PAPER • OPEN ACCESS

Photo- and hadrodisintegration constraints on massive relics decaying into neutrinos

View the article online for updates and enhancements.

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Photo- and hadrodisintegration constraints on massive relics decaying into neutrinos

S. Bianco [ID],[a] P.F. Depta [ID],[b] J. Frerick [ID],[a,c] T. Hambye [ID],[d] M. Hufnagel [ID][d] and K. Schmidt-Hoberg [ID][a,e]

[a] *Deutsches Elektronen-Synchrotron DESY,*
Notkestr. 85, 22607 Hamburg, Germany

[b] *Max-Planck-Institut für Kernphysik,*
Saupfercheckweg 1, 69117 Heidelberg, Germany

[c] *Dipartimento di Fisica, Sapienza Università di Roma,*
Piazzale Aldo Moro 5, 00185, Roma, Italy

[d] *Service de Physique Théorique, Université Libre de Bruxelles,*
C.P. 225, B-1050 Brussels, Belgium

[e] *CP3-Origins, University of Southern Denmark,*
Campusvej 55, DK-5230 Odense M, Denmark

E-mail: sara.bianco@desy.de, frederik.depta@mpi-hd.mpg.de, jonas.frerick@uniroma1.it, thomas.hambye@ulb.be, marco.hufnagel@ulb.be, kai.schmidt-hoberg@desy.de

ABSTRACT: We perform a detailed study of the cosmological constraints on the decay of a relic particle ϕ into neutrinos, $\phi \to \nu\bar{\nu}$, in particular those arising from the observed light-element abundances in the early Universe. Several processes are relevant, including final-state radiation associated with the decay, as well as subsequent interactions of the injected neutrinos with the thermal background neutrinos or between themselves. All processes generically contribute to the production of electromagnetic and often also hadronic material and may therefore induce late-time photodisintegration and hadrodisintegration reactions, i.e. the destruction of light elements that have previously been formed during BBN. Here, we examine this scenario with a Monte-Carlo inspired probabilistic approach rather than Boltzmann techniques, taking into account all of these different reactions as well as their interplay. We find the resulting constraints to be very significant, covering a broad range of previously unexplored masses and lifetimes of the relic source particle.

JCAP11(2025)072

Contents

1 Introduction

Many beyond the Standard Model (SM) scenarios suggest the existence of new unstable particles that were present during the early Universe. This is, for example, the case in dark-sector (DS) models, where in addition to dark matter (DM), other particles are present, often mediating interactions between the DM and the SM or within the DS itself. Specific examples include DM scenarios where communication with the SM proceeds via a portal interaction [1], or where the exchange of a light mediator induces sizeable DM self-interactions [2–7].

These additional particles will in general significantly impact the cosmological evolution and the associated observables. In particular, if they are unstable and decay with a lifetime shorter than the age of the Universe, they will generally affect predictions for Big Bang Nucleosynthesis (BBN) as well as for the Cosmic Microwave Background (CMB), resulting in constraints on their abundance, mass and lifetime. The strongest constraints typically arise if the decay proceeds into electromagnetic (EM) material (such as photons or charged leptons) and/or hadronic material (such as quarks or gluons), see e.g. [8, 9]. Thus, to avoid too strong constraints in various scenarios (see e.g. [6, 7]), one could assume that the decay instead proceeds predominantly into neutrinos. However, this does not imply that the constraints will become completely irrelevant, because a fraction of the neutrino energy will unavoidably still be converted into electromagnetic and/or hadronic material. In fact, it is known that heavy new particles decaying into neutrinos, inevitably produce additional on- or off-shell EW gauge bosons on top of the primary neutrinos. The effects of the resulting shower have been studied before, in particular for the EM case (see e.g. [10–15]). Other works that study the direct or indirect injection of neutrinos into the primordial plasma, include [16–19].

In this work, we perform a detailed study of the decay of a relic particle ϕ into neutrinos, $\phi \to \nu\bar{\nu}$, carefully evaluating the limits arising from the observed light-element abundances, while also including effects that have previously been neglected in the literature. We concentrate on lifetimes larger than about 10^4 s, focusing in particular on the late-time disintegration of the light elements previously synthesised during BBN. In addition to the effect of the electroweak (EW) shower from final-state radiation (FSR) mentioned above, we also properly take into account the effect of subsequent interactions of the injected neutrinos, specifically for masses below or around the EW scale. Of particular relevance are interactions with a cosmic background neutrino, which we refer to as *thermal scattering* (see [10, 16, 20] for early work), as well as interactions among the injected neutrinos themselves, which we refer to as *non-thermal scattering*. While thermal scatterings may readily induce electromagnetic material, non-thermal scatterings typically have a larger energy available and therefore also result in hadronic injections. All processes may therefore induce late-time photodisintegration reactions, while FSR and non-thermal scatterings additionally result in hadrodisintegration as well. Here, we carefully examine this scenario, using the most recent nuclear abundances and rates, and taking into account all the different reactions as well as their interplay and their impact on the cosmological parameters. In particular, we track the evolving phase-space distribution of the injected neutrinos with a Monte-Carlo inspired probabilistic approach and also take into account that for masses $m_\phi \lesssim m_{W,Z}$, the primary decay and the emission process in FSR no longer factorise. We point out that non-thermal scatterings, while being computationally rather challenging and therefore largely ignored in the literature, can be

very relevant and dominate the resulting limits in significant regions of parameter space. We also stress that hadrodisintegration reactions induced by FSR often improve the resulting limits by many orders of magnitude compared to those from photodisintegration largely considered in the literature. Overall we find that, even for the case where neutrinos are injected into the primordial plasma — which are after all electromagnetically and colour neutral particles — the resulting constraints can be very stringent, covering a broad range in mass and lifetime of the relic source particle.

This work is structured as follows: in section 2, we begin with an introductory discussion of the various processes that are relevant for neutrino injections, and of the corresponding physical effects. We then proceed in section 3 with a more detailed description of the thermal and non-thermal scattering processes, both contributing to the neutrino cascade. The aim of that section is to gain an intuitive understanding of the processes and scales at play during the injection of neutrinos into the primordial plasma. In section 4, we briefly discuss some BBN basics, including photo- and hadrodisintegration. The more technical aspects of our work are covered in section 5. There, we describe our numerical framework and explain in detail how we perform the actual calculations. We interface our code with the extended version `v2.0.0-dev` (currently on a different branch on `GitHub`) of the publicly available code `ACROPOLIS` [21], which not only enables the calculation of the evolution of the light-element abundances resulting from photodisintegration reactions but also takes into account the effects from hadrodisintegration (so far only for specific scenarios and certain assumptions). We dedicate section 6 to the discussion of the resulting limits, showing several parameter planes and the effect of a mixed branching ratio into e^+e^- before concluding in section 7. For the interested reader, we provide several appendices with more details. Throughout this work, we use natural units $\hbar = c = k_{\rm B} = 1$.

2 Basics of neutrino injections

In this work, we perform a detailed study of late-time neutrino injections into the primordial plasma and evaluate the corresponding limits that arise from changes to the light-element abundances produced during BBN. Specifically, we consider a relic particle ϕ with mass m_ϕ that decays into neutrinos with a lifetime $\tau_\phi \gtrsim 10^4\,{\rm s}$, meaning that — at the time of the decay — the light-element abundances have already reached their asymptotic values as predicted by standard, i.e. *thermal*, BBN. However, even for such lifetimes, the latter values may still deviate from those in the SM, since the expansion history of the Universe is altered due to the presence of the relic ϕ. Additionally, once thermal BBN has concluded at $t < 10^4\,{\rm s} \lesssim \tau_\phi$, the subsequent decay of ϕ and the following neutrino cascade then induce additional photo- and hadrodisintegration reactions, which lead to a second phase of *non-thermal* nucleosynthesis, which may further alter the previously fixed abundances. We postpone a detailed discussion of decays with shorter lifetimes, i.e. $\tau_\phi \lesssim 10^4\,{\rm s}$, to future work.[1] Such early injections will

---

[1]The case of short lifetimes for the decay $\phi \to \nu\bar\nu$ has been studied with some simplifying assumptions in [22]. In particular, interconversions of protons into neutrons are considered, which alter the abundances of the light nuclei already during thermal nucleosynthesis. However, for the relic masses considered in our work, such interactions would be dominated by deep inelastic scattering between nucleons and neutrinos [23], adding significant complications to the process. We are not aware of such a study.
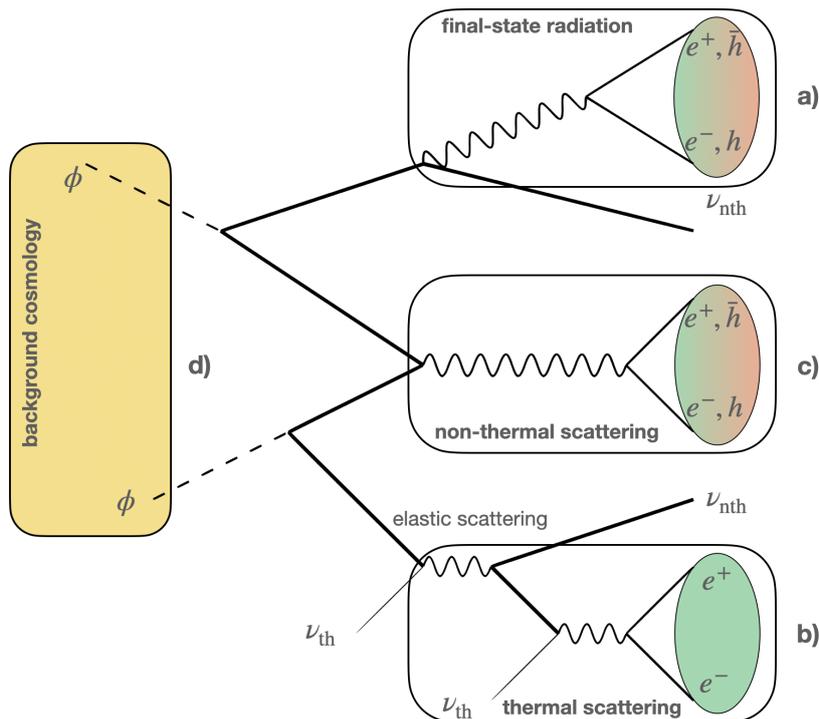
Figure 1. Graphical representation of a possible neutrino cascade, where a fraction of the energy ends up in electromagnetic and hadronic material. We briefly describe the physics of the various sub-processes in the text.

generically already change the reactions *during thermal* BBN, which adds an additional complication to the calculation.

In this section, we present the basic processes that may lead to a change of the light-element abundances, e.g. via the injection of electromagnetic and/or hadronic material, as well as their implications.

2.1 Relevant processes

The processes that are relevant for our analysis are summarised in figure 1, where we show a possible realisation of a neutrino cascade with different sub-processes. We will discuss them one-by-one in the following:

a) *Final-state radiation.* As depicted in sub-process a) of figure 1, in order to produce electromagnetic and/or hadronic material, a neutrino is not necessarily required to interact with the medium. Instead, the decay of the source particle can produce a virtual neutrino emitting a W- or Z-boson, which then leads to an EW shower that may produce photons and electromagnetically charged particles, as well as hadrons such as pions, protons, and neutrons. These radiative corrections, which are most relevant for larger ϕ masses, are already well known and have previously been studied for electromagnetic final states. Notably, the spectrum of final-state particles can be determined using `PYTHIA8.3` [24, 25] as presented in [15]. Here, we follow the latter work but also extract the hadronic component of the shower using the same approach. Given the range of

possible final states, this process can lead to both photo- and hadrodisintegration. For smaller masses, FSR becomes suppressed and we find that thermal and/or non-thermal scattering processes (see below) are more relevant. Note that in [10], the authors also consider both components of the FSR. However, they treat the hadronic branching ratio as a free parameter instead of using a robust SM prediction.

b) *Thermal scattering.* As depicted in sub-process b) of figure 1, each injected neutrino can also interact with a neutrino of the cosmic neutrino background (CνB) to produce a pair of light, charged SM particles, in particular e^+e^- pairs. For the lifetimes and thus temperatures considered in this work, the production of heavier particles such as muons or pions is kinematically suppressed for relic masses below the TeV-scale, meaning that we do not consider them here. In addition, each injected neutrino may scatter elastically, once or several times, with any CνB neutrino. As we argue below, these scatterings do not typically affect the distribution of final-state electrons and positrons.

For most of the relevant parameter space, we find that the thermal interaction rate of the injected neutrinos is well below the Hubble rate, in which case a fairly simple analytical description of these reactions can be achieved (cf. section 3.2). In particular, one can show that the injected neutrinos typically scatter at most once during their propagation through the thermal background, which constitutes an important simplification for our numerical treatment in section 5. Earlier works have already considered this effect in various ways [10, 16, 20].

c) *Non-thermal scattering.* As depicted in sub-process c) of figure 1, two injected neutrinos can also undergo interactions among each other, thereby producing a pair of SM particles. In this case, both initial neutrinos have an energy much larger than the background temperature, unlike in the case of thermal scattering. This implies that for injected neutrinos with an energy of the order of the EW scale, all possible final states are usually available. After their production, the produced SM particles may either decay or hadronise, and the resulting shower can again be simulated using `PYTHIA8.3`. In this work, we assume that all unstable particles except for neutrons decay immediately, which we find to be a good approximation. After the shower, we therefore end up with *(i)* electromagnetic material, which induces subsequent photodisintegration reactions, *(ii)* secondary neutrinos, which are at much smaller energies than the initial ones and can therefore be neglected, and *(iii)* protons and neutrons, which induce subsequent hadrodisintegration reactions.

Let us stress that, when aiming for a conservative limit, it is generally important to take into account both hadronic and EM final states. This is because hadrodisintegration tends to overproduce deuterium via the destruction of helium-4, while photodisintegration often leads to a direct destruction of deuterium. Consequently, both effects go in different directions and therefore may cancel each other. Prior works [26–28] also discussed the effect of non-thermal neutrino scatterings. However, they make use of a significantly simplified setup while relying on long-outdated observations.
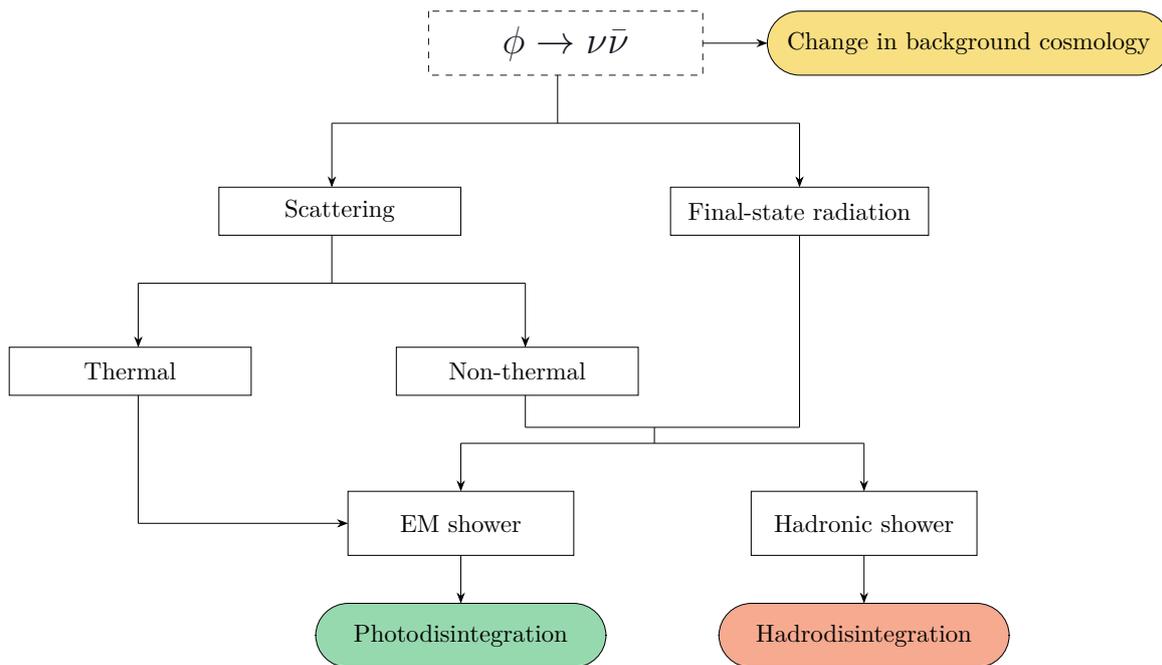
Figure 2. Flowchart of the processes considered in this work.

d) *Background cosmology.* As depicted in sub-process d) of figure 1, and in addition to the previously discussed effects leading to photo- and hadrodisintegration, the extra energy density associated with the long-lived relic ϕ and the injected neutrinos also leads to a change of the cosmological evolution, which in turn affects BBN and the resulting light-element abundances. On a similar note, this also leads to a change in the effective number of relativistic degrees of freedom, generally parameterised by the effective number of neutrinos $N_{\rm eff}$, which in turn can be constrained by CMB measurements.

2.2 A first look at the resulting limits

In figure 2, we provide a condensed visualisation of the procedure that we use to calculate the limits on neutrino injections. As part of this flowchart, we emphasise the individual branches that we discussed in the previous section, but also the fact that there is an interference between the different effects as indicated by the convergence of the tree before the boxes labelled "EM shower" and "Hadronic shower". Before entering into the details of the analysis, it is useful to qualitatively discuss the physical effects of the various processes. Here, we do so by showing in advance the results obtained in this work for two fixed masses, i.e. $m_\phi = 100\,{\rm GeV}$ and $m_\phi = 500\,{\rm GeV}$. By fixing the mass to one of these values and scanning over the abundance and the lifetime of ϕ, we obtain the limits presented in figure 3. These figures show the total limits that we obtain by fully including all effects, as well as the partial limits that we obtain by retaining only individual effects of the various processes above.[2] In the process, we also separate the effects of EM and hadronic injections. Note that this section heavily relies on various concepts that are only explained briefly, but will be discussed in more

---

[2]This separation is somewhat artificial, since a true separation is basically impossible due to the rich interplay between the different effects.

detail later. However, whenever this is the case, we provide forward or literature references to aid the reader.

We start by discussing the results for $m_\phi = 100\,\text{GeV}$. For this choice of mass, one of the least important contributions comes from the EM component of the EW shower, originating from *final-state radiation* during the decay process (cf. e.g. [15]) (dashed orange). This is not surprising, as this contribution is suppressed for values of $m_\phi/2$ below the EW scale, i.e. when the emitted gauge boson is forced to be off-shell (cf. figure 4). These constraints are particularly weak for short lifetimes, since any EM injections that happen too early will additionally be shut off as all photons with energy $E > m_e^2/(22T)$ get rapidly depleted due to e^+e^- pair production on the background photons (cf. e.g. [29, 30] and eq. (4.9)). This becomes particularly evident from the sharp decrease in sensitivity towards $\tau_\phi \sim 10^4\,\text{s}$. Overall, the resulting (photodisintegration) constraints are mainly driven by an underproduction of deuterium ($\tau_\phi \lesssim 3 \times 10^6\,\text{s}$) or an overproduction of helium-3 ($\tau_\phi \gtrsim 3 \times 10^6\,\text{s}$). In addition to this, the hadrons produced in the shower may induce sufficient hadrodisintegration reactions (dotted orange), which in-turn significantly effect the light-element abundances. For small lifetimes, $\tau_\phi < 10^7\,\text{s}$, the corresponding limits turn out to be orders of magnitude more stringent than the ones originating purely from EM material. This is due to efficient hadrodisintegration induced by energetic neutrons. In fact, neutrons are completely stopped by their interactions with the background photons only for injections at $t \lesssim 10^2\,\text{s}$ [31], i.e. much earlier than any lifetime considered in this work. Consequently, at the temperatures we are interested in, neutrons experience only negligible energy loss compared to other particles (cf. e.g. [31]). This allows them to disintegrate helium-4 at earlier times leading to deuterium overproduction (cf. section 4.3 and appendix D), when other particles would dissipate their energy too quickly and thus impose only very weak constraints. For large lifetimes, $\tau_\phi > 10^7\,\text{s}$, the hadrodisintegration effects turn out to be comparable to the EM ones. In both cases, the limits stem mainly from an overproduction of deuterium, and there is no cancellation between both effects. In conclusion, the total FSR limit (solid orange) is stronger than the individual ones. Note that for such large lifetimes, the effect of the hadronic part is plateauing around $\tau_\phi \sim 10^9$ s.

When it comes to *thermal scattering*, hadronic injections are irrelevant, as the corresponding final states are kinematically impossible due to the small energy of thermal neutrinos (cf. section 3.2.1). However, the same is not true for EM injection (solid red). For large lifetimes, the corresponding constraint rapidly gets much weaker than the one from FSR (solid orange). This behaviour is expected, as scattering processes depend quadratically on the number density of the initial-state particles and hence are suppressed quadratically due to cosmic expansion, whereas FSR is only suppressed linearly. For $\tau_\phi \lesssim 10^7\,\text{s}$, these injections lead to similar constraints than the ones from hadrons injected via FSR (dotted orange). Like in the case of FSR, the overall limits mainly stem from either an underproduction of deuterium ($\tau_\phi \lesssim 3 \times 10^6\,\text{s}$) or an overproduction of helium-3 ($\tau_\phi \gtrsim 3 \times 10^6\,\text{s}$).

Regarding *non-thermal scattering*, the corresponding bounds from photodisintegration (dashed blue) are very similar to the ones from thermal scattering for $\tau_\phi < 10^7\,\text{s}$. For longer lifetimes, however, they do not decouple as fast as the thermal ones and thus become much stronger. While interactions between two non-thermal neutrinos also suffer from the
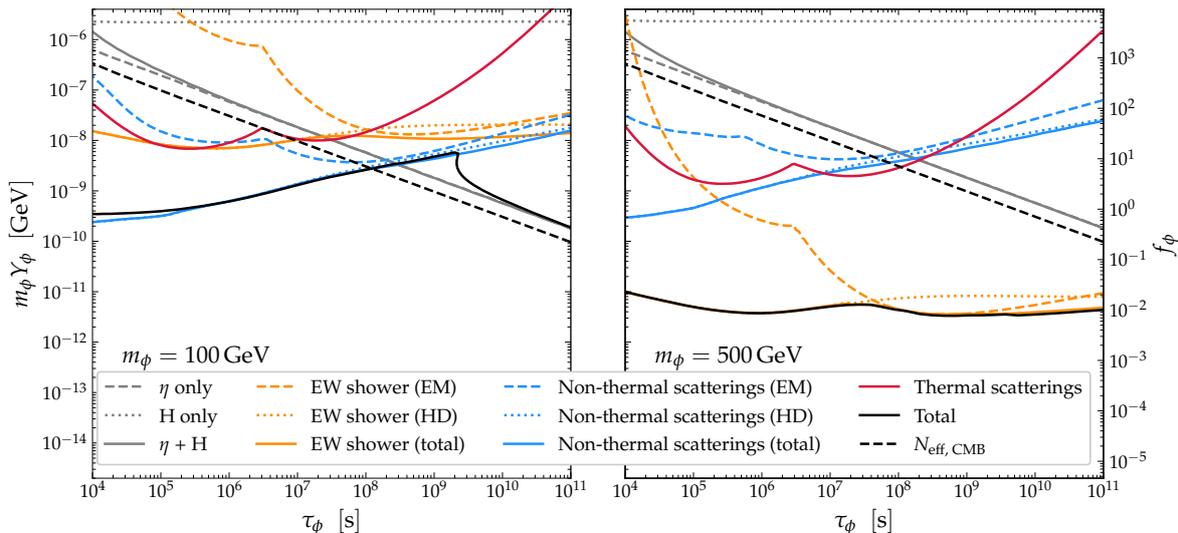
Figure 3. *Left:* BBN limits for $m_\phi = 100\,\text{GeV}$ in the $\tau_\phi - m_\phi Y_\phi$ parameter plane. Besides the overall limit (solid black), we also show the individual contributions that arise from the effects of the EW shower, i.e. FSR, (orange), thermal scattering (red), non-thermal scattering (blue), and the modified background cosmology (grey). For comparison, we also indicate the CMB N_{eff} limit (dashed black). *Right:* same, but for $m_\phi = 500\,\text{GeV}$.

aforementioned dilution suppression, the non-relativistic relic which sources the neutrinos, dilutes slower than the background neutrinos, giving a relative boost to the impact of the non-thermal scattering. The same is true for the hadronic component (dotted blue) when considering large lifetimes. Similar to FSR, the contributions from EM and hadronic injections add up in this regime, thus leading to slightly stronger bounds (solid blue). For lifetimes $\tau_\phi < 10^8\,\text{s}$, the hadrodisintegration limits from non-thermal scattering are significantly stronger than any other limit. As a result, they dominate the overall limit over a large range of parameter space and therefore are essentially identical to the combined limit (solid black). For short lifetimes, $\tau_\phi \lesssim 10^5\,\text{s}$, we further observe that the individual limit due to non-thermal scattering is even stronger than the total limit. However, this effect is rather artificial and stems from the (non-perfect) separation of the different effects: when thermal scatterings are turned off, less neutrinos scatter thermally into e^+e^- or $\nu\nu$, which increases the pool of potential particles that can undergo non-thermal scattering. Since the latter mechanism is dominant in this regime, this leads to somewhat stronger limits.

Finally, the limits are also affected by the modified *background cosmology*, which can manifest in the form of two different effects. On the one hand (dashed grey), a change in N_{eff} leads to a different best-fit value for the baryon-to-photon ratio η (cf. (4.5)), which has a direct influence on the deuterium abundance, which decreases (increases) for larger (smaller) values of η (cf. e.g. section 4.1). On the other hand (dotted grey), the additional energy density of ϕ and the non-thermal neutrinos leads to a change of the Hubble rate H, which has a direct effect on the BBN products (cf. e.g. [32]). For the lifetimes considered in this work, i.e. those that lie sufficiently past the end of thermal BBN, the resulting bound can effectively be interpreted as an N_{eff} bound from BBN, since it probes the additional energy density (as

encoded in H) that is present during this time. In general, this effect is comparatively weak, and its inclusion changes the overall limit that originates from both cosmological effects (solid grey) only below $\tau_\phi \sim 10^6$ s. Notably, the total limit becomes weaker in this region, due to a cancellation between an η-driven deuterium underproduction and an H-driven deuterium overproduction. Overall, the limits arising from the modified background cosmology are the dominant BBN limits for very large lifetimes, $\tau_\phi > 2 \times 10^9$ s. Thus, the full BBN constraint (solid black), which takes into account all effects simultaneously, goes from a regime where it is dominated by the non-thermal hadrodisintegration effect at small lifetimes (see above) to a regime where it is dominated by the cosmological BBN limit at large lifetimes. Interestingly, in between these two regimes, the full BBN limit follows the one from non-thermal scattering for almost one order of magnitude in lifetime at around $\tau_\phi \sim 10^9$ s. This is a prime example of a cancellation between different contributions which would have been invisible when only showing each contribution individually. In this region, the aforementioned underproduction of deuterium due to the change in η is (over-)compensated for by the overproduction of deuterium induced by non-thermal scattering reactions.

For comparison, we also show the CMB limit that arises due to the presence of an additional radiation component during recombination, parameterised in terms of the effective number of neutrinos $N_{\rm eff}$ (dashed black). Numerically, we find that the resulting constraint is well parameterised by $f_\phi(\tau_\phi/10^4\,{\rm s}) \simeq 680$ (also cf. appendix A). Deviations from this naive formula only arise for large masses, when the non-thermal neutrinos lose a significant amount of their energy due to the injection of EM/hadronic material. Overall, we find that the CMB limits are stronger than the one from BBN only for lifetime above 10^8 s. By also considering BBN constraints, it is therefore possible to exclude much larger parts of parameter space.

So far we have discussed the resulting constraints for $m_\phi = 100\,{\rm GeV}$. For smaller masses, the behaviour is qualitatively similar. In particular, the total BBN constraint displays the same general behaviour, dominated by the same two regimes just discussed above with a similar cancellation feature in between (cf. figure 10). However, for larger masses, important differences arise, as illustrated in right panel of figure 3 for $m_\phi = 500\,{\rm GeV}$. Most importantly, both constraints arising from EM and hadronic FSR injections become much more stringent. While the qualitative behaviour remains similar to the one for $m_\phi = 100\,{\rm GeV}$, the strength of the limit increases by as much as three orders of magnitude. As a result, the limits from FSR injections come to dominate all other constraints. In particular, the hadronic contribution (dotted orange) dominates for lifetime below $\sim 10^7$ s, while becoming comparable to the EM one (dashed orange) above. Notably, taking into account the effects from hadrodisintegration improves the constraints by several orders of magnitude for small lifetimes. Overall, the combined constraint (solid black) depends only mildly on the lifetime. Finally, most other limits do not change qualitatively. While the bounds arising from the background cosmology (solid grey) remain very similar, the ones from thermal scattering (solid red) increase significantly but, as before, are never dominant.

3 Details of neutrino injections

In this section, we provide a more rigorous description of the simplified picture given in section 2.1. This includes, among other aspects, discussions of neutrino oscillations, kinematic

thresholds, and details on the various scattering processes. We want to provide, as far as possible, some analytical intuition for the processes and some of the interesting effects we found, before we introduce our numerical framework in section 5. Note that in this section, we do not give an explicit discussion of FSR, details of which can be found in [15].

3.1 Neutrino oscillations

Once produced, and potentially before undergoing further interactions, the injected neutrinos may oscillate. Depending on the temperature at the time of injection, the timescale for a neutrino to oscillate can be much shorter than the one for scattering. Indeed, for the range of lifetimes considered in this work, the oscillation time (see e.g. [33])

$$t_{\text{osc}} \sim \frac{T}{\Delta m^2} \sim 10^{-5} \, \text{s} \left(\frac{T}{\text{MeV}} \right) \tag{3.1}$$

is much shorter than both the Hubble time and the interaction time (also cf. eqs. (3.6) and (3.7) below). The only timescale smaller than this, is the one for FSR. For the present discussion, the important conclusion is that neutrino oscillations are basically instantaneous, so that all flavours will be present basically immediately after injection / FSR, even if the initial decay produces only one flavour. We therefore utilise the following prescription: if a neutrino of a given flavor $\alpha \in \{e, \mu, \tau\}$ is injected into the plasma, it is instead treated as a superposition of all three flavor eigenstates $\beta \in \{e, \mu, \tau\}$ weighted by the oscillation probabilities encoded in the PMNS matrix (a more detailed description can be found in section 5)

$$(\mathbf{M}_{\text{osc}})_{\alpha\beta} = \sum_{i=1}^{3} |U_{\alpha i}|^2 |U_{\beta i}|^2 \approx \begin{pmatrix} 0.55 & 0.17 & 0.28 \\ 0.17 & 0.45 & 0.37 \\ 0.28 & 0.37 & 0.35 \end{pmatrix} . \tag{3.2}$$

This matrix can be applied to the spectrum of decay products for arbitrary branching ratios, but also to the spectrum resulting from elastic scattering reactions to determine the distribution of outgoing neutrinos in each case. In this work, we largely concentrate on injections of the form $\phi \rightarrow \bar{\nu}_e \nu_e$ for concreteness (cf. figure 3 above), noting that other choices will hardly affect the final results (cf. section 6).

3.2 Thermal scattering of the form $\nu \nu_{\text{th}} \rightarrow e^+ e^-$

3.2.1 Thresholds and timescales

We begin the discussion of thermal interactions with an investigation of the kinematic thresholds and the interaction rates, before providing a more rigorous analytical treatment.

