



Weaker yet again: mass spectrum-consistent cosmological constraints on the neutrino lifetime

Joe Zhiyu Chen^{1,a}, Isabel M. Oldengott^{2,b}, Giovanni Pierobon^{1,c}, Yvonne Y. Y. Wong^{1,d}

¹ Sydney Consortium for Particle Physics and Cosmology, School of Physics, The University of New South Wales, Sydney, NSW 2052, Australia

² Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain, Chemin du cyclotron, 2, Louvain-la-Neuve 1348, Belgium

Received: 25 March 2022 / Accepted: 11 June 2022
© The Author(s) 2022

Abstract We consider invisible neutrino decay $\nu_H \rightarrow \nu_l + \phi$ in the ultra-relativistic limit and compute the neutrino anisotropy loss rate relevant for the cosmic microwave background (CMB) anisotropies. Improving on our previous work which assumed *massless* ν_l and ϕ , we reinstate in this work the daughter neutrino mass m_{ν_l} in a manner consistent with the experimentally determined neutrino mass splittings. We find that a nonzero m_{ν_l} introduces a *new* phase space factor in the loss rate Γ_T proportional to $(\Delta m_v^2/m_{\nu_H}^2)^2$ in the limit of a small squared mass gap between the parent and daughter neutrinos, i.e., $\Gamma_T \sim (\Delta m_v^2/m_{\nu_H}^2)^2 (m_{\nu_H}/E_\nu)^5 (1/\tau_0)$, where τ_0 is the ν_H rest-frame lifetime. Using a general form of this result, we update the limit on τ_0 using the Planck 2018 CMB data. We find that for a parent neutrino of mass $m_{\nu_H} \lesssim 0.1$ eV, the new phase space factor *weakens* the constraint on its lifetime by up to a factor of 50 if Δm_v^2 corresponds to the atmospheric mass gap and up to 10^5 if the solar mass gap, in comparison with naïve estimates that assume $m_{\nu_l} = 0$. The revised constraints are (i) $\tau^0 \gtrsim (6 \rightarrow 10) \times 10^5$ s and $\tau^0 \gtrsim (400 \rightarrow 500)$ s if only one neutrino decays to a daughter neutrino separated by, respectively, the atmospheric and the solar mass gap, and (ii) $\tau^0 \gtrsim (2 \rightarrow 6) \times 10^7$ s in the case of two decay channels with one near-common atmospheric mass gap. In contrast to previous, naïve limits which scale as $m_{\nu_H}^5$, these mass spectrum-consistent τ_0 constraints are remarkably independent of the parent mass and open up a swath of parameter space within the projected reach of IceCube and other neutrino telescopes in the next two decades.

1 Introduction

Precision measurements of the cosmic microwave background (CMB) anisotropies and the large-scale matter distribution have in the past two decades yielded some of the tightest constraints on particle physics. The most widely known amongst these are upper limits on the neutrino mass sum, which at $\sum m_\nu \lesssim 0.12$ eV for a one-parameter extension of the standard Λ CDM model [1],¹ supersedes at face value current laboratory β -decay end-point spectrum measurements by a factor of 30 [2]. Of lesser prominence albeit equal significance are CMB constraints on the neutrino lifetime.

Relativistic invisible neutrino decay occurring on time scales of CMB formation leave an interesting signature on the CMB anisotropies. In the cosmological context, Standard-Model neutrinos decouple at temperatures of $T \sim 1$ MeV and free-stream thereafter. Free-streaming could be inhibited, however, if non-standard collisional processes such as neutrino decay and especially its concurrent inverse decay—the latter of which is kinematically feasible only in the limit of an ultra-relativistic neutrino ensemble—should become efficient at $T \lesssim 1$ MeV. At CMB formation times, the absence of neutrino free-streaming prevents the transfer of power from the monopole (density) and dipole (velocity divergence) fluctuations of the neutrino fluid to higher multipole moments (e.g., anisotropic stress) [3]. Such a suppression of higher-order anisotropies manifests itself predominantly in the CMB primary anisotropies as an enhancement of power at multipoles $\ell \gtrsim 200$ in the CMB temperature power spectrum, which can be exploited to place constraints on the non-standard interaction(s) responsible for the phenomenon [4–16].

^a e-mail: zhiyu.chen@unsw.edu.au

^b e-mail: isabel.oldengott@uclouvain.be

^c e-mail: g.pierobon@unsw.edu.au (corresponding author)

^d e-mail: yvonne.y.wong@unsw.edu.au

¹ 95% credible limit derived from a 7-parameter fit to the data combination Planck TT,TE,EE+lowE+lensing+BAO [1].

Following this line of argument, measurements of the Planck mission (2018 data release) currently constrain the rest-frame neutrino lifetime to $\tau_0 \gtrsim (4 \times 10^5 \rightarrow 4 \times 10^6) (m_\nu/50 \text{ meV})^5 \text{ s}$, assuming invisible decay of a parent neutrino not exceeding $m_\nu \simeq 0.2 \text{ eV}$ in mass into a massless particles [9].² This lower bound is to be compared with laboratory limits obtained from analyses of the neutrino disappearance rate in JUNO and KamLAND+JUNO, $\tau_0/m_\nu > 7.5 \times 10^{-11} \text{ (s/eV)}$ and $\tau_0/m_\nu > 1.1 \times 10^{-9} \text{ (s/eV)}$ for a normal and an inverted mass ordering respectively [19], as well as astrophysical limits from solar neutrinos $\tau_0/m_\nu \gtrsim 10^{-5} \rightarrow 10^{-3} \text{ (s/eV)}$ [20, 21], SN1987A $\tau_0 \gtrsim (1 \rightarrow 10) \text{ s}$ [22], and IceCube $\tau_0/m_\nu \gtrsim 2.4 \times 10^3 \text{ (s/eV)}$ [23].

Central to the extraction of τ_0 from cosmological data is the neutrino anisotropy loss rate or transport rate Γ_T , particularly how this rate relates to τ_0 itself and other fundamental properties of the particles participating in the interaction. Some of us have recently considered this issue via a rigorous computation of the transport rate from the full Boltzmann hierarchies incorporating the two-body decay $\nu_H \leftrightarrow \nu_l + \phi$ and its inverse process induced by a Yukawa interaction [9]. For decay into *massless* daughter particles, Ref. [9] finds a transport rate $\Gamma_T \sim (m_\nu/E_\nu)^5 (1/\tau_0)$ under the assumption of linear response. This result is in contrast to the rate $\Gamma_T \sim (m_\nu/E_\nu)^3 (1/\tau_0)$ employed in previous analyses [4, 5, 7, 8], which is based upon a heuristic but (as we shall show) incorrect random walk argument.

While representing a considerable step forward, the result of Ref. [9] is nonetheless incomplete in that the mass difference between the parent neutrino ν_H and daughter neutrino ν_l must respect the neutrino mass spectra allowed by neutrino oscillations experiments. Global three-flavour analyses of oscillation data currently find the squared mass differences to be [24–26]

$$\Delta m_{21}^2 = 7.50_{-0.20}^{+0.22} \times 10^{-5} \text{ eV}^2, \quad (1)$$

$$|\Delta m_{31}^2|_N = 2.55_{-0.03}^{+0.02} \times 10^{-3} \text{ eV}^2, \quad (2)$$

$$|\Delta m_{31}^2|_I = 2.45_{-0.03}^{+0.02} \times 10^{-3} \text{ eV}^2, \quad (3)$$

where $|\Delta m_{31}^2|_N$ applies to a normal mass ordering (NO; $m_3 > m_2 > m_1$), and $|\Delta m_{31}^2|_I$ to an inverted mass ordering (IO; $m_2 > m_1 > m_3$). Relative to the benchmark parent neutrino mass of 0.2 eV , these squared mass differences are clearly small enough that the assumption of a massless daughter neutrino may become untenable and consequently alter the relationship between the transport rate Γ_T and the rest-frame decay rate τ_0 away from $\Gamma_T \sim (m_\nu/E_\nu)^5 (1/\tau_0)$.

² It is also possible to invoke invisible neutrino decay in the non-relativistic limit as a means to relax cosmological neutrino mass bounds. See, e.g., Refs. [17, 18] for recent studies. The rest-frame lifetimes required are typically of order $0.01 \rightarrow 0.1$ of the universe's age, leading to a phenomenology that differs completely from the ultra-relativistic case considered in this work.

In this work, we improve upon the original calculation of Ref. [9] to explicitly allow for a massive daughter neutrino ν_l . We furthermore extend the computation of the anisotropic stress (i.e., multipole $\ell = 2$) damping rate to $\ell > 2$ multipole moments—for compatibility with the neutrino Boltzmann hierarchy [27] solved in the CMB code CLASS [28]—and put the generalisation of the calculation to other interaction structures (besides Yukawa) on firmer grounds. We find that, in the presence of a massive ν_l , a consistent transport rate Γ_T is suppressed by an extra phase space factor proportional to $(\Delta m_\nu^2/m_\nu^2)^2$ in the limit of a small squared mass difference between the parent and daughter neutrinos, i.e., $\Gamma_T \sim (\Delta m_\nu^2/m_\nu^2)^2 (m_\nu/E_\nu)^5 (1/\tau_0)$, and this factor is *in addition to* the phase space suppression ($\sim \Delta m_\nu^2/m_\nu^2$) already inherent in the lifetime τ_0 .

Using a general version of this result valid for all Δm_ν^2 , we fit the Planck 2018 CMB data to derive a new set of cosmological constraints on the neutrino lifetime as functions of the lightest neutrino mass (i.e., m_1 in NO and m_3 in IO) that are consistent with mass difference measurements. Readers interested in a quick number can jump directly to Fig. 3 where the new constraints on τ_0 are presented. Here, we note that for a parent neutrino of $m_\nu \lesssim 0.1 \text{ eV}$, the new phase space suppression factor weakens the CMB constraint on its lifetime by as much as a factor of 50 if Δm_ν^2 corresponds to the atmospheric mass gap and by five orders of magnitude if the solar mass gap, compared with naïve limits that do not account for the daughter neutrino mass.

The paper is organised as follows. We begin with a description of the physical system in Sect. 2. We then introduce the basic equations of motion for the individual ν_H , ν_l and ϕ component in Sect. 3, before condensing them into a single set of effective equations of motion for the combined neutrino+ ϕ system in Sect. 4. Using standard statistical inference techniques presented in Sect. 5, we derive in Sect. 6 new CMB constraints on the neutrino lifetime that are consistent with the experimentally determined neutrino mass splittings. We conclude in Sect. 7. Four appendices contain respectively full expressions of the collision integrals (Appendix 1), computational details of the $\ell \geq 2$ effective damping rates (Appendix 1), a revised random walk picture explaining the $\propto (m_\nu/E_\nu)^5$ dependence of the transport rate and why the old $\propto (m_\nu/E_\nu)^3$ rate is incorrect (Appendix 1), and summary statistics of our parameter estimation analyses (Appendix 1).

2 The physical system

We consider the two-body decay of a parent neutrino ν_H into a massive daughter neutrino ν_l and a massless particle ϕ , whose transition probability is described by a Lorentz-invariant squared matrix element $|\mathcal{M}|^2$. Previously in Ref. [9] we had assumed ϕ to be a scalar and the decay described

by a Yukawa interaction, motivated by Majoron models [29–31] in which a new Goldstone boson of a spontaneously broken global $U(1)_{B-L}$ symmetry couples to neutrinos. Another possibility arises within gauged $L_i - L_j$ models, where a new vector boson plays the role of the massless state [32–35]. Models of neutrinos decaying on cosmological timescales can be found in Refs. [36–39]. Here, in order to keep our analysis as general as possible, apart from the quantum statistical requirement that ϕ must be boson, we shall leave its nature and the nature of the interaction unspecified.

Because cosmological observables do not measure neutrino spins, it is not possible to define a reference direction in the decaying neutrino's rest-frame. Thus, for an ensemble of neutrinos with randomly aligned spins, we can effectively consider the decay to be isotropic in the decaying neutrino's rest-frame, with a decay rate $\Gamma_{\text{dec}}^0 \equiv 1/\tau_0$ related to the squared matrix element via

$$\Gamma_{\text{dec}}^0 = \frac{\Delta(m_{\nu H}, m_{\nu l}, m_\phi)}{16\pi m_{\nu H}^3} |\mathcal{M}|^2, \quad (4)$$

where

$$\Delta(x, y, z) \equiv \sqrt{[x^2 - (y+z)^2][x^2 - (y-z)^2]} \quad (5)$$

is the Källén function, and it is understood that $|\mathcal{M}|^2$ needs to be evaluated at particle energies and momenta consistent with conservation laws.

In the cosmological context, the decay process can be expected to come into effect around scale factors $a \sim a_{\text{dec}}$, where a_{dec} is defined via $\Gamma_{\text{dec}}(a_{\text{dec}}) = H(a_{\text{dec}})$, with $\Gamma_{\text{dec}} = \langle m_{\nu H}/E_{\nu H} \rangle \Gamma_{\text{dec}}^0$ the ensemble-averaged Lorentz-boosted decay rate. We are interested in the decay of ν_H occurring (i) up to the time of recombination but substantially after neutrino decoupling, i.e., $10^{-10} \ll a_{\text{dec}} \lesssim 10^{-3}$, and (ii) while the bulk of the ν_H ensemble is ultra-relativistic. Two observations are in order:

1. The ultra-relativistic requirement guarantees that the inverse decay process $\nu_l + \phi \rightarrow \nu_H$ is kinematically allowed and kicks in on roughly the same time scale as ν_H decay, enabling ν_H , ν_l , and ϕ to form a coupled system whose combined anisotropic stress can be suppressed through collisional damping. The same requirement also effectively limits the parent neutrino mass to the ballpark $m_{\nu H} \lesssim 3T_\nu(a_{\text{rec}}) \sim 3 \times (4/11)^{1/3} T_\gamma(a_{\text{rec}}) \sim 0.2$ eV, if the phenomenology of anisotropic stress loss is to remain unambiguously applicable up to the recombination era when the temperature of the photon bath drops to $T_\gamma(a_{\text{rec}}) \sim 0.1$ eV.
2. Since decay and inverse decay occur by construction only after neutrino decoupling, our scenario does not generate an excess of radiation. Rather, the (ν_H, ν_l, ϕ) -system merely shares the energy that was in the original ν_H and

ν_l populations. Consequently, in the time frame of interest cosmology is unaffected at the background level, and the phenomenology of the (ν_H, ν_l, ϕ) -system *vis-à-vis* the CMB primary anisotropies is at leading order limited to anisotropic stress loss.

These observations tell us that, as far as the CMB primary anisotropies are concerned, the (ν_H, ν_l, ϕ) -system can be modelled at leading order as a single massless fluid whose energy density is equivalent to that of the original ν_H and ν_l populations at $a < a_{\text{dec}}$. What remains to be specified is the rate at which the fluid's anisotropic stress and higher-order multipole moments are lost via decay and inverse decay; this is the subject of Sects. 3 and 4.

3 Basic equations of motion

A complete set of evolution equations for the individual ν_H , ν_l , and ϕ phase space density under the influence of gravity and decay/inverse decay has been presented in Ref. [9]. We briefly describe the relevant ones below.

We work in the synchronous gauge defined by the invariant $ds^2 = a^2(\tau) [-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j]$, and split the Fourier-transformed phase space density of species i into a homogeneous and isotropic component and a perturbed part, i.e., $f_i(\mathbf{k}, \mathbf{q}, \tau) = \bar{f}_i(|\mathbf{q}|, \tau) + F_i(\mathbf{k}, \mathbf{q}, \tau)$, where τ is conformal time, $\mathbf{q} \equiv |\mathbf{q}|\hat{\mathbf{q}}$ the comoving 3-momentum, and $\mathbf{k} \equiv |\mathbf{k}|\hat{\mathbf{k}}$ is the Fourier wave vector conjugate to the comoving spatial coordinates $\mathbf{x}$ of the line element.

Because for the scenario at hand the (ν_H, ν_l, ϕ) -system behaves at leading order essentially as a single massless fluid whose energy density remains constant in the time frame of interest, the evolution of the individual homogeneous phase space density $\bar{f}_i(|\mathbf{q}|, \tau)$ is not a main concern. In fact, in Sect. 4 we shall approximate all $\bar{f}_i$'s with equilibrium distributions in order to derive closed-form expressions for the effective damping rates.

