comparisons of the errors magnitudes, as well as by noticing that even a strong reduction in the systematic error does not increase the discrimination power significantly, as then the measurement becomes limited by statistics.

We do not show the variation of other parameters explicitly here, because their influence can be inferred from the variations of the parameters discussed. The variation of the collider energy yields similar changes as shown in Figure 3, namely that the closer the energy is to the mass of the Z', the better the discrimination power will be. A variation of the luminosity solely affects the magnitude of the statistical errors. Therefore, a decrease in luminosity will lead to larger errors and similar behavior as shown in Figure 5.

Appendix B: Test statistics derivation

In this section, we discuss the origin of the shift in our χ^2-distributions by the number of observables, as it appears in Eq. (9) and Eq. (11). Let us call the pseudo-measurement of an observable for a given reference model $\hat{O}_i$ and the corresponding theory prediction O_i with the error ΔO_i and assume that they follow the normal distribution

$$f(\hat{O}_i) = \exp\left[-\frac{(\hat{O}_i - O_i)^2}{\Delta O_i \sqrt{2\pi}}\right] \frac{1}{\Delta O_i \sqrt{2\pi}}. \tag{B.1}$$

This means that the expectation value for $\hat{O}_i$ is

$$\langle \hat{O}_i \rangle = O_i \tag{B.2}$$

and the variance

$$\langle (\hat{O}_i - O_i)^2 \rangle = \Delta O_i^2. \tag{B.3}$$

We can also expand the variance:

$$\langle (\hat{O}_i - O_i)^2 \rangle = \langle \hat{O}_i^2 \rangle - 2\langle \hat{O}_i \rangle O_i + O_i^2 = \langle \hat{O}_i^2 \rangle - O_i^2, \tag{B.4}$$

and therefore rewrite:

$$\langle \hat{O}_i^2 \rangle = O_i^2 + \Delta O_i^2. \tag{B.5}$$

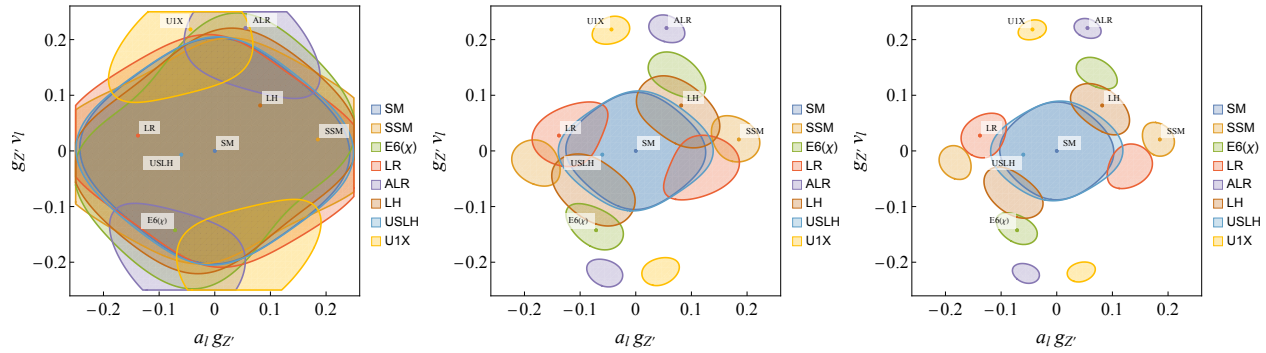
If we now determine the expectation value for χ^2 to test a model with observables O_i^{test} and insert Eq. (B.5), we find

$$\langle \chi^2 \rangle = \sum_{i=1}^{n_{ob}} \left\langle \frac{(\hat{O}_i - O_i^{\text{test}})^2}{\Delta O_i^2} \right\rangle = \sum_{i=1}^{n_{ob}} \frac{(O_i - O_i^{\text{test}})^2}{\Delta O_i^2} + n_{ob}, \tag{B.6}$$

where the shift by n_{ob} appears due to the second term in Eq. (B.5) and explains why it is included in Eq. (9) and Eq. (11). This is equivalent to the outcome when averaging over a large number of pseudo-measurements when drawing values from Eq. (B.1) randomly using a Monte Carlo algorithm.

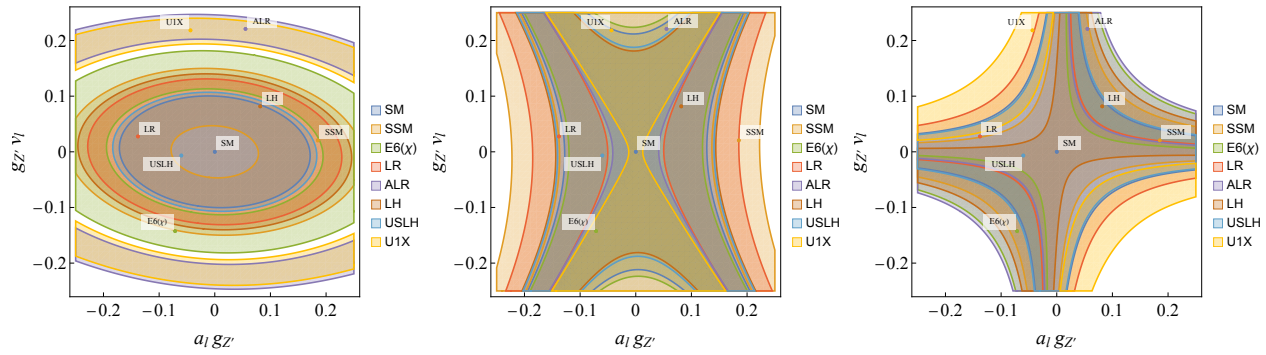
Appendix C: Resolution power of individual observables

In Figure 6, we show the resolution power of the individual observables σ_{tot}, A_{FB} and A_{LR}. The total cross section gives elliptic resolution regions while the two asymmetries give distinct hyperbolic regions. This can be understood by the dependence of the observables on either the sum of the squares of axial and vector couplings (σ_{tot}), the difference of the squares (A_{FB}) or their product (A_{LR}) in the Born approximation [14]. In comparison to the approximate bounds for the observables used in [14], we see slight asymmetries in the bounding regions as a result of using the full Born amplitudes. They are nevertheless in very good agreement with the approximate bounds.



(a) $\Delta_{\mathrm{sys},i} = 5\%$ (b) $\Delta_{\mathrm{sys},i} = 1\%$ (c) $\Delta_{\mathrm{sys},i} = 0.1\%$

Fig. 5: Resolution power for different systematic errors. Inputs: $\mathscr{L}_{\mathrm{int}} = 10\,\mathrm{ab}^{-1}$, $E_{\mathrm{CM}} = 10\,\mathrm{TeV}$, $M_{Z'} = 30\,\mathrm{TeV}$, $P_{\mathrm{eff}} = 80\%$.



(a) σ_{tot} (b) A_{FB} (c) A_{LR}

Fig. 6: Resolution power of individual observables. Inputs: $\mathscr{L}_{\mathrm{int}} = 10\,\mathrm{ab}^{-1}$, $E_{\mathrm{CM}} = 10\,\mathrm{TeV}$, $M_{Z'} = 30\,\mathrm{TeV}$.

Appendix D: Asymmetries

Here we present expressions for left-right and forward-backward asymmetries in terms of individual interference diagrams (see Fig. 1) for a high-energy machine (*i.e.* we neglect the masses of the fermions). The forward-backward asymmetry reads

$$A_{FB}^{f}(s) = \frac{1}{8\,\pi\,s\,\sigma^{f}(s)}\Big\{2g_Z^4 a_\mu^{SM} a_f^{SM} v_\mu^{SM} v_f^{SM}|\chi_Z|^2 + 2g_{Z'}^4 a_\mu a_f v_\mu v_f |\chi_{Z'}|^2 + q_\mu q_f e^2\Big(g_Z^2 a_\mu^{SM} a_f^{SM}\mathrm{Re}(\chi_\gamma\chi_Z^*) + g_{Z'}^2 a_\mu a_f \mathrm{Re}(\chi_\gamma\chi_{Z'}^*)\Big) + g_Z^2 g_{Z'}^2 (a_\mu^{SM} v_\mu + a_\mu v_\mu^{SM})(a_f^{SM} v_f + a_f v_f^{SM})\mathrm{Re}(\chi_Z\chi_{Z'}^*)\Big\}, \tag{D.7}$$

the left-right asymmetry is

$$\begin{aligned} A_{LR}^{f}(s) = \frac{1}{12\,\pi\,s\,\sigma^{f}(s)}\Big\{ & g_Z^4\Big(a_\mu^{SM} v_\mu^{SM}\Big[(a_f^{SM})^2 + (v_f^{SM})^2\Big] + a_f^{SM} v_f^{SM}\Big[(a_\mu^{SM})^2 + (v_\mu^{SM})^2\Big]\Big)|\chi_Z|^2 \\ & + g_{Z'}^4\Big(a_f v_f\Big[a_\mu^2 + v_\mu^2\Big] + a_\mu v_\mu\Big[a_f^2 + v_f^2\Big]\Big)|\chi_{Z'}|^2 \\ & + e^2 q_\mu q_f\Big[g_Z^2(a_f^{SM} v_\mu^{SM} + a_\mu^{SM} v_f^{SM})\mathrm{Re}(\chi_\gamma\chi_Z^*) + g_{Z'}^2(a_f v_\mu + a_\mu v_f)\mathrm{Re}(\chi_\gamma\chi_{Z'}^*)\Big] \\ & + g_Z^2 g_{Z'}^2\Big[(a_\mu^{SM} a_\mu + v_\mu^{SM} v_\mu)(a_f^{SM} v_f + v_f^{SM} a_f) + (a_f^{SM} a_f + v_f^{SM} v_f)(a_\mu^{SM} v_\mu + v_\mu^{SM} a_\mu)\Big]\mathrm{Re}(\chi_Z\chi_{Z'}^*)\Big\}, \end{aligned} \tag{D.8}$$

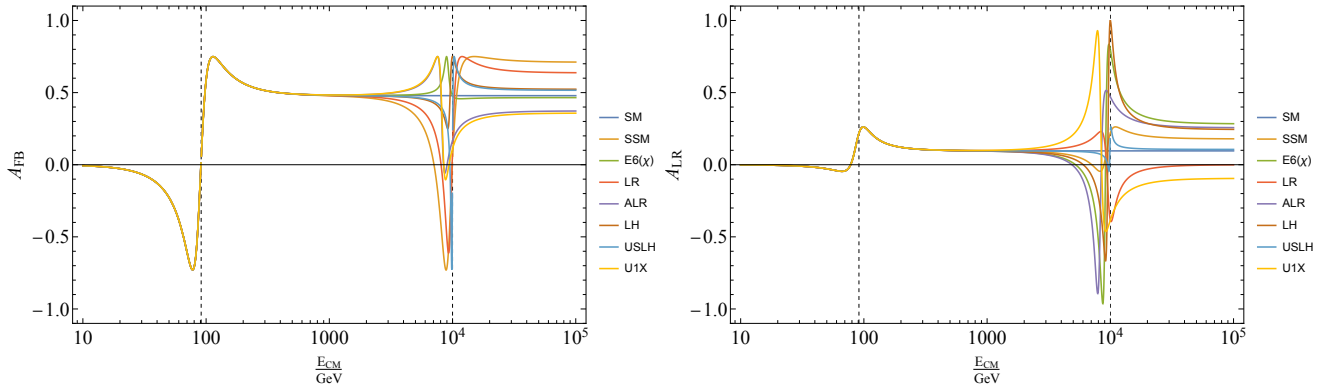


Fig. 7: We show our results for the forward-backward and left-right asymmetry that enter our χ^2-analysis as function of the c.o.m. energy. The dashed lines lie at the mass poles at $M_Z = 91.1882\,\text{GeV}$ and $M_{Z'} = 10\,\text{TeV}$.

and for massless fermions, it holds that $A_{pol} = A_{LR}$. The total cross section in this limit is given by

$$\begin{aligned}\sigma^f(s) = \frac{1}{12\pi s}\Big\{ & g_Z^4\left[(a_\mu^{SM})^2+(v_\mu^{SM})^2)\right]\left[(a_f^{SM})^2+(v_f^{SM})^2\right]|\chi_Z|^2 + g_{Z'}^4(a_\mu^2+v_\mu^2)(a_f^2+v_f^2)|\chi_{Z'}|^2 + e^4 q_\mu^2 q_f^2|\chi_\gamma|^2 \\ & + 2e^2 q_\mu q_f\left[g_Z^2 v_f^{SM} v_\mu^{SM}\mathrm{Re}(\chi_\gamma\chi_Z^*) + g_{Z'}^2 v_f v_\mu \mathrm{Re}(\chi_\gamma\chi_{Z'}^*)\right] \\ & + 2g_Z^2 g_{Z'}^2\left[a_f^{SM}a_f + v_f^{SM}v_f\right]\left[a_\mu^{SM}a_\mu + v_\mu^{SM}v_\mu\right]\mathrm{Re}(\chi_Z\chi_{Z'}^*)\Big\}.\end{aligned} \tag{D.9}$$

Here, $\chi_{Z^{(\prime)}}$ are the gauge bosons propagators times a factor of s to make them dimensionless, *i.e.* the ratio of the photon propagator to considered gauge boson,

$$\chi_{Z^{(\prime)}} = \frac{s}{s + i\Gamma_{Z^{(\prime)}}M_{Z^{(\prime)}} - M_{Z^{(\prime)}}^2}\,, \tag{D.10}$$

and $\chi_\gamma = 1$. This way of writing illustrates which interference of amplitude times conjugate amplitude is each term's origin when squaring the diagrams in Fig. 1. The distributions are shown in Fig. 7. Note that the formulae above do not hold for final state muons because one would have additional t-channel contributions there.

References

1. R. L. Workman, et al., Review of Particle Physics, PTEP 2022 (2022) 083C01. doi:10.1093/ptep/ptac097.
2. A. Leike, The Phenomenology of extra neutral gauge bosons, Phys. Rept. 317 (1999) 143–250. arXiv:hep-ph/9805494, doi:10.1016/S0370-1573(98)00133-1.
3. P. Langacker, The Physics of Heavy Z' Gauge Bosons, Rev. Mod. Phys. 81 (2009) 1199–1228. arXiv:0801.1345, doi:10.1103/RevModPhys.81.1199.
4. M. Aaboud, et al., Search for resonances in diphoton events at $\sqrt{s}$=13 TeV with the ATLAS detector, JHEP 09 (2016) 001. arXiv:1606.03833, doi:10.1007/JHEP09(2016)001.
5. G. Aad, et al., Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, JHEP 03 (2020) 145. arXiv:1910.08447, doi:10.1007/JHEP03(2020)145.
6. G. Aad, et al., Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector, Phys. Lett. B 796 (2019) 68–87. arXiv:1903.06248, doi:10.1016/j.physletb.2019.07.016.
7. A. M. Sirunyan, et al., Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s}$ = 13 TeV, JHEP 05 (2020) 033. arXiv:1911.03947, doi:10.1007/JHEP05(2020)033.
8. G. Aad, et al., Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Eur. Phys. J. C 80 (12) (2020) 1165. arXiv:2004.14636, doi:10.1140/epjc/s10052-020-08554-y.
9. A. M. Sirunyan, et al., Search for resonant and non-resonant new phenomena in high-mass dilepton final states at $\sqrt{s}$ = 13 TeV, JHEP 07 (2021) 208. arXiv:2103.02708, doi:10.1007/JHEP07(2021)208.

10. A. Tumasyan, et al., Search for heavy resonances decaying to WW, WZ, or WH boson pairs in the lepton plus merged jet final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV, Phys. Rev. D 105 (3) (2022) 032008. arXiv:2109.06055, doi:10.1103/PhysRevD.105.032008.
11. V. M. Lozano, R. M. S. Seoane, J. Zurita, Z'-explorer 2.0: Reconnoitering the dark matter landscape, Comput. Phys. Commun. 288 (2023) 108729. arXiv:2109.13194, doi:10.1016/j.cpc.2023.108729.
12. Prospects for searches for heavy Z' and W' bosons in fermionic final states with the ATLAS experiment at the HL-LHC (2018).
13. M. Cvetic, S. Godfrey, Discovery and identification of extra gauge bosons, 1995, pp. 383–415. arXiv:hep-ph/9504216, doi:10.1142/9789812830265_0007.
14. A. Leike, S. Riemann, Z' search in e^+e^- annihilation, Z. Phys. C 75 (1997) 341–348. arXiv:hep-ph/9607306, doi:10.1007/s002880050477.
15. S. Godfrey, P. Kalyniak, A. Tomkins, Distinguishing between models with extra gauge bosons at the ILC, in: 2005 International Linear Collider Physics and Detector Workshop and 2nd ILC Accelerator Workshop, 2005. arXiv:hep-ph/0511335.
16. P. Osland, A. A. Pankov, A. V. Tsytrinov, Identification of extra neutral gauge bosons at the International Linear Collider, Eur. Phys. J. C 67 (2010) 191–204. arXiv:0912.2806, doi:10.1140/epjc/s10052-010-1272-z.
17. V. V. Andreev, G. Moortgat-Pick, P. Osland, A. A. Pankov, N. Paver, Discriminating Z' from Anomalous Trilinear Gauge Coupling Signatures in e+e- \to W+W- at ILC with Polarized Beams, Eur. Phys. J. C 72 (2012) 2147. arXiv:1205.0866, doi:10.1140/epjc/s10052-012-2147-2.
18. T. Han, P. Langacker, Z. Liu, L.-T. Wang, Diagnosis of a New Neutral Gauge Boson at the LHC and ILC for Snowmass 2013 (8 2013). arXiv:1308.2738.
19. S. Godfrey, T. Martin, Z' Discovery Reach at Future Hadron Colliders: A Snowmass White Paper, in: Snowmass 2013: Snowmass on the Mississippi, 2013. arXiv:1309.1688.
20. D. Kapukchyan, T. M. P. Tait, Sensitivity of a Future High Energy e^+e^- Collider to Z' Bosons, J. Phys. G 41 (2014) 075011. arXiv:1312.3377, doi:10.1088/0954-3899/41/7/075011.
21. A. . A. . Pankov, A. . V. Tsytrinov, Model identification of new heavy Z' bosons at ILC with polarized beams, J. Phys. Conf. Ser. 938 (1) (2017) 012059. doi:10.1088/1742-6596/938/1/012059.
22. A. Gulov, Y. Moroz, Optimal observables for Z' models in annihilation leptonic processes, Phys. Rev. D 98 (11) (2018) 115014. arXiv:1711.02853, doi:10.1103/PhysRevD.98.115014.
23. Z. Lu, H. Li, Z.-L. Han, Z.-G. Si, L. Zhao, Phenomenology of Heavy Neutral Gauge Boson at Muon Collider (12 2023). arXiv:2312.17427.
24. S. Jindariani, et al., Promising Technologies and R&D Directions for the Future Muon Collider Detectors (3 2022). arXiv:2203.07224.
25. N. Bartosik, et al., Simulated Detector Performance at the Muon Collider (3 2022). arXiv:2203.07964.
26. D. Stratakis, et al., A Muon Collider Facility for Physics Discovery (3 2022). arXiv:2203.08033.
27. C. Aime, et al., Muon Collider Physics Summary (3 2022). arXiv:2203.07256.
28. J. de Blas, et al., The physics case of a 3 TeV muon collider stage (3 2022). arXiv:2203.07261.
29. 2023 P5 Report: Exploring the Quantum Universe, https://www.usparticlephysics.org/2023-p5-report/ (2023).
30. C. Accettura, et al., Towards a muon collider, Eur. Phys. J. C 83 (9) (2023) 864, [Erratum: Eur.Phys.J.C 84, 36 (2024)]. arXiv:2303.08533, doi:10.1140/epjc/s10052-023-11889-x.
31. 2020 Update of the European Strategy for Particle Physics, CERN Council, Geneva, 2020. doi:10.17181/ESU2020.
32. H. Georgi, S. L. Glashow, Unity of All Elementary Particle Forces, Phys. Rev. Lett. 32 (1974) 438–441. doi:10.1103/PhysRevLett.32.438.
33. J. C. Pati, A. Salam, Lepton Number as the Fourth Color, Phys. Rev. D 10 (1974) 275–289, [Erratum: Phys.Rev.D 11, 703–703 (1975)]. doi:10.1103/PhysRevD.10.275.
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. Slansky, Group Theory for Unified Model Building, Phys. Rept. 79 (1981) 1–128. doi:10.1016/0370-1573(81)90092-2.
36. P. Langacker, M. Plumacher, Flavor changing effects in theories with a heavy Z' boson with family nonuniversal couplings, Phys. Rev. D 62 (2000) 013006. arXiv:hep-ph/0001204, doi:10.1103/PhysRevD.62.013006.
37. S. Baek, J. H. Jeon, C. S. Kim, B0(s) - anti-B0(s) Mixing in Leptophobic Z-prime Model, Phys. Lett. B 641 (2006) 183–188. arXiv:hep-ph/0607113, doi:10.1016/j.physletb.2006.08.041.
38. E. Ma, Neutrino masses in an extended gauge model with E(6) particle content, Phys. Lett. B 380 (1996) 286–290. arXiv:hep-ph/9507348, doi:10.1016/0370-2693(96)00524-2.

39. E. Keith, E. Ma, Efficacious extra U(1) factor for the supersymmetric standard model, Phys. Rev. D 54 (1996) 3587–3593. arXiv:hep-ph/9603353, doi:10.1103/PhysRevD.54.3587.
40. V. Barger, P. Langacker, H.-S. Lee, Primordial nucleosynthesis constraints on Z' properties, Phys. Rev. D 67 (2003) 075009. arXiv:hep-ph/0302066, doi:10.1103/PhysRevD.67.075009.
41. J.-h. Kang, P. Langacker, T.-j. Li, Neutrino masses in supersymmetric SU(3)(C) x SU(2)(L) x U(1)(Y) x U(1)-prime models, Phys. Rev. D 71 (2005) 015012. arXiv:hep-ph/0411404, doi:10.1103/PhysRevD.71.015012.
42. S. F. King, S. Moretti, R. Nevzorov, Theory and phenomenology of an exceptional supersymmetric standard model, Phys. Rev. D 73 (2006) 035009. arXiv:hep-ph/0510419, doi:10.1103/PhysRevD.73.035009.
43. R. N. Mohapatra, J. C. Pati, A Natural Left-Right Symmetry, Phys. Rev. D 11 (1975) 2558. doi:10.1103/PhysRevD.11.2558.
44. B. de Carlos, J. R. Espinosa, Cold dark matter candidate in a class of supersymmetric models with an extra U(1), Phys. Lett. B 407 (1997) 12–21. arXiv:hep-ph/9705315, doi:10.1016/S0370-2693(97)00747-8.
45. S. Nakamura, D. Suematsu, Supersymmetric extra U(1) models with a singlino dominated LSP, Phys. Rev. D 75 (2007) 055004. arXiv:hep-ph/0609061, doi:10.1103/PhysRevD.75.055004.
46. V. Barger, P. Langacker, I. Lewis, M. McCaskey, G. Shaughnessy, B. Yencho, Recoil Detection of the Lightest Neutralino in MSSM Singlet Extensions, Phys. Rev. D 75 (2007) 115002. arXiv:hep-ph/0702036, doi:10.1103/PhysRevD.75.115002.
47. H.-S. Lee, K. T. Matchev, S. Nasri, Revival of the thermal sneutrino dark matter, Phys. Rev. D 76 (2007) 041302. arXiv:hep-ph/0702223, doi:10.1103/PhysRevD.76.041302.
48. G. Belanger, A. Pukhov, G. Servant, Dirac Neutrino Dark Matter, JCAP 01 (2008) 009. arXiv:0706.0526, doi:10.1088/1475-7516/2008/01/009.
49. J. L. Hewett, T. G. Rizzo, Low-Energy Phenomenology of Superstring Inspired E(6) Models, Phys. Rept. 183 (1989) 193. doi:10.1016/0370-1573(89)90071-9.
50. V. D. Barger, W.-Y. Keung, E. Ma, Sequential W and Z Bosons, Phys. Lett. B 94 (1980) 377–380. doi:10.1016/0370-2693(80)90900-4.
51. R. W. Robinett, J. L. Rosner, Prospects for a Second Neutral Vector Boson at Low Mass in SO(10), Phys. Rev. D 25 (1982) 3036, [Erratum: Phys.Rev.D 27, 679 (1983)]. doi:10.1103/PhysRevD.27.679.
52. G. Altarelli, B. Mele, M. Ruiz-Altaba, Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders, Z. Phys. C 45 (1989) 109, [Erratum: Z.Phys.C 47, 676 (1990)]. doi:10.1007/BF01556677.
53. Y. Achiman, B. Stech, Quark-lepton symmetry and mass scales in an e6 unified gauge model, Physics Letters B 77 (4-5) (1978) 389–393.
54. D. London, J. L. Rosner, Extra Gauge Bosons in E(6), Phys. Rev. D 34 (1986) 1530. doi:10.1103/PhysRevD.34.1530.
55. R. N. Mohapatra, G. Senjanovic, Neutrino Mass and Spontaneous Parity Nonconservation, Phys. Rev. Lett. 44 (1980) 912. doi:10.1103/PhysRevLett.44.912.
56. 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.
57. E. Ma, Particle dichotomy and left-right decomposition of e_6 superstring models, Phys. Rev. D 36 (1987) 274–276. doi:10.1103/PhysRevD.36.274.
URL https://link.aps.org/doi/10.1103/PhysRevD.36.274
58. M. Ashry, S. Khalil, Phenomenological aspects of a TeV-scale alternative left-right model, Physical Review D 91 (1) (jan 2015). doi:10.1103/physrevd.91.015009.
URL https://doi.org/10.1103%2Fphysrevd.91.015009
59. N. Arkani-Hamed, A. G. Cohen, E. Katz, A. E. Nelson, The Littlest Higgs, JHEP 07 (2002) 034. arXiv:hep-ph/0206021, doi:10.1088/1126-6708/2002/07/034.
60. T. Han, H. E. Logan, B. McElrath, L.-T. Wang, Phenomenology of the little Higgs model, Phys. Rev. D 67 (2003) 095004. arXiv:hep-ph/0301040, doi:10.1103/PhysRevD.67.095004.
61. W. Kilian, J. Reuter, The Low-energy structure of little Higgs models, Phys. Rev. D 70 (2004) 015004. arXiv:hep-ph/0311095, doi:10.1103/PhysRevD.70.015004.
62. M. Schmaltz, The Simplest little Higgs, JHEP 08 (2004) 056. arXiv:hep-ph/0407143, doi:10.1088/1126-6708/2004/08/056.
63. W. Kilian, D. Rainwater, J. Reuter, Pseudo-axions in little Higgs models, Phys. Rev. D 71 (2005) 015008. arXiv:hep-ph/0411213, doi:10.1103/PhysRevD.71.015008.
64. W. Kilian, D. Rainwater, J. Reuter, Distinguishing little-Higgs product and simple group models at the LHC and ILC, Phys. Rev. D 74 (2006) 095003, [Erratum: Phys.Rev.D 74, 099905 (2006)]. arXiv:hep-ph/0609119, doi:10.1103/PhysRevD.74.095003.
65. S. Oda, N. Okada, D.-s. Takahashi, Classically conformal U(1)' extended standard model and Higgs vacuum stability, Phys. Rev. D 92 (1) (2015) 015026. arXiv:

1504.06291, doi:10.1103/PhysRevD.92.015026.
66. S. Iso, N. Okada, Y. Orikasa, Classically conformal b–l extended standard model, Physics Letters B 676 (1-3) (2009) 81–87.
67. H. Fritzsch, P. Minkowski, Unified Interactions of Leptons and Hadrons, Annals Phys. 93 (1975) 193–266. doi:10.1016/0003-4916(75)90211-0.
68. A. de Rújula, H. Georgi, S. L. Glashow, Trinification of All Elementary Particle Forces, in: Fifth Workshop on Grand Unification, 1984.
69. W. Kilian, J. Reuter, Unification without doublet-triplet splitting, Phys. Lett. B 642 (2006) 81–84. arXiv:hep-ph/0606277, doi:10.1016/j.physletb.2006.09.033.
70. F. Braam, A. Knochel, J. Reuter, An Exceptional SSM from E6 Orbifold GUTs with intermediate LR symmetry, JHEP 06 (2010) 013. arXiv:1001.4074, doi:10.1007/JHEP06(2010)013.
71. F. Braam, J. Reuter, A Simplified Scheme for GUT-inspired Theories with Multiple Abelian Factors, Eur. Phys. J. C 72 (2012) 1885. arXiv:1107.2806, doi:10.1140/epjc/s10052-012-1885-5.
72. T. G. Rizzo, Gauge Kinetic Mixing in the E_6SSM, Phys. Rev. D 85 (2012) 055010. arXiv:1201.2898, doi:10.1103/PhysRevD.85.055010.
73. 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.
74. T. Appelquist, H.-C. Cheng, B. A. Dobrescu, Bounds on universal extra dimensions, Phys. Rev. D 64 (2001) 035002. arXiv:hep-ph/0012100, doi:10.1103/PhysRevD.64.035002.
75. T. Han, J. D. Lykken, R.-J. Zhang, On Kaluza-Klein states from large extra dimensions, Phys. Rev. D 59 (1999) 105006. arXiv:hep-ph/9811350, doi:10.1103/PhysRevD.59.105006.
76. A. Heister, et al., Measurement of the tau polarization at LEP, Eur. Phys. J. C 20 (2001) 401–430. arXiv:hep-ex/0104038, doi:10.1007/s100520100689.
77. P. Abreu, et al., A Precise measurement of the tau polarization at LEP-1, Eur. Phys. J. C 14 (2000) 585–611. doi:10.1007/s100520000363.
78. K. Mekala, D. Jeans, J. Tian, J. Reuter, A. F. Żarnecki, Precise measurement of light-quark electroweak couplings at future colliders, presented at the Epiphany'24 conference, https://indico.cern.ch/event/1288528/contributions/5718576/ (2024).
79. I. Zurbano Fernandez, et al., High-Luminosity Large Hadron Collider (HL-LHC): Technical design report 10/2020 (12 2020). doi:10.23731/CYRM-2020-0010.
80. The International Linear Collider Technical Design Report - Volume 2: Physics (6 2013). arXiv:1306.6352.
81. H. Abramowicz, et al., The International Linear Collider Technical Design Report - Volume 4: Detectors (6 2013). arXiv:1306.6329.
82. 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.
83. W. Abdallah, et al., CEPC Technical Design Report – Accelerator (v2) (12 2023). arXiv:2312.14363.
84. A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report (10 2012). doi:10.5170/CERN-2012-007.
85. Physics and Detectors at CLIC: CLIC Conceptual Design Report (2 2012). arXiv:1202.5940, doi:10.5170/CERN-2012-003.
86. A. Abada, et al., FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3, Eur. Phys. J. ST 228 (4) (2019) 755–1107. doi:10.1140/epjst/e2019-900087-0.
87. 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.
88. J. Tang, et al., Concept for a Future Super Proton-Proton Collider (7 2015). arXiv:1507.03224.
89. M. Bogomilov, et al., Demonstration of cooling by the Muon Ionization Cooling Experiment, Nature 578 (7793) (2020) 53–59. arXiv:1907.08562, doi:10.1038/s41586-020-1958-9.
90. J. Chen, T. Han, B. Tweedie, Electroweak Splitting Functions and High Energy Showering, JHEP 11 (2017) 093. arXiv:1611.00788, doi:10.1007/JHEP11(2017)093.
91. T. Han, Y. Ma, K. Xie, High energy leptonic collisions and electroweak parton distribution functions, Phys. Rev. D 103 (3) (2021) L031301. arXiv:2007.14300, doi:10.1103/PhysRevD.103.L031301.
92. F. Garosi, D. Marzocca, S. Trifinopoulos, LePDF: Standard Model PDFs for high-energy lepton colliders, JHEP 09 (2023) 107. arXiv:2303.16964, doi:10.1007/JHEP09(2023)107.
93. W. Kilian, T. Ohl, J. Reuter, Whizard—simulating multi-particle processes at lhc and ilc, The European Physical Journal C 71 (9) (Sep. 2011). doi:10.1140/epjc/s10052-011-1742-y.
URL http://dx.doi.org/10.1140/epjc/s10052-011-1742-y
94. M. Moretti, T. Ohl, J. Reuter, O'mega: An optimizing matrix element generator (2001). arXiv:hep-ph/0102195.

95. N. D. Christensen, C. Duhr, B. Fuks, J. Reuter, C. Speckner, Introducing an interface between WHIZARD and FeynRules, Eur. Phys. J. C 72 (2012) 1990. arXiv:1010.3251, doi:10.1140/epjc/s10052-012-1990-5.
96. C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, T. Reiter, UFO - The Universal FeynRules Output, Comput. Phys. Commun. 183 (2012) 1201–1214. arXiv:1108.2040, doi:10.1016/j.cpc.2012.01.022.
97. L. Darmé, et al., UFO 2.0: the 'Universal Feynman Output' format, Eur. Phys. J. C 83 (7) (2023) 631. arXiv:2304.09883, doi:10.1140/epjc/s10052-023-11780-9.
98. P. M. Bredt, W. Kilian, J. Reuter, P. Stienemeier, NLO electroweak corrections to multi-boson processes at a muon collider, JHEP 12 (2022) 138. arXiv:2208.09438, doi:10.1007/JHEP12(2022)138.

Prospects for measuring time variation of astrophysical neutrino sources at dark matter detectors

Yi Zhuang,[1] Louis E. Strigari,[1] Lei Jin,[2] and Samiran Sinha[3]

[1]*Department of Physics and Astronomy, Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA*

[2]*Department of Mathematics and Statistics, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA*

[3]*Department of Statistics, Texas A&M University, College Station, Texas 77843, USA*

(Dated: February 29, 2024)

We study the prospects for measuring the time variation of solar and atmospheric neutrino fluxes at future large-scale Xenon and Argon dark matter detectors. For solar neutrinos, a yearly time variation arises from the eccentricity of the Earth's orbit, and, for charged current interactions, from a smaller energy-dependent day-night variation to due flavor regeneration as neutrinos travel through the Earth. For a 100-ton Xenon detector running for 10 years with a Xenon-136 fraction of $\lesssim 0.1\%$, in the electron recoil channel a time-variation amplitude of about 0.8% is detectable with a power of 90% and the level of significance of 10%. This is sufficient to detect time variation due to eccentricity, which has amplitude of $\sim 3\%$. In the nuclear recoil channel, the detectable amplitude is about 10% under current detector resolution and efficiency conditions, and this generally reduces to about 1% for improved detector resolution and efficiency, the latter of which is sufficient to detect time variation due to eccentricity. Our analysis assumes both known and unknown periods. We provide scalings to determine the sensitivity to an arbitrary time-varying amplitude as a function of detector parameters. Identifying the time variation of the neutrino fluxes will be important for distinguishing neutrinos from dark matter signals and other detector-related backgrounds, and extracting properties of neutrinos that can be uniquely studied in dark matter experiments.

arXiv:2402.18454v1 [hep-ph] 28 Feb 2024

I. INTRODUCTION

Over the past several decades, direct dark matter detection experiments have made tremendous progress in constraining weak-scale particle dark matter [1, 2]. Future larger-scale detectors will be sensitive to not only particle dark matter, but also astrophysical neutrinos and various other rare-event phenomenology [3]. The most prominent of the neutrino signals are from the Sun, the atmosphere, and the diffuse supernova neutrino background (DSNB) [4]. Understanding these signals has important implications for the future of particle dark matter searches, but also for understanding the nature of the sources and the properties of neutrinos [5].

Various methods have been proposed to distinguish neutrinos and a possible dark matter signal. These include exploiting the energy distribution of nuclear recoils between neutrinos and dark matter [6, 7], the differences in arrival directions [8], and the differences in the periodicities of the signal [9, 10]. New physics in the neutrino sector may also change the nature of the predicted neutrino signal [11–13], and provide a method to discriminate from dark matter.

Here we examine the time variations of the neutrino signals in more detail and study the prospects for measuring these time variations. For solar neutrinos, the time variation of the flux is due to the eccentricity of the Earth's orbit and the day-night effect. The former is independent of the neutrino flavor, while the latter, which results from neutrino interactions with the matter as they pass through the Earth, is flavor dependent. Both effects are present in a dark matter experiment through the nuclear recoil and electron recoil channels. Beyond solar neutrinos, there may be detectable time variation of other components of the astrophysical neutrino flux. Due to the solar cycle, there is a time variation of the atmospheric neutrino flux which is $\sim 10-30\%$ [14] depending on the detector location. We present the first estimates of the detectability of the time variation of the atmospheric neutrino flux given realistic future detector configurations.

Identifying the time variation for the different neutrino sources is important for properly extracting the signal and distinguishing it from dark matter [4, 15]. In addition, it is important to characterize these signals to constrain neutrino properties. Previous studies of neutrinos in dark matter experiments have only considered the time variation of solar neutrinos as being due to the eccentricity of the Earth's orbit, and for idealistic values of the nuclear recoil threshold [9]. In addition to examining more realistic and updated detector configurations, we consider the prospects for measuring time variation of solar neutrinos using the neutrino-electron elastic scattering channel for the first time.

Solar neutrino experiments have previously searched for time variations in their signals, including experiments that have successfully established time variation due to the eccentricity of the Earth's orbit [16, 17] and the day-night effect [18]. These experiments each used a range of statistical technqiues to identify the time variable signal. As part of our analysis, we rigorously compare the different statistical methodologies for extracting the time-varying signal. We present results for the sensitivity of given experiments to time-varying amplitudes, and quantify the prospects for signal extraction as a function of experimental sensitivity and background levels.

This paper is organized as follows. In Section II, we briefly describe the signals and the models for the detector efficiency. In Section III we describe the periodic signals used in this work, and also we summarize the previous experiments that have searched for neutrino periodicity. Next, in Section IV, we review the statistical methodologies used in our analysis. In Section V, we describe the simulation strategy, compare statistical methods and introduce the signal-to-noise ratio as a convenient estimation tool. Then in Section VI, we present our resulting projections, and in Section VII, the discussion and conclusions.

II. EVENT RATES AT DIFFERENT DETECTORS

A. Theoretical calculation

Figures 1 and 2 show the electron and nuclear recoil spectra for the solar, atmospheric, and DSNB spectra for Xenon and Argon targets. The nuclear recoil spectrum uses the neutral current coherent elastic neutrino-nucleus scattering (CEνNS) channel, and the electron recoil channel uses charged current neutrino-electron elastic scattering (ES). We refer to previous literature for details of these calculations [19]. Here we simply highlight the components of the spectrum as a function of recoil energy to get a sense of the recoil threshold required to detect each component. For Solar neutrinos, we used the high metallicity model for the normalization [20]. We also show the appropriate experimental background components [21].

In addition to the energy dependence shown in Figure 1 and 2, for several of the neutrino components there is a time variation to the flux. For solar neutrinos, since the time variation due to the eccentricity of the Earth's orbit affects all flavors, it will be present in both the CEνNS and the ES channels. For charged current detection channels, the day-night effect due to oscillations is present in the ES channel. For atmospheric neutrinos, the time variation from the solar modulation of the atmospheric neutrino

flux [14] affects all flavors and is detectable through CEνNS. Of the spectra shown in Figure 2, the only essentially steady-state component is the DSNB, which has a cosmological origin [22]. In the section below we describe our parameterization of each of these time-varying components in more detail.

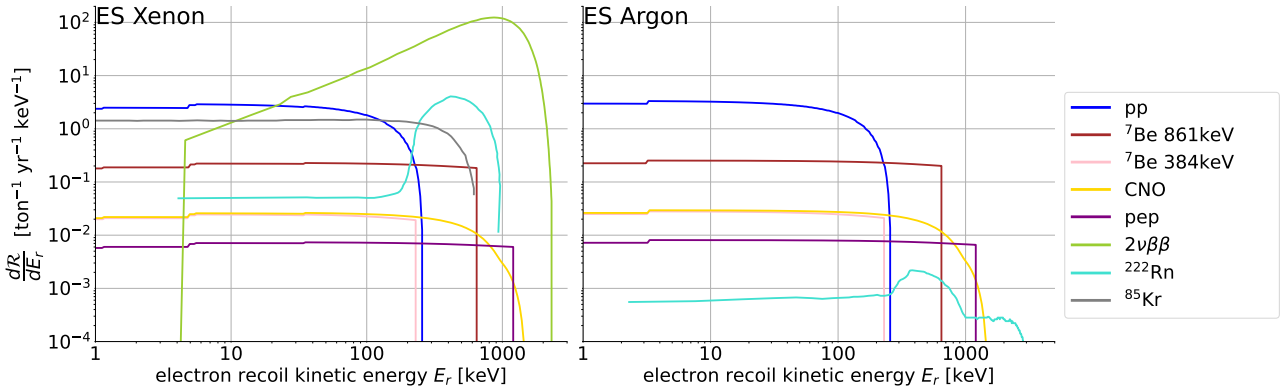


FIG. 1. Neutrino-electron elastic scattering (ES) spectra for Xenon (left) and Argon (right) for solar and experimental background components. For ES Argon, we consider ^{222}Rn as the background [23].

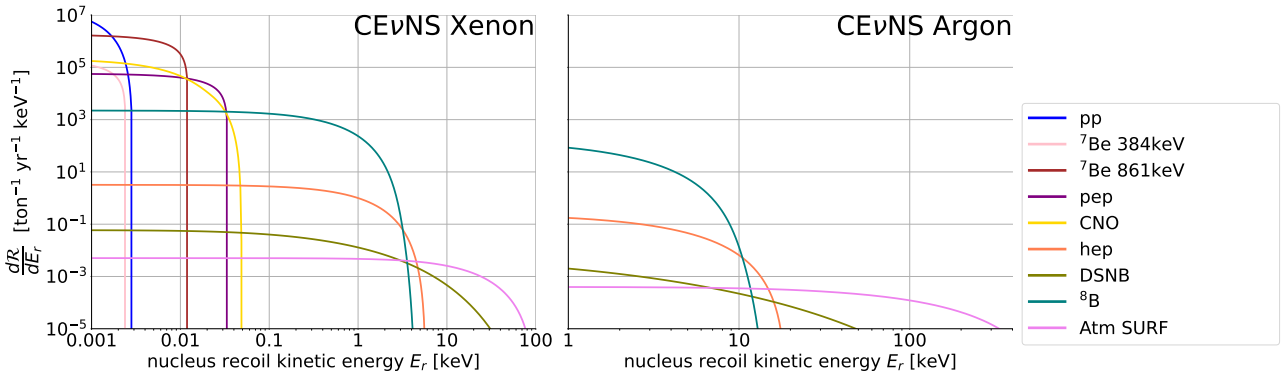


FIG. 2. Neutrino-nucleus scattering (CEνNS) spectra for Xenon (left) and Argon (right). Shown are the components of the solar, atmospheric, and DSNB spectra. The atmospheric spectra are shown for the SURF detector location.

B. Adding Resolution and Detector Efficiency

To estimate the detector efficiency and resolution in Xenon, we use the Noble Element Simulation Technique (NEST) [24] code. Neutrinos (or dark matter particles) interact with the gas in the detector, producing a scintillation signal, S1, and ionization electrons, which then drift along the electric field to produce a scintillation signal, S2. The NEST code simulates the detection of events in the space of S1 and S2. For the NEST configurations, we compare several possible options. Similar to previous studies, we consider the all enhanced parameters "Xe100t-5"[1] [25]. We also use the specific experimental files LZ[2] and G3[3]. For all cases, we do not select a specific region of interest (ROI), so the signal encompass all available space for electron and nuclear recoils.

With this setup, we estimate the detector efficiency as a function of the true nuclear or electron recoil energy. To obtain the efficiency, we simulate 10^7 random recoil energies uniformly over the range $[0-100]$ keV within the detector. After processing through NEST, each simulated energy corresponds to a specific S1 and S2. We bin the detected events in recoil energy space, count the number of events with both valid S1/S2 [26], and divide the valid counts by the total number of events in each energy bin.

[1] https://zenodo.org/records/3653516"Detector_Xe100t_5.hh"
[2] https://github.com/NESTCollaboration/nest/blob/master/include/Detectors/LZ_SR1.hh
[3] https://github.com/NESTCollaboration/nest/blob/master/include/Detectors/Detector_G3.hh

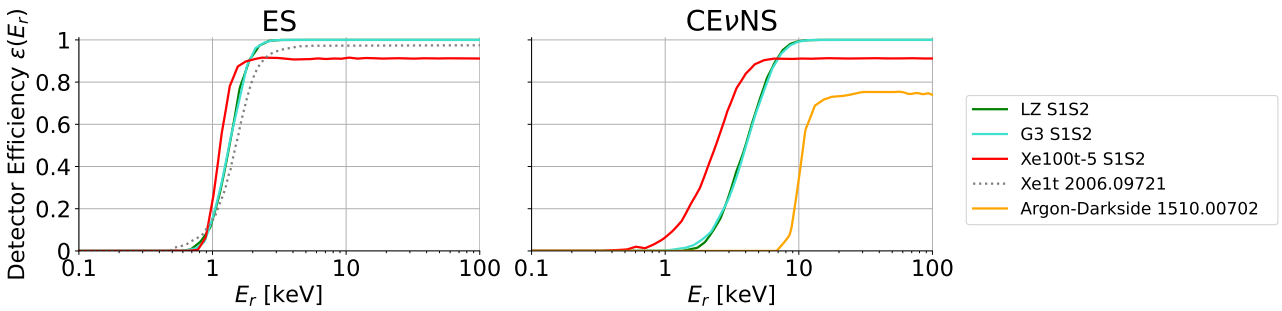


FIG. 3. Different detector efficiency models from NEST simulations ("LZ", "G3", "Xe100t-5") and current experiments ("Argon-DarkSide", "Xe1T"). *Left panel*: Solid curves are "LZ", "G3" and "Xe100t-5" for Xenon elastic scattering (ES). The dotted curve is the efficiency from Xenon1T [26]. *Right panel*: Solid curves are "LZ", "G3" and "Xe100t-5" for the Xenon nuclear recoil CEνNS channel. The orange solid curve is the efficiency from DarkSide-50 [27] for nuclear recoils in Argon.

The results for the ES and CEνES efficiencies are shown in Figure 3. For comparison, we show the ES efficiency used by Xenon1T [26], which is similar to our "Xe100t-5", LZ and G3 curves, with the differences between them likely explained by using different detector parameters, including the ROI and the recoil energies simulated. For Xenon, the electron recoil efficiency drops to zero rapidly below ~ 1 keV, and is nearly 100% above this energy.

To obtain the event rate modified by resolution and detector efficiencies, we integrate from a threshold energy, E_{thrd}, to an endpoint energy, E_{end}, and obtain $\mathcal{R} = \int_{E_{thrd}}^{E_{end}} (d\mathcal{R}/dE_r) dE_r$, where

$$\frac{d\mathcal{R}}{dE_r} = \varepsilon(E_r) \int dE_r^{'} \frac{d\mathcal{R}(E_r^{'})}{dE_r^{'}} \frac{1}{\sqrt{2\pi\sigma^2(E_r^{'})}} e^{-\frac{1}{2}\frac{(E_r - E_r^{'})^2}{\sigma^2(E_r^{'})}}$$

where $E_r^{'}$ is the true recoil energy, E_r is the detected recoil energy, $\sigma(E_r^{'})$ is the resolution at the true recoil energy, and $\varepsilon(E_r)$ is the detector efficiency at the detected recoil energy.

Different resolution and efficiency models for different nuclear targets that we use for our analysis are summarized in Tables I and II. For Xenon our resolution and efficiency models rely on NEST simulations as described above. For our Argon ES resolution and efficiency model, we choose $E_{thrd} = 100$ keV in order to determine how our results are affected by choosing a threshold characteristic of current experiments.

channel	target	resolution model $\sigma(E_r)$	efficiency model $\varepsilon(E_r)$	E_{thrd} [keV]	E_{end} [keV]	signal components [ton^{-1} yr^{-1}]	background components [ton^{-1} yr^{-1}]
ES	Xe	ideal	ideal	1	650	pp, ^{7}Be 861, ^{7}Be 384, pep, CNO	$f_{2\nu\beta\beta} \times 2\nu\beta\beta$, ^{85}Kr, ^{222}Rn
		$\left(0.31\sqrt{\frac{E_r}{\text{keV}}} + 0.0035\frac{E_r}{\text{keV}}\right)$ keV [26]	Xe100t-5	0	790		
	Ar	ideal	ideal	1	3400		^{222}Rn
		$0.1E_r$	/	100	3400		

TABLE I. Target, resolution model, detector efficiency model, recoil energy threshold, recoil energy endpoint, components experiencing time variation, and components that have constant rate constant in time in the ES channel. $f_{2\nu\beta\beta}$ is the fraction of the remaining $2\nu\beta\beta$ background after depletion of ^{136}Xe.

channel	target	resolution model $\sigma(E_r)$	efficiency model $\varepsilon(E_r)$	E_{thrd} [keV]	E_{end} [keV]	signal components [ton^{-1} yr^{-1}]	background components [ton^{-1} yr^{-1}]
CEνNS	Xe	ideal	ideal	1	4	^{8}B	/
		$\left(0.31\sqrt{\frac{E_r}{\text{keV}}} + 0.0035\frac{E_r}{\text{keV}}\right)$ keV [26]	Xe100t-5 LZ G3	0	9		
	Ar	ideal	ideal	1	13		
		$0.1E_r$	Darkside	0	29		

TABLE II. Same as Table I except for the CEνNS channel.

Experiment	Signals of interest	$\mathcal{T}$	Δ t	scanned frequency range $[f_{min}, f_{max}]$	results
Homestake [28]	quasi-biennial periodicity $\frac{1}{2.2}$ yr^{-1}, Rieger periodicity $\frac{1}{157}$ day^{-1}			[0, 20] yr^{-1}	no evidence
	solar rotation frequency			[12.4, 13.1] yr^{-1}	peak at 12.88 yr^{-1} significant at the 3% level
Sudbury	blind search	572.2 calendar days (D_2O) 762.7 calendar days (salt) [29, 30]	1 day	[$\frac{1}{10}$, 365.25] yr^{-1}	peak at 0.296 day^{-1} (D_2O), 0.971 day^{-1} (salt), 0.417 day^{-1} (combined)
				[$\frac{1}{10}$, 182.625] yr^{-1}	peak at 0.408 day^{-1} (D_2O), 0.429 day^{-1} (salt), 0.413 day^{-1} (combined)
	7% amplitude 9.43 yr^{-1} Kamioka result			[9.33, 9.53] yr^{-1}	best-fit amplitude 1.3% (combined)
	eccentricity 1 yr^{-1}				best-fit eccentricity ε=0.0143
Borexino	diurnal modulation 1 day^{-1}	740.88 live days [31] Phase-I			day-night asymmetry = 0.001
	eccentricity 1 yr^{-1}	1456 astro days [32] Phase-II	30.43 days		best-fit eccentricity ε = 0.0174
			1 day 7 days		peak at 1 yr^{-1}
	eccentricity 1 yr^{-1}	$\sim$ 10 yrs [16] Phase-II+III	30 days		best-fit eccentricity ε = 0.0184
	Carrington rotation 13.4 yr^{-1}, diurnal modulation 1 day^{-1}		8 hrs	[1, 547] yr^{-1}	day-night asymmetry = 0.003
Kamioka	blind search	1871 elapsed days SK-1 [17, 33–36]	10 days	[0.0002, 0.0987] day^{-1} [33]	peak at 0.0726 day^{-1}
				[0, 50] yr^{-1} [17]	peak at 9.42 yr^{-1}
				[0, 100] yr^{-1} [34] [0, 40] yr^{-1} [34]	peak at 26.57 yr^{-1} alias peak at 9.42 yr^{-1}
			5 days	[0.0002, 0.19187] day^{-1} [33]	peak at 0.1197 day^{-1} 0.0726 day^{-1} removed
				[0, 50] yr^{-1} [17, 35]	alias peak of 9.42 yr^{-1} at 26.52yr^{-1} possible solar rotation
	eccentricity 1 yr^{-1} [36]		1.5 months		best-fit eccentricity ε=0.021
	Day-Night difference [36]				day-night asymmetry = -0.021, -0.017
	blind search	5803 live days [37] SK-I,II, III, IV	5 days	[4e^{-4}, 0.2] day^{-1}	peak at 0.126 day^{-1}
				[10^{-6}, 0.2] day^{-1}	peak at 0.143 day^{-1}
	eccentricity 1 yr^{-1}		$\frac{365.25}{12}$ days		best-fit eccentricity ε=0.0153
IceCube [38]	temperature dependent variation of atmospheric neutrino flux	May 2012 - May 2017	1 day 30 days		high statistics at 140 day^{-1} in the South region

TABLE III. Summary of previous experimental results searching for time variation of solar neutrinos. The columns are 1) Experiment, 2) signal of interest experiment searched for, 3) experimental exposure, 4) time binning used in the analysis, 5) scanned frequency range, 6) results of the analysis, along with specific comments where appropriate. Boxes are left blank in the experimental exposure and time binning columns for Homestake because that information was not provided. Boxes are left blank in the scanned frequency range column because in these cases the signal is fit for given the known eccentricity, and a scan of frequency is not required.

III. PHYSICAL PERIODIC SIGNALS OF INTEREST

A. Solar yearly modulation

The slight eccentricity, ε, to the Earth's orbit around the Sun induces a yearly variation in the neutrino flux. The time dependence of the flux is parameterized as [39]

$$\phi(t) = \frac{R_\odot}{4\pi r^2(t)} \approx \frac{R_\odot}{4\pi r_0^2}\left[1 + 2\varepsilon\cos\left(\frac{2\pi t}{P}\right)\right], \tag{1}$$

where P denotes the periodicity measured in years, $r(t)$ is the distance between the Earth and Sun at time t measured in years (i.e., $0 \leq t \leq 1$), $r_0 = 1$ AU, and $R_\odot$ is the neutrino production rate in the Sun per unit time. The second term, $2\varepsilon\cos(2\pi t/P)$, in Equation 1 is responsible for the small time variation. The amplitude of the flux variation is approximately 0.03342, and has been measured by several experiments, including recently by Borexino [16]. For standard assumptions for the dark matter velocity distribution, this time variation is out of phase with the standard modulation signal predicted from dark matter and has been proposed as a means to separate neutrinos from dark matter signals [9].

B. Solar daily modulation

A second time variation in the solar neutrino flux is the day night effect, which is due to the regeneration of the electron neutrino flux from matter effects as electron neutrinos pass through the Earth. The day-night asymmetry is detectable for experiments sensitive to the electron neutrino flux through charged current interactions. Defining N_D and N_N as the number of day and night events, the day-night asymmetry, A_{DN}, is parametrized as [16]

$$A_{DN} = \frac{N_D - N_N}{0.5(N_D + N_N)} = \frac{2A_{d,DN}}{\sqrt{2}R_\odot},$$

where $A_{d,DN}$ is the amplitude of the daily modulation. Super-Kamiokande has now established the day-night effect for high energy ^{8}B neutrinos [18] with a measured amplitude of $\lesssim 3\%$ that is consistent with the LMA-MSW solution.

At lower energies where solar neutrinos transition are due to vacuum oscillations, the LMA-MSW solution predicts that the day-night modulation is much smaller, $\lesssim 0.1\%$. The best limits on the day-night modulation at these energies come from Borexino measurements of the ^{7}Be component [16]. Borexino is sensitive to electron recoils $\gtrsim 150$ keV, so is not sensitive to the modulation for the lower energy pp component. Since Xenon and Argon-based dark matter experiments are sensitive to lower threshold , $\sim$ few keV electron recoils, they will extend limits on solar neutrino time variation and day-night effects to lower energies than have been previously studied. In our analysis, we simply compare the limits on daily modulation from dark matter detectors to the limits on the modulation obtained from Borexino.

C. Atmospheric neutrinos

We also study the prospects for identifying the time variation of the atmospheric neutrino flux. Atmospheric neutrinos are detected through the CEνNS channel, with a modulation amplitude that depends on the detector location. We model the atmospheric time variation with a sinusoidal function in a similar manner to Equation 1. In this case, the period is given by the 11-year period of the Solar cycle, and the amplitudes are given by the maximum difference between the Solar minimum and maximum fluxes calculated in Ref. [14]. The flux differs for different detector locations, which we consider in our analysis below. For example for the detector locations China Jinping Underground Laboratory (CJPL), Kamioka, Laboratori Nazionali del Gran Sasso (LNGS), the Sanford Underground Research Facility (SURF), and SNOlab, the maximum amplitudes for time variation are $A_{atm} = [0.0459, 0.0382, 0.0461, 0.1475, 0.1327]$, respectively, where A_{atm} and ω are defined through

$$\begin{aligned}
\phi_{atm}(t) &= \bar{\phi} A_{atm}\sin(\omega t), \\
\bar{\phi} &= \frac{\phi_{09} + \phi_{14}}{2}, \\
A_{atm} &= \frac{\phi_{09} - \phi_{14}}{2\phi_{09}}, \\
\omega &= \frac{2\pi}{11\text{yr}}.
\end{aligned}$$

The quantities ϕ_{09} and ϕ_{14} are the fluxes in the year of 2009 and 2014, corresponding to Solar min and Solar max, respectively.

D. Previous Experimental Methods and Results

Over the past several decades, multiple experiments have searched for possible periodic signals in neutrino data. To this point, the eccentricity has been measured by Borexino, and the day-night modulation due to charged current interactions has been established by SK. An upper limit of the diurnal modulation has been set by Borexino.

A typical data analysis procedure is to bin the observed data in time bins. Then, a Fourier transform is performed, a range of frequencies is scanned, and then peaks that correspond to possible periodic signals are found. The peaks are checked to determine if the power of the peak exceeds a certain threshold (false-alarm probability). The specific details of turning time-binned data into the periodogram, the error of the observed data, and the terminology describing the methods differ among the experimental analyses.

In Table III, we summarize the results from experiments at Homestake, Kamioka, Sudbury, Borexino, and IceCube. We refer to each paper for more details on their analysis methods. We characterize the experiments in terms of their run time $\mathcal{T}$, assumed time binning Δt, frequency range scanned, the periodic signal of interests, and the results. These summarized results provide a point of comparison for our future projections for dark matter detectors.

IV. STATISTICAL METHODS FOR TIME-VARYING ANALYSIS

This section establishes our statistical methods for detecting solar and atmospheric neutrinos, specifically describing our methods for constructing the likelihood function. In this case, the null hypothesis H_0 is defined as a signal with a constant rate in time, while the alternative, H_1, is defined as a signal whose rate of events varies in time with a period, P, and an amplitude, A_d.

We use the following models under H_0 and H_1 to formulate our likelihood ratio test. Let n_i be the number of observed events in the time bin $(t_i, t_i + \Delta t)$. We assume that

$$n_i \equiv n(t_i) \sim \begin{cases} \text{Pois}(\lambda \Delta t) & \text{under } H_0, \\ \text{Pois}(\lambda_i) & \text{under } H_1, \end{cases}$$

where the expected number of events in the time bin $(t_i, t_i + \Delta t)$ under the assumption of H_1 is

$$\lambda_i \equiv \lambda(t_i) = \lambda \int_{t_i}^{t_i+\Delta t} \{1 + A\cos(\omega u) + B\sin(\omega u)\} du.$$

In the above formulation, λ, A, and B are unknown parameters, and ω is an unknown frequency. In λ_i, the background signal is $\lambda \Delta t$ and the time-varying signal is $\lambda \int_{t_i}^{t_i+\Delta t} \{A\cos(\omega u) + B\sin(\omega u)\} du$. Note that H_0 is a special case of H_1 as we obtain the null model by setting $A = B = 0$ in λ_i.

First, we assume that ω is known for a gentle exposition of the methodology. To test H_0 against H_1, we use the generalized likelihood ratio (GLR) method. The term generalized is used when at least one of H_0 and H_1 is a composite hypothesis; in our case, both are composite hypotheses. [30] used the GLR method for identifying the periodicity of ^{8}B solar neutrino flux released by the SNO Collaboration. However, unlike [30], we obtain the expected number of events within a time bin by integrating the time-varying instantaneous rate function, which is an appropriate method, especially when the bin width is not very small. Moreover, besides applying to ^{8}B solar neutrinos, we use this model for all components of the solar and atmospheric neutrino flux. The GLR statistic is

$$\text{GLR} = \frac{\max_\lambda \prod_{i=1}^{N_t} \frac{\lambda^{n_i} e^{-\lambda}}{n_i!}}{\max_{\lambda,A,B} \prod_{i=1}^{N_t} \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!}} = \frac{\max_\lambda \prod_{i=1}^{N_t} \lambda^{n_i} \exp(-\lambda)}{\max_{\lambda,A,B} \prod_{i=1}^{N_t} \lambda_i^{n_i} \exp(-\lambda_i)}. \tag{2}$$

The numerator of (2) is

$$(\widehat{\lambda}_0 \Delta t)^{\sum_{i=1}^{N_t} n_i} \exp(-\widehat{\lambda}_0 \Delta t N_t),$$

where $\widehat{\lambda}_0 = \sum_{i=1}^{N_t} n_i / N_t \Delta t$ denotes the maximum likelihood estimator of λ under H_0. The denominator of (2) is

$$\prod_{i=1}^{N_t} \widehat{\lambda}_i^{n_i} \exp(-\widehat{\lambda}_i),$$

where $\widehat{\lambda}_i = \widehat{\lambda}_1 \int_{t_i}^{t_i+\Delta t}\{1+\widehat{A}\cos(\omega u)+\widehat{B}\sin(\omega u)\}du$ with $\widehat{\lambda}_1$, $\widehat{A}$ and $\widehat{B}$ are the solution of the following gradient equations (obtained by differentiating the log-likelihood function under H_1)

$$0 = \sum_{i=1}^{N_t}\left[\frac{n_i}{\lambda} - \int_{t_i}^{t_i+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du\right], \tag{3}$$

$$0 = \sum_{i=1}^{N_t}\left\{\frac{n_i\int_{t_i}^{t_i+\Delta t}\cos(\omega u)du}{\int_{t_i}^{t_i+\Delta t}\{1+A\cos(\omega t_i)+B\sin(\omega u)\}du} - \lambda\int_{t_i}^{t_i+\Delta t}\cos(\omega u)du\right\}, \tag{4}$$

$$0 = \sum_{i=1}^{N_t}\left\{\frac{n_i\int_{t_i}^{t_i+\Delta t}\sin(\omega u)du}{\int_{t_i}^{t_i+\Delta t}\{1+A\cos(\omega t_i)+B\sin(\omega u)\}du} - \lambda\int_{t_i}^{t_i+\Delta t}\sin(\omega u)du\right\}. \tag{5}$$

Equation (3) yields

$$\lambda = \frac{\sum_{j=1}^{N_t} n_j}{\sum_{j=1}^{N_t}\int_{t_j}^{t_j+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du},$$

and using this expression in (4) and (5) we obtain

$$\begin{aligned}
0 &= \sum_{i=1}^{N_t}\left[\frac{n_i}{\int_{t_i}^{t_i+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du} - \frac{\sum_{j=1}^{N_t} n_j}{\sum_{j=1}^{N_t}\int_{t_j}^{t_j+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du}\right]\int_{t_i}^{t_i+\Delta t}\cos(\omega v)dv,\\
0 &= \sum_{i=1}^{N_t}\left[\frac{n_i}{\int_{t_i}^{t_i+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du} - \frac{\sum_{j=1}^{N_t} n_j}{\sum_{j=1}^{N_t}\int_{t_j}^{t_j+\Delta t}\{1+A\cos(\omega u)+B\sin(\omega u)\}du}\right]\int_{t_i}^{t_i+\Delta t}\sin(\omega v)dv.
\end{aligned} \tag{6}$$

To solve (6), we assume that data n_i's and the frequency ω are known, and replace $\int_{t_i}^{t_i+\Delta t}\sin(\omega v)dv$ by $[\cos(\omega t_i)-\cos\{\omega(t_i+\Delta t)\}]/\omega$ and $\int_{t_i}^{t_i+\Delta t}\cos(\omega v)dv$ by $[\sin\{\omega(t_i+\Delta t)\}-\sin(\omega t_i)]/\omega$. The likelihood spectrum at ω is

$$S(\omega) = -\ln(\mathrm{GLR}) = \left[\sum_{i=1}^{N_t}\left\{n_i\ln(\widehat{\lambda}_i)-\widehat{\lambda}_i\right\} - \ln(\widehat{\lambda}_0\Delta t)\sum_{i=1}^{N_t} n_i + N_t\widehat{\lambda}_0\Delta t\right], \tag{7}$$

where the expression of $\widehat{\lambda}_i$ is given before equation (3). For a given value of ω, this is the key statistic used for the GLR test.

We can assume that the frequency ω is known or unknown when analyzing data. If ω is known and N_t is large, then to test the hypotheses (for detecting the signal), one can reject H_0 when $2S(\omega) > \chi^2_{2,\alpha}$, where $\chi^2_{2,\alpha}$ denotes the upper α percentile of the Chi-square distribution with 2 degrees of freedom. For an unknown ω, we define the test statistic as

$$S_{\max} = \max_{\omega\in\Omega} S(\omega),$$

where Ω is set of set of M distinct frequencies. In this case, we reject H_0 when $S_{\max} > S_{\max,\alpha}$, where $S_{\max,\alpha}$ denotes the upper α percentile of the $S_{\max}$ distribution. Since the analytical form of the distribution is unknown, the percentile $S_{\max,\alpha}$ is determined by a numerical procedure given in the next section.

Alternative to the GLR method, we also consider the Lomb-Scargle method [40], hereafter referred to as LS. The periodogram at a given frequency ω is

$$LS(\omega) = \frac{1}{2\sigma^2}\left(\frac{\left[\sum_{i=1}^{N_t} n_i\cos\{\omega(t_i-\tau)\}\right]^2}{\sum_{i=1}^{N_t}\cos^2\{\omega(t_i-\tau)\}} + \frac{\left[\sum_{i=1}^{N_t} n_i\sin\{\omega(t_i-\tau)\}\right]^2}{\sum_{i=1}^{N_t}\sin^2\{\omega(t_i-\tau)\}}\right)$$

with $\tau = (1/2\omega)\tan^{-1}\left\{\sum_{i=1}^{N_t}\sin(2\omega t_i)/\sum_{i=1}^{N_t}\cos(2\omega t_i)\right\}$ and $\sigma^2 = \sum_{i=1}^{N_t} n_i^2/(N_t-1)$. Like the GLR case, for known ω, we reject H_0 if $2L(\omega) > \chi^2_{2,\alpha}$ and for the unknown ω we reject H_0 if $L_{\max} = \max_{\omega\in\Omega} L(\omega) > L_{\max,\alpha}$. This $L_{\max,\alpha}$ is the upper α

percentile point of $L_{\max}$. This percentile (or the critical value of this test) can be calculated analytically or numerically. Ref. [30] approximated the distribution of $L_{\max}$ by the following probability density function

$$f_X(x|\nu) = \nu(1-e^{-x})^{\nu-1}e^{-x}, x>0, \tag{8}$$

where ν is a function of the number of scanned frequencies in Ω. Following [33], we take $\nu = \mathcal{M}_{scan}$, and the details on how to calculate $\mathcal{M}_{scan}$ is given in the next section. To be specific, Equation (8) is the density of the maximum of ν independent standard exponential (χ^2_2 /2) random variables. Therefore, we can determine the upper α percentile of $LS_{\max}$ from the density given in Equation (8). The upper α percentile, denoted by $LS_{\max,\alpha,1}$ is $-\log\{1-(1-\alpha)^{1/\nu}\}$. We use $LS_{\max,\alpha,2}$ to denote the percentile determined by a simulation technique.

V. SYNTHETIC DATA SIMULATION STRATEGY

A. General simulation strategy

We generated data under H_0 and H_1. The counts n_i over the time bin $[t_i, t_i+\Delta t)$ follow

$$\begin{array}{ll} \text{Pois}\{(\mathcal{R}_s+\mathcal{R}_b)\mathcal{D}\Delta t\} & \text{under } H_0, \\ \text{Pois}\{\mathcal{R}_b\mathcal{D}\Delta t + \mathcal{R}_s\mathcal{D}(\Delta t + A_d\int_{t_i}^{t_i+\Delta t}\sin\left(\frac{2\pi}{P}u-\Psi\right)du)\} & \text{under } H_1. \end{array}$$

Here $\mathcal{R}_s$ is a constant signal event rate and $\mathcal{R}_b$ is a constant rate for the background noise. Here, we have chosen not to bin in recoil energy, so $\mathcal{R} = \mathcal{R}_s + \mathcal{R}_b$ denotes the event rate at any given time integrated over the entire recoil energy range and summing over all flux components. For analyzing (testing of hypotheses) simulated data, we employed the GLR and LS methods.

Note that we use a value of ω for data simulation. But at the analyses stage, ω is assumed to be once known and then an unknown parameter. For unknown ω, the statistic $S_{\max}$ and the critical value of $S_{\max}$ are determined as follows. We take Ω as a set of $N_{\text{entire}} = n_o\mathcal{T}/\Delta t$ evenly-spaced frequencies on the entire range for the scanned frequencies $[f_{min}, 1/\Delta t]$ for a given run time $\mathcal{T}$, width of the time interval Δt, and over-sample factor n_o [41], so $\Omega = \{\omega_1, \ldots, \omega_{N_{entire}}\}$. We set $n_o = 10$ for the Monte Carlo simulation. We use the same Ω for both $S_{\max}$ and $L_{\max}$. Another concept needed for deriving the null distribution of $L_{\max}$ is the effective number of scanned frequencies over $[f_{min}, 1/\Delta t]$. It is $\mathcal{M}_{entire} = N_{entire}/f_{adj}$, where f_{adj} is an empirical result for the scanning density. Note that we are interested in the frequency range $[f_{min}, f_{max}]$ with $f_{max} = \min(1/\Delta t, 1\,\text{day}^{-1})$, and $f_{min} = 1/365.25\,\text{day}^{-1}$ for P = 1 yr or 1 day, $f_{min} = 1/11\,\text{yrs}^{-1}$ for P = 11 yr. Then we over scan $N_{entire} \times f_{scan}/f_{entire}$ on $[f_{min}, f_{max}]$ where $f_{entire} = 1/\Delta t - f_{min}$ and $f_{scan} = f_{max} - f_{min}$. Specifically, to be consistent with the frequency range of interest, f_{adj} for different Δt and $\mathcal{T}$ are showed in Appendix VIII. The number of independent frequencies scanned in a smaller frequency range [33] is $\mathcal{M}_{scan} = \mathcal{M}_{entire} \times f_{scan}/f_{entire}$, and this $\mathcal{M}_{scan}$ is used to approximate the distribution of $L_{\max}$ given in (8).

1. *Critical value $S_{\max,\alpha}$ from the distribution of $S_{\max}$*

We found that the empirical density of $S_{\max}$ is somewhat close to (8) for some choice of ν, which need not be the same as the number of frequencies of Ω. However, it isn't easy to prove that the density of $S_{\max}$ has the form (8) because1) for any two arbitrary frequencies ω and ω', both $S(\omega)$ and $S(\omega')$ are calculated on the same observed data, just at two different frequencies, so they need not be independent, and 2) finding the distribution of the maximum of dependent χ^2 random variables is not an easy task. Therefore, the threshold $S_{\max,\alpha}$ estimated after fitting (8) to the $S_{\max}$ data, will be an approximation of the upper α percentile of $S_{\max}$'s distribution. For each 5000 simulated dataset under H_0, we computed $S_{\max}$, which helps to obtain the empirical distribution of $S_{\max}$. We then fit model (8) to the $S_{\max}$ values denoted by $\{S_{\max,r}, r = 1, \ldots, 5000\}$ and obtain the maximum likelihood estimates (MLE) of ν by maximizing the following log-likelihood function:

$$\sum_{r=1}^{5000}[\log(\nu) + (\nu-1)\log\{1-\exp(-S_{\max,r})\} - S_{\max,r}].$$

Let us denote the MLE of ν by $\widehat{\nu}$. Since the cumulative distribution function of the probability density (8) is $F(h) = \{1-\exp(-h)\}^{\nu}$, we set $\{1-\exp(-S_{\max,\alpha})\}^{\widehat{\nu}} = 1-\alpha$, and solving this equation we obtain the estimate $S_{\max,\alpha,1} = -\log\{1-(1-\alpha)^{1/\widehat{\nu}}\}$. Next, we use the upper αth percentile of the 5000 $S_{\max}$ values as the second estimate of $S_{\max,\alpha}$, call it $S_{\max,\alpha,2}$.

$\mathcal{T}$	Δt	$\mathcal{D}$	average number	LS			GLR		
years	days	tons	of events per bin	$\mathcal{M}_{scan}$	$LS_{\max,\alpha,1}$	$LS_{\max,\alpha,2}$	$\widehat{\nu}$	$S_{\max,\alpha,1}$	$S_{\max,\alpha,2}$
10	30	100	821	202	7.56	7.24	133	7.14	7.39
10	10	100	273	608	8.66	8.4	463	8.39	8.49
10	5	100	136	1217	9.35	9.23	946	9.1	9.26
1/6	0.2	100	5	404	8.25	8.05	329	8.05	8.14
1/6	0.2	500	27	404	8.25	8.03	328	8.04	8.16
1/6	0.2	2000	109	404	8.25	8.12	325	8.03	8.22

TABLE IV. This table shows the threshold values of the test statistic derived from 1) the Lomb-Scargle periodogram and 2) the GLR, for different Δt, $\mathcal{T}$, $\mathcal{M}_{scan}$ and when the detector's size is assumed to be $\mathcal{D} = 100$ tons, $\mathcal{R}_s = 100\,\text{ton}^{-1}\,\text{yr}^{-1}$ and $\mathcal{R}_b = 0\,\text{ton}^{-1}\,\text{yr}^{-1}$, and $\alpha = 10\%$. The data are generated under H_0. Under each method two estimates of threshold are presented, a) model based corresponds to the one with suffix 1 and b) empirical distribution based, corresponds to the one with suffix 2. All results are based on simulation with 5000 replications.

2. *Results between GLR and LS methods under H_0*

In our simulation, we set $\alpha = 0.1$. The scanned frequencies were taken between $[f_{min}, f_{max}]$ with a bin-width of Δt, run-time $\mathcal{T}$, and the detector size $\mathcal{D} = 100$ ton. Computation of the threshold under the Lomb-Scargle method was much faster as it did not involve any equation solving, even when the number of frequencies was in the order of 10^5. In contrast, the GLR-based approach was time-consuming. Specifically, the computing time of GLR was proportional to the number of scanned frequencies and the length of n_i. We used scipy.root [42] and astropy.timeseries.LombScargle [43] of Python for GLR and LS calculations, respectively. Table IV contains the critical values for the GLR and LS methods when $\mathcal{T}$ = 10 yrs.

3. *Simulation procedure under H_1*

The power of the test is the probability of rejecting H_0 when data are generated under the alternative hypothesis H_1. The simulated data under H_1 is a Poisson distribution with a time-varying rate determined by a given amplitude A_d and period P. We choose $\Psi = P/4$ such that the simulated time variation starts at the extreme value.

We are mainly interested in the periodic signals in the neutrino flux from the Sun with a period of $P = 1$ day (due to Earth rotating itself), $P = 365.25$ days (due to Earth orbiting around the Sun), and atmosphere $P = 11$ yrs (due to the 11 yr solar cycle, the changing strength of the solar wind). All three periods should in principle be observable. In the data generation process, we consider $P = 1$ day and 365.25 days with $\Delta t < 1$ day; $P = 365.25$ days with 1 day $\leq \Delta t < 365.25$ days; $P = 11$ yrs when $\Delta t < 1$ yr and $\mathcal{T} \geq 11$ yrs.

When $n_i \overset{ind}{\sim} \text{Poisson}(\lambda_i), i = 1, \ldots, N_t$, the power of the tests is calculated under two scenarios

1. Known period P: Under such cases when the periods are known, instead of searching for all possible physical and theoretical peaks by finding a peak as is being done in actual experiments, we calculate the statistics $S(\omega_{known})$ and $LS(\omega_{known})$ at that known $\omega_{known} = 2\pi/P$, instead of estimating ω. The powers of the tests are

$$\text{Power} = \begin{cases} \text{pr}\{2S(\omega_{known}) > \chi^2_{2,\alpha}\}, & \text{for GLR,} \\ \text{pr}\{2LS(\omega_{known}) > \chi^2_{2,\alpha}\}, & \text{for LS.} \end{cases}$$

2. Unknown period P: Experimentally, possible periodic signals are searched by finding peaks over a wide frequency range. The scanned frequency range and frequency number are mentioned in the third paragraph of this section. In this case, S_{max} may be located at the alias of the input frequency instead of at the input frequency. The powers of the tests are

$$\text{Power} = \begin{cases} \text{pr}(S_{\max} > S_{\max,\alpha}), & \text{for GLR,} \\ \text{pr}(LS_{\max} > LS_{\max,\alpha}), & \text{for LS.} \end{cases}$$

We define $\beta = 1 - \text{Power}$ as the probability of failing to reject the null hypothesis when the alternative is true (i.e., the type II error). For the empirical power (power obtained via simulation), we first fix $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t)$. Then change A_d over a grid of values. For every choice of $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t, A_d)$, we generate Poisson data $\{n_1, \ldots, n_{N_t}\}$ under H_1. We calculate $S(\omega_{known})$ or $LS(\omega_{known})$ or S_{max} or LS_{max} for this simulated data. We repeat this procedure 5000 times, then calculate the

proportion of times these quantities is larger than $\chi^2_{2,\alpha}$ or $S_{\max,\alpha}$ or $LS_{\max,\alpha}$ accordingly. This proportion is an estimate of the power (Eqn 9). This procedure is repeated for every combination of the parameters. Next we draw the power curves against A_d for a fixed value of $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t)$ and we denote the amplitude A_d as $A_{d,\alpha,\beta}$ when the probability of type-I error rate of a hypothesis test being α and type-II error being β.

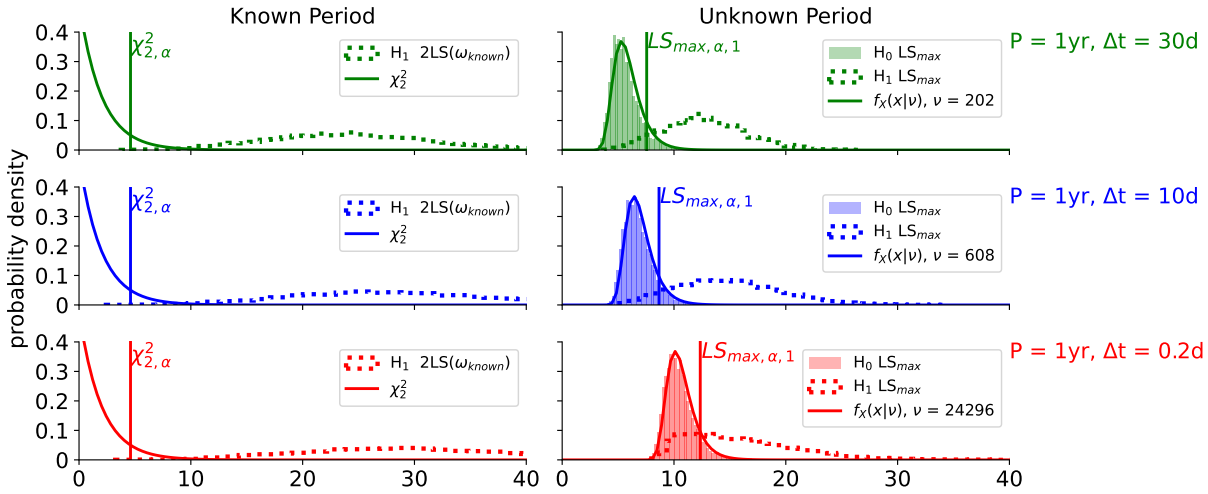


FIG. 4. Example of histograms for the assumptions of Known Periods (left) and Unknown Periods (right) under the null hypothesis H_0 and the alternative hypothesis H_1. Assumed parameters are runtime $\mathcal{T}$ = 10 yrs, time binnings $\Delta t = 0.2, 10, 30$ days, period of $P = 365.25$ days, amplitude $A_d = 0.024$, $\mathcal{R}_s = 100$ ton^{-1} yr^{-1}, and no experimental backgrounds $\mathcal{R}_b = 0$ton^{-1} yr^{-1}. *Left column*: The solid curve is the χ^2_2 distribution, the vertical line is the critical value $\chi^2_{2,\alpha}$ with α = 0.1. The dotted histogram in each panel is the distribution of 5000 simulated values of $2LS_{max}$ under the alternative hypothesis. *Right column*: The filled histogram in each panel is the empirical distribution of 5000 simulated values of $LS_{\max}$ under the null hypothesis H_0 using the LS method as described in the text. The solid curve is $f_X(x|\nu)$ with $\nu = \mathcal{M}_{scan}$, the vertical line is the critical value $LS_{\max,\alpha,1}$ with $\alpha = 0.1$. The dotted histogram in each panel is the distribution of 5000 simulated values of $LS(\omega_{known})$ under the alternative hypothesis with $\omega_{known} = 2\pi/P$ at the detection period $P = 365.25$ days.

Figure 4 shows an example when the detecting period $P = 365.25$ days using the run time $\mathcal{T} = 10$ yrs, and binned in $\Delta t = 0.2, 10, 30$ days. The left histograms are under the "Known Period" condition, where H_0 is the χ^2_2 distribution and H_1 is $2LS(\omega_{known})$. The right histograms are under the "Unknown Period" condition , where H_0 and H_1 are LS_{max}. The amplitude for H_1 is $A_d = 0.024$. The vertical line in each panel is the critical value $\chi^2_{2,\alpha}$ (left column) and $LS_{\max,\alpha,1}$ (right column) when $\alpha = 0.1$. The power $(1-\beta)$ is the proportion of $2LS(\omega_{known})$ (or $LS_{\max}$) greater than the threshold $\chi^2_{2,\alpha}$ (or $LS_{\max,\alpha,1}$) out of the 5000 realizations under H_1, where $\omega_{known} = 2\pi/P$.

4. *Comparing $A_{d,\alpha,\beta}$ between GLR and LS methods*

For a fixed value of $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t)$, we want to compare $A_{d,\alpha,\beta}$ between LS and GLR methods, under "Known Period" and "Unknown Period" scenarios. We also test the performance for a small time bin and a small number of events [16]. For these purposes, we set $\mathcal{R}_s = 100$ ton^{-1} yr^{-1} and $\mathcal{R}_b = 0$ ton^{-1} yr^{-1}. We consider $\mathcal{T} = 1/6$ yrs with $\Delta t = 0.2$ days, and $\mathcal{T} = 10$ yrs with $\Delta t = 10, 30$ days, and control the number of events in each time bin by changing $\mathcal{D}$. One example is shown in Figure 5, and results are shown in Table V. Under the "Known Period" scenarios, the results from GLR and LS are similar. However, under the unknown period scenario, the power of GLR is generally better than the LS method, especially when the number of events in time bins is small. Both methods become equivalent in power as the number of events in each bin increases. We adopt LS in the rest of the analyses as it is computationally more efficient than GLR, especially for small time bin Δt and long run time $\mathcal{T}$.

B. Estimating $A_{d,\alpha,\beta}$ for arbitrary ($\mathcal{R}_s$, $\mathcal{R}_b$, $\mathcal{D}$) and fixed ($\mathcal{T}$,Δt, P) via the signal-to-noise ratio

In the previous section, $A_{d,\alpha,\beta}$ is obtained via a simulation process for a given $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t)$. This process is time intensive. It is desirable to obtain an analytical expression for the variation of $A_{d,\alpha,\beta}$ when ($\mathcal{R}_s$, $\mathcal{R}_b$, $\mathcal{D}$) take on different values. The signal rate $\mathcal{R}_s$ depends on flux uncertainty, $\sin\theta^2_w$, nuclei, physical process, energy threshold E_{thrd}, detector efficiency $\epsilon(E_r)$ and resolution $\sigma(E_r)$. The background rate $\mathcal{R}_b$ depends on the background depletion and physical process to generate the

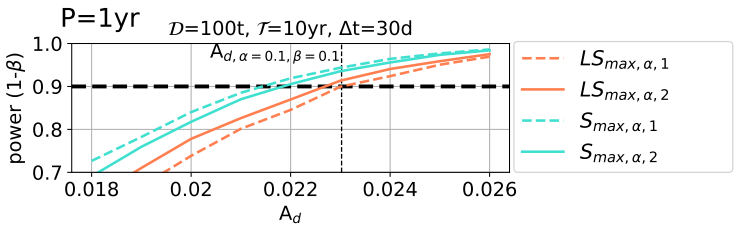


FIG. 5. Power curves (the probability of rejecting H_0) for the GLR and Lomb-Scargle methods when $\alpha = 0.1$ and $P = 1$ yr and under the "Unknown Period" scenario. The critical values are $LS_{\max,\alpha,1}$, $LS_{\max,\alpha,2}$ for LS and $S_{\max,\alpha,1}$ and $S_{\max,\alpha,2}$ for GLR. The vertical line is $A_{d,\alpha=0.1,\beta=0.1}$ under the LS method and when the critical value is $LS_{\max,\alpha,1}$.

$\mathcal{T}$	Δt	Period	$\mathcal{D}$	average number of	$A_{d,\alpha,\beta}$					
					"Known Period " scenario		"Unknown Period" scenario			
years	days	days	tons	events per bin	LS	GLR	$LS_{\max,\alpha,1}$	$LS_{\max,\alpha,2}$	$S_{\max,\alpha,1}$	$S_{\max,\alpha,2}$
10	30	365.25	100	821	0.015	0.015	0.023	0.023	0.021	0.022
10	10	365.25	100	273	0.015	0.015	0.023	0.023	0.023	0.023
1/6	0.2	1	100	5	0.12	0.12	0.195	0.193	0.188	0.189
1/6	0.2	1	500	27	0.055	0.054	0.087	0.086	0.084	0.084
1/6	0.2	1	2000	109	0.027	0.027	0.043	0.043	0.042	0.042

TABLE V. This table shows the amplitude $A_{d,\alpha,\beta}$ when $\alpha = 0.1$, $\beta = 0.1$ for scenarios 1) Known Period and 2) Unknown Period, and methods 1) the Lomb-Scargle periodogram and 2) the GLR, for different $\mathcal{D}$, $\mathcal{T}$, Δt and when $\mathcal{R}_s = 100$ ton^{-1} yr^{-1}, $\mathcal{R}_b = 0$ ton^{-1} yr^{-1}. Under the Known Period scenario, the critical value is $\chi^2_{2,\alpha}$ for LS and GLR. Under the Unknown Period scenario, the critical values are $LS_{\max,\alpha,1}$, $LS_{\max,\alpha,2}$ for LS and $S_{\max,\alpha,1}$ and $S_{\max,\alpha,2}$ for GLR.

background. The detector size $\mathcal{D}$ depends on the experiment. In this section, after obtaining $A_{d,\alpha,\beta}$ for $(\mathcal{R}_b, \mathcal{R}_s, \mathcal{D}, \mathcal{T}, P, \Psi, \Delta t)$, we find an efficient way to obtain $A'_{d,\alpha,\beta}$ for $(\mathcal{R}_b', \mathcal{R}_s', \mathcal{D}', \mathcal{T}, P, \Psi, \Delta t)$ so that the test maintains the same power.

We set $\alpha = 0.1$ and $(1-\beta) = 0.9$ for the "90% detection", then the corresponding detectable amplitude is denoted by $A_{d,\alpha=0.1,\beta=0.1}$. The signal for the detectable amplitude $A_{d,\alpha,\beta}$ in the time bin $[t_i, t_i+\Delta t]$ under H_1 is $\mathcal{R}_s \mathcal{D} A_{d,\alpha,\beta} \int_{t_i}^{t_i+\Delta t} \sin(2\pi u/P)\, du$. Here α, β, P, Δt, and $\mathcal{T}$ are assumed to be fixed (not changing). Also, the background is $\sqrt{(\mathcal{R}_s + \mathcal{R}_b)\mathcal{D}\Delta t}$, representing the Poisson standard deviation under the null model for the number of occurrences in one interval. Since the time-dependent parameters ($\mathcal{T}$, Δt, P) are fixed, the critical value stays the same during the scaling. In addition, the signal function deviated from the background on interval $[t_i, t_i+\Delta t]$ is $A_{d,\alpha,\beta}\mathcal{R}_s\mathcal{D}\int_{t_i}^{t_i+\Delta t} \sin(2\pi u/P)\, du$. Component $\int_{t_i}^{t_i+\Delta t} \sin(2\pi u/P)\, du$ determines the shape of the signal on $[0, T]$. When time-dependent parameters ($\mathcal{T}$, Δt, P) are fixed, the shape of this signal function over $[0, \mathcal{T}]$ may remain the same, but its magnitude may be different, depending on the parameters ($A_{d,\alpha,\beta}$, $\mathcal{R}_s$, $\mathcal{D}$). In this case, the power of the tests only depends on the signal's magnitude, because the shape's effect is the same. The squared signal-to-noise ratio (SNR^2) is

$$SNR^2\big|_{\mathcal{R}_s,\mathcal{R}_b,\mathcal{D}} = A^2_{d,\alpha,\beta} \times \mathcal{R}_s \times \mathcal{D} \times \frac{1}{(1+\mathcal{R}_b/\mathcal{R}_s)} \times F(P, \mathcal{T}, \Delta t),$$

where $F(P, \mathcal{T}, \Delta t)$ is some suitable effect due to the shape on the squared signal-to-noise ratio. Note that the signal-to-noise ratio for amplitude $A_{d,\alpha,\beta}$ when there is no constant background ($\mathcal{R}_s, \mathcal{R}_b = 0, \mathcal{D}$), is

$$SNR^2\big|_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}} = A^2_{d,\alpha,\beta} \times \mathcal{R}_s \times \mathcal{D} \times F(P, \mathcal{T}, \Delta t).$$

Therefore, for any $(\mathcal{R}_s, \mathcal{R}_b, \mathcal{D})$,

$$SNR^2|_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}} = SNR^2|_{\mathcal{R}_s,\mathcal{R}_b>0,\mathcal{D}} \times \left(1 + \frac{\mathcal{R}_b}{\mathcal{R}_s}\right). \tag{9}$$

Next,

$$\begin{aligned} SNR^2\big|_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}} &= A^2_{d,\alpha,\beta}\times\mathcal{R}_s\times\mathcal{D}\times F(P,\mathcal{T},\Delta t)\\ &= A^2_{d,\alpha,\beta}\times\mathcal{R}_s\left(\frac{\mathcal{R}_s^{'}}{\mathcal{R}_s^{'}}\right)\times\mathcal{D}\left(\frac{\mathcal{D}^{'}}{\mathcal{D}^{'}}\right)\times F(P,\mathcal{T},\Delta t)\\ &= \left(A_{d,\alpha,\beta}\sqrt{\frac{\mathcal{R}_s}{\mathcal{R}_s^{'}}}\sqrt{\frac{\mathcal{D}}{\mathcal{D}^{'}}}\right)^2\times\mathcal{R}_s^{'}\times\mathcal{D}^{'}\times F(P,\mathcal{T},\Delta t). \end{aligned} \tag{10}$$

Now, suppose that for another set of values of $(\mathcal{R}_s^{'},\mathcal{R}_b^{'}=0,\mathcal{D}^{'})$ the signal-to-noise ratio is the same as of $(\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D})$ while all other parameters except the amplitude are fixed for both cases, i.e.,

$$SNR^2\big|_{\mathcal{R}_s^{'},\mathcal{R}_b^{'}=0,\mathcal{D}^{'}} = A^{'2}_{d,\alpha,\beta}\times\mathcal{R}_s^{'}\times\mathcal{D}^{'}\times F(P,\mathcal{T},\Delta t) = SNR^2\big|_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}}. \tag{11}$$

Now, using (9) and (11) we obtain

$$SNR^2\big|_{\mathcal{R}_s^{'},\mathcal{R}_b^{'}>0,\mathcal{D}^{'}}\times\left(1+\frac{\mathcal{R}_b^{'}}{\mathcal{R}_s^{'}}\right) \;=\; SNR^2\big|_{\mathcal{R}_s^{'},\mathcal{R}_b^{'}=0,\mathcal{D}^{'}} = SNR^2\big|_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}}, \tag{12}$$

and using (10) in (12) we have

$$(A^{'}_{d,\alpha,\beta})^2\times\mathcal{R}_s^{'}\times\mathcal{D}^{'}\times\frac{1}{(1+\mathcal{R}_b^{'}/\mathcal{R}_s^{'})}\times F(P,\mathcal{T},\Delta t) \;=\; \left(A_{d,\alpha,\beta}\sqrt{\frac{\mathcal{R}_s}{\mathcal{R}_s^{'}}}\sqrt{\frac{\mathcal{D}}{\mathcal{D}^{'}}}\right)^2\times\mathcal{R}_s^{'}\times\mathcal{D}^{'}\times F(P,\mathcal{T},\Delta t),$$

which yields the amplitude for other setting $(\mathcal{R}_s^{'},\mathcal{R}_b^{'}=0,\mathcal{D}^{'})$

$$A^{'}_{d,\alpha,\beta} = A_{d,\alpha,\beta}\times\sqrt{\frac{\mathcal{R}_s}{\mathcal{R}_s^{'}}}\sqrt{\frac{\mathcal{D}}{\mathcal{D}^{'}}}\times\sqrt{1+\frac{\mathcal{R}_b^{'}}{\mathcal{R}_s^{'}}}. \tag{13}$$

As an example, for $\mathcal{T}=10$ yrs, $\Delta t=10$ days, $P=365.25$ days, $\mathcal{R}_s=100$ ton^{-1} yr^{-1}, $\mathcal{R}_b=0$ ton^{-1} yr^{-1}, $\mathcal{D}=100$ tons, the detectable amplitude with a power of 90% and a level of significance 10% is $A_{d,\alpha=0.1,\beta=0.1}=0.02357377$ (results in Figure 5). Using Equation (13), we can now obtain the detectable amplitude $A^{'}_{d,\alpha=0.1,\beta=0.1}=0.1292$ for $\mathcal{D}^{'}=100$ tons, $\mathcal{R}_s^{'}=474.367$ ton^{-1} yr^{-1} (ES Xenon resolution +efficiency) and $\mathcal{R}_b^{'}=50060.5$ ton^{-1} yr^{-1} (all backgrounds). This calculation is done without running a time-consuming simulation. The detectable amplitudes for other possible values of $\mathcal{R}_s^{'}$ and $\mathcal{R}_b^{'}$ are given in the "signal components" and "background components" column in Table I II. Moreover, when the eccentricity $A^{'}_{d,\alpha=0.1,\beta=0.1}=A_{ecc}=0.03342$, using Equation (13) we obtain the detector's size $\mathcal{D}^{'}$

$$\mathcal{D}^{'} = \mathcal{D}\left(A_{d,\alpha=0.1,\beta=0.1}\times\frac{\sqrt{1+\mathcal{R}_b^{'}/\mathcal{R}_s^{'}}}{A_{ecc}}\right)^2\left(\frac{\mathcal{R}_s}{\mathcal{R}_s^{'}}\right)$$

interpreted as the detector size required for the eccentricity to be 90% detected.

Figures 8 and 9 show the amplitudes for different combinations of $(\mathcal{T},\Delta t,P,\alpha,\beta)$. For example, running simulations we obtain $A_{d,\alpha=0.1,\beta=0.1}=0.0291161$, $A_{d,\alpha=0.05,\beta=0.1}=0.03$ for $\mathcal{T}=10$ yrs, $\Delta t=0.2$ days, $P=1$ day, $\mathcal{R}_s=100$ ton^{-1} yr^{-1}, $\mathcal{R}_b=0$ ton^{-1} yr^{-1}, $\mathcal{D}=100$ tons. Then using Equation (13) we obtain $A^{'}_{d,\alpha=0.1,\beta=0.1}$, $A^{'}_{d,\alpha=0.05,\beta=0.1}$ for the same $\mathcal{T}$, Δt, P, and different choices of $(\mathcal{R}_s^{'},\mathcal{R}_b^{'},\mathcal{D}^{'})$. We also consider the cases, ($\mathcal{T}=10$ yrs, $\Delta t=0.2$ days, $P=1$ day), ($\mathcal{T}=10$ yrs, $\Delta t=10$ days, $P=1$ yr) and ($\mathcal{T}=10$ yrs, $\Delta t=30$ days, $P=1$ yr).

C. Estimating $A_{d,\alpha,\beta}$ to have the required power with different Δt and $\mathcal{T}$

We now fix $\mathcal{R}_s=100$ ton^{-1} yr^{-1}, $\mathcal{R}_b=0$ ton^{-1} yr^{-1}, $\mathcal{D}=100$ tons and explore the relation between the amplitude $A_{d,\alpha,\beta}$ and the time-dependent parameters $(\mathcal{T},\Delta t,P)$. For a given $(\mathcal{T},\Delta t)$, we scan a series of frequencies (P) to obtain $A_{d,\alpha,\beta,P}$ for each P.

Then we find the P range where $A_{d,\alpha,\beta}$ is roughly constant $A_{d,\alpha,\beta,P} \approx \overline{A}_{d,\alpha,\beta}$ where we use $\overline{A}_{d,\alpha,\beta}$ to denote the average of the amplitudes over different Ps.

The simulations show how different T and Δt affect the strength of signal $A_{d,\alpha,\beta}$ to maintain 90% power of the test. With the requirement that $P \gg \Delta t$, $\int_{t_i}^{t_i+\Delta t} \sin(\omega v) dv \approx \Delta t \sin(\omega t_i)$. Therefore, when $\mathcal{R}_b = 0$, the signal noise ratio $SNR^2_{\mathcal{R}_s,\mathcal{R}_b=0,\mathcal{D}}$ is approximated by $A^2_{d,\alpha,\beta} \times \mathcal{R}_s \times \mathcal{D} \times \Delta t \times F_{sin}(\mathcal{T},P)$, where $F_{sin}(\mathcal{T},P)$ represents the effect on the signal-to-noise ratio due to different parameters T, and P, considering a signal with the shape of $sin(\omega t)$. The sample size of observations is $N_t = T/\Delta t$. The sample can differ with $\mathcal{T}$ and Δt. When all other parameters are fixed, but Δt decreases, the signal-to-noise ratio decreases, while the sample size N_t increases. Consequently, the expected power shall remain the same.