The threshold energy of injected neutrinos E_{ee}, required to produce an electron-positron pair while scattering off a background neutrino at temperature T, is approximately given by (here, we use that $T_\nu \sim \mathcal{O}(1)T$)

$$E_{ee} \sim \frac{m_e^2}{T} \gtrsim 26 \, \text{MeV} \left(\frac{10 \, \text{keV}}{T} \right) , \tag{3.3}$$

with m_e being the electron mass. Consequently, since we focus on neutrino injections at $T \lesssim 10 \, \text{keV}$, the injected neutrinos must have a kinetic energy well above the one of the

Process	Cross-section coefficient $\Sigma(s)$
$\nu_i \bar{\nu}_i \to \nu_j \bar{\nu}_j$	1
$\nu_i \nu_i \to \nu_i \nu_i$	6
$\nu_i \nu_j \to \nu_i \nu_j$	3
$\nu_i \bar{\nu}_j \to \nu_i \bar{\nu}_j$	1
$\nu_j \bar{\nu}_j \to \nu_j \bar{\nu}_j$	4
$\nu_e \bar{\nu}_e \to e\bar{e}$	$\left[\left(1 + 2\frac{m_e^2}{s}\right)\left(1 + 4s_w^2 + 8s_w^4\right) - 3\frac{m_e^2}{s}\right]\sqrt{1 - \frac{4m_e^2}{s}}$
$\nu_\mu \bar{\nu}_\mu \to e\bar{e}$	$\left[\left(1 + 2\frac{m_e^2}{s}\right)\left(1 - 4s_w^2 + 8s_w^4\right) - 3\frac{m_e^2}{s}\right]\sqrt{1 - \frac{4m_e^2}{s}}$
$\nu_\tau \bar{\nu}_\tau \to e\bar{e}$	$\left[\left(1 + 2\frac{m_e^2}{s}\right)\left(1 - 4s_w^2 + 8s_w^4\right) - 3\frac{m_e^2}{s}\right]\sqrt{1 - \frac{4m_e^2}{s}}$

Table 1. Cross-section coefficients $\Sigma(s)$ (cf. eq. (3.5)) for all available processes in the small energy regime (Fermi theory). These expressions have been adapted and generalised from table 1 in [22]. Here, $i, j \in \{e, \mu, \tau\}$ with $i \neq j$, and we use the shorthand notation $\sin \theta_w \equiv s_w$ for the sine of the Weinberg angle.

background neutrinos in order for this production to be feasible. For smaller energies, non-thermal neutrinos can no longer annihilate into electron-positron pairs, meaning that they no longer need to be tracked for the neutrino cascade. However, they do still contribute to N_{eff}. For the production of heavier particles, e.g. $\mu^+\mu^-$ or $\pi^+\pi^-$, the respective threshold energies $E_{\mu\mu}$ and $E_{\pi\pi}$ must instead be significantly larger

$$E_{\mu\mu/\pi\pi} \sim \frac{m_{\mu/\pi}^2}{T} \gtrsim (1.1/1.8) \text{ TeV} \left(\frac{10\,\text{keV}}{T}\right) , \tag{3.4}$$

with similar values (up to factors of $\mathcal{O}(1)$) being obtained also for the mixed production of, e.g. $e^+\mu^-$. For such high energies, the overall limit is dominated by FSR (cf. sub-process (a) in figure 1), meaning that these processes are not relevant to determine the overall limit. We will therefore not consider them further in the following.

For the purpose of this work, any neutrino that undergoes scattering with the neutrino background can therefore either produce an electron-positron pair or a pair of (potentially different) non-thermal neutrinos (cf. section 3.3). In general, we find that — for $T \lesssim 10\,\text{keV}$ — the corresponding centre-of-mass energy s is always small compared to the EW scale as described by the Z-boson mass m_Z, i.e. $s \ll m_Z^2$. This implies that the various processes can be described well within the effective Fermi theory without the need to invoke the full EW theory. Within this approximation, the cross-section for any of the available scattering reactions can be parameterised as

$$\sigma(s) = \Sigma(s)\frac{G_F^2 s}{6\pi} . \tag{3.5}$$

Here, G_F is the Fermi constant and $\Sigma(s)$ is a numerical coefficient, which differs by process (cf. table 1). For $s \gg m_e^2$, which is true for most parts of parameter space, $\Sigma(s) \simeq \text{const.}$ holds, meaning that all cross-sections are approximately linear in s.

Depending on the energy of the injected neutrino, the interaction time $t_{\rm th}$ corresponding to thermal scattering can be either faster or slower than the Hubble time t_H. In general, these times can be approximated as

$$t_H \sim \frac{1}{H(T)} \sim \frac{m_P}{T^2} \sim 1\,s \left(\frac{T}{\rm MeV}\right)^{-2} , \qquad (3.6)$$

$$t_{\rm th} \sim \frac{1}{G_F^2 T^4 E} \sim 10^{-3}\,{\rm s} \left(\frac{T}{\rm MeV}\right)^{-4} \left(\frac{E}{10\,{\rm GeV}}\right)^{-1} , \qquad (3.7)$$

with the Planck mass m_P. For the last expression, we used that roughly $t_{\rm th} \sim 1/(\sigma_{\rm th}\bar{n}_\nu)$ with the number density $\bar{n}_\nu \sim T^3$ of thermal neutrinos and the thermal scattering cross-section $\sigma_{\rm th} \propto G_F^2 s \propto G_F^2 E T_\nu$ with $T_\nu \propto T$ within effective Fermi theory (also cf. eqs. (3.9)–(3.11) below for a more detailed calculation). Notably, the interaction time $t_{\rm th}$ depends on the energy E of the non-thermal neutrino, meaning that for certain choices of E, the interaction time can be much shorter than the Hubble time. Specifically, $t_{\rm th} < t_H$ is true for

$$E \gtrsim 100\,{\rm GeV} \left(\frac{10\,{\rm keV}}{T}\right)^2 . \qquad (3.8)$$

Consequently, a significant amount of scatterings per Hubble time only occurs for heavy ϕ with lifetimes $\tau_\phi \sim 10^4\,$s corresponding to a neutrino injection at $T \sim 10\,$keV. In this case, the injected neutrinos basically scatter instantaneously, i.e. as soon as they are produced. However, this does not imply that the neutrinos efficiently thermalise, since each scattering reaction reduces the energy of the final-state neutrino, making thermal scattering progressively less efficient and potentially spoiling eq. (3.8) before thermal equilibrium can be established.

For most of the relic masses and lifetimes of interest, the thermal interaction rate is below the Hubble rate and interactions will generally not take place instantaneously (if at all), which is why we concentrate on this regime in the analytical discussion below.

3.2.2 A simplified analytical approach

In this subsection, we expand on the previous discussion by providing a simplified analytical approach for handling EM injection due to reactions of the form $\nu\nu_{\rm th} \to e^+ e^-$, with a non-thermal (i.e. injected) neutrino ν and a thermal neutrino $\nu_{\rm th}$. In the process, we provide a simplified analytical treatment to gain some intuition for this process and its implications. Afterwards, we discuss the shortcomings of this treatment, thus motivating the more sophisticated treatment outlined in section 5, which is used for the final analysis.

Let us start by considering a single neutrino injected at time $t_{\rm inj}$ corresponding to the temperature $T_{\rm inj}$.[3] Assuming a reaction $X + Y \to Z$ among massless initial-state particles, the interaction rate Γ_Z describing the number of reactions happening per time, is given by (this expression is obtained by simplifying the corresponding collision operator in the Boltzmann equation)

$$\Gamma_Z(T, E) = \frac{g_Y}{16\pi^2 E^2} \int_0^\infty {\rm d}\epsilon\, f_Y(\epsilon) \int_0^{4E\epsilon} {\rm d}s\, s \cdot \sigma_Z(s) . \qquad (3.9)$$

---

[3]For simplicity, we neglect neutrino oscillations in the following (simplified) analytical discussion.

Here, $\sigma_Z(s)$ is the cross-section of the reaction, E is the energy of X and ϵ is the energy of Y, with the latter particles featuring a phase-space distribution f_Y and g_Y degrees of freedom. Consequently, the corresponding rate $\Gamma_{ee}(T, E)$ for the process $\nu\bar{\nu}_{\mathrm{th}} \to e^+e^-$ with $X = \nu$, $Y = \bar{\nu}_{\mathrm{th}}$ and $Z = e^+e^-$ is given by

$$\Gamma_{ee}(T, E) = \frac{g_\nu}{16\pi^2 E^2} \int_0^\infty \mathrm{d}\epsilon\, f_{\nu,\mathrm{th}}(\epsilon) \int_0^{4E\epsilon} \mathrm{d}s\, s \cdot \sigma_{ee}(s)\,, \tag{3.10}$$

where $\sigma_{ee}(s)$ is the cross-section defined in eq. (3.5) with the correct expression for $\Sigma(s)$ from table 1. Assuming for simplicity that $m_e = 0$, in which case $\Sigma(s) = \mathrm{const.} \equiv \Sigma_\infty$, we have $\sigma_{ee}(s) \equiv \Sigma_\infty G_F^2 s/(6\pi)$ and consequently

$$\Gamma_{ee}(T, E) \overset{m_e=0}{\simeq} \frac{g_\nu}{16\pi^2 E^2} \int_0^\infty \mathrm{d}\epsilon\, f_{\nu,\mathrm{th}}(\epsilon) \int_0^{4E\epsilon} \mathrm{d}s\, \frac{\Sigma_\infty G_F^2 s^2}{6\pi} = \frac{7\pi}{90} \Sigma_\infty G_F^2 E T_\nu^4\,. \tag{3.11}$$

Note that in this expression, Σ_∞ assumes different values depending on the flavours of the initial-state neutrinos participating in the reaction (cf. table 1).

As already stated above — in order to gain a simplified analytical understanding —, we consider the case where the Hubble rate defines the shortest timescale of the problem (when ignoring FSR and neutrino oscillations). Under these assumptions, the energy of each injected neutrino is modified mainly by redshift, meaning that at a given redshift z, the remaining energy of the neutrino is given by $E(z) = E_{\mathrm{inj}}(z/z_{\mathrm{inj}})$, where E_{inj} and z_{inj} refer to the neutrino energy and the redshift at the time of injection. By enforcing the condition $E(z)T \lesssim m_e^2$, we therefore find that a neutrino can no longer produce electron-positron pairs at temperatures $T \lesssim (m_e^2/E_{\mathrm{inj}})(z_{\mathrm{inj}}/z)$. This is also the point at which $\sqrt{s} \sim 2\,m_e$, meaning that our assumption in eq. (3.11), i.e. $\Sigma(s) \sim \mathrm{const.}$, breaks down. However, we find that — for sufficiently heavy relics — this problem is of little relevance, since the corresponding temperatures are small enough to no longer allow for efficient energy injection anyway, e.g. because of dilution effects or because the relic has already fully decayed. Nevertheless, it is worth noting that the analytical treatment of this subsection technically only applies to sufficiently heavy relics.

To calculate the total fraction of energy $\zeta_{\mathrm{em}}(t_{\mathrm{inj}})$ that is transferred into the plasma by a neutrino injected at time t_{inj}, let us consider an infinitesimal time step $\mathrm{d}t$. In this case, the differential energy fraction $\mathrm{d}\zeta_{\mathrm{em}}$ injected during the interval $\mathrm{d}t$ is given by

$$\mathrm{d}\zeta_{\mathrm{em}} \simeq \frac{E(t)}{E_{\mathrm{inj}}} \Gamma_{ee}(t)\mathrm{d}t\,. \tag{3.12}$$

Here, $\Gamma_{ee}(t)\mathrm{d}t$ is the probability for a non-thermal neutrino with energy $E(t)$ to scatter into e^+e^- during the interval $t \to t+\mathrm{d}t$, in which case it transfers a fraction $E(t)/E_{\mathrm{inj}}$ of its original energy E_{inj} into EM material. Using the time-temperature relation $\mathrm{d}T/\mathrm{d}t = -H(T)T$, we then find

$$\frac{\mathrm{d}\zeta_{\mathrm{em}}}{\mathrm{d}T} = \frac{\mathrm{d}\zeta_{\mathrm{em}}}{\mathrm{d}t} \times \frac{\mathrm{d}t}{\mathrm{d}T} = -\frac{E(T)}{E_{\mathrm{inj}}} \Gamma_{ee}(T) \times \frac{1}{H(T)T} = -\left(\frac{\Gamma_{ee}}{H}\right)_{\mathrm{inj}} \frac{T^3}{T_{\mathrm{inj}}^4}\,. \tag{3.13}$$

In the last step, we have used $\Gamma_{ee}(T) \propto E(T)T^4$ according to eq. (3.11), $H(T) \propto T^2$ and $E(T) = E_{\mathrm{inj}}(z/z_{\mathrm{inj}}) = E_{\mathrm{inj}}(T/T_{\mathrm{inj}})$. By integrating this expression from $T = T_{\mathrm{inj}}$ to $T \to 0$,
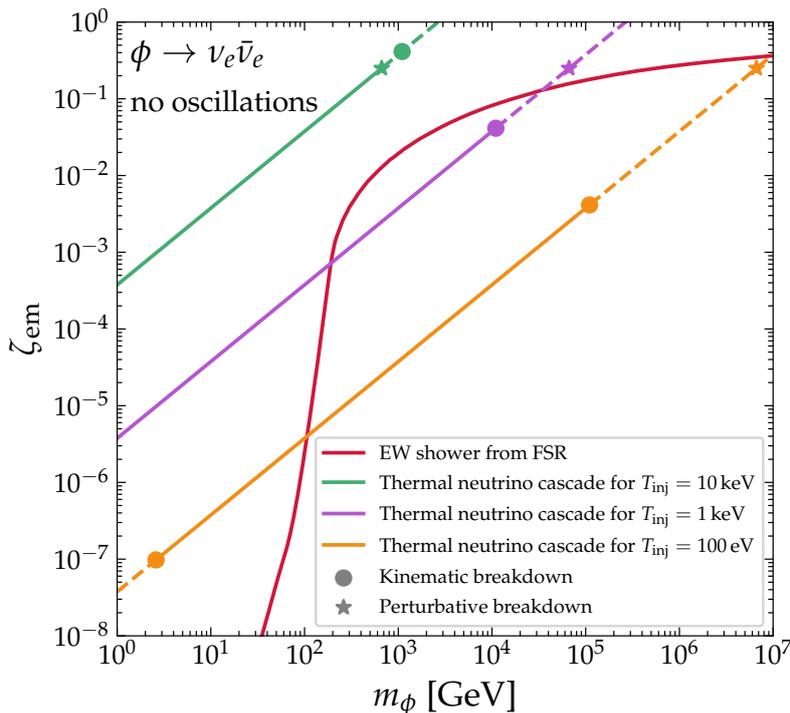
Figure 4. Comparison of the fraction $\zeta_{\rm em}$ of energy that is injected in the form of EM material as a function of the relic mass m_ϕ within *(i)* the context of a relic ϕ that decays into two neutrinos, which subsequently cause an EW shower due to FSR (red, cf. [15]), and *(ii)* our simplified analytical description of the neutrino cascade after injection with $T_{\rm inj} = 10\,{\rm keV}, 1\,{\rm keV}, 100\,{\rm eV}$ (green, purple, orange). We further indicate the values of $m_\phi = 2E_{\rm inj}$ for which electron-positron production becomes kinematically inaccessible (lower circle), muons become kinematically available (upper circle), or the annihilation rate approaches the Hubble rate (star). In all of these cases, our simplified analytical description breaks down.

we therefore find

$$\zeta_{\rm em}(T_{\rm inj}) = \int_{T_{\rm inj}}^{0} {\rm d}T \, \frac{{\rm d}\zeta_{\rm em}}{{\rm d}T} \simeq \int_{0}^{T_{\rm inj}} {\rm d}T \, \left(\frac{\Gamma_{ee}}{H}\right)_{\rm inj} \frac{T^3}{T_{\rm inj}^4} = \frac{1}{4}\left(\frac{\Gamma_{ee}}{H}\right)_{\rm inj}. \tag{3.14}$$

Notably, this result demonstrates that the total fraction of the injected energy in eq. (3.14) is smaller than the naive estimate $(\Gamma_{ee}/H)_{\rm inj}$ by a factor 4, courtesy to the energy loss due to redshifting before scattering. Additionally, given that the scattering reactions of the neutrinos usually happen on timescales larger than the Hubble time (see above), the explicit polynomial dependence on T in eq. (3.13) — contrary to an expression $\propto \delta(T - T_{\rm inj})$ —, shows that the energy transfer into EM material is *not instantaneous*, but rather happens over a prolonged period of time; a key feature that is unique to neutrinos. In contrast, particles that are either electromagnetically or colour charged, would dump their energy instantaneously, thus leading to an immediate change of the light-element abundances. Non-thermal neutrinos, on the other hand, can still be around until well-after the lifetime of the relic to induce disintegration processes.

In figure 4, we visualise these results by showing the energy fraction $\zeta_{\rm em}$ that was derived within this framework using the electron-neutrino annihilation cross-section of table 1

in the $m_e = 0$ limit. We find that $\zeta_{\rm em}$ scales linearly with the mass of the relic ϕ as expected from eq. (3.14). For comparison, we also indicate the values of $\zeta_{\rm em}$ that are obtained when considering only the EW shower from FSR as calculated in [15]. Notably, for small masses or early injections, thermal scattering dominates and therefore is expected to lead to a strengthening of the resulting constraints. Additionally, we also indicate the kinematic threshold (lower circle) below which electron-positron production becomes inaccessible according to eq. (3.3). While we did not take into account the kinematic suppression just above the threshold, we find that this would have an effect only for low-energy, late-time injections as shown in the figure for $T_{\rm inj} = 100\,{\rm eV}$. However, there also exists a second kinematic breakdown (upper circle), which is driven by the production of particles heavier than electrons, specifically muons. As outlined in eq. (3.4), this effect only plays a role for very heavy particles or very early injections. Therefore, this effective cut-off never becomes dominant in our work.

In addition to the kinematic breakdowns, there also exists a perturbative breakdown for our description (star), which we estimate by solving $\Gamma_{ee}(T_{\rm inj})/H(T_{\rm inj}) = 1$ for $T_{\rm inj}$. Beyond this threshold, eq. (3.14) is no longer valid, as it does not account for multi-scatterings. However, we find that for masses beyond $1\,{\rm TeV}$, the limits are always dominated by the EW shower from FSR, meaning that neither the perturbative nor the upper kinematic threshold play any significant role.

Note that — besides the shortcomings already discussed above and the fact that we only consider a single neutrino species — for the purpose of describing the non-thermal interactions (see below), we need information on the full energy distribution of the high-energy neutrinos, which is impossible to obtain within this simplified approach. Thus, we will ultimately need to abandon this analytical description for a more rigorous numerical treatment. A thorough discussion of this approach can be found in section 5.

3.3 Thermal scattering of the form $\nu\nu_{\rm th} \to \nu\nu$

So far, we have assumed that — before scattering into e^+e^-, the non-thermal neutrinos only lose energy due to redshift. However, they may also scatter with a CνB to produce a pair of (potentially different) secondary neutrinos instead. In this case, both neutrinos redistribute their initial energy between each other, meaning that the thermal background neutrino becomes non-thermal. At first sight, it seems that such reactions could have a large effect on the subsequent production of e^+e^- pairs, since table 1 shows that scattering into neutrinos is about 5 to 10 times faster than scattering into e^+e^-. However, a closer look reveals that this is not the case. In fact, the annihilation rate only depends linearly on the energy of the individual particles, which implies that — at a given temperature — the annihilation rate of both final-state, non-thermal neutrinos is the same as the annihilation rate that the single initial-state neutrino would have had, if it did not scatter before, i.e. (cf. eq. (3.11) with $\Sigma = {\rm const.} \equiv \Sigma_\infty$ from table 1)

$$\Gamma_{ee}(T, E_1) + \Gamma_{ee}(T, E_2) = \frac{7\pi}{90}\Sigma_\infty G_F^2 T_\nu^4 \left(E_1 + E_2\right) = \Gamma_{ee}(T, E_1 + E_2)\,. \qquad (3.15)$$

Consequently, for thermal scattering, the total energy that is injected in the form of e^+e^- pairs does not depend on how that energy is distributed among the different neutrinos.[4] Since we assume that interactions of the form $\nu\nu_{\text{th}} \to \nu\nu$ do not significantly alter the thermal neutrino abundance, it is reasonable to disregard these interactions. In other words, it is justified to consider the energy injection induced by a single neutrino, as discussed in section 3.2. However, the introduction of non-thermal collisions will void this argument due to the breakdown of Fermi theory, as discussed in the next section.

3.4 Non-thermal scattering between two high-energy neutrinos

Naively, one could expect that scattering reactions between two injected, high-energy neutrinos are negligible due to their comparatively small number density (this has been assumed e.g. in [10, 16, 20]). However, as already emphasised in section 2, in the context of this work, we instead find that such interactions are highly relevant over large parts of parameter space due to their larger centre-of-mass energy, thus leading to scattering that occurs on resonance with much higher cross-section. In this work, we therefore utilise a rather complete treatment of non-thermal neutrino scattering to calculate the resulting constraints.

While the small number density of high-energy neutrinos suppresses interactions among them, this suppression can be compensated for by a significant increase in the scattering cross-section, since the larger centre-of-mass energy (compared to thermal scattering) makes the Z-resonance in the s-channel diagrams of the annihilation processes kinematically available for large parts of parameter space. The larger centre-of-mass energy also implies the existence of additional final-states (as opposed to only $\nu\nu$ and e^+e^-), leading, in particular, to hadronic particles, which have a strong influence on the resulting limits. To the best of our knowledge, the inevitable injection of high-energy hadrons from non-thermal scatterings has not yet been considered in the literature, but still provides the leading constraints for relatively small lifetimes and masses below $\sim 100\,\text{GeV}$ (see above). In figure 5, we plot the total cross-sections (including all possible two-body final states) as a function of the centre-of-mass energy $\sqrt{s}$ for various initial-state neutrino flavours. As expected, the resulting curves approach the sum of all cross-sections in table 1 asymptotically for small $\sqrt{s}$.[5] In the figure, we further indicate the mass thresholds for the charged lepton, as well as the Z-boson mass, thus highlighting the sharp transitions at threshold, e.g. at $\sqrt{s} = 2m_e$ and $\sqrt{s} = m_Z$. In fact, for each lepton l_i, there exist two thresholds, i.e. the standard annihilation threshold from $\nu_i\bar{\nu}_i \to l_i\bar{l}_i$ at $\sqrt{s} = 2m_{l_i}$ and the mixed threshold(s) from $\nu_i\bar{\nu}_j \to l_i\bar{l}_j$ at $\sqrt{s} = (m_{l_i} + m_{l_j})$.

To quantify the effect of the non-thermal interactions, in figure 6, we additionally show the average fraction $\langle\zeta_{\text{em}}^{\text{n-th}}\rangle$ of EM energy that is injected into the plasma due to one reaction happening at a centre-of-mass energy $\sqrt{s}$. These results have been obtained by utilising `PYTHIA8.3` to simulate the particles shower induced by the final-state particles (see section 5 below for more information).[6] We find that in the low-energy regime, $\langle\zeta_{\text{em}}^{\text{n-th}}\rangle$ reaches a plateau

---

[4]As discussed in the previous section, this result is only valid well above the e^+e^- production threshold.

[5]Note that the given cross-sections have been calculated at leading order, which is why processes with charge conjugated initial states, like e.g. $\nu_e\nu_e$ and $\bar{\nu}_e\bar{\nu}_e$, feature the same cross-section. Also, again at leading order, the cross-sections for $\nu_e\nu_\mu$, $\nu_e\nu_\tau$ and $\nu_\mu\nu_\tau$ are identical. We therefore do not show all combinations.

[6]Note that this fraction is the same in the cosmic rest frame and the centre-of-mass frame due to the isotropic distribution of the injected particles.
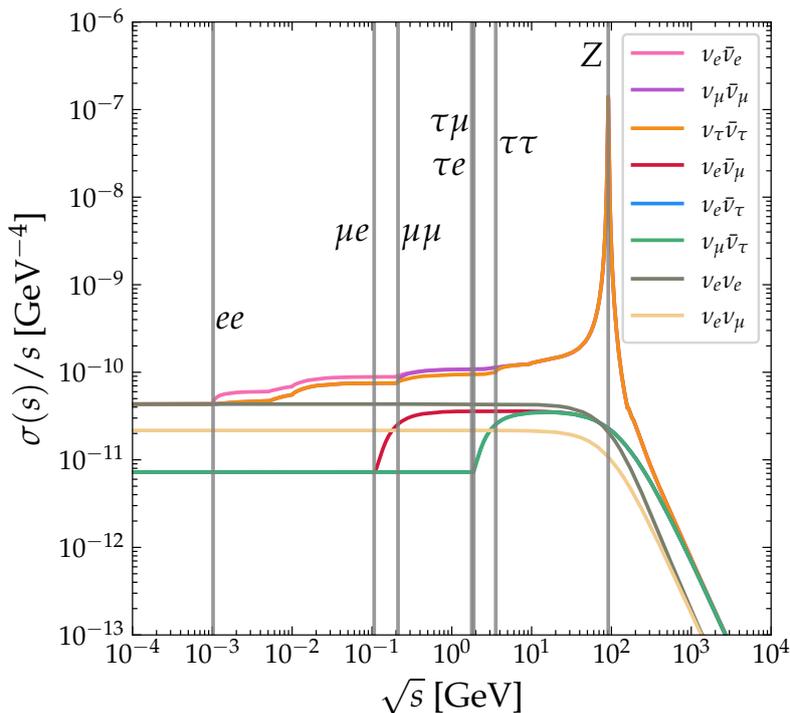
Figure 5. The total cross-section for different initial-states and all possible two-body final states. Note that only interactions between neutrinos and anti-neutrinos can go on resonance, while the remaining processes consistently drop below the naive Fermi scaling at high energies. All data has been extracted by running `MadGraph5` [34]. Note that the contributions for $\nu_e\bar{\nu}_\tau$ (blue) and $\nu_\mu\bar{\nu}_\tau$ (green), while strictly speaking slightly different, are virtually indistinguishable in the plot.

in agreement with our discussion on thermal scattering (cf. table 1).[7] Overall, we find that once kinematically allowed, the amount of EM material injected by neutrino collisions of different flavour (red, blue, green), does not vary much with energy and ultimately converges to a constant value at very high energies. This result is expected, since such combinations of neutrinos only undergo t-channel scattering and hence do not feel the Z-resonance. For interactions with identical flavours in the initial state, there instead exist small variations with energy. Interestingly, all corresponding lines (orange, purple, pink) converge around the resonance, at which point the final states are dominated by Z-boson decay, irrespective of the initial-state configuration. For higher energies, the lines diverge again, since the large phase-space favours non-resonant interactions over the kinematically suppressed resonance.

For centre-of-mass energies around the Z-pole, the resonance provides the dominant contribution for non-thermal scattering. In particular, for a collision of two high-energy neutrinos, the resonance effectively selects the scattering angle in such a way to put the interaction on resonance (as the probabilities for other configurations are much smaller). However, in this work, we still take into account the full energy distribution (which we assume to be isotropic) and the exact cross-sections, without relying on a narrow-width

---

[7]We have interpolated between the `PYTHIA8.3` results and those from Fermi theory for $\sqrt{s} < 1\,\mathrm{GeV}$. Moreover, in the plot, we only consider combinations that can have non-neutrino final states at leading order.
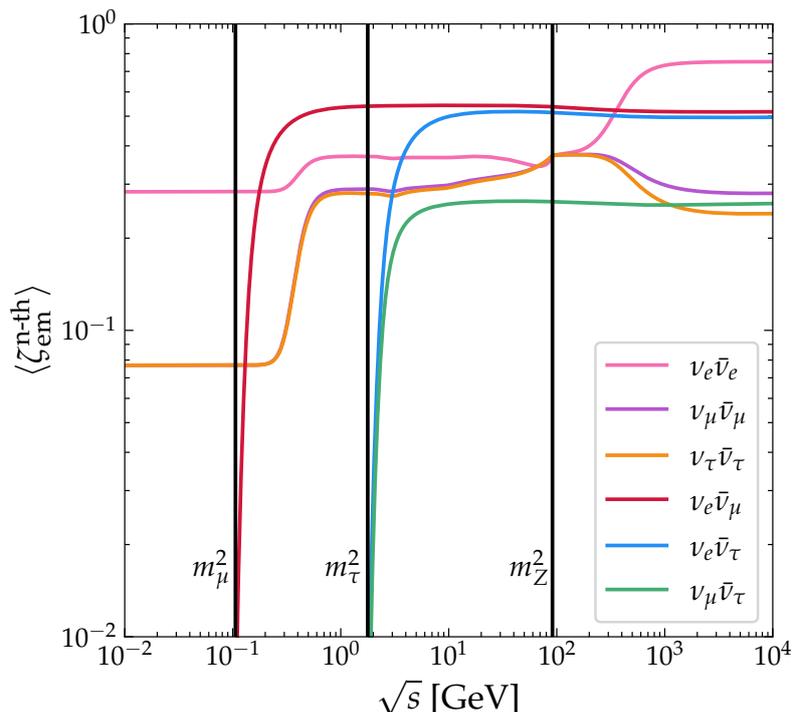
Figure 6. The average fraction of energy injected in the form of EM material from collisions of two high-energy neutrinos, as calculated via `PYTHIA8.3`. We only show neutrino-antineutrino combinations, since — at leading order — only those combinations feature non-neutrino final states due to lepton-number conservation.

approximation. In particular, this means that scattering is also allowed, albeit suppressed, in the energy range below the resonance (cf. section 5 for more details).

Regarding the injection of hadronic material, the distribution of final states obtained by `PYTHIA8.3` reveals that non-thermal interactions on average inject less than one nucleon per scattering, with an average kinetic energy of a few percent of the sum of the initial neutrino energies. At first glance, this seems to be a negligible contribution. However, in section 4.3 we will emphasise that hadronic injections should still be handled with care. In fact, we find that this contribution yields the dominant constraints for significant parts of parameter space. We provide more details on this contribution in section 5.

At this point, it is worth noting that the inclusion of non-thermal scattering reactions significantly complicates the calculation. Many of the arguments that were used for the simplified approach in section 3.2.2 rely on the linearity of the underlying processes, among other assumptions. However, many of these simplifications are no longer justified for non-thermal scattering, since *(i)* non-thermal, high-energy collisions inject both EM and hadronic material, *(ii)* the corresponding interaction rate is inherently quadratic in the amount of non-thermal neutrinos, and *(iii)* Fermi theory breaks down for high-energy collisions at and beyond the resonance. As a result, the amount of injected EM (and hadronic) material is no longer independent of whether the neutrinos underwent $\nu\nu_{\rm th} \to \nu\nu$ scatterings before, as no equivalent of eq. (3.15) exists for non-thermal scattering. Therefore, it is crucial to take into account the temporal evolution of the energy distribution of non-thermal neutrinos. However,

for the parameter space considered in this work, we find that $H \gg \Gamma_{\nu\nu_{\rm th} \to \nu\nu}$, meaning that two injected neutrinos usually only scatter once and therefore are not much affected by previous (non-existent) scattering reactions. However, scatterings of the form $\nu\nu_{\rm th} \to \nu\nu$ will still deplete the number of high-energy neutrinos. In practice, we remove each neutrino from the spectrum that scatters in this way, as a second scattering of the final-state particles is unlikely. This approximation, which considerably simplifies the simulation, is justified as long as the aforementioned relation between interaction rate and Hubble rate is true.

3.5 Interactions with other particles

Collisions between the injected neutrinos and the relic ϕ (such as $\nu\phi \to \nu\phi$) are highly suppressed due to the small coupling corresponding to lifetimes $\tau_\phi > 10^4\,{\rm s}$ and can therefore be neglected in our calculation. Additionally, neutrinos do not interact directly with the background photons (apart from loop-corrections), and their interaction with the background electrons/positrons, like e.g. $\nu e^\pm \to \nu e^\pm$, is suppressed by the small baryon-to-photon ratio. The same argument applies to direct interactions between neutrinos and any light nuclei N, i.e. neutrinodisintegration of the form $\nu N \to \ldots$ (cf. appendix B for a numerical demonstration). However, the neutrino-induced interconversion between neutrons and protons during and shortly after thermal BBN, e.g. $\nu p \to n e^+$, can play an important role [22]. Nevertheless, we find that for the relatively long-lived relics considered in this work, the limits presented in [22] are subdominant, meaning that the inclusion of this effect would not change our results.

4 A brief overview of thermal and non-thermal BBN

In this section, we provide a concise review of the standard approach for deriving bounds on BSM models using BBN results. We begin by introducing the observational light-element abundances and by briefly discussing the correlation between $N_{\rm eff}$ and η. Next, we summarise the main features of photodisintegration and present the basic framework in which ACROPOLIS operates. Finally, we describe the hadrodisintegration formalism, which we have incorporated into `v2.0.0-dev` of ACROPOLIS for the purpose of this study.

