PROBING HIDDEN SECTORS WITH EARLY UNIVERSE COSMOLOGY

BY

PRANJAL RALEGANKAR

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Gilbert Holder, Chair
Associate Professor Jessie Shelton, co-Director of Research
Associate Professor Peter Adshead, co-Director of Research
Assistant Professor Yonatan Kahn

Abstract

There is no fundamental requirement for all particles to be charged under the Standard Model gauge symmetries. Consequently, there could naturally be a hidden sector of particles that have gone undetected and dark matter could reside in it. A hidden sector of particles could naturally have a different temperature than the plasma formed by the Standard Model particles in the early universe. Such a hidden sector can alter early universe cosmology from the assumed behavior in a variety of ways, which we explore in this thesis. First, we discuss how both hidden sector and Standard Model sector can be populated via inflaton decays after inflation. Next, we explore how thermally decoupled hidden sector with a massive lightest particle ($m \gg$ MeV) can enhance the abundance of sub-Earth mass dark matter microhalos today. We then explore how hidden sectors with a massless lightest particle can cause an inhomogeneous distribution of helium to hydrogen ratios if the hidden sector never had any interaction with the Standard Model sector. If instead the hidden sector had some interaction, then we show how the constraints on dark radiation energy density can be employed to constrain the interaction strength between the two sectors.

To my mother, father, and sister.

Acknowledgments

However well-motivated be the models I have studied in this thesis, there is an unfortunate possibility that they are not the objective reality of our universe. So perhaps this maybe the only section with objectively true statements, where I describe the very real impact the people around me have had in making me grow both as a researcher and as a person. Although, similar to abelian gauge groups, faithfully representing the impact of people around me in a finite space (or dimensions) is mathematically impossible. So forgive me as I provide a very incomplete list of the people and the ways they have impacted me in the last five years.

The biggest contribution to my growth as a researcher has been from my advisors: Jessie Shelton and Peter Adshead. For me their advising style struck a perfect balance between freedom and hands-on guidance. Among the variety of things they have guided me on, their contribution in helping me grow as an academic writer stands out distinctly for me. I am especially grateful to the patience they showed as I struggled to improve my writing and also for reigning in my penchant of making overly strong statements and generalizations. They also have a special skill in suggesting projects that are both doable and challenging, which I have only come to recognize and appreciate in hindsight. The well-chosen projects helped me become confident in my research abilities, which motivated me to continue pursuing a career in academia.

I would not have even started my academic pursuit in cosmology if not for the initial mentorship by Vikram Rentala. Despite me not having a background in physics, he accepted to mentor my bachelor's thesis in cosmology. Multiple afternoons of discussion about fundamentals of cosmology with Vikram sparked my interest in this field and motivated to pursue PhD in it. I am also grateful for his continued mentorship as I try to climb the academic career ladder.

Other than my advisors, Adrienne Erickek and Gil Holder have had the next biggest impact on my PhD research. Adrienne's expertise on cosmological perturbations provided much needed guidance for two of my papers. I am especially grateful to her for going out of her way to support me for several of my postdoc applications. I am thankful to Gil in helping me understand the nuance involved in scientific data analysis and for informal advising over the years.

Teaching has been a big part of my PhD, even though its contribution is not reflected directly in my thesis. In particular, my experience as a teaching assistant for the particle physics course stands out among others and I would like to thank Yonatan Kahn for this experience.

The friendships I made during the past five years played a crucial role in making me enjoy my PhD journey. I would like to thank Suryanarayana Sankagiri, Krishnakant Saboo, and Vaishnavi Subramaniyan for being like a family to me in the last three years. I want to thank Shivang Agarwal and Abhishek Sharma for being constant companions through highs and lows since the beginning of my PhD. The company of Prakruthi Burra, Chirag Shetty and the others mentioned above have been the sole reason for making the pandemic bearable, and sometimes even enjoyable. I am grateful to Suraj Sanjay Jog, Unnat Jain, Anshika Gupta, Ahmed Sameer Khan, Konik Kothari, Anand Bhattad, Ashok Vardhan, Deepika Chillar, Anirudh

Choudhary, Sanket Kanjalkar, and Anirudh Jain for providing a me a sense of community to belong to.

I am also grateful to the community at physics department in UIUC, for providing a caring and socially lively environment. Thanks to the several long walks with Sai Naga Manoj Paladugu, where we shared our research struggles, I realized that I am not alone in my struggles. The company of Abid Khan, Mark Hirsbrunner, Vincent Hickl, Tyler Salners, Shaun Lahert, Amber Lennox, and Weizhen Jia helped me adjust to the grad student life in my initial years of PhD. I was fortunate to have Zachary Weiner and Kaloian Lozanov as officemates, both of whom helped make the office a fun place. No acknowledgement section in a UIUC physics department's thesis can be complete without thanking Lance Cooper and Wendy Wimmer for their unwavering support.

Finally, I want to acknowledge my parents, Pallavi Ralegankar and Purananand Ralegankar, as well as my sister, Prachi Ralegankar, for providing unconditional support to me for the past 28 years. Research endeavor in theoretical physics requires a certain shielding from the day-to-day troubles of life, and I believe my family's over caring of their youngest member enabled me to pursue this career path.

Table of contents

List of Abbreviations

BBN	Big Bang nucleosynthesis.
CMB	Cosmic microwave background.
HS	Hidden sector.
SM	Standard model.
EMDE	Early matter dominated era.
ECDE	Early cannibal dominated era.
LHS	Left hand side.
RHS	Right hand side.

Chapter 1

Introduction

The Standard Model of particle physics has had resounding success in explaining the nature of fundamental particles that form the building blocks of all known matter [1]. However, a host of observational data from galactic rotation curves to the cosmic microwave background (CMB) signals the existence of an unknown form of matter comprising more than 85% of the matter in our Universe [2, 3]. While this dark matter could be a single particle or a family of particles missing from the Standard Model of particle physics, to date no signs of non-gravitational interaction of dark matter with the Standard Model have been detected. Further, the traditional weakly interacting massive particle (WIMP) paradigm for dark matter is under ever-increasing pressure from the dearth of observational signatures in collider, direct, and indirect detection experiments [4, 5].