The simulation results showing how $A_{d,\alpha,\beta}$ depends on different parameters as a function of P are given in Figures 6 7. Each panel is the $A_{d,\alpha=0.1,\beta=0.1,P}$ for a series of scanned frequencies (P) under a given $(\mathcal{T},\Delta t)$. $A_{d,\alpha=0.1,\beta=0.1,P}$ is roughly constant when Δt is small enough such that the reached A_d is the same as the input A_d, $10\Delta t < P < \mathcal{T}/4$ when $\Delta t \geq 10$ days, and $20\Delta t < P < \mathcal{T}/4$ when $\Delta t < 10$ days. In addition, we found an interesting empirical relationship that by an equation $y = ax^2 + bx + c$, where $y = \log(\overline{A}^2_{d,\alpha,\beta}\mathcal{R}_s\mathcal{D}\Delta t)$ and x depends on the sample size of observation N_t and the critical value. Specifically, $x = \log\{\tau(\mathcal{M}_{scan})/N_t\}$, where $\tau(\mathcal{M}_{scan}) = max(1, \log(\mathcal{M}_{scan}))$. Under the "Known Period scenario", $\mathcal{M}_{scan} = 1$ and $x = \log(1/N_t)$. For a given α, the critical value is constant under the "Known Period scenario", and for "Unknown Period scenario", with $\mathcal{M}_{scan} \gg 1$, the critical value $LS_{max,\alpha,2} \approx CONST + \log(M_{scan})$. The corresponding x and the fitted a, b, c for different parameters are showed in Table VI

	x_{min}	x_{max}	α	β	a	b	c
Known Period	-10.218	-4.291	0.1	0.1	3.004e-03	1.049	3.251
			0.05	0.1	5.168e-03	1.077	3.529
Unknown Period	-7.5	-2.755	0.1	0.1	2.344e-02	1.309	3.037
			0.05	0.1	2.598e-02	1.345	3.224

TABLE VI. Fitted values of a, b, c for the empirical relationship $y = ax^2 + bx + c$ within $[x_{min}, x_{max}]$ assuming $\mathcal{R}_s$ = 100 ton^{-1} yr^{-1}, $\mathcal{R}_b$ = 0, $\mathcal{D}$ = 100 ton.

Using the above empirical relations, to summarize the procedure to find $A_{d,\alpha,\beta}$, first, we select the proper Δt and $\mathcal{T}$ based on possible P such that $A_{d,\alpha,\beta,P} \approx \overline{A}_{d,\alpha,\beta}$. Adjust $\mathcal{M}_{scan}$ such that $LS_{max,\alpha,1} \approx LS_{max,\alpha,2}$ to ensure H_0 is consistent with the model, and obtain $\mathcal{M}_{scan}$ and x, then obtain $\overline{A}_{d,\alpha,\beta} = \sqrt{\frac{e^{ax^2+bx+c}}{\mathcal{R}_s\mathcal{D}\Delta t}}$ using parameters in Table VI and obtain

$$\overline{A}'_{d,\alpha,\beta} = \overline{A}_{d,\alpha,\beta} \times \sqrt{\frac{\mathcal{R}_s}{\mathcal{R}_s'}} \sqrt{\frac{\mathcal{D}}{\mathcal{D}'}} \sqrt{1 + \frac{\mathcal{R}_b'}{\mathcal{R}_s'}} \tag{14}$$

via Eqn 13. Then, the nearby amplitudes are scanned for more accurate simulations. The detector size $\mathcal{D}'$ can be obtained similarly via $\mathcal{D}' = \mathcal{D} \times \left(\frac{\overline{A}_{d,\alpha,\beta}}{A_{ecc}}\right)^2 \frac{\mathcal{R}_s}{\mathcal{R}_s'} \left(1 + \frac{\mathcal{R}_b'}{\mathcal{R}_s'}\right)$.

Figure 11 13 15 are the application of this method to estimate $\overline{A}'_{d,\alpha=0.1,\beta=0.1}$ as a function of runtime $\mathcal{T}$. We choose Δt = 0.05 days for P = 1 day, and Δt = 10 days for P = 1 yr and 11 yrs, such that $A'_{d,\alpha,\beta,P} \approx \overline{A}'_{d,\alpha,\beta}$ for these periods. The minimum runtime $\mathcal{T}$ when $\overline{A}'_{d,\alpha=0.1,\beta=0.1}$ meets $A_{ecc} = 0.03342$ or $A_{d,DN,max} = 0.00891$ or A_{atm} can be interpreted as the above physical amplitudes are the 90% detectable amplitudes. Though the empirical results are fitting within $[x_{min}, x_{max}]$, the results from simulating the nearby run time $\mathcal{T}$ is consistent with these estimated $\mathcal{T}$ when $x < x_{min}$ or $x > x_{max}$.

VI. RESULTS

We now move on to presenting the results for the time-varying analysis, focusing first on solar neutrinos. As above, we consider the electron neutrino scattering channel (ES) and nuclear recoil channels (CEvNS) separately, and the results are split into "Known Period" and "Unknown Period" scenarios. Starting with the ES channel, we take the end-point energy of ^{7}Be and calculate $R_\odot$ in Equation 1 by summing all the components of the solar neutrino flux (pp, CNO, ^{7}Be, and pep) over a range [1 keV, 650 keV] for the ideal resolution + ideal efficiency case and [0 keV, 790 keV] for the smear+efficiency case, since the contribution outside of this range is dominated by $2\nu\beta\beta$. The flux within these ranges is dominated by pp neutrinos.

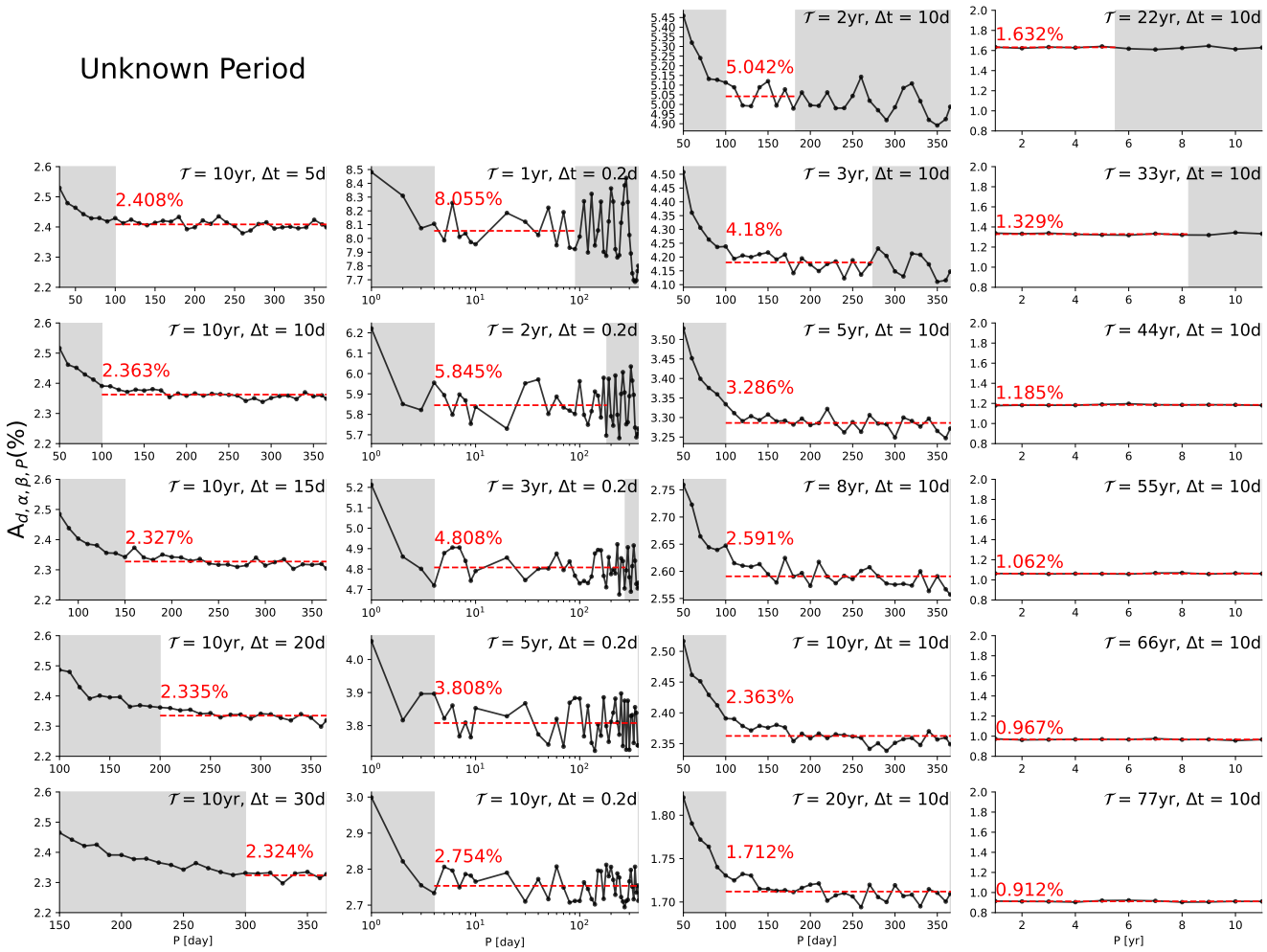


FIG. 6. The "Unknown Period" scenario plot for 90% detection ($\alpha = 0.1$, $\beta = 0.1$) amplitude $A_{d,\alpha,\beta,P}$ as a function of P for different ($\mathcal{T}$, Δt) and when $\mathcal{R}_s = 100$ ton^{-1} yr^{-1}, $\mathcal{R}_b = 0$ ton^{-1} yr^{-1} and $\mathcal{D} = 100$ tons in the simulation study. *1st column:* $\mathcal{T} = 10$ years, Δt varies from 5 to 30 days. *2nd column:* $\Delta t = 0.2$ days and $\mathcal{T}$ varies from 1 to 10 years. *3rd column:* $\Delta t = 10$ days and $\mathcal{T}$ varies from 2 to 20 years. *4th column:* $\Delta t = 10$ days and $\mathcal{T}$ varies from 22 to 77 years. In each panel, the white region is where the amplitude $A_{d,\alpha,\beta,P}$ can be approximated by the average of the amplitudes over different frequencies $\overline{A}_{d,\alpha,\beta}$, and the red horizontal line is $\overline{A}_{d,\alpha,\beta}$.

We first show the results as a function of different depletion of backgrounds for ES, and as a function of different energy thresholds for CEνNS, under the fiducial case of a 100-ton detector, with a run time of 10 years and a fixed time-bin Δt (Figure 8 9). Then we show the results as a function of detector run-time for a fixed time-bin Δt, detector size $\mathcal{D}$ for ES and CEνNS (Figures 10 11 12 13 14 15) .

Figure 8 shows the required amplitude at which the probability of detecting the signal is $1-\beta = 0.9$, at $\alpha = 0.1$. Indicated on the figure as the upper horizontal line is the modulation amplitude induced by the eccentricity. We see that for a $2\nu\beta\beta$ fraction $\lesssim 10\%$ at the assumed exposure the experiment will be sensitive to the time variation from eccentricity. Also shown as the lower horizontal lines are the upper limits on the day-night modulation of the ^{7}Be component from Borexino, $A_d = A_{DN}/\sqrt{2} \approx 0.009$ [16]. This shows that a $2\nu\beta\beta$ fraction in Xenon at a level $\sim 1\%$ has a detectable amplitude with 90% power similar to the best-fit amplitude A_d from Borexino [4]. The Xenon measurements are sensitive to the lower energy pp component, for which there are no strong bounds on the time variation.

Moving on to the CEνNS channel using the ^{8}B flux, Figure 9 shows the required amplitude for the 90% detection at $\alpha = 0.1$, as a function of nuclear recoil energy threshold. Shown are the results for different assumptions of the efficiency, in one case

[4] Although these two amplitudes are similar, they cannot be compared directly since the Borexino result comes from fitting the binned-data and our result comes from the likelihood ratio test.

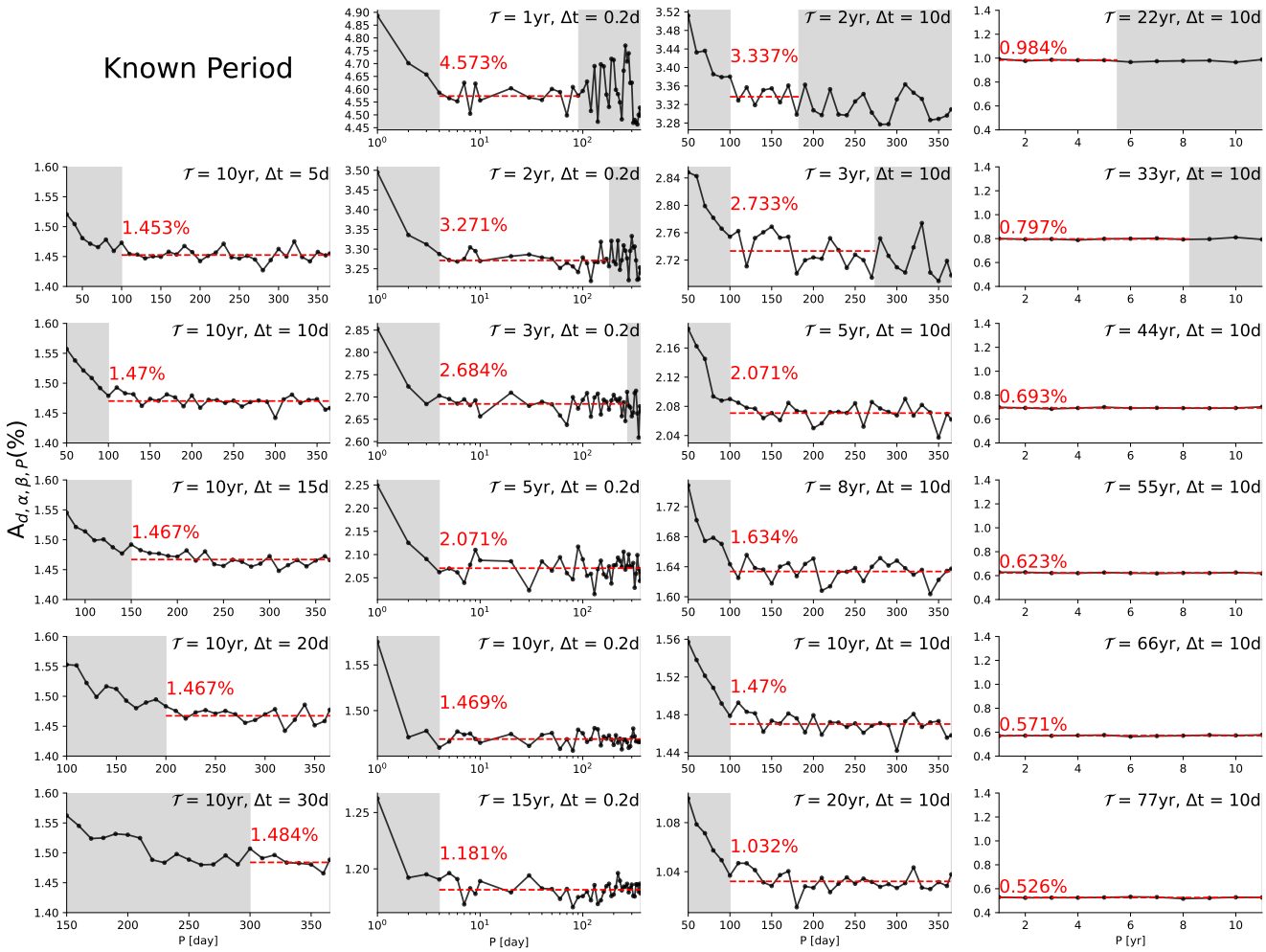


FIG. 7. Same as Figure 6 but for known periods.

showing both perfect efficiency and in the remaining cases using the detector efficiency and energy resolution modeled from NEST simulations (Figure 3). From Figure 9 we see that the eccentricity is clearly detectable for the assumed exposure of 100 ton and 10 years for the ideal case. For the case with the efficiency and resolution corrections added, for thresholds $\lesssim 1$ keV there is a saturation and the eccentricity is nearly detectable given the specifications used. Note that since ^{8}B neutrinos are detected via a neutral current interaction, there is no sensitivity to the day-night effect in the CEνNS channel (there could be sensitivity if radiative corrections were detectable [44]).

Figure 10 shows the power of detecting daily and yearly time variation through electron recoils as a function of detector runtime, for an assumed 100-ton detector size. We see that even for large 2νββ fractions, the eccentricity will be detectable at "90%" chance under the Known Period scenario with an exposure of $\lesssim 10$ years. Here we choose a time binning of $\Delta t = 10$ days, though we find that our results are relatively insensitive to the time binning. Similarly, under the Known Period scenario, the Borexino sensitivity can be achieved at "90%" chance for run times ~ 7 years when background includes only 2νββ and is depleted to 0.1%.

Figure 11 shows the "90% detection" amplitude of detecting time variation through electron recoils as a function of detector runtime for Δt = 0.05 days and an assumed 50-ton exposure. The 3% eccentricity will be detectable with 90% power under the Known Period scenario within ~ 13 years for large 2νββ fractions and other backgrounds, while 2νββ needs to be depleted down to $\lesssim 0.1\%$ to achieve the similar result under the "Unknown Period scenario". For daily variations, a nearly-fully depleted detector will be able to match the Borexino upper limits on daily time variation for a run time of $\lesssim 13$ years under the Known Period scenario.

Figure 12 shows the power of detecting time variation of ^{8}B as a function of detector runtime, for an assumed detector size of 100 ton. Here we see that for an idealized detector, yearly time modulation is detectable at "90%" for a runtime of ~ 3 years

under the Known Period scenario and ~ 5 years under the Unknown Period scenario. We note that this is similar to the time for detection of eccentricity using the electron recoil channel, for a low 2νββ fraction. For the "Xe100t-5" efficiency model, the detection timescale increases to $\lesssim 15$ years.

Figure 13 shows amplitude for 90% detection of time variation of an arbitrary amplitude for ^{8}B as a function of detector runtime for Δt = 10 days and an assumed 50-ton or 100-ton exposure. From this we again see that the eccentricity will be detectable with 90% power under the Known Period scenario within $\lesssim 10$ years for idealized detector or the resolution+"Xe100t-5" efficiency detector. Under the Unknown Period scenario, an idealized detector is needed to achieve the similar run time for a 50-ton or 100-ton detector.

We finally checked the prospects for detection of time variation of atmospheric neutrinos. The resulting power for detecting the modulation of atmospheric neutrinos from Section III C at each location (the amplitude A_{atm} is 0.0459 at CJPL, 0.0382 at Kamioka, 0.0461 at LNGS, 0.1475 at SURF and 0.1327 at SNOlab) is shown in Figure 14, for the case of an ideal detector with no efficiency or energy resolution correction. Given that the signal is weaker than the signal for solar neutrinos, in this case we consider larger detector volumes of 200, 600, and 1000 ton, and plot the power as a function of the experimental run time. Figure 15 shows the "90%" amplitude of detecting time variation of atmospheric neutrinos as a function of detector runtime for Δt = 30 days and an assumed 200-ton or 600-ton exposure. Though the experimental volumes considered in Figures 14 15 are larger than that in our solar neutrino analysis, they provide a sense of what will be required to detect the atmospheric time variation.

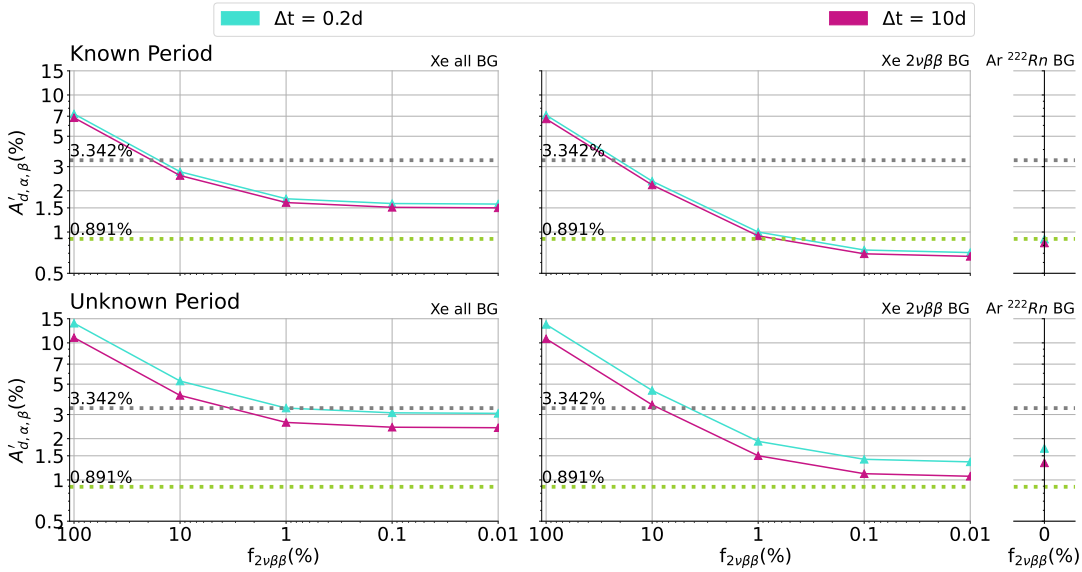


FIG. 8. 90% detection (α = 0.1, β = 0.1) amplitude $A'_{d,\alpha,\beta}$ for solar components through electron scattering obtained from Equation 13 as a function of 2νββ fraction $f_{2\nu\beta\beta}$. Dashed lines are the amplitude of yearly modulation $A_{ecc} = 0.03342$ and the day-night modulation $A_{d,DN,max} = 0.00891$. The label "all BG" includes 2νββ, ^{222}Rn, and ^{85}Kr backgrounds. The assumed runtime is $\mathcal{T}$ = 10 years, and the detector size $\mathcal{D}$ = 100 ton. The time binnings and periods are $\Delta t = 0.2$ days for P = 1 day, $\Delta t = 10$ days for P = 1 yr. The curves show the spectrum including energy resolution and "Xe100t-5" detector efficiency for Xenon (Figure 3), and 10 % energy resolution with recoil energy threshold 100 keV only for Argon. Each curve is for a different time binning, Δt, as indicated. The top horizontal line shows the expected amplitude due to the eccentricity, and the bottom horizontal line shows the bound on the amplitude of the day-night modulation from Borexino.

VII. CONCLUSIONS

In this paper we have studied the prospects for detecting time variation of solar and atmospheric neutrinos in future large-scale Xenon and Argon dark matter detectors. We have developed rigorous statistical methods for detecting a time-varying signal, and compared the statistical methods under the assumptions that we are searching for a known periodicity or an unknown periodicity.

For time varying signals, we have focused on detecting solar neutrinos through ES and CEνNS. In the ES channel, for a 100-ton Xenon detector running for 10 years with a Xenon-136 fraction of $\lesssim 0.1\%$, a time-variation amplitude of $\sim 1\%$ is detectable for the experimental configurations that we study. This is sufficient to detect time variation due to eccentricity. For this same run time, Xenon experiments will achieve similar sensitivity to the Borexino experiment for daily modulations that will be induced

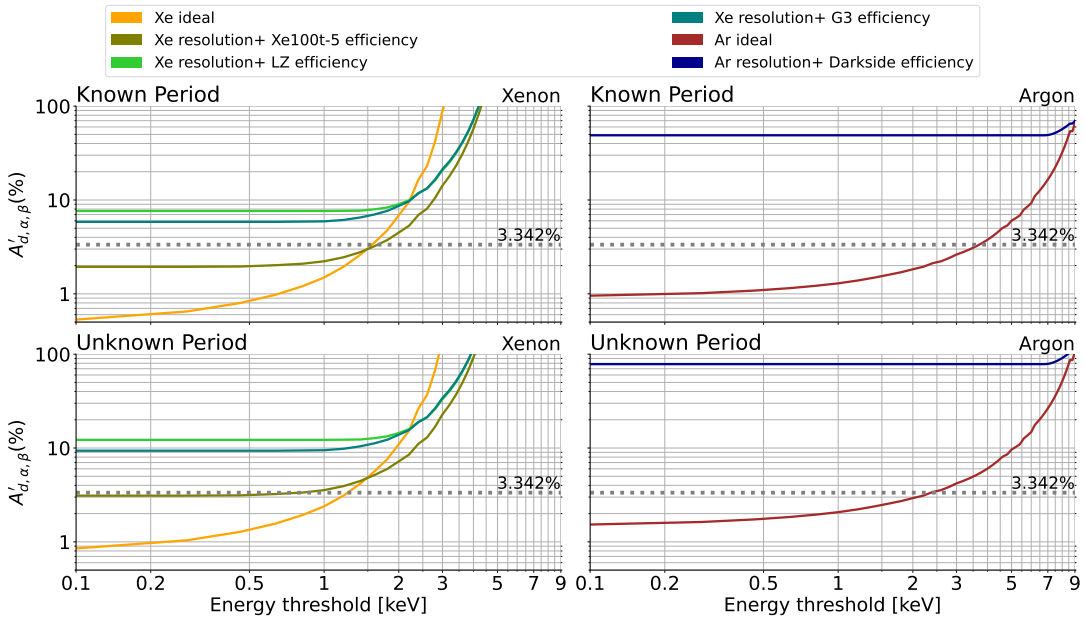


FIG. 9. 90% detection ($\alpha = 0.1$, $\beta = 0.1$) amplitude $A'_{d,\alpha,\beta}$ for ^{8}B solar components through CEvNS obtained from Equation 13 as a function of recoil energy threshold, using $\Delta t = 10$ days. The assumed runtime is $\mathcal{T} = 10$ years, and the detector size is $\mathcal{D} = 100$ tons. Curves are shown for different detector and efficiency models as indicated. The dashed line is the amplitude of yearly modulation $A_{ecc} = 0.03342$.

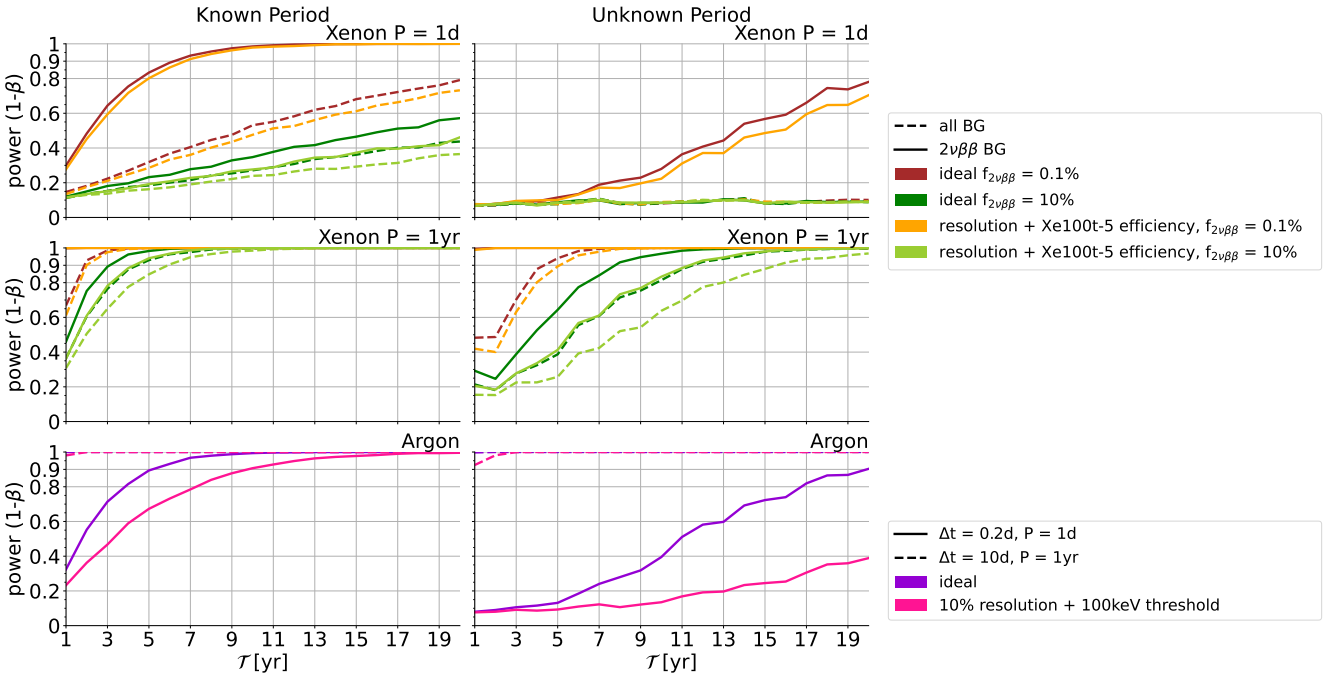


FIG. 10. The power for detecting time variation of solar components (summing over pp, ^{7}Be, CNO and pep components) through electron scattering, under the assumption of the "Known Period" (left panel) and "Unknown Period" (right panel) scenarios described in the text. We take $\alpha = 0.1$, and plot the results as a function of detector runtime $\mathcal{T}$. The yearly modulation has an amplitude of $A_{ecc} = 0.03342$, due to the eccentricity of the Earth orbit, and the daily modulation has an amplitude of $A_{d,DN,max} = 0.00891$, which matches the Borexino upper limits on the daily modulation. The assumed detector size is $\mathcal{D} = 100$ ton. *Top row*: Xenon detection of $A_{d,DN,max}$ for a time binning of $\Delta t = 0.2$ days. *Middle row*: Xenon detection of A_{ecc} with $\Delta t = 10$ days. *Bottom row*: Argon detection of A_{ecc} with $\Delta t = 10$ days, and $A_{d,DN,max}$ with $\Delta t = 0.2$ days. For Xenon, "all BG" includes 2$\nu\beta\beta$, ^{222}Rn and ^{85}Kr. For Argon, the background includes ^{222}Rn.

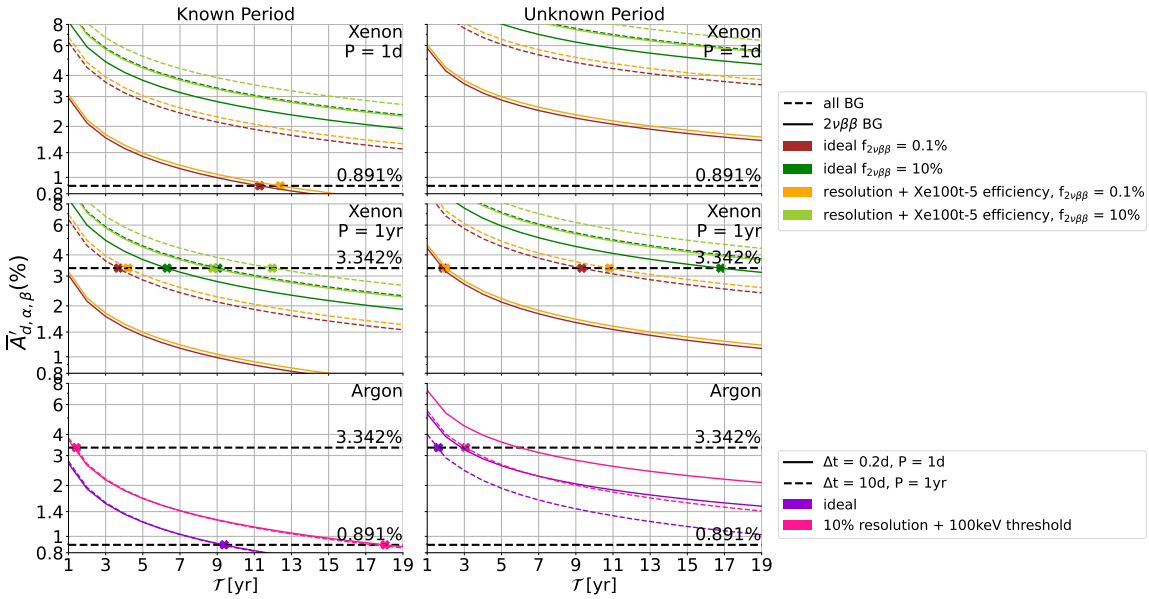


FIG. 11. The detectable amplitude for solar components through electron scattering with 90% power and 10% level of significance ($\alpha = 0.1$, $\beta = 0.1$), estimated from Equation 14 as a function of detector runtime $\mathcal{T}$ for $\mathcal{D}' = 50$ tons, under "Known Period" and "Unknown Period" scenarios. Dashed lines are the amplitude of yearly modulation $A_{ecc} = 0.03342$ and the day-night modulation $A_{d,DN,max} = 0.00891$. *Top row*: Xenon detection of $A_{d,DN,max}$ for a time binning of $\Delta t = 0.05$ days. *Middle row*: Xenon detection of A_{ecc} with $\Delta t = 10$ days. *Bottom row*: Argon detection of A_{ecc} with $\Delta t = 10$ days, and of $A_{d,DN,max}$ with $\Delta t = 0.05$ days. Points indicate the run time $\mathcal{T}$ when $\overline{A}'_{d,\alpha=0.1,\beta=0.1}$ reaches A_{ecc} or $A_{d,DN,max}$. For Xenon, "all BG" includes 2νββ, ^{222}Rn and ^{85}Kr. For Argon, the background includes ^{222}Rn.

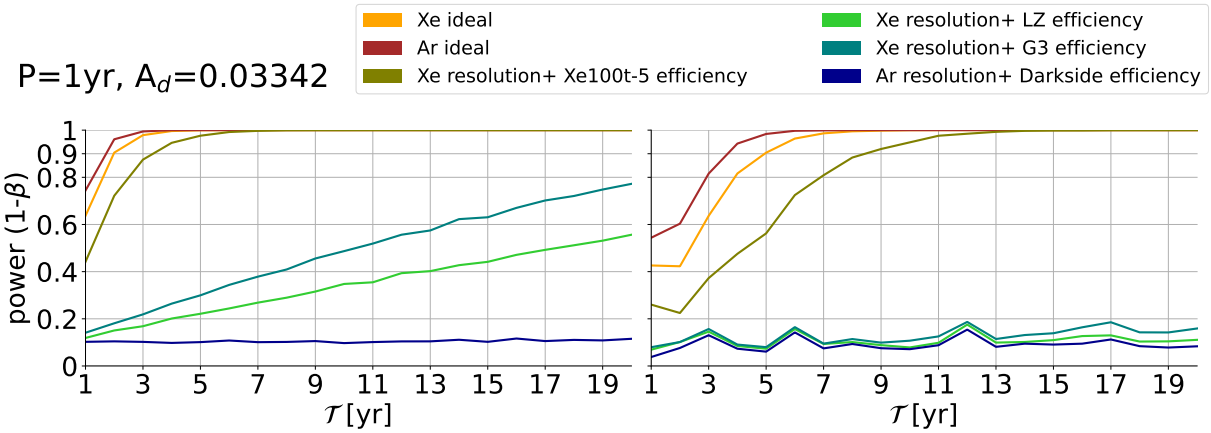


FIG. 12. The power of detecting time variation of ^{8}B through CEνNS under "Known Period" (left column) and "Unknown Period" (right column) scenarios with $\alpha = 0.1$, as a function of detector runtime $\mathcal{T}$. The assumed detector size is 100 ton. For Xenon assumptions are for an ideal detector and the "Xe100t-5", "LZ" and "G3" detector efficiencies, and for Argon an ideal detector and the "Darkside" detector efficiency.

from the propagation of electron neutrinos through the Earth. Both of these results will hold if the 2νββ fraction can be reduced to $\lesssim 0.01$.

In the CEνNS channel, we find that for a threshold of 1 keV, a detector with ideal efficiency and energy resolution will be sensitive the a yearly variation in the flux. For a detector with efficiency and resolution similar to current Xenon experiments, detectors will approach sensitivity to the eccentricity for thresholds $\lesssim 1$ keV. In comparison to previous studies which considered the time variation in the nuclear recoil channel due to the eccentricity [9], we add the more realistic detector scenarios to our modeling. The detector modeling effects most strongly affect the CEνNS channel, whereas the ES channel is most strongly affected by the 2νββ fraction, as discussed above.

Dark matter experiments provide the unique possibility to measure the time variation of the solar neutrino flux through multiple

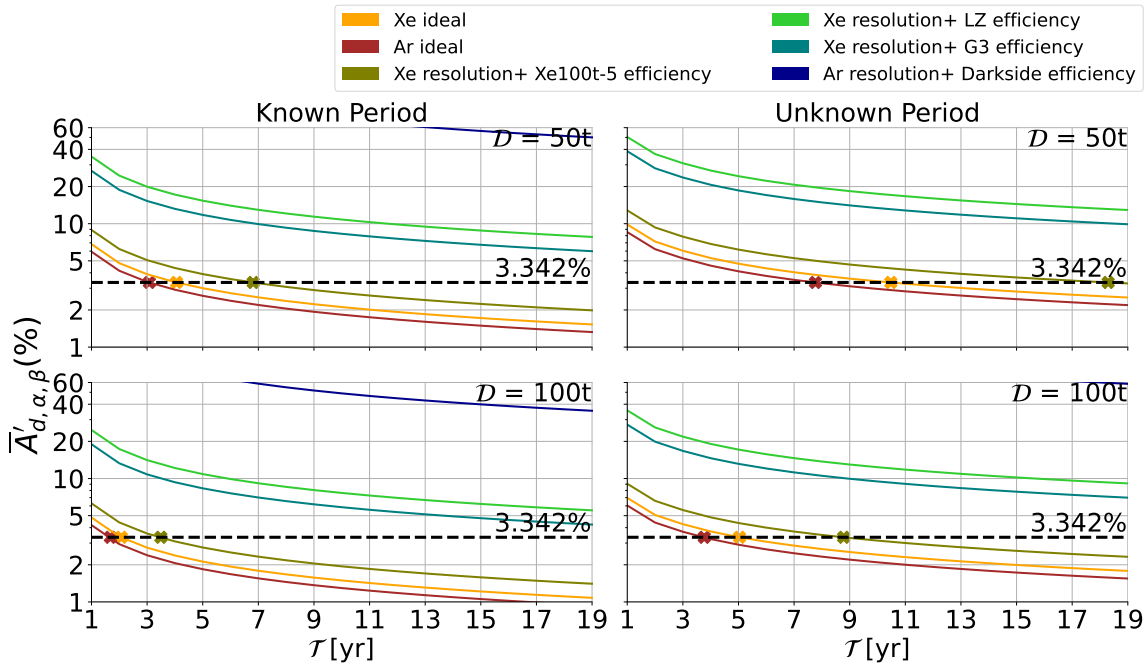


FIG. 13. The detectable amplitude for ^{8}B neutrinos through CEvNS with 90% power and 10% level of significance ($\alpha = 0.1, \beta = 0.1$), estimated from Equation 14 as a function of detector runtime $\mathcal{T}$, for $\mathcal{D}' = 50, 100$ tons, under "Known Period" and "Unknown Period" scenarios, when $\Delta t = 10$ days. The dashed line is the amplitude of yearly modulation $A_{ecc} = 0.03342$. Crosses indicate the run time $\mathcal{T}$ when $\overline{A}'_{d,\alpha=0.1,\beta=0.1}$ reaches A_{ecc}. Configurations of ($\mathcal{R}_s'$, $\mathcal{R}_b'$, $\mathcal{D}'$) from different detector target, resolution, and efficiency are indicated.

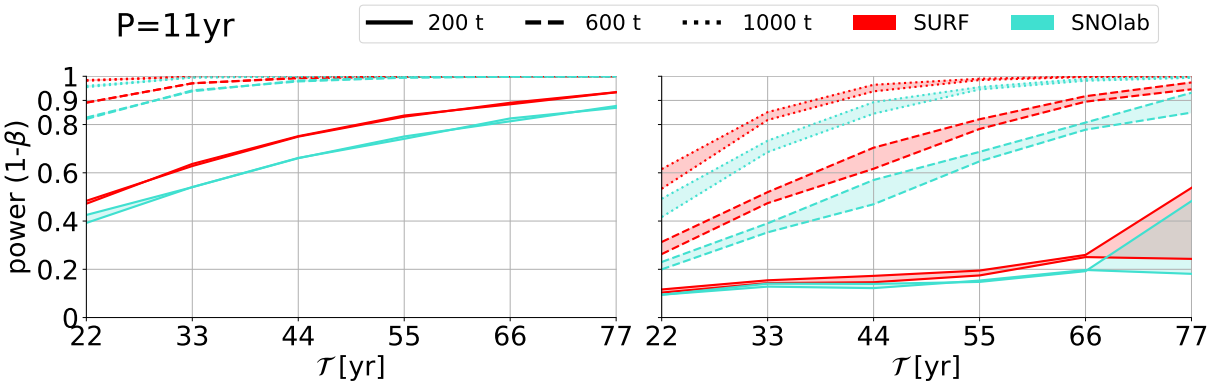


FIG. 14. Power for detecting atmospheric neutrino time variation at SURF and SNOlab through nuclear recoils under "Known Period" (left panel) and "Unknown Period" (right panel) scenarios for $\alpha = 0.1$, as a function of detector runtime $\mathcal{T}$. The assumed detector sizes $\mathcal{D}$ are 200, 600, 1000 tons. Bands indicate the differences between Δt = 10 days and Δt = 30 days for assumptions of an ideal detector.

detection channels. Detecting eccentricity in both channels would be important because it confirms the solar origin of the signal. Further, it provides a test for the long term stability of the detector, and provides a means to identify possible time-varying signals of unknown origin.

ACKNOWLEDGEMENTS

Y. Z. and L. E. S. are supported by the DOE Grant No. DE-SC0010813.

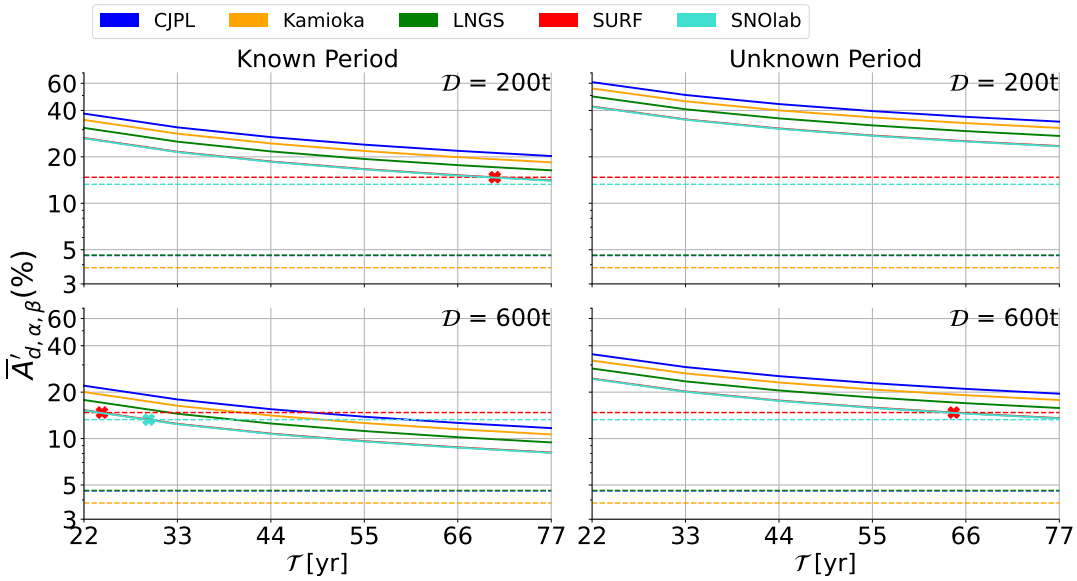


FIG. 15. The detectable amplitude for atmospheric neutrinos through CEvNS with 90% power and 10% level of significance ($\alpha = 0.1$, $\beta = 0.1$), estimated via Equation 14 as a function of detector runtime $\mathcal{T}$, for $\mathcal{D}' = 200,600$ tons, under "Known Period" and "Unknown Period" scenarios, when $\Delta t = 30$ days. The dashed horizontal lines are A_{atm} in Section III C, and each color is a different location as indicated. Crosses indicate the run time $\mathcal{T}$ when $\overline{A}'_{d,\alpha=0.1,\beta=0.1}$ reaches A_{atm} at the corresponding location. A xenon detector with ideal resolution and efficiency is assumed.

VIII. APPENDIX

Table VII shows the critical values $LS_{max,\alpha,1}$ and $LS_{max,\alpha,2}$ for Δt = 0.2 days and $\mathcal{T}$ from 1 yr to 20 yrs.

Table VIII shows the critical values $LS_{max,\alpha,1}$ and $LS_{max,\alpha,2}$ for Δt = 5 days and $\mathcal{T}$ from 1 yr to 22 yrs.

Table IX shows the critical values $LS_{max,\alpha,1}$ and $LS_{max,\alpha,2}$ for Δt = 10 days and $\mathcal{T}$ from 1 yr to 22 yrs.

$\mathcal{T}$ [yr]	$\mathcal{M}_{entire}$	$\mathcal{M}_{scan}$	$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.02$	
			$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$
1	12175	2429	10.05	9.83	10.77	10.56	11.7	11.37
2	24350	4859	10.74	10.42	11.46	11.25	12.39	12.27
3	36525	7288	11.14	10.98	11.86	11.71	12.8	12.87
4	48700	9718	11.43	11.18	12.15	11.89	13.08	12.68
5	60875	12148	11.66	11.39	12.38	12.15	13.31	13.07
6	73050	14577	11.84	11.73	12.56	12.47	13.49	13.28
7	85225	17007	11.99	12.05	12.71	12.84	13.64	13.84
8	97400	19437	12.13	11.89	12.85	12.7	13.78	13.79
9	109575	21866	12.24	12.01	12.96	12.76	13.89	13.58
10	121750	24296	12.35	12.14	13.07	12.87	14.0	13.83
11	133925	26726	12.44	12.34	13.16	13.11	14.1	14.0
12	146100	29155	12.53	12.49	13.25	13.24	14.18	14.33
13	158275	31585	12.61	12.62	13.33	13.32	14.26	14.24
14	170450	34015	12.68	12.78	13.4	13.53	14.34	14.46
15	182625	36444	12.75	12.54	13.47	13.23	14.41	14.2
16	194800	38874	12.82	12.61	13.54	13.33	14.47	14.21
17	206975	41304	12.88	12.66	13.6	13.37	14.53	14.22
18	219150	43733	12.94	12.77	13.66	13.51	14.59	14.3
19	231325	46163	12.99	12.75	13.71	13.45	14.64	14.53
20	243500	48593	13.04	12.97	13.76	13.69	14.69	14.65

TABLE VII. Critical values for Δt = 0.2 days, f_{scan} = [1/365.25 day^{-1}, 1 day^{-1}], n_o = 10, f_{adj}=1.5

f_{adj}	$\mathcal{T}$ [yr]	$\mathcal{M}_{entire}$	$\mathcal{M}_{scan}$	$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.02$	
				$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$
7	1	104	104	6.9	6.79	7.61	7.44	8.55	8.39
6	2	243	243	7.74	7.68	8.46	8.43	9.4	9.53
6	3	365	365	8.15	8.66	8.87	9.92	9.8	12.29
6	4	487	487	8.44	8.55	9.16	9.32	10.09	10.66
6	5	608	608	8.66	8.44	9.38	9.14	10.31	9.96
6	6	730	730	8.84	9.31	9.56	10.56	10.49	12.83
6	7	852	852	9.0	8.96	9.72	9.67	10.65	10.62
6	8	974	974	9.13	9.14	9.85	9.94	10.78	10.92
6	9	1095	1095	9.25	9.14	9.97	9.85	10.9	10.84
6	10	1217	1217	9.35	9.23	10.07	9.98	11.01	11.03
6	11	1339	1339	9.45	9.48	10.17	10.3	11.1	11.34
6	12	1461	1461	9.54	10.0	10.26	11.05	11.19	13.2
6	13	1582	1582	9.62	9.58	10.34	10.4	11.27	11.26
6	14	1704	1704	9.69	9.71	10.41	10.52	11.34	11.64
6	15	1826	1826	9.76	9.83	10.48	10.59	11.41	11.56
6	16	1948	1948	9.82	9.94	10.54	10.68	11.48	11.63
6	17	2069	2069	9.89	10.12	10.61	10.9	11.54	11.98
6	18	2191	2191	9.94	9.8	10.66	10.59	11.59	11.51
6	19	2313	2313	10.0	9.91	10.72	10.63	11.65	11.57
6	20	2435	2435	10.05	10.03	10.77	10.77	11.7	11.68
6	21	2556	2556	10.1	10.16	10.82	10.87	11.75	12.0
6	22	2678	2678	10.14	10.12	10.86	10.83	11.79	11.9

TABLE VIII. Critical values for Δt = 5 days, $f_{scan} = [1/365.25\,\text{day}^{-1},\ 1/5\,\text{day}^{-1}]$, $n_o = 10$

f_{adj}	$\mathcal{T}$ [yr]	$\mathcal{M}_{entire}$	$\mathcal{M}_{scan}$	$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.02$	
				$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$	$LS_{max,\alpha,1}$	$LS_{max,\alpha,2}$
7	1	52	52	6.2	5.76	6.92	6.42	7.85	7.28
7	2	104	104	6.9	6.88	7.61	7.62	8.55	8.63
6	3	182	182	7.45	8.47	8.17	10.26	9.11	13.55
6	4	243	243	7.74	7.57	8.46	8.35	9.4	9.42
6	5	304	304	7.97	7.66	8.69	8.32	9.62	9.35
6	6	365	365	8.15	9.07	8.87	10.64	9.8	14.11
6	7	426	426	8.3	8.1	9.02	8.83	9.96	9.66
6	8	487	487	8.44	8.42	9.16	9.21	10.09	10.35
6	9	547	547	8.55	8.35	9.27	9.03	10.21	9.91
6	10	608	608	8.66	8.4	9.38	9.2	10.31	10.11
6	11	669	669	8.76	8.68	9.48	9.45	10.41	10.61
6	12	730	730	8.84	9.92	9.56	12.1	10.49	16.08
6	13	791	791	8.92	8.86	9.64	9.55	10.58	10.43
6	14	852	852	9.0	8.86	9.72	9.55	10.65	10.64
6	15	913	913	9.07	8.96	9.79	9.67	10.72	10.69
6	16	974	974	9.13	9.23	9.85	10.06	10.78	11.24
6	17	1034	1034	9.19	9.38	9.91	10.21	10.84	11.11
6	18	1095	1095	9.25	9.12	9.97	9.88	10.9	10.79
6	19	1156	1156	9.3	9.2	10.02	10.0	10.95	10.89
6	20	1217	1217	9.35	9.33	10.07	10.03	11.01	11.04
6	21	1278	1278	9.4	9.28	10.12	10.04	11.05	11.07
6	22	1339	1339	9.45	9.45	10.17	10.23	11.1	11.23

TABLE IX. Critical values for Δt = 10 days, $f_{scan} = [1/365.25\,\text{day}^{-1},\ 1/10\,\text{day}^{-1}]$, $n_o = 10$

[1] E. Aprile et al. (XENON), Phys. Rev. Lett. **121**, 111302 (2018), 1805.12562.
[2] J. Aalbers et al. (LUX-ZEPLIN) (2022), 2207.03764.
[3] J. Aalbers et al. (2022), 2203.02309.
[4] J. Billard, L. Strigari, and E. Figueroa-Feliciano, Phys. Rev. D **89**, 023524 (2014), 1307.5458.
[5] B. Dutta and L. E. Strigari, Ann. Rev. Nucl. Part. Sci. **69**, 137 (2019), 1901.08876.

[6] J. B. Dent, B. Dutta, J. L. Newstead, and L. E. Strigari, Phys. Rev. D **93**, 075018 (2016), 1602.05300.
[7] J. B. Dent, B. Dutta, J. L. Newstead, and L. E. Strigari, Phys. Rev. D **95**, 051701 (2017), 1607.01468.
[8] C. A. J. O'Hare, A. M. Green, J. Billard, E. Figueroa-Feliciano, and L. E. Strigari, Phys. Rev. D **92**, 063518 (2015), 1505.08061.
[9] J. H. Davis, JCAP **03**, 012 (2015), 1412.1475.
[10] S. Sassi, A. Dinmohammadi, M. Heikinheimo, N. Mirabolfathi, K. Nordlund, H. Safari, and K. Tuominen, Phys. Rev. D **104**, 063037 (2021), 2103.08511.
[11] D. G. Cerdeño, M. Fairbairn, T. Jubb, P. A. N. Machado, A. C. Vincent, and C. Bœhm, JHEP **05**, 118 (2016), [Erratum: JHEP 09, 048 (2016)], 1604.01025.
[12] D. Aristizabal Sierra, B. Dutta, S. Liao, and L. E. Strigari, JHEP **12**, 124 (2019), 1910.12437.
[13] D. Aristizabal Sierra, V. De Romeri, L. J. Flores, and D. K. Papoulias, JCAP **01**, 055 (2022), 2109.03247.
[14] Y. Zhuang, L. E. Strigari, and R. F. Lang, Phys. Rev. D **105**, 043001 (2022), 2110.14723.
[15] C. A. J. O'Hare, Phys. Rev. Lett. **127**, 251802 (2021), 2109.03116.
[16] S. Appel et al. (Borexino) (2022), 2204.07029.
[17] G. Ranucci, Phys. Rev. D **73**, 103003 (2006), hep-ph/0511026.
[18] K. Abe et al. (Super-Kamiokande), Phys. Rev. D **94**, 052010 (2016), 1606.07538.
[19] Y. Zhuang, L. E. Strigari, L. Jin, and S. Sinha (2023), 2307.13792.
[20] W. C. Haxton, R. G. Hamish Robertson, and A. M. Serenelli, Ann. Rev. Astron. Astrophys. **51**, 21 (2013), 1208.5723.
[21] J. L. Newstead, L. E. Strigari, and R. F. Lang, Phys. Rev. D **99**, 043006 (2019), 1807.07169.
[22] A. M. Suliga, J. F. Beacom, and I. Tamborra, Phys. Rev. D **105**, 043008 (2022), 2112.09168.
[23] D. Franco et al., JCAP **08**, 017 (2016), 1510.04196.
[24] M. Szydagis, N. Barry, K. Kazkaz, J. Mock, D. Stolp, M. Sweany, M. Tripathi, S. Uvarov, N. Walsh, and M. Woods, arXiv e-prints arXiv:1106.1613 (2011), 1106.1613.
[25] J. L. Newstead, R. F. Lang, and L. E. Strigari, Phys. Rev. D **104**, 115022 (2021), 2002.08566.
[26] E. Aprile et al. (XENON), Phys. Rev. D **102**, 072004 (2020), 2006.09721.
[27] P. Agnes, L. Agostino, I. F. M. Albuquerque, T. Alexander, A. K. Alton, K. Arisaka, H. O. Back, B. Baldin, K. Biery, G. Bonfini, et al., Phys. Rev. D **93**, 081101 (2016), 1510.00702.
[28] P. Sturrock, G. Walther, and S. Wheatland, The Astrophysical Journal **491**, 409 (2009).
[29] B. Aharmim, S. N. Ahmed, A. E. Anthony, E. W. Beier, A. Bellerive, M. Bergevin, S. D. Biller, M. G. Boulay, M. G. Bowler, Y. D. Chan, et al., Phys. Rev. D **72**, 052010 (2005), hep-ex/0507079.
[30] G. Ranucci and M. Rovere, Phys. Rev. D **75**, 013010 (2007), hep-ph/0605212.
[31] G. Bellini et al. (Borexino), Phys. Lett. B **707**, 22 (2012), 1104.2150.
[32] M. Agostini, K. Altenmüller, S. Appel, V. Atroshchenko, D. Basilico, G. Bellini, J. Benziger, D. Bick, G. Bonfini, L. Borodikhina, et al., Astroparticle Physics **92**, 21 (2017), 1701.07970.
[33] J. Yoo, Y. Ashie, S. Fukuda, Y. Fukuda, K. Ishihara, Y. Itow, Y. Koshio, A. Minamino, M. Miura, S. Moriyama, et al., Phys. Rev. D **68**, 092002 (2003), hep-ex/0307070.
[34] P. A. Sturrock, Astrophys. J. **594**, 1102 (2003), hep-ph/0304073.
[35] P. A. Sturrock, D. O. Caldwell, J. D. Scargle, and M. S. Wheatland, arXiv e-prints hep-ph/0501205 (2005), hep-ph/0501205.
[36] J. Hosaka, K. Ishihara, J. Kameda, Y. Koshio, A. Minamino, C. Mitsuda, M. Miura, S. Moriyama, M. Nakahata, T. Namba, et al., Phys. Rev. D **73**, 112001 (2006), hep-ex/0508053.
[37] K. Abe, C. Bronner, Y. Hayato, K. Hiraide, K. Hosokawa, K. Ieki, M. Ikeda, J. Kameda, Y. Kanemura, R. Kaneshima, et al., arXiv e-prints arXiv:2311.01159 (2023), 2311.01159.
[38] P. Heix, S. Tilav, C. Wiebusch, and M. Zöcklein, arXiv e-prints arXiv:1909.02036 (2019), 1909.02036.
[39] Borexino Collaboration, G. Bellini, J. Benziger, D. Bick, G. Bonfini, D. Bravo, M. B. Avanzini, B. Caccianiga, L. Cadonati, F. Calaprice, et al., arXiv e-prints arXiv:1308.0443 (2013), 1308.0443.
[40] J. D. Scargle, Astrophys. J. **263**, 835 (1982).
[41] J. T. VanderPlas, The Astrophysical Journal Supplement Series **236**, 16 (2018).
[42] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., Nature Methods **17**, 261 (2020).
[43] A. M. Price-Whelan, B. Sipőcz, H. Günther, P. Lim, S. Crawford, S. Conseil, D. Shupe, M. Craig, N. Dencheva, A. Ginsburg, et al., The Astronomical Journal **156**, 123 (2018).
[44] N. Mishra and L. E. Strigari, Phys. Rev. D **108**, 063023 (2023), 2305.17827.

arXiv:2402.18410v1 [hep-ph] 28 Feb 2024

Imprints of new physics operators in the semileptonic $B \to a_1(1260)\ell^-\bar{\nu}_\ell$ process in SMEFT approach

Manas Kumar Mohapatra[1],* Dhiren Panda[1],† and Rukmani Mohanta[1]‡

[1] *School of Physics, University of Hyderabad, Hyderabad - 500046, India*

Abstract

At present, there are several measurements of B decays that exhibit discrepancies with the predictions of the Standard Model, and suggest the presence of new physics in $b \to s$ and $b \to c(u)$ quark level transitions. Motivated by the prospects of the ongoing high-luminosity B factories, we study the exclusive $B \to a_1(1260)\ell^-\bar{\nu}_\ell$ process within the Standard Model Effective Field Theory (SMEFT) formalism, to understand the sensitivity of new physics. The new physics parameters are constrained by using the experimental branching fractions of the (semi)leptonic $B \to \ell\bar{\nu}$ and $B \to (\pi, \rho, \omega)\ell\bar{\nu}$ processes (where $\ell = e, \mu, \tau$) which undergo $b \to u\ell\bar{\nu}$ quark level transitions. We then perform a comprehensive angular analysis of the exclusive $B \to a_1(1260)\ell^-\bar{\nu}_\ell$ process in the Standard Model and in the presence of various new physics operators. We also provide the predictions and comment on various observables, such as branching ratio, forward-backward asymmetry, and the test of lepton flavor non universality of the $B \to a_1(1260)\ell^-\bar{\nu}_\ell$ channel.

*Electronic address: manasmohapatra12@gmail.com
†Electronic address: pandadhiren530@gmail.com
‡Electronic address: rmsp@uohyd.ac.in

I. INTRODUCTION

The search for physics beyond the Standard model (SM) in B meson decays has not only attracted considerable attention in recent times but also anticipated to remain as an active area of research for the coming years. With thorough and careful investigation of various B decays, it might be possible to get the smoking gun signals of new physics (NP) from B sector. The LHC at CERN, particularly LHCb experiment plays a pivotal role complementary to the Belle II experiment at KEK, and the results from these two experiments impart the physicists to have a clear understanding of the nature of b quark decays. In this regard, the semileptonic B meson decays mediated by flavor changing neutral current ($b \to s$) and charged current ($b \to c/u$) transitions play a crucial role in probing the sensitivity of NP beyond the SM. Though, several anomalies are observed in these decay modes, so far none of the measurements are statistically significant enough to provide an unambiguous signal of NP. The future up-gradation of LHC with large data samples and improved precision can reduce the systematic errors in the existing measurements. The most interesting observables in $b \to s\ell\ell$ transitions, which were in the limelight for quite sometime for providing the unequivocal hints of NP, are the lepton flavor universality (LFU) violating observables R_K and R_{K^*}, defined as

$$R_{K^{(*)}} = \frac{\mathcal{B}(B \to K^{(*)}\mu\mu)}{\mathcal{B}(B \to K^{(*)}ee)} \,. \tag{1}$$

Recently, the updated results from LHCb [1, 2] confirmed that the measured values of these observables are consistent with their SM predictions, which are of order unity. However, there exists a variety of other observables in $b \to s\ell\ell$ transition, e.g., the celebrated P_5' observable, branching fractions of several decay modes, which manifest a few sigma deviations with their SM results. The LHCb [3, 4] and ATLAS [5] collaborations display a 3.3σ deviation from the SM prediction in the measurement of P_5'. The branching ratio of $B_s \to \phi\,\mu^-\mu^+$ decay mode indicates at 3.3σ [6, 7] in $q^2 \in [1.1, 6.0]$ GeV^2. Moreover, the measurement of the $R_{K_S^0}$ and $R_{K^{*+}}$ [8], also deviate from their SM predictions at 1.4σ and 1.5σ, respectively. Therefore, the present scenario does not completely rule out the presence of NP in the FCNC mediated transitions $b \to s\ell\ell$.

Analogously, for the charged current interaction processes mediated through $b \to c\ell\nu$ transition, the lepton flavour universality violation observable is defined as $R_D = \mathcal{B}(B \to D\tau\nu)/\mathcal{B}(B \to D(e,\mu)\nu)$, which has a very precise SM prediction [9–12], calculated using the form factors by lattice QCD approach. In Mariond 2019, the measurement of R_D, announced by Belle Collaboration [13], was consistent with the previous measurement. However, its world average value determined by HFLAV group [14] still deviates at the level of 1.4σ from the SM prediction. Additionally, the average of various measurements of $R_{D^*} = \mathcal{B}(B \to D^*\tau\nu)/\mathcal{B}(B \to D^*(e,\mu)\nu)$ from BaBar, Belle and LHCb

experimental measurements shows a tension at the level of 2.5σ [14]. On the other hand, the measured value of the τ polarization fraction $P_\tau^{D^*}$ and the longitudinal polarization fraction of D^* meson [15–17] reported by Belle collaboration make a difference from their SM predictions at 1.6σ and 1.5σ level, respectively. Additionally, the LHCb collaboration in 2017 [18] measured the LFU violating observable $R_{J/\psi} = \mathcal{B}(B \to J/\psi\tau\nu)/\mathcal{B}(B \to J/\psi(e,\mu)\nu)$ which lies above the SM prediction at the level of 1.8σ [19, 20]. With these intriguing set of results in the $b \to c\ell\nu$ quark level processes, it is worth exploring other similar deviations to understand the sensitivity of physics beyond the SM in charged current transitions, particularly those undergoing the quark level transition pertaining to $b \to u\ell\nu$ channels. Several mild discrepancies between the SM predictions and the experimental measurements have been observed in various $b \to u\ell\nu$ mediated transitions. A few processes are measured at the B-factories, although these $b \to u$ quark level transitions are CKM suppressed as compared to $b \to c$ transitions. The measured branching fraction of the leptonic $B \to \tau\nu$ process by Belle and BaBar [21–24], is not in good agreement with its SM value [25, 26]. An upper bound on the branching fraction of $B \to \pi\tau\nu$ has been reported to be 2.5×10^{-4} by Belle collaboration [27]. Additionally, the branching ratios of the exclusive $B \to \ell\nu$ and $B \to (\pi, \rho, \omega)\ell\nu$ ($\ell = \mu, e$) decays still show mild deviations from their SM results. Inspired by these set of differences between the measured values and the SM expectations, we exploit $B \to a_1\ell\nu$ mode (where a_1 is an axial vector meson) in this work. The observation of the charmless hadronic $B^0 \to a_1(1260)\pi$ decay channel by BaBar and Belle collaborations [28–31] helps us to perform a detailed theoretical study of the exclusive semileptonic $B \to a_1\ell\nu$ decay process. The study of the fully differential angular distribution of this channel is quite interesting as the meson a_1 decays into $\rho\pi$, which provides important source of information in the SM as well as in its possible extensions. Additionally, the ρ meson in the $a_1 \to \rho\pi$ process involves both the longitudinal ($\rho_{||}$) and transverse polarization ($\rho_\perp$), which also motivates the experimental search for this mode. Many analyses have been performed to account for the $b \to u\ell\nu$ transitions, see, e.g., Refs. [32–39]. Recently, the $B \to A\ell\nu$ (where A is an axial vector meson) channels have been investigated in Ref. [40–42]. A huge data samples, $(772 \pm 11) \times 10^6$ $B\bar{B}$ [43] and $(6.53 \pm 0.66) \times 10^6$ $B_s\bar{B}_s$ [44] pairs at $\Upsilon(4S)$ and $\Upsilon(5S)$ resonances, respectively are collected at Belle experiment. This statistics can be increased by collecting more data by Belle II experiment. Therefore, in principle, the $B \to a_1\ell\nu$ decay mode can be easily accessible in B factory experiments in near future.

In this work, our aim is to explore the consequences of a model independent effective theory formalism, the so called the Standard Model Effective Field Theory (SMEFT) approach on the exclusive semileptonic $B \to a_1\ell\nu$ decay mode. The BSM (beyond the Standard model) physics effects can be constructed from the various SMEFT operators and can be analysed by performing fit to the associated new physics couplings. Our fit anatomy includes the experimental measurements of the branching fractions of the

leptonic $B \to \ell\bar{\nu}$ and the semileptonic $B \to (\pi, \rho, \omega)\ell\bar{\nu}$ processes (where $\ell = e, \mu, \tau$). We mainly study the angular coefficient structure in the SM as well as in the presence of SMEFT NP operators. In addition, we also provide the predictions, and comment on the branching fraction and the lepton flavour universality violation observable defined by $\mathcal{R}_{a_1} = \mathcal{B}(B \to a_1\tau\nu)/\mathcal{B}(B \to a_1(e,\mu)\nu)$.

The remainder of the paper is structured as follows. In section II, we recapitulate the SMEFT Lagrangian with a list of most relevant dimension six operators for $b \to u\ell\nu$ transitions. In section III, we discuss the fully differential decay distributions of $B \to a_1(\to \rho_{||}\pi)$ and $B \to a_1(\to \rho_{\perp}\pi)$, by exploiting different set of angular coefficient functions in terms of new scalar, pseudoscalar, vector and tensor operators weighted by real couplings. A detailed analysis of angular coefficient functions are studied in presence of various NP couplings in section IV. We also focus on the branching ratio and the LFU violating observable in this section. Finally, we conclude our work in section V.

II. THEORETICAL FRAMEWORK

In the absence of direct evidence of new particles close to electroweak scale at the Large Hadron Collider, emphasis has been given to look for indirect evidence of the existence of new particles at scale $\Lambda_{\rm NP}$, that surpasses the electroweak scale. The SMEFT approach, one of the most efficient platforms, allows to investigate the possible NP hints in the b-hadron decays [45–47]. In this framework, the new physics effects at the scale $\Lambda_{\rm NP}$ comprises a set of higher dimensional operators that are suppressed with the NP energy scale. The form of these operators are built with the SM fields and respects the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ [48, 49]. In this work, we focus on the $b \to u\ell\nu$ processes within this formalism by considering the dimension-6 operators. However, it would be very important to analyze the NP effects indirectly on the SM low energy processes. To understand the model-independent effects of heavy NP scale, the SMEFT effective Lagrangian at mass dimension-6 are expressed as [50]

$$\mathcal{L}_{\rm eff} = \mathcal{L}_{\rm SM} + \sum_{Q_i = Q_i^\dagger} \frac{\tilde{C}_i}{\Lambda^2} Q_i + \sum_{Q_i \neq Q_i^\dagger} \left(\frac{\tilde{C}_i}{\Lambda^2} Q_i + \frac{\tilde{C}_i^*}{\Lambda^2} Q_i^\dagger \right) . \tag{2}$$

The relevant SMEFT dimension-six operators contributing to $b \to u\ell\nu$ processes Q_i, obtained by integrating out the heavy NP particles, are given as follow

$$\begin{aligned}
&Q_{lq}^{(3)} = (\bar{\ell}_p \gamma_\mu \sigma_a \ell_r)(\bar{q}_s \gamma^\mu \sigma^a q_t), \quad && Q_{ledq} = (\bar{\ell}_p^j e_r)(\bar{d}_s q_{tj}), && Q_{lequ}^{(1)} = (\bar{\ell}_p^j e_r)\varepsilon_{jk}(\bar{q}_s^k u_t), \\
&Q_{lequ}^{(3)} = (\bar{\ell}_p^j \sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar{q}_s^k \sigma^{\mu\nu} u_t), && Q_{\phi q}^{(3)} = (\phi^\dagger i \overleftrightarrow{D}^a_\mu \phi)(\bar{q}_p \sigma_a \gamma^\mu q_r), && Q_{\phi ud} = (\tilde{\phi}^\dagger i D_\mu \phi)(\bar{u}_p \gamma^\mu d_r), \\
&Q_{\ell q}^{(1)} = (\bar{\ell}_p \gamma_\mu \ell_r)(\bar{q}_s \gamma^\mu q_t), && Q_{\phi q}^{(1)} = (\phi^\dagger i \overleftrightarrow{D}_\mu \phi)(\bar{q}_p \gamma^\mu q_r), &&
\end{aligned} \tag{3}$$

where $\overleftrightarrow{D}_\mu = D_\mu - \overleftarrow{D}_\mu$, $\overleftrightarrow{D}^a_\mu = \sigma^a D_\mu - \overleftarrow{D}_\mu \sigma^a$, with σ^a as the Pauli matrices and ϵ_{jk} are the totally antisymmetric tensor in $SU(2)_L$ space. The quark "q" and the lepton field "ℓ" shown in the above equation are doublets under $SU(2)_L$ group whereas the fermions u, d and e correspond to the right handed singlet fields. Out of eight relevant operators, the $(Q^{(3)}_{\phi q}, Q_{\phi ud})$ and $Q^{(3)}_{\phi \ell}$ are responsible for the modification of left and right handed W boson couplings with quarks and leptons, respectively. In this analysis, we focus only on the operators $Q^{(3)}_{lq}$, Q_{ledq}, $Q^{(3)}_{lequ}$ and $Q^{(3)}_{\phi q}$ contributing to $b \to u\ell\nu$ transitions [51, 52].

At low energy, the most general weak effective Hamiltonian governing to $b \to u\ell\nu$ transitions is given as [52–56]

$$\mathcal{H}_{\rm eff} = \mathcal{H}^{\rm SM}_{\rm eff} + \frac{4G_F}{\sqrt{2}} V_{ub} \sum_{i,\ell} C^{(\ell)}_i O^{(\ell)}_i + \text{h.c.}\,, \tag{4}$$

where $O^{(\ell)}_i$ and $C^{(\ell)}_i$ are the local effective operators and the Wilson coefficients encoding the NP contributions respectively. The effective mass dimension-six operators are given as

$$\begin{aligned}
O^{(\ell)}_{V_L} &= (\bar{u}_L \gamma^\mu b_L)(\bar{\ell}_L \gamma_\mu \nu_{\ell L}), & O^{(\ell)}_{V_R} &= (\bar{u}_R \gamma^\mu b_R)(\bar{\ell}_L \gamma_\mu \nu_{\ell L}),\\
O^{(\ell)}_{S_L} &= (\bar{u}_R b_L)(\bar{\ell}_R \nu_{\ell L}), & O^{(\ell)}_{S_R} &= (\bar{u}_L b_R)(\bar{\ell}_R \nu_{\ell L}),\\
O^{(\ell)}_{T} &= (\bar{u}_R \sigma^{\mu\nu} b_L)(\bar{\ell}_R \sigma_{\mu\nu} \nu_{\ell L}),
\end{aligned} \tag{5}$$

where the $O^{(\ell)}_{V_{L,R}}$, $O^{(\ell)}_{S_{L,R}}$ and $O^{(\ell)}_T$ are respectively known are vector, scalar and tensor operators. Now, in terms of the dim-6 SMEFT operators, the above Wilson coefficients can be expressed as follows [50, 56]

$$C^{(\ell)}_{V_L} = -\frac{V_{ud}}{V_{ub}} \frac{v^2}{\Lambda^2} \left[\tilde{C}^{(3)}_{\ell q}\right]_{\ell\ell 13}, \qquad C^{(\ell)}_{V_R} = \frac{1}{2V_{ub}} \frac{v^2}{\Lambda^2} \left[\tilde{C}_{\phi ud}\right]_{13}, \tag{6}$$

$$C^{(\ell)}_{S_L} = -\frac{1}{2V_{ub}} \frac{v^2}{\Lambda^2} \left[\tilde{C}^{(1)}_{\ell equ}\right]^*_{\ell\ell 31}, \qquad C^{(\ell)}_{S_R} = -\frac{V_{ud}}{2V_{ub}} \frac{v^2}{\Lambda^2} \left[C_{\ell edq}\right]^*_{\ell\ell 31}, \tag{7}$$

$$C^{(\ell)}_{T} = -\frac{1}{2V_{ub}} \frac{v^2}{\Lambda^2} \left[\tilde{C}^{(3)}_{\ell equ}\right]^*_{\ell\ell 31}. \tag{8}$$

Here, Λ is the NP cut-off scale expected to be at TeV scale and the VEV of SM Higgs field ϕ is taken as $v = 246$ GeV. The above NP contributions are obtained by matching the operators given in Eq (3) to the weak effective Hamiltonian in Eq. (4) by assuming the down quark mass basis to be diagonal. Now, we explore how these NP couplings affect the $b \to u\ell\nu$ decay modes explicitly in the SMEFT formalism. Here we discuss the exclusive $B \to \ell\nu$, and $B \to (\pi, \rho, \omega)\ell\nu$ decay modes and investigate how the sensitivity of NP couplings arising from these processes impact the $B \to a_1(\to \rho\pi)\ell\nu$ channel. Now, from the most general effective Hamiltonian governing to $b \to u\ell\nu$ transition [53, 54], one can obtain the branching ratio of the pure leptonic $B \to \ell\nu$ decays. This braching fraction is sensitive to axial and pseudoscalar coefficients and is given as [50, 57]

$$\frac{\mathcal{B}(B \to \ell\nu)}{\mathcal{B}(B \to \ell\nu)_{\rm SM}} = \left| 1 - C_A + \frac{m^2_B}{m_l (m_b + m_u)} C_P \right|^2. \tag{9}$$

where $C_A = C_{V_R}^{(\ell)} - C_{V_L}^{(\ell)}$ and $C_P = C_{S_R}^{(\ell)} - C_{S_L}^{(\ell)}$. Similarly, the differential decay width for the semileptonic $B \to P\ell\nu$ decay mode, sensitive to vector, scalar and tensor Wilson coefficients, is given by [58, 59]

$$\begin{aligned}\frac{d\Gamma(B\to P\ell\nu)/dq^2}{d\Gamma(B\to P\ell\nu)^{\rm SM}/dq^2} &= \left|1+C_V^{(\ell)}\right|^2\left[\left(1+\frac{m_l^2}{2q^2}\right)H_{V,0}^{s\,2}+\frac{3}{2}\frac{m_\ell^2}{q^2}\,H_{V,t}^{s\,2}\right]\\ &+\frac{3}{2}|C_S^{(\ell)}|^2\,H_S^{s\,2}+8|C_T^{\ell}|^2\left(1+\frac{2m_\ell^2}{q^2}\right)H_T^{s\,2}\\ &+3{\rm Re}[(1+C_V^{(\ell)})(C_S^{(\ell)^*})]\frac{m_\ell}{\sqrt{q^2}}\,H_S^s H_{V,t}^s\\ &-12{\rm Re}[(1+C_V^{(\ell)})C_T^{(\ell)*}]\frac{m_\ell}{\sqrt{q^2}}\,H_T^s H_{V,0}^s\,,\end{aligned} \tag{10}$$

where $C_V^{(\ell)} \equiv C_{V_L}^{(\ell)} + C_{V_R}^{(\ell)}$ and the scalar coefficient $C_S^{(\ell)} \equiv C_{S_L}^{(\ell)} + C_{S_R}^{(\ell)}$. The hadronic matrix elements $H_{(V,0),(V,t),S,T}^{s,2}$ are parameterized by the form factors $f_+(q^2)$, $f_0(q^2)$ and $f_T(q^2)$. Now for the $B \to V\ell\nu$ (where V denotes the vector meson) process, the differential decay width relative to SM is given as [58, 59]

$$\begin{aligned}\frac{d\Gamma(B\to V\ell\nu)/dq^2}{d\Gamma(B\to V\ell\nu)^{\rm SM}/dq^2} &= \left(|1+C_{V_L}^{(\ell)}|^2+|C_{V_R}^{(\ell)}|^2\right)\left[\left(1+\frac{m_l^2}{2q^2}\right)\left(H_{V,+}^2+H_{V,-}^2+H_{V,0}^2\right)+\frac{3}{2}\frac{m_\ell^2}{q^2}\,H_{V,t}^2\right]\\ &-2{\rm Re}[(1+C_{V_L}^{(\ell)})C_{V_R}^{(\ell)*}]\left[\left(1+\frac{m_\ell^2}{2q^2}\right)\left(H_{V,0}^2+2H_{V,+}H_{V,-}\right)+\frac{3}{2}\frac{m_\ell^2}{q^2}\,H_{V,t}^2\right]\\ &+\frac{3}{2}|C_{S_R}^{(\ell)}-C_{S_L}^{(\ell)}|^2\,H_S^2+8|C_T^{\ell}|^2\left(1+\frac{2m_\ell^2}{q^2}\right)\left(H_{T,+}^2+H_{T,-}^2+H_{T,0}^2\right)\\ &+3{\rm Re}[(1-(C_{V_R}^{(\ell)}-C_{V_L}^{(\ell)}))(C_{S_R}^{(l)*}-C_{S_L}^{(\ell)*})]\frac{m_\ell}{\sqrt{q^2}}\,H_S H_{V,t}\\ &-12{\rm Re}[(1+C_{V_L}^{(\ell)})C_T^{(\ell)*}]\frac{m_\ell}{\sqrt{q^2}}\left(H_{T,0}H_{V,0}+H_{T,+}H_{V,+}-H_{T,-}H_{V,-}\right)\\ &+12{\rm Re}[C_{V_R}^{(\ell)}C_T^{(\ell)*}]\frac{m_\ell}{\sqrt{q^2}}\left(H_{T,0}H_{V,0}+H_{T,+}H_{V,-}-H_{T,-}H_{V,+}\right)\,,\end{aligned} \tag{11}$$

where the details of the hadronic matrix elements in terms of the form factors $A_{0,1,2}(q^2)$, $V(q^2)$ and $T_{1,2,3}(q^2)$ can be found in Ref. [59].

III. SENSITIVITY OF NEW PHYSICS

A. Constraints on the SMEFT coefficients

In this subsection, we discuss the constraint analysis for the Wilson coefficients of SMEFT approach. We consider the branching fractions of the $B \to \ell\nu$, $B \to \pi\ell\nu$ and $B \to V\ell\nu$ processes where the vector meson $V = (\omega, \rho)$, and the lepton $\ell = (e, \mu, \tau)$. For numerical

computation, we use various input parameters such as mass of quarks, leptons, mesons, CKM matrix element, Fermi coupling constant G_F, and the lifetime of B meson etc as listed in Table I. For hadronic inputs such as the form factors of $B \to \pi\ell\nu$ transition, we consider

Parameter	Value	Parameter	Value (GeV)	Parameter	Value (GeV)
G_F	1.166×10^{-5} GeV^{-2}	m_u	0.003	m_B	5.279
$(\tau_{B^\pm}, \tau_{B^0})$	$(1.638, 1.519) \times 10^{-12}$ s	m_c	1.270	m_π	0.139
(V_{ub}, V_{ud})	(0.0038, 0.976)	m_b	4.18	m_ρ	0.775
f_B	0.1905	m_e	0.005	m_ω	0.782
$\mathcal{B}_{a_1\to\rho\pi}$	0.6019	m_τ	1.776	m_{a_1}	1.230

TABLE I: Values of the input parameters [60] used in our analysis.

the BCL parametrization from Ref. [61]

$$f_+(q^2) = \frac{1}{(1-q^2/m_{B^*}^2)} \sum_{n=0}^{N-1} a_n^+ \Big[z^n - (-1)^{n-N} \frac{n}{N}\, z^N \Big] , \qquad f_0(q^2) = \sum_{n=0}^{N-1} a_n^0 z^n . \tag{12}$$

Here the mass of B^* meson is considered as 5.325 GeV, and $b_n^{+,0}$ are the expansion coefficients. The expansion parameter z in the above equation is defined as

$$z \equiv z(q^2) = \frac{\sqrt{t_+ - q^2} - \sqrt{t_+ - t_0}}{\sqrt{t_+ - q^2} + \sqrt{t_+ - t_0}} , \tag{13}$$

where $t_+ = (m_B + m_\pi)^2$ and $t_0 = (m_B + m_\pi)(\sqrt{m_B} - \sqrt{m_\pi})^2$. Also the expansion coefficients are extracted from the combined fit of the data obtained from the $B \to \pi\ell\bar{\nu}_\ell$ q^2 distribution and the lattice results [62, 63] which are given as below

$$\begin{aligned} &a_0^+ = 0.419 \pm 0.013, \quad a_1^+ = -0.495 \pm 0.054, \quad a_2^+ = -0.43 \pm 0.13, \quad a_3^+ = 0.22 \pm 0.31, \\ &a_0^0 = 0.510 \pm 0.019, \quad a_1^0 = -1.700 \pm 0.082, \quad a_2^0 = -1.53 \pm 0.19, \quad a_3^0 = 4.52 \pm 0.83. \end{aligned} \tag{14}$$

The form factor $f_T(q^2)$ is related to $f_+^{B\to\pi}(q^2)$ through the the following relation [42],

$$f_+^{B\to\pi}(q^2) = \frac{m_B}{m_B + m_\pi} f_T^{B\to\pi}(q^2). \tag{15}$$

We use the LCSR method [64] to compute the form factor for $B \to \rho$ and $B \to \omega$ transitions. The formula is given by

$$F_i(q^2) = P_i(q^2) \sum_k \alpha_k^i \, \big[z(q^2) - z(0) \big]^k , \tag{16}$$

where $P_i(q^2) = 1/(1 - q^2/m_{R,i}^2)$. The details of the expansion coefficients α_k^i are provided in Ref. [64]. Now using the available data of these processes, we perform a naive χ^2 analysis to constraint the NP SMEFT coefficients. The relevant χ^2 is defined as

$$\chi^2(\tilde{C}^{\rm NP}) = \sum_i \frac{\Big(\mathcal{O}_i^{\rm Th}(\tilde{C}^{\rm NP}) - \mathcal{O}_i^{\rm Exp} \Big)^2}{(\Delta\mathcal{O}_i^{\rm Exp})^2 + (\Delta\mathcal{O}_i^{\rm SM})^2}, \tag{17}$$

where $\mathcal{O}_i^{\rm Th}$ and $\mathcal{O}_i^{\rm Exp}$ represent the theoretical values and the measured central value of the observables, respectively. The denominator represents the errors associated with the theoretical and experimental values. From this analysis, we obtained the SMEFT new physics couplings. For completeness, we discuss for both μ and τ modes, and shown the details of the fit values in Table - II.

SMEFT couplings	Best fit (μ mode)	Best fit (τ mode)
$[\tilde{C}^{(3)}_{\ell q}]_{\ell\ell 13}$	0.013	0.163
$[\tilde{C}^{(3)}_{\ell equ}]_{\ell\ell 31}$	-0.0008	-0.043
$[\tilde{C}^{(1)}_{\ell equ}]_{\ell\ell 31}$	-0.004	-0.080
$[\tilde{C}_{\ell edq}]_{\ell\ell 31}$	0.005	-0.015
$([\tilde{C}^{(3)}_{\ell q}]_{\ell\ell 13}, [\tilde{C}^{(1)}_{\ell equ}]_{\ell\ell 31})$	(0.016, 0.001)	(-0.015, -0.080)
$([\tilde{C}^{(3)}_{\ell q}]_{\ell\ell 13}, [\tilde{C}_{\ell edq}]_{\ell\ell 31})$	(0.015, 0.004)	(-0.050,0.100)
$([\tilde{C}^{(3)}_{\ell q}]_{\ell\ell 13}, [\tilde{C}^{(3)}_{\ell equ}]_{\ell\ell 31})$	(0.113, 0.003)	(-0.030, 0.150)
$([\tilde{C}^{(1)}_{\ell equ}]_{\ell\ell 31}, [\tilde{C}^{(3)}_{\ell equ}]_{\ell\ell 31})$	(-0.004, -0.001)	(0.060, -0.050)
$([\tilde{C}_{\ell edq}]_{\ell\ell 31}, [\tilde{C}^{(3)}_{\ell equ}]_{\ell\ell 31})$	(0.006, -0.001)	(0.113, -0.045)
$([\tilde{C}^{(1)}_{\ell equ}]_{\ell\ell 31}, [\tilde{C}_{\ell edq}]_{\ell\ell 31})$	(0.002, 0.0015)	(-0.050,-0.055)

TABLE II: Best fit values of SMEFT coefficients obtained from $B \to \ell\nu$ and $B \to (\pi, V)\ell\nu$ $(V = \rho, \omega)$ decay channels.

We also provide the constraint regions of all NP coefficients with the benchmark point as the χ^2 best-fit value in Fig. 1. Now, we discuss the comprehensive analysis of the angular coefficients of $B \to a_1\ell\nu$ process.

B. Differential Angular Analysis of $B \to a_1(1260)\ell\nu$ Process

It is quite interesting to visualize the $a_1(\rho\pi)$ channel in the exclusive $B \to a_1(1260)\ell\nu$ decay mode as the ρ meson is transversely and longitudinally polarized. The four dimensional

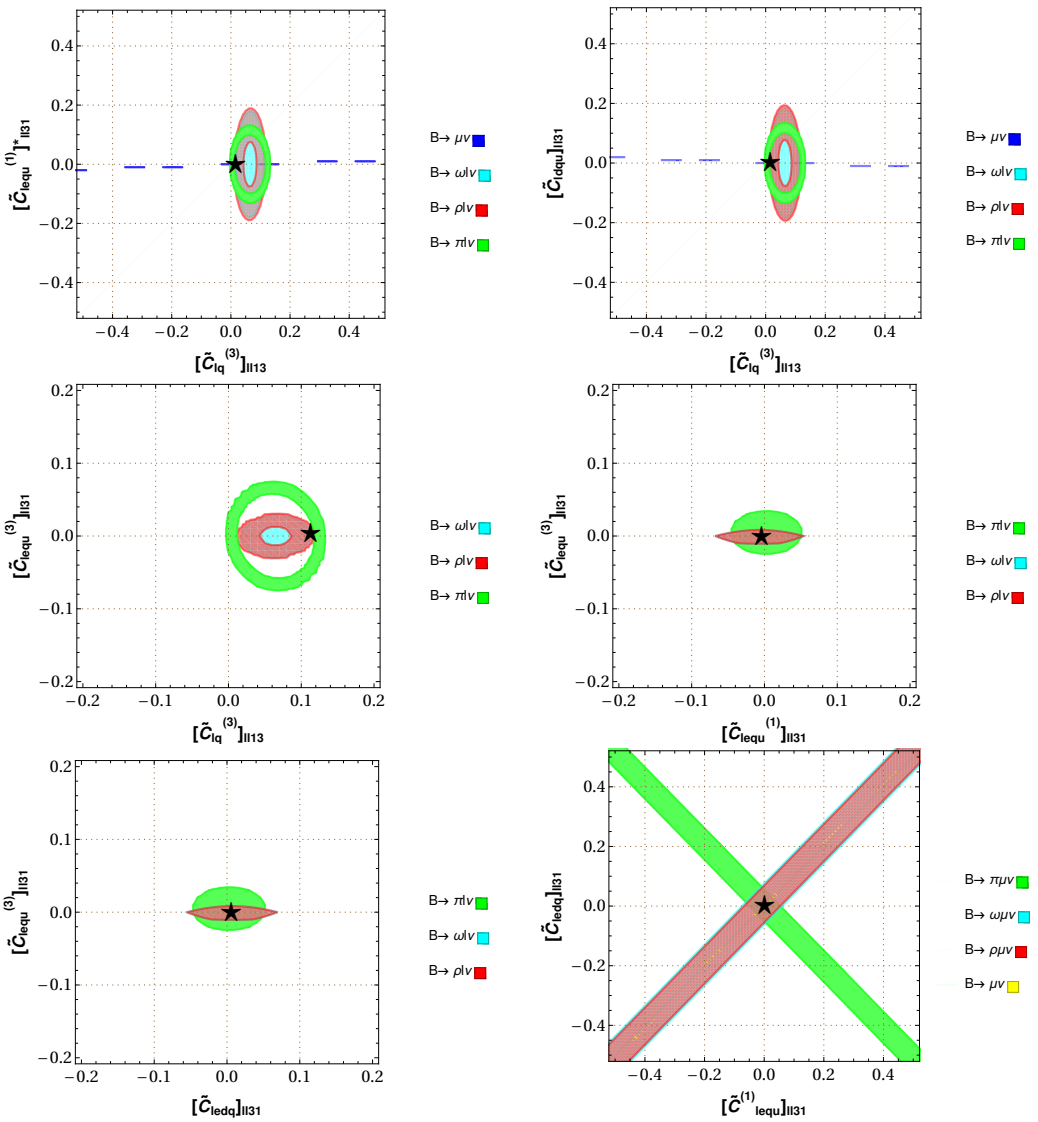


FIG. 1: Allowed regions for all the possible 2-D couplings from $b \to u\mu\nu$ available data. The colors distinguish various decay modes shown in the right side of each panel. The black star represents the corresponding best-fit value.

differential decay distribution amplitude is given as follows [41]

$$\begin{aligned}
\frac{d^4\Gamma(\bar{B} \to a_1(\to \rho_{\|(\perp)}\pi)\ell^-\bar{\nu}_\ell)}{dq^2\, d\cos\theta\, d\phi\, d\cos\theta_V} &= \mathcal{N}_{a_1}^{\|(\perp)}|\vec{p}_{a_1}|\left(1-\frac{m_\ell^2}{q^2}\right)^2\Big\{I_{1s,\|(\perp)}^{a_1}\sin^2\theta_V + I_{1c,\|(\perp)}^{a_1}(3+\cos 2\theta_V) \\
&+ \left(I_{2s,\|(\perp)}^{a_1}\sin^2\theta_V + I_{2c,\|(\perp)}^{a_1}(3+\cos 2\theta_V)\right)\cos 2\theta \\
&+ I_{3,\|(\perp)}^{a_1}\sin^2\theta_V\sin^2\theta\cos 2\phi + I_{4,\|(\perp)}^{a_1}\sin 2\theta_V\sin 2\theta\cos\phi \\
&+ I_{5,\|(\perp)}^{a_1}\sin 2\theta_V\sin\theta\cos\phi \\
&+ \left(I_{6s,\|(\perp)}^{a_1}\sin^2\theta_V + I_{6c,\|(\perp)}^{a_1}(3+\cos 2\theta_V)\right)\cos\theta \\
&+ I_{7,\|(\perp)}^{a_1}\sin 2\theta_V\sin\theta\sin\phi\Big\}\ .
\end{aligned} \tag{18}$$

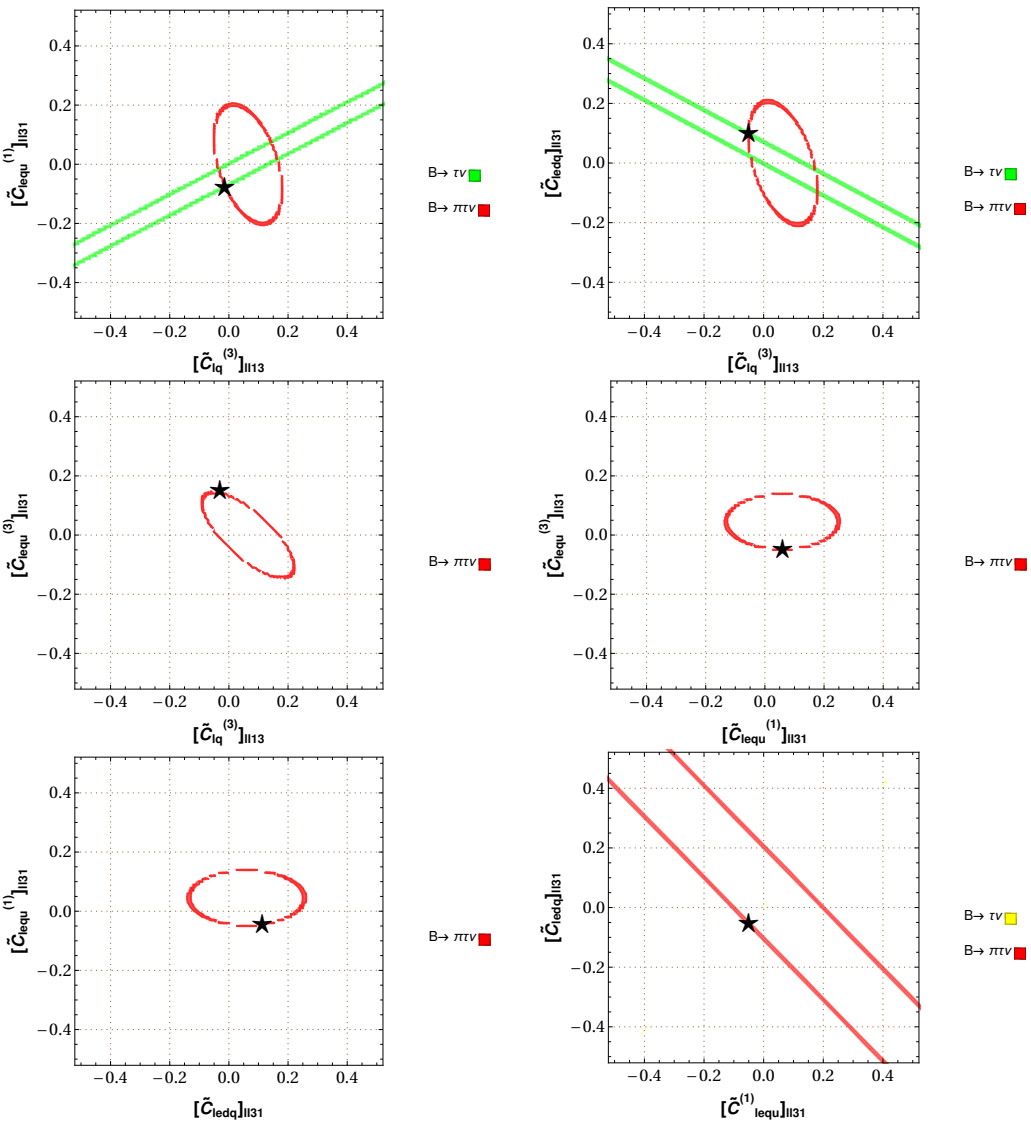


FIG. 2: Similar as Fig. 1 for $b \to u\tau\nu$ channels.

The symbol $\perp$ and $\parallel$ refer to the transverse and longitudinal polarizations of ρ meson. The expressions of the coefficients $\mathcal{N}_{a_1}^{\parallel(\perp)}$ and $|\vec{p}_{a_1}|$ can be read as

$$\mathcal{N}_{a_1}^{\parallel(\perp)} = \frac{3G_F^2|V_{ub}|^2\mathcal{B}(a_1 \to \rho_{\parallel(\perp)}\pi)}{128(2\pi)^4 m_B^2}, \qquad |\vec{p}_{a_1}| = \sqrt{\lambda(m_B^2, m_{a_1}^2, q^2)}/2m_B \,. \tag{19}$$

The angular coefficient functions I_i^{a1} in terms of the scalar, vector and tensor couplings are given as [41]

$$\begin{aligned} I_i &= |1+\epsilon_V|^2\, I_i^{SM} + |\epsilon_S|^2 I_i^{NP,S} + |\epsilon_T|^2 I_i^{NP,T} + 2\,\mathrm{Re}\left[\epsilon_S(1+\epsilon_V^*)\right] I_i^{INT,S} \\ &+ 2\,\mathrm{Re}\left[\epsilon_T(1+\epsilon_V^*)\right] I_i^{INT,T} + 2\,\mathrm{Re}\left[\epsilon_S\epsilon_T^*\right] I_i^{INT,ST}, \qquad (i=1,\ldots 6), \\ I_7 &= 2\,\mathrm{Im}\left[\epsilon_X(1+\epsilon_V^*)\right] I_7^{INT,X} + 2\,\mathrm{Im}\left[\epsilon_T(1+\epsilon_V^*)\right] I_7^{INT,T} + 2\,\mathrm{Im}\left[\epsilon_X\epsilon_T^*\right] I_7^{INT,XT}. \end{aligned} \tag{20}$$

Here the NP couplings ϵ_i $(i = V, S, T)$ are given as

$$\epsilon_V = C_{V_L}, \quad \epsilon_S = C_{S_R} + C_{S_L} \quad \epsilon_T = C_T\,. \tag{21}$$

The angular coefficient functions I_i^{SM} represent the SM contributions whereas the other coefficients such as $I_i^{NP,(S,T)}$ and $I_i^{INT,(S,T,ST)}$, expressed in terms of helicity amplitudes, are the terms corresponding to NP and interference of NP with SM contributions respectively. The detailed expressions are collected in Tables V–IX of Appendix B, together with the relations of the helicity amplitudes to the hadronic form factors. Here we examine the several angular coefficient functions discussed above using the SMEFT approach. Now, the several q^2 dependent observables such as branching fraction, leptonic forward-backward asymmetry in the $B \to a_1 \ell\nu$ decay mode can be obtained as [41]

$$\begin{aligned}
\mathcal{B}(q^2) &= \tau_B \frac{16}{9}\pi \frac{d\Gamma_{(\|,\perp)}}{dq^2}(q^2) = \tau_B \mathcal{N}_{a_1} |\vec{p}_{a_1}| \Big(1 - \frac{m_l^2}{q^2}\Big)^2 \Big(3I_{1s(\|,\perp)}^{a_1} + 12\, I_{1c(\|,\perp)}^{a_1} - I_{2s(\|,\perp)}^{a_1} - 4I_{2c(\|,\perp)}^{a_1}\Big)\,, \\
\mathcal{A}_{FB}^{l}(q^2) &= \left[\int_0^1 d\cos\,\theta\, \frac{d^2\Gamma}{dq^2 d\cos\,\theta} - \int_{-1}^0 d\cos\,\theta\, \frac{d^2\Gamma}{dq^2 d\cos\,\theta}\right] \Big/ \frac{d\Gamma}{dq^2} \\
&= \frac{8\pi}{3} \frac{\mathcal{N}_{a_1} |\vec{P}_{a_1}|\, \left(1 - \frac{m_l^2}{q^2}\right)^2 \Big(I_{6s} + 4I_{6c}\Big)}{d\Gamma/dq^2}
\end{aligned} \tag{22}$$

Alongside, we also discuss the lepton non universality (LNU) observable as,

$$\mathcal{R}_{a_1}(q^2) = \frac{d\Gamma(B \to a_1 \tau\nu)/dq^2}{d\Gamma(B \to a_1 (e,\mu)\nu)/dq^2}. \tag{23}$$

In order to study the $B \to a_1 \ell\nu$ process, we adopt the dipole paramerization for the form factors from Ref. [35], given as

$$F(q^2) = \frac{F(0)}{1 - a(q^2/m_B^2) + b(q^2/m_B^2)^2}\,. \tag{24}$$

The values of the form factors at $q^2 = 0$, i.e $F(0)$, and other relevant parameters a and b are given in Table III.