Of greater interest to us is the perturbed component $F_i(\mathbf{k}, \mathbf{q}, \tau)$. Following standard practice [27], we decompose F_i in terms of a Legendre series,

$$F_i(|\mathbf{k}|, |\mathbf{q}|, x) \equiv \sum_{\ell=0}^{\infty} (-i)^\ell (2\ell+1) F_{i,\ell}(|\mathbf{k}|, |\mathbf{q}|) P_\ell(x), \quad (6)$$

$$F_{i,\ell}(|\mathbf{k}|, |\mathbf{q}|) \equiv \frac{i^\ell}{2} \int_{-1}^1 dx F_i(|\mathbf{k}|, |\mathbf{q}|, x) P_\ell(x),$$

where $x \equiv \hat{\mathbf{k}} \cdot \hat{\mathbf{q}}$, and $P_\ell(x)$ is a Legendre polynomial of degree ℓ . Then, the equations of motion for the Legendre moments $F_{i,\ell}$ can be written as

$$\begin{aligned}
\dot{F}_{i,0} &= -\frac{|\mathbf{q}||\mathbf{k}|}{\epsilon_i} F_{i,1} + \frac{1}{6} \frac{\partial \bar{f}_i}{\partial \ln |\mathbf{q}|} \dot{h} + \left(\frac{df_i}{d\tau} \right)_{C,0}^{(1)}, \\
\dot{F}_{i,1} &= \frac{|\mathbf{q}||\mathbf{k}|}{\epsilon_i} \left(-\frac{2}{3} F_{i,2} + \frac{1}{3} F_{i,0} \right) + \left(\frac{df_i}{d\tau} \right)_{C,1}^{(1)}, \\
\dot{F}_{i,2} &= \frac{|\mathbf{q}||\mathbf{k}|}{\epsilon_i} \left(-\frac{3}{5} F_{i,3} + \frac{2}{5} F_{i,1} \right) \\
&\quad - \frac{\partial \bar{f}_i}{\partial \ln |\mathbf{q}|} \left(\frac{2}{5} \dot{\eta} + \frac{1}{15} \dot{h} \right) + \left(\frac{df_i}{d\tau} \right)_{C,2}^{(1)}, \\
\dot{F}_{i,\ell>2} &= \frac{|\mathbf{k}|}{2\ell+1} \frac{|\mathbf{q}|}{\epsilon_i} [\ell F_{i,\ell-1} - (\ell+1) F_{i,\ell+1}] \\
&\quad + \left(\frac{df_i}{d\tau} \right)_{C,\ell}^{(1)}. \tag{7}
\end{aligned}$$

Here, $\cdot \equiv \partial/\partial\tau$, $\epsilon_i(|\mathbf{q}|) \equiv (|\mathbf{q}|^2 + a^2 m_i^2)^{1/2}$ is the comoving energy of the particle species i of mass m_i , $h \equiv \delta^{ij} h_{ij}(\mathbf{k}, \tau)$ and $\eta = \eta(\mathbf{k}, \tau) \equiv -k^i k^j h_{ij}/(4k^2) + h/12$ are the trace and traceless components of h_{ij} in Fourier space respectively, and $(df_i/d\tau)_{C,\ell}^{(1)}$ represents the ℓ th Legendre moment of the linear-order Boltzmann collision integral.

The linear-order decay/inverse decay collision integrals have been derived in Ref. [9] assuming a Yukawa interaction. However, following the arguments of Sect. 2 and using the fact that the squared matrix element is Lorentz-invariant, it is straightforward to generalise the integrals of Ref. [9] to any interaction structure. The same argument also allows us to recast the collision integrals in favour of the rest-frame decay rate Γ_{dec}^0 as a fundamental parameter (over the squared matrix element $|\mathcal{M}|^2$). The interested reader can find the full expressions in Eqs. (A.1)–(A.3) in Appendix 1.

The equations of motion (7) together with the collision integrals (A.1)–(A.3) can be programmed directly into and solved by a CMB code such as CLASS [28]. In practice, however, numerical solutions particularly in the ultra-relativistic limit can be highly non-trivial, because of the vastly different time scales at play: the basic decay time scale $1/\Gamma_{\text{dec}}$ and the emergent transport time scale $1/\Gamma_{\text{T}} \sim (E_\nu/m_\nu)^4 (1/\Gamma_{\text{dec}})$. In Sect. 4 we show how these equations of motion can be condensed in the ultra-relativistic limit into a single set of equations for the combined (ν_H, ν_l, ϕ) -system under certain reasonable simplifying assumptions.

4 Effective equations of motion for the neutrino+ ϕ system

We are interested in the evolution of perturbations in the combined (ν_H, ν_l, ϕ) -system in the ultra-relativistic limit. To this end we define the integrated Legendre moment,

$$\begin{aligned}
\mathcal{F}_\ell(|\mathbf{k}|) &\equiv \frac{(2\pi^2 a^4)^{-1}}{\bar{\rho}_{\nu\phi}} \sum_{i=\nu_H, \nu_l, \phi} g_i \\
&\quad \times \int d|\mathbf{q}| |\mathbf{q}|^2 \epsilon_i \left(\frac{|\mathbf{q}|}{\epsilon_i} \right)^\ell F_{i,\ell}(|\mathbf{q}|), \tag{8}
\end{aligned}$$

where g_i is the internal degree of freedom of the i th particle species, and $\bar{\rho}_{\nu\phi}$ represents the total background energy density of the (ν_H, ν_l, ϕ) -system which we take to be ultra-relativistic, i.e., $\bar{\rho}_{\nu\phi} \propto a^{-4}$, and equal to the energy density in the pre-decay ν_H and ν_l populations.

Summing and integrating the equations of motion, Eq. (7), as per the definition (8) and taking the ultra-relativistic limit everywhere *except* the collision terms, we obtain

$$\begin{aligned}
\dot{\mathcal{F}}_0 &= -|\mathbf{k}| \mathcal{F}_1 - \frac{2}{3} \dot{h}, \\
\dot{\mathcal{F}}_1 &= \frac{|\mathbf{k}|}{3} (\mathcal{F}_0 - 2\mathcal{F}_2), \\
\dot{\mathcal{F}}_2 &= \frac{|\mathbf{k}|}{5} (2\mathcal{F}_1 - 3\mathcal{F}_3) + \frac{4}{15} \dot{h} + \frac{8}{5} \dot{\eta} + \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,2}, \\
\dot{\mathcal{F}}_{\ell>2} &= \frac{|\mathbf{k}|}{2\ell+1} [\ell \mathcal{F}_{\ell-1} - (\ell+1) \mathcal{F}_{\ell+1}] + \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}, \tag{9}
\end{aligned}$$

where

$$\begin{aligned}
\left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell} &\equiv \frac{(2\pi^2 a^4)^{-1}}{\bar{\rho}_{\nu\phi}} \sum_{i=\nu_H, \nu_l, \phi} g_i \\
&\quad \times \int d|\mathbf{q}| |\mathbf{q}|^2 \epsilon_i \left(\frac{|\mathbf{q}|}{\epsilon_i} \right)^\ell \left(\frac{df_i}{d\tau} \right)_{C,\ell}^{(1)}(|\mathbf{q}|) \tag{10}
\end{aligned}$$

are the effective collision integrals constructed from the individual collision integrals (A.1)–(A.3). Note that the effective collision integrals for $\ell = 0, 1$ integrated moments must vanish exactly because of energy and momentum conservation in the (ν_H, ν_l, ϕ) -system.

4.1 Effective collision integrals

Equations (9) and (10) are not yet in a closed form. To put them into a closed form, several simplifying approximations to the effective collision integrals are in order [9]:

1. We assume equilibrium Maxwell-Boltzmann statistics, meaning that the background phase space density of each particle species can be described at any instant by $\bar{f}_i(|\mathbf{q}|) = e^{-(\epsilon_i - \mu_i)/T_0}$, with a common comoving temperature T_0 and chemical potentials satisfying $\mu_{\nu H} = \mu_{\nu l} + \mu_\phi$. Physically, this means we take the background equilibration rate Γ_{dec} to be much larger than the loss rates of the anisotropic stress and higher-order anisotropies we are interested to compute. As demonstrated in Ref. [9],

relativistic ν_H decay and inverse decay attain a steady-state/quasi-equilibrium regime in which the above equilibrium conditions are reasonably well satisfied.

2. We take the individual species' Legendre moments $F_{i,\ell}(|\mathbf{k}|, |\mathbf{q}|)$ to be separable functions of $|\mathbf{k}|$ and $|\mathbf{q}|$, given by [12]

$$F_{i,\ell}(|\mathbf{k}|, |\mathbf{q}|) \simeq -\frac{1}{4} \frac{d\tilde{f}_i}{d \ln |\mathbf{q}|} \mathcal{F}_{i,\ell}(|\mathbf{k}|), \quad (11)$$

where $\mathcal{F}_{i,\ell}$ is the analogue of the integrated Legendre moment $\mathcal{F}_\ell$ of Eq. (8) but for the i th species alone. In the absence of collisions, this so-called separable *ansatz* (11) is exact in the ultra-relativistic limit and reflects the fact that, for massless particles, gravity is achromatic and all $|\mathbf{q}|$ -modes evolve in the same way. For collisional systems near equilibrium, i.e., where the momentum-dependence of the background phase space densities is largely constant in time, the *ansatz* also appears to capture the evolution of $\mathcal{F}_{i,\ell \leq 2}(|\mathbf{k}|)$ well in the ultra-relativistic limit, across a range of $|\mathbf{k}|$ -modes and redshifts [12].

3. In standard cosmology, all $\ell \geq 2$ anisotropies are induced by gravity after a Fourier $\mathbf{k}$ -mode enters the horizon. Gravity induces the same perturbation contrast $\mathcal{F}_{i,\ell}$ in all particle species at the same point in space, which immediately leads us to our third approximation

$$\mathcal{F}_{\nu_H,\ell}(|\mathbf{k}|) \simeq \mathcal{F}_{\nu_l,\ell}(|\mathbf{k}|) \simeq \mathcal{F}_{\phi,\ell}(|\mathbf{k}|) \simeq \mathcal{F}_\ell(|\mathbf{k}|), \quad (12)$$

where $\mathcal{F}_\ell(|\mathbf{k}|)$ is identically the ℓ th integrated Legendre moment of Eq. (8).

Note that by making these reasonable assumptions, we have essentially recast the problem of establishing the effective transport rate into a relaxation time calculation in terms of the system's linear response to a small external perturbation.

Under the above approximations and assuming further that $m_\phi = 0$, explicit evaluation of the effective collision integral (10) yields

$$\left(\frac{d\mathcal{F}}{d\tau}\right)_{C,\ell} = -\alpha_\ell a \tilde{\Gamma}_{\text{dec}} \left(\frac{am_{\nu H}}{T_0}\right)^4 \times \mathcal{F} \left(\frac{am_{\nu H}}{T_0}\right) \Phi \left(\frac{m_{\nu l}}{m_{\nu H}}\right) \mathcal{F}_\ell. \quad (13)$$

See Appendix 1 for details of the derivation. Here, the rate $\tilde{\Gamma}_{\text{dec}}$ is approximately the boosted decay rate and is defined as in Ref. [9] via

$$\begin{aligned} \tilde{\Gamma}_{\text{dec}} &\equiv \frac{(4\pi^2 a^4)^{-1}}{\bar{\rho}_{\nu\phi}} am_{\nu H} T_0^3 g_{\nu H} e^{\mu_{\nu H}/T_0} \Gamma_{\text{dec}}^0 \\ &= \frac{1}{12} \frac{(a^4 \bar{\rho}_{\nu H})}{(a^4 \bar{\rho}_{\nu\phi})} \frac{am_{\nu H}}{T_0} \Gamma_{\text{dec}}^0, \end{aligned} \quad (14)$$

where $\bar{\rho}_{\nu H} \equiv (3g_{\nu H} T_0^4/a^4 \pi^2) e^{\mu_{\nu H}/T_0}$ is the homogeneous, background energy density of ν_H in the ultra-relativistic limit.

The dimensionless function $\mathcal{F}(x)$ in Eq. (13) is a decreasing function of x given by

$$\mathcal{F}(x) = \frac{1}{2} e^{-x} \left[-1 + x - e^x (x^2 - 2) \Gamma(0, x) \right], \quad (15)$$

where $\Gamma(0, x)$ denotes the incomplete gamma function. For $x \sim 10^{-10} \rightarrow 0.1$, $\mathcal{F}(x)$ evaluates to $\sim 20 \rightarrow 1$. Beyond $x \sim 1$, $\mathcal{F}(x)$ quickly drops to zero, reflecting the fact that after ν_H becomes non-relativistic, inverse decay must also shut down so that the combined neutrino+ ϕ fluid reverts to a free-streaming system. In order to bypass the incomplete gamma function in our numerical solution, we find it convenient to use the small- x expansion

$$\begin{aligned} \mathcal{F}_{\text{sm}}(x) &\simeq -\frac{1}{2} + 2x - x^2 - \frac{x^3}{9} + \frac{x^4}{96} - \frac{x^5}{900} \\ &\quad + \frac{x^6}{8640} + (\gamma + \log x) \left(\frac{x^2}{2} - 1 \right) + \mathcal{O}(x^7), \end{aligned} \quad (16)$$

where $\gamma \simeq 0.57721$ is Euler-Mascheroni constant. The expansion is accurate to $< 7\%$ for $x \leq 2$. For $x > 2$, we switch to the function $\mathcal{F}_{\text{lg}}(x) \simeq 2.4 e^{-x}/x^{1.5}$, which approximates Eq. (15) to $< 10\%$ at $2 < x < 17$.

Relative to the result of Ref. [9], the collision integral (13) also contains two new elements. Firstly, the phase space factor,

$$\Phi(y) = \frac{1}{1-y^2} \left(1 - y^4 + 4y^2 \ln y \right), \quad (17)$$

allows for the possibility of a nonzero $m_{\nu l}$ and takes on values $1 \rightarrow 0$ in the region $0 \leq y < 1$ (or, equivalently, $0 \leq m_{\nu l} < m_{\nu H}$). In the limit of a small squared mass gap, i.e., $\Delta m_v^2 \equiv m_{\nu H}^2 - m_{\nu l}^2 \ll m_{\nu H}^2$, it reduces to $\Phi(m_{\nu l}/m_{\nu H}) \rightarrow (1/3)(\Delta m_v^2/m_{\nu H}^2)^2$ and traces its origin to the fact that the momentum carried by each decay product in the rest frame of ν_H is in fact $\Delta m_v^2/(2m_{\nu H})$, not $m_{\nu H}/2$ as in the case of massless decay products. Secondly, the ℓ -dependent coefficients α_ℓ are given by the expression

$$\alpha_\ell = \frac{1}{32} \left(3\ell^4 + 2\ell^3 - 11\ell^2 + 6\ell \right) \quad (18)$$

for all $\ell \geq 0$. These coefficients increase with ℓ , reflecting the fact that, for a fixed decay opening angle, the decay and inverse decay processes become more efficient at wiping out anisotropies on progressively smaller angular scales.

Lastly, we highlight again that the effective collision rate in Eq. (13) has in total five powers of $am_{\nu H}/T_0$, one of which is the Lorentz boost appearing in the boosted decay rate (14). This result is in drastic contrast with the much larger $\propto (am_{\nu H}/T_0)^3$ rate used in previous works [4, 5, 7, 8], which we stress is *not* an *ab initio* result, but rather one based

upon a heuristic random walk argument. Appendix 1 shows that this heuristic argument is in fact incomplete, and that a consistent random walk picture accounting for all same-order effects must yield the much smaller $\propto (am_{\nu H}/T_0)^5$ rate derived from first principles in this work.

4.2 Three-state to two-state approximation

The decay scenario presented thus far concerns only two of the three Standard-Model neutrino mass states. Without model-specific motivations to exclude one particular mass state from the decay interaction, one might generally expect all three mass states to couple similarly at the Lagrangian level and participate in the decay/inverse decay process as parent particles and/or decay products. Then, kinematics largely determine the decay rates, and we can deduce from the measured neutrino squared mass differences that the rest-frame decay rates $\Gamma_{i \rightarrow j}^0$ for the processes $\nu_i \rightarrow \nu_j + \phi$ follow the patterns

$$\Gamma_{3 \rightarrow 1}^0 \simeq \Gamma_{3 \rightarrow 2}^0 \gg \Gamma_{2 \rightarrow 1}^0 \quad (19)$$

assuming a normal mass ordering, and

$$\Gamma_{2 \rightarrow 3}^0 \simeq \Gamma_{1 \rightarrow 3}^0 \gg \Gamma_{2 \rightarrow 1}^0 \quad (20)$$

assuming an inverted mass ordering. All IO mass spectra and those NO mass spectra consistent with $m_1 \gtrsim 0.03$ eV satisfy the condition (19) or (20) at better than 10%. For $m_1 \lesssim 0.03$ eV in NO, the ratio $\Gamma_{2 \rightarrow 1}^0/\Gamma_{3 \rightarrow 2}^0$ rises to ~ 0.25 while $\Gamma_{3 \rightarrow 1}^0/\Gamma_{3 \rightarrow 2}^0$ approaches 0.85 [9]. Thus, Eq. (19) is still a reasonably well satisfied.

If the two larger (boosted) decay rates satisfy the equilibrium condition $\Gamma_{i \rightarrow j} \gtrsim H$, then, irrespective of the mass ordering, we can expect the phenomenology of the three-flavour system to be similar to the two-flavour case discussed in Sect. 2, save for the understanding that it is now the ν_1 , ν_2 , ν_3 , and ϕ that form a single massless fluid system. It remains to determine the effective collision integral and hence the $\ell \geq 2$ anisotropy damping rates, which we approximate as follows.