An alternative — and arguably more generic — scenario is that there exists a whole sector of particles, one of them comprising the DM today, that do not interact with Standard Model gauge forces: the strong and the electroweak force, but might have their own dark gauge forces. Such hidden sectors are also predicted in several classes of string compactification theories that attempt to unify the Standard Model of physics with gravity [6]. Moreover, exotic dynamics in a hidden sector can lead to substantial changes in dark matter properties relative to a traditional WIMP, which might explain the null results in the terrestrial experiments while also motivating new strategies for detection (e.g., [7–15]).

In standard cosmology, all the Standard Model particles equilibrate with each other to form a thermal plasma in the early universe (within ∼1 second of the birth of the universe), and this thermal state washes out any information of the prior history. Consequently, we have little to no information about the state of the universe prior to Big Bang nucleosynthesis (BBN), which occurs when the Standard Model plasma is at a temperature of around $T_{SM} \sim$ MeV. However, if a hidden sector of particles exists, then they may not equilibrate with the Standard Model particles in the early universe. This opens the window for the information of the early universe to remain preserved until today and thus be observable. In this thesis, we explore different potentially observable ways an out-of-equilibrium hidden sector can cause deviation from the assumed cosmology prior to BBN. As the particles in the early universe are at much higher temperatures than today, the cosmological imprints we study allow us to probe the properties of hidden sector at very high energies, $E >$ MeV.

A key parameter on which any observable in hidden sector cosmology depends on is the relative temperature between the hidden sector particles and the Standard Model particles.[1] This temperature ratio between the

[1] Although we consider the hidden sector to not thermalize with the Standard Model plasma, we will be primarily interested

two sectors is determined by the processes that populate the two sectors at the very beginning of the universe. In traditional cosmology, the Standard Model particles are considered to be populated by the decay of the inflaton field. Here the inflaton field is a spin-0 field that causes an epoch of accelerated expansion right after the birth of the universe, for which we have compelling evidence from the observations of the CMB spectrum [16]. A natural mechanism to populate the hidden sector would be to have the inflaton field to also decay into hidden sector particles. In Chapter 3, which is based on the work in Ref. [17], we study how the temperature ratio between the hidden sector and the Standard Model sector depend on the inflaton couplings with the two sectors.

After the temperature of the hidden sector, the mass and couplings of the lightest hidden sector particle are the next most cosmologically relevant parameters. This is because at any given hidden sector temperature, all hidden sector particles with masses larger than the temperature would typically have annihilated or decayed into lighter hidden sector particles. As the late time behavior of hidden sector has maximum impact on cosmological observables, the behavior of lightest hidden sector particle has the most phenomenological relevance.

For instance, if the lightest sector particle has a large mass, $m \gg$ MeV, then it can generically cause an early matter dominated era (EMDE) prior to BBN. Such an EMDE enhances the abundance of dark matter micro-halos with masses smaller than $10^{-6}\ M_\odot$ [18, 19]. A large abundance of micro-halos would boost the DM annihilation signals and thus can noticeably change predictions for indirect detection experiments like Fermi-LAT [18–20]. These halos might also be detected by pulsar timing arrays [21, 22] or by their impact on stellar microlensing within galaxy clusters [23–25].

In chapter 4, based on the work in Refs. [26, 27], we study a scenario where the lightest hidden sector particle has a large mass ($m \gg$ MeV) as well as strong number changing self interactions. In this case, once the temperature of the hidden sector falls below the mass of the lightest particle, its strong number-changing self-interactions can cause the particles to annihilate among themselves. This self-annihilation converts the particles' rest mass energy into thermal energy, leading to a logarithmic evolution of the temperature of its plasma, $T \propto 1/\log(a)$ [7]. This self-annihilating phase of evolution is commonly known as cannibalism and the particle that undergoes such an evolution is known as a cannibal. We show that when the lightest hidden sector particle is a cannibal, then there can be an early cannibal-dominated era (ECDE) prior to BBN. Similar to EMDE, we show that an ECDE also enhances the population of dark matter microhalos today and we study how the cannibal particle properties are imprinted in the microhalo spectrum.

If instead the lightest hidden sector particle is massless, then the lightest particle contributes to dark radiation today. The fluctuations in the CMB spectrum are sensitive to the energy density in dark radiation, and as no dark radiation has been observed, the energy density in dark radiation is severely constrained [3]. In chapter 4, which represents the content in Ref. [28], we show how one can convert the constraint on dark radiation density to a constraint on interactions between Standard Model sector and hidden sector that might be active prior to BBN.

While there has been no evidence of dark radiation in current measurements of CMB, it is possible that we might detect dark radiation in future with improved CMB measurements. If we do observe dark radiation, then a natural question one might ask is whether the particle behind dark radiation was always out of contact with Standard Model particles or if it had some contact in the early universe. We can distinguish between the non-interacting and weak-interacting scenario by checking if the density fluctuations of dark radiation is in or out of sync with the density fluctuations of photons. In chapter 5, which represents the work in Ref. [29],

in the regime where the internal hidden sector interactions are strong enough to have internally thermalized the hidden sector.

we show that an out-of-sync fluctuation, also known as isocurvature, produces different amounts of helium to hydrogen ratio (He/H) in different parts of the universe during BBN. Consequently, we explain how an absence of excess variance in the measurements of He/H from different galaxies can be used to constrain isocurvature between dark radiation and photons.

Before proceeding with the detailed discussion of above topics, we provide a brief review of the thermal history of the Standard Model particles during and prior to BBN in chapter 2. We highlight the primary physics that determine the evolution of the thermal plasma in an expanding universe and thus lay the ground work for rest of the thesis.