After obtaining the NP couplings from the constraint analysis, we discuss the angular coefficients of the differential 4-dimensional distribution amplitude as well as the observables of the $B \to a_1 \ell\nu$ decay mode in the SM and in the presence of new SMEFT coefficients. We discuss the transverse and longitudinal analysis of the given process as per the behavior of ρ meson.

IV. RESULTS AND DISCUSSIONS

After obtaining the values of NP effective coefficients by exploiting the $B \to \ell\bar{\nu}$ and $B \to (\pi, \rho, \omega)\ell\bar{\nu}$ processes, we investigate the q^2 distribution of the differential decay rate

F	$F(0)$	a	b
A^{Ba_1}	$0.26^{+0.06+0.00+0.03}_{-0.05-0.01-0.03}$	$1.72^{+0.05}_{-0.05}$	$0.66^{+0.07}_{-0.06}$
$V_0^{Ba_1}$	$0.34^{+0.07+0.01+0.08}_{-0.07-0.02-0.08}$	$1.73^{+0.05}_{-0.06}$	$0.66^{+0.06}_{-0.08}$
$V_1^{Ba_1}$	$0.43^{+0.10+0.01+0.05}_{-0.09-0.01-0.05}$	$0.75^{+0.05}_{-0.05}$	$-0.12^{+0.05}_{-0.02}$
$V_2^{Ba_1}$	$0.13^{+0.03+0.00+0.00}_{-0.03-0.01-0.00}$	$--$	$--$
$T_1^{Ba_1}$	$0.34^{+0.08+0.00+0.05}_{-0.07-0.01-0.05}$	$1.69^{+0.06}_{-0.05}$	$0.61^{+0.08}_{-0.05}$
$T_2^{Ba_1}$	$0.34^{+0.08+0.00+0.05}_{-0.07-0.01-0.05}$	$0.71^{+0.07}_{-0.05}$	$-0.16^{+0.03}_{-0.02}$
$T_3^{Ba_1}$	$0.30^{+0.07+0.05+0.05}_{-0.06-0.01-0.05}$	$1.60^{+0.06}_{-0.05}$	$0.53^{+0.06}_{-0.04}$

TABLE III: $B \to a_1$ form factor inputs [35].

for the $B \to a_1 \ell \nu$ decay mode along with the angular coefficient functions. In this formalism, the new physics analysis includes the scalar, vector and tensor operators, which affect the semileptonic B transitions. To understand the sensitivity of NP operators, we analyze the longitudinal and transverse contributions in SM as well as in the SMEFT approach. Below, we provide a comprehensive analysis of the longitudinal and transverse angular coefficients of $B \to a_1 \ell \nu$ channel.

A. Analysis of longitudinal angular coefficients of $B \to a_1 \ell \nu$ process

In this subsection, we discuss all the longitudinal angular coefficients $I_i^{(a_1 \to \rho_{||} \pi)}$ of the $B \to a_1 \ell \nu$ decay mode in the SM as well as in SMEFT approach. We categorize the NP analysis in two ways such as $(S+V)$ and $(S+V+T)$. The former includes both scalar and vector contributions whereas the later one bears the scalar, vector and tensor effective coefficients. In order to visualize the behaviour of the given decay mode with different leptonic final states, we probe for μ and τ modes separately, which are depicted in Fig. 3 and Fig. 4, respectively. The detailed analysis is presented below.

1. *μ mode behavior*

From the four-dimensional angular distribution of the decay width given in Eq. (18), the complete set of longitudinal angular coefficients in SM and in the presence of various NP induced SMEFT operators can be seen in Eq. (20). Now, we discuss the q^2 behavior of the angular coefficinets $I^{a_1}_{(1s,||)}(q^2)$, $I^{a_1}_{(2s,||)}(q^2)$, $I^{a_1}_{(2c,||)}(q^2)$, $I^{a_1}_{(3,||)}(q^2)$, $I^{a_1}_{(4,||)}(q^2)$, $I^{a_1}_{(5,||)}(q^2)$, $I^{a_1}_{(6s,||)}(q^2)$, $I^{a_1}_{(6c,||)}(q^2)$ and $I^{a_1}_{(1c,||)}(q^2)$ (in units of GeV^4), which are depicted below in Fig. 3. As our objective is to nurture the presence of real coefficients, the behavior of $I_7^{a_1}$ being imaginary is not considered in our analysis. The coefficient functions $I^{a_1}_{(1s,||)}$, $I^{a_1}_{(2s,||)}$, $I^{a_1}_{(2c,||)}$, $I^{a_1}_{(3,||)}$, $I^{a_1}_{(4,||)}$,

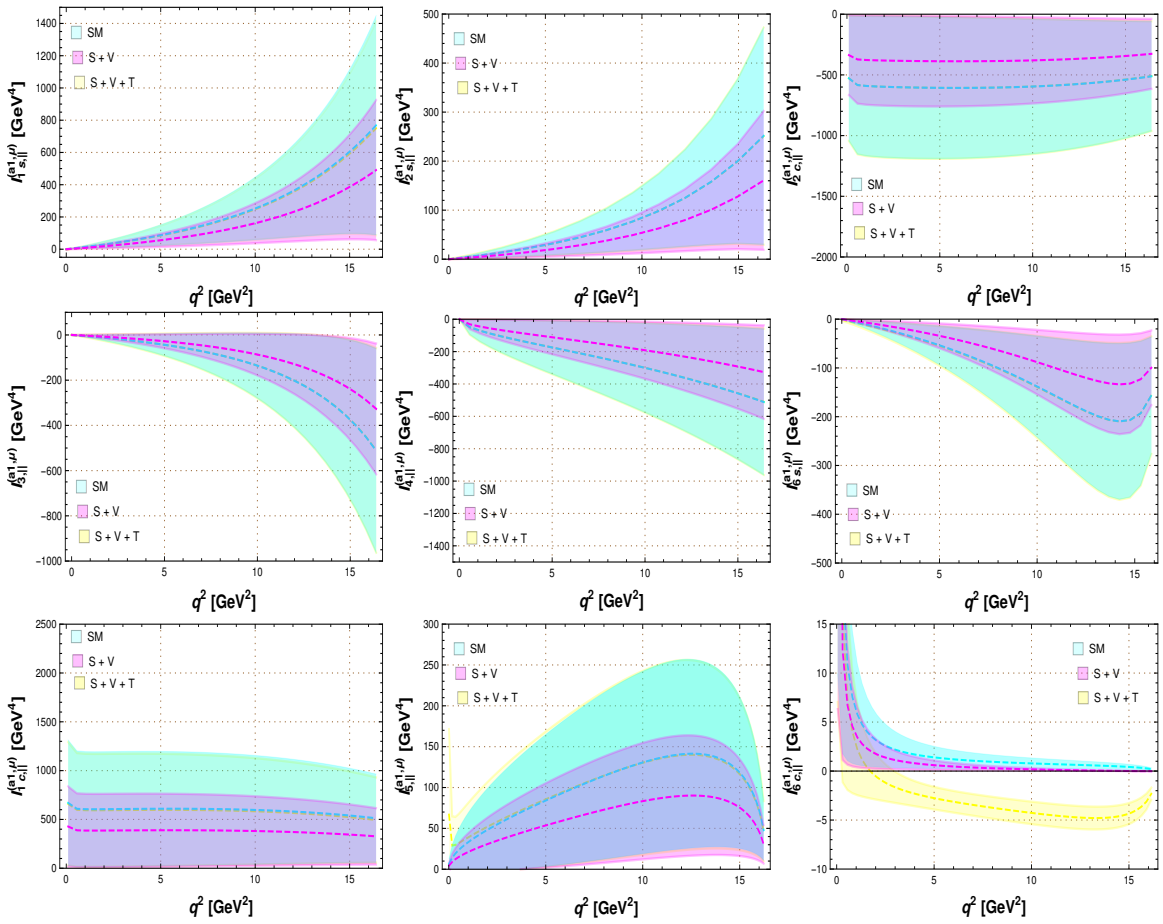


FIG. 3: The q^2 dependency of the angular coefficients of $B \to a_1(\rho_{||}\pi)\mu\bar{\nu}$ decay mode in SM (cyan), $S+V$ (magenta) and $S+V+T$ (yellow).

$I^{a_1}_{(6s,||)}$, and $I^{a_1}_{(1c,||)}$ are independent of the tensor coefficient $\tilde{C}_T$, providing the only the scalar and vector effects. Among these, the $I^{a_1}_{(6s,||)}$ displays a significant deviation in the presence of NP couplings whereas the others lie within 1σ of the SM contribution. The cumulative behavior of scalar, vector and tensor couplings on $I^{a_1}_{(5,||)}$ are deviated slightly from the SM value in the low q^2 region. However, the impact of NP couplings without tensor contribution are clearly remarked from the SM contribution, wherein the central value lies away from the 1σ bound of SM. In bottom-right panel of Fig. 3, the variation of $I^{a_1}_{(6s,||)}$ w.r.t q^2 is shown in the presence of the SM as well as in new physics. We observe that all the distributions with the inclusion of NP are reduced can be distinguishable. However, the effect of NP with tensor coefficient provides a large deviation from the standard model contribution in $q^2 \in [1.5,16]$ GeV^2.

2. τ mode behavior

After the study of μ mode analysis in $B \to a_1 \ell \nu$ process, we focus on the τ lepton behavior depicted in Fig. 4 as shown below. All the angular functions, both in SM, and in presence of $(S+V)$ and $(S+V+T)$ are shown in this figure. As we mentioned before, the NP is dependent on the $\tilde{C}_V$ and $\tilde{C}_S$ SMEFT couplings. In Fig. 4, the red color represents

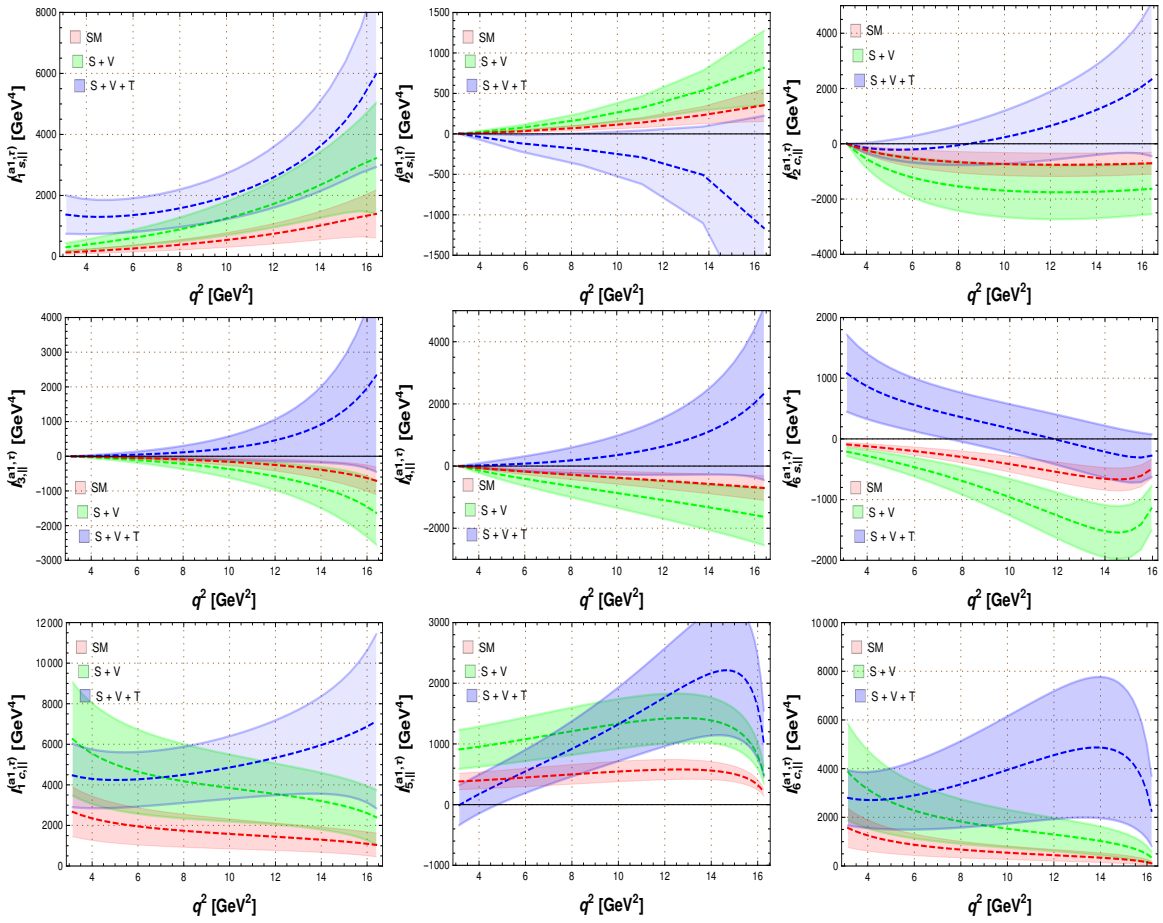


FIG. 4: The q^2 dependency of the angular coefficients of $B \to a_1(\rho_{||}\pi)\tau\bar{\nu}$ decay mode in SM (red), $S+V$ (green) and $S+V+T$ (blue).

the SM contribution whereas the NP contribution without (with) the tensor coupling is shown in green (blue) color. The angular functions deviate significantly in the presence of the benchmark values of the NP coefficients. We found that no such spectacular deviation is observed in the functions except $I^{a_1}_{6s,||}$, $I^{a_1}_{1c,||}$, $I^{a_1}_{5,||}$ and $I^{a_1}_{6c,||}$. These four functions are very sensitive to $\tilde{C}_V$ and $\tilde{C}_S$. However, the central value of the NP contribution remains away from the 1σ uncertainty of SM. Now, inclusion of the NP tensor coefficient in addition to $\tilde{C}_S$ and $\tilde{C}_V$, significantly enhanced the values of the angular coefficients from their SM results. The 1σ band of the NP contributions away from the 1σ uncertainty of the SM. In other words, though the deviation from the SM predictions in the functions is observed in the NP

scenarios, it is, however, more pronounced in the τ mode of $B \to a_1 \ell \nu$ process. Additionaly, a large hadronic uncertainty in the τ mode is obtained in the functions $I^{a_1}_{1s,||}$, $I^{a_1}_{1c,||}$ and $I^{a_1}_{6c,||}$ as compared to others. Similarly, while dealing with μ mode, the functions have large 1σ uncertainty as compared to the SM.

B. Analysis of transverse angular coefficients in $B \to a_1 \ell \nu$ process

In this subsection, we discuss all the transverse angular coefficients $I_i^{(a_1 \to \rho_\perp \pi)}$ of the $B \to a_1 \ell \nu$ decay mode in SM and in the SMEFT approach. Similar to the longitudinal analysis of $B \to a_1 \ell \nu$ channel, we also study the new physics effects in two scenarios, i.e., $(S+V)$ and $(S+V+T)$. Analogous to the longitudinal analysis, we also explore the given process for μ and τ leptonic modes. The details of the transverse analysis of $B \to a_1 \mu \nu$ and $B \to a_1 \tau \nu$ channel are depicted in Fig. 5 and Fig. 6, respectively, and are discussed below.

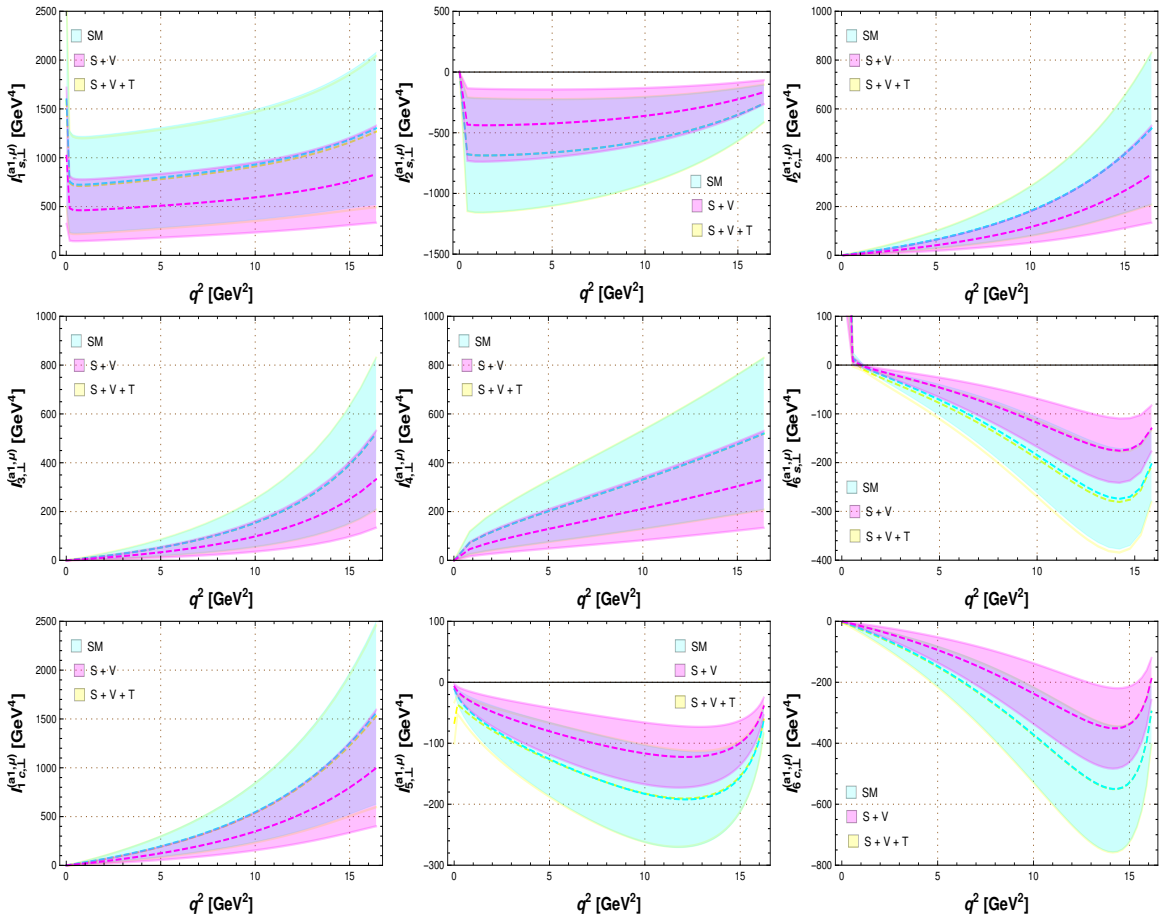


FIG. 5: The q^2 dependency of the angular coefficients of $B \to a_1(\rho_\perp \pi)\mu\bar{\nu}$ decay mode in SM (cyan), $S+V$ (magenta) and $S+V+T$ (yellow).

1. μ mode behavior

Here, we study the muonic behavior of the $B \to a_1 \ell\nu$ channel. The representation of the colors shown in Fig. 5 are same as the longitudinal case discussed in Fig. 3. One can observe that the coefficient functions $I^{a_1}_{1s,\perp}$, $I^{a_1}_{2c,\perp}$, $I^{a_1}_{3,\perp}$, $I^{a_1}_{4,\perp}$ and $I^{a_1}_{1c,\perp}$ are suppressed whereas the others such as $I^{a_1}_{2s,\perp}$, $I^{a_1}_{6s,\perp}$, $I^{a_1}_{5,\perp}$ and $I^{a_1}_{6c,\perp}$ are enhanced in the presence of both scalar and vector NP couplings. However, the presence of the coupling associated with the tensor operator Q_T does not induce any significant deviation in the angular coefficients from their SM values. Even the central SM values are almost same as the NP contributions with tensor effective coefficient. However, the functions $I^{a_1}_{6s,\perp}$ and $I^{a_1}_{5,\perp}$ show a slight differential behavior between the SM and the $(S+V+T)$ scenario.

2. τ mode behavior

In this subsection, we perform the transverse analysis of the angular coefficient functions of the $B \to a_1 \tau\nu$ decay mode in the presence of various NP couplings. Here also the scalar, vector and tensor operators involve in the new physics analysis. The details of the study has been shown in Fig. 6. The color representation of the plots depicted in this figure are same as the longitudinal analysis discussed in Fig. 4.

In the light of the NP coefficients of scalar and vector operators, the angular functions $I^{a_1}_{1s,\perp}$, $I^{a_1}_{2c,\perp}$, $I^{a_1}_{3,\perp}$, $I^{a_1}_{4,\perp}$, $I^{a_1}_{6s,\perp}$ and $I^{a_1}_{1c,\perp}$ get enhanced whereas the functions $I^{a_1}_{2s,\perp}$, $I^{a_1}_{5,\perp}$ and $I^{a_1}_{6c,\perp}$ shifted to lower values with large 1σ uncertainties. Now in order to have an insight of the tensor new physics coupling in addition to both $\tilde{C}_V$ and $\tilde{C}_S$, the angular coefficient functions I_i ar found to have deviations remarkably. The tensor coupling even shifts the functions $I^{a_1}_{1s,\perp}$, $I^{a_1}_{2s,\perp}$, $I^{a_1}_{6s,\perp}$, $I^{a_1}_{1c,\perp}$ and $I^{a_1}_{6c,\perp}$ to higher values as compared to the both scalar and vector contributions. However, the angular functions $I^{a_1}_{2c,\perp}$, $I^{a_1}_{3,\perp}$ and $I^{a_1}_{4,\perp}$ and $I^{a_1}_{5,\perp}$ shift downward and significantly deviate from the SM values.

After the analysis of the angular coefficient functions for $B \to a_1(\to \rho_{||}\pi)\ell\nu$ and $B \to a_1(\to \rho_{\perp}\pi)\ell\nu$ channels, we now proceed to explore other observables of the exclusive $B \to a_1 \ell\nu$ decay mode. In particular, we will focus on the branching ratio ($\mathcal{B}$) and the LFU violating observable $\mathcal{R}_{a_1}$. The q^2 variation of the branching ratios are shown in Fig. 7. The colors representing different contributions are presented in the plot legends. The left panel of the figure depicts $B \to a_1 \mu\nu$ channel. One can notice from the figure that, the branching fraction is enhanced with respect to the SM result in the presence of the operators without tensor. However, with the inclusion of tensor coupling, it remains consistent with its SM value. On the other hand, the right panel of the Fig. 7 represents the branching ratio of $B \to a_1 \tau\nu$ mode. In the presence of $(S+V)$ and $(S+V+T)$ type NP contributions, the branching ratio shows an appreciable deviation. In the lower left (right) panel, we display

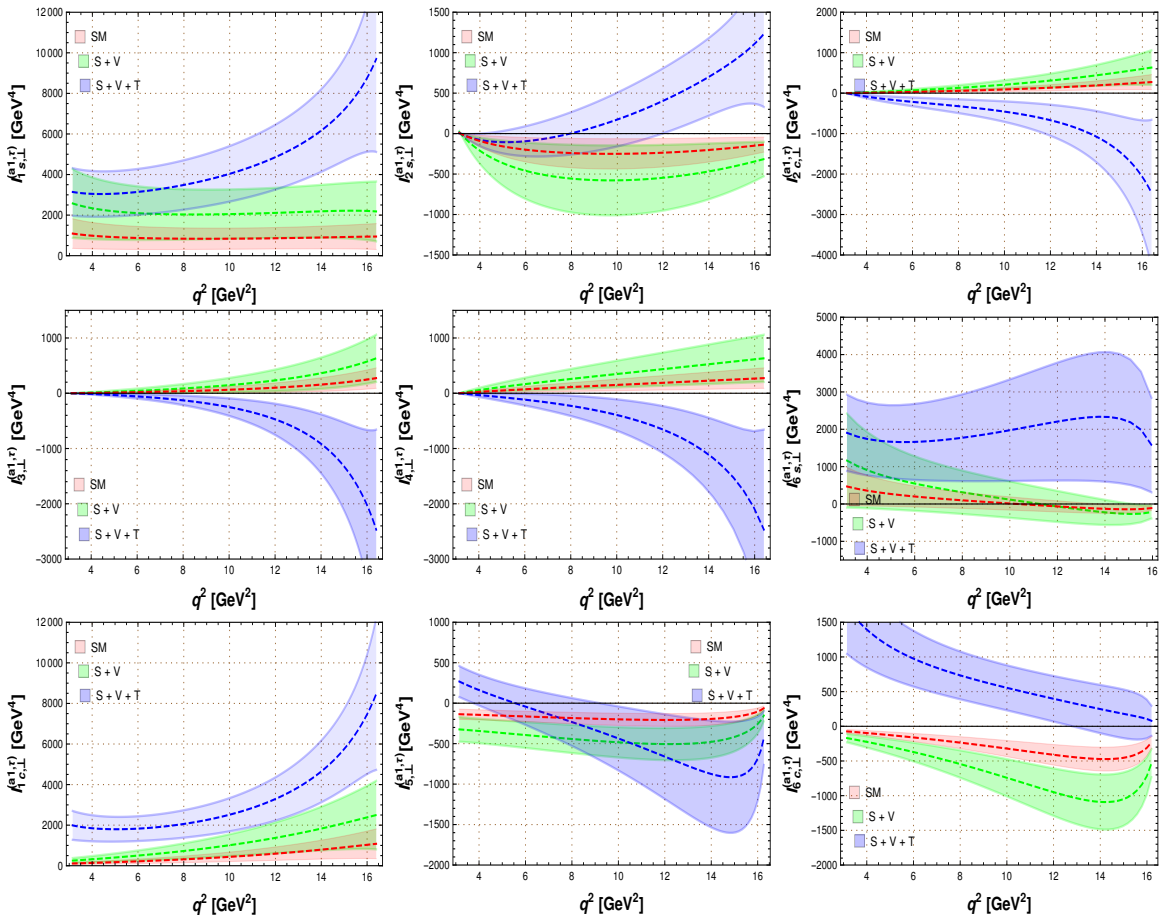


FIG. 6: The q^2 dependency of the angular coefficients of $B \to a_1(\rho_\perp \pi)\tau\bar{\nu}$ decay mode in SM (red), $S+V$ (green) and $S+V+T$ (blue).

the variations of the the forward-backward asymmetry of $B \to a_1\mu\nu$ ($B \to a_1\tau\nu$) mode. It should be noted that, the NP contribution to forward-backward asymmetry is almost negligible except for $S+V+T$ scenario for $B \to a_1\tau\nu$.

The predictions of the LFU violating observable in addition to the $\mathcal{B}$ and A_{FB} are coherently presented in Fig. 8. For completeness, the outcomes of this investigation are summarized below.

- $\mathcal{B}$: The simultaneous presence of the scalar and vector operators reduce the branching ratio (μ mode) while in the presence of $(S+V+T)$ couplings, it becomes consistent with the SM prediction. On the other hand, the branching fraction (τ mode) is enhanced in both of the NP scenarios, and has significant deviation from the SM.

- A_{FB} : We observe the zero crossing of the A_{FB} (μ mode) at $q^2 \simeq 1$ GeV^2 in the SM. In the presence of NP with tensor coupling, this has been slightly shifted to $q^2 \simeq 0.8$ GeV^2. On the other hand, for the τ mode the zero crossing occurs at $q^2 \simeq 11.5$ GeV^2 in the presence of NP without tensor operator. However, when the tensor coupling is

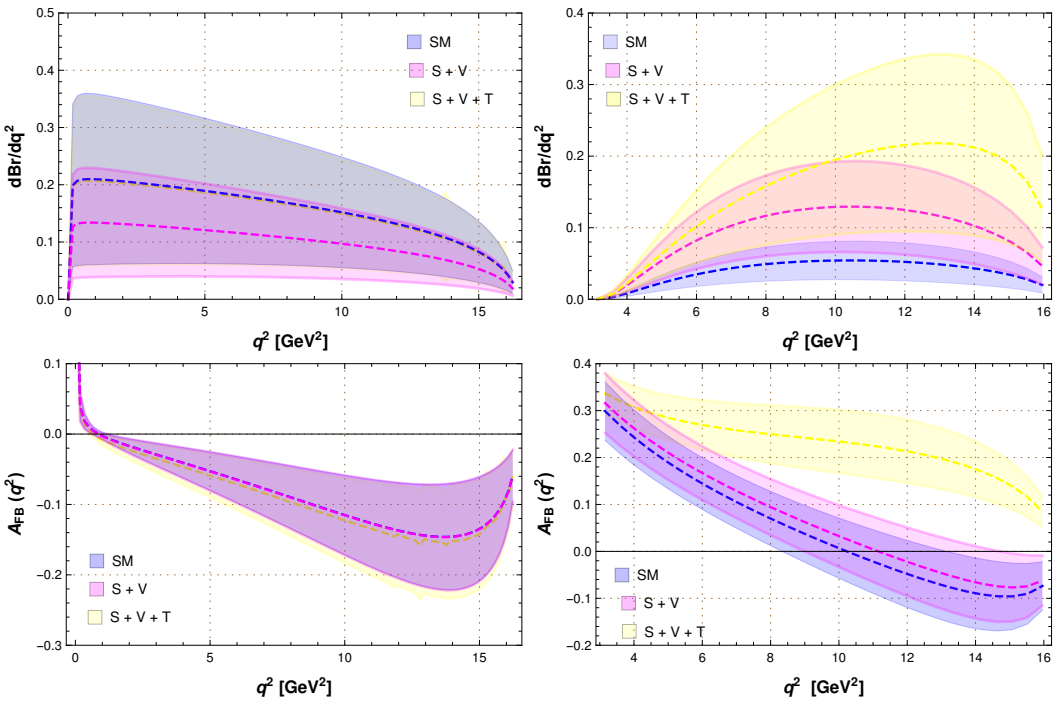


FIG. 7: The branching ratio (in units of 10^{-4}) and FB asymmetry of $B \to a_1 \ell \bar{\nu}$ for $\ell = \mu$ (left) and $\ell = \tau$ (right).

Observable \ Prediction	SM	NP with $(S+V)$	NP with $(S+V+T)$
$\mathcal{B}_\mu(\times 10^{-4})$	(2.053 ± 0.862)	(1.311 ± 0.536)	(2.018 ± 0.854)
$\mathcal{B}_\tau(\times 10^{-4})$	(0.700 ± 0.264)	(1.610 ± 0.531)	(2.585 ± 0.721)
A_{FB}^{μ}	(-0.068 ± 0.060)	(-0.070 ± 0.070)	(-0.073 ± 0.086)
A_{FB}^{τ}	(0.037 ± 0.065)	(0.060 ± 0.088)	(0.241 ± 0.115)
$\mathcal{R}_{a_1}$	(0.341 ± 0.192)	(1.228 ± 0.645)	(1.280 ± 0.650)

TABLE IV: The predictions of the observables of $B \to a_1 \ell \nu$ process in SM and in SMEFT approach.

included, we don't see any such zero-crossing.

- $\mathcal{R}_{a_1}$: We provide the prediction for the LFU violating observable $\mathcal{R}_{a_1}$, both in the SM and in the presence of new physics scenarios in Fig. 8 as well as in Table IV. From Table IV, one can observe a significant discrepancy between the SM and NP contributions.

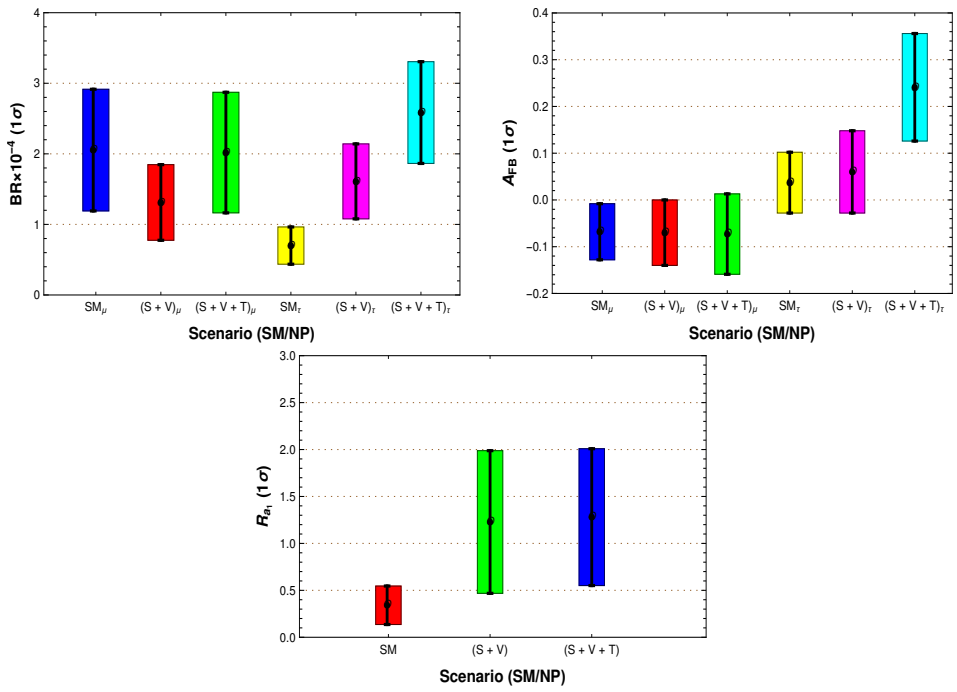


FIG. 8: Predictions of the $\mathcal{B}$, A_{FB} and R_{a_1} (1σ range) of $B \to a_1 \ell \bar{\nu}$ process, respectively for the allowed solutions corresponding to (S, V) and (S, V, T) operators.

V. CONCLUSION

Inspired by the anomalies present in the charged-current mediated $b \to c\ell\nu$ transitions, we have studied the CKM suppressed semileptonic $B \to a_1 \ell\nu$ decay mode, which is mediated through $b \to u\ell\nu$ transition, in the SMEFT framework. Considering the WET Lagrangian describing the $b \to u\ell\nu$ transitions, we correlate the NP couplings in terms of of the SMEFT couplings. We then constrained the parameter space for the new couplings by using the experimental available data of the branching fractions of $B \to \ell\bar{\nu}$ and $B \to (\pi, \rho, \omega)\ell\bar{\nu}$ processes in the presence of individual as well as mixed couplings of SMEFT Wilson coefficients. Using these constrained new couplings, we investigate the branching fractions, longitudinal and transverse angular coefficients as well as the forward-backward asymmetry parameters of $B \to a_1 \ell\nu$ processes for both muonic and tauonic final states for various NP scenarios. We found that these observables show appreciable deviations fro their SM predictions in the presence of NP operators as summarized in Table IV. It should be emphasized that with the inclusion of $S + V$ type NP contributions, the branching fraction of $B \to a_1 \mu\nu$ is reduced from its SM value while for $S + V + T$ type it remains consistent with the SM result. On the other hand, for tauonic mode the branching fraction enhanced significantly from SM for both type of NP scenarios. The forward-backward asymmetry parameter remains more or less consistent with the SM result. To summarize, we have performed a comprehensive study of $B \to a_1 \ell\nu$ decay mode and found that the branching fractions both in the SM as

well as in the NP scenarios are within the sensitivity reach of currently running Belle-II and LHCb experiments. The precise determination of this branching ratio would establish or rule out the possible role of new physics in $b \to u\ell\nu$ transition.

Acknowledgments

MKM would like to acknowledge IoE PDRF, University of Hyderabad for the financial support. DP acknowledges the support of Prime Minister's Research Fellowship, Government of India. RM would like to acknowledge University of Hyderabad IoE project grant no. RC1-20-012.

Appendix A: Hadronic Matrix Elements and the Helicity amplitudes

The hadronic matrix elements for $B \to a_1$ transition are written in terms of the form factors as follows [42],

$$\begin{aligned}\langle a_1(p',\epsilon)|\bar{u}\gamma_\mu(1-\gamma_5)b|\bar{B}(p)\rangle &= \frac{2A^{B\to a_1}(q^2)}{m_B+m_{a_1}}i\epsilon_{\mu\nu\alpha\beta}\epsilon^{*\nu}p^\alpha p'^\beta \\ &+ \Big\{(m_B+m_{a_1})\left[\epsilon^*_\mu - \frac{(\epsilon^*\cdot q)}{q^2}q_\mu\right]V_1^{B\to a_1}(q^2) \\ &- \frac{(\epsilon^*\cdot q)}{m_B+m_{a_1}}\left[(p+p')_\mu - \frac{m_B^2-m_{a_1}^2}{q^2}q_\mu\right]V_2^{B\to a_1}(q^2) \\ &+ (\epsilon^*\cdot q)\frac{2m_{a_1}}{q^2}q_\mu V_0^{B\to a_1}(q^2)\Big\} \qquad \text{(A1)}\end{aligned}$$

with the condition $V_0^{B\to a_1}(0) = \dfrac{m_B+m_{a_1}}{2m_{a_1}}V_1^{B\to a_1}(0) - \dfrac{m_B-m_{a_1}}{2m_{a_1}}V_2^{B\to a_1}(0)$, and

$$\langle a_1(p',\epsilon)|\bar{u}b|\bar{B}(p)\rangle = \frac{2m_{a_1}}{m_b-m_u}(\epsilon^*\cdot q)V_0^{B\to a_1}(q^2) \qquad \text{(A2)}$$

$$\begin{aligned}\langle a_1(p',\epsilon)|\bar{u}\sigma_{\mu\nu}b|\bar{B}(p)\rangle &= i\,T_0^{B\to a_1}(q^2)\frac{\epsilon^*\cdot q}{(m_B+m_{a_1})^2}(p_\mu p'_\nu - p_\nu p'_\mu) \\ &+ i\,T_1^{B\to a_1}(q^2)(p_\mu\epsilon^*_\nu - \epsilon^*_\mu p_\nu) + i\,T_2^{B\to a_1}(q^2)(p'_\mu\epsilon^*_\nu - \epsilon^*_\mu p'_\nu) \qquad \text{(A3)}\end{aligned}$$

$$\begin{aligned}\langle a_1(p',\epsilon)|\bar{u}\sigma_{\mu\nu}\gamma_5 b|\bar{B}(p)\rangle &= T_0^{B\to a_1}(q^2)\frac{\epsilon^*\cdot q}{(m_B+m_{a_1})^2}\epsilon_{\mu\nu\alpha\beta}p^\alpha p'^\beta \\ &+ T_1^{B\to a_1}(q^2)\epsilon_{\mu\nu\alpha\beta}p^\alpha\epsilon^{*\beta} + T_2^{B\to a_1}(q^2)\epsilon_{\mu\nu\alpha\beta}p'^\alpha\epsilon^{*\beta} \qquad \text{(A4)}\end{aligned}$$

The helicity amplitudes required for the $B \to a_1 \ell \nu$ process are given as follows,

$$\begin{aligned} H_0^{a_1} &= \frac{-(m_B+m_{a_1})^2(m_B^2-m_{a_1}^2-q^2)V_1(q^2)+\lambda(m_B^2,\,m_{a_1}^2,\,q^2)V_2(q^2)}{2m_{a_1}(m_B+m_{a_1})\sqrt{q^2}} \\ H_\pm^{a_1} &= \frac{-(m_B+m_{a_1})^2 V_1(q^2) \pm \sqrt{\lambda(m_B^2,\,m_{a_1}^2,\,q^2)}A(q^2)}{m_B+m_{a_1}} \\ H_t^{a_1} &= \frac{\sqrt{\lambda(m_B^2,\,m_{a_1}^2,\,q^2)}}{\sqrt{q^2}}\,V_0(q^2)\ . \end{aligned} \tag{A5}$$

For the tensor operator, the associated helicity amplitudes are given as [65]:

$$\begin{aligned} H_+^{NP} &= \frac{1}{\sqrt{q^2}}\left\{\left[m_B^2-m_{a_1}^2+\lambda^{1/2}(m_B^2,m_{a_1}^2,q^2)\right](T_1^{B\to a_1}+T_2^{B\to a_1})+q^2(T_1^{B\to a_1}-T_2^{B\to a_1})\right\} \\ H_-^{NP} &= \frac{1}{\sqrt{q^2}}\left\{\left[m_B^2-m_{a_1}^2-\lambda^{1/2}(m_B^2,m_{a_1}^2,q^2)\right](T_1^{B\to a_1}+T_2^{B\to a_1})+q^2(T_1^{B\to a_1}-T_2^{B\to a_1})\right\} \\ H_L^{NP} &= 4\Big\{\frac{\lambda(m_B^2,m_{a_1}^2,q^2)}{m_{a_1}(m_B+m_{a_1})^2}\,T_0^{B\to a_1}+2\frac{m_B^2+m_{a_1}^2-q^2}{m_{a_1}}\,T_1^{B\to a_1}+4m_{a_1}\,T_2^{B\to a_1}\Big\}\ . \end{aligned} \tag{A6}$$

Appendix B: Angular coefficient functions $I_i^{SM,NP}$ (Longitudinal and Transverse)

The details of the angular coefficients in the SM are given as follows:

i	$I_{i,\|\|}$	$I_{i,\perp}$
$I_{1s}^{a_1}$	$\frac{1}{2}(H_+^2+H_-^2)(m_\ell^2+3q^2)$	$2H_t^2m_\ell^2+H_0^2(m_\ell^2+q^2)+\frac{1}{4}(H_+^2+H_-^2)(m_\ell^2+3q^2)$
$I_{1c}^{a_1}$	$4H_t^2m_\ell^2+2H_0^2(m_\ell^2+q^2)$	$\frac{1}{2}(H_+^2+H_-^2)(m_\ell^2+3q^2)$
$I_{2s}^{a_1}$	$-\frac{1}{2}(H_+^2+H_-^2)(m_\ell^2-q^2)$	$[H_0^2-\frac{1}{4}(H_+^2+H_-^2)](m_\ell^2-q^2)$
$I_{2c}^{a_1}$	$2H_0^2(m_\ell^2-q^2)$	$-\frac{1}{2}(H_+^2+H_-^2)(m_\ell^2-q^2)$
$I_3^{a_1}$	$2H_+H_-(m_\ell^2-q^2)$	$-H_+H_-(m_\ell^2-q^2)$
$I_4^{a_1}$	$H_0(H_++H_-)(m_\ell^2-q^2)$	$-\frac{1}{2}H_0(H_++H_-)(m_\ell^2-q^2)$
$I_5^{a_1}$	$-2H_t(H_++H_-)m_\ell^2-2H_0(H_+-H_-)q^2$	$H_t(H_++H_-)m_\ell^2+H_0(H_+-H_-)q^2$
$I_{6s}^{a_1}$	$2(H_+^2-H_-^2)q^2$	$-4H_tH_0m_\ell^2+(H_+^2-H_-^2)q^2$
$I_{6c}^{a_1}$	$-8H_tH_0m_\ell^2$	$2(H_+^2-H_-^2)q^2$
$I_7^{a_1}$	0	0

TABLE V: Angular coefficient functions in the four dimensional $\bar{B} \to a_1(\rho\pi)\ell^-\bar{\nu}_\ell$ decay distribution given in Eq.18, in SM.

The details of the angular coefficients in the new physics are given as follows:

TABLE VI: Angular coefficient functions for $\bar{B} \to a_1(\rho\pi)\ell^-\bar{\nu}_\ell$: NP term with S operator, interference SM-NP with S operator, and NP-NP interference with S and T operators, Eq.(20).

i	$I_{i,\|\|}^{NP,S}$	$I_{i,\|\|}^{INT,S}$	$I_{i,\|\|}^{INT,ST}$
$I_{1s}^{a_1}$	0	0	0
$I_{1c}^{a_1}$	$4H_t^2\frac{q^4}{(m_b-m_u)^2}$	$4H_t^2\frac{m_\ell q^2}{m_b-m_u}$	0
$I_{2s}^{a_1}$	0	0	0
$I_{2c}^{a_1}$	0	0	0
$I_3^{a_1}$	0	0	0
$I_4^{a_1}$	0	0	0
$I_5^{a_1}$	0	$-H_t(H_+ + H_-)\frac{m_\ell q^2}{m_b-m_u}$	$-2H_t(H_+^{NP} + H_-^{NP})\frac{(q^2)^{3/2}}{m_b-m_u}$
$I_{6s}^{a_1}$	0	0	0
$I_{6c}^{a_1}$	0	$-4H_tH_0\frac{m_\ell q^2}{m_b-m_u}$	$-H_t\,H_L^{NP}\frac{(q^2)^{3/2}}{m_b-m_u}$
$I_7^{a_1}$	0	$-H_t(H_+ - H_-)\frac{m_\ell q^2}{m_b-m_u}$	$-2H_t(H_+^{NP} - H_-^{NP})\frac{(q^2)^{3/2}}{m_b-m_u}$

i	$I_{i,\perp}^{NP,S}$	$I_{i,\perp}^{INT,S}$	$I_{i,\perp}^{INT,ST}$
$I_{1s}^{a_1}$	$2H_t^2\frac{q^4}{(m_b-m_u)^2}$	$2H_t^2\frac{m_\ell q^2}{m_b-m_u}$	0
$I_{1c}^{a_1}$	0	0	0
$I_{2s}^{a_1}$	0	0	0
$I_{2c}^{a_1}$	0	0	0
$I_3^{a_1}$	0	0	0
$I_4^{a_1}$	0	0	0
$I_5^{a_1}$	0	$\frac{1}{2}H_t(H_+ + H_-)\frac{m_\ell q^2}{m_b-m_u}$	$H_t(H_+^{NP} + H_-^{NP})\frac{(q^2)^{3/2}}{m_b-m_u}$
$I_{6s}^{a_1}$	0	$-2H_tH_0\frac{m_\ell q^2}{m_b-m_u}$	$-H_t\,H_L^{NP}\frac{(q^2)^{3/2}}{2(m_b-m_u)}$
$I_{6c}^{a_1}$	0	0	0
$I_7^{a_1}$	0	$\frac{1}{2}H_t(H_+ - H_-)\frac{m_\ell q^2}{m_b-m_u}$	$H_t(H_+^{NP} - H_-^{NP})\frac{(q^2)^{3/2}}{m_b-m_u}$

TABLE VII: Similar as Table VI with transverse coefficients.

i	$I^{NP,T}_{i,\parallel}$	$I^{INT,T}_{i,\parallel}$
$I^{a_1}_{1s}$	$2[(H^{NP}_+)^2+(H^{NP}_-)^2](3m_\ell^2+q^2)$	$4(H^{NP}_+H_+ + H^{NP}_-H_-)m_\ell\sqrt{q^2}$
$I^{a_1}_{1c}$	$\frac{1}{8}(H^{NP}_L)^2(m_\ell^2+q^2)$	$H^{NP}_L H_0 m_\ell\sqrt{q^2}$
$I^{a_1}_{2s}$	$2[(H^{NP}_+)^2+(H^{NP}_-)^2](m_\ell^2-q^2)$	0
$I^{a_1}_{2c}$	$-\frac{1}{8}(H^{NP}_L)^2(m_\ell^2-q^2)$	0
$I^{a_1}_{3}$	$-8H^{NP}_+H^{NP}_-(m_\ell^2-q^2)$	0
$I^{a_1}_{4}$	$-\frac{1}{2}H^{NP}_L(H^{NP}_+ + H^{NP}_-)(m_\ell^2-q^2)$	0
$I^{a_1}_{5}$	$-H^{NP}_L(H^{NP}_+ - H^{NP}_-)m_\ell^2$	$-\frac{1}{4}[H^{NP}_L(H_+-H_-)+8H^{NP}_+(H_t+H_0)$
		$+8H^{NP}_-(H_t-H_0)]m_\ell\sqrt{q^2}$
$I^{a_1}_{6s}$	$8[(H^{NP}_+)^2-(H^{NP}_-)^2]m_\ell^2$	$4(H^{NP}_+H_+ - H^{NP}_-H_-)m_\ell\sqrt{q^2}$
$I^{a_1}_{6c}$	0	$-H^{NP}_L H_t m_\ell\sqrt{q^2}$
$I^{a_1}_{7}$	0	$-\frac{1}{4}[H^{NP}_L(H_++H_-)-8H^{NP}_+(H_t+H_0)$
		$+8H^{NP}_-(H_t-H_0)]m_\ell\sqrt{q^2}$

TABLE VIII: Angular coefficient functions for $\bar{B}\to a_1(\rho\pi)\ell^-\bar{\nu}_\ell$: NP term with T operator and interference SM-NP with T operator.

i	$I^{NP,T}_{i,\perp}$	$I^{INT,T}_{i,\perp}$
$I^{a_1}_{1s}$	$[(H^{NP}_+)^2+(H^{NP}_-)^2](3m_\ell^2+q^2)$	$\frac{1}{2}[4(H^{NP}_+H_+ + H^{NP}_-H_-)$
	$+\frac{1}{16}(H^{NP}_L)^2(m_\ell^2+q^2)$	$+H^{NP}_L H_0]m_\ell\sqrt{q^2}$
$I^{a_1}_{1c}$	$2[(H^{NP}_+)^2+(H^{NP}_-)^2](3m_\ell^2+q^2)$	$4(H^{NP}_+H_+ + H^{NP}_-H_-)m_\ell\sqrt{q^2}$
$I^{a_1}_{2s}$	$[(H^{NP}_+)^2+(H^{NP}_-)^2](m_\ell^2-q^2)$	0
	$-\frac{1}{16}(H^{NP}_L)^2(m_\ell^2-q^2)$	
$I^{a_1}_{2c}$	$2[(H^{NP}_+)^2+(H^{NP}_-)^2](m_\ell^2-q^2)$	0
$I^{a_1}_{3}$	$4H^{NP}_+H^{NP}_-(m_\ell^2-q^2)$	0
$I^{a_1}_{4}$	$\frac{1}{4}H^{NP}_L(H^{NP}_+ + H^{NP}_-)(m_\ell^2-q^2)$	0
$I^{a_1}_{5}$	$\frac{1}{2}H^{NP}_L(H^{NP}_+ - H^{NP}_-)m_\ell^2$	$\frac{1}{8}[H^{NP}_L(H_+-H_-)+8H^{NP}_+(H_t+H_0)$
		$+8H^{NP}_-(H_t-H_0)]m_\ell\sqrt{q^2}$
$I^{a_1}_{6s}$	$4[(H^{NP}_+)^2-(H^{NP}_-)^2]m_\ell^2$	$-\frac{1}{2}[-4(H^{NP}_+H_+ - H^{NP}_-H_-)+H^{NP}_L H_t]m_\ell\sqrt{q^2}$
$I^{a_1}_{6c}$	$8[(H^{NP}_+)^2-(H^{NP}_-)^2]m_\ell^2$	$4(H^{NP}_+H_+ - H^{NP}_-H_-)m_\ell\sqrt{q^2}$
$I^{a_1}_{7}$	0	$\frac{1}{8}[H^{NP}_L(H_++H_-)-8H^{NP}_+(H_t+H_0)$
		$+8H^{NP}_-(H_t-H_0)]$

TABLE IX: Similar as Table VIII with transverse coefficients.

[1] R. Aaij *et al.*, "Test of lepton universality in $b\to s\ell^+\ell^-$ decays," *Phys. Rev. Lett.*, vol. 131, no. 5, p. 051803, 2023.

[2] R. Aaij *et al.*, "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays," *Phys. Rev. D*, vol. 108, no. 3, p. 032002, 2023.

[3] R. Aaij *et al.*, "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$," *Phys. Rev. Lett.*, vol. 111, p. 191801, 2013.

[4] R. Aaij *et al.*, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity," *JHEP*, vol. 02, p. 104, 2016.

[5] M. Aaboud *et al.*, "Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *JHEP*, vol. 10, p. 047, 2018.

[6] R. Aaij *et al.*, "Branching Fraction Measurements of the Rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$- Decays," *Phys. Rev. Lett.*, vol. 127, no. 15, p. 151801, 2021.

[7] R. Aaij *et al.*, "Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$," *JHEP*, vol. 09, p. 179, 2015.

[8] R. Aaij *et al.*, "Tests of lepton universality using $B^0 \to K_S^0\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays," *Phys. Rev. Lett.*, vol. 128, no. 19, p. 191802, 2022.

[9] J. A. Bailey *et al.*, "B→D$\ell\nu$ form factors at nonzero recoil and —V$_{cb}$— from 2+1-flavor lattice QCD," *Phys. Rev. D*, vol. 92, no. 3, p. 034506, 2015.

[10] H. Na, C. M. Bouchard, G. P. Lepage, C. Monahan, and J. Shigemitsu, "$B \to Dl\nu$ form factors at nonzero recoil and extraction of $|V_{cb}|$," *Phys. Rev. D*, vol. 92, no. 5, p. 054510, 2015. [Erratum: Phys.Rev.D 93, 119906 (2016)].

[11] S. Aoki *et al.*, "Review of lattice results concerning low-energy particle physics," *Eur. Phys. J. C*, vol. 77, no. 2, p. 112, 2017.

[12] D. Bigi and P. Gambino, "Revisiting $B \to D\ell\nu$," *Phys. Rev. D*, vol. 94, no. 9, p. 094008, 2016.

[13] A. Abdesselam *et al.*, "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method," 4 2019.

[14] Y. S. Amhis *et al.*, "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021," *Phys. Rev. D*, vol. 107, no. 5, p. 052008, 2023.

[15] S. Hirose *et al.*, "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$," *Phys. Rev. Lett.*, vol. 118, no. 21, p. 211801, 2017.

[16] S. Hirose *et al.*, "Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle," *Phys. Rev. D*, vol. 97, no. 1, p. 012004, 2018.

[17] A. Abdesselam *et al.*, "Measurement of the D^{*-} polarization in the decay $B^0 \to D^{*-}\tau^+\nu_\tau$," in *10th International Workshop on the CKM Unitarity Triangle*, 3 2019.

[18] R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$," *Phys. Rev. Lett.*, vol. 120, no. 12, p. 121801, 2018.

[19] R. Dutta, "Exploring R_D, R_{D^*} and $R_{J/\Psi}$ anomalies," 10 2017.

[20] R. Dutta and A. Bhol, "$B_c \to (J/\psi, \eta_c)\tau\nu$ semileptonic decays within the standard model and beyond," *Phys. Rev. D*, vol. 96, no. 7, p. 076001, 2017.

[21] I. Adachi *et al.*, "Evidence for $B^- \to \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle," *Phys. Rev. Lett.*, vol. 110, no. 13, p. 131801, 2013.

[22] B. Kronenbitter *et al.*, "Measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$ decays with the semileptonic tagging method," *Phys. Rev. D*, vol. 92, no. 5, p. 051102, 2015.

[23] B. Aubert *et al.*, "A Search for $B^+ \to \ell^+ \nu_\ell$ Recoiling Against $B^- \to D^0 \ell^- \bar{\nu} X$," *Phys. Rev. D*, vol. 81, p. 051101, 2010.

[24] J. P. Lees *et al.*, "Evidence of $B^+ \to \tau^+ \nu$ decays with hadronic B tags," *Phys. Rev. D*, vol. 88, no. 3, p. 031102, 2013.

[25] M. Bona *et al.*, "An Improved Standard Model Prediction Of BR(B —> tau nu) And Its Implications For New Physics," *Phys. Lett. B*, vol. 687, pp. 61–69, 2010.

[26] J. Charles *et al.*, "Predictions of selected flavour observables within the Standard Model," *Phys. Rev. D*, vol. 84, p. 033005, 2011.

[27] P. Hamer *et al.*, "Search for $B^0 \to \pi^- \tau^+ \nu_\tau$ with hadronic tagging at Belle," *Phys. Rev. D*, vol. 93, no. 3, p. 032007, 2016.

[28] B. Aubert *et al.*, "Observation of B0 meson decays to a+(1)(1260) pi-," in *32nd International Conference on High Energy Physics*, 8 2004.

[29] B. Aubert *et al.*, "Observation of B0 Meson Decay to a+-(1)(1260) pi-+," *Phys. Rev. Lett.*, vol. 97, p. 051802, 2006.

[30] B. Aubert *et al.*, "Measurements of CP-Violating Asymmetries in $B^0 \to$ a+-(1) (1260) $\pi^\mp$ decays," *Phys. Rev. Lett.*, vol. 98, p. 181803, 2007.

[31] K. Abe *et al.*, "Measurement of the Branching Fraction for B0 —> a+-(1)(1260) pi-+ with 535 Million B anti-B Pairs," in *5th Conference on Flavor Physics and CP Violation*, 6 2007.

[32] T. M. Aliev and M. Savci, "Semileptonic B —> a(1) lepton neutrino decay in QCD," *Phys. Lett. B*, vol. 456, pp. 256–263, 1999.

[33] Z.-G. Wang, "Analysis of the $B \to a_1(1260)$ form-factors with light-cone QCD sum rules," *Phys. Lett. B*, vol. 666, pp. 477–482, 2008.

[34] K.-C. Yang, "Form-Factors of B(u,d,s) Decays into P-Wave Axial-Vector Mesons in the Light-Cone Sum Rule Approach," *Phys. Rev. D*, vol. 78, p. 034018, 2008.

[35] R.-H. Li, C.-D. Lu, and W. Wang, "Transition form factors of B decays into p-wave axial-vector mesons in the perturbative QCD approach," *Phys. Rev. D*, vol. 79, p. 034014, 2009.

[36] N. Rajeev and R. Dutta, "Impact of vector new physics couplings on $B_s \to (K, K^*)\tau\nu$ and $B \to \pi\tau\nu$ decays," *Phys. Rev. D*, vol. 98, no. 5, p. 055024, 2018.

[37] S. Sahoo, A. Ray, and R. Mohanta, "Model independent investigation of rare semileptonic $b \to u l \bar{\nu}_l$ decay processes," *Phys. Rev. D*, vol. 96, no. 11, p. 115017, 2017.

[38] R. Dutta, "Predictions of $B_c \to (D, D^*)\tau\,\nu$ decay observables in the standard model," *J. Phys. G*, vol. 46, no. 3, p. 035008, 2019.

[39] R. Dutta, "$\Lambda_b \to (\Lambda_c, p)\,\tau\,\nu$ decays within standard model and beyond," *Phys. Rev. D*, vol. 93, no. 5, p. 054003, 2016.

[40] X.-W. Kang, T. Luo, Y. Zhang, L.-Y. Dai, and C. Wang, "Semileptonic B and B_s decays involving scalar and axial-vector mesons," *Eur. Phys. J. C*, vol. 78, no. 11, p. 909, 2018.

[41] P. Colangelo, F. De Fazio, and F. Loparco, "Probing New Physics with $\bar{B} \to \rho(770)\,\ell^-\bar{\nu}_\ell$ and $\bar{B} \to a_1(1260)\,\ell^-\bar{\nu}_\ell$," *Phys. Rev. D*, vol. 100, no. 7, p. 075037, 2019.

[42] P. Colangelo, F. De Fazio, and F. Loparco, "Probes of Lepton Flavor Universality in $b \to u$ Transitions," *Particles*, vol. 3, no. 1, pp. 145–163, 2020.

[43] Y. B. Li *et al.*, "Evidence of a structure in $\bar{K}^0\Lambda_c^+$ consistent with a charged $\Xi_c(2930)^+$, and updated measurement of $\bar{B}^0 \to \bar{K}^0\Lambda_c^+\bar{\Lambda}_c^-$ at Belle," *Eur. Phys. J. C*, vol. 78, no. 11, p. 928, 2018.

[44] B. Pal *et al.*, "Observation of the decay $B_s^0 \to K^0\overline{K}^0$," *Phys. Rev. Lett.*, vol. 116, no. 16, p. 161801, 2016.

[45] S. Weinberg, "Effective Gauge Theories," *Phys. Lett. B*, vol. 91, pp. 51–55, 1980.

[46] S. R. Coleman, J. Wess, and B. Zumino, "Structure of phenomenological Lagrangians. 1.," *Phys. Rev.*, vol. 177, pp. 2239–2247, 1969.

[47] C. G. Callan, Jr., S. R. Coleman, J. Wess, and B. Zumino, "Structure of phenomenological Lagrangians. 2.," *Phys. Rev.*, vol. 177, pp. 2247–2250, 1969.

[48] W. Buchmuller and D. Wyler, "Effective Lagrangian Analysis of New Interactions and Flavor Conservation," *Nucl. Phys. B*, vol. 268, pp. 621–653, 1986.

[49] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian," *JHEP*, vol. 10, p. 085, 2010.

[50] A. Greljo, J. Salko, A. Smolkovič, and P. Stangl, "SMEFT restrictions on exclusive b $\to$ u$\ell\nu$ decays," *JHEP*, vol. 11, p. 023, 2023.

[51] J. Aebischer, A. Crivellin, M. Fael, and C. Greub, "Matching of gauge invariant dimension-six operators for $b \to s$ and $b \to c$ transitions," *JHEP*, vol. 05, p. 037, 2016.

[52] M. González-Alonso, J. Martin Camalich, and K. Mimouni, "Renormalization-group evolution of new physics contributions to (semi)leptonic meson decays," *Phys. Lett. B*, vol. 772, pp. 777–785, 2017.

[53] V. Cirigliano, J. Jenkins, and M. Gonzalez-Alonso, "Semileptonic decays of light quarks beyond the Standard Model," *Nucl. Phys. B*, vol. 830, pp. 95–115, 2010.

[54] T. Bhattacharya, V. Cirigliano, S. D. Cohen, A. Filipuzzi, M. Gonzalez-Alonso, M. L. Graesser, R. Gupta, and H.-W. Lin, "Probing Novel Scalar and Tensor Interactions from (Ultra)Cold Neutrons to the LHC," *Phys. Rev. D*, vol. 85, p. 054512, 2012.

[55] J. Aebischer, M. Fael, C. Greub, and J. Virto, "B physics Beyond the Standard Model at One Loop: Complete Renormalization Group Evolution below the Electroweak Scale," *JHEP*, vol. 09, p. 158, 2017.

[56] E. E. Jenkins, A. V. Manohar, and P. Stoffer, "Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching," *JHEP*, vol. 03, p. 016, 2018.

[57] P. Biancofiore, P. Colangelo, and F. De Fazio, "On the anomalous enhancement observed in

$B \to D^{(*)}\tau\bar{\nu}_\tau$ decays," *Phys. Rev. D*, vol. 87, no. 7, p. 074010, 2013.

[58] M. Tanaka and R. Watanabe, "New physics in the weak interaction of $\bar{B} \to D^{(*)}\tau\bar{\nu}$," *Phys. Rev. D*, vol. 87, no. 3, p. 034028, 2013.

[59] Y. Sakaki, M. Tanaka, A. Tayduganov, and R. Watanabe, "Testing leptoquark models in $\bar{B} \to D^{(*)}\tau\bar{\nu}$," *Phys. Rev. D*, vol. 88, no. 9, p. 094012, 2013.

[60] R. L. Workman *et al.*, "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, 2022.

[61] C. Bourrely, I. Caprini, and L. Lellouch, "Model-independent description of B —> pi l nu decays and a determination of —V(ub)—," *Phys. Rev. D*, vol. 79, p. 013008, 2009. [Erratum: Phys.Rev.D 82, 099902 (2010)].

[62] J. A. Bailey *et al.*, "$|V_{ub}|$ from $B \to \pi\ell\nu$ decays and (2+1)-flavor lattice QCD," *Phys. Rev. D*, vol. 92, no. 1, p. 014024, 2015.

[63] J. A. Bailey *et al.*, "$B \to \pi\ell\ell$ form factors for new-physics searches from lattice QCD," *Phys. Rev. Lett.*, vol. 115, no. 15, p. 152002, 2015.

[64] A. Bharucha, D. M. Straub, and R. Zwicky, "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules," *JHEP*, vol. 08, p. 098, 2016.

[65] P. Colangelo and F. De Fazio, "Scrutinizing $\overline{B} \to D^*(D\pi)\,\ell^-\overline{\nu}_\ell$ and $\overline{B} \to D^*(D\gamma)\,\ell^-\overline{\nu}_\ell$ in search of new physics footprints," *JHEP*, vol. 06, p. 082, 2018.

arXiv:2402.18254v1 [hep-ph] 28 Feb 2024

Photon quantum kinetic equations and collective modes in an axion background

Marc Comadran[1,2,3*] and Cristina Manuel[1,2†]

[1] *Instituto de Ciencias del Espacio (ICE, CSIC) C. Can Magrans s.n., 08193 Cerdanyola del Vallès, Spain and*

[2] *Institut d'Estudis Espacials de Catalunya (IEEC), Ed. RDIT, Campus UPC, 08860 Castelldefels (Barcelona), Spain and*

[3] *Departament de Física Quàntica i Astrofísica and Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franqués 1, 08028 Barcelona, Spain*

(Dated: February 29, 2024)

We develop a quantum kinetic theory for photons in the presence of an axion background and in the collisioness limit. In deriving the classical regime of our quantum kinetic equations, we observe that they capture well known features of axion electrodynamics. By projecting the Wigner function onto a polarization basis, relating the Wigner matrix function with the Stokes parameters, we establish the dispersion relations and transport equations for each polarization space component. Additionally, we investigate how the axion background affects the dispersion relations of photon collective modes within an electron-positron plasma at equilibrium temperature T. While the plasmon remains unaffected, we find that the axion background breaks the degeneracy of transverse collective modes at order $eg_{a\gamma}T(\partial a)$, where e represents the electron charge, $g_{a\gamma}$ denotes the photon-axion coupling, and ∂a represents the scale associated with variations in the axion field.

* mcomadca60@alumnes.ub.edu

† cmanuel@ice.csic.es

I. INTRODUCTION

The exploration of axion electrodynamics holds interest in various domains of physics. Originally conceived within the realm of high-energy physics to address the so called CP problem of QCD [1–4], the term axion is nowadays more generically used to refer to a broader class of light pseudoscalar particles, regardless or not they are related to QCD. Axions naturally manifest in extensions of the Standard Model of particle physics, thus deserving serious consideration as potential candidates for dark matter. There are several intense experimental programs to search for these elusive particles, both in the laboratory an in astrophysical scenarios (see, for example, [5, 6], and references therein). Concurrently, analogous axion-photon couplings manifest in certain condensed matter systems [7], giving account of topological magnetoelectric phenomena.

In this article we will consider photon properties in a plasma when there is also an axion background. A plasma is typically characterized by either a temperature T and/or chemical potential μ. We will consider that the axion wavelength is much lesser than any of these scales T, μ which describe the medium. The interactions between axions and photons are described by the Lagrangian

$$\mathcal{L}_{\rm int} = \frac{1}{4} g_{a\gamma} a F_{\mu\nu} \tilde{F}^{\mu\nu} \ . \tag{1}$$

Here $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$ is the dual of the electromagnetic tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, being $\epsilon^{\mu\nu\rho\sigma}$ the Levi-Civita tensor, the fields $A^\mu(x)$ and $a(x)$ are associated to photons and axions respectively, and $g_{a\gamma}$ is the axion-photon coupling [6]. This coupling suggests that an axion background acts as an effective chiral medium. There has been a renewed interest in the last decade on chiral media, with the discovery of a variety of new quantum chiral transport phenomena [8], such as the chiral magnetic effect. A clear parallelism among axion electrodynamics and chiral media has been stressed [9].

Quantum field theory methods have been developed for the study of relativistic plasmas [10]. It is by now well understood that the particle fields of different energy scales have to be treated differently (see, for example [11]). Effective field theories have been designed for that purpose. For momenta scales of the order of the temperature T or higher, the photons are treated as quasiparticles, which are more efficiently described with transport equations. For momenta much lower than the temperature T the photon fields are then described as classical background fields, whose properties are then modified by the medium. Collective modes then

emerge for those scales, with the appearance of the so called plasmon mode [12, 13], which is absent in vacuum.

The purpose of this article is to study how an axion background modifies both the photon transport equation and also the collective modes of a thermal plasma. We will assume that the time and spatial scales of variation of the axion are much less than the photon momentum. While our study would be valid for astrophysical and cosmological plasmas, it can be also of used for other condensed matter systems. We will use well established quantum field theory methods to study these two effects. Similar photon transport equations have been derived [14, 15]. We will comment later on differences with Ref. [14], while in Ref. [15] the axions are considered as quasiparticles and not as a background. Therefore there the axions interact with the photons through the collision term, allowing for the conversion of axions into photons in the background of a magnetic field, as those commonly found in astrophysical plasmas. In this work, we ignore these processes, while they could be easily incorporated by including a proper collision term to our transport equations.

The paper is structured as follows; in Sec. II we derive quantum kinetic equations for photons in the presence of an axion background using the Keldysh Schiwgner formalism and in Sec. III we address the effect of the axion background onto the collective modes of photons. Finally in Sec. IV we discuss our findings and make contact with other approaches found in the literature. We also give some details on the derivation of the kinetic equations in Appendix A. Appendix B is devoted to provide the transport equations in a linear polarization basis.

Let us establish our notations and conventions. We use the signature $\mathrm{diag}(g^{\mu\nu}) = (1,-1,-1,-1)$, and the normalization $\epsilon^{0123} = -\epsilon_{0123} = 1$ for the Levi-Civita tensor. Rising or lowering spatial indices produces a minus sign e.g if $A^\mu = (A^0, A^i)$ is a vector in Minkowski space, we have $A^i = -A_i$. Boldface letters will be used to denote three dimensional vectors $A^i = \boldsymbol{A}$. The four gradient is $\partial_\mu = (\partial_0, \partial_i)$, where $\partial_i = \nabla$. The Minkowski product between A^μ and B^μ is defined as $A \cdot B = g_{\mu\nu} A^\mu B^\mu = A^0 B^0 - \boldsymbol{A} \cdot \boldsymbol{B}$. Natural units $\hbar = c = k_B = 1$ are used throughout.

II. QUANTUM KINETIC EQUATIONS FOR PHOTONS IN AN AXION BACKGROUND

The Keldysh-Schwinger formulation of quantum field theory has become a well-established tool to study relativistic plasmas, whether they are at or far to thermal equilibrium [16]. In

the Keldysh-Schwigner formalism, non-equilibrium Green functions are defined as 2×2 matrix in the complex, closed time path contour, see e.g Ref. [17] for a recent review. In the case of photons one defines

$$\boldsymbol{G}^{\mu\nu}(x,y) = \begin{pmatrix} G^{c,\mu\nu}(x,y) & G^{<,\mu\nu}(x,y) \\ G^{>,\mu\nu}(x,y) & G^{a,\mu\nu}(x,y) \end{pmatrix} = \begin{pmatrix} \langle \mathcal{T} A^\mu(x) A^\nu(y)\rangle & \langle A^\nu(y) A^\mu(x)\rangle \\ \langle A^\mu(x) A^\nu(y)\rangle & \langle \tilde{\mathcal{T}} A^\mu(x) A^\nu(y)\rangle \end{pmatrix} , \tag{2}$$

where $\mathcal{T}$ and $\tilde{\mathcal{T}}$ denote time and anti-time ordering along the complex path, respectively, and $\langle \ldots \rangle$ stands for average over an ensemble of states. In order to derive kinetic equations it is sufficient to study the dynamics of the lesser (or greater) component of the Green function $G^{<,\mu\nu}(x,y)$ (or $G^{>,\mu\nu}(x,y)$), as this is related to the photon phase space density after a Wigner transformation.

A relevant problem that emerges in the present formulation is that the components of the Green function in Eq. (2) are not gauge invariant quantities, and they contain non-transverse degrees of freedom. There are various ways to circumvent this inconvenience, the usual prescription is to impose gauge fixing conditions, although there are other approaches, such as defining a gauge-invariant two-point Green function [18]. In a forthcoming publication, we will explore alternative possibilities by employing effective field theory techniques. In this work, we adopt the former and impose the Lorentz gauge $\partial^\mu A_\mu = 0$, which leads to the following gauge conditions for the lesser component of the Green function

$$\partial_{x,\mu} G^{<,\mu\nu}(x,y) = 0 \ , \tag{3a}$$

$$\partial_{y,\nu} G^{<,\mu\nu}(x,y) = 0 \ . \tag{3b}$$

The equations of motion obeyed by each component of the Green function can be deduced from the Kadanoff-Baym equations [19]. In the collisionless limit and allowing photons to interact with the axion background through the coupling in Eq. (1), the Kadanoff-Baym equations take the simple form

$$\left(g_{\mu\lambda}\Box - g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}(\partial^\alpha a)\partial^\beta\right)_x G^{<,\mu\nu}(x,y) = g^\nu_\lambda \delta(x-y) \ , \tag{4a}$$

$$\left(g_{\mu\nu}\Box - g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}(\partial^\alpha a)\partial^\beta\right)_y G^{<,\lambda\mu}(x,y) = g^\lambda_\nu \delta(x-y) \ . \tag{4b}$$

Where $\Box = \partial^\mu\partial_\mu$ and the suffix $(\ldots)_x$ indicates that all operators act on the coordinate x^μ. The second equation above, in which the operators act on y^μ, is easily obtained by considering the equation of motion obeyed by the greater two-point Green function, renaming $x^\mu \leftrightarrow y^\mu$ and using the property

$$G^{>,\nu\mu}(y,x) = G^{<,\mu\nu}(x,y) \ . \tag{5}$$

The phase space density, also called Wigner function, is defined as the Wigner transform of the Green function

$$G^{<,\mu\nu}(X,q) = \int d^4s\, e^{iq\cdot s} G^{<,\mu\nu}(X+s/2, X-s/2) \ , \tag{6}$$

where $X^\mu = (x^\mu + y^\mu)/2$ and $s^\mu = x^\mu - y^\mu$ are the center and relative coordinates, respectively. Let us recall, as said in the introduction, that we assume that $|\partial_\mu a|/|a| \ll q_\mu$. Adding and subtracting the Wigner transformed equations of motion of Eqs. (4a-4b) one finds the collisionless equations

$$\left(q^2 - \frac{\partial^2}{4}\right) G^{\mu\nu} + \frac{g_{a\gamma}}{2}\left(\epsilon^{\mu\rho\alpha\beta} A_{\alpha\beta} G_\rho^{\ \nu} + \epsilon^{\nu\rho\alpha\beta} A^*_{\alpha\beta} G^\mu_{\ \rho}\right) = 0 \ , \tag{7a}$$

$$(iq\cdot\partial) G^{\mu\nu} + \frac{g_{a\gamma}}{2}\left(\epsilon^{\mu\rho\alpha\beta} A_{\alpha\beta} G_\rho^{\ \nu} - \epsilon^{\nu\rho\alpha\beta} A^*_{\alpha\beta} G^\mu_{\ \rho}\right) = 0 \ , \tag{7b}$$

which have to be complemented with the Wigner transformed gauge conditions of Eqs. (3a-3b)

$$\left(\frac{1}{2}\partial_\mu - iq_\mu\right) G^{\mu\nu} = \left(\frac{1}{2}\partial_\mu + iq_\mu\right) G^{\nu\mu} = 0 \ , \tag{8}$$

and the additional condition, necessary to eliminate the residual ambiguity of the Lorentz gauge

$$u_\mu G^{\mu\nu} = u_\mu G^{\nu\mu} = 0 \ , \tag{9}$$

where u^μ is a time-like vector representing the velocity of the reference frame, satisfying $u^2 = 1$. When writing the above equations, we used the notation $\partial^\mu = \partial/\partial X_\mu$ and dropped the lesser symbol and the arguments of the Wigner function to enlighten the notation, $G^{\mu\nu} = G^{<,\mu\nu}(X,q)$. Additionally, we established

$$A_{\alpha\beta} G_\rho^{\ \nu} \equiv \sum_{n=0}^{\infty} \frac{(-i\Delta)^n}{n!} \partial_\alpha a(X) \left(\frac{1}{2}\partial_\beta - iq_\beta\right) G_\rho^{\ \nu}(X,q) \ , \tag{10}$$

being $\Delta = \frac{1}{2}\frac{\partial}{\partial q}\cdot\frac{\partial}{\partial X}$. For details regarding the derivation of the kinetic equations, we refer to Appendix A. In the absence of an axion background, we reproduce the photon quantum kinetic equations found in Refs. [20–22]. The Keldysh-Schwinger formalism could also be used to reproduce the collision term in the photon transport equation. We will not add a collision term in this reference, but refer to [15] for that purpose.

A. Classical transport equation in the polarization basis

It is convenient to obtain the classical limit of the photon transport equations, which are easier to handle and are enough to study the physics at large scales. For that purpose, one

should carry out a gradient expansion, assuming that variations of the Wigner function on the scale associated to the center coordinate are much less than those on the relative coordinate i.e $\partial_X^\mu \ll \partial_s^\mu$. That this corresponds to a classical limit can be seen by going to momentum space in the relative coordinate and restoring $\hbar$, as one then assumes $\hbar\partial_X^\mu \ll q^\mu$ when performing the gradient expansion. In a thermal plasma, as most photons have momenta of the order of the temperature T, this simply implies to look for variations at scales larger than the inverse of the temperature. Thus, we perform a gradient expansion of Eqs. (7a-7b) and the Wigner function

$$G^{\mu\nu}(X,q) = G^{\mu\nu}_{(0)}(X,q) + G^{\mu\nu}_{(1)}(X,q) + \mathcal{O}(\partial_X^2) \ , \tag{11}$$

where $G^{\mu\nu}_{(0)}(X,q)$ should be understood as the classical limit of the Wigner function, while higher order terms in the gradient expansion correspond to the quantum corrections. As we will only consider the classical limit in this work, we drop the subindex (0), and understand that we only keep the $G^{\mu\nu}_{(0)}$ in the expansion. In the classical limit the kinetic equations read

$$q^2 G^{\mu\nu} - \frac{ig_{a\gamma}}{2}(\partial_\alpha a) q_\beta \left(\epsilon^{\mu\rho\alpha\beta} G_\rho{}^\nu - \epsilon^{\nu\rho\alpha\beta} G^\mu{}_\rho\right) = 0 \ , \tag{12a}$$

$$(iq\cdot\partial) G^{\mu\nu} - \frac{ig_{a\gamma}}{2}(\partial_\alpha a) q_\beta \left(\epsilon^{\mu\rho\alpha\beta} G_\rho{}^\nu + \epsilon^{\nu\rho\alpha\beta} G^\mu{}_\rho\right) = 0 \ , \tag{12b}$$

and the classical Wigner function is constrained by

$$q_\mu G^{\mu\nu} = q_\mu G^{\nu\mu} = u_\mu G^{\mu\nu} = u_\mu G^{\nu\mu} \ . \tag{13}$$

After imposing the condition that the Wigner function is orthogonal to u^μ, the resulting framework is essentially identical to adopting the Coulomb gauge [20]. Yet another way to eliminate the residual gauge ambiguity of the Lorentz gauge is to project the Wigner function into the physical space, using transverse projectors [24].

It is convenient to write the transport equation in a polarization basis. If we introduce a two dimensional basis of polarization vectors (e.g ϵ_a^μ with $a = \{1,2\}$) satisfying, $\epsilon_a^* \cdot \epsilon_b = \delta_{ab}$, and $\epsilon_a \cdot u = \epsilon_a \cdot q = 0$, then the Wigner function can be expressed as

$$G^{\mu\nu}(X,q) = \sum_{a,b=1,2} \epsilon_a^{*\mu} \epsilon_b^\nu \, G^{ab}(X,q) \ . \tag{14}$$

Projecting Eqs. (12a-12b) onto the polarization basis we easily obtain the kinetic equations obeyed by the polarization space components of the Wigner function

$$q^2 G^{ab} - \frac{ig_{a\gamma}}{2}\epsilon_{\mu\nu\alpha\beta}(\partial^\alpha a) q^\beta \left(\epsilon_c^{*\mu}\epsilon_a^\nu G^{cb} - \epsilon_c^\mu \epsilon_b^{*\nu} G^{ac}\right) = 0 \ , \tag{15a}$$

$$(iq\cdot\partial) G^{ab} - \frac{ig_{a\gamma}}{2}\epsilon_{\mu\nu\alpha\beta}(\partial^\alpha a) q^\beta \left(\epsilon_c^{*\mu}\epsilon_a^\nu G^{cb} + \epsilon_c^\mu \epsilon_b^{*\nu} G^{ac}\right) = 0 \ , \tag{15b}$$

where summation over repeated indices should be understood. It is simpler to solve the transport equations in a circular polarization basis, since the polarization space components of the Wigner function then decouple. Explicitly, we introduce the polarization basis vectors $\epsilon_a^\mu = \{\epsilon_+^\mu, \epsilon_-^\mu\}$, characterized by the properties

$$\epsilon_\pm^* \cdot \epsilon_\pm = 0 \ , \quad \epsilon_\pm^* \cdot \epsilon_\mp = 1 \ , \quad (\epsilon_\pm^\mu)^* = \epsilon_\mp^\mu \ . \tag{16}$$

Let us elaborate on the physical interpretation of the Wigner function components in the circular polarization basis. Since $G^{\mu\nu}$ is invariant under basis rotations, i.e $\epsilon_\pm^{\mu'} \to e^{\pm i\theta}\epsilon_\pm^\mu$, the polarization space components of the Wigner function transform as [24]

$$G^{\pm\pm'} \to G^{\pm\pm} \ , \quad G^{\pm\mp'} \to e^{\pm 2i\theta} G^{\pm\mp} \ , \tag{17}$$

which reveals that $G^{\pm\pm}$ and $G^{\pm\mp}$ have null ($s = 0$) and integer ($s = \pm 2$) spin respectively. Moreover, the polarization space components of the Wigner function can be directly related to the Stokes parameters [23], their relation in the circular basis is

$$G^{ab} = \begin{pmatrix} G^{++} & G^{+-} \\ G^{-+} & G^{--} \end{pmatrix} = \begin{pmatrix} G^I - G^V & G^Q - iG^U \\ G^Q + iG^U & G^I + G^V \end{pmatrix} \ . \tag{18}$$

Hence, the diagonal components of the Wigner function $G^{\pm\pm}$ relate to the intensity G^I and degree of circular polarization G^V of the photon ensemble, while the off-diagonal components $G^{\pm\mp}$, decomposed into the Stokes parameters G^Q and G^U, give information on the polarization phases and are related to the so called E and B polarization modes [24].

The kinetic equations for each polarization space component of the Wigner function in the circular basis can be simplified after using the identity [25]

$$i(\epsilon_-^{*\mu}\epsilon_-^\nu - \epsilon_+^{*\mu}\epsilon_+^\nu) = \frac{1}{\kappa}\epsilon^{\mu\nu\alpha\beta} u_\beta q_\alpha \ , \tag{19}$$

where we defined $\kappa = \sqrt{(u\cdot q)^2 - q^2}$. Hence, for the diagonal components G^{++} and G^{--}, corresponding to right and left handed circularly polarized photons respectively, we find

$$\left(q^2 \pm \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]\right) G^{\pm\pm} = 0 \ , \tag{20a}$$

$$iq\cdot\partial\, G^{\pm\pm} = 0 \ . \tag{20b}$$

As for the off-diagonal components G^{+-} and G^{-+}, reflecting the correlation of different polarization in the photon ensemble, we get

$$q^2\, G^{\pm\mp} = 0 \ , \tag{21a}$$

$$\left(iq\cdot\partial \pm \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]\right) G^{\pm\mp} = 0 \ . \tag{21b}$$

The form of Eqs. (20a-21a) suggest that the general structure of the Wigner function is

$$G^{ab}(X,q) = 4\pi\delta(Q^2_{ab})\mathrm{sgn}(u\cdot q) f^{ab}(X,q) \ , \quad \{a,b\} = \{+,-\} \ , \tag{22}$$

where $\mathrm{sgn}(x)$ is the sign function, the quantities Q^2_{ab} govern the dispersion relations and $f^{ab}(X,q)$ are the off-shell distribution functions for each polarization space component of the Wigner function. The transport equations obeyed by the on-shell distribution functions are derived finding the dispersion relations and imposing the resulting on-shell conditions. Let us show how this is realized in the rest frame of the medium $u^\mu = (1,\mathbf{0})$. In this frame, the dispersion relations are obtained as solutions to the equations

$$Q^2_{\pm\pm} = \omega^2 - |\boldsymbol{q}|^2 \pm g_{a\gamma}\left[|\boldsymbol{q}|\partial_0 a + \omega(\hat{\boldsymbol{q}}\cdot\nabla a)\right] = 0 \ , \tag{23a}$$

$$Q^2_{\pm\mp} = \omega^2 - |\boldsymbol{q}|^2 = 0 \ , \tag{23b}$$

and the transport equations read

$$\left(i\omega\partial_0 + i\boldsymbol{q}\cdot\nabla\right) G^{\pm\pm} = 0 \ , \tag{24a}$$

$$\left(i\omega\partial_0 + i\boldsymbol{q}\cdot\nabla \pm g_{a\gamma}\left[|\boldsymbol{q}|\partial_0 a + \omega(\hat{\boldsymbol{q}}\cdot\nabla a)\right]\right) G^{\pm\mp} = 0 \ , \tag{24b}$$

where we used the notation $q^\mu = (\omega,\boldsymbol{q})$ for the photon momentum and defined $\hat{\boldsymbol{q}} = \boldsymbol{q}/|\boldsymbol{q}|$. Then, we see that the presence of the axion induces a different dispersion relation for the diagonal components of the Wigner function G^{++} and G^{--}, while both off-diagonal components G^{+-} and G^{-+} obey a free dispersion relation, unaffected by the axion. Explicitly, solving Eqs. (23a-23b) we find the following dispersion relations

$$\omega_{\pm\pm}(\boldsymbol{q}) \approx |\boldsymbol{q}| \mp \frac{1}{2} g_{a\gamma}(\partial_0 a + \hat{\boldsymbol{q}}\cdot\nabla a) \ , \tag{25a}$$

$$\omega_{\pm\mp}(\boldsymbol{q}) = |\boldsymbol{q}| \ . \tag{25b}$$

The first equation above coincides with the result first found by Harari and Sikivie [26], note that we approximated $\omega_{\pm\pm}(\boldsymbol{q})$ at linear order in $g_{a\gamma}$. There are also negative energy solutions

$$\widetilde{\omega}_{\pm\pm}(\boldsymbol{q}) \approx -|\boldsymbol{q}| \pm \frac{1}{2} g_{a\gamma}(\partial_0 a - \hat{\boldsymbol{q}}\cdot\nabla a) \ , \tag{26a}$$

$$\widetilde{\omega}_{\pm\mp}(\boldsymbol{q}) = -|\boldsymbol{q}| \ . \tag{26b}$$

Imposing the on-shell conditions dictated by Eqs. (25a-25b) onto Eqs. (24a-24b) respectively leads to the transport equations obeyed by the on-shell distribution functions, that we formally

define as

$$f^{\pm\pm}(X,\boldsymbol{q}) = f^{\pm\pm}(X,q)\Big|_{q_0=\omega_{\pm\pm}(\boldsymbol{q})} \,, \tag{27a}$$

$$f^{\pm\mp}(X,\boldsymbol{q}) = f^{\pm\mp}(X,q)\Big|_{q_0=|\boldsymbol{q}|} \,, \tag{27b}$$

for the positive energy solutions. Thus, at first order in the gradient expansion and at linear order in $g_{a\gamma}$ we find

$$\left(i\partial_0 + i\hat{\boldsymbol{q}}\cdot\nabla\right) f^{\pm\pm}(X,\boldsymbol{q}) = 0 \,, \tag{28a}$$

$$\left(i\partial_0 + i\hat{\boldsymbol{q}}\cdot\nabla \pm g_{a\gamma}\left(\partial_0 a + \hat{\boldsymbol{q}}\cdot\nabla a\right)\right) f^{\pm\mp}(X,\boldsymbol{q}) = 0 \,. \tag{28b}$$

Please note that the effective velocity appearing in Eq. (28a), is $\boldsymbol{v}_{\rm eff} = \hat{\boldsymbol{q}} + \mathcal{O}(\partial_X)$, which consistently ignores the effects of the axion, as it would enter as quantum effect, at second order in the gradient expansion. Similar distribution functions and transport equations can be defined for the negative energy solutions.

The conditions we assumed in this work are equivalent to those carried out in the so called eikonal approximation [27]. It has been argued that in the eikonal approximation there is no chiral bending of light in the presence of an axion background in vacuum [27], as the index of refraction of both left and right handed waves is one in this approximation. Please note that in this gradient expansion we found that the dispersion law of the right/left handed photons might be also written down as $(q \pm \frac{g_{a\gamma}}{2}\partial a)^2 \approx 0$, so that the photons travel at the speed of light, as also found in [27].

The photon current associated to the polarized photons can be defined as

$$J^{\mu,ab}(X) = (n^{ab},\, \boldsymbol{j}^{ab}) = \int \frac{d^4q}{(2\pi)^4} q^\mu G^{ab}(X,q) \,. \tag{29}$$

In a plasma at thermal equilibrium, and in the frame at rest with the medium, the right and left handed photon distribution function is the Bose-Einstein distribution function $f_{\rm B}(\omega) = 1/(e^{\omega/T}-1)$. Then, by direct computation one finds that the difference between equilibrium densities of right and left handed photons is proportional to the temporal variation of the axion field

$$n^{++} - n^{--} = \frac{g_{a\gamma}T^2}{3}\partial_0 a + \mathcal{O}(g_{a\gamma}^2) \,. \tag{30}$$

Similarly, one finds

$$\boldsymbol{j}^{++} - \boldsymbol{j}^{--} = \frac{g_{a\gamma}T^2}{9}\nabla a + \mathcal{O}(g_{a\gamma}^2) \,, \tag{31}$$

such that the spatial gradient of the axion field induces a difference in the right and left-handed photon currents. On the ohter hand, from Eq. (28b) we see that the axion induces a rotation on the E and B modes, if there are such polarizations modes. If $X_0^\mu = (t_0, \boldsymbol{X}_0)$ denotes an initial coordinate, then one finds that at a final state $X_f^\mu = (t_f, \boldsymbol{X}_f)$

$$f^{\pm\mp}(X_f, \boldsymbol{q}) = f^{\pm\mp}(X_0, \boldsymbol{q}) \exp\left\{\mp i g_{a\gamma}[a(X_f) - a(X_0)]\right\} \ . \tag{32}$$

As these components have spin 2, the angle of rotation of these modes is $g_{a\gamma}(\Delta a)/2$. This accounts for the rotation of the polarization vector first discussed in [30, 31]. In particular, if the initial configuration only contains E polarization modes, the axion background induces the appeareance of B-modes. This is an effect that has also already been discussed in the literature [28, 29], and that our transport equation properly encodes.

III. COLLECTIVE MODES OF PHOTONS IN AN AXION BACKGROUND

The momentum of most quasiparticles that constitute an electromagnetic plasma is of the order of the equilibrium temperature T and/or the chemical potential μ. Collective modes then emerge as perturbations whose typical momentum, that we denote by Q_μ, is much lesser than those scales, of the order of the Debye mass $Q_\mu \sim m_D \ll T, \mu$. In the previous section, we assumed that the variations of the axion field are much less than the momentum of the photons $|\partial_\mu a|/|a| \ll q_\mu \sim T, \mu$, which allowed us to treat the axion as a background field and the photons as quasiparticles. We also emphasized that if the variations of the axion field are of the same order of the photon momentum, then axions should also be treated as quasiparticles. The interaction between the collective modes and the axion field has also to be addressed differently according to the hierarchy between their typical scales. We will assume that the variations of the axion field are much less than the momentum of the collective modes $|\partial_\mu a|/|a| \ll Q_\mu \sim m_D$, so that the axion still can be effectively described as a background field.

In the absence of an axion background or any CP violating effect in the medium, there is a transverse collective mode that is degenerate and a longitudinal collective mode, the so called plasmon, which is absent in vacuum. The impact of the axion background on the dispersion relations of collective modes is reflected in the dynamics of the dressed propagator $\hat{G}^{\mu\nu}(x, y)$, whose inverse reads in momentum space

$$\hat{G}^{-1}_{\mu\nu}(Q) = -Q^2 g_{\mu\nu} + \Pi_{\mu\nu}(Q) + i g_{a\gamma} \epsilon_{\mu\nu\alpha\beta} (\partial^\alpha a) Q^\beta \ . \tag{33}$$

Within the frame of reference in which the medium is in motion with velocity u^μ, one can establish three independent projectors which are orthogonal to both Q^μ and u^μ as [25]

$$P_T^{\mu\nu} = \tilde{g}^{\mu\nu} - P_L^{\mu\nu} \ , \quad P_L^{\mu\nu} = \frac{\tilde{u}^\mu \tilde{u}^\nu}{\tilde{u}^2} \ , \quad P_P^{\mu\nu} = \frac{i}{\kappa} \epsilon^{\mu\nu\alpha\beta} Q_\alpha u_\beta \ , \tag{34}$$

where

$$\tilde{u}^\mu = \tilde{g}^{\mu\nu} u_\nu \ , \quad \tilde{g}^{\mu\nu} = g^{\mu\nu} - \frac{Q^\mu Q^\nu}{Q^2} \ , \tag{35}$$

and κ was given below Eq. (19) (now one should replace $q^\mu \to Q^\mu$). The projectors in Eq. (34) are called transverse, longitudinal and parity odd projectors respectively. They satisfy the properties

$$\begin{gathered} P_L^2 = 1 \ , \quad P_T^2 = 2 \ , \quad P_P^2 = -2 \ , \\ P_L^{\mu\nu} P_{T,\mu\nu} = P_L^{\mu\nu} P_{P,\mu\nu} = P_T^{\mu\nu} P_{P,\mu\nu} = 0 \ . \end{gathered} \tag{36}$$

As the effect of a parity odd source, either due to the axion background or the medium, is to split the otherwise degenerate right and left circular polarization modes of the photon [25], it is convenient to introduce additional right (+) and left (−) projectors

$$P_h^{\mu\nu} = \frac{1}{2}(P_T^{\mu\nu} + h P_P^{\mu\nu}) \ , \quad h = \pm \ , \tag{37}$$

in terms of which, the polarization tensor can be decomposed as

$$\Pi^{\mu\nu} = \sum_{h=\pm} P_h^{\mu\nu} \left(\Pi_T + h \, \Pi_P\right) - \frac{Q^2}{\kappa^2} P_L^{\mu\nu} \Pi_L \ . \tag{38}$$

Note that if $\Pi_P = 0$ one recovers the usual decomposition of the polarization tensor into its transverse and longitudinal component. The inverse of the dressed propagator in Eq. (33) can be decomposed in terms of the projectors $P_+^{\mu\nu}, P_-^{\mu\nu}$ and $P_L^{\mu\nu}$ too and then inverted, which gives

$$\hat{G}^{\mu\nu}(Q) = -\frac{\kappa^2}{Q^2} \frac{P_L^{\mu\nu}}{\kappa^2 + \Pi_L} - \sum_{h=\pm} \frac{P_h^{\mu\nu}}{Q^2 - \Pi_T + h \left(\Pi_P + \frac{g_{a\gamma}}{\kappa} \left[(Q \cdot \partial a)(u \cdot Q) - Q^2 (u \cdot \partial a)\right]\right)} \ . \tag{39}$$

The poles in the dressed propagator above determine the dispersion relations obeyed by the collective modes within the medium. It is worth mentioning that the longitudinal collective mode remains unaffected by the interaction of photons with the axion background. The projectors used to decompose the dressed propagator can be related to the polarization vectors

$$P_+^{\mu\nu} = -\epsilon_+^{*\mu} \epsilon_+^\nu \ , \quad P_-^{\mu\nu} = -\epsilon_-^{*\mu} \epsilon_-^\nu \ , \quad P_L^{\mu\nu} = -\epsilon_L^\mu \epsilon_L^\nu \ , \tag{40}$$

where we introduced the longitudinal polarization vector as [25]

$$\epsilon_L^\mu = \frac{\tilde{u}^\mu}{\sqrt{-\tilde{u}^2}} \ . \tag{41}$$

Thus, we may write the dressed propagator in terms of the polarization basis vectors

$$\hat{G}^{\mu\nu}(Q) = \frac{\kappa^2}{Q^2}\frac{\epsilon_L^\mu \epsilon_L^\nu}{\kappa^2 + \Pi_L} + \sum_{h=\pm} \frac{\epsilon_h^{*\mu}\epsilon_h^\nu}{Q^2 - \Pi_T + h\left(\Pi_P + \frac{g_{a\gamma}}{\kappa}\left[(Q\cdot\partial a)(u\cdot Q) - Q^2(u\cdot\partial a)\right]\right)} \cdot \quad (42)$$

Let us now assume a QED plasma at very high temperature, such that one can neglect the electron mass. The Debye mass is then $m_D^2 = e^2T^2/3$ where e is the electron charge so that the momentum of the collective modes is of order $Q_\mu \sim eT$. The values of the longitudinal Π_L and transverse Π_T polarization tensors components are then well-known, and given by the so called hard thermal loop (HTL) expressions [10]. Certainly, the axion background also affects the polarization tensor through its coupling to electrons. We ignore those contributions here, but they would be required for a more complete analysis. Then, assuming $\Pi_P = 0$ and moving to the rest frame of the plasma $u^\mu = (1, \mathbf{0})$, the dispersion relations associated to the two transverse collective modes are obtained as solutions to the equations

$$\omega^2 - |\boldsymbol{q}|^2 - \frac{m_D^2\omega^2}{2|\boldsymbol{q}|^2}\left[1 - \left(1 - \frac{|\boldsymbol{q}|^2}{\omega^2}\right)\frac{\omega}{2|\boldsymbol{q}|}\ln\left(\frac{\omega + |\boldsymbol{q}|}{\omega - |\boldsymbol{q}|}\right)\right] \pm g_{a\gamma}\left[|\boldsymbol{q}|\partial_0 a + \omega(\hat{\boldsymbol{q}}\cdot\nabla a)\right] = 0 \ , \quad (43)$$

where we used the notation $Q^\mu = (\omega, \boldsymbol{q})$. The effect of the axion background then comes in modifications of order $eTg_{a\gamma}\partial a$ to the dispersion laws. While these have to be solved numerically, it is possible to find simple analytical solutions in some cases. For instance, in the long wavelength limit $m_D \gg |\boldsymbol{q}|$ we find the solutions

$$\omega_\pm^2 \approx \frac{m_D^2}{3} \mp \frac{g_{a\gamma}m_D}{\sqrt{3}}(\hat{\boldsymbol{q}}\cdot\nabla a) \ , \quad (44)$$

implying that right and left handed circularly polarized collective modes oscillate with different plasma frequencies, and at this expansion order also different from that of the plasmon mode, which is $\omega_L = m_D/\sqrt{3} \equiv \omega_{\rm pl}$ [10]. On the other hand, in the regime $m_D \ll |\boldsymbol{q}| \ll T$ one finds

$$\omega_\pm^2 \approx |\boldsymbol{q}|^2 + m_D^2 \mp g_{a\gamma}(\partial_0 a + \hat{\boldsymbol{q}}\cdot\nabla a) \ , \quad (45)$$

so that the axion produces a different shift on the effective asymptotic masses for right and left handed modes. It is also interesting to study the limit $\omega \ll |\boldsymbol{q}|$, as in this regime, and due to Landau damping, there is an additional family of poles in the transverse modes which are purely imaginary. Let us assume without loss of generality that $\Im(\omega) > 0$ where $\Im$ denotes the imaginary part, then one finds the solutions

$$\omega_\pm = -i\frac{4|\boldsymbol{q}|^3}{\pi m_D^2}\left(1 \mp \frac{g_{a\gamma}\partial_0 a}{|\boldsymbol{q}|}\right) = -i\gamma_\pm \ , \quad (46)$$

This has the same form of the chiral instabilities that are found in chiral media characterized by an imbalance in the population of right and left handed fermions [33]. In fact, it has the same form, after identifying the chiral chemical potential μ_5 with $g_{a\gamma}\partial_0 a/2\alpha$ [33], [34]. The collective modes evolve in time as $\exp(-i\omega_\pm t) \sim \exp(-\gamma_\pm t)$, and they would become unstable if $\gamma_\pm$ becomes negative. As the sign of $\partial_0 a$ can be either positive or negative, this leads to the conditions $\pm g_{a\gamma}\partial_0 a > |\boldsymbol{q}|$, however, in this article we assumed that $|\partial_\mu a|/|a| \ll Q_\mu$, as the axion is treated as a background field. Therefore, taking this assumption into account and due to the smallness of the axion-photon coupling constant $g_{a\gamma}$, we conclude that $\gamma_\pm$ remains positive and there are no unstable modes in this case.