4.1 The light-element abundances

In this work, we use the latest recommended values for the observed light-element abundances for $\mathcal{Y}_p$, D/^{1}H as reported in [35] and for ^{3}He/D (only as an upper limit) as reported in [36]:

$$\mathcal{Y}_p = (2.45 \pm 0.03) \times 10^{-1}\,, \tag{4.1}$$

$$\mathrm{D}/^1\mathrm{H} = (2.547 \pm 0.029) \times 10^{-5}\,, \tag{4.2}$$

$$^3\mathrm{He}/\mathrm{D} = (8.3 \pm 1.5) \times 10^{-1}\,. \tag{4.3}$$

Following the usual convention for the light-elements abundance, we use $N/N' \equiv Y_N/Y_{N'}$ with $N, N' \in \{^1\mathrm{H}, \mathrm{D}, {}^3\mathrm{He}, {}^7\mathrm{Li}\}$, $Y_N \equiv n_N/n_b$, and n_b being the baryon number density. For ^{4}He, we instead use the common parameterisation $\mathcal{Y}_p \equiv 4Y_{^4\mathrm{He}}$. In addition to the elements mentioned above, the primordial lithium abundance is inferred to be [35]

$$^7\mathrm{Li}/^1\mathrm{H} = (1.6 \pm 0.3) \times 10^{-10}\,. \tag{4.4}$$

However, this measurement is known to be in contradiction with the value predicted by standard BBN, a discrepancy commonly referred to as the (cosmological) lithium problem.[8] In this work, we do not try to resolve this tension and therefore (conservatively) take into account only the limits from the other abundances.

To calculate the evolution of the different abundances theoretically during thermal BBN (potentially with a modified background cosmology), we utilise a modified version of `AlterBBN` [39, 40], denoted as `AlterAlterBBN` in the following,[9] which can handle arbitrary background cosmologies and further incorporates the most recent set of nuclear reaction rates [41–45]. As part of the calculation, we utilise the value of the baryon-to-photon ratio η obtained by the latest Planck measurements [46]. Notably, it is well known that η has a strong correlation with N_{eff}, which we incorporate by following the prescription presented in [47]. More precisely, the latest Planck data, i.e. the 95% confidence region ellipse (Planck TT, TE, EE+lowE+lensing+BAO) in the $\Omega_b h^2 - N_{\text{eff}}$ plane in figure 26 of [46], implies that for a given value of N_{eff}, the best-fit value of η is given by

$$\eta_{N_{\text{eff}}} = \overline{\eta} + r\sigma_\eta \frac{N_{\text{eff}} - \overline{N}_{\text{eff}}}{\sigma_{N_{\text{eff}}}} \,, \tag{4.5}$$

with

$$\overline{\eta} = 6.128 \times 10^{-10}\,, \quad \sigma_\eta = 4.9 \times 10^{-12}\,, \quad \overline{N}_{\text{eff}} = 2.991\,, \quad \sigma_{N_{\text{eff}}} = 0.169\,, \quad r = 0.677\,. \tag{4.6}$$

Since $\text{D}/^1\text{H}$ exhibits a rather strong dependence on η, the above uncertainty further introduces an additional error on the deuterium abundance, and consequently the total experimental uncertainty $\sigma_{\text{D}/^1\text{H}}^{\text{exp}}$ becomes

$$\sigma_{\text{D}/^1\text{H}}^\eta = \left| \frac{\text{d}(\text{D}/^1\text{H})}{\text{d}\eta} \sigma_\eta \sqrt{1 - r^2} \right|_{\eta = \eta_{N_{\text{eff}}}} \approx 0.024 \times 10^{-5}\,, \tag{4.7}$$

$$\Rightarrow \quad \sigma_{\text{D}/^1\text{H}}^{\text{exp}} = \sqrt{\left(\sigma_{\text{D}/^1\text{H}}^{\text{obs}}\right)^2 + \left(\sigma_{\text{D}/^1\text{H}}^\eta\right)^2} \approx 0.035 \times 10^{-5}\,, \tag{4.8}$$

with $\sigma_{\text{D}/^1\text{H}}^{\text{obs}} = 0.029 \times 10^{-5}$ from eq. (4.2) above. For the other elements, the uncertainty induced by η is much smaller and thus negligibly compared to the one from observations. In this work, we use these measurements to calculate constraints at 95% C.L., identical to the procedure used in [48] (also cf. this paper for more information).

4.2 A quick look at photodisintegration

The particles originating from neutrino collisions, e.g. among others the e^+e^- pairs discussed in section 3.2, feature an energy E_0 that is much larger than the one of thermal species, i.e. $E_0 \gg T$. If the injected particles are electromagnetic, i.e. electrons, positrons and photons, they can efficiently scatter on the background photons, thus inducing an EM cascade, which ultimately produces a spectrum of non-thermal photons. If the energy of the particles initiating this cascade is above the electron pair-creation threshold, i.e. $E_0 > E_{ee}^{\text{th}} \simeq m_e^2/(22T)$, any

---

[8]Currently, an astrophysical solution to this discrepancy appears most likely [37, 38].

[9]For the corresponding source code, see https://github.com/hep-mh/alteralterbbn.

photon produced during this cascade with an energy above this threshold is rapidly depleted. In this case, it is well known that the cascade produces a universal spectrum of non-thermal photons, which is well approximated by [29][10]

$$f_{\gamma,\text{univ}}(T, E) \simeq \frac{S_{\text{em}}(T)}{\Gamma_\gamma(T, E)} \times \begin{cases} K_0 (E/E_X)^{-3/2} & , \quad E < E_X \,, \\ K_0 (E/E_X)^{-2} & , \quad E_X < E < E_{ee}^{\text{th}} \,, \\ 0 & , \quad E > E_{ee}^{\text{th}} \,. \end{cases} \tag{4.9}$$

Here, $K_0 = E_0 E_X^{-2} \left[2 + \ln(E_{ee}^{\text{th}}/E_X)\right]^{-1}$, $E_X = m_e^2/(80T)$, Γ_γ is the total scattering rate of non-thermal photons with the background, and S_{em} is the source-term describing the total amount of injected EM material.

Note that in the context of photodisintegration, the distribution function f_γ is usually defined to be differential in energy instead of momentum f_γ, with both of them being related via

$$f_\gamma(T, E) = g_\gamma \frac{Ep}{2\pi^2} f_\gamma(T, p) \,. \tag{4.10}$$

Here, $g_\gamma = 2$ is the number of photon degrees of freedom.

To evaluate eq. (4.9) for our scenario, we have to compute the source term S_{em} originating from the neutrino cascade. If all particles are injected with the same energy E_0,

$$S_{\text{em}}(T) = \sum_{x=e^\pm, \gamma} \frac{\mathrm{d}n_x^{\text{inj}}}{\mathrm{d}t} = \sum_{x=e^\pm, \gamma} \frac{1}{E_0} \frac{\mathrm{d}\rho_x^{\text{inj}}}{\mathrm{d}t} \,, \tag{4.11}$$

where the different $\mathrm{d}n_x^{\text{inj}}/\mathrm{d}t$ ($\mathrm{d}\rho_x^{\text{inj}}/\mathrm{d}t$) denote the number (energy) densities of injected particles per time interval originating from the neutrino cascade. Notably, the first expression for S_{em} in eq. (4.11) is indeed only true if all particles are injected with the same energy E_0; however, the second expression remains valid even if particles of different energies $E_0' < E_0$ contribute to the same universal spectrum with maximal injection energy E_0, which is the case for our scenario with $E_0 = m_\phi/2$. A more detailed calculation of this quantity is presented in section 5 below. At this point, let us note that S_{em} receives independent contributions from all relevant processes, meaning that in our case

$$S_{\text{em}}(T) = S_{\text{em}}^{\text{dcy}}(T) + S_{\text{em}}^{\text{fsr}}(T) + S_{\text{em}}^{\text{th}}(T) + S_{\text{em}}^{\text{n-th}}(T) \,. \tag{4.12}$$

Here, $S_{\text{em}}^{\text{fsr}}(T)$, $S_{\text{em}}^{\text{th}}(T)$, and $S_{\text{em}}^{\text{n-th}}(T)$ are the source terms originating solely from FSR, thermal, and non-thermal scattering, respectively (a detailed calculation of the different source terms is provided in section 5 below). Finally, $S_{\text{em}}^{\text{dcy}}(T)$ encodes potential direct contributions from the decay of the relic, e.g. if the latter also decays into electron-positron pairs with a certain branching ratio BR_{ee}. We explore this possibility in more detail below.

---

[10]We made sure that the universal spectrum is always a good approximation for the parameters considered in this work. Therefore, we do not have to explicitly solve the equations governing the EM cascade to obtain the spectrum (cf. e.g. [49, 50]).

After the EM cascade, the non-thermal photons originating from the EM cascade, i.e. the ones encoded in $f_{\gamma,\text{univ}}$, will initiate photodisintegration reactions, e.g. $\gamma D \to np$ among others. The Boltzmann equations governing these reactions can be written as [21] (dropping the t/T dependence for convenience)

$$\left[\frac{dn_X}{dt}\right]_{\text{photo}} = \sum_j n_j N_{j\gamma\to X} \int_0^\infty dE\, f_\gamma(E)\sigma_{j\gamma\to X}(E) - n_X \sum_{j'} \int_0^\infty dE\, f_\gamma(E)\sigma_{X\gamma\to j'}(E)\,. \tag{4.13}$$

Here, $f_\gamma = f_{\gamma,\text{univ}}$, X is any of the light nuclei under consideration, $N_{j\gamma\to X}$ is the number of X nuclei that is produced in the reaction $j\gamma \to X$, and $\sigma_{j\gamma\to X}$ ($\sigma_{X\gamma\to j'}$) are the cross-sections for the different reactions creating (destroying) X, which can be found in [21]. Given an expression for $S_{\text{em}}(T)$, we could in principle use the most recent stable version `v1.3.2` or the updated version `v2.0.0-dev` of `ACROPOLIS` to calculate $f_{\gamma,\text{univ}}$ and to solve eq. (4.13) in order to obtain the final light-element abundances after photodisintegration. However, the neutrino cascade also injects hadronic material and the hadrodisintegration reactions usually happen at similar timescales as the photodisintegration reactions. Consequently, both effects are generally intertwined and eq. (4.13) must be extended by additional reactions in order to handle both effects simultaneously.

4.3 A quick look at hadrodisintegration

Overall, the hadrons produced in the neutrino cascade lead to significant effects, even though their production is generically subject to a stronger kinematic suppression.

As already mentioned above, non-thermal scattering reactions (but also final-state radiation) usually produce less than one nucleon per interaction, with kinetic energies at the GeV-scale, i.e. large enough to destroy ^{4}He in a nucleon-nucleus interaction. This usually leads to a *production* of deuterium. However, since photodisintegration mainly causes a *destruction* of deuterium, there might be cancellations between both effects, which is why it is imperative to consider both of them simultaneously.

In fact, despite the smaller number of injected hadrons (compared to EM material), hadrodisintegration is still important due to the following two reasons:

1. The cross-section for the destruction of ^{4}He via a proton or neutron ($p/n + {}^4\text{He} \to \text{D} + \dots$) is about two orders of magnitude bigger than the one for the destruction of D via a photon ($\gamma + \text{D} \to \dots$)

2. ^{4}He is more than three orders of magnitude more abundant than D, which further boosts the relative importance of the hadrodisintegration rate.

Therefore, even if the ratio between injected hadrons and photons is $\mathcal{O}(10^{-5})$, hadrodisintegration effects can still be as significant as photodisintegration effects. In addition, for small lifetimes, $\tau_\phi \lesssim 10^4\,$s, high-energy photons are efficiently depleted before undergoing any disintegration reactions, which is not true for neutrons.

The effect of injecting hadrons during or after thermal BBN has already been extensively discussed in [31, 51], and more recently in [52, 53]. However, since there does not exist any public code, in this work, we perform an independent calculation based on these publications. Further details on our implementation can be found in appendix D, where we present a concise overview of the most important calculational steps. In this appendix, we also point out slight discrepancies between our calculation and the one in the literature, as well as how this changes the final disintegration processes.

While the details of our calculation are delegated to the appendix, let us point out how eq. (4.13) is modified in the presence of hadrodisintegration reactions. To this end, we introduce two parameters $\xi_X^{n/p}[n_j](T, K)$ for each element $X \in \{p, n, \mathrm{D}, \mathrm{T}, {}^3\mathrm{He}, {}^4\mathrm{He}\}$ (cf. appendix D for details on how to calculate these parameters). This quantity encodes the amount of X particles that are produced (or destroyed if negative) due to the injection of a single neutron/proton with kinetic energy K at a given temperature T. Moreover, it depends explicitly on the abundance of the present light elements, as indicated by the argument n_j. Given this quantity, the Boltzmann equation governing hadrodisintegration can be written as (again dropping the t/T dependence for convenience)

$$\left[\frac{\mathrm{d}n_X}{\mathrm{d}t}\right]_{\mathrm{hadro}} = \sum_{y=n,p} \int_0^\infty \mathrm{d}K \; \xi_X^y[n_j](K) \frac{\mathrm{d}^2 n_y^{\mathrm{inj}}}{\mathrm{d}t\mathrm{d}K}(K). \tag{4.14}$$

Here, $\mathrm{d}^2 n_{n/p}^{\mathrm{inj}}/(\mathrm{d}t\mathrm{d}K)$ is the number density of neutrons/protons that are injected per time and kinetic-energy interval. Combining this equation with eq. (4.13), the total Boltzmann equation governing non-thermal BBN (NBBN) is consequently given by

$$\left[\frac{\mathrm{d}n_X}{\mathrm{d}t}\right]_{\mathrm{NBBN}} = \left[\frac{\mathrm{d}n_X}{\mathrm{d}t}\right]_{\mathrm{photo}} + \left[\frac{\mathrm{d}n_X}{\mathrm{d}t}\right]_{\mathrm{hadro}}. \tag{4.15}$$

At this point, let us note that in order to solve this equation, we further make the following two simplifications:

1. Both processes that inject hadrons, i.e. non-thermal scattering and FSR, inject an equal amount of protons and neutrons. We checked that this statement holds true within our setup to a reasonable accuracy.

2. Both processes inject hadrons of a single, *average* kinetic energy, which however still depends on the initial energy and the time of injection (cf. figure 9 below for details). This is in line with other simplifications we make in section 5.

Under these assumptions, eq. (4.14) simplifies to

$$\left[\frac{\mathrm{d}n_X}{\mathrm{d}t}\right]_{\mathrm{hadro}} \simeq \sum_{y=n,p} \left(\xi_X^y[n_j](K_{\mathrm{hd}}^{\mathrm{fsr}}) \frac{\mathrm{d}n_{\mathrm{hd}}^{\mathrm{fsr}}}{\mathrm{d}t} + \xi_X^y[n_j](K_{\mathrm{hd}}^{\mathrm{n\text{-}th}}) \frac{\mathrm{d}n_{\mathrm{hd}}^{\mathrm{n\text{-}th}}}{\mathrm{d}t} \right). \tag{4.16}$$

Here, $\mathrm{d}n_{\mathrm{hd}}^{\mathrm{n\text{-}th}}/\mathrm{d}t$ ($\mathrm{d}n_{\mathrm{hd}}^{\mathrm{fsr}}/\mathrm{d}t$) is the number density of injected hadrons per time interval that are produced with an average kinetic energy $K_{\mathrm{hd}}^{\mathrm{n\text{-}th}}$ ($K_{\mathrm{hd}}^{\mathrm{fsr}}$) due to non-thermal scattering (FSR). A detailed calculation of these quantities is presented in section 5 below. Note again, that the right-hand side of this equation explicitly depends on the light-element abundances

n_j via the parameters ξ_X^y. Finally, let us note that similar to eq. (4.12), we can also define a hadronic source term

$$S_{\rm hd}(T) = S_{\rm hd}^{\rm fsr}(T) + S_{\rm hd}^{\rm n\text{-}th}(T)\,, \tag{4.17}$$

which receives contributions from FSR and non-thermal scattering reactions. While not directly entering the corresponding Boltzmann equation, this quantity is still relevant for the handling of the neutrino cascade below (cf. eq. (5.34)).

For the parameters considered in this work, eq. (4.15) for NBBN decouples from the Boltzmann equation for thermal, i.e. standard, BBN (SBBN). In practice, we therefore *(i)* track the neutrino cascade to calculate the modified Hubble rates as well as the quantities $S_{\rm em}$, $K_{\rm hd}^{\rm fsr}$, etc. (see section 5 below for more information), *(ii)* use the modified background cosmology as an input to solve the equations for SBBN with the help of `AlterAlterBBN`, and *(iii)* use the resulting abundances as initial conditions for our updated version of `ACROPOLIS`, which then solves eq. (4.15) with the expressions from eqs. (4.13) and (4.14), the latter of which has been simplified specifically for our scenario.

At this point, let us also briefly comment on how we handle the theoretical errors entering eq. (4.15). Specifically, we consider *(i)* the errors on the nuclear reaction rates used in `AlterAlterBBN`, and *(ii)* the errors on our calculation of ξ_X^y, which amount to $\sim 20\% \equiv \epsilon_\xi$ (cf. appendix D). Regarding *(i)*, we run `AlterAlterBBN` three times with the high, mean and low values of the nuclear reaction rates, thus obtaining three sets of final SBBN abundances $Y_X^{\Gamma\,{\rm high},0}$, $Y_X^{\Gamma\,{\rm mean},0}$, and $Y_X^{\Gamma\,{\rm low},0}$, respectively. We then solve eq. (4.15) once for each set of these initial abundances, leading to the final abundances $Y_X^{\Gamma\,{\rm high}}$, $Y_X^{\Gamma\,{\rm mean}}$, and $Y_X^{\Gamma\,{\rm low}}$. Regarding *(ii)*, we solve eq. (4.15) with the initial condition $Y_X^{\Gamma\,{\rm mean},0}$ and $\xi_X^y \{(1+\epsilon_\xi), 1, (1-\epsilon_\xi)\}$ to obtain three sets of final abundances $Y_X^{\xi\,{\rm high}}$, $Y_X^{\xi\,{\rm mean}}$, and $Y_X^{\xi\,{\rm low}}$. We then calculate the theoretical error on the different abundances Y_X as the minimal difference between all six values presented above. Finally, we combine the theoretical error calculated in this way with the observational error on the abundances presented in section 4.1 to obtain the overall exclusion limits.

A note on anti-nucleons. In our analysis, we do not include the effect of anti-baryons, which are subleading. Hadrodisintegration reactions caused by anti-nucleons have been studied in [52] (henceforth KKMT18). Using figure 6 of KKMT18, we find that the effect of anti-neutrons (denoted by "$\xi_{H_i,\bar{n}}$") becomes bigger than the effect of protons (denoted by "$\xi_{H_i,p}$") only at very high energies, which are not relevant for the parameters considered in this work. Since all of our interactions conserve baryon number, high-energy anti-baryons and baryons are always produced in equal numbers. In this work, we explicitly take into account the fact that the non-thermal collisions respect baryon-number conservation. In practice, we subtract the injected baryons from the final abundance of hydrogen to make sure that it does not bias the baryon-to-photon ratio.

A note on proton-neutron conversions. Protons and neutrons are not the only hadrons injected via the cascade. As shown in [31] (henceforth KKM05), injected pions also contribute to a change in the light-element abundances. However, this happens more indirectly via

the interconversion of protons and neutrons during SBBN, i.e. for $T \gtrsim 0.1\,\mathrm{MeV}$ or $t \lesssim 100\,\mathrm{s}$, which is below the lifetimes we consider in this work.

4.4 Interlude: the CMB N_{eff} constraint

In the literature, constraints from BBN are often compared to the one originating from N_{eff} at the time of recombination. In the context of this work, the contribution of the additional non-thermal neutrinos to the effective number of neutrinos at the time t_{rec} of recombination can be written as

$$N_{\mathrm{eff}} = \frac{\rho_\nu^{\mathrm{th}}(t_{\mathrm{rec}}) + \rho_\nu^{\mathrm{n\text{-}th}}(t_{\mathrm{rec}})}{2\frac{7}{8}\frac{\pi^2}{30}\left(\frac{4}{11}\right)^{4/3} T(t_{\mathrm{rec}})^4} \equiv [3 + \Delta N_{\mathrm{eff}}(t_{\mathrm{rec}})] \left(\frac{11}{4}\right)^{4/3} \left(\frac{T_\nu(t_{\mathrm{rec}})}{T(t_{\mathrm{rec}})}\right)^4 . \qquad (4.18)$$

In this expression, which defines ΔN_{eff}, we have split the energy density of the neutrinos ρ_ν into a thermal and a non-thermal part, with the former one being sourced by standard cosmology while the latter one arises solely from the decay of the dark-sector particles.[11] Like the one for BBN from H and η, the limits considered in this section also derive from a change in the background cosmology, with the Planck mission [46] constraining ΔN_{eff} to $\Delta N_{\mathrm{eff}}(t_{\mathrm{rec}}) < 0.29$. In appendix A, we provide a simplified parameterisation of the resulting constraint in the context of this work.

5 Overview of the numerical treatment

5.1 Introductory remarks

This section is dedicated to a thorough discussion of the computational approach that is used to track the neutrino cascade, as well as the resulting injections of EM and hadronic material. As already mentioned above, since we only consider lifetimes sufficiently beyond the end of thermal BBN, we can effectively "factorise" the calculation into three different steps:

- *Neutrino cascade.* In a first step, we track the evolution and decay of the relic ϕ, as well as the secondary production of EM and hadronic material due to the injected neutrinos. Here, we use the properties of the relic, e.g. its mass m_ϕ and its lifetime τ_ϕ, as input, while returning as an output *(i)* the parameters describing the modified background cosmology, e.g. H, $\mathrm{d}T/\mathrm{d}t$ etc., and *(ii)* the parameters quantifying the EM and hadronic injections, e.g. S_{em}, $K_{\mathrm{hd}}^{\mathrm{fsr}}$, etc. A summary of the different input and output parameters can be found in table 2.

- *Thermal nucleosynthesis (SBBN).* In a second step, the modified cosmological history, specifically H and $\mathrm{d}T/\mathrm{d}t$, is used as an input for `AlterAlterBBN`, which solves the Boltzmann equations for SBBN and returns the resulting light-element abundances.

---

[11]Naively, additional relativistic neutrinos will always increase N_{eff}, provided that interactions with the thermal bath (or even with other non-thermal neutrinos) do not substantially deplete their energy density. However, this is not necessarily the case, as shown in [54], where interactions of neutrinos with the thermal neutrino background were considered. These interactions result in the reheating of the neutrino bath, which, in turn, partially rethermalises with the electromagnetic (EM) sector, leading to a decrease in N_{eff}. However, in our work this effect does not play a role.

- *Non-thermal nucleosynthesis (NBBN).* In a final step, the modified background cosmology, the parameters describing the EM and hadronic injections, as well as the abundances obtained from `AlterAlterBBN`, are used as an input for our modified version of `ACROPOLIS`, which calculates the late-time modification of the previously produced light elements due to photo- and hadrodisintegration reactions. Note that, in order to use `ACROPOLIS` (cf. [21]), it is required to implement a model encoding the desired setup. As to not unnecessarily clutter the codebase of `ACROPOLIS`, we implement this model as part of an independent code called `Xena`,[12] which can be used in conjunction with `v.2.0.0-dev` of `ACROPOLIS`.

In this section, we mainly focus on the first step of the calculation, i.e. the neutrino cascade, since details on the other steps can be found in the literature. Specifically, *(i)* [39] discusses the equations governing SBBN, as well as details on their solution within `AlterBBN`, which is used as the basis for `AlterAlterBBN`, *(ii)* [21] discusses the inner workings of `ACROPOLIS`, including details on how to handle the process of photodisintegration, and *(iii)* appendix D — which is largely based on [31, 52] — thoroughly discusses the hadronic cascade leading to hadrodisintegration as implemented in the modified version `v.2.0.0-dev` of `ACROPOLIS`, further including several references to the literature.

To determine the modified background cosmology as well as the parameters describing the EM/hadronic injections, we need to track the temporal evolution of both the relic particles ϕ and the injected neutrinos. Usually, this is done by solving the appropriate set of coupled Boltzmann equations. In simple scenarios or under simplifying assumptions, this set of equations can be solved semi-analytically. However, in more general cases, this task may be difficult to achieve even numerically. This is also true for the scenario at hand, which features coupled integro-differential equations. We therefore refrain from using the usual Boltzmann approach. Instead, we employ a *probability-based* ansatz for tracking the evolution of the particles participating in the neutrino cascade. Regarding the handling of thermal scattering processes, we find that a regular Monte-Carlo approach works well, since the target particles feature a predefined thermal distribution (also cf. [17, 55] for a Monte Carlo approach focusing on smaller lifetimes and smaller masses, i.e. $\tau_\phi \lesssim 1\,\mathrm{s}$ and $m_\phi \lesssim 2\,\mathrm{GeV}$). However, when also considering non-thermal scattering reactions, which are not considered in [17, 55], we find that a full Monte-Carlo approach is computationally unfeasible. Therefore, we instead opt for the procedure described in this section, which is still inspired by a Monte-Carlo solution, but instead utilises averages over the distributions of certain quantities in order to reduce the computationally expensive task of drawing each quantity from a full distribution.

In a nutshell, we therefore proceed as follows: we discretise the time evolution over small individual time steps $\mathrm{d}t$, which are calculated based on the most relevant time-scales of the problem, i.e. the Hubble scale and the time-scales for (non-)thermal scattering. In every step, we calculate the probability for the relic to decay as well as the probabilities for (non-)thermal scattering to occur. Based on these probabilities, we then handle the EW shower from FSR as well as the various scattering reactions in order to calculate the amount of injected EM/hadronic material using averaged quantities, while redshifting the abundance

---

[12]For the corresponding source code, see https://github.com/hep-mh/xena.

Variable	Description	Unit
Input parameters		
Y_ϕ	Abundance of ϕ	—
m_ϕ	Mass of ϕ	MeV
τ_ϕ	Lifetime of ϕ	s
BR_{ee}	Branching ratio into e^+e^-	-
One-dimensional output parameters		
f_ϕ	Fractional abundance of ϕ, $\frac{\Omega_\phi}{\Omega_{\mathrm{DM}}}$	—
N_{eff}	Effective number of neutrinos	—
η	Baryon-to-photon ratio	—
Two-dimensional output parameters		
t	Time	s
T	Temperature	MeV
$\frac{\mathrm{d}T}{\mathrm{d}t}$	Time-temperature relation	MeV^2
T_ν	Neutrino temperature	MeV
H	Hubble rate	MeV
n_γ	Photon number density	MeV^3
S_{em}	Electromagnetic source term, eq. (4.11)	MeV^4
S_{hd}	Hadronic source term	MeV^4
$\frac{\mathrm{d}n_{\mathrm{hd}}^{\mathrm{n\text{-}th}}}{\mathrm{d}t}$	Injected hadrons per time from non-th. scattering, eq. (4.14)	MeV^4
$K_{\mathrm{hd}}^{\mathrm{n\text{-}th}}$	Avg. kinetic energy of hadrons from non-th. scattering, eq. (4.14)	MeV
$\frac{\mathrm{d}n_{\mathrm{hd}}^{\mathrm{fsr}}}{\mathrm{d}t}$	Injected hadrons per time interval from FSR, eq. (4.14)	MeV^4
$K_{\mathrm{hd}}^{\mathrm{fsr}}$	Avg. kinetic energy of hadrons from FSR, eq. (4.14)	MeV

Table 2. Description of the input/output parameters used for our numerical treatment. Some of the output quantities are also shown for a specific benchmark point in figure 9.

and energy of all neutrinos that have not scattered within the given time step.[13] This way, we also track the temporal evolution of the relic particles and the non-thermal neutrinos, thus quantifying their influence on the background cosmology.

As shown in table 2, the input for the simulation comprises the mass m_ϕ, the lifetime τ_ϕ, and the abundance $Y_\phi = n_\phi/s$ of the relic, with the latter quantity featuring the total entropy density s, which is approximately equal to the one of the SM s_{sm}. Additionally, we may also parameterise the abundance of ϕ via the parameter $f_\phi \equiv \Omega_\phi^0/\Omega_{\mathrm{DM}}^0$. Here, Ω_ϕ^0 is the fraction of energy relative to the critical energy density that ϕ would have today if it would not decay, and $\Omega_{\mathrm{DM}}^0 = 0.26$. In the case of negligible entropy production, there exists a linear, one-to-one relation between $m_\phi Y_\phi$ and f_ϕ. This is the case for all relevant parts of parameter space, which is why in the following, we often use $m_\phi Y_\phi$ and f_ϕ interchangeably.

---

[13]As a conservative assumption, we neglect rescattered particles. See also the discussion in section 3.4.

Using the above quantities, the number of ϕ particles per unit volume is given by

$$n_\phi \simeq Y_\phi s_{\mathrm{sm}} = \tilde{n}_\phi e^{-t/\tau_\phi} = \frac{\Omega_\phi^0 \rho_{\mathrm{crit}}^0}{(m_\phi R^3)} e^{-t/\tau_\phi} \,. \tag{5.1}$$

Here, $\tilde{n}_\phi$ is the 'would-be' number density of ϕ if it was not decaying, R is the scale factor (we fix the scale factor to be equal to 1 today), and ρ_{crit}^0 is the critical energy density. Additionally, we also consider the case of a non-vanishing branching ratio of ϕ directly into electron-positron pairs, which introduces an additional input parameter BR_{ee}. We discuss the impact of this parameter in more detail in section 6.5. Finally, let us note that for the evaluation of other cosmological quantities, we assume no additional dark-sector contributions beyond ϕ. Regarding the different input parameters, in the following, we present a concise summary of the parameter space covered in this work:

- **m_ϕ.** We restrict our calculation to $m_\phi \in [1\,\mathrm{GeV}, 10\,\mathrm{TeV}]$. The lower boundary is motivated by the fact that the limits weaken significantly with mass and for $m_\phi < 1\,\mathrm{GeV}$ are only determined by the modified cosmology. Specifically, the kinematic threshold for the thermal production of electrons ($ET \gtrsim m_e^2$) is usually not met for sub-GeV relics.

- **τ_ϕ.** We focus on lifetimes $\tau_\phi \in [10^4, 10^{11}]\,\mathrm{s}$. Most notably, the lower limit is motivated by the breakdown of our simplifying assumptions, i.e. neutrinos may also scatter on background electrons or into muons, SBBN and NBBN no longer factorise, etc. Instead, the upper limit is motivated by the fact that for larger lifetimes, other measurements provide more powerful bounds, e.g. those from CMB anisotropies or spectral distortions [8, 9, 15].

- **$f_\phi \,|\, Y_\phi$.** The upper limit of values to consider for f_ϕ depends on τ_ϕ as it is determined by the N_{eff} constraints, which gives the dominant limit for sufficiently large values of f_ϕ, thus making other constraints irrelevant. In this context, we have confirmed that our numerical results, i.e. specifically the resulting light-element abundances, are calculated consistently well above the N_{eff} limit.

- **BR_{ee}.** While we consider the full range $\mathrm{BR}_{ee} \in [0, 1]$, our main focus remains on pure neutrino injections, i.e. $\mathrm{BR}_{ee} = 0$.

5.2 Algorithmic implementation of the neutrino cascade

To aid the reader, in figure 7 we provide a visual representation of the algorithmic implementation discussed on this section. It is instructive to consult this figure while reading through this section.

