

Improved cosmological limits on Z' models with light right-handed neutrinos

Tim Herbermann ^a and Manfred Lindner ^a

^aMax-Planck-Institut für Kernphysik,
Saupfercheckweg 1, 69117 Heidelberg, Germany

E-mail: tim.herbermann@mpi-hd.mpg.de, lindner@mpi-hd.mpg.de

ABSTRACT: We improve limits on Z' extensions of the Standard Model (SM) with light right-handed neutrinos. The presence of shared gauge interactions between the light right-handed neutrinos and other SM fermions allows for production of ν_R in the early Universe and we use the excess in the effective number of neutrino species ΔN_{eff} to place limits. Our benchmark model is a minimal gauged $U(1)_{B-L}$ that often arises as a building block in other models, and we discuss applicability to more general $U(1)$ extensions. We devise an improved Monte Carlo integration scheme convenient for implementation of generic integrated Boltzmann equations with minimal simplifying assumptions. We sketch our numerical implementation in detail for future reference. Using the new ACT DR6 limit $\Delta N_{\text{eff}} < 0.17$, we improve constraints on the gauge coupling for $1 \text{ GeV} < m_{Z'} < 100 \text{ TeV}$ by orders of magnitude and find the strongest limits thus far, surpassing even current and future colliders, and explore the potential of future CMB experiments to test $U(1)$ extensions up to the GUT scale. We perform a detailed analysis of the robustness of cosmological limits within standard and non-standard thermal histories and find that a strong first order phase transition, early dark energy or early matter domination could dilute ν_R abundances beyond detection. We investigate the effect of reheating on ν_R -genesis and provide results and prescriptions to apply our bounds to non-standard thermal histories. Limits are generically weakened for reheating $T_{\text{reh}} \ll m_{Z'}$. Our results suggest that projected limits on Z' with Dirac neutrinos can only be accommodated for in non-standard thermal histories, thus limiting the options to include dark matter candidates or Dirac leptogenesis.

KEYWORDS: particle physics - cosmology connection, cosmological neutrinos, physics of the early universe

ARXIV EPRINT: [2505.04695](https://arxiv.org/abs/2505.04695)

Contents

1	Introduction	1
2	Dirac neutrinos in the ΔN_{eff} landscape	3
3	Minimal gauged $B - L$ with Dirac neutrinos	4
4	Boltzmann equations	5
4.1	General framework	6
4.2	Finite temperature effects	7
4.3	Resonances	8
5	Results	11
5.1	Standard cosmology	11
5.2	Non-standard cosmologies	13
6	Conclusion	18
A	Integrating collision operators	20
A.1	$2 \rightarrow 2$ processes	20
A.2	$1 \rightarrow 2$ processes	21
A.3	Performance and other implementation details	22
B	Matrix elements, cross sections and decay rates	22

1 Introduction

The nature of neutrinos is one of the big open questions in neutrino physics. While the Standard Model (SM) of particle physics predicts massless neutrinos, neutrino oscillations and the implied non-vanishing neutrino masses are an experimental reality (see e.g. [1]). What these observations do not yet reveal is whether the active neutrinos are of Majorana or Dirac type. A possible observation of neutrinoless double beta decay [2] will clarify the situation — with profound consequences beyond neutrino mass model building. As far as cosmology is concerned, one major difference between Dirac and Majorana neutrinos is that the former come with light right-handed neutrinos that could give an extra contribution to the total radiation budget of the Universe.

Cosmological excess radiation is conventionally normalized to the energy density of one active neutrino flavor, $N_{\text{eff}} = (8/7) (11/4)^{4/3} \rho_{\text{rad}}/\rho_{\gamma}$. The theoretical SM expectation is $N_{\text{eff}} = 3.044$ [3–6],¹ and deviations from this expectation are usually parametrized as

¹Current theoretical predictions prefer this value, with some concurrent studies preferring $N_{\text{eff}} = 3.045$. We adopt the aforementioned value, but note that this small difference will not affect our conclusions in a meaningful way.

$\Delta N_{\text{eff}} = N_{\text{eff}} - 3.044$. This is a powerful probe of any SM extension that introduces new light degrees of freedom.

If the extra degrees are in chemical equilibrium in the early Universe, their relic abundance and contribution to the effective number of neutrinos can readily be estimated by their decoupling temperatures as [7]

$$\Delta N_{\text{eff}} \simeq 0.027 g_x \left(\frac{106.75}{g_*(T_{\text{dec}})} \right)^{4/3}, \quad (1.1)$$

where g_x is the (effective) internal degrees of freedom for a light species x . The presence of three ν_R that left equilibrium with the SM at the electroweak scale or above would contribute a significant $\Delta N_{\text{eff}} \simeq 0.14$. However, it is possible that interactions are too feeble to ever thermalize light right-handed neutrinos, and akin to the often studied freeze-in production of dark matter [8], a significant relic abundance can still be realized [9, 10].

Bounds on Dirac neutrinos in context of Z' extensions more generally and gauged $U(1)_{B-L}$ in particular have been studied before, with limits based on the Planck best fits [11] by adopting either simplified scattering rate approaches or more sophisticated calculations involving Boltzmann equations. The considered models constitute both minimal extensions [7, 10, 12–14] and extended models that may include dark matter candidates, or address other problems [15–26]. It is possible that extended sectors also affect ΔN_{eff} in non-trivial ways, even reducing it in some circumstances [27, 28]. Such studies typically assume a standard thermal history, a plausible but strong assumption that deserves scrutiny and that was already partially relaxed in [14].

In what follows, we present an improved calculation of the relic abundance of right-handed neutrinos under the benchmark assumption that $B - L$ is a gauged symmetry. The particle model we consider is minimal, but arises as a building block in all models that have Dirac neutrinos and gauged $B - L$.² The ν_R abundance that arises from this building block can thus be considered an irreducible contribution that deserves extra scrutiny. Similar structures will naturally arise in all models that introduce Dirac neutrinos that share gauge interactions with at least a subset of the SM. Therefore, results we find within our benchmark study will approximately hold also for any other Z' extension that couples ν_R and SM fermions.

We calculate the excess radiation from ν_R by solving the respective Boltzmann equations. Additional focus is given to the cosmological assumptions and we analyze in detail the effects of non-standard thermal histories. The framework we develop is exact on the level of integrated Boltzmann equations with only minimal additional assumptions. Key part of our computation is the efficient Monte Carlo integration of the relevant collision operators, and we incorporate in particular leading thermal corrections and the effects of the final state statistics from the finite density environment. While it is not expected that these effects give rise to significant corrections for ν_R -genesis, they are straightforward to include in Monte Carlo approaches. The ansatz we sketch should prove useful for the study of other systems of Boltzmann equations beyond production of ν_R , where such effects can be significant.

In section 2, we summarize the current experimental state of ΔN_{eff} , and how it relates to Dirac neutrino model building. We introduce the minimal model for this study in section 3.

²While we only consider a Dirac nature of neutrinos here, our results also apply to light right-handed neutrinos in the Majorana case as long as $m_{\nu_R} < T_{\text{CMB}}$ and the ν_R contribute as cosmological radiation.

Then, we develop an almost approximation-free approach to integrated Boltzmann equations in section 4, drawing on the ansatz in [9] and developing it further. We discuss our improved limits and forecasts in section 5, including a comprehensive discussion of their robustness and required modifications beyond the standard cosmological picture. We conclude in section 6.

2 Dirac neutrinos in the ΔN_{eff} landscape

If neutrinos are Dirac particles, the light right-handed extra degrees of freedom contribute to the radiation budget of the Universe. Excess radiation is constrained by precise measurements of the angular scale of acoustic peaks in the cosmic microwave background (CMB) and its damping tail [29], and by the primordial abundance of elements from Big Bang Nucleosynthesis (BBN). Proposals to use either CMB features or primordial element abundances to constrain extra radiation are old and have been considered e.g. in [30–32]. Since then, the bounds on the effective number of neutrino species ΔN_{eff} tightened significantly — with Planck indicating $\Delta N_{\text{eff}} < 0.285$ at 95% C.L. [7, 11], cosmology provided valuable input for models that predict new light degrees of freedom.

Recently, the data release 6 of the Atacama Cosmology Telescope (ACT) [33, 34] has improved on these limits. Utilizing the full primary CMB data from ACT in combination with large and medium scale data from Planck, the limit is pushed down to $\Delta N_{\text{eff}} < 0.17$ at 95% C.L., the strongest limit on relativistic relics so far.

These results are complimented by joint CMB+BBN analyses, which combine data sets across the cosmological history and find agreeable limits of $\Delta N_{\text{eff}} < 0.180$ [35] and $\Delta N_{\text{eff}} < 0.163$ [36] at 95% C.L. respectively. An incomplete subset combination of SPT-3G polarization and lensing data, when combined with Planck, ACT and BAO, finds $N_{\text{eff}} = 2.86 \pm 0.13$, $N_{\text{eff}} = 2.83 \pm 0.13$, and $N_{\text{eff}} = 2.89 \pm 0.23$ when allowing for extended $\Lambda\text{CDM} + N_{\text{eff}}$, $\Lambda\text{CDM} + N_{\text{eff}} + \sum m_\nu$, and $\Lambda\text{CDM} + N_{\text{eff}} + Y_p$ fits respectively [37], giving an indication that cosmological data continues to disfavor large amounts of extra radiation and future searches will continue to tighten constraints on SM extensions with new light degrees of freedom.

Forecasted sensitivities of ongoing and planned experiments surpass these constraints, and Simons Observatory (SO) [38] and SPT-3G [39] are projected to reach around $\Delta N_{\text{eff}} < 0.12$ at 95% C.L. [7], and CMB-S4 [40] anticipates to reach $\Delta N_{\text{eff}} < 0.06$. Futuristic proposals like CMB-HD [41] are designed with the goal of breaching the important $\Delta N_{\text{eff}} = 0.027$ benchmark of a single thermalized Nambu-Goldstone boson. We conclude that current cosmological bounds are close to the important benchmark of three fully thermalized ν_R , and with the indication that ongoing or upcoming next-generation experiments will push below this important threshold soon.

Assuming that the Dirac mass is generated by the SM Higgs, and that the interactions of the right-handed partners are restricted to Yukawa interactions, current neutrino mass bounds imply $\Delta N_{\text{eff}} < 10^{-11}$ [9], a contribution too small to ever be within experimental reach. However, it is generally expected that the neutrino sector encompasses larger extensions. First, to explain the smallness of neutrino masses when compared to the SM fermions, and second, to preserve their Dirac nature if applicable. The method of choice is usually by means of a (global or local) symmetry that forbids the presence of a Majorana mass term. While it is perfectly fine to protect the Dirac nature by means of a global symmetry, the

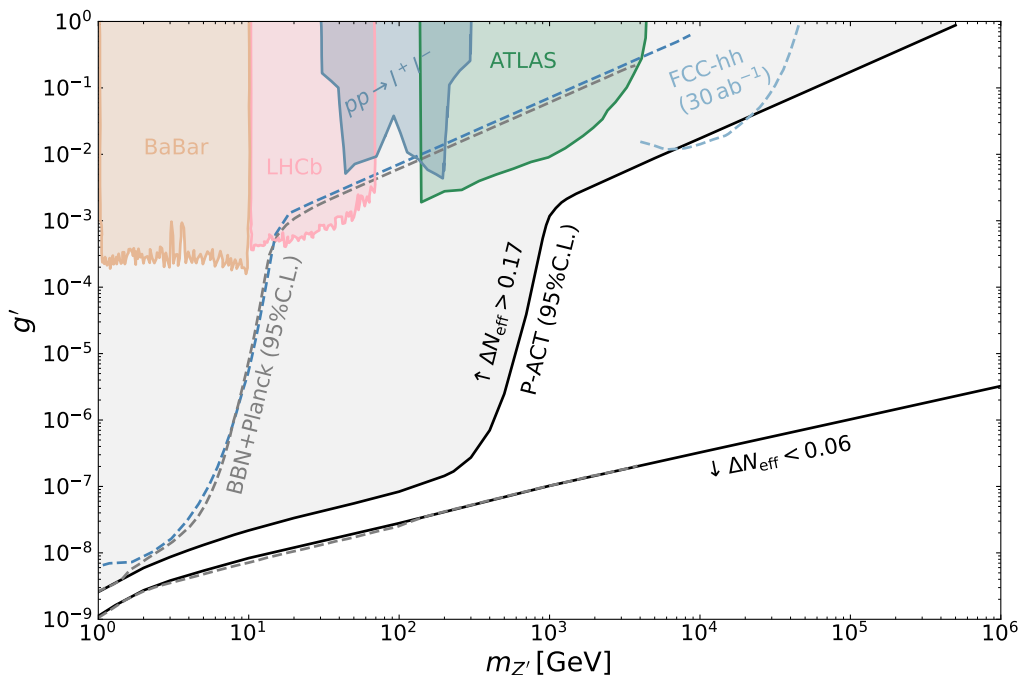


Figure 1. Constraints on $U(1)_{B-L}$ gauge bosons in presence of light right-handed neutrinos. The black solid curves are updated contours of ΔN_{eff} . The upper one corresponds to $\Delta N_{\text{eff}} = 0.17$, excluded by ACT [34] and disfavored by current BBN+CMB combinations [35, 36]. The lower curve is $\Delta N_{\text{eff}} = 0.06$, which serves as a benchmark for future CMB experiments (see section 2). We also show previous forecasts and Planck+BBN limits from References [10](grey dashed) and [7](blue dashed). We show complimentary limits from ATLAS [43], limits from the Drell-Yan process in ATLAS [42, 44], a recasting of LHCb limits [42, 45], and dark photon decays from BaBar [42, 46]. For comparison, we show the projected reach of FCC-hh for Z' searches from [47].