IV. DISCUSSION

We have developed a quantum kinetic theory in the collisionless limit for photons with the presence of an axion background, which is summarized in Eqs.(7-9). Performing a gradient expansion of the operators and the Wigner function (or phase-space distribution), we derived their classical limit and projected the resulting equations on a basis of polarization vectors, yielding Eqs.(12-13). A considerable advantage of this last projection is that the components of the Wigner function in polarization space can be directly related to the Stokes parameters, thus having a clear physical interpretation. Then, using a circular polarization basis of vectors, we derived the transport equations obeyed by the on-shell distribution functions $f^{\pm\pm}$ and $f^{\pm\mp}$ in the rest frame of the medium, given by Eqs.(28a-28b) respectively, which is the central result of this article. Those equations properly encode features of axion electrodynamics, such that right and left handed circular polarized photons obey different dispersion relations, or the phase rotation of the polarization plane, which have been explored before in the literature.

A similar transport equation for photons in an axion background was derived in Ref.[14], there the authors derived a transport equation for the Stokes parameters in a time-dependent axion background. Our treatment is more general and fully covariant, and allows for the incorporation of different sort of corrections. For the comparison with Ref.[14] see Appendix B.

There are several ways in which our work could be extended, for instance, we could include the effects of collisions of photons with the quasiparticles of the thermal bath or compute quantum corrections to our classical transport equations. Another interesting generalization would be to consider that photons propagate through a non flat space time, as considered

in Ref.[24], as the presence of the axion background could provide new sources of B mode polarization.

We have also addressed the effects of the axion background on the photon collective oscillations within the medium through the photon-axion interaction. As expected, the axion background breaks the degeneracy of right and left handed circular polarized collective modes, while the plasmon remains unaffected. The contribution of the axion background is of order $eg_{a\gamma}T(\partial a)$, where ∂a is the scale associated to the variations of the axion field. We also considered limiting cases for the dispersion relations of the transverse collective modes; in the regime $m_D \gg |\boldsymbol{q}|$ the axion produces a shift on the oscillation frequencies of right and left handed polarized collective modes, see Eq.(44), while in the regime $m_D \ll |\boldsymbol{q}| \ll T$ the axion modifies their effective asymptotic masses (cf. Eq.(45)).

It has been argued that specific photon modes in chiral media may not propagate and experience instabilities [9]. However, we have shown in this work that if the chiral media consists of an axion background, both right and left handed photons are propagating modes, since the assumption that the axion field acts as a background prevents those instabilities to occur. We stress that the situation would change if the axions are considered as quanta, interacting with photons through the Lagrangian of Eq.(1). A similar reasoning and conclusion applies for the collective modes, that we have elaborated in Sec.II.

A relevant scenario, that we have not considered in this work, is when the variations of the axion field are comparable to the momentum scale of the collective modes, as in that case interactions between axions and collective modes can occur, leading to interesting phenomena and possible windows for detection [35–38].

ACKNOWLEDGEMENTS

We thank Diego Blas for discussions. This work was supported by Ministerio de Ciencia, Investigación y Universidades (Spain) MCIN/AEI/10.13039/501100011033/ FEDER, UE, under the projects PID2019-110165GB-I00 and PID2022-139427NB-I00, by Generalitat de Catalunya by the project 2021-SGR-171 (Catalonia). This work was also partly supported by the Spanish program Unidad de Excelencia Maria de Maeztu CEX2020-001058-M, financed by MCIN/AEI/10.13039/501100011033.

Appendix A: Kinetic equations

In this section we give some details on how to derive the kinetic equations Eqs. (7a-7b). We start by defining the Wigner transform of the sum and difference of the equations of motion given by Eqs. (4a-4b) as

$$\begin{aligned}(I_{\pm})_{\lambda\nu} = \int d^4s\, e^{iq\cdot s}\big\{ \left(g_{\mu\lambda}\Box - g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}(\partial^\alpha a)\partial^\beta\right)_x G^\mu{}_\nu(x,y) \\ \pm \left(g_{\mu\nu}\Box - g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}(\partial^\alpha a)\partial^\beta\right)_y G_\lambda{}^\mu(x,y)\big\}\ .\end{aligned} \tag{A1}$$

Then, we move from configuration space variables x^μ and y^μ to Wigner space variables X^μ and s^μ using the relations

$$X^\mu = \frac{x^\mu + y^\mu}{2}\ , \quad s^\mu = x^\mu - y^\mu\ , \tag{A2a}$$

$$\partial_x^\mu = \frac{1}{2}\partial_X^\mu + \partial_s^\mu\ , \quad \partial_y^\mu = \frac{1}{2}\partial_X^\mu - \partial_s^\mu\ . \tag{A2b}$$

Doing so, we can write

$$\begin{aligned}(I_{\pm})_{\lambda\nu} = \int d^4s\, e^{iq\cdot s}\Bigg\{\left[g_{\mu\lambda}\left(\partial_s\cdot\partial_X + \partial_s^2 + \frac{1}{4}\partial_X^2\right) - g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}A^{\alpha\beta}(X,s)\right]G^\mu{}_\nu(X+s/2, X-s/2) \\ \pm\left[g_{\mu\nu}\left(-\partial_s\cdot\partial_X + \partial_s^2 + \frac{1}{4}\partial_X^2\right) - g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}A^{\alpha\beta}(X,-s)\right]G_\lambda{}^\mu(X+s/2, X-s/2)\Bigg\}\ .\end{aligned} \tag{A3}$$

Where we defined the following operator, acting on the Wigner function

$$A^{\alpha\beta}(X,s) = \left(\frac{1}{2}\partial_X^\alpha + \partial_s^\alpha\right)a(X+s/2)\left(\frac{1}{2}\partial_X^\beta + \partial_s^\beta\right)\ . \tag{A4}$$

Now we perform a gradient expansion of the axion field

$$a(X+s/2) = \sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{s\cdot\partial_X}{2}\right)^n a(X)\ , \tag{A5}$$

and after the Wigner transformation, we find

$$\begin{aligned}(I_{\pm})_{\lambda\nu} = \Bigg\{\left[g_{\mu\lambda}\left(-iq\cdot\partial_X - q^2 + \frac{1}{4}\partial_X^2\right) - g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}A^{\alpha\beta}(X,q)\right]G^\mu{}_\nu(X,q) \\ \pm\left[g_{\mu\nu}\left(iq\cdot\partial_X - q^2 + \frac{1}{4}\partial_X^2\right) - g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}A^{*,\alpha\beta}(X,q)\right]G_\lambda{}^\mu(X,q)\Bigg\}\ .\end{aligned} \tag{A6}$$

Where now $A^{\alpha\beta}(X,q)$ is given by Eq.(10), in which we neglected the arguments for simplicity. So finally, we find for the dispersion relation

$$\begin{aligned}(I_{+})_{\lambda\nu} = \left(-2q^2 + \frac{1}{2}\partial_X^2\right)G_{\lambda\nu}(X,q) \\ -g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}A^{\alpha\beta}(X,q)G^\mu{}_\nu(X,q) - g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}A^{*,\alpha\beta}(X,q)\,G_\lambda{}^\mu(X,q) = 0\ .\end{aligned} \tag{A7}$$

While for the transport equation

$$\begin{aligned}(I_-)_{\lambda\nu} &= -2i(q\cdot\partial_X)G_{\lambda\nu}(X,q)\\ -g_{a\gamma}\epsilon_{\mu\lambda\alpha\beta}A^{\alpha\beta}(X,q)G^{\mu}{}_{\nu}(X,q) &+ g_{a\gamma}\epsilon_{\mu\nu\alpha\beta}A^{*,\alpha\beta}(X,q)\,G_{\lambda}{}^{\mu}(X,q) = 0\ .\end{aligned} \tag{A8}$$

Now we divide both Eqs. (A7-A8) by a factor of -2, also rising the indices in the Wigner function and the Levi-Civita tensor using the property $\epsilon_{\mu\nu\rho\sigma} = -\epsilon^{\mu\nu\rho\sigma}$ we reach to

$$\begin{aligned}\left(q^2 - \frac{1}{4}\partial_X^2\right)G^{\lambda\nu} - \frac{g_{a\gamma}}{2}\left(\epsilon^{\mu\lambda\alpha\beta}A_{\alpha\beta}G_{\mu}{}^{\nu} + \epsilon^{\mu\nu\alpha\beta}A^*_{\alpha\beta}\,G^{\lambda}{}_{\mu}\right) = 0\ ,\\ (iq\cdot\partial_X)\,G^{\lambda\nu} - \frac{g_{a\gamma}}{2}\left(\epsilon^{\mu\lambda\alpha\beta}A_{\alpha\beta}G_{\mu}{}^{\nu} - \epsilon^{\mu\nu\alpha\beta}A^*_{\alpha\beta}\,G^{\lambda}{}_{\mu}\right) = 0\ ,\end{aligned} \tag{A9}$$

which exactly give Eqs. (7a-7b) after relabeling some indices and using the antisymmetry property of the Levi-Civita tensor.

Appendix B: Classical transport equation in the linear polarization basis

In this section we give the classical kinetic Eqs. (12a-12b) in a linear polarization basis and then write them in terms of the Stokes parameters. We start by introducing a linear polarization basis vectors $\epsilon_a^\mu = \{\epsilon_1^\mu, \epsilon_2^\mu\}$, satisfying $\epsilon_a^*\cdot\epsilon_b = \delta_{ab}$ and $(\epsilon_a^\mu)^* = \epsilon_a^\mu$. The relation between the components of the Wigner function and the Stokes parameters in a linear polarization basis is [23]

$$G^{ab} = \begin{pmatrix} G^{11} & G^{12}\\ G^{21} & G^{22}\end{pmatrix} = \begin{pmatrix} G^I + G^Q & G^U - iG^V\\ G^U + iG^V & G^I - G^Q\end{pmatrix}\ , \tag{B1}$$

so that, after using the identity

$$\epsilon_1^\mu\epsilon_2^\nu - \epsilon_2^\mu\epsilon_1^\nu = \frac{1}{\kappa}\epsilon^{\mu\nu\alpha\beta}u_\beta q_\alpha\ , \tag{B2}$$

the dispersion laws for the Stokes parameters may be written as

$$q^2G^I - \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]G^V = 0\ , \tag{B3a}$$

$$q^2G^V - \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]G^I = 0\ , \tag{B3b}$$

$$q^2G^Q = 0\ , \tag{B3c}$$

$$q^2G^U = 0\ . \tag{B3d}$$

A problem with this formulation, that was not discussed in Ref. [14] is that the equations for G^I and G^V are coupled, so that they can not be treated individually as propagating modes.

Instead, the propagating modes are the combinations $G^I \mp G^V$, corresponding to right and left handed photons respectively, whose dispersion relations can be obtained by solving

$$\left(q^2 \pm \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]\right)\left(G^I \mp G^V\right) = 0 \ , \tag{B4}$$

which yields to the solutions of Eqs. (25a-26a) in the rest frame. The transport equations in the linear polarization basis read

$$(q\cdot\partial)\left(G^I \mp G^V\right) = 0 \ , \tag{B5a}$$

$$(q\cdot\partial)G^Q + \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]G^U = 0 \ , \tag{B5b}$$

$$(q\cdot\partial)G^U - \frac{g_{a\gamma}}{\kappa}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]G^Q = 0 \ . \tag{B5c}$$

The coupled equations for G^Q and G^U can be solved by applying the operator $(q\cdot\partial)$ on each equation and neglecting terms with two derivatives acting on the axion field. Thus, we find

$$(q\cdot\partial)^2 G^Q + \frac{g_{a\gamma}^2}{\kappa^2}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]^2 G^Q = 0 \ , \tag{B6a}$$

$$(q\cdot\partial)^2 G^U - \frac{g_{a\gamma}^2}{\kappa^2}\left[(q\cdot\partial a)(u\cdot q) - q^2(u\cdot\partial a)\right]^2 G^U = 0 \ . \tag{B6b}$$

Now, the general structure of the Stokes parameters is $G^{Q,U}(X,q) = 4\pi\delta(q^2)\mathrm{sgn}(u\cdot q) f^{Q,U}(X,q)$, being $f^{Q,U}(X,q)$ the corresponding off-shell distribution functions. Moving to the rest frame and imposing the on-shell condition, we reach to

$$\left[(v\cdot\partial)^2 + \Omega^2\right] f^Q(X,\boldsymbol{q}) = 0 \ , \tag{B7a}$$

$$\left[(v\cdot\partial)^2 - \Omega^2\right] f^U(X,\boldsymbol{q}) = 0 \ , \tag{B7b}$$

where $f^{Q,U}(X,\boldsymbol{q})$ are the on-shell distribution functions for the positive energy solutions, also we introduced the velocity vector $v^\mu = (1,\hat{\boldsymbol{q}})$ and the frequency $\Omega = g_{a\gamma}(v\cdot\partial a)$. Note that Ω coincides with the frequency defined in Ref. [14] when neglecting the gradient of the axion field (there is a difference of a factor of 2 due to our distinct definition of the dual of the electromagnetic tensor $\tilde{F}^{\mu\nu}$).

[1] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440-1443 (1977) doi:10.1103/PhysRevLett.38.1440

[2] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791-1797 (1977) doi:10.1103/PhysRevD.16.1791

[3] S. Weinberg, Phys. Rev. Lett. **40**, 223-226 (1978) doi:10.1103/PhysRevLett.40.223

[4] F. Wilczek, Phys. Rev. Lett. **40**, 279-282 (1978) doi:10.1103/PhysRevLett.40.279

[5] C. B. Adams, N. Aggarwal, A. Agrawal, R. Balafendiev, C. Bartram, M. Baryakhtar, H. Bekker, P. Belov, K. K. Berggren and A. Berlin, *et al.* [arXiv:2203.14923 [hep-ex]].

[6] A. Caputo and G. Raffelt, [arXiv:2401.13728 [hep-ph]].

[7] A. Sekine and K. Nomura, J. Appl. Phys. **129**, no.14, 141101 (2021) doi:10.1063/5.0038804 [arXiv:2011.13601 [cond-mat.mes-hall]].

[8] D. E. Kharzeev, Prog. Part. Nucl. Phys. **75**, 133-151 (2014) doi:10.1016/j.ppnp.2014.01.002 [arXiv:1312.3348 [hep-ph]].

[9] Z. Qiu, G. Cao and X. G. Huang, Phys. Rev. D **95**, no.3, 036002 (2017) doi:10.1103/PhysRevD.95.036002 [arXiv:1612.06364 [cond-mat.mes-hall]].

[10] M. L. Bellac, Cambridge University Press, 2011, ISBN 978-0-511-88506-8, 978-0-521-65477-7 doi:10.1017/CBO9780511721700

[11] D. F. Litim and C. Manuel, Phys. Rept. **364**, 451-539 (2002) doi:10.1016/S0370-1573(02)00015-7 [arXiv:hep-ph/0110104 [hep-ph]].

[12] V. P. Silin, Sov. Phys. JETP **11**, no.5, 1136-1140 (1960)

[13] H. A. Weldon, Phys. Rev. D **26**, 1394 (1982) doi:10.1103/PhysRevD.26.1394

[14] S. Shakeri and F. Hajkarim, JCAP **04** (2023), 017 doi:10.1088/1475-7516/2023/04/017 [arXiv:2209.13572 [hep-ph]].

[15] J. I. McDonald, B. Garbrecht and P. Millington, JCAP **12**, 031 (2023) doi:10.1088/1475-7516/2023/12/031 [arXiv:2307.11812 [hep-ph]].

[16] P. Danielewicz, Annals Phys. **152**, 239-304 (1984) doi:10.1016/0003-4916(84)90092-7

[17] S. Mrowczynski, B. Schenke and M. Strickland, Phys. Rept. **682** (2017), 1-97 doi:10.1016/j.physrep.2017.03.003 [arXiv:1603.08946 [hep-ph]].

[18] D. Vasak, M. Gyulassy and H. T. Elze, Annals Phys. **173** (1987), 462-492 doi:10.1016/0003-4916(87)90169-2

[19] L. P. Kadanoff, G. Baym, The Benjamin/Cumming Publishing Company, Inc., Reading-Massachussetts, 1981.

[20] K. Hattori, Y. Hidaka, N. Yamamoto and D. L. Yang, JHEP **02** (2021), 001 doi:10.1007/JHEP02(2021)001 [arXiv:2010.13368 [hep-ph]].

[21] X. G. Huang, P. Mitkin, A. V. Sadofyev and E. Speranza, JHEP **10**, 117 (2020) doi:10.1007/JHEP10(2020)117 [arXiv:2006.03591 [hep-th]].

[22] S. Lin, Phys. Rev. D **105**, no.7, 076017 (2022) doi:10.1103/PhysRevD.105.076017 [arXiv:2109.00184 [hep-ph]].

[23] J. D. Jackson, Wiley, 1998, ISBN 978-0-471-30932-1

[24] M. Beneke and C. Fidler, Phys. Rev. D **82** (2010), 063509 doi:10.1103/PhysRevD.82.063509 [arXiv:1003.1834 [astro-ph.CO]].

[25] J. F. Nieves and P. B. Pal, Phys. Rev. D **39** (1989), 652 [erratum: Phys. Rev. D **40** (1989), 2148] doi:10.1103/PhysRevD.39.652

[26] D. Harari and P. Sikivie, Phys. Lett. B **289** (1992), 67-72 doi:10.1016/0370-2693(92)91363-E

[27] D. Blas, A. Caputo, M. M. Ivanov and L. Sberna, Phys. Dark Univ. **27**, 100428 (2020) doi:10.1016/j.dark.2019.100428 [arXiv:1910.06128 [hep-ph]].

[28] G. C. Liu, S. Lee and K. W. Ng, Phys. Rev. Lett. **97**, 161303 (2006) doi:10.1103/PhysRevLett.97.161303 [arXiv:astro-ph/0606248 [astro-ph]].

[29] G. C. Liu and K. W. Ng, Phys. Dark Univ. **16**, 22-25 (2017) doi:10.1016/j.dark.2017.02.004 [arXiv:1612.02104 [astro-ph.CO]].

[30] S. M. Carroll, G. B. Field and R. Jackiw, Phys. Rev. D **41**, 1231 (1990) doi:10.1103/PhysRevD.41.1231

[31] S. M. Carroll, Phys. Rev. Lett. **81**, 3067-3070 (1998) doi:10.1103/PhysRevLett.81.3067 [arXiv:astro-ph/9806099 [astro-ph]]. Copy to ClipboardDownload

[32] C. Pitrou, Astropart. Phys. **125** (2021), 102494 doi:10.1016/j.astropartphys.2020.102494 [arXiv:1902.09456 [astro-ph.CO]].

[33] Y. Akamatsu and N. Yamamoto, Phys. Rev. Lett. **111** (2013), 052002 doi:10.1103/PhysRevLett.111.052002 [arXiv:1302.2125 [nucl-th]].

[34] S. Carignano and C. Manuel, Phys. Rev. D **99**, no.9, 096022 (2019) doi:10.1103/PhysRevD.99.096022 [arXiv:1811.06394 [hep-ph]].

[35] G. G. Raffelt, Phys. Rev. D **37** (1988), 1356 doi:10.1103/PhysRevD.37.1356

[36] N. V. Mikheev, D. A. Rumyantsev and Y. E. Shkol'nikova, Zh. Eksp. Teor. Fiz. **90** (2009), 668-671 doi:10.1134/S0021364009210036

[37] A. Caputo, A. J. Millar and E. Vitagliano, Phys. Rev. D **101** (2020) no.12, 123004 doi:10.1103/PhysRevD.101.123004 [arXiv:2005.00078 [hep-ph]].

[38] M. D. Redell, S. Mukherjee and W. C. Lee, Phys. Rev. B **93** (2016) no.24, 241201 doi:10.1103/PhysRevB.93.241201 [arXiv:1603.06810 [cond-mat.str-el]].

PROCEEDINGS OF SCIENCE

Ultra-light Dark Matter Search with Astrophysical Neutrino Flavour

Carlos A. Argüelles,[a] **Kareem Farrag**[b] **and Teppei Katori**[c,*]

[a] *Harvard University, Department of Physics and Laboratory for Particle Physics and Cosmology, Cambridge, MA 02138, USA*

[b] *Chiba University, Department of Physics and Institute for Global Prominent Research, Chiba, 263-8522, Japan*

[b] *King's College London, Department of Physics, London, WC2R 2LS, UK*

E-mail: carguelles@g.harvard.edu, k.r.h.farrag@chiba-u.jp, teppei.katori@kcl.ac.uk

The ultra-light dark matter is a new class of dark matter candidates. Unlike traditional dark matter particle candidates, the ultra-light dark matter behaves like a classical field, which saturates the entire Milky Way galaxy with a coherent oscillation. If such dark matter exists and couples with neutrinos, properties of astrophysical neutrinos propagating in the Milky Way would be modified and detectable by neutrino telescopes such as IceCube. Meantime, IceCube looked for quantum-gravity-motivated effects from the astrophysical neutrino flavour information. IceCube did not find evidence of new physics, and they set limits on neutrino - Lorentz violating field couplings. Here, we investigate if these results can be applied to limit the ultra-light dark matter couplings with neutrinos. It is found that the strong limits of neutrino-dark matter couplings in low dark matter mass regions can be obtained in this approach.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*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:2402.18126v1 [hep-ph] 28 Feb 2024

1. IceCube Lorentz violation limits

Recently, the IceCube collaboration set the most stringent limits on Lorentz violating effects in the neutrino sector using the astrophysical neutrino flavour data [1, 2]. For this, 7.5 years of the high-energy starting event (HESE) data sample is used [3]. Since the data statistics are limited, astrophysical neutrino flavour is represented by the flavour ratio [4–7] where the fraction of electron, muon, and tau neutrinos ($\nu_e : \nu_\mu : \nu_\tau$) are fit from the data. They are understood as the sum of neutrino and anti-neutrino contributions with integrating the assumed astrophysical neutrino spectrum. Couplings between neutrinos and Lorentz violating spacetime, if existed, can deviation the flavour data from the standard astrophysical neutrino flavour prediction to anomalous value [8]. The analysis used effective operators derived from the Standard-Model Extension (SME) framework [9] assuming all effects are isotropic. For example, under the SME, the effective Hamiltonian of astrophysical neutrinos with dimension-three Lorentz violating operators can be written explicitly in the flavour basis,

$$H_{eff} \sim \frac{1}{2E}\cdot\begin{pmatrix} m^2_{ee} & m^2_{e\mu} & m^2_{\tau e} \\ m^{2*}_{e\mu} & m^2_{\mu\mu} & m^2_{\mu\tau} \\ m^{2*}_{\tau e} & m^{2*}_{\mu\tau} & m^2_{\tau\tau} \end{pmatrix} + \begin{pmatrix} \mathring{a}^{(3)}_{ee} & \mathring{a}^{(3)}_{e\mu} & \mathring{a}^{(3)}_{\tau e} \\ a^{(3)}_{e\mu} & a^{(3)}_{\mu\mu} & a^{(3)}_{\mu\tau} \\ a^{(3)}_{\tau e} & a^{(3)}_{\mu\mu} & \mathring{a}^{(3)}_{\tau\tau} \end{pmatrix}. \quad (1)$$

Here, the first term is the neutrino mass matrix [10], and the second term is the dimension-three isotropic Lorentz violating term [11]. Although they did not find the evidence of presence of these terms in the current IceCube flavor data, they set order magnitude stronger limits on certain types of couplings. For example, the dimension-three Lorentz violating operator in $\tau - \tau$ sector is constrained down to $\mathring{a}^{(3)}_{\tau\tau} < 2\times10^{-26}$ GeV with the Bayes factor greater than 10. Interestingly, these limits can be re-interpreted to constrain other physics [12]. In these proceedings, we investigate the flavour effect of neutrino - dark matter couplings [13–24] from the IceCube Lorentz violation limits.

2. Ultra-Light Dark Matter

Ultra-light dark matter is a class of dark matter models where the mass of dark matter particles is as low as 10^{-22} eV and they can still be the dominant source of dark matter [25, 26]. The searches of the weakly interaction massive particles (WIMPs) use elastic scatterings between WIMPs and nuclear targets [27–29]. On the other hand, ultra-light dark matter behaves as a classical field and it may saturate in our galaxy with maintaining the coherent oscillation. In this scenario, if they interact with neutrinos, the propagation of astrophysical neutrinos in our galaxy may be affected. Although interactions are so small and it cannot affect the energy spectrum, the effect may be detectable from anomalous flavour mixings inconsistent with the standard astrophysical scenario.

To set up our analysis, we make several assumptions. First, we assume dark matter fields and astrophysical neutrino couplings are described by vector or axial vector operators [13–21], or dark matter fields themselves are vector or axial vector fields (such as neutrinophilic axion [22], vector dark matter [23, 24]), because these interactions make matter potential terms in vacuum and they can be identified with Lorentz violating operators in Eq. 1. Second, we assume a constant dark matter density, $\rho = 0.3$ GeV/m^3, across the whole Milky Way galaxy to simplify the problem. The

IceCube HESE astrophysical neutrino data sample is a diffuse sample and it contains neutrinos from all directions. The ultra-light dark matter search with neutrinos can be improved by assuming more realistic dark matter density profiles, such as NFW potential [30]. Third, we assume the flavour conversion is adiabatic [31]. This condition is energy-dependent and it is broken for fast oscillating dark matter field. Under these conditions, we can write down the dimension-three $\tau - \tau$ sector isotropic SME coefficient in terms of the neutrino - dark matter coupling. By using the dark matter with mass m_a and coupling with tau neutrino g_τ,

$$\mathring{a}^{(3)}_{\tau\tau} = g_\tau \frac{\sqrt{2\rho}}{m_a} \cos(m_a t) \ . \tag{2}$$

And fourth, we assume the oscillation of the ultra-light dark matter field is fast enough compared with the time scale for neutrinos to cross the Milky Way. The mass of the dark matter with the de Broglie length of the radius of the Milky Way is around 4×10^{-28} eV. Although such mass is not forbidden as a candidate of the galactic dark matter [26], we focus on the region greater than 10^{-22} eV where the ultra-light dark matter can be the dominant source of the galactic dark matter. In this scenario, astrophysical neutrinos crossing the Milky Way experience many oscillations of the dark matter field. Furthermore, neutrinos propagate different distances in the Milky Way, and the phase of the oscillation is different for neutrinos detected by IceCube. The combination of these effects makes the couplings between neutrinos and dark matter fields to be averaged out, and the expectation value becomes zero. However, interactions of neutrinos and the oscillating dark matter fields spread the astrophysical neutrino flavour at arrival and they are still detectable [20, 32]. Now, we investigate this effect on the flavour triangle of the IceCube astrophysical neutrino flavour.

3. Flavor ratio from neutrino - dark matter coupling

Fig. 1 represent the flavour triangle of the IceCube HESE sample [3, 7]. Each point in this space shows the fraction of electron, muon, and tau neutrinos ($\nu_e : \nu_\mu : \nu_\tau$). A "bowtie" shape at the central region around $(\nu_e : \nu_\mu : \nu_\tau) \sim (1 : 1 : 1)$ is the expected flavour ratio from the standard astrophysical neutrino scenario where flavour ratios at the productions are all possible combinations of electron neutrinos and muon neutrinos, or $(\nu_e : \nu_\mu : \nu_\tau) \sim (x : 1 - x : 0)$ with $x = [0, 1]$. In other words, the standard astrophysical neutrino production scenario corresponds to the right side of this triangle plot. Then, the neutrino standard model Hamiltonian, Eq. 1 without the second term, will bring them to this bowtie region by the standard flavour mixings. If the data flavour ratio is outside of this region, there is a chance that is due to new physics.

The shaded regions represent the 68% and 95% data contours from the IceCube HESE 7.5-yr data analysis [3, 7]. These contours enclose this astrophysical standard scenario phase space completely, meaning we would not find any new physics which deviates slightly from there even if it exist. Therefore, future improved analysis, larger statistics and better flavour sensitivity, will be essential to look for new physics through the astrophysical neutrino flavour [12].

By adding the second term of Eq. 1, the predicted flavour distribution is drastically changed. Fig. 2 includes nonzero $\mathring{a}^{(3)}_{\tau\tau}$ matrix element with the size of the IceCube limit (2×10^{-26} GeV) to see the effect in the flavour triangle plot. The predicted flavour distribution changes from a bowtie

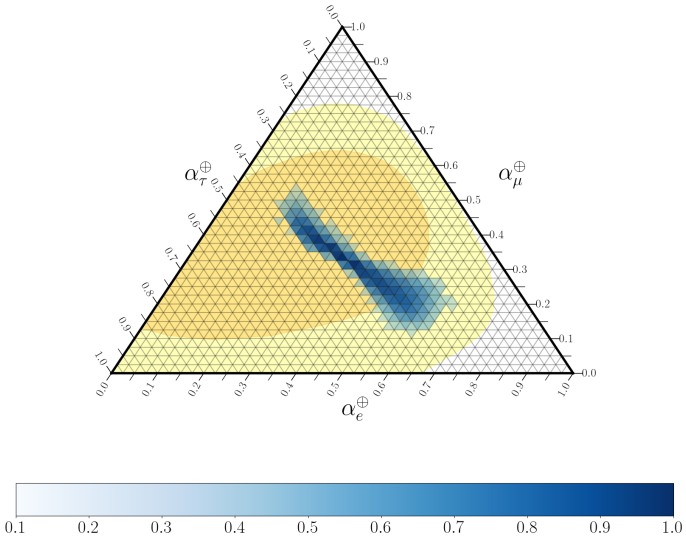


Figure 1: Astrophysical neutrino flavour without new physics. Each point in this figure is the flavour ratio of astrophysical neutrinos at detection. The histogram is the expected flavour ratio for the standard neutrino model with the standard astrophysical neutrino production scenarios, and it is enclosed by the 68% and 95% contours from the IceCube data [7].

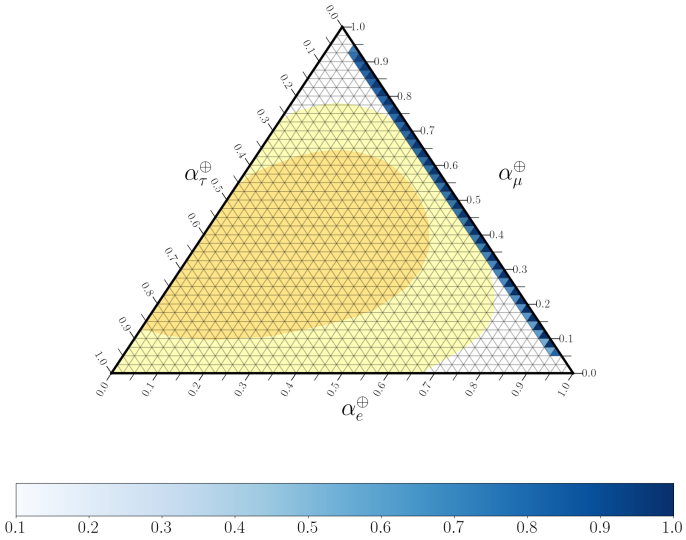


Figure 2: Astrophysical neutrino flavour with static dark matter potential, $g_\tau\sqrt{2\rho}/m_a = 2\times10^{-26}$ GeV. The value of dark matter is chosen because this value was rejected from the IceCube Lorentz violation analysis [1]. Unlike Fig. 1, predicted astrophysical neutrino flavour is mostly outside of the 68% and 95% data contours [7].

shape in Fig. 1 to a line on the right side, and most of the points go to the outside of the data contours. This is consistent that a precise analysis found that this scenario is rejected in the IceCube with Bayes factor > 10.

In terms of ultra-light dark matter, Eq. 2, Fig. 2 corresponds to the presence of neutrino - dark

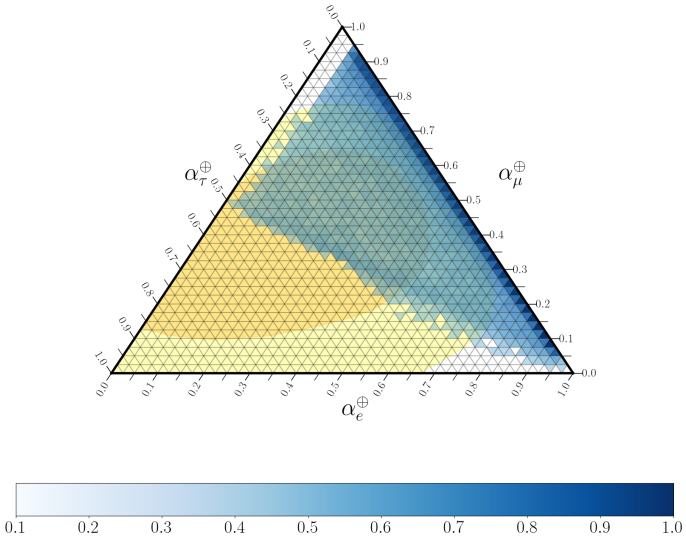


Figure 3: Astrophysical neutrino flavour with oscillating dark matter potential, Eq. 3, with $g_\tau\sqrt{2\rho}/m_a = 2\times10^{-26}$ GeV. The value of the dark matter field amplitude is chosen from the IceCube Lorentz violation analysis [1]. Unlike Fig. 2, some of predicted astrophysical neutrino flavour predictions are enclosed by the 68% and 95% data contours [7].

matter coupling, with the dark matter amplitude $g_\tau\sqrt{2\rho}/m_a = 2\times10^{-26}$ GeV but no time-varying oscillation. The interpretation of the IceCube data is that such dark matter model is rejected, therefore, we set the limit of the combination of the neutrino - dark matter coupling and the dark matter mass,

$$\frac{g_\tau}{m_a} < 10^{-14}\ eV^{-1}\ . \tag{3}$$

This limit is stronger than the current and future limits from accelerator-based, atmospheric, and solar neutrino experiments [23, 24]. This is because the expected effect scales with energy, and the high-energy astrophysical neutrinos have higher energy than other known neutrinos and strongly constrain neutrino - dark matter couplings. However, such limit may not be applicable if the ultra-light dark matter field is time varying within our galaxy.

Fig 3 shows the flavour ratio with Eq. 1, but here nonzero $\mathring{a}^{(3)}_{\tau\tau}$ is replaced with Eq. 2 to simulate the time-dependent neutrino - dark matter coupling. If we include the time variation of the dark matter field, since the potential is zero on average, one would expect that the dark matter potential will not affect the astrophysical neutrino flavour ratio. However, the time variation of the field adds the smearing feature on the flavour triangle [20, 32], and it spreads the predicted flavour ratio distribution in this triangle. The difference between Fig. 2 and Fig. 3 means these two models receive different constraints from the astrophysical flavour data. In general, we expect a weaker limit for the oscillating dark matter models than the static ultra-light dark matter models.

4. Outlook

Here, we briefly present a new way to look for ultra-light dark matter from the IceCube astrophysical neutrino flavour. In particular, we investigate the way to translate the recent IceCube Lorentz violation limits [1, 2] to the limits of neutrino - ultra-light dark matter couplings. The expected limits from this analysis are likely to be very strong, and future neutrino telescopes, such as IceCube-Gen2 [33, 34], can perform an improved ultra-light dark matter search with astrophysical neutrino flavour.

Acknowledgements

CAA is supported by the Faculty of Arts and Sciences of Harvard University, and the Alfred P. Sloan Foundation, USA. KF is supported by the JSPS and KAKENHI, Japan. TK is supported by UKRI STFC, UK. We thank the organizers of the ICRC2023 conference for hosting this largest-ever ICRC conference with great fun.

References

[1] **IceCube** Collaboration, R. Abbasi *et al.* *Nature Phys.* **18** no. 11, (2022) 1287–1292.

[2] **IceCube** Collaboration, C. A. Argüelles, and others *PoS* **ICRC2023** (these proceedings) 1225.

[3] **IceCube** Collaboration, R. Abbasi *et al.* *Phys. Rev. D* **104** (2021) 022002.

[4] **IceCube** Collaboration, M. G. Aartsen *et al.* *Phys. Rev. Lett.* **114** no. 17, (2015) 171102.

[5] **IceCube** Collaboration, M. G. Aartsen *et al.* *Astrophys. J.* **809** no. 1, (2015) 98.

[6] **IceCube** Collaboration, M. G. Aartsen *et al.* *Phys. Rev. D* **99** no. 3, (2019) 032004.

[7] **IceCube** Collaboration, R. Abbasi *et al.* *Eur. Phys. J. C* **82** no. 11, (2022) 1031.

[8] C. A. Argüelles, T. Katori, and J. Salvado *Phys. Rev. Lett.* **115** (Oct, 2015) 161303.

[9] A. Kostelecky and M. Mewes *Phys. Rev. D* **85** (2012) 096005.

[10] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou *JHEP* **09** (2020) 178.

[11] **IceCube** Collaboration, M. G. Aartsen *et al.* *Nature Phys.* **14** no. 9, (2018) 961–966.

[12] C. A. Argüelles *et al.* *Eur. Phys. J. C* **83** no. 1, (2023) 15.

[13] A. Berlin *Phys. Rev. Lett.* **117** (Nov, 2016) 231801.

[14] M. M. Reynoso and O. A. Sampayo *Astropart. Phys.* **82** (2016) 10–20.

[15] P. F. de Salas, R. A. Lineros, and M. Tórtola *Phys. Rev. D* **94** no. 12, (2016) 123001.

[16] R. W. Rasmussen, L. Lechner, M. Ackermann, M. Kowalski, and W. Winter *Phys. Rev. D* **96** no. 8, (2017) 083018.

[17] Y. Farzan and S. Palomares-Ruiz *Phys. Rev. D* **99** no. 5, (2019) 051702.

[18] A. Y. Smirnov and X.-J. Xu *JHEP* **12** (2019) 046.

[19] S. Karmakar, S. Pandey, and S. Rakshit *JHEP* **10** (2021) 004.

[20] M. Losada, Y. Nir, G. Perez, and Y. Shpilman *JHEP* **04** (2022) 030.

[21] T. Gherghetta and A. Shkerin. arXiv:2305.06441.

[22] G.-Y. Huang and N. Nath *Eur. Phys. J. C* **78** no. 11, (2018) 922.

[23] D. Brzeminski, S. Das, A. Hook, and C. Ristow. arXiv:2212.05073.

[24] G. Alonso-Álvarez, K. Bleau, and J. M. Cline *Phys. Rev. D* **107** no. 5, (2023) 055045.

[25] W. Hu, R. Barkana, and A. Gruzinov *Phys. Rev. Lett.* **85** (2000) 1158–1161.

[26] D. J. E. Marsh *Phys. Rept.* **643** (2016) 1–79.

[27] **XENON** Collaboration, E. Aprile *et al.* *Phys. Rev. Lett.* **121** no. 11, (2018) 111302.

[28] **PandaX-4T** Collaboration, Y. Meng *et al.* *Phys. Rev. Lett.* **127** no. 26, (2021) 261802.

[29] **LZ** Collaboration, J. Aalbers *et al.* arXiv:2207.03764.

[30] J. F. Navarro, C. S. Frenk, and S. D. M. White *Astrophys. J.* **462** (1996) 563–575.

[31] E. Akhmedov, J. Kopp, and M. Lindner *JCAP* **09** (2017) 017.

[32] L. Hamaide, H. Müller, and D. J. E. Marsh *Phys. Rev. D* **106** (Dec, 2022) 123509.

[33] **IceCube-Gen2** Collaboration, M. G. Aartsen *et al.* *J. Phys. G* **48** no. 6, (2021) 060501.

[34] **IceCube-Gen2** Collaboration, A. Ishihara, and others *PoS* **ICRC2023** (these proceedings) 994.

The collisional energy loss of a heavy-quark in a semi-quark-gluon plasma

Qianqian Du[a,b], Mudong Du[a], and Yun Guo[a,b,*]
[a] *Department of Physics, Guangxi Normal University, Guilin, 541004, China*
[b] *Guangxi Key Laboratory of Nuclear Physics and Technology, Guilin, 541004, China*

By utilizing a background field effective theory, we compute the collisional energy loss of a heavy-quark moving through a semi-quark-gluon plasma characterized by nontrivial holonomy for Polyakov loops. We consider the elastic scatterings between the incident heavy-quark and the thermal partons with both hard and soft momentum transfers. Our results show that the heavy-quark energy loss is significantly suppressed in the semi-quark-gluon plasma due to a background field that is self-consistently generated in the effective theory. On the other hand, the suppression strongly depends on the temperature of the plasma which becomes negligible above $2 \sim 3$ times the critical temperature. For a realistic coupling constant, ignoring a relatively weak dependence on the heavy-quark velocity, the suppression on the collisional energy loss can be approximated by an overall factor determined solely by the background field.

arXiv:2402.18004v1 [hep-ph] 28 Feb 2024

I. INTRODUCTION

Under extremely hot and dense conditions created in the heavy-ion experiments, the confined states of hadrons are broken, resulting in a new state of matter known as quark-gluon plasma (QGP) [1, 2]. Studying the equation of state is a crucial aspect of understanding the physical properties of the strongly interacting matter. At asymptotically high temperatures, perturbative QCD works very well. Especially, the hard-thermal-loop (HTL) resummed perturbation theory is believed to be a powerful tool which provides reliable predictions on the thermodynamics of the QGP down to several times the critical temperature T_d [3, 4]. On the other hand, the hadron resonance gas model can be used to simulate the confined hadronic phase at temperatures below T_d. However, there exists an intermediate region termed as semi-QGP [5], from the critical temperature to a few times that, where the physics is of particular interest. For high energy heavy-ion experiments carried out at the Large Hadron Collider (LHC), and especially at Relativistic Heavy Ion Collider (RHIC), the temperatures probed are not far from the critical temperature. Unfortunately, neither the hadron resonance gas nor the HTL resummed perturbation theory can be a reliable theoretical means to explore this intermediate region.

The partial deconfinement in a semi-QGP is characterized by nontrivial holonomy for Polyakov loops. As the order parameter for the deconfining phase transition in $SU(N)$ gauge theories, the Polyakov loop is non-zero but less than unity in the semi-QGP region according to the Lattice simulations [6, 7]. This behavior can be realized by introducing a classical background field for the time-like component of the vector potential A_0. The resulting matrix models [8, 9], which can be considered as a background field effective theory, give rise to a deconfining phase transition through a competition between two different contributions making up the effective potential in this model. One is a perturbative term, favoring the completely deconfined vacuum, and the other is a non-perturbative term, driving the system to confinement. In addition, the predicted thermodynamics is also in a good agreement with the Lattice data. Given the success made by the matrix models, it is reasonable to adopt such an effective theory to study some other physical quantities in a semi-QGP.

In this work, we will focus on the collisional energy loss $-dE/dx$ of an energetic heavy-quark moving through a hot and dense QCD plasma which is only partially deconfined. As observed in the relativistic heavy-ion experiments, the suppression of high transverse momentum hadrons can be explained by the energy loss of high energy partons in the plasma. This phenomenon is called the jet quenching, an excellent hard probe which provides a novel window to unravel the fascinating properties of the deconfined nuclear matter [10–13]. Furthermore, unlike the light quarks which lose energy mainly through the gluon radiation, the collisional energy loss of a heavy-quark is comparable to the radiated energy loss because of the dead-cone effect related to the large quark mass [14–17]. Therefore, investigating the collisional energy loss induced by the elastic scatterings between the incident heavy-quark and thermal medium partons can help us to understand the measured heavy-flavor spectra in nucleus-nucleus collisions. After the first detailed calculation of $-dE/dx$ for a heavy-quark in the QCD plasma made by Braaten and Thoma [18, 19], many attentions have been paid to this research topic and there were various theoretical developments over the past thirty years, see [20–30] and references therein.

* Corresponding author: yunguo@mailbox.gxnu.edu.cn

It is certainly interesting to consider the collisional energy loss of a heavy-quark by using a background field effective theory when the non-perturbative physics plays an important role and may affect the final results significantly. Our calculation aims to provide quantitative estimates on $-dE/dx$ in a semi-QGP and also to assess the influence of a non-zero background field that is self-consistently generated within this effective theory. As a first step towards a full QCD analysis, we only concentrate on a gluonic plasma because including dynamical quarks in the effective theory is currently an open question. The rest of the paper is organized as follows. In Sec. II, we give the basic definition of the collisional energy loss in QCD and review the theoretical approaches to calculate the hard and soft contributions to $-dE/dx$ in the perturbation theory. In Sec. III, after introducing the background field effective theory of a semi-QGP, we discuss how the resummation of the usual hard thermal loops will be modified in the presence of a background field. In addition, the hard and soft contributions to the energy loss are separately derived and the corresponding analytical results are obtained in the weak coupling limit. We show our numerical results of $-dE/dx$ for a charm and bottom quark in Sec. IV where detailed discussions on the heavy-quark velocity/momentum dependence of the collisional energy loss and its suppression in the semi-QGP are also presented. Finally, a summary and outlook can be found in Sec. V.

II. THE DEFINITION OF COLLISIONAL ENERGY LOSS IN QCD

When a high energy heavy-quark propagates through a plasma consisting of light quarks and gluons in thermal equilibrium, it may lose energy by interacting with the thermal partons. The rate of energy loss per distance travelled is given by [18]

$$-\frac{dE}{dx} \;=\; \frac{1}{v}\int_{m_Q}^{\infty} dE^{'}(E-E^{'})\frac{d\Gamma(E)}{dE^{'}}\,, \tag{1}$$

where E and $\mathbf{v}=\mathbf{p}/E$ denote the energy and velocity of the incident heavy-quark with mass m_Q. In addition, the interaction rate $\Gamma(E)$ reads

$$\begin{aligned}\Gamma(E) \;&=\; \frac{1}{2E}\int\frac{d^3\mathbf{p}'}{(2\pi)^3 2E'}\int\frac{d^3\mathbf{k}}{(2\pi)^3 2k}n(k)\int\frac{d^3\mathbf{k}'}{(2\pi)^3 2k'}\big(1-n(k')\big)\\ &\times\; (2\pi)^4\delta^4(P+K-P'-K')\left(\frac{1}{2}\sum_{\text{spins}}\frac{1}{3}\sum_{\text{color}}|\mathcal{M}|^2\right).\end{aligned} \tag{2}$$

In the above equation, $P=(E,\mathbf{p})$ and $P'=(E',\mathbf{p}')$ are the four-momenta of the incoming and outgoing heavy-quark, respectively. The four-momenta of the medium partons which scatter off of the heavy-quark are denoted by $K=(k,\mathbf{k})$ and $K'=(k',\mathbf{k}')$. For quark-gluon scattering, the phase space is weighted by a Bose-Einstein distribution $n(k)=(e^{k/T}-1)^{-1}$ and a factor $1-n(k')$ accounting for the Bose enhancement.

The matrix element $\mathcal{M}$ can be determined by computing the corresponding Feynman diagrams for the elastic scattering processes. It is straightforward to obtain the tree-level contributions as shown in Fig. 1 where, however, an infrared divergence stemming from the t-channel diagram emerges when integrating over the transferred momentum $\mathbf{q}=\mathbf{p}-\mathbf{p}'$. Following Braaten and Yuan [31], one can introduce an arbitrary scale q^* for the momentum transferred by which one defines the so-called hard and soft contributions to $-dE/dx$. The hard contributions account for scatterings with momentum transfer larger than q^* and can be expressed as

$$\begin{aligned}-\Big(\frac{dE}{dx}\Big)^{Qg(t)}_{\text{hard}} \;&=\; \frac{(4\pi)^3\alpha_s^2}{v}\int\frac{d^3\mathbf{k}}{(2\pi)^3}\frac{n(k)}{k}\int\frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{1+n(k')}{k'}\delta(\omega-\mathbf{v}\cdot\mathbf{q})\\ &\times\; \theta(q-q^*)\,\frac{4\omega}{(\omega^2-q^2)^2}\left[(k-\mathbf{v}\cdot\mathbf{k})^2+\frac{1-v^2}{2}(\omega^2-q^2)\right],\end{aligned} \tag{3}$$

where $\alpha_s=g^2/(4\pi)$ and $\omega\equiv E-E'$. Unless otherwise stated, we set the number of colors $N=3$ in the following. For the quark-gluon scattering, contributions from the s- and u-channels are finite for small momentum exchange, therefore, there is no need to introduce the cutoff q^*. The corresponding result reads

$$-\Big(\frac{dE}{dx}\Big)^{Qg(s+u)}_{\text{hard}} = \frac{(4\pi)^3\alpha_s^2}{2v}\int\frac{d^3\mathbf{k}}{(2\pi)^3}\frac{n(k)}{k}\int\frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{1+n(k')}{k'}\delta(\omega-\mathbf{v}\cdot\mathbf{q})\frac{\omega(1-v^2)^2}{(k-\mathbf{v}\cdot\mathbf{k})^2}\,. \tag{4}$$

To obtain the above equations, we consider the thermal partons with typical momentum k, $k'\sim T$ and assume the mass and velocity of the heavy-quark satisfy $m_Q\gg T$ and $v\gg T/E$. Consequently, contributions suppressed by

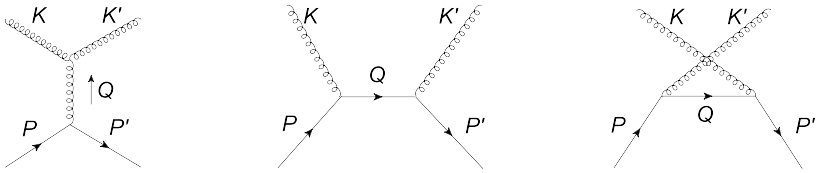


FIG. 1. Tree level Feynman diagrams for the scattering process $Qg \to Qg$.

T/m_Q or T/p have been dropped in $|\mathcal{M}|^2$. Notice that under the above assumptions, interferences between t- and s/u-channel become negligible and the energy conservation in Eq. (2) reduces to $\delta(\omega - \mathbf{v}\cdot\mathbf{q})$. Further simplification can be made by realizing the fact that in Eqs. (3) and (4), terms involving the product $n_B(k)n_B(k')$ are odd under the interchange $\mathbf{k} \leftrightarrow \mathbf{k}'$, and thus don't contribute to the energy loss after integrating over the momenta.

The interaction rate Γ in Eq. (2) can be also expressed in terms of heavy-quark self-energy [18]. For the soft contribution to $-dE/dx$, the above mentioned infrared divergence can be eliminated by using the HTL resummed gluon propagator in the calculation of the heavy-quark self-energy. It has been shown that [32] the resulting soft contribution to $-dE/dx$ is equivalent to the energy loss obtained in the classical plasma physics where $-dE/dx$ is caused by the Lorentz force due to the chromoelectric field induced by the incident heavy-quark, namely [20, 33]

$$\left(\frac{dE}{dt}\right)_{\rm soft} = {\rm Re}\int d^3\mathbf{x}\; \mathbf{J}^a_{\rm ext}(t,\mathbf{x})\cdot\mathbf{E}^a_{\rm ind}(t,\mathbf{x})\,, \tag{5}$$

where the current $\mathbf{J}^a_{\rm ext}(t,\mathbf{x}) = \mathsf{c}^a\mathbf{v}\,\delta^{(3)}(\mathbf{x}-\mathbf{v}t)$ represents the propagation of the heavy-quark with color charge c^a defined by $\sum \mathsf{c}^a\mathsf{c}^a = g^2C_F$. For $SU(3)$, the Casimir invariant $C_F = 4/3$. Using the Vlasov-Maxwell equations, the induced chromoelectric field in the momentum space can be written as

$$E^{i,a}_{\rm ind}(Q) = i\omega\big[\Delta^{ij}(Q) - \Delta^{ij}_0(Q)\big]J^{j,a}_{\rm ext}(Q)\,, \tag{6}$$

with $Q = (\omega, \mathbf{q})$. In the above equation, $\Delta^{ij}(Q)$ denotes the HTL resummed gluon propagator in the temporal axial gauge while $\Delta^{ij}_0(Q)$ is the corresponding temperature-independent free propagator. Explicitly, we have

$$\begin{aligned}
\Delta^{ij}(Q) &= \frac{1}{\omega^2 - q^2 - m_D^2\Pi_T(\hat\omega)}A^{ij} + \frac{1}{\omega^2 - m_D^2\Pi_L(\hat\omega)}B^{ij}\,,\\
\Delta^{ij}_0(Q) &= \frac{1}{\omega^2 - q^2}A^{ij} + \frac{1}{\omega^2}B^{ij}\,,
\end{aligned} \tag{7}$$

where the transverse and longitudinal structure functions read

$$\begin{aligned}
\Pi_T(\hat\omega) &= \frac{\hat\omega^2}{2}\left(1 - \frac{\hat\omega^2-1}{2\hat\omega}\ln\frac{\hat\omega+1}{\hat\omega-1}\right),\\
\Pi_L(\hat\omega) &= \hat\omega^2\left(-1 + \frac{\hat\omega}{2}\ln\frac{\hat\omega+1}{\hat\omega-1}\right),
\end{aligned} \tag{8}$$

and the two projectors are defined as

$$A^{ij} = \delta^{ij} - q^iq^j/q^2\,, \qquad B^{ij} = q^iq^j/q^2\,. \tag{9}$$

In addition, $\hat\omega = \omega/q$ and the Debye screening mass $m_D^2 = Ng^2T^2/3$. Combining the above equations, we arrive at the following form for the soft contribution to the energy loss

$$-\left(\frac{dE}{dx}\right)_{\rm soft} = \frac{16\pi\alpha_s}{3v}{\rm Im}\int\frac{d^3\mathbf{q}}{(2\pi)^3}(\mathbf{v}\cdot\mathbf{q})v^i\big[\Delta^{ij}(Q) - \Delta^{ij}_0(Q)\big]v^j\theta(q^*-q)\Big|_{\omega=\mathbf{v}\cdot\mathbf{q}}\,, \tag{10}$$

where the cutoff q^* regulates the logarithmic divergence for large momentum transfer. Such an unphysical behavior appears because Eq. (5) is only valid for soft processes with a typical momentum transfer $q \sim gT$.

III. THE COLLISIONAL ENERGY LOSS OF A HEAVY-QUARK IN A SEMI-QGP

In this section, we study the collisional energy loss of a heavy-quark in a semi-QGP where the focus will be put on the influence of a background field $\mathcal{Q}$ on the corresponding $-dE/dx$. In Sec. III A, we review some basics of

the effective theory adopted in our calculation and derive the $\mathcal{Q}$-dependent resummed gluon propagator. The hard and soft contributions to $-dE/dx$ in the presence of a background field will be discussed in Secs. III B and III C, respectively.

A. Effective theory of a semi-QGP

According to the previous discussions, it is possible to introduce a classical background field $A_0^{\rm cl}$ to describe the nontrivial Polyakov loop in the deconfining phase transition for $SU(N)$ gauge theories. The background field is assumed to be constant in spacetime and given by a diagonal matrix in color space,

$$(A_0^{\rm cl})_{ab} = \frac{1}{g}\mathcal{Q}^a\delta_{ab}\,. \tag{11}$$

Furthermore, it satisfies the traceless condition $\sum_{a=1}^{N}\mathcal{Q}^a = 0$ with $a, b = 1, \cdots, N$.

The Wilson line in the temporal direction is defined as

$$\mathbf{L} = \mathcal{P}\exp\Big(ig\int_0^\beta A_0^{\rm cl}\,d\tau\Big)\,, \tag{12}$$

where $\mathcal{P}$ denotes the time ordering, τ is the imaginary time and $\beta \equiv 1/T$ is the inverse temperature. Accordingly, the gauge invariant Polyakov loop takes the following form

$$\ell = \frac{1}{N}{\rm Tr}\,\mathbf{L}\,. \tag{13}$$

One can also define higher loops, $(1/N){\rm Tr}\,\mathbf{L}^n$, and there are $N-1$ independent loops for $SU(N)$.

After taking into account the classical background field, the effective potential in general $SU(N)$ gauge theories can be computed in the perturbation theory. The one-loop result is

$$\mathcal{V}_{\rm pt} = \frac{2\pi^2T^4}{3}\sum_{ab}\mathcal{P}^{ab,ba}B_4(|\mathsf{q}^{ab}|)\,. \tag{14}$$

The dependence on the background field $\mathcal{Q}$ is given by the fourth Bernoulli polynomials $B_4(x) = x^2(1-x)^2 - 1/30$. For later use, we also need the second Bernoulli polynomials $B_2(x) = x^2 - x + 1/6$. Notice that the argument x should be understood as $x - [x]$ with $[x]$ the largest integer less than x, which is nothing but the modulo function. In the above equation, a and b run from 1 to N when summing over the color indices. In addition, we define $\mathsf{q}^{ab} = \mathsf{q}^a - \mathsf{q}^b$ with $\mathsf{q}^a \equiv \mathcal{Q}^a/(2\pi T)$ and the projection operator $\mathcal{P}^{ab,cd}$ in the double line basis reads

$$\mathcal{P}^{ab,cd} = \mathcal{P}^{ab}_{dc} = \delta^a_d\delta^b_c - \frac{1}{N}\delta^{ab}\delta_{cd}\,. \tag{15}$$

It is more convenient to perform the calculation in the double line basis when a background field is present. In this basis, color indices in the fundamental representation are denoted by $a, b, \cdots = 1, 2, \cdots, N$, while in the adjoint representation, they are denoted by a pair of fundamental indices, ab. The generators of the fundamental representation are given by $(t^{ab})_{cd} = \frac{1}{\sqrt{2}}\mathcal{P}^{ab}_{cd}$, normalized as ${\rm Tr}\,(t^{ab}t^{cd}) = \frac{1}{2}\mathcal{P}^{ab,cd}$. More details about the double line basis can be found in Refs. [34, 35].

The equation of motion for the background field based on Eq. (14) indicates that the system is always in a fully deconfined phase with vanishing background field, and no phase transition could happen. It can also be shown that after including the two-loop perturbative corrections [36–40], one still has $\mathcal{Q} = 0$ as the vacuum. To drive the system to confinement, a nonperturbative contribution needs to be added to the effective potential. It is expected to play an important role near the critical temperature where the confined vacuum is favored. One possible way to generate such a contribution is to include a mass scale in the dispersion relation for gauge bosons, $\omega_k = \sqrt{k^2 + M^2}$, then expand the resulting effective potential in the high temperature limit $M \ll T$. The leading order term in the expansion is nothing but the perturbative $\mathcal{V}_{\rm pt}$ as given in Eq. (14), and the next-to-leading order term appears as the desired nonperturbative contribution in the effective potential which takes the following form [8]

$$\mathcal{V}_{\rm npt} = \frac{M^2T^2}{2}\sum_{ab}\mathcal{P}^{ab,ba}B_2(|\mathsf{q}^{ab}|)\,. \tag{16}$$

An alternative way to generate the above nonperturbative contribution is to embed a two-dimensional ghost field isotropically in four dimensions [9]. In this case, the mass scale in Eq. (16) should be considered as an upper limit for the transverse momentum of the embedded fields. It indicates that the ghost field is two dimensional at short distances, however, for distances larger than the scale of confinement, it becomes fully four-dimensional field.

The combination of Eqs. (14) and (16) leads to an effective theory for studying the semi-QGP. One can derive the equation of motion for the classical background field in this theory and a non-zero and temperature dependent background field arises. At low temperatures, the solution $\mathsf{q}^a = (N-2a+1)/(2N)$ minimizes the effective potential[1] which corresponds to the confining vacuum with vanishing Polyakov loop. When the temperature gets very large, $\mathsf{q}^a \sim M^2/T^2$ indicates a perturbative vacuum where the Polyakov loop approaches to unity. By requiring the phase transition to occur at the critical temperature T_d, the mass scale M which is the only parameter in the effective theory can be determined. For the $SU(3)$ gauge theory, by parameterizing the background field as

$$\vec{\mathsf{q}} = (\mathsf{q}, 0, -\mathsf{q})\,, \tag{17}$$

the explicit T-dependence of the background field in the deconfined phase can be determined from its equation of motion as the following

$$\mathsf{q} = \frac{1}{36}\big(9 - \sqrt{81 - 80(T_d/T)^2}\big)\,, \tag{18}$$

where $M/T_d = 2\sqrt{10}\pi/9$ has been used. It is worth noting that besides the nontrivial behavior of the Polyakov loop in the deconfining phase transition, the above effective theory is also capable of qualitatively reproducing the Lattice simulations on the thermodynamics for the $SU(N)$ gauge theories. Further improvements have been proposed, leading to various versions of the matrix models for deconfinement [41–44].

In this work, we will investigate the influence of the background field on the heavy-quark collisional energy loss when it passes through the hot and dense medium. For the hard scatterings, only thermal distributions of the medium partons are affected by the background field. The resulting Bose-Einstein distribution functions are given by [5, 45]

$$n(k_0, \mathsf{q}^{ab}) = \begin{cases} \dfrac{1}{e^{|k_0|/T - 2\pi i \mathsf{q}^{ab}} - 1} \equiv n_+^{ab}(k_0) & \text{for} \quad k_0 > 0 \\ \dfrac{1}{e^{|k_0|/T + 2\pi i \mathsf{q}^{ab}} - 1} \equiv n_-^{ab}(k_0) & \text{for} \quad k_0 < 0 \end{cases}. \tag{19}$$

On the other hand, for scatterings with soft gluon exchange, it becomes necessary to use the HTL resummed gluon propagator which regulates the infrared divergence. In the presence of a background field, the resummed propagator obtained from the Dyson-Schwinger equation gets modified through the $\mathcal{Q}$-dependent gluon self-energy.

Within the framework of perturbation theory, it was found that the HTL approximated gluon self-energy is not transverse due to the appearance of an anomalous term at nonzero $\mathcal{Q}$ [34, 46]. In addition, this anomalous term also results in an ill-defined gluon resummed propagator [47] as well as an unexpected discontinuity in the free energy at higher order in the coupling constant [48, 49]. These known issues further justify the necessity of employing an effective theory to properly incorporate the effect of a background field. After taking into account the contributions from the aforementioned two-dimensional ghosts, the anomalous term is completely cancelled and the resulting gluon self-energy has the same Lorentz structure as its counterpart at $\mathcal{Q} = 0$, and thus preserves the transversality [9].

Based on the Dyson-Schwinger equation, the HTL resummed gluon propagator in the effective theory was first computed in [47] where the covariant gauge was used. For our purpose, we carry out a similar calculation in the temporal axial gauge and the $N^2 - N$ off-diagonal components in color space are given by

$$\Delta_{ij}^{ab,cd}(Q, \vec{\mathsf{q}}) \overset{a\neq b}{=\!=} \delta^{ad}\delta^{bc}\left[\frac{1}{\omega^2 - q^2 - (\mathcal{M}_D^2)^{ab}(\vec{\mathsf{q}})\Pi_T(\hat{\omega})}A_{ij} + \frac{1}{\omega^2 - (\mathcal{M}_D^2)^{ab}(\vec{\mathsf{q}})\Pi_L(\hat{\omega})}B_{ij}\right], \tag{20}$$

where the $\mathcal{Q}$-modified screening mass can be written as

$$(\mathcal{M}_D^2)^{ab}(\vec{\mathsf{q}}) = \left[\frac{3}{N}\sum_{e=1}^{N}\Big(B_2(|\mathsf{q}^{ae}|) + B_2(|\mathsf{q}^{eb}|)\Big) + \frac{3M^2}{4\pi^2T^2}\right]m_D^2\,, \tag{21}$$

[1] This is true under the straight line ansatz where the background fields q^a with $a = 1, 2, \cdots, N$ have constant spacing and automatically satisfy the traceless condition.

with $a \neq b$ and the contribution $\sim M^2$ comes from the two-dimensional ghost field. As compared to Eq. (7), it turns out that the influence of the background field on the transverse and longitudinal gluon self-energy amounts to the same modification on the screening mass. It is simply a $\mathcal{Q}$-dependent multiplication as given by the terms in the square bracket in Eq. (21).

In addition, the diagonal components of the HTL resummed gluon propagator, after multiplying by the projection operator and summing over the color indices, read

$$\sum_{ab}\mathcal{P}^{aa,bb}\Delta_{ij}^{aa,bb}(Q,\vec{\mathsf{q}})=\sum_{\sigma=1}^{N-1}\left[\frac{1}{\omega^2-q^2-\left(\mathcal{M}_D^2\right)^{\sigma}(\vec{\mathsf{q}})\Pi_T(\hat{\omega})}A_{ij}+\frac{1}{\omega^2-\left(\mathcal{M}_D^2\right)^{\sigma}(\vec{\mathsf{q}})\Pi_L(\hat{\omega})}B_{ij}\right]. \tag{22}$$

Similarly, the influence of a non-zero background field is fully encoded in the modified screening masses which can be written as

$$\left(\mathcal{M}_D^2\right)^{\sigma}(\vec{\mathsf{q}})=\left(1+\frac{3M^2}{4\pi^2T^2}+\frac{6}{N}\lambda^{\sigma}\right)m_D^2\,, \tag{23}$$

where terms in the bracket represent the background field modifications and λ^{σ} with $\sigma=1,2,\cdots,N-1$ denote a set of $\mathcal{Q}$-dependent functions. For $SU(3)$ gauge theory, the explicit form of λ^{σ} can be obtained as

$$\vec{\lambda}=\left(3\mathsf{q}^2-3\mathsf{q},\,9\mathsf{q}^2-5\mathsf{q}\right). \tag{24}$$

In fact, only the color sum $\sum_{abcd}\mathcal{P}^{ab,cd}\Delta_{ij}^{ab,cd}$ is relevant for studying the soft scattering processes. Combining the above results of the diagonal and off-diagonal contributions, we arrive at the following compact form

$$\sum_{abcd}\mathcal{P}^{ab,cd}\Delta_{ij}^{ab,cd}(Q,\vec{\mathsf{q}})=\sum_{ab/\sigma}\left[\frac{1}{\omega^2-q^2-\left(\mathcal{M}_D^2\right)^{ab/\sigma}(\vec{\mathsf{q}})\Pi_T(\hat{\omega})}A_{ij}+\frac{1}{\omega^2-\left(\mathcal{M}_D^2\right)^{ab/\sigma}(\vec{\mathsf{q}})\Pi_L(\hat{\omega})}B_{ij}\right], \tag{25}$$

where $\sum_{ab/\sigma}=\sum_{ab}+\sum_{\sigma}$ denotes a sum over both diagonal and off-diagonal gluons and for general $SU(N)$, $a,b=1,2,\cdots,N$ with $a\neq b$ and $\sigma=1,2,\cdots,N-1$. For vanishing background field, it reduces to[2]

$$\sum_{abcd}\mathcal{P}^{ab,cd}\Delta_{ij}^{ab,cd}(Q,\vec{\mathsf{q}})\overset{\mathcal{Q}=0}{=\!=}(N^2-1)\Big[\frac{1}{\omega^2-q^2-m_D^2\Pi_T(\hat{\omega})}A^{ij}+\frac{1}{\omega^2-m_D^2\Pi_L(\hat{\omega})}B^{ij}\Big], \tag{26}$$

which indicates that the N^2-1 gluons have equal contribution to the above result. For non-zero background field, on the other hand, both the N^2-N off-diagonal gluons and $N-1$ diagonal gluons acquire different $\mathcal{Q}$-dependent modifications, and thus become distinguishable by their associated screening masses. Notice that according to our parametrization Eq. (17), there are only two different screening masses for the 6 off-diagonal gluons in $SU(3)$ gauge theory.

B. The hard scattering processes in the background field

In this subsection, we calculate the hard contributions to the collisional energy loss in the background field. As already mentioned before, the influence of a non-zero $\mathcal{Q}$ is reflected in the distributions of the thermal partons as given in Eq. (19). A special feature of such a $\mathcal{Q}$-dependent distribution is that the gluons are denoted by a pair of the color indices, ab. Accordingly, the color structures of the squared matrix element $|\mathcal{M}|^2$ need to be re-computed in the double line basis. We list the corresponding Feynman rules in the axial gauge in Fig. 2.

It is found that for non-vanishing background field, the energy loss $-dE/dx$ due to quark-gluon scattering can be obtained from Eqs. (3) and (4) by the replacement

$$\begin{aligned}&4\int\frac{d^3\mathbf{k}}{(2\pi)^3}\frac{n(k)}{k}\int\frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{1-n(k')}{k'}\to\\&\sum_{abcdef}\frac{1}{6}f^{ab,cd,ef}f^{ba,dc,fe}\int\frac{d^3\mathbf{k}}{(2\pi)^3}\frac{\left(n_+^{ab}(k)+n_-^{ab}(k)\right)/2}{k}\int\frac{d^3\mathbf{k}'}{(2\pi)^3}\frac{1-\left(n_+^{cd}(k')+n_-^{cd}(k')\right)/2}{k'}\,,\end{aligned} \tag{27}$$

[2] Temperature independent contributions $\sim M^2$ in Eqs. (21) and (23) are negligible in the high temperature limit where the background field vanishes.

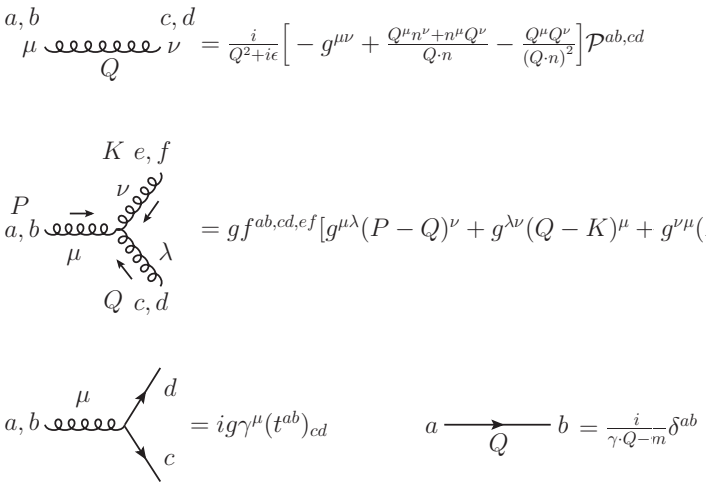


FIG. 2. The Feynman rules in the axial gauge relevant to our calculations. Here, n^μ is a unit vector and the structure constant $f^{ab,cd,ef} = i(\delta^{ad}\delta^{cf}\delta^{eb} - \delta^{af}\delta^{cb}\delta^{ed})/\sqrt{2}$.

where 4 is the color factor in zero background field. Analogous to the $\mathcal{Q} = 0$ case, the above form indicates that terms involving the product $n^{ab}_{\pm}(k)n^{cd}_{\pm}(k')$ don't contribute to the energy loss. This can be seen by interchanges of the momenta $\mathbf{k} \leftrightarrow \mathbf{k}'$ and the color indices $ab \leftrightarrow cd$. Thus, for non-zero background field, we arrive at the following expressions for the hard contributions to $-dE/dx$,

$$-\left(\frac{dE}{dx}\right)^{Qg(t)}_{\rm BF,hard} = \frac{2\pi g^4}{v}\sum_{a,b=1}^{3}\left(1-\frac{1}{3}\delta^{ab}\right)\int\frac{d^3\mathbf{k}}{(2\pi)^3 k}\frac{n^{ab}_+(k)+n^{ab}_-(k)}{2}\int\frac{d^3\mathbf{k}'}{(2\pi)^3 k'}\delta(\omega-\mathbf{v}\cdot\mathbf{q}) \times \frac{\omega}{(\omega^2-q^2)^2}\left[(k-\mathbf{v}\cdot\mathbf{k})^2+\frac{1-v^2}{2}(\omega^2-q^2)\right]\theta(q-q^*)\,, \tag{28}$$

$$-\left(\frac{dE}{dx}\right)^{Qg(s+u)}_{\rm BF,hard} = \frac{\pi g^4}{4v}\sum_{a,b=1}^{3}\left(1-\frac{1}{3}\delta^{ab}\right)\int\frac{d^3\mathbf{k}}{(2\pi)^3 k}\frac{n^{ab}_+(k)+n^{ab}_-(k)}{2}\int\frac{d^3\mathbf{k}'}{(2\pi)^3 k'}\delta(\omega-\mathbf{v}\cdot\mathbf{q})\frac{\omega(1-v^2)^2}{(k-\mathbf{v}\cdot\mathbf{k})^2}\,. \tag{29}$$

The energy loss due to the s- and u-channel scatterings can be obtained analytically which reads

$$-\left(\frac{dE}{dx}\right)^{Qg(s+u)}_{\rm BF,hard} = \alpha_s^2\pi T^2\sum_{ab}\left(1-\frac{1}{3}\delta^{ab}\right)B_2(|\mathsf{q}^{ab}|)f_1(v) = \frac{\alpha_s^2 T^2}{\pi}\sum_{n=1}^{\infty}\frac{|{\rm Tr}\,\mathbf{L}^n|^2-1}{n^2}f_1(v)\,, \tag{30}$$

where

$$f_1(v) = \frac{1}{v}-\frac{1-v^2}{2v^2}\ln\frac{1+v}{1-v}\,. \tag{31}$$

To get the above result, we used

$$\sum_{ab}\int_0^\infty k\,n^{ab}_{\pm}(k)dk = T^2\sum_{n=1}^{\infty}\frac{|{\rm Tr}\,\mathbf{L}^n|^2}{n^2} = \pi^2T^2\sum_{ab}B_2(|\mathsf{q}^{ab}|)\,. \tag{32}$$

Our results show that for the s- and u-channel scatterings, the modification of a background field is simply given by the factor $\sum_{ab}(1-\delta^{ab}/3)B_2(|\mathsf{q}^{ab}|)$. In the case of $SU(3)$, it equals $4(1-3\mathsf{q})^2/3$ which reduces to $4/3$ for vanishing background field as expected. However, integrals appearing in the t-channel scattering turn out to be more involved. To proceed further, we make the following change of the integral variables

$$\frac{1}{2(2\pi)^2}\int\frac{d^3\mathbf{k}}{k}\int\frac{d^3\mathbf{k}'}{k'} \to \int_{\frac{1+v}{2}q^*}^{\infty}dk\int_{q^*}^{\frac{2k}{1+v}}qdq\int_{-vq}^{vq}d\omega + \int_{\frac{1+v}{2}q^*}^{\infty}dk\int_{\frac{2k}{1+v}}^{\frac{2k}{1-v}}qdq\int_{q-2k}^{vq}d\omega + \int_{\frac{1-v}{2}q^*}^{\frac{1+v}{2}q^*}dk\int_{q^*}^{\frac{2k}{1-v}}qdq\int_{q-2k}^{vq}d\omega\,. \tag{33}$$

The above integral intervals are determined to meet the constraint condition that $-1 \leq \cos\theta \leq 1$ where θ denotes the angle between $\mathbf{k}$ and $\mathbf{k}'$ and $\cos\theta = (k^2 + (\omega+k)^2 - q^2)/(2k(k+w))$. Further constraints due to the two theta functions $\theta(q-q^*)$ and $\theta(v^2q^2-\omega^2)$ have also been taken into account in Eq. (33). The second theta function arises after averaging the integrand over the direction of $\mathbf{v}$ which is valid because the energy loss is independent of the incident direction of the heavy-quark in an isotropic medium.

The cutoff q^* was introduced to eliminate the infrared divergence associated with soft momentum transfer. In fact, in the limit $q^* \to 0$, only the first integral over q in Eq. (33) becomes divergent, and thus needs to be regulated. Consequently, a simplified version of Eq. (33) can be obtained as

$$\frac{1}{2(2\pi)^2}\int\frac{d^3\mathbf{k}}{k}\int\frac{d^3\mathbf{k}'}{k'} \to \int_0^\infty dk\int_{q^*}^{\frac{2k}{1+v}} qdq\int_{-vq}^{vq} d\omega + \int_0^\infty dk\int_{\frac{2k}{1+v}}^{\frac{2k}{1-v}} qdq\int_{q-2k}^{vq} d\omega\,, \tag{34}$$

which coincides with the variable change as used in [18] and leads to an analytical form for the energy loss even for non-zero background field. The result can be expressed as

$$\begin{aligned} -\left(\frac{dE}{dx}\right)^{Qg(t)}_{\rm BF,hard} &= \frac{2\alpha_s^2}{\pi}\sum_{ab}\left(1-\frac{1}{3}\delta^{ab}\right)\int_0^\infty kdk\,\frac{n_+^{ab}(k)+n_-^{ab}(k)}{2}\left[f_1(v)\left(\ln(k/q^*)-\frac{1}{2}\right)+f_2(v)\right] \\ &= \frac{2\alpha_s^2T^2}{\pi}\sum_{n=1}^\infty\frac{|{\rm Tr}\,\mathbf{L}^n|^2-1}{n^2}\left[f_1(v)\left(\ln\frac{T}{q^*}+\frac{1}{2}-\gamma-\ln n\right)+f_2(v)\right], \end{aligned} \tag{35}$$

where

$$\begin{aligned} f_2(v) &= \left[\frac{1}{2}-\frac{1}{2}\ln(1-v^2)+\ln 2\right]f_1(v) - \frac{1-v^2}{4v^2}\left[{\rm Li}_2\left(\frac{1+v}{2}\right)\right. \\ &- \left.{\rm Li}_2\left(\frac{1-v}{2}\right)+\frac{1}{2}\ln\frac{1-v^2}{4}\ln\frac{1+v}{1-v}\right] - \frac{2}{3}v\,, \end{aligned} \tag{36}$$

and ${\rm Li}_2(x)$ is the polylogarithm. As we can see in the limit $q^* \to 0$, the above result has a logarithmic divergence $\sim \ln q^*$ and the background field modification on this divergent part is the same as that on the s- and u-channel contributions. On the other hand, for the finite part, besides the term $\sim k$ in the integral over k, there is also a term $\sim k\ln k$ which leads to a more involved $\mathcal{Q}$-dependence of the energy loss which cannot be described simply by the Bernoulli polynomial.

It can also be shown that the hard contribution in Eq. (28) is always positive when varying q^*. However, because of the simplification used in Eq. (34), Eq. (35) may lead to a negative energy loss at large q^*. Therefore, Eq. (34) is only a good approximation when $q^* \ll T$. For an arbitrary cutoff, one should adopt Eq. (33) to change the integral variables for mathematical rigour. Thus, Eq. (28) can be written as

$$\begin{aligned} -\left(\frac{dE}{dx}\right)^{Qg(t)}_{\rm BF,hard} &= \frac{3\alpha_s^2}{\pi v^2}\sum_{ab}\left(1-\frac{1}{3}\delta^{ab}\right)\iiint_{\rm sum} dk\,dq\,d\omega\,\frac{n_+^{ab}(k)+n_-^{ab}(k)}{2}\,\frac{\omega}{q^2} \\ &\times \left[\frac{k^2+k\omega+\omega^2/4}{q^2}-\frac{1-v^2}{q^2-\omega^2}\,\frac{k^2+k\omega+q^2}{3}-\frac{v^2}{12}\right], \end{aligned} \tag{37}$$

where we use the shorthand notation $\iiint_{\rm sum} dk\,dq\,d\omega$ to denote the sum of the three triple integrals as given in Eq. (33). Consequently, the integral over k needs to be carried out numerically. Similar to the $\mathcal{Q}=0$ case discussed in [20], the energy loss based on Eq. (37) is always positive for an arbitrary cutoff and agrees with Eq. (35) in the small q^* region.

C. The soft scattering processes in the background field

The soft contributions to the collisional energy loss can be studied based on Eq. (10) where the influence of the background field amounts to a modification on the gluon propagator. In the double line basis, the gluon propagator is not diagonal in the color space and summing over the color indices becomes nontrivial. In order to explicitly show the color structures, we rewrite Eq. (10) as

$$-\left(\frac{dE}{dx}\right)_{\rm BF,soft} = \sum_{abcd}\frac{4\pi\alpha_s}{v}\frac{1}{2N}\mathcal{P}^{ab,cd}\,{\rm Im}\int\frac{d^3\mathbf{q}}{(2\pi)^3}(\mathbf{v}\cdot\mathbf{q})v^i\left[\Delta_{ij}^{ab,cd}(Q,\mathfrak{q})-(\Delta_0)_{ij}^{ab,cd}(Q)\right]v^j\theta(q^*-q)\Big|_{\omega=\mathbf{v}\cdot\mathbf{q}}. \tag{38}$$

In the above equation, $(\Delta_0)_{ij}^{ab,cd}(Q)$ is not the usual bare propagator in perturbation theory. It actually corresponds to the zero temperature limit of the resummed propagator in the effective theory and can be written as

$$(\Delta_0)_{ij}^{ab,cd}(Q)=\mathcal{P}^{ab,cd}\left[\frac{1}{\omega^2-q^2-\frac{3g^2M^2}{4\pi^2}\Pi_T(\hat{\omega})}A_{ij}+\frac{1}{\omega^2-\frac{3g^2M^2}{4\pi^2}\Pi_L(\hat{\omega})}B_{ij}\right]. \tag{39}$$

The usual bare propagator can be recovered by setting $M^2=0$ in the above equation which has no contribution to the energy loss because of a vanishing imaginary part. However, the propagator as given in Eq. (39) leads to an extra contribution which as we will see, is important to ensure the cancellation of the cutoff dependence in the weak coupling limit although such a contribution is temperature independent.

In the limit $\mathcal{Q}\to 0$, the resummed gluon propagation $\Delta_{ij}^{ab,cd}(Q,\mathsf{q})$ is also proportional to $\mathcal{P}^{ab,cd}$ and the sum over the color indices in Eq. (38) can be easily carried out which leads to the following color factor

$$\sum_{abcd}\frac{1}{2N}\mathcal{P}^{ab,cd}\mathcal{P}^{ab,cd}=C_F=\frac{4}{3}, \tag{40}$$

for $SU(3)$. On the other hand, when $\mathcal{Q}\neq 0$, based on the resummed gluon propagator as given in Sec. III A, the following compact form can be found for the collisional energy loss due to soft scatterings,

$$\begin{aligned}-\left(\frac{dE}{dx}\right)_{\rm BF,soft} &= \frac{\alpha_s}{12\pi v^2}\sum_{ab/\sigma/s}{\rm Im}\int_{-v}^{+v}d\hat{\omega}\frac{v^2-\hat{\omega}^2}{\hat{\omega}}\left[\frac{\hat{\omega}^2}{(1-\hat{\omega}^2)^2}F_T^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})\ln\frac{(1-\hat{\omega}^2)(q^*)^2+F_T^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})}{F_T^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})}\right.\\ &+\left.\frac{1}{v^2-\hat{\omega}^2}F_L^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})\ln\frac{F_L^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})-(q^*)^2\hat{\omega}^2}{F_L^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})}\right],\end{aligned} \tag{41}$$

where we used $d^3\mathbf{q}=2\pi q^2dqd\hat{\omega}/v$ and the integral over q has been done analytically. Furthermore, the shorthand notation of the color sums is defined as

$$\sum_{ab/\sigma/s}=\sum_{ab}+\sum_{\sigma}-\sum_{s}, \tag{42}$$

where the first two sums on the right hand side have already appeared in Eq. (25). Notice the minus sign for the last sum which is related to the contribution associated with $(\Delta_0)_{ij}^{ab,cd}(Q)$ in Eq. (38). In the high temperature limit where $\mathcal{Q}\to 0$, such a contribution becomes negligible as compared to the contributions from the first two sums. However, near the critical temperature, the three sums in Eq. (42) give comparable contributions to the energy loss. In addition, we have

$$F_{T/L}^{ab/\sigma/s}(\hat{\omega},\vec{\mathsf{q}})=(\mathcal{M}_D^2)^{ab/\sigma/s}(\vec{\mathsf{q}})\Pi_{T/L}(\hat{\omega}), \tag{43}$$

where $(\mathcal{M}_D^2)^s=3g^2M^2/(4\pi^2)$ is independent on the color index s and for $SU(N)$, $s=1,2,\cdots,N^2-1$. Therefore, the purpose of introducing such a trivial sum over s is merely to make the form of Eq. (41) compact.

By assuming $q^*\gg gT\sim m_D$, one can expand the logarithm in Eq. (41). Keeping only the leading order terms in the expansion, an analytical expression for the q^*-dependent part can be obtained which in the case of $SU(3)$ reads

$$\frac{\alpha_s}{12}\sum_{ab/\sigma/s}(\mathcal{M}_D^2)^{ab/\sigma/s}f_1(v)\ln q^*=\frac{8\pi\alpha_s^2T^2}{3}f_1(v)(1-3\mathsf{q})^2\ln q^*. \tag{44}$$

Furthermore, the q^*-dependent part of the hard contribution in Eq. (35) can be written as

$$-\frac{2\alpha_s^2T^2}{\pi}\sum_{n=1}^{\infty}\frac{|{\rm Tr}\,\mathbf{L}^n|^2-1}{n^2}f_1(v)\ln q^* = -\frac{8\pi\alpha_s^2T^2}{3}f_1(v)(1-3\mathsf{q})^2\ln q^*. \tag{45}$$

Clearly, the cutoff dependence cancels between the hard and soft contributions which is similar to the case of zero background field as studied in Ref. [18]. Notice that the above cancellation doesn't depend on the details of the background field, namely, a specified temperature dependence as determined by Eq. (18) is not needed to make the cancellation happen. Here, contributions from the last sum in Eq. (42) are important to ensure the exact cancellation.

One can also get the full leading order result of the soft contributions when expanding Eq. (41) with $q^* \gg gT$, which is given by

$$-\left(\frac{dE}{dx}\right)_{\rm BF,soft} = \frac{8\pi\alpha_s^2 T^2}{3}(1-3\mathsf{q})^2\Big[f_1(v)\ln\left(q^*/m_D\right) - f_3(v)\Big] - \frac{\alpha_s}{24}f_1(v)\sum_{ab/\sigma/s}\left(\mathcal{M}_D^2\right)^{ab/\sigma/s}\ln\left(\left(\mathcal{M}_D^2\right)^{ab/\sigma/s}/m_D^2\right). \tag{46}$$

In the high temperature limit where $\mathsf{q}\to 0$, the second line in the above equation can be neglected and we get the corresponding result at zero background field [19]

$$-\left(\frac{dE}{dx}\right)_{\rm BF=0,soft} = \frac{8\pi\alpha_s^2 T^2}{3}\Big[f_1(v)\ln\left(q^*/m_D\right) - f_3(v)\Big], \tag{47}$$

where

$$f_3(v) = \frac{1}{v^2}\int_0^v d\hat{\omega}\,\hat{\omega}^2\Big[\frac{v^2-\hat{\omega}^2}{2(1-\hat{\omega}^2)^2}\big(G_T(\hat{\omega})+\ln(\pi\hat{\omega}/4)\big) + \big(G_L(\hat{\omega})+\ln(\pi\hat{\omega}/2)\big)\Big], \tag{48}$$

with

$$G_{T/L}(\hat{\omega}) = Q_{T/L}(\hat{\omega})\Big[\frac{\pi}{2} - \arctan\big(Q_{T/L}(\hat{\omega})\big)\Big] + \frac{1}{2}\ln\left(1+Q_{T/L}^2(\hat{\omega})\right), \tag{49}$$

and

$$Q_T(\hat{\omega}) = \frac{1}{\pi}\Big[\ln\big(\frac{1+\hat{\omega}}{1-\hat{\omega}}\big) + \frac{2\hat{\omega}}{1-\hat{\omega}^2}\Big], \qquad Q_L(\hat{\omega}) = \frac{1}{\pi}\Big[\frac{2}{\hat{\omega}} - \ln\big(\frac{1+\hat{\omega}}{1-\hat{\omega}}\big)\Big]. \tag{50}$$

The above result also indicates that the background field modification on the soft contributions is not simply an overall factor $(1-3\mathsf{q})^2$ due to the appearance of the second line in Eq. (46). Similar behavior has already been found in the t-channel contribution in the hard scatterings, see Eq. (35).

We should point out that the simplification used in Eq. (34) is valid when $q^* \ll T$, while the expansion of Eq. (41) is ensured to be convergent if $q^* \gg gT$. Therefore, the above cancellation of the cutoff dependence requires a very small coupling constant in order to meet the assumption $gT \ll q^* \ll T$ which is also the prerequisite to use the energy loss obtained in Eqs. (35) and (46). Similar to the negative energy loss obtained from Eq. (35) for very large q^*, Eq. (46) also leads to a negative result when the cutoff becomes too small as compared to the screening mass. On the other hand, when extrapolating to a realistic value of the coupling constant, the above assumption cannot be well satisfied. Consequently, for an arbitrary cutoff, we need to use Eq. (33) to carry out the integrals for the hard contributions and evaluate the soft contributions by Eq. (41) without any expansion. The total energy loss which needs to be determined numerically does show a q^*-dependence. Such an ambiguity can be eliminated by using the principle of minimum sensitivity [20], and thus the cutoff q^* is fixed by minimizing the total energy loss with respect to q^*.

Finally, let's focus on the $\mathcal{Q}$-dependence of the collisional energy loss based on our results in the weak coupling limit. For the hard processes, since $q^* \ll T$, the dominant t-channel contribution comes from the term $\sim \ln(T/q^*)$ in Eq. (35). Therefore, the influence of the background field amounts to a q-dependent factor reading $(1-3\mathsf{q})^2$ for $SU(3)$[3]. For the soft processes, keeping only the logarithmically enhanced contribution $\sim \ln(q^*/m_D)$ with $q^* \gg m_D$, the same conclusion can be drawn from Eq. (46). As a result, the q^*-independent total energy loss equals the factor $(1-3\mathsf{q})^2$ times the corresponding result at vanishing background field. In the weak coupling limit, the dominant contribution in the total energy loss is $\sim \ln(1/g)$ which comes from the soft processes. Therefore, a proper treatment on the soft momentum exchange with the hard thermal loop resummation is necessary.

IV. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we present our numerical results for the collisional energy loss of a heavy-quark in a semi-QGP which consists of the hard contributions from Eqs. (30) and (37) as well as the soft contributions from Eq. (41). As

[3] The same is true for the s- and u-channel contributions. However, there is no logarithmic enhancement in Eq. (30), therefore, as compared to the t-channel contribution, contributions from s- and u-channels can be neglected.

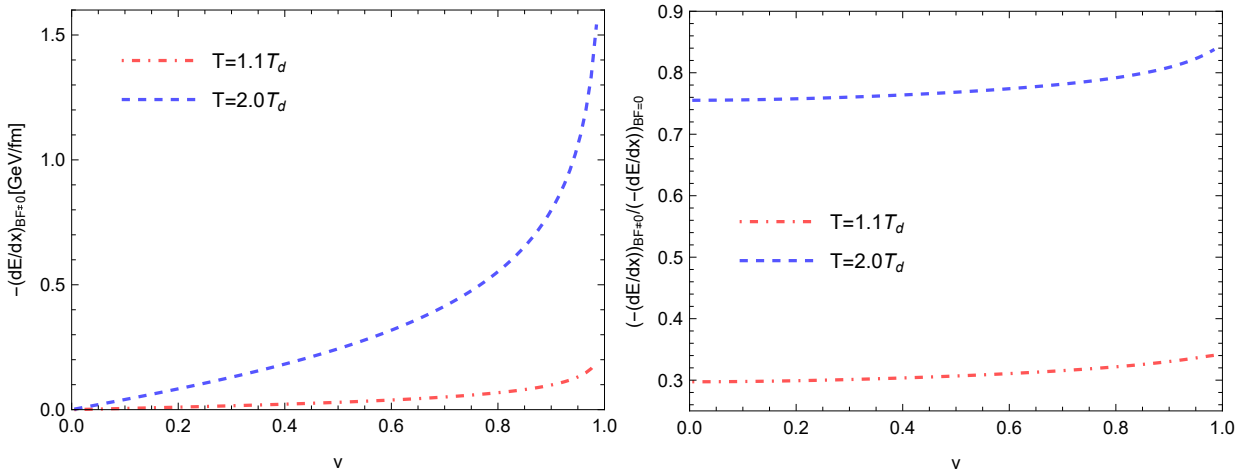


FIG. 3. Collisional energy loss in a background field (left) and the ratio of the energy loss with and without the background field (right) as a function of the heavy-quark velocity v at $T = 1.1T_d$ and $T = 2.0T_d$.

already mentioned above, the cutoff dependence of the total energy loss can be fixed through the following equation

$$-\frac{d}{dq^*}\left(\frac{dE}{dx}\right)_{\rm BF}\Big|_{q^*=q^{\rm pms}} = 0\,, \quad \text{with} \quad -\left(\frac{dE}{dx}\right)_{\rm BF} \equiv -\left(\frac{dE}{dx}\right)_{\rm BF,soft} - \left(\frac{dE}{dx}\right)_{\rm BF,hard}\,, \tag{51}$$

and thus the resulting $-dE/dx$ serves as a lower bound for the heavy-quark energy loss. We consider the collisional energy loss above the critical temperature $T_d = 270\,\text{MeV}$ for the $SU(3)$ gauge theory. Like the zero background field case, our numerical evaluations suggest that the energy loss shows a weak q^*-dependence as we decrease the coupling constant and eventually the energy loss coincides with the result given by the sum of Eqs. (30), (35) and (46) in the weak coupling limit where the q^*-dependence vanishes. For simplicity, we fix the coupling constant $\alpha_s = 0.3$, corresponding to its typical value in the semi-QGP region.

Given the temperature dependence of the background field in Eq. (18), the collisional energy loss depends only on the temperature of the plasma T and the velocity of the incident heavy-quark v. In the left plot of Fig. 3, we show $-dE/dx$ as a function of v at different temperatures. As we can see, the collisional energy loss increases with increasing velocity and a very rapid growth is observed in the large v region. Qualitatively, this coincides with the behavior as observed at vanishing background field[4]. It is also found that the heavy-quark energy loss is very sensitive to the temperature of the plasma, a significant enhancement appears when the medium moves from a region close to the critical temperature T_d into the perturbative deconfining region where we choose $T/T_d = 2.0$, corresponding to a rather small background field according to Eq. (18). A more direct way to see the influence of the background field is to consider the ratio of the energy loss with and without $\mathcal{Q}$ which can be found in the right plot of Fig. 3. In general, the collisional energy loss is reduced in the presence of a background field. For a given temperature, this ratio, or in other words, the suppression of the energy loss is insensitive to the velocity as long as v is not very large. Therefore, considering small and intermediate incident velocities, the collisional energy loss in a background field can be well approximated with the corresponding energy loss at $\mathcal{Q} = 0$ multiplied by an overall T-dependent factor. Notice that without considering the background field, the temperature dependence of the energy loss is very simple, *i.e.*, $-dE/dx \sim T^2$, when a fixed coupling is used. This is obviously not true for non-zero background field because the above mentioned factor changes with T significantly. On the other hand, a notable increase of the energy loss ratio can be found in the large v region[5].