We work in natural units which set $\hbar = c = 1$, while retaining factors of the reduced Planck mass, $M_{\rm Pl} = 1/\sqrt{8\pi G_N} \approx 2.435 \times 10^{18}$ GeV.

Chapter 2

Early Universe thermal cosmology

In this chapter we review the standard cosmology of the universe within the few seconds after its birth, when all the Standard Model (SM) particles were in a hot thermal plasma. We begin in section 2.1 by first reviewing the cosmological equations that determine the evolution of this early universe. Using these equations, we show why the early universe is primarily dominated by relativistic particles of the SM and that the contribution of dark matter is negligible. In section 2.2, we review the Boltzmann equations that describe how particles maintain thermal equilibrium in an expanding universe. Using these equations, we explain why the SM particles form a relativistic plasma in the early universe. Subsequently, we focus on the dynamics of the SM plasma in section 2.3. In particular, we review 1) the evolution of the number of particles in the SM plasma as the plasma cools, 2) how the matter-antimatter asymmetry determines the proton abundance, 3) the decoupling of neutrinos, and finally 4) the formation of helium and deuterium nuclei at Big Bang nucleosynthesis. In section 2.4, we discuss a possible minimal origin of dark matter from the SM plasma. Finally, in section 2.5 we discuss how an out-of-equilibrium hidden sector might be produced from the SM plasma in the early universe via renormalizeable interactions and highlight the attractor nature of the system of equations governing the hidden sector.

2.1 Equations of motion in a homogeneous and isotropic universe

Einstein's theory of general relativity describes how the metric of spacetime depends on the matter content at a given location. When the theory is applied to the whole universe along with the cosmological principle that the universe on large scales is isotropic and homogeneous, we find that our universe must dynamically expand with time.

To capture this expansion we work in the comoving coordinates, x^i, which remain fixed with time, while the physical coordinates are given by ax^i, where a is the scale factor of the expanding universe. The metric for an expanding flat universe is then given by

$$ds^2 = dt^2 - a^2(dx^2 + dy^2 + dz^2), \tag{2.1}$$

where t denotes time. More generally, the universe can also have an overall curvature. However, as the current observations place stringent limit on the existence of such curvature [3], we only focus on the spatially flat universe.

The energy momentum tensor for an isotropic and homogeneous universe is given by

$$T^{\mu}_{\nu} = \begin{bmatrix} -\rho & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{bmatrix}, \tag{2.2}$$

where ρ and $\mathcal{P}$ are the total energy density and pressure of the matter in the universe, respectively.

Using the metric in eq. (2.1) and the energy momentum tensor in eq. (2.2) in the Einstein equations gives us the Friedmann equation:

$$\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{1}{\sqrt{3}M_{\mathrm{Pl}}}\sqrt{\rho} \equiv H, \tag{2.3}$$

where H is called the Hubble rate, and M_{Pl} is the reduced Planck mass. This equation highlights how the expansion rate of the universe depends on the energy content of the universe.

The energy density of the total matter in the universe also evolves due to the expansion of the universe. The evolution can be determined through the energy momentum conservation equation, $\nabla_{\mu}T^{\mu\nu} = 0$, which yields,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + 3H(\rho + \mathcal{P}) = 0. \tag{2.4}$$

As H, ρ and $\mathcal{P}$ are all positive, the above equation implies that the energy density of the universe dilutes as the universe expands. Conversely, in the early universe when the scale factor was much smaller, the matter content was compressed to much higher densities.

To complete the coupled set of equations given by eq. (2.4) and eq. (2.3), we need to specify the relationship between $\mathcal{P}$ and ρ. This relationship is commonly parameterized through the equation-of-state,

$$w = \frac{\mathcal{P}}{\rho}. \tag{2.5}$$

The equation of state so defined is better analyzed for individual particles comprising the total matter content in the universe.

For any particle i in a homogeneous and isotropic universe, their phase space distribution function f_i is only a function of the particle's energy. Both the particle's energy density and pressure can be evaluated using the distribution function via

$$\rho_i = g_i \int \frac{d^3p}{(2\pi)^3} E f_i(E), \qquad \mathcal{P}_i = g_i \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} f_i(E), \tag{2.6}$$

where g_i is the number of degrees of freedom of particle i. One can see that for relativistic particles, i.e. for particles' whose distribution function is heavily weighted at $p \gg m$, the equation of state is simply $w_i = 1/3$. Conversely, for non-relativistic particles we have $w_i \ll 1$. Thus, from eq. (2.4) we can see that for relativistic particles $\rho \propto 1/a^4$, while for non-relativistic particles, $\rho \propto 1/a^3$. Hence, at sufficiently early times, we expect the energy density of relativistic particles to dominate the universe.

Note that even if a particle is non-relativistic today, we expect them to become relativistic at sufficiently early times. To see why, we need to look at the geodesic equation of motion of individual particles in

expanding spacetime. For a particle with four momenta $p^\mu = (E, \vec{p})$, the geodesic equation yields,

$$|\vec{p}| \propto 1/a. \tag{2.7}$$

Consequently, for $a \ll 1$ in the early universe, we can expect all particles to become relativistic, $|\vec{p}| \gg m$.

The redshifting of momenta also explains the difference in the evolution of ρ between relativistic and non-relativistic particles. In particular, the energy of a relativistic particle also redshifts as $E = p \propto 1/a$, but that of a non-relativistic particle stays constant, $E = m$. As the energy density of particles can be approximately expressed as $\rho \sim \langle E \rangle n$, where $\langle E \rangle$ is the average energy and n is the number density, we can see that for non-relativistic matter $\rho \propto 1/a^3$ because the number density dilutes as $1/a^3$. Meanwhile, for relativistic matter $\rho \propto 1/a^4$ because the average energy of the particle is also redshifting in an expanding universe.