possibility of quantum gravity to break global symmetries would suggest that a protection by means of a gauge symmetry is the preferred option. Among the most popular and well motivated options is the promotion of the accidental $U(1)_{B-L}$ symmetry of the SM to a local one. This unavoidably necessitates the introduction of ν_R to cancel gauge anomalies, thus making the presence of right-handed neutrinos a requirement for internal consistency. We note here that $B-L$ extensions of the SM, although well motivated, are already subject to strong laboratory constraints even beyond the Dirac hypothesis (e.g. [42]).

3 Minimal gauged $B-L$ with Dirac neutrinos

Our benchmark model for Z' extensions is a gauged $U(1)_{B-L}$, and we adopt a minimal setup. The derived constraints will either be exact in the case that additional model constituents are feebly coupled, or conservative, as extra interactions tend to further boost production of ν_R . Therefore, we introduce three right-handed partners ν_R to the SM neutrino and

augment the Lagrangian density by

$$\mathcal{L} \supset \frac{1}{4} F'_{\mu\nu} F'^{\mu\nu} + \frac{1}{2} m_{Z'}^2 Z'_\mu Z'^\mu + g' Z'_\mu \sum_{f \in \text{SM}} Q_{B-L}^{(f)} \bar{f} \gamma^\mu f + g' Z'_\mu Q_{B-L}^{(\nu_R)} \bar{\nu}_R \gamma^\mu \nu_R, \quad (3.1)$$

and made explicit the interactions with SM fermions (including the left neutrino) and the light right-handed neutrinos. The SM $B - L$ gauges are $1/3$ for quarks and -1 for leptons. In the absence of other fermions charged under $U(1)_{B-L}$, we are restricted to choose either $Q_{B-L}^{(\nu_R)} = -1$ or $Q_{B-L}^{(\nu_R)} = -4, -4, 5$ to cancel gauge anomalies. We remain agnostic about the origin of the Dirac mass for neutrinos. For the -1 choice of gauge charges, it is possible to write down Yukawa terms with the Higgs doublet. The Dirac nature of the neutrino is then protected by the gauge symmetry and the model as it stands can be considered complete up to the origin of the Z' mass. Whether it is realized as a Stueckelberg mass, or if an extended scalar sector breaks $\Delta(B - L) \neq 2$ and induces a mass term for the Z' is largely irrelevant to obtain robust constraints on this class of models. In the case of $Q_{B-L}^{(\nu_R)} = -4, -4, 5$, a Dirac mass from the Higgs only is not possible. This necessarily implies that such a model is incomplete already on the level of ν_R mass generation. Still, it would be possible to derive constraints that are conservative, or exact if the extended sector is largely decoupled from the SM and ν_R production. We will not explicitly compute limits for this particular case, but approximate limits can be obtained from appropriate rescaling of the gauge coupling. Likewise, other Z' extensions can also be approximately constrained by rescaling.

If $B - L$ remains unbroken, the Lagrangian in Equation 3.1 should contain kinetic mixing terms $\mathcal{L} \supset \frac{\epsilon}{2} Y_{\mu\nu} F'^{\mu\nu}$, since such a mixing term will inevitably be generated from radiative corrections. As far as mixing to the photon is concerned, exact cancellations from destructive interference are fine-tuned and not stable under running, and cancellations from mixing to the Z cannot be exact due to the chiral nature of the latter. Thus, if $B - L$ remains unbroken and kinetic mixing is necessarily generated, it can at most provide a minor correction of the effective coupling that we probe, and the effect can again mostly be absorbed in a rescaling of gauge coupling.

4 Boltzmann equations

The abundance of cosmological relics is conveniently calculated in the framework of Boltzmann equations (e.g. [48]). For our purposes, it is not necessary to study Boltzmann equations on phase space level, as results of integrated Boltzmann equations for number densities or energy densities usually give excellent results compared to phase space solutions (see e.g. [49]). In the following, we sketch an alternative numerical approach to integrated Boltzmann equations without further approximations by means of Monte Carlo integration of the collision operator. This is complementary to available numerical schemes (e.g. [50]) that also aim to include relativistic effects and quantum statistics. While such schemes are well suited for computational efficiency, we can allow to make a few drawbacks on performance but enjoy a more straightforward implementation. We draw in particular on Reference [9], and develop the Monte Carlo approach further. More details on the numerical implementation are given in appendix A.

4.1 General framework

We simplify the problem by treating the right-handed neutrinos and anti-neutrinos as a common species with $g_{\nu_R} = 2 \times 3$ internal degrees of freedom [9, 10, 51]. As long as Z' couples (approximately) universal to ν_R , this is a good approximation. The Boltzmann equation for the energy density of the collective ν_R species and the SM are given by

$$\begin{aligned}\frac{d\rho_{\nu_R}}{dt} &= -4H\rho_{\nu_R} + \mathcal{C}_{\text{SM}\rightarrow\nu_R}^{(\rho)}, \\ \frac{d\rho_{\text{SM}}}{dt} &= -3H(\rho_{\text{SM}} + P_{\text{SM}}) - \mathcal{C}_{\text{SM}\rightarrow\nu_R}^{(\rho)}.\end{aligned}\tag{4.1}$$

Here, the integrated collision operator describes the energy transfer rate from the SM plasma to the ν_R sector. The Hubble rate is $H^2 = \frac{8\pi G}{3}(\rho_{\text{SM}} + \rho_{\nu_R})$. In its most general form, the collision operator can be written as a sum over any number of initial states i and final states f that contain at least one $\nu_R(\bar{\nu}_R)$

$$\begin{aligned}\mathcal{C}_{i\rightarrow f}^{(\rho)} &= \sum_{i,f\in\nu_R} \int d\Pi_{i_1}\dots d\Pi_{i_n} d\Pi_{f_1}\dots d\Pi_{f_m} (2\pi)^4 \delta^{(4)} \left(\sum p_i - \sum p_f \right) E_{\nu_R} \\ &\times \left(|\mathcal{M}_{i\rightarrow f}|^2 f_{i_1}\dots f_{i_n} \bar{f}_{f_1}\dots \bar{f}_{f_m} - |\mathcal{M}_{f\rightarrow i}|^2 f_{f_1}\dots f_{f_m} \bar{f}_{i_1}\dots \bar{f}_{i_n} \right).\end{aligned}\tag{4.2}$$

Here we introduced several shorthands, that we will use frequently. We denote the Lorentz invariant phase space element as $d\Pi_i = \frac{d^3 p_i}{(2\pi)^3 2E_i}$, the phase space density of x is $f_x = f_x(E_x)$, and $\bar{f}_x = 1 \pm f_x$ is the final state phase space and the sign depends on whether particle x is a boson (fermion). We will further assume CP symmetry such that $|\mathcal{M}_{i\rightarrow f}|^2 = |\mathcal{M}_{f\rightarrow i}|^2$ holds. Note that in our convention, the matrix element already contains the symmetry factor $1/n!$ for identical particles in the initial and final states [48]. We restrict the analysis to $2 \leftrightarrow 2$ and $2 \leftrightarrow 1$ processes, but a generalization to $n \rightarrow m$ processes would be straightforward. Further, we assume that the SM remains in thermal equilibrium at all times.³

If the SM maintains equilibrium, then

$$\rho_{\text{SM}} = \frac{\pi^2}{30} g_{(\rho)} T^4, \quad P_{\text{SM}} = \frac{\pi^2}{90} g_{(P)} T^4,\tag{4.3}$$

where we allow for the small difference between $g_{(\rho)}$ and $g_{(P)}$ that occurs during phase transitions or if a constituent becomes non-relativistic and disappears from the plasma [52], and we use the numerical values from Reference [53].

To close the system of equations, we prescribe equilibrium phase space distributions. By assumption, we can incorporate equilibrium distributions for all SM particles. We make the same prescription for ν_R , and we compute the temperature from inverting the energy density of ν_R under the equilibrium assumption [9, 10]. For as long as the right-handed neutrinos are chemically coupled to the SM or maintain internal equilibrium, this approach is exact. During the decoupling process from the SM plasma, the distribution maintains its equilibrium shape up to spectral distortions. In the opposite limit, where ν_R never attains

³It would be interesting to consider cases where e.g. resonantly enhanced processes could cause brief violations of the equilibrium assumption. This would need to be studied on the phase space level and is beyond the scope of this work.

chemical equilibrium with the SM, the backreaction and final state occupation is suppressed and insensitive to the underlying phase space density, thus making this choice a reasonable approximation in both limits. Only an intermediate regime, close to thermalization of ν_R is sensitive to the precise phase space distribution, but our prescription can be understood as a smooth interpolation between the two limiting cases.

We integrate the collision operator numerically without any additional approximations. This makes the result exact on the level of integrated Boltzmann equations and the phase space density prescription we adopt. This can be done by performing a subset of the integrals in the collision operator analytically, and then evaluating the remaining integrals numerically. Since the integrals are still highly dimensional, it is advantageous to resort to Monte Carlo integration [9]. The details of the numerical implementation are given in appendix A, and we find an acceptable performance given the ease of implementation. This complements previous approaches to production of light right-handed neutrinos in the early Universe, where the quantum statistics of final states in particular are often neglected and simplified prescriptions to include the backreaction are employed [9, 10, 54–56], or complete but rather complicated schemes to include also the quantum statistics of final states [50, 57]. We stress again that with straightforward numerical integration we can consistently include final state statistics and the backreaction as it appears in the collision operator. However, in the scenario we study here we do not expect these corrections to be significant.

We integrate the set of Boltzmann equations Equation 4.1 from an initial reheating temperature T_{reh} up until the energy transfer is completed ($\mathcal{C}/H \ll 1$). Then, ΔN_{eff} can be inferred from the relation [9]

$$\Delta N_{\text{eff}} \simeq \frac{4}{7} g(\rho) \left(\frac{10.75}{g(s)} \right)^{4/3} \frac{\rho_{\nu_R}}{\rho_{\text{SM}}}, \quad (4.4)$$

which holds for all temperatures after the ν_R abundance is set and ν_R and the SM are no longer in contact.