In Fig. 4, we show the velocity dependence of the cutoff $q^{\rm pms}$ as determined by Eq. (51). No matter if a background field is taken into account or not, at a given temperature, $q^{\rm pms}$ exhibits a very weak v-dependence for small and intermediate velocities. However, a large increase of $q^{\rm pms}$ is found when the velocity gets large. Furthermore, for non-zero background field, $q^{\rm pms}/T$ changes when varying the temperature T at a fixed velocity. This is very different from the $\mathcal{Q} = 0$ case where $q^{\rm pms}/T$ becomes a constant. We can expect that in the high temperature limit, $q^{\rm pms}/T$

[4] The results without considering the background field correspond to those obtained in the perturbative theory. Besides setting $\mathcal{Q} = 0$, the mass scale M^2 in Eqs. (21) and (23) is also zero, namely, the $N^2 - 1$ gluons have the same screening mass m_D for $SU(N)$.

[5] In fact, our numerical results show that the energy loss ratio has a turning point at a velocity very close to 1, which is not visible in Fig. 3. However, in the ultra-relativistic limit, finite quark mass corrections become important there and for a realistic charm or bottom quark, calculation beyond our approximation is needed which we will discuss in the following.

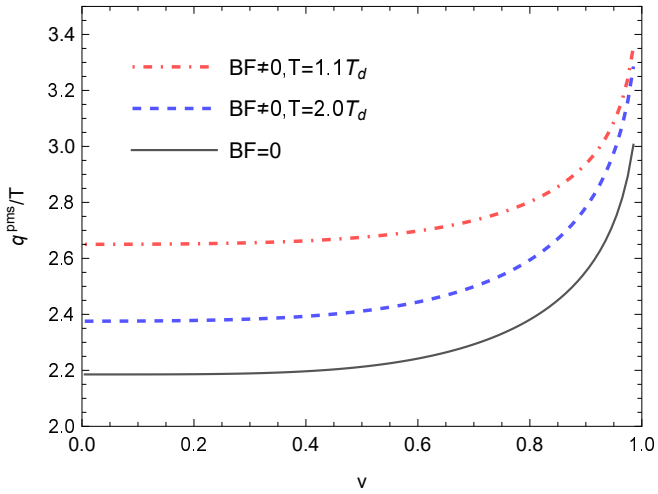


FIG. 4. The velocity dependence of $q^{\rm pms}/T$ with and without the background field.

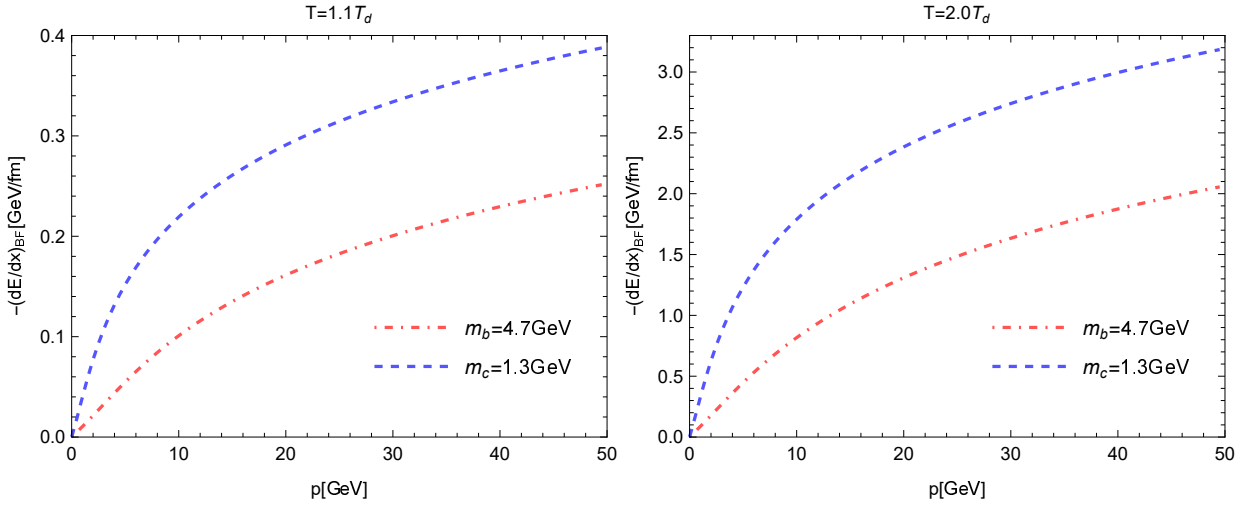


FIG. 5. Energy loss of the bottom ($m_b = 4.7$MeV) and charm quark ($m_c = 1.3$MeV) in a background field as a function of momentum p at $T = 1.1T_d$ (left), and $T = 2.0T_d$ (right).

would be identical to its counterpart at vanishing background field. Therefore, the background field has a significant impact on the cutoff near the deconfinement point, which gradually diminishes as the temperature increases.

It should be noted that in principle the above results are applicable for infinitely large quark mass because of the approximations used in our calculation. For a charm and bottom quark, finite mass corrections become important for very small ($v \to 0$) and large ($v \to 1$) velocities. On the one hand, when the incident velocity is smaller than the thermal velocity $\sim (T/m_Q)^{1/2}$, the heavy-quark is expected to gain energy from a collision. As a result, $-dE/dx$ must change sign at sufficiently small v [33]. On the other hand, a more accurate treatment on the energy loss is required for ultra-relativistic energies $E \gg m_Q^2/T$ where the largest momentum transfer used in Eq. (33) is no longer valid[6] and some dropped terms in the squared matrix element may also contribute significantly [24]. A crossover energy was estimated to be $\sim 1.8m_Q^2/T$ in Ref. [31] for QCD at vanishing background field. Accordingly, there exists an upper bound of the heavy-quark velocity given by $1 - 0.28(T/m_Q)^2$, above which finite mass corrections start to play a role. However, with the typical temperatures achieved in the high energy heavy-ion experiments, we can expect that for both charm and bottom quark our results are reliable up to very large incident velocities.

In order to see the flavor dependence of the collisional energy loss, we show in Fig. 5 $-dE/dx$ for the charm and bottom quark as a function of the momentum p at different temperatures. The momentum p of the heavy-quark can

[6] The largest momentum transfer $2k/(1-v)$ in Eq. (33) is only an approximation, which becomes exact when $E \ll m_Q^2/T$.

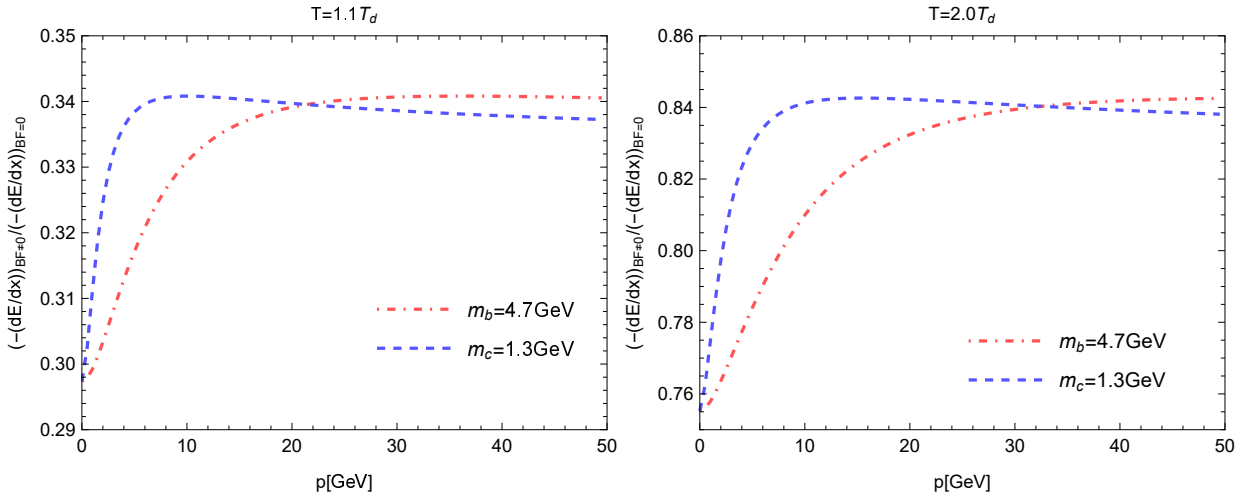


FIG. 6. The ratio of the energy loss with and without the background field as a function of the heavy-quark momentum p for charm and bottom at $T = 1.1T_d$ (left) and $T = 2.0T_d$ (right).

be related to its velocity through the equation $p = vm_Q/\sqrt{1-v^2}$ and we choose $m_c = 1.3\,\mathrm{GeV}$ and $m_b = 4.7\,\mathrm{GeV}$ in the numerical evaluations. Similar to that found in previous literature without considering the background field, the collisional energy loss at $\mathcal{Q} \neq 0$ grows quickly with increasing momentum and a charm quark loses more energy as compared to the heavier bottom quark. This conclusion can be understood from the fact that the energy loss increases monotonically with increasing v according to Fig. 3. Recall that for a given p, the incident quark with larger quark mass has a smaller velocity, and thus one can naturally expect that the collisional energy loss of a charm quark is more pronounced as compared to a bottom quark carrying the same momentum. In Fig. 6 we also plot the energy loss ratio as a function of the heavy-quark momentum. It can be shown that before it gets saturated at some large momentum p, the energy loss ratio has a quick increase in the small momentum region where the background field has a stronger impact on the suppression of $-dE/dx$ for a bottom quark. In principle, the observed p-dependence of the energy loss ratio can be also obtained from the corresponding v-dependence. Here, we should mention that an approximate constant ratio existing for $v \lesssim 0.5$ as shown in Fig. 3 is not contradictory to the results given in Fig. 6. In fact, such an observation indicates that for a charm quark, the energy loss ratio is almost independent on the momentum up to $p \sim 0.8\,\mathrm{GeV}$, which becomes $p \sim 2.7\,\mathrm{GeV}$ for a bottom quark. Due to a rather narrow region, this behavior becomes invisible in Fig. 6. On the other hand, the saturation of the energy loss ratio in the large momentum region is actually related to the very weak v-dependence of the momentum when it gets large. Finally, it is worth pointing out that for a given temperature, although the energy loss ratio shows a dependence on the heavy-quark velocity or momentum, its magnitude doesn't change dramatically when varying p or v.

As the background field strongly depends on the temperature, the most direct way to see the influence of the background field on the collisional energy loss is to study the energy loss ratio as a function of the temperature T. The corresponding results are shown in Fig. 7 where we consider two different heavy-quark velocities. Our results suggest that the background field suppresses $-dE/dx$ and the suppression is very significant in the entire semi-QGP region, from T_d to about $2 \sim 3T_d$. Especially, at the critical temperature, the heavy-quark only losses $\sim 10\%$ of the energy that would be lost in the case where $\mathcal{Q} = 0$. This can be attributed to the partial ionization of the plasma which corresponds to a reduced number density of the thermal partons in the presence of a background field. On the other hand, the energy loss ratio increases quickly with the increasing temperature, and above the semi-QGP region, non-zero $\mathcal{Q}$ only shows a very weak influence on the heavy-quark energy loss. This behavior can be understood by looking at the T-dependence of the background field. According to Eq. (18), the magnitude of the background field drops very quickly with increasing T and becomes negligible at temperatures larger than $2 \sim 3T_d$. In addition, there is only a moderate difference between the dashed curve for $v = 0.3$ and the dot-dashed curve for $v = 0.9$, therefore, the energy loss ratio has a weak dependence on the heavy-quark velocity at any given temperature. This is actually consistent with our above findings.

We also compare our results with the energy loss ratio obtained in Ref. [26] where the authors computed the collisional energy in a semi-QGP with the perturbation theory. They focused on the ultra-relativistic limit where $v \to 1$ and considered tree level Feynman diagrams with bare propagators. The infrared divergence associated with the t-channel contribution was regulated by choosing the Debye mass m_D to be the cutoff, and thus no resummation was considered for the soft momentum exchange. Consequently, the modification of the background field on the collisional energy loss turns out to be very simple, which is entirely encoded in the q-dependent factor $(1-3\mathsf{q})^2$ for

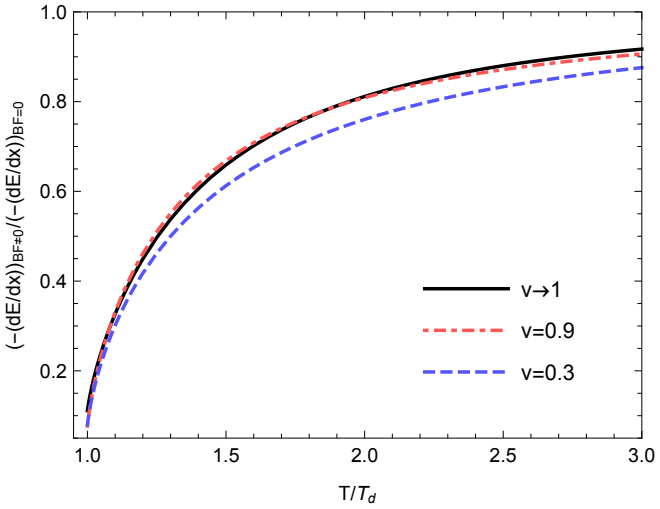


FIG. 7. The ratio of the energy loss with and without the background field as a function of the temperature for different heavy-quark velocities. The result in the ultra-relativistic limit $v \to 1$ comes from Ref. [26] where the values of the background field are based on Eq. (18).

$SU(3)$. This is exactly the same as what we found based on the result of $-dE/dx$ in the weak coupling limit[7]. The corresponding result in [26] is also presented. However, in a perturbative theory, non-zero background field cannot be generated self-consistently from the equation of motion. Therefore, the values of the background field were extracted from the Lattice simulations on the Polyakov loop in [26]. For a direct comparison, the solid curve in Fig. 7 is obtained with the background field determined by Eq. (18). It is interesting to see an agreement between the results from [26] and our results at large velocities even for a realistic coupling constant. As a result, we can expect a smooth transition to the ultra-relativistic limit $v \to 1$ where in principle our result is no longer reliable due to the finite quark mass. In addition, ignoring the weak dependence on the heavy-quark velocity, the energy loss ratio can be approximated by an overall factor $(1 - 3\mathsf{q})^2$ and according to our numerical results, the corresponding error is less than 10% when a realistic coupling constant is considered.

V. SUMMARY AND OUTLOOK

In this work, by utilizing a background field effective theory, we calculated the collisional energy loss of a heavy-quark propagating through a semi-QGP characterized by a non-trivial Polyakov loop. For temperatures close to the critical temperature, non-perturbative physics played an important role on the physical observables and such an effective theory which well described the thermodynamics of the medium turned out to be an ideal tool to study the relevant problems in a semi-QGP. We considered the energy loss due to elastic scatterings between the incident heavy-quark and the medium partons and showed that a non-zero background field, self-consistently generated from the equation of motion in the effective theory, had an important influence on both hard and soft contributions to the collisional energy loss. For hard scatterings, it was sufficient to use the bare propagator to compute the squared matrix element, while the thermal distribution functions of the medium partons were modified in the presence of a background field. For soft scatterings, it was necessary to take into account the HTL resummation in order to regulate the infrared divergence. Accordingly, the $\mathcal{Q}$-modified resummed gluon propagator which carried the correct screening information in the semi-QGP became the crucial ingredient to evaluate the energy loss associated with the soft processes.

Our results demonstrated that in the weak coupling limit, the collisional energy loss was independent on the momentum cutoff q^* which has been introduced to separate the hard and soft contributions to $-dE/dx$. This was exactly the same as what happened at $\mathcal{Q} = 0$ where the cutoff dependence was cancelled between the hard and soft contributions. Furthermore, when $\mathcal{Q} \neq 0$, the cancellation didn't rely on the explicit form of the background field which was considered as a function of the temperature. It was also found that in the limit $g \to 0$, keeping only the

[7] In the ultra-relativistic limit, the leading logarithmic contribution from t-channel is $\sim \ln(ET/m_D^2)$. In addition, the u-channel contribution cannot be neglected because it has a logarithmic enhancement $\sim \ln(ET/m_Q^2)$. However, in both cases, the modification due to the background field is simply given by $(1 - 3\mathsf{q})^2$.

leading logarithmic contribution $\sim \ln(1/g)$ in the collisional energy loss, the background field modification on $-dE/dx$ became very simple and the total energy loss at $\mathcal{Q} \neq 0$ was just given by a $\mathcal{Q}$-dependent factor multiplied by the corresponding energy loss at $\mathcal{Q} = 0$.

For a realistic coupling constant, on the other hand, the cutoff dependence existed in the collisional energy loss and $-dE/dx$ had a notable dependence on q^* when the coupling constant became large. We determined the cutoff by minimizing the total energy loss with respect to q^*, and thus the corresponding $-dE/dx$ could be considered as a lower bound for the heavy-quark energy loss. The numerical results suggested that although sharing some common properties with that at $\mathcal{Q} = 0$, the collisional energy loss was reduced in the presence of a non-zero $\mathcal{Q}$. Although the suppression was not sensitive to the heavy-quark velocity/momentum, it strongly depended on the temperature of the medium. Therefore, different from a simple $\sim T^2$ dependence at vanishing background field, $-dE/dx$ was expected to have a more involved T-dependence in a semi-QGP. In addition, the mass hierarchy of the energy loss, *i.e.*, a charm quark losing more energy than a bottom quark was also observed at non-zero $\mathcal{Q}$. With a given temperature, the background field led to a stronger suppression effect for a bottom quark although the magnitudes of the energy loss ratio for a charm and bottom quark didn't differ significantly from each other. Finally, as the temperature approached to the critical temperature from above, the fast decrease of the energy loss ratio indicated a significant reduction of the collisional energy loss. Especially, the energy loss of a heavy-quark was only $\sim 10\%$ of that in a vanishing background field when $T = T_d$. However, the influence of the background field became negligible above $2 \sim 3T_d$ where the saturated energy loss ratio got very close to one. Here, an interesting finding was that the simple form $(1 - 3\mathfrak{q})^2$ describing the $\mathcal{Q}$-modification on $-dE/dx$ in the weak coupling limit could be approximately applied to a moderate coupling constant if we ignored the weak velocity dependence of the energy loss ratio.

The background field effective theory is a useful theoretical tool to study the physics in a semi-QGP which is probably the most interesting region of the hot and dense medium being explored in high energy heavy-ion physics. Besides the heavy-quark collisional energy loss discussed in this work, it can be also adopted to study other relevant observables, such as the photon and dilepton production, the radiated energy loss and so on. However, a full QCD analysis requires a non-trivial generalization of the effective theory where one needs to non-perturbatively include the contributions from the thermal fermions. Further work along this line needs to be carried out in the future.

ACKNOWLEDGEMENTS

This work is supported by the NSFC of China under Project Nos. 12065004 and 12305135, by Guangxi Natural Science Foundation under Grant No. 2023GXNSFBA026027, and by the Central Government Guidance Funds for Local Scientific and Technological Development, China (No. Guike ZY22096024).

[1] B. A. Freedman and L. D. McLerran, Fermions and Gauge Vector Mesons at Finite Temperature and Density. 3. The Ground State Energy of a Relativistic Quark Gas, Phys. Rev. D **16**, 1169 (1977).
[2] E. V. Shuryak, Quark-Gluon Plasma and Hadronic Production of Leptons, Photons and Psions, Phys. Lett. B **78**, 150 (1978).
[3] J. O. Andersen, L. E. Leganger, M. Strickland, and N. Su, Three-loop HTL QCD thermodynamics, JHEP **08**, 053, arXiv:1103.2528 [hep-ph].
[4] N. Haque, A. Bandyopadhyay, J. O. Andersen, M. G. Mustafa, M. Strickland, and N. Su, Three-loop HTLpt thermodynamics at finite temperature and chemical potential, JHEP **05**, 027, arXiv:1402.6907 [hep-ph].
[5] Y. Hidaka and R. D. Pisarski, Suppression of the Shear Viscosity in a "semi" Quark Gluon Plasma, Phys. Rev. D **78**, 071501 (2008), arXiv:0803.0453 [hep-ph].
[6] S. Gupta, K. Huebner, and O. Kaczmarek, Renormalized Polyakov loops in many representations, Phys. Rev. D **77**, 034503 (2008), arXiv:0711.2251 [hep-lat].
[7] A. Mykkanen, M. Panero, and K. Rummukainen, Casimir scaling and renormalization of Polyakov loops in large-N gauge theories, JHEP **05**, 069, arXiv:1202.2762 [hep-lat].
[8] P. N. Meisinger, T. R. Miller, and M. C. Ogilvie, Phenomenological equations of state for the quark gluon plasma, Phys. Rev. D **65**, 034009 (2002), arXiv:hep-ph/0108009.
[9] Y. Hidaka and R. D. Pisarski, Effective models of a semi-quark-gluon plasma, Phys. Rev. D **104**, 074036 (2021), arXiv:2009.03903 [hep-ph].
[10] X.-N. Wang and M. Gyulassy, Gluon shadowing and jet quenching in A + A collisions at s**(1/2) = 200-GeV, Phys. Rev. Lett. **68**, 1480 (1992).
[11] M. Gyulassy, I. Vitev, X.-N. Wang, and B.-W. Zhang, Jet quenching and radiative energy loss in dense nuclear matter, , 123 (2004), arXiv:nucl-th/0302077.

[12] G.-Y. Qin and X.-N. Wang, Jet quenching in high-energy heavy-ion collisions, Int. J. Mod. Phys. E **24**, 1530014 (2015), arXiv:1511.00790 [hep-ph].

[13] E. Wang and X.-N. Wang, Jet tomography of dense and nuclear matter, Phys. Rev. Lett. **89**, 162301 (2002), arXiv:hep-ph/0202105.

[14] Y. L. Dokshitzer and D. E. Kharzeev, Heavy quark colorimetry of QCD matter, Phys. Lett. B **519**, 199 (2001), arXiv:hep-ph/0106202.

[15] B.-W. Zhang, E. Wang, and X.-N. Wang, Heavy quark energy loss in nuclear medium, Phys. Rev. Lett. **93**, 072301 (2004), arXiv:nucl-th/0309040.

[16] M. G. Mustafa, Energy loss of charm quarks in the quark-gluon plasma: Collisional versus radiative, Phys. Rev. C **72**, 014905 (2005), arXiv:hep-ph/0412402.

[17] S. Wicks, W. Horowitz, M. Djordjevic, and M. Gyulassy, Elastic, inelastic, and path length fluctuations in jet tomography, Nucl. Phys. A **784**, 426 (2007), arXiv:nucl-th/0512076.

[18] E. Braaten and M. H. Thoma, Energy loss of a heavy fermion in a hot plasma, Phys. Rev. D **44**, 1298 (1991).

[19] E. Braaten and M. H. Thoma, Energy loss of a heavy quark in the quark - gluon plasma, Phys. Rev. D **44**, R2625 (1991).

[20] P. Romatschke and M. Strickland, Energy loss of a heavy fermion in an anisotropic QED plasma, Phys. Rev. D **69**, 065005 (2004), arXiv:hep-ph/0309093.

[21] P. Romatschke and M. Strickland, Collisional energy loss of a heavy quark in an anisotropic quark-gluon plasma, Phys. Rev. D **71**, 125008 (2005), arXiv:hep-ph/0408275.

[22] C. P. Herzog, A. Karch, P. Kovtun, C. Kozcaz, and L. G. Yaffe, Energy loss of a heavy quark moving through N=4 supersymmetric Yang-Mills plasma, JHEP **07**, 013, arXiv:hep-th/0605158.

[23] M. Djordjevic, Collisional energy loss in a finite size QCD matter, Phys. Rev. C **74**, 064907 (2006), arXiv:nucl-th/0603066.

[24] S. Peigne and A. Peshier, Collisional Energy Loss of a Fast Muon in a Hot QED Plasma, Phys. Rev. D **77**, 014015 (2008), arXiv:0710.1266 [hep-ph].

[25] S. Peigne and A. Peshier, Collisional energy loss of a fast heavy quark in a quark-gluon plasma, Phys. Rev. D **77**, 114017 (2008), arXiv:0802.4364 [hep-ph].

[26] S. Lin, R. D. Pisarski, and V. V. Skokov, Collisional energy loss above the critical temperature in QCD, Phys. Lett. B **730**, 236 (2014), arXiv:1312.3340 [hep-ph].

[27] M. Elias, J. Peralta-Ramos, and E. Calzetta, Heavy quark collisional energy loss in the quark-gluon plasma including finite relaxation time, Phys. Rev. D **90**, 014038 (2014), arXiv:1404.7790 [hep-ph].

[28] C. Han, D.-f. Hou, B.-f. Jiang, and J.-r. Li, Jet energy loss in quark-gluon plasma: Kinetic theory with a Bhatnagar-Gross-Krook collisional kernel, Eur. Phys. J. A **53**, 205 (2017).

[29] M. Yousuf Jamal and V. Chandra, Energy loss of heavy quarks in the isotropic collisional hot QCD medium, Eur. Phys. J. C **79**, 761 (2019), arXiv:1907.12033 [nucl-th].

[30] J. Peng, K. Yu, S. Li, W. Xiong, F. Sun, and W. Xie, Unraveling collisional energy loss of a heavy quark in quark-gluon plasma, (2024), arXiv:2401.10644 [hep-ph].

[31] E. Braaten and T. C. Yuan, Calculation of screening in a hot plasma, Phys. Rev. Lett. **66**, 2183 (1991).

[32] M. H. Thoma, Collisional energy loss of fast charged particles in relativistic plasmas, J. Phys. G **26**, 1507 (2000), arXiv:hep-ph/0003016.

[33] M. H. Thoma and M. Gyulassy, Quark Damping and Energy Loss in the High Temperature QCD, Nucl. Phys. B **351**, 491 (1991).

[34] Y. Hidaka and R. D. Pisarski, Hard thermal loops, to quadratic order, in the background of a spatial 't Hooft loop, Phys. Rev. D **80**, 036004 (2009), [Erratum: Phys.Rev.D 102, 059902 (2020)], arXiv:0906.1751 [hep-ph].

[35] P. Cvitanovic, Group theory for Feynman diagrams in non-Abelian gauge theories, Phys. Rev. D **14**, 1536 (1976).

[36] C. P. Korthals Altes, Constrained effective potential in hot QCD, Nucl. Phys. B **420**, 637 (1994), arXiv:hep-th/9310195.

[37] A. Dumitru, Y. Guo, and C. P. Korthals Altes, Two-loop perturbative corrections to the thermal effective potential in gluodynamics, Phys. Rev. D **89**, 016009 (2014), arXiv:1305.6846 [hep-ph].

[38] U. Reinosa, J. Serreau, M. Tissier, and N. Wschebor, Two-loop study of the deconfinement transition in Yang-Mills theories: SU(3) and beyond, Phys. Rev. D **93**, 105002 (2016), arXiv:1511.07690 [hep-th].

[39] J. Maelger, U. Reinosa, and J. Serreau, Perturbative study of the QCD phase diagram for heavy quarks at nonzero chemical potential: Two-loop corrections, Phys. Rev. D **97**, 074027 (2018), arXiv:1710.01930 [hep-ph].

[40] Y. Guo and Q. Du, Two-loop perturbative corrections to the constrained effective potential in thermal QCD, JHEP **05**, 042, arXiv:1810.13090 [hep-ph].

[41] A. Dumitru, Y. Guo, Y. Hidaka, C. P. K. Altes, and R. D. Pisarski, How Wide is the Transition to Deconfinement?, Phys. Rev. D **83**, 034022 (2011), arXiv:1011.3820 [hep-ph].

[42] A. Dumitru, Y. Guo, Y. Hidaka, C. P. K. Altes, and R. D. Pisarski, Effective Matrix Model for Deconfinement in Pure Gauge Theories, Phys. Rev. D **86**, 105017 (2012), arXiv:1205.0137 [hep-ph].

[43] Y. Guo, Matrix Models for Deconfinement and Their Perturbative Corrections, JHEP **11**, 111, arXiv:1409.6539 [hep-ph].

[44] R. D. Pisarski and V. V. Skokov, Chiral matrix model of the semi-QGP in QCD, Phys. Rev. D **94**, 034015 (2016), arXiv:1604.00022 [hep-ph].

[45] K. Furuuchi, Confined phase in real time formalism and the fate of the world behind the horizon, Phys. Rev. D **73**, 046004 (2006), arXiv:hep-th/0510056.

[46] Y. Wang, Q. Du, and Y. Guo, Real-time hard-thermal-loop gluon self-energy in a semiquark-gluon plasma, Phys. Rev. D **106**, 054033 (2022), arXiv:2207.06039 [hep-ph].

[47] Y. Guo and Z. Kuang, Resummed gluon propagator and Debye screening effect in a holonomous plasma, Phys. Rev. D **104**, 014015 (2021), arXiv:2009.01516 [hep-ph].
[48] C. P. Korthals Altes, H. Nishimura, R. D. Pisarski, and V. V. Skokov, Conundrum for the free energy of a holonomous gluonic plasma at cubic order, Phys. Lett. B **803**, 135336 (2020), arXiv:1911.10209 [hep-th].
[49] C. P. Korthals Altes, H. Nishimura, R. D. Pisarski, and V. V. Skokov, Free energy of a Holonomous Plasma, Phys. Rev. D **101**, 094025 (2020), arXiv:2002.00968 [hep-ph].

arXiv:2402.17974v1 [hep-ph] 28 Feb 2024

All-heavy pentaquarks in a nonrelativistic potential quark model

Zhi-Biao Liang[1], Feng-Xiao Liu[2,3] *, and Xian-Hui Zhong[1,3] †

1) Department of Physics, Hunan Normal University, and Key Laboratory of Low-Dimensional Quantum Structures and Quantum Control of Ministry of Education, Changsha 410081, China

2) Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China and

3) Synergetic Innovation Center for Quantum Effects and Applications (SICQEA), Hunan Normal University, Changsha 410081, China

In a nonrelativistic potential quark model framework, we carry out a serious calculation of the all-heavy pentaquarks by adopting the explicitly correlated Gaussian method. A complete mass spectrum for the $1S$ states is obtained. For the $ccccc\bar{c}$, $cccc\bar{b}$, $bbbbc\bar{c}$, $bbbbb\bar{b}$, $ccbbc\bar{c}$, and $bbccb\bar{b}$ systems, the obtained states are compact and lie far above the lowest dissociation baryon-meson threshold. While, in the $cccbc\bar{c}$, $cccbb\bar{b}$, $bbbcc\bar{c}$, and $bbbcb\bar{b}$ systems with $\{123\}4\bar{5}$ symmetry, the two low-lying configurations with $J^P = 5/2^-$ and $3/2^-$ have a typical molecular structure due to the special role of the color-Coulomb interactions, they may be good candidates of stable states below the dissociation baryon-meson thresholds.

I. INTRODUCTION

Searching for genuine exotic multiquark states beyond the conventional meson ($q\bar{q}$) and baryon (qqq) states has been one of the most important initiatives since the establishment of quark model in 1964 [1–3]. Since the discovery of $X(3872)$ by Belle in 2003 [4], many tetraquark candidates, such as the series hidden-charmed/bottom XYZ states [5], the doubly-charmed state $T_{cc}(3875)^+$ [6, 7], and charmed-strange states [8, 9], have been observed in experiments. Furthermore, the exotic P_c [10–12] and P_{cs} [13] states as candidates of pentaquark states were also reported by the LHCb collaboration. All of these observed exotic states contains two or three light quarks, it is very difficult to determine whether they are hadronic molecular states or genuine multiquark states due to the role of light-meson exchanges. This dilemma should be largely alleviated for the all-heavy multiquarks. They are most likely to be genuine multiquark states since there is no light-meson exchange potential, which is often needed by the formation molecules. Thus, the study of the all-heavy multiquarks may provide an interesting way for establishing the genuine multiquark states.

Impressively, some all-heavy multiquark states have been observed in LHC experiments. In 2020, the LHCb Collaboration observed a narrow structure $X(6900)$ together a broad structure ranging from 6.2 to 6.8 GeV in the di-J/ψ invariant mass spectrum [14]. Later in 2022, the $X(6900)$ was confirmed in the same final state by both the ATLAS [15] and CMS [16] collaborations. Moreover, in the lower mass region the CMS measurements show that a clear resonance $X(6600)$ lies in the di-J/ψ spectrum. These clear structures may be evidences for genuine all-charmed tetraquark $cc\bar{c}\bar{c}$ states. The discovery of $cc\bar{c}\bar{c}$ states has demonstrated the powerful abilities of LHC in productions of fully-heavy hadrons, and also indicates that the all-heavy pentaquark should be exist. Thus, one may expect to observe some all-heavy pentaquarks in forthcoming experiments.

Stimulated by these, some relative studies of the all-heavy pentaquarks have been carried out within various models in recent years. The all-charmed and -bottom pentaquarks were studied with the QCD sum rules [17, 18]. For the baryon-meson type, the masses of $ccccc\bar{c}$ and $bbbbb\bar{b}$ are predicted to be ~ 7.41 GeV and ~ 21.60 GeV, respectively in Ref. [17], which are notably smaller than those of 7.93 ± 0.15 GeV and 23.91 ± 0.15 GeV predicted for the diquark-diquark-antiquark type in Ref. [18]. In Ref. [19], the mass spectra of the $1S$-wave all-heavy pentaquarks were systematically studied with the simple chromomagnetic interaction (CMI) model. The masses of $ccccc\bar{c}$ and $bbbbb\bar{b}$ are predicted to be ~ 7.9 GeV and ~ 23.8 GeV, respectively, which are slightly above the lowest dissociation baryon-meson mass threshold. While, there may exist a stable $J^P = 3/2^-$ $bbccb\bar{b}$ states with a mass of ~ 17.4 MeV. In Ref. [20], the all-heavy pentaquarks were further studied within the MIT bag model by including the chromomagnetic interaction. In this framework, the masses of $ccccc\bar{c}$ and $bbbbb\bar{b}$ are predicted to be ~ 8.2 GeV and ~ 24.8 GeV, and no stable states below the dissociation baryon-meson thresholds are found.

In the recent two years, the all-heavy pentaquarks have been studied with more comprehensive potential quark models, where besides the chromomagnetic interaction, the confining and Coulomb-like potentials are also included in the calculations. In Refs. [21, 22], the authors studied the all-heavy pentaquarks by using the resonating group method (RGM), in which two-cluster approximation is adopted. They obtained one possible stable $J^P = 1/2^-$ $ccccc\bar{c}$ state with a mass of ~ 7.9 GeV, and two possible stable $bbbbb\bar{b}$ states with $J^P = 1/2^-$ and $J^P = 3/2^-$ in the mass range of ~ 23.8 GeV, and several possible stable candidates in the charmed-bottom pentaquarks. In Ref. [23], considering the baryon-meson and diquark-diquark-antiquark configurations, the authors obtained several narrow resonances above 8.0 GeV and 24.0 GeV for the $ccccc\bar{c}$ and $bbbbb\bar{b}$ pentaquarks, respectively, based on the Gaussian expansion method combined with a complex-scaling range approach. Recently, a more serious dynamical calculation beyond the cluster approximation has been carried out by using the variational method with the trial spatial wave function in a simple Gaussian form [24]. No stable pentaquarks states below the dissociation baryon-meson thresholds are found. The masses for the ground $ccccc\bar{c}$ and $bbbbb\bar{b}$ states are predicted

*E-mail: liufx@ihep.ac.cn
†E-mail: zhongxh@hunnu.edu.cn

to be $\sim$ 8.2 GeV and $\sim$ 24.2 GeV, respectively, which are consistent with those obtained with the trial function in the most general Gaussian form [25]. However, due to the simple Gaussian basis adopted in Ref [24], no stable solutions are obtained for many states in the all-heavy pentaquarks containing unequal mass quarks, such as $cccb\bar{b}$, $cccb\bar{c}$, $bbbc\bar{c}$, and $bbbc\bar{b}$.

In this work, we carry out a dynamical calculation of the mass spectra of the $1S$-wave all-heavy pentaquarks with a nonrelativistic potential quark model (NRPQM). This model is based on the Hamiltonian of the Cornell model [26], whose parameters have been well determined based on the successful description of the heavy quarkonium spectra [27–29], and triply charmed and bottom baryon spectra [30]. This model has been extended to study the mass spectra of all-heavy tetraquarks by our group in Refs. [31–33]. To solve the five-body problem accurately, we adopt the explicitly correlated Gaussian (ECG) method [34, 35], in which a variational trial spatial wave function is expanded with the most general nondiagonal Gaussian basis functions associated to the Jacobi coordinates. This is one of the most powerful approaches currently used for calculating the properties of few-body systems, and offers great flexibility, and high accuracy [35]. Based on the ECG numerical method, the mass spectrum of charmed-strange tetraquarks was successfully predicted by our group, recently [36]. The main purposes of this work are (i) to give a reliable prediction of the mass spectra for the fully-heavy pentaquarks based on our serious studies; (ii) to reveal the inner structure special properties of the pentaquark states due to the complex symmetry and dynamics; (iii) to assess the differences of the predictions from different potential models with different numerical methods.

As follows, we first give a brief introduction to our framework. We then give our numerical results and discussions for the S-wave all-heavy pentaquarks states in Sec. III. Finally, a summary is given in Sec. IV.

II. FRAMEWORK

In this section, we first give a brief review of the nonrelativistic potential quark model. Then, we introduce the quark model classification of the pentaquarks based on the symmetries. Finally, we give a brief introduction to the ECG method what we adopt in the calculations.

A. Hamiltonian

In NRPQM, the Hamiltonian for a pentaquark system is given by

$$H = \sum_{i}^{5}(m_i + T_i) - T_{cm} + \sum_{i<j}^{5} V_{ij}(r_{ij}), \tag{1}$$

where m_i and T_i represent the mass and kinetic energy of the i-th quark, respectively; T_{cm} represents the kinetic energy of the center of mass of the pentaquark system; $r_{ij} = |\boldsymbol{r_i} - \boldsymbol{r_j}|$ is the distance between the i-th and j-th quarks; while V_{ij} is an effective potential between them. In this work, the effective potential V_{ij} is adopted a widely used form,

$$\begin{aligned} V_{ij}(r_{ij}) = &-\frac{3}{16}\left(\boldsymbol{\lambda}_i \cdot \boldsymbol{\lambda}_j\right)\left(br_{ij} - \frac{4}{3}\frac{\alpha_{ij}}{r_{ij}} + C_0\right) \\ &- \frac{\alpha_{ij}}{4}\left(\boldsymbol{\lambda}_i \cdot \boldsymbol{\lambda}_j\right) \cdot \left\{\frac{\pi}{2} \cdot \frac{\sigma_{ij}^3 e^{-\sigma_{ij}^2 r_{ij}^2}}{\pi^{3/2}} \cdot \frac{16}{3m_i m_j}\left(\mathbf{S}_i \cdot \mathbf{S}_j\right)\right\}, \end{aligned} \tag{2}$$

where $\boldsymbol{\lambda}_{i,j}$ are the color operators acting on the i, j-th quarks, $\boldsymbol{S}_{i,j}$ represent the spin operators of the i, j-th quarks; while C_0 represents the zero-point energy. The parameters b and α_{ij} denote the strength of the confinement and strong coupling of the one-gluon-exchange potential, respectively. In this work, the parameter set $\{b, \alpha_{ij}, C_0, \sigma_{ij}, m_i\}$ is taken the same as that in our previous works [30, 33] and are listed in Table I. These parameters were determined through the fitting of meson spectrum. Finally, it should be mentioned that in this work we only consider the low-lying $1S$-wave pentaquark states without any orbital excitations, thus, the spin-orbit and tensor potentials are not included.

TABLE I: The potential model parameters.

m_c/m_b(GeV)	1.483/4.852
$\alpha_{cc}/\alpha_{bc}/\alpha_{bb}$	0.5461/0.5021/0.4311
$\sigma_{cc}/\sigma_{bc}/\sigma_{bb}$(GeV)	1.1384/1.3000/2.3200
b(GeV^2)	0.1425
C_0(GeV)	0

B. Pentaquark configuration

The pentaquark configuration can be expressed as a product the flavor, spatial, spin, and color parts, i.e.,

$$|\Psi_{5q}\rangle = |flavor\rangle \otimes |spatial\rangle \otimes |spin\rangle \otimes |color\rangle. \tag{3}$$

In the flavor space, the available configurations for all-heavy pentaquarks are "$ccccc\bar{c}$, $cccc\bar{b}$, $bbbbb\bar{b}$, $bbbb\bar{c}$, $cccb\bar{b}$, $cccb\bar{c}$, $bbbc\bar{c}$, $bbbc\bar{b}$, $ccbb\bar{c}$, $bbcc\bar{b}$". The spin and color parts can be constructed according to the permutation symmetry. For the low-lying $1S$ pentaquark states without any excitations, their spatial wave functions are symmetrical when exchanging the coordinates of any two quarks. To obtain the spatial part of the pentaquarks, one should solve the Schröinger equation.

Due to the pentaquark system containing identical quarks, it must satisfy the Pauli principle. Considering the permutation symmetry of identical quarks, the pentaquark configurations can be classified into three categories, which are denoted by

$$\begin{aligned} \{1234\}\bar{5} &: \{cccc\}\bar{c}, \{cccc\}\bar{b}, \{bbbb\}\bar{b}, \{bbbb\}\bar{c}, \\ \{123\}4\bar{5} &: \{ccc\}b\bar{b}, \{ccc\}b\bar{c}, \{bbb\}c\bar{c}, \{bbb\}c\bar{b}, \\ \{12\}\{34\}\bar{5} &: \{cc\}\{bb\}\bar{c}, \{bb\}\{cc\}\bar{b}. \end{aligned} \tag{4}$$

The quarks in {} should satisfy the requirements of the permutation symmetry.

For a pentaquark system, one can construct three colorless configurations (color singlets) in the color space based on the SU(3)-group representation theory. Their representations with Young tableaux are given by [39]

$$C_1 = \begin{array}{|c|c|}\hline 1&4\\\hline 2\\\cline{1-1} 3\\\cline{1-1}\end{array}_{\,3} \otimes (5)_{\bar{3}},\ C_2 = \begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array}_{\,3} \otimes (5)_{\bar{3}},\ C_3 = \begin{array}{|c|c|}\hline 1&3\\\hline 2\\\cline{1-1} 4\\\cline{1-1}\end{array}_{\,3} \otimes (5)_{\bar{3}}. \quad (5)$$

By using the C-G coefficients of SU(3) group [37], one can explicitly write out the three color configurations, which are can be found in the literature, such as Ref. [20].

For a pentaquark system, the total spin quantum numbers are possibly $J = 5/2, 3/2, 1/2$. Based on the SU(2) symmetry, one can construct the configurations with $J = 5/2, 3/2, 1/2$ in the spin space. Their representations with Young tableaux are given by

$$J = \frac{5}{2} : S_1 = \begin{array}{|c|c|c|c|}\hline 1&2&3&4\\\hline\end{array}5. \quad (6)$$

$$J = \frac{3}{2} : S_2 = \begin{array}{|c|c|c|c|}\hline 1&2&3&4\\\hline\end{array}\,5, \qquad S_3 = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4\\\cline{1-1}\end{array}\,5,$$
$$S_4 = \begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2\\\cline{1-1}\end{array}\,5, \qquad S_5 = \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3\\\cline{1-1}\end{array}\,5. \quad (7)$$

$$J = \frac{1}{2} : S_6 = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5\\\cline{1-2}\end{array}, \quad S_7 = \begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2&5\\\cline{1-2}\end{array}, \quad S_8 = \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3&5\\\cline{1-2}\end{array},$$
$$S_9 = \begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\cline{1-2}\end{array}\,5, \quad S_{10} = \begin{array}{|c|c|}\hline 1&3\\\hline 2&4\\\cline{1-2}\end{array}\,5. \quad (8)$$

Combining the C-G coefficient of SU(2) group, one can obtain the spin wave function $\psi_{S_i}|JJ_z\rangle$ corresponding to a special Young tableau.

With the configurations of the color and spin spaces, one can further construct the configurations of the spin⊗color space. For the low-lying $1S$-wave pentaquark configurations, both the spatial and flavor parts are symmetric when one exchanges any two identical quarks, thus, the spin⊗color part should be antisymmetric. To get the coupling configurations in the spin⊗color space, we need the C-G coefficients of S_4 group, which are taken from Ref. [38].

For a pentaquark system with $\{1234\}\bar{5}$ symmetry, there are two configurations with $J^P = \frac{3}{2}^-$ and $\frac{1}{2}^-$:

$$1S_{\frac{3}{2}^-}(\{1234\}\bar{5}) = \frac{1}{\sqrt{3}}(C_1S_3 + C_2S_4 - C_3S_5),$$
$$1S_{\frac{1}{2}^-}(\{1234\}\bar{5}) = \frac{1}{\sqrt{3}}(C_1S_6 + C_2S_7 - C_3S_8).$$

It should be mentioned that there is no $J^P = \frac{5}{2}^-$ configurations. In this case, the $\{1234\}\bar{5}$ symmetry requires the color configurations are fully antisymmetric since the spin configurations are fully symmetric, however, the color configurations given in Eq. (5) cannot satisfy this requirement.

For a pentaquark system with $\{123\}4\bar{5}$ symmetry, one can obtain a $J^P = \frac{5}{2}^-$ configuration,

$$1S_{\frac{5}{2}^-}(\{123\}4\bar{5}) = C_1S_1,$$

three $J^P = \frac{3}{2}^-$ configurations,

$$1S_{\frac{3}{2}^-}(\{123\}4\bar{5})_1 = \sqrt{\frac{1}{3}}(C_2S_4 - C_3S_5 + C_1S_3),$$
$$1S_{\frac{3}{2}^-}(\{123\}4\bar{5})_2 = \sqrt{\frac{1}{6}}(-C_2S_4 + C_3S_5 + 2C_1S_3),$$
$$1S_{\frac{3}{2}^-}(\{123\}4\bar{5})_3 = C_1S_2,$$

and three $J^P = \frac{1}{2}^-$ configurations,

$$1S_{\frac{1}{2}^-}(\{123\}4\bar{5})_1 = \sqrt{\frac{1}{3}}(C_2S_7 - C_3S_8 + C_1S_6),$$
$$1S_{\frac{3}{2}^-}(\{123\}4\bar{5})_2 = \sqrt{\frac{1}{6}}(-C_2S_7 + C_3S_8 + 2C_1S_6),$$
$$1S_{\frac{3}{2}^-}(\{123\}4\bar{5})_3 = \sqrt{\frac{1}{2}}(C_2S_{10} - C_3S_9).$$

For a pentaquark system with $\{12\}\{34\}\bar{5}$ symmetry, there is a $J^P = \frac{5}{2}^-$ configuration,

$$1S_{\frac{5}{2}^-}(\{12\}\{34\}\bar{5}) = \sqrt{\frac{2}{3}}C_3S_1 - \sqrt{\frac{1}{3}}C_1S_1,$$

four $J^P = \frac{3}{2}^-$ configurations,

$$1S_{\frac{3}{2}^-}(\{12\}\{34\}\bar{5})_1 = \sqrt{\frac{1}{3}}(C_2S_4 - C_3S_5 + C_1S_3),$$
$$1S_{\frac{3}{2}^-}(\{12\}\{34\}\bar{5})_2 = \sqrt{\frac{1}{3}}(C_1S_5 - C_3S_3) - \sqrt{\frac{1}{6}}(C_2S_4 + C_3S_5),$$
$$1S_{\frac{3}{2}^-}(\{12\}\{34\}\bar{5})_3 = \frac{\sqrt{2}}{6}(3C_2S_4 + C_3S_5 - 2C_1S_3) + \frac{1}{3}(C_1S_5 - C_3S_3),$$
$$1S_{\frac{3}{2}^-}(\{12\}\{34\}\bar{5})_4 = \sqrt{\frac{2}{3}}C_3S_2 - \sqrt{\frac{1}{3}}C_1S_2,$$

and four $J^P = \frac{1}{2}^-$ configurations,

$$1S_{\frac{1}{2}^-}(\{12\}\{34\}\bar{5})_1 = \sqrt{\frac{1}{3}}(C_2S_7 - C_3S_8 + C_1S_6),$$
$$1S_{\frac{1}{2}^-}(\{12\}\{34\}\bar{5})_2 = \sqrt{\frac{1}{3}}(C_1S_8 - C_3S_6) - \sqrt{\frac{1}{6}}(C_2S_7 + C_3S_8),$$
$$1S_{\frac{1}{2}^-}(\{12\}\{34\}\bar{5})_3 = \frac{\sqrt{2}}{6}(3C_2S_7 + C_3S_8 - 2C_1S_6) + \frac{1}{3}(C_1S_8 - C_3S_6),$$
$$1S_{\frac{1}{2}^-}(\{12\}\{34\}\bar{5})_4 = \sqrt{\frac{2}{3}}C_3S_9 - \sqrt{\frac{1}{3}}C_1S_9.$$

To know more technical details for constructing the pentaquark configurations in the spin⊗color space, one can refer to the works [19, 20, 24, 39].

C. Numerical method

To solve the five-body problem accurately, we adopt the ECG method [34, 35]. It is a well-established variational method to solve quantum few-body problems in molecular, atomic, and nuclear physics. The spatial part of the wave function for a pentaquark system without any angular momenta is expanded in terms of ECG basis set. Such a basis function can be expressed as

$$\psi(\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3,\boldsymbol{r}_4,\boldsymbol{r}_5)=\exp\left[-\sum_{i<j}^{5}a_{ij}\boldsymbol{r}_{ij}^2\right], \tag{9}$$

where a_{ij} are variational parameters. For a pentaquark system with $\{1234\}\bar{5}$ symmetry, one has two independent variational parameters, i.e., $a_{12}=a_{13}=a_{14}=a_{23}=a_{24}=a_{34}=a$ and $a_{15}=a_{25}=a_{35}=a_{45}=b$, due to the symmetry of identical quarks. Similarly, for a pentaquark system with $\{123\}4\bar{5}$-symmetry, there are four variational parameters, $a_{12}=a_{13}=a_{23}=a$, $a_{14}=a_{24}=a_{34}=b$, $a_{15}=a_{25}=a_{35}=c$ and $a_{45}=d$. While for a pentaquark system with $\{12\}\{34\}\bar{5}$ symmetry, there are also five independent variational parameters, $a_{12}=a$, $a_{13}=a_{14}=a_{23}=a_{24}=b$, $a_{15}=a_{25}=c$, $a_{34}=d$ and $a_{35}=a_{45}=f$.

It is convenient to use a set of the Jacobi coordinates $\boldsymbol{\xi}=(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\boldsymbol{\xi}_3,\boldsymbol{\xi}_4)$ instead of position vectors $\boldsymbol{r}_i$ $(i=1-5)$. For example, one can take a set of Jacobi coordinates as follows,

$$\begin{aligned}
\boldsymbol{\xi_1}&=\boldsymbol{r_1}-\boldsymbol{r_2},\\
\boldsymbol{\xi_2}&=\boldsymbol{r_3}-\frac{m_1\boldsymbol{r_1}+m_2\boldsymbol{r_2}}{m_1+m_2},\\
\boldsymbol{\xi_3}&=\boldsymbol{r_4}-\boldsymbol{r_5},\\
\boldsymbol{\xi_4}&=\frac{m_1\boldsymbol{r_1}+m_2\boldsymbol{r_2}+m_3\boldsymbol{r_3}}{m_1+m_2+m_3}-\frac{m_4\boldsymbol{r_4}+m_5\boldsymbol{r_5}}{m_4+m_5}.
\end{aligned} \tag{10}$$

Then, the correlated Gaussian basis function ψ can be rewritten as

$$G(\boldsymbol{\xi},A)=\exp(-\boldsymbol{\xi}^T A\boldsymbol{\xi}), \tag{11}$$

where A is a 4×4 matrix, which is related to the variational parameters. The spatial part of the trial wave function can be expanded with a set of correlated Gaussians:

$$\Psi(\boldsymbol{\xi},A)=\sum_{k=1}^{N}C_k G(\boldsymbol{\xi},A_k), \tag{12}$$

where N is the number of Gaussian basis functions. The accuracy of the trial function depends on the number N and the nonlinear parameter matrix A_k. In our calculations, following the method of Ref. [40], we let the variational parameters form a geometric progression. For example, for a variational parameter a, we take

$$a_i=\frac{1}{2\left(a_1 q^{i-1}\right)^2}\quad(i=1,\cdots,n_{max}). \tag{13}$$

The Gaussian size parameters $\{a_1,q,n_{max}\}$ will be determined through the variation method. In the calculations, the final results should be stable and independent with these parameters.

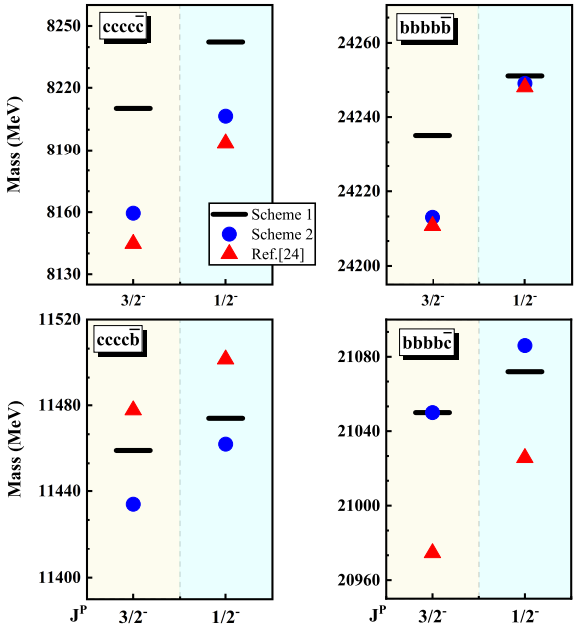


FIG. 1: Spectrum for all-heavy pentaquarks with $\{1234\}\bar{5}$ symmetry. The black short solid lines stand for the predictions based on the potential model of the present work with the ECG numerical method (Scheme 1). The solid circles and triangles stand for the predictions based on the potential model of Ref. [24], which are solved with the ECG numerical method in this work (Scheme 2) and with the simple diagonal Gaussian basis in Ref. [24], respectively.

For a given pentaquark configuration, one can work out the Hamiltonian matrix elements,

$$H_{kk'}=\langle\psi_{CS}G(\boldsymbol{\xi},A_k)|H|\psi_{CS}G(\boldsymbol{\xi},A_{k'})\rangle, \tag{14}$$

where ψ_{CS} is the spin-color wave function. Then, by solving the generalized matrix eigenvalue problem,

$$\sum_{k'=1}^{N}(H_{kk'}-EN_{kk'})C_{k'}=0, \tag{15}$$

one can obtain the eigenenergy E, and the expansion coefficients $\{C_k\}$. Where $N_{kk'}$ is an overlap factor defined by $N_{kk'}=\langle G(\boldsymbol{\xi},A_k)|G(\boldsymbol{\xi},A_{k'})$.

Finally, it should be mentioned that in the calculations we can obtained stable solutions when we take $N=n^a_{max}\times n^b_{max}=8\times8=64$ for the system with $\{1234\}\bar{5}$ symmetry, $N=n^a_{max}\times n^b_{max}\times n^c_{max}\times n^d_{max}=5\times5\times5\times5=625$ for the system with $\{123\}4\bar{5}$ symmetry, while $N=n^a_{max}\times n^b_{max}\times n^c_{max}\times n^d_{max}\times n^f_{max}=4\times4\times4\times4\times4=1024$ for the system with $\{12\}\{34\}\bar{5}$ symmetry.

III. RESULTS AND DISCUSSIONS

A. States with $\{1234\}\bar{5}$ symmetry

In the all-heavy pentaquarks, the $ccccc\bar{c}$, $cccc\bar{b}$, $bbbb\bar{c}$, and $bbbb\bar{b}$ systems have the $\{1234\}\bar{5}$ symmetry. Considering the

TABLE II: Predicted mass spectra of $1S$ states for the $ccccc\bar{c}$, $bbbbb\bar{b}$, $ccccb\bar{b}$, and $bbbbc\bar{c}$ systems compared with those of other works. The unit is MeV. "Scheme 1" stands for the results obtained with the ECG method based on the potential model of the present work, while "Scheme 2" stands for the results obtained with the ECG method based on the potential model of Ref. [24].

	J^P	Configuration	Scheme 1	Scheme 2	[24]	[19]	[20]	[23]	[21]	[18]	[17]
$ccccc\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}5$)	8210	8159	8145	7864	8229	8095			7410^{+270}_{-310}
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}5$)	8242	8206	8193	7949	8262	8045	7892	7930 ± 150	
$bbbbb\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}5$)	24235	24213	24211	23775	24761	24035	23748		21600^{+730}_{-220}
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}5$)	24251	24249	24248	23821	24770	24035	23810	23910±150	
$ccccb\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}5$)	11459	11434	11478	11130	11569				
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}5$)	11474	11462	11502	11177	11582				
$bbbbc\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}5$)	21050	21050	20975	20652	21472				
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}5$)	21072	21086	21026	20699	21491				

TABLE III: The average contributions of each part of the Hamiltonian (in MeV) and the root mean square radii (in fm) for the $1S$-wave all-heavy pentaquark configurations with the $\{1234\}\bar{5}$ symmetry.

				Each part contribution of Hamiltonian				RMS Radius	
	J^P	Configuration	Mass	$\langle T\rangle$	$\langle V^{Conf}\rangle$	$\langle V^{Coul}\rangle$	$\langle V^{SS}\rangle$	$r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34}$	$r_{15}, r_{25}, r_{35}, r_{45}$
$ccccc\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}\bar{5}$)	8210	886	860.7	-977.8	26.1	0.5248	0.5176
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}\bar{5}$)	8242	846	879.8	-954.9	56.3	0.5327	0.5337
$ccccb\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}\bar{5}$)	11459	858	800	-1017	34	0.5135	0.4413
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}\bar{5}$)	11474	842	807	-1007	48	0.5172	0.4467
$bbbbc\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}\bar{5}$)	21050	919	548	-1317	9	0.2988	0.3864
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}\bar{5}$)	21072	889	557	-1296	30	0.3017	0.3957
$bbbbb\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-($\{1234\}\bar{5}$)	24235	931	467	-1436	14	0.2842	0.2839
	$1/2^-$	$1S_{\frac{1}{2}}$-($\{1234\}\bar{5}$)	24251	905	473	-1416	29	0.2870	0.2890

permutation symmetry of identical quarks, there are two states with $J^P = 3/2^-$ and $J^P = 1/2^-$ for each pentaquark system. The mass splitting between them is very small ($\sim 15-30$ MeV), the $J = 1/2^-$ slightly lies above the $J = 3/2^-$ state. The predicted mass spectrum has been given in Table II and also shown in Fig. 1.

For the $ccccc\bar{c}$ system, the masses of the ground ($1S$) states are predicted to be ~ 8.2 GeV, which is ~ 300 MeV above the mass threshold of $J/\psi\Omega_{ccc}(3/2^+)$ according to the predictions of the triply heavy baryon spectrum [30]. For the $bbbbb\bar{b}$ system, the masses of the $1S$ states are predicted to be ~ 24.2 GeV, which is ~ 340 MeV above the mass threshold of $\Upsilon(1S)\Omega_{bbb}(3/2^+)$. For the $ccccb\bar{b}$ system, the masses of the $1S$ states are predicted to be ~ 11.5 GeV, which is ~ 400 MeV above the mass threshold of $B_c\Omega_{ccc}(3/2^+)$. While, for the $bbbbc\bar{c}$ system, the masses of the two $1S$ states are predicted to be ~ 21.1 GeV, which is ~ 400 MeV above the mass threshold of $B_c\Omega_{bbb}(3/2^+)$.

Similar studies have been carried out within another potential model approach in a recent work by An *et al.* [24]. For a comparison, we listed their results in Table II and also shown in Fig. 1. It is seen that our predicted mass spectra are consistent with those in Ref. [24]. The differences of the predictions between ours and those in Ref. [24] are within 100 MeV. To further see the numerical method dependency of the predictions, by using the same Hamiltonian of Ref. [24] we calculate the mass spectrum with the ECG method as well. The results are listed in Table II and shown in Fig. 1. From Fig. 1, one can see that the results for the $ccccc\bar{c}$ and $bbbbb\bar{b}$ systems obtained with the ECG method (solid circles) are very close to those (solid triangles) obtained with the simple diagonal Gaussian basis in Ref. [24]. However, for the $ccccb\bar{b}$ and $bbbbc\bar{c}$ systems containing unequal mass quarks, there is a considerable difference of $\sim 40-80$ MeV in the results obtained with the two different numerical methods. Some other studies about these pentaquarks can be found in the literature based on the MIT bag model [20], CMI model [19], Lattice-QCD inspired quark model [23], chiral quark model/quark delocalization color screening model [21], and QCD sum rule approach [17, 18], the results are collected in Table II for a comparison. There are large discrepancies (about $100-1000$ MeV) in the predictions from various models.

Moreover, to know some details of the inner structure of the predicted pentaquark states, we calculate the root mean square

TABLE IV: Predicted mass spectra of $1S$ states for the $cccb\bar{c}$, $cccb\bar{b}$, $bbbc\bar{b}$ and $bbbc\bar{c}$ systems.

	$cccb\bar{c}$		$cccb\bar{b}$		
J^P	Mass	Eigenvector	Mass	Eigenvector	Configuration
$5/2^-$	(11160)	(1)	(14322)	(1)	$(1S_{\frac{5}{2}}\text{-}(\{123\}4\bar{5}))$
$3/2^-$	$\begin{pmatrix}11124\\11349\\11482\end{pmatrix}$	$\begin{pmatrix}-0.02&-0.09&1\\0.57&0.82&0.08\\-0.82&0.57&0.03\end{pmatrix}$	$\begin{pmatrix}14282\\14522\\14737\end{pmatrix}$	$\begin{pmatrix}-0.04&-0.08&1\\0.56&0.83&0.09\\-0.83&0.56&0.02\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_1\\1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_2\\1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_3\end{pmatrix}$
$1/2^-$	$\begin{pmatrix}11356\\11419\\11506\end{pmatrix}$	$\begin{pmatrix}0.46&0.85&0.26\\-0.23&-0.17&0.96\\-0.86&0.50&-0.12\end{pmatrix}$	$\begin{pmatrix}14534\\14645\\14745\end{pmatrix}$	$\begin{pmatrix}0.53&0.84&0.06\\-0.07&-0.03&1\\-0.84&0.54&-0.05\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_1\\1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_2\\1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_3\end{pmatrix}$
	$bbbc\bar{b}$		$bbbc\bar{c}$		
J^P	Mass	Eigenvector	Mass	Eigenvector	Configuration
$5/2^-$	(20767)	(1)	(17540)	(1)	$(1S_{\frac{5}{2}}\text{-}(\{123\}4\bar{5}))$
$3/2^-$	$\begin{pmatrix}20731\\20951\\21120\end{pmatrix}$	$\begin{pmatrix}-0.04&-0.09&1\\0.54&0.84&0.10\\-0.84&0.54&0.01\end{pmatrix}$	$\begin{pmatrix}17465\\17740\\17913\end{pmatrix}$	$\begin{pmatrix}-0.08&-0.14&0.99\\0.55&0.82&0.16\\-0.83&0.55&0.01\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_1\\1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_2\\1S_{\frac{3}{2}}\text{-}(\{123\}4\bar{5})_3\end{pmatrix}$
$1/2^-$	$\begin{pmatrix}20964\\21051\\21128\end{pmatrix}$	$\begin{pmatrix}0.52&0.86&0.05\\-0.05&-0.03&1\\-0.86&0.52&-0.03\end{pmatrix}$	$\begin{pmatrix}17761\\17837\\17928\end{pmatrix}$	$\begin{pmatrix}0.51&0.85&0.10\\-0.08&-0.7&0.99\\-0.86&0.52&-0.03\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_1\\1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_2\\1S_{\frac{1}{2}}\text{-}(\{123\}4\bar{5})_3\end{pmatrix}$

TABLE V: Predicted mass spectra of $1S$ states for the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems.

	$ccbb\bar{c}$		$bbcc\bar{b}$		
J^P	Mass	Eigenvector	Mass	Eigenvector	Configuration
$5/2^-$	(14691)	1	(17831)	1	$(1S_{\frac{5}{2}}\text{-}(\{12\}\{34\}\bar{5}))$
$3/2^-$	$\begin{pmatrix}14584\\14625\\14651\\14665\end{pmatrix}$	$\begin{pmatrix}-0.16&-0.01&-0.39&0.91\\0.52&0.85&-0.04&0.09\\0.19&0.11&-0.89&-0.41\\-0.82&0.51&0.25&-0.04\end{pmatrix}$	$\begin{pmatrix}17815\\17835\\17856\\17880\end{pmatrix}$	$\begin{pmatrix}0.03&0.05&-0.30&0.95\\0.40&0.91&-0.05&-0.08\\0.37&-0.09&0.88&0.27\\-0.84&0.40&0.36&0.11\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{3}{2}}\text{-}(\{12\}\{34\}\bar{5})_1\\1S_{\frac{3}{2}}\text{-}(\{12\}\{34\}\bar{5})_2\\1S_{\frac{3}{2}}\text{-}(\{12\}\{34\}\bar{5})_3\\1S_{\frac{3}{2}}\text{-}(\{12\}\{34\}\bar{5})_4\end{pmatrix}$
$1/2^-$	$\begin{pmatrix}14585\\14606\\14655\\14684\end{pmatrix}$	$\begin{pmatrix}0.12&0.77&-0.29&-0.56\\-0.15&-0.58&-0.09&-0.79\\-0.17&-0.15&-0.94&0.24\\-0.97&0.22&0.14&0.01\end{pmatrix}$	$\begin{pmatrix}17815\\17819\\17858\\17891\end{pmatrix}$	$\begin{pmatrix}-0.13&-0.53&-0.15&-0.83\\-0.19&-0.82&0.10&0.53\\0.25&-0.06&0.95&-0.18\\-0.94&0.22&0.26&-0.04\end{pmatrix}$	$\begin{pmatrix}1S_{\frac{1}{2}}\text{-}(\{12\}\{34\}\bar{5})_1\\1S_{\frac{1}{2}}\text{-}(\{12\}\{34\}\bar{5})_2\\1S_{\frac{1}{2}}\text{-}(\{12\}\{34\}\bar{5})_3\\1S_{\frac{1}{2}}\text{-}(\{12\}\{34\}\bar{5})_4\end{pmatrix}$

(RMS) radii between two quarks, i.e., $\sqrt{\langle r_{ij}^2\rangle}$. The results are listed in Tables III. It is found that $1S$ states with $\{1234\}\bar{5}$ symmetry should have a compact structure. The RMS radii for the $ccccc\bar{c}$, $cccc\bar{b}$, $bbbbc\bar{c}$, and $bbbbb\bar{b}$ systems are in the ranges of $(0.51, 0.54)$, $(0.44, 0.52)$, $(0.29, 0.40)$, and $(0.28, 0.29)$ fm, respectively. The $bbbbb\bar{b}$ system is much more compact than $ccccc\bar{c}$ due to the much stronger attractive color-Coulomb potential.

We further analyze the contributions from each part of the Hamiltonian for the pentaquark states. The results are listed in Table III. It shows that the averaged kinetic energy $\langle T\rangle$ is nearly a stable value $\sim$ 900 MeV for all of the states. The color-Coulomb potential contributes to a large negative value $\langle V^{Coul}\rangle$, about $-950, -1000, -1300, -1400$ MeV for the $ccccc\bar{c}$, $cccc\bar{b}$, $bbbbc\bar{c}$, and $bbbbb\bar{b}$ systems, respectively. The linear confining potential $\langle V^{Conf}\rangle$ has the same order of magnitude as $\langle T\rangle$, which cannot be neglected. For the $ccccc\bar{c}$ system, the value of confining potential $\langle V^{Conf}\rangle$, $\sim$ 870 MeV, is comparable with that of $\langle T\rangle$. While for the $bbbbb\bar{b}$ system, the confining potential $\langle V^{Conf}\rangle$, $\sim$ 470 MeV, is about one half of the $ccccc\bar{c}$ system. It is interesting to find that for a pentaquark system the value of $\langle T\rangle + \langle V^{Conf}\rangle + \langle V^{Coul}\rangle$ for the $J^P = 1/2^-$ state is equal to that of the $J^P = 3/2^-$ state, although each parts for these two states are different. From this point of view, the mass splitting between the $J^P = 1/2^-$ and $3/2^-$ states is caused by the chromo-magnetic interaction.

As a whole, in the $ccccc\bar{c}$, $bbbbb\bar{b}$, $cccc\bar{b}$, and $bbbbc\bar{c}$ systems,

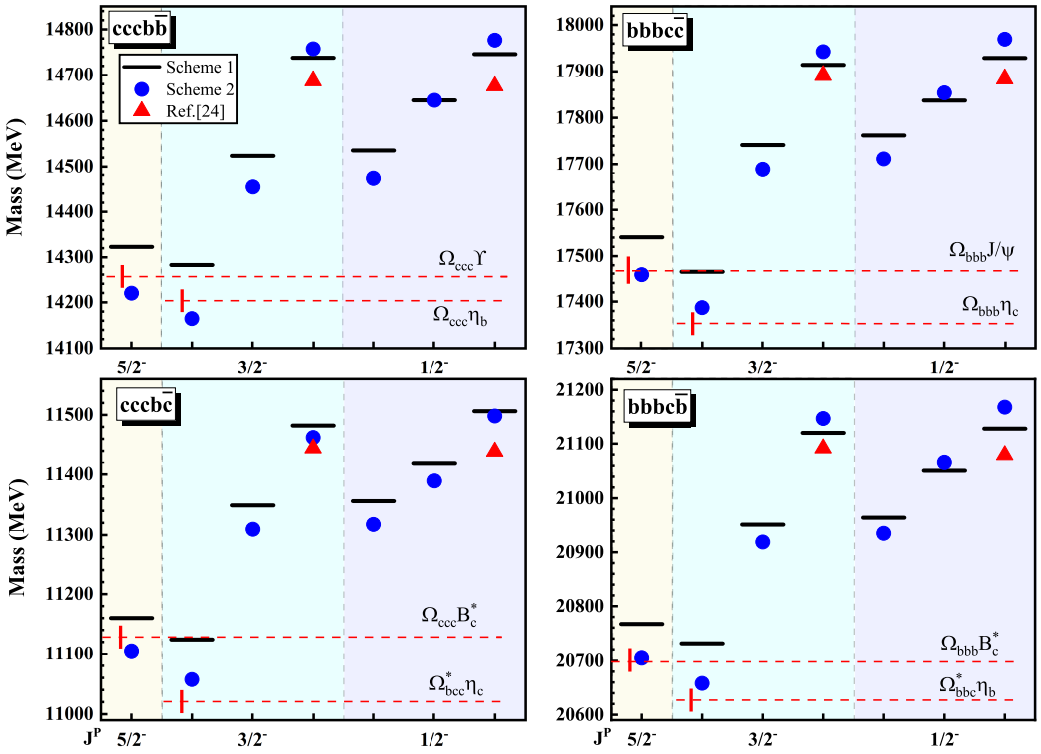


FIG. 2: Spectrum for all-heavy pentaquarks with $\{123\}4\bar{5}$ symmetry. The black short solid lines stand for the predictions based on the potential model of the present work with the ECG numerical method. The solid circles and triangles stand for the predictions based on the potential model of Ref. [24], which are obtained with the ECG numerical method in this work and with the simple diagonal Gaussian basis in Ref. [24], respectively. The red dashed lines represent the lowest $1S$-wave dissociation baryon-meson thresholds for the two low-lying states with $J^P = 5/2^-$ and $3/2^-$ in each system. The short red solid lines represent the uncertainties according to the Lattice predictions of the triply heavy baryons [41].

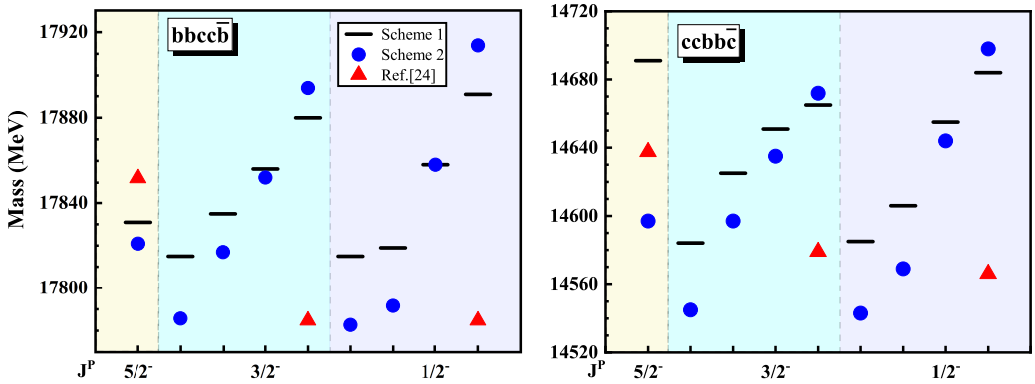


FIG. 3: Spectrum for all-heavy pentaquarks with $\{12\}\{34\}\bar{5}$ symmetry. The short solid lines stand for the predictions based on the potential model of the present work with the ECG numerical method. The solid circles and triangles stand for the predictions based on the potential model of Ref. [24], which are obtained with the ECG numerical method in this work and with the simple diagonal Gaussian basis in Ref. [24], respectively.

there are no stable states below the dissociation baryon-meson thresholds. The $1S$ states should have a compact structure. For each pentaquark system, the mass splitting between the $J^P = 1/2^-$ and $3/2^-$ states is no more than 30 MeV.

B. States with $\{123\}4\bar{5}$ symmetry

In the all-heavy pentaquarks, the $cccb\bar{b}$, $cccb\bar{c}$, $bbbc\bar{c}$, and $bbbcb\bar{b}$ systems have the $\{123\}4\bar{5}$ symmetry. Considering

the permutation symmetry of identical quarks, for each pentaquark system there are seven $1S$ states, one with $J^P = 5/2^-$, three with $J^P = 3/2^-$, and three with $J^P = 1/2^-$. The predicted mass spectrum has been given in Table IV and also shown in Fig. 2. From Table IV, one can see that there exist large mixing between the two configurations $1S\,_{J^P}(\{123\}4\bar{5})_1$ and $1S\,_{J^P}(\{123\}4\bar{5})_2$ with the same spin-parity J^P numbers. The splitting between the two mixed states are notably increased to $\sim 150-200$ MeV due to the configuration mixing effect, while without this effect the splitting between the two pure states is $\sim 50-90$ MeV. The $1S\,_{1/2^-}(\{123\}4\bar{5})_3$ and $1S\,_{3/2^-}(\{123\}4\bar{5})_3$ are nearly two pure configurations. For a given pentaquark system, the lowest state is the $J^P = 3/2^-$ configuration $1S\,_{\frac{3}{2}^-}(\{123\}4\bar{5})_3$, which is slightly ($\sim 40-60$ MeV) lower than the $J^P = 5/2^-$ state. The masses of the other five states are notably larger than these two low-lying states, the mass gap reaches up to about $200-500$ MeV.

The masses of the $1S$ states for the $cccb\bar{c}$, $cccb\bar{b}$, $bbbc\bar{c}$, and $bbbc\bar{b}$ systems scatter in the range of $\sim (11.1, 11.5)$, $\sim (14.3, 14.8)$, $\sim (17.5, 17.9)$, and $\sim (20.7, 21.1)$ GeV, respectively. In Ref. [24], the authors given their predictions of the masses for the highest $J^P = 1/2^-$ and $J^P = 3/2^-$ states. For a comparison, we shown their results (denoted by the solid triangles) in Fig. 2 as well. It is seen that these predictions in Ref [24] are generally consistent with ours. However, in Ref [24], no stable solutions for the other low-lying states are obtained by taking a simple Gaussian form for the trial spatial wave function. By using the same Hamiltonian adopted in Ref. [24], we calculate the mass spectrum with the ECG method as well. We obtain stable solutions for all of the $1S$ states, which are shown in Fig. 2 with solid circles as the scheme 2. It is found that mass spectra predicted with the two different potential models adopted in Ref. [24] (scheme 2) and present work (scheme 1) are similar, it shows only a small model dependency. The difference for most of the predictions between the two potential models is no more than 40 MeV. Some other studies about these pentaquarks can be found in the literature based on the MIT bag model [20], chromomagnetic interaction model [19], and chiral quark model [22]. It is interesting to find that the four $J^P = 5/2^-$ states for the $cccb\bar{c}$, $cccb\bar{b}$, $bbbc\bar{c}$, and $bbbc\bar{b}$ systems reported in Ref. [22] are comparable with our predictions.

To know about the inner structure of the pentaquark states, in Tables VI we further give the root mean square (RMS) radii between two quarks. It is found that the five high-lying $1S$ states with $J^P = 1/2^-$ and $J^P = 3/2^-$ for a given pentaquark system are compact states, their RMS radii are no more than 0.52 fm. While the two low-lying $J^P = 3/2^-$ and $J^P = 5/2^-$ states have a typical molecular structure. From Table VI, one can see that in these states the three identical quarks $Q_1Q_2Q_3$ and the two non-identical quarks $Q_4\bar{Q}_5$ form two compact clusters, respectively. The RMS radii between two quarks in each cluster are no more than 0.5 fm. However, the RMS radii between two compact clusters reaches up to a fairly large value, ~ 2 fm. The molecular structure of the $J^P = 5/2^-$ states is also found in Ref. [22].

The contributions from each part of the Hamiltonian for the pentaquark states are listed in Table VI of the appendix. Taking the $cccb\bar{c}$ system as an example, from the Table VI, one can see that the kinetic energy $\langle T\rangle \sim (900, 1000)$ MeV, the confining potential $\langle V^{Conf}\rangle \sim (700, 800)$ MeV, and the color-Coulomb potential $\langle V^{Coul}\rangle \sim -(1000, 1300)$ MeV have the same order of magnitude. The sums of these contributions for different states are usually very different. Thus, one cannot give a reliable prediction of the mass splitting when only considering the chromo-magnetic interaction. It should be pointed out that for the two low-lying $J^P = 3/2^-$ and $J^P = 5/2^-$ states, the color-Coulomb potential $\langle V^{Coul}\rangle$ is much more attractive ($\sim 200-300$ MeV) than that for the other five high-lying states, which leads to the fairly large gap between the two low-lying states and the five high-lying states.

The molecular structure of the low-lying $J^P = 3/2^-$ and $J^P = 5/2^-$ states is also mainly caused by the color-Coulomb interactions. In these two states, the color-Coulomb interactions between quarks in the colored subsystems $Q_4\bar{Q}_5$ and $Q_1Q_2Q_3$ are very strong. The color-Coulomb potentials from $Q_4\bar{Q}_5$ and $Q_1Q_2Q_3$ are about $\sim 5-6$ and $\sim 2-3$ times stronger than those for the other states, respectively. Thus, the subsystems $Q_4\bar{Q}_5$ and $Q_1Q_2Q_3$ forms two compact clusters. However, the color-Coulomb interactions between the two clusters are zero. Due to no attractive color-Coulomb interactions, the distance between the two clusters $Q_4\bar{Q}_5$ and $Q_1Q_2Q_3$ become very large.

Except for the special molecular structures, the two low-lying states with $J^P = 5/2^-$ and $J^P = 3/2^-$ in a given pentaquark system may be good candidates of stable states below the dissociation baryon-meson thresholds what we are looking for. If taking the masses for the triply-heavy baryons predicted by Lattice QCD [41], from Fig. 2, one can see that for a given pentaquark system the masses of the low-lying states with $J^P = 5/2^-$ and $J^P = 3/2^-$ are close to their lowest S-wave strong decay channels. With the potential model of the present work (scheme 1), the predicted masses of the two low-lying states with $J^P = 5/2^-$ and $J^P = 3/2^-$ are about a few tens MeV larger than the thresholds of their lowest S-wave strong decay channels, while with the potential model of Ref. [24] (scheme 2), the predicted masses are much more closer to or slightly below these thresholds. Considering the model uncertainty, the two low-lying $J^P = 5/2^-$ and $J^P = 3/2^-$ states with $\{123\}4\bar{5}$ symmetry may lie below the dissociation baryon-meson thresholds. Similar feature of the $J^P = 5/2^-$ states is also predicted in Ref. [22].

As a whole, a complete spectrum for the $1S$-wave all-heavy pentaquarks with $\{123\}4\bar{5}$ symmetry is given based on the potential models. For the $J^P = 1/2^-$ and $3/2^-$ states, there exist large mixing between the two different color-spin configurations $1S\,_{J^P}(\{123\}4\bar{5})_1$ and $1S\,_{J^P}(\{123\}4\bar{5})_2$. Besides the compact structures, there also exist two molecular structures for each pentaquark system, which correspond to the two low-lying $J^P = 5/2^-$ and $J^P = 3/2^-$ states. They may be candidates of the stable states below the dissociation baryon-meson thresholds.

C. States with $\{12\}\{34\}\bar{5}$ symmetry

In the all-heavy pentaquarks, the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems have the $\{12\}\{34\}\bar{5}$ symmetry. Considering the permutation symmetry of identical quarks, for each pentaquark system there are nine $1S$ states, one with $J^P = 5/2^-$, four with $J^P = 3/2^-$, and four with $J^P = 1/2^-$. The predicted mass spectrum has been given in Table V and also shown in Fig. 3. For the $J^P = 3/2^-$ and $1/2^-$ states, there is sizeable mixing in the different color-spin configurations. The masses of the $1S$ states for the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems are predicted to be in the ranges of $\sim (14.6, 14.7)$ and $\sim (17.8, 17.9)$ GeV, respectively, which are included in the mass range of the five high-lying states for the $cccb\bar{b}$ ($bbbc\bar{c}$) system. These $1S$ states are far above the dissociation baryon-meson thresholds.

By adopting the same Hamiltonian of Ref. [24], we calculate the mass spectrum with the ECG method as well. The results (scheme 2) are shown with solid circles in Fig. 3 for comparing with our potential model results (scheme 1). It is found that two different potential models give very similar results. The difference for most of the predictions is no more than 50 MeV. It should be mentioned that in Ref. [24], the authors did not give the whole mass spectrum of the $1S$ states for the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems due to the limitation of their numerical method, where a simple diagonal Gaussian basis is adopted. The discrepancy between the results obtained with the diagonal Gaussian basis (solid triangles) [24] and the ECG method (solid circles) can reach up to $\sim$ 100 MeV. Thus, to deal with the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems containing unequal mass quarks, the ECG method is suggested. Some other studies about these pentaquarks can be found in the literature based on the MIT bag model [20], CMI model [19], and chiral quark model [22]. There are large discrepancies (about 300 – 800 MeV) in the predictions from various models.

To know about the inner structure of the pentaquark states, in Table VII we further give the root mean square (RMS) radii between two quarks. It is found that the $1S$ states of the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems are compact states, the RMS radii are predicted to be in the range of $\sim 0.28 - 0.51$ fm.

The contributions from each part of the Hamiltonian for the pentaquark states are given in Table VII as well. Taking the $ccbb\bar{c}$ system as an example, from the table one can see that the kinetic energy $\langle T\rangle \sim 900$ MeV, the confining potential $\langle V^{Conf}\rangle \sim 700$ MeV, and the color-Coulomb potential $\langle V^{Coul}\rangle \sim -(1100, 1200)$ MeV have the same order of magnitude. By summing these contributions, $\langle T\rangle+\langle V^{Conf}\rangle+\langle V^{Coul}\rangle$, for each $1S$ state, one obtains three nearly degenerate values 443, 472, and 486 MeV. These values together with the small spin-spin potential $\langle V^{SS}\rangle$ determine the mass splittings between the different configurations.

As a whole, a complete spectrum for the $1S$-wave all-heavy pentaquarks $ccbb\bar{c}$ and $bbcc\bar{b}$ with $\{12\}\{34\}\bar{5}$ symmetry is given based on the potential models. For the $J^P = 1/2^-$ and $3/2^-$ states, there exists sizeable mixing between the different color-spin configurations. The $1S$ pentaquark states in the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems should be compact structures, and far above the dissociation baryon-meson thresholds.

IV. SUMMARY

In this work, we carry out a dynamical calculation of the mass spectra for the all-heavy pentaquarks with a nonrelativistic potential model. To precisely treat a five-body system, we apply the ECG method, in which a variational trial spatial wave function is expanded with nondiagonal Gaussians associated to the Jacobi coordinates. A complete mass spectrum for the $1S$ states is obtained.

In the all-heavy pentaquarks with $\{1234\}\bar{5}$ symmetry, the masses of the $1S$-wave states for $ccccc\bar{c}$ and $bbbbb\bar{b}$ are predicted to be in the ranges of $\sim$ 8.2 and $\sim$ 24.2 GeV, respectively; while the masses for $cccc\bar{b}$ and $bbbbc\bar{c}$ are predicted to be in the ranges of $\sim$ 11.5 and $\sim$ 21.1 GeV, respectively. These pentaquark states have a compact structure, and lie far above the lowest dissociation baryon-meson mass threshold.

For the $cccbc\bar{c}$, $cccbb\bar{b}$, $bbbcc\bar{c}$, and $bbbcb\bar{b}$ systems with $\{123\}4\bar{5}$ symmetry, the masses of the $1S$ states scatter in the range of $\sim (11.1, 11.5)$, $\sim (14.3, 14.8)$, $\sim (17.5, 17.9)$, and $\sim (20.7, 21.1)$ GeV, respectively. For each system, the five high-lying states with $J^P = 1/2^-$ and $3/2^-$ are compact states, and lie far above the lowest dissociation baryon-meson mass threshold. However, the two low-lying states with $J^P = 5/2^-$ and $3/2^-$ have a typical molecular structure, $(Q_1Q_2Q_3)$-$(Q_4\bar{Q}_5)$, due to the special role of the color-Coulomb interactions. They may be good candidates of stable states below the dissociation baryon-meson thresholds what we are looking for.

For the $ccbb\bar{c}$ and $bbcc\bar{b}$ systems with $\{12\}\{34\}\bar{5}$ symmetry, the masses of the $1S$ states scatter in the ranges of $\sim (14.6, 14.7)$ and $\sim (17.8, 17.9)$ GeV, respectively. These pentaquark states have a compact structure, and lie far above the lowest dissociation baryon-meson mass threshold.

Finally, it should be emphasized that with a serious consideration of the five-body problem, the results obtained from different potential models should be similar. To reliably deal with the all-heavy pentaquarks containing unequal mass quarks, it is necessary to be expanded the spatial wave functions with the most general Gaussian form rather than a simple diagonal Gaussian form.

Acknowledgements

We thank Yan-Rui Liu for part contributions to this work. We also thank Hong-Tao An for useful discussions about the pentaquark configurations. This work is supported by the National Natural Science Foundation of China under Grants Nos.12175065 and 12235018.

[1] M. Gell-Mann, A Schematic Model of Baryons and Mesons, Phys. Lett. **8**, 214-215 (1964).

[2] G. Zweig, An SU(3) model for strong interaction symmetry and

TABLE VI: The average contributions of each part of the Hamiltonian (in MeV) and the root mean square radii (in fm) for the $1S$-wave all-heavy pentaquark configurations with the $\{123\}4\bar{5}$ symmetry.

				Each part contribution of Hamiltonian					RMS Radius			
	J^P	Configuration	Mass	$\langle T\rangle$	$\langle V^{Conf}\rangle$	$\langle V^{Coul}\rangle$	$\langle V^{SS}\rangle$	r_{12},r_{13},r_{23}	r_{14},r_{24},r_{34}	r_{15},r_{25},r_{35}	r_{45}
	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{123\}4\bar{5})$	11160	962	681	-1302	35	0.4846	2.2071	2.2156	0.3267
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_1$	11439	886	789	-1038	19	0.5088	0.4410	0.5018	0.4334
$cccb\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_2$	11390	913	771	-1092	14	0.5000	0.4706	0.5285	0.3810
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_3$	11126	1000	673	-1332	1	0.4847	2.2563	2.2637	0.3134
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_1$	11470	847	806	-1015	48	0.5164	0.4471	0.5168	0.4462
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_2$	11395	906	773	-1087	20	0.4993	0.4702	0.5284	0.3861
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_3$	11416	885	781	-1047	11	0.5145	0.4332	0.5025	0.4769
	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{123\}4\bar{5})$	14322	936	646	-1447	34	0.5103	1.7147	1.7146	0.2292
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_1$	14670	873	718	-1101	27	0.4953	0.4256	0.4232	0.3303
$cccb\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_2$	14587	927	695	-1201	14	0.4927	0.4507	0.4514	0.2667
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_3$	14284	991	638	-1494	-5	0.5104	1.7257	1.7256	0.2184
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_1$	14685	853	726	-1088	41	0.4993	0.4291	0.4288	0.3355
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_2$	14595	915	697	-1191	21	0.4924	0.4510	0.4516	0.2713
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_3$	14645	879	712	-1112	13	0.4945	0.4179	0.4197	0.3793
	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{123\}4\bar{5})$	17540	1058	506	-1584	38	0.2491	2.0011	1.9992	0.4195
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_1$	17859	895	634	-1202	9	0.3180	0.4093	0.4039	0.4723
$bbbc\bar{c}$	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_2$	17787	953	606	-1291	-3	0.2898	0.4327	0.4350	0.4439
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_3$	17472	1140	484	-1642	-31	0.2490	2.1234	2.1214	0.3847
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_1$	17884	862	646	-1180	34	0.3211	0.4145	0.4138	0.4849
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_2$	17806	929	614	-1275	16	0.2900	0.4373	0.4391	0.4568
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_3$	17837	893	631	-1202	-6	0.3368	0.3964	0.3996	0.4854
	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{123\}4\bar{5})$	20767	1071	454	-1676	27	0.2580	1.9594	1.9473	0.3269
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_1$	21071	865	566	-1267	16	0.3058	0.3950	0.3180	0.3943
$bbbc\bar{b}$	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_2$	20998	946	536	-1374	-1	0.2835	0.4035	0.3327	0.3613
		$1S_{\frac{3}{2}}$-$(\{123\}4\bar{5})_3$	20733	1110	446	-1706	-8	0.2581	2.0506	2.0402	0.3135
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_1$	21084	848	572	-1255	29	0.3077	0.3982	0.3214	0.3998
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_2$	21008	934	539	-1365	9	0.2836	0.4062	0.3336	0.3668
		$1S_{\frac{1}{2}}$-$(\{123\}4\bar{5})_3$	21051	847	568	-1250	-6	0.3210	0.3860	0.3223	0.4096

its breaking. Version 1, CERN-TH-401.

[3] G. Zweig, An SU(3) model for strong interaction symmetry and its breaking. Version 2, CERN-TH-412.

[4] S. K. Choi *et al.* [Belle], Observation of a narrow charmonium-like state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays, Phys. Rev. Lett. **91**, 262001 (2003).

[5] R. L. Workman *et al.* [Particle Data Group], Review of Particle Physics, PTEP **2022**, 083C01 (2022).

[6] R. Aaij *et al.* [LHCb], Study of the doubly charmed tetraquark T_{cc}^{+}, Nature Commun. **13**, 3351 (2022).

[7] R. Aaij *et al.* [LHCb], Observation of an exotic narrow doubly charmed tetraquark, Nature Phys. **18**, 751-754 (2022).

[8] R. Aaij *et al.* [LHCb], First Observation of a Doubly Charged Tetraquark and Its Neutral Partner, Phys. Rev. Lett. **131**, 041902 (2023).

[9] R. Aaij *et al.* [LHCb], A model-independent study of resonant structure in $B^{+} \to D^{+}D^{-}K^{+}$ decays, Phys. Rev. Lett. **125**, 242001 (2020).

[10] R. Aaij *et al.* [LHCb], Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^{-} p$ Decays, Phys. Rev. Lett. **115**, 072001 (2015).

[11] R. Aaij *et al.* [LHCb], Model-independent evidence for $J/\psi p$

TABLE VII: The average contributions of each part of the Hamiltonian (in MeV) and the root mean square radii (in fm) for the $1S$-wave all-heavy pentaquark configurations with the $\{12\}\{34\}\bar{5}$ symmetry.

				Each part contribution of Hamiltonian				RMS Radius				
	J^P	Configuration	Mass	$\langle T\rangle$	$\langle V^{Conf}\rangle$	$\langle V^{Coul}\rangle$	$\langle V^{SS}\rangle$	r_{12}	$r_{13}, r_{14}, r_{23}, r_{24}$	r_{15}, r_{25}	r_{34}	r_{35}, r_{45}
$ccbb\bar{c}$	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{12\}\{34\}\bar{5})$	14619	903	702	-1162	23	0.4780	0.4493	0.5043	0.2902	0.4227
	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_1$	14652	898	709	-1121	14	0.4918	0.4220	0.4860	0.3299	0.4168
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_2$	14636	908	704	-1140	11	0.4813	0.4258	0.4943	0.3083	0.4206
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_3$	14642	910	704	-1128	3	0.4923	0.4115	0.4833	0.3440	0.4196
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_4$	14595	935	690	-1182	-1	0.4725	0.4432	0.4920	0.2881	0.4131
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_1$	14680	860	724	-1098	41	0.4990	0.4277	0.5000	0.3336	0.4287
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_2$	14598	960	685	-1172	-27	0.4722	0.4178	0.4760	0.3042	0.4055
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_3$	14649	900	708	-1122	10	0.4942	0.4129	0.4868	0.3451	0.4227
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_4$	14602	926	694	-1176	6	0.4733	0.4428	0.4970	0.2883	0.4175
$bbcc\bar{b}$	$5/2^-$	$1S_{\frac{5}{2}}$-$(\{12\}\{34\}\bar{5})$	17831	895	626	-1231	19	0.2824	0.4300	0.3142	0.4647	0.4205
	$3/2^-$	$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_1$	17869	884	636	-1194	20	0.3162	0.4081	0.3156	0.4781	0.4061
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_2$	17842	909	626	-1220	5	0.2967	0.4093	0.3130	0.4662	0.4081
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_3$	17855	904	629	-1205	5	0.3284	0.3975	0.3174	0.4767	0.3998
		$1S_{\frac{3}{2}}$-$(\{12\}\{34\}\bar{5})_4$	17819	913	621	-1243	7	0.2808	0.4264	0.3103	0.4617	0.4154
	$1/2^-$	$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_1$	17885	864	643	-1179	35	0.3189	0.4116	0.3204	0.4822	0.4118
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_2$	17822	937	617	-1239	-15	0.2939	0.4045	0.3070	0.4611	0.4007
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_3$	17859	898	631	-1201	9	0.3292	0.3984	0.3186	0.4778	0.4012
		$1S_{\frac{1}{2}}$-$(\{12\}\{34\}\bar{5})_4$	17818	913	620	-1244	6	0.2806	0.4246	0.3113	0.4608	0.4157

contributions to $\Lambda_b^0 \to J/\psi p K^-$ decays, Phys. Rev. Lett. **117**, 082002 (2016).

[12] R. Aaij *et al.* [LHCb], Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$, Phys. Rev. Lett. **122**, 222001 (2019).

[13] R. Aaij *et al.* [LHCb], Observation of a $J/\psi\Lambda$ Resonance Consistent with a Strange Pentaquark Candidate in $B^- \to J/\psi\Lambda\bar{p}$ Decays, Phys. Rev. Lett. **131**, 031901 (2023).

[14] R. Aaij *et al.* [LHCb], Observation of structure in the J/ψ-pair mass spectrum, Sci. Bull. **65**, 1983-1993 (2020).

[15] G. Aad *et al.* [ATLAS], Observation of an Excess of Dicharmonium Events in the Four-Muon Final State with the ATLAS Detector, Phys. Rev. Lett. **131**, 151902 (2023).

[16] A. Hayrapetyan *et al.* [CMS], Observation of new structure in the J/ψJ/ψ mass spectrum in proton-proton collisions at $\sqrt{s}$ = 13 TeV, [arXiv:2306.07164 [hep-ex]].

[17] J. R. Zhang, Fully-heavy pentaquark states, Phys. Rev. D **103**, 074016 (2021).

[18] Z. G. Wang, Analysis of the fully-heavy pentaquark states via the QCD sum rules, Nucl. Phys. B **973**, 115579 (2021).

[19] H. T. An, K. Chen, Z. W. Liu and X. Liu, Fully heavy pentaquarks, Phys. Rev. D **103**, 074006 (2021).

[20] W. X. Zhang, H. T. An and D. Jia, Masses and magnetic moments of exotic fully heavy pentaquarks, Eur. Phys. J. C **83**, 727 (2023).

[21] Y. Yan, Y. Wu, X. Hu, H. Huang and J. Ping, Fully heavy pentaquarks in quark models, Phys. Rev. D **105**, 014027 (2022).

[22] Y. Yan, Y. Wu, H. Huang, J. Ping and X. Zhu, Prediction of charmed-bottom pentaquarks in quark model, Eur. Phys. J. C **83**, 610 (2023).

[23] G. Yang, J. Ping and J. Segovia, Fully charm and bottom pentaquarks in a lattice-QCD inspired quark model, Phys. Rev. D **106**, 014005 (2022).

[24] H. T. An, S. Q. Luo, Z. W. Liu and X. Liu, Fully heavy pentaquark states in constituent quark model, Phys. Rev. D **105**, 074032 (2022).

[25] M. C. Gordillo and J. M. Alcaraz-Pelegrina, Asymptotic mass limit of large fully heavy compact multiquarks, Phys. Rev. D **108**, 054027 (2023).

[26] E. Eichten, K. Gottfried, T. Kinoshita, K. D. Lane and T. M. Yan, Charmonium: The Model, Phys. Rev. D **17**, 3090 (1978) [erratum: Phys. Rev. D **21**, 313 (1980)].

[27] Q. Li, M. S. Liu, Q. F. Lü, L. C. Gui and X. H. Zhong, Canonical interpretation of $Y(10750)$ and $\Upsilon(10860)$ in the Υ family, Eur. Phys. J. C **80**, 59 (2020).

[28] W. J. Deng, H. Liu, L. C. Gui and X. H. Zhong, Charmonium spectrum and their electromagnetic transitions with higher multipole contributions, Phys. Rev. D **95**, 034026 (2017).

[29] Q. Li, M. S. Liu, L. S. Lu, Q. F. Lü, L. C. Gui and X. H. Zhong, Excited bottom-charmed mesons in a nonrelativistic quark model, Phys. Rev. D **99**, 096020 (2019).