From observations of our universe, we find that the matter in our universe is comprised of SM particles as well as an unknown dark matter. While the present day dark matter abundance is about five times larger than the abundance of SM particles, in the early universe the energy density of SM particles dominates over that of dark matter. In particular, experiments indicate that the energy density of neutrinos and photons overcome the dark matter density at $a_{\rm eq}/a_0 = 1/3411$ [3], where a_0 is the present day scale factor and $a_{\rm eq}$ is the scale factor at matter-radiation equality. As we go further back in time, the photons and neutrinos reach sufficiently high energies to pair-produce the rest of the SM particles at relativistic energies and form a plasma. In this thesis we are primarily concerned in the early history of the universe, when the plasma of SM particles was at temperatures larger than $T \gtrsim 0.1$ MeV.

At such early times, the Hubble rate is predominantly determined by the energy density of SM particles, and the Friedmann equation yields,

$$H = \frac{1}{\sqrt{3} M_{\rm Pl}} \sqrt{\rho_{SM}} \propto 1/a^2. \tag{2.8}$$

On cosmological timescales, only those particle interactions can have a significant impact whose rates are larger than the Hubble rate. Consequently, the evolution of H is critical in determining when various non-gravitational interactions become important cosmologically.

2.2 Equilibrium in an expanding universe

In this section we discuss how particles maintain thermal equilibrium in an expanding universe and highlight the conditions necessary to maintain thermal equilibrium. Using these conditions, we then argue why the SM particles come into thermal equilibrium in the early universe. Towards the end, we discuss how the distribution function of individual particles evolve once they fall out-of-equilibrium due to the expansion of the universe.

2.2.1 The Boltzmann equation

We begin by first considering the evolution of the single particle phase space distribution function, $f(\vec{x}, \vec{p}, t)$, in an expanding universe. The distribution function, f, quantifies the number of particles in a volume element

of phase space, $\frac{d^3\vec{p}}{(2\pi)^3}d^3\vec{x}$, such that the number density is determined by

$$n(\vec{x},t) = \int \frac{d^3\vec{p}}{(2\pi)^3} f(\vec{x},\vec{p},t). \tag{2.9}$$

Note that f is a Lorentz scalar because both the phase space element, $d^3\vec{x}d^3\vec{p}$, and the total number of particles $N = \int \frac{d^3\vec{p}}{(2\pi)^3}d^3\vec{x}f$ are Lorentz scalars.

In an isotropic and homogeneous universe, the distribution function is only a function of particle energy, $f(E,t)$. The evolution equation for such a distribution function in an expanding spacetime is given by the following Boltzmann equation,

$$\frac{\partial f}{\partial t} - H\frac{\vec{p}^2}{E}\frac{\partial f}{\partial E} = \mathcal{C}[f], \tag{2.10}$$

where $\mathcal{C}$ is the collision operator that quantifies the change in particle number occupying momenta $\vec{p}$ due to collisions with other particles. The collision operator is given by

$$\begin{aligned}\mathcal{C}[f] = -\sum_{\text{pr}} \frac{1}{2E}\int d\Pi_a d\Pi_b ... d\Pi_i d\Pi_j ... |\mathcal{M}_p|^2 S(2\pi)^4\delta^4\left(p + \sum_i p_i - \sum_a p_a\right) \\ \times\Big[f(p)f_a(p_a)f_b(p_b)...[1\pm f_i(p_i)][1\pm f_j(p_j)]... - f_i(p_i)f_j(p_j)...[1\pm f(p)][1\pm f_a(p_i)]...\Big],\end{aligned} \tag{2.11}$$

where the summation goes over all processes of form $\psi + i + j + .. \to a + b$ such that the final state does not have the original ψ particle occupying momenta $|\vec{p}|$. Above

$$d\Pi_i = \frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} = \frac{d^4p_i}{(2\pi)^3}\delta(p_i^2 - m_i^2) \tag{2.12}$$

is the Lorentz-invariant phase space element of the involved particles, $|\mathcal{M}_p|^2$ is the spin-summed matrix element for the given processes, S is the symmetry factor to account for identical particles in the initial and final states, and the δ^4 term is a Dirac delta enforcing energy-momentum conservation. In the square brackets in eq. (2.11) we are subtracting the number of times the reverse processes occurs from the number of times forward process occurs. Here for simplicity we have assumed that $|\mathcal{M}_p|^2$ is the same for reverse and forward process, which is only true for processes that are CP invariant. The factors of $f_a(p_a)$ denote the probability of finding particle a with momentum p_a while the factors of $[1\pm f_i(p_i)]$ indicate Bose-enhancement or Pauli-suppression by final state particle i depending on whether it is a boson or fermion.

If the interaction rate of a given process is much larger than the Hubble rate, then on time scales much shorter than the expansion rate of the universe, the Boltzmann equation for all the involved particles would be driven to an equilibrium state where the square brackets in eq. (2.11) for the given process becomes zero. The detailed balance of the square brackets is satisfied only if

$$f = f_{eq} \equiv \frac{1}{\exp{(E-\mu)/T} \pm 1}, \tag{2.13}$$

and if $\mu + \mu_i + \mu_j + ... = \mu_a + .. + \mu_b$. Above, the plus (minus) sign is if the particle is a fermion (boson), T is the temperature of the system, and μ is the chemical potential of the particle.

In the limit $f = f_{eq}$, the collision terms on the RHS of the Boltzmann equation (eq. (2.10)) are set to

zero. Assuming f_{eq} also satisfies the LHS of the Boltzmann equation, requires

$$\frac{\partial f_{eq}}{\partial t} - H\frac{\vec{p}^2}{E}\frac{\partial f_{eq}}{\partial E} = 0. \tag{2.14}$$

Substituting the equilibrium distribution yields

$$EH\left(\frac{d\ln(aT)}{d\ln a} - \frac{m^2}{E^2}\right) + T\frac{d(\mu/T)}{dt} = 0. \tag{2.15}$$

For relativistic particles, we can see that $f = f_{eq}(p, T, \mu)$ with $T \propto 1/a$ and $\mu/T =$ constant satisfies both LHS and RHS of the Boltzmann equations independently. For non-relativistic particles the LHS and the RHS cannot be be independently satisfied. Any non-zero values from the LHS of the Boltzmann equation for non-relativistic particles is canceled by the collision term on the RHS by having small distortions around the equilibrium distribution.