For the purpose of this section, we describe the abundance of any particle x at a given time t via the parameter $a_x(t) \equiv N_x(t)/N_\phi^{\mathrm{tot}}$, where $N_x(t)$ is the number of x particles at time t inside a comoving volume element, and N_ϕ^{tot} is the total number of ϕ particles prior to any decay processes, again within a comoving volume element. This definition is convenient, since $a_x(t)$ is not affected by Hubble expansion. By choosing some reference time $t_0 \ll \tau_\phi$ — see the
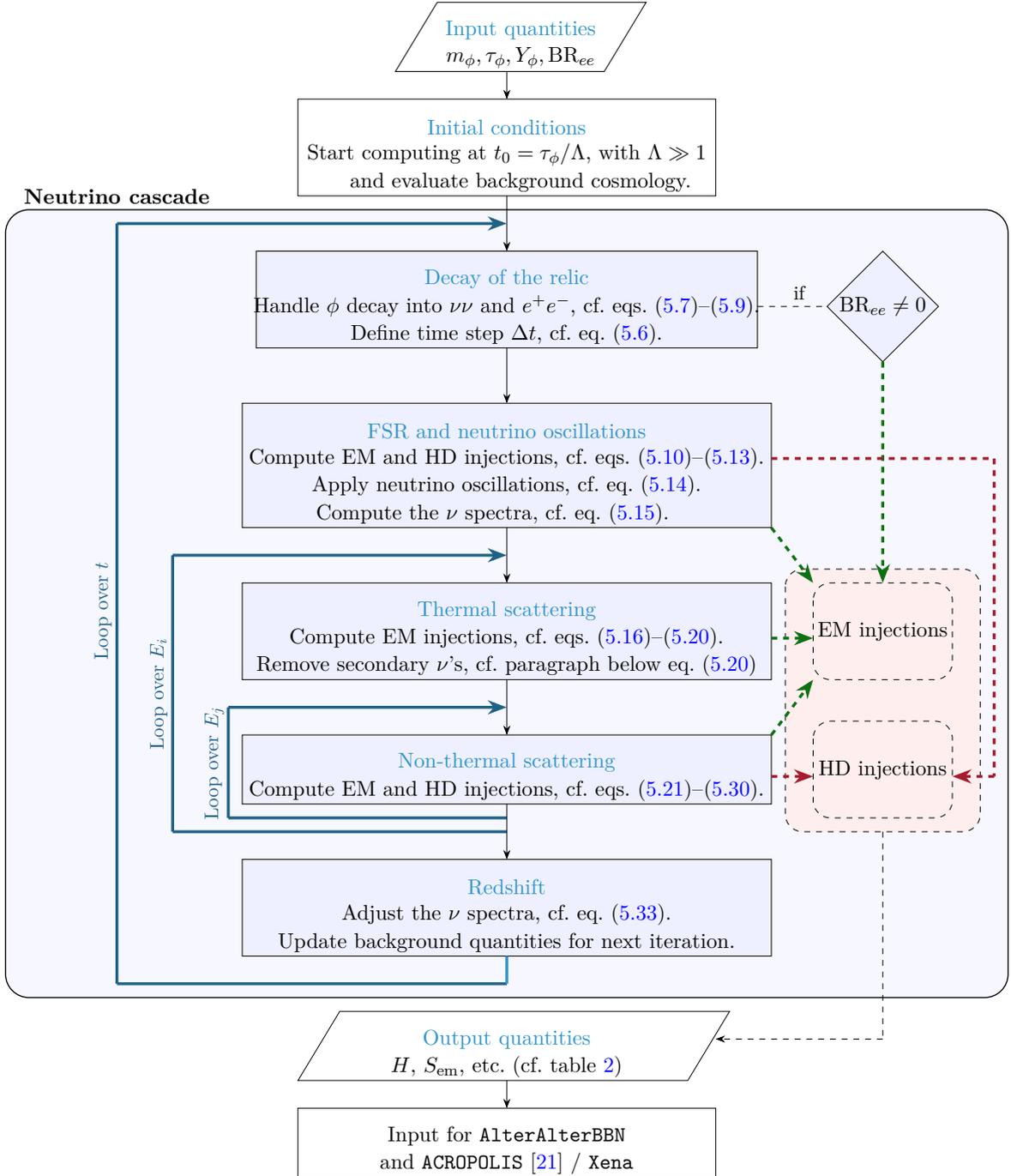
Figure 7. Flowchart of the code developed in the context of this study to determine the injection of EM and hadronic material from the (neutrino) cascade induced by the decay of a relic particle.

initial conditions below — with the scale factor $R_0 \equiv R(t_0)$ and $n_{\phi,0} \equiv n_\phi(t_0)$, the number density $n_x(t)$ of x at any other time $t \neq t_0$ can then be calculated from $a_x(t)$ via the relation

$$n_x(t) = a_x(t)n_{\phi,0}[R_0/R(t)]^3 \,. \tag{5.2}$$

For the neutrinos, we therefore end up with six different quantities, which we conveniently write in vector notation (here and in the following, a bar under a quantity denotes a vector in neutrino space)

$$\underline{a}_\nu(t) \equiv (a_{\nu_e}, a_{\nu_\mu}, a_{\nu_\tau}, a_{\bar{\nu}_e}, a_{\bar{\nu}_\mu}, a_{\bar{\nu}_\tau})^T(t) \,. \tag{5.3}$$

Using eq. (5.2), we therefore find $\underline{n}_\nu(t) = \underline{a}_\nu(t)n_{\phi,0}[R_0/R(t)]^3$. At this point, let us note that we ultimately want to quantify the effects of the different neutrino interactions. Regarding the non-thermal scattering processes specifically (cf. eq. (3.9) with $f_Y = f_\nu$),[14] this requires the knowledge of the neutrino spectra $\underline{f}_\nu(t, E)$ at any given time t. As part of the code, we encode said spectrum in terms of a list $D_\nu(t)$, which — for a given time t — stores discrete pairs of the form $[E : \underline{a}_\nu(t, E)]$. Consequently, $D_\nu(t)$ can usually be represented as[15]

$$D_\nu(t) = \left\{ \left[E_1 : \underline{a}_\nu(t, E_1)\right], \left[E_2 : \underline{a}_\nu(t, E_2)\right], \cdots \right\} \,. \tag{5.4}$$

Such a list encodes a spectrum of particles with discrete energies E_i and associated abundances $\underline{a}_\nu(t, E_i)$, i.e.

$$\underline{f}_\nu(t, E) = 2\pi^2 \sum_i \underbrace{\underline{a}_\nu(t, E_i)n_{\phi,0}[R_0/R(t)]^3}_{\equiv \underline{n}_\nu(t, E_i)} \frac{\delta(E - E_i)}{E_i^2} \,. \tag{5.5}$$

Here, $\underline{n}_\nu(t, E_i)$ are the individual number densities of neutrinos with fixed energy $\sim E_i$ at time t, which, when summed over E_i, yield the total number densities of neutrinos $\underline{n}_\nu(t) = \int \underline{f}_\nu(t, E)\mathrm{d}^3p/(2\pi)^3 = \sum_{E_i} \underline{n}_\nu(t, E_i)$. In order to reduce visual clutter, we drop the time dependence of most quantities in the following.

For visualisation, in the left panel of figure 8 (left) we show the time evolution of the neutrino spectrum for a benchmark point close to the exclusion limit, i.e. $m_\phi = 100\,\mathrm{GeV}$, $\tau_\phi = 10^5\,\mathrm{s}$ and $m_\phi Y_\phi = 10^{-7}\,\mathrm{GeV}$. Here, we focus on the temporal evolution close to the lifetime of ϕ, i.e. at times $t/\tau_\phi \in \{0.1, 0.5, 1, 5, 10\}$, in order to capture the most interesting effects. Specifically, we compare the full numerical evolution ("full", solid line) with the spectrum that is obtained from only ϕ decay and redshifting ("RS only", dashed line). Moreover, we compare the latter result with a corresponding analytical calculation, which has been obtained under the assumption of a vanishing scattering rate (dotted line, cf. appendix C). Overall, we find that the analytical calculation captures the full evolution very well, as long as neutrino interactions can be neglected. However, we also find that the neutrino spectrum is depleted for those particles that have been injected at an early time (left part of the

---

[14]Some interactions such as FSR and neutrino oscillations are independent of the spectrum.

[15]For performance reasons, we assume that two energies E and E' are equal if $|E' - E|/E \ll 1$. In this case, we only add one element with energy $E \sim E'$ and an abundance vector that is the sum of the two individual vectors.
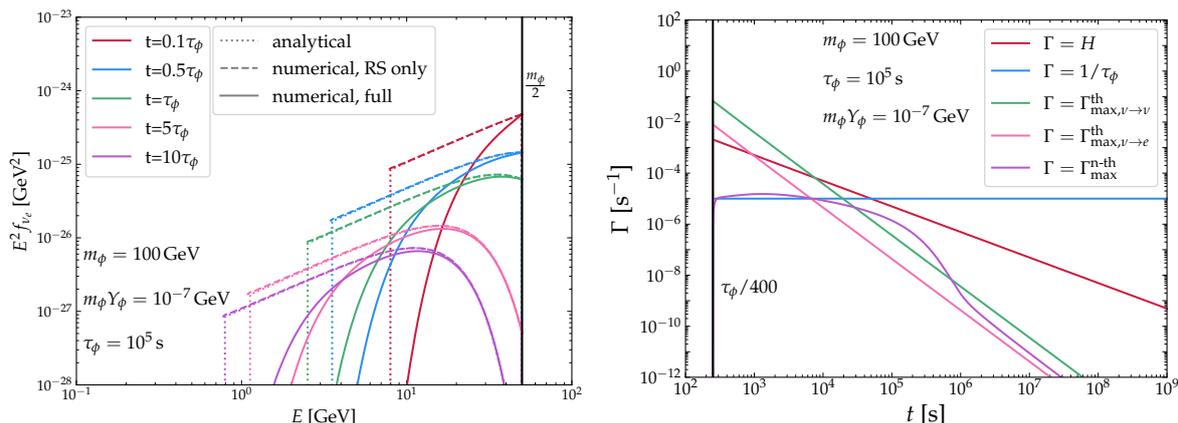
Figure 8. *Left:* phase-space distribution function of the injected neutrinos from $\phi \to \nu_e \bar{\nu}_e$ for different ratios $t/\tau_\phi \in \{0.1, 0.5, 1, 5, 10\}$, taking into account all effects (solid), or only the redshift (dashed). For comparison, the latter case is also compared to the corresponding analytical approximation from appendix C (dotted). Here, we conservatively only show those particles that have never undergone any scattering reactions up to the given time. *Right:* time evolution of a selection of rates. Note that all scattering rates are taken for ν_e and we use their maximum values as introduced in eq. (5.6).

curve). Note that in our approximation, we remove all neutrinos from the spectrum that have previously scattered. However, in reality these neutrinos would not vanish completely, but their energy would be moved to lower energy bins (cf. section 3.3). Numerically, we find that for $t = \tau_\phi/10$, roughly 40% of all injected neutrinos did scatter before. Considering later times, the depletion is less efficient, e.g. for $t = 10\tau_\phi$ almost 90% of all injected neutrinos did not yet scatter and hence are still present in the spectrum. Note that the cutoff in the spectrum towards small energies comes from the fact that we only track the spectrum starting from some initial time t_0, which is discussed in the next section. However, due to the suppression of the spectrum for such energies, this approach is fully justified.

Initial conditions. Ultimately, we want to start our calculation at some time $t_0 = \tau_\phi/\Lambda$ with $\Lambda \gg 1$, i.e. before the heavy relic starts to decay in significant amounts. Given a value of Λ, we therefore have to first determine the corresponding temperatures $T_0 \equiv T(t_0)$ and $T_{\nu,0} \equiv T_\nu(t_0)$. To this end, let us note that $\rho_\phi \propto R^{-3}$ and $\rho_{\rm sm} \propto R^{-4}$, meaning that for large temperatures, say $T \gg T_{\rm eq}$ with $T_{\rm eq}$ being the temperature at which $\rho_{\rm sm} = \rho_\phi$, the energy density of the relic becomes negligible compared to the one of the SM. Consequently, above some temperature $T_i > \max\{m_t, T_{\rm eq}\}$ with the mass m_t of the top quark, the Hubble rate obeys $H(T) \propto \sqrt{\rho_{\rm sm}(T)} \propto T^2$. For the parameters considered in this work, we find that $T_i = 10^4$ GeV is a valid choice. At this temperature, $T = T_\nu$, $H \propto T^2$ and $dT/dt = -HT$, which implies $t_i = t(T_i) = [2H(T_i)]^{-1}$ as well as $T_{\nu,i} = T_\nu(T_i) = T_i$. Finally, to go from t_i to t_0, we implement the following procedure, starting from the temperature $T = T_i$:

(i) Calculate the energy density of the relic.

(ii) Calculate the Hubble rate $H(T) = \sqrt{\rho_\phi(T) + \rho_{\rm sm}(T)}/(\sqrt{3}M_{\rm pl})$.

(iii) Define a small time step $\Delta t = \tilde{\varepsilon}/H$ with $\tilde{\varepsilon} \ll 1$ (we find that $\tilde{\varepsilon} = 10^{-4}$ gives a good compromise between speed and accuracy).

(iv) Calculate the respective time-temperature relations $\mathrm{d}T/\mathrm{d}t$ and $\mathrm{d}T_\nu/\mathrm{d}t$ with the previously determined Hubble rate and use them to update the time $t \to t + \Delta t$, the temperature $T \to T + (\mathrm{d}T/\mathrm{d}t)\Delta t$, as well as the neutrino temperature $T_\nu \to T_\nu + (\mathrm{d}T_\nu/\mathrm{d}t)\Delta t$.

(v) Repeat until $t \geq t_0$.

By following this procedure, we obtain a consistent set of initial conditions $(t_0, T_0, T_{\nu,0})$ with $t_0 \simeq \tau_\phi/\Lambda$. In this work, we specifically set $\Lambda = 400$, noting that larger values do not significantly change our results but are computationally much more expensive. The value of t_0 is then used as the reference value for $n_{\phi,0}$ and R_0 from the previous paragraph, which implies $a_\phi(t_0) \simeq 1$ and $D_\nu(t_0) = \varnothing$. Setting $a_\phi(t_0) \simeq 1$ again insinuates that we neglect all potential decays until this time, which we find to be a good approximation, since the fraction of particles that decay before, e.g. $|1 - e^{1/\Lambda}| \sim 0.3\%$ for $\Lambda = 400$, is sufficiently small.

Interlude. As part of the following steps, we calculate the relevant parameters describing the EM and hadronic injections, i.e. S_em in eq. (4.9), $K_\mathrm{hd}^\mathrm{fsr}$ in eq. (4.14), etc. To provide the reader with a graphical representation of these parameters, in figure 9, we show the resulting quantities (top, bottom left) for a single benchmark point with $m_\phi = 100\,\mathrm{GeV}$, $\tau_\phi = 10^5\,\mathrm{s}$ and $m_\phi Y_\phi = 10^{-7}\,\mathrm{GeV}$. Note that the initial increase of the different quantities close to the initial condition at $t = t_0 = \tau_\phi/400$ (black line) is not physical, but instead stems from the fact that we do not calculate these quantities for earlier times, i.e. $S_\mathrm{em}(t < t_0) \equiv 0$, etc. Therefore, each of them first needs to initialise to their correct physical value for $t > t_0$. However, we checked that this artefact has no influence on our results, since the light-element abundances during NBBN only start to significantly change at $t \sim \tau_\phi$ when $1/(2t) \simeq H \simeq \Gamma_\phi = 1/\tau_\phi$, at which point all quantities have already assumed their actual values. In fact, for $t < \tau_\phi$, the terms on the right-hand sides of eqs. (4.13) and (4.14) are suppressed by a factor $\Gamma_\phi/H \sim t/\tau_\phi$. In the same figure, we also show several quantities describing the modified background cosmology (bottom right). To complement these plots, in the right panel of figure 8, we further show a compilation of the relevant rates for the same benchmark point.

Decay of the relic. Starting from the initial conditions determined in the last paragraph, we now track the decay of the relic ϕ, as well as the subsequent neutrino cascade. Like before, we do this by following the evolution of the individual particles over small time steps Δt. Consequently, each step of the iteration can be assigned a unique set of values T, T_ν, t, a_ϕ and $D_\nu = D_\nu(t)$, while we track the neutrinos that will be present in the next step, i.e. at $t + \Delta t$, in a separate list $D_\nu(t + \Delta t)$ (with $D_\nu(t + \Delta t) = \varnothing$ at the start of each step). To determine Δt for a given temperature T, however, it is no longer sufficient to only consider the Hubble rate $H(T)$, since potentially relevant scattering reactions add another timescale to the problem. Therefore, we also calculate the maximal thermal scattering rate $\Gamma_\mathrm{th}^\mathrm{max}(T)$ and the maximal non-thermal scattering rate $\Gamma_\mathrm{n\text{-}th}^\mathrm{max}(T)$,[16] and define

$$\Delta t = \frac{\varepsilon}{\max\left(\Gamma_\mathrm{th}^\mathrm{max} + \Gamma_\mathrm{n\text{-}th}^\mathrm{max}, H\right)} \tag{5.6}$$

---

[16]More precisely, referencing eqs. (5.17) and (5.22) below, we use $\Gamma_\mathrm{th}^\mathrm{max} = \max \underline{\Gamma}_{\nu\to\nu}^\mathrm{th}$ and $\Gamma_\mathrm{n\text{-}th}^\mathrm{max} = \max \underline{\Gamma}_{\nu\to X}^\mathrm{n\text{-}th}$.

with $\varepsilon \ll 1$. In this work, we find that $\varepsilon = 10^{-2}$ gives a good compromise between speed and accuracy. While errors on the individual quantities can be $\sim 1\%$–10% for this choice — specifically for small lifetimes $\tau_\phi \sim 10^4\,\mathrm{s}$ when many scatterings happen per Hubble time —, such differences are barely noticeable in the final plots, which span many orders of magnitude. Given a relic abundance a_ϕ, the number Δa_ϕ of ϕ particles that decay during the time step Δt is then given by

$$\Delta a_\phi = a_\phi \left| 1 - e^{-\Delta t/\tau_\phi} \right| . \tag{5.7}$$

These decays lead to the two-body injection of electrons/positrons and neutrinos with energy $E_0 = m_\phi/2$ according to the branching ratios BR_{ee} and $\mathrm{BR}_{\nu\nu} \equiv 1 - \mathrm{BR}_{ee}$, respectively. The injected electron-positron pairs directly contribute to the amount of injected electromagnetic material and therefore change $S_{\mathrm{em}}(t)$ in eq. (4.12) by (cf. eq. (4.11) with $E_0 = m_\phi/2$ and $\dot{\rho}_{e\pm}^{\mathrm{inj}} \simeq -\mathrm{BR}_{ee}\dot{\rho}_\phi = \mathrm{BR}_{ee}m_\phi|\dot{n}_\phi|$)

$$S_{\mathrm{em}}(t) \supset \mathrm{BR}_{ee} \times 2 \times \frac{\Delta n_\phi}{\Delta t} = S_{\mathrm{em}}^{\mathrm{dcy}} \tag{5.8}$$

with $\Delta n_\phi = \Delta a_\phi n_{\phi,0} R^3/R_0^3 > 0$. The injected neutrinos, however, will undergo additional interactions (see below). As discussed above, to handle some of these reactions, we need to know the different neutrino spectra, meaning that all injected neutrinos must be tracked as part of $D_\nu(t + \Delta t)$. At face value, and assuming for simplicity that the relic only decays into (anti-)electron neutrinos, this would imply that for the next step we have to add

$$D_\nu(t + \Delta t) \supset \left[m_\phi/2 : \Delta a_\nu \right] \qquad \text{with} \qquad \Delta \underline{a}_\nu \equiv \mathrm{BR}_{\nu\nu} \times (\Delta a_\phi, 0, 0, \Delta a_\phi, 0, 0)^T . \tag{5.9}$$

However, since both neutrino oscillations and the EW shower due to FSR can change the composition and energy of the injected neutrinos before any scattering reactions become important, it is instructive to first handle these effects, before adding any decay products to the spectrum.

Final-state radiation. Immediately after getting injected, i.e. on time-scales much smaller than those of the proceeding scattering reactions and even oscillations, the neutrinos lose some of their energy via FSR, which further leads to the injection of both EM and hadronic material. In this work, we quantify this process via the *average* fractions $\zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi)$ and $\zeta_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$ of the original neutrino energy that ends up in the form of EM and hadronic material, respectively. Using these quantities, the *average* energy of each injected neutrino after the shower is then given by

$$E_0^{\mathrm{post\text{-}fsr}}(m_\phi) = \frac{m_\phi}{2} \left[1 - \zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi) - \zeta_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi) \right] . \tag{5.10}$$

The dissipated energy, in turn, gets distributed between $S_{\mathrm{em}}(t)$ in eq. (4.12) and $S_{\mathrm{hd}}(t)$ in eq. (4.17), i.e.

$$S_{\mathrm{em}}(t) \supset \mathrm{BR}_{\nu\nu} \times \zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi) \times \frac{2\Delta n_\phi}{\Delta t} = S_{\mathrm{em}}^{\mathrm{fsr}} \tag{5.11}$$

$$S_{\mathrm{hd}}(t) \supset \mathrm{BR}_{\nu\nu} \times \zeta_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi) \times \frac{2\Delta n_\phi}{\Delta t} = S_{\mathrm{hd}}^{\mathrm{fsr}} . \tag{5.12}$$

For correctly handling the hadronic injections, we further need to know the amount of hadrons that this process injects per time interval, i.e. $\mathrm{d}n_{\mathrm{hd}}^{\mathrm{fsr}}/\mathrm{d}t$ in eq. (4.14), as well as the corresponding distribution of kinetic energies (cf. section 4.3). As stated above, we quantify this contribution via the simplifying assumption that all hadrons are injected with the *average* kinetic energy $K_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$ from eq. (4.14), while setting

$$\frac{\mathrm{d}n_{\mathrm{hd}}^{\mathrm{fsr}}}{\mathrm{d}t}(t) = \mathrm{BR}_{\nu\nu} \times N_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi) \times \frac{\Delta n_\phi}{\Delta t}\,, \tag{5.13}$$

where $N_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$ is the *average* number of protons/neutrons that get injected by FSR following a single decay. Conceptually, the four relevant functions $\zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi)$, $\zeta_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$, $N_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$, and $K_{\mathrm{hd}}^{\mathrm{fsr}}(m_\phi)$ are obtained by sampling $\mathcal{O}(10^6)$ $\nu_e\bar{\nu}_e$ events, which are then used as an input for the simulation of the EW shower via the `VINCIA` module of `PYTHIA8.3` [24, 25]. All quantities are then determined as the respective averages over all simulated events.[17] This procedure is analogous to the one described in section 2.1 of [15], with the only difference being that we additionally extract information on the hadronic fragments of the shower (for more information, please consult the aforementioned paper).[18] For reference, the three latter quantities are presented in figure 19 of appendix F, while $\zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi)$ has already been shown in figure 4. Finally, let us note that the EW shower can also produce secondary neutrinos, which in principle need to be added to $D_\nu(t+\Delta t)$. However, we find that their contribution is usually much smaller than the one of the primary neutrinos, meaning that they do not lead to a significant injection of EM or hadronic energy. Consequently, we neglect them in the following.

Neutrino oscillations. After the EW shower from FSR, but still before any scattering reactions, the injected neutrinos further undergo oscillations. Extending the neutrino-oscillation matrix $\mathbf{M}_{\mathrm{osc}}$ from eq. (3.2) to six dimensions ($\mathbf{M}_{\mathrm{osc}}$ acts in the same way on neutrinos and anti-neutrinos), this process changes the flavour-composition of the original injection with associated abundance $\Delta\underline{a}_\nu$ according to the relation

$$\Delta\underline{a}_\nu^{\mathrm{osc}} = \mathbf{M}_{\mathrm{osc}}\Delta\underline{a}_\nu\,. \tag{5.14}$$

Consequently, to incorporate the effects of the FSR and neutrino oscillations, we replace eq. (5.9) by

$$D_\nu(t + \Delta t) \supset \left[E_0^{\mathrm{post\text{-}fsr}}(m_\phi) : \Delta\underline{a}_\nu^{\mathrm{osc}}\right]. \tag{5.15}$$

Thermal scattering. Given a neutrino spectrum $D_\nu(t)$, any neutrino with energy $E_i \in D_\nu(t)$ can participate in thermal scattering reactions with the background neutrinos. At this point, let us stress again that for the values of m_ϕ and τ_ϕ considered in this work with $E_i \leq m_\phi/2$, the maximally available centre-of-mass energy $s_{\mathrm{max}}^{\mathrm{th}} \sim m_\phi T_\nu(\tau_\phi)$ from thermal scattering off a background neutrino with temperature T_ν, is always well below the Z-resonance and

---

[17]As already indicated above, for both protons and neutrons, the quantities $N_{\mathrm{hd}}^{\mathrm{fsr}}$ and $K_{\mathrm{hd}}^{\mathrm{fsr}}$ are equal at the $\mathcal{O}(0.1\%)$ level, which is why it is justified to use identical values for both nucleons.

[18]In practice, we also need to fix the nature of the relic to run the shower. Following the cited paper, we have decided on a vector particle. For a scalar particle, we have checked that both EM and hadronic injection vary on the level of $\mathcal{O}(1)$ which leads to mildly different bounds.

even the $e\mu$ threshold. Consequently, given a neutrino ν_j with energy E_i that scatters off any thermal background (anti-)neutrino $\nu_{\rm th}$, the only available channels are $\nu_j\nu_{\rm th} \to \nu\nu$ and $\nu_j\nu_{\rm th} \to e^+e^-$.[19] Using $\Gamma_0^{\rm th}(E_i, T_\nu) = (7\pi/90)G_F^2 E_i T_\nu^4$ (cf. eq. (3.11) without the process-dependent factor Σ_∞), the corresponding rates are given by

$$\Gamma_{\nu_j\to\nu}^{\rm th}(E_i, T_\nu) \equiv {}^{10}\!/\!_3\Gamma_0^{\rm th}(E_i, T_\nu) \qquad \text{and} \qquad \Gamma_{\nu_j\to e}^{\rm th}(E_i, T_\nu) = F_j(E_i)\Gamma_0^{\rm th}(E_i, T_\nu). \quad (5.16)$$

Here, $F_j(E_i)$ is the analogue of Σ_∞ in eq. (3.11), but it has been obtained by performing the integral $\int \mathrm{d}s\, s \cdot \sigma_{ee}(s)$ for a *non-vanishing* electron-mass. Consequently, $F_j \neq$ const. and $F_j(E_i \ll m_e) = 0$. Also, the index j indicates that F_j differs depending on the initial-state flavour (cf. table 1). In vector notation (cf. eq. (5.3)), we can therefore write

$$\underline{\Gamma}_{\nu\to\nu}^{\rm th} \equiv {}^{10}\!/\!_3\Gamma_0^{\rm th}(1, 1, 1, 1, 1, 1)^T \qquad \text{and} \qquad \underline{\Gamma}_{\nu\to e}^{\rm th} \equiv \Gamma_0^{\rm th}(F_e, F_\mu, F_\tau, F_e, F_\mu, F_\tau)^T. \quad (5.17)$$

Given a neutrino with energy E_i, the probability for a thermal scattering reaction to happen within a time interval Δt therefore is

$$\Delta\underline{P}_{\nu\to\nu/e}^{\rm th}(E_i) = \underline{\Gamma}_{\nu\to\nu/e}^{\rm th}(E_i)\Delta t. \quad (5.18)$$

Consequently, the total number $\Delta n_{\nu\to\nu/e}^{\rm th}$ of neutrinos with energy E_i that scatter into other neutrinos / electron-positron pairs is given by

$$\Delta n_{\nu\to\nu/e}^{\rm th}(E_i) = \Delta\underline{P}_{\nu\to\nu/e}^{\rm th}(E_i) \cdot \underline{n}_\nu(E_i). \quad (5.19)$$

On the one hand, any neutrino that scatters into e^+e^- transfers all of its energy E_i into the electromagnetic sector, which implies that the corresponding S-factor $S_{\rm em}(t)$ in eq. (4.12) receives a contribution of the form

$$S_{\rm em}(t) \supset \frac{E_i}{E_0}\frac{\Delta n_{\nu\to e}^{\rm th}(E_i)}{\Delta t} = \frac{E_i}{E_0}\left[\underline{\Gamma}_{\nu\to e}^{\rm th}(E_i) \cdot \underline{n}_\nu(E_i)\right] \equiv \Delta S_{\rm em}^{\rm th}(E_i) \qquad \forall E_i \in D_\nu(t). \quad (5.20)$$

Notably, such a contribution exists for each $E_i \in D_\nu(t)$, meaning that the total change of $S_{\rm em}(t)$ is equal to the sum of each individual contribution, i.e. $S_{\rm em}^{\rm th} = \sum_{E_i} \Delta S_{\rm em}^{\rm th}(E_i)$. On the other hand, any neutrino with energy E_i that scatters into $\nu\nu$ effectively leads to the conversion into two, i.e. one additional, non-thermal neutrinos with average energy $E_i/2$ and abundance $\Delta\underline{P}_{\nu\to\nu}^{\rm th}(E_i) \odot \underline{a}_\nu(E_i)$,[20] which — in the next time step — can themselves interact with the thermal neutrino background. Therefore, a full treatment would imply that the final-state particles be added to $D_\nu(t + \Delta t)$. However, we find that the contribution of these secondary neutrinos is negligible for most lifetimes considered in this work. This is because their abundance $\propto \Delta\underline{P}_{\nu\to\nu}^{\rm th} \ll 1$ is sufficiently small compared to the one of neutrinos that do not scatter at all and instead undergo redshift (cf. eq. (5.33) below). In the following, we therefore neglect such neutrinos by not adding them to $D_\nu(t + \Delta t)$.[21]

---

[19]Thermal reactions therefore do not inject hadrons, meaning that $S_{\rm hd}$ and $\mathrm{d}n_{\rm hd}/\mathrm{d}t$ remain unaltered. However, if we were to consider $\tau_\phi < 10^4\,$s, additional channels like $\nu\nu_{\rm th} \to e\mu$ and $\nu\nu_{\rm th} \to \pi\pi$ would eventually open up, which would lead to the injection of hadronic material and thus to a change of $S_{\rm hd}$ and $\mathrm{d}n_{\rm hd}/\mathrm{d}t$.

[20]Here, the operator $\odot$ denotes the Hadamard product, i.e. the different entries of the two vectors are multiplied element-wise to form a new vector of the same dimension.

[21]Technically, this is only true for large lifetimes, when the number of scattering reactions per Hubble time is always $\ll 1$. However, this assumption breaks down for $\tau_\phi \lesssim 10^5\,$s, and we make small errors for $\tau_\phi \to 10^4\,$s. Nevertheless, even in this case, we are conservative, since additional particles would only lead to an increase of injected material. We postpone a rigorous treatment of this additional regime to future work.
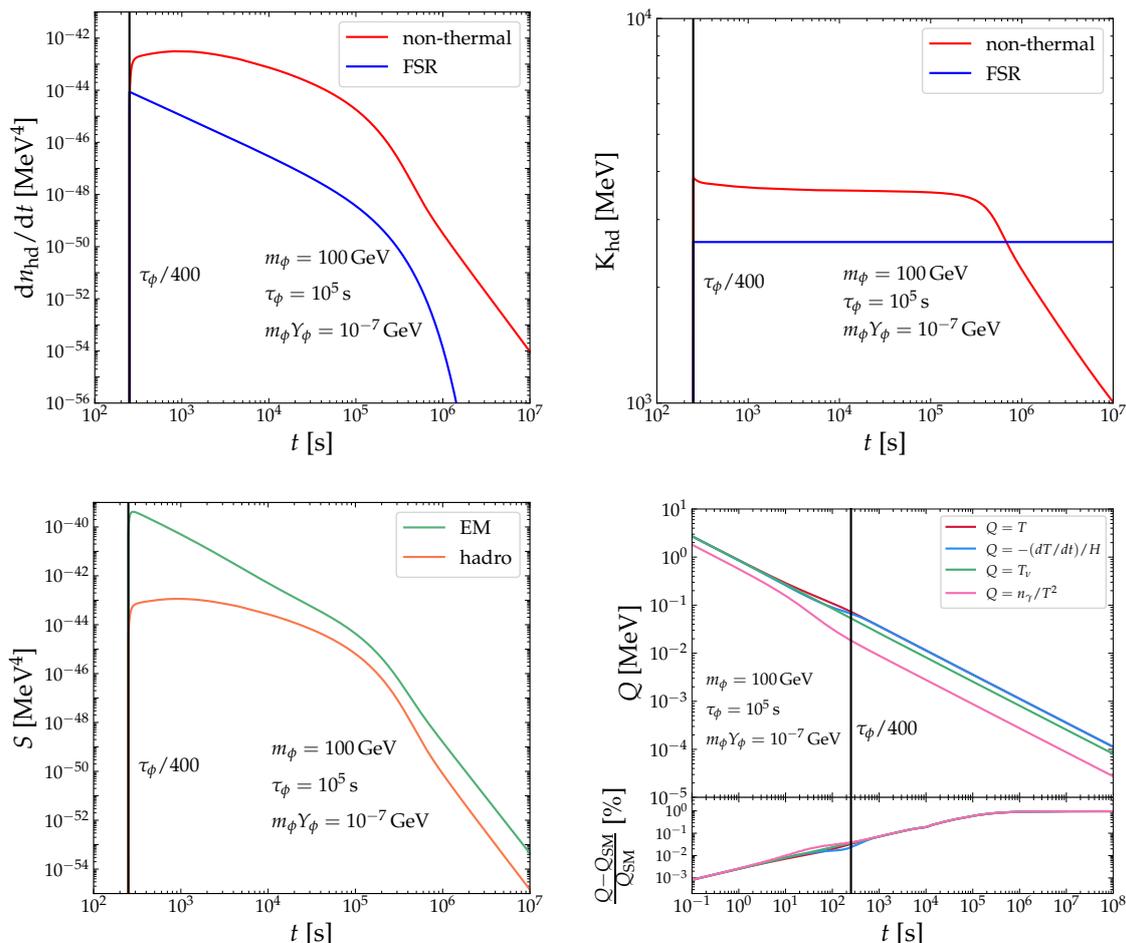
Figure 9. Time dependence of the main quantities resulting from the neutrino cascade. This includes *(i)* the injection rate of hadrons $\frac{\mathrm{d}n_{\mathrm{hd}}}{\mathrm{d}t}$ (upper left) and the average kinetic energy of these particles K_{hd} (upper right) from non-thermal scattering (red) and FSR (blue), and *(ii)* the EM (green) and hadronic (orange) source terms $S_{\mathrm{em/hd}}$ (lower left). Additionally, we show several quantities Q describing the modified background cosmology (lower right), specifically by comparing them to their values Q_{SM} in the SM.