4.2 Finite temperature effects

In a dense plasma, the dispersion relation of particles is modified by finite temperature corrections. While a full treatment at the level of thermal quantum field theory is beyond the scope of our approach here, the leading effect is often captured by introducing thermal masses [55, 57, 58]. We adopt this prescription here to account for leading thermal effects. To this end, all masses are replaced by the corresponding thermal masses in a plasma. Before electroweak symmetry breaking (EWSB), the thermal masses for the fermions are chiral,

giving different corrections to left- and right-handed fields [58]

$$m_{l,L}^2 = \frac{m_Z^2 + 2m_W^2 + m_l^2}{2v_H^2} T^2, \quad (4.5)$$

$$m_{l,R}^2 = \frac{m_Z^2 - m_W^2 + \frac{1}{2}m_l^2}{2v_H^2} T^2, \quad (4.6)$$

$$m_{q,L}^2 = \frac{1}{6}g_s^2 T^2 + \frac{3m_W^2 + \frac{1}{9}(m_Z^2 - m_W^2) + m_u^2 + m_d^2}{2v_H^2}, \quad (4.7)$$

$$m_{u,R}^2 = \frac{1}{6}g_s^2 T^2 + \frac{\frac{4}{9}(m_Z^2 - m_W^2) + \frac{1}{2}m_u^2}{2v_H^2}, \quad (4.8)$$

$$m_{d,R}^2 = \frac{1}{6}g_s^2 T^2 + \frac{\frac{1}{9}(m_Z^2 - m_W^2) + \frac{1}{2}m_d^2}{2v_H^2}, \quad (4.9)$$

and after EWSB, the Higgs induced mass dominates but is corrected by [57]

$$\Delta m_l^2 = \frac{1}{8}e^2 T^2, \quad (4.10)$$

$$\Delta m_q^2 = \frac{1}{6}g_s^2 T^2. \quad (4.11)$$

Note that in principle we would have to add the contributions from the new Z' boson, whenever it is part of the plasma. However, throughout this study we are always in regimes where either $T \ll m_{Z'}$, or $g' \ll 1$ (or both), such that the Z' induced corrections are negligible. Likewise, gauge bosons receive an effective thermal mass which for the Z' is [57]

$$m_{Z'}^2 = \frac{1}{6}g'^2 T^2 (N_V + N_S + \frac{N_F}{2}), \quad (4.12)$$

where $N_{S,F,V}$ is the number of scalar, fermionic and vector fields in the bath and coupled to Z' and again, we are studying regions of parameter space where the thermal correction is small.

4.3 Resonances

Some extra care is needed for s-channel resonances, in particular if the mediator that goes on a resonance is also in thermal equilibrium, as issues of double counting arise [55, 57, 59]. To model the s-channel resonance, we adopt the Breit-Wigner form of the propagator by making the replacement

$$\frac{1}{(s - m_{Z'}^2)^2} \rightarrow \frac{1}{(s - m_{Z'}^2)^2 + m_{Z'}^2 \Gamma_{Z'}^2}. \quad (4.13)$$

In the parameter space of interest, $\Gamma_{Z'} \ll m_{Z'}$ holds and it is convenient to use the narrow width approximation (NWA). In the NWA

$$\frac{1}{(s - m_{Z'}^2)^2 + m_{Z'}^2 \Gamma_{Z'}^2} \rightarrow \frac{\pi}{m_{Z'} \Gamma_{Z'}} \delta(s - m_{Z'}^2), \quad (4.14)$$

and the on-shell production of mediators and their subsequent decay is resonantly enhanced. In case of a scalar resonance, the amplitude may be written as a sum of the resonant part with

the Breit-Wigner prescription in the NWA, and the product of the amplitudes for production from inverse decays and subsequent decays. Following Reference [57], this would correspond to

$$\left| \mathcal{M}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} \right|^2 = \frac{\left| \mathcal{M}_{f\bar{f} \rightarrow Z'} \right|^2 \left| \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R} \right|^2}{m_{Z'} \Gamma_{Z'}} \pi \delta(s - m_{Z'}^2). \quad (4.15)$$

For vector resonances, we have the added complication of spin correlations, and the amplitude can not be factorized exactly as above [57, 60, 61]. However, we can introduce the decorrelated matrix element, see e.g. [61], to find a NWA prescription. First, we observe that for a vector in the s-channel,

$$\begin{aligned} \left| \mathcal{M}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} \right|^2 &\propto \left| \mathcal{M}_{f\bar{f} \rightarrow Z'}^\mu \left(g_{\mu\nu} - \frac{k^\mu k^\nu}{m_{Z'}^2} \right) \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R}^\nu \right|^2 \\ &= \left| \sum_\lambda \mathcal{M}_{f\bar{f} \rightarrow Z'}^\mu \varepsilon_\mu^*(k, \lambda) \varepsilon_\nu(k, \lambda) \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R}^\nu \right|^2. \end{aligned} \quad (4.16)$$

The NWA for a vector resonance corresponds to making the replacement

$$\left| \sum_\lambda \mathcal{M}_{f\bar{f} \rightarrow Z'}^\mu \varepsilon_\mu^*(k, \lambda) \varepsilon_\nu(k, \lambda) \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R}^\nu \right|^2 \rightarrow \frac{1}{3} \sum_{\lambda, \lambda'} \left| \mathcal{M}_{f\bar{f} \rightarrow Z'}^\mu \varepsilon_\mu^*(k, \lambda) \right|^2 \left| \varepsilon_\nu(k, \lambda) \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R}^\nu \right|^2, \quad (4.17)$$

which is correct up to neglecting spin correlations between the initial state and final state. It has been shown that for sufficiently inclusive rates such as total cross sections, spin correlations have a negligible contribution [61].⁴ We show explicitly in appendix B that the exact matrix element with NWA propagator, and the decorrelated matrix element give the same resonant cross section in vacuum. Since we assume that the phase space distributions of different spin states are identical for any species, we do not expect to introduce additional correlations.

Then, as a prescription in the resonant regime, we write

$$\begin{aligned} \mathcal{C}_{f\bar{f} \leftrightarrow \nu_R \bar{\nu}_R} &= \int d\Pi_1 d\Pi_2 d\Pi_3 d\Pi_4 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4) \frac{\left| \mathcal{M}_{f\bar{f} \rightarrow Z'} \right|^2 \left| \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R} \right|^2}{2E_{Z'} m_{Z'} \Gamma_{Z'}} \\ &\quad \times \pi \delta(E_1 + E_2 - E_{Z'}) \left(f_1 f_2 \bar{f}_3 \bar{f}_4 - f_3 f_4 \bar{f}_1 \bar{f}_2 \right) \\ &\equiv \mathcal{C}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} - \mathcal{C}_{f\bar{f} \leftarrow \nu_R \bar{\nu}_R}. \end{aligned} \quad (4.18)$$

Here, the indices 1, 2 correspond to the initial state fermions f, \bar{f} , and indices 3, 4 denote the final state $\nu_R, \bar{\nu}_R$.

As was shown in reference [57], if we focus on the forward collision operator $\mathcal{C}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R}$ and neglect the backreaction, but assume that the mediator is in thermal equilibrium with

⁴In our case production proceeds from a thermal bath, and all possible spin configurations are assumed to be equally likely.

the SM, we can write the expression as

$$\begin{aligned}
\mathcal{C}_{f\bar{f}\rightarrow\nu_R\bar{\nu}_R} &= \int d\Pi_1 d\Pi_2 \frac{2\pi \delta^{(1)}(E_1 + E_2 - E_{Z'})}{4E_{Z'}\Gamma_{Z'}m_{Z'}} \left| \mathcal{M}_{Z'\rightarrow f\bar{f}} \right|^2 f_1 f_2 \\
&\quad \times \int d\Pi_3 d\Pi_4 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4) \left| \mathcal{M}_{Z'\rightarrow\nu_R\bar{\nu}_R} \right|^2 \bar{f}_3 \bar{f}_4 \\
&= \int d\Pi_3 d\Pi_4 d\Pi_{Z'} (2\pi)^4 \delta^{(4)}(p_3 + p_5 - p_{Z'}) \left| \mathcal{M}_{Z'\rightarrow\nu_R\bar{\nu}_R} \right|^2 f_3 f_4 \bar{f}_{Z'} \frac{\hat{\Gamma}_{Z'\rightarrow f\bar{f}}}{f_{Z'}\Gamma_{Z'}} ,
\end{aligned} \tag{4.19}$$

where to arrive in the last line we expand with $\frac{\bar{f}_{Z'}}{f_{Z'}}$ and complete the momentum conserving delta function for Z' with a dummy integration over $p_{Z'}$. We also exploited the equilibrium assumption between f and Z' , which allows us to interchange $f_{Z'}\bar{f}_1\bar{f}_2 = f_1f_2\bar{f}_{Z'}$. For convenience, we define the shorthand

$$\hat{\Gamma}_{Z'\rightarrow f\bar{f}} = \frac{1}{2m_{Z'}} \int d\Pi_1 d\Pi_2 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_{Z'}) \left| \mathcal{M}_{Z'\rightarrow f\bar{f}} \right|^2 (1 - f_1)(1 - f_2) . \tag{4.20}$$

which can be interpreted as a decay width that is corrected by occupation of final states.

The general expectation is that if f and Z' are in thermal equilibrium, production of ν_R should be equivalent to considering decays of Z' from the thermal bath, or in other words that Equation 4.19 coincides with

$$\mathcal{C}_{Z'\rightarrow\nu_R\bar{\nu}_R} = \int d\Pi_1 d\Pi_2 d\Pi_{Z'} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_{Z'}) \left| \mathcal{M}_{Z'\rightarrow f\bar{f}} \right|^2 f_1 f_2 \bar{f}_{Z'} \tag{4.21}$$

which implies $\Gamma_{Z'} = \frac{1}{f_{Z'}} \sum_f \hat{\Gamma}_{Z'\rightarrow f\bar{f}}$ which we can understand as the appropriate Breit-Wigner width in a plasma with thermalized mediator, in accordance with the result of [57]. In the presence of non-equilibrium processes, e.g. decays into final states that are not part of the equilibrium plasma, a minor modification is necessary (see discussion in [57]). Essentially, the distribution of Z' deviates from an equilibrium distribution (its distribution is no longer set by equilibrating decays and inverse decays of plasma constituents, but also the out-of-equilibrium final states). This can be captured by adding these invisible decays to $\Gamma_{Z'}$ and rescaling Equation 4.21 accordingly.

If the abundance of ν_R is set by freeze-in from the resonant contribution, but Z' is in equilibrium with the SM, it suffices to study the decay of on-shell mediators from the plasma. Since the off-resonance contribution is suppressed in higher orders of the gauge coupling, production is dominated by the resonance, or equivalently, the decay. We will study the latter for the case that Z' is thermalized, as it is particularly convenient.

If Z' is not in equilibrium with the SM, we start again from Equation 4.18. We perform the same computation as before, i.e. considering only the forward contribution, but this time we enforce $f_{Z'} \ll 1$, as the mediator abundance is suppressed, and readily arrive at

$$\mathcal{C}_{f\bar{f}\rightarrow\nu_R\bar{\nu}_R} = \int d\Pi_1 d\Pi_2 d\Pi_{Z'} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_{Z'}) \left| \mathcal{M}_{f\bar{f}\rightarrow Z'} \right|^2 \hat{\text{Br}}_{Z'\rightarrow\nu_R\bar{\nu}_R} . \tag{4.22}$$

Again, the hat indicates that $\hat{\text{Br}}_{Z'\rightarrow\nu_R\bar{\nu}_R}$ is the branching ration in medium that includes final state occupations.

Neglecting the final state statistics remains correct as far as the frozen-in species is concerned explicitly. However, in the case of resonant production without a mediator in equilibrium, the production has a dependence on the branching ratio in medium — which is affected by the final state occupation of SM particles and constitutes a correction due to final state statistics. Thus, we observe an indirect final state dependence for the frozen-in species from the medium dependent decay width into SM fermions which can have profound consequences, however in the situations we consider this effect is expected to be negligibly small (see also discussion in Reference [50]).

To summarize, in the regime where the mediator is thermalized and the explicit $2 \rightarrow 2$ contribution is suppressed compared to the resonant part, we study decays instead of the $2 \rightarrow 2$ process, and we avoid the issue of double counting. If the mediator is no longer thermalized, we use Equation 4.22 for freeze-in production, as the abundance of ν_R is also suppressed. We see here that such a prescription explicitly arises from treating the resonant part of the $2 \rightarrow 2$, while bearing in mind that the production occurs in a plasma background, thus making the possible double counting issue manifest.

Lastly, we highlight that the approach and the numerical scheme we develop here and in appendix A are general, and can easily be adapted to other systems of integrated Boltzmann equations, e.g. dark matter relic abundance calculations, and provide an alternative to either other approximate techniques or rather involved complete inclusions of final state effects. The ease of implementation compared to exact reduced integral expression as in e.g. [50, 54, 56, 57] would make such an approach lucrative if the problem is not limited by computational resources. Moreover, extending the numerical integration beyond $2 \rightarrow 2$ processes is straightforward and we can keep the consistent inclusion of relativistic and spin-statistic effects in this case.