Thus one cannot find the temperature evolution of a plasma that also have non-relativistic particles just by looking at the Boltzmann equation of a single particle. The temperature evolution can be found by applying energy-momentum conservation for the total plasma, eq. (2.4), and expressing the energy density and pressure in terms of the temperature of the plasma. In the next subsection we provide the expressions for these densities and then using these expressions we find the evolution of temperature of the Standard Model plasma in section 2.3.1.

2.2.2 List of equilibrium parameters

In this subsection, we tabulate the expression for energy, pressure and number density in terms of temperature and chemical potential.

For relativistic bosons with negligible chemical potential, we have

$$\rho = \frac{g\pi^2}{30}T^4, \qquad n = \frac{g\zeta(3)}{\pi^2}T^3, \qquad \mathcal{P} = \frac{\rho}{3}. \tag{2.16}$$

For relativistic fermions with negligible chemical potential, we have

$$\rho = \frac{7}{8}\frac{g\pi^2}{30}T^4, \qquad n = \frac{3}{4}\frac{g\zeta(3)}{\pi^2}T^3, \qquad \mathcal{P} = \frac{\rho}{3}. \tag{2.17}$$

If we approximate the equilibrium distribution of particles as Maxwell-Boltzmann distribution, $f_{eq} = e^{(\mu-E)/T}$, then we obtain

$$\rho = e^{\mu/T}\frac{gm^4}{2\pi^2}\left(\frac{xK_1(x) + 3K_2(x)}{x^2}\right), \qquad n = e^{\mu/T}\frac{gm^3}{2\pi^2}\frac{K_2(x)}{x}, \qquad \mathcal{P} = nT, \tag{2.18}$$

where $x = m/T$ and K_n is the modified Bessel function of the second kind. In the limit, $T \gg m$ or $x \ll 1$, ρ and n simplify to

$$\rho = e^{\mu/T}\frac{3gT^4}{\pi^2}, \qquad n = e^{\mu/T}\frac{gT^3}{\pi^2}. \tag{2.19}$$

Whereas in the limit $T \ll m$ we obtain

$$\rho = mn, \qquad\qquad n = e^{(\mu-m)/T} g \left(\frac{mT}{2\pi}\right)^{3/2}. \tag{2.20}$$

Note that the distribution of both bosons and fermions approximate to Maxwell-Boltzmann distribution in the limit, $T \ll m$.

For Standard Model particles, the chemical potential is always negligible, $\mu \ll T$, when the particles are relativistic. Since particles and antiparticles must have opposite chemical potential in equilibrium, the excess of particles over antiparticles for $\mu \ll T$ and $T \gg m$ is

$$n_b - \bar{n}_b = \frac{gT^3}{3}\frac{\mu}{T} \qquad\qquad n_f - \bar{n}_f = \frac{gT^3}{6}\frac{\mu}{T}, \tag{2.21}$$

where the first relation is for bosons and the second one is for fermions.

While we do not see large chemical potentials in the Standard Model, it is possible for fermions to have a significant chemical potential even when they are relativistic because of Pauli-exclusion. In the limit $\mu \gg \{T, m\}$, when fermions are degenerate, we find

$$\rho = \frac{1}{8\pi^2} g\mu^4 \qquad\qquad n = \frac{1}{6\pi^2} g\mu^3 \qquad\qquad \mathcal{P} = \frac{\rho}{3}. \tag{2.22}$$

For bosons, the chemical potential has an upper bound, $\mu < m$. In the limit $\mu \to m$, any new particles added to the system fill the zero momentum state and lead to the formation of a Bose-condensate. Thus, the condition $\mu \gg T$ can never be achieved for a relativistic boson.

2.2.3 Condition for thermalization

So far we have been interested in the regimes where the rates at which the thermalizing interactions occur are much larger than the expansion rate of the universe. To see why that condition is generically satisfied in the early universe, let us consider the process: $e^+ + e^- \to 2\gamma$, which maintains electrons and positrons in thermal equilibrium. Integrating the Boltzmann equation for electrons (eq. (2.10)) with $\int d^3p/(2\pi)^3$ yields,[1]

$$\frac{dn_{e^-}}{dt} + 3Hn_{e^-} = -\int d\Pi_{e^-} d\Pi_{e^+} d\Pi_{\gamma_1} d\Pi_{\gamma_2} |\mathcal{M}|^2 S (2\pi)^4 \delta^4(p_{e^-} + p_{e^+} - p_{\gamma_1} - p_{\gamma_2}) \\ \times \Big[f_{e^-}(p_{e^-}) f_{e^+}(p_{e^+}) [1 + f_\gamma(p_{\gamma_1})][1 + f_\gamma(p_{\gamma_2})] - f_\gamma(p_{\gamma_1}) f_\gamma(p_{\gamma_2}) \Big]. \tag{2.23}$$

For simplicity, we have ignored the Pauli-suppression from electrons above. Ignoring final state quantum statistical effects of particles amounts to setting their equilibrium distribution to be Maxwell-Boltzmann, $f_{eq} = e^{(\mu-E)/T}$, as can be seen using detailed balance. When $e^+ + e^- \to 2\gamma$ process is in equilibrium, the chemical potentials of electron and positron satisfy, $\mu_{e^-} + \mu_{e^+} = 0$ because the photons do not have a chemical potential.[2] As the process goes out of equilibrium, we consider the distribution function of electron and

[1] Integrating the the Boltzmann equation eq. (2.10) with $\int d^3p\, E/(2\pi)^3$ returns the Boltzmann equation for energy density, ρ, given in eq. (2.4), except on the RHS we have collision terms that describe energy transfer from the concerned particle to other particles.