[30] M. S. Liu, Q. F. Lü and X. H. Zhong, Triply charmed and bottom baryons in a constituent quark model, Phys. Rev. D **101**, 074031 (2020).

[31] M. S. liu, F. X. Liu, X. H. Zhong and Q. Zhao, Full-heavy tetraquark states and their evidences in the LHCb di-J/ψ spectrum, [arXiv:2006.11952 [hep-ph]].
[32] F. X. Liu, M. S. Liu, X. H. Zhong and Q. Zhao, Higher mass spectra of the fully-charmed and fully-bottom tetraquarks, Phys. Rev. D **104**, 116029 (2021).
[33] M. S. Liu, Q. F. Lü, X. H. Zhong and Q. Zhao, All-heavy tetraquarks, Phys. Rev. D **100**, 016006 (2019).
[34] K. Varga and Y. Suzuki, Precise Solution of Few Body Problems with Stochastic Variational Method on Correlated Gaussian Basis, Phys. Rev. C **52**, 2885-2905 (1995).
[35] J. Mitroy, S. Bubin, W. Horiuchi, Y. Suzuki, L. Adamowicz, W. Cencek, K. Szalewicz, J. Komasa, D. Blume and K. Varga, Theory and application of explicitly correlated Gaussians, Rev. Mod. Phys. **85**, 693-749 (2013).
[36] F. X. Liu, R. H. Ni, X. H. Zhong and Q. Zhao, Charmed-strange tetraquarks and their decays in the potential quark model, Phys. Rev. D **107**, 096020 (2023).
[37] T. A. Kaeding, Tables of SU(3) isoscalar factors, Atom. Data Nucl. Data Tabl. **61**, 233-288 (1995).
[38] F. Stancu and S. Pepin, Isoscalar factors of the permutation group, Few Body Syst. **26**, 113-133 (1999).
[39] W. Park, A. Park, S. Cho and S. H. Lee, $P_c(4380)$ in a constituent quark model, Phys. Rev. D **95**, 054027 (2017).
[40] E. Hiyama, Y. Kino and M. Kamimura, Gaussian expansion method for few-body systems, Prog. Part. Nucl. Phys. **51**, 223-307 (2003).
[41] Z. S. Brown, W. Detmold, S. Meinel and K. Orginos, Charmed bottom baryon spectroscopy from lattice QCD, Phys. Rev. D **90**, 094507 (2014).

Nucleons and vector mesons in a confining holographic QCD model

Alfonso Ballon-Bayona,[a] **Adão S. da Silva Junior**[a]

[a] *Instituto de Física, Universidade Federal do Rio de Janeiro, Caixa Postal 68528, RJ 21941-972, Brazil.*

E-mail: aballonb@if.ufrj.br, adaossj@pos.if.ufrj.br

arXiv:2402.17950v1 [hep-ph] 28 Feb 2024

Abstract: We present a simple holographic QCD model that provides a unified description of vector mesons and nucleons in a confining background based on Einstein-dilaton gravity. For the confining background we consider analytical solutions of the Einstein-dilaton equations where the dilaton is a quadratic function of the radial coordinate far from the boundary. We build actions for the 5d gauge field and the 5d Dirac field dual to the 4d flavor current and the 4d nucleon interpolator respectively. In order to obtain asymptotically linear Regge trajectories we impose for each sector the condition that the effective Schrödinger equation has a potential that grows quadratically in the radial coordinate far from the boundary. For the vector mesons we show that this condition is automatically satisfied by a 5d Yang-Mills action minimally coupled to the metric and the dilaton. For the nucleons we find that the mass term of the 5d Dirac action needs to be generalised to include couplings to the metric and the dilaton. Using Sturm-Liouville theory we obtain a spectral decomposition for the hadronic correlators consistent with large N_c QCD. Our setup contains only three parameters: the mass scale associated with confinement, the 5d gauge coupling and the 5d Dirac coupling. The last two are completely fixed by matching the correlators at high energies to perturbative QCD. We calculate masses and decay constants and compare our results against available experimental data. Our model can be thought of as a consistent embedding of soft wall models in Einstein-dilaton gravity.

Contents

1 Introduction **1**

2 Confining holographic QCD models from Einstein-dilaton gravity **3**
- 2.1 The action 3
- 2.2 The field equations 4
- 2.3 Confining holographic QCD models 5
- 2.4 Conformal symmetry breaking and confinement 6

3 Vector mesons in confining holographic QCD **8**
- 3.1 The 4d flavour currents 8
- 3.2 The 5d action and field equations 9
- 3.3 VEVs of the 4d flavour currents 11
- 3.4 The on-shell action, the bulk to boundary propagator and the two point function 12
- 3.5 Spectral decomposition for the bulk to boundary propagator 13
- 3.6 The 4d current correlator 15
- 3.7 Spectrum of vector mesons 15
- 3.8 Wave functions and vector meson decay constants 17

4 Nucleons in confining holographic QCD **19**
- 4.1 The 4d nucleon operator 19
- 4.2 The 5d action and field equations 21
- 4.3 VEV of the 4d nucleon operator 23
- 4.4 The on-shell action, the bulk to boundary propagator and the two point function 25
- 4.5 Spectral decomposition for the bulk to boundary propagator 27
- 4.6 The 4d nucleon correlator 29
- 4.7 Spectrum of nucleons 30
- 4.8 Wave functions and nucleon decay constants 32

5 Conclusions **34**

A Sturm-Liouville theory and the spectral decomposition **36**

B The Proca Field Propagator **38**

C Vector mesons and nucleons in the soft wall model **39**
- C.1 Vector mesons 39
- C.2 Nucleons 40

D Vector mesons and nucleons in the hard wall model **41**
D.1 Vector mesons 41
D.2 Nucleons 42

1 Introduction

The origin of hadron masses and its relation to confinement is one of the challenging problems in quantum chromodynamics (QCD) due to the necessity of non-perturbative techniques. An important mechanism for mass generation in QCD is related to the spontaneous breaking of chiral symmetry, see for example [1]. The order parameter of spontaneous chiral symmetry breaking is the quark condensate defined as the VEV (vacuum expectation value) of the quark mass operator, i.e. $\langle \bar{q}q \rangle$. Another important quantity associated with mass generation, confinement and the QCD vacuum is the so-called gluon condensate defined as the VEV of the Yang-Mills operator, i.e. $\langle \mathrm{Tr} F^2 \rangle$. The gluon and quark condensates are both related to the QCD trace anomaly and the vacuum energy density of QCD, see for instance [2].

Several approaches have emerged in an attempt to describe QCD in the non-perturbative regime, such as the Nambu-Jona-Lasinio model [3, 4], the linear sigma model [5], chiral perturbation theory (for a review see [6]), lattice QCD [7] and QCD sum rules [8] (for a review see [9]). The Nambu-Jona-Lasinio and the linear sigma models describe spontaneous chiral symmetry breaking, generating a mass scale. Chiral perturbation theory is a systematic approach that exploits the approximate chiral symmetry of QCD at low energies and allows describing some properties of hadrons. Lattice QCD is a numerical approach based on the discretisation of spacetime, enabling the calculation of correlation functions of QCD operators through numerical simulations on a lattice. This approach allows for the investigation of non-perturbative properties of hadrons, such as masses. Lastly, the QCD sum rules approach considers correlation functions of composite operators built from quark fields and then uses operator product expansions (OPE) and spectral functions to estimate hadronic properties such as masses and decay constants. The composite operators are usually known as interpolating fields (this terminology is also used in lattice QCD).

The AdS/CFT correspondence is an alternative approach for investigating QCD in the strong coupling regime. This conjecture establishes a duality between string theories on $AdS_{d+1} \times M$ (AdS is Anti-de Sitter spacetime and M is a compact space) and conformal field theories (CFT) in d dimensions [10–12]. In this work, we will restrict ourselves to the particular example of the duality that relates IIB string theory on $AdS_5 \times S^5$ to the $\mathcal{N} = 4$ Super Yang-Mills theory in four dimensions. After the conjecture was proposed, some models emerged that became known as AdS/QCD, which aim to capture low-energy aspects of QCD by breaking the conformal symmetry. These models incorporate QCD properties such as confinement and chiral symmetry breaking.

There are two main approaches in AdS/QCD: the bottom-up and top-down. The bottom-up approach aims to capture QCD properties mapping the deformation of the

CFT_4 to deformations of the AdS_5 space. In this approach, a minimal set of 5d fields is introduced to describe the dynamics of 4d operators similar to those appearing in real QCD. The actions are usually minimal models for the 5d fields that reproduce the symmetries of the dual 4d operators. Examples of models within this approach include the hard wall model [13–16], the soft wall model [17], and the Einstein-dilaton models [18–23]. In the hard wall model, specific boundary conditions are imposed on the AdS space, while the soft wall model introduces a scalar field, known as the dilaton, in the action. The Einstein-dilaton model, distinct from the other two, is consistent with Einstein's equations and allows for a description of a non-trivial vacuum in the 4d dual theory. A more rigorous approach in AdS/QCD is the top-down approach, which aligns with string theory principles and introduces the breaking of conformal symmetry and supersymmetry through a setup of D-branes. Models within this approach include the D3/D7 model [24] and the D4/D8 model [11, 25].

An important test for AdS/QCD is the description of hadronic masses and decay constants and its relation to confinement. In this work we will focus on the description of light vector mesons and nucleons using the bottom-up approach. Light vector mesons have been investigated previously in the bottom-approach using hard wall models [15, 26], soft wall models [17, 27] and metric deformations in AdS [28–30] considering a 5d Yang-Mills action. There has also been some progress on the description of light vector mesons in holographic models inspired by string theory [19, 31, 32] and models based on Einstein-dilaton gravity [33, 34]. Nucleons have been investigated previously in holographic QCD following two approaches. In the first approach one builds a 5d Dirac action for the 5d Dirac field dual to a 4d nucleon interpolator, see [35, 36] for the hard wall model, [37–39] for the soft wall model and [28, 30] for AdS deformations. The second approach consists of mapping 5d solitons to 4d skyrmions, see for example [40–46]. We also noticed some recent progress on the description of fermionic states qualitatively similar to baryons considering a fermionic action for Dp/Dq brane models in string theory [47, 48].

In this paper, we present a simple 5d holographic QCD model that provides a unified description of light vector mesons and nucleons in a confined background, the latter arising from Einstein-dilaton gravity. Our model contains only three parameters; two of them are 5d coupling constants that are fixed matching the result for the two-point correlators at high energies to perturbative QCD, the third parameter is the mass scale associated with confinement which can be fixed matching for instance the mass of the $\rho(770)$ meson to the mass of the fundamental vector meson state, i.e. m_{ρ^0}. We calculate the spectrum of light vector mesons and nucleons as well as their decay constants. In order to provide a clean comparison to previous models and experimental data we will present our results dividing all the observables by the appropriate power of m_{ρ^0}. We impose for vector mesons and nucleons a condition that guarantees an asympotically linear spectrum, namely that the Schrödinger effective potential grows quadratically in the radial coordinate far from the boundary. We find for vector mesons that this condition is automatically satisfied a 5d Yang-Mills action minimally coupled to the metric and dilaton. For the nucleons we find that the mass term of the 5d Dirac action needs to be extended to include non-minimal couplings to the metric and dilaton. We use Sturm-Liouville theory to obtain

spectral decompositions for the two-point correlation functions associated with the 4d flavour current and the 4d nucleon interpolator (Ioffe current). We show that the spectral decompositions for the hadronic correlators are consistent with QCD in the large N_c limit. This in turn allows us to obtain a holographic dictionary for the decay constants of vector mesons and nucleons valid for a general class of holographic models based on Einstein-dilaton gravity.

The organisation of this paper is as follows: In section 2 we review the action and field equations of Einstein-dilaton gravity and present two analytical solutions that satisfy the confinement criterion. These two concrete backgrounds will be used in the rest of the paper. In section 3, we investigate the light vector mesons. We describe the 5d action, the field equations, the holographic dictionary for the VEV of the 4d flavour current and the bulk to boundary propagator. Using Sturm-Liouville theory we obtain a spectral decomposition for the current correlator and finally solving the Schrödinger equation we obtain the vector meson spectrum and decay constants. In section 4, we investigate the nucleons. We present the Dirac action, the field equations, the VEV of the 4d nucleon operator and the bulk to boundary propagator. Using Sturm-Liouville theory we find the spectral decomposition for the nucleon correlator and finally solving the Schrödinger equations we obtain the spectrum of nucleons and the nucleon decay constants. Our conclusions are presented in section 5 and additional material is described in four appendices. Appendix A briefly reviews the Sturm-Liouville theory and the spectral decomposition. The Proca field propagator associated with vector mesons is described in appendix B. The vector mesons and nucleons in the soft wall model and hard wall model are described in appendices C and D respectively.

2 Confining holographic QCD models from Einstein-dilaton gravity

2.1 The action

Holographic QCD models based on Einstein-dilaton gravity are described by the following action in the string frame [19, 22, 23]

$$S_s = \sigma \int_{\mathcal{M}} d^5x \sqrt{-g_s}\, e^{-2\Phi_s} \left[R_s + \mathcal{L}^s_\Phi\right] . \tag{2.1}$$

In this expression, $\sigma = 1/16\pi G_5$, where G_5 represents the five-dimensional Newton's constant, R_s is the Ricci scalar, Φ is the dilaton and $\mathcal{L}^s_\Phi$ is the dilaton lagrangian express by

$$\mathcal{L}^s_\Phi = 4\partial_\mu\Phi\partial^\mu\Phi + \ell^{-2}V_s(\Phi) \tag{2.2}$$

The subscript "s" in the previous equations indicates that they are written in the string frame. The parameter ℓ is the AdS radius. We omit here an additional surface term that is required from the variational principle [1].

[1] This term, usually called the Gibbons - Hawking - York boundary term, is also important in the study of vacuum energy in holographic QCD.

In Einstein-dilaton gravity, there are two interesting frames: the string frame and the Einstein frame. We can write the action (2.1) in the Einstein frame using the following transformations

$$g_{mn} = g^s_{mn}\, e^{-\frac{4}{3}\Phi}\,, \tag{2.3}$$

$$V(\Phi) = V_s\, e^{\frac{4}{3}\Phi}\,. \tag{2.4}$$

Plugging (2.3) and (2.4) into (2.1), the action in Einstein frame becomes

$$S_E = \sigma \int d^5x\, \sqrt{-g}\, [R + \mathcal{L}_\Phi]\ , \tag{2.5}$$

where

$$\mathcal{L}_\Phi = -\frac{4}{3} g^{mn} \partial_m \Phi \partial_n \Phi + \ell^{-2} V(\Phi)\,. \tag{2.6}$$

2.2 The field equations

By varying the action (2.5) with respect to the metric, we obtain

$$R_{mn} - \frac{R}{2} g_{mn} = \frac{1}{2\sigma} T_{mn}\,, \tag{2.7}$$

$$\nabla^2 \Phi + \frac{3}{8\ell^2} \frac{dV}{d\Phi} = 0\,, \tag{2.8}$$

where the tensorial equation in (2.7) corresponds to the Einstein equations in the presence of scalar matter and (2.8) is the generalization of the Klein-Gordon equation in curved space. The energy-momentum tensor T_{mn}, is given by

$$T_{mn} = \sigma \left[\frac{8}{3} \partial_m \Phi \partial_n \Phi + g_{mn} \mathcal{L}_\Phi \right] \tag{2.9}$$

and the Einstein equations can also be written in the Ricci form

$$R_{mn} = \frac{4}{3} \partial_m \Phi \partial_n \Phi - \frac{1}{3\ell^2} g_{mn} V\,. \tag{2.10}$$

We now consider the following ansatz for the 5d metric:

$$ds^2 = \frac{1}{\zeta(z)^2} \left[-dt^2 + dx_i^2 + dz^2 \right] \tag{2.11}$$

This metric preserves Poincaré symmetry. Plugging this ansatz into the Einstein-dilaton equations we find the following field equations:

$$\zeta'' - \frac{4}{9} \zeta \Phi'^2 = 0\,, \tag{2.12}$$

$$\ell^{-2} V - \zeta^5 (\zeta^{-3})'' = 0\,, \tag{2.13}$$

$$\frac{8}{3} \zeta^2 \left[\Phi'' - 3 \frac{\zeta'}{\zeta} \Phi' \right] + \ell^{-2} \frac{dV}{d\Phi} = 0\,. \tag{2.14}$$

where $' = d/dz$. The equation (2.14) comes from the scalar differential equation (2.8) or from the Bianchi identity $\nabla_n T^{mn} = 0$. This equation is not independent because it can be obtained from equations (2.12) and (2.13). The inverse scale factor $\zeta(z)$ is usually written in terms of the warp factor $A(z)$ using the relation

$$\zeta(z) = \exp(-A(z)). \tag{2.15}$$

The warp factor in the string frame takes the form

$$A_s(z) = A(z) + \frac{2}{3}\Phi = -\ln\zeta + \frac{2}{3}\Phi\,. \tag{2.16}$$

2.3 Confining holographic QCD models

In this work we will consider holographic QCD models where the dilaton is quadratic far from the boundary (infrared regime), i.e.

$$\Phi(z \to \infty) = kz^2\,. \tag{2.17}$$

Near the boundary we only impose that the metric is asymptotically AdS, i.e.

$$\zeta(z \to 0) = \frac{z}{\ell}\,. \tag{2.18}$$

The IR asymptotic behaviour (2.17) was originally proposed by Karch, Katz, Son and Stephanov [17] as a condition that guarantees approximate linear Regge trajectories for mesons. It was later proven by Gursoy, Kiritsis and Nitti [19] that this asymptotic behaviour is compatible with the confinement criterion and also leads to a linear spectrum for glueballs, see also [23]. Later in this section we will use the warp factor in the string frame $A_s(z)$ to show that models satisfying this asymptotic behaviour satisfy the confinement criterion developed in [49].

In order to build concrete models we consider two simple analytical solutions of the Einstein-dilaton equations that satisfy the conditions (2.17) and (2.18).

The first model is given by

$$\Phi_I(z) = kz^2 \quad , \quad \zeta_I(z) = \Gamma(5/4)\left(\frac{3}{k}\right)^{1/4}\frac{\sqrt{z}}{\ell}\,I_{1/4}\left(\frac{2}{3}kz^2\right)\,. \tag{2.19}$$

In this case we considered a simple ansatz for the dilaton field $\Phi(z)$ and found the inverse scale factor $\zeta(z)$ using the Einstein-dilaton equation (2.12). This model was investigated by Huang and Li in [22].

The second model is given by

$$\Phi_{II}(z) = \frac{1}{2}\sqrt{k}z\sqrt{9+4kz^2} + \frac{9}{4}\sinh^{-1}\left(\frac{2}{3}\sqrt{k}z\right) \quad , \quad \zeta_{II}(z) = \frac{z}{\ell}\exp\left(\frac{2}{3}kz^2\right)\,. \tag{2.20}$$

In this case we took a simple ansatz for the inverse scale factor $\zeta(z)$ and use the Einstein-dilaton equation (2.12) to find the dilaton $\Phi(z)$. This model was proposed by Gursoy, Kiritsis, and Nitti in [19] as a simple analytical model for describing confinement [2].

[2] A similar analytical model was proposed earlier in the string frame as a phenomenological approach for the quark-antiquark potential without actually solving the Einstein-dilaton equations [50].

The behaviour of the inverse scale factor, in the Einstein frame, for models I and II is shown on the left panel of figure 1. In both cases the inverse scale factor behaves as $\zeta(z) = z/\ell + \dots$ at small z (AdS asymptotics) and becomes $\exp(\frac{2}{3}kz^2 + \dots)$ at very large z. The dilaton field $\Phi(z)$ is displayed on the right panel of figure 1. In model I the dilaton field is always $\Phi(z) = kz^2$ whilst in model II it evolves from $\Phi(z) = 3\sqrt{k}z + \dots$ at small z to $\Phi(z) = kz^2$ at large z.

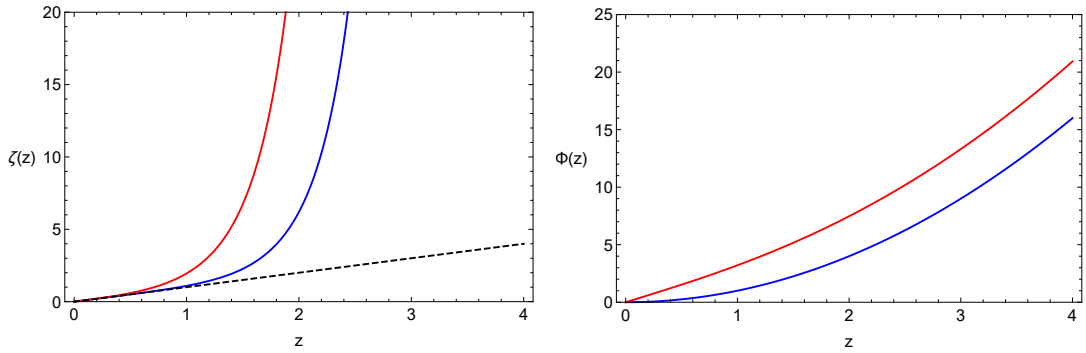


Figure 1. Left panel: Inverse scale factor $\zeta(z)$ in the Einstein frame for model I (blue) and model II (red). The black dashed line represents the AdS limit $\zeta(z) = z$. Right panel: Dilaton field $\Phi(z)$ for model I (blue) and model II (red). The plots were obtained in units of $k = 1$ and $\ell = 1$.

Using the Einstein-dilaton equation (2.13) we can reconstruct the dilaton potential $V(\Phi)$ for models I and II. This is shown in figure 2 where we also show the limit $V(\Phi \to 0) = 12$ that corresponds to the negative cosmological constant for AdS space.

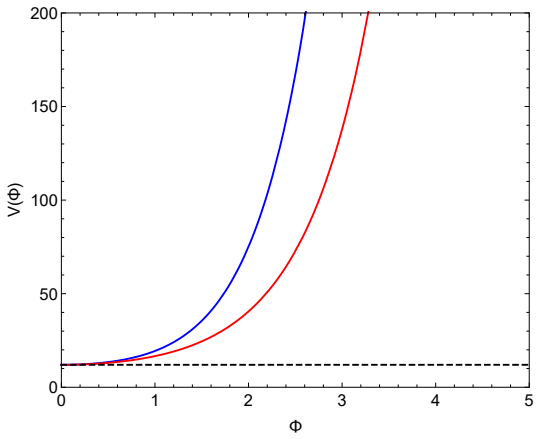


Figure 2. Dilaton potential in the Einstein frame for model I (blue) and model II (red). The black dashed line represents the AdS limit $V = 12$ when Φ goes to zero.

2.4 Conformal symmetry breaking and confinement

The models presented in the previous section describe an explicit breaking of conformal symmetry and guarantee confinement. In this section we briefly describe the confinement criterion discussed in [19] for Einstein-dilaton models based on the general criterion found

in [49]. The behaviour of the potential energy of a heavy quark-antiquark pair, described by a rectangular Wilson loop, for a review see [51], when the distance between them is large is given by

$$E(L) = \mu\, f(z^\star) L \tag{2.21}$$

where $E(L)$ is the potential energy of the quark-antiquark pair as a function of the distance L, μ is the fundamental tension of the string, and $f = \exp(2A_s)$ is a function of the string warp factor A_s. A non-zero minimum for f, located at $z = z^\star$, guarantees a non-zero string tension for the quark-antiquark potential. Note that confinement in Einstein-dilaton models involves the string frame warp factor

$$A_s(z) = A(z) + \frac{2}{3}\Phi(z)\,. \tag{2.22}$$

As explained previously in this section, in order to guarantee a linear spectrum for mesons and glueballs, the dilaton field must behave at large z as

$$\Phi(z \to \infty) = k z^2 + \dots\,, \tag{2.23}$$

and this in turn implies that the Einstein frame warp factor should behave as

$$A(z \to \infty) = -\frac{2}{3} k z^2 + \dots\,. \tag{2.24}$$

The dots in the equations above represent subleading terms for $\Phi(z)$ and $A(z)$. As described in [19], at large z the dilaton and warp factor should satisfy the condition

$$\Phi(z) + \frac{3}{2} A(z) = \frac{3}{4} \ln |A'(z)| + \dots \quad (z \to \infty)\,. \tag{2.25}$$

This in turn implies that the string frame warp factor behaves at large z as

$$A_s(z \to \infty) = \frac{1}{2} \ln |A'(z)| + \dots = \frac{1}{2} \ln(\sqrt{k} z) + \dots\,. \tag{2.26}$$

On the other hand AdS asymptotics at small z implies that

$$A_s(z \to 0) = -\ln(z/\ell) + \dots\,. \tag{2.27}$$

These results when applied to the function $f = \exp(2A_s)$ imply that this function behaves at large z as $f(z \to \infty) = \sqrt{k} z + \dots$ and at small z at $f(z \to 0) = (\ell/z)^2 + \dots$. Then the function $f(z)$ is non-monotonic in z and possesses a minimum at some $z = z^*$. This is shown in figure 3 where we plot the function $f = \exp(2A_s)$ for models I and II.

We finish this section with an important remark. In the holographic QCD models presented here conformal symmetry breaking and confinement are driven by a single (infrared) mass scale $\sqrt{k}$. If we set k to zero the dilaton vanishes and we recover the AdS space and conformal symmetry. The analog of this situation in QCD is the presence of a gluon condensate associated with a non-zero trace for the stress energy tensor and conformal symmetry breaking (the QCD trace anomaly). In fact, the problem of generation of hadron masses is expected to be understood in terms of a non-trivial stress energy tensor.

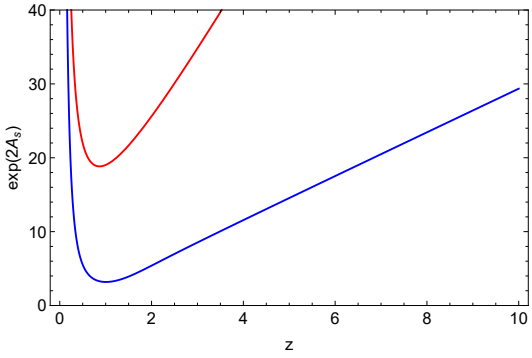


Figure 3. The function $f = \exp(2A_s)$ as a function of z for model I (blue) and model II (red). The plot was done in units of $k = 1$ and $\ell = 1$.

In the following sections we will incorporate vector mesons and nucleons in models I and II. We will obtain a unified description of vector meson and nucleon masses in terms of the single mass scale $\sqrt{k}$. When investigating the spectrum of vector mesons and nucleons we will focus on mass ratios, since those are independent of the choice of k. We will also find that the two point correlation functions for the vector meson and nucleon interpolating fields satisfy a spectral decomposition consistent with QCD in the large N limit. We will use this decompostion to extract the decay constants of vector mesons and nucleons.

3 Vector mesons in confining holographic QCD

In this section, we will describe vector mesons in confining holographic QCD models based on Einstein-dilaton gravity. Firstly, we will present the 5d action and the equations of motion in both coordinate and momentum space. The VEVs and their connections with 4d currents are discussed. Subsequently, we will study the on-shell action and the bulk to boundary propagator, allowing us to obtain the two-point function. The spectral decomposition for the bulk to boundary propagator is described using Sturm-Lioiuville theory in order to obtain a spectral decomposition for the two-point function consistent with QCD in the large N_c limit. Lastly, we will obtain the spectrum and decay constants of vector mesons for the Einstein-dilaton models I and II described in section 2, comparing with previous models and available experimental data.

3.1 The 4d flavour currents

Consider vector mesons in large N_c QCD with $N_f = 2$ flavors. The flavour (isospin) currents responsible for creation of vector meson states can be written as

$$J^{\mu,a}(x) = \bar{q}(x)\gamma^{\mu}T^{a}q(x)\,, \tag{3.1}$$

where $q(x)$ is the quark doublet with components $u(x)$ and $d(x)$ and T^a, with $a = 1, 2, 3$, are the generators of the $SU(2)$ group. For simplicity we assume flavour (isospin) symmetry $m_u = m_d$ so that the flavour current is conserved. The matrix element for this current

when applied to the vacuum and a vector meson state can be written as

$$\langle 0|J^{\mu,a}(0)|V^{n,b}(p,\lambda)\rangle = F_{v^n}\,\epsilon^{\mu}(p,\lambda)\delta^{ab}\,, \tag{3.2}$$

where $\epsilon^{\mu}(p,\lambda)$ is the polarisation of the vector meson state. The coupling F_{v^n} is associated with the probability amplitude of creating a particular vector meson state from the vacuum. It can be related directly to the weak decay of vector mesons and for this reason it is known as the vector meson decay constant. Since the flavour current is conserved its conformal dimension is equal to its canonical dimension, so $\Delta = 3$ at all RG energy scales. Thus the decay constant F_{v^n} has dimension of mass squared [3]. In large N_c QCD the correlation function for two flavour currents admits the spectral decomposition [52, 53]

$$\langle J^{\mu,a}(q)J^{\nu,b}(q')\rangle = \delta^4(q-q')\delta^{ab}P_{\perp}^{\mu\nu}(q)\sum_n \frac{F_{v^n}^2}{q^2+m_{v^n}^2}\,, \tag{3.3}$$

where

$$P_{\perp}^{\mu\nu}(q) = \eta^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\,, \tag{3.4}$$

is the transverse projector which appear in propagators of massive spin 1 states (vector mesons). The result in (3.3) was obtained previously for some particular holographic QCD models [15, 26, 27]. In this section we will show that a general class of holographic QCD models based on Einstein-dilaton gravity lead to current correlators that satisfy the spectral decomposition (3.3).

3.2 The 5d action and field equations

We start with a set of 5d gauge fields $V_m^a(z,x)$ dual to the 4d flavour currents $J_\mu^a(x)$. In order to describe the spectrum of vector mesons we only need a 5d action quadratic on these fields. Assuming a minimal coupling to the metric and dilaton field, the action can be written as

$$S_V = -\int d^4x\,dz \frac{1}{4g_5^2}\sqrt{-g_s}\,e^{-\Phi}v_{mn}^a{}^2\,, \tag{3.5}$$

where $v_{mn}^a = \partial_m V_n^a - \partial_n V_m^a$ are the (Abelian) field strengths [4], the 5d metric g_{mn}^s is in the string frame and the index $a = 1,2,3$ is implicitly summed. The gauge coupling is fixed as $g_5^2 = 12\pi^2/N_c$ in order to reproduce the perturbative QCD result for the current correlator at small distances [15]. The action in (3.5) can be obtained from the vectorial sector of holographic models of chiral symmetry breaking after expanding at quadratic order the 5d Yang-Mills-Higgs action associated with the breaking of $SU(2)_L \times SU(2)_R$ chiral symmetry, see for example [15, 17, 34, 54, 55].

As described in the previous section, in holographic QCD models based on Einstein-dilaton gravity the string frame metric can be written as

$$g_{mn}^s = e^{2A_s(z)}\eta_{\hat{m}\hat{n}}\,, \tag{3.6}$$

[3] The meson states are normalised as $\langle\vec{p}|\vec{q}\rangle = 2E_{\vec{p}}(2\pi)^3\delta^3(\vec{p}-\vec{q})$.

[4] Non-Abelian terms are of cubic or higher order on the fields V_m^a and are relevant only to describe interactions.

where $A_s(z)$ is the string frame warp factor and the indices $(\hat{m}, \hat{n})$ correspond to coordinates in the $5d$ flat metric. Then the action in (3.5) becomes

$$S_2 = -\int d^4x \int dz \, \frac{1}{4g_5^2} e^{A_s - \Phi} v^a_{\hat{m}\hat{n}}{}^2 \,. \tag{3.7}$$

Varying the action (3.7), in order to get the field equations, we will have both the contribution of the bulk action and the boundary

$$\delta S_2 = \delta S_2^{\text{Bulk}} + \delta S_2^{\text{Bdy}} \tag{3.8}$$

where

$$\delta S_2^{\text{Bulk}} = \int d^4x \int dz \, \delta V^a_{\hat{n}} \partial_{\hat{m}} \left(\frac{1}{g_5^2} e^{A_s - \Phi} v_a^{\hat{m}\hat{n}} \right) , \tag{3.9}$$

and

$$\delta S_2^{\text{Bdy}} = -\int d^4x \int dz \, \partial_{\hat{m}} \left(\frac{1}{g_5^2} e^{A_s - \Phi} v_a^{\hat{m}\hat{n}} \delta V^a_{\hat{n}} \right) . \tag{3.10}$$

Imposing periodic boundary conditions in the x^μ coordinates, the boundary term reduces to

$$\delta S_2^{\text{Bdy}} = -\int d^4x \left(\frac{1}{g_5^2} e^{A_s - \Phi} v_a^{\hat{z}\hat{\mu}} \delta V^a_{\hat{\mu}} \right)^{z \to \infty}_{z = \epsilon} . \tag{3.11}$$

As described in the previous section, in this work we consider holographic QCD models where the dilaton is quadratic far from the boundary, c.f. (2.23). The string frame warp factor in that case becomes logarithmic far from the boundary, c.f. (2.26). Using these results we conclude that the surface term at $z \to \infty$ will vanish due to the presence of $e^{-\Phi}$. Imposing Dirichlet boundary condition for the fields $V^c_{\hat{\mu}}$ at the boundary $z = \epsilon$ one guarantees that $\delta S_2^{\text{Bdy}} = 0$.

The vanishing of δS_2^{Bulk} leads us to the Euler-Lagrange equations

$$\partial_{\hat{m}} \left(e^{A_s - \Phi} v_a^{\hat{m}\hat{n}} \right) = 0 \,. \tag{3.12}$$

These can be understood as a generalization of Maxwell equations for the fields $V^a_{\hat{m}}$ in a background with metric g^s_{mn}, given in (3.6) and a dilaton $\Phi(z)$. These equations are invariant under the gauge transformation

$$V_{\hat{m},a} \to V_{\hat{m},a} - \partial_{\hat{m}} \lambda^a_V \,. \tag{3.13}$$

We can write (3.12) in terms of the coordinates z and $\hat{\mu}$ decomposing the gauge field $V^a_{\hat{m}} = (V^a_z, V^a_{\hat{\mu}})$ and the derivatives $\partial_{\hat{m}} = (\partial_z, \partial_{\hat{\mu}})$. In this way, the equation (3.12) written in components is expressed as

$$\begin{aligned} &\left[\partial_z + A_s' - \Phi' \right] \left(\partial_z V^{\hat{\mu},a} - \partial^{\hat{\mu}} V^a_z \right) + \Box V^{\hat{\mu},a} - \partial^{\hat{\mu}} \left(\partial_{\hat{\nu}} V^{\hat{\nu},a} \right) = 0 \,, \\ &\Box V^a_z - \partial_z \left(\partial_{\hat{\mu}} V^{\hat{\mu},a} \right) = 0 \,. \end{aligned} \tag{3.14}$$

The gauge symmetry (3.13) allows us to define $V_z^a = 0$. The quadri-dimensional vector $V_{\hat{\mu}}$ admit the Lorentz decomposition

$$V_{\hat{\mu},a} = V_{\hat{\mu},a}^{\perp} + \partial_{\hat{\mu}}\xi^a \tag{3.15}$$

where $V_{\hat{\mu},a}^{\perp}$ is the transverse vector field and ξ^a are massless scalar fields not present in QCD. Since it is not possible to find normalisable modes for these fields we can set ξ^a to zero.

Using these results the equations (3.14) reduce to

$$\Big[\partial_z + A_s' - \Phi'\Big]\partial_z V_{\perp}^{\hat{\mu},a} + \Box V_{\perp}^{\hat{\mu},a} = 0 \tag{3.16}$$

where $V_{\perp}^{\hat{\mu},a}$ is the physical field that describe the vector mesons. Taking the 4d Fourier transform one obtains

$$\Big[\partial_z + A_s' - \Phi'\Big]\partial_z V_{\perp}^{\hat{\mu},a} - q^2 V_{\perp}^{\hat{\mu},a} = 0 \tag{3.17}$$

3.3 VEVs of the 4d flavour currents

In this subsection the vacuum expectation values (VEVs) of the 4d flavour currents. We start by writing the boundary term (3.11) as

$$\delta S_2^{Bdy} = \int d^4x \Big[\frac{1}{g_5^2} e^{A_s - \Phi} v_a^{\hat{z}\hat{\mu}} \delta V_{\hat{\mu}}^a\Big]_{z=\epsilon}. \tag{3.18}$$

As described in the previous susbsection, the surface term at $z \to \infty$ vanishes due to the dilaton asymptotic behaviour. At small z (near the boundary) we can approximate the metric by the AdS metric and solve the equation (3.16). We find that the vector gauge field can be expanded at small z as

$$V_{\hat{\mu},c}(x,z) = V_{\hat{\mu},c}^{(0)}(x) + \cdots + V_{\hat{\mu},c}^{(2)}(x) z^2 + \dots, \tag{3.19}$$

where $V_{\hat{\mu},c}^{(0)}(x)$ are the 4d external sources and $V_{\hat{\mu},c}^{(2)}(x)$ are the VEV coefficients. The VEV of the flavour currents responsible for the creation of vector mesons, according to the holographic dictionary, is given by

$$\langle J^{\hat{\mu},a}(x)\rangle = \frac{\delta S_2^{o-s}}{\delta V_{\hat{\mu},a}^{(0)}(x)} = \frac{\delta S_2^{Bdy}}{\delta V_{\hat{\mu},a}^{(0)}(x)} = \frac{1}{g_5^2}\Big[e^{A_s-\Phi} v_a^{\hat{z}\hat{\mu}}\Big]_{z=\epsilon} = \frac{1}{g_5^2}\Big[e^{A_s-\Phi}\partial_z V^{\hat{\mu},a}\Big]_{z=\epsilon}, \tag{3.20}$$

where the last equality holds for the gauge $V_z^a = 0$. Note that it is possible to write the VEVs in (3.20) in terms of the VEV coefficients $V_{\hat{\mu},c}^{(2)}(x)$. In this work it will be sufficient to use the result (3.20). Later in this section we will derive a Sturm-Liouville expansion for the vector fields that will lead to a spectral decomposition for the correlator of flavour currents.

3.4 The on-shell action, the bulk to boundary propagator and the two point function

In this subsection we will write the on-shell action in terms of bulk to boundary propagator and the 4d sources. We will establish the connection between the bulk to boundary propagator, the VEV of the 4d flavour currents and the correlator of flavour currents.

First we evaluate the action in (3.7) on-shell and find

$$S_2^{o-s} = S_{2,Bdy}^{o-s} + S_{2,Bulk}^{o-s}\,, \tag{3.21}$$

where

$$S_{2,Bdy}^{o-s} = -\int d^4x \int dz\, \partial_{\hat{m}}\Big[\frac{1}{2g_5^2} e^{A_s-\Phi} v_a^{\hat{m}\hat{n}} V_{\hat{n}}^a\Big]\,, \tag{3.22}$$

and

$$S_{2,Bulk}^{o-s} = \int d^4x \int dz\, V_{\hat{n}}^a \partial_{\hat{m}}\Big(\frac{1}{2g_5^2} e^{A_s-\Phi} v_a^{\hat{m}\hat{n}}\Big) = 0\,. \tag{3.23}$$

We remind the reader that the indices $(\hat{m},\hat{n})$ are raised or lowered using the 5d flat metric $\eta_{\hat{m}\hat{n}}$. As expected, the on-shell action becomes a surface term. Using again periodic boundary conditions for the x^μ coordinates and the condition that the surface term at $z\to\infty$ vanishes, due to the asymptotic behavior of the dilaton, the on-shell action is reduced to

$$S_2^{o-s} = \int d^4x \Big[\frac{1}{2g_5^2} e^{A_s-\Phi}(\partial_z V_a^{\hat{\mu}}) V_{\hat{\mu}}^a\Big]_{z=\epsilon}\,, \tag{3.24}$$

where we also used the gauge condition $V_z^a = 0$. We can define the bulk to boundary propagator in real space by the relation

$$V_{\hat{\mu}}^a(z,x) = \int d^4y\, K_{\hat{\mu}\hat{\nu}}^{ab}(z,x;y) V_b^{\hat{\nu},0}(y). \tag{3.25}$$

where $K_{\hat{\mu}\hat{\nu}}^{cd}(z,x;y)$ is the bulk to boundary propagator (in real space) and $V_b^{\hat{\nu},0}(y)$ is the 4d external source. Plugging (3.25) into (3.24) yields

$$S_2^{o-s} = \int d^4x \int d^4y \Big\{\frac{1}{2g_5^2} V_c^{\hat{\mu},0}(x)\Big[e^{A_s-\Phi}\partial_z K_{\hat{\mu}\hat{\nu}}^{cd}(z,x;y)\Big]_{z=\epsilon} V_d^{\hat{\nu},0}(y)\Big\}\,, \tag{3.26}$$

The VEV of the flavour currents (3.20) can also be expressed in terms of the bulk to boundary propagator:

$$\langle J_{\hat{\mu},c}(x)\rangle = \frac{1}{g_5^2}\Big[e^{A_s-\Phi}\partial_z V_{\hat{\mu},c}\Big]_{z=\epsilon} = \frac{1}{g_5^2}\int d^4y \Big[e^{A_s-\Phi}\partial_z K_{\hat{\mu}\hat{\nu}}^{cd}(z,x;y)\Big]_{z=\epsilon} V_d^{\hat{\nu},0}(y)\,. \tag{3.27}$$

Varying the on-shell action in (3.26) we obtain

$$\delta S_2^{o-s} = \int d^4x\, \langle J_{\hat{\mu},c}(x)\rangle \delta V_c^{\hat{\mu},0}\,, \tag{3.28}$$

as expected. Note that we used the $x \leftrightarrow y$ symmetry in the bulk to boundary propagators. According to the holographic dictionary the correlator of flavour currents in real space corresponds to

$$G^{cd}_{\hat{\mu}\hat{\nu}}(x-y) = \langle J_{\hat{\mu},c}(x) J_{\hat{\nu},d}(y)\rangle = \frac{\delta S_2^{o-s}}{\delta V_c^{\hat{\mu},0}(x)\delta V_d^{\hat{\nu},0}(y)} = \frac{1}{g_5^2}\Big[e^{A_s-\Phi}\partial_z K^{cd}_{\hat{\mu}\hat{\nu}}(z,x;y)\Big]_{z=\epsilon} \tag{3.29}$$

The relation between the VEV and the source is expressed through the two-point function, given by

$$\langle J_{\hat{\mu},c}(x)\rangle = \int d^4y\, G^{cd}_{\hat{\mu}\hat{\nu}}(x-y) V_d^{\hat{\nu},0}(y)\,. \tag{3.30}$$

3.5 Spectral decomposition for the bulk to boundary propagator

In subsection 3.2 we saw that the vector mesons are described by a transverse vector field. This implies that the bulk to boundary propagator in momentum space takes the form

$$\tilde{K}^{ab}_{\hat{\mu}\hat{\nu}}(z,q) = P_\perp^{\mu\nu}(q)\delta^{ab}\, V(z,q)\,, \tag{3.31}$$

where $P_\perp^{\mu\nu}(q)$ is the transverse projector, defined in (3.4), and $V(z,q)$ a scalar function that carries all the information of the the bulk to boundary propagator in momentum space.

In momentum space the 5d gauge fields can be written as

$$\tilde{V}^a_{\hat{\mu}}(z,q) = \tilde{K}^{ab}_{\hat{\mu}\hat{\nu}}(z,q)\tilde{V}_b^{\hat{\nu},0}(q)\,. \tag{3.32}$$

Using these relations in the field equation (3.17) we obtain

$$\Big[\Big(\partial_z + A_s' - \Phi'\Big)\partial_z - q^2\Big]V(z,q) = 0\,, \tag{3.33}$$

which is an ordinary second order differential equation for the bulk to boundary propagator. It is convenient to rewrite this equation as

$$\Big[\partial_z\Big(e^{A_s-\Phi}\partial_z\Big) - q^2 e^{A_s-\Phi}\Big]V(z,q) = 0\,. \tag{3.34}$$

This equation can be writen as a Sturm-Liouville equation

$$\Big[\mathcal{L} + \lambda\, r(z)\Big]y(z) = f(z) \quad , \quad \mathcal{L} = \partial_z\Big(p(z)\partial_z\Big) - s(z)\,, \tag{3.35}$$

where we identify

$$p(z) = e^{A_s-\Phi}\ ,\ s(z) = 0\ ,\ \lambda = -q^2\ ,\ r(z) = e^{A_s-\Phi} \text{ and } f(z) = 0\,(\text{homogeneous})\,. \tag{3.36}$$

The Sturm-Liouville theory is briefly described in appendix A. In the non-homogeneous case, i.e. $f(z) \neq 0$, we can define the Green's function by the equation

$$\Big[\mathcal{L} + \lambda\, r(z)\Big]G(z;z') = \delta(z-z')\,. \tag{3.37}$$

Now we will define an infinite set of eigenfunctions, $v^n(z)$, that obey the eigenvalue equation

$$\Big[\mathcal{L} + \lambda_n r(z)\Big]v^n(z) = 0\,, \tag{3.38}$$

or

$$\left[\partial_z\left(e^{A_s-\Phi}\partial_z\right)+m_{v^n}^2 e^{A_s-\Phi}\right]v^n(z)=0 \tag{3.39}$$

where $\lambda_n = m_{v^n}^2$ are the eigenvalues. Note that these Sturm-Liouville modes are essentially the normalisable modes in holographic QCD. Indeed, these modes satisfy the orthonormality condition

$$\int dz\, e^{A_s-\Phi}v^m(z)v^n(z)=\delta^{mn}\,, \tag{3.40}$$

and the Green's function admits the spectral decomposition

$$G(z;z')=-\sum_n \frac{v^n(z)v^n(z')}{q^2+m_{v^n}^2}\,. \tag{3.41}$$

For more details see appendix A.

We can find a relation between the bulk to boundary propagator $V(z,q)$, corresponding to the homogeneous solution, can be written in terms of the Green's function, associated with the non-homogeneous solution as follows.

Multiplyng both sides of (3.37) by $V(q,z)$, integrating over z and using (3.34) we obtain

$$V(z',q)=\left[e^{A_s(z)-\Phi(z)}(V(z,q)\partial_z G(z;z')-G(z;z')\partial_z V(z,q))\right]_{z=\epsilon}^{z\to\infty} \tag{3.42}$$

For a dilaton that is quadratic at large z, it is possible to show that the surface term at $z\to\infty$ vanishes so we end up with the relation

$$V(z',q)=-\left[e^{A_s-\Phi}\partial_z G(z;z')\right]_{z=\epsilon}\,, \tag{3.43}$$

where we also used the boundary condition $V(\epsilon,q)=1$. Substituting the spectral decomposition (3.41) in (3.43), we find

$$V(z',q)=\sum_n c_n(q^2)v^n(z')\,, \tag{3.44}$$

where

$$c_n(q^2)=\frac{\left[e^{A_s-\Phi}\partial_z v^n(z)\right]_{z=\epsilon}}{q^2+m_{v^n}^2}\,. \tag{3.45}$$

Using this result and the orthonormality condition (3.40) we obtain

$$\int dz\, e^{A_s-\Phi}v^m V(z,q)=c_m(q^2) \tag{3.46}$$

replacing this result in (3.44) we obtain the completeness relation for the normalisable (Sturm-Liouville) modes

$$\sum_n e^{A_s-\Phi}v^n(z)v^n(z')=\delta(z-z')\,. \tag{3.47}$$

Plugging (3.44) into (3.31) the tensorial bulk to boundary propagator becomes

$$\tilde{K}^{ab}_{\hat{\mu}\hat{\nu}}(z,q)=P_\perp^{\mu\nu}(q)\delta^{ab}\sum_n c_n(q^2)v^n(z)\,. \tag{3.48}$$

3.6 The 4d current correlator

The 2-point current correlator in real space was obtained in (3.29) from the bulk to boundary propagator. In momentum space it takes the form

$$G^{ab}_{\hat{\mu}\hat{\nu}}(q) = \frac{1}{g_5^2}\Big[e^{A_s-\Phi}\partial_z \tilde{K}^{ab}_{\hat{\mu}\hat{\nu}}(z,q)\Big]_{z=\epsilon} \tag{3.49}$$

Using the spectral decomposition (3.48) with the coefficients (3.45), the 2-point function becomes

$$G^{ab}_{\hat{\mu}\hat{\nu}}(q) = \left(\eta_{\hat{\mu}\hat{\nu}} - \frac{q_{\hat{\mu}}q_{\hat{\nu}}}{q^2}\right)\delta^{ab}\sum_n \frac{F^2_{v^n}}{q^2+m^2_{v^n}}\,, \tag{3.50}$$

where the coefficients F_{v^n} are defined as

$$F_{v^n} = \frac{1}{g_5}\Big[e^{A_s-\Phi}\partial_z v_n(z)\Big]_{z=\epsilon}\,. \tag{3.51}$$

The F_{v^n} can be interpreted as probability amplitudes associated with the creation of vector mesons from the vacuum. They are commonly known as vector meson decay constants because they are relevant for describing the weak decay of vector mesons.

The result in (3.50) is very general for holographic QCD models based on Einstein-dilaton gravity. It is consistent with (3.3), which is the spectral decompostion for a current correlator in large N_c QCD. Note the appearance in (3.50) of 4d propagators for the vector mesons. The vector meson propagator can be obtained as a particular case of the Proca propagator, as described in appendix B.

3.7 Spectrum of vector mesons

To obtain the spectrum of vector mesons we need to solve the eigenvalue problem for the normalisable (Sturm-Liouville) modes

$$\left[\partial_z\left(e^{A_s-\Phi}\partial_z\right) + m^2_{v^n}e^{A_s-\Phi}\right]v^n(z) = 0 \tag{3.52}$$

We can write this equation in the form of a Schrödinger equation considering the Bogoliubov transformation

$$v^n = e^{-B_V}\psi_{V^n} \quad \text{where} \quad B_V = \frac{1}{2}\left(A_s-\Phi\right)\,. \tag{3.53}$$

Plugging (3.53) into (3.52), we find the following Schrödinger equation

$$\Big[\partial_z^2 + m^2_{v^n} - V_V\Big]\psi_{V^n} = 0\,, \tag{3.54}$$

V_V is expressed by

$$V_V = B_V'' + B_V'^2\,. \tag{3.55}$$

From the Schrödinger equation, we can derive the mass spectrum and the wave functions associated with the normalisable (Sturm-Liouville) modes. Notice that the dominant contribution to the Schrödinger potential at large z (far from the boundary) is given by the

dilaton which is quadratic in z for large z. This in turn guarantees the condition that the Schrödinger potential is quadratic in z at large z leading to asymptotically linear Regge trajectories.

Figure 4 displays our numerical results for the Schrödinger potential for vector mesons in the Einstein-dilaton models I and II (blue and red lines respectively), compared against the soft wall model (black dashed line). Note that the potentials of the Einstein-dilaton models are very similar to the potential in the soft wall model. The main difference between them is that model I (model II) displays a miminum at lower (higher) energy than the soft wall model. One may conclude from this analysis that model I (model II) leads to a lower (higher) mass for the fundamental state ρ_0. However, the mass of the fundamental state also depends on the infrared parameter k which can be fixed differently for each model. In this work we will consider only dimensionless mass ratios so that we do not need to fix the infrared parameter k.

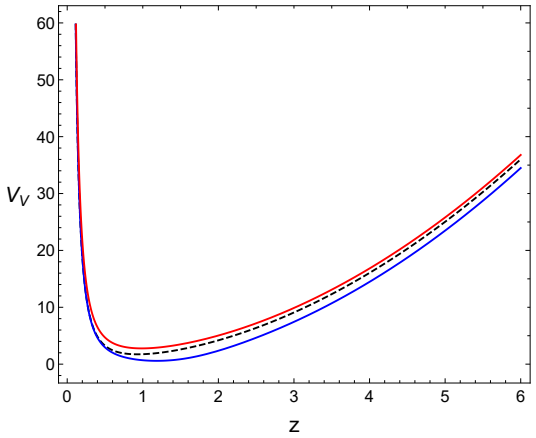


Figure 4. Schrödinger potentials for vector mesons in Einstein- dilaton models I and II (blue and red solid lines) and soft wall model (black dashed line).

Asymptotic solution and numerical integration

In order to find the spectrum of vector mesons we need to solve the differential equation (3.52) or equivalently the Schrödinger equation (3.54). We first find the asymptotic solution at small z

$$v^n(z) = N_{V^n} z^2 + \dots \quad , \quad \text{or} \quad \psi_{V^n}(z) = N_{V^n} z^{3/2} + \dots \,, \tag{3.56}$$

where N_{V^n} is a constant necessary for the normalisation condition

$$\int dz\, \psi_{V^n}(z)^2 = 1\,. \tag{3.57}$$

The eigenvalues of the problem can be obtained integrating numerically either (3.52) or (3.54) and imposing the following behaviour at large z

$$\lim_{z\to\infty} \sqrt{z}\, \psi_{V^n}(z) = 0\,, \tag{3.58}$$

which guarantees that the solution is normalisable. The numerical procedure, commonly known as the shooting method, consists of shooting the value of m_{V^n} until one finds a solution that satisfies the condition (3.58). In this way one finds a discrete set of eigenvalues corresponding to the vector meson masses.

Spectrum

We present in table 1 our results for the spectrum of vector mesons in the Einstein-dilaton models I and II described in section 2. As described above, we consider only dimensionless mass ratios so that we can compare different models without fixing the infrared parameter k. We show in table 1 our results for the mass ratios m_{ρ^n}/m_{ρ^0} for the first five excited states, i.e. $n = 1, .., 5$. The mass of the fundamental state m_{ρ^0} can later be fixed to the corresponding experimental value fixing the infrared parameter k. We compare our results for models I and II with previous results obtained using the soft wall and hard wall models and also against experimental results.

Ratio	Model I	Model II	Soft wall	Hard wall	Experimental
m_{ρ^1}/m_{ρ^0}	1.591	1.34	1.414	2.295	1.652 ± 0.048
m_{ρ^2}/m_{ρ^0}	2.015	1.611	1.732	3.598	1.888 ± 0.032
m_{ρ^3}/m_{ρ^0}	2.365	1.843	2	4.903	2.216 ± 0.026
m_{ρ^4}/m_{ρ^0}	2.67	2.049	2.236	6.209	2.443 ± 0.072
m_{ρ^5}/m_{ρ^0}	2.944	2.236	2.45	7.514	2.727 ± 0.265

Table 1. Ratio of vector meson masses m_{ρ^n}/m_{ρ^0} for the first excited states $n = 1, .., 5$ in the Einstein-dilaton models I and II, the soft wall model and the hard wall model, compared against experimental results. The experimental result for m_{ρ^1} was taken from [56] and the experimental results for the other states were obtained from PDG [57], including the mass of the fundamental state $m_{\rho^0} = 0.776 \pm 0.001\,\text{GeV}$.

Figure 5 shows the behaviour of the squared masses of vector mesons as a function of the radial excitation number in the Einstein-dilaton models I and II (blue and red solid lines) and the soft wall model (black dashed line), compared against experimental data (orange dots and error bars). As expected, the Einstein-dilaton models I and II lead to approximately linear Regge trajectories whilst the Regge trajectory in the soft wall model is exactly linear. The main difference between the Einstein-dilaton models and the soft wall model is that the masses of excited states grow faster (slower) in model I (model II) than in the soft wall model leading to a higher (smaller) slope. Note that the Einstein-dilaton model I and the soft wall model provide results that are closer to the experimental data.

3.8 Wave functions and vector meson decay constants

Besides calculating the spectrum, it is important to investigate the vector meson wave functions. This allows us to identify the emergence of the fundamental state V^0 and the excited states V^n with $n = 1, 2, \ldots$ by a comparison with normal modes in wave mechanics. From the small z behaviour of the vector meson wave functions we will also be able to extract the vector meson decay constants F_{V^n}.

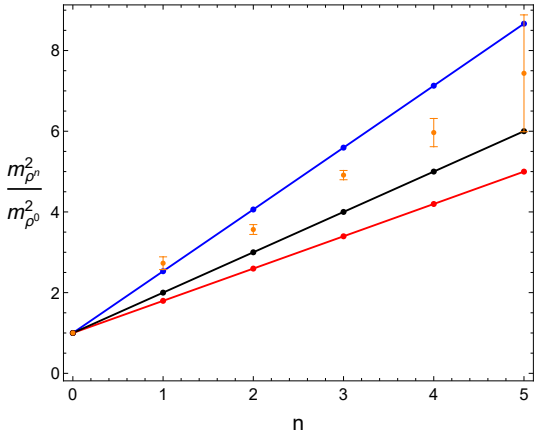


Figure 5. Dimensionless squared mass ratios $m^2_{\rho^n}/m^2_{\rho^0}$ for vector mesons as a function of the radial excitation number n in the Einstein-dilaton models I and II (blue and red solid lines and dots) and the soft wall model (black solid line and dots), compared against experimental data (orange dots and error bars).

Figure 6 illustrates the behavior of the vector meson wave functions in Einstein-dilaton models I and II (blue and red solid curves) and the soft wall model (black dashed curve). Note that the wave functions in models I and II are not very different to the wave functions in the soft wall model. The discrepancy occurs at small and intermediate values of z. This is expected because the Einstein-dilaton models affect the differential equation (3.52) through the dilaton and the AdS space deformation while in the soft wall model the AdS space is not deformed. At large z the quadratic dependence of the dilaton field is expected to be the dominant contribution to the differential equation which is the same as in the soft wall model.

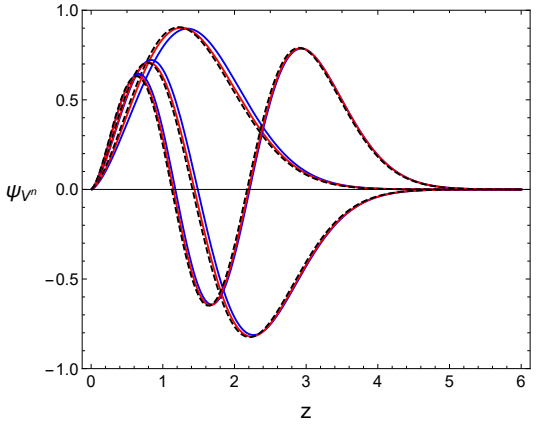


Figure 6. Wave functions of vector mesons in the Einstein- dilaton models I and II (blue and red solid curves) and the soft wall model (black dashed curve).

We finally evaluate the vector meson decay constant as follows

$$F_{v^n} = \frac{1}{g_5}\Big[e^{A_s-\Phi}\partial_z v_n(z)\Big]_{z=\epsilon} = \frac{2}{g_5}N_{V^n}\,, \tag{3.59}$$

where we used the small z behaviour of the normalisable mode (3.56), the AdS asymptotic behaviour for A_s and the property that the dilaton vanishes at the AdS boundary. The normalisation contants N_{V^n} are calculated numerically using the normalisation condition (3.57). In table 2 we present our results for the dimensionless ratios $\sqrt{F_{\rho^n}}/m_{\rho^0}$ for Einstein-dilaton models I and II, compared against the soft wall model and the hard wall model. For the fundamental state we also compare against the experimental result. We conclude that, although the Einstein-dilaton model I provides a better result than the soft wall model, the hard wall model still provides the best result. We would like to remark that the results for the vector meson decay constants in the case of excited states are theoretical predictions from holographic QCD. In particular, we note that all holographic QCD models predict that the vector meson decay constants grow with the radial excitation number. We hope that these predictions will be tested in the near future.

Ratio	Model I	Model II	Soft wall	Hard wall	Experimental
$\sqrt{F_{\rho^0}}/m_{\rho^0}$	0.3719	0.283	0.3355	0.4246	0.446 ± 0.0019
$\sqrt{F_{\rho^1}}/m_{\rho^0}$	0.4704	0.3407	0.3989	0.7946	-
$\sqrt{F_{\rho^2}}/m_{\rho^0}$	0.5298	0.3798	0.4415	1.114	-
$\sqrt{F_{\rho^3}}/m_{\rho^0}$	0.5741	0.41	0.4744	1.405	-
$\sqrt{F_{\rho^4}}/m_{\rho^0}$	0.61	0.4351	0.5017	1.677	-

Table 2. Dimensionless ratios $\sqrt{F_{\rho^n}}/m_{\rho^0}$ for vector meson decay constants in the Einstein-dilaton models I and II, the soft wall model and the hard wall model, compared against the experimental result. The experimental result was obtained using $\sqrt{F_{\rho^0}} = 0.3462 \pm 0.0014\,\text{GeV}$ [58] and $m_{\rho^0} = 0.776 \pm 0.001\,\text{GeV}$ [57]. .

4 Nucleons in confining holographic QCD

In this section, we describe 1/2 spin baryons, more specifically the nucleons (proton and neutron). We first present the so-called Ioffe currents which are spinorial operators associated with the creation of nucleons and describe the spectral decomposition for the nucleon correlator in large N_c QCD. Next, we present the 5d action for the Dirac spinor dual to the nucleon operator and derive the equations of motion. Subsequently, we study the VEV of nucleon operators and from the on-shell action we obtain the two-point nucleon correlation function. Next, we investigate the spectral decomposition for the bulk to boundary propagator using the Sturm-Liouville theory and we find a spectral decomposition for the nucleon correlator consistent with large N_c QCD. Finally, we obtain the spectrum and decay constants for nucleons for the Einstein dilaton models I and II described in section 2 and compare against previous models and available experimental data.

4.1 The 4d nucleon operator

Consider nucleons in large N_c QCD with $N_f = 2$ flavours. For simplicity we consider isospin symmetry, i.e. $m_u = m_d$. The creation of nucleon states can be described by

nucleon operators built from the quark fields. For the case of the proton, the nucleon operator takes the form of the Ioffe current [59, 60]

$$\mathcal{O}(x) = \epsilon_{abc}\left(u_a^T(x)C\gamma_\mu u_b(x)\right)\gamma_5\gamma^\mu d_c(x)\,, \tag{4.1}$$

where u and d are the quark fields, a, b, and c are color indices, and C is the charge conjugation operator. The operator (4.1) has $I_z = 1/2$ corresponding to proton states. A similar operator can be constructed for the neutron states ($I_z = -1/2$) replacing the uud structure by a ddu structure. The matrix element for the nucleon operator when applied to the vacuum and a nucleon state can be written as

$$\langle 0|\,\mathcal{O}(0)\,|N^n(p)\rangle = \lambda_{N^n}\,u^n(p) \tag{4.2}$$

where $u^n(p)$ is the Dirac spinor corresponding to the nucleon state. The coupling λ_{N^n} is associated with the probability amplitude of creating a particular nucleon state from the vacuum. Although there is no direct connection of these couplings to the weak decay of the neutron we will nevertheless call them nucleon decay constants. If the conformal dimension Δ of the nucleon operator $\mathcal{O}$ is equal to the canonical dimension we have $\Delta = 9/2$ and the nucleon decay constant λ_{N^n} has dimension of mass cubed [5]. If we take into account the effect of the anomalous dimension one would obtain $\Delta < 9/2$ and λ_n would have dimension $M^{\Delta-3/2}$. In this work we will investigate the spectrum of nucleons for the cases $\Delta = 9/2$ and $\Delta = 7/2$ using holographic QCD based on Einstein-dilaton gravity.

In large N_c QCD the nucleon correlator admits the following spectral decomposition [52, 61]

$$\langle\mathcal{O}(q)\bar{\mathcal{O}}(q')\rangle = i\delta^4(q-q')\sum_n \lambda_{N^n}^2\frac{(i\not{q}+m_{N^n})}{q^2+m_{N^n}^2}\,. \tag{4.3}$$

On the right hand side we identify the Dirac propagators associated with the different nucleon states. In holographic QCD we are interested on the two point correlation function of the right part of the nucleon correlator, namely

$$\begin{aligned}\langle\mathcal{O}_R(q)\bar{\mathcal{O}}_R(q')\rangle &= P_R\langle\mathcal{O}(q)\bar{\mathcal{O}}(q')\rangle P_L \\ &= \delta^4(q-q')(-P_R\not{q})\sum_n\frac{\lambda_{N^n}^2}{q^2+m_{N^n}^2}\,,\end{aligned} \tag{4.4}$$

where

$$P_{R/L} = \frac{1}{2}\left(\mathbb{1}\pm\gamma^5\right)\,, \tag{4.5}$$

are the right and left chiral projectors. The result in (4.4) was obtained previously in the soft wall model [38]. In this section we will show that a general class of holographic QCD models based on Einstein-dilaton gravity lead to nucleon correlators that satisfy the spectral decomposition (4.4).

[5] The nucleon states are normalised as $\langle\vec{p}|\vec{q}\rangle = 2E_{\vec{p}}\,(2\pi)^3\delta^3(\vec{p}-\vec{q})$ and the Dirac spinors as $\bar{u}^r(p)u^s(p) = 2m\,\delta^{rs}$.

4.2 The 5d action and field equations

We start with a 5d Dirac field $\psi(z,x)$ dual to the 4d nucleon operator $\mathcal{O}$. The dynamics of the 5d Dirac field can be obtained coupling the Dirac spinor to a background given by Einstein-dilaton gravity. The generalised 5d Dirac action action in the string frame can be written as

$$S_F = G_F \int d^5x\sqrt{-g_s}\, e^{-\Phi}\Big(\frac{i}{2}\bar{\psi}\not{D}\psi + \text{c.c.} - i\tilde{m}\bar{\psi}\psi\Big) + \Delta S\,, \tag{4.6}$$

where ψ and $\bar{\psi}$ are the Dirac spinor and its adjoint respectively, with $\bar{\psi} = \psi^\dagger\Gamma^{\hat{0}}$. We have included a surface term ΔS necessary for the variational principle. The coupling $\tilde{m}$ is a generalisation of the mass term that may include first derivatives of the metric and the dilaton. The 5d coupling constant G_F will be determined later when comparing the result for the two point nucleon correlator at high energies with the perturbative QCD result.

The covariant derivative in the Feynman notation is given by

$$\not{D} = \Gamma^n D_n \tag{4.7}$$

where the (curved space) gamma matrices, Γ^n, and the covariant derivative, D_n, explicitly are

$$\Gamma^n = e^n_{\hat{a}}\Gamma^{\hat{a}}\,, \tag{4.8}$$

$$D_n = \partial_n + \frac{1}{8}\omega_n^{\hat{a}\hat{b}}\left[\Gamma_{\hat{a}},\Gamma_{\hat{b}}\right] = \partial_n + \frac{1}{4}\omega_n^{\hat{a}\hat{b}}\Gamma_{\hat{a}\hat{b}}\,. \tag{4.9}$$

The quantities $e^n_{\hat{a}}$ and $\omega_n^{\hat{a}\hat{b}}$ are the veilbein and spin connection respectively and $\Gamma^{\hat{a}}$ are the gamma matrices in 5d flat space. For the string frame metric given in (3.6) they take the form

$$e^n_{\hat{a}} = e^{-A_s(z)}\delta^n_{\hat{a}}\,, \tag{4.10}$$

$$\omega_n^{\hat{a}\hat{b}} = e^{m\hat{a}}\nabla_n e^{\hat{b}}_m\,. \tag{4.11}$$

Note that m,n are tensorial indices associated with the curved space g_s^{mn} whilst $\hat{a},\hat{b}$ are tensorial indices associated with the tangent flat space $\eta_{\hat{a}\hat{b}}$. The gamma matrices in the tangent space satisfy the Clifford algebra

$$\{\Gamma^{\hat{a}},\Gamma^{\hat{b}}\} = 2\eta^{\hat{a}\hat{b}}\mathbb{1}\,. \tag{4.12}$$

The coupling $\exp(-\Phi)$ in the Dirac action can be absorbed in the following redefinition of the Dirac spinor

$$\psi \to e^{\Phi/2}\psi\,. \tag{4.13}$$

Plugging (4.13) into action (4.6) we obtain

$$S_F = G_F \int d^5x\sqrt{-g_s}\Big(\frac{i}{2}\bar{\psi}\not{D}\psi + \text{c.c.} - i\tilde{m}\bar{\psi}\psi\Big) + \Delta S\,. \tag{4.14}$$

To find the equation of motion we first note that the only non-vanishing components of the spin connection are

$$\omega_{\mu}^{\hat{z}\hat{\nu}} = -\omega_{\mu}^{\hat{\nu}\hat{z}} = -A_s' \delta_{\mu}^{\hat{\nu}} \,. \tag{4.15}$$

Using (4.8), (4.9), (4.10) and (4.15), the Dirac operator acting on the Dirac field takes the form

$$\not{D}\psi = \Gamma^n D_n \psi = e^{-A_s} \left(\Gamma^{\hat{a}} \partial_{\hat{a}} + 2A_s' \Gamma^{\hat{z}} \right) \psi \,, \tag{4.16}$$

where the indices $\hat{a} = (\hat{z}, \hat{\mu})$ are contracted using the $5d$ Minkowski metric $\eta_{\hat{a}\hat{b}}$. Writing the action (4.14) in terms of the operator (4.16) we find

$$S_F = G_F \int d^5x \, e^{4A_s} \left(\frac{i}{2} \bar{\psi} \Gamma^{\hat{a}} \partial_{\hat{a}} \psi - \frac{i}{2} (\partial_{\hat{a}} \bar{\psi}) \Gamma^{\hat{a}} \psi - i e^{A_s} \tilde{m} \bar{\psi} \psi \right) + \Delta S \,. \tag{4.17}$$

The field equations are found by varying the action (4.17) with respect to ψ and $\bar{\psi}$. We obtain

$$\left(\Gamma^{\hat{a}} \partial_{\hat{a}} + 2A_s' \Gamma^{\hat{z}} - e^{A_s} \tilde{m} \right) \psi = 0 \,, \tag{4.18}$$

$$\bar{\psi} \left(\overleftarrow{\partial_{\hat{a}}} \Gamma^{\hat{a}} + 2A_s' \Gamma^{\hat{z}} + e^{A_s} \tilde{m} \right) = 0 \,. \tag{4.19}$$

It's interesting to write the equation (4.18) in terms of left and right chiralities of the Dirac field. Thus, in the decomposition $\psi = \psi_R + \psi_L$, the left and right components are given by

$$\psi_{R/L} = \frac{1}{2} \left(\mathbb{1} \pm \Gamma^{\hat{z}} \right) \psi = P_{R/L} \psi \tag{4.20}$$

$$\bar{\psi}_{R/L} = \bar{\psi} \frac{1}{2} \left(\mathbb{1} \mp \Gamma^{\hat{z}} \right) = \bar{\psi} P_{L/R} \tag{4.21}$$

where P_R and P_L are the right and left chiral projectors. The left and right spinors are eigenstates of the chirality operator, $\Gamma^{\hat{z}} = \gamma^5$,

$$\Gamma^{\hat{z}} \psi_{R/L} = \pm \psi_{R/L} \,. \tag{4.22}$$

Plugging (4.20), (4.21) and (4.22) in the Dirac equation, (4.18), we arrive at the following system of coupled equations

$$\not{\partial} \psi_L = - \left(\partial_z + 2A_s' - e^{A_s} \tilde{m} \right) \psi_R \,, \tag{4.23}$$

$$\not{\partial} \psi_R = \left(\partial_z + 2A_s' + e^{A_s} \tilde{m} \right) \psi_L \,, \tag{4.24}$$

and a similar system for their adjoints. Acting on the right with the operator $\not{\partial}$ in (4.23) and (4.24), we obtain the decoupled second-order differential equations

$$\Box \psi_{R/L} = - \left(\partial_z + 2A_s' \pm e^{A_s} \tilde{m} \right) \left(\partial_z + 2A_s' \mp e^{A_s} \tilde{m} \right) \psi_{R/L} \tag{4.25}$$

The general solutions for the left and right Dirac fields can be written as

$$\psi_{R/L}(x, z) = \int d^4q \, e^{iq \cdot x} \, F_{R/L}(q, z) \alpha_{R/L}(q) \tag{4.26}$$

where $F_{R/L}(q,z)$ are the bulk to boundary propagators in momentum space for the right and left chiralities whilst $\alpha_{R/L}(q)$ are left and right spinorial sources in the 4d field theory. Plugging (4.26) into (4.25), we obtain the equation for the bulk to boundary propagator

$$\Big[\partial_z^2 + 4A_s'\partial_z + 2A_s'' + 4A_s'^2 \mp \partial_z\Big(e^{A_s}\tilde{m}\Big) - e^{2A_s}\tilde{m}^2 + Q^2\Big]F_{R/L} = 0\,, \tag{4.27}$$

where $Q = \sqrt{-q^2}$.

Alternatively, we can expand the right and left Dirac fields in terms of 4d modes as follows

$$\psi_{R/L}(x,z) = \sum_n f^n_{R/L}(z)\alpha^n_{R/L}(x)\,. \tag{4.28}$$

The 4d modes $\alpha^n_{R/L}(x)$ satisfy the coupled equations

$$\not{\partial}\alpha^n_{R/L} = m_{N^n}\,\alpha^n_{L/R}\,, \tag{4.29}$$

which are equivalent to the Dirac equation

$$(\not{\partial} - m_{N^n})\alpha^n = 0\,, \tag{4.30}$$

for the 4d Dirac spinor modes $\alpha^n(x) = \alpha^n_R(x) + \alpha^n_L(x)$. Using these results in (4.23) we find that the normalisable modes $f^n_{R/L}(q,z)$ obey the system of coupled equations

$$\Big(\partial_z + 2A_s' \mp e^{A_s}\tilde{m}\Big)f^n_{R/L} = \mp m_{N^n}\,f^n_{L/R}\,, \tag{4.31}$$

The second order decoupled equations for these normalisable modes take the form

$$\Big[\partial_z^2 + 4A_s'\partial_z + 2A_s'' + 4A_s'^2 \mp \partial_z\Big(e^{A_s}\tilde{m}\Big) - e^{2A_s}\tilde{m}^2 + m^2_{N^n}\Big]f^n_{R/L} = 0\,. \tag{4.32}$$

Note that the equations (4.32) can be thought as the eigenvalue equations associated with the bulk to boundary propagator satisfying equation (4.27).

In subsection 4.5 we will apply the Sturm-Liouville theory to the equations (4.27) in order to arrive at a spectral decomposition for the bulk to boundary propagator and in subsection 4.6 we will obtain the spectral decomposition of the 4d nucleon correlator. In subsection 4.7 we will use the equations (4.32) to find the spectrum of nucleons. But first we will obtain in the following two subsections the VEV of the 4d nucleon operator as well as the dictionary for the nucleon correlator in terms of the bulk to boundary propagator.

4.3 VEV of the 4d nucleon operator

In this subsection we obtain the holographic dictionary for the VEV of the right projection of the nucleon operator, namely

$$\langle\mathcal{O}_R(x)\rangle = P_R\langle\mathcal{O}(x)\rangle\,, \tag{4.33}$$

from the 5d action. The key observation is that this operator couples to a left spinorial source $\alpha_L(x)$ as

$$\int d^4x\,\Big(\bar{\alpha}_L(x)\langle\mathcal{O}_R(x)\rangle + c.c.\Big)\,. \tag{4.34}$$

The 4d spinorial source $\alpha_L(x)$ will appear as the leading term coefficient in the small z (UV) expansion of the 5d left spinor field $\psi_L(x,z)$. As described at the beginning of the section, the nucleon operator $\langle \mathcal{O}_R(x) \rangle$ has conformal dimension Δ. We will consider the cases $\Delta = 9/2$ (canonical dimension) and $\Delta = 7/2$ (including anomalous dimension). The 4d source $\alpha_L(x)$ have conformal dimension $4 - \Delta$.

Let us start with the variation of the action

$$\delta S_F = \delta S^{bulk} + \delta S^{Bdy}\,, \tag{4.35}$$

where

$$\delta S^{Bulk} = G_F \int d^5x \left(i e^{4A_s} \delta\bar{\psi} (\Gamma^{\hat{a}} \partial_{\hat{a}} + 2A_s' \Gamma^{\hat{z}} - \tilde{m} e^{A_s}) \psi + c.c. \right) = 0\,, \tag{4.36}$$

and

$$\delta S^{Bdy} = G_F \int d^5x\, \partial_{\hat{a}} \left(-\frac{i}{2} e^{4A_s} \delta\bar{\psi} \Gamma^{\hat{a}} \psi + c.c. \right) + \delta(\Delta S)\,. \tag{4.37}$$

Imposing periodic boundary condition in the x^μ coordinates and using the property that the spinor field solution decays fast enough at $z \to \infty$ vanishes, the on-shell variation reduces to

$$\delta S_F = G_F \int d^4x \left(\frac{i}{2} e^{4A_s} \delta\bar{\psi} \Gamma^{\hat{z}} \psi + \ c.c. \right)_{z=\epsilon} + \delta(\Delta S)\,. \tag{4.38}$$

Decomposing the Dirac spinor in their chiralities we obtain

$$\delta S_F = G_F \int d^4x \left(\frac{i}{2} e^{4A_s} \delta\bar{\psi}_L \psi_R - \frac{i}{2} e^{4A_s} \delta\bar{\psi}_R \psi_L \right)_{z=\epsilon} + c.c. + \delta(\Delta S)\,. \tag{4.39}$$

The left and right chiralities of the Dirac field are coupled, which means that it is impossible to fix them simultaneously. As a result, we need to select one of the chiralities. In order to fix the left component, we define the surface term ΔS as

$$\begin{aligned} \Delta S &= G_F \int d^4x \left(\sqrt{-\gamma} \frac{i}{2} \bar{\psi}\psi \right)_{z=\epsilon} \\ &= G_F \int d^4x \left(\frac{i}{2} e^{4A_s} (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) \right)_{z=\epsilon}\,. \end{aligned} \tag{4.40}$$

Varying this surface term we obtain

$$\delta(\Delta S) = G_F \int d^4x \left(\frac{i}{2} e^{4A_s} (\delta\bar{\psi}_L \psi_R + \delta\bar{\psi}_R \psi_L) \right)_{z=\epsilon} + c.c\,. \tag{4.41}$$

Plugging (4.41) into (4.39) we obtain the final result for the variation of the action

$$\begin{aligned} \delta S_F &= G_F \int d^4x \left(i\, e^{4A_s} \delta\bar{\psi}_L \psi_R \right)_{z=\epsilon} + c.c. \\ &= \int d^4x \left(\delta\bar{\psi}_L \Pi_R + \bar{\Pi}_R \delta\psi_L \right)_{z=\epsilon}\,, \end{aligned} \tag{4.42}$$

where we introduced the conjugate momenta

$$\Pi_R = i\, G_F e^{4A_s}\psi_R \quad , \quad \bar{\Pi}_R = i\, G_F e^{4A_s}\bar{\psi}_R\,. \tag{4.43}$$

Note from (4.42) that fixing the left spinor at the boundary is now consistent with the variational principle. Solving at small z (near the AdS boundary) the second order differential equations (4.25) for the left and right components we find

$$\begin{aligned}\psi_L(x,z) &= \alpha_L(x) z^{2-m} + \cdots + \beta_L(x) z^{3+m} + \dots\,,\\ \psi_R(x,z) &= \alpha_R(x) z^{3-m} + \cdots + \beta_R(x) z^{2+m} + \dots\,,\end{aligned} \tag{4.44}$$

where m is the constant mass which is the asymptotic value of the 5d mass coupling $\tilde{m}(z)$ in the limit $z \to 0$ (near the AdS boundary). The 4d spinors $\alpha_L(x)$ and $\alpha_R(x)$ are the source coefficients associated with the non-normalisable sector of the 5d spinors $\psi_L(x,z)$ and $\psi_R(x,z)$ respectively. The 4d spinors $\beta_L(x)$ and $\beta_R(x)$ are the VEV coefficients corresponding to the normalisable sector of the 5d spinors $\psi_L(x,z)$ and $\psi_R(x,z)$ respectively.

As described above, we take $\alpha_L(x)$ as the only independent 4d source. Note that it has conformal dimension $2+m$ since the 5d spinor has conformal dimension zero near the AdS boundary. This source couples to the operator O_R of conformal dimension $\Delta = 2+m$ so we can find the VEV of this operator using the holographic dictionary. From the action variation in (4.42) we obtain

$$\begin{aligned}\langle \mathcal{O}_R\rangle = \frac{\delta S_F}{\delta \bar{\alpha}_L} &= \left(z^{2-m}\Pi_R\right)_{z=\epsilon}\\ &= i\, G_F \left(z^{2-m} e^{4A_s}\psi_R\right)_{z=\epsilon}\end{aligned} \tag{4.45}$$

Using (4.24) we can write the result in (4.45) in terms of the left spinor:

$$\langle \mathcal{O}_R\rangle = i\, G_F \left(z^{2-m} e^{4A_s} \frac{\not{\partial}}{\partial^2}(\partial_z + 2A_s' + e^{A_s}\tilde{m})\psi_L \right)_{z=\epsilon}\,. \tag{4.46}$$

The VEV, according to the results (4.45) and (4.46), is the one-point function in the presence of the 4d source $\bar{\alpha}_L$. In the next subsection we will obtain the two-point function of the nucleon operator from the bulk to boundary propagator and will find a relation with the VEV.

4.4 The on-shell action, the bulk to boundary propagator and the two point function

The on-shell and the bulk to boundary propagator of the Dirac field allow us to obtain the VEV and the two-point function for nucleons. Our starting point is the Dirac action in (4.17) with the additional surface term given in (4.40). Evaluating this action on-shell we obtain

$$S_F^{o-s} = S_{Bulk}^{o-s} + S_{Bdy}^{o-s}\,, \tag{4.47}$$

where the bulk action is given by

$$S^{o-s}_{Bulk} = G_F \int d^5x\, e^{4A_s} \Big(\frac{i}{2}\bar{\psi}(-2A_s'\Gamma^{\hat{z}} + e^{A_s}\tilde{m})\psi - \frac{i}{2}\bar{\psi}(-2A_s'\Gamma^{\hat{z}} - e^{A_s}\tilde{m})\psi - ie^{A_s}\tilde{m}\bar{\psi}\psi \Big) = 0\,, \tag{4.48}$$

and the boundary action is

$$\begin{aligned} S^{o-s}_{Bdy} = \Delta S^{o-s} &= G_F \int d^4x \left(\frac{i}{2} e^{4A_s} \bar{\psi}_L \psi_R + c.c. \right)_{z=\epsilon} \\ &= G_F \int d^4x \left(\frac{i}{2} e^{4A_s} \bar{\psi}_L \frac{\not{\partial}}{\partial^2} (\partial_z + 2A_s' + e^{A_s}\tilde{m})\psi_L + c.c. \right)_{z=\epsilon} . \end{aligned} \tag{4.49}$$

Note that in (4.48) we used the equations (4.18)-(4.19) for the Dirac field and in (4.49) we used (4.24).

The bulk to boundary propagator written in coordinate space can be expressed by the following relation

$$\psi_L(z,x) = \int d^4y\, F_L(z,x;y)\, \alpha_L(y)\,, \tag{4.50}$$

where $\psi_L(z,x)$ is the left component of the Dirac field in 5d, $F_L(z,x;y)$ is a real scalar representing bulk to boundary propagator and $\alpha_L(y)$ is the 4d left spinorial source. Substituting (4.50) in (4.49) and (4.48), the on-shell action becomes

$$S^{o-s}_F = G_F \int d^4x \int d^4y \Big(\frac{i}{2} \bar{\alpha}_L(x) \left(z^{2-m} e^{4A_s} \frac{\not{\partial}_{x-y}}{\partial^2} (\partial_z + 2A_s' + e^{A_s}\tilde{m}) F_L(z,x;y) \right)_{z=\epsilon} \alpha_L(y) + c.c. \Big)\,. \tag{4.51}$$

where we also used the asymptotic behavior (4.44).

Note that the VEV in (4.46) can be written in terms of the bulk to boundary propagator as

$$\langle \mathcal{O}_R(x) \rangle = i\, G_F \int d^4y \left(z^{2-m} e^{4A_s} \frac{\not{\partial}_{x-y}}{\partial^2} (\partial_z + 2A_s' + e^{A_s}\tilde{m}) F_L(z,x;y) \right)_{z=\epsilon} \alpha_L(y)\,. \tag{4.52}$$

Varying the on-shell action we obtain

$$\delta S^{o-s}_F = \int d^4x \left(\delta\bar{\alpha}_L(x) \langle \mathcal{O}_R(x) \rangle + c.c. \right)\,, \tag{4.53}$$

as expected. Varying once more we obtain the two-point function

$$\begin{aligned} \Gamma_R(x-y) = \langle \mathcal{O}_R(x) \bar{\mathcal{O}}_R(y) \rangle &= \frac{\delta S^{o-s}_F}{\delta\bar{\alpha}_L(x)\delta\alpha_L(y)} = P_R \frac{\delta \langle \bar{\mathcal{O}}_R(y) \rangle}{\delta\bar{\alpha}_L(x)} \\ &= i G_F P_R \frac{\not{\partial}_{x-y}}{\partial^2} \left(z^{2-m} e^{4A_s} (\partial_z + 2A_s' + e^{A_s}\tilde{m}) F_L(z,x;y) \right)_{z=\epsilon}\,. \end{aligned} \tag{4.54}$$

The relation between the one-point and two-point functions is

$$\langle \mathcal{O}_R(x) \rangle = \int d^4y \, \Gamma_R(x-y) \, \alpha_L(y) \, . \tag{4.55}$$

The equation (4.50) can be written in momentum space as

$$\psi_L(z,q) = F_L(z,q) \, \alpha_L(q) \, . \tag{4.56}$$

The VEV (4.52) in momentum space takes the form

$$\langle \mathcal{O}_R(q) \rangle = \Gamma_R(q) \, \alpha_L(q) \, , \tag{4.57}$$

where

$$\Gamma_R(q) = -G_F P_R \frac{\not{q}}{Q^2} \left(z^{2-m} e^{4A_s} (\partial_z + 2A_s' + e^{A_s} \tilde{m}) F_L(z,q) \right)_{z=\epsilon} \, . \tag{4.58}$$

We end this subsection fixing the coupling constant G_F that characterises the 5d Dirac action. In order to do that we evaluate the correlator (4.58) in the limit $q^2 \to \infty$ (UV). In this limit the 4d theory becomes conformal and the bulk to boundary propagator can be approximated by the (analytical) solution corresponding to 5d AdS space. For m half-integer we find that

$$\Gamma_R(q) = G_F a_m P_R \not{q} \, q^{2m-1} \ln q^2 \, , \tag{4.59}$$

where

$$a_m = \frac{(-1)^{m-1/2}}{2^{2m} \Gamma(m+1/2)^2} \, . \tag{4.60}$$

For $m = 5/2$ we have $\Delta = 9/2$ which is the canonical dimension of the nucleon operator. In this case we can compare against the perturbative QCD result [60]

$$\Gamma_R(q) = \frac{1}{64\pi^4} P_R \not{q} \, q^4 \ln q^2 \quad \text{(perturbative QCD)} \, , \tag{4.61}$$

and obtain

$$G_F = \frac{2}{\pi^4} \, . \tag{4.62}$$

In the following subsections we will obtain a spectral decomposition for the bulk to boundary propagator using Sturm-Liouville theory. From this result we will finally obtain the spectral decomposition of the nucleon correlator. The spectrum of nucleons then will be obtained from the eigenalue problem and the nucleon decay constants will be extracted from the coefficients of the spectral decomposition.

4.5 Spectral decomposition for the bulk to boundary propagator

In this subsection, we will use Sturm-Liouville theory to find a spectral decomposition for the bulk to boundary propagator. We will proceed in a similar way as in the case of vector mesons, described in subsection 3.5.

We start writing the equation in (4.27) for the left bulk to boundary propagator in the following form:

$$\left[\left(\partial_z + 4A_s' \right) \partial_z + \theta_L + Q^2 \right] F_L(q,z) = 0 \tag{4.63}$$

where

$$\theta_L = 2A_s'' + 4A_s'^2 + \partial_z(e^{A_s}\tilde{m}) - e^{2A_s}\tilde{m}^2\,. \tag{4.64}$$

Rewriting (4.63) as

$$\left[\partial_z\left(e^{4A_s}\partial_z\right) + e^{4A_s}\theta_L + Q^2 e^{4A_s}\right] F_L(q,z) = 0, \tag{4.65}$$

we identify this equation with the Sturm-Liouville equation

$$\Big[\mathcal{L}_L + \lambda\, r(z)\Big] y(z) = f(z) \quad , \quad \mathcal{L}_L = \partial_z\Big(p(z)\partial_z\Big) - s_L(z)\,, \tag{4.66}$$

where

$$p(z) = r(z) = e^{4A_s}, \quad s_L(z) = -e^{4A_s}\theta_L, \quad \lambda = Q^2 \quad \text{and } f(z) = 0\,\text{(homogeneous)}. \tag{4.67}$$

The Sturm-Liouville theory is briefly described in appendix A. The Green's function $G_L(z;z')$ corresponding to the non-homogeneous case satisfies the equation

$$\left[\mathcal{L}_L + \lambda\, r(z)\right] G_L(z;z') = \delta(z-z') \tag{4.68}$$

Again, it is convenient to define an infinite set of eigenfunctions $f_L^n(z)$ by the eigenvalue equation

$$\left[\mathcal{L}_L + \lambda_n r(z)\right] f_L^n(z) = 0\,, \tag{4.69}$$

or

$$\left[\partial_z\left(e^{4A_s}\partial_z\right) + e^{4A_s}\theta_L + \lambda_n e^{4A_s}\right] f_L^n(z) = 0\,, \tag{4.70}$$

where $\lambda_n = m_{N^n}^2$ are the eigenvalues. Comparing this equation with (4.32) we see that the Sturm-Liouville modes are the normalisable modes in holographic QCD. These modes satisfy the orthonormality condition

$$\int dz\, e^{4A_s} f_L^m(z) f_L^n(z) = \delta^{mn}\,, \tag{4.71}$$

and the Green's function admits the following spectral decomposition

$$G_L(z;z') = \sum_n \frac{f_L^n(z) f_L^n(z')}{q^2 + m_{N^n}^2}\,. \tag{4.72}$$

For more details see appendix A.

Multiplying both sides of (4.68) by $F_L(q,z)$, integrating over z and using (4.65) we obtain the relation between the bulk to boundary propagator and the Green's function:

$$F_L(q,z') = \Big[e^{4A_s(z)}\Big(F_L(q,z)\partial_z G_L(z;z') - G_L(z;z')\partial_z F_L(q,z)\Big)\Big]_{z=\epsilon}^{\infty}\,. \tag{4.73}$$

Assuming that $F_L(q,z)$ and $G_L(z;z')$ vanish sufficiently fast in the limit $z\to\infty$ and using the spectral decomposition (4.72) we obtain

$$F_L(q,z')=\sum_n \frac{f_{L,n}(z')}{q^2+m_{N^n}^2}\Big[e^{4A_s(z)}\Big(F_L(q,z)f'_{L,n}(z)-f_{L,n}(z)F'_L(q,z)\Big)\Big]_{z=\epsilon} \tag{4.74}$$

From (4.26) and (4.44) we see that $F_L(q,z)$ behaves as z^{2-m} at small z. Using also the asymptotic behaviour for the warp factor the equation (4.74) reduces to

$$F_L(q,z')=\sum_n \frac{f_n m_{N^n} f_{L,n}(z')}{q^2+m_{N^n}^2}\,, \tag{4.75}$$

where we also used the coupled equations (4.31) and the coefficients f_n are defined as

$$f_n=\Big[z^{-2-m}f_{R,n}(z)\Big]_{z=\epsilon}\,. \tag{4.76}$$

In the next subsection we will relate these coefficients correspond to the nucleon decay constants. Lastly, it is easy to show that the Sturm-Liouville modes satisfy the completeness relation

$$\sum_n e^{4A_s} f_{L,n}(z)f_{L,n}(z')=\delta(z-z')\,. \tag{4.77}$$

For more details see appendix A. In the following subsection we will obtain the spectral decomposition forthe 4d nucleon correlator and show the compatibility with the spectral decomposition expected in large N_c QCD.

4.6 The 4d nucleon correlator

In subsection 4.4 we obtained the holographic dictionary (4.58) that relates the 4d nucleon correlator $\Gamma_R(q)$ to the 5d bulk to boundary propagator $F_L(z,q)$. In subsection 4.5 we obtained the spectral decomposition (4.75) for the bulk to boundary propagator. Then using (4.58) and (4.75) we finally obtain the spectral decomposition for the nucleon correlator:

$$\Gamma_R(q)=-P_R\frac{\not{q}}{Q^2}\sum_n \frac{\lambda_{N^n}^2 m_{N^n}^2}{q^2+m_{N^n}^2}\,, \tag{4.78}$$

where

$$\lambda_{N^n}=\sqrt{G_F}f_n=\sqrt{G_F}\Big[z^{-2-m}f_{R,n}(z)\Big]_{z=\epsilon}\,, \tag{4.79}$$

and we also used the coupled equations (4.31). It is interesting the correlator in (4.78) as

$$\Gamma_R(q)=-P_R\not{q}\left(\frac{1}{Q^2}\sum_n \lambda_{N^n}^2+\sum_n\frac{\lambda_{N^n}^2}{q^2+m_{N^n}^2}\right)\,. \tag{4.80}$$

The first term in (4.80) diverges. This UV divergence is expected since we have worked with the original on-shell action without introducing holographic renormalisation. Subtracting this UV divergence we obtain the renormalised correlator

$$\Gamma_R^{\rm ren}(q)=-P_R\not{q}\sum_n\frac{\lambda_{N^n}^2}{q^2+m_{N^n}^2}\,. \tag{4.81}$$

This final result (4.81) for the nucleon correlator is valid for a general class of holographic QCD models based on Einstein-dilaton gravity and it is consistent with the spectral decomposition (4.4) obtained in large N_c QCD. The coefficients λ_{N^n} defined in (4.79) are therefore identified with the nucleon decay constants.

4.7 Spectrum of nucleons

In this subsection we obtain the spectrum of nucleons solving the eigenvalue equation (4.32) for the normalisable modes. Before doing that it is interesting to rewrite (4.32) as Schrödinger equations and investigate the corresponding Schrödinger potentials.

Using the Bogoliubov transformation

$$f^n_{R/L}(z) = e^{-2A_s(z)}\xi^n_{R/L}(z)\,, \tag{4.82}$$

in (4.32) we obtain

$$\left[-\partial_z^2 + V_{R/L}\right]\xi^n_{R/L} = m^2_{N^n}\xi^n_{R/L} \tag{4.83}$$

where the Schrödinger potential $V_{R/L}$ are given by

$$V_{R/L} = \pm\partial_z\left(e^{A_s}\tilde{m}\right) + (e^{A_s}\tilde{m})^2\,. \tag{4.84}$$

Motivated by the Schrödinger potential potential behavior of vector mesons (3.55), which is a combination of the warp factor and dilation derivatives, we postulate the following mass coupling for our model

$$\tilde{m} = e^{-A_s}\left(\frac{1}{2}\Phi' - mA_s'\right)\,. \tag{4.85}$$

The coefficients were fixed in order to recover on the one hand the 5d constant mass m in the AdS limit and on the other hand to guarantee a quadratic behaviour for the Schrödinger potential at large z compatible with the soft wall model. The latter is a necessary requirement for obtaining asymptotically linear Regge trajectories for the nucleons, i.e. $m_n^2 \sim n$ at large n [6].

Figure 7 shows the results for the Schrödinger potentials V_R (left panel) and V_L (right panel) for the nucleons in the case $m = 3/2$. The blue and red lines represent the results for the Einstein-dilaton models I and II respectively whereas the black dashed lines represent the results for the soft wall model. Figure 8 shows the results for the Schrödinger potentials in the case $m = 5/2$. Note that the Schrödinger potentials for the soft wall model present a minumum at a higher value compared with the minima for the Einstein-dilaton models I and II. Note that this effect is enhanced as we go from the case $m = 3/2$ to the case $m = 5/2$.

Asymptotic solution and numerical integration

To find the spectrum of nucleons we need to solve the eigenvalue equations (4.32) or equivalently the Schrödinger equations (4.83). As expected, the eigenvalues for the left and right sector are the same since these two sectors are coupled.

[6] This can be easily checked using a WKB approximation.

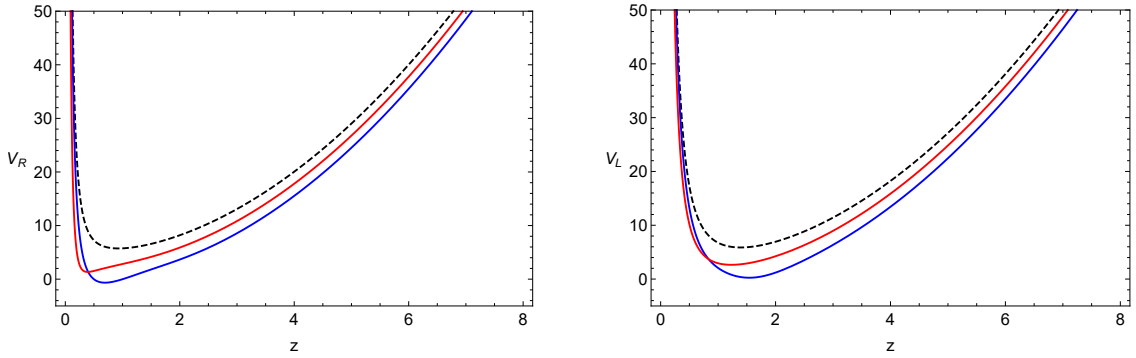


Figure 7. Schrödinger potentials V_R (left panel) and V_L (right panel) for Einstein-dilaton models I and II (blue and red lines) and the soft wall model (black dashed lines) in the case $m = 3/2$.