5 Results

5.1 Standard cosmology

Before discussing the results in more detail, let us briefly stress again the explicit and implicit assumptions we made. Neutrinos are Dirac in nature, and their Dirac nature is in particular protected by $U(1)_{B-L}$, implying that if this symmetry is broken, $\Delta(B-L) \neq 2$. Moreover, we have an implicit assumption on the reheating temperature. We either consider thermalized ν_R that leaves equilibrium at some point, or freeze-in production which peaks at $T \sim m_{Z'}/3$. In either case, for the result to be applicable we need the Universe to reheat to $T_{\text{reh}} \sim m_{Z'}$, as the bulk of ν_R production will occur for temperatures smaller than the mass.

We show the improved limits and forecasts in figure 1. The current experimental bound of $\Delta N_{\text{eff}} \simeq 0.17$ implies that if the three ν_R have been in thermal equilibrium with the SM, they must have decoupled at a temperature of about $T_{\text{dec}} \sim 40 \text{ GeV}$ (see Equation 1.1). For the freeze-out limit, we compared the result of our exact computation including thermal masses with simplified prescriptions for the backreaction term [9, 25], no final state statistics, and vacuum masses. The corrections we find are on a percent level.

If thermal equilibrium has never been established, the relic abundance will be set by freeze-in production which peaks around $T \sim m_{Z'}/3$. The qualitative behavior of the updated

limits (black curve in figure 1) is analogous to previous studies [7, 10] for larger values of ΔN_{eff} . For mediator masses $m_{Z'} \gg 1$ TeV, the decoupling temperature is below the resonant regime $T \lesssim m_{Z'}$, and so for the abundance to saturate the current limits on ΔN_{eff} , the ν_R inevitably thermalize at high temperatures and the relic abundance is subsequently determined by the decoupling process near $T \sim 40$ GeV. As we transition to smaller masses, the peak of resonant production is similar to or falls below the estimated decoupling temperature. Consequently, the relic abundance cannot be set by decoupling alone anymore, but is given predominantly by the freeze-in from inverse decays at lower temperatures. Due to resonant enhancement, the limits become significantly stronger, as we fully transition to the freeze-in regime.

We also provide a refinement on projected experimental reaches. Here we focus on the expected sensitivity of the original proposal for CMB-S4 ($\Delta N_{\text{eff}} < 0.06$), and also compare to previously determined forecasts from [10], which approximates final state statistics and the backreaction term. In addition, we include thermal masses as a leading thermal correction. As discussed in section 4, the out of equilibrium production of mediators becomes indirectly sensitive to final state statistics. While decays into ν_R are not Pauli-blocked, decay to SM fermions are to some extent. Consequently, the branching ratio receives an in medium enhancement, production of ν_R is boosted. The effect is small and combined with thermal masses the corrections we find generally do not effect results beyond percent level, which is consistent with uncertainty estimates.

We note also that, as we approach smaller values of the mass and small gauge couplings, the total decay width becomes comparable to or smaller than the Hubble rate, $\Gamma_{Z'} \lesssim H$, and the Z' becomes cosmologically long-lived. After production, they redshift as matter and slower than the ambient radiation bath. Thus, the energy injection at decay is larger compared to prompt decays close to production [10, 13, 62]. We have also considered this explicit production of Z' by explicitly tracking their evolution as non-relativistic matter before decaying into SM final states and right-handed neutrinos. For the masses and couplings we consider, the effect is negligibly small and typically on a sub percent level, in accordance with the estimates from [10]. Nevertheless, we note a disagreement of up to $\sim 10\%$ above $m_{Z'} \sim 1$ GeV that also persists if we make the same approximations as [10] and consistently include the delayed decay. We therefore attribute this difference to numerical difficulties.

Forecasted limits we provide can be extrapolated to in principle arbitrarily high gauge boson masses. Indeed, for $m_{Z'} \sim \Lambda_{\text{GUT}} \sim 10^{16}$ GeV with $g' \sim \mathcal{O}(1)$ is at the edge of what seems to be detectable in the near future. In turn this would imply that any natural $g' \sim \mathcal{O}(1)$ realization below the GUT scale is accessible to us, and absence of detection would pose serious problems for large classes of models that preserves the Dirac nature of the neutrino by means of a gauge symmetry. We note here that such a conclusion makes strong assumptions on the thermal history. We revisit and relax these assumptions in the next section.

The updated limits and forecasts presented here will also hold as good approximations to other U(1) extensions of the SM. Prerequisite is the existence of light ν_R , and that ν_R and at least a subset of SM fermions are subject to the same new gauge interaction. Then, approximate limits on such models may be obtained by making an appropriate rescaling of gauge couplings.

5.2 Non-standard cosmologies

For now, let us assume that the Universe has reheated to the aforementioned $T \sim m_{Z'}$ or higher, and that the abundance of right-handed neutrinos has been set. If couplings are large, and ν_R remains coupled to the SM, then whatever change we make to the SM plasma will propagate also to the ν_R fluid. Only changes to the SM plasma after decoupling of the right-handed neutrinos will affect the ratio $\rho_{\nu_R}/\rho_{\text{SM}}$, and thus ΔN_{eff} . Of course it is feasible that ν_R -genesis falls into a non-standard episode of the cosmological history. While we do not study this case in detail, we argue that for many scenarios such a modified history during production is typically accompanied by a period of reheating of the SM. Thus, the relic abundance from this period is diluted, and the dominant part is given by subsequent production after reheating.

As we argue, the most significant changes to the naive predictions of ΔN_{eff} come from entropy injection after the initial ν_R abundance is set. Then, after the end of entropy injection the relations $\Delta N_{\text{eff}} \propto \rho_{\nu_R}/\rho_{\text{SM}}$ and $\rho \propto s^{4/3}$ hold.⁵ We define a dilution factor $D = S_{\text{after}}/S_{\text{before}}$, and this definition implies

$$\Delta N_{\text{eff}} \rightarrow \Delta N_{\text{eff}} \times \left(\frac{D_{s,\nu_R}}{D_{s,\text{SM}}} \right)^{4/3}. \quad (5.1)$$

In most cases, there is negligible entropy injection into the ν_R sector, and so $D_{s,\nu_R} = 1$. Equation 5.1 provides a convenient way to rescale limits we find when there is additional entropy being produced. We provide approximate expressions for the dilution factors in various modified thermal histories, which enables the recasting of our limits.

We can summarize a large class of modified cosmologies by introducing an additional fiducial energy density ρ_ϕ . To this end we augment Equation 4.1

$$\frac{d\rho_\phi}{dt} + 3H(1 + w_\phi)\rho_\phi = -\Gamma_\phi\rho_\phi \quad (5.2)$$

$$\frac{d\rho_{\text{SM}}}{dt} + 3H(\rho_{\text{SM}} + P_{\text{SM}}) = (1 - \kappa)\Gamma_\phi\rho_\phi - \mathcal{C}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} \quad (5.3)$$

$$\frac{d\rho_{\nu_R}}{dt} + 4H\rho_{\nu_R} = \mathcal{C}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} + \kappa\Gamma_\phi\rho_\phi. \quad (5.4)$$

To fully describe the cosmology, it is not enough to only specify the new parameters, namely the equation of state w , the decay rate Γ_ϕ , the relative fractions injected into the respective species κ . We also need to the normalization of ρ_ϕ relative to the SM, i.e. by giving the explicit ratio at some initial or reference scale factor/time. In particular, it contains variations of dark energies ($w_\phi = -1$), non-relativistic species or coherent oscillations with a potential phase of early matter domination ($w_\phi = 0$), or additional dark radiation ($w_\phi = 1/3$). Starting from this parametrization, we provide dilution factors for selected benchmark scenarios.

5.2.1 Extra degrees of freedom

The first situation we consider is the presence of additional thermalized degrees of freedom, collectively labeled as ϕ here. In the language of our augmented cosmological fluid equations,

⁵If we are for the moment agnostic about the subtle differences in the relativistic degrees of freedom.

we consider any number of extra degrees of freedom captured in ρ_ϕ as radiation-like and tightly coupled to the SM plasma. Then, we need to study ν_R production relative to the SM + ϕ plasma. Every degree of freedom freezing out from the plasma increases the SM temperature courtesy of entropy conservation. The dilution factor amounts to

$$D_s = \left(\frac{g_{\text{SM}}^{(s)} + g_\phi^{(s)}}{g_{\text{SM}}^{(s)}} \right). \quad (5.5)$$

Even if we double the degrees of freedom, which is roughly what we would expect e.g. from Supersymmetry, we only reduce ΔN_{eff} by a little more than half. We conclude that additional degrees of freedom between the electroweak scale and end of ν_R production, unless very numerous, do not significantly reduce the prospects of detection.

5.2.2 Phase transitions and other vacuum energies

We consider a vacuum like component of energy that eventually decays via an intermediate stage that redshifts as matter. This parametrization captures many models of early dark energies, but also phase transitions in the early Universe. We largely follow the analytic estimates from [63], and only make the substitutions necessary for our discussion. We define $T_i(a_i)$ as the temperature (scale factor) at which ϕ starts to dominate, i.e. $\rho_\phi = \rho_{\text{SM}}$. The entropy injection phase is characterized by the non-adiabatic evolution of the SM plasma and starts at T_e . Eventually, we once more reach an adiabatic radiation-like SM at T_r which we define approximately from the time at which $\Gamma_\phi = H$ holds. Instead, if the decay is triggered at some T_e and $\Gamma_\phi \gg H(T_e)$ holds, the vacuum energy decays very efficiently. Energy is promptly transferred to the respective sectors. Taking results from [63], we omit a possible entropy transfer to the ν_R sector ($\kappa = 0$) and we find for the instantaneous entropy injection

$$D_s = \frac{T_r^3}{T_e^3} = \left(1 + \frac{30\Delta V}{\pi^2 g_{(\rho)} T_e^4} \right)^{3/4} \simeq (1 + \alpha)^{3/4}, \quad (5.6)$$

where $\alpha = \Delta V / \rho_{\text{SM}}(T_e)$. Thus, only significant supercooling in a phase transition, or a period of significant vacuum energy domination with subsequent reheating could cause a meaningful dilution of ν_R .

The case of an extended decay phase is slightly more involved. Again, adapting the solution from [63], the dilution factor if the vacuum energy dominates prior to its decay, in this case reads

$$D_s = \left(\frac{90}{8\pi^3 g_{(\rho)}(T_r)} \right)^{3/4} \left(\frac{2}{5} \Gamma_\phi M_P \right)^{1/2} \left(\frac{8\pi\Delta V}{3} \right) \frac{1}{T_e^3}, \quad (5.7)$$

which e.g. applies to a phase transition, where the energy density transfers into oscillations of the scalar field around the new minimum and the subsequent decay of the oscillations is slow. In most cases, the prompt injection estimate gives a reasonable indication if reheating of the SM is significant.

Thus, in order for a phase transition or other forms of early dark energy to have a meaningful impact on primordial ν_R -genesis, they need to be the dominant energy density

prior to reheating. Only then is a dilution of ν_R relative to the SM that exceeds $\mathcal{O}(1)$ feasible from sufficient reheating. Should this occur, we can just consider the production of ν_R after reheating instead and the limits that cast on ν_R are consequently weakened.

Notably, a BSM induced first order electroweak phase transition or modified QCD transitions are often expected to fall within the $\alpha < 1$ regime, so modifications to these transitions will not challenge the robustness of our limits (see e.g. the discussion of models in [64]).

5.2.3 Early matter dominated era

We sketch the changes introduced to cosmology by an era of early matter domination (EMD). Here, we remain agnostic to what caused the deviation from an early radiation-like Universe, and instead adopt an appropriate phenomenological description. Non-standard expansion histories, and in particular EMD, are widely studied in the literature, and we refer to refs. [65, 66] for reviews on the subject.

As we consider a matter component that dominates the total energy budget of the Universe, we set $w = 0$ in our phenomenological model. We define T_{eq} as the temperature, at which $\rho_\phi = \rho_{\text{rad}} \simeq \rho_{\text{SM}}$ holds, and it defines the onset of the matter era. During the matter dominated era and when energy transfer from the matter sector to the radiation sector is inefficient, the SM continues to evolve with negligible energy injection. Since both, a possible ν_R population and the SM plasma continue to evolve adiabatically when energy injection is negligible, they scale as $\rho \propto a^{-4}$. It follows that the Hubble rate evolves as $H \propto a^{-3/2} \propto T_{\text{SM}}^{3/2}$ in this case [67].