[2] Since the number of photons is not conserved in Standard Model processes, the photon chemical potential is driven to zero when photon number changing processes are in equilibrium.

positron to become of the form,

$$f_{e^\pm}(p_{e^\pm}) = e^{\tilde{\mu}_\pm/T} f_{eq,e^\pm}(p_{e^\pm}) = \frac{n_{e^\pm}}{n_{eq,e^\pm}} f_{eq,e^\pm}, \tag{2.24}$$

where $n_{eq,e^\pm}$ is the number density of electrons/positrons if they are in thermal equilibrium, $f = f_{eq}$. Doing so simplifies the collision term on the RHS to

$$\frac{dn_{e^-}}{dt} + 3Hn_{e^-} = \langle\sigma v\rangle(T)[n_{eq,e^-} n_{eq,e^+} - n_{e^-} n_{e^+}], \tag{2.25}$$

where

$$\langle\sigma v\rangle(T) = \frac{1}{n_{eq,e^-} n_{eq,e^+}} \int d\Pi_{e^-} d\Pi_{e^+} d\Pi_{\gamma_1} d\Pi_{\gamma_2} |\mathcal{M}|^2 S (2\pi)^4 \delta^4(p_{e^-} + p_{e^+} - p_{\gamma_1} - p_{\gamma_2}) f_{eq,\gamma}(p_{\gamma_1}) f_{eq,\gamma}(p_{\gamma_2}). \tag{2.26}$$

From eq. (2.25) it is clear that that electron positron annihilation rate can maintain electrons in thermal equilibrium if

$$\frac{\langle\sigma v\rangle(T) n_{eq,e^+}}{H} > 1. \tag{2.27}$$

Now let us see how this ratio evolves with the scale factor, a, as we go back in time. Sufficiently early when the universe is radiation dominated, we know that the Hubble rate evolves as $H \propto 1/a^2$. In contrast, the equilibrium number density of electrons and positrons grows exponentially for $T < m_e$ because $n_{eq,e^\pm} \propto e^{-m_e/T}$ and T increases with decreasing scale factor. Because of the electromagnetic gauge interactions, the cross-section $\langle\sigma v\rangle$ for electrons is large enough for the ratio in eq. (2.27) to exceed one at some $T < m_e$. Consequently, as we go back in time and T approaches m_e, the photons obtain significant energy to pair produce electrons and positrons and form a thermal bath where electrons and positrons are in equilibrium. While here we concentrated on electrons, similar analysis also holds for most of the other particles in the SM and thus we expect all of them to come into thermal equilibrium at some point in the early universe.[3]

2.2.4 Thermal decoupling

Apart from neutrinos and photons, all other particles in the SM fall out of thermal equilibrium once the temperature falls below the mass of the particle. After the particles become non-relativistic, their equilibrium number density is Boltzmann suppressed, $n_{eq} \propto e^{-m/T}$, because while the particles undergo annihilations into lighter particles in the SM plasma, the lighter particles do not have sufficient energy to produce the massive particles in the reverse process. As n_{eq} falls exponentially with temperature, $\langle\sigma v\rangle(T) n_{eq}$ soon falls below the Hubble rate. Once that happens, the annihilations stop and particle number density is conserved from that point onwards. The deviation from the equilibrium number density is encaptured by particle developing a chemical potential, $n = e^{\tilde{\mu}/T} n_{eq}$, that no longer satisfies chemical balance equation imposed by by the particle number changing interactions. Thus, a particle falls out of chemical equilibrium whenever

[3]Interestingly, even much earlier in the history of the universe, $T > 10^{16}$ GeV, the SM particles are expected to fall out of thermal equilibrium. This is because when all the involved particles are relativistic, the only dimensionful quantity that can set interaction rate is the temperature of the plasma, $\Gamma \propto T$. Consequently, if we extrapolate sufficiently back in time, the Hubble rate, $H \propto T^2$, exceeds all inter particle interaction rate. Here we do not focus on such early history because one generically expects new physics to come into play and alter this naive extrapolation.

its number changing interaction rate falls below the Hubble rate. We revisit this departure from chemical equilibrium in more detail in the context of WIMP dark matter in section 2.4.

Note that even if the particles have fallen out of chemical equilibrium they can still maintain the same temperature with rest of the plasma via elastic scattering processes. For electrons and positrons this occurs via Compton scattering, $e + \gamma \to e + \gamma$. One can check that detailed balance for such elastic scattering processes enforce particles to have the same temperature, but allow the particles to have an arbitrary value of chemical potential. When these scattering processes are larger than the Hubble rate, the particles are said to be in kinetic equilibrium. Almost always one finds that the particles fall out of kinetic equilibrium much after chemical equilibrium.

For particles that fall out of chemical and kinetic equilibrium after they become non-relativistic, their distribution remains that of a thermal particle. This is because, in the absence of any interaction, the momentum of particles redshifts as $p \propto 1/a$. Consequently, the particles maintain their Maxwell-Boltzmann distribution, $\propto e^{-p^2/(2mT)}$, by having its effective temperature redshift as $T \propto 1/a^2$. Note that since the comoving number density of particles if fixed, their chemical potential must asymptote to the particle mass such that $e^{(\mu-m)/T}$ remains constant in eq. (2.20).

For particles that fall out of chemical and kinetic equilibrium while they are relativistic, for e.g. neutrinos and photons, their effective temperature redshifts as $T \propto 1/a$ even if they become non-relativistic at some later point. This is because, they inherit the distribution of relativistic particles at decoupling, which is only a function of particle momenta, p. As the momentum of particles redshifts as $p \propto 1/a$, their distribution too redshifts such that we can define a temperature that evolves as $T \propto 1/a$ and a chemical potential that remains constant relative to temperature, $\mu/T =$ constant.