Non-thermal scattering. In addition to the thermal background neutrinos, each neutrino with energy $E_i \in D_\nu(t)$ can also scatter non-thermally with any other neutrinos with energy $E_j \in D_\nu(t)$. As already stressed above, while reactions of this form are suppressed by the abundance of non-thermal neutrinos, they can still be relevant if not dominant, due to the fact that the maximal centre-of-mass energy $s_{\max}^{\text{n-th}} = 4E_i E_j$ can be much larger than $s_{\max}^{\text{th}}$. In particular, the scattering can be resonantly enhanced. To calculate this effect, let us consider the scattering of two neutrinos $p, q \in \{\nu_e, \nu_\mu, \nu_\tau, \bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau\}$ with energies E_i and E_j, respectively. By inserting the spectrum from eq. (5.5) for q into eq. (3.9), we find that — in our framework — the scattering rate for any combinations p, q is given by

$$\Gamma_{pq}^{\text{n-th}}(E_i, E_j) = \frac{n_q(E_j)}{8E_i^2 E_j^2} \underbrace{\int_0^{4E_i E_j} \mathrm{d}s \, s \cdot \sigma_{pq}(s)}_{\equiv \gamma_{pq}^{\text{n-th}}(E_i, E_j)}, \tag{5.21}$$

– 36 –

which can be interpreted as a 6x6 matrix in neutrino space. Consequently, as long as $4E_i E_j > m_Z^2$ and the scattering proceeds via an *s*-channel diagram, the corresponding rate is resonantly enhanced. As a result, Fermi theory is no longer applicable to determine the cross-section, which is why we use `MadGraph5` [34] instead for this task (cf. figure 5). Using eq. (5.21), we can calculate the rate $\underline{\Gamma}^{\text{n-th}}_{\nu \to X}(E_i)$ for a given neutrino with energy E_i to scatter off any other available non-thermal neutrino. In vector notation, we find

$$\left[\underline{\Gamma}^{\text{n-th}}_{\nu \to X}(E_i) \right]_p = \sum_q \sum_{E_j} \Gamma^{\text{n-th}}_{pq}(E_i, E_j) \,. \tag{5.22}$$

Consequently, the scattering probability (analogues to eq. (5.18)) is given by

$$\Delta \underline{P}^{\text{n-th}}_{\nu \to X}(E_i) = \underline{\Gamma}^{\text{n-th}}_{\nu \to X}(E_i) \Delta t \,. \tag{5.23}$$

To employ a treatment similar to the one used for the EW shower from FSR, we further need to calculate the *average* energy fractions $\zeta^{\text{n-th}}_{\text{em}}(s)$ and $\zeta^{\text{n-th}}_{\text{hd}}(s)$ of injected EM and hadronic material, the *average* number $N^{\text{n-th}}_{\text{hd}}(s)$ of injected protons/neutrons, as well as the *average* kinetic energy $K^{\text{n-th}}_{\text{hd}}(s)$ of the injected nucleons, all of which now explicitly depend on the centre-of-mass energy s for which the scattering happens. Similar to the procedure described before, we determine these quantities by using `MadGraph5` to sample $\mathcal{O}(10^6)$ events of the form $pq \to X$ with any two-particle final-state X for a range of different centre-of-mass energies s. Afterwards, we use the resulting events as an input for `PYTHIA8.3` in order to further simulate the subsequent particle shower.[22] The different quantities $\zeta^{\text{n-th}}_{\text{em}}(s)$ etc., are then again determined as the respective averages over all simulated events for a given value of s.[23] Notably, the *average* kinetic energy $\tilde{K}^{\text{n-th}}_{\text{hd}}(s)$ obtained this way, is given in the centre-of-mass frame. However, we are interested in the value of this quantity in the co-moving reference frame, which we can obtain by means of the relation (cf. appendix E)

$$K^{\text{n-th}}_{\text{hd}}(s) = \frac{\tilde{K}^{\text{n-th}}_{\text{hd}}(s) + m_y}{\sqrt{s}} (E_i + E_j) - m_y \equiv y^{\text{n-th}}_{\text{hd}}(s)(E_i + E_j) - m_y \tag{5.24}$$

with the mass m_y of the nucleon $y \in \{p, n\}$.

At this point, it is important to note that two neutrinos with energy E_i and E_j can scatter with any centre-of-mass energy $0 < s < 4E_i E_j$, depending on the angle under which the collision happens. To account for this effect, we further *average* all quantities over all possible values of s. To this end, we define the operator[24]

$$\langle Q \rangle_{pq}(E_i, E_j) = \frac{1}{\gamma^{\text{n-th}}_{pq}(E_i, E_j)} \int_0^{4E_i E_j} \mathrm{d}s \, s \cdot Q_{pq}(s) \cdot \sigma_{pq}(s) \,, \tag{5.25}$$

which determines the *average* value of any quantity $Q(s)$, weighted by the probability for the scattering reaction to happens with any possible centre-of-mass energy $0 < s < 4E_i E_j$.

---

[22]We assume that all unstable particles except neutrons decay during the shower, i.e. we neglect any potential energy loss of these particles due to scatterings with the background plasma prior to their decay. We checked that this does not lead to a sizeable error.

[23]Since we simulate events with all possible two-body final states, this average is also an average over all possible values of X.

[24]Each quantity is therefore averaged twice: once over their distribution originating from the shower for a given value of s and once over all possible values of s for which the particles can scatter in the first place.

The resulting function $\langle \zeta_{\text{em}}^{\text{n-th}} \rangle$ has already been presented in figure 6. We refer the reader to figure 19 of appendix F for a visualisation of the remaining quantities. Using the definition above, we approximate the contributions to the individual S-factors from non-thermal scattering via the relations[25] (cf. eq. (5.20))

$$S_{\text{em}}(t) \supset \frac{E_i + E_j}{E_0} \sum_{p,q} \langle \zeta_{\text{em}}^{\text{n-th}} \rangle_{pq}(E_i, E_j) \Gamma_{pq}^{\text{n-th}}(E_i, E_j) n_p(E_i) \equiv \Delta S_{\text{em}}^{\text{n-th}}(E_i, E_j) \qquad (5.26)$$

$$\forall E_i, E_j \in D_\nu(t) \text{ with } E_j \leq E_i$$

and

$$S_{\text{hd}}(t) \supset \frac{E_i + E_j}{E_0} \sum_{p,q} \langle \zeta_{\text{hd}}^{\text{n-th}} \rangle_{pq}(E_i, E_j) \Gamma_{pq}^{\text{n-th}}(E_i, E_j) n_p(E_i) \equiv \Delta S_{\text{hd}}^{\text{n-th}}(E_i, E_j) \qquad (5.27)$$

$$\forall E_i, E_j \in D_\nu(t) \text{ with } E_j \leq E_i \,.$$

Here, we sum over all possible initial states pq, but only consider energy combinations with $E_j \leq E_i$ in order to avoid double-counting. Overall, the total change of $S_{\text{em/hd}}(t)$ is equal to the sum of each individual contribution, i.e. $S_{\text{em/hd}}^{\text{n-th}} = \sum_{E_i, E_j \leq E_i} \Delta S_{\text{em/hd}}^{\text{n-th}}(E_i, E_j)$. In a similar way, the different contributions to $\mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t$ can be approximated as[26]

$$\frac{\mathrm{d}n_{\text{hd}}^{\text{n-th}}}{\mathrm{d}t}(t) \supset \sum_{p,q} \langle N_{\text{hd}}^{\text{n-th}} \rangle_{pq}(E_i, E_j) \Gamma_{pq}^{\text{n-th}}(E_i, E_j) n_p(E_i) \equiv \Delta \frac{\mathrm{d}n_{\text{hd}}^{\text{n-th}}}{\mathrm{d}t}(E_i, E_j)$$

$$\forall E_i, E_j \in D_\nu(t) \text{ with } E_j \leq E_i \,. \qquad (5.28)$$

At this point, it is important to note that each contribution $\Delta \mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t(E_i, E_j)$ fundamentally has a slightly different *average* kinetic energy according to eq. (5.24), i.e.

$$K_{\text{hd},pq}^{\text{n-th}}(E_i, E_j) = \langle y_{\text{hd}}^{\text{n-th}} \rangle_{pq}(E_i, E_j)(E_i + E_j) - m_p \,. \qquad (5.29)$$

However, instead of handling each contribution separately, we combine them all into one contribution $\mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t = \sum_{E_i, E_j \leq E_i} \Delta \mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t(E_i, E_j)$ with the corresponding *averaged* kinetic energy

$$K_{\text{hd}}^{\text{n-th}} = \frac{\sum_{E_i, E_j \leq E_i} \sum_{p,q} K_{\text{hd},pq}^{\text{n-th}}(E_i, E_j) \Delta \mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t(E_i, E_j)}{\mathrm{d}n_{\text{hd}}^{\text{n-th}}/\mathrm{d}t} \,. \qquad (5.30)$$

Finally, let us note that non-thermal scattering reactions can also inject secondary neutrinos, which may undergo additional scattering reactions in the next time step. However, as before, we find that their contribution is not relevant, which is why we do not add them to $D_\nu(t+\Delta t)$. We can observe this in particular from figure 8 (right panel): we see that even for this benchmark point which is already robustly excluded (cf. figure 3) the *maximum* non-thermal scattering rate is below the Hubble rate. Thus, any neutrino injection from a non-thermal

---

[25]For each reaction, a fraction $\langle \zeta_{\text{em/hd}}^{\text{n-th}} \rangle_{pq}(E_i, E_j)$ of the total energy $E_i + E_j$ is transformed into EM/hadronic material, while the number of reactions happening per time and comoving volume element is given by $\Gamma_{pq}^{\text{n-th}}(E_i, E_j) n_p(E_i)$.

[26]We handle this quantity separately from $\mathrm{d}n_{\text{hd}}^{\text{fsr}}/\mathrm{d}t$, since both have different average kinetic energies.

scattering process is very unlikely to undergo further interactions. We therefore estimate the uncertainty of neglecting secondary neutrinos from non-thermal interactions to be small relative to the large effect of EM and hadronic injections in these collisions. Should the mass of the relic be so high that secondary neutrinos would experience significant FSR, we are in a regime where the initial EW shower is dominant anyways so that the impact of non-thermal scattering is irrelevant.

Redshift. Any particle that does not scatter within the time interval Δt, will simply redshift. The probability for this to happen is given by

$$\Delta \underline{P}^{\mathrm{rs}}_{\nu \to \nu}(E_i) = 1 - \Delta \underline{P}^{\mathrm{th}}_{\nu \to e}(E_i) - \Delta \underline{P}^{\mathrm{th}}_{\nu \to \nu}(E_i) - \Delta \underline{P}^{\mathrm{n\text{-}th}}_{\nu \to X}(E_i) \,. \tag{5.31}$$

To incorporate this process, each particle with initial energy E_i is assigned a new energy

$$E_i^{\mathrm{rs}}(E_i) = E_i \times \frac{R(t)}{R(t + \Delta t)} \,. \tag{5.32}$$

Finally, we add each new energy with the corresponding abundance $\underline{P}^{\mathrm{rs}}_{\nu \to \nu}(E_i) \odot \underline{a}_\nu(E_i)$ to the spectrum $D_\nu(t + \Delta t)$,

$$D_\nu(t + \Delta t) \supset \left[E_i^{\mathrm{rs}}(E_i) : \Delta \underline{P}^{\mathrm{rs}}_{\nu \to \nu}(E_i) \odot \underline{a}_\nu(E_i) \right] \qquad \forall E_i \in D_\nu(t) \,. \tag{5.33}$$

Consequently, all of these particles will be processed again in the next time step, albeit with a slightly smaller energy. As a reminder, let us note that we do not take into account the energy redistribution of neutrino-to-neutrino scatterings, and consequently the redshift is the only mechanism to populate different energies in the spectrum.

The next iteration step. After handling all processes, we update all relevant quantities in order to prepare the next iteration step. For this, we set $t \to t + \Delta t$ with

- $T(t + \Delta t) \simeq T(t) + \frac{\mathrm{d}T}{\mathrm{d}t}(t)\Delta t$

- $T_\nu(t + \Delta t) \simeq T_\nu(t) + \frac{\mathrm{d}T_\nu}{\mathrm{d}t}(t)\Delta t$

- $a_\phi(t + \Delta t) \simeq \max\{a_\phi(t) - \Delta a_\phi(t), 0\}$

In the last step, we made sure that $a_\phi < 0$ never happens numerically. Moreover, we replace $D_\nu(t) \to D_\nu(t + \Delta t)$ and update the energy density of neutrinos by means of the following relation

$$\rho_\nu(t + \Delta t) \simeq \underbrace{\rho_\nu(t) \left[\frac{R(t)}{R(t + \Delta t)} \right]^4}_{\text{Redshift}} \underbrace{+ m_\phi \Delta a_\phi(t)}_{\phi \text{ decay}} \underbrace{- E_0 S_{\mathrm{em}}(t)\Delta t}_{\text{EM inj.}} \underbrace{- E_0 S_{\mathrm{hd}}(t)\Delta t}_{\text{hadronic inj.}} \,. \tag{5.34}$$

Afterwards we repeat the procedure with the new set of quantities t, T, T_ν, a_ϕ, ρ_ν, and D_ν. Finally, we stop the iteration once $T < T_f \equiv 10^{-2}T(\tau_\phi)$, i.e. long after the decay of the relic particle has concluded.

Post-processing. At the end of the iterative procedure, we still have to calculate some derived quantities. Assuming that we stop at a temperature T_f, we first calculate the effective number of additional neutrino species via the equation

$$\Delta N_{\text{eff}} = \frac{\rho_\nu(T_f)}{\rho_{\text{ref}}(T_f)} \tag{5.35}$$

with $\rho_{\text{ref}}(T_f) = 2 \cdot {}^{7}/_{8} \cdot {}^{\pi^2}/_{30} \cdot \left({}^{4}/_{11}\right)^{4/3} \cdot T_f^4$. Afterwards, we use this result to determine the best-fit value for the baryon-to-photon ratio η according to eq. (4.5).

Finally, let us note that we technically also calculate the parameter f_ϕ only at the end of the cascade, since there might not be a one-to-one relation between f_ϕ and $m_\phi Y_\phi$ due to significant entropy injection. In this case

$$f_\phi = m_\phi Y_\phi s_{\text{sm}}(T_{\text{cmb}})/\rho_{\text{dm}}(T_{\text{cmb}}), \tag{5.36}$$

with the SM entropy density s_{sm} at the CMB temperature T_{cmb} and the corresponding energy density of DM $\rho_{\text{dm}}(T_{\text{cmb}}) = 8.095895 \times 10^{-47} \times \Omega_{\text{dm}} h^2$. However, for all parameter points considered in this work, we find that $s_{\text{sm}}(T_{\text{cmb}})/\rho_{\text{dm}}(T_{\text{cmb}}) = \text{const.} \approx 2.306 \times 10^9$, meaning that $f_\phi \propto m_\phi Y_\phi$ and consequently f_ϕ is already known from the start for a given value of $m_\phi Y_\phi$. Based on this observation, in the following, we will therefore often use f_ϕ and $m_\phi Y_\phi$ interchangeably to parameterise the pre-decay abundance of the decaying relic, noting that Y_ϕ generally is a more fundamental input parameter.

6 Results and discussion

In this section, we present the limits derived from our numerical framework discussed above. Here, we mainly consider the case $\phi \to \bar{\nu}_e \nu_e$ with $\text{BR}_{ee} = 0$; however, we will later also comment on other flavours as well as cases with $\text{BR}_{ee} \neq 0$. Consequently, the relevant parameter space is three-dimensional, featuring f_ϕ, or equivalently $m_\phi Y_\phi$, m_ϕ, and τ_ϕ. To provide maximal coverage of the parameter space, our discussion below is split into three sections covering the three possible combinations of two parameters each.

6.1 The $\tau_\phi - f_\phi$ ($m_\phi Y_\phi$) parameter plane

Let us first discuss the resulting limits in the $\tau_\phi - f_\phi$ parameter plane for a selection of fixed masses m_ϕ. Since we already discussed the two particular choices $m_\phi = 100\,\text{GeV}$ and $m_\phi = 500\,\text{GeV}$ at the end of section 2 (cf. figure 3), we will refrain from a redundant discussion of the individual contributions to the total limits, and instead focus on the general trend of how the limits change with m_ϕ.

In figure 10, we show the resulting constraints for different relic masses between $2\,\text{GeV}$ and $10\,\text{TeV}$ (solid, different colours). For smaller masses, i.e. $m_\phi \lesssim 2\,\text{GeV}$, we find that the limits become subleading to the ones from N_{eff} and are therefore irrelevant. This is due to the fact that *(i)* protons and neutrons become kinematically unavailable, and *(ii)* photodisintegration becomes less efficient. In fact, the threshold for the production of nucleons — which induce the leading constraints for large parts of parameter space — is roughly $m_\phi \sim 2\,\text{GeV}$. Moreover, when it comes to the efficiency of photodisintegration, we find that the EW shower from FSR
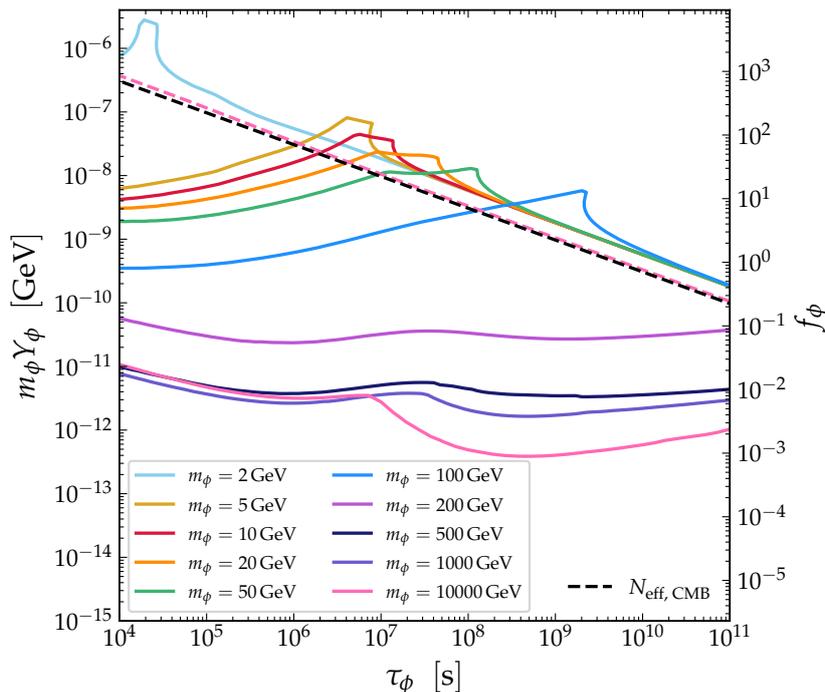
Figure 10. Overall BBN constraints originating from the decay $\phi \to \nu_e \bar{\nu}_e$ for different masses m_ϕ (solid, different colours). In addition, we also show the corresponding N_{eff} constraint for $m_\phi = 10\,\text{TeV}$ (pink dashed), noting that this constraint only deviates from the simple power law from eq. (A.5) described in appendix A (black dashed) for large masses and small lifetimes.

is suppressed below the EW scale, while thermal collisions at the GeV-scale only produce very few electron-positron pairs, even for high temperatures (cf. figure 4). Additionally, even though the fraction of injected EM material (cf. figure 6) due to non-thermal collisions is still relatively high for $\sqrt{s} = \mathcal{O}(\text{GeV})$, the corresponding cross-section (cf. figure 5) is already deep within the effective Fermi regime and therefore quickly decouples for small neutrino energies and thus relic masses. Consequently, for such small masses, we find an absence of efficient disintegration reactions, meaning that the light-element abundances are mainly changed by the modified cosmology, i.e. due to the change in the baryon-to-photon ratio. We therefore refrain from considering masses below $m_\phi = 2\,\text{GeV}$.

At this point, let us quickly note that all bounds that are driven by scattering reactions, decouple for sufficiently large lifetimes due to dilution effects, which implies that the modified background cosmology inevitable takes over the total limit at some point. In particular, this becomes apparent by the fact that the constraints for all masses $m_\phi \leq 200\,\text{GeV}$ asymptotically approach a line parallel to the one from the CMB N_{eff} limit (dashed), albeit weaker by a factor ~ 2. Regarding the N_{eff} constraint, we show two different lines: one for the "naive" expectation (dashed black), assuming that all neutrinos end up in the form of additional radiation, as well as the consistently-calculated one for $m_\phi = 10\,\text{TeV}$ (dashed pink), in which case a relevant portion of the neutrino energy ends up in the form of EM or hadronic radiation, thus leading to a slightly smaller energy density of neutrinos at recombination and therefore to weaker limits. The N_{eff} constraints for all other masses lie between these two

cases. In general, we find that for larger values of m_ϕ, the resulting constraints outperform the ones from N_{eff} for a larger range of lifetimes. Also, in the transition region, where the bounds transition from being dominated by the background cosmology to being dominated by disintegration reactions (e.g. at $\tau_\phi \sim 10^9\,\text{s}$ for $m_\phi = 100\,\text{GeV}$), we find a weakening of the limits due to a generic cancellation of both effects, as already discussed in section 2.

Starting at small masses, as soon as kinematically available, the hadronic injections from non-thermal scattering come to rapidly dominate the bounds for small lifetimes, thus pushing them beyond the ones from N_{eff}. This effect becomes apparent from the massive jump between $m_\phi = 2\,\text{GeV}$ (light blue) and $m_\phi = 5\,\text{GeV}$ (gold). However, beyond this boost, a further increase in mass up to $m_\phi \sim 50\,\text{GeV}$ (green) only leads to a mild further strengthening of the limits, as the energy $E < m_\phi/2$ of the injected neutrinos remains below the Z-resonance. However, when going past the resonance, i.e. from $m_\phi = 50\,\text{GeV}$ (green) to $m_\phi = 100\,\text{GeV}$ (blue), the limits strengthen by another factor ~ 10, showcasing the fact that the hadronic injections from non-thermal scattering receive a massive resonance boost. However, beyond this threshold, the rate for non-thermal interactions is no longer expected to increase drastically, since the corresponding cross-section (cf. figure 5) saturates. This effect can also be observed when comparing the left and right panels of figure 3 in section 2. Let us further stress that the bounds from hadrodisintegration do not become constant for $\tau_\phi < 10^4\,\text{s}$, but are expected to vanish for $\tau_\phi \lesssim 10^2\,\text{s}$, when all injected hadrons are efficiently stopped before initiating a hadronic cascade.[27]

Notably, pushing the mass beyond the EW scale, further enables efficient EM/hadronic injection due to FSR (also cf. figure 4), which causes another jump between $m_\phi = 100\,\text{GeV}$ (blue) and $m_\phi = 200\,\text{GeV}$ (purple) (also cf. figure 17 for the individual contributions for $m_\phi = 200\,\text{GeV}$, similar to figure 3). In fact, for sufficiently large masses, $m_\phi > 200\,\text{GeV}$ (purple), the limits originating from FSR outperform the ones from N_{eff} for the full range of lifetimes considered in this work. For even larger masses, the constraints still slightly strengthen with mass, but also start to slowly saturate towards the TeV-scale. For such masses, the limits become slightly stronger for intermediate lifetimes, $\tau_\phi \sim 10^8\,\text{s}$–$10^{10}\,\text{s}$, causing a small bump in an otherwise fairly flat limit. This can be understood from the fact that — in this regime — the EM contribution outperforms the (only mildly lifetime-dependent) hadronic one.

For the remainder of this section, let us discuss the asymptotic behaviour of the different limits. To this end, let us first fix the mass and the abundance of ϕ, thus quantifying the asymptotic behaviour for large lifetimes. In this case, when increasing the lifetime — at some point — the hadronic injections will consist solely of protons, since the neutrons with lifetime $\tau_n \approx 880\,\text{s} \ll \tau_\phi$ effectively decay immediately. In addition, for sufficiently large lifetimes, the energy-loss mechanism of protons becomes inefficient due to the low energy of the background particles, meaning that all protons enter the hadrodisintegration processes with their (potentially high) initial injection energy. As long as the hadrodisintegration rate is still above the Hubble rate, the hadronic cascade will therefore proceed until all protons (as well as other fragments) are scattered down to energies beyond which they can no longer disintegrate ^4He. As a result, the different parameters ξ_X^p, i.e. the numbers of

---

[27]Note that for $\tau_\phi < 10^2\,\text{s}$, other effects like the pion-induced interconversion of protons and neutrons become important.

additional nuclei produced due to the injection of a *single* proton, become independent of the injection time. Moreover, the number of injected protons only depends on the initial abundance ϕ, i.e. — for a fixed abundance — there is no dilution effect and the same amount of nucleons is injected for each individual decay, independent of the time. Therefore, the limits arising from hadronic injections are expected to approximately reach a plateau at large lifetimes. Considering EM injections instead, photodisintegration is well parameterised by the universal spectrum (cf. eq. (4.9)), which is always temperature dependent. Therefore, the asymptotic behaviour is non-trivial, and numerically we find that the limits from EM injections weaken for sufficiently large lifetimes.

Additionally, we can also quantify the asymptotic behaviour for large masses — for which FSR again dominates — by fixing the lifetime and the abundance of ϕ. Increasing m_ϕ for fixed τ_ϕ and $f_\phi \propto m_\phi Y_\phi$, leads to a linear reduction of the "no-decay" number density $\tilde{n}_\phi \propto f_\phi/m_\phi$ of ϕ (cf. eq. (5.1)). For the terms on the right-hand side of eq. (4.13) describing photodisintegration, this loss in number density is compensated for by an increase of $f_{\gamma,\text{univ}}$ with $E_0 \sim m_\phi$ (cf. eq. (4.9)), such that $f_{\gamma,\text{univ}} \propto S_\text{em}E_0 \sim \tilde{n}_\phi \zeta_\text{em}^\text{fsr}(m_\phi)E_0 \sim f_\phi m_\phi^{-1}\zeta_\text{em}^\text{fsr}(m_\phi)m_\phi \sim f_\phi \zeta_\text{em}^\text{fsr}(m_\phi)$. Hence, for large masses, the limits still increase but only slowly according to the (mild) increase of $\zeta_\text{em}^\text{fsr}$ with mass (cf. figure 4). Specifically, this behaviour is relevant for $\tau_\phi \gtrsim 10^7$ s, for which the EM component of FSR dominates the limits for large masses. However, for the terms on the right-hand side of eq. (4.14) describing hadrodisintegration, such a compensation does not exist for large m_ϕ. In fact, both N_hd^fsr and thus $\mathrm{d}n_\text{hd}^\text{fsr}/\mathrm{d}t$ (cf. figure 19) as well as ξ_X^y (cf. figure 16) only mildly (but not linearly) depend on $\sqrt{s} \sim E \sim m_\phi$, i.e. not enough to counteract the loss of abundance. Once the loss in sensitivity due to the smaller abundance overtakes the increase in sensitivity due to the slightly larger values of N_hd^fsr and ξ_X^y, the limits therefore undergo a turn-around, i.e. they start to weaken with mass after an initial phase of strengthening. Indeed, this behaviour is relevant for $\tau_\phi \lesssim 10^7$ s, in which case the limits are dominated by the hadronic component of FSR. Specifically, the results obtained above are well represented in the figure (cf. the change between $m_\phi = 1$ TeV and $m_\phi = 10$ TeV) and also in agreement with the results of [52].

Based on the discussion above, we naturally find that the EM contribution from FSR slowly grows towards an asymptotic value (cf. figure 4). In particular, this contribution is a good proxy for the full limits, as shown in figure 18 of appendix F.

6.2 The $m_\phi - \tau_\phi$ parameter plane

We now turn to the discussion of the $m_\phi - \tau_\phi$ parameter plane for a selection of fixed abundances. In figure 11, we show the corresponding results for six different choices of f_ϕ (different colours). As already discussed above, the N_eff bound is effectively independent of the particle mass and therefore manifests itself as a horizontal line, becoming stronger for a larger abundance. We show this bound as a dashed line, whereas the ones stemming from BBN are shown as solid lines. Moreover, the intersection of both constraints is marked by a star and the excluded region is shaded for clarity. We see that abundances above $f_\phi = 10^2$ are largely excluded for $m_\phi \gtrsim 1$ GeV, except for a small region, for which EM and hadronic injections are suppressed. Overall, reducing the initial abundances of ϕ relaxes the resulting limits. We find that for every abundance, there exists a value of τ_ϕ beyond which the N_eff constraints start
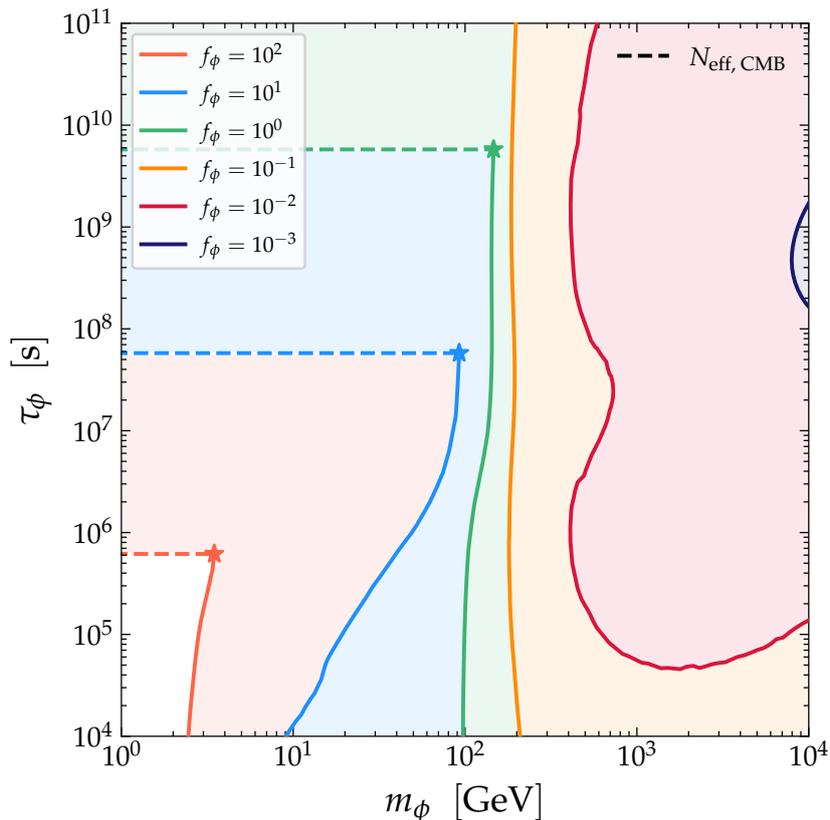
Figure 11. Overall BBN constraints originating from the decay $\phi \to \nu_e \bar{\nu}_e$ for different initial abundances f_ϕ (solid, different colours). In addition, we also show the corresponding N_{eff} constraints (dashed, different colours), as well as the point beyond which the latter ones become dominant (star). For clarity, we further shade the excluded part of parameter space.

to dominate, which is due to the different redshift behaviour of matter (ϕ) and radiation (ν). The parameter space above this line is fully excluded. For smaller lifetimes, the constraints are dominated by NBBN. Moreover, we find that small abundances are excluded only for large masses and intermediate lifetimes, as expected from the results in figure 10.