Consequently, the ratio of ρ_{ν_R} to ρ_{SM} remains unaffected if ν_R -genesis completed before the onset of the matter dominated era. Production during such an era will be different. For example, for a freeze-in like production, the ratio \mathcal{C}/H is important, and the modified scaling of the Hubble rate compared to a standard radiation era affects the yield. We do not study this case in more detail here. As we will see below, any meaningful period of early matter domination with subsequent reheating will usually dilute such a primordial ν_R abundance. We note that this type of production has been studied extensively for dark matter relic abundances (e.g. [68]), and at least for UV-sensitive freeze-in, production from this regime is found to be negligible [63].

When entropy injection is efficient and the energy injected into the SM plasma is no longer negligible, its evolution deviates from adiabatic expansion. Assuming that ϕ continues to scale like matter and dominates the total energy density, it is possible to show that $\rho_{\text{SM}} \propto a^{-3/2}$ and $T \propto g_{(\rho)}^{-1/4} a^{-3/8}$. Analogous relations will hold for ν_R if decays of ϕ inject energy. In the following, we assume negligible changes for ν_R . Moreover, we define the end of EMD by the relation $H(T_r) \simeq \Gamma_\phi$, as decays become efficient when the decay rate exceeds the Hubble rate and the entire energy density stored in the matter component is then quickly transferred to radiation. Then, a dilution factor depending on T_{eq} and T_r can be inferred [67]

$$D_s = \frac{T_{\text{eq}} g_{(\rho)}(T_{\text{eq}}) g_{(s)}(T_r)}{T_r g_{(\rho)}(T_r) g_{(s)}(T_{\text{eq}})} \simeq \frac{T_{\text{eq}}}{T_r}. \quad (5.8)$$

Thus, dilution is approximately given by the ratio of temperatures between the onset of the matter dominated era, and its end. Unless the matter dominated era was very brief,⁶ dilution generically yields $D_s \gg 1$. Therefore, we become practically insensitive to primordial abundances for typical periods of extended matter domination. Only production towards the end of entropy injection may leave an imprint, similar to our discussion for the vacuum dominated Universe. As a conservative estimate, production starting from the second reheating temperature T_r can be used.

5.2.4 Low reheating temperature

The last case we consider is that of a significantly lowered reheating temperature. This is relevant for two reasons: first, an outright low reheating temperature coming from inflation is possible. Afterall, the most stringent bounds on the reheating temperature come from not spoiling the success of BBN at $T_{\text{BBN}} \sim \text{MeV}$. Moreover, from the previous considerations we infer that many modified cosmologies lead to a (partial) reheating at lower temperatures. As far as primordial ν_R abundances are concerned, even a rather mild reheating can significantly suppress ΔN_{eff} detection prospects, so studying production from $T_{\text{reh}} < m_{Z'}$ is important.

For masses relevant in collider searches, this has also been considered in [14], but we extend this to the entire mass range up to the GUT scale and we utilize our improved calculation scheme for ν_R -genesis. To mimic the effect of low reheating, we start integrating the relevant equations not from $T \gg m_{Z'}$ as in the previous section, but rather from a chosen value $T_{\text{reh}} < m_{Z'}$. Doing this simplifies the problem considerably, as we omit the challenging and chaotic phase of reheating and possible ν_R produced in the reheating process itself. In that sense, the results we provide are conservative, as we capture only production after reheating completed. In particular, direct reheating contributions, and possible primordial relics that have not fully diluted away in the reheating process are neglected.

We show the results from reduced reheating temperatures in figure 2 for selected values of $m_{Z'}$. We focus on near- and far-future collider reaches, and selected values between the anticipated futuristic collider reaches and the GUT scale. Limits remain stringent for reheating close to the mass of the gauge boson, as this regime is dominated by the resonantly enhanced inverse decay and subsequent decays of Z' . We then observe a weakening in limits, as the on-shell production becomes kinematically impossible after reheating. For even lower reheating temperatures of $T_{\text{reh}} \lesssim 10^{-2} m_{Z'}$, the relation between gauge coupling and reheating temperature becomes universal. This is the UV-sensitive freeze-in of $f\bar{f} \rightarrow \nu_R \bar{\nu}_R$ via an effective Four-Fermi interaction. Production is determined by the highest temperature reached, and relic energy densities in this regime are controlled approximately by the combination $g^4 T_{\text{reh}}^3 M_P / m_{Z'}^4$. We find for a benchmark value $\Delta N_{\text{eff}} \sim 0.1$ the approximate empirical relations

$$g \sim 1.8 \left(\frac{T_{\text{reh}}/m_{Z'}}{10^{-3}} \right)^{-3/4} \left(\frac{m_{Z'}}{10^9 \text{ GeV}} \right), \quad (5.9)$$

$$\frac{T_{\text{reh}}}{m_{Z'}} \sim 0.05 \left(\frac{m_{Z'}}{10^9 \text{ GeV}} \right)^{1/3} \left(\frac{g'}{0.1} \right)^{-4/3}. \quad (5.10)$$

⁶Here we gloss over the fact that for a very brief matter era, the approximations needed for this result will break down. Our conclusions are not affected by this.

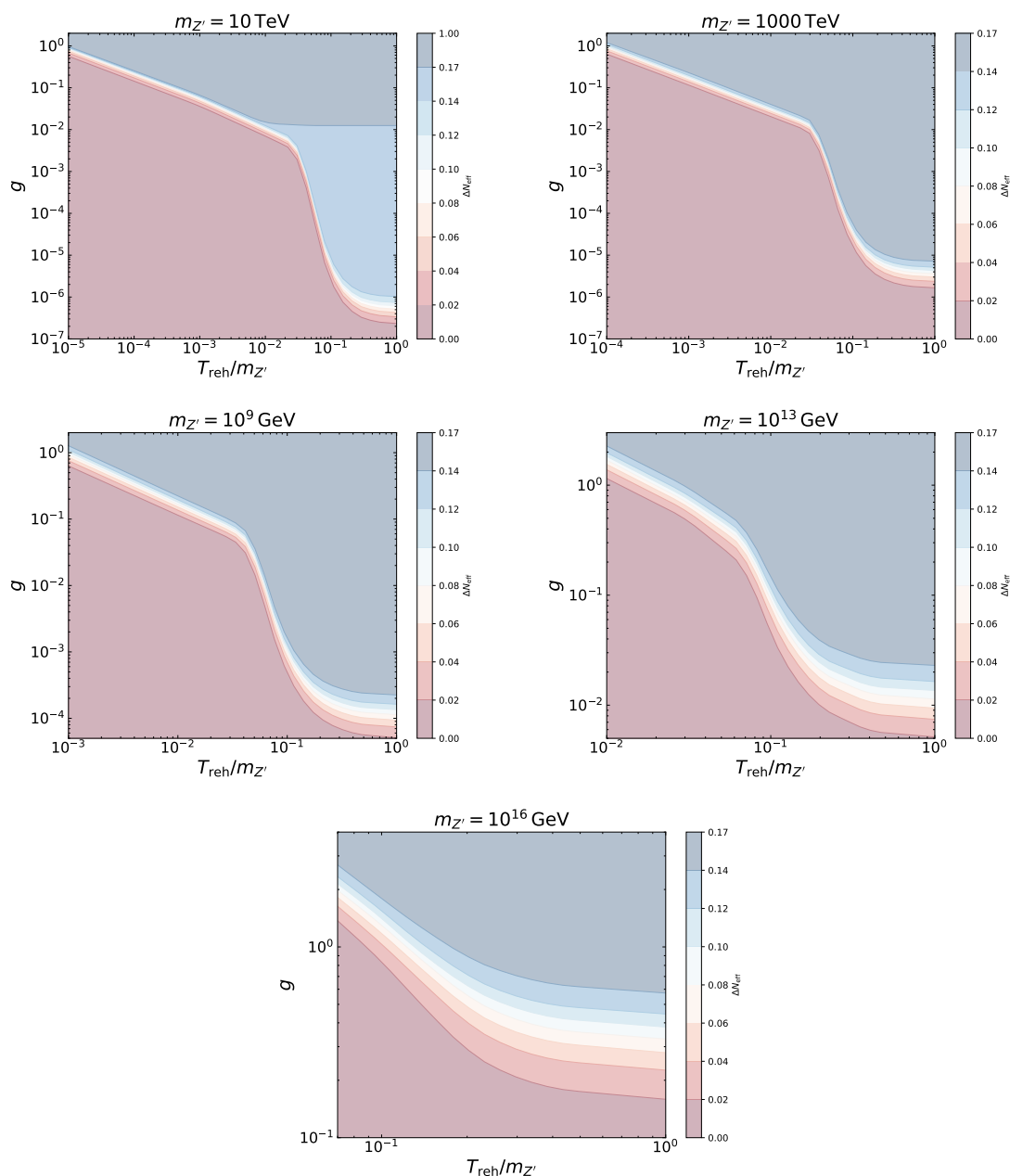


Figure 2. The effect of a lowered reheating temperature for selected values of the Z' mass. For masses up to thousands of TeV, limits can be avoided by having a significantly lowered reheating temperature only. Especially for $m_{Z'} \sim \text{TeV}$ and $g \sim \mathcal{O}(1)$ we can still have a Z' within collider reach, if there was a significant reheating event as low as temperatures $T_{\text{reh}} \lesssim 100 \text{ MeV}$. For larger masses, strong limits on $\mathcal{O}(1)$ gauge coupling may be avoided by having (inflationary or secondary) reheating at temperatures a few orders of magnitude below the gauge boson mass. We note that for scales as high as the GUT scale, if the Universe actually reheated to such high temperatures, a seemingly natural realization with order 1 gauge coupling becomes testable. However, reheating to just an order of magnitude below the mass will suppress any signal from ΔN_{eff} , and GUT scale Z' with Dirac neutrinos may still avoid. We note that the necessary reheating temperature to produce detectable amounts of ν_R at the GUT scale is in tension with upper limits on the scale of inflation.

These convenience expressions give a good first estimate as to an extended model can evade bounds if the reheating temperature is lowered.

We observe that, in particular for extremely heavy Z' , the strong limits we would get from extrapolating the forecasts in figure 1 can be weakened substantially. For scales as high as the GUT scale, we observe that a testable ΔN_{eff} contribution requires reheating temperatures close to the upper bounds on the scale of inflation [69]. Below that, strong bounds on heavy Z' with $\mathcal{O}(1)$ gauge couplings can be relaxed if reheating is only two to three orders of magnitude below the gauge boson mass. As we progress to smaller masses in the 10 – 1000 TeV range, evading current and projected limits for natural choices of gauge coupling requires going to very small reheating temperatures significantly below the temperature of EWSB. These ranges could in principle be considered testable in the laboratory, and so the possible implications on the reheating temperature are particularly interesting to consider here.

This is especially interesting if we consider our minimal model as a building block of more complex models that aim to incorporate other phenomena, such as including suitable dark matter candidates. The irreducible ν_R contribution force such models to potentially low reheating temperatures, with implications for dark matter production, phase transitions, or other cosmological observables. We note here in particular possible consequences for Dirac leptogenesis [70]. The requirement that ν_R and the SM are decoupled during the era in which sphalerons are active implies that future experiments will strongly constrain the feasibility of Dirac leptogenesis in models that include gauged $B - L$ or similar Z' extensions. Unless it seems desirable to go to sufficiently small gauge coupling, such bounds on leptogenesis can again only be avoided by lowering the reheating temperature — which in return narrows the window on creating the initial lepton asymmetry and the efficiency of the sphaleron process in models with Dirac leptogenesis considerably.

6 Conclusion

Models that predict neutrinos to be Dirac particles remain well motivated and widely studied. Such models often include new gauged $U(1)$ symmetries, and the combination of a Z' with Dirac neutrinos is strongly constrained by cosmology. In particular, gauged extension provide an explanation of the Dirac nature of the neutrino, and among possible extensions, a gauged $U(1)_{B-L}$ provides a useful benchmark case. The gauge symmetry protects Dirac nature of the light neutrinos even if broken, as long as $\Delta(B - L) \neq 2$.