Normal mass ordering Here, we assume the two lightest mass states with masses m_1 and m_2 to be effectively degenerate and set $\Gamma_{2 \rightarrow 1}^0 = 0$. Then, following the recipe laid down in Ref. [9], the equations of motion can be condensed by multiplying by a factor of 2 (i) the ν_H and ϕ collision integrals (A.1) and (A.3), and (ii) all momentum-integrated quantities pertaining to ν_l . This immediately leads to an effective collision integral

$$\left(\frac{d\mathcal{F}}{d\tau}\right)_{C,\ell} = 2 \frac{(2\pi^2 a^4)^{-1}}{\bar{\rho}_{\nu\phi}} \sum_{i=\nu_H, \nu_l, \phi} g_i \times \int d|\mathbf{q}| |\mathbf{q}|^2 \epsilon_i \left(\frac{|\mathbf{q}|}{\epsilon_i}\right)^\ell \left(\frac{df_i}{d\tau}\right)_{C,\ell}^{(1)}(|\mathbf{q}|), \quad (21)$$

that differs from the original definition (10) and hence Eq. (13) by an overall factor of 2.

Inverted mass ordering Here, we treat the two heaviest mass states as effectively degenerate and again set $\Gamma_{2 \rightarrow 1}^0 = 0$. Then, multiplying by a factor of 2 (i) the ν_l and ϕ collision integrals (A.2) and (A.3), and (ii) all momentum-integrated quantities of ν_H , it is straightforward to show the same overall factor of 2 appears again in front of the effective collision integral as in the normal mass ordering case in Eq. (21).

4.3 Effective temperature and density ratio

We close this section with a brief discussion of the effective comoving temperature T_0 and the ν_H background energy density $\bar{\rho}_{\nu H}$ after equilibration, both of which appear in the effective collision integral (13).

The ν_H background energy density appears in the energy density ratio $(a^4 \bar{\rho}_{\nu H})/(a^4 \bar{\rho}_{\nu\phi})$, where we remind the reader that $\bar{\rho}_{\nu\phi}$ stands for the total energy density in the combined (ν_H, ν_l, ϕ) -system and is, leaving aside dilution due to expansion, equal to the pre-decay energy density in ν_H and ν_l . This energy density ratio generally evolves with time, but at a rate much smaller than the Hubble expansion H in the steady-state/quasi-equilibrium regime of our assumptions; Ref. [9] finds numerically a fairly constant

$$\frac{(a^4 \bar{\rho}_{\nu H})}{(a^4 \bar{\rho}_{\nu\phi})} \sim \frac{1}{3} \rightarrow \frac{1}{2} \quad (22)$$

in the time frame of interest (i.e., where $\Gamma_{\text{dec}} \gg H$ and the ultra-relativistic approximation is well satisfied). In our analysis, we take this ratio to be $1/3$.

The temperature T_0 is the effective comoving temperature of the combined (ν_H, ν_l, ϕ) -system, and should be lower than the pre-decay neutrino comoving temperature $T_{\nu,0} \simeq 1.95 \text{ K} \simeq 1.68 \times 10^{-4} \text{ eV}$. To estimate T_0 , we note that energy conservation imposes the relation $a_{\text{pre}}^4 (\bar{\rho}_{\nu H, \text{pre}} + \bar{\rho}_{\nu l, \text{pre}}) = a_{\text{post}}^4 (\bar{\rho}_{\nu H, \text{post}} + \bar{\rho}_{\nu l, \text{post}} + \bar{\rho}_\phi)$ between the pre-decay and post-decay energy densities. Assuming Maxwell-Boltzmann statistics (i.e., $\bar{\rho} = 3T_0 \bar{n}/a$), this leads to

$$\frac{T_0}{T_{\nu,0}} = \frac{a_{\text{pre}}^3 (\bar{n}_{\nu H, \text{pre}} + \bar{n}_{\nu l, \text{pre}})}{a_{\text{post}}^3 (\bar{n}_{\nu H, \text{post}} + \bar{n}_{\nu l, \text{post}} + \bar{n}_\phi)}. \quad (23)$$

The two-body decay also stipulates number conservation laws, namely, $a_{\text{pre}}^3 \bar{n}_{\nu H, \text{pre}} = a_{\text{post}}^3 (\bar{n}_{\nu H, \text{post}} + \bar{n}_\phi)$ and $a_{\text{pre}}^3 \bar{n}_{\nu l, \text{pre}} = a_{\text{post}}^3 (\bar{n}_{\nu l, \text{post}} - \bar{n}_\phi)$, which, upon substitution into Eq. (23), yields

$$\begin{aligned} \frac{T_0}{T_{\nu,0}} &= 1 - \frac{\bar{n}_\phi}{\bar{n}_{\nu H, \text{post}} + \bar{n}_{\nu l, \text{post}} + \bar{n}_\phi} \\ &\simeq 1 - \frac{\bar{\rho}_\phi}{\bar{\rho}_{\nu\phi}}. \end{aligned} \quad (24)$$

Numerically, Ref. [9] finds the energy density ratio to be at most $\bar{\rho}_\phi/\bar{\rho}_{\nu\phi} \sim 0.1$ in the steady-state regime of interest. Therefore, in this analysis, we take T_0 to be the same as the pre-decay neutrino comoving temperature $T_{\nu,0}$.

5 Statistical inference

Having derived the effective equations of motion in a way consistent with neutrino oscillation data, we wish now to put constraints on the neutrino rest-frame lifetime τ_0 using cosmological data.

Before we describe our data analysis, observe that the effective collision integral (13) can be broken down in terms of its dependence on the scale factor a into

$$\left(\frac{d\mathcal{F}}{d\tau}\right)_{C,\ell} = -\alpha_\ell a^6 Y \mathcal{F}(aX) \mathcal{F}_\ell. \quad (25)$$

Here,

$$X \equiv \frac{m_{\nu H}}{T_0} \simeq 298 \left(\frac{m_{\nu H}}{0.05 \text{ eV}}\right) \left(\frac{T_{\nu,0}}{T_0}\right) \quad (26)$$

is an effective mass parameter, while

$$\begin{aligned} Y &\equiv \frac{C}{12} \frac{(a^4 \bar{\rho}_{\nu H})}{(a^4 \bar{\rho}_{\nu\phi})} \Phi\left(\frac{m_{\nu l}}{m_{\nu H}}\right) \left(\frac{m_{\nu H}}{T_0}\right)^5 \Gamma_{\text{dec}}^0 \\ &\simeq 6.55 C \times 10^{10} \left(\frac{(a^4 \bar{\rho}_{\nu H})}{(a^4 \bar{\rho}_{\nu\phi})} / \frac{1}{3}\right) \left(\frac{T_{\nu,0}}{T_0}\right)^5 \\ &\quad \times \Phi\left(\frac{m_{\nu l}}{m_{\nu H}}\right) \left(\frac{m_{\nu H}}{0.05 \text{ eV}}\right)^5 \Gamma_{\text{dec}}^0 \end{aligned} \quad (27)$$

is an effective transport rate, with $C = 1$ for the one parent/one daughter scenario (“one decay channel”), and $C = 2$ if we wish to use the three-state to two-state approximation to describe the case of two parents/one daughter or one parent/two daughters (“two decay channels”). Both X and Y contain only quantities that are taken to be constant in time in our analysis (see Sect. 4.3), and together these two effective parameters determine completely the phenomenology of the effective collision integral (13).

5.1 Model parameter space

To derive credible limits on the rest-frame lifetime τ_0 , we consider two one-variable extensions to the standard six-variable Λ CDM fit, scenarios A and B, whose parameter spaces are spanned respectively by

$$\begin{aligned} \theta_A &= \{\omega_b, \omega_c, \theta_s, \tau_{\text{reio}}, n_s, \ln(10^{10} A_s), Y\}, \\ \theta_A^{\text{fixed}} &= \{N_{\text{fs}} = 1, N_{\text{int}} = 2.0440, X = X_{\text{scan}}\}, \end{aligned} \quad (28)$$

and

$$\begin{aligned} \theta_B &= \{\omega_b, \omega_c, \theta_s, \tau_{\text{reio}}, n_s, \ln(10^{10} A_s), Y\}, \\ \theta_B^{\text{fixed}} &= \{N_{\text{fs}} = 0, N_{\text{int}} = 3.0440, X = X_{\text{scan}}\}. \end{aligned} \quad (29)$$

In both scenarios, the variables are the physical baryon density parameter ω_b , the physical cold dark matter density parameter ω_c , the angular sound horizon θ_s , the optical depth to reionisation τ_{reio} , the spectral index of the primordial curvature power spectrum n_s and its amplitude A_s at the pivot scale $k_{\text{piv}} = 0.05 \text{ Mpc}^{-1}$, and the effective transport rate Y . The scenarios differ in the choices of fixed parameters in the neutrino sector.

Scenario A Here, the numbers of free-streaming and interacting neutrinos are always fixed at $N_{\text{fs}} = 1$ and $N_{\text{int}} = 2.0440$ respectively (such that N_{fs} and N_{int} add up to the standard $N_{\text{eff}}^{\text{SM}} = 3.0440$ [40, 41]), while the effective mass parameter X is fixed at a value X_{scan} to be discussed below. We model all neutrinos as massless states in CLASS, in keeping with the limits of validity of our effective equations of motion (13). Planck 2018 CMB data alone currently constrain the neutrino mass sum to $\sum m_\nu \lesssim 0.3 \text{ eV}$ (95%C.I.) in a Λ CDM+ m_ν 7-variable fit [1]; in combination with non-CMB data, the bound tightens to $\sum m_\nu \lesssim 0.12 \text{ eV}$ (95%C.I.) [1]. Therefore, as long as we fit CMB data only and cap the parent neutrino mass at $m_{\nu H} \simeq 0.1 \text{ eV}$ in the interpretation of our fit outcomes, our modelling is internally consistent. Note that this is a lower cap than the limit $m_{\nu H} \lesssim 0.2 \text{ eV}$ established in Sect. 2, which, we remind the reader, stems from requiring that the neutrinos stay ultra-relativistic up to recombination.³

The mass cap also sets an upper limit on the value of X_{scan} that can be unambiguously explored with our modelling. Specifically, for the choice of $T_0 = T_{\nu,0}$, we are limited to the parameter range

$$X_{\text{scan}} \in [52, 595], \quad (30)$$

where the lower end corresponds to the smallest possible mass gap consistent with oscillation measurements, i.e., $\sqrt{\Delta m_{21}^2} = 0.0087 \text{ eV}$, and the upper limit is the mass cap

³ We do not consider the case of one *massive* free-streaming and two massless interacting neutrinos, because given current sensitivities, any bound on the neutrino mass derived under these assumptions cannot have an internally consistent interpretation that is also compatible with the mass spectra established by oscillations experiments. To see this, suppose for instance the fit returns an upper limit of 0.3 eV on the mass of the free-streaming neutrino. Then, by oscillation measurements the other two (interacting) neutrinos must also have masses in the 0.3 eV ball park, which immediately conflicts with the $m_{\nu H} \lesssim 0.2 \text{ eV}$ limit of validity of our simplified modelling. Furthermore, adding up the three masses gives a neutrino mass sum of 0.9 eV—not the 0.3 eV indicated by the modelling and hence the fit—and this is unlikely to be compatible with CMB data. In other words, the 0.3 eV mass limit we started with in this example is actually meaningless.

Table 1 All decay and free-streaming (FS) scenarios considered in this work, and their corresponding mass ordering (NO or IO), decay mass gap, and minimum allowed parent neutrino mass $m_{\nu H}$. We use the convention $m_3 > m_2 > m_1$ for normal mass ordering and $m_2 > m_1 > m_3$ for inverted mass ordering, and a common value $\Delta m_{21}^2 = 7.50 \times 10^{-5}$. Note that numerically we do not distinguish between $|\Delta m_{31}^2|_{\text{N}}$, $|\Delta m_{31}^2|_{\text{I}}$, $\Delta m_{32}^2|_{\text{N}} \equiv |\Delta m_{31}^2|_{\text{N}} - \Delta m_{21}^2$, and $|\Delta m_{23}^2|_{\text{I}} \equiv |\Delta m_{31}^2|_{\text{I}} + \Delta m_{21}^2$: the difference between them is less than 5% [24–26]. This allows us to lump four physically distinct cases into one single scenario A1 and to use a single representative mass splitting $|\Delta m_{31}^2| = 2.5 \times 10^{-3} \text{ eV}^2$. The same is in principle also true for two physical cases of scenario B. Nonetheless, we distinguish between B1 and B2, because the former comprises two decay modes, while the latter has only one: this makes the definitions of the particle lifetime differ by a factor of two between the two cases

		FS	Decay	Gap	Min $m_{\nu H}^2$
Scenario A: one decay channel					
A1	NO	ν_1	$\nu_3 \rightarrow \nu_2$	$\Delta m_{32}^2 _{\text{N}}$	$ \Delta m_{31}^2 _{\text{N}}$
		ν_2	$\nu_3 \rightarrow \nu_1$	$ \Delta m_{31}^2 _{\text{N}}$	$\Delta m_{31}^2 _{\text{N}}$
	IO	ν_2	$\nu_1 \rightarrow \nu_3$	$ \Delta m_{31}^2 _{\text{I}}$	$ \Delta m_{31}^2 _{\text{I}}$
		ν_1	$\nu_2 \rightarrow \nu_3$	$\Delta m_{23}^2 _{\text{I}}$	$\Delta m_{23}^2 _{\text{I}}$
A2	NO	ν_3	$\nu_2 \rightarrow \nu_1$	Δm_{21}^2	Δm_{21}^2
A3	IO	ν_3	$\nu_2 \rightarrow \nu_1$	Δm_{21}^2	$\Delta m_{23}^2 _{\text{I}}$
Scenario B: two decay channels					
B1	NO	–	$\nu_3 \rightarrow \nu_2, \nu_1$	$ \Delta m_{31}^2 _{\text{N}}$	$ \Delta m_{31}^2 _{\text{N}}$
B2	IO	–	$\nu_1, \nu_2 \rightarrow \nu_3$	$ \Delta m_{31}^2 _{\text{I}}$	$ \Delta m_{31}^2 _{\text{I}}$

discussed immediately above. Then, with appropriate post-processing, scenario A can be interpreted to represent six physically distinct cases, summarised in Table 1.

Scenario B This scenario has $N_{\text{int}} = 3.0440$ interacting neutrinos in the three-state to two-state approximation and no free-streaming neutrinos. It represents two physically distinct cases, as summarised in Table 1. Again, as with scenario A, to consistently interpret the fit outcomes requires that we limit the parent neutrino masses to $m_{\nu H} \lesssim 0.1 \text{ eV}$. This in turn limits the range of effective mass X to

$$X_{\text{scan}} \in [298, 595] \quad (31)$$

for the choice of $T_0 = T_{\nu,0}$, where the lower bound corresponds to the mass gap $\sqrt{|\Delta m_{31}^2|} = 0.05 \text{ eV}$.

Figure 1 shows a selection of CMB TT power spectra computed using our implementation of the effective equations of motion (25) in CLASS [28],⁴ relative to the prediction for a reference Λ CDM cosmology. For a fixed effective mass parameter $X = 298$ or, equivalently, $m_{\nu H} = 0.05 \text{ eV}$ assuming $T_0 = T_{\nu,0}$, the top panel contrasts the signatures of ultra-relativistic neutrino decay for a range of effective transport rates Y in both scenarios A and B. Observe that for the same value of Y , the signature of decay in scenario

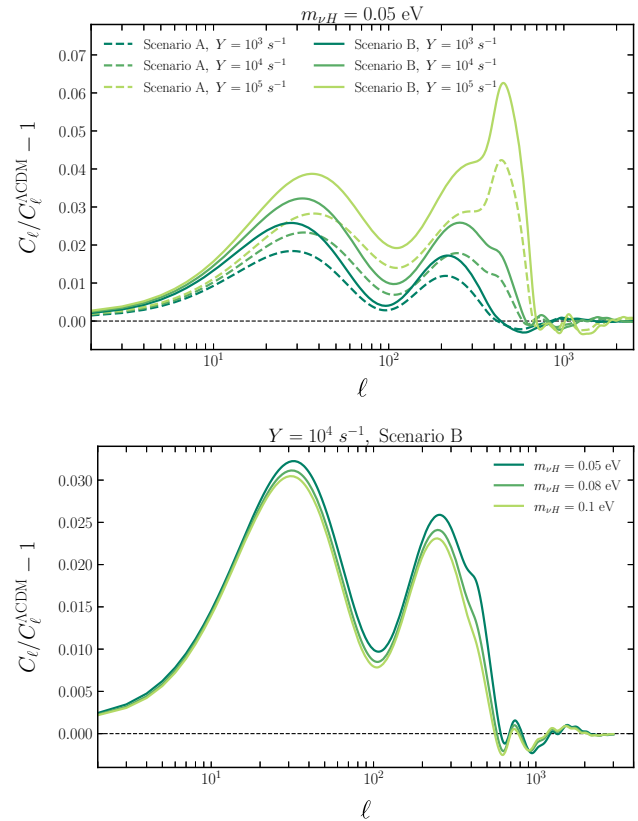


Fig. 1 CMB TT power spectra of ultra-relativistic invisible neutrino decay scenarios relative to the Planck 2018 best-fit Λ CDM cosmology. In each case, we vary only the effective mass parameter X (Eq. (26)) and effective transport rate Y (Eq. (27)), while keeping all other cosmological parameters fixed at their best-fit Λ CDM values. *Top*: Variations between scenarios A (solid line) and B (dashed line) for a range of Y values, with X fixed at $X = 298$ or, equivalently, $m_{\nu H} = 0.05 \text{ eV}$ assuming $T_0 = T_{\nu,0}$. *Bottom*: Variations in scenario B with respect to the decaying neutrino mass $m_{\nu H}$, with $T_0 = T_{\nu,0}$ and at a fixed the effective transport rate $Y = 10^4 \text{ s}^{-1}$

A is approximately a factor of 2/3 times that of scenario B. This is consistent with expectation, as only two out of three neutrinos interact in the former scenario, while all three neutrinos participate in the interaction in the latter case. We can therefore expect constraints on Y to be stronger in scenario B than in scenario A.