2.3 Cosmological evolution of the Standard Model plasma

Having discussed the evolution of distribution function for individual particles, we now turn our attention to the combined evolution of a thermal plasma formed by Standard Model (SM) particles. We first discuss the temperature evolution of this plasma as the particles in the SM become non-relativistic. Next, we explain how the matter-antimatter asymmetry causes the particles to have a small chemical potential in the early universe and how it determines the relic abundance of nucleons. We then discuss the physics behind neutrino decouupling, which marks the start of thermal decoupling of the SM plasma. Finally, we describe the formation helium and deuterium nuclei in the early universe, whose abundances today provide the only direct evidence of the thermal state of SM particles in the early universe.

2.3.1 Degrees of freedom

Earlier we discussed how the temperature evolution of plasma consisting of non-relativistic particles cannot be described by looking at Boltzmann equations of individual particles. In the context of SM, apart from photons and neutrinos, all other particles become non-relativistic when they are still part of the SM plasma. Consequently, to describe the temperature evolution of the plasma we need to evolve the combined energy density of the plasma in an expanding universe using eq. (2.4).

We begin by first expressing energy density of the SM plasma, ρ_{SM}, in terms of the relativistic degrees of

freedom in the plasma and its temperature, T_{SM},

$$\rho_{SM} = \sum_i g_i \int \frac{d^3p}{(2\pi)^3} E \frac{1}{e^{E/T_{SM}} + \xi_i} \equiv \frac{g_* \pi^2}{30} T_{SM}^4, \tag{2.28}$$

where g_* is the effective degrees of freedom in the SM plasma,

$$g_* = \sum_i g_i \frac{30}{\pi^2} \int \frac{d^3p}{(2\pi)^3} E \frac{1}{e^{E/T_{SM}} + \xi_i}. \tag{2.29}$$

In the above definition of g_* we have neglected the contribution from the chemical potential, $\mu_i \ll T$. As discussed in the previous section, the chemical potential can become significantly large once particles become non-relativistic and fall out of chemical equilibrium. However, as the abundance of particles is exponentially suppressed in the process, their contribution to total SM energy is negligible compared to other particles that are relativistic. Note that relativistic particles in thermal equilibrium are allowed to have significant chemical potential, $\mu > T$, but in the context of SM we find that $\mu \ll T$ when particles are relativistic. We elaborate more on the chemical potentials in the SM in the next subsection.

At any given temperature, we can approximately calculate g_* by considering contribution from particles with masses $m_i < T_{SM}$. Doing so yields,

$$g_* \approx \sum_{i=\text{bosons}} g_i + \frac{7}{8} \sum_{i=\text{fermions}} g_i. \tag{2.30}$$

Note that the contribution of fermions and bosons to g_* is slightly different because they follow different thermal distributions. Thus at $T_{SM} \gg 200$ GeV, when all the SM particles are relativistic, we find

$$g_* = (2 + 2 \times 8 + 2 \times 3 + 4) + \frac{7}{8} \times 3 \left[4 + 2 + 4 \times 2 \times 3\right] = 106.75, \tag{2.31}$$

where we have counted photons, eight gluons, three SU(2) gauge bosons, and the complex Higgs boson doublet in the first bracket. In the second bracket we have three generations of fermions, with four degrees of freedom for each of the charged leptons and the quark doublets but only 2 degrees of freedom for left handed neutrinos. Additionally, the quark doublet has been multiplied by three to take into account the three colour degrees of freedom. As the SM plasma cools and particles become non-relativistic, g_* decreases.

In the left panel of figure 2.1 we show the evolution of g_* with photon temperature, T_γ. The largest change occurs at $T_\gamma \sim 200$ MeV where QCD phase transition occurs and all quarks become bounded into hadrons and mesons. At $T_{EW} = 160$ GeV the electroweak phase transition occurs [30].

Similar to g_*, we can also define an effective degrees of freedom for pressure, g_{*p}, via

$$g_{*p} = \sum_i g_i \frac{90}{\pi^2} \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} \frac{1}{e^{E/T_{SM}} + \xi_i}. \tag{2.32}$$

Then we can rewrite $\mathcal{P}_{SM} = g_{*p} \pi^2 T_{SM}^4 / 90$. In the left panel of figure 2.1 we also show the evolution of g_{*p} with temperature. Note that away from mas thresholds, $g_{*p} = g_*$ as expected.

By replacing ρ_{SM} and $\mathcal{P}_{SM}$ in terms of g_*, g_{*p} and T_{SM}, we can find the evolution of T_{SM} with scale factor using the Boltzmann equation, eq. (2.4). Alternatively, we can also find the evolution without integrating the