The model we adopt in this study is minimal, but arises as a building block in many non-minimal realizations of gauged $U(1)_{B-L}$ with Dirac neutrinos, or similar structures for more general Z' extensions. We then compute contributions of the light-right handed neutrino to the total radiation budget of the Universe. We sketch and employ an integration scheme for evaluating collision operators, as they appear in integrated Boltzmann equations without any additional assumptions. The scheme proves to be computational reliable and efficient, and in principle enables convenient implementations of a variety of processes. Notably, we can compute results without making simplifying assumptions on final state statistics or the backreaction, but only have to adopt a prescription for the underlying phase space density, which in many situations can be physically motivated. While for this study these effects

do not give rise to significant corrections, when applied to different problems our scheme will make the inclusion of such effects relatively easy.

Using the new constraint on the extra radiation $\Delta N_{\text{eff}} < 0.17$ (95% C.L.) from the latest ACT DR6 data release, we update the respective limits in the $m_{Z'} - g'$ plane. The new limits are orders of magnitude stronger than previous CMB limits from Planck, and surpass collider limits by a similar margin and on all scales we can test for now. Even the forecasted reach of the FCC is barely competitive compared to the updated cosmological limit, and also only for a narrow range of gauge boson masses. Notably, we provide for the first time CMB-only limits in the otherwise inaccessible range of $m_{Z'} > 4 \text{ TeV}$, and excluding gauge boson masses of $m_{Z'} \sim 100 \text{ TeV}$ for $g' = \mathcal{O}(1)$.

We show the potential of future CMB experiments by quoting results for a benchmark of $\Delta N_{\text{eff}} = 0.06$ and find our forecasts to be consistent with previous results. We provide exact solutions to the integrated Boltzmann equations and find that corrections to simplified approaches are generically of $\mathcal{O}(1\%)$, which is consistent with previous uncertainty estimates. When extrapolated to higher masses, the forecasts suggest that even GUT scale realizations of $B - L$ with $g' \sim \mathcal{O}(1)$ are testable, although careful examination of the underlying assumptions makes this unlikely. However, natural realizations of gauged $U(1)_{B-L}$ with Dirac neutrinos will be severely constrained by future observations on all scales up to Λ_{GUT} within a standard thermal history.

The limits we find depend on assumptions on the underlying cosmology, namely that the Universe reached temperatures $T \sim m_{Z'}$ and higher, and that the thermal history proceeded to be radiation-like and without major entropy generating events between creation of ν_R and BBN. We discuss at length the effect of modifications to this assumptions. Notably, we consider the presence of extra degrees of freedom in the plasma that eventually freeze-out, as well as entropy injection coming from phase transitions, early dark energies, and a period of early matter domination. We provide recipes to translate our bounds to scenarios of a modified thermal history. Constraints are robust when facing typical non-standard electroweak or QCD phase transitions, but note that a period of early matter domination generically leads to significant reheating that will dilute a relic ν_R beyond detection. The dominant contribution is expected to come from the UV-sensitive, post-reheating production of ν_R in those scenarios.

We consider the effect of lowered reheating temperatures in detail, providing modified limits for selected benchmark points and show that constraints can be partially evaded if the Universe initially reheats to $T_{\text{reh}} \ll m_{Z'}$ or faces a second period of reheating, e.g. from an early matter era. In particular, scenarios close to the GUT scale that are seemingly testable in the near future when extrapolating our limits are shown to evade constraints, mostly because the reheating temperatures required are in tension with upper bounds on the reheating temperature from inflation. Taking reheating into account, natural realizations $g \sim \mathcal{O}(1)$ across all mass scales will evade cosmological limits only when the reheating temperature is lowered substantially. This provides additional constraints on extended models that require reaching certain temperatures to achieve dark matter production or Dirac leptogenesis. In particular, finding a Z' coupled to Dirac neutrinos at current or future colliders would imply a profound deviation from the early radiation dominated Universe.

An important result is that, cosmological bounds exceed laboratory and collider bounds by a significant margin on all mass scales. Mass scales unconstrained by collider and laboratory in the hundreds of TeV regime are being tested by cosmology, and even higher scales seem within reach. Future experiments have the ability to provide tremendous improvements on limits on Dirac neutrinos subject to a shared gauge interaction with SM particles. The limits are generic and apply to a wide class of models that involve a gauged $B - L$ (and with small modifications other generic gauged $U(1)$ extensions) as a building block, assuming neutrinos are Dirac particles. Should future experiments prevail and provide their forecasted limits without any clear sign of excess radiation, it would provide a strong indication towards the viability of Dirac neutrino models. Likewise, if the Dirac nature of neutrinos were be established by other means, we face a strong hint towards a non-standard thermal history of the early Universe.

A Integrating collision operators

We compute the integrals of the collision operator using a Monte Carlo approach. To this end, we utilize the VEGAS framework [74, 75], which through adaptive sampling achieves fast convergence of integrals and enables us to integrate collision operators efficiently. In the following, we show the implementation steps of $2 \rightarrow 2$ and the $1 \rightarrow 2$ collision operators. We comment briefly on subsequent implementation steps, and hope that the steps and comments presented here prove useful to future implementations of similar strategies.

A.1 $2 \rightarrow 2$ processes

We reduce this collision operator from 12 to 5 dimensions, similar to [76] although we deviate for some steps. We start with an integral of an arbitrary function of the momenta over the 4-particle phase space

$$C_{(F)} = \int \frac{d^3 p_1}{(2\pi)^3 2E_1} \frac{d^3 p_2}{(2\pi)^3 2E_2} \frac{d^3 p_3}{(2\pi)^3 2E_3} \frac{d^3 p_4}{(2\pi)^3 2E_4} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4) F(p_i). \quad (\text{A.1})$$

It is always possible go to a coordinate system in which $\vec{p}_1 = (0, 0, p_1)^T$. Angular integrations of the first momentum are then trivial and we replace $d^3 p_1 = 4\pi p_1^2 dp_1$. Then, $\vec{p}_2 = p_2(\sin \beta, 0, \cos \beta)^T$, and $\vec{p}_3 = p_3(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^T$, with respect to the fixed direction of \vec{p}_1 . Performing the trivial azimuthal integration of p_2 yields

$$C_{(F)} = 2(2\pi)^2 \int_0^\infty \frac{dp_1 p_1^2}{(2\pi)^3 2E_1} \int_{-1}^1 d \cos \beta \int_0^\infty \frac{dp_2 p_2^2}{(2\pi)^3 2E_2} \int_0^{2\pi} d\phi \int_{-1}^1 d \cos \theta \quad (\text{A.2})$$

$$\int_0^\infty \frac{dp_3 p_3^2}{(2\pi)^3 2E_3} \int \frac{d^3 p_4}{(2\pi)^3 2E_4} (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4) \times F(p_i). \quad (\text{A.3})$$

We continue by eliminating the spatial part of the δ -distribution by integrating p_4 , leaving $\delta(E_1 + E_2 - E_3 - E_4)$. Since spatial momentum conservation has been enforced, we understand E_4 as a shorthand for

$$E_4 = \sqrt{m_4^2 + (\vec{p}_1 + \vec{p}_2 - \vec{p}_3)^2} \quad (\text{A.4})$$

$$= (m_4^2 + p_1^2 + p_2^2 + p_3^2 + 2p_1 p_2 \cos \beta - 2p_1 p_3 \cos \theta - 2p_2 p_3 \cos \phi) \quad (\text{A.5})$$

$$- 2p_2 p_3 (\sin \beta \sin \theta \cos \phi + \cos \beta \cos \theta)^{1/2}. \quad (\text{A.6})$$

Moreover we write the δ -distribution as

$$\delta(E_1 + E_2 - E_3 - E_4) = \frac{E_1 + E_2 - E_3}{2p_2 p_3 \sin \beta \sin \theta} \delta(\cos \phi - \cos \phi_c), \quad (\text{A.7})$$

where

$$\cos \phi_c = \frac{m_1^2 + m_2^2 + m_3^2 - m_4^2 + 2(E_1 E_2 - E_1 E_3 - E_2 E_3) - 2p_1 p_2 \cos \beta}{2p_2 p_3 \sin \beta \sin \theta} \quad (\text{A.8})$$

The integral over ϕ is symmetric around $\phi = \pi$, so we replace $\int_0^{2\pi} d\phi = 2 \int_0^\pi d\phi$. When the dust settles, we arrive at our reduced collision integral

$$\mathcal{C}_{(F)} = 4(2\pi)^4 \int_0^\infty \frac{dp_1 p_1^2}{(2\pi)^3 2E_1} \int_0^\infty \frac{dp_2 p_2^2}{(2\pi)^3 2E_2} \int_0^\infty \frac{dp_3 p_3^2}{(2\pi)^3 2E_3} \int_{-1}^1 d \cos \beta \int_{-1}^1 d \cos \theta \quad (\text{A.9})$$

$$\times \frac{1}{2E_4} \frac{E_1 + E_2 - E_3}{2p_2 p_3 \sin \beta \sin \theta} \times \Theta(0 \leq \cos^2 \phi_c \leq 1) \times F(\vec{p}_1, \vec{p}_2, \vec{p}_3, \vec{p}_4), \quad (\text{A.10})$$

where we completed all the previously discussed steps and also evaluated the ϕ integration, i.e. E_4 and \vec{p}_4 are to be understood as functions of the remaining integration variables, and in particular $\cos \phi = \cos \phi_c$ from the transformed δ -distribution is enforced. We restrict the integration domain to the physically allowed region by means of an indicator function Θ . In principle, we could explicitly find the boundaries of the physically allowed integration domain (see e.g. [76]), but for our purposes this is not necessary.

The integrations $d \cos \beta$ and $d \cos \theta$ are already suitable for a Monte Carlo integration. For the spatial momentum, we make variable transformations $p_i = -\Lambda_i \log x_i$, where x_i becomes our new integration variable and Λ_i is a dimensional constant. We find good convergence behavior for a variety of choices, but typically use $\Lambda_i = \mathcal{O}(1 - 10) T_{\text{SM}}$ for fast convergence (see also Reference [9]). With these variable substitutions, Equation A.10 is the collision integral as we have implemented. We explicitly verified the correctness of the implementation against the well-known semi-analytic result from Gelmini and Gondolo [77].

A.2 1 \rightarrow 2 processes

The integral is 9 dimensional, and can be reduced to 2, which would be perfectly suitable for traditional numerical integration. To illustrate the convenience of the Monte Carlo approach, we also perform the integration of (inverse) decays with Monte Carlo sampling. The steps are essentially the same as before, just with the added convenience of fewer integrals to handle.

We are once more free to choose the coordinate system. We make the same choice as before and use $\vec{p}_1 = p_1(0, 0, 1)^T$. As before, \vec{p}_1 and \vec{p}_2 form a plane, so it is possible to write $\vec{p}_2 = p_2(\sin \theta, 0, \cos \theta)^T$ with trivial azimuthal dependence, i.e. we can pick the alignment of the plane to our convenience. The spatial part of the δ -distribution eliminates $d^3 p_3$, and now $E_3 = \sqrt{m_3^2 + (\vec{p}_1 - \vec{p}_2)^2}$ holds. For the remaining δ -distribution, we write it in terms of the angle $\delta(\cos \theta - \cos \theta_c)$, where now

$$\cos \theta_c = \frac{m_3^2 + p_1^2 + p_2^2 - (E_1 - E_2)^2}{2p_1 p_2}, \quad (\text{A.11})$$

holds. We restrict the integration domain to the physically allowed space by using an indicator function $\Theta(0 \leq \cos^2 \phi \leq 1)$, but could equally well just derive an analytic expression for the integration domain. Putting everything together, we arrive at

$$\mathcal{C}_{(F)} = 8\pi^2 \int_0^\infty \frac{dp_1 p_1^2}{(2\pi)^3 2E_1} \int_0^\infty \frac{dp_2 p_2^2}{(2\pi)^3 2E_2} \times \frac{1}{2E_3} \frac{E_3}{p_1 p_2} \times F(p_i). \quad (\text{A.12})$$

The remaining two integrations are again performed numerically after making the substitution $p_i = -\Lambda_i \log x_i$. We have also verified this collision integral against analytic decay rates when distributions are Maxwell-Boltzmann (see also Reference [9]).

A.3 Performance and other implementation details

By first performing few adaptive integration steps of just below 1000 evaluations, we can adapt the **VEGAS** integration routine to the integrand. The actual evaluation uses few integration steps but a larger number of evaluations. Using the internal uncertainty estimate of **VEGAS** and after optimizing some **VEGAS** configurations, we can achieve a relative error of 0.1% for evaluation times on a single core in the 10 – 100 ms range (Precise values depend also on the underlying particle models). With suitable parallelization, an effective evaluation of many integrals, e.g. to provide an interpolation grid, remains feasible at manageable computational load.