The bottom panel of Fig. 1 shows variations in the decay signature with respect to the effective mass parameter X , or $m_{\nu H}$ at a fixed T_0 , for a fixed effective transport rate $Y = 10^4 \text{ s}^{-1}$ in scenario B. Here, we see that the signature is weaker for larger values of X . This is because X appears only in the prefactor $\mathcal{F}(aX)$ in the effective equation of motion (25). Since $\mathcal{F}(aX)$ is a decreasing function of its argument, for a fixed evolution history $a(t)$, larger values of X will generally lead to smaller prefactors and hence weaker signatures in the CMB anisotropies. This also means that we can expect constraints on Y to weaken with increasing X .

⁴ Our modified version of CLASS is available for download at https://github.com/gpirobon/Class_NuDecay.

5.2 Data and analysis

We compute the CMB temperature and polarisation anisotropies for a large sample of parameter values in the parameter spaces defined in Eqs. (28) and (29) using our modified version of CLASS. We test these outputs against the Planck 2018 data [1] using

1. The TTTEEE likelihood at $\ell \geq 30$,
2. The Planck low- ℓ temperature+polarisation likelihood, and
3. The Planck lensing likelihood,

a combination labelled “TTTEEE+lowE+lensing” in Ref. [1]. The parameter spaces of Eqs. (28) and (29) are sampled as Monte Carlo Markov Chains (MCMC) generated with the package MONTEPYTHON-3 [42]. We construct credible intervals using GETDIST [43].

The prior probability densities employed in the analysis are always flat and linear in each parameter directions, with prior boundaries as follows:

- For the variables ω_b , ω_c , θ_s , n_s , and $\ln(10^{10} A_s)$, the priors are unbounded at both ends.
- For the optical depth to reionisation, we impose $\tau_{\text{reio}} \in [0.01, \infty)$.
- For the effective transport rate, we use $Y \in [0, \infty)$ in scenario A and $Y \in [0, 4000]$ in scenario B.

We generate in excess of 250,000 samples in each case (and over a million samples in some cases). Convergence of the chains is defined by the Gelman-Rubin convergence criterion $R - 1 < 0.015$.

6 Results and discussions

Figure 2 shows the 1D marginal posteriors for the effective transport rate Y for a range of fixed effective mass parameter values $X = X_{\text{scan}}$ in both scenarios A (top panel) and B (bottom panel). The corresponding one-sided 95% credible limits on Y and the equivalent free-streaming redshift z_{fs} in each case are summarised in Table 2. This redshift is defined via the condition

$$\Gamma_{\text{T}}(z_{\text{fs}}) = H(z_{\text{fs}}), \quad (32)$$

where we have identified the transport rate Γ_{T} with the $\ell = 2$ damping rate, i.e.,

$$\Gamma_{\text{T}} \equiv a^5 Y \mathcal{F}(aX). \quad (33)$$

Physically, z_{fs} characterises the time at which the combined (ν_H, ν_l, ϕ) -system switches from a free-streaming gas to a tightly-coupled fluid [5].

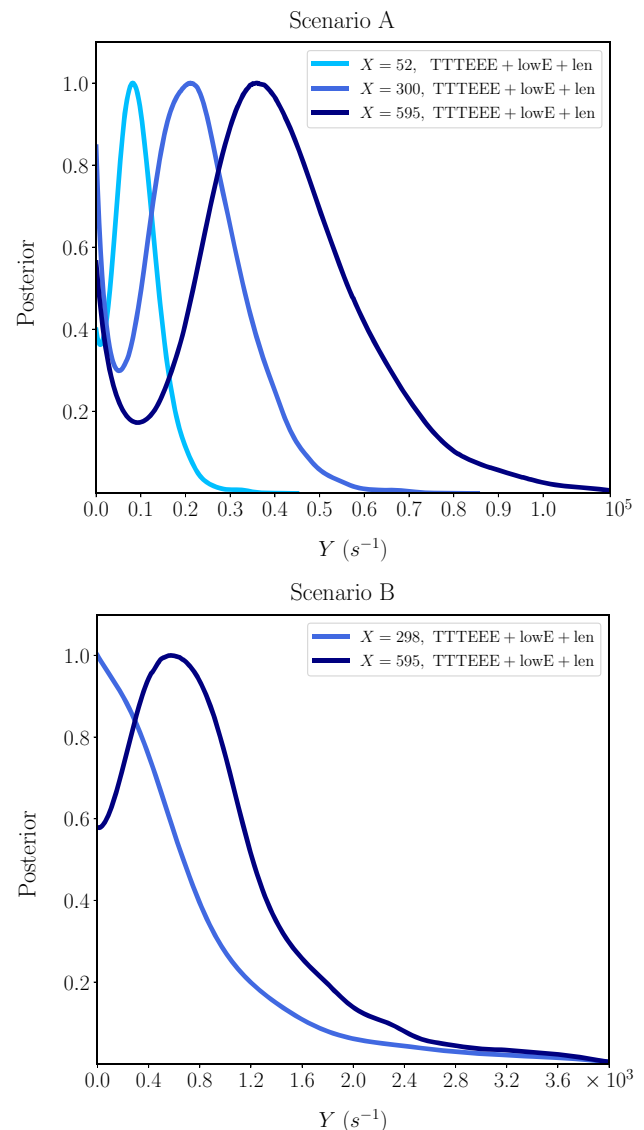


Fig. 2 1D marginalised posteriors for the effective transport rate Y (Eq. (27)) obtained from our MCMC analysis for a range of fixed effective mass values $X = X_{\text{scan}}$ in scenarios A (top panel) and B (bottom panel), whose parameter spaces are defined in Eqs. (28) and (29) respectively. The data used are the TTTEEE+lowE+lensing combination from the Planck 2018 data release [1]

Table 2 Marginalised one-sided 95% credible limits on the effective transport rate Y and the equivalent free-streaming redshift z_{fs} (Eq. (32)) for fixed values of the effective mass parameter X (and their equivalent parent neutrino mass $m_{\nu H}$ assuming $T_0 = T_{\nu,0}$) in scenarios A and B

	X		Y (s^{-1})	z_{fs}
	Fixed	$m_{\nu H}$ (eV)	95% C.L.	95% C.L.
A	52	0.0087	< 18611.32	< 2457
	300	0.05	< 41490.24	< 2450
	595	0.1	< 76221.34	< 2476
B	298	0.05	< 2286.08	< 1509
	595	0.1	< 2287.58	< 1338

Firstly, we observe that within each scenario, the constraint on Y weakens with increasing X and that for the same X , the scenario B constraint is typically an order of magnitude tighter than its scenario A counterpart. Both trends are consistent with expectations (see Sect. 5.1). Likewise, the constraints on the free-streaming redshift z_{fs} are independent of the choice of X within each scenario—to better than 1% in scenario A and about 10% in scenario B—as one would expect from the definition (32).

Having established these z_{fs} bounds also allows us to roughly map our bounds on Y in Table 2 to a simple expression

$$Y \lesssim 5.3 \times 10^4 / \mathcal{F}(X/2460) \text{ s}^{-1} \quad (34)$$

in scenario A, and

$$Y \lesssim 2 \times 10^3 / \mathcal{F}(X/1500) \text{ s}^{-1} \quad (35)$$

in scenario B. In the case of the latter, our new constraint $z_{\text{fs}} \lesssim 1500$ is clearly tighter than the previous limit $z_{\text{fs}} \lesssim 1965$ [7]. The difference is likely attributable to improved precision of the Planck 2018 CMB data over the 2013 data used in Ref. [7], although of course the different time dependences of the transport rates assumed in the analyses— $\propto a^3$ in [7] vs our $\propto a^5 \mathcal{F}(aX)$ in Eq. (33)—as well as the assumption of instantaneous recoupling in Ref. [7] could also have played a role.

Secondly, while we have quoted in Table 2 one-sided 95% credible limits on Y and z_{fs} , it is evident in Fig. 2 that in almost all cases the 1D posterior for Y peaks at a nonzero value. Indeed, Ref. [8] also found a similar feature, despite using a transport rate with a $\propto a^3$ time dependence [as opposed to our $\propto a^5 \mathcal{F}(aX)$ rate]. Reference [7] likewise found a shifted peak in their analysis of the Planck 2013 data for $z_{\text{fs}} > 0$, notwithstanding the assumption of instantaneous recoupling.

However, in all cases the peak shift of either Y or z_{fs} from zero is statistically insignificant (i.e., $< 2\sigma$) and appears to be driven by CMB polarisation data [8]. We therefore dwell no further on the subject, except to note that the shifted Y peak is accompanied by a marginal increase in the inferred scalar spectral index n_s —a trend also observed in Ref. [8]—and consequently the small-scale RMS fluctuation σ_8 , in comparison with their $Y = 0$ counterparts derived from the same data combination. There are otherwise no strong degeneracies between Y and other base Λ CDM parameters. The interested reader may consult Appendix 1 for more summary statistics of our parameter estimation analyses. Here, we conclude in view of the degeneracy directions that adding baryon acoustics oscillations measurements to the analysis is unlikely to improve the bound on Y , while a low-redshift

small-scale σ_8 measurement could conceivably tighten the bound.⁵

6.1 Constraints on the neutrino lifetime

To reinterpret our marginalised one-sided 95% credible limits on the effective transport rate Y (Table 2) as limits on the rest-frame neutrino lifetime τ_0 , we first interpolate the constraints on Y as a function of the effective mass parameter X (or, equivalently, the mass of the decaying neutrino $m_{\nu H}$) within each scenario. Then, using Eq. (27) these upper limits on Y constraints can be converted into lower bounds on $\tau_0 = 1/\Gamma_{\text{dec}}^0$ for any neutrino mass spectrum desired. Alternatively, we can use Eq. (27) to directly translate the approximate limits of Eqs. (34) and (35) into

$$\tau_0 \gtrsim 1.2 \times 10^6 \mathcal{F} \left[0.12 \left(\frac{m_{\nu H}}{0.05 \text{ eV}} \right) \right] \times \Phi \left(\frac{m_{\nu l}}{m_{\nu H}} \right) \left(\frac{m_{\nu H}}{0.05 \text{ eV}} \right)^5 \text{ s} \quad (36)$$

for scenario A, and

$$\tau_0 \gtrsim 6.6 \times 10^7 \mathcal{F} \left[0.2 \left(\frac{m_{\nu H}}{0.05 \text{ eV}} \right) \right] \times \Phi \left(\frac{m_{\nu l}}{m_{\nu H}} \right) \left(\frac{m_{\nu H}}{0.05 \text{ eV}} \right)^5 \text{ s} \quad (37)$$

for scenario B2.

Note that in the case of B1, the lifetime of ν_3 needs to be computed from $\tau_0 = 1/(2\Gamma_{\text{dec}}^0)$, because of its two distinct decay modes. See Table 1.

Figure 3 shows the limits on τ_0 constructed from interpolation for the physical cases listed in Table 1, as functions of the lightest neutrino masses m_1 (NO) or m_3 (IO). Also displayed in the same plots are the naïve constraints on τ_0 we would have obtained had we set the mass of the daughter neutrino to zero in the phase space factor Φ . For reference we also show current constraints from SN1987A [22] and IceCube [23], as

⁵ Note that we have had to apply a prior on the effective transport parameter Y in scenario B of $Y \in [0, 4000]$, as opposed to allowing Y to be unbounded, i.e., $Y \in [0, \infty)$, as in scenario A. This is because, had we not applied the restrictive prior, the Y parameter in the case of $X = 298$ would have exhibited a strong but accidental degeneracy with the spectral index n_s in the high- Y tail. The tail region is extensive, with χ^2 values hovering around $\Delta\chi^2 \simeq 20$ worse than the best-fit, but presents no meaningful second peak. Had we used low-redshift σ_8 measurements this tail would likely have been cut off naturally. This tail feature is not present in the $X = 595$ case. But we have nonetheless applied the same restrictive prior on Y here for consistency within scenario B. Without the prior the 1D marginalised 95% bound on Y in the $X = 595$ case would relax from $Y \lesssim 2288$ to $Y \lesssim 2546$.

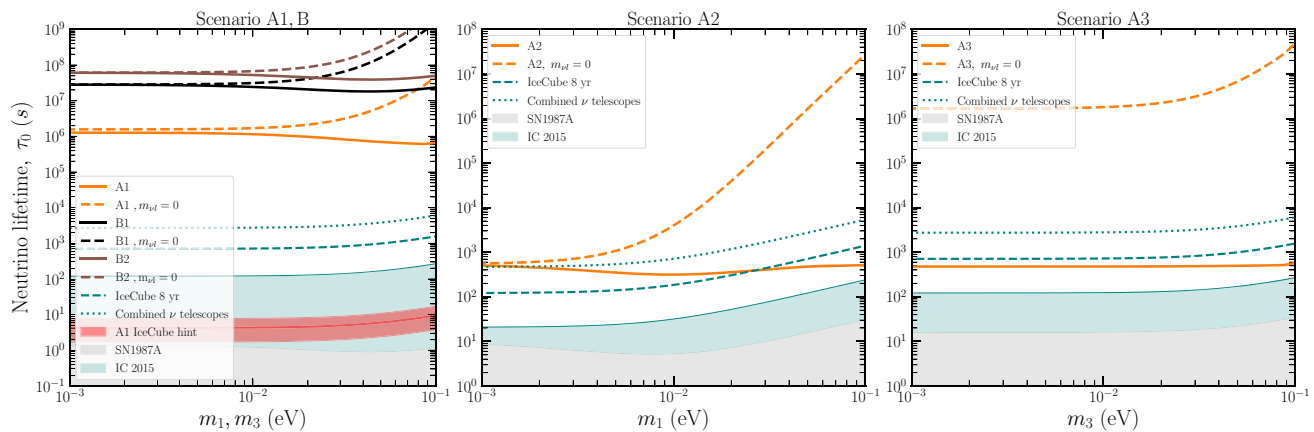


Fig. 3 Lower bounds on the neutrino lifetime converted from the 95% credible limits on the effective transport rate Y (Table 2) for the three single-decay-channel cases of scenario A and the two two-decay-channel cases of scenario B (see Table 1), as functions of the lightest neutrino mass m_1 (NO) or m_3 (IO). Orange (black) solid lines denote the revised bounds of this work for scenarios A (B), while the dotted lines represent the naïve constraints we would have obtained had we assumed a massless daughter neutrino $m_{\nu l} = 0$ in the phase space factor $\Phi(m_{\nu l}/m_{\nu H})$ (Eq. (17)). The interpretation of these bounds in conjunction with Table 1 is unambiguous in scenario A. In scenario B1 the bounds apply to ν_3 , while in B2 they apply to both ν_1 and ν_2 assuming equal lifetimes. We additionally show current bounds from

SN1987A [22] and from the IceCube 2015 data analysed in Ref. [23] (“IC 2015”) assuming the flavour composition of a full pion decay scenario, i.e., $(f_e, f_\mu, f_\tau) = (1, 2, 0)$, at the source, as well as the projected reach of future IceCube runs (“IceCube 8 yr”) and a combination of future neutrino telescopes (including Baikal-GVD, KM3NeT, P-ONE, TAMBO, and IceCube-Gen2), the latter of which also assumes improved oscillations parameter measurements from JUNO, DUNE, and Hyper-Kamiokande. See Ref. [23] for details. Also displayed in red in the left panel is the invisible decay scenario proposed in Ref. [44] that would solve the tension between cascade and track data sets of IceCube

well as projected constraints from future measurements by IceCube and other neutrino telescopes [23].⁶

Relative to our previous estimate of $\tau_0 \gtrsim (4 \times 10^5 \rightarrow 4 \times 10^6 (m_{\nu H}/0.05 \text{ eV})^5 \text{ s}$ [9], the new naïve (i.e., $m_{\nu l} = 0$) constraints are comparable to the high end at $m_{\nu H} \simeq 0.05 \text{ eV}$ in scenario A (one decay channel) and a factor of 50 tighter than the high end in scenario B (two decay channels) at the same mass point. These differences can be attributed to (i) different prefactors in the relation between the effective transport rate and the rest-frame decay rate (Eq. (27)), and (ii) the fact that the old estimate had been obtained from a rough con-

version of the z_{fs} bound of Ref. [7]—itself derived from older CMB data—using the $\Gamma_T(z_{\text{fs}}) = H(z_{\text{fs}})$ condition. Furthermore, because the constraints on Y weaken with m_{ν} [due to $\mathcal{F}(am_{\nu}/T_0)$ in Eq. (25)], the scaling of the naïve τ_0 bounds with $m_{\nu H}$ in fact follows more closely $\propto m_{\nu H}^4$ than $\propto m_{\nu H}^5$ in the mass range of interest (as can be seen most clearly in the middle panel of Fig. 3).