6.3 The $m_\phi - f_\phi$ parameter plane

Let us now discuss the $m_\phi - f_\phi$ parameter plane for a selection of fixed lifetimes. In figure 12, we show the resulting constraints for eight different choices of τ_ϕ (different colours). Here, the total BBN bounds obtained in this work are shown as solid lines, while the ones originating from N_{eff} are shown with dashed lines. Again, the transition points between the two are marked by a star. Note that we do not show $m_\phi < 1\,\text{GeV}$, since all bounds are weaker than the one from N_{eff} below this threshold, in accordance with our previous discussion. We find that particles with larger lifetimes require larger masses in order to outperform the N_{eff} limits. Moreover, there exist two points beyond which the limits strengthen substantially. The first one is at $m_\phi \sim 2\,\text{GeV}$ when hadronic final states becomes kinematically available. The second one is for masses close to the EW scale, at which point *(i)* the Z-resonance can be hit during non-thermal scattering and *(ii)* injection due to FSR becomes efficient. We
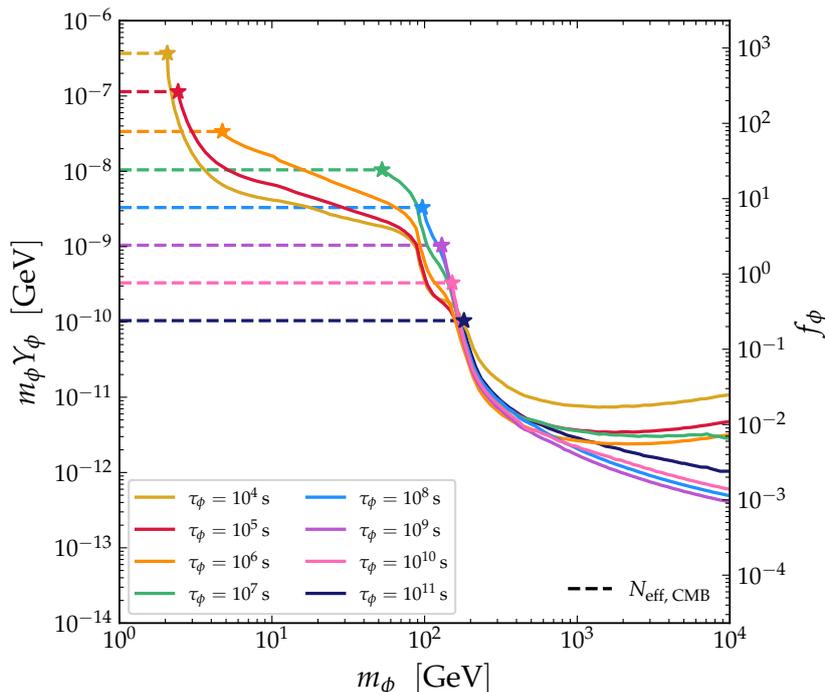
Figure 12. Overall BBN constraints originating from the decay $\phi \to \nu_e \bar{\nu}_e$ for different lifetimes τ_ϕ (solid, different colours). In addition, we also show the corresponding N_{eff} constraints (dashed, different colours), as well as the point beyond which the latter ones become dominant (star).

observe that for $m_\phi \gtrsim 500\,\text{GeV}$ and $\tau_\phi \gtrsim 10^8\,\text{s}$, the limits are generically well-described by effectively only considering the EM component of the EW shower from FSR (cf. figure 18 and appendix F for a more detailed discussion). However, for smaller lifetimes, the hadronic component of the EW shower from FSR is dominant for heavy relics, which explains the different asymptotic behaviour.

6.4 Results for other neutrino flavours

Let us now briefly discuss how the limits change when other neutrino flavours are initially injected in the ϕ decay. In figure 13, we show the resulting constraints in the $\tau_\phi - f_\phi$ plane for $m_\phi = 100\,\text{GeV}$, and *(i)* the three cases of single-flavour injections $\phi \to \bar{\nu}_e \nu_e$ (dark blue), $\phi \to \bar{\nu}_\mu \nu_\mu$ (green), $\phi \to \bar{\nu}_\tau \nu_\tau$ (orange), as well as *(ii)* the "democratic case" in which all neutrino-flavours are injected with equal branching ratios. We find that deviations among the different cases are very small, meaning that the bounds for $\phi \to \bar{\nu}_e \nu_e$ (which we mostly show throughout this paper) are also valid for the other cases to a very good approximation.

6.5 Results for $\text{BR}_{ee} \neq 0$

Up to this point, we have tried to keep this work as model-agnostic as possible. However, one may wonder how the resulting limits for $\phi \to \nu\nu$ compare to those for $\phi\phi \to e^+ e^-$, specifically since SU(2) invariance in the SM naively enables both channels with equal branching ratios. In figure 14, we therefore present the resulting limits for $m_\phi \in \{10, 100\}\,\text{GeV}$ (red, blue) and different choices of $\text{BR}_{ee} \neq 0$. Specifically, we show the case of pure neutrino injection

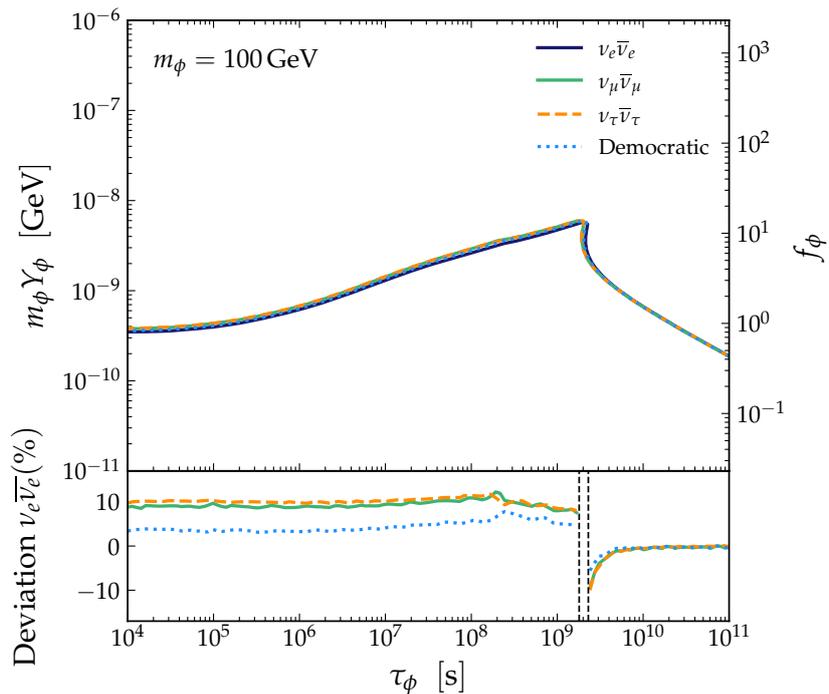**Figure 13.** Absolute and relative comparison of the total BBN limits for $m_\phi = 100$ GeV, assuming different initial neutrino flavours in the decay. For the relative comparison, we leave out a narrow region where there is no one-to-one correspondence between τ_ϕ and $m_\phi Y_\phi$.
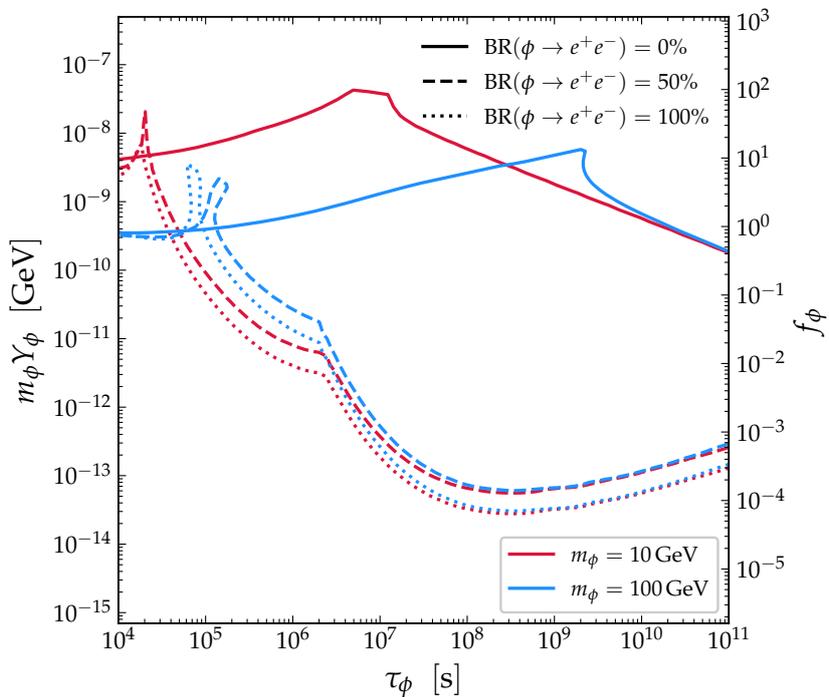

Figure 14. Comparison of the resulting BBN limits for $m_\phi \in \{10, 100\}$ GeV (red, blue) and different values of BR_{ee} (different linestyles). Specifically, we show the three cases $\mathrm{BR}_{ee} = 0$ (solid), $\mathrm{BR}_{ee} = 0.5$ (dashed), and $\mathrm{BR}_{ee} = 1$ (dotted).

with $\mathrm{BR}_{ee} = 0$ (solid), pure EM injection with $\mathrm{BR}_{ee} = 1$ (dotted), and mixed injections with $\mathrm{BR}_{ee} = 0.5$ (dashed). As expected, pure EM injection yield stronger limits for most lifetimes, since the direct injection of EM material (dotted) leads to much more efficient photodisintegration.

However, for small lifetimes, electromagnetic injection ceases to be effective and (even suppressed) hadronic injections become relevant. In this regime pure neutrino injections (solid) lead to comparably strong constraints, mainly due to the presence of hadronic injections arising from non-thermal scatterings. In addition, neutrinos also profit from the *delayed* injection of EM/hadronic material during the neutrino cascade, meaning that a relic with $\tau_\phi = 10^4\,\mathrm{s}$ still injects a significant amount of material at $t \gg \tau_\phi$ due to the fact that scattering is rare in large regions of the parameter space. Therefore, neutrinos can push beyond the usual cutoff at $\tau_\phi \sim 10^4\,\mathrm{s}$ ($\tau_\phi \sim 10^2\,\mathrm{s}$) below which the constraints from photodisintegration (hadrodisintegration) vanish. For the case of electrons it turns out that contributions from hadronic FSR are still relevant even for these rather small masses due to the emission of virtual photons — a contribution which is not present for the neutrinos.[28] Note that there now also exists a transition region between $\tau_\phi = 10^4\,\mathrm{s}$ and $\tau_\phi = 10^5\,\mathrm{s}$, for which a cancellation between the deuterium production due to the hadronic injection and the deuterium destruction due to the EM injection weakens the limits from electron injections significantly.

When considering the SU(2) symmetric case (dashed), the limits are typically weaker than the ones for pure injections by a factor 2. Again there is a cancellation among the EM and hadronic contributions. Note however that the cancellation is displaced significantly w.r.t. the case of $\mathrm{BR}_{ee} = 1$ for $m_\phi = 100\,\mathrm{GeV}$ as the cancellation is very sensitive to the relative strength of hadronic and EM contributions. Note that for much larger masses, $m_\phi \gg 100\,\mathrm{GeV}$, and for large lifetimes, the limits for pure neutrino and pure EM injections are expected to exhibit a very similar behaviour, since — in this case — also the limits for $\mathrm{BR}_{ee} = 0$ are dominated by EM injections, specifically from FSR. Consequently, the ratio of both limits is then equal to the ratio of the injected EM energy. Since for $\mathrm{BR}_{ee} = 1$ basically 100% of the energy is released in the form of EM material, this ratio is given by the red line in figure 4. Consequently, the ratio approaches a value close to unity asymptotically for large masses, implying that the bounds for pure neutrino injections are only marginally weaker in this regime. For smaller lifetimes, the same statement is not true, since hadronic injections complicate the matter. In conclusion, we therefore find that neutrinos still lead to important constraints, even if the relic can also directly decay into electron-positron pairs. However, there also exist well-motivated models for which BR_{ee} is naturally either small or zero, some of which are discussed in the following section.

6.6 A note on model-building

While the previous section discusses a valid point, injected neutrinos do not necessarily imply an equal amount or even any primary charged leptons. Let us briefly discuss some potential scenarios, for which a relic ϕ can dominantly decay into two neutrinos. A well known example of such a scenario is the Majoron framework. In its standard minimal version, the key interaction is of the form $y_N \phi \overline{N^c} N$ with a sterile neutrino N and $\phi = s e^{i\theta}$,

---

[28]We thank the anonymous referee for pointing out the relevance of this contribution.

where θ is the pseudoscalar Majoron field, a pseudo-Goldstone boson resulting from the breaking of a global Abelian $B - L$ symmetry by the vacuum expectation value of the field s (see [56, 57] and e.g. [58, 59]). The decay into light neutrinos proceeds through seesaw mixing between the heavy sterile neutrino and ordinary neutrinos at a rate $\Gamma_\phi \sim m_\theta y_N^2 m_\nu^2/(16\pi m_N^2)$, with m_ν being the neutrino mass scale. In this scenario, a suppressed decay into e^+e^- is also induced at one loop, with an extra relative factor $\sim (m_e^2 m_N^2/m_W^4)/(4\pi^2)^2$ in the decay width. Specifically, this decay channel is relatively enhanced by the mass of the seesaw-state, but involves extra chirality flip and loop suppression factors, so that it is suppressed for $m_N \lesssim 10^9$ GeV. The range of lifetimes considered in this work, $\tau_\phi = (10^4 - 10^{12})$ s, typically requires a seesaw scale $m_N \sim (10^4 - 10^8 \text{ GeV}) \cdot y_N$. Obviously, in this case the Majoron is not the DM particle, as often considered, but instead corresponds to the unstable relic ϕ assumed in this work. Overall, we see that for such a model, the BBN bounds on the neutrino channel are stronger than the BBN ones on the charged lepton channel in a significant part of the parameter space.

Another option is to consider the coupling of a gauge boson to two sterile neutrinos, also followed by seesaw neutrino mixings into ordinary neutrinos [60]. In this case, both the decay into light neutrinos and into e^+e^- are suppressed by four powers of the heavy sterile neutrino, with an extra loop suppression for the charged lepton channel, so that here too the primary decay channel is into light neutrinos.

Yet another straightforward example of a scenario leading to a decay into neutrinos, without — in this case — any production of primary charged leptons, is to consider the source particle as the neutral component of a scalar triplet of hypercharge 2, $\Delta = (\delta^{++}, \delta^+, \delta^0)$. In this case, the two charged scalar triplet components, which are heavier due to EW symmetry breaking, disappear quickly from the thermal bath by decaying into the neutral component. The remaining neutral component decays into $\nu\nu$ from a ΔLL Yukawa interaction. The relic density of such triplets, prior to their decay, is the one of a hypercharge 2 scalar triplet DM candidate, see [61, 62], that is to say of a typical WIMP. From the EW annihilations, the relic density is below (above) the observed DM relic density for masses below (above) 2.1 TeV [63] (and can also be below it around the Higgs boson resonance). Scalar quartic couplings can further increase the annihilation, i.e. suppress the relic density [62]. In this framework, there is no decay into charged leptons due to lepton-number conservation.

Phenomenologically very similar to the decays of a relic into two neutrinos as studied in this work are decays into one neutrino and one dark-sector state. In the latter case, it is natural that no associated charged leptons are produced. An example of such a setup would be a supersymmetric sneutrino, which decays into one neutrino and one gravitino, a scenario previously studied in [10]. Another option is to consider the decay of a scalar or a gauge boson into two not so heavy sterile neutrinos ν_s. In this case, the leading decay, if kinematically allowed, would be $\phi \to \bar{\nu}_s \nu_s$, but mixing with SM neutrinos would also induce e.g. $\phi \to \bar{\nu}_s \nu$, which is much less suppressed than the decay into charged leptons. In this case, there would be a strong limit from N_{eff} considerations, but for sufficiently large mixings, the limits derived in this work may also be relevant. If the sterile neutrino mass is close to the mass of ϕ, the decay $\phi \to \bar{\nu}_s \nu$ could also be the leading decay channel.

7   Conclusion

In this work, we have performed a detailed study of the decay of a relic particle ϕ into neutrinos, $\phi \to \nu\bar{\nu}$, concentrating on lifetimes larger than about 10^4 seconds. At this time, primordial nucleosynthesis has basically completed, but late-time injections of EM or hadronic material may still disintegrate the light elements previously synthesised. The case of neutrino injections is of particular interest, as it is expected to be less constrained compared to injection of other SM particles. We have shown that, although these constraints are generically weaker, they are still much stronger than naively expected, covering large parts of the parameter space defined by the mass m_ϕ, the lifetime τ_ϕ, and the abundance f_ϕ of the relic particle ϕ.

Once produced, the neutrinos interact very little with the abundant photons, baryons, and electrons that are present at the time. However, they can still interact significantly with the numerous background neutrinos that are present in the primordial plasma (*thermal scattering*). Moreover, two injected neutrinos can also interact among each other (*non-thermal scattering*). While the injection of electron-positron pairs due to thermal scattering can lead to relevant limits, we find that non-thermal scattering further results in hadronic injections, which we show to be very important, as they dominate the resulting limits by several orders of magnitude in significant regions of parameter space. Specifically, this effect dominates the constraints for masses up to $\sim 150\,\mathrm{GeV}$ and (depending on the mass) lifetimes up to 10^7–10^9 s (cf. figures 3 and 10). In fact, hadronic injections from non-thermal scattering are dominant in this region for three main reasons:

a) non-thermal scatterings, unlike thermal ones, profit from the fact that they involve two energetic neutrinos in the initial state, leading to a larger centre-of-mass energy, thus allowing the production of heavier particles due to a larger cross-section from resonant scattering close to the Z-pole;

b) the nucleons produced via non-thermal scattering feature a hadrodisintegration cross-section for destroying helium-4 (and producing deuterium) that is two orders of magnitude larger than the one for photodisintegration of deuterium;

c) helium-4 is more than three orders of magnitude more abundant than deuterium.

Instead, for masses up to $\sim 100\,\mathrm{GeV}$ and larger lifetimes, the dominant constraint on the light-element abundances stems from the modification of the baryon-to-photon ratio η. However, in this regime, the CMB N_{eff} bounds dominate over the ones from BBN. In between these two regimes, there is a cancellation of both effects, which leads to weaker constraints. Interestingly, for small lifetimes, the constraints coming from neutrino injections can be more stringent than the ones coming from electron-positron injection, as the effects of a pure EM cascade is suppressed at early times.

For masses beyond the EW scale, the effects of the EW shower from FSR quickly become dominant. In this region, we further find that — for lifetimes that are not too large — the induced hadrons lead to an improvement of the constraints by several orders of magnitude compared to the injection of only EM material. In fact, hadrons dominate the limits for lifetimes up to 10^7 s. For larger lifetimes, photodisintegration from EM material due to FSR dominates. Both EM and hadronic contributions from FSR lead to an upper bound

on the source-particle abundance, which is relatively independent of the lifetime and mass, namely $f_\phi \leq 10^{-2}$–10^{-3}.

The derivation of our results was made possible by a new, Monte-Carlo inspired probabilistic approach, which can be used to track the relevant part of the evolving phase-space distribution of the injected neutrinos. This is essential in order to faithfully evaluate the effects of non-thermal scattering, which previously had been largely ignored in the literature due to its computationally challenging nature. In this context, an important, simplifying insight is that any neutrino that scatters once, no longer has to be tracked as part of the spectrum. This is particularly true in case the corresponding interaction rates are smaller than the Hubble rate, so that neutrinos typically scatter only once (if at all).

In this work, we have concentrated on a "model-agnostic" approach regarding the late-time injection of energetic neutrinos and included generic effects which had previously been neglected in the literature. We have briefly commented on a number of different particle-physics scenarios, which result in a predominant injection of neutrinos (possibly in addition to truly decoupled states) and therefore are expected to lead to constraints that are very similar to the ones presented in this work. Overall, we found that, even for this seemingly innocent case, the resulting constraints from photo- and hadrodisintegration can be very stringent, much stronger than the indirect ones from a new radiation component parameterised by $N_{\rm eff}$.

Acknowledgments

A   A simplified calculation of the CMB $N_{\rm eff}$ limit

In this appendix, we present a simplified calculation of the CMB $N_{\rm eff}$ constraint within our scenario. In the main text and specifically for eq. (4.18), we assumed that ϕ had already decayed fully into neutrinos. However, to track the cosmological evolution and to estimate the resulting limit, it is useful to also take into account the energy density in ϕ prior to its decay. Using both the energy density $\rho_\nu^{\rm n\text{-}th}(t)$ of the ultra-relativistic non-thermal neutrinos, as well as the energy density $\rho_\phi(t)$ of the non-relativistic relic, we can define a combined DM energy density $\rho_{\rm DM}(t)$ by means of the simplified relation

$$\rho_\nu^{\rm n\text{-}th}(t) + \rho_\phi(t) \equiv \begin{cases} f_\phi \rho_{\rm DM}(t) & \text{for} \quad t \leq \tau_\phi \\ f_\phi \rho_{\rm DM}(t) \left(\frac{R(\tau_\phi)}{R(t)} \right) & \text{for} \quad t > \tau_\phi \end{cases} . \tag{A.1}$$

Taking into account the transition from radiation to matter domination at the time $t_{\rm eq}$ of matter-radiation equality, we find at the time $t_{\rm rec}$ of recombination

$$\rho_\nu^{\rm n\text{-}th}(t_{\rm rec}) \simeq f_\phi \rho_{\rm DM}(t_{\rm rec}) \left(\frac{R(\tau_\phi)}{R(t_{\rm eq})} \right) \left(\frac{R(t_{\rm eq})}{R(t_{\rm rec})} \right) \simeq f_\phi \rho_{\rm DM}(t_{\rm rec}) \frac{\tau_\phi^{1/2} t_{\rm eq}^{1/6}}{t_{\rm rec}^{2/3}} . \tag{A.2}$$

Using this relation, we can derive an approximate limit on the combination $f_\phi \tau_\phi^{1/2}$, assuming that none of the neutrinos dissipate any of their energy in the form of EM/hadronic material.[29] Taking for simplicity $N_{\rm eff,sm} = 3$, since the difference to the actual value is much smaller than what is allowed by the limit on $\Delta N_{\rm eff}$, and by approximating $T_\nu = (4/11)^{1/3} T$, we find

$$\Delta N_{\rm eff} \simeq \frac{\rho_\nu^{\rm n\text{-}th}(t_{\rm rec})}{2\frac{7}{8}\frac{\pi^2}{30}T_{\nu,\rm rec}^4} < 0.29 \tag{A.3}$$

$$\Rightarrow \quad f_\phi \tau_\phi^{1/2} \lesssim 1.4 \cdot 10^6 \, {\rm s}^{1/2} \frac{\rho_\nu^{\rm th}}{3\rho_{\rm DM}}\Big|_{t=t_{\rm rec}} = 1.4 \cdot 10^6 \, {\rm s}^{1/2} \frac{\Omega_\nu^{\rm th} h^2}{3\Omega_{\rm DM} h^2}\Big|_{t=t_{\rm today}} z_{\rm rec} \simeq 7 \cdot 10^4 \, {\rm s}^{1/2}$$

$$\Rightarrow \quad f_\phi \left(\frac{\tau_\phi}{10^4 \, {\rm s}}\right)^{1/2} \simeq 700 \,. \tag{A.4}$$

Here, $t_{\rm eq} \approx 50000 \, {\rm yrs}$, $t_{\rm rec} \approx 380000 \, {\rm yrs}$, and $z_{\rm rec} \approx 1100$. Moreover, we used the fact that $2\frac{7}{8}\frac{\pi^2}{30}T_{\nu,\rm rec}^4 \simeq \rho_\nu^{\rm th}(t_{\rm rec})/3$ which can then be related to $\Omega_\nu^{\rm th} h^2$ [46]. In fact, this approximation is very close to what is obtained by a full numerical treatment, i.e.

$$f_\phi \left(\frac{\tau_\phi}{10^4 \, {\rm s}}\right)^{1/2} \simeq 680 \,, \tag{A.5}$$

which we use e.g. for the black line in figure 10.

Finally, let us note that as part of our numerical analysis, we calculate a more consistent value of $N_{\rm eff}$, taking into account also the fact that the non-thermal neutrinos lose some of their energy in the form of EM/hadronic material. Eq. (A.5) is therefore used only for comparison and reference.

B   On the relevance of neutrinodisintegration reactions

In this appendix, we show that the abundance of injected neutrinos is not significantly impacted by deep inelastic scattering (DIS) between the neutrinos and the background nucleons.[30] To this end, we compare the interaction rate of this process $\sim \langle \sigma_n v \rangle n_n$ with the one for scattering off thermal background neutrinos $\sim \langle \sigma_\nu v \rangle n_\nu$, i.e.

$$\frac{\langle \sigma_n v \rangle n_n}{\langle \sigma_\nu v \rangle n_\nu} \sim \frac{10^{-38} \, {\rm cm}^2 \frac{E}{\rm GeV} \eta T^3}{G_F^2 E T^4} \sim \frac{10^{-11}\eta}{10^{-10}\frac{T}{\rm GeV}} \sim 10^2 \eta \frac{\rm MeV}{T} \sim 10^{-8}\frac{\rm MeV}{T} \,. \tag{B.1}$$

Here, we utilised the DIS cross-section from [23, 64] as well as $n_n \sim \eta T^3$. In conclusion, we find that — for all relevant temperatures $T \lesssim 10^{-2} \, {\rm MeV}$ —, it is indeed safe to ignore the scattering reactions between non-thermal neutrinos and the background nuclei. However, we note that similar to the quasi-elastic scattering studied in [22], sizeable neutrino injections during thermal BBN could change this conclusion. Nevertheless, this problem is irrelevant for this work specifically, since we focus on $\tau_\phi \geq 10^4 \, {\rm s}$.

---

[29]In our numerical analysis, we take into account that a fraction of the energy ends up in the form of EM/hadronic material.

[30]For neutrino energies beyond a few GeV, DIS scattering dominates over elastic scattering.

C A simplified analytical expression for the neutrino spectrum

In this appendix, we briefly discuss a simplified analytical expression for the neutrino spectrum that we found during the early stages of the project. After developing a numerical implementation which is both efficient and more precise, we refrained from using this expression in more detail. However, we still find that this expression can give some instructive insights, specifically in the context of showcasing the agreement between the numerical and the analytical result in figure 8.

In the following, we sketch the different steps that are required to derive this simplified analytical approximation, thereby highlighting all important approximations. Assuming that all scattering rates are negligible w.r.t. the Hubble expansion, and by using $H \simeq 1/2t$, the evolution of the neutrino spectrum is determined by the following simplified Boltzmann equation

$$\frac{\partial f_\nu}{\partial t} - \frac{E}{2t}\frac{\partial f_\nu}{\partial E} \simeq \frac{4\pi^2}{E^2}\frac{n_\phi}{\tau_\phi}\delta(E - E_0)\,. \tag{C.1}$$

Here, we used the fact that the initial particle ϕ is non-relativistic and features a lifetime τ_ϕ as well as an abundance n_ϕ, the latter of which is defined in eq. (5.1). Moreover, $E_0 = m_\phi/2$. This equation can be solved analytically, which yields

$$f_\nu(t, E) \simeq \frac{8\pi^2}{E_0^5}\frac{n_\phi\left(tE^2/E_0^2\right)}{\tau_\phi}tE^2\Theta(E_0 - E)\Theta\left(E - E_0\sqrt{t_0/t}\right)\,. \tag{C.2}$$

In this expression, the upper limit in energy, as indicated by the first Heaviside function, is determined by the maximally possible neutrino energy $E_0 = m_\phi/2$. The lower limit, as indicated by the second Heaviside function, is less trivial and accounts for the "starting time" t_0 of the neutrino injection. Technically, such a starting time is in fact artificial, since decays always happen in some capacity, even well before $t = \tau_\phi$. Here, and specifically for figure 8, we set $t_0 = \tau_\phi/\Lambda$ as in section 5 for consistency. However, let us note that even in a more realistic scenario, the interactions of the neutrinos with the background will eventually deplete the spectrum, thus leading to an effective cutoff, a feature that can e.g. be observed in the full numerical result in figure 8. Finally, let us note that the spectrum further depends on the number density n_ϕ evaluated at $t(E/E_0)^2$. This factor naturally shuts off the injections due to the exponential suppression of the relic abundance beyond its lifetime. Additionally, the redshift factor in the argument ensures that only those neutrinos injected at a later time suffer from this suppression, while neutrinos injected at earlier times have already lost some of their energy E due to redshift, i.e. $E \ll E_0$, and therefore are not suppressed.

D A more detailed look at the hadronic cascade

We provide a concise summary of the physical processes driving the hadronic cascade in this appendix. In combination with the Boltzmann equation in eq. (4.15), the numerical framework presented in this section forms the basis for the implementation of the hadrodisintegration reactions in v2.0.0-dev of ACROPOLIS. In particular, we focus on the calculation of the parameters $\xi_X^y[n_j](T, K)$ for $y \in \{p, n\}$ and $X \in \{p, n, \mathrm{D}, \mathrm{T}, {}^3\mathrm{He}, {}^4\mathrm{He}\}$.

In a nutshell, energetic hadrons injected into the thermal plasma can scatter off the background particles, thus initiating a hadronic cascade. In this context, the most important processes are

1. EM scattering with background particles, i.e. primarily electrons e^- and photons γ, which leads to continuous energy loss,

2. decay of the unstable hadrons, i.e. primarily neutrons in our case,

3. hadronic scattering with background nuclei, i.e. primarily protons p_{BG} and helium-4 nuclei α_{BG}.

Notably, when describing these processes, we can safely neglect the expansion of the Universe, since the Hubble rate is always much smaller than the rates of any of the processes above.[31]

Before delving into a detailed description of the various processes, let us first point out some important approximations that we employ as part of our calculation. In KKM05, one of the main goals was to derive strictly conservative constraints originating solely from the destruction of ^{4}He and the subsequent production of D, T and ^{3}He. Since we follow this publication closely, many of the following simplifications are also based on this principle. However, it is worth noting that — in the larger context of this work — the injection of energetic nucleons is only one of the important effects, and the final light-element abundances are also impacted by a modified background cosmology and EM injections. As discussed already in the main text, hadronic and EM injections can have opposite effects and thus lead to cancellation effects, which can make some exclusion bounds overly optimistic and no longer conservative. However, in practise, we find that such cancellations are only relevant for very small parts of parameter space, meaning that the following approximations can still largely be considered conservative:

Target particles. As shown in KKM05, the nuclear cross-sections are of similar orders-of-magnitude for all light elements, meaning that the corresponding scattering probabilities differ only due to the different abundances of the targets. Therefore, we consider only the two most abundant nuclei as targets, i.e. p_{BG} and α_{BG}. The impact of any other element is generically suppressed by their small abundance.