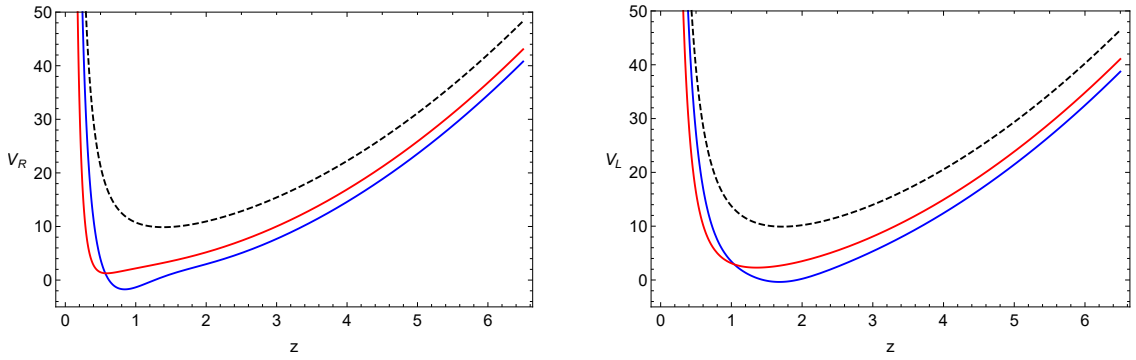


Figure 8. Schrödinger potentials V_R (left panel) and V_L (right panel) for Einstein-dilaton models I and II (blue and red lines) and the soft wall model (black dashed lines) in the case $m = 5/2$.

The eigenvalues and eigenfunctions are found numerically. For the numerical integration, we use for the initial conditions the asymptotic solution at small z

$$f_R^n = N_R^n z^{2+m} \quad , \quad f_L^n = N_L^n z^{3+m} \, . \tag{4.86}$$

The normalisation constants N_R^n and N_L^n can be obtained imposing the condition that the eigenfunctions ξ_R^n and ξ_L^n, defined in (4.82), are normalised to 1. The numerical integration is carried from small z to large z where we impose the asymptotic behaviour

$$\lim_{z\to\infty} \sqrt{z}\, \xi_{R/L}^n = 0 \, . \tag{4.87}$$

Using this shooting method we find the set of eigenvalues m_{N^n} corresponding to the 4d nucleon masses.

Spectrum

In tables 3 and 4 we present our results for the nucleon masses for the Einstein-dilaton models I and II and compare them against the soft wall model, the hard wall model as well as the experimental results. Table 3 displays the results when the conformal dimension is fixed as $\Delta = 7/2$ whereas table 4 corresponds to the case $\Delta = 9/2$. The former takes into account the possible contribution from the anomalous dimension whilst the latter sets

the anomalous dimension to zero. The results for the soft wall model and the hard wall model were obtained following [38] and [36] respectively. We briefly review those works in appendices C and D.

From our analysis we conclude that the Einstein-dilaton model I provide the results that are closest to the experimental data in both cases $\Delta = 7/2$ and $\Delta = 9/2$. This can also be seen in figure 9 where we plot the squared masses of the first six nucleon states as a function of the radial excitation number for the Einstein-dilaton models I and II (blue and red solid lines with dots) and the soft wall model (black dashed line with dots). As expected, the Regge trajectories in the Einstein-dilaton models I and II are approximately linear whilst the Regge trajectory in the soft wall model is exactly linear.

Ratio	Model I	Model II	Soft wall	Hard wall	Experimental [57]
m_{N^0}/m_{ρ^0}	0.987	0.988	1.414	1.593	1.209 ± 0.002
m_{N^1}/m_{ρ^0}	1.623	1.339	1.732	2.917	1.856 ± 0.039
m_{N^2}/m_{ρ^0}	2.053	1.613	2	4.23	2.204 ± 0.039
m_{N^3}/m_{ρ^0}	2.403	1.847	2.236	5.54	2.423 ± 0.065
m_{N^4}/m_{ρ^0}	2.707	2.054	2.449	6.849	2.706 ± 0.065

Table 3. Nucleon masses divided by the mass of the ρ_0 meson in the case $\Delta = 7/2$ ($m = 3/2$) in the Einstein-dilaton models, the soft wall model and the hard wall model compared against the experimental results from PDG [57].

Ratio	Model I	Model II	Soft wall	Hard wall	Experimental [57]
m_{N^0}/m_{ρ^0}	0.896	0.952	1.732	2.136	1.209 ± 0.002
m_{N^1}/m_{ρ^0}	1.593	1.314	2	3.5	1.856 ± 0.039
m_{N^2}/m_{ρ^0}	2.04	1.595	2.236	4.832	2.204 ± 0.039
m_{N^3}/m_{ρ^0}	2.399	1.833	2.449	6.153	2.423 ± 0.065
m_{N^4}/m_{ρ^0}	2.708	2.043	2.646	7.468	2.706 ± 0.065

Table 4. Nucleon masses divided by the mass of the ρ_0 meson in the case $\Delta = 9/2$ ($m = 5/2$) in the Einstein-dilaton models, the soft wall model and the hard wall model compared against the experimental results from PDG [57].

4.8 Wave functions and nucleon decay constants

Figures 10 and 11 display the normalised eigenfunctions $f_R^n(z)$ and $f_L^n(z)$, representing the nucleon states, for the cases $\Delta = 7/2$ ($m = 3/2$) and $\Delta = 9/2$ ($m = 5/2$) respectively. The blue and red solid lines correspond to Einstein-dilaton models I and II respectively whilst the black dashed line represents the soft wall model. These results confirm that the first nucleon masses obtained in the previous subsection correspond to the fundamental state and the first excited states.

Previously in this section we obtained the holographic dictionary for the nucleon decay constants (4.79). We can finally evaluate this formula using the normalised eigenfunctions

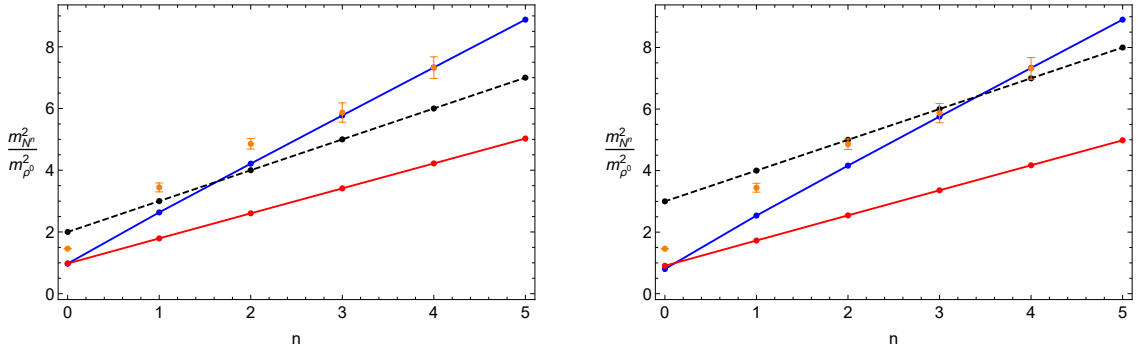


Figure 9. **Left:** Dimensionless squared mass ratios $m^2_{N^n}/m^2_{\rho^0}$ in the case $\Delta = 7/2$ ($m = 3/2$) for nucleons in the Einstein-dilaton models I and II (blue and red solid lines with dots) and the soft wall model (black dashed line with dots), compared against experimental data (orange dots and error bars). **Right:** Same as Left but this time $\Delta = 9/2$ ($m = 5/2$).

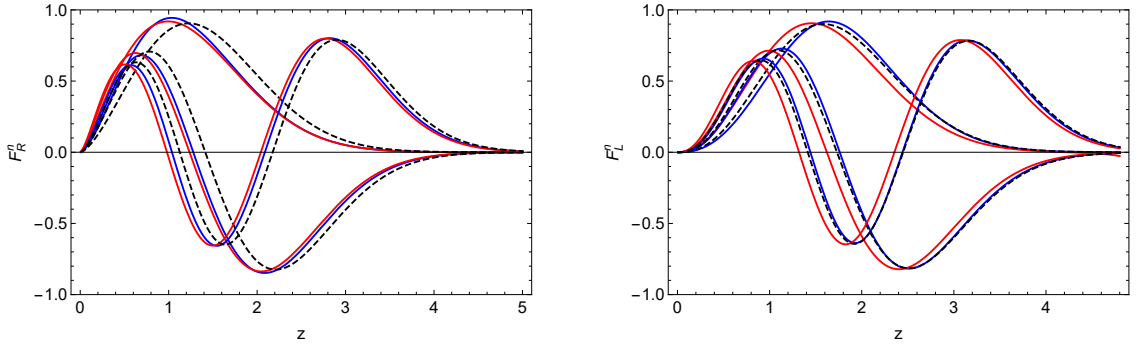


Figure 10. Normalised wave functions $F^n_R(z)$ (left panel) and $F^n_L(z)$ (right panel) for the Einstein-dilaton models I (blue), II (red) and the soft wall model (black dashed) for the case $\Delta = 7/2$ ($m = 3/2$).

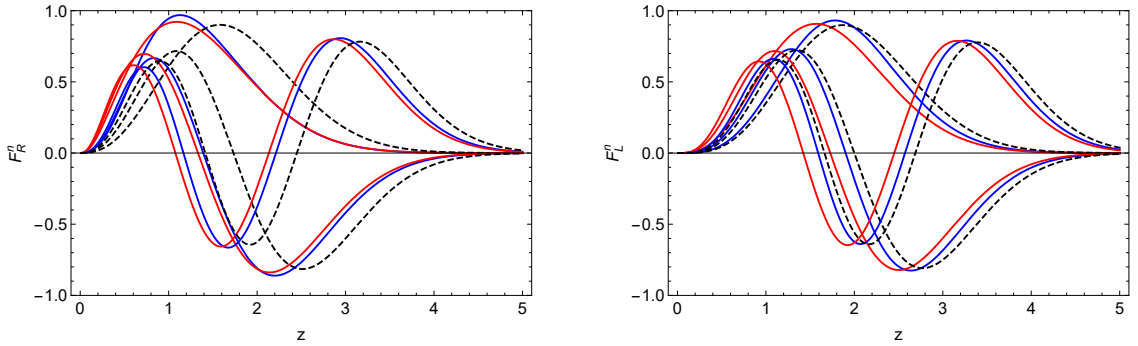


Figure 11. Normalised wave functions $F^n_R(z)$ (left panel) and $F^n_L(z)$ (right panel) for the Einstein-dilaton models I (blue), II (red) and the soft wall model (black dashed) for the case $\Delta = 9/2$ ($m = 5/2$).

and obtain

$$\lambda_{N^n} = \sqrt{G_F}\Big[z^{-2-m} f_{R,n}(z)\Big]_{z=\epsilon} = \sqrt{G_F} N^n_R\,, \tag{4.88}$$

where N_R^n is the normalisation constant in the right sector. The coupling constant G_F in the fermionic sector was already fixed in (4.62) in order to reproduce the perturbative QCD result for the correlation function.

We display in tables 5 and 6 our results for the nucleon decay constants in the cases $\Delta = 7/2$ ($m = 3/2$) and $\Delta = 9/2$ ($m = 5/2$) respectively. In the latter case we also present the result for the fundamental state in lattice QCD obtained in [62] using a nucleon operator similar the one presented in (4.1). Note that the Einstein-dilaton model I and the soft wall model provide results that are closer to lattice QCD. It is interesting to note that holographic QCD models provide results for the excited nucleon states which, as far as we are concerned, are not available in other non-perturbative approaches. In particular, all the holographic QCD models predic that the nucleon decay constants grow with the radial excitation number.

Ratio	Model I	Model II	Soft wall	Hard wall
$\lambda_{N^0}/m_{\rho^0}^{\alpha}$	0.1108	0.09835	0.0507	0.1667
$\lambda_{N^1}/m_{\rho^0}^{\alpha}$	0.1302	0.1158	0.0716	0.4096
$\lambda_{N^2}/m_{\rho^0}^{\alpha}$	0.1519	0.1284	0.0877	0.7138
$\lambda_{N^3}/m_{\rho^0}^{\alpha}$	0.1708	0.1388	0.1013	1.069
$\lambda_{N^4}/m_{\rho^0}^{\alpha}$	0.1878	0.1478	0.1133	1.469

Table 5. Nucleon decay constants λ_{N^n} divided by $m_{\rho_0}^{\alpha}$ with $\alpha = \Delta - 3/2 = m + 1/2$ in the case $\Delta = 7/2$ ($m = 3/2$) in the Einstein-dilaton models, the soft wall model and the hard wall model.

Ratio	Model I	Model II	Soft wall	Hard wall	Lattice QCD [62]
$\lambda_{N^0}/m_{\rho^0}^{\alpha}$	0.1055	0.158	0.01791	0.1414	0.05778 ± 0.0107
$\lambda_{N^1}/m_{\rho^0}^{\alpha}$	0.1201	0.1906	0.03102	0.4755	-
$\lambda_{N^2}/m_{\rho^0}^{\alpha}$	0.1462	0.2172	0.04387	1.058	-
$\lambda_{N^3}/m_{\rho^0}^{\alpha}$	0.172	0.2409	0.05664	1.931	-
$\lambda_{N^4}/m_{\rho^0}^{\alpha}$	0.1973	0.2627	0.06937	3.129	-

Table 6. Nucleon decay constants λ_{N^n} divided by $m_{\rho_0}^{\alpha}$ with $\alpha = \Delta - 3/2 = m + 1/2$ in the case $\Delta = 9/2$ ($m = 5/2$) in the Einstein-dilaton models, the soft wall model and the hard wall model. In this case we also compare against the lattice QCD result.

5 Conclusions

In this paper, we built a bottom-up holographic QCD model that provides a unified description of vector mesons and nucleons in a confining background based on Einstein-dilaton gravity. Our model has three parameters: g_5^2 associated with the two-point correlation function in the vector meson sector, G_F associated with the two-point correlation function in the nucleon sector and k the infrared mass scale associated with the spectrum. We fixed g_5^2 and G_F using the perturbative QCD results for the hadronic correlators in the high

energy regime. Since we worked with dimensionless ratios for the hadronic masses and decay constants we did not need to fix the constant k.

To investigate the spectral decomposition of the hadronic correlators and the associated decay constants, we applied the Sturm-Liouville theory inspired by previous works [15, 26, 27]. The Sturm-Liouville theory allowed us to find spectral decompositions for the 5d bulk to boundary propagators and the 4d correlation functions. We showed that the latter are compatible with QCD in the large N_c limit. We obtained a holographic dictionary for the hadronic decay constants that is valid for a general class of holographic models based on Einstein-dilaton gravity. The Sturm-Liouville also led naturally to the completeness relation for the normalisable modes, identified as the Sturm-Liouville modes.

In table 1 we presented our results for the spectrum of vector mesons in terms of the ratio between the masses of excited states, i.e. m_{ρ^n} with $n \geq 1$, and the mass of the ground state m_{ρ^0} for the Einstein-dilaton models I and II, discussed in section 2. We compared our results with the soft wall model, the hard wall model and experimental data. We concluded that the Einstein-dilaton model I and the soft wall model provide results that are the closest to the experimental results. We presented our results for the nucleon spectrum in Eintein-dilaton models I and II for the cases $\Delta = 7/2$ and $\Delta = 9/2$ in tables 3 and 4 respectively. Our results for the nucleon spectrum were presented in terms of the ratios between the masses of the nucleon states m_{N^n} with $n = 0, 1, \ldots$ (ground state and excited states) relative to the mass of the vector meson ground state m_{ρ^0}. We compared our results with the soft wall model, the hard wall model and experimental data and concluded that the Einstein-dilaton model I provide the best results compared to experimental data.

In table 2 we presented our results for the vector meson decay constants F_{ρ^n} in terms of the dimensionless ratios $\sqrt{F_{\rho^n}}/m_{\rho^0}$ for $n = 0, 1, \ldots$ (ground state and excited states). We compared our results for the Einstein-dilaton models I and II with the soft wall and hard wall model. For the ground state case ($n = 0$) we also compared against the only available experimental data and concluded that the hard wall model still provides the closest result. We presented our results for the nucleon decay constants λ_{N^n} for the cases $\Delta = 7/2$ and $\Delta = 9/2$ in tables 5 and 6 respectively. The results were presented in terms of the dimensionless ratios $\lambda_{N^n}/m_\rho^\alpha$ with $\alpha = \Delta - 3/2$. For the case $\Delta = 9/2$ (canonical dimension) and $n = 0$ (nucleon ground stated) we also compared against the lattice QCD result and concluded that the Einstein-dilaton model I and the soft wall model provide results that are closer to lattice QCD. It is important to remark that holographic QCD models are capable of predicting the decay constants of excited states. In particular, we noted that for both light vector mesons and nucleons the decay constants increase with the radial excitation number. We hope that in the future, as more experimental results on hadronic decay constants become available, these findings can be further tested.

A natural continuation of this work would be to investigate the spectrum and decay constants of the Delta baryons that have spin and isospin 3/2 in the context of holographic QCD models based on Einstein-dilaton gravity. Some works have already been developed using the hard wall model and the soft wall model [35, 63, 64]. We also want to apply the Sturm-Liouville theory in that case to obtain the spectral decomposition for the correlators of Delta baryon operators. We are also interested in studying the strong couplings between

vector mesons and baryons, the electromagnetic and the gravitational form factors. Last but not least, we want to investigate the effects of chiral symmetry breaking on the mass generation of nucleons and vector mesons. We intend to develop these works in the near future.

Acknowledgments

The work of the author A.B-B is partially funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), grant No. 314000/2021-6, and Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES, Brazil), Finance Code 001. The work of the author A.S.S. Jr has financial support from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq).

A Sturm-Liouville theory and the spectral decomposition

The Sturm-Liouville theory can be described, for example, by a non-homogeneous one-dimensional second-order differential equation [65, 66]

$$\frac{d}{dz}\left(p(z)\frac{dy}{dz}\right) - s(z)y + \lambda r(z)y = f(z) \tag{A.1}$$

where $y(z)$, $p(z)$, $s(z)$, $r(z)$, and $f(z)$ are functions of z and λ is a constant parameter. In the homogeneous case we takes $f(z) = 0$ in (A.1) equation. From the two first terms in (A.1), we can define the Sturm-Liouville operator,

$$\mathcal{L} \equiv \frac{d}{dz}\left(p(z)\frac{d}{dz}\right) - s(z). \tag{A.2}$$

This is a second-order self-adjoint operator with eigenvalue λ. Rewriting the equation (A.1) in terms of $\mathcal{L}$, we have

$$[\mathcal{L} + \lambda\, r(z)]\, y = f(z). \tag{A.3}$$

We will be particularly interested in the solution of the homogeneous case $f(z) = 0$, we will call this solution $y_0(z)$. We can obtain the Green's functions that satisfy (A.3) starting from

$$[\mathcal{L} + \lambda\, r(z)]\, G(z; z') = \delta(z - z') \tag{A.4}$$

where $G(z; z')$ is the Green function that must obey some boundary condition. We can expand the Green's functions into a series of eigenfunctions expressed as

$$G(z; z') = \sum_n a_n(z')\varphi_n(z) \tag{A.5}$$

Plugging (A.5) into (A.4) and imposing the orthonormality condition

$$\int dz\, r(z)\, \varphi_m(z)\varphi_n(z) = \delta_{mn} \tag{A.6}$$

we have

$$a_n(z') = \frac{\varphi_n(z')}{\lambda - \lambda_n} \tag{A.7}$$

Substituting (A.7) in (A.5), we obtain the spectral decomposition for the Green's function

$$G(z; z') = \sum_n \frac{\varphi_n(z)\varphi_n(z')}{\lambda - \lambda_n}. \tag{A.8}$$

The eigenfunctions obey the equation

$$\left[\mathcal{L} + \lambda_n r(z)\right] \varphi_n(z) = 0. \tag{A.9}$$

We can relate the homogeneous solution $y_0(z)$ to the Green's function $G(z; z')$ as follows. Multiplying both sides of (A.4) by $y_0(z)$, integrating by parts twice over z and using the homogeneous equation for $y_0(z)$ we obtain

$$y_0(z') = \left[r(z)\big(y_0(z)\partial_z G(z; z') - G(z; z')\partial_z y_0(z)\big)\right]_{z=z_i}^{z=z_f} \tag{A.10}$$

Note that this result has the form of a Wronskian in the z variable for the functions $y_0(z)$ and $G(z, z')$. The limits of integration z_i and z_f depend on the boundary conditions of the problem.

Plugging the spectral decomposition (A.8) into (A.10) we find the expansion

$$y_0(z') = \sum_n \frac{\varphi_n(z')}{\lambda - \lambda_n}\left[r(z)\big(y_0(z)\partial_z\varphi_n(z) - \varphi_n(z)\partial_z y_0(z)\big)\right]_{z=z_i}^{z=z_f} \equiv \sum_n \alpha_n \varphi_n(z')\,. \tag{A.11}$$

Using the orthonormality condition, we have

$$\int dz\, r(z)\varphi_m(z)y_0(z) = \alpha_m\,. \tag{A.12}$$

From (A.11) and (A.12) we obtain

$$y_0(z') = \sum_n \left[\int dz\, r(z)\varphi_n(z)y_0(z)\varphi_n(z')\right] \tag{A.13}$$

$$= \int dz\, y_0(z) \left[\sum_n r(z)\varphi_n(z)\varphi_n(z')\right] \tag{A.14}$$

$$= \int dz\, \delta(z - z')y_0(z) \tag{A.15}$$

In this way, we find the completeness relation

$$\sum_n r(z)\varphi_n(z)\varphi_n(z') = \delta(z - z'). \tag{A.16}$$

B The Proca Field Propagator

In this appendix, we briefly discuss the Proca field with an additional term that acts as a Lagrange multiplier, see for example [67, 68].

Consider the Proca Lagrangian for a massive spin 1 particle in 4 dimensions

$$\mathcal{L} = -\frac{1}{4}(F_{\mu\nu})^2 + \frac{1}{2}m^2 A_\mu^2 - \left(1 - \frac{1}{\chi}\right)(\partial_\mu A^\mu)^2 \tag{B.1}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{B.2}$$

being $F_{\mu\nu}$ the field strength usual, A^μ is the gauge field, m is the mass and χ the Lagrange multiplier. The equations of motions obtained from (B.1) written in momentum space takes the form

$$\left[\eta^{\mu\nu}\left(k^2 + m^2\right) - \left(1 - \frac{1}{\chi}\right)k^\mu k^\nu\right] A_\mu = 0 \tag{B.3}$$

To obtain the two-point correlation function, let's write the above operator in (B.3) as

$$\eta^{\mu\nu}(k^2 + m^2) - k^\mu k^\nu + \frac{1}{\chi}k^\mu k^\nu = \left(\eta^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right)\left(k^2 + m^2\right) + \left(\frac{k^\mu k^\nu}{k^2}\right)\frac{1}{\chi}\left(k^2 + \chi m^2\right) \tag{B.4}$$

The Proca propagator is obtained by inverting the terms that multiply the projectors in (B.4)

$$\langle A^\mu(k)A^\nu(-k)\rangle = \frac{-i}{k^2 + m^2}\left(\eta^{\mu\nu} - \frac{k^{\mu\nu}}{k^2}\right) + \frac{-i\chi}{k^2 + \chi m^2}\left(\frac{k^\mu k^\nu}{k^2}\right) \tag{B.5}$$

$$= \frac{-i}{k^2 + m^2}\left[\eta^{\mu\nu} - \frac{k^\mu k^\nu}{k^2 + \chi m^2}(1 + \chi)\right] \tag{B.6}$$

From the result (B.6) above, we can take some limits. For $\chi = 0$, we have

$$\langle A^\mu(k)A^\nu(-k)\rangle\Big|_{\chi=0} = \frac{-i}{k^2 + m^2}\left[\eta^{\mu\nu} - \frac{k^\mu k^\nu}{k^2}\right] \tag{B.7}$$

For $\chi \to \infty$,

$$\langle A^\mu(k)A^\nu(-k)\rangle\Big|_{\chi=0} = \frac{-i}{k^2 + m^2}\left[\eta^{\mu\nu} - \frac{k^\mu k^\nu}{m^2}\right] \tag{B.8}$$

Lastly, for $\chi = -1$

$$\langle A^\mu(k)A^\nu(-k)\rangle\Big|_{\chi=0} = \frac{-i\eta^{\mu\nu}}{k^2 + m^2}\,. \tag{B.9}$$

C Vector mesons and nucleons in the soft wall model

The soft wall model, initially proposed in the seminal paper by Karch, Katz, Son, and Stephanov in [17], has been demonstrated to effectively capture the Regge trajectories of various particles, including vector mesons and nucleons. In order to break the conformal symmetry, the soft wall model incorporates a dilaton field $\Phi(z)$ into its action. In this appendix we provide a succinct overview of the field equations, normalisable solutions, spectrum and decay constants of vector mesons and nucleons within the framework of the soft wall model.

C.1 Vector mesons

In section 3 we study the vector mesons using the Einstein-dilaton model. We can get the vector meson results for the soft wall model from the equations of the Einstein-dilaton model by considering the AdS limit. We start with the equation (3.17)

$$\Big[\partial_z + A_s' - \Phi'\Big]\partial_z V_\perp^{\hat{\mu},a} - q^2 V_\perp^{\hat{\mu},a} = 0\,. \tag{C.1}$$

In the soft wall, the warp factor is $A_s = -\ln z$ and the dilaton remains quadratic, $\Phi = kz^2$. For simplicity, let's consider $k = 1$ in our equations. It is convenient to write the vector field as $V_\perp^{\hat{\mu},a} = \eta_\perp^{\hat{\mu}} v(q^2, z)$ where $\eta_\perp^{\hat{\mu}}$ is a transverse polarisation vector, i.e. $q_{\hat{\mu}}\eta_\perp^{\hat{\mu}} = 0$. Writing the vector field this way, the above equation reduces to

$$\left[z^2\partial_z^2 - (1+2z^2)z\partial_z\right] v - q^2 z^2 v = 0 \tag{C.2}$$

This differential equation has an analytical solution that can be written as

$$v(Q,z) = c_1\, z^2 U\left(1 - \frac{Q^2}{4}, 2, z^2\right) + c_2\, z^2 M\left(1 - \frac{Q^2}{4}, 2, z^2\right) \tag{C.3}$$

where $Q = \sqrt{-q^2}$, c_1 and c_2 are constant coefficients whereas $U(a,b,x)$ and $M(a,b,x)$ are the Tricomi and Kummer confluent hypergeometric functions. To guarantee the regularity of the solution far from the boundary, we must take $c_2 = 0$. In this way, the solution reduces to the Tricomi function

$$v(Q,z) = c_1\, z^2 U\left(1 - \frac{Q^2}{4}, 2, z^2\right)\,. \tag{C.4}$$

In order to obtain the normalisable solution the first argument of the Tricomi function in (C.4) has to be $-n$ with n a non-negative integer. This leads to the spectrum of the vector mesons

$$m_{v_n}^2 = 4(n+1) \quad \text{with} \quad n = 0, 1, 2, 3, ... \tag{C.5}$$

The normalisable solutions take the form

$$v_n(z) = N_{v^n} z^2 L_n^1(z^2) \quad \text{with} \quad n = 0, 1, 2, 3, ... \tag{C.6}$$

where $L_n^k(x)$ are the associated Laguerre polynomials and

$$N_{v^n} = \sqrt{\frac{2}{n+1}} \tag{C.7}$$

are the normalisation constants that can be obtained using the condition

$$\int dz\, e^{A_s - \Phi} v_n^2 = 1\,. \tag{C.8}$$

For small z (near the boundary) the normalisable solution takes the form

$$v_n(z) = c_{2,n}\, z^2 + c_{4,n}\, z^4 + \ldots\,, \tag{C.9}$$

where

$$c_{2,n} = (n+1)N_{v^n} \quad \text{and} \quad c_{4,n} = -\frac{1}{2}n(n+1)N_{v^n} = -\frac{n}{2}c_{2,n}\,. \tag{C.10}$$

The decay constants can be obtained from (3.51) considering $A_s = -\ln z$, reproducing the results of [27, 55]

$$f_{v^n} = \frac{1}{g_5}\left[e^{A_s - \Phi}\partial_z v_n(z)\right]_{z\to 0} = \frac{2c_{2,n}}{g_5} = \frac{2(n+1)N_{v^n}}{g_5} = \frac{1}{g_5}\sqrt{8(n+1)}\,. \tag{C.11}$$

C.2 Nucleons

In this appendix, we obtain the results for the soft wall model as a particular case of the results of section 4 for the Einstein-dilaton models. The starting point is the equation (4.27) given by

$$\left[\partial_z^2 + 4A_s'\partial_z + 2A_s'' + 4A_s'^2 \mp \partial_z\left(e^{A_s}\tilde{m}\right) - e^{2A_s}\tilde{m}^2 + Q^2\right]F_{R/L} = 0\,. \tag{C.12}$$

Considering the warp factor as $A_s = -\ln z$ and $\tilde{m} = m + \Phi$ with $\Phi = kz^2$, we obtain the equation for the soft wall model

$$\left[(z\partial_z)^2 - 5z\partial_z + 6 + z^2(Q^2 \mp k) \pm -(m + kz^2)^2\right] F_{R/L} = 0 \tag{C.13}$$

Again, for simplicity we will take $k = 1$. The general solution of this equation that is regular at large z (far from the boundary) takes the form

$$\begin{aligned} F_R(z) &= e^{-z^2/2}(z^2)^{1+\frac{m}{2}}\left[d_1 U(m + \frac{1}{2} - \frac{Q^2}{4}, m + \frac{1}{2}, z^2)\right], \\ F_L(z) &= e^{-z^2/2}(z^2)^{\frac{3}{2}+\frac{m}{2}}\left[c_1 U(m + \frac{1}{2} - \frac{Q^2}{4k}, m + \frac{3}{2}, z^2)\right], \end{aligned} \tag{C.14}$$

where d_1 and c_1 are constant coefficients. By arguments similar to the previous section, the spectrum of nucleons is given by

$$m_{N^n}^2 = 4n + 4m + 2\,, \quad n = 0, 1, 2, 3, \ldots \tag{C.15}$$

where m is usually chosen as $m = 3/2$ ($\Delta = 7/2$) or $m = 5/2$ ($\Delta = 9/2$). The normalisable solutions are expressed by

$$f^n_{R/L}(z) = N_{R/L} e^{-z^2/2} z^{5/2+(m\mp 1/2)} L_n^{m\mp 1/2}(z^2)\,, \tag{C.16}$$

with the normalisation constants

$$N_R = \sqrt{\frac{2\,\Gamma(n+1)}{\Gamma(n+m+1/2)}} \quad , \quad N_L = \sqrt{\frac{2\,\Gamma(n+1)}{\Gamma(n+m+3/2)}}\,, \tag{C.17}$$

obtained from the normalisation condition

$$\int dz\, e^{4A(z)} f^m_{R/L}(z) f^n_{R/L}(z) = \delta^{mn} \tag{C.18}$$

Using the holographic dctionary (4.79), the nucleon decay constants in the soft wall model take the form

$$\begin{aligned} \lambda_{N^n} &= \sqrt{G_F} f_n = \sqrt{G_F}\Big[z^{-2-m} f_{R,n}(z)\Big]_{z=\epsilon} \\ &= \sqrt{G_F} N_R L_n^{m-1/2}(0) = \frac{2\sqrt{G_F}}{N_R \Gamma(m+1/2)}\,. \end{aligned} \tag{C.19}$$

which is compatible with [38].

D Vector mesons and nucleons in the hard wall model

In the context of holographic QCD models in the bottom-up approach, the hard wall model is the pioneer. This model was proposed by Polchinski and Strassler in the study of glueball scattering in the fixed angle regime [13]. Further investigations of glueballs [14], mesons and chiral symmetry breaking [15] showed that the hard wall model constitutes a very good toy model for investigating hadronic physics. The model consists of cutting the AdS space limiting the holographic coordinate to the region $0 < z \leq z_0$ and imposing boundary conditions for the 5d fields. By slicing the AdS space, the conformal symmetry is broken and this allows a mass gap to be generated. In this appendix, we briefly review the field equations and solutions, the spectrum and the decay constants for the case of vector mesons and nucleons.

D.1 Vector mesons

In order to describe vector mesons in the hard wall model, we consider $A_s = -\ln z$ and $\Phi = 0$ in equation (3.17) and take the vector field as $V_\perp^{\hat{\mu},a} = \eta_\perp^{\hat{\mu}} v(q^2, z)$. This reduces the equation to

$$\left[z^2\partial_z^2 - z\partial_z\right] v - q^2 z^2 v = 0\,. \tag{D.1}$$

This differential equation has analytical solutions, that can be written as

$$v(Q, z) = z\left[c_1\, J_1(Qz) + c_2\, Y_1(Qz)\right]\,, \tag{D.2}$$

where $J_m(x)$ and $Y_m(x)$ are Bessel functions of the first and second kind respectively and $Q = \sqrt{-q^2}$. The normalisable solution is given by

$$v_n(z) = N_{v^n} z J_1(Q_n z) \tag{D.3}$$

The spectrum of the vector mesons can be obtained by imposing a Neumann boundary condition at the hard wall $z = z_0$. For simplicity, we will work in units where $z_0 = 1$. The condition at the hard wall becomes

$$J_0(Q_n) = 0 \tag{D.4}$$

and the spectrum is given by

$$m_{V^n} = Q_n = j_{0,n} \quad n = 1, 2, \dots , \tag{D.5}$$

where $j_{0,n}$ are zeros of the Bessel function $J_0(x)$. The behavior of the normalisable solution for small z is

$$v_n(z) = c_{2,n} z^2 + c_{4,n} z^4 + ... \tag{D.6}$$

where we can identify the coefficients as

$$c_{2,n} = \frac{1}{2} N_{v^n} Q_n, \quad c_{4,n} = -\frac{Q_n^2}{8} c_{2,n} \, . \tag{D.7}$$

Using the orthonormality condition

$$\int_0^{z_0} dz \, e^{A_s} v_m v_n = \delta_{mn} \tag{D.8}$$

we obtain the normalisation constant

$$N_{v^n} = \frac{\sqrt{2}}{J_1(j_{0,n})} \, . \tag{D.9}$$

The holographic dictionary for the decay constants can be written as, see for instance [26, 55],

$$f_{v^n} = \frac{1}{g_5} \Big[e^{A_s} \partial_z v_n(z) \Big]_{z \to 0} = \frac{2 c_{2,n}}{g_5} = \frac{N_{v^n} Q_n}{g_5} \, . \tag{D.10}$$

D.2 Nucleons

The nucleons can be described in the hard wall model from (4.27) considering $A_s = -\ln z$ and $\Phi = 0$, thus $\tilde{m} = m$. In this way, the equation (4.27) reduces to

$$\Big\{ (z\partial_z)^2 - 5 z \partial_z + 6 \pm m - m^2 + Q^2 z^2 \Big\} F_{R/L} = 0 \, , \tag{D.11}$$

whose solution is

$$F_{R/L}(z) = c_1 z^{5/2} J_{m \mp \frac{1}{2}}(Qz) + c_2 z^{5/2} Y_{m \mp \frac{1}{2}}(Qz) \tag{D.12}$$

The normalisable solutions are given by

$$f^n_{R/L}(z) = N^n_{R/L} z^{5/2} J_{m\mp\frac{1}{2}}(Q_n z)\,. \tag{D.13}$$

Again we will work fix the hard wall position as $z_0 = 1$ for simplicity. In the hard wall model there are two possible boundary conditions for the normalisable solutions in the nucleon sector, fixing either f^n_R (model I) or f^n_L (model I) at the hard wall. We will be interested only in model I because model II allows for a zero mode not present in the nucleon spectrum [36, 38]. Fixing f^n_R at the hard wall $z = z_0$ we obtain

$$J_{m-\frac{1}{2}}(Q_n) = 0 \tag{D.14}$$

The spectrum of nucleons becomes

$$Q_n = j_{1,n} \quad \text{when} \quad m = 3/2\,, \tag{D.15}$$

$$Q_n = j_{2,n} \quad \text{when} \quad m = 5/2\,, \tag{D.16}$$

where $j_{1,n}$ and $j_{2,n}$ are zeros of the Bessel functions $J_1(x)$ and $J_2(x)$, respectively. The normalisation condition is given by

$$\int dz\, e^{4A_s} f^m_{R/L}(z) f^n_{R/L}(z) = \delta^{mn} \tag{D.17}$$

Using this condition we find the normalisation constants

$$N_{R/L} = \frac{\sqrt{2}}{J_{m+\frac{1}{2}}(Q_n)}\,. \tag{D.18}$$

Using the holographic dictionary (4.79) we find for the hard wall model that

$$\lambda_n = \sqrt{G_F}\Big[z^{-2-m} f_{R,n}(z)\Big]_{z=\epsilon} = \sqrt{G_F} N_{R/L} \frac{2^{1/2-m} Q_n^{m-1/2}}{\Gamma(m+1/2)}\,, \tag{D.19}$$

which is compatible with [38].

References

[1] V. A. Miransky, Dynamical symmetry breaking in quantum field theories. World Scientific, 1994.

[2] E. V. Shuryak, The QCD vacuum, hadrons and the superdense matter, vol. 71. 10, 2004.

[3] Y. Nambu and G. Jona-Lasinio, "Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1.," Phys. Rev. **122** (1961) 345–358.

[4] Y. Nambu and G. Jona-Lasinio, "Dynamical model of elementary particles based on an analogy with superconductivity. II.," Phys. Rev. **124** (1961) 246–254.

[5] M. Gell-Mann and M. Levy, "The axial vector current in beta decay," Nuovo Cim. **16** (1960) 705.

[6] S. Scherer, "Introduction to chiral perturbation theory," Adv. Nucl. Phys. **27** (2003) 277, arXiv:hep-ph/0210398.

[7] K. G. Wilson, "Confinement of Quarks," Phys. Rev. D **10** (1974) 2445–2459.

[8] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, "QCD and Resonance Physics. Theoretical Foundations," Nucl. Phys. B **147** (1979) 385–447.

[9] P. Colangelo and A. Khodjamirian, "QCD sum rules, a modern perspective," arXiv:hep-ph/0010175.

[10] J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity," Adv. Theor. Math. Phys. **2** (1998) 231–252, arXiv:hep-th/9711200.

[11] E. Witten, "Anti-de Sitter space and holography," Adv. Theor. Math. Phys. **2** (1998) 253–291, arXiv:hep-th/9802150.

[12] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, "Gauge theory correlators from noncritical string theory," Phys. Lett. B **428** (1998) 105–114, arXiv:hep-th/9802109.

[13] J. Polchinski and M. J. Strassler, "Hard scattering and gauge / string duality," Phys. Rev. Lett. **88** (2002) 031601, arXiv:hep-th/0109174.

[14] H. Boschi-Filho and N. R. F. Braga, "Gauge / string duality and scalar glueball mass ratios," JHEP **05** (2003) 009, arXiv:hep-th/0212207.

[15] J. Erlich, E. Katz, D. T. Son, and M. A. Stephanov, "QCD and a holographic model of hadrons," Phys. Rev. Lett. **95** (2005) 261602, arXiv:hep-ph/0501128.

[16] L. Da Rold and A. Pomarol, "Chiral symmetry breaking from five dimensional spaces," Nucl. Phys. B **721** (2005) 79–97, arXiv:hep-ph/0501218.

[17] A. Karch, E. Katz, D. T. Son, and M. A. Stephanov, "Linear confinement and AdS/QCD," Phys. Rev. D **74** (2006) 015005, arXiv:hep-ph/0602229.

[18] U. Gursoy and E. Kiritsis, "Exploring improved holographic theories for QCD: Part I," JHEP **02** (2008) 032, arXiv:0707.1324 [hep-th].

[19] U. Gursoy, E. Kiritsis, and F. Nitti, "Exploring improved holographic theories for QCD: Part II," JHEP **02** (2008) 019, arXiv:0707.1349 [hep-th].

[20] S. S. Gubser and A. Nellore, "Mimicking the QCD equation of state with a dual black hole," Phys. Rev. D **78** (2008) 086007, arXiv:0804.0434 [hep-th].

[21] R.-G. Cai, S. He, and D. Li, "A hQCD model and its phase diagram in Einstein-Maxwell-Dilaton system," JHEP **03** (2012) 033, arXiv:1201.0820 [hep-th].

[22] D. Li and M. Huang, "Dynamical holographic QCD model for glueball and light meson spectra," JHEP **11** (2013) 088, arXiv:1303.6929 [hep-ph].

[23] A. Ballon-Bayona, H. Boschi-Filho, L. A. H. Mamani, A. S. Miranda, and V. T. Zanchin, "Effective holographic models for QCD: glueball spectrum and trace anomaly," Phys. Rev. D **97** no. 4, (2018) 046001, arXiv:1708.08968 [hep-th].

[24] A. Karch and E. Katz, "Adding flavor to AdS / CFT," JHEP **06** (2002) 043, arXiv:hep-th/0205236.

[25] T. Sakai and S. Sugimoto, "Low energy hadron physics in holographic QCD," Prog. Theor. Phys. **113** (2005) 843–882, arXiv:hep-th/0412141.

[26] H. R. Grigoryan and A. V. Radyushkin, "Form Factors and Wave Functions of Vector Mesons in Holographic QCD," Phys. Lett. B **650** (2007) 421–427, arXiv:hep-ph/0703069.

[27] H. R. Grigoryan and A. V. Radyushkin, "Structure of vector mesons in holographic model with linear confinement," Phys. Rev. D **76** (2007) 095007, arXiv:0706.1543 [hep-ph].

[28] H. Forkel, M. Beyer, and T. Frederico, "Linear square-mass trajectories of radially and orbitally excited hadrons in holographic QCD," JHEP **07** (2007) 077, arXiv:0705.1857 [hep-ph].

[29] W. de Paula, T. Frederico, H. Forkel, and M. Beyer, "Dynamical AdS/QCD with area-law confinement and linear Regge trajectories," Phys. Rev. D **79** (2009) 075019, arXiv:0806.3830 [hep-ph].

[30] E. Folco Capossoli, M. A. Martín Contreras, D. Li, A. Vega, and H. Boschi-Filho, "Hadronic spectra from deformed AdS backgrounds," Chin. Phys. C **44** no. 6, (2020) 064104, arXiv:1903.06269 [hep-ph].

[31] I. Iatrakis, E. Kiritsis, and A. Paredes, "An AdS/QCD model from tachyon condensation: II," JHEP **11** (2010) 123, arXiv:1010.1364 [hep-ph].

[32] D. Areán, I. Iatrakis, M. Järvinen, and E. Kiritsis, "The discontinuities of conformal transitions and mass spectra of V-QCD," JHEP **11** (2013) 068, arXiv:1309.2286 [hep-ph].

[33] S. He, S.-Y. Wu, Y. Yang, and P.-H. Yuan, "Phase Structure in a Dynamical Soft-Wall Holographic QCD Model," JHEP **04** (2013) 093, arXiv:1301.0385 [hep-th].

[34] A. Ballon-Bayona, T. Frederico, L. A. H. Mamani, and W. de Paula, "Dynamical holographic QCD model for spontaneous chiral symmetry breaking and confinement," Phys. Rev. D **108** no. 10, (2023) 106016, arXiv:2308.07503 [hep-ph].

[35] G. F. de Teramond and S. J. Brodsky, "Hadronic spectrum of a holographic dual of QCD," Phys. Rev. Lett. **94** (2005) 201601, arXiv:hep-th/0501022.

[36] D. K. Hong, T. Inami, and H.-U. Yee, "Baryons in AdS/QCD," Phys. Lett. B **646** (2007) 165–171, arXiv:hep-ph/0609270.

[37] S. J. Brodsky and G. F. de Teramond, "AdS/CFT and Light-Front QCD," Subnucl. Ser. **45** (2009) 139–183, arXiv:0802.0514 [hep-ph].

[38] Z. Abidin and C. E. Carlson, "Nucleon electromagnetic and gravitational form factors from holography," Phys. Rev. D **79** (2009) 115003, arXiv:0903.4818 [hep-ph].

[39] T. Gutsche, V. E. Lyubovitskij, I. Schmidt, and A. Vega, "Dilaton in a soft-wall holographic approach to mesons and baryons," Phys. Rev. D **85** (2012) 076003, arXiv:1108.0346 [hep-ph].

[40] H. Hata, T. Sakai, S. Sugimoto, and S. Yamato, "Baryons from instantons in holographic QCD," Prog. Theor. Phys. **117** (2007) 1157, arXiv:hep-th/0701280.

[41] K. Hashimoto, T. Sakai, and S. Sugimoto, "Holographic Baryons: Static Properties and Form Factors from Gauge/String Duality," Prog. Theor. Phys. **120** (2008) 1093–1137, arXiv:0806.3122 [hep-th].

[42] A. Pomarol and A. Wulzer, "Baryon Physics in Holographic QCD," Nucl. Phys. B **809** (2009) 347–361, arXiv:0807.0316 [hep-ph].

[43] A. Cherman, T. D. Cohen, and M. Nielsen, "Model Independent Tests of Skyrmions and

Their Holographic Cousins," Phys. Rev. Lett. **103** (2009) 022001, arXiv:0903.2662 [hep-ph].

[44] P. Sutcliffe, "Holographic Skyrmions," Mod. Phys. Lett. B **29** no. 16, (2015) 1540051.

[45] Y. Hayashi, T. Ogino, T. Sakai, and S. Sugimoto, "Stringy excited baryons in holographic quantum chromodynamics," PTEP **2020** no. 5, (2020) 053B04, arXiv:2001.01461 [hep-th].

[46] M. Järvinen, E. Kiritsis, F. Nitti, and E. Préau, "The V-QCD baryon: numerical solution and baryon spectrum," JHEP **05** (2023) 081, arXiv:2212.06747 [hep-th].

[47] R. Abt, J. Erdmenger, N. Evans, and K. S. Rigatos, "Light composite fermions from holography," JHEP **11** (2019) 160, arXiv:1907.09489 [hep-th].

[48] T. Nakas and K. S. Rigatos, "Fermions and baryons as open-string states from brane junctions," JHEP **12** (2020) 157, arXiv:2010.00025 [hep-th].

[49] Y. Kinar, E. Schreiber, and J. Sonnenschein, "Q anti-Q potential from strings in curved space-time: Classical results," Nucl. Phys. B **566** (2000) 103–125, arXiv:hep-th/9811192.

[50] O. Andreev and V. I. Zakharov, "Heavy-quark potentials and AdS/QCD," Phys. Rev. D **74** (2006) 025023, arXiv:hep-ph/0604204.

[51] A. V. Ramallo, "Introduction to the AdS/CFT correspondence," Springer Proc. Phys. **161** (2015) 411–474, arXiv:1310.4319 [hep-th].

[52] E. Witten, "Baryons in the 1/n Expansion," Nucl. Phys. B **160** (1979) 57–115.

[53] D. T. Son and M. A. Stephanov, "QCD and dimensional deconstruction," Phys. Rev. D **69** (2004) 065020, arXiv:hep-ph/0304182.

[54] A. Ballon-Bayona and L. A. H. Mamani, "Nonlinear realization of chiral symmetry breaking in holographic soft wall models," Phys. Rev. D **102** no. 2, (2020) 026013, arXiv:2002.00075 [hep-ph].

[55] A. Ballon-Bayona, L. A. H. Mamani, and D. M. Rodrigues, "Spontaneous chiral symmetry breaking in holographic soft wall models," Phys. Rev. D **104** no. 12, (2021) 126029, arXiv:2107.10983 [hep-ph].

[56] **OBELIX** Collaboration, A. Bertin et al., "Study of anti-p p –> 2pi+ 2pi- annihilation from S states," Phys. Lett. B **414** (1997) 220–228.

[57] **Particle Data Group** Collaboration, R. L. Workman et al., "Review of Particle Physics," PTEP **2022** (2022) 083C01.

[58] J. F. Donoghue, E. Golowich, and B. R. Holstein, Dynamics of the standard model, vol. 2. CUP, 2014.

[59] B. L. Ioffe, "Calculation of Baryon Masses in Quantum Chromodynamics," Nucl. Phys. B **188** (1981) 317–341. [Erratum: Nucl.Phys.B 191, 591–592 (1981)].

[60] T. D. Cohen, R. J. Furnstahl, D. K. Griegel, and X.-m. Jin, "QCD sum rules and applications to nuclear physics," Prog. Part. Nucl. Phys. **35** (1995) 221–298, arXiv:hep-ph/9503315.

[61] D. B. Leinweber, "QCD sum rules for skeptics," Annals Phys. **254** (1997) 328–396, arXiv:nucl-th/9510051.

[62] D. B. Leinweber, "Nucleon properties from unconventional interpolating fields," Phys. Rev. D **51** (1995) 6383–6393, arXiv:nucl-th/9406001.

[63] H. C. Ahn, D. K. Hong, C. Park, and S. Siwach, "Spin 3/2 Baryons and Form Factors in AdS/QCD," Phys. Rev. D **80** (2009) 054001, arXiv:0904.3731 [hep-ph].

[64] N. J. Huseynova, "Coupling Constant of a Vector Meson with Delta Baryons in the Soft-Wall Ads/QCD Model," Russ. Phys. J. **63** no. 5, (2020) 860–866.

[65] G. B. Arfken and H. J. Weber, "Mathematical methods for physicists," 1999.

[66] E. Butkov, Mathematical physics. Addison-Wesley, 1968.

[67] C. Itzykson and J. B. Zuber, Quantum Field Theory. International Series In Pure and Applied Physics. McGraw-Hill, New York, 1980.

[68] M. E. Peskin and D. V. Schroeder, An Introduction to quantum field theory. Addison-Wesley, Reading, USA, 1995.

arXiv:2402.17859v1 [hep-ph] 27 Feb 2024

OU-HEP-240301

Weak scale supersymmetry emergent from the string landscape

Howard Baer[1*], Vernon Barger[2†], Dakotah Martinez[1§] and Shadman Salam[3§]

[1]*Homer L. Dodge Department of Physics and Astronomy, University of Oklahoma, Norman, OK 73019, USA*

[2]*Department of Physics, University of Wisconsin, Madison, WI 53706 USA*

[3]*Department of Mathematics and Natural Sciences, Brac University, Dhaka 1212, Bangladesh*

Abstract

Superstring flux compactifications can stabilize all moduli while leading to an enormous number of vacua solutions, each leading to different $4-d$ laws of physics. While the string landscape provides at present the only plausible explanation for the size of the cosmological constant, it may also predict the form of weak scale supersymmetry which is expected to emerge. Rather general arguments suggest a power-law draw to large soft terms, but these are subject to an anthropic selection of not-too-large a value for the weak scale. The combined selection allows one to compute relative probabilities for the emergence of supersymmetric models from the landscape. Models with weak scale naturalness appear most likely to emerge since they have the largest parameter space on the landscape. For finetuned models such as high scale SUSY or split SUSY, the required weak scale finetuning shrinks their parameter space to tiny volumes, making them much less likely to appear compared to natural models. Probability distributions for sparticle and Higgs masses from natural models show a preference for Higgs mass $m_h \sim 125$ GeV with sparticles typically beyond present LHC limits, in accord with data. From these considerations, we briefly describe how natural SUSY is expected to be revealed at future LHC upgrades. This article is a contribution to the Special Edition of the journal *Entropy* honoring Paul Frampton on his 80th birthday.

*Email: baer@ou.edu

†Email: barger@pheno.wisc.edu

§Email: dakotah.s.martinez-1@ou.edu

§Email: ext.shadman.salam@bracu.ac.bd

1 Introduction

Superstring theory provides the most promising avenue for unifying the Standard Model with gravity, but at the cost of requiring 6 or 7 extra spatial dimensions [1–5]. The low energy limit $E < m_P$ (where m_P is the reduced Planck mass) of string theory, once Kaluza-Klein modes are integrated out, is expected to be $10-d$ supergravity (SUGRA). The $10-d$ SUGRA theory is then assumed to be compactified to a tiny $6-d$ space K tensored with our usual $4-d$ (approximately) Minkowski spacetime M_4: $M_{10} = M_4 \times K$. Originally, K was taken to be a $6-d$ compact Ricci-flat Kähler manifold with special holonomy [6]; such a Calabi-Yau manifold admits a conserved Killing spinor leading to a remnant $N = 1$ supersymmetry (SUSY) on M_4.

The cosmological constant (CC) problem remained a thorny issue until the early 2000s when it was realized that string flux compactifications could lead to an enormous number of vacuum states each with different $4-d$ laws of physics, and in particular, different Λ_{CC} values [7]. Such large numbers of vacuum states ($N_{vac} \sim 10^{500}$ is an oft-quoted number [8]) provided a setting for Weinberg's anthropic solution to the CC problem [9]. But if the landscape [10] of string vacua provides a solution to the CC problem, might it also enter into other naturalness problems, such as the m_{weak}/m_P (or related, m_{SUSY}/m_P) hierarchy problems (where $m_P \simeq 2.4 \times 10^{18}$ GeV)?

In this contribution to the volume of the journal *Entropy* honoring Paul Frampton on his 80th birthday, we address this question. Here, we will put forward arguments in favor not only of weak scale SUSY as emergent from the string landscape, but indeed of a special form of weak scale SUSY (WSS)– SUSY with radiatively-driven naturalness [11, 12], or stringy-natural SUSY [13]. The specific form of WSS predicts at present that a light SUSY Higgs boson with mass $m_h \simeq 125$ GeV should emerge, whilst sparticles masses are at present somewhat or well beyond reach of the CERN Large Hadron Collider (LHC) [14]. It also allows us to predict a variety of SUSY signatures which may allow for SUSY discovery at LHC luminosity upgrades over the upcoming years. Perhaps most important of these are the soft isolated opposite-sign dileptons+MET which arise from light higgsino pair production [15] and which recoil against a hard initial state jet radiation [16–19]. At present, both ATLAS [20] and CMS [21] with 139 fb^{-1} seem to have 2σ excesses in this channel, and an associated monojet signal may also be emerging [22].

2 Approximate supersymmetric vacua from string theory

The main motivation for SUSY is that it provides a 'tHooft technical naturalness solution to the gauge hierarchy problem via the cancellation of quadratic divergences associated with the Higgs sector. This is true for SUSY breaking at any energy scale below m_P, since in the limit of $m_{SUSY} \to 0$, the model becomes more (super)symmetric. Thus, SUSY provides a technically natural solution to the so-called Big Hierarchy problem.

Specific motivation for weak scale SUSY comes from the Little Hierarchy problem and what we call *practical naturalness [23, 24]:* an observable $\mathcal{O}$ is practically natural if all independent contributions to $\mathcal{O}$ are comparable to or less than $\mathcal{O}$. For the case of WSS, we can relate the

weak scale $m_{weak} \sim m_{W,Z,h} \sim 100$ GeV to the weak scale soft SUSY breaking terms and SUSY conserving μ term via the scalar potential minimization conditions:

$$\frac{m_Z^2}{2} = \frac{m_{H_d}^2 + \Sigma_d^d - (m_{H_u}^2 + \Sigma_u^u)\tan^2\beta}{\tan^2\beta - 1} - \mu^2 \sim -m_{H_u}^2 - \Sigma_u^u(\tilde{t}_{1,2}) - \mu^2 \qquad (1)$$

where $m_{H_{u,d}}^2$ are the soft SUSY breaking Higgs masses and the $\Sigma_{u,d}^{u,d}$ contain an assortment of loop corrections to the scalar potential (explicit formulae are included in Ref's [12]). A measure of practical naturalness Δ_{EW} can be defined which compares the largest (absolute) contribution to the right-hand-side of Eq. 1 to $m_Z^2/2$. Requiring $\Delta_{EW} \lesssim 30$ fulfills the practical naturalness condition. From Eq. 1, we see immediately that $m_{H_u}^2$ must be driven to *small* negative values at the weak scale while the μ term must also be $\mu \sim 100 - 350$ GeV. The latter condition means the higgsinos are usually the lightest SUSY particles, and the only ones required to be $\sim m_{weak}$. The other sparticles enter via the $\Sigma_{u,d}^{u,d}$ terms and hence are suppressed by loop factors, and so can live in the TeV or beyond range. We shall see shortly that practical naturalness is closely linked to selection of SUSY models on the landscape.

On the theory side, we expect the $4-d$ vacua emergent from the landscape to often contain some remnant SUSY.

- *Remnant spacetime SUSY:* In Ref. [25], Acharya argues that all stable, Ricci-flat manifolds in dimensions < 11 have special holonomy, and consequently a conserved Killing spinor. If so, then some remnant spacetime SUSY should exist in the $4-d$ low-energy effective field theory (LE-EFT).

- *EW stability:* A problem for the Standard Model to be the low-energy effective field theory for $E < m_P$ is electroweak stability in that the Higgs quartic term λ may evolve to negative values at $E > m_{GUT}$ leading to a runaway scalar potential. For $m_t \sim 173.2$ GeV, the SM is just on the edge of metastability/runaway [26, 27]. The Minimal Supersymmetric Standard Model (MSSM) has no such problem since for the MSSM the quartic couplings involve the gauge couplings which are always positive.

- *Landscape vacua stability:* In Ref's [28, 29], Dine *et al.* ask what sort of conditions can stabilize landscape deSitter vacua against decay to AdS vacua, leading to a big crunch. The presence of SUSY leads to absolutely stable vacua whilst the presence of approximate (broken) SUSY leads to (metastable) vacua decay rates $\Gamma \sim m_P e^{-m_P^2/m_{3/2}^2}$ far beyond the age of the universe.

- *Hierarchy of scales:* While a hierarchy of scales is typically hard to come by in many BSM models, SUSY models allow for dynamical SUSY breaking [30] where the SUSY breaking scale m_{hidden} is gained via dimensional transmutation $m_{hidden} \sim m_P e^{-8\pi^2/bg^2}$ and where the soft terms are developed as $m_{soft} \sim m_{hidden}^2/m_P$ under gravity-mediation.[1]

- *Harmony:* Witten emphasizes that consistent QFTs exist for spin-0, 1/2, 1, 3/2 and 2. The graviton is the physical spin-2 particle and the spin-3/2 Rarita-Schwinger gravitino

[1]Here, we only consider gravity-mediation since gauge mediation leads to tiny trilinear soft terms A which then require unnatural top-squark contributions Σ_u^u to gain $m_h \sim 125$ GeV [31].

field would exist as the superpartner of the graviton, thus filling out all allowed spin states.

In addition, WSS is motivated experimentally by a variety of measurements.

- The measured values of the gauge couplings unify under MSSM RG evolution while not under most other BSM extensions, including the SM itself [32].
- The measured top-quark mass is large enough to seed the required radiative breakdown of EW symmetry [33].
- The measured value of $m_h \simeq 125$ GeV falls squarely into the range allowed by the MSSM: $m_h \lesssim 130$ GeV [34].
- Precision EW corrections tend to prefer the (heavy spectra) MSSM over the SM [35].

It is often complained, with good reason, that gravity mediation has its own flavor and CP problems, the former arising from operators such as $\int d^4\theta S^\dagger S Q_i^\dagger Q_j/m_P^2$ where S is a hidden sector superfield obtaining a SUSY breaking vev $F_S \sim 10^{11}$ GeV and the Q_i are visible sector chiral superfields with generation index $i,j = 1-3$. Since no symmetry forbids such flavor mixing, then FCNCs are expected to be large in gravity-mediated SUSY breaking (historically, this strongly motivated the search for flavor conserving models such as gauge-mediation and sometimes anomaly-mediation.) It is pointed out in Ref. [36] that the landscape provides its own decoupling/quasi-degeneracy solution to the SUSY flavor and CP problems by pulling first/second generation matter scalars to a flavor-independent upper bound in the 20-40 TeV range.

For these reasons, we will assume a so-called "fertile patch" or friendly neighborhood [37] of the string landscape: those vacua which include the MSSM as the LE-EFT and where only the CC and the magnitude of the weak scale scan within the landscape. In this case, Yukawa couplings and gauge couplings are instead fixed by string dynamics rather than environmental selection. This leads to predictive landscape models [37]: if the CC is too large, then large scale structure will not form which seems required for complexity to emerge (the structure principle, leading to Weinberg's successful prediction of Λ_{CC}). Only the magnitude of the weak scale scans. If $m_{weak}^{PU} \gtrsim 4m_{weak}^{OU}$, then the down-up quark mass difference becomes so large that neutrons are no longer stable in nuclei and the only atoms formed in the early universe are Hydrogen. If $m_{weak}^{PU} \lesssim 0.5m_{weak}^{OU}$, then we get a universe with only neutrons. This is the atomic principle [38], since complex nuclei are also apparently needed for complexity to emerge in any pocket universe (PU) within the greater multiverse (and where OU refers to m_{weak} in our universe).

3 Natural SUSY from the landscape

It is emphasized by Douglas that the CC scans independently of the SUSY breaking scale in the landscape [39]. For the SUSY breaking scale, we expect the vacua to be distributed as

$$dN_{vac} \sim f_{SUSY} \cdot f_{EWSB} \cdot dm_{SUSY}^2 \tag{2}$$

where m_{SUSY} is the overall hidden sector SUSY breaking scale expected to be $\sim 10^{11}$ GeV such that the scale of soft terms is given by $m_{soft} \sim m_{SUSY}^2/m_P$.

3.1 Distribution of soft breaking terms on the landscape

How is f_{SUSY} distributed? Douglas [39] emphasizes that there is nothing in string theory to favor any particular SUSY breaking vev over another, and hence m_{soft} would be distributed as a power-law

$$f_{SUSY} \sim m_{soft}^{2n_F+n_D-1} \tag{3}$$

where n_F is the total number of (complex-valued) F-breaking fields and n_D is the total number of (real-valued) D-breaking fields contributing to the overall scale of SUSY breaking $m_{SUSY}^4 = \sum_i |F_i|^2 + \sum_\alpha D_\alpha^2$. The prefactor of 2 in the exponent comes since the F_i are distributed randomly as complex numbers. For the textbook case of SUSY breaking via a single F term, then we expect $f_{SUSY} \sim m_{soft}^1$, *i.e.* a *linear* draw to large soft terms. If more hidden fields contribute to the overall SUSY breaking scale, then the draw to large soft terms will be a stronger power-law.

While the overall SUSY breaking scale is distributed as a power-law, the different functional dependence [40–42] of the soft terms on the hidden sector SUSY breaking fields means that gaugino masses, the trilinear soft terms and the various scalar masses will effectively scan independently on the landscape [43]. Now it is an *advantage* that different scalar mass-squared terms scan independently (as expected in SUGRA) since the first/second generation scalars get pulled to much higher values than 3rd generation scalars, while the two Higgs soft masses are also non-universal and scan independently. This situation is borne out in Nilles *et al.* mini-landscape where different fields gain different soft masses due to their different geographical locations on the compactification manifold [44]. In terms of gravity mediation, then the so-called n-extra-parameter non-universal Higgs model (NUHMn) with parameters [45, 46]

$$m_0(i),\ m_{H_u},\ m_{H_d},\ m_{1/2}, A_0,\ \tan\beta \qquad (NUHM4) \tag{4}$$

provides the proper template. Since the matter scalars fill out a complete spinor rep of $SO(10)$, we assume each generation $i = 1{-}3$ is unified to $m_0(i)$. Also, for convenience one may ultimately trade m_{H_u} and m_{H_d} for the more convenient weak scale parameters m_A and μ. One may also build (and scan separately) the natural anomaly-mediated SUSY breaking model [47, 48] (nAMSB) and the natural generalized mirage mediation model [49] (nGMM).

3.2 The ABDS window

The anthropic selection on the landscape comes from f_{EWSB}. This involves a rather unheralded prediction of the MSSM: the value of the weak scale in terms of soft SUSY breaking parameters and μ, as displayed in Eq. 1. However, in the multiverse, here we rely on the existence of a friendly neighborhood [37] wherein the LE-EFT contains the MSSM but where only dimensionful quantities such as Λ_{CC} and $v_u^2 + v_d^2$ scan, whilst dimensionless quantities like gauge and Yukawa couplings are determined by dynamics. This assumption leads to *predictivity* as we shall see.

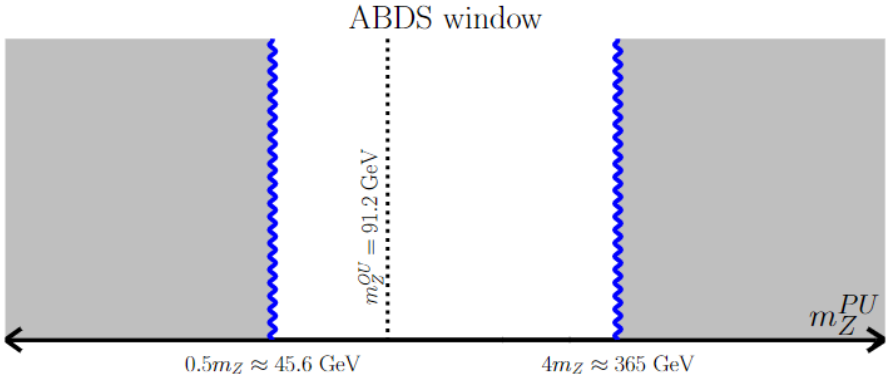


Figure 1: The ABDS-allowed window within the range of m_Z^{PU} values.

Under these assumptions, then we ask what conditions lead to complex nuclei, atoms as we know them, and hence the ability to generate complex lifeforms in a pocket universe? For different values of soft terms, frequently one is pushed into a weak scale scalar potential with charge-or-color breaking minima (CCB) where one or more charged or colored scalar mass squared is driven tachyonic (i.e., $m^2 < 0$). Such CCB minima must be vetoed. Also, for too large of values of $m_{H_u}^2$, then its value is *not* driven to negative values and EW symmetry is not broken. These we label as "no EWSB" and veto them as well. In practice, we must check boundedness of the scalar potential from below in the vacuum stability conditions and that the origin of field space has been destabilized at tree-level.

At this point, we are left with (MS)SM vacua where EW symmetry is properly broken, but where $m_{weak} \sim m_{W,Z,h}$ is at a different value from what we see in our universe. Here, we rely on the prescient analysis of Agrawal, Barr, Donoghue and Seckel (ABDS) [38, 50]. If the derived value of the weak scale is bigger than ours by a factor $(2-5)$, then the light quark mass difference $m_d - m_u$ becomes so large that neutrons are no longer stable in the nucleus and nuclei with $Z \gg N$ are not bound; such pocket universes would have nuclei of single protons only, and would be chemically inert. Likewise, if the PU value of the weak scale is a factor ~ 0.5 less than our measured value, then one obtains a universe with only neutrons– also chemically inert. The ABDS window of allowed values is that

$$0.5m_{weak}^{OU} < m_{weak}^{PU} \lesssim 4m_{weak}^{OU} \tag{5}$$

where we take the $(2-5)m_{weak}^{OU}$ to be $\sim 4m_{weak}^{OU}$ for definiteness, which is probably a conservative value. It is very central to our analysis and so is displayed in Fig. 1. Our anthropic condition f_{EWSB} is then that the scalar potential lead to minima with not only appropriate EWSB, but also that the derived value of the weak scale lie within the ABDS window. Vacua not fulfilling these conditions must be vetoed.[2]

[2]Early papers on this topic used instead a naturalness "penalty" of $f_{EWSB} \sim m_{weak}^2/m_{SUSY}^2$; this condition would allow for many of the vacua which are forbidden by our approach.

3.3 EW natural SUSY emergent from the landscape

The next goal is to build a toy simulation of our friendly neighborhood of the string landscape. We can generate the soft terms of Eq. 4 according to a power-law selection, usually taken to be $n = 2n_F + n_D - 1 = 1$ (linear draw). While Eq. 3 favors the largest possible soft terms, the anthropic veto f_{EWSB} places an upper bound on such terms because usually large soft terms lead to too large a value of m_{weak}^{PU} beyond the ABDS window. The trick is to take the upper bound on scan limits beyond the upper bound posed by f_{EWSB}. However, in some cases larger soft terms are *more* apt to lie within the ABDS window. A case in point is $m_{H_u}^2$: the smaller its value, the deeper negative it runs to unnatural values at the weak scale, while as it gets larger, then it barely runs negative: EW symmetry is barely broken. As its high scale value becomes even larger, it doesn't run negative by m_{weak}, and EW symmetry is not broken– such vacua are vetoed. Also, for small A_0, the $\Sigma_u^u(\tilde{t}_{1,2})$ terms can be large. When A_0 becomes large negative, then canellations occur in $\Sigma_u^u(\tilde{t}_{1,2})$ such that these loop corrections then lie within the ABDS window: large negative weak scale A terms make $\Sigma_u^u(\tilde{t}_{1,2})$ more natural while raising the light Higgs mass $m_h \sim 125$ GeV.

A plot of the weak scale values of m_{H_u} and μ is shown in Fig. 2 (taken from Ref. [51]) for the case where all radiative corrections– some negative and some positive [52]– lie within the ABDS window. The ABDS window lies between the red and green curves. Imagine playing darts with this target, trying to land your dart within the ABDS window. There is a large region to the lower-left where both m_{H_u} and μ are $\lesssim 350$ GeV which leads to PUs with complexity. Alternatively, if you want to land your dart at a point with $\mu \sim 1000$ GeV, then the target space has pinched off to a tiny volume: the target space is finetuned and your dart will almost never land there. The EW natural SUSY models live in the lower-left ABDS window while finetuned SUSY models with large Δ_{EW} lie within the extremely small volume between the red and green curves in the upper-right plane. This tightly-constrained region is labeled by split SUSY [53], high scale SUSY [54] and minisplit [55].

It is often said that landscape selection offers an alternative to naturalness and allows for finetuned SUSY models. After all, isn't the CC finetuned? However, from Fig. 2 we see that models with EW naturalness (low Δ_{EW}) have a far greater relative probability to emerge from landscape selection than do finetuned SUSY models.

In Fig. 3 (from Ref. [51]), we perform a numerical exercise to generate high scale SUSY soft terms in accord with an $n = 1$ draw in Eq. 3. The green dots are viable vacua states with appropriate EWSB and m_{weak}^{PU} within the ABDS window. While some dots do land in the finetuned region, the bulk of points lie within the EW natural SUSY parameter space.

An alternative view is gained from Fig. 4 from Ref. [56]. Here, we compute contributions to the scalar potential within a variety of SUSY models including RNS (radiatively-driven natural SUSY [11]), CMSSM [57], G_2MSSM [58], high scale SUSY [59], spread SUSY [60], minisplit [55], split SUSY [53] and the SM with cutoff $\Lambda = 10^{13}$ TeV, indicative of the neutrino see-saw scale [61]. The x-axis is either the SM μ parameter or the SUSY μ parameter while the y-axis is the calculated value of m_Z within the PU. The ABDS window is the horizontal blue-shaded region. For μ distributed as equally likely at all scales (integrates to a log), then the length of the x-axis interval leading to m_Z^{PU} within the ABDS window can be regarded as a relative probability measure P_μ for the model to emerge from the landscape. There is a

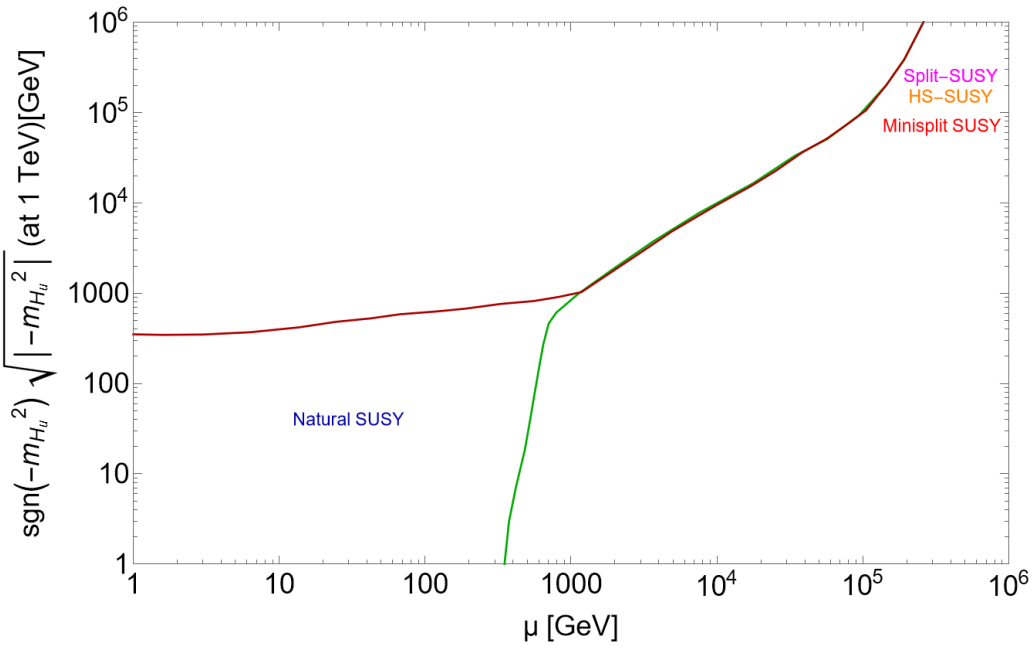


Figure 2: The μ^{PU} vs. $\sqrt{-m_{H_u}^2(weak)}$ parameter space in a toy model ignoring radiative corrections to the Higgs potential. The region between red and green curves leads to $m_{weak}^{PU} < 4m_{weak}^{OU}$ so that the atomic principle is satisfied.

substantial interval for the RNS model, but for finetuned SUSY models, the interval is typically much more narrow than the width of the printed curves. We can see that finetuned models have only a tiny range of μ values which allow habitation within the ABDS window. Using the magic of algebra, the width of the μ intervals can be computed, and the results are in Table 1. Here, P_μ is to be considered as a *relative* probability. From the Table, we see that the SM is about 10^{-27} times likely to emerge as compared to RNS. Minisplit is $10^{-4} - 10^{-8}$ times likely to emerge (depending on the version of minisplit). Even the once-popular CMSSM model is $\sim 10^{-3}$ times likely than RNS to emerge from the landscape.

4 Radiatively-driven natural SUSY

Along with probability distributions for models to emerge from the landscape, one can compute probability distributions for sparticle and Higgs mass values from particular models given an assumed value of n in f_{SUSY}. Here, we use a linear draw, $n = 1$, to large soft terms with the NUHM4 model as the LE-EFT. We capture non-finetuned models by requiring $\Delta_{EW} \lesssim 30$, *i.e.*, that the largest independent contribution to m_Z lies within the ABDS window. These models have radiatively-driven naturalness (RNS) where RG running drives various soft terms to natural values at the weak scale.

The distribution for the light Higgs mass is shown in Fig. 5 (taken from Ref. [64]). We see for $n = 1$ that the blue distribution rises to a maximum at $m_h \sim 125$ GeV. This is where A_t is large enough to yield cancellations in the $\Sigma_u^u(\tilde{t}_{1,2})$ terms, but also lifts m_h up to ~ 125

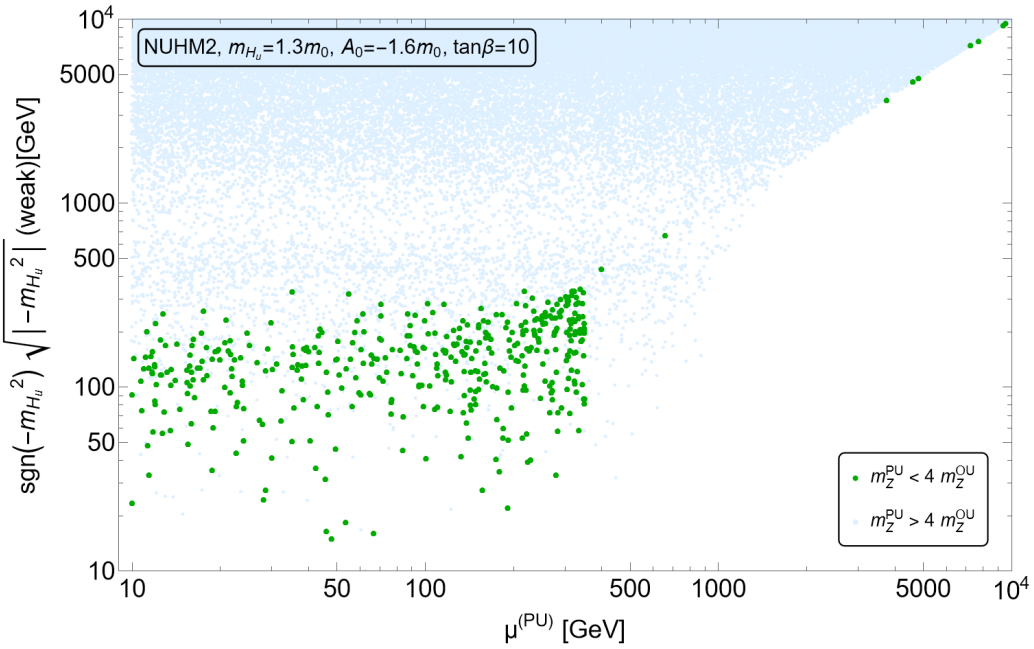


Figure 3: The value of $m_{H_u}(weak)$ vs. μ^{PU} The green points denote vacua with appropriate EWSB and with $m_{weak}^{PU} < 4m_{weak}^{OU}$ so that the atomic principle is satisfied. Blue points have $m_{weak}^{PU} > 4m_{weak}^{OU}$.

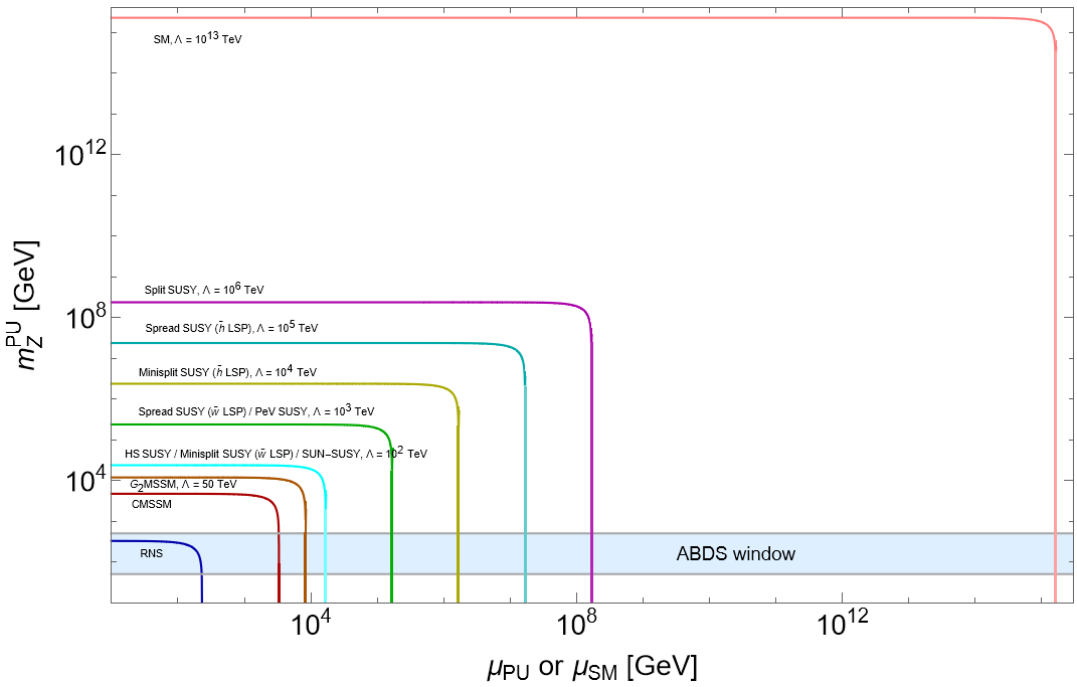


Figure 4: Values of m_Z^{PU} vs. μ_{PU} or μ_{SM} for various natural (RNS) and unnatural SUSY models and the SM. The ABDS window extends here from $m_Z^{PU} \sim 50 - 500$ GeV.

model	$\tilde{m}(1,2)$	$\tilde{m}(3)$	gauginos	higgsinos	m_h	P_μ
SM	-	-	-	-	-	$7\cdot 10^{-27}$
CMSSM ($\Delta_{EW}=2641$)	~ 1	~ 1	~ 1	~ 1	$0.1-0.13$	$5\cdot 10^{-3}$
PeV SUSY	$\sim 10^3$	$\sim 10^3$	~ 1	$1-10^3$	$0.125-0.155$	$5\cdot 10^{-6}$
Split SUSY	$\sim 10^6$	$\sim 10^6$	~ 1	~ 1	$0.13-0.155$	$7\cdot 10^{-12}$
HS-SUSY	$\gtrsim 10^2$	$\gtrsim 10^2$	$\gtrsim 10^2$	$\gtrsim 10^2$	$0.125-0.16$	$6\cdot 10^{-4}$
Spread ($\tilde{h}$LSP)	10^5	10^5	10^2	~ 1	$0.125-0.15$	$9\cdot 10^{-10}$
Spread ($\tilde{w}$LSP)	10^3	10^3	~ 1	$\sim 10^2$	$0.125-0.14$	$5\cdot 10^{-6}$
Mini-Split ($\tilde{h}$LSP)	$\sim 10^4$	$\sim 10^4$	$\sim 10^2$	~ 1	$0.125-0.14$	$8\cdot 10^{-8}$
Mini-Split ($\tilde{w}$LSP)	$\sim 10^2$	$\sim 10^2$	~ 1	$\sim 10^2$	$0.11-0.13$	$4\cdot 10^{-4}$
SUN-SUSY	$\sim 10^2$	$\sim 10^2$	~ 1	$\sim 10^2$	0.125	$4\cdot 10^{-4}$
G_2MSSM	$30-100$	$30-100$	~ 1	~ 1	$0.11-0.13$	$2\cdot 10^{-3}$
RNS/landscape	$5-40$	$0.5-3$	~ 1	$0.1-0.35$	$0.123-0.126$	1.4

Table 1: A survey of some unnatural and natural SUSY models along with general expectations for sparticle and Higgs mass spectra in TeV units. We also show relative probability measure P_μ for the model to emerge from the landscape. For RNS, we take $\mu_{min}=10$ GeV.

GeV via maximal stop mixing [11]. For comparison, we also show the orange histogram for $n=-1$ where soft terms are equally favored at any mass scale. Here, the distribution peaks at $m_h\sim 118$ GeV with hardly any probability at $m_h\sim 125$ GeV.

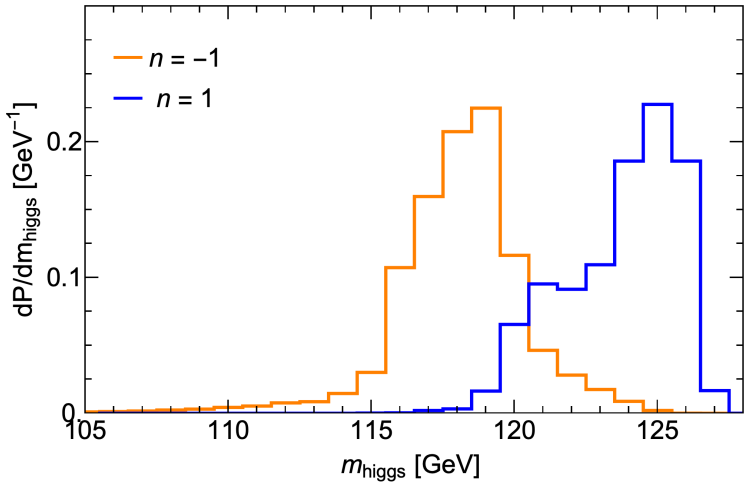


Figure 5: Probability distributions for the light Higgs scalar mass m_h from the $f_{SUSY}=m_{soft}^{\pm 1}$ distributions of soft terms in the string landscape with $\mu=150$ GeV.

In Fig. 6, we show the corresponding probability distribution for the gluino mass. Here, for $n=1$ the curve begins around $m_{\tilde{g}}\sim 1$ TeV and reaches a broad maximum around 3-4 TeV, while petering out beyond $m_{\tilde{g}}\sim 6$ TeV. The present LHC Run 2 limit from ATLAS/CMS [62,63] is $m_{\tilde{g}}\gtrsim 2.2$ TeV from searches within the simplified model context. From the plot, we see that

LHC is only beginning to probe the expected range of $m_{\tilde{g}}$ values from the landscape.

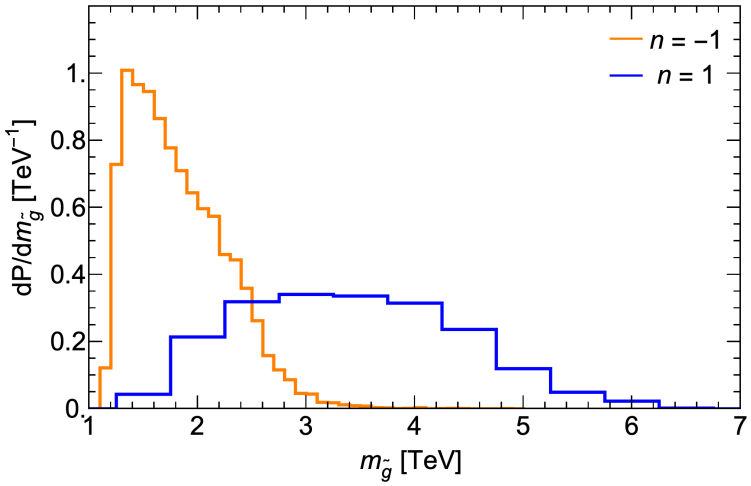


Figure 6: Probability distribution for $m_{\tilde{g}}$ from the $f_{SUSY} = m_{soft}^{\pm 1}$ distributions of soft terms in the string landscape with $\mu = 150$ GeV.

Other sparticle and Higgs mass distributions from the landscape are shown in Ref's [14,64], and they are typically beyond or even well-beyond present LHC limits. For instance, light top-squarks are expected around $m_{\tilde{t}_1} \sim 1 - 2.5$ TeV whilst first/second generation squarks and sleptons are expected near $m_{\tilde{q},\tilde{\ell}} \sim 10 - 30$ TeV. From this point of view, LHC is at present seeing exactly what the string landscape predicts.

5 Conclusions

Theoretical arguments suggest many models which include a remnant spacetime SUSY to populate the string landscape of $4-d$ vacua. We assume a friendly neighborhood of the landscape populated by the MSSM as the LE-EFT, but where the CC and also the soft SUSY breaking gaugino masses, scalar mases and A-terms scan via a power-law draw to large values. Landscape selection of soft terms then allows for a derived value of the weak scale which must lie within the ABDS window in order for the atomic principle to be obeyed, leading to complex nuclei and hence atoms which are needed for complexity.

Under the landscape selection of soft SUSY breaking terms, one expects radiative natural SUSY, or RNS, to be much more prevalent than finetuned SUSY models such as CMSSM, G2MSSM, high scale SUSY, split SUSY or minisplit SUSY. This is evident because in RNS, where all contributions to the weak scale lie within the ABDS window, there is a much larger volume of scan space leading to $m_{weak} \in ABDS$. Alternatively, if even one contribution to the weak scale lies outside the ABDS window, then the remaining volume of parameter space leading to $m_{weak} \in ABDS$ shrinks to tiny values, and is relatively less likely. This is borne out by toy simulations of the string landscape and also allows for a relative probability measure P_μ for different models to emerge from the landscape. For instance, $P_\mu(RNS) \sim 1.4$ compared to

for instance $P_\mu(HS\text{-}SUSY) \sim 6 \times 10^{-4}$. Finally, we show probability distributions of the light Higgs mass and gluinos, showing that the present LHC is seeing what one would expect from the string landscape. New SUSY signals especially from higgsino pair production could arise within the next few years at LHC. With all these beautiful results, we anticipate that Paul will begin to work on landscape SUSY as well ;).

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DE-SC-0009956 and DE-SC-0017647. VB gratefully acknowledges support from the William F. Vilas estate.

References

[1] M. B. Green, J. H. Schwarz, E. Witten, SUPERSTRING THEORY. VOL. 1: INTRODUCTION, Cambridge Monographs on Mathematical Physics, 1988.

[2] M. B. Green, J. H. Schwarz, E. Witten, SUPERSTRING THEORY. VOL. 2: LOOP AMPLITUDES, ANOMALIES AND PHENOMENOLOGY, 1988.

[3] J. Polchinski, String theory. Vol. 1: An introduction to the bosonic string, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2007. doi:10.1017/CBO9780511816079.

[4] J. Polchinski, String theory. Vol. 2: Superstring theory and beyond, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2007. doi:10.1017/CBO9780511618123.

[5] A. Hebecker, Naturalness, String Landscape and Multiverse: A Modern Introduction with Exercises, Vol. 979 of Lecture Notes in Physics, 2021. doi:10.1007/978-3-030-65151-0.

[6] P. Candelas, G. T. Horowitz, A. Strominger, E. Witten, Vacuum configurations for superstrings, Nucl. Phys. B 258 (1985) 46–74. doi:10.1016/0550-3213(85)90602-9.

[7] R. Bousso, J. Polchinski, Quantization of four form fluxes and dynamical neutralization of the cosmological constant, JHEP 06 (2000) 006. arXiv:hep-th/0004134, doi:10.1088/1126-6708/2000/06/006.

[8] S. Ashok, M. R. Douglas, Counting flux vacua, JHEP 01 (2004) 060. arXiv:hep-th/0307049, doi:10.1088/1126-6708/2004/01/060.

[9] S. Weinberg, Anthropic Bound on the Cosmological Constant, Phys. Rev. Lett. 59 (1987) 2607. doi:10.1103/PhysRevLett.59.2607.

[10] L. Susskind, The Anthropic landscape of string theory (2003) 247–266arXiv:hep-th/0302219.

[11] H. Baer, V. Barger, P. Huang, A. Mustafayev, X. Tata, Radiative natural SUSY with a 125 GeV Higgs boson, Phys. Rev. Lett. 109 (2012) 161802. arXiv:1207.3343, doi:10.1103/PhysRevLett.109.161802.

[12] H. Baer, V. Barger, P. Huang, D. Mickelson, A. Mustafayev, X. Tata, Radiative natural supersymmetry: Reconciling electroweak fine-tuning and the Higgs boson mass, Phys. Rev. D 87 (11) (2013) 115028. arXiv:1212.2655, doi:10.1103/PhysRevD.87.115028.

[13] H. Baer, V. Barger, S. Salam, Naturalness versus stringy naturalness (with implications for collider and dark matter searches, Phys. Rev. Research. 1 (2019) 023001. arXiv:1906.07741, doi:10.1103/PhysRevResearch.1.023001.

[14] H. Baer, V. Barger, H. Serce, K. Sinha, Higgs and superparticle mass predictions from the landscape, JHEP 03 (2018) 002. arXiv:1712.01399, doi:10.1007/JHEP03(2018)002.

[15] H. Baer, V. Barger, P. Huang, Hidden SUSY at the LHC: the light higgsino-world scenario and the role of a lepton collider, JHEP 11 (2011) 031. arXiv:1107.5581, doi:10.1007/JHEP11(2011)031.

[16] Z. Han, G. D. Kribs, A. Martin, A. Menon, Hunting quasidegenerate Higgsinos, Phys. Rev. D 89 (7) (2014) 075007. arXiv:1401.1235, doi:10.1103/PhysRevD.89.075007.

[17] H. Baer, A. Mustafayev, X. Tata, Monojet plus soft dilepton signal from light higgsino pair production at LHC14, Phys. Rev. D 90 (11) (2014) 115007. arXiv:1409.7058, doi:10.1103/PhysRevD.90.115007.

[18] C. Han, D. Kim, S. Munir, M. Park, Accessing the core of naturalness, nearly degenerate higgsinos, at the LHC, JHEP 04 (2015) 132. arXiv:1502.03734, doi:10.1007/JHEP04(2015)132.

[19] H. Baer, V. Barger, D. Sengupta, X. Tata, New angular and other cuts to improve the Higgsino signal at the LHC, Phys. Rev. D 105 (9) (2022) 095017. arXiv:2109.14030, doi:10.1103/PhysRevD.105.095017.

[20] G. Aad, et al., Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, Phys. Rev. D 101 (5) (2020) 052005. arXiv:1911.12606, doi:10.1103/PhysRevD.101.052005.

[21] A. Tumasyan, et al., Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP 04 (2022) 091. arXiv:2111.06296, doi:10.1007/JHEP04(2022)091.

[22] D. Agin, B. Fuks, M. D. Goodsell, T. Murphy, Monojets reveal overlapping excesses for light compressed higgsinos (11 2023). arXiv:2311.17149.

[23] H. Baer, V. Barger, M. Savoy, Upper bounds on sparticle masses from naturalness or how to disprove weak scale supersymmetry, Phys. Rev. D 93 (3) (2016) 035016. arXiv:1509.02929, doi:10.1103/PhysRevD.93.035016.

[24] H. Baer, V. Barger, D. Martinez, S. Salam, Practical naturalness and its implications for weak scale supersymmetry, Phys. Rev. D 108 (3) (2023) 035050. arXiv:2305.16125, doi:10.1103/PhysRevD.108.035050.

[25] B. S. Acharya, Supersymmetry, Ricci Flat Manifolds and the String Landscape, JHEP 08 (2020) 128. arXiv:1906.06886, doi:10.1007/JHEP08(2020)128.

[26] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, A. Strumia, Higgs mass and vacuum stability in the Standard Model at NNLO, JHEP 08 (2012) 098. arXiv:1205.6497, doi:10.1007/JHEP08(2012)098.

[27] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio, A. Strumia, Investigating the near-criticality of the Higgs boson, JHEP 12 (2013) 089. arXiv:1307.3536, doi:10.1007/JHEP12(2013)089.

[28] M. Dine, G. Festuccia, A. Morisse, K. van den Broek, Metastable Domains of the Landscape, JHEP 06 (2008) 014. arXiv:0712.1397, doi:10.1088/1126-6708/2008/06/014.

[29] M. Dine, G. Festuccia, A. Morisse, The Fate of Nearly Supersymmetric Vacua, JHEP 09 (2009) 013. arXiv:0901.1169, doi:10.1088/1126-6708/2009/09/013.

[30] M. Dine, J. D. Mason, Supersymmetry and Its Dynamical Breaking, Rept. Prog. Phys. 74 (2011) 056201. arXiv:1012.2836, doi:10.1088/0034-4885/74/5/056201.

[31] H. Baer, V. Barger, D. Mickelson, M. Padeffke-Kirkland, SUSY models under siege: LHC constraints and electroweak fine-tuning, Phys. Rev. D 89 (11) (2014) 115019. arXiv:1404.2277, doi:10.1103/PhysRevD.89.115019.

[32] S. Dimopoulos, S. Raby, F. Wilczek, Supersymmetry and the Scale of Unification, Phys. Rev. D 24 (1981) 1681–1683. doi:10.1103/PhysRevD.24.1681.

[33] L. E. Ibanez, G. G. Ross, SU(2)-L x U(1) Symmetry Breaking as a Radiative Effect of Supersymmetry Breaking in Guts, Phys. Lett. B 110 (1982) 215–220. doi:10.1016/0370-2693(82)91239-4.

[34] P. Slavich, et al., Higgs-mass predictions in the MSSM and beyond, Eur. Phys. J. C 81 (5) (2021) 450. arXiv:2012.15629, doi:10.1140/epjc/s10052-021-09198-2.

[35] S. Heinemeyer, W. Hollik, D. Stockinger, A. M. Weber, G. Weiglein, Precise prediction for M(W) in the MSSM, JHEP 08 (2006) 052. arXiv:hep-ph/0604147, doi:10.1088/1126-6708/2006/08/052.

[36] H. Baer, V. Barger, D. Sengupta, Landscape solution to the SUSY flavor and CP problems, Phys. Rev. Res. 1 (3) (2019) 033179. arXiv:1910.00090, doi:10.1103/PhysRevResearch.1.033179.

[37] N. Arkani-Hamed, S. Dimopoulos, S. Kachru, Predictive landscapes and new physics at a TeV (1 2005). arXiv:hep-th/0501082.

[38] V. Agrawal, S. M. Barr, J. F. Donoghue, D. Seckel, Anthropic considerations in multiple domain theories and the scale of electroweak symmetry breaking, Phys. Rev. Lett. 80 (1998) 1822–1825. arXiv:hep-ph/9801253, doi:10.1103/PhysRevLett.80.1822.

[39] M. R. Douglas, Statistical analysis of the supersymmetry breaking scale (5 2004). arXiv:hep-th/0405279.

[40] S. K. Soni, H. A. Weldon, Analysis of the Supersymmetry Breaking Induced by N=1 Supergravity Theories, Phys. Lett. B 126 (1983) 215–219. doi:10.1016/0370-2693(83)90593-2.

[41] V. S. Kaplunovsky, J. Louis, Model independent analysis of soft terms in effective supergravity and in string theory, Phys. Lett. B 306 (1993) 269–275. arXiv:hep-th/9303040, doi:10.1016/0370-2693(93)90078-V.

[42] A. Brignole, L. E. Ibanez, C. Munoz, Towards a theory of soft terms for the supersymmetric Standard Model, Nucl. Phys. B 422 (1994) 125–171, [Erratum: Nucl.Phys.B 436, 747–748 (1995)]. arXiv:hep-ph/9308271, doi:10.1016/0550-3213(94)00068-9.

[43] H. Baer, V. Barger, S. Salam, D. Sengupta, String landscape guide to soft SUSY breaking terms, Phys. Rev. D 102 (7) (2020) 075012. arXiv:2005.13577, doi:10.1103/PhysRevD.102.075012.

[44] H. P. Nilles, P. K. S. Vaudrevange, Geography of Fields in Extra Dimensions: String Theory Lessons for Particle Physics, Mod. Phys. Lett. A 30 (10) (2015) 1530008. arXiv:1403.1597, doi:10.1142/S0217732315300086.

[45] J. R. Ellis, T. Falk, K. A. Olive, Y. Santoso, Exploration of the MSSM with nonuniversal Higgs masses, Nucl. Phys. B 652 (2003) 259–347. arXiv:hep-ph/0210205, doi:10.1016/S0550-3213(02)01144-6.

[46] H. Baer, A. Mustafayev, S. Profumo, A. Belyaev, X. Tata, Direct, indirect and collider detection of neutralino dark matter in SUSY models with non-universal Higgs masses, JHEP 07 (2005) 065. arXiv:hep-ph/0504001, doi:10.1088/1126-6708/2005/07/065.

[47] H. Baer, V. Barger, D. Sengupta, Anomaly mediated SUSY breaking model retrofitted for naturalness, Phys. Rev. D 98 (1) (2018) 015039. arXiv:1801.09730, doi:10.1103/PhysRevD.98.015039.

[48] H. Baer, V. Barger, J. Bolich, J. Dutta, D. Sengupta, Natural anomaly mediation from the landscape with implications for LHC SUSY searches, Phys. Rev. D 109 (3) (2024) 035011. arXiv:2311.18120, doi:10.1103/PhysRevD.109.035011.