However, once a finite $m_{\nu l}$ has been accounted for via the new phase space factor Φ in accordance with oscillation measurements, the lifetime bounds on the decaying neutrino relax significantly in all physical cases. Specifically, if the parent-daughter neutrino mass gap corresponds to the atmospheric mass splitting (i.e., scenarios A1 and B), then irrespective of the neutrino mass ordering we see that the bound on τ_0 at any one mass point weakens by up to a factor of 50 relative to the naïve bound for a lightest neutrino mass m_1 (NO) or m_3 (IO) not exceeding 0.1 eV.

Even more dramatic relaxations can be seen in those cases in which the mass gap is determined by the solar mass splitting, i.e., scenarios A2 and A3. Here, the τ_0 constraints weaken by four to five orders relative to the naïve bounds in the case of IO (i.e., scenario A3) across the whole m_3 range of interest. In the case of NO (i.e., scenario A2), the relaxation also reaches five orders of magnitude, although for a smaller range of m_1 .

⁶ The present and projected constraints from IceCube and other neutrino telescopes shown in Fig. 3 assume (i) normal mass ordering, (ii) two decaying neutrinos ν_2 and ν_3 , and (iii) rest-frame lifetime-to-mass ratios satisfying $\tau_3/m_3 = \tau_2/m_2$. These assumptions imply a hierarchy of the associated transport rates, i.e., $\Gamma_{T,3 \rightarrow 1}, \Gamma_{T,3 \rightarrow 2} \gg \Gamma_{T,2 \rightarrow 1}$, and are hence largely consistent with the modelling of the NO cases of our scenarios A1 (if only one of $\nu_3 \rightarrow \nu_1$ and $\nu_3 \rightarrow \nu_2$ is present) and B1 (if both are present). Less clear, however, is how to reconcile these assumptions with those behind our scenarios A2 and A3, as well as the IO cases of scenarios A1 and B2. Nonetheless, the decay modelling of Ref. [23] suggests that switching the model assumption from two decaying and one stable neutrinos to one decaying and two stable neutrinos will not likely have a drastic effect [i.e., no more than $\mathcal{O}(1)$] on the resulting constraints on τ_0 . We therefore opt to show the neutrino telescope constraints in the middle and right panels of Fig. 3 despite the inconsistency, but caution the reader that these constraints must be taken as indicative only.

Interestingly, inclusion of the new phase space factor Φ also translates into revised lifetime constraints that are remarkably independent of the neutrino mass. We have seen previously in Sect. 4.1 that Φ asymptotes to $(1/3)(\Delta m_v^2/m_{vH}^2)^2$ in the limit $\Delta m_v^2 \equiv m_{vH}^2 - m_{vl}^2 \ll m_{vH}^2$. Since the naïve bounds on τ_0 scale effectively as m_{vH}^4 in the mass range of interest, we see immediately that including Φ must reduce the dependence of the final τ_0 bounds to $\propto (\Delta m_v^2)^2$, which, depending on the neutrino pair participating in the decay, is always fixed by either the solar or the atmospheric squared neutrino mass splitting. Indeed, we find in scenarios A2 and A3 a fairly mass-independent revised constraint of $\tau_0 \gtrsim (400 \rightarrow 500)$ s, while in scenario A1 the bound is some $(|\Delta m_{31}^2|/\Delta m_{21}^2)^2 \sim 10^3$ times tighter at $\tau_0 \gtrsim (6 \rightarrow 10) \times 10^5$ s. The two scenario B bounds are tighter still at $\tau_0^0 \gtrsim (2 \rightarrow 6) \times 10^7$ s, but apply to two decay channels with a near-common atmospheric mass gap.

In the case of A2 and A3, we note while bearing in mind the caveats of Footnote 6 that our revised CMB constraints on τ_0 are merely a factor of a few more stringent than current bounds derived in Ref. [23] from the 2015 IceCube measurements of the flavour composition of TeV–PeV-energy astrophysical neutrinos [45], assuming the flavour ratio of a full pion decay scenario, i.e., $(f_e, f_\mu, f_\tau) = (1, 2, 0)$, at the source. Indeed, according to the projections of Ref. [23], the sensitivity to the neutrino lifetime of IceCube alone based on a combination 8 years of starting events and through-going track will likely already supersede our CMB constraints entirely in scenario A3 and partially in scenario A2 in the mass range presented. Together with improved measurements of the neutrino mixing parameters by JUNO [46], DUNE [47], and Hyper-Kamiokande [48], even the A2 CMB constraints could become be entirely overtaken in the next two decades [23] by data collected at an array of future neutrino telescopes such as Baikal-GVD [49], KM3NeT [50], P-ONE [51], TAMBO [52], and IceCube-Gen2 [53]. In light of this possibility, it would be interesting to see if the neutrino telescope constraints on and projected sensitivities to τ_0 would vary significantly if derived under model assumptions more consistent with our scenarios A2 and A3. Likewise, it remains to be determined if future CMB measurements on a comparable timeline such as CMB-S4 [54] will be competitive in this space.

Lastly, one may be tempted to use Eqs. (36) and (37) to extrapolate our lifetime limits to parent neutrino mass values beyond our scan range, i.e., $m_{vH} \gtrsim 0.1$ eV. Leaving aside that such large masses might run into problems with CMB bounds, it is interesting to note the function $\mathcal{F}(x)$ naturally shuts down the anisotropy loss when the condition of ultra-relativistic decay cannot be satisfied, which translates in turn into a rapid deterioration of the lifetime bounds at large values of m_{vH} . In fact, for $m_{vH} \gtrsim 1$ eV in scenario A and $m_{vH} \gtrsim 0.5$ eV in scenario B, we expect the lifetime

bounds to deteriorate as $\sim \exp(-x)/x^{0.5}$, where $x = x_A \simeq 0.12 \times (m_{vH}/0.05 \text{ eV})$ and $x = x_B \simeq 0.2 \times (m_{vH}/0.05 \text{ eV})$ for scenarios A and B respectively. In other words, at such large m_{vH} values, there are no neutrino lifetime constraints from free-streaming requirements.

7 Conclusions

We have considered in this work invisible neutrino decay $\nu_H \rightarrow \nu_l + \phi$ and its inverse process in the ultra-relativistic limit, and derived the effective Boltzmann hierarchy and associated ℓ th anisotropy loss rate (or, equivalently, the transport rate Γ_T) for the coupled (ν_H, ν_l, ϕ) -system relevant for CMB anisotropy computations. Relative to our previous work [9] which assumed both decay products to be massless, we have in this work allowed the daughter neutrino ν_l to remain massive.

We find that a nonzero m_{vl} introduces a new phase space suppression factor $\Phi(m_{vl}/m_{vH})$ (Eq. (17)) in the transport rate, which in the limit of a small squared mass gap, $\Delta m_v^2 \equiv m_{vH}^2 - m_{vl}^2 \ll m_{vH}^2$, asymptotes to $(1/3)(\Delta m_v^2/m_{vH}^2)^2$. Thus, the transport rate Γ_T is related to the rest-frame lifetime τ_0 of the parent neutrino ν_H roughly as $\Gamma_T \sim (\Delta m_v^2/m_{vH}^2)^2 (m_{vH}/E_v)^5 (1/\tau_0)$, where the factor $(m_{vH}/E_v)^5$ was established previously in Ref. [9]. Note that while we have derived our transport rates rigorously from first principles under a reasonable set of assumptions, it is possible to attribute the various components that make up the expressions to physical quantities such as the opening angle and the momenta of the decay products. We have dedicated Appendix 1 to interpreting our transport rate in terms of a random walk using these physical quantities, as well as explaining why the old random walk argument of Ref. [4] is incomplete and hence yields a transport rate that is too large. We emphasise that the new phase space factor Φ is *in addition* to the phase space suppression already inherent in the lifetime τ_0 , and traces its origin to the momenta carried by the decay products in the decay rest frame.

Implementing our effective Boltzmann hierarchy (9) and transport rates (25) into the CMB code CLASS [28], we have performed a series of MCMC fits to the CMB temperature and polarisation anisotropy measurements by the Planck mission (2018 data release) and determined a new set of constraints on the neutrino lifetime τ_0 in a manner consistent with the neutrino mass splittings established by oscillation experiments. These new constraints are presented in Fig. 3 as functions of the lightest neutrino mass, i.e., m_1 in the normal mass ordering ($m_3 > m_2 > m_1$) and m_3 in the inverted mass ordering ($m_2 > m_1 > m_3$). For a parent neutrino mass up to about $m_{vH} \simeq 0.1$ eV, we find that the presence of the new phase space factor weakens the CMB constraint on its lifetime by as much as a factor of 50 if the daughter neutrino mass m_{vl}

is separated from $m_{\nu H}$ by the atmospheric mass gap, and by up to five orders of magnitude if separated by the solar mass gap, in comparison with naïve limits that assume $m_{\nu l} = 0$.

The revised constraints are (i) $\tau^0 \gtrsim (6 \rightarrow 10) \times 10^5$ s and $\tau^0 \gtrsim (400 \rightarrow 500)$ s if only one neutrino decays to a daughter neutrino separated by, respectively, the atmospheric and the solar mass gap, and (ii) $\tau^0 \gtrsim (2 \rightarrow 6) \times 10^7$ s in the case of two decay channels with one near-common atmospheric mass gap. Relative to existing constraints derived from other probes, these CMB limits are, despite significant revision, still globally the most stringent. However, we note that the relaxation of the bounds has also opened up a swath of parameter space likely within the projected reach of IceCube and other telescopes in the next two decades [23]. This may in turn call for closer scrutiny of the neutrino anisotropy loss modelling upon which the derivation of the CMB constraints of this work is based. For example, where cosmological and astrophysical constraints meet, a more consistent treatment of neutrinos masses and quantum statistics may be required to accurately delineate the viable decay parameter space. We defer this investigation to a future work.

Finally, we remind the reader that while we have consistently accounted for the daughter neutrino mass according to experimentally measured mass splittings, we have assumed $m_\phi = 0$ in the determination of the lifetime constraints. This assumption need not be true, and a choice of m_ϕ satisfying $m_{\nu H} > m_\phi \gg m_{\nu l}$ in the regions (i) $m_1, m_3 \lesssim 10^{-2}$ eV in scenarios A1 and B and (ii) $m_1 \lesssim 2 \times 10^{-3}$ eV in A2 may in fact also relax the cosmological constraints on τ_0 presented in Fig. 3. (Phase space considerations limit the range of permissible m_ϕ values to $m_\phi < m_{\nu H} - m_{\nu l}$ and hence their effects on the τ_0 constraints where a finite $m_{\nu l}$ already produces a significant suppression, except for a few finely-tuned values in the vicinity of $m_\phi = \Delta m_{\nu}$.) Such a relaxation may be similarly modelled using the phase space factor Φ (Eq. (17)) with $m_{\nu l} \rightarrow m_\phi$, although the resulting limits on τ_0 would depend strongly on the choice of mass for the hypothetical ϕ particle. We leave the exploration of m_ϕ -dependent neutrino lifetime constraints as an exercise to the interested reader.

Acknowledgements We thank Jan Hamann for useful discussions on random walks. JZC acknowledges support from an Australian Government Research Training Program Scholarship. IMO is supported by Fonds de la recherche scientifique (FRS-FNRS). GP is supported by a UNSW University International Postgraduate Award. Y³W is supported in part by the Australian Government through the Australian Research Council's Future Fellowship (project FT180100031). This research is enabled by the Australian Research Council's Discovery Project (project DP170102382) funding scheme, and includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This work can be fully reproduced using a modified version of the CLASS code (online repository provided) and the publicly available Planck 2018 data.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

Appendix A: Collision integrals

The individual linear-order decay/inverse decay collision integrals for the (ν_H, ν_l, ϕ) -system have been derived in Ref. [9] assuming a Yukawa interaction. Here, we present a generalised version of the result valid for any interaction structure, recast following the arguments of Sect. 2 such that the rest-frame decay rate Γ_{dec}^0 (rather than the squared matrix element $|\mathcal{M}|^2$) appears in the integrals as a fundamental parameter.

Decomposed in terms of a Legendre series, the ℓ th moments of the three collision integrals are given respectively by

$$\begin{aligned} \left(\frac{df_{\nu H}}{d\tau} \right)_{C,\ell}^{(1)}(|\mathbf{q}_1|) &= \frac{2a^2 m_{\nu H}^3 \Gamma_{\text{dec}}^0}{\Delta(m_{\nu H}, m_{\nu l}, m_\phi) \epsilon_1 |\mathbf{q}_1|} \left\{ -F_{\nu H,\ell}(|\mathbf{q}_1|) \int_{q_{2-}^{(\nu H)}}^{q_{2+}^{(\nu H)}} d|\mathbf{q}_2| \right. \\ &\quad \times \frac{|\mathbf{q}_2|}{\epsilon_2} \Omega_1 + \int_{q_{2-}^{(\nu H)}}^{q_{2+}^{(\nu H)}} d|\mathbf{q}_2| \frac{|\mathbf{q}_2|}{\epsilon_2} P_\ell(\cos \alpha^*) F_{\nu l,\ell}(|\mathbf{q}_2|) \Omega_2 \\ &\quad \left. + \int_{q_{3-}^{(\nu H)}}^{q_{3+}^{(\nu H)}} d|\mathbf{q}_3| \times \frac{|\mathbf{q}_3|}{\epsilon_3} P_\ell(\cos \beta^*) F_{\phi,\ell}(|\mathbf{q}_3|) \Omega_3 \right\}, \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} \left(\frac{df_{\nu l}}{d\tau} \right)_{C,\ell}^{(1)}(|\mathbf{q}_2|) &= \frac{2a^2 m_{\nu H}^3 \Gamma_{\text{dec}}^0}{\Delta(m_{\nu H}, m_{\nu l}, m_\phi) \epsilon_2 |\mathbf{q}_2|} \\ &\quad \left\{ \int_{q_{1-}^{(\nu l)}}^{q_{1+}^{(\nu l)}} d|\mathbf{q}_1| \frac{|\mathbf{q}_1|}{\epsilon_1} P_\ell(\cos \alpha^*) F_{\nu H,\ell}(|\mathbf{q}_1|) \Omega_1 \right. \\ &\quad - F_{\nu l,\ell}(|\mathbf{q}_2|) \int_{q_{1-}^{(\nu l)}}^{q_{1+}^{(\nu l)}} d|\mathbf{q}_1| \frac{|\mathbf{q}_1|}{\epsilon_1} \Omega_2 \\ &\quad \left. - \int_{q_{3-}^{(\nu l)}}^{q_{3+}^{(\nu l)}} d|\mathbf{q}_3| \frac{|\mathbf{q}_3|}{\epsilon_3} P_\ell(\cos \gamma^*) F_{\phi,\ell}(|\mathbf{q}_3|) \Omega_3 \right\}, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} \left(\frac{df_\phi}{d\tau} \right)_{C,\ell}^{(1)}(|\mathbf{q}_3|) &= \frac{4a^2 m_{\nu H}^3 \Gamma_{\text{dec}}^0}{\Delta(m_{\nu H}, m_{\nu l}, m_\phi) \epsilon_3 |\mathbf{q}_3|} \end{aligned}$$

$$\left\{ \int_{q_{1-}^{(\phi)}}^{q_{1+}^{(\phi)}} d|\mathbf{q}_1| \frac{|\mathbf{q}_1|}{\epsilon_1} P_\ell(\cos \beta^*) F_{vH,\ell}(|\mathbf{q}_1|) \Omega_1 \right. \\ \left. - \int_{q_{2-}^{(\phi)}}^{q_{2+}^{(\phi)}} d|\mathbf{q}_2| \times \frac{|\mathbf{q}_2|}{\epsilon_2} P_\ell(\cos \gamma^*) F_{vI,\ell}(|\mathbf{q}_2|) \Omega_2 \right. \\ \left. - F_{\phi,\ell}(|\mathbf{q}_3|) \int_{q_{1-}^{(\phi)}}^{q_{1+}^{(\phi)}} d|\mathbf{q}_1| \frac{|\mathbf{q}_1|}{\epsilon_1} \Omega_3 \right\}. \quad (\text{A.3})$$

Here,

$$\cos \alpha^* = \frac{2\epsilon_1\epsilon_2 - a^2(m_{vH}^2 + m_{vI}^2 - m_\phi^2)}{2|\mathbf{q}_1||\mathbf{q}_2|}, \\ \cos \beta^* = \frac{2\epsilon_1\epsilon_3 - a^2(m_{vH}^2 - m_{vI}^2 + m_\phi^2)}{2|\mathbf{q}_1||\mathbf{q}_3|}, \quad (\text{A.4}) \\ \cos \gamma^* = \frac{2\epsilon_2\epsilon_3 - a^2(m_{vH}^2 - m_{vI}^2 - m_\phi^2)}{2|\mathbf{q}_2||\mathbf{q}_3|}$$

originate purely from kinematics; the phase space factors including quantum statistics and their corresponding no-quantum-statistics approximations (q_s) are

$$\Omega_1(|\mathbf{q}_2|, |\mathbf{q}_3|) = 1 - \bar{f}_{vI}(|\mathbf{q}_2|) + \bar{f}_\phi(|\mathbf{q}_3|) \xrightarrow{q_s} 1, \\ \Omega_2(|\mathbf{q}_1|, |\mathbf{q}_3|) = \bar{f}_{vH}(|\mathbf{q}_1|) + \bar{f}_\phi(|\mathbf{q}_3|) \xrightarrow{q_s} \bar{f}_\phi(|\mathbf{q}_3|), \\ \Omega_3(|\mathbf{q}_1|, |\mathbf{q}_2|) = \bar{f}_{vI}(|\mathbf{q}_2|) - \bar{f}_{vH}(|\mathbf{q}_1|) \xrightarrow{q_s} \bar{f}_{vI}(|\mathbf{q}_2|); \quad (\text{A.5})$$

and

$$q_{i\pm}^{(j)} = \left| \frac{\epsilon_j \Delta(m_k, m_i, m_j) \pm |\mathbf{q}_j| \sqrt{\Delta^2(m_k, m_i, m_j) + 4m_i^2 m_j^2}}{2m_j^2} \right| \quad (\text{A.6})$$

are the integration limits, with the understanding that $i \neq j \neq k$, and the mapping $(1, 2, 3) \leftrightarrow (v_H, v_I, \phi)$ applies.