Some processes can usually be added and jointly integrated, thus drastically reducing the number of integral evaluations even in the presence of many processes that contribute. In the case study here, we may define for the process $f\bar{f} \rightarrow \nu_R \bar{\nu}_R$ as

$$\text{Process}(f) \propto \left| \mathcal{M}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R} \right|^2 \left(f_1 f_2 \bar{f}_3 \bar{f}_4 - f_3 f_4 \bar{f}_1 \bar{f}_2 \right) \quad (\text{A.13})$$

where we omitted all the factors coming from variable transformations in Equation A.10. The integration is just over angles and dimensionless 3-momenta, meaning $\text{Process}(f)$ with all its omitted prefactors contains the full information about a given channel. Then, it proves convenient to use Equation A.10 not for a single $\text{Process}(f)$, but rather $\sum_f \text{Process}(f)$, leading to a single evaluation of Equation A.10 at the cost of a moderately more expensive integrand.

We would also like to note that in many situations parameters can be factored out, e.g. powers of couplings for many processes. Then, the collision operator only needs to be precomputed once, which again reduces the number of evaluations needed.

B Matrix elements, cross sections and decay rates

Here we list all relevant matrix elements. Expressions involving γ -matrices have been evaluated using **FeynCalc** [78, 79].

For the (inverse) decays we have

$$\left| \mathcal{M}_{Z' \rightarrow f\bar{f}} \right|^2 = 4 \left(2m_f^2 + m_{Z'}^2 \right). \quad (\text{B.1})$$

$$\left| \mathcal{M}_{Z' \rightarrow \nu_R \bar{\nu}_R} \right|^2 = 2m_{Z'}^2. \quad (\text{B.2})$$

The total decay width in vacuum is given by

$$\Gamma_{Z'} = \frac{g^2}{12\pi} m_{Z'} \left(3 + \sum_{f \neq \nu}^{m_{Z'} > 2m_f} \frac{Q_{B-L}^2}{N_C(f)} \left(1 + \frac{2m_f}{m_{Z'}} \right) \sqrt{1 - \frac{4m_f^2}{m_{Z'}^2}} \right), \quad (\text{B.3})$$

where the factor 3 comprises the contribution of left- and right-handed neutrinos. We neglect the complications of hadronization and assume decay into free quarks only.

For the process $f\bar{f} \leftrightarrow \nu_R \bar{\nu}_R$, we find

$$|\mathcal{M}_{f\bar{f} \rightarrow \nu_R \bar{\nu}_R}|^2 = g_{Z'}^4 N_C(f) Q_f^2 Q_{\nu_R}^2 \frac{4(2m_f^4 + 2m_f^2(s-t-u) + t^2 + u^2)}{(m_{Z'}^2 - s)^2}. \quad (\text{B.4})$$

The prescription for the resonant case is discussed in the main text. In the following, we show explicitly that in vacuum, the NWA and the explicit matrix element give the same cross section for resonant production.

We write the differential cross section as

$$\frac{d\sigma}{dt} = \frac{1}{64\pi s} \frac{1}{|\vec{p}_{1,\text{cm}}|^2} |\mathcal{M}|^2, \quad (\text{B.5})$$

replace $u = 2m_f^2 - t - s$, and integrate the Mandelstam t with boundaries $-(p_{1,\text{cm}} \mp p_{3,\text{cm}})$, where $p_{1,\text{cm}} = \sqrt{s/4 - m_f^2}$ and $p_{3,\text{cm}} = \sqrt{s}/2$. Then, after also multiplying by 3 for the generations of ν_R in the final state, we find

$$\sigma = \frac{g^4}{2\pi} \frac{Q_f^2 Q_{\nu_R}^2}{N_C(f)} \sqrt{\frac{s}{s - 4m_f^2}} \left(1 + \frac{2m_f^2}{s} \right) \frac{s}{(s - m_{Z'}^2)^2 + \Gamma_{Z'}^2 m_{Z'}^2}. \quad (\text{B.6})$$

Now if we approximate the resonance as a delta distribution, we readily arrive at

$$\sigma_{\text{res}} = \frac{g^4}{2\pi} \frac{Q_f^2 Q_{\nu_R}^2}{N_C(f)} \sqrt{\frac{m_{Z'}^2}{m_{Z'}^2 - 4m_f^2}} \left(1 + \frac{2m_f^2}{m_{Z'}^2} \right) \frac{\pi m_{Z'}}{\Gamma_{Z'}} \delta(s - m_{Z'}^2). \quad (\text{B.7})$$

This result is to be compared with the cross section we find from the NWA prescription to judge whether spin correlations are relevant for the total cross section. Using Equation 4.17 (The factor 1/3 here is canceled by the three generations of ν_R), the t integration is now trivial, and we find

$$\sigma_{\text{NWA}} = \frac{1}{64\pi s} \frac{1}{|\vec{p}_{1,\text{cm}}|^2} g_{Z'}^2 \frac{Q_f^2}{N_C(f)} g_{Z'}^2 Q_{\nu_R}^2 8m_{Z'}^2 (2m_f^2 + m_{Z'}^2) \frac{\pi}{m_{Z'} \Gamma_{Z'}} \sqrt{s} \sqrt{s - 4m_f^2} \delta(s - m_{Z'}^2), \quad (\text{B.8})$$

where the last two square root factors come from the integration. After replacing $s = m_{Z'}^2$ and simplifying, it is evident that $\sigma_{\text{NWA}} = \sigma_{\text{res}}$ holds, i.e. the total cross sections of the approaches are the same. Therefore, spin correlations do not affect the total cross section in vacuum, and since the distributions we assume throughout this study do not explicitly depend on spin, we expect this to hold also for interactions in medium.

Acknowledgments

The authors thank Salvador Centelles Chuliá, Thede de Boer, and Yi Chung for helpful discussions and Paul Frederik Depta for valuable comments on the numerical implementation. TH acknowledges support by the IMPRS-PTFS. This research work made extensive use of NumPy [71], SciPy [72] and Matplotlib [73].

References

- [1] C. Giganti, S. Lavignac and M. Zito, *Neutrino oscillations: The rise of the PMNS paradigm*, *Prog. Part. Nucl. Phys.* **98** (2018) 1 [[arXiv:1710.00715](#)] [[INSPIRE](#)].
- [2] M.J. Dolinski, A.W.P. Poon and W. Rodejohann, *Neutrinoless Double-Beta Decay: Status and Prospects*, *Ann. Rev. Nucl. Part. Sci.* **69** (2019) 219 [[arXiv:1902.04097](#)] [[INSPIRE](#)].
- [3] J. Froustey, C. Pitrou and M.C. Volpe, *Neutrino decoupling including flavour oscillations and primordial nucleosynthesis*, *JCAP* **12** (2020) 015 [[arXiv:2008.01074](#)] [[INSPIRE](#)].
- [4] M. Escudero Abenza, *Precision early universe thermodynamics made simple: N_{eff} and neutrino decoupling in the Standard Model and beyond*, *JCAP* **05** (2020) 048 [[arXiv:2001.04466](#)] [[INSPIRE](#)].
- [5] P.F. de Salas and S. Pastor, *Relic neutrino decoupling with flavour oscillations revisited*, *JCAP* **07** (2016) 051 [[arXiv:1606.06986](#)] [[INSPIRE](#)].
- [6] K. Akita and M. Yamaguchi, *A precision calculation of relic neutrino decoupling*, *JCAP* **08** (2020) 012 [[arXiv:2005.07047](#)] [[INSPIRE](#)].
- [7] K.N. Abazajian and J. Heeck, *Observing Dirac neutrinos in the cosmic microwave background*, *Phys. Rev. D* **100** (2019) 075027 [[arXiv:1908.03286](#)] [[INSPIRE](#)].
- [8] L.J. Hall, K. Jedamzik, J. March-Russell and S.M. West, *Freeze-In Production of FIMP Dark Matter*, *JHEP* **03** (2010) 080 [[arXiv:0911.1120](#)] [[INSPIRE](#)].
- [9] X. Luo, W. Rodejohann and X.-J. Xu, *Dirac neutrinos and N_{eff} . Part II. The freeze-in case*, *JCAP* **03** (2021) 082 [[arXiv:2011.13059](#)] [[INSPIRE](#)].
- [10] P. Adshead, P. Ralegankar and J. Shelton, *Dark radiation constraints on portal interactions with hidden sectors*, *JCAP* **09** (2022) 056 [[arXiv:2206.13530](#)] [[INSPIRE](#)].
- [11] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [[arXiv:1807.06209](#)] [[INSPIRE](#)].
- [12] J. Heeck, *Unbroken $B - L$ symmetry*, *Phys. Lett. B* **739** (2014) 256 [[arXiv:1408.6845](#)] [[INSPIRE](#)].
- [13] H. Esseili and G.D. Kribs, *Cosmological implications of gauged $U(1)_{B-L}$ on ΔN_{eff} in the CMB and BBN*, *JCAP* **05** (2024) 110 [[arXiv:2308.07955](#)] [[INSPIRE](#)].
- [14] L. Caloni, P. Stengel, M. Lattanzi and M. Gerbino, *Constraining UV freeze-in of light relics with current and next-generation CMB observations*, *JCAP* **10** (2024) 106 [[arXiv:2405.09449](#)] [[INSPIRE](#)].
- [15] A. Biswas, D.K. Ghosh and D. Nanda, *Concealing Dirac neutrinos from cosmic microwave background*, *JCAP* **10** (2022) 006 [[arXiv:2206.13710](#)] [[INSPIRE](#)].
- [16] K.S. Babu, X.-G. He, M. Su and A. Thapa, *Naturally light Dirac and pseudo-Dirac neutrinos from left-right symmetry*, *JHEP* **08** (2022) 140 [[arXiv:2205.09127](#)] [[INSPIRE](#)].

- [17] D. Borah, S. Mahapatra, D. Nanda and N. Sahu, *Type II Dirac seesaw with observable ΔN_{eff} in the light of W -mass anomaly*, *Phys. Lett. B* **833** (2022) 137297 [[arXiv:2204.08266](#)] [[INSPIRE](#)].
- [18] M. Berbig, *S.M.A.S.H.E.D.: Standard Model Axion Seesaw Higgs inflation Extended for Dirac neutrinos*, *JCAP* **11** (2022) 042 [[arXiv:2207.08142](#)] [[INSPIRE](#)].
- [19] S. Mishra, N. Narendra, P.K. Panda and N. Sahoo, *Scalar dark matter and radiative Dirac neutrino mass in an extended $U(1)_{B-L}$ model*, *Nucl. Phys. B* **981** (2022) 115855 [[arXiv:2112.12569](#)] [[INSPIRE](#)].
- [20] D. Borah et al., *Effective theory of light Dirac neutrino portal dark matter with observable ΔN_{eff}* , *Phys. Rev. D* **112** (2025) 055010 [[arXiv:2502.10318](#)] [[INSPIRE](#)].
- [21] M. Berbig, *Freeze-In of radiative keV-scale neutrino dark matter from a new $U(1)_{B-L}$* , *JHEP* **09** (2022) 101 [[arXiv:2203.04276](#)] [[INSPIRE](#)].
- [22] D. Mahanta and D. Borah, *Low scale Dirac leptogenesis and dark matter with observable ΔN_{eff}* , *Eur. Phys. J. C* **82** (2022) 495 [[arXiv:2101.02092](#)] [[INSPIRE](#)].
- [23] C. Han, M.L. López-Ibáñez, B. Peng and J.M. Yang, *Dirac dark matter in $U(1)_{B-L}$ with the Stueckelberg mechanism*, *Nucl. Phys. B* **959** (2020) 115154 [[arXiv:2001.04078](#)] [[INSPIRE](#)].
- [24] N. Das and D. Borah, *Light Dirac neutrino portal dark matter with gauged $U(1)_{B-L}$ symmetry*, *Phys. Rev. D* **109** (2024) 075045 [[arXiv:2312.06777](#)] [[INSPIRE](#)].
- [25] P. Adshead, Y. Cui, A.J. Long and M. Shamma, *Unraveling the Dirac neutrino with cosmological and terrestrial detectors*, *Phys. Lett. B* **823** (2021) 136736 [[arXiv:2009.07852](#)] [[INSPIRE](#)].
- [26] D.K. Ghosh, P. Ghosh, S. Jeusun and R. Srivastava, *N_{eff} at CMB challenges $U(1)_X$ light gauge boson scenarios*, *Phys. Rev. D* **110** (2024) 075032 [[arXiv:2404.10077](#)] [[INSPIRE](#)].
- [27] S. Safi, M. Farhang, O. Mena and E. Di Valentino, *Semiblind reconstruction of the history of effective number of neutrinos using CMB data*, *Phys. Rev. D* **110** (2024) 103513 [[arXiv:2404.01457](#)] [[INSPIRE](#)].
- [28] I.R. Wang and X.-J. Xu, *Imprints of light dark matter on the evolution of cosmic neutrinos*, *JCAP* **05** (2024) 050 [[arXiv:2312.17151](#)] [[INSPIRE](#)].
- [29] Z. Hou et al., *How Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail*, *Phys. Rev. D* **87** (2013) 083008 [[arXiv:1104.2333](#)] [[INSPIRE](#)].
- [30] G. Steigman, K.A. Olive and D.N. Schramm, *Cosmological Constraints on Superweak Particles*, *Phys. Rev. Lett.* **43** (1979) 239 [[INSPIRE](#)].
- [31] A.D. Dolgov, *Neutrinos in cosmology*, *Phys. Rept.* **370** (2002) 333 [[hep-ph/0202122](#)] [[INSPIRE](#)].
- [32] A.D. Dolgov and Y.B. Zeldovich, *Cosmology and Elementary Particles*, *Rev. Mod. Phys.* **53** (1981) 1 [[INSPIRE](#)].
- [33] ACT collaboration, *The Atacama Cosmology Telescope: DR6 Constraints on Extended Cosmological Models*, [arXiv:2503.14454](#) [[INSPIRE](#)].
- [34] ACT collaboration, *The Atacama Cosmology Telescope: DR6 Power Spectra, Likelihoods and Λ CDM Parameters*, [arXiv:2503.14452](#) [[INSPIRE](#)].
- [35] T.-H. Yeh, J. Shelton, K.A. Olive and B.D. Fields, *Probing physics beyond the standard model: limits from BBN and the CMB independently and combined*, *JCAP* **10** (2022) 046 [[arXiv:2207.13133](#)] [[INSPIRE](#)].
- [36] B.D. Fields, K.A. Olive, T.-H. Yeh and C. Young, *Big-Bang Nucleosynthesis after Planck*, *JCAP* **03** (2020) 010 [*Erratum ibid.* **11** (2020) E02] [[arXiv:1912.01132](#)] [[INSPIRE](#)].