[49] H. Baer, V. Barger, H. Serce, X. Tata, Natural generalized mirage mediation, Phys. Rev. D 94 (11) (2016) 115017. arXiv:1610.06205, doi:10.1103/PhysRevD.94.115017.

[50] V. Agrawal, S. M. Barr, J. F. Donoghue, D. Seckel, Viable range of the mass scale of the standard model, Phys. Rev. D 57 (1998) 5480–5492. arXiv:hep-ph/9707380, doi:10.1103/PhysRevD.57.5480.

[51] H. Baer, V. Barger, D. Martinez, S. Salam, Radiative natural supersymmetry emergent from the string landscape, JHEP 03 (2022) 186. arXiv:2202.07046, doi:10.1007/JHEP03(2022)186.

[52] H. Baer, V. Barger, D. Martinez, Comparison of SUSY spectra generators for natural SUSY and string landscape predictions, Eur. Phys. J. C 82 (2) (2022) 172. arXiv:2111.03096, doi:10.1140/epjc/s10052-022-10141-2.

[53] N. Arkani-Hamed, S. Dimopoulos, Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC, JHEP 06 (2005) 073. arXiv:hep-th/0405159, doi:10.1088/1126-6708/2005/06/073.

[54] V. Barger, J. Jiang, P. Langacker, T. Li, Non-canonical gauge coupling unification in high-scale supersymmetry breaking, Nucl. Phys. B 726 (2005) 149–170. arXiv:hep-ph/0504093, doi:10.1016/j.nuclphysb.2005.08.007.

[55] A. Arvanitaki, N. Craig, S. Dimopoulos, G. Villadoro, Mini-Split, JHEP 02 (2013) 126. arXiv:1210.0555, doi:10.1007/JHEP02(2013)126.

[56] H. Baer, V. Barger, D. Martinez, S. Salam, Fine-tuned vs. natural supersymmetry: what does the string landscape predict?, JHEP 09 (2022) 125. arXiv:2206.14839, doi:10.1007/JHEP09(2022)125.

[57] G. L. Kane, C. F. Kolda, L. Roszkowski, J. D. Wells, Study of constrained minimal supersymmetry, Phys. Rev. D 49 (1994) 6173–6210. arXiv:hep-ph/9312272, doi:10.1103/PhysRevD.49.6173.

[58] B. S. Acharya, K. Bobkov, G. L. Kane, J. Shao, P. Kumar, The G(2)-MSSM: An M Theory motivated model of Particle Physics, Phys. Rev. D 78 (2008) 065038. arXiv:0801.0478, doi:10.1103/PhysRevD.78.065038.

[59] V. Barger, N. G. Deshpande, J. Jiang, P. Langacker, T. Li, Implications of Canonical Gauge Coupling Unification in High-Scale Supersymmetry Breaking, Nucl. Phys. B 793 (2008) 307–325. arXiv:hep-ph/0701136, doi:10.1016/j.nuclphysb.2007.10.013.

[60] L. J. Hall, Y. Nomura, Spread Supersymmetry, JHEP 01 (2012) 082. arXiv:1111.4519, doi:10.1007/JHEP01(2012)082.

[61] F. Vissani, Do experiments suggest a hierarchy problem?, Phys. Rev. D 57 (1998) 7027–7030. arXiv:hep-ph/9709409, doi:10.1103/PhysRevD.57.7027.

[62] G. Aad, et al., Search for squarks and gluinos in final states with one isolated lepton, jets, and missing transverse momentum at $\sqrt{s} = 13$ with the ATLAS detector, Eur. Phys. J. C 81 (7) (2021) 600, [Erratum: Eur.Phys.J.C 81, 956 (2021)]. arXiv:2101.01629, doi:10.1140/epjc/s10052-021-09748-8.

[63] T. C. Collaboration, et al., Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum, JHEP 10 (2019) 244. arXiv:1908.04722, doi:10.1007/JHEP10(2019)244.

[64] H. Baer, V. Barger, S. Salam, D. Sengupta, Mini-review: Expectations for supersymmetry from the string landscape, in: Snowmass 2021, 2022. arXiv:2202.11578.

Flavor-Violating ALPs, Electron $g-2$, and the Electron-Ion Collider

Hooman Davoudiasl*,[1] Roman Marcarelli†,[1,2] and Ethan T. Neil‡[2]

[1] *High Energy Theory Group, Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*

[2] *Department of Physics, University of Colorado, Boulder, Colorado 80309, USA*

We revisit the possibility that light axion-like particles (ALPs) with lepton flavor violating couplings could give significant contributions to the electron's anomalous magnetic moment $g_e - 2$. Unlike flavor diagonal lepton-ALP couplings, which are exclusively axial, lepton flavor violating couplings can have arbitrary chirality. Focusing on the e-τ ALP coupling, we find that the size of the contribution to $g_e - 2$ depends strongly on the chirality of the coupling. A significant part of the parameter space for which such a coupling can explain experimental anomalies in $g_e - 2$ can be probed at the Electron-Ion Collider, which is uniquely sensitive to the chirality of the coupling using the polarization of the electron beam.

arXiv:2402.17821v1 [hep-ph] 27 Feb 2024

I. INTRODUCTION

Axion-like particles (ALPs) are a generic type of new particle that can occur in a wide variety of scenarios for physics beyond the Standard Model (BSM). Such particles commonly arise as pseudo-Nambu-Goldstone bosons associated with new physics [1, 2]. Their pseudo-Goldstone nature means that ALPs can naturally be very light compared to the scale of new physics, with very weak interactions suppressed by the same large mass scale.

Within the general theory space, an intriguing possibility is the presence of significant lepton-flavor-violating (LFV) couplings for an ALP. Lepton flavor is almost perfectly conserved in the Standard Model (SM), up to neutrino flavor mixing, so that lepton flavor violation is a sensitive probe of new physics. At the same time, existing anomalies in precision lepton physics, in particular in the anomalous magnetic moments of the electron and muon, motivate the exploration of new particles coupled to leptons. Previous studies of the phenomenology of LFV ALPs include [3–12].

While there is significant attention on present tensions between theory and experiment in the anomalous magnetic moment of the muon [13], the anomalous magnetic moment of the electron g_e-2 is also showing possible tensions with the SM, although the situation is somewhat less clear. The latest measurement of $g_e - 2$ [14] is in tension with the SM prediction of either $+2.2\sigma$ or -3.7σ [15], depending on whether the fine-structure constant α obtained from Rb [16] or Cs [17] is used as input.

Contributions from lepton-flavor violating ALP couplings to $g_e - 2$ have been calculated and studied in the literature previously [5, 6, 10]. In this work, we focus on the interplay between the magnitude of the e-τ ALP coupling and its chirality, which has a significant impact on the $g_e - 2$ contribution. We find a substantial region of parameter space where a single e-τ coupling with a GeV-scale ALP and the correct chirality can explain the current experimental anomalies in $g_e - 2$ without being constrained by other experiments, but with discovery potential in future searches at the EIC [9]. The EIC is also uniquely well-equipped to probe the chiral structure of such a coupling directly by utilizing the polarization of its electron beam.

II. MODEL SETUP

We will consider an ALP-e-τ interaction of the form

$$\mathcal{L}_{\rm int} = \frac{\partial_\mu a}{\Lambda}\bar{\tau}\gamma^\mu\left(V_{\tau e} + A_{\tau e}\gamma_5\right)e + \text{H.C.}, \tag{1}$$

where $V_{\tau e}$ and $A_{\tau e}$ are complex constants in general. We can decompose this term into a more useful form by defining

$$\begin{aligned} \phi_V &\equiv \arg V_{\tau e}, & \phi_A &\equiv \arg A_{\tau e} \\ \phi &\equiv \phi_A - \phi_V, & \tan\theta &\equiv \left(|V_{\tau e}|/|A_{\tau e}|\right), \\ C_{\tau e} &\equiv \sqrt{|A_{\tau e}|^2 + |V_{\tau e}|^2}. \end{aligned} \tag{2}$$

Then, the interaction can be written as

$$\mathcal{L}_{\rm int} = \frac{C_{\tau e}e^{i\phi_V}}{\Lambda}\partial_\mu a\,\bar{\tau}\gamma^\mu\left(\sin\theta + e^{i\phi}\cos\theta\gamma_5\right)e + \text{H.C.} \tag{3}$$

Note that if the coupling were lepton flavor diagonal, hermiticity would require $\theta = \phi_A = \phi_V = 0$. Hence, the presence of flavor violation allows for both parity (P) violation (via θ) and charge/parity (CP) violation (via ϕ and ϕ_V). The overall phase ϕ_V does not contribute to any of the physical processes we consider below.

For simplicity, we will study the case in which the only significant coupling is $C_{\tau e}$. As explored in Ref. [8], other bounds from precision flavor physics are relevant for ALPs heavier than m_τ only when lepton-flavor diagonal couplings are also present and sufficiently strong.

*email: hooman@bnl.gov

†email: roman.marcarelli@colorado.edu

‡email: ethan.neil@colorado.edu

III. CONTRIBUTIONS TO ELECTRON EDM AND $g-2$

LFV ALPs which couple to electrons and taus can contribute to the electric and magnetic dipole moments of the electron. The electric and magnetic dipole moment can be extracted from the electron-photon interaction vertex Γ^μ, which can be decomposed into the general form [18]

$$\Gamma^\mu = \gamma^\mu F_1(q^2) + \frac{i\sigma^{\mu\nu}}{2m_e} q_\nu F_2(q^2) + \frac{i\sigma^{\mu\nu}}{2m_e} q_\nu \gamma_5 F_3(q^2) + \frac{1}{2m_e}\left(q^\mu - \frac{q^2}{2m_e}\gamma^\mu\right)\gamma_5 F_4(q^2), \quad (4)$$

where $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$, $q = p_2 - p_1$ is the external photon momentum, p_1 is the initial momentum of the electron, p_2 is the final momentum of the electron; for an on-shell amplitude $q^2 = 0$. The form factors F_i are all real [18]. Each form factor corresponds to a property of the electron: F_1 yields the electric charge, F_2 the magnetic dipole moment, F_3 the electric dipole moment, and F_4 the anapole moment. In particular, the electric and magnetic moments are given by

$$a_e = \frac{g_e - 2}{2} = F_2(0); \qquad d_e = -\frac{e}{2m_e} F_3(0). \quad (5)$$

In the limit $m_\tau \gg m_e$, the contributions to these dipole moments from the interaction (3) are

$$\Delta a_e = -\frac{m_e^2 C_{\tau e}^2}{16\pi^2\Lambda^2}\left(f(x_\tau) + \frac{m_\tau}{m_e} g(x_\tau)\cos 2\theta\right), \quad (6)$$

$$\Delta d_e = -\frac{e}{2m_e}\frac{m_\tau m_e C_{\tau e}^2}{16\pi^2\Lambda^2} g(x_\tau)\sin 2\theta \sin\phi, \quad (7)$$

where $x_\tau \equiv m_a^2/m_\tau^2$ and f and g are given by

$$f(x) = \frac{2x^2(2x-1)}{(x-1)^4}\log x - \frac{5 - 19x + 20x^2}{3(x-1)^3}, \quad (8)$$

$$g(x) = \frac{2x^2}{(x-1)^3}\log x + \frac{1-3x}{(x-1)^2}. \quad (9)$$

These results were obtained using PackageX [19] in Mathematica and are in agreement with Refs. [5, 6] in the appropriate limits. Our result for Δd_e reproduces the formula from Ref. [20] for the contribution from flavor-violating lepton-ALP couplings.

The first term in the contribution to the magnetic dipole moment (6) is often neglected for $m_\tau \gg m_e$, but we emphasize that this is not the case for chiral interactions, where $\theta = \pm\pi/4$. In particular, we find that for chiral or near-chiral interactions, the form factor F_2 is suppressed by m_e/m_τ but is still non-zero. The electric dipole moment, on the other hand, is proportional to $\sin 2\theta \sin\phi$, so it is only non-zero when there is CP violation ($\theta \neq 0$ and $\phi \neq 0$).

For $m_a = 10$ GeV, as a representative value for our study, we find that $\phi \lesssim 10^{-7}$ if $C_{\tau e} = 1$, $\Lambda = 1$ TeV, and θ is near $\pi/4$, in light of the experimental bound $|d_e| < 4.1 \times 10^{-30}$ e cm [21]. Hence, it is justified to assume $\phi = 0$ for our analysis, given the very stringent limit from the electron EDM. For a more detailed discussion, see Refs. [10, 20, 22].

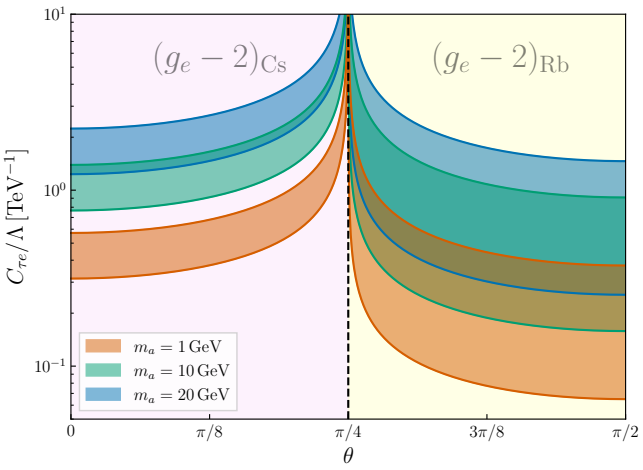


FIG. 1: 2σ explanations for Δa_e(Cs) (pink-shaded region, $\theta \lesssim \pi/4$) and for Δa_e(Rb) (yellow-shaded region, $\theta \gtrsim \pi/4$) at $m_a = 1$, 10 and 20 GeV. Which anomaly is explained is dependent on the sign of Δa_e, which in turn is dependent on the parity-violating angle θ. To good approximation, the sign flips at $\theta = \pi/4$, but there are $O(m_e/m_\tau) \sim 10^{-4}$ corrections. Digitized plot data for this and other figures are available as ancillary files on arXiv.org.

Notably, the sign of Δa_e (Eq. (6)) is dependent on the value of the parity-violating angle θ; the shift is negative for $\cos 2\theta \gtrsim 0$ and positive for $\cos 2\theta \lesssim 0$ (up to factors of m_e/m_τ). This is interesting in light of the most recent experimental determination of a_e, where the sign of the discrepancy with the SM depends on which experimental measurement of the fine-structure constant α is used. In particular, if one uses the measurement from rubidium α(Rb) [16], there is a $+2.2\sigma$ deviation

$$\Delta a_e({\rm Rb}) = (34 \pm 16) \times 10^{-14}, \quad (10)$$

whereas if one uses the measurement from cesium α(Cs) [17], there is a -3.7σ deviation

$$\Delta a_e({\rm Cs}) = (-101 \pm 27) \times 10^{-14} \quad (11)$$

from the SM expectation. A flavor-violating ALP interaction of the form considered here can explain an anomalous contribution to Δa_e of either sign.

IV. RESULTS

It is straightforward to compute the value of $C_{\tau e}$ that would explain one of these anomalies for a given angle θ and ALP mass m_a. The result is shown in Fig. 1 for $m_a = 1$, 10, and 20 GeV and $0 < \theta < \pi$. Explanations

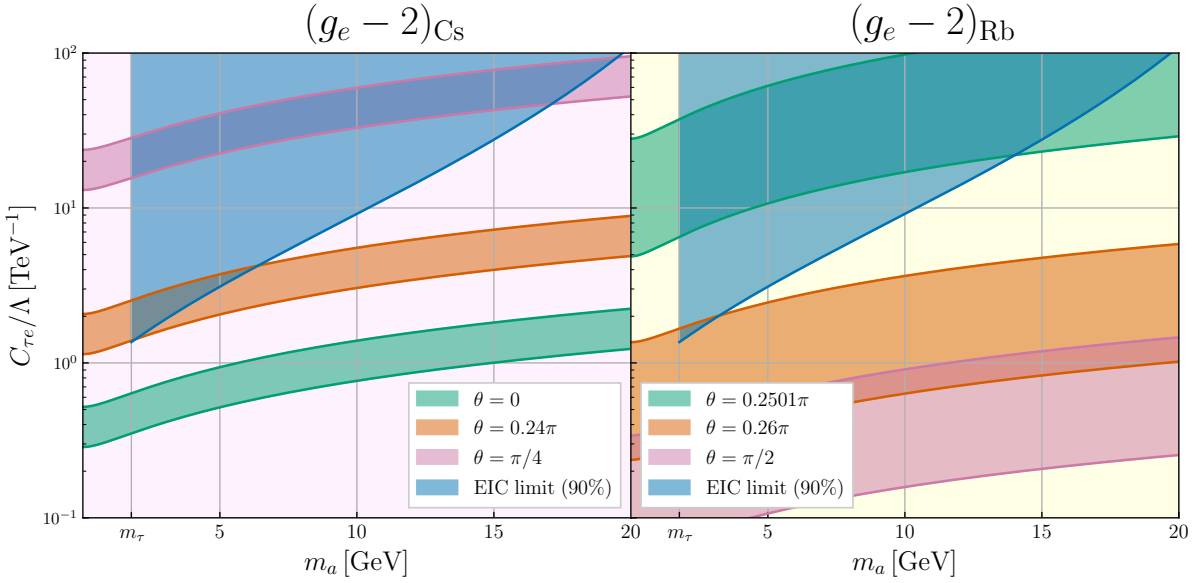


FIG. 2: 2σ explanations for Δa_e(Cs) anomaly (left, pink-shaded region) and Δa_e(Rb) anomaly (right, yellow-shaded region). Limits placed on $C_{e\tau}$ at the EIC at 90% from Ref. [8] are shown in blue. When the ALP e-τ coupling is chiral or near-chiral ($\theta \approx \pi/4$), the EIC can probe an explanation for either the cesium or rubidium anomaly.

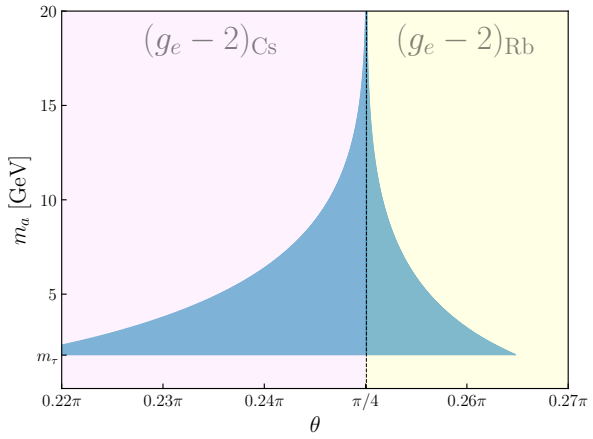


FIG. 3: The masses m_a and parity-violating angles θ for which the EIC can probe a 2σ explanation to the $g_e - 2$ anomaly using α(Cs) (pink shaded region) and α(Rb) (yellow shaded region).

for Δa_e(Rb) (positive) are in the yellow-shaded region whereas explanations for Δa_e(Cs) (negative) are in the pink-shaded region. As the figure shows, explanations for the Δa_e anomaly can be found for $\mathcal{O}(1)$ or smaller $C_{\tau e}$ with $\Lambda \sim 1$ TeV over a wide range of ALP masses and PV angles. Notably, the magnitude of the coupling required to explain either anomaly grows for a near-chiral coupling ($|\theta| \to \pi/4$).

As detailed in Ref. [9], the EIC is in a unique position to probe the coupling $C_{\tau e}$, especially when the diagonal ALP couplings are relatively small. Chiral new physics is of particular interest at the EIC due to its polarization capabilities, and is well-motivated due to the asymmetry between left- and right-handed fields already present in the SM. For an example of a UV BSM setup which has ALPs with chiral LFV interactions, see Ref. [23].

In Fig. 2, we plot the 90% confidence level projected exclusions on $C_{\tau e}$ from the EIC with $\mathcal{L} = (100/A)$ fb^{-1}, discussed in Ref. [8], and the corresponding explanations to the cesium anomaly (left, pink-shaded background) and the rubidium anomaly (right, yellow-shaded background). We see that the EIC is able to probe near-chiral ALP-lepton couplings which offer explanations to either $g_e - 2$ anomaly. With a better τ efficiency and a dedicated background analysis to improve on the limits from Ref. [9], the EIC may be able to probe a wider range of parity violating (PV) couplings.

If the PV angle is near-chiral, the EIC's polarization control of the electron beam may allow determination of the angle θ. The EIC is expected to achieve over 70% polarization of the electron beam [24]. Given production cross-sections dependent on the electron beam polarization σ_L and σ_R, and a polarization fraction p, the left-right asymmetry observable at the EIC is

$$r_{LR} = \frac{[p\sigma_L + (1-p)\sigma_R] - [(1-p)\sigma_L + p\sigma_R]}{\sigma_L + \sigma_R} \tag{12}$$

$$= (2p-1)\frac{\sigma_L - \sigma_R}{\sigma_L + \sigma_R}. \tag{13}$$

For this ALP model with PV angle θ, this corresponds to a left-right asymmetry

$$r_{LR}(\theta) = (2p-1)\sin 2\theta. \tag{14}$$

This is extremal at $\theta = \pm\pi/4$. If a flavor-violating ALP signal were to be discovered at the EIC, probing the left-right asymmetry of the signal to determine the PV angle θ would give an opportunity to directly verify the connection between such a signal and its contribution to $g_e - 2$.

V. CONCLUDING REMARKS

Axion-like particles (ALPs) can arise in a variety of models and may, on general grounds, have flavor violating interactions. In this work, we considered lepton flavor violation and in particular $e - \tau$ transitions mediated by ALPs. We also accounted for the possibility that ALPs can have parity violating interactions, akin to the electroweak sector of the SM. With these assumptions, we found that ALPs can explain present hints for electron $g_e - 2$ deviations whose sign depends on the choice of the precision value of the fine structure constant α that is used as input for the theory prediction.

It is not yet clear which, if any, of the $g_e - 2$ deviations would survive once the experimental extraction of α is disambiguated. However, assuming that the size of the possible deviation remains at the current levels, we showed that the EIC can constrain the requisite $e - \tau$ coupling, for nearly chiral interactions and ALP masses above m_τ, up to ~ 20 GeV. Probing the smaller $e - \tau$ coupling which is needed to explain $g_e - 2$ deviations for PV angles θ away from $\pi/4$ will require improvements to the EIC search, or complementary searches in other future experiments for $e - \tau$ FV ALPs.

In addition, the e-beam polarization capabilities of the EIC can provide a unique probe of the parity violating angle that parameterizes departures from the purely chiral limit, over the above mass range. We noted that ALP couplings can also include a CP violating phase, but the severe bounds on the electric dipole moment of the electron imply that such a phase must be extremely tiny, at the level of $\mathcal{O}(10^{-7})$ or less.

Open questions, like the identity of dark matter, make a compelling case for new physics, yet allow a vast array of possibilities for where it may reside. This motivates consideration of a broad set of ideas and the venues where they could be tested. The capabilities of the EIC can lend themselves to searches for some of the signals that arise in well-motivated frameworks, like that of the ALP interactions studied here; further work along this direction hence seems warranted.

Acknowledgements

We thank Marvin Schnubel and Sebastian Trojanowski for discussions. This work is supported by the U.S. Department of Energy under Grant Contracts DE-SC0012704 (H. D.) and DE-SC0010005 (E. N. and R. M.). This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664.

[1] J. Jaeckel and A. Ringwald, Ann. Rev. Nucl. Part. Sci. **60**, 405 (2010), 1002.0329.
[2] I. G. Irastorza and J. Redondo, Prog. Part. Nucl. Phys. **102**, 89 (2018), 1801.08127.
[3] G. B. Gelmini, S. Nussinov, and T. Yanagida, Nucl. Phys. B **219**, 31 (1983).
[4] K. Tsumura and L. Velasco-Sevilla, Phys. Rev. D **81**, 036012 (2010), 0911.2149.
[5] M. Bauer, M. Neubert, S. Renner, M. Schnubel, and A. Thamm, Phys. Rev. Lett. **124**, 211803 (2020), 1908.00008.
[6] C. Cornella, P. Paradisi, and O. Sumensari, JHEP **01**, 158 (2020), 1911.06279.
[7] M. Endo, S. Iguro, and T. Kitahara, JHEP **06**, 040 (2020), 2002.05948.
[8] H. Davoudiasl, R. Marcarelli, N. Miesch, and E. T. Neil, Phys. Rev. D **104**, 055022 (2021), 2105.05866.
[9] H. Davoudiasl, R. Marcarelli, and E. T. Neil, JHEP **02**, 071 (2023), 2112.04513.
[10] M. Bauer, M. Neubert, S. Renner, M. Schnubel, and A. Thamm, JHEP **09**, 056 (2022), 2110.10698.
[11] Y. Jho, S. Knapen, and D. Redigolo, JHEP **10**, 029 (2022), 2203.11222.
[12] S. Knapen, K. Langhoff, T. Opferkuch, and D. Redigolo (2023), 2311.17915.
[13] T. Aoyama et al., Phys. Rept. **887**, 1 (2020), 2006.04822.
[14] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse, Phys. Rev. Lett. **130**, 071801 (2023), 2209.13084.
[15] H. Davoudiasl and P. P. Giardino (2023), 2311.12112.
[16] L. Morel, Z. Yao, P. Cladé, and S. Guellati-Khélifa, Nature **588**, 61 (2020).
[17] R. H. Parker, C. Yu, W. Zhong, B. Estey, and H. Müller, Science **360**, 191 (2018), 1812.04130.
[18] M. Nowakowski, E. A. Paschos, and J. M. Rodriguez, Eur. J. Phys. **26**, 545 (2005), physics/0402058.
[19] H. H. Patel, Comput. Phys. Commun. **218**, 66 (2017), 1612.00009.
[20] L. Di Luzio, R. Gröber, and P. Paradisi, Phys. Rev. D **104**, 095027 (2021), 2010.13760.
[21] T. S. Roussy et al., Science **381**, adg4084 (2023), 2212.11841.
[22] L. Di Luzio, H. Gisbert, G. Levati, P. Paradisi, and P. Sørensen (2023), 2312.17310.
[23] H. Davoudiasl, P. P. Giardino, E. T. Neil, and E. Rinaldi, Phys. Rev. D **96**, 115003 (2017), 1709.01082.
[24] R. Abdul Khalek et al., Nucl. Phys. A **1026**, 122447 (2022), 2103.05419.

arXiv:2402.17820v1 [hep-ph] 27 Feb 2024

An Axion Pulsarscope

Mariia Khelashvili,[1,2,3,*] Mariangela Lisanti,[1,4,†] Anirudh Prabhu,[5,‡] and Benjamin R. Safdi[6,7,§]

[1] *Department of Physics, Princeton University, Princeton, NJ 08544, USA*
[2] *Bogolyubov Institute for Theoretical Physics of the NAS of Ukraine, Metrolohichna Str. 14-b, Kyiv, 03143, Ukraine*
[3] *Goethe Universität, Max-von-Laue Str. 1, Frankfurt am Main, 60438, Germany*
[4] *Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA*
[5] *Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA*
[6] *Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA*
[7] *Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

(Dated: February 29, 2024)

Electromagnetic fields surrounding pulsars may source coherent ultralight axion signals at the known rotational frequencies of the neutron stars, which can be detected by laboratory experiments (*e.g.*, pulsarscopes). As a promising case study, we model axion emission from the well-studied Crab pulsar, which would yield a prominent signal at $f \approx 29.6$ Hz regardless of whether the axion contributes to the dark matter abundance. We estimate the relevant sensitivity of future axion dark matter detection experiments such as DMRadio-GUT, Dark SRF, and CASPEr, assuming different magnetosphere models to bracket the uncertainty in astrophysical modeling. For example, depending on final experimental parameters, the Dark SRF experiment could probe axions with any mass $m_a \ll 10^{-13}$ eV down to $g_{a\gamma\gamma} \sim 3 \times 10^{-13}$ GeV^{-1} with one year of data and assuming the vacuum magnetosphere model. These projected sensitivities may be degraded depending on the extent to which the magnetosphere is screened by charge-filled plasma. The promise of pulsar-sourced axions as a clean target for direct detection experiments motivates dedicated simulations of axion production in pulsar magnetospheres.

Introduction.— The existence of axions is predicted by numerous well-motivated extensions to the Standard Model [1, 2]. While interesting in their own right, these ultralight pseudoscalar bosons can potentially account for the dark matter (DM), as is the case for the quantum chromodynamics (QCD) axion [3–10]. Given their ubiquity in theoretical models, a broad laboratory program exists to search for axions (see [11–13]). This includes haloscope experiments, which search for axion DM in the Milky Way [14–28], helioscope experiments, which search for axions produced by the Sun [29, 30], and "light shining through walls" experiments, which aim to produce axions directly in the lab [31, 32]. This Letter proposes a fourth alternative—the *pulsarscope*—and demonstrates that planned experiments can successfully search for coherent axion signals emitted by nearby pulsars.

Axions are generically expected to couple to Standard Model fields through dimension-five operators, motivating efforts to detect them in the laboratory. For example, the axion field, $a(x)$, couples to electromagnetism through the operator $\mathcal{L} \supset -g_{a\gamma\gamma} a F\tilde{F}/4 = g_{a\gamma\gamma} a \mathbf{E}\cdot\mathbf{B}$, where $g_{a\gamma\gamma}$ is the coupling constant, F is the quantum electrodynamics (QED) field strength, and $\mathbf{E}$ ($\mathbf{B}$) is the electric (magnetic) field. The axion may also couple to Standard Model fermions f through the operator $\mathcal{L} \supset (g_{aff}/2m_f)\partial_\mu a \bar{f}\gamma^\mu\gamma^5 f = -(g_{aff}/m_f)\boldsymbol{\nabla} a \cdot \mathbf{S}$ (in the non-relativistic limit), where $\mathbf{S}$ is the fermion spin operator, m_f is the fermion mass, and g_{aff} is the coupling constant. These operators may induce observable signatures in laboratory experiments, such as axion-to-photon conversion (*e.g.*, [29, 30, 33–38]) or spin precession (*e.g.*, [39–45]).

Rapidly-rotating neutron stars (NSs), or pulsars, can serve as axion factories because their magnetospheres may possess regions of large, un-screened accelerating electric fields ($\mathbf{E}\cdot\mathbf{B} \neq 0$) [46–48]. In non-axisymmetric pulsar magnetospheres, ultralight axions are efficiently radiated at the rotation frequency of the pulsar, Ω [46]. As illustrated in Fig. 1, this relativistic axion signal travels towards Earth, where it may be detected by a ground-based experiment. While their densities are generally much smaller than the DM density near Earth [12], pulsar-sourced axions benefit from large coherence times and have known frequencies.

We estimate the sensitivity of proposed axion DM detection experiments to pulsar-sourced axions. We consider two extreme models of the pulsar magnetosphere to bracket the uncertainty on the expected signal. The leading sensitivity to low-mass axions is obtained from the CASPEr-wind [40–43], Dark SRF [49, 50], and DMRadio-GUT [51, 52] experiments. Most optimistically, these experiments could probe previously unexplored regions of axion parameter space for axion masses $m_a \ll 10^{-13}$ eV; however, if the magnetospheres are heavily screened, then the sensitivities may be subdominant relative to current constraints.

Axion Radiation from Pulsars.— The axion field sourced by a pulsar is determined by the distribution of $\mathbf{E}\cdot\mathbf{B}$ in the magnetosphere according to the Klein-Gordon equation, $(\Box + m_a^2)a(x) = g_{a\gamma\gamma}\mathbf{E}\cdot\mathbf{B}$. In the limit where the axion mass $m_a \to 0$, this equation is analogous to that for the electric potential, interpreted as $a(x)$, in Lorenz gauge with charge density $\rho_{\rm eff} = g_{a\gamma\gamma}\mathbf{E}\cdot\mathbf{B}$. Thus, one can calculate the radiated axion signal given a model

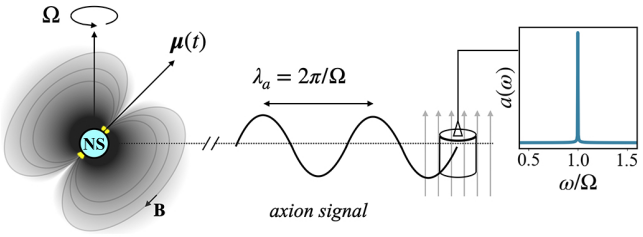


FIG. 1. A neutron star (NS) rotating with frequency Ω emits axions from gap regions where $\mathbf{E}\cdot\mathbf{B}\neq 0$ (shown in grey for the VDM model and in yellow for the PC model, not to scale). Relativistic axions emitted by the NS may convert to detectable signals in laboratory experiments (*pulsarscopes*). The resulting signal is highly coherent, with a prominent peak in frequency at the pulsar rotation frequency, Ω.

for $\mathbf{E}\cdot\mathbf{B}$ exterior to a NS.

We consider two limiting scenarios for $\mathbf{E}\cdot\mathbf{B}$. The first is the Vacuum Dipole Model (VDM) [53–58]. In this case, the exterior of the NS is assumed to be vacuum and the magnetic field is described by a rotating dipole with magnetic moment $\boldsymbol{\mu}(t) = B_0 R_{\rm NS}^3 \hat{\mu}(t)$, where B_0 is the surface magnetic field and $R_{\rm NS}$ the NS radius. The unit vector $\hat{\mu}$ is misaligned from the rotation axis by an angle θ_m and rotates around the axis with frequency Ω. The NS is well-approximated as a perfect conductor, therefore $\mathbf{E}\cdot\mathbf{B}=0$ in the interior. The exterior $\mathbf{E}\cdot\mathbf{B}$ is determined by solving Laplace's equation with conducting boundary conditions imposed on the NS surface [53].

In the VDM, it follows that

$$\mathbf{E}\cdot\mathbf{B} = \frac{|\boldsymbol{\mu}|^2\Omega}{4r^5}\Big[\cos\theta\sin^2\theta_m + \frac{1}{2}\sin\theta\sin(2\theta_m)\times \big(\Omega r\cos(\phi-\Omega t) - \sin(\phi-\Omega t)\big)\Big]\,, \tag{1}$$

with (θ,ϕ) the spherical coordinates in the frame in which $\hat{z}$ is aligned with the NS's rotation axis. The first term on the right-hand-side of (1) does not have time dependence and thus does not radiate axions, while the second term gives sub-dominant radiated power relative to the third by an amount proportional to $\Omega R_{\rm NS} \ll 1$. The third term in (1) shows that the radiation will be at angular frequency Ω with dipole moment

$$|\mathbf{p}_{\rm eff}| = \left|\int d^3x\,\mathbf{x}\rho_{\rm eff}(\mathbf{x})\right| \approx \frac{\pi}{6} g_{a\gamma\gamma}\frac{|\boldsymbol{\mu}|^2\Omega}{R_{\rm NS}}\sin(2\theta_m)\,. \tag{2}$$

Using the dipole radiation formula, this suggests a radiated power in axions of the form

$$\begin{aligned} P_a &\approx \frac{\pi}{432} g_{a\gamma\gamma}^2 B_0^4 \Omega^6 R_{\rm NS}^{10}\sin^2(2\theta_m) \\ &\approx 1.4\times 10^{35}\,\frac{\rm erg}{\rm s}\left(\frac{g_{a\gamma\gamma}}{10^{-12}\ {\rm GeV}^{-1}}\right)^2\left(\frac{\Omega}{2\pi\times 30\ {\rm Hz}}\right)^6 \\ &\times\left(\frac{R_{\rm NS}}{14\ {\rm km}}\right)^{10}\left(\frac{B_0}{8.5\times 10^{12}\ {\rm G}}\right)^4\sin^2(2\theta_m)\,. \end{aligned} \tag{3}$$

The fiducial pulsar parameters above correspond to those of the Crab pulsar, as motivated below. The Supplementary Material (SM) generalizes (3) to include the non-zero axion mass (see also [46]).

For sufficiently rapid pulsar rotation, $\mathbf{E}\cdot\mathbf{B}$ is strong enough to liberate charges from the NS surface, which populate the magnetosphere and screen the electric field in the process. However, even in active pulsars, $\mathbf{E}\cdot\mathbf{B}$ is not expected to be screened everywhere. An example of a partially-screened magnetosphere is the disk-dome or electrosphere solution found in slowly-rotating NSs [59–61]. These solutions have screened magnetospheres extending out to distances of order $R_{\rm NS}$ from the NS surface, beyond which the VDM applies. Referring to (2), excluding distances up to $xR_{\rm NS}$ from the NS center, with $x\sim$ few, reduces the radiated power by a factor of x.

A lower bound on the axion emission power may be computed assuming the magnetosphere is mostly screened. The production of screening $e^\pm$ pairs occurs in gap regions with non-vanishing $\mathbf{E}\cdot\mathbf{B}$, such as those that arise near the magnetic poles of the pulsar. Pulsar rotation defines a boundary, called the light cylinder (at radius $R_{\rm LC}=1/\Omega$ in natural units), beyond which particles cannot co-rotate with the NS. Charged particles escape the magnetosphere along open field lines (defined as those that do not close within the light cylinder), leading to charge starvation above the polar cap (PC), defined as the region on the NS surface that contains the footpoints of all open field lines. The dearth of plasma above the PC opens a gap with $\mathbf{E}\cdot\mathbf{B}\neq 0$. As the gap grows, it becomes unstable to runaway $e^\pm$ pair production, the dynamics of which have been modeled semi-analytically in [48, 62] and from first-principles kinetic plasma simulations in [63, 64]. All timescales associated with screening of the electric field are much smaller than the rotational period of the pulsar. Thus, when considering axion radiation at the rotational frequency, the gap may be treated as a point particle with "axion charge"

$$Q_a = g_{a\gamma\gamma}\int_{\rm gap} d^3x\,\langle\mathbf{E}\cdot\mathbf{B}(\mathbf{x},t)\rangle_t\,, \tag{4}$$

where $\langle\cdot\rangle_t$ represents the time average. We assume that the gap is screened for one light crossing time and model the gap as a cylinder with radius $r_{\rm PC} = R_{\rm NS}\sqrt{\Omega R_{\rm NS}}$ and height h. Note that the gap radius may be derived by equating the magnetic flux through the PC region, $\pi B_0 r_{\rm PC}^2$, and the flux through the light cylinder [65]. At low altitude, $\ell \ll R_{\rm NS}$ with ℓ the height above the NS surface, the electric field in the gap is $E_\parallel(\ell) = 2\Omega B_0\ell$, which corresponds to a voltage drop across the gap of $\Phi_{\rm PC} = \Omega B_0 h^2$ [66]. With this geometry, the axion charge is $Q_a = \pi g_{a\gamma\gamma} B_0^2 \Omega r_{\rm PC}^2 h^2$.

The height of the gap is limited by the vacuum breakdown; as h increases, the voltage drop increases and the gap becomes unstable to pair production. Most conservatively, we may set h to be the mean free path of a pair-

producing photon [67–69], which is approximated by [66]

$$h \sim 7\ \mathrm{m} \left(\frac{\Omega}{2\pi \times 30\ \mathrm{Hz}} \frac{B_0}{8.5 \times 10^{12}\ \mathrm{G}}\right)^{-4/7} . \quad (5)$$

The true gap height may be larger than that in (5) by a factor $N_{\rm gen}$, which represents the number of generations of pair production required to screen the gap. Assuming the charge density grows exponentially during the cascade, $N_{\rm gen}$ depends logarithmically on the ratio between the current supplied by the magnetosphere and the Goldreich-Julian current density $j_{\rm GJ} \sim 2\Omega B_0$, giving $N_{\rm gen} \sim$ few. Modeling the current supply of the magnetosphere requires global simulations, and so we assume $N_{\rm gen} = 1$ in this work to be conservative.

A NS with a dipole magnetic field has two antipodal PCs with opposite "axion charge." Taking the model of two oppositely charged point masses rotating rigidly and inclined at an angle θ_m relative to the rotation axis, the radiated power in the dipole approximation is

$$\begin{aligned} P_a &\approx \frac{1}{3\pi} R_{\rm NS}^2 \Omega^4 Q_a^2 \sin^2\theta_m \\ &\approx 10^{20}\ \frac{\mathrm{erg}}{\mathrm{s}} \left(\frac{g_{a\gamma\gamma}}{10^{-12}\ \mathrm{GeV}^{-1}}\right)^2 \left(\frac{\Omega}{2\pi \times 30\ \mathrm{Hz}}\right)^{\frac{40}{7}} \quad (6) \\ &\times \left(\frac{R_{\rm NS}}{14\ \mathrm{km}}\right)^8 \left(\frac{B_0}{8.5 \times 10^{12}\ \mathrm{G}}\right)^{12/7} \sin^2(\theta_m) , \end{aligned}$$

for $m_a \ll \Omega$. See SM for the generalization to non-zero axion mass. Equations (3) and (6) fully bracket the current uncertainties on the emitted power in axions.

Among pulsars in the ATNF catalog [70], the Crab pulsar provides the highest axion power, by a factor of ≈ 20 in the VDM, and is within a factor of two of the most efficient axion-emitting pulsar in the PC model. The spin-down parameters of the Crab are measured directly from timing of the radio beam: $P = 0.0338238880741$ s ($\Omega/2\pi = 29.5649038871$ Hz) and $\dot{P} = 4.1958812 \times 10^{-13}$ (on 15/1/2024) [70, 71]. Its radius, $R_{\rm Crab}$, may be determined by observations of the Crab nebula [72]: $R_{\rm Crab} = 14$–15 km, where the spread arises due to uncertainty in the NS equation of state. Self-consistently incorporating the measurement of the pulsar radius, mass, and moment of inertia with the equation for vacuum-filled pulsar spin down gives $B_{\rm Crab} > 8.5\ (6.9) \times 10^{12}$ G for $R_{\rm Crab} = 14\ (15)$ km [73]. These inferred magnetic field values are roughly consistent with those from analytic spin-down models studied in [74], $B_{\rm Crab} = (5.0$–$8.3) \times 10^{12}$ G. Motivated by these analyses, we fix the relevant Crab pulsar parameters to be $R_{\rm NS} = 14$ km and $B_0 = 8.5 \times 10^{12}$ G, which are conservative within the aforementioned observational constraints.

Detecting Pulsar-Sourced Axions.— The pulsar-induced axion field $a(\mathbf{x}, t)$ on Earth is analogous to that of DM (see, *e.g.*, [75]) with a few exceptions. First, the local energy density in axions due to the pulsar is given by $\rho_{\rm PS} = P_a/(4\pi D^2)$, with D the distance to the pulsar. The terrestrial energy density of pulsar-sourced axions, $\rho_{\rm PS}$, is in general much less than the local DM density, $\rho_{\rm DM}$. For example, if the full spin-down power of the Crab pulsar were emitted in axions, the ratio of energy densities would be $\rho_{\rm PS}/\rho_{\rm DM} \sim 10^{-13}$. Despite the orders-of-magnitude suppression in the local axion density, pulsar-sourced axions possess several key differences relative to DM axions that facilitate their detection.

While DM axions are non-relativistic, with characteristic momenta $k \sim 10^{-3}\omega$ and $\omega \sim m_a$, pulsar-induced axions are generically relativistic, with $k = \sqrt{\omega^2 - m_a^2} \sim \omega$ for $m_a \ll \omega \approx \Omega$. (Note that axions are only produced by the pulsar if $m_a < \Omega$.) This aspect of the pulsar-induced axion signal is useful for experiments such as CASPEr-wind that rely on the axion field's spatial gradient.

DM axions are expected to be ultra-narrow in frequency space, with the axion field only having non-trivial support over a spread in frequencies $\delta\omega/\omega \sim 10^{-6}$, but the pulsar-induced axion signals may be even narrower. Assuming that the un-screened parts of magnetosphere co-rotate with the NS, then the axion signal has a natural line width set by the pulsar spin-down parameters: $\delta\omega/\omega = \dot{P}$, where $\dot{P}$ is the observed rate of change of the period. Most observed pulsars have $\dot{P} \ll 10^{-10}$, giving nearly monochromatic axion signals. For example, the Crab pulsar has $\dot{P} \approx 4.2 \times 10^{-13}$ as mentioned previously, making the Crab-induced axion signal millions of times narrower than the expected DM signal. (Note that every few years the Crab undergoes small sporadic "glitches," resulting in changes in the frequency and amplitude of the signal; we assume the observing window does not contain any such major timing anomalies.) By causality, deviations from co-rotation are expected near the light cylinder. On the other hand, in both the VDM model and the PC model, along with the electrosphere model, the axion emission is predominantly produced near the NS surface and far from the light cylinder. Understanding the extent to which magnetosphere drag may affect the signal line width requires dedicated simulations that include general relativistic effects, which are beyond the scope of this work.

Another key aspect of the pulsar-induced axion signal is that it is generated at known frequencies, unlike in the case of DM. Given that the mass of the DM axion is currently unknown, experiments must scan over millions of different possibilities. On the other hand, the Crab signal is at the known frequency $f \approx 29.6$ Hz. The frequency of the signal is modulated by known effects (*e.g.* pulsar spin down and motion of the Earth), requiring real-time adjustment of the observing frequency.

We now follow the formalism in [75, 76] to project the sensitivity of key upcoming axion DM experiments to the signal from the Crab pulsar. We concentrate on experiments like Dark SRF and CASPEr-wind, though in the

SM we discuss DMRadio-GUT, which may yield comparable sensitivity relative to a Dark SRF experiment.

Superconducting radio frequency (SRF) cavities have demonstrated quality factors $Q \gtrsim 10^{11}$ [77, 78], making them ideal resonators to look for highly monochromatic axion signals. The Dark SRF approach (called "heterodyne detection") to low-frequency axion detection involves the axion driving a transition between two nearly degenerate modes of an SRF cavity [49, 50]. In this approach, the axion field drives a transition from a cavity "pump mode" of frequency $\omega_0 \sim$ GHz to a nearly-degenerate "signal mode" of frequency $\omega_1 \approx \omega_0 + \omega_a$, where ω_a is the frequency of the axion field. The signal is measured by a readout waveguide coupled to the signal mode [49, 50].

We envision measuring the power in the signal mode at equal time intervals over a total integration time t and then computing the power spectral density (PSD) of the data in frequency space. Since the axion coherence time ($\tau_a \gtrsim$ kyr) is much larger than t, the axion signal is expected to lie entirely in a single frequency bin, labeled k^*. The power in bin k^*, obtained by integrating the PSD over the bin (see SM for details), is

$$S_{k^*} \simeq \pi^2 \left(g_{a\gamma\gamma}\eta_{10}B_p\right)^2 V_{\rm cav} \left(\frac{Q_1}{\omega_1}\right) \rho_{\rm PS} \,, \qquad (7)$$

where η_{10} is a dimensionless mode overlap factor, B_p is the magnetic field amplitude of the pump mode, $V_{\rm cav}$ is the cavity volume, and Q_1 is the loaded quality factor of the signal mode. At the Crab pulsar frequency, the dominant noise source is thermal. The noise power in bin k^* is (see SM)

$$B_{k^*} \simeq 4\pi T_{\rm cav} \left(\frac{Q_1}{Q_{\rm int}}\right) \Delta\omega \,, \qquad (8)$$

where $T_{\rm cav}$ is the cavity temperature, $Q_{\rm int}$ is the intrinsic quality factor of the cavity, and $\Delta\omega = 2\pi/t$ is the frequency bin width. The probability of observing a given PSD in bin k^* is exponentially distributed with mean $S_{k^*} + B_{k^*}$ for fixed signal parameters; this implies that the expected 95% upper limit under the null hypothesis may be determined by setting $S_{k^*}/B_{k^*} \approx 8.48$ [79]. Figure 2 (left panel) shows this projection, adopting fiducial parameters for the Crab pulsar as described previously and fiducial detector parameters: $\eta_{10} = 1$, $B_p = 0.2$ T, $V_{\rm cav} = 1\,{\rm m}^3$, $\omega_0 \approx \omega_1 = 2\pi\times 100$ MHz, $Q_{\rm int} \approx Q_1 \approx 10^{12}$, $T_{\rm cav} = 1.8$ K, and $t = 1$ year as in [49, 50].

Figure 2 (left panel) compares the expected sensitivity to existing constraints from the CAST experiment [80], which looks for axions produced in the Sun. The shaded grey region also illustrates the current constraints from a number of astrophysical axion probes, including searches for spectral distortions of astrophysical sources with Chandra X-ray data [81–84] and super star cluster searches for axions with NuSTAR X-ray data [85]. Note that some of the astrophysical limits shown may also be subject to theoretical uncertainties (*e.g.*, [86, 87]). Assuming the VDM (labeled 'Vacuum'), the Dark SRF projected sensitivity would be world leading, improving upon the CAST bounds on $g_{a\gamma\gamma}$ by more than two orders of magnitude. In the electrosphere model, the axion power is lower than predicted by the VDM by a factor of $x \sim$ few, which leads to a modest drop in sensitivity of $x^{1/4}$. Assuming the conservative PC model, with radiated power given in (6), the projected sensitivity does not pass that of the CAST experiment. This motivates more detailed modeling of the Crab pulsar magnetosphere to determine precisely where the sensitivity lies.

Figure 2 (left panel) also illustrates the projected sensitivity of the planned DMRadio-GUT experiment [52], assuming the experiment takes data with a quality factor of 2×10^7 [52] for one year with a resonant readout at the Crab frequency. See the SM for further details.

The right panel of Fig. 2 shows the projected sensitivity of the CASPEr-wind experiment [40–43] under both the VDM and PC magnetosphere models, assuming $m_a \ll 10^{-13}$ eV. CASPEr-wind searches for the axion through its coupling to the nucleon: $H_{\rm int} = \gamma_N \boldsymbol{B}_a \cdot \boldsymbol{S}_N$, with γ_N the nucleon gyromagnetic ratio and $\boldsymbol{B}_a$ an effective magnetic field that is equal to $\boldsymbol{B}_a = -g_{aNN}\boldsymbol{\nabla} a/\gamma_N$. CASPEr-wind, and also co-magnetometer experiments such as [88, 89], use nuclear magnetic resonance techniques with a transverse magnetometer to search for spin precession of a magnetized sample due to the axion interaction Hamiltonian; the effect is resonantly enhanced when the axion frequency matches the Larmor frequency $\omega_0 = \gamma_N B_E$, with B_E the static external magnetic field that aligns the spins.

Measuring the transverse magnetic field over $t =$ 1 year, the axion signal contributes to a single frequency bin because $t \ll \tau_a$. On the other hand, the width of the resonance is limited by the transverse spin relaxation time T_2, which is expected to be much less than t. As in the SRF case, one may compute the expected signal contribution within the frequency bin k^* and compare it to the expected background noise, which arises predominantly from spin projection and magnetometer noise for CASPEr-wind [79]. (See the SM for details.)

Figure 2 (right panel) shows the projected 95% upper limits in the space of $|g_{a\gamma\gamma}|$ and $|g_{aNN}|$. (Note that the observed signal is proportional to the combination $g^2_{a\gamma\gamma}g^2_{aNN}$.) The projected sensitivity is shown for two different assumptions regarding the CASPEr-wind type experiment (see, *e.g.*, [79] for the justification of these choices): the more conservative case takes a Xe-129 target, which has a nuclear magnetic moment of $0.78\mu_N$, with μ_N the nuclear magneton, along with the number density $n \approx 1.3 \times 10^{22}$ cm^{-3}. The more optimistic scenario assumes a Helium-3 target with a nuclear magnetic moment $\sim 2.1\mu_N$ and $n \approx 2.8 \times 10^{22}$ cm^{-3}. For both, the volume is 1 m^3, spin relaxation time is $T_2 = 100$ sec, and

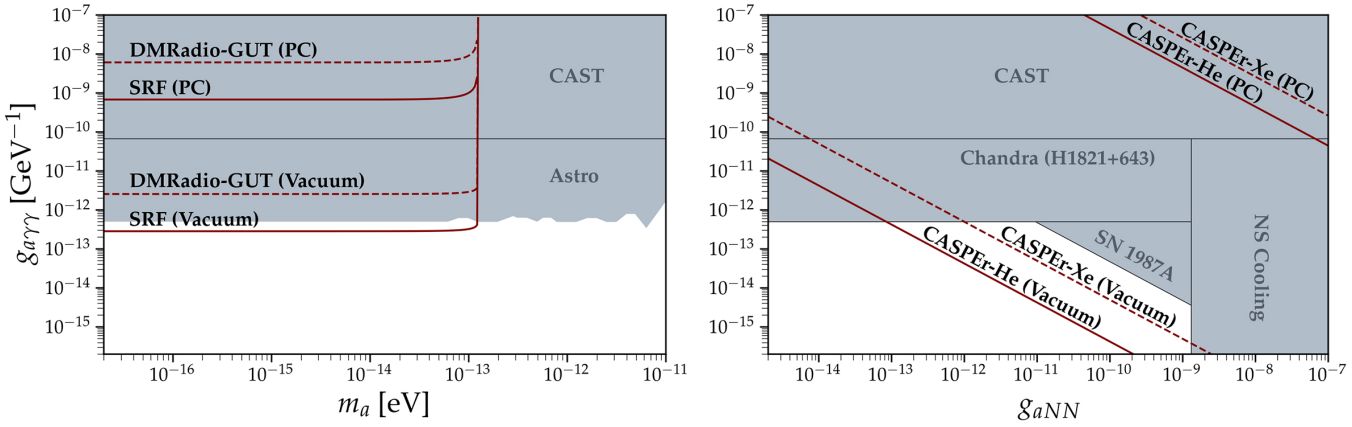


FIG. 2. **Left:** The projected sensitivity to the axion-photon coupling $g_{a\gamma\gamma}$ as a function of the axion mass, m_a, from one-year dedicated searches with the proposed Dark SRF axion experiment and the DMRadio-GUT experiment for the axion-induced signal from the Crab pulsar at $f \approx 29.6$ Hz. We illustrate projected 95% upper limits under the null hypothesis assuming both the VDM magnetosphere ('Vacuum') and for the PC model, where the axions are only produced in small gaps near the NS poles. The grey shaded regions represent current constraints on this parameter space, as summarized in the main text. **Right:** As in the left panel, but for the future CASPER-wind experiment, which searches for the axion-nucleon coupling g_{aNN}. This figure assumes $m_a \ll 10^{-13}$ eV and illustrates the projected 95% upper limits in the $g_{a\gamma\gamma} - g_{aNN}$ plane, for two different target materials.

the effective SQUID loop area is 0.3 cm^2.

We compare the expected upper limits from CASPEr-wind to the existing constraints on this parameter space on $g_{a\gamma\gamma}$ alone along with the NS cooling constraints on g_{aNN} [90]. Additionally, the non-observation of gamma-rays from SN1987A that would have been produced by nucleon bremsstrahlung (with g_{aNN}) and then converted to gamma-rays in the Galactic magnetic fields (with $g_{a\gamma\gamma}$) gives a constraint on the combination $g_{a\gamma\gamma} \times g_{aNN}$ [91, 92], which is also illustrated. Both the Xe and He targets could probe previously unexplored regions of parameter space for the VDM magnetosphere.

Discussion.— We proposed a new target for axion DM detection experiments: pulsar-sourced axions. These relativistic axions are produced at known frequencies; for example, the signal from the Crab pulsar is around $f \approx 29.6$ Hz. Upcoming searches for axions from the Crab pulsar may probe previously-unexplored regions of low-mass axion parameter space. Global particle-in-cell (PIC) plasma simulations of the Crab magnetosphere along the lines of [93–95] are needed to better model the distribution of $\mathbf{E} \cdot \mathbf{B}$ in the magnetosphere and compute the axion luminosity and line width more accurately.

It is also possible that better pulsar candidates exist that are not known at present. These could include high-field, high-period pulsars that are not beamed towards Earth or are otherwise obscured to electromagnetic radiation. Young pulsars born after supernova, including possible future supernova, may also be superior targets. Transient axion signals could also arise from gamma-ray bursts and magnetar glitches.

The axion emission from pulsars may also directly affect the observed pulsar quantities such as the period and magnetic field. For example, the spin-down luminosity of the Crab pulsar is $L_{\rm spin-down} = 4.5 \times 10^{38}$ erg/s [70]. The axion-induced emission formula in (3) surpasses the spin-down luminosity for axion-photon couplings larger than $g_{a\gamma\gamma} \sim 5 \times 10^{-11}$ GeV^{-1}. This strongly suggests a conservative upper limit of $|g_{a\gamma\gamma}| \lesssim 5 \times 10^{-11}$ GeV^{-1}, assuming the VDM model, for $m_a \lesssim 10^{-13}$ eV. On the other hand, the limit can likely be substantially improved by studying the back-reaction of the axion emission more carefully on the pulsar dynamics.

Another class of astrophysically-produced low-frequency axions that could be detectable with terrestrial DM detectors are signals narrow in the time domain. For example, we estimate that NS-NS inspirals and NS-black-hole inspirals may yield detectable but short-lived axion signals, which would be coincident in time with gravitational wave signals, depending on the magnetic properties of the NSs and the distance to the event. We leave detailed estimates of this signal to future work, which could most optimistically add axions as an additional messenger to the growing field of multi-messenger astronomy.

Acknowledgments— The authors acknowledge A. Berlin, S. Chaudhuri, J. Foster, H. Hakobyan, Y. Kahn, N. Rodd, A. Sushkov, L. Winslow, and S. Witte for helpful discussions. ML and MK are supported by the Department of Energy (DOE) under Award Number DE-SC0007968. ML also acknowledges support from the Simons Investigator in Physics Award. AP acknowledges support from the Princeton Center for Theoretical Science. BRS is supported in part by the DOE Early Career

Grant DESC0019225 and in part by a Sloan Research Fellowship. MK is grateful to Princeton University for their hospitality. The work presented in this paper was performed on computational resources managed and supported by Princeton Research Computing. This research also made extensive use of the publicly available codes `IPython` [96], `matplotlib` [97], `Jupyter` [98], `NumPy` [99], and `SciPy` [100].

* khelashvili@fias.uni-frankfurt.de
† mlisanti@princeton.edu
‡ prabhu@princeton.edu
§ brsafdi@berkeley.edu

[1] P. Svrcek and E. Witten, JHEP **2006**, 051–051 (2006).
[2] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, Phys. Rev. D **81**, 123530 (2010).
[3] J. Preskill, M. B. Wise, and F. Wilczek, Physics Letters B **120**, 127 (1983).
[4] L. Abbott and P. Sikivie, Phys. Lett. B **120**, 133 (1983).
[5] M. Dine and W. Fischler, Phys. Lett. B **120**, 137 (1983).
[6] Planck Collaboration, Aghanim, N., *et al.*, A&A **641**, A6 (2020).
[7] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[8] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).
[9] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
[10] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
[11] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, Ann. Rev. Nucl. Part. Sci. **65**, 485 (2015), arXiv:1602.00039 [hep-ex].
[12] C. B. Adams *et al.*, in *Snowmass 2021* (2022) arXiv:2203.14923 [hep-ex].
[13] B. R. Safdi, PoS **TASI2022**, 009 (2024), arXiv:2303.02169 [hep-ph].
[14] C. Hagmann, P. Sikivie, N. S. Sullivan, and D. B. Tanner, Phys. Rev. D **42**, 1297 (1990).
[15] C. Hagmann *et al.* (ADMX), Phys. Rev. Lett. **80**, 2043 (1998), arXiv:astro-ph/9801286.
[16] S. J. Asztalos *et al.* (ADMX), Phys. Rev. D **64**, 092003 (2001).
[17] S. J. Asztalos *et al.* (ADMX), Phys. Rev. Lett. **104**, 041301 (2010), arXiv:0910.5914 [astro-ph.CO].
[18] N. Du *et al.* (ADMX), Phys. Rev. Lett. **120**, 151301 (2018), arXiv:1804.05750 [hep-ex].
[19] T. Braine *et al.* (ADMX), Physical Review Letters **124** (2020), 10.1103/physrevlett.124.101303.
[20] R. Bradley *et al.*, Rev. Mod. Phys. **75**, 777 (2003).
[21] S. J. Asztalos *et al.*, Phys. Rev. D **69**, 011101 (2004).
[22] T. M. Shokair *et al.*, International Journal of Modern Physics A **29**, 1443004 (2014).
[23] B. M. Brubaker *et al.* (HAYSTAC), Phys. Rev. Lett. **118**, 061302 (2017).
[24] L. Zhong *et al.* (HAYSTAC), Physical Review D **97** (2018), 10.1103/physrevd.97.092001.
[25] K. M. Backes *et al.* (HAYSTAC), Nature **590**, 238 (2021).
[26] J. Choi, S. Ahn, B. Ko, S. Lee, and Y. Semertzidis, NIM-A **1013**, 165667 (2021).
[27] N. Crescini *et al.* (QUAX), Phys. Rev. Lett. **124**, 171801 (2020), arXiv:2001.08940 [hep-ex].
[28] B. T. McAllister *et al.*, "The organ experiment: An axion haloscope above 15 ghz," (2017), arXiv:1706.00209 [physics.ins-det].
[29] P. Sikivie, Phys. Rev. Lett. **51**, 1415 (1983), [Erratum: Phys.Rev.Lett. 52, 695 (1984)].
[30] V. Anastassopoulos *et al.* (CAST), Nature Phys. **13**, 584 (2017), arXiv:1705.02290 [hep-ex].
[31] K. Ehret *et al.* (ALPS), Phys. Lett. **B689**, 149 (2010), arXiv:1004.1313 [hep-ex].
[32] M. Diaz Ortiz, J. Gleason, H. Grote, A. Hallal, M. Hartman, H. Hollis, K.-S. Isleif, A. James, K. Karan, T. Kozlowski, A. Lindner, G. Messineo, G. Mueller, J. Põld, R. Smith, A. Spector, D. Tanner, L.-W. Wei, and B. Willke, Physics of the Dark Universe **35**, 100968 (2022).
[33] N. Du *et al.* (ADMX), Phys. Rev. Lett. **120**, 151301 (2018), arXiv:1804.05750 [hep-ex].
[34] T. Braine *et al.* (ADMX), Phys. Rev. Lett. **124**, 101303 (2020), arXiv:1910.08638 [hep-ex].
[35] C. Bartram *et al.* (ADMX), Phys. Rev. D **103**, 032002 (2021), arXiv:2010.06183 [astro-ph.CO].
[36] C. Bartram, T. Braine, E. Burns, R. Cervantes, N. Crisosto, N. Du, H. Korandla, G. Leum, P. Mohapatra, T. Nitta, L. J. Rosenberg, G. Rybka, J. Yang, J. Clarke, I. Siddiqi, A. Agrawal, A. V. Dixit, M. H. Awida, A. S. Chou, M. Hollister, S. Knirck, A. Sonnenschein, W. Wester, J. R. Gleason, A. T. Hipp, S. Jois, P. Sikivie, N. S. Sullivan, D. B. Tanner, E. Lentz, R. Khatiwada, G. Carosi, N. Robertson, N. Woollett, L. D. Duffy, C. Boutan, M. Jones, B. H. LaRoque, N. S. Oblath, M. S. Taubman, E. J. Daw, M. G. Perry, J. H. Buckley, C. Gaikwad, J. Hoffman, K. W. Murch, M. Goryachev, B. T. McAllister, A. Quiskamp, C. Thomson, and M. E. Tobar (ADMX Collaboration), Phys. Rev. Lett. **127**, 261803 (2021).
[37] J. L. Ouellet *et al.* (ABRACADABRA), Phys. Rev. Lett. **122**, 121802 (2019), arXiv:1810.12257 [hep-ex].
[38] C. P. Salemi *et al.*, Phys. Rev. Lett. **127**, 081801 (2021), arXiv:2102.06722 [hep-ex].
[39] P. W. Graham and S. Rajendran, Phys. Rev. D **88**, 035023 (2013), arXiv:1306.6088 [hep-ph].
[40] D. Budker, P. W. Graham, M. Ledbetter, S. Rajendran, and A. O. Sushkov, Physical Review X **4** (2014), 10.1103/physrevx.4.021030.
[41] D. F. Jackson Kimball *et al.*, Springer Proc. Phys. **245**, 105 (2020), arXiv:1711.08999 [physics.ins-det].
[42] A. Garcon, J. W. Blanchard, G. P. Centers, N. L. Figueroa, P. W. Graham, D. F. J. Kimball, S. Rajendran, A. O. Sushkov, Y. V. Stadnik, A. Wickenbrock, T. Wu, and D. Budker, Science Advances **5** (2019), 10.1126/sciadv.aax4539.
[43] T. Wu, J. W. Blanchard, G. P. Centers, N. L. Figueroa, A. Garcon, P. W. Graham, D. F. J. Kimball, S. Rajendran, Y. V. Stadnik, A. O. Sushkov, A. Wickenbrock, and D. Budker, Phys. Rev. Lett. **122**, 191302 (2019).
[44] C. Gao, W. Halperin, Y. Kahn, M. Nguyen, J. Schütte-Engel, and J. W. Scott, Phys. Rev. Lett. **129**, 211801 (2022), arXiv:2208.14454 [hep-ph].
[45] J. W. Foster, C. Gao, W. Halperin, Y. Kahn, A. Mande, M. Nguyen, J. Schütte-Engel, and J. W. Scott, (2023), arXiv:2310.07791 [hep-ph].
[46] B. Garbrecht and J. I. McDonald, JCAP **07**, 044 (2018),

arXiv:1804.04224 [astro-ph.CO].
[47] A. Prabhu, Physical Review D **104** (2021), 10.1103/physrevd.104.055038.
[48] D. Noordhuis, A. Prabhu, S. J. Witte, A. Y. Chen, F. Cruz, and C. Weniger, Phys. Rev. Lett. **131**, 111004 (2023), arXiv:2209.09917 [hep-ph].
[49] A. Berlin, R. T. D'Agnolo, S. A. R. Ellis, C. Nantista, J. Neilson, P. Schuster, S. Tantawi, N. Toro, and K. Zhou, Journal of High Energy Physics **2020** (2020), 10.1007/jhep07(2020)088.
[50] A. Berlin, R. T. D'Agnolo, S. A. R. Ellis, and K. Zhou, Phys. Rev. D **104**, L111701 (2021).
[51] L. Brouwer *et al.* (DMRadio), Phys. Rev. D **106**, 103008 (2022), arXiv:2204.13781 [hep-ex].
[52] L. Brouwer *et al.* (DMRadio), Phys. Rev. D **106**, 112003 (2022), arXiv:2203.11246 [hep-ex].
[53] A. J. Deutsch, Annales d'Astrophysique **18**, 1 (1955).
[54] F. Hoyle, J. V. Narlikar, and J. A. Wheeler, Nature **203**, 914 (1964).
[55] F. Pacini, Nature **216**, 567 (1967).
[56] F. Pacini, Nature **219**, 145 (1968).
[57] J. P. Ostriker and J. E. Gunn, Astrophys. J. **157**, 1395 (1969).
[58] D. B. Melrose and R. Yuen, The Astrophysical Journal **745**, 169 (2012).
[59] J. Pétri, J. Heyvaerts, and S. Bonazzola, Astronomy and Astrophysics **384**, 414 (2002).
[60] A. Philippov and M. Kramer, Annual Review of Astron and Astrophys **60**, 495 (2022).
[61] F. Cruz, T. Grismayer, A. Y. Chen, A. Spitkovsky, R. A. Fonseca, and L. O. Silva, (2023), arXiv:2309.04834 [astro-ph.HE].
[62] E. A. Tolman, A. A. Philippov, and A. N. Timokhin, (2022), arXiv:2202.01303 [astro-ph.HE].
[63] F. Cruz, T. Grismayer, A. Y. Chen, A. Spitkovsky, and L. O. Silva, Astrophysical Journal Letters **919**, L4 (2021), arXiv:2108.11702 [astro-ph.HE].
[64] F. Cruz, T. Grismayer, S. Iteanu, P. Tortone, and L. O. Silva, Phys. Plasmas **29**, 052902 (2022), arXiv:2204.03766 [astro-ph.HE].
[65] P. Goldreich and W. H. Julian, Astrophys. J. **157**, 869 (1969).
[66] M. A. Ruderman and P. G. Sutherland, Astrophys. J. **196**, 51 (1975).
[67] A. N. Timokhin and A. K. Harding, Astrophys. J. **810**, 144 (2015), arXiv:1504.02194 [astro-ph.HE].
[68] A. N. Timokhin and A. K. Harding, Astrophys. J. **871**, 12 (2019), arXiv:1803.08924 [astro-ph.HE].
[69] A. Caputo, S. J. Witte, A. A. Philippov, and T. Jacobson, (2023), arXiv:2311.14795 [hep-ph].
[70] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs, Astronomical Journal **129**, 1993 (2005), arXiv:astro-ph/0412641 [astro-ph].
[71] A. G. Lyne, R. S. Pritchard, and F. Graham Smith, Monthly Notices of the Royal Astronomical Society **265**, 1003 (1993), https://academic.oup.com/mnras/article-pdf/265/4/1003/3173877/mnras265-1003.pdf.
[72] M. Bejger and P. Haensel, Astronomy and Astrophysics **396**, 917 (2002), arXiv:astro-ph/0209151 [astro-ph].
[73] A. Philippov, A. Tchekhovskoy, and J. G. Li, Monthly Notices of the Royal Astronomical Society **441**, 1879–1887 (2014).
[74] F. F. Kou and H. Tong, Monthly Notices of the Royal Astronomical Society **450**, 1990 (2015), https://academic.oup.com/mnras/article-pdf/450/2/1990/3076208/stv734.pdf.
[75] J. W. Foster, N. L. Rodd, and B. R. Safdi, Phys. Rev. D **97**, 123006 (2018), arXiv:1711.10489 [astro-ph.CO].
[76] J. A. Dror, H. Murayama, and N. L. Rodd, Physical Review D **103** (2021), 10.1103/physrevd.103.115004.
[77] A. Romanenko, A. Grassellino, A. C. Crawford, D. A. Sergatskov, and O. Melnychuk, Applied Physics Letters **105** (2014), 10.1063/1.4903808.
[78] S. Posen, G. Wu, A. Grassellino, E. Harms, O. Melnychuk, D. Sergatskov, N. Solyak, A. Romanenko, A. Palczewski, D. Gonnella, and T. Peterson, Physical Review Accelerators and Beams **22** (2019), 10.1103/physrevaccelbeams.22.032001.
[79] J. A. Dror, S. Gori, J. M. Leedom, and N. L. Rodd, Phys. Rev. Lett. **130**, 181801 (2023).
[80] M. Arik *et al.* (CAST), Phys. Rev. D **92**, 021101 (2015).
[81] J. S. Reynés, J. H. Matthews, C. S. Reynolds, H. R. Russell, R. N. Smith, and M. C. D. Marsh, Mon. Not. Roy. Astron. Soc. **510**, 1264 (2021), arXiv:2109.03261 [astro-ph.HE].
[82] C. S. Reynolds, M. C. D. Marsh, H. R. Russell, A. C. Fabian, R. Smith, F. Tombesi, and S. Veilleux, Astrophys. J. **890**, 59 (2020), arXiv:1907.05475 [hep-ph].
[83] M. C. D. Marsh, H. R. Russell, A. C. Fabian, B. P. McNamara, P. Nulsen, and C. S. Reynolds, JCAP **12**, 036 (2017), arXiv:1703.07354 [hep-ph].
[84] D. Wouters and P. Brun, Astrophys. J. **772**, 44 (2013), arXiv:1304.0989 [astro-ph.HE].
[85] C. Dessert, J. W. Foster, and B. R. Safdi, Phys. Rev. Lett. **125**, 261102 (2020), arXiv:2008.03305 [hep-ph].
[86] M. Libanov and S. Troitsky, Phys. Lett. B **802**, 135252 (2020), arXiv:1908.03084 [astro-ph.HE].
[87] J. H. Matthews, C. S. Reynolds, M. C. D. Marsh, J. Sisk-Reynés, and P. E. Rodman, Astrophys. J. **930**, 90 (2022), arXiv:2202.08875 [astro-ph.HE].
[88] J. Lee, M. Lisanti, W. A. Terrano, and M. Romalis, Physical Review X **13** (2023), 10.1103/physrevx.13.011050.
[89] I. M. Bloch, G. Ronen, R. Shaham, O. Katz, T. Volansky, and O. Katz, Science Advances **8** (2022), 10.1126/sciadv.abl8919.
[90] M. Buschmann, C. Dessert, J. W. Foster, A. J. Long, and B. R. Safdi, Phys. Rev. Lett. **128**, 091102 (2022), arXiv:2111.09892 [hep-ph].
[91] A. Payez, C. Evoli, T. Fischer, M. Giannotti, A. Mirizzi, and A. Ringwald, JCAP **02**, 006 (2015), arXiv:1410.3747 [astro-ph.HE].
[92] C. A. Manzari, Y. Park, I. Savoray, and B. R. Safdi, (2024 to-appear).
[93] A. Spitkovsky, The Astrophysical Journal **648**, L51 (2006).
[94] C. Kalapotharakos and I. Contopoulos, Astronomy and Astrophysics **496**, 495 (2009), arXiv:0811.2863 [astro-ph].
[95] J. Pétri, Monthly Notices of the Royal Astronomical Society **424**, 605 (2012).
[96] F. Pérez and B. E. Granger, Computing in Science and Engineering **9**, 21 (2007).
[97] J. D. Hunter, Computing in Science & Engineering **9**, 90 (2007).
[98] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla,

and C. Willing, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt (IOS Press, 2016) pp. 87 – 90.

[99] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Nature **585**, 357 (2020).

[100] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nature Methods **17**, 261 (2020).

[101] N. N. Bogolyubov and D. V. Shirkov, *Introduction To the Theory of Quantized Fields*, Vol. 3 (1959).

[102] E. Economou, *Green's Functions in Quantum Physics*, Springer Series in Solid-State Sciences (Springer Berlin Heidelberg, 2006).

[103] M. Li and R. Ruffini, Physics Letters A **116**, 20 (1986).

[104] D. E. Krause, H. T. Kloor, and E. Fischbach, Phys. Rev. D **49**, 6892 (1994).

[105] Y. Kahn, B. R. Safdi, and J. Thaler, Phys. Rev. Lett. **117**, 141801 (2016).

[106] J. N. Benabou, J. W. Foster, Y. Kahn, B. R. Safdi, and C. P. Salemi, Phys. Rev. D **108**, 035009 (2023), arXiv:2211.00008 [hep-ph].

[107] S. M. Anton, J. S. Birenbaum, S. R. O'Kelley, V. Bolkhovsky, D. A. Braje, G. Fitch, M. Neeley, G. C. Hilton, H. M. Cho, K. D. Irwin, F. C. Wellstood, W. D. Oliver, A. Shnirman, and J. Clarke, Phys. Rev. Lett. **110**, 147002 (2013).

[108] L. Brouwer, S. Chaudhuri, H.-M. Cho, J. Corbin, W. Craddock, C. Dawson, A. Droster, J. Foster, J. Fry, P. Graham, R. Henning, K. Irwin, F. Kadribasic, Y. Kahn, A. Keller, R. Kolevatov, S. Kuenstner, A. Leder, D. Li, J. Ouellet, K. Pappas, A. Phipps, N. Rapidis, B. Safdi, C. Salemi, M. Simanovskaia, J. Singh, E. van Assendelft, K. van Bibber, K. Wells, L. Winslow, W. Wisniewski, and B. Y. and, Physical Review D **106** (2022), 10.1103/physrevd.106.103008.

Supplementary Material for An Axion Pulsarscope

Mariia Khelashvili, Mariangela Lisanti, Anirudh Prabhu, and Benjamin R. Safdi

I. AXION EMISSION IN A PULSAR'S ELECTROMAGNETIC FIELD

The derivation of the axion power in the vacuum dipole model (VDM) and the polar cap (PC) model in the main text assumes the axion is massless. This section generalizes the derivation to finite mass $m_a \leq \Omega$, where Ω is the pulsar's rotational frequency. The evolution of the axion field exterior to the pulsar is governed by the Klein-Gordon equation, $(\Box + m_a^2)a(x) = g_{a\gamma\gamma}\mathbf{E}\cdot\mathbf{B}$, where $\mathbf{E}$ ($\mathbf{B}$) is the electric (magnetic) field in the pulsar magnetosphere. The Klein-Gordon equation with a source can be solved using the retarded Green function [101, 102]:

$$G(\mathbf{x},\tau) = \frac{\theta(\tau)}{2\pi}\left(\delta(s^2) - \frac{m_a}{2s}\theta(s^2)\,J_1(m_a s)\right), \qquad s^2 = \tau^2 - \mathbf{x}^2\,, \tag{S1}$$

where $J_1(x)$ is the Bessel function, θ is the Heaviside function, and δ is the delta function. The solution can be simplified for the case of localized periodic sources using the method of [103, 104]. In this case, the radiation from the pulsar magnetosphere is sourced by the dipole moment of $\mathbf{E}\cdot\mathbf{B}$ to leading-order (see, *e.g.*, (2) in the main text). We briefly summarize the calculations of axion dipole radiation and then apply the results to the two magnetosphere models considered in this work.

Assuming that the source periodically rotates at the pulsar rotation frequency Ω, the axion-field solution takes the form [104]

$$a(\mathbf{r},t) = \sum_{m=-\infty}^{\infty} e^{-im\Omega t} a_m(\mathbf{r})\,. \tag{S2}$$

In general, periodically rotating sources radiate at the fundamental frequency Ω and its integer multiples. For any realistic pulsar, however, the near-surface region rotates with velocity $v \sim \Omega R_{\rm NS} \ll 1$, where $R_{\rm NS}$ is the neutron star's radius, and radiated power at higher frequencies is highly suppressed compared to the fundamental frequency by powers of v. Thus, we only consider radiation at the fundamental frequency:

$$a(\mathbf{r},t) = e^{-i\Omega t}a_\Omega(\mathbf{r}) + e^{i\Omega t}a_{-\Omega}(\mathbf{r})\,. \tag{S3}$$

The corresponding Fourier component of the charge density, $\rho_{\rm eff}(\mathbf{r},t) \equiv g_{a\gamma\gamma}\mathbf{E}\cdot\mathbf{B}(\mathbf{r},t)$, is

$$\rho_{\pm\Omega}(\mathbf{r}) = \frac{1}{T}\int_0^T dt\, e^{\pm i\Omega t}\rho_{\rm eff}(\mathbf{r},t)\,, \tag{S4}$$

where $T = 2\pi/\Omega$ is the source rotation period. The axion field obeys the equation of motion:

$$(\nabla^2 + k^2)a_{\pm\Omega}(\mathbf{r}) = -\rho_{\pm\Omega}(\mathbf{r})\,, \quad k = \sqrt{\Omega^2 - m_a^2}\,. \tag{S5}$$

We may solve this equation using the outgoing-wave Green function [104],

$$G(\mathbf{r},\mathbf{r}') = \frac{1}{4\pi}\frac{e^{\pm ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,, \tag{S6}$$

leading to

$$a_{\pm\Omega}(\mathbf{r}) = \frac{1}{4\pi}\int d^3r'\frac{e^{\pm ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\rho_{\pm\Omega}(\mathbf{r}')\,. \tag{S7}$$

In the long-wavelength ($k|\mathbf{r}'| \ll 1$), far-field ($|\mathbf{r}| \gg |\mathbf{r}'|$) limit, the solution takes the form

$$a_{\pm\Omega}(\mathbf{r}) = \frac{e^{\pm ikr}}{4\pi r}\left(\int d^3r'\rho_{\pm\Omega}(\mathbf{r}') \pm ik\hat{n}\cdot\int d^3r'\rho_{\pm\Omega}(\mathbf{r}')\mathbf{r}' + \cdots\right) = \frac{e^{\pm ikr}}{4\pi r}\left[Q_{\pm\Omega} \pm ik\hat{n}\cdot\mathbf{p}_{\pm\Omega} + \cdots\right], \tag{S8}$$

where $\hat{n}$ is the unit vector in the direction of $\mathbf{r}$, and $Q_{\pm\Omega}$ ($\mathbf{p}_{\pm\Omega}$) is the monopole (dipole) moment of $\rho_{\pm\Omega}$, given by

$$Q_{\pm\Omega} = \int d^3r' \rho_{\pm\Omega}(\mathbf{r}') \,, \qquad \mathbf{p}_{\pm\Omega} = \int d^3r' \rho_{\pm\Omega}(\mathbf{r}') \mathbf{r}' \,. \tag{S9}$$

In the VDM and PC models, the monopole term vanishes and the leading-order term is that of the dipole. The time-dependent solution for the outgoing axion wave radiated by the dipole is

$$a(\mathbf{r},t) = \frac{ik}{4\pi r} \left(\hat{n} \cdot \mathbf{p}_{+\Omega}\, e^{-i(\Omega t - kr)} - \hat{n} \cdot \mathbf{p}_{-\Omega}\, e^{i(\Omega t - kr)} \right) . \tag{S10}$$

The radiated axion power follows:

$$P_a = \int \mathbf{\Pi} \cdot d\boldsymbol{\sigma} \,, \qquad \langle \mathbf{\Pi} \rangle_T = \langle \dot{a}^* \boldsymbol{\nabla} a \rangle_T = \langle \dot{a}^*_{\Omega}(\mathbf{r},t) \boldsymbol{\nabla} a_{\Omega}(\mathbf{r},t) + \dot{a}^*_{-\Omega}(\mathbf{r},t) \boldsymbol{\nabla} a_{-\Omega}(\mathbf{r},t) \rangle_T \,, \tag{S11}$$

where $\langle .. \rangle_T$ represents the time-average over the rotational period. For a real source, $\mathrm{Im}\, \rho(\mathbf{r},t) = 0$ and the Fourier components are $\rho^*_{-\Omega} = \rho_{\Omega}$, therefore $\mathbf{p}^*_{-\Omega} = \mathbf{p}_{\Omega}$ and the total axion dipole radiation becomes:

$$\langle P_a \rangle_T = \frac{1}{8\pi^2} \Omega k^3 \iint d\omega_n |\mathbf{p}_{\Omega} \cdot \hat{n}|^2 \,. \tag{S12}$$

The integration is performed over the solid angle ω_n, and the final expression includes the Fourier component of the dipole moment:

$$\mathbf{p}_{\Omega} = \frac{1}{T} \int_0^T dt\, e^{i\Omega t} \int d^3r \rho_{\rm eff}(\mathbf{r},t) \mathbf{r} \,. \tag{S13}$$

A. Vacuum Dipole Model

For a fully unscreened magnetosphere, the distribution of effective charge density, $g_{a\gamma\gamma} \mathbf{E} \cdot \mathbf{B}$, in the pulsar's magnetosphere is given by (1). The radiation of axions in this scenario can be modeled by a rotating dipole with effective dipole moment (2). From (S13),

$$\hat{n} \cdot \mathbf{p}_{\Omega} \approx \frac{i\pi}{12} g_{a\gamma\gamma} \frac{|\boldsymbol{\mu}|^2 \Omega}{R_{\rm NS}} \sin(2\theta_m) \sin\theta_n e^{i\phi_n} \,, \tag{S14}$$

where $|\boldsymbol{\mu}| = B_0 R^3_{\rm NS}$, B_0 is the surface magnetic field of the pulsar, θ_m is the angle between the pulsar's magnetic dipole axis and its rotation axis, and (θ_n, ϕ_n) are the angles in spherical coordinates that define the direction of $\hat{n}$. From the dipole radiation power (S12), it follows that

$$\langle P_a \rangle_T = \frac{\pi}{432} g^2_{a\gamma\gamma} B^4_0 \Omega^6 R^{10}_{\rm NS} \sin^2(2\theta_m) \left(1 - \frac{m_a^2}{\Omega^2} \right)^{3/2} \,, \tag{S15}$$

which reduces to (3) in the limit of a massless axion.

B. Polar Cap Model

We now discuss the model of a mostly-screened magnetosphere, where the axion is sourced only in the small polar cap (PC) regions above the magnetic poles. The size of the PC, $r_{\rm pc} \ll R_{\rm NS} \ll 1/\Omega$, can be neglected, so we model the PCs as two rotating point-like sources with opposite charges ($Q^N_a = -Q^S_a$) and opposite velocity and position ($\mathbf{v}^N_{\rm pc} = -\mathbf{v}^S_{\rm pc}$ and $\mathbf{r}^N_{\rm pc} = -\mathbf{r}^S_{\rm pc} = \mathbf{q}(t)$), where N (S) represents the PC at the northern (southern) pole. The effective charge density for the point-like PC model is

$$\rho(\mathbf{r},t) = Q_a \left[\delta(\mathbf{r} - \mathbf{q}(t)) - \delta(\mathbf{r} + \mathbf{q}(t)) \right] \,, \tag{S16}$$

where

$$\mathbf{q}(t) = R_{\rm NS}\left(\sin\theta_m\left(\cos(\Omega t)\hat{\mathbf{x}} + \sin(\Omega t)\hat{\mathbf{y}}\right) + \cos\theta_m\hat{\mathbf{z}}\right) . \tag{S17}$$

We can find the Fourier component of the dipole moment (S13):

$$\hat{n}\cdot\mathbf{p}_\Omega = Q_a R_{\rm NS}\sin\theta_m\sin\theta_n e^{-i\phi_n} , \tag{S18}$$

where θ_n, ϕ_n are spherical coordinate angles of the observer. Finally, the resulting axion radiation power is:

$$\langle P_a\rangle_T = \frac{1}{3\pi}R_{\rm NS}^2\Omega^4 Q_a^2\sin^2\theta_m\left(1-\frac{m_a^2}{\Omega^2}\right)^{3/2} , \tag{S19}$$

which is the same as (6) in the ultra-relativistic limit ($m_a \ll \Omega$).

II. DARK SRF PROJECTION

This section provides details for computing the sensitivity of superconducting radio frequency (SRF) cavities to relativistic monochromatic axion signals, using the conventions described in [49, 50]. The search for relativistic axions does not exactly map onto the search for non-relativistic axion DM, since spatial gradients cannot be ignored in the former case. In particular, the wave equation takes the form

$$\left(\frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2\right)\mathbf{E} = g_{a\gamma\gamma}\left((\boldsymbol{\nabla}a\cdot\boldsymbol{\nabla})\mathbf{B}_p - \dot{a}\frac{\partial\mathbf{B}_p}{\partial t}\right) , \tag{S20}$$

where a is the axion field and $\mathbf{B}_p$ is the magnetic field associated with the pump mode. The first term on the right-hand side in (S20) can be neglected as it is incapable of exciting resonant modes of a cavity [76]. The axion field drives power into the signal mode, with the following power spectral density (PSD) (see *e.g.*, [49]):

$$S_{\rm sig}(\omega) = \frac{\omega_1}{4Q_1}(g_{a\gamma\gamma}\eta_{10}B_p)^2 V_{\rm cav}\,\frac{\omega^2(\omega-\omega_0)^2 S_a(\omega-\omega_0)}{(\omega^2-\omega_1^2)^2+(\omega\omega_1/Q_1)^2} , \tag{S21}$$

where S_a is the PSD of the axion signal, ω_0 and ω_1 are the frequencies of the pump and signal modes, respectively, Q_1 is the quality factor of the signal mode, η_{10} is the mode overlap factor, and $V_{\rm cav}$ is the volume of the cavity. The PSD is defined using the following normalization:

$$\int S_a(\omega)d\omega = (2\pi)^2\frac{\rho_{\rm PS}}{\Omega^2} , \tag{S22}$$

where $\rho_{\rm PS}$ is the local axion energy density and Ω is the frequency of the axion signal. Note the distinction with the DM search in which the frequency is equal to the axion mass.

Since the axion coherence time is much greater than the integration time of the experiment, the axion signal PSD falls in a single frequency bin. The signal and noise powers measured in a frequency bin are determined by integrating their respective PSDs over the frequencies in the bin. On resonance, the splitting between the two modes is equal to the frequency of the axion signal: $\omega_1 = \omega_0 + \Omega$. The signal PSD may be approximated as a delta function, giving a total power of

$$\int_{\omega_1-\Delta\omega/2}^{\omega_1+\Delta\omega/2} S_{\rm sig}(\omega)d\omega = \pi^2(g_{a\gamma\gamma}\eta_{10}B_p)^2\left(\frac{Q_1}{\omega_1}\right)V_{\rm cav} , \tag{S23}$$

using the normalization of (S22). To obtain the sensitivity, we compute the total noise power within the frequency bin into which the axion signal falls. For the optimistic parameters presented in [49, 50], the dominant noise source at the relevant frequency is thermal noise with total bin power

$$\int_{\omega_1-\Delta\omega/2}^{\omega_1+\Delta\omega/2}\left(\frac{Q_1}{Q_{\rm int}}\right)\frac{4\pi T_{\rm cav}(\omega\omega_1/Q_1)^2}{(\omega^2-\omega_1^2)^2+(\omega\omega_1/Q_1)^2}d\omega = 4\pi T_{\rm cav}\left(\frac{Q_1}{Q_{\rm int}}\right)\Delta\omega. \tag{S24}$$

III. DMRADIO-GUT PROJECTION

To estimate the projected sensitivity for DMRadio-GUT, we follow the formalism in [105]. The detector consists of a large toroidal magnet and an external pickup loop inductively coupled to a SQUID magnetometer that measures the magnetic flux through the toroid. In the presence of a static magnetic field, the axion field generates an effective charge density, $\rho_{a,\rm eff} = -g_{a\gamma\gamma}\boldsymbol{\nabla} a\cdot\mathbf{B}$ (not to be confused with the axion energy density), and current density, $\mathbf{J}_{a,\rm eff} = g_{a\gamma\gamma}\dot{a}\mathbf{B}$. In axion DM experiments, the charge density is ignored as it is gradient-suppressed. For relativistic sources, such as pulsar-sourced axions, however, $|\boldsymbol{\nabla} a| \sim \dot{a}$, and the charge and current densities are comparable. However, the effect of the charge density is suppressed relative to that of the current density by a factor of $\Omega L_p \ll 1$, where Ω is the frequency of the axion signal and L_p is the inductance of the pickup loop. The magnetic flux through the pickup loop due to the oscillating axion field is

$$\Phi_{\rm pickup} = g_{a\gamma\gamma} B_{\rm max}\sqrt{2\rho_{\rm DS}} V_B \cos(\Omega t)\,, \tag{S25}$$

where $B_{\rm max}$ is the peak magnetic flux in the toroid, $\rho_{\rm DS}$ is the axion density in the detector, and V_B is the effective volume containing the magnetic field. The flux through the inductively-coupled SQUID is [105]

$$\Phi_{\rm SQUID} = \frac{\alpha}{2}\sqrt{\frac{L}{L_p}}\Phi_{\rm pickup}\,, \tag{S26}$$

where $\alpha \approx 0.7$ and L is the inductance of the SQUID, which we take to be $L \approx 1$ nH.

An axion DM search could employ either broadband or resonant readout, with the former being preferred at low frequency. The reason is that for a resonant search, the need to scan over DM mass significantly reduces the interrogation time in each frequency bin. In contrast, the signal frequency of pulsar-sourced axions is known to very high precision, set by the precision of timing measurements of pulsars. Thus, we focus on resonant readout with an LC circuit. The bandwidth of the readout circuit is $\Delta\omega_{\rm RO} = \omega/Q_0$, where $Q_0 = 1/(\omega RC)$ is the quality factor of the circuit and C (R) is the capacitance (resistance) of the capacitor, respectively.

We can model the system as an RLC circuit, with the pickup loop placed at the center of the toroid playing the role of the inductor. On resonance, the resistance of the circuit is related to the intrinsic quality factor through $R = \Omega L_p/Q_0$. The axion field drives an AC voltage in the readout circuit with PSD [106]

$$S^a_{VV}(\omega) = \pi\Omega^2 g^2_{a\gamma\gamma} B^2_{\rm max} c^2_{\rm pu} V_B^2 \rho_{\rm DS}\,\delta(\omega-\Omega)\,, \tag{S27}$$

where $c_{\rm pu}$ quantifies the fraction of the axion-sourced magnetic field that couples to the resonant circuit. We have assumed that the axion signal is sufficiently narrow that its spectrum is well-approximated by a delta function. In reality, the axion signal will be spread over the finite bandwidth of the circuit $\Delta\omega = \Omega/Q_0$. On resonance, the induced current PSD is $S^a_{II} = S^a_{VV}/R^2$, where R is the resistance defined above. One way to parameterize the noise contributions is with a "noise temperature," $k_B T_N(\omega) = N(\omega)\hbar\omega$, where k_B is the Boltzmann constant and $\hbar$ is the reduced Planck constant. $N(\omega)$ receives contributions from thermal noise, amplifier noise, and $1/f$ noise. The noise PSD may be written as follows [106]

$$S^{\rm noise}_{II}(\omega) = \frac{Q_0}{\pi\Omega L_p}(2k_B T_{\rm LC} + \eta_A\hbar\omega)\,, \tag{S28}$$

where $T_{\rm LC}$ is the temperature of the LC circuit and η_A describes how far the amplifier is from operating at the standard quantum limit (SQL), with $\eta_A = 1$ corresponding to the SQL (1/2 from quantum-limited amplifier noise and 1/2 from zero-point thermal fluctuations). For a pulsar-sourced axion search with $\Omega/(2\pi) \sim 30$ Hz, the contribution of $1/f$ noise cannot be neglected. In keeping with the convention of (S28), we absorb the contribution of $1/f$ noise into η_A. The flux PSD of $1/f$ noise is given by $S_\Phi(\omega) = A^2_\Phi(\omega/2\pi\times{\rm Hz})^\xi$, where $\xi \approx -1$ and, for the low temperatures considered here, $A^2_\Phi \lesssim (10^{-6}\Phi_0)^2/{\rm Hz}$, where $\Phi_0 = h/2e$ is the magnetic flux quantum, with h being the Planck constant and e being the magnitude of the electron charge [107]. On resonance, the contribution of $1/f$ noise to η_A is approximately 10^3. Since $T_{\rm LC} \gg \eta_A\Omega$, the dominant noise source is thermal. The signal and thermal noise at frequency $\omega_a = \Omega$ are

$$P_{\rm sig} = Q_0\frac{\Omega\Phi^2_{\rm pickup}}{2L_p}\,, \quad P_{\rm noise} = k_B T_{\rm LC}\times\frac{2\pi}{T_{\rm int}}\,, \tag{S29}$$

changed this equation. Check if you agree. where $T_{\rm int}$ is the integration time. We compute the sensitivity by setting $P_{\rm sig} = 8.48\times P_{\rm noise}$ (as in the main text) and use the following fiducial parameters roughly in line with those of DMRadio-GUT [108]: $B_{\rm max} = 16$ T, $c_{\rm pu} = 0.1$, $V_B = 10$ m^3, $L_p = \pi R = \pi\times 2.2$ m, $T = 10$ mK, $Q_0 = 2\times 10^7$, and $T_{\rm int} = 1$ year.

IV. CASPER PROJECTION

This section discusses detection of pulsar-sourced axions through their coupling to fermions. For non-relativistic nuclei, the axion field acts as an effective magnetic field with interaction Hamiltonian $H_{\rm int} = \gamma_{_N} \mathbf{B}_a \cdot \mathbf{S}_N$, where $\gamma_{_N}$ is the gyromagnetic ratio of the nucleon, $\mathbf{B}_a = -g_{aNN} \boldsymbol{\nabla} a/\gamma_{_N}$ is the effective axion magnetic field, and $\mathbf{S}_N$ is the nucleon spin operator. Several experiments, including CASPEr [40] and co-magnetometer experiments [88, 89], use nuclear magnetic resonance techniques to look for axion DM. This section computes the sensitivity of a CASPEr-wind-type experiment to pulsar-sourced axions. The starting point is a macroscopic sample of nuclear spins aligned using a strong, static magnetic field, $\mathbf{B}_E = B_E \hat{z}$. We assume the apparatus is aligned such that $\mathbf{B}_a = B_{a,0} \hat{x} \cos(\Omega t - k_a x)$, where $k_a^2 = \Omega^2 - m_a^2$. The axion magnetic field causes the spins to precess about $\hat{z}$, leading to an oscillating transverse magnetization of the sample. This effect is resonantly enhanced when the axion frequency matches the Larmor frequency, $\omega_0 = \gamma_{_N} B_E$. The width of the resonance is $\Delta\omega \simeq 1/T_2$, where T_2 is the transverse spin relaxation time. The evolution of the transverse magnetization, M_x, is described by the Bloch equations, which give

$$\ddot{M}_x + \frac{2}{T_2} \dot{M}_x + \omega_0^2 M_x = \gamma_{_N} M_0 \omega_0 B_{a,0} \cos \Omega t \,, \tag{S30}$$

where M_0 is the initial magnetization of the system and $B_{a,0} = -g_{aNN} k_a \sqrt{2\rho_a}/(\Omega \gamma_{_N})$. We have taken the system to be located at $x = 0$, assume that the size of the apparatus is much less than the axion wavelength, and work in the limit where $T_2 \ll T_{\rm int} \ll \tau_a$, with $T_{\rm int}$ the integration time and τ_a the coherence time of the signal. The coherence time of the axion signal is set by the spin-down parameters of the pulsar being observed, $P/\dot{P}$, which is in general much longer than a year (taken to be the fiducial integration time). To compute the sensitivity, we envision taking N measurements with interval $\Delta t = T_{\rm int}/N$. In frequency space, this corresponds to bins with frequency $\omega_n = 2\pi n/T_{\rm int}$. Since $\tau_a \gg T_{\rm int}$, the signal lies entirely within one bin with PSD

$$P_a = g_{aNN}^2 \rho_a k_a^2 M_0^2 \frac{T_2^2 T_{\rm int}}{2\Omega^2} \,. \tag{S31}$$

The axion-driven transverse magnetic field is measured using a SQUID magnetometer. The main sources of noise come from: thermal effects, spin projection noise, and quantum noise from the SQUID. As in [79], we assume the system temperature is sufficiently low that thermal noise is subdominant. Spin projection noise arises due to the intrinsic uncertainty in spin measurement along a given direction. The PSDs of spin projection and magnetometer noise on resonance are, respectively, [79]

$$P_{\rm SP} = \frac{\gamma_{_N}^2 n J T_2}{2V} \,, \quad P_{\rm SQUID} = \frac{1}{A_{\rm eff}^2} \frac{(10^{-6} \Phi_0)^2}{\rm Hz} \,, \tag{S32}$$

where n and V are the number density and volume of the sample, respectively, J is the nuclear spin, and $A_{\rm eff}$ is the effective area of the SQUID pickup loop.