Appendix B: $\ell \geq 2$ effective collision integrals

We outline in this appendix the reduction of the $\ell \geq 2$ collision integrals (10) to the form (13). Firstly, we note that the effective collision integral (10) is a sum of three terms, which we label

$$\left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell} \equiv \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(vH)} + \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(vI)} + \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(\phi)}. \quad (\text{B.1})$$

Using the v_H collision integral (A.1) and setting $m_\phi = 0$, the v_H component evaluates straightforwardly to

$$\left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(vH)}$$

$$= a \tilde{\Gamma}_{\text{dec}} \left(\frac{m_{vH}^2}{m_{vH}^2 - m_{vI}^2} \right) T_0^{-4} \mathcal{F}_\ell \int d|\mathbf{q}_1| \frac{|\mathbf{q}_1|^{\ell+1}}{\epsilon_1^\ell} e^{-\epsilon_1/T_0} \\ \times \left\{ \int_{q_{2-}^{(vH)}}^{q_{2+}^{(vH)}} d|\mathbf{q}_2| \frac{|\mathbf{q}_2|}{\epsilon_2} \left[-\frac{|\mathbf{q}_1|^2}{\epsilon_1} + \frac{|\mathbf{q}_2|^2}{\epsilon_2} P_\ell(\cos \alpha^*) \right] \right. \\ \left. + \int_{q_{3-}^{(vH)}}^{q_{3+}^{(vH)}} d|\mathbf{q}_3| |\mathbf{q}_3| P_\ell(\cos \beta^*) \right\}, \quad (\text{B.2})$$

where $\tilde{\Gamma}_{\text{dec}}$ is defined in Eq. (14) and is approximately the boosted decay rate.

Observe that the ℓ -dependence of the double integral (B.2) is contained entirely in (i) $|\mathbf{q}_1|^\ell/\epsilon_1^\ell$, and (ii) the Legendre polynomials. Upon a small parameter expansion in $a^2 m_{vH}^2$ and $a^2 m_{vI}^2$ to $\mathcal{O}(a^4 m_{vH}^4, a^4 m_{vI}^4, a^4 m_{vH}^2 m_{vI}^2)$, the former becomes a quadratic polynomial in ℓ :

$$\left(\frac{|\mathbf{q}_1|}{\epsilon_1} \right)^\ell \simeq 1 - \ell \frac{a^2 m_{vH}^2}{2|\mathbf{q}_1|^2} + \ell(\ell+2) \frac{a^4 m_{vI}^4}{8|\mathbf{q}_1|^4} + \dots \quad (\text{B.3})$$

For the benchmark $m_{vH} \simeq 0.2$ eV and a typical comoving momentum of $|\mathbf{q}_1| \simeq 3T_{v,0} \simeq 5 \times 10^{-4}$ eV, this expansion is accurate at recombination ($z = z_{\text{rec}} \simeq 1100$) to 0.2% and 11% for $\ell = 2$ and $\ell = 10$, respectively, and improves to 0.03% and 1% at $z \simeq 1500$. For the largest m_{vH} value actually analysed in this work, i.e., $m_{vH} \simeq 0.1$ eV, sub-10% accuracy is possible even for $\ell = 20$ at $z \simeq z_{\text{rec}}$. The small parameter expansion therefore appears well justified in the time frame of interest.

On the other hand, the arguments of the Legendre polynomials expand to

$$\cos \alpha^* \simeq 1 + \frac{a^2 m_{vH}^2}{2|\mathbf{q}_1|^2} \left(1 - \frac{|\mathbf{q}_1|}{|\mathbf{q}_2|} \right) + \frac{a^2 m_{vI}^2}{2|\mathbf{q}_2|^2} \left(1 - \frac{|\mathbf{q}_2|}{|\mathbf{q}_1|} \right) \\ + \frac{a^4 m_{vH}^2 m_{vI}^2}{4|\mathbf{q}_1|^2 |\mathbf{q}_2|^2} - \frac{a^4 m_{vH}^4}{8|\mathbf{q}_1|^4} - \frac{a^4 m_{vI}^4}{8|\mathbf{q}_2|^4} + \dots, \\ \cos \beta^* \simeq 1 + \frac{a^2 m_{vH}^2}{2|\mathbf{q}_1|^2} \left(1 - \frac{|\mathbf{q}_1|}{|\mathbf{q}_3|} \right) \\ + \frac{a^2 m_{vI}^2}{2|\mathbf{q}_1||\mathbf{q}_3|} - \frac{a^4 m_{vH}^4}{8|\mathbf{q}_1|^4} + \dots, \quad (\text{B.4})$$

while the Legendre polynomials themselves expand again to polynomials in ℓ :

$$P_\ell(1+x) \simeq 1 + \frac{1}{2} \ell(\ell+1)x \\ + \frac{1}{16} (\ell-1)\ell(\ell+1)(\ell+2)x^2 + \dots \quad (\text{B.5})$$

Substituting Eq. (B.4) into Eq. (B.5) and keeping terms to $\mathcal{O}(a^4 m_{vH}^4, a^4 m_{vI}^4, a^4 m_{vH}^2 m_{vI}^2)$, we immediately find in combination with Eq. (B.3) that the double integral (B.2)

must be a quartic polynomial in ℓ with ℓ -independent coefficients. We shall not write out the polynomial in full here, but suffice it to say that this is the origin of quartic ℓ -dependence of the α_ℓ coefficients (18). As with Eq. (B.3), for the benchmark $m_{vH} \simeq 0.2$ eV and typical comoving momenta $|\mathbf{q}_1| \simeq |\mathbf{q}_2| \simeq 3T_{v,0} \simeq 5 \times 10^{-4}$ eV, the small parameter expansion (B.5) is sub-percent accurate at $z \simeq z_{\text{rec}}$ for $\ell \leq 5$ and incurs at most an error of 15% for $\ell = 10$. It is thus a very good approximation.

Similarly to Eq. (B.8), using the v_l collision integral (A.2) under the assumption of $m_\phi = 0$ yields

$$\begin{aligned} & \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(vl)} \\ &= a \tilde{\Gamma}_{\text{dec}} \left(\frac{m_{vH}^2}{m_{vH}^2 - m_{vl}^2} \right) T_0^{-4} \\ & \times \frac{g_{vl}}{g_{vH}} \mathcal{F}_\ell \int d|\mathbf{q}_2| \frac{|\mathbf{q}_2|^{\ell+1}}{\epsilon_2^\ell} \\ & \times \left\{ \int_{q_{1-}^{(vl)}}^{q_{1+}^{(vl)}} d|\mathbf{q}_1| \frac{|\mathbf{q}_1|}{\epsilon_1} \left[\frac{|\mathbf{q}_1|^2}{\epsilon_1} P_\ell(\cos \alpha^*) - \frac{|\mathbf{q}_2|^2}{\epsilon_2} \right] e^{-\epsilon_1/T_0} \right. \\ & \left. - \int_{q_{3-}^{(vl)}}^{q_{3+}^{(vl)}} d|\mathbf{q}_3| |\mathbf{q}_3| P_\ell(\cos \gamma^*) e^{-\epsilon_1/T_0} \right\}. \end{aligned} \quad (\text{B.6})$$

To progress further, following Ref. [9] we swap the order of integration, so that the final momentum integration to perform is always over $|\mathbf{q}_1| \in [0, \infty)$. The swap between the $|\mathbf{q}_1|$ and $|\mathbf{q}_2|$ double integral is straightforward, and entails setting the new integration limits as per

$$\int d|\mathbf{q}_2| \int_{q_{1-}^{(vl)}}^{q_{1+}^{(vl)}} d|\mathbf{q}_1| = \int d|\mathbf{q}_1| \int_{q_{2-}^{(vH)}}^{q_{2+}^{(vH)}} d|\mathbf{q}_2|. \quad (\text{B.7})$$

To apply the same trick to the $|\mathbf{q}_2|$ and $|\mathbf{q}_3|$ pair, however, we first need to change of the integration variable $|\mathbf{q}_3|$ to $|\mathbf{q}_1|$ via $|\mathbf{q}_3| \rightarrow \epsilon_1 - \epsilon_2$. Then, recognising that integration limits can also be written as Heaviside functions, we can establish the relation [9]

$$\begin{aligned} & \Theta(|\mathbf{q}_3| - q_{3-}^{(vl)}) \Theta(q_{3+}^{(vl)} - |\mathbf{q}_3|) \Big|_{|\mathbf{q}_3| \rightarrow \epsilon_1 - \epsilon_2} \\ &= \Theta(1 - \cos^2 \gamma^*) \Big|_{|\mathbf{q}_3| \rightarrow \epsilon_1 - \epsilon_2} \\ &= \Theta(1 - \cos^2 \alpha^*) \\ &= \Theta(|\mathbf{q}_2| - q_{2-}^{(vH)}) \Theta(q_{2+}^{(vH)} - |\mathbf{q}_2|). \end{aligned} \quad (\text{B.8})$$

Note that, while we have assumed $m_\phi = 0$ to simplify our calculations, the relation (B.8) follows from kinematics and holds for all mass values *provided* we define the change of variable more generally as $\epsilon_3 \rightarrow \epsilon_1 - \epsilon_2$. Thus, Eq. (B.6) becomes

$$\left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(vl)}$$

$$\begin{aligned} &= a \tilde{\Gamma}_{\text{dec}} \left(\frac{m_{vH}^2}{m_{vH}^2 - m_{vl}^2} \right) T_0^{-4} \mathcal{F}_\ell \int d|\mathbf{q}_1| \frac{|\mathbf{q}_1|^{\ell+1}}{\epsilon_1^\ell} e^{-\epsilon_1/T_0} \\ & \times \int_{q_{2-}^{(vH)}}^{q_{2+}^{(vH)}} d|\mathbf{q}_2| \frac{\epsilon_1^{\ell-2}}{|\mathbf{q}_1|^{\ell-2}} \frac{|\mathbf{q}_2|^{\ell+1}}{\epsilon_2^\ell} \left[P_\ell(\cos \alpha^*) \right. \\ & \left. - \frac{|\mathbf{q}_2|^2 \epsilon_1}{\epsilon_2 |\mathbf{q}_1|^2} - \frac{\epsilon_1}{|\mathbf{q}_1|^2} (\epsilon_1 - \epsilon_2) P_\ell(\cos \gamma^*) \Big|_{|\mathbf{q}_3| \rightarrow \epsilon_1 - \epsilon_2} \right], \end{aligned} \quad (\text{B.9})$$

where we have used $g_{vl} = g_{vH} = 2$. Again, the term $|\mathbf{q}_2|^\ell / \epsilon_2^\ell$ can be expanded in the small parameters $a^2 m_{vH}^2$ and $a^2 m_{vl}^2$ to $\mathcal{O}(a^4 m_{vH}^4, a^4 m_{vl}^4, a^4 m_{vH}^2 m_{vl}^2)$ as per Eq. (B.3) save for the replacements $m_1 \rightarrow m_2$ and $|\mathbf{q}_1| \rightarrow |\mathbf{q}_2|$, while the argument of the second Legendre polynomial expands to

$$\begin{aligned} \cos \gamma^* &\simeq 1 - \frac{a^2 m_{vH}^2}{2|\mathbf{q}_2||\mathbf{q}_3|} + \frac{a^2 m_{vl}^2}{2|\mathbf{q}_2|^2} \left(1 + \frac{|\mathbf{q}_2|}{|\mathbf{q}_3|} \right) \\ &\quad - \frac{a^4 m_{vl}^4}{8|\mathbf{q}_2|^4} + \dots \end{aligned} \quad (\text{B.10})$$

Then, by Eq. (B.5) we can again expect the ℓ -dependence of the double integral (B.9) to simplify to that of a quartic polynomial in ℓ .

Finally, the ϕ component of the effective collision integral can be similarly established from the ϕ collision integral (A.3) following the same procedure as above:

$$\begin{aligned} & \left(\frac{d\mathcal{F}}{d\tau} \right)_{C,\ell}^{(\phi)} \\ &= a \tilde{\Gamma}_{\text{dec}} \left(\frac{m_{vH}^2}{m_{vH}^2 - m_{vl}^2} \right) T_0^{-4} \mathcal{F}_\ell \int d|\mathbf{q}_1| \frac{|\mathbf{q}_1|^{\ell+1}}{\epsilon_1^\ell} e^{-\epsilon_1/T_0} \\ & \times \int_{q_{3-}^{(vH)}}^{q_{3+}^{(vH)}} d|\mathbf{q}_3| \frac{\epsilon_1^{\ell-2}}{|\mathbf{q}_1|^{\ell-2}} |\mathbf{q}_3| \left[P_\ell(\cos \beta^*) - \frac{|\mathbf{q}_3| \epsilon_1}{|\mathbf{q}_1|^2} \right. \\ & \left. - \frac{\epsilon_1}{|\mathbf{q}_1|^2} \frac{(\epsilon_1 - |\mathbf{q}_3|)^2 - a^2 m_{vl}^2}{(\epsilon_1 - |\mathbf{q}_3|)} P_\ell(\cos \gamma^*) \Big|_{\epsilon_2 \rightarrow \epsilon_1 - |\mathbf{q}_3|} \right], \end{aligned} \quad (\text{B.11})$$

with $g_\phi = 1$ as an input. A quartic polynomial in ℓ again follows from a small parameter expansion in $a^2 m_{vH}^2$ and $a^2 m_{vl}^2$ to $\mathcal{O}(a^4 m_{vH}^4, a^4 m_{vl}^4, a^4 m_{vH}^2 m_{vl}^2)$.

Then, collecting all terms (B.2), (B.9), and (B.11) yields the effective collision integral (13).

Appendix C: A revised random walk picture of isotropisation

Appendix C.1: Preliminaries

Consider the process $v_H \rightarrow v_l + \phi$ in the rest-frame of v_H , which we denote $\mathcal{S}'$. Suppose the decay products are emitted along the y -axis. Assuming $m_\phi = 0$, the decay products have

momentum

$$p^* = \frac{m_{\nu H}^2 - m_{\nu l}^2}{2m_{\nu H}} \equiv \frac{\Delta m_v^2}{2m_{\nu H}}. \quad (\text{C.1})$$

Boosting the system in the x -direction to the laboratory frame $\mathcal{S}$ at an ultra-relativistic speed $v_{\nu H}$, we find that the decay product i has momentum components in the x - and y -directions given respectively by

$$\begin{aligned} p_x^{(i)} &= \gamma_{\nu H} v_{\nu H} \sqrt{p^{*2} + m_i^2}, \\ p_y^{(i)} &= \pm p^*, \end{aligned} \quad (\text{C.2})$$

where $\gamma_{\nu H} \equiv (1 - v_{\nu H}^2)^{-1/2} = E_{\nu H}/m_{\nu H}$ is the Lorentz factor, and $\tan \theta_i = p_y^{(i)}/p_x^{(i)}$ gives the angle of the decay trajectory relative to the initial direction of ν_H . For the massless ϕ particle, this angle always evaluates $|\theta_\phi| \simeq \gamma_{\nu H}^{-1}$ in the limit $\gamma_{\nu H} \gg 1$, while for a massive daughter neutrino ν_l we find a suppressed emission angle of $|\theta_{\nu l}| \simeq [\Delta m_v^2/(2m_{\nu H}^2)] \gamma_{\nu H}^{-1}$ in the same limit. The latter, or more directly, p^* in Eq. (C.1), is the origin of the new phase space factor (17) in the effective transport rate (13).