Projectile particles. The most important hadrons that are injected by non-thermal scattering of the neutrinos are p and n, with additional (high-energy) nucleons being produced during the various steps of the hadronic cascade. Heavier final-state nuclei resulting from the cascade often have a smaller kinetic energy based on the prescription used for the energy transfer (also cf. appendix C of KKM05). Their impact on the subsequent evolution of the cascade is therefore subleading compared to the one of neutrons and protons. As a result, we assume that only protons and neutrons act as projectile particles. However, in accordance with KKM05, we do use a simplified treatment to estimate if the produced nuclei survive or are disintegrated by either a CMB photon or

---

[31]We estimate this by comparing the Hubble rate $H \sim 1\,\mathrm{s}^{-1}(T/\mathrm{MeV})^2$ with a typical nuclear interaction rate $\Gamma \sim n_p\sigma_{\mathrm{nuc}} \sim 10^8\,\mathrm{s}^{-1}(T/\mathrm{MeV})^3 \gg H$. Consequently, for any temperature before recombination, the hadronic cascade is well approximated as instantaneous w.r.t. to cosmic expansion.

a background proton (see discussion below). If the survival condition is not met, we assume that the particle is broken up into its separate nucleons with equal momenta, e.g. $p_{\text{BG}} + \text{T} \rightarrow p + 2n$, with the latter ones being reinjected into the cascade if their kinetic energy is above $20\,\text{MeV}$.

Energy range. Experimental measurements for the cross-sections of hadronic processes where the kinetic energy of the projectile is higher than $\sim 20\,\text{GeV}$ are not available, except for some data on pp and np scatterings.[32] However, based on the available data, KKM05 estimates the cross-sections for high-energy hadronic scattering to become approximately constant beyond $20\,\text{GeV}$, an approximation that we also employ in this work. Nevertheless, we note that the energy of nucleons injected from FSR and/or non-thermal scattering is usually only a few % of the initial energy and thus — for the parameters considered in this work — rarely much above the extrapolation threshold.

Isospin symmetry. Even below ~ 20 GeV, some relevant experimental data is not available. Therefore, we invoke isospin symmetry to approximate the missing cross-sections whenever necessary and possible.[33] More precisely, we consider all processes shown in table 3,[34] but use separate cross-sections for protons and neutrons only for the reactions in the upper part of the table, while using the same (isospin-symmetric) cross-section for protons and neutrons in the lower part. Although a comprehensive analysis of the full nuclear interaction network would yield more precise predictions for the effects of hadrodisintegration, we still expect that all dominant effects are captured well within this approximation. In particular, since we anyway assume a 20% uncertainty on the nuclear reaction rates.

D.1 Parametrisation of the hadron spectrum

In this section, we introduce the scheme that we use to track the spectrum/amount of the individual nuclei that are present during the cascade. Here, the numerical framework is inspired by the discussion in appendix D of KKM05. First, we introduce a vector $\underline{A}$ of the form

$$\underline{A} = \begin{pmatrix} \underline{N}_p \\ \underline{N}_n \\ \Delta N_p \\ \Delta N_n \\ \Delta N_{\text{D}} \\ \Delta N_{\text{T}} \\ \Delta N_{^3\text{He}} \\ \Delta N_{^4\text{He}} \end{pmatrix} . \tag{D.1}$$

In this expression, $\underline{N}_p$ and $\underline{N}_n$ are N-dimensional vectors describing the distribution of protons and neutrons over N energy bins that are equidistant in log-space, e.g. $N_{p,i}$ is the number of

---

[32]In general, we use the nuclear cross-sections presented in KKM05.

[33]EM corrections are typically subleading in the energy range of interest.

[34]Note that we denote charged pions, neutral pions, and multiple pions in the final state all by π for simplicity.

protons in the ith energy bin. Here, the minimal kinetic energy $K_{\min} \approx 20\,\mathrm{MeV}$ equals the threshold energy for the disintegration of helium-4, and the maximal kinetic energy $K_{\max}$ corresponds to the highest possible injection energy of the hadrons. Moreover, the six different values ΔN_X count the relative number of X particles that have been produced/destroyed due to the hadronic injection. Consequently, $\underline{A}$ is a $2N + 6$ dimensional vector. Note that ΔN_p and ΔN_n must be considered in addition to the spectra $\underline{N}_p$ and $\underline{N}_n$, as they also track nucleons with kinetic energy $K < K_{\min}$. *In the following, we label the different entries starting at an index 0 as this directly maps to the numerical implementation, and we use the same convention for the indices of any of the matrices discussed below.*

The evolution of $\underline{A}$ throughout the cascade is then handled by two evolution matrices, i.e. one matrix $\mathbf{M}_{\mathrm{eloss}}$ describing the energy loss of the available hadrons up until their next scattering reaction, and one matrix $\mathbf{M}_{\mathrm{scat}}$ describing *one* of those proceeding hadronic scattering reactions. After *one* step of the cascade, we thus have $\underline{A}_0 \to \mathbf{M}_{\mathrm{scat}}\mathbf{M}_{\mathrm{eloss}}\underline{A}_0$ for some initial vector $\underline{A}_0$. As an example, let us assume that a single proton with kinetic energy K is injected into the plasma. The initial vector $\underline{A}_0$ would then only have two non-zero entries, namely $\Delta N_p = 1$ and $N_{p,j} = 1$, where j denotes the energy bin to which K belongs to. During the cascade, i.e. by applying $\mathbf{M}_{\mathrm{scat}}\mathbf{M}_{\mathrm{eloss}}$ repeatedly, many other entries of $\underline{A}$, including ΔN_{D} etc., will assume non-zero values. Ultimately, the calculation converges, leading to a final vector $\underline{A}_f = \lim_{n\to\infty}(\mathbf{M}_{\mathrm{scat}}\mathbf{M}_{\mathrm{eloss}})^n\underline{A}_0$, which can be used to extract ξ_X^y. We discuss the calculation of the two transition matrices in the respective sections below.

D.2 Electromagnetic energy loss

For the calculation of the hadronic cascade, it is of particular importance to perform a careful analysis of the energy loss of the injected hadrons due to EM interactions (point 1 above): if the hadrons lose much of their energy before scattering hadronically, they might no longer have enough energy to initiate any disintegration reactions. However, if their energy loss is small, they can instead scatter energetically with the background nuclei, thus potentially destroying them in the process. In this subsection, we therefore first discuss the relevant processes leading to such EM energy loss. An injected hadron H_i can interact with the background electrons and photons, mainly via:

- Coulomb/magnetic moment scattering $H_i + e^\pm \to H_i + e^\pm$;

- Compton scattering $H_i + \gamma \to H_i + \gamma$;

- Bethe-Heitler (BH) scattering $H_i + \gamma \to H_i + e^+ + e^-$;

- Photo-pion process $H_i + \gamma \to H_i' + \pi$.

Here, H_i' is a potentially different hadron. For charged hadrons, e.g. protons, EM energy loss is dominated by Coulomb scattering, whereas for neutral hadrons, e.g. neutrons, magnetic-moment interactions are most relevant. In general, we can express the energy-loss rate as (cf. appendix B of KKM05 for a collection of the individual expressions)

$$\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t} = \left(\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t}\right)_{\mathrm{Coulomb}} + \left(\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t}\right)_{\mathrm{Compton}} + \left(\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t}\right)_{\mathrm{BH}} + \left(\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t}\right)_{\mathrm{photo\text{-}pion}} . \quad (\mathrm{D.2})$$

Process	$i = n$	$i = p$	Reaction Type
(i, p_{BG}; 1)	$n + p_{BG} \rightarrow n + p$	$p + p_{BG} \rightarrow p + p$	elastic
(i, p_{BG}; 2)	$n + p_{BG} \rightarrow n + p + \pi$	$p + p_{BG} \rightarrow p + p + \pi$	inelastic
(i, p_{BG}; 3)	$n + p_{BG} \rightarrow n + n + \pi$	$p + p_{BG} \rightarrow p + n + \pi$	inelastic
(i, p_{BG}; 4)	$n + p_{BG} \rightarrow p + p + \pi$	$p + p_{BG} \rightarrow n + p + \pi$	inelastic
(i, p_{BG}; 5)	$n + p_{BG} \rightarrow p + p + \pi$	$p + p_{BG} \rightarrow n + n + \pi$	inelastic

Process	$i = n$	$i = p$	Reaction Type
(i, α; 1)	$n + \alpha_{BG} \rightarrow n + \alpha$	$p + \alpha_{BG} \rightarrow p + \alpha$	elastic
(i, α; 2)	$n + \alpha_{BG} \rightarrow \mathrm{D} + \mathrm{T}$	$p + \alpha_{BG} \rightarrow \mathrm{D} + {}^3\mathrm{He}$	inelastic
(i, α; 3)	$n + \alpha_{BG} \rightarrow 2n + {}^3\mathrm{He}$	$p + \alpha_{BG} \rightarrow p + n + {}^3\mathrm{He}$	inelastic
(i, α; 4)	$n + \alpha_{BG} \rightarrow p + n + \mathrm{T}$	$p + \alpha_{BG} \rightarrow 2p + \mathrm{T}$	inelastic
(i, α; 5)	$n + \alpha_{BG} \rightarrow n + 2\mathrm{D}$	$p + \alpha_{BG} \rightarrow p + 2\mathrm{D}$	inelastic
(i, α; 6)	$n + \alpha_{BG} \rightarrow p + 2n + \mathrm{D}$	$p + \alpha_{BG} \rightarrow 2p + n + \mathrm{D}$	inelastic
(i, α; 7)	$n + \alpha_{BG} \rightarrow 2p + 3n$	$p + \alpha_{BG} \rightarrow 3p + 2n$	inelastic
(i, α; 8)	$n + \alpha_{BG} \rightarrow n + \alpha + \pi$	$p + \alpha_{BG} \rightarrow p + \alpha + \pi$	inelastic

Table 3. *Top:* hadronic processes involving a background proton p_{BG}. *Bottom:* hadronic processes involving a background helium-4 nuclei α_{BG}. Both tables are taken from KKM05; however, following KKMT18, we added additional inelastic processes in the upper part.

Following KKM05, we then define the quantity

$$R_{H_i + A_j \rightarrow A_k} \left(E_{H_i}^{(\mathrm{in})}, E'_{H_i}; T \right) \equiv \int_{E_{H_i}^{(\mathrm{in})}}^{E'_{H_i}} \Gamma_{H_i + A_j \rightarrow A_k} \left(\frac{\mathrm{d}E_{H_i}}{\mathrm{d}t} \right)^{-1} \mathrm{d}E_{H_i} , \qquad (\mathrm{D.3})$$

which can be interpreted as the number of hadronic interactions from table 3 of the type $H_i + A_j \rightarrow A_k$ that happen during the time it takes for H_i to electromagnetically downscatter from the initial energy $E_{H_i}^{(\mathrm{in})}$ to the final energy E'_{H_i}. Here, $A_{j/k}$ are generic initial and final states of the hadronic interactions in table 3, and $\Gamma_{H_i + A_j \rightarrow A_k}$ are the corresponding interaction rates. Since we only consider p_{BG} and α_{BG} to be relevant targets, we have $A_j \in \{p_{\mathrm{BG}}, \alpha_{\mathrm{BG}}\}$. Moreover, for the hadronic cascade $H_i \in \{p, n\}$, whereas $H_i \in \{p, n, \mathrm{D}, \mathrm{T}, {}^3\mathrm{He}, {}^4\mathrm{He}\}$ for the calculation of the survival probability (see below). Finally, a list of all possible (multi-particle) states A_k is encoded in table 3, which labels interactions as $(i, j; k)$. Note that the notation used in eq. (D.3) is somewhat simplified, since only T is specified as an external parameter. However, in reality, the hadronic cascade also depends on the primordial helium-4 abundance $\mathcal{Y}_p$ and the baryon-to-photon ratio η via the interaction rates $\Gamma_{H_i + A_j \rightarrow A_k} \propto n_{A_j}$.[35] We keep this dependence implicit in our notation to avoid visual clutter.

---

[35]In terms of final quantities $\xi_X^y[n_j](T, K)$, this manifests in terms of the dependence on $[n_j]$, i.e. $Y_j = n_j/n_b$ with the baryon number density $n_b \sim \eta n_\gamma$.

Using eq. (D.3), we estimate the total energy loss by summing over all hadronic processes, i.e.

$$R^{H_i}\left(E^{(\mathrm{in})}_{H_i}, E'_{H_i}; T\right) \equiv \sum_{j,k} R_{H_i + A_j \to A_k}\left(E^{(\mathrm{in})}_{H_i}, E'_{H_i}; T\right). \tag{D.4}$$

Here, the quantity $R^{H_i}(E^{(\mathrm{in})}_{H_i}, E'_{H_i}; T)$ parameterises the total number of hadronic scattering reactions that occur before the initial energy $E^{(\mathrm{in})}_{H_i}$ of the hadron H_i is reduced to the final energy E'_{H_i} due to EM interactions with the thermal plasma. This parameter allows determining whether an energetic hadron is stopped before engaging in hadronic interactions. On the one hand, if $R^{H_i}(E^{(\mathrm{in})}_{H_i}, E^{\mathrm{th}}_{H_i}; T) \lesssim 1$ with the smallest threshold energy $E^{\mathrm{th}}_{H_i}$ of any disintegration reaction involving H_i, the hadron H_i is stopped before being able to induce any hadrodisintegration reaction. On the other hand, if $R^{H_i}(E^{(\mathrm{in})}_{H_i}, E^{\mathrm{th}}_{H_i}; T) > 1$,[36] the hadrons are not fully stopped and can therefore alter the primordial light-element abundances. Using R^{H_i}, we can therefore determine the energy $\tilde{E}^{(R=1)}_{H_i}$ of the energetic hadron H_i before it engages in disintegration reactions, i.e. by solving

$$R^{H_i}\left(E^{(\mathrm{in})}_{H_i}, \tilde{E}^{(R=1)}_{H_i}; T\right) = 1. \tag{D.5}$$

For stable hadrons, the interaction rate coincides with the scattering rate, $\Gamma_{H_i + A_j \to A_k} = n_{A_j} \sigma_{H_i + A_j \to A_k} \beta_{H_i}$. However, for unstable hadrons like neutrons, there exists an additional contribution from the decay rate, e.g. $\Gamma_{n \to p} = (\gamma_n \tau_n)^{-1}$ with $\gamma_n = E_n/m_n$ for neutrons.[37]

Finally, the equations presented above can be used to construct the matrix $\mathbf{M}_{\mathrm{eloss}}$ by implementing the following procedure: for any given entry $j < N$ ($N \le j < 2N$) of $\underline{A}$, we set $H_i = p$ ($H_i = n$) and $E^{(\mathrm{in})}_{H_i} = K_j$, where K_j is the central value of the kinetic energy bin corresponding to the entry j.[38] Afterwards, we solve eq. (D.5) to obtain $\tilde{E}^{(R=1)}_{H_i} \equiv K_k$, corresponding to a different energy bin $k(j)$, which explicitly depends on j, i.e. the energy and identity of the incoming particle, but also other parameters like the temperature T (cf. eq. (D.5)). Based on this result, the corresponding entries of the transfer matrix are then given by

$$(\mathbf{M}_{\mathrm{eloss}})_{ij} = \delta_{i,k(j)} \qquad \text{for} \quad j < 2N \quad \text{if} \quad K_k > K_{\min}. \tag{D.6}$$

In addition, since the EM energy loss does not lead to the production or destruction of any nuclei, we have

$$(\mathbf{M}_{\mathrm{eloss}})_{ij} = \delta_{i,j} \qquad \text{for} \quad 2N \le j < 2N + 6. \tag{D.7}$$

Note that the so-defined matrix is strictly triangular, since the energy-loss mechanism never increases the energy of the initial particle.

---

[36]In this case, the interpretation as the number of interactions starts to break down, since hadronic scattering will also contribute to the energy loss.

[37]In KKM05, the neutron decay is incorporated into the energy loss instead; however, we checked that this leads to only small numerical deviations in the final result.

[38]Note that for $j' < N$ both $j = j'$ and $j = j' + N$ share the same value $K_{j'} = K_{j'+N}$, with $A_{j'}$ ($A_{j'+N}$) counting the number of protons (neutrons) with this energy.
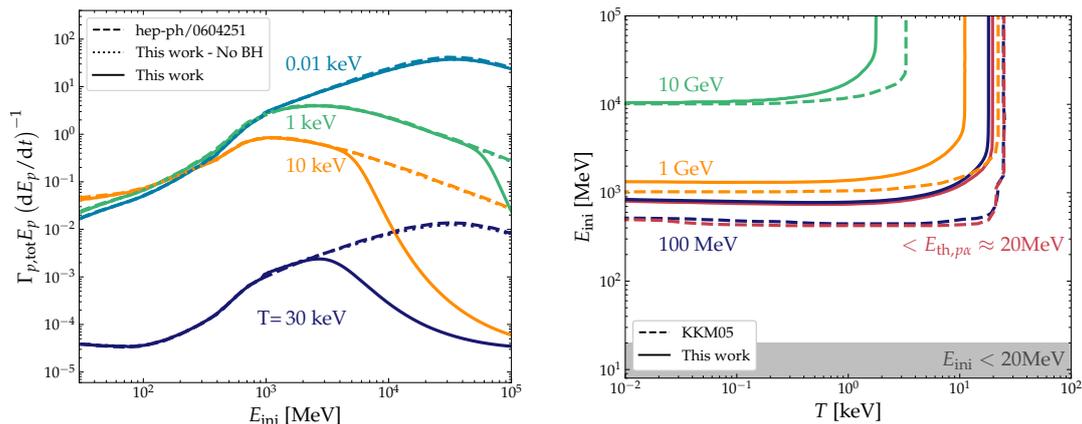
Figure 15. *Left:* the energy-loss term in eq. (D.2) as obtained from our implementation based on KKM05 (dash-dotted). For comparison, we also show the results from figure 1 of [51] (dashed), which we are able to reproduce with good precision. For higher energies, the agreement gets spoiled, since the Bethe-Heitler process becomes dominant, which has been ignored in [51] (see the solid line). *Right:* reproduction of figure 15 in KKM05. The discrepancy is more pronounced for smaller energies and higher temperatures.

A note on [31]. As it turns out, we do not seem to be able to reproduce the stopping power of charged particles shown in figures 15–20 of KKM05. This is despite the fact that the corresponding expressions used to calculate the stopping power in KKM05 are in agreement with other literature results, e.g. in [51]. In figure 15 (right), we illustrate this difference by showing the analogue of figure 15 in KKM05, i.e. the contours of constant $\tilde{E}_p^{(R=1)}$ from eq. (D.5) for protons as a function of the temperature T and the injection energy $E_{\rm ini} = E_p^{\rm (in)}$. Most notably, we find that the minimum starting energy a particle can have in order to interact before dropping below the disintegration threshold for helium-4 is significantly higher in our calculation.

To analyse and understand this discrepancy, we have cross-checked our own implementation of $\mathrm{d}E_p/\mathrm{d}t$ from eq. (D.2) with the findings of [51], in particular figure 1 of that work. To this end, in figure 15 (left), we show a comparison between our results and those from [51] for the quantity $\Gamma_{p,\rm tot} E_p (\mathrm{d}E_p/\mathrm{d}t)^{-1}$ where $\Gamma_{p,\rm tot}$ is the total scattering rate of protons. We find that our implementation, which after all is based on KKM05, accurately matches the results of [51], except for the high-energy regime, for which the authors of [51] do not take into account the contribution from Bethe-Heitler scattering. However, artificially removing this terms from our calculation, again leads to a good match between the two results.[39] Overall, we therefore conclude that our implementation is sound, while the discrepancy most likely comes from a difference in the Coulomb terms, as this would explain why our results coincide with those of KKM05 in the case of neutron. However, since our results for $\mathrm{d}E_p/\mathrm{d}t$ agree with those in [51] — while KKM05 does not show a similar plot — we trust our own implementation and therefore use it for this work.

---

[39]Note the visible (but numerically irrelevant) bump in the curves around $1\,{\rm GeV}$, which comes from transition between the ultra- and non-relativistic regime in the expressions of KKM05.

D.3 Hadronic scattering reactions

As already discussed earlier, if $R^{H_i}(E_{H_i}^{(\text{in})}, E_{H_i}^{\text{th}}; T) > 1$, hadronic scattering reactions become relevant. In this case, the α_{BG} targets can get destroyed (cf. table 3), thus producing additional D, T and/or ^{3}He nuclei, as well as additional protons and neutrons, which are potentially able to destroy further elements in the next cascade step. To fully consider this effect, it is imperative to follow the different daughter particles throughout the hadronic cascade. In this section, we provide a brief description of the numerical treatment that we employ for this task. For more details, we again refer the reader to the reference papers KKM05 and KKMT18.

Ultimately, the full treatment of the hadronic cascade proceeds in two distinct steps. At first, protons and neutrons lose part of the energy due to EM interactions until $R^{H_i} = 1$.[40] Afterwards, the different projectile particles with energy E_i undergo scattering reactions with the possible target particles, whereby the probabilities $P_{(i,j;k)}(E_i; T)$ for each of the processes (i, j, k) in table 3 to happen is proportional to the corresponding interaction rate, i.e.

$$P_{(i,j;k)}(E_i; T) \equiv \frac{\Gamma_{H_i + A_j \to A_k}(E_i; T)}{\sum_{A_j = p_{\text{BG}}, \alpha_{\text{BG}}} \sum_{A_k} \Gamma_{H_i + A_j \to A_k}(E_i; T) + \Gamma_{H_i \to \dots}(E_i; T)}. \tag{D.8}$$

Here, k labels the set of final-state particles, while i and j label the initial-state projectile and target particle, respectively. Note that we also include the decay rate $\Gamma_{H_i \to \dots}$ in the denominator, which, however, is only non-zero for unstable particles, i.e. neutrons. The nuclear cross-sections driving these rates can be found in KKM05.

For describing the redistribution of energy due to these processes, which is imperative for calculating $\mathbf{M}_{\text{scat}}$, we track the spectra of all final-state particle $A_{k,l}$, where the index l differentiates the different nuclei in the multi-particle state A_k. To this end, for a given initial state, we calculate the energies of the resulting final-states particles according to the prescription used in appendix C of KKM05. Notably, if $A_{k,l}$ is neither a proton nor a neutron, it is important to check if the produced nucleus survives. To this end, we perform a simplified test by evaluating the equations

$$E_{A_{k,l}}^{(R=1)} \le E_{A_{k,l}}^{\text{th},p} \qquad \text{and} \qquad \sqrt{3TE_{A_{k,l}}} \le Q_{A_{k,l}}. \tag{D.9}$$

Here, $E_{A_{k,l}}$ is the energy of the particle $A_{k,l}$, $Q_{A_{k,l}}$ is the threshold energy for photodisintegration of $A_{k,l}$ due to a CMB photon and $E_{A_{k,l}}^{\text{th},p}$ is the threshold energy for hadrodisintegration of $A_{k,l}$ due to a background proton. If at least one of these conditions is violated, we assume that the nucleus is fully destroyed and we re-inject its nucleons into the cascade assuming equipartition of their momentum. Consequently, the survival probability of $A_{k,l}$ is given by

$$P_{A_{j,k}}^{\text{surv}}(E_{A_{j,k}}; T) = \begin{cases} 0, & \text{if eq. (D.9) is violated} \\ 1, & \text{otherwise} \end{cases}. \tag{D.10}$$

As a result, even if the final-state particles in table 3 do not contain neutrons or protons, the reinjection of nucleons due to the destruction of heavier nuclei can still give rise to additional projectile particles.

---

[40]Notably, some particles are slowed down below threshold without ever starting a cascade.

Based on this setup, we calculate the different entries $(\mathbf{M}_{\text{scat}})_{ij}$ of the transition matrix in the following way

1. Split the transition matrix into different components for the different nuclear reactions, weighted by their respective scattering probability, i.e.

$$\mathbf{M}_{\text{scat}} = \sum_{i,j,k} P_{(i,j;k)} \mathbf{M}_{\text{scat}}^{(i,j;k)} \,. \tag{D.11}$$

2. For a given process $(i,j;k)$, consider each energy bin α with central kinetic energy K_α for the projectile particle and calculate the spectra $f_p(K; K_\alpha)$ and $f_n(K; K_\alpha)$ of the outgoing protons and neutrons according to appendix C of KKM05. Afterwards, set the different entries $(\mathbf{M}_{\text{scat}}^{(i,j;k)})_{\beta\alpha}$ with $K_{\min} < K_\beta < K_\alpha$ such that the entries $\underline{N}_p$ and $\underline{N}_n$ of $\underline{A}' = \mathbf{M}_{\text{scat}}^{(i,j;k)}\underline{A}$ — with $\underline{A}$ encoding the state directly after the injection of a single proton or neutron with energy K_α — are distributed according to the spectra $f_p(K_\beta; K_\alpha)$ and $f_n(K_\beta; K_\alpha)$. Additionally, set the entries $(\mathbf{M}_{\text{scat}}^{(i,j;k)})_{\beta\alpha}$ with $2N \le \beta < 2N + 2$ such that ΔN_p and ΔN_n of $\underline{A}'$ represent the change in the number of protons and neutrons.

3. For each process $(i,j;k)$, also calculate the energies of the outgoing D, T, ^{3}He, and ^{4}He. For each nucleus that survives (cf. eq. (D.9)), set the entries of $(\mathbf{M}_{\text{scat}}^{(i,j;k)})_{\beta\alpha}$ with $2N+2 \le \beta < 2N + 6$ such that ΔN_{D} etc. reflect the change in nuclei (including the destruction of a potential α_{BG} in the initial state). Instead, if the nucleus does not survive, ensure that the individual nucleon fragments are represented in the different entries of $\underline{A}'$.

Notably, this is only a sketch for the calculation of $\mathbf{M}_{\text{scat}}$. For more details on the exact implementation, we also refer the reader to the publicly-available source code of `v2.0.0-dev` of `ACROPOLIS`.

D.4 Extracting ξ_X^y

Following the previous discussion, we are interested in extracting the parameters ξ_X^y for the Boltzmann equation (4.14).

As already mentioned before, we can use $\mathbf{M}_{\text{eloss}}$ and $\mathbf{M}_{\text{scat}}$ to calculate the final vector $\underline{A}_f$ at the end of the cascade, i.e.

$$\underline{A}_f = \lim_{n\to\infty} (\mathbf{M}_{\text{scat}}\mathbf{M}_{\text{eloss}})^n \underline{A}_0 \equiv \mathbf{T}\underline{A}_0 \,. \tag{D.12}$$

Here, n denotes the number of cascade steps and $\underline{A}_0$ is the initial vector describing the injection (cf. section D.1). Since both $\mathbf{M}_{\text{eloss}}$ and $\mathbf{M}_{\text{scat}}$, are strictly triangular, the same is true also for the final transfer matrix $\mathbf{T}$. For $n \to \infty$, the matrix $\mathbf{T}$ is mostly zero, except for the entries T_{ij} with $i \in [2N, \ldots, (2N + 6) - 1]$ and $j \in [1, \ldots, 2N - 1]$ (*remember that we start indexing at 0*). In practice, we set n to a sufficiently large number, such that all other entries might not be zero but negligible. In fact, since the matrix multiplication is very fast, we can easily ensure convergence.[41]

---

[41]For best performance, note that

$$(\mathbf{M}_{\text{scat}}\mathbf{M}_{\text{eloss}})^{2^n} = \left[(\mathbf{M}_{\text{scat}}\mathbf{M}_{\text{eloss}})^2\right]^n \,,$$

meaning that n multiplications are sufficient to simulate 2^n cascade steps.

Given the transfer matrix $\mathbf{T}$, we can then extract the different parameters ξ_X^y from eq. (4.14). More precisely, we can identify (omitting the T dependence)

$$\xi_p^p(K_i) = T_{2N+0,2i}, \qquad \xi_p^n(K_i) = T_{2N+0,2i+N} \tag{D.13}$$

$$\xi_n^p(K_i) = T_{2N+1,2i}, \qquad \xi_n^n(K_i) = T_{2N+1,2i+N} \tag{D.14}$$

$$\xi_D^p(K_i) = T_{2N+2,2i}, \qquad \xi_D^n(K_i) = T_{2N+2,2i+N} \tag{D.15}$$

$$\xi_T^p(K_i) = T_{2N+3,2i}, \qquad \xi_T^n(K_i) = T_{2N+3,2i+N} \tag{D.16}$$

$$\xi_{^3\mathrm{He}}^p(K_i) = T_{2N+4,2i}, \quad \xi_{^3\mathrm{He}}^n(K_i) = T_{2N+4,2i+N} \tag{D.17}$$

$$\xi_{^4\mathrm{He}}^p(K_i) = T_{2N+5,2i}, \quad \xi_{^4\mathrm{He}}^n(K_i) = T_{2N+5,2i+N}. \tag{D.18}$$

Here, $i \in [0, \ldots, N-1]$ runs over the different energy bins with central energy K_i. Using this identification, we can implement the hadronic cascade into `ACROPOLIS` by means of eq. (4.15). For consistency, we further explicitly check that baryon number is conserved. More precisely, we demand

$$\sum_X B_X \xi_X^y = 1, \tag{D.19}$$

with the baryon number B_X of the nucleus X and $y \in \{n, p\}$, since the relation must be fulfilled for both nucleons separately. This equation is based on the interpretation of ξ_X^y as the number of produced/destroyed nuclei X per single injection of the initial nucleon y.

To conclude this part of the appendix, let us briefly discuss the results we obtain for one benchmark value of KKMT18, i.e. for $T = 4\,\mathrm{keV}$, $\eta = 6.1\cdot10^{-10}$, and $\mathcal{Y}_p = 0.25$. Here, we focus on the production of D and the destruction of ^{4}He, since we expect these contributions to have the largest impact on our final results. In the left part of figure 16, we show the deuterium yield (blue-green) per injected hadron for both neutrons (solid) and protons (dashed), and compare it to the corresponding results in KKMT18 (red). We find excellent agreement for the case of injected neutrons in the low-energy regime. For neutrons of higher energies as well as for injected protons in general, we instead observe small deviations, due to the discrepancy we find in the handling of the energy loss (cf. figure 15). This discrepancy becomes especially clear for the case of injected protons, for which the aforementioned difference in the Coulomb term yields a much stronger stopping power in the context of our own calculation. Therefore, injected protons are significantly less efficient in causing hadrodisintegration reactions and higher energies are required to cause a non-zero impact on the nuclear abundances. For neutrons of small energy, the discrepancy in the energy-loss terms only plays a sub-leading role, as they mainly interact via magnetic-moment scattering. Consequently, the (otherwise leading) Coulomb term is irrelevant. However, at higher energies, conversions from protons into neutrons (also cf. the upper part of table 3) become important (as also analysed in KKMT18), thus leading again to deviations in the results. Specifically, the protons produced by these conversion processes will experience the same increased stopping power that already caused the deviations in the case of injected protons. Therefore, the efficiency of neutron injection also decreases for higher energies, for which the inelastic (conversion) processes become more pronounced as shown in figures 11 and 12 of KKM05. This argument is fairly general and therefore directly applies also to the destruction of ^{4}He, as shown in the right
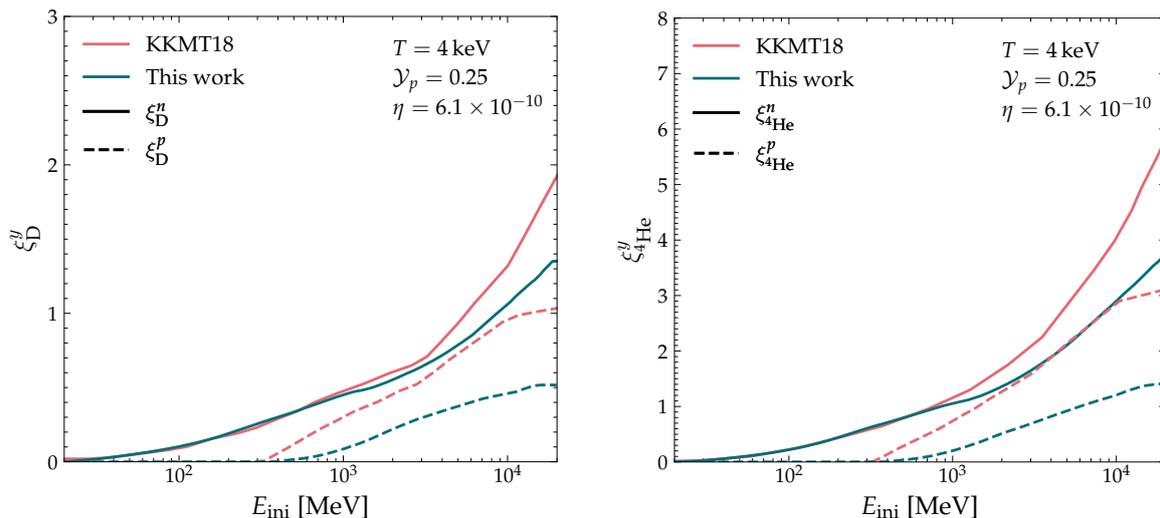
Figure 16. *Left:* deuterium yield (blue-green) per injected baryon for fixed temperature compared to the literature results (red). *Right:* the same on the left but for ^{4}He. Note that in this panel, we have flipped the sign of $\xi^y_{^4\mathrm{He}}$ in order for it to be positive.

part of figure 16. For better comparability, we have flipped the sign of our $\xi^y_{^4\mathrm{He}}$ in this plot, since it is negative for our convention while KKMT18 defines it as positive.