- [37] SPT-3G collaboration, *Cosmology from CMB lensing and delensed EE power spectra using 2019–2020 SPT-3G polarization data*, *Phys. Rev. D* **111** (2025) 083534 [[arXiv:2411.06000](#)] [[INSPIRE](#)].
- [38] SIMONS OBSERVATORY collaboration, *The Simons Observatory: Astro2020 Decadal Project Whitepaper*, *Bull. Am. Astron. Soc.* **51** (2019) 147 [[arXiv:1907.08284](#)] [[INSPIRE](#)].
- [39] SPT-3G collaboration, *SPT-3G: A Next-Generation Cosmic Microwave Background Polarization Experiment on the South Pole Telescope*, *Proc. SPIE Int. Soc. Opt. Eng.* **9153** (2014) 91531P [[arXiv:1407.2973](#)] [[INSPIRE](#)].
- [40] K. Abazajian et al., *CMB-S4 Science Case, Reference Design, and Project Plan*, [arXiv:1907.04473](#) [[INSPIRE](#)].
- [41] N. Sehgal et al., *CMB-HD: An Ultra-Deep, High-Resolution Millimeter-Wave Survey Over Half the Sky*, *Bull. Am. Astron. Soc.* **51** (2019) 1 [[arXiv:1906.10134](#)] [[INSPIRE](#)].
- [42] M. Escudero, S.J. Witte and N. Rius, *The dispirited case of gauged $U(1)_{B-L}$ dark matter*, *JHEP* **08** (2018) 190 [[arXiv:1806.02823](#)] [[INSPIRE](#)].
- [43] ATLAS collaboration, *Search for new high-mass phenomena in the dilepton final state using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **10** (2017) 182 [[arXiv:1707.02424](#)] [[INSPIRE](#)].
- [44] ATLAS collaboration, *Measurement of the Drell-Yan triple-differential cross section in pp collisions at $\sqrt{s} = 8\text{ TeV}$* , *JHEP* **12** (2017) 059 [[arXiv:1710.05167](#)] [[INSPIRE](#)].
- [45] LHCb collaboration, *Search for Dark Photons Produced in 13 TeV pp Collisions*, *Phys. Rev. Lett.* **120** (2018) 061801 [[arXiv:1710.02867](#)] [[INSPIRE](#)].
- [46] BABAR collaboration, *Search for a Dark Photon in e^+e^- Collisions at BaBar*, *Phys. Rev. Lett.* **113** (2014) 201801 [[arXiv:1406.2980](#)] [[INSPIRE](#)].
- [47] W. Liu, S. Kulkarni and F.F. Deppisch, *Heavy neutrinos at the FCC-hh in the $U(1)_{B-L}$ model*, *Phys. Rev. D* **105** (2022) 095043 [[arXiv:2202.07310](#)] [[INSPIRE](#)].
- [48] E.W. Kolb and M.S. Turner, *The Early Universe*, Taylor and Francis (2019) [[DOI:10.1201/9780429492860](#)] [[INSPIRE](#)].
- [49] K. Ala-Mattinen, M. Heikinheimo, K. Kainulainen and K. Tuominen, *Momentum distributions of cosmic relics: Improved analysis*, *Phys. Rev. D* **105** (2022) 123005 [[arXiv:2201.06456](#)] [[INSPIRE](#)].
- [50] G. Bélanger et al., *micrOMEGAs5.0 : Freeze-in*, *Comput. Phys. Commun.* **231** (2018) 173 [[arXiv:1801.03509](#)] [[INSPIRE](#)].
- [51] X. Luo, W. Rodejohann and X.-J. Xu, *Dirac neutrinos and N_{eff}* , *JCAP* **06** (2020) 058 [[arXiv:2005.01629](#)] [[INSPIRE](#)].
- [52] L. Husdal, *On Effective Degrees of Freedom in the Early Universe*, *Galaxies* **4** (2016) 78 [[arXiv:1609.04979](#)] [[INSPIRE](#)].
- [53] M. Laine and M. Meyer, *Standard Model thermodynamics across the electroweak crossover*, *JCAP* **07** (2015) 035 [[arXiv:1503.04935](#)] [[INSPIRE](#)].
- [54] G. Arcadi, O. Lebedev, S. Pokorski and T. Toma, *Real Scalar Dark Matter: Relativistic Treatment*, *JHEP* **08** (2019) 050 [[arXiv:1906.07659](#)] [[INSPIRE](#)].
- [55] V. De Romeri, D. Karamitros, O. Lebedev and T. Toma, *Neutrino dark matter and the Higgs portal: improved freeze-in analysis*, *JHEP* **10** (2020) 137 [[arXiv:2003.12606](#)] [[INSPIRE](#)].

- [56] O. Lebedev and T. Toma, *Relativistic Freeze-in*, *Phys. Lett. B* **798** (2019) 134961 [[arXiv:1908.05491](#)] [[INSPIRE](#)].
- [57] T. Bringmann, S. Heeba, F. Kahlhoefer and K. Vangsnes, *Freezing-in a hot bath: resonances, medium effects and phase transitions*, *JHEP* **02** (2022) 110 [[arXiv:2111.14871](#)] [[INSPIRE](#)].
- [58] P. Elmfors, K. Enqvist and I. Vilja, *Thermalization of the Higgs field at the electroweak phase transition*, *Nucl. Phys. B* **412** (1994) 459 [[hep-ph/9307210](#)] [[INSPIRE](#)].
- [59] H.A. Weldon, *Simple Rules for Discontinuities in Finite Temperature Field Theory*, *Phys. Rev. D* **28** (1983) 2007 [[INSPIRE](#)].
- [60] T. Bringmann, F. Calore, A. Galea and M. Garny, *Electroweak and Higgs Boson Internal Bremsstrahlung: General considerations for Majorana dark matter annihilation and application to MSSM neutralinos*, *JHEP* **09** (2017) 041 [[arXiv:1705.03466](#)] [[INSPIRE](#)].
- [61] C.F. Uhlemann and N. Kauer, *Narrow-width approximation accuracy*, *Nucl. Phys. B* **814** (2009) 195 [[arXiv:0807.4112](#)] [[INSPIRE](#)].
- [62] S.-P. Li and X.-J. Xu, *N_{eff} constraints on light mediators coupled to neutrinos: the dilution-resistant effect*, *JHEP* **10** (2023) 012 [[arXiv:2307.13967](#)] [[INSPIRE](#)].
- [63] C. Benso, F. Kahlhoefer and H. Mansour, *Dark matter phase-in: producing feebly-interacting particles after a first-order phase transition*, *JHEP* **07** (2025) 195 [[arXiv:2504.10593](#)] [[INSPIRE](#)].
- [64] C. Caprini et al., *Detecting gravitational waves from cosmological phase transitions with LISA: an update*, *JCAP* **03** (2020) 024 [[arXiv:1910.13125](#)] [[INSPIRE](#)].
- [65] R. Allahverdi et al., *The First Three Seconds: a Review of Possible Expansion Histories of the Early Universe*, *Open J. Astrophys.* **4** (2021) astro.2006.16182 [[arXiv:2006.16182](#)] [[INSPIRE](#)].
- [66] B. Batell et al., *Conversations and deliberations: Non-standard cosmological epochs and expansion histories*, *Int. J. Mod. Phys. A* **40** (2025) 2530004 [[arXiv:2411.04780](#)] [[INSPIRE](#)].
- [67] M. Dutra and Y. Wu, *EvoEMD: cosmic evolution with an early matter-dominated era*, *Phys. Dark Univ.* **40** (2023) 101198 [[arXiv:2111.15665](#)] [[INSPIRE](#)].
- [68] M. Drees and F. Hajkarim, *Dark Matter Production in an Early Matter Dominated Era*, *JCAP* **02** (2018) 057 [[arXiv:1711.05007](#)] [[INSPIRE](#)].
- [69] PLANCK collaboration, *Planck 2018 results. X. Constraints on inflation*, *Astron. Astrophys.* **641** (2020) A10 [[arXiv:1807.06211](#)] [[INSPIRE](#)].
- [70] K. Dick, M. Lindner, M. Ratz and D. Wright, *Leptogenesis with Dirac neutrinos*, *Phys. Rev. Lett.* **84** (2000) 4039 [[hep-ph/9907562](#)] [[INSPIRE](#)].
- [71] C.R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [[arXiv:2006.10256](#)] [[INSPIRE](#)].
- [72] P. Virtanen et al., *SciPy 1.0—Fundamental Algorithms for Scientific Computing in Python*, *Nature Meth.* **17** (2020) 261 [[arXiv:1907.10121](#)] [[INSPIRE](#)].
- [73] J.D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comput. Sci. Eng.* **9** (2007) 90 [[INSPIRE](#)].
- [74] G.P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration*, *J. Comput. Phys.* **27** (1978) 192 [[INSPIRE](#)].
- [75] G.P. Lepage, *Adaptive multidimensional integration: VEGAS enhanced*, *J. Comput. Phys.* **439** (2021) 110386 [[arXiv:2009.05112](#)] [[INSPIRE](#)].
- [76] T. Bringmann et al., *Minimal sterile neutrino dark matter*, *Phys. Rev. D* **107** (2023) L071702 [[arXiv:2206.10630](#)] [[INSPIRE](#)].

- [77] P. Gondolo and G. Gelmini, *Cosmic abundances of stable particles: Improved analysis*, *Nucl. Phys. B* **360** (1991) 145 [[INSPIRE](#)].
- [78] R. Mertig, M. Bohm and A. Denner, *FEYN CALC: Computer algebraic calculation of Feynman amplitudes*, *Comput. Phys. Commun.* **64** (1991) 345 [[INSPIRE](#)].
- [79] V. Shtabovenko, R. Mertig and F. Orellana, *FeynCalc 9.3: New features and improvements*, *Comput. Phys. Commun.* **256** (2020) 107478 [[arXiv:2001.04407](#)] [[INSPIRE](#)].