In the case of *two massless* decay products, Ref. [4] argues that, because of the small decay opening angle, $\theta = \theta_\phi = \theta_{\nu l} \simeq \gamma_{\nu H}^{-1} \ll 1$, it takes $N \simeq 1/\theta^2$ number of decay/inverse decay events to turn a momentum vector by 90° . This estimate comes from a standard random walk argument generalised to momentum space, where the random walker has a mean square displacement of $\sim (p_{\nu H} \theta_i)^2$ and *equal probabilities* of going $+\theta_i$ or $-\theta_i$ after each decay or inverse decay event, while maintaining its absolute momentum $p_{\nu H}$. Assuming further that each event occurs over a timescale of $\sim 1/\Gamma_{\text{dec}}$, where $\Gamma_{\text{dec}} = \gamma_{\nu H}^{-1} \Gamma_{\text{dec}}^0$ is the boosted decay rate, one immediately establishes that the time required to execute this 90-degree turn must be

$$T_2 \sim \theta^{-2} \gamma / \Gamma_{\text{dec}}^0. \quad (\text{C.3})$$

Reference [4] identifies T_2 with the timescale over which the quadrupole of the (ν_H, ν_l, ϕ) -system is lost through decay/inverse decay, in drastic contrast with the much longer isotropisation timescale,

$$T_4 \sim \theta^{-4} \gamma / \Gamma_{\text{dec}}^0, \quad (\text{C.4})$$

established from our relaxation time/linear response calculation in the same $m_{\nu l} = 0$ limit [9].

It is not completely clear how one might generalise T_2 in Eq. (C.3) to the case of a massive ν_l , since the two decay products must now be emitted at different angles. One possibility might be to identify the emission angle θ with the geometric mean, $\theta \equiv \pm \sqrt{|\theta_{\nu l} \theta_\phi|}$. At minimum, this choice of θ yields a T_4 from Eq. (C.4) to within $\mathcal{O}(1)$ factors of the timescale obtained via rigorous calculations in this work (Eq. (13)). This identification also appears to be the only sensible choice if we must boil what is in principle a two-parameter problem

(for two decay products) down to a description in terms of one angle θ alone. Our task in this appendix, therefore, is to reconcile the two timescales T_2 and T_4 , and to supply a random walk interpretation for the latter in terms of θ .

Appendix C.2: Coverage versus isotropisation

To reconcile the two timescales T_2 and T_4 , and to understand why it is the latter, not the former, that gives the isotropisation timescale, we observe first of all that as a representation of isotropisation of an anisotropic system via scattering, the random walk picture as presented above in fact contains a number of inconsistencies. For one, the same reasoning would lead us to conclude that the momentum vector of a single particle can be turned by 180° in time $\sim 4 \times T_2$. This is in itself not an improbable outcome. It makes no physical sense, however, to associate this outcome with the timescale over which the *dipole* of the (ν_H, ν_l, ϕ) -system is lost through decay/inverse decays, because as an isolated system, momentum conservation always preserves the system's dipole. This simple consideration illustrates that one needs to be extremely cautious when identifying the behaviour of *one single particle* with changes to the *bulk properties* of a system. They are *a priori* not the same thing.

Next, we note that ν_H decay always happens and isotropisation is limited only by inverse decay. However, in writing down Γ_{dec} as a universal *inverse* decay rate without qualification, there is an implicit assumption that the scattering centres effecting the inverse decay are themselves isotropic in momentum space, such that the probability of ν_H being emitted in the $+\theta_i$ or $-\theta_i$ direction is everywhere the same. This assumption is not true for our (ν_H, ν_l, ϕ) -system, which is constantly subject to tidal gravitational forces that generate $\mathcal{O}(10^{-5})$ local momentum anisotropies. These anisotropies give rise to anisotropic probabilities in the emission direction of ν_H , which in turn bias the random walk. At face value this bias is tiny—of order 10^{-5} . However, we must keep in mind that it has the same physical origin as the momentum anisotropy we wish to wipe by inverse decay, i.e., the bias and the momentum anisotropy are of *exactly the same order of magnitude*. A consistent picture of isotropisation therefore *cannot* ignore the bias, and understanding how it influences the random walks is key to explaining why the isotropisation timescale is T_4 , not T_2 .

To simplify the discussion, let us separate the system into a population of “walkers” and a population of “background” particles that recombine with the walkers to form ν_H . As illustrated in Fig. 4, when a walker encounters a background with a quadrupole anisotropy, strict angular and momentum requirements for an inverse decay to happen favour the resulting ν_H to be emitted in a direction as closely aligned as possible with the principal axis of anisotropy. The alignment is however never exact, as ν_H is always emitted at an angle

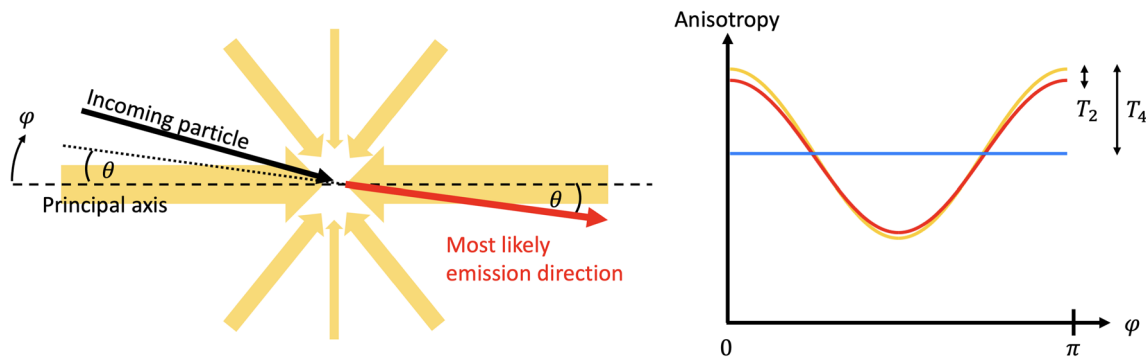


Fig. 4 Schematic illustrating how a quadrupole anisotropy in the background population can enhance the probability of v_H being emitted in a direction as aligned as possible with the principal axis of anisotropy. Observe in this example that the emitted v_H is misaligned with the principal axis of anisotropy by an angle $\sim \theta$. This small misalignment leads to a distribution of random walkers after one coverage time T_2 (Eq. (C.3)) that is slightly dispersed relative to the distribution of the background population, i.e., the walker population (red line in the right

plot) is slightly isotropised relative to the background population (yellow line) after one coverage time T_2 . Then, by repeatedly updating the background anisotropy with the new walker population anisotropy (i.e., we “swap” the roles of the walker and the background populations) after every T_2 , we will eventually after $M \sim 1/\theta^2$ updates end up with a system that is completely isotropised (blue line). The total time this isotropisation process takes is $T_4 \sim M \times T_2$ (Eq. (C.4))

$\sim \theta$ relative to both the trajectory of the walker and of the background particle that combines with it. In other words, we can think of the random walkers as “samplers” of an “emission direction function” $\mathcal{E}$ that has a gradient in the angular direction following the anisotropy of the background, but with a small isotropic dispersion over the emission angle θ —isotropic because the inverse decay interaction itself is invariant under rotation; any asymmetry in the v_H emission direction is due solely to the background anisotropies. This small isotropic dispersion is important, as it is precisely this dispersion that allows the $\ell \geq 2$ anisotropies of the system to be lost through decay/inverse decay, while preserving the dipole.

Now, as mentioned above, anisotropies in the cosmological context are typically no more than $\sim 10^{-5}$. We can therefore still expect a random walker to cover all angular directions in $N \sim 1/\theta^2$ steps, or, equivalently, in a time T_2 : for this reason we shall call T_2 the *coverage time*. It is also over this timescale that the random walker “forgets” where it has come from. However, if we were to ask, what would happen to a whole population of random walkers after one coverage time T_2 , then the answer would be we would find a walker population with no memory of its initial configuration but that is now distributed in a way that mimics the emission direction function $\mathcal{E}$, itself dictated by the underlying anisotropy of the background. Thus, we have not managed to wipe out the system’s anisotropy over time T_2 ; we have merely sampled the emission direction function $\mathcal{E}$ and hence turned the random walker population into a slightly less anisotropic version of the background. At this point the reader may find our description reminiscent of a Markov Chain. It is indeed a Markov Chain, where the emission direction function $\mathcal{E}$ plays the role of the likelihood function, and T_2 is the burn-in time.

We can easily extend this picture to include updates to the emission direction function $\mathcal{E}$ as the system loses its anisotropy, in order to estimate the isotropisation timescale, i.e., the time it takes to flatten $\mathcal{E}$. In reality the update must of course take place continuously, as the loss of anisotropy is itself a continuous process. In our discretised picture with one walker and one background population, however, we can mimic the update by swapping the roles of the walkers and the background after every coverage time T_2 . That is, at every swap the new emission direction function $\mathcal{E}_{\text{new}}$ is an isotropically smeared version of the old function $\mathcal{E}_{\text{old}}$ over the emission angle θ given by

$$\mathcal{E}_{\text{new}}(\varphi) = \int d\varphi' \mathcal{E}_{\text{old}}(\varphi') f(\varphi' - \varphi, \theta), \quad (\text{C.5})$$

where $f(\varphi, \theta)$ is a normalised filter function of width $\sim \theta$. Then, for the quadrupole anisotropy of Fig. 4—given by $\mathcal{E}_{\text{old}} = (3 \cos^2 \varphi - 1)/2$ —and assuming a top-hat filter $f(\varphi, \theta) = \Theta(\varphi + \theta)\Theta(\theta - \varphi)/(2\theta)$, we see immediately that at its maximum point $\mathcal{E}_{\text{old}}(\varphi = 0) = 1$, the new emission direction function $\mathcal{E}_{\text{new}}(\varphi = 0)$ must reduce to

$$\begin{aligned} \mathcal{E}_{\text{new}}(\varphi = 0) &= \frac{1}{4} + \frac{3 \sin(2\theta)}{8\theta} \\ &\simeq 1 - \frac{\theta^2}{2}, \quad \theta \ll 1 \end{aligned} \quad (\text{C.6})$$

after the first swap, i.e., it reduces by an amount $\theta^2/2$; had we used a normalised Gaussian filter of width θ , we would have obtained a similar reduction of $3\theta^2/2$. Thus, we arrive at the conclusion that the number of swaps required to flatten $\mathcal{E}$ must be $M \sim 1/\theta^2$. Then, together with the coverage timescale T_2 from Eq. (C.3), we find an isotropisation

timescale of $\sim M \times T_2 \sim \theta^{-4} \gamma / \Gamma_{\text{dec}}^0$, which is precisely T_4 of Eq. (C.4).

In conclusion, the random walk argument employed in Ref. [4] to support an isotropisation timescale of T_2 (Eq. (C.3)) carries a strong but unjustified assumption that the background on which a random walker scatters is *already isotropised*. Once this assumption is dropped, T_2 is no longer sufficient time for isotropisation, and should be reinterpreted as the coverage time for a random walker to survey the system's configuration once and to forget its previous state, while the anisotropy decreases only by a small amount θ^2 . The true isotropisation timescale is a much longer T_4 (Eq. (C.4)), stemming from the need for each random walker to survey and forget its own previous state another $M \sim 1/\theta^2$ times, before all walkers will agree on their final configurations.

Appendix D: Parameter estimation

We provide in this section summary statistics of our Markov Chains from Sect. 5.2. Table 3 shows the mean and 1D marginalised credible intervals for the variables of Eqs. (28) (scenario A) and (29) (scenario B), as well as for two derived parameters, the reduced Hubble expansion rate h and the RMS fluctuation on 8 Mpc σ_8 , for all choices of fixed $X = X_{\text{scan}}$ considered. For comparison we also provide the parameter estimates in a vanilla Λ CDM fit to the same data combination taken from Ref. [1]. Figures 5 and 6 show the corresponding 2D marginalised contours.

Table 3 Mean and 1D marginalised 68% credible intervals for the base Λ CDM parameters and the marginalised one-sided 95% credible limits on the effective transport rate Y derived from the Planck 2018 CMB data combination TTTEEE+lowE+lensing [1], for fixed values of the effective mass parameter X in scenarios A and B. The vanilla Λ CDM estimates are taken from Ref. [1]

	Scenario A		
	$X = 52$	$X = 300$	$X = 595$
ω_b	0.02249 ± 0.00016	0.02248 ± 0.00016	0.02249 ± 0.00016
ω_c	0.1208 ± 0.0013	0.1208 ± 0.0013	0.1208 ± 0.0012
$100\theta_s$	1.04178 ± 0.00029	1.04181 ± 0.00030	1.04179 ± 0.00030
τ_{reio}	0.0561 ± 0.0074	0.0560 ± 0.0080	0.0567 ± 0.00755
n_s	0.9747 ± 0.0045	0.9751 ± 0.0049	0.9759 ± 0.0049
$\ln(10^{10} A_s)$	3.0424 ± 0.0145	3.0431 ± 0.0154	3.0443 ± 0.0149
$Y(s^{-1})$	< 18611	< 41636	< 74448
h	0.6768 ± 0.0557	0.6768 ± 0.0568	0.6768 ± 0.0538
σ_8	0.8276 ± 0.0063	0.8280 ± 0.0064	0.8287 ± 0.0062
	Scenario B		Λ CDM [1]
	$X = 298$	$X = 595$	
ω_b	0.02229 ± 0.00015	0.02227 ± 0.00015	0.02237 ± 0.00015
ω_c	0.1200 ± 0.00122	0.1201 ± 0.0012	0.1200 ± 0.0012
$100\theta_s$	1.04188 ± 0.00030	1.04191 ± 0.00029	1.04092 ± 0.00031
τ_{reio}	0.0545 ± 0.0072	0.0539 ± 0.0072	0.0544 ± 0.0073
n_s	0.9690 ± 0.0044	0.9674 ± 0.0043	0.9649 ± 0.0042
$\ln(10^{10} A_s)$	3.0424 ± 0.0138	3.0423 ± 0.0140	3.044 ± 0.014
$Y(s^{-1})$	< 2286	< 2288	–
h	0.6780 ± 0.5592	0.6778 ± 0.5592	0.6736 ± 0.0054
σ_8	0.8245 ± 0.0060	0.8243 ± 0.0060	0.8111 ± 0.0060

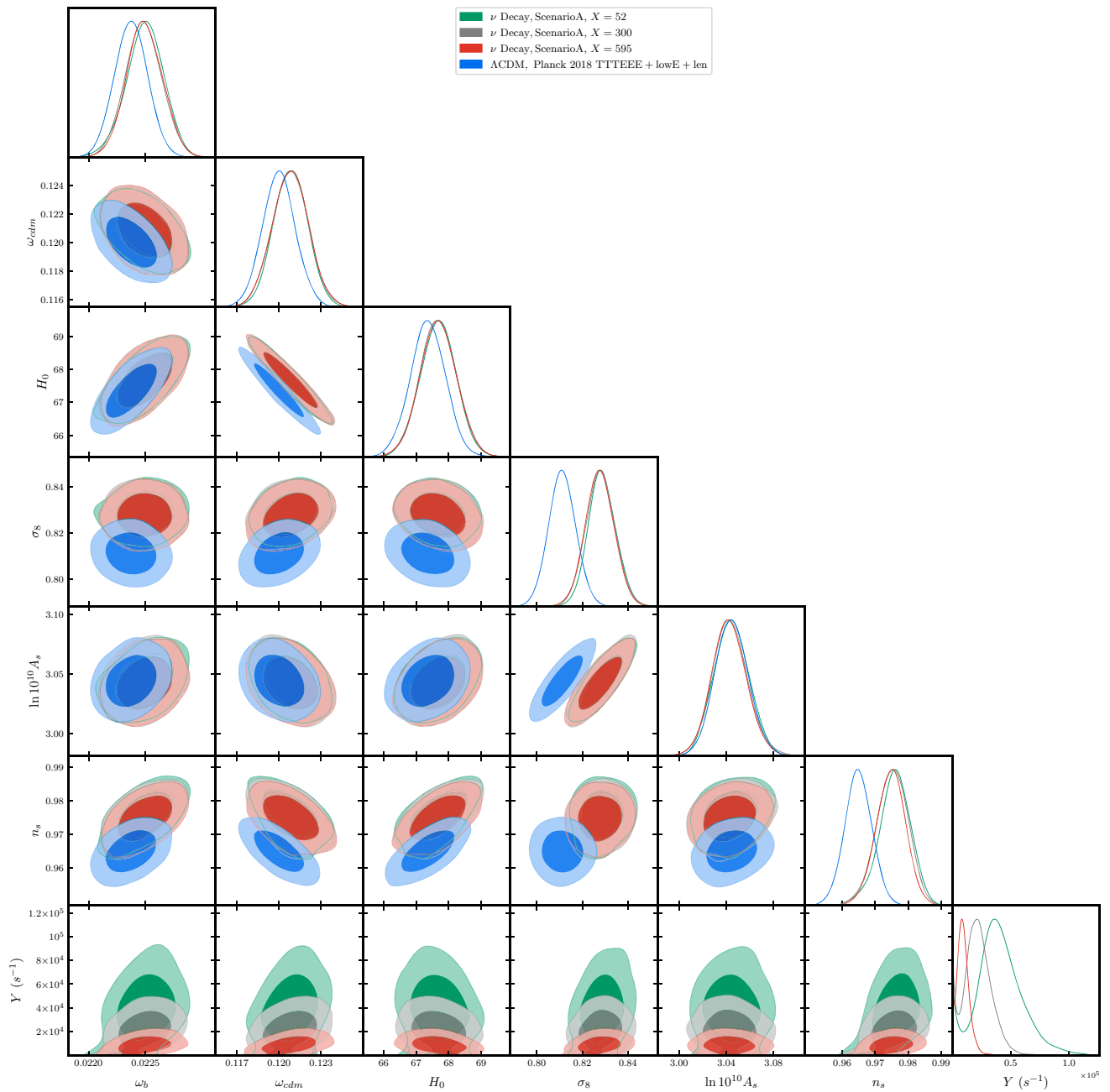


Fig. 5 2D marginalised 68% and 95% contours for the variables of Scenario A, for fixed values of the effective mass parameter X . We show also the corresponding contours from a vanilla Λ CDM fit