Finally, let us note that we employ one last approximation as part of the numerical implementation. In general, $\xi^y_X[n_j]$ depends non-linearly on the number density of the available background nuclei n_j. However, we have confirmed that to a reasonable approximation, some combinations $\xi^y_X[n_j]/n_A$ are basically constant. In fact, this statement is true specifically for $A = p$ in case $X \in \{p, n\}$ and $A = {}^4\mathrm{He}$ if $X \in \{\mathrm{D}, \mathrm{T}, {}^3\mathrm{He}, {}^4\mathrm{He}\}$. Therefore, we can approximate

$$\xi^y_X[n_j](T, K) \simeq \frac{\xi^y_X[n^0_j](T, K)}{n^0_{A(X)}} \times n_{A(X)}(T) \,, \tag{D.20}$$

with $A(X) = p$ if $X \in \{p, n\}$ and $A(X) = {}^4\mathrm{He}$ otherwise, which transforms eq. (4.14) from a non-linear to a linear differential equation. Here, $n^0_{j/A(X)}$ are the corresponding initial abundances at the start of NBBN. Using this approximation, within `ACROPOLIS` eq. (4.14) can therefore be solved in the same manner as eq. (4.13) (see the `ACROPOLIS` manual for more information).

E Boosting between the cosmic rest frame and the centre-of-mass frame

In this appendix, we derive the formula used in eq. (5.24). To this end, we introduce two 4-momenta for the two initial-state neutrinos of the interaction, which we parametrise by the corresponding scattering angle θ between both neutrinos in the cosmic rest frame, i.e.

$$p_1 = \begin{pmatrix} E_1 \\ 0 \\ 0 \\ E_1 \end{pmatrix} \quad \text{and} \quad p_2 = \begin{pmatrix} E_2 \\ 0 \\ E_2 \sin\theta \\ E_2 \cos\theta \end{pmatrix} . \tag{E.1}$$

Here, E_1 and E_2 are the energies of the initial-state neutrinos. At this point, it is useful to introduce the Mandelstam variable

$$s = (p_1 + p_2)^2 = 2E_1 E_2 \left(1 - \cos\theta\right) \quad \Leftrightarrow \quad \cos\theta = 1 - \frac{s}{2E_1 E_2}, \tag{E.2}$$

for $m_\nu = 0$, which is independent of the reference frame.[42] Given the 4-momenta from above, the total 4-momentum of the system in the cosmic rest frame is given by

$$p_1 + p_2 = \begin{pmatrix} E_1 + E_2 \\ 0 \\ E_2 \sin\theta \\ E_1 + E_2 \cos\theta \end{pmatrix}. \tag{E.3}$$

Since we handle non-thermal scattering in the centre-of-mass (CM) frame, we have to boost $p_1 + p_2$ into this exact frame in order to use the results of PYTHIA8.3. To this end, we first determine the corresponding Lorentz transformation Λ_z in z-direction. Specifically, the boost velocity β_z in this direction is given by

$$\beta_z = \frac{E_1 + E_2 \cos\theta}{E_1 + E_2}, \tag{E.4}$$

which yields

$$\Lambda_z(p_1 + p_2) = \begin{pmatrix} \sqrt{(1 - \cos\theta)\, E_2\, (2E_1 + E_2\, (1 + \cos\theta))} \\ 0 \\ E_2 \sin\theta \\ 0 \end{pmatrix}. \tag{E.5}$$

Similarly, for the Lorentz transformation Λ_y in y-direction the boost is given by

$$\beta_y = \sqrt{\frac{(1 + \cos\theta)E_2}{2E_1 + E_2(1 + \cos\theta)}}, \tag{E.6}$$

as it leads to the desired result

$$\Lambda_y \Lambda_z(p_1 + p_2) = \begin{pmatrix} \sqrt{2E_1 E_2 \left(1 - \cos\theta\right)} \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{E.7}$$

We now consider the 4-momentum of a generic baryon in the centre-of-mass frame resulting from non-thermal scattering, i.e.

$$p_{b,\mathrm{cm}} = \begin{pmatrix} E_b \\ p_b \sin\theta_b \cos\phi_b \\ p_b \sin\theta_b \sin\phi_b \\ p_b \cos\theta_b \end{pmatrix}. \tag{E.8}$$

---

[42]In our convention, scattering with $\theta = 0$ is impossible and therefore not considered in the following.

Here, E_b and p_b are the energy and momentum of the baryon, while θ_b and ϕ_b are two angles describing its traversal direction. By applying the inverse transformation $\Lambda_z^{-1}\Lambda_y^{-1}$, we obtain the corresponding 4-momentum in the cosmic rest frame,

$$p_{b,\mathrm{lab}} = \Lambda_z^{-1}\Lambda_y^{-1} p_{b,\mathrm{cm}}\,, \tag{E.9}$$

which still depends on the angles defined in the CM frame. Assuming an isotropic distribution of injected baryons, we perform the average

$$\begin{aligned}
\langle p_{b,\mathrm{lab}}\rangle &= \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}\mathrm{d}\cos\theta_b\mathrm{d}\phi_b\,\Lambda_z^{-1}\Lambda_y^{-1}p_{b,\mathrm{cm}}\\
&= \frac{\Lambda_z^{-1}\Lambda_y^{-1}}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}\mathrm{d}\cos\theta_b\mathrm{d}\phi_b\,p_{b,\mathrm{cm}}\\
&= \Lambda_z^{-1}\Lambda_y^{-1}\begin{pmatrix} E_b \\ 0 \\ 0 \\ 0 \end{pmatrix}\,,
\end{aligned} \tag{E.10}$$

where we have used the linearity of the Lorentz transformation. Ultimately, we are only interested in the temporal component of this vector,

$$E_{b,\mathrm{lab}} = \frac{E_1 + E_2}{\sqrt{s}} E_b\,, \tag{E.11}$$

which is the exact expression used in eq. (5.24).

F   A collection of supplementary figures

In this appendix, we collect some figures that supplement our selection of plots in the main text. While we briefly comment on each of them here, we refrain from a complete discussion due to the auxiliary nature of these figures.

In figure 17, we show the analogue of figure 3 but for intermediate case $m_\phi = 200\,\mathrm{GeV}$. For this mass, the resulting constraints are just at the transition between being dominated by non-thermal scattering for smaller masses and FSR for higher masses. Qualitatively, we find it to be very similar to the cases that have already been discussed in greater detail in section 2.2.

In figure 18, we compare in the $m_\phi - f_\phi$ plane *(i)* the numerical results obtained in this work (full lines), and *(ii)* the resulting EM limits by assuming only FSR (dotted lines). We find that for particles heavier than the EW scale that are sufficiently long-lived, case *(ii)* provides an excellent approximation, as all other effects are subleading.[43] Moreover, we extend the line from case *(ii)* to masses beyond the ones considered in this work, in order to provide a projection for even heavier long-lived particles.

Finally, in figure 19, we show some of the auxiliary quantities that have been used in section 5. More precisely, we visualise ζ_{hd} (upper left), N_{hd} (upper right), and K_{hd} (lower

---

[43]This statement is mildly lifetime-dependent, because the EM part of the EW shower from FSR has a maximum contribution at around $\tau_\phi \sim 10^8$-10^9 s and for values close to this peak, the dominance of the limits happens at even smaller masses. This becomes apparent by comparing the lines of $\tau_\phi = 10^9$ s and $\tau_\phi = 10^{11}$ s.
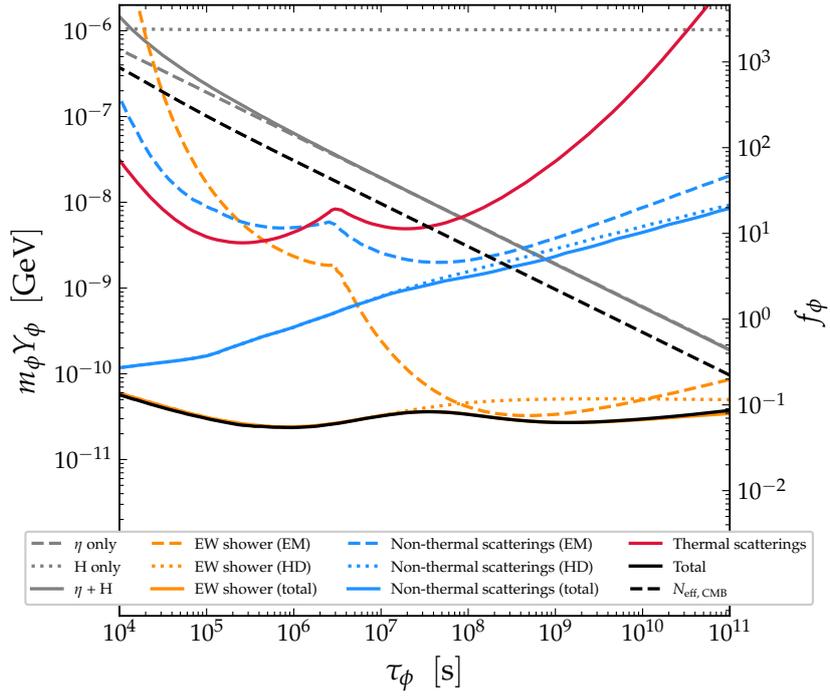
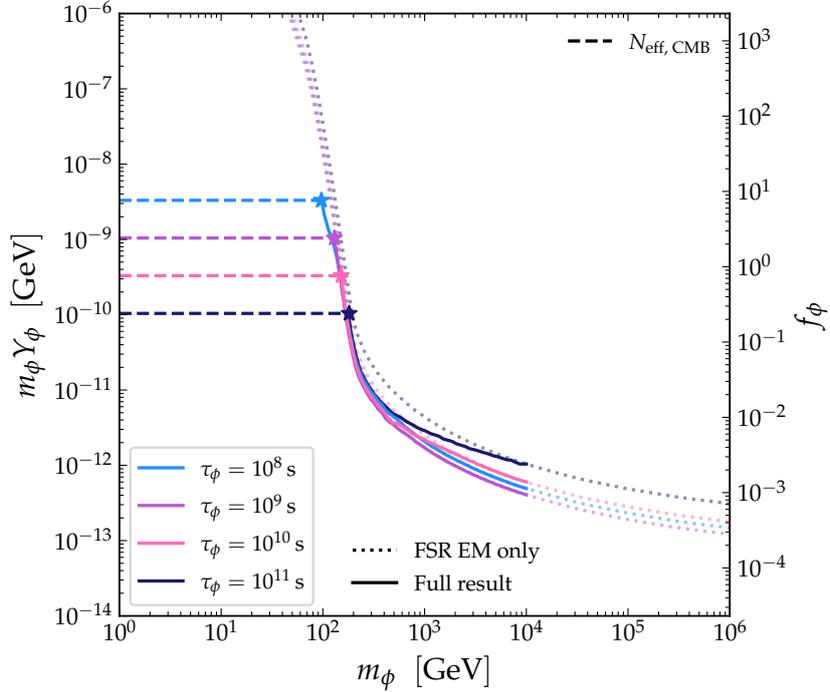**Figure 17.** Similar to figure 3, but for $m_\phi = 200\,\mathrm{GeV}$.

Figure 18. Comparison between the full numerical result for different (large) lifetimes with the one for direct e^+e^- injections from FSR, i.e. suppressed by a factor $\zeta_{\mathrm{em}}^{\mathrm{fsr}}(m_\phi)$.
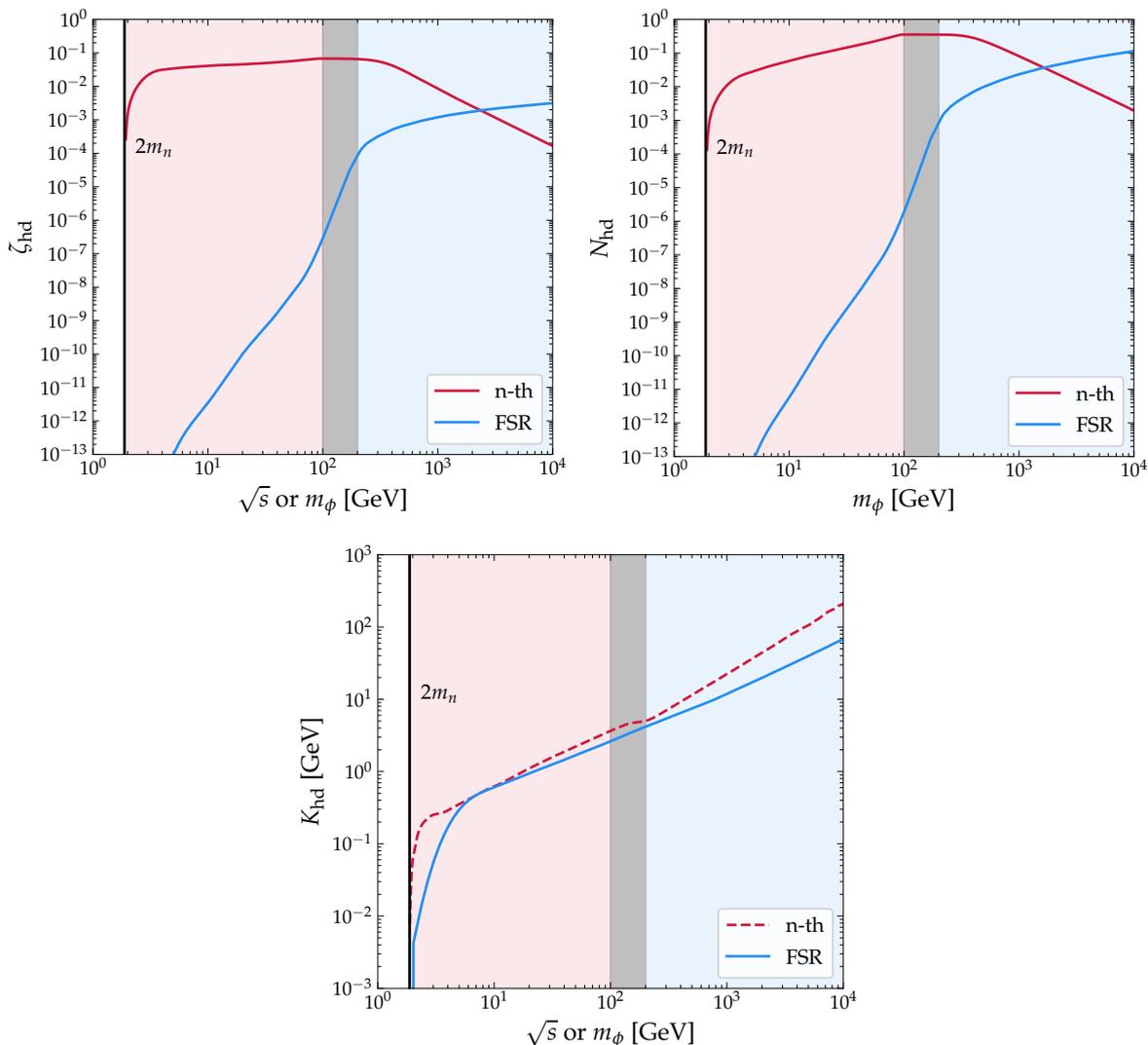
Figure 19. Additional quantities describing the hadronic injection (cf. section 5), i.e. ζ_{hd} (upper left), N_{hd} (upper right), and K_{hd} (lower centre), for non-thermal scattering (red, $x = \sqrt{s}$) and FSR (blue, $x = m_\phi$). In the case of non-thermal scattering, the y-axis for a given quantity Q_{hd} technically shows the averaged quantity $\langle Q_{\mathrm{hd}}^{\mathrm{n\text{-}th}} \rangle$. We shade the regions of energy/mass where a given effect is dominant with the corresponding colour, with the grey band indicating the transition region where both effects can be equally important. Note that for K_{hd} from non-thermal scattering, we show the value in the centre-of-mass frame, since the corresponding value in the cosmic rest frame explicitly depends on the energies of the initial-state particles (cf. eq. (5.24)).

centre) originating from both non-thermal scattering (red, $x = \sqrt{s}$) and FSR (blue, $x = m_\phi$). In general, we find that $\zeta_{\mathrm{hd}}^{\mathrm{fsr}}$ and $N_{\mathrm{hd}}^{\mathrm{fsr}}$ behave qualitatively similar to $\zeta_{\mathrm{em}}^{\mathrm{fsr}}$ from figure 4. In addition, the average kinetic energies $K_{\mathrm{hd}}^{\mathrm{fsr}}$ and $\langle K_{\mathrm{hd}}^{\mathrm{n\text{-}th}} \rangle$ grow roughly linearly beyond the nucleon threshold. Instead, for the quantities $\langle \zeta_{\mathrm{hd}}^{\mathrm{n\text{-}th}} \rangle$ and $\langle N_{\mathrm{hd}}^{\mathrm{n\text{-}th}} \rangle$ from non-thermal scattering, we observe larger contributions around the EW scale with a strong decline beyond it.

References

[1] B. Batell, M. Pospelov and A. Ritz, *Exploring Portals to a Hidden Sector Through Fixed Targets*, *Phys. Rev. D* **80** (2009) 095024 [`arXiv:0906.5614`] [InSPIRE].

[2] J.L. Feng, M. Kaplinghat and H.-B. Yu, *Halo Shape and Relic Density Exclusions of Sommerfeld-Enhanced Dark Matter Explanations of Cosmic Ray Excesses*, *Phys. Rev. Lett.* **104** (2010) 151301 [`arXiv:0911.0422`] [InSPIRE].

[3] M.R. Buckley and P.J. Fox, *Dark Matter Self-Interactions and Light Force Carriers*, *Phys. Rev. D* **81** (2010) 083522 [`arXiv:0911.3898`] [InSPIRE].

[4] A. Loeb and N. Weiner, *Cores in Dwarf Galaxies from Dark Matter with a Yukawa Potential*, *Phys. Rev. Lett.* **106** (2011) 171302 [`arXiv:1011.6374`] [InSPIRE].

[5] S. Tulin, H.-B. Yu and K.M. Zurek, *Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure*, *Phys. Rev. D* **87** (2013) 115007 [`arXiv:1302.3898`] [InSPIRE].

[6] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg and P. Walia, *Strong constraints on self-interacting dark matter with light mediators*, *Phys. Rev. Lett.* **118** (2017) 141802 [`arXiv:1612.00845`] [InSPIRE].

[7] T. Hambye and L. Vanderheyden, *Minimal self-interacting dark matter models with light mediator*, *JCAP* **05** (2020) 001 [`arXiv:1912.11708`] [InSPIRE].

[8] T.R. Slatyer, *Energy Injection And Absorption In The Cosmic Dark Ages*, *Phys. Rev. D* **87** (2013) 123513 [`arXiv:1211.0283`] [InSPIRE].

[9] V. Poulin, J. Lesgourgues and P.D. Serpico, *Cosmological constraints on exotic injection of electromagnetic energy*, *JCAP* **03** (2017) 043 [`arXiv:1610.10051`] [InSPIRE].

[10] T. Kanzaki, M. Kawasaki, K. Kohri and T. Moroi, *Cosmological Constraints on Neutrino Injection*, *Phys. Rev. D* **76** (2007) 105017 [`arXiv:0705.1200`] [InSPIRE].

[11] T.R. Slatyer and C.-L. Wu, *General Constraints on Dark Matter Decay from the Cosmic Microwave Background*, *Phys. Rev. D* **95** (2017) 023010 [`arXiv:1610.06933`] [InSPIRE].

[12] P. Stöcker, M. Krämer, J. Lesgourgues and V. Poulin, *Exotic energy injection with ExoCLASS: Application to the Higgs portal model and evaporating black holes*, *JCAP* **03** (2018) 018 [`arXiv:1801.01871`] [InSPIRE].

[13] H. Liu, G.W. Ridgway and T.R. Slatyer, *Code package for calculating modified cosmic ionization and thermal histories with dark matter and other exotic energy injections*, *Phys. Rev. D* **101** (2020) 023530 [`arXiv:1904.09296`] [InSPIRE].

[14] M. Cirelli et al., *PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection*, *JCAP* **03** (2011) 051 [*Erratum ibid.* **10** (2012) E01] [`arXiv:1012.4515`] [InSPIRE].

[15] T. Hambye, M. Hufnagel and M. Lucca, *Cosmological constraints on the decay of heavy relics into neutrinos*, *JCAP* **05** (2022) 033 [`arXiv:2112.09137`] [InSPIRE].

[16] S.K. Acharya and R. Khatri, *Constraints on N_{eff} of high energy non-thermal neutrino injections upto $z \sim 10^8$ from CMB spectral distortions and abundance of light elements*, *JCAP* **11** (2020) 011 [`arXiv:2007.06596`] [InSPIRE].

[17] M. Ovchynnikov and V. Syvolap, *Primordial Neutrinos and New Physics: Novel Approach to Solving the Neutrino Boltzmann Equation*, *Phys. Rev. Lett.* **134** (2025) 101003 [`arXiv:2409.15129`] [InSPIRE].

[18] K. Akita et al., *New Physics Decaying into Metastable Particles: Impact on Cosmic Neutrinos*, *Phys. Rev. Lett.* **134** (2025) 121001 [`arXiv:2411.00892`] [ɪɴSPIRE].

[19] K. Akita et al., *Dynamics of metastable standard model particles from long-lived particle decays in the MeV primordial plasma*, *Phys. Rev. D* **111** (2025) 063542 [`arXiv:2411.00931`] [ɪɴSPIRE].

[20] M. Kawasaki and T. Moroi, *Gravitino decay into a neutrino and a sneutrino in the inflationary universe*, *Phys. Lett. B* **346** (1995) 27 [`hep-ph/9408321`] [ɪɴSPIRE].

[21] P.F. Depta, M. Hufnagel and K. Schmidt-Hoberg, *ACROPOLIS: A generiC fRamework fOr Photodisintegration Of LIght elementS*, *JCAP* **03** (2021) 061 [`arXiv:2011.06518`] [ɪɴSPIRE].

[22] S. Chang et al., *Constraining MeV to 10 GeV Majorons by big bang nucleosynthesis*, *Phys. Rev. D* **110** (2024) 015019 [`arXiv:2401.00687`] [ɪɴSPIRE].

[23] J.A. Formaggio and G.P. Zeller, *From eV to EeV: Neutrino Cross Sections Across Energy Scales*, *Rev. Mod. Phys.* **84** (2012) 1307 [`arXiv:1305.7513`] [ɪɴSPIRE].

[24] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [`arXiv:1410.3012`] [ɪɴSPIRE].

[25] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, *SciPost Phys. Codeb.* **2022** (2022) 8 [`arXiv:2203.11601`] [ɪɴSPIRE].

[26] J.A. Frieman and G.F. Giudice, *Cosmologically Benign Gravitinos at the Weak Scale*, *Phys. Lett. B* **224** (1989) 125 [ɪɴSPIRE].

[27] J. Gratsias, R.J. Scherrer and D.N. Spergel, *Indirect photofission of light elements from high-energy neutrinos in the early universe*, *Phys. Lett. B* **262** (1991) 298 [ɪɴSPIRE].

[28] A.A. de Laix and R.J. Scherrer, *Improved cosmological constraints on neutrino producing decaying particles*, *Phys. Rev. D* **48** (1993) 562 [ɪɴSPIRE].

[29] M. Kawasaki and T. Moroi, *Electromagnetic cascade in the early universe and its application to the big bang nucleosynthesis*, *Astrophys. J.* **452** (1995) 506 [`astro-ph/9412055`] [ɪɴSPIRE].

[30] M. Hufnagel, K. Schmidt-Hoberg and S. Wild, *BBN constraints on MeV-scale dark sectors. Part II. Electromagnetic decays*, *JCAP* **11** (2018) 032 [`arXiv:1808.09324`] [ɪɴSPIRE].

[31] M. Kawasaki, K. Kohri and T. Moroi, *Big-Bang nucleosynthesis and hadronic decay of long-lived massive particles*, *Phys. Rev. D* **71** (2005) 083502 [`astro-ph/0408426`] [ɪɴSPIRE].

[32] M. Hufnagel, K. Schmidt-Hoberg and S. Wild, *BBN constraints on MeV-scale dark sectors. Part I. Sterile decays*, *JCAP* **02** (2018) 044 [`arXiv:1712.03972`] [ɪɴSPIRE].

[33] M. Escudero, *Neutrino decoupling beyond the Standard Model: CMB constraints on the Dark Matter mass with a fast and precise N_{eff} evaluation*, *JCAP* **02** (2019) 007 [`arXiv:1812.05605`] [ɪɴSPIRE].

[34] J. Alwall et al., *MadGraph 5: Going Beyond*, *JHEP* **06** (2011) 128 [`arXiv:1106.0522`] [ɪɴSPIRE].

[35] PARTICLE DATA GROUP collaboration, *Review of particle physics*, *Phys. Rev. D* **110** (2024) 030001 [ɪɴSPIRE].

[36] J. Geiss and G. Gloeckler, *Isotopic Composition of H, He and Ne in the Protosolar Cloud*, in *Solar System History from Isotopic Signatures of Volatile Elements*, R. Kallenbach et al. eds., Springer Netherlands (2003), p. 3–18 [`DOI:10.1007/978-94-010-0145-8_1`].

[37] A. Korn, *The ups and downs of inferred cosmological lithium*, *EPJ Web Conf.* **297** (2024) 01007 [`arXiv:2406.09974`] [ɪɴSPIRE].

[38] GALAH collaboration, *The GALAH Survey: A new constraint on cosmological lithium and Galactic lithium evolution from warm dwarf stars*, *Mon. Not. Roy. Astron. Soc.* **497** (2020) L30 [arXiv:2006.05173] [inSPIRE].

[39] A. Arbey, *AlterBBN: A program for calculating the BBN abundances of the elements in alternative cosmologies*, *Comput. Phys. Commun.* **183** (2012) 1822 [arXiv:1106.1363] [inSPIRE].

[40] A. Arbey, J. Auffinger, K.P. Hickerson and E.S. Jenssen, *AlterBBN v2: A public code for calculating Big-Bang nucleosynthesis constraints in alternative cosmologies*, *Comput. Phys. Commun.* **248** (2020) 106982 [arXiv:1806.11095] [inSPIRE].

[41] S. Gariazzo, P. F. de Salas, O. Pisanti and R. Consiglio, *PArthENoPE revolutions*, *Comput. Phys. Commun.* **271** (2022) 108205 [arXiv:2103.05027] [inSPIRE].

[42] O. Pisanti, G. Mangano, G. Miele and P. Mazzella, *Primordial Deuterium after LUNA: concordances and error budget*, *JCAP* **04** (2021) 020 [arXiv:2011.11537] [inSPIRE].

[43] I. Tišma et al., *Experimental cross section and angular distribution of the $^2H(p,\gamma)^3He$ reaction at Big-Bang nucleosynthesis energies*, *Eur. Phys. J. A* **55** (2019) 137 [inSPIRE].

[44] A. Tumino et al., *New Determination of the $^2H(d,p)^3H$ and $^2H(d,n)^3He$ Reaction Rates at Astrophysical Energies*, *Astrophys. J.* **785** (2014) 96.

[45] V. Mossa et al., *The baryon density of the Universe from an improved rate of deuterium burning*, *Nature* **587** (2020) 210 [inSPIRE].

[46] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [arXiv:1807.06209] [inSPIRE].

[47] P.F. Depta, M. Hufnagel and K. Schmidt-Hoberg, *Robust cosmological constraints on axion-like particles*, *JCAP* **05** (2020) 009 [arXiv:2002.08370] [inSPIRE].

[48] P.F. Depta, M. Hufnagel and K. Schmidt-Hoberg, *Updated BBN constraints on electromagnetic decays of MeV-scale particles*, *JCAP* **04** (2021) 011 [arXiv:2011.06519] [inSPIRE].

[49] V. Poulin and P.D. Serpico, *Loophole to the Universal Photon Spectrum in Electromagnetic Cascades and Application to the Cosmological Lithium Problem*, *Phys. Rev. Lett.* **114** (2015) 091101 [arXiv:1502.01250] [inSPIRE].

[50] V. Poulin and P.D. Serpico, *Nonuniversal BBN bounds on electromagnetically decaying particles*, *Phys. Rev. D* **91** (2015) 103007 [arXiv:1503.04852] [inSPIRE].

[51] K. Jedamzik, *Big bang nucleosynthesis constraints on hadronically and electromagnetically decaying relic neutral particles*, *Phys. Rev. D* **74** (2006) 103509 [hep-ph/0604251] [inSPIRE].

[52] M. Kawasaki, K. Kohri, T. Moroi and Y. Takaesu, *Revisiting Big-Bang Nucleosynthesis Constraints on Long-Lived Decaying Particles*, *Phys. Rev. D* **97** (2018) 023502 [arXiv:1709.01211] [inSPIRE].

[53] L. Angel, G. Arcadi, M.M.A. Paixão and F.S. Queiroz, *Updated BBN Bounds on Hadronic Injection in the Early Universe: The Gravitino Problem*, arXiv:2501.09120 [inSPIRE].

[54] A. Boyarsky, M. Ovchynnikov, N. Sabti and V. Syvolap, *When feebly interacting massive particles decay into neutrinos: The Neff story*, *Phys. Rev. D* **104** (2021) 035006 [arXiv:2103.09831] [inSPIRE].

[55] M. Ovchynnikov and V. Syvolap, *How new physics affects primordial neutrinos decoupling: Direct simulation Monte Carlo approach*, *Phys. Rev. D* **111** (2025) 063527 [arXiv:2409.07378] [inSPIRE].

[56] Y. Chikashige, R.N. Mohapatra and R.D. Peccei, *Are There Real Goldstone Bosons Associated with Broken Lepton Number?*, *Phys. Lett. B* **98** (1981) 265 [INSPIRE].

[57] J. Schechter and J.W.F. Valle, *Neutrino Decay and Spontaneous Violation of Lepton Number*, *Phys. Rev. D* **25** (1982) 774 [INSPIRE].

[58] C. Garcia-Cely and J. Heeck, *Neutrino Lines from Majoron Dark Matter*, *JHEP* **05** (2017) 102 [arXiv:1701.07209] [INSPIRE].

[59] M. Frigerio, T. Hambye and E. Masso, *Sub-GeV dark matter as pseudo-Goldstone from the seesaw scale*, *Phys. Rev. X* **1** (2011) 021026 [arXiv:1107.4564] [INSPIRE].

[60] R. Coy and T. Hambye, *Neutrino lines from DM decay induced by high-scale seesaw interactions*, *JHEP* **05** (2021) 101 [arXiv:2012.05276] [INSPIRE].

[61] M. Cirelli, N. Fornengo and A. Strumia, *Minimal dark matter*, *Nucl. Phys. B* **753** (2006) 178 [hep-ph/0512090] [INSPIRE].

[62] T. Hambye, F.-S. Ling, L. Lopez Honorez and J. Rocher, *Scalar Multiplet Dark Matter*, *JHEP* **05** (2009) 066 [*Erratum ibid.* **05** (2010) 066] [arXiv:0903.4010] [INSPIRE].

[63] S. Bottaro, *Closing the Window on WIMP Dark Matter*, Ph.D. Thesis, Scuola Normale Superiore, Pisa, Italy (2022) [https://hdl.handle.net/11384/132646].

[64] M. Sajjad Athar, A. Fatima and S.K. Singh, *Neutrinos and their interactions with matter*, *Prog. Part. Nucl. Phys.* **129** (2023) 104019 [arXiv:2206.13792] [INSPIRE].