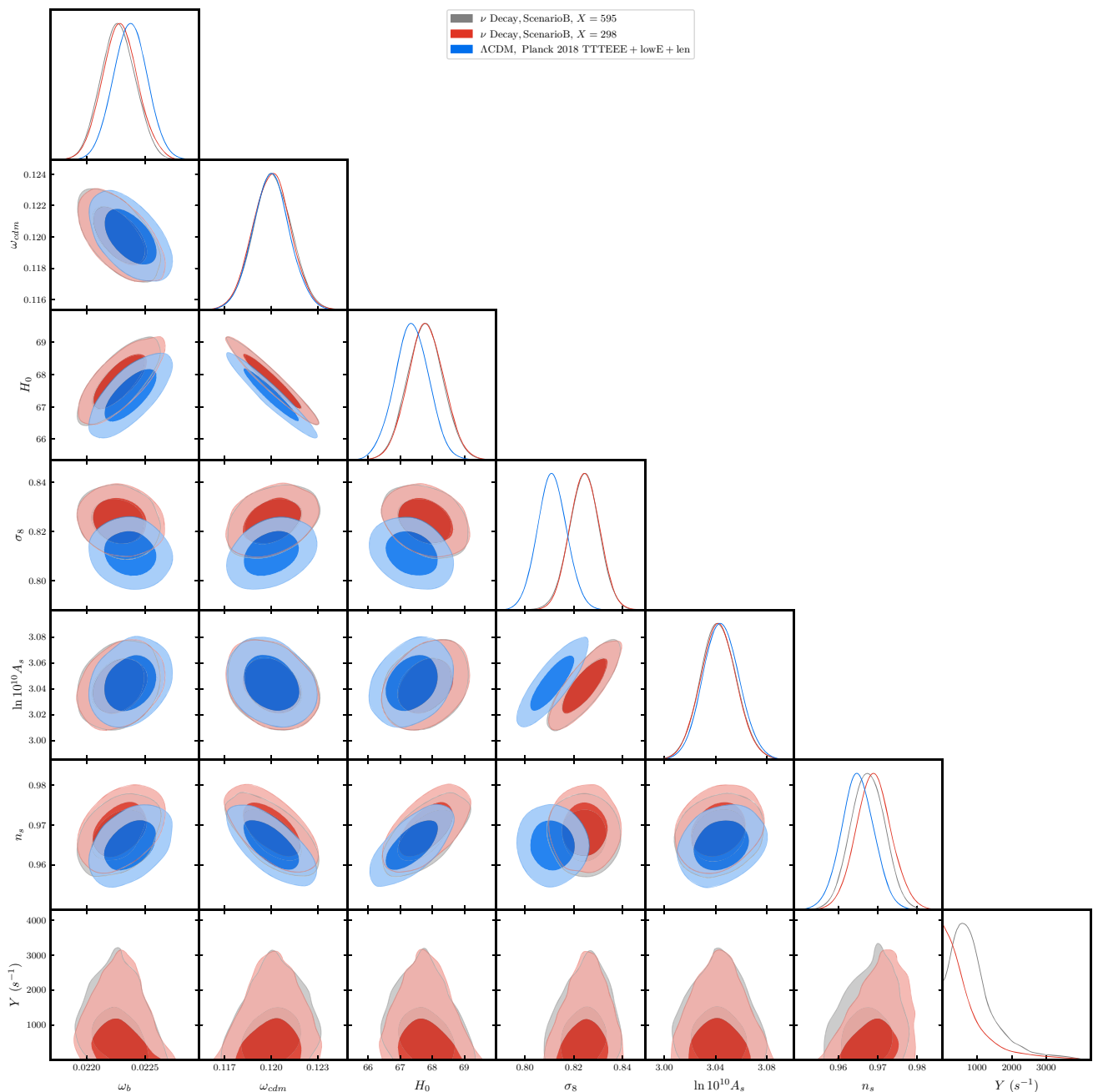


Fig. 6 Same as Fig. 5, but for scenario B

References

1. Planck Collaboration, N. Aghanim et al., Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* **641**, A6 (2020). [arXiv:1807.06209](#) [astro-ph.CO]. [Erratum: *Astron. Astrophys.* 652, C4 (2021)]
2. KATRIN Collaboration, M. Aker et al. Improved upper limit on the neutrino mass from a direct kinematic method by KATRIN. *Phys. Rev. Lett.* **123**(22), 221802 (2019). [arXiv:1909.06048](#) [hep-ex]
3. S. Hannestad, Structure formation with strongly interacting neutrinos - Implications for the cosmological neutrino mass bound. *JCAP* **02**, 011 (2005). [arXiv:astro-ph/0411475](#)
4. S. Hannestad, G. Raffelt, Constraining invisible neutrino decays with the cosmic microwave background. *Phys. Rev. D* **72**, 103514 (2005). [arXiv:hep-ph/0509278](#) [hep-ph]
5. A. Basboll, O.E. Bjælde, S. Hannestad, G.G. Raffelt, Are cosmological neutrinos free-streaming? *Phys. Rev. D* **79**, 043512 (2009). [arXiv:0806.1735](#) [astro-ph]
6. F.-Y. Cyr-Racine, K. Sigurdson, Limits on neutrino-neutrino scattering in the early universe. *Phys. Rev. D* **90**(12), 123533 (2014). [arXiv:1306.1536](#) [astro-ph.CO]
7. M. Archidiacono, S. Hannestad, Updated constraints on non-standard neutrino interactions from Planck. *JCAP* **1407**, 046 (2014). [arXiv:1311.3873](#) [astro-ph.CO]

8. M. Escudero, M. Fairbairn, Cosmological constraints on invisible neutrino decays revisited. *Phys. Rev. D* **100**(10), 103531 (2019). [arXiv:1907.05425](#) [hep-ph]
9. G. Barenboim, J.Z. Chen, S. Hannestad, I.M. Oldengott, T. Tram, Y.Y.Y. Wong, Invisible neutrino decay in precision cosmology. *JCAP* **03**, 087 (2021). [arXiv:2011.01502](#) [astro-ph.CO]
10. L. Lancaster, F.-Y. Cyr-Racine, L. Knox, Z. Pan, A tale of two modes: Neutrino free-streaming in the early universe. *JCAP* **07**, 033 (2017). [arXiv:1704.06657](#) [astro-ph.CO]
11. I. M. Oldengott, C. Rampf, Y. Y. Y. Wong, Boltzmann hierarchy for interacting neutrinos I: formalism. *JCAP* **1504**(04), 016 (2015). [arXiv:1409.1577](#) [astro-ph.CO]
12. I. M. Oldengott, T. Tram, C. Rampf, Y. Y. Y. Wong, “Interacting neutrinos in cosmology: exact description and constraints. *JCAP* **1711**(11), 027 (2017). [arXiv:1706.02123](#) [astro-ph.CO]
13. C. D. Kreisch, F.-Y. Cyr-Racine, O. Doré, Neutrino puzzle: Anomalies, interactions, and cosmological tensions. *Phys. Rev. D* **101**(12), 123505 (2020). [arXiv:1902.00534](#) [astro-ph.CO]
14. F. Forastieri, M. Lattanzi, P. Natoli, Cosmological constraints on neutrino self-interactions with a light mediator. *Phys. Rev. D* **100**(10), 103526 (2019). [arXiv:1904.07810](#) [astro-ph.CO]
15. G. Barenboim, P. B. Denton, I. M. Oldengott, Constraints on inflation with an extended neutrino sector. *Phys. Rev. D* **99**(8), 083515 (2019). [arXiv:1903.02036](#) [astro-ph.CO]
16. J. Venzor, G. Garcia-Arroyo, A. Pérez-Lorenzana, J. De-Santiago, Massive neutrino self-interactions with a light mediator in cosmology. [arXiv:2202.09310](#) [astro-ph.CO]
17. C. S. Lorenz, L. Funcke, M. Löffler, E. Calabrese, Reconstruction of the neutrino mass as a function of redshift. *Phys. Rev. D* **104**(12), 123518 (2021). [arXiv:2102.13618](#) [astro-ph.CO]
18. G. F. Abellán, Z. Chacko, A. Dev, P. Du, V. Poulin, Y. Tsai, Improved cosmological constraints on the neutrino mass and lifetime. [arXiv:2112.13862](#) [hep-ph]
19. Y. P. Porto-Silva, S. Prakash, O. L. G. Peres, H. Nunokawa, H. Minakata, Constraining visible neutrino decay at KamLAND and JUNO. *Eur. Phys. J. C* **80**(10), 999 (2020). [arXiv:2002.12134](#) [hep-ph]
20. SNO Collaboration, B. Aharmim et al., Constraints on neutrino lifetime from the sudbury neutrino observatory. *Phys. Rev. D* **99**(3), 032013 (2019). [arXiv:1812.01088](#) [hep-ex]
21. L. Funcke, G. Raffelt, E. Vitagliano, Distinguishing dirac and majorana neutrinos by their decays via nambu-goldstone bosons in the gravitational-anomaly model of neutrino masses. *Phys. Rev. D* **101**(1), 015025 (2020). [arXiv:1905.01264](#) [hep-ph]
22. M. Kachelriess, R. Tomas, J.W.F. Valle, Supernova bounds on Majoron emitting decays of light neutrinos. *Phys. Rev. D* **62**, 023004 (2000). [arXiv:hep-ph/0001039](#)
23. N. Song, S.W. Li, C.A. Argüelles, M. Bustamante, A.C. Vincent, The future of high-energy astrophysical neutrino flavor measurements. *JCAP* **04**, 054 (2021). [arXiv:2012.12893](#) [hep-ph]
24. P.F. de Salas, D.V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C.A. Ternes, M. Tórtola, J.W.F. Valle, 2020 global reassessment of the neutrino oscillation picture. *JHEP* **02**, 071 (2021). [arXiv:2006.11237](#) [hep-ph]
25. I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, A. Zhou, The fate of hints: updated global analysis of three-flavor neutrino oscillations. *JHEP* **09**, 178. [arXiv:2007.14792](#) [hep-ph]
26. F. Capozzi, E. Di Valentino, E. Lisi, A. Marrone, A. Melchiorri, A. Palazzo, Global constraints on absolute neutrino masses and their ordering. *Phys. Rev. D* **95**(9), 096014 (2017). [arXiv:2003.08511](#) [hep-ph]. [Addendum: *Phys. Rev. D* **101**, 116013 (2020)]
27. C.-P. Ma, E. Bertschinger, Cosmological perturbation theory in the synchronous and conformal Newtonian gauges. *Astrophys. J.* **455**, 7–25 (1995). [arXiv:astro-ph/9506072](#) [astro-ph]
28. D. Blas, J. Lesgourgues, T. Tram, The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes. *JCAP* **1107**, 034 (2011). [arXiv:1104.2933](#) [astro-ph.CO]
29. G.B. Gelmini, M. Roncadelli, Left-handed neutrino mass scale and spontaneously broken lepton number. *Phys. Lett. B* **99**, 411–415 (1981)
30. Y. Chikashige, R.N. Mohapatra, R. Peccei, Are there real goldstone bosons associated with broken lepton number? *Phys. Lett. B* **98**, 265–268 (1981)
31. J. Schechter, J. Valle, Neutrino decay and spontaneous violation of lepton number. *Phys. Rev. D* **25**, 774 (1982)
32. J. A. Dror, Discovering leptonic forces using nonconserved currents. *Phys. Rev. D* **101**(9), 095013 (2020). [arXiv:2004.04750](#) [hep-ph]
33. M. Ekhterachian, A. Hook, S. Kumar, Y. Tsai, Bounds on gauge bosons coupled to nonconserved currents. *Phys. Rev. D* **104**(3), 035034 (2021). [arXiv:2103.13396](#) [hep-ph]
34. C.-H. Chen, C.-W. Chiang, C.-W. Su, Ultra-high-energy neutrino scattering in an anomalous U(1) effective field theory. *Phys. Lett. B* **827**, 136988 (2022). [arXiv:2110.07517](#) [hep-ph]
35. G. Alonso-Álvarez, J.M. Cline, Sterile neutrino dark matter catalyzed by a very light dark photon. *JCAP* **10**, 041 (2021). [arXiv:2107.07524](#) [hep-ph]
36. G.B. Gelmini, J.W.F. Valle, Fast invisible neutrino decays. *Phys. Lett. B* **142**, 181–187 (1984)
37. A.S. Joshipura, S.D. Rindani, Fast neutrino decay in the minimal seesaw model. *Phys. Rev. D* **46**, 3000–3007 (1992). [arXiv:hep-ph/9205220](#)
38. E.K. Akhmedov, A.S. Joshipura, S. Ranfone, J.W.F. Valle, Left-right symmetry and neutrino stability. *Nucl. Phys. B* **441**, 61–75 (1995). [arXiv:hep-ph/9501248](#)
39. M. Escudero, J. Lopez-Pavon, N. Rius, S. Sandner, Relaxing cosmological neutrino mass bounds with unstable neutrinos. [arXiv:2007.04994](#) [hep-ph]
40. J. Froustey, C. Pitrou, M.C. Volpe, Neutrino decoupling including flavour oscillations and primordial nucleosynthesis. *JCAP* **12**, 015 (2020). [arXiv:2008.01074](#) [hep-ph]
41. J.J. Bennett, G. Buldgen, P.F. De Salas, M. Drewes, S. Gariazzo, S. Pastor, Y.Y.Y. Wong, Towards a precision calculation of N_{eff} in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED. *JCAP* **04**, 073 (2021). [arXiv:2012.02726](#) [hep-ph]
42. T. Brinckmann, J. Lesgourgues, MontePython 3: boosted MCMC sampler and other features. *Phys. Dark Univ.* **24**, 100260 (2019). [arXiv:1804.07261](#) [astro-ph.CO]
43. A. Lewis, GetDist: a Python package for analysing Monte Carlo samples. [arXiv:1910.13970](#) [astro-ph.IM]
44. P. B. Denton, I. Tamborra, Invisible neutrino decay could resolve IceCube’s track and cascade tension. *Phys. Rev. Lett.* **121**(12), 121802 (2018). [arXiv:1805.05950](#) [hep-ph]
45. IceCube Collaboration, M. G. Aartsen et al., A combined maximum-likelihood analysis of the high-energy astrophysical neutrino flux measured with IceCube. *Astrophys. J.* **809**(1), 98 (2015). [arXiv:1507.03991](#) [astro-ph.HE]
46. JUNO Collaboration, F. An et al., Neutrino physics with JUNO. *J. Phys. G* **43**(3), 030401 (2016). [arXiv:1507.05613](#) [physics.ins-det]
47. DUNE Collaboration, B. Abi et al., Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I Introduction to DUNE. *JINST* **15**(08), T08008 (2020). [arXiv:2002.02967](#) [physics.ins-det]
48. Hyper-Kamiokande Collaboration, K. Abe et al., Hyper-kamiokande design report. [arXiv:1805.04163](#) [physics.ins-det]
49. Baikal-GVD Collaboration, A. D. Avrorin et al., The Baikal-GVD neutrino telescope: First results of multi-messenger study. *PoS ICRC2019*, 1013 (2020). [arXiv:1908.05450](#) [astro-ph.HE]

50. KM3Net Collaboration, S. Adrian-Martinez et al., Letter of intent for KM3NeT 2.0. *J. Phys. G* **43**(8), 084001 (2016). [arXiv:1601.07459](#) [astro-ph.IM]
51. P-O.N.E. Collaboration, M. Agostini et al., The Pacific Ocean neutrino experiment. *Nature Astron.* **4**(10), 913–915 (2020). [arXiv:2005.09493](#) [astro-ph.HE]
52. A. Romero-Wolf et al., An andean deep-valley detector for high-energy tau neutrinos. In: Latin American Strategy Forum for Research Infrastructure. 2, 2020. [arXiv:2002.06475](#) [astro-ph.IM]
53. IceCube-Gen2 Collaboration, M. G. Aartsen et al., IceCube-Gen2: the window to the extreme Universe. *J. Phys. G* **48**(6), 060501 (2021). [arXiv:2008.04323](#) [astro-ph.HE]
54. K. Abazajian et al., CMB-S4 Science Case, Reference Design, and Project Plan. [arXiv:1907.04473](#) [astro-ph.IM]