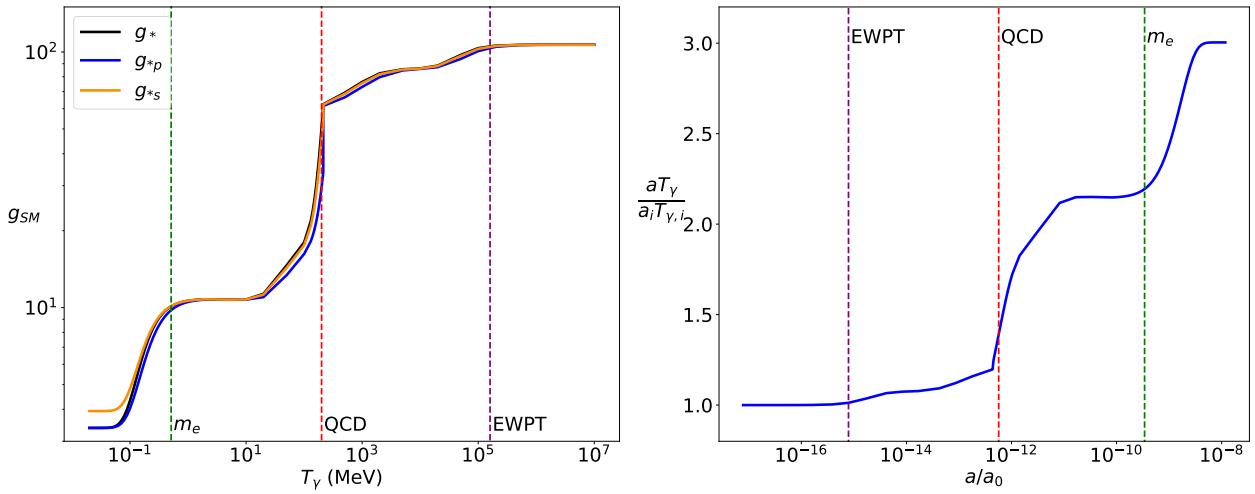


Figure 2.1: **Left Panel:** Effective degrees of freedom in SM plasma as a function of temperature of photon. The black, blue and orange lines correspond to degrees of freedom in energy, pressure and entropy density, respectively. **Right Panel:** Evolution of comoving photon temperature as a function of scale factor of the universe. The comoving temperature has been normalized with $T_{\gamma,i} = 10$ TeV and the scale factor has been normalized with the scale factor today, a_0. The vertical green dashed lines corresponds to $T_\gamma = m_e = 0.511$ MeV. The vertical red and purple dashed lines mark when QCD and electroweak phase transition occur.

Boltzmann equation by noting that in the limit $\mu_i = 0$,

$$\frac{\partial \mathcal{P}_i}{\partial T_i} = \frac{\rho_i + \mathcal{P}_i}{T_i}. \tag{2.33}$$

Then using the above equation along with the Boltzmann equation, one can show that

$$a^3 \left(\frac{\rho_{SM} + \mathcal{P}_{SM}}{T_{SM}} \right) = \text{const.} \equiv a^3 s_{SM}, \tag{2.34}$$

where s is the entropy density. Defining g_{*s} as the effective degrees of freedom in the entropy such that,

$$s_{SM} = \frac{2\pi^2}{45} g_{*s} T_{SM}^3, \qquad g_{*s} = \frac{3}{4}(g_* + g_{*p}/3), \tag{2.35}$$

we can obtain the evolution of T_{SM} simply via,

$$T_{SM} = T_{SM,i} \frac{a_i}{a} \left(\frac{g_{*s}(T_{SM,i})}{g_{*s}(T_{SM})} \right)^{1/3}. \tag{2.36}$$

In the left panel of figure 2.1 we also show the evolution of g_{*s} with T_γ and in the right panel we show the evolution of aT_γ with scale factor.

Note that the while T_{SM} is always decreasing as the universe expands, T_{SM} cools more slowly than $1/a$ when a SM particle becomes non-relativistic. This reduction in cooling occurs because as a SM particle becomes non-relativistic, it annihilates or decays into lighter SM particles. Consequently, the rest mas energy of the particle is converted into heat which slightly counters the cooling due to expanding universe. The heat from particle annihilations/decay is large enough such that $\rho_{SM} a^4 \propto g_*(aT_{SM})^4$ also increases whenever a particle becomes non-relativistic.

After neutrinos decouple around $T_{SM} \sim 1$ MeV, there is no common temperature for the SM plasma. Below this temperature, we define the effective degrees of freedom in the SM relative to the photon temperature, T_γ. Note that as $T_\gamma = T_{SM}$ before neutrino decoupling, one can use the entropy conservation equation to find the evolution of T_γ with a even through neutrino decoupling.

With the evolution of T_{SM} given by eq. (2.36), we can find the evolution of the Hubble rate via

$$H = 4.3 \times 10^{-22} \text{ MeV} \left(\frac{g_*(T_{SM})}{10}\right)^{1/2} \left(\frac{T_{SM}}{\text{MeV}}\right)^2 . \tag{2.37}$$

2.3.2 Chemical potentials and baryon asymmetry

In the previous subsection, we assumed that the chemical potential of SM particles is negligible. For particles in the SM whose number is not conserved, for e.g. the photon, this assumption is automatically true because their chemical potential is driven to zero by number changing interactions. However, there are various particle numbers that are conserved in the SM and thus thermal equilibrium alone is not sufficient to drive all the chemical potentials to zero.

For instance, in quantum electrodynamics, the total electron number is conserved, $(n_{e^-} - n_{e^+})a^3 = Q$ =constant. So when electrons and positrons are in thermal equilibrium, their chemical potentials are related by $\mu_{e^-} = -\mu_{e^+}$ and can have a non-zero value set by Q. The value of Q itself has to be specified either from an initial condition or from observations. The individual chemical potential of electron and positron vanish only if $Q = 0$.

The SM has several such conserved quantities: hypercharge, colour, SU(2) charge, baryon number, and lepton number. Additionally, Lepton number for each generation is also approximately conserved in the early universe because the neutrino masses are negligible. Consequently, when all the SM processes are active, there are only seven independent chemical potentials and chemical potentials of all involved particles can be written in terms of these seven chemical potentials. For hypercharge, colour, and SU(2), we observe no net charge in the universe today and consequently we can set the chemical potential corresponding to those charges to zero. Thus the number of independent non-zero chemical potentials are further reduced to four. However, we do observe an abundance of nucleons over anti nucleons today, i.e. a non-zero value of baryon number. There is also an equivalent abundance of electrons over positrons to make the total charge of the universe zero. Since the abundance of neutrinos over anti-neutrinos is very poorly constrained, we cannot comment if there is a net non-zero value of Lepton number as well.

The total baryon number is defined as

$$B = \frac{n_b - \bar{n}_b}{s} \tag{2.38}$$

where n_b and $\bar{n}_b$ are the number density of particles carrying $+1$ and -1 baryon number, respectively. Their difference is divided by the entropy density of the SM because both the numerator and denominator evolve as $1/a^3$, keeping B constant. From present day observations we find

$$B = \frac{n_p(a_0) + n_n(a_0)}{s(T_0)} \approx \frac{\rho_b(a_0)}{m_p s(T_0)} = 7.6 \times 10^{-11} \left(\frac{\Omega_b h^2}{0.02}\right), \tag{2.39}$$

where the subscript 0 denotes the value of the parameters today. In the above approximation we have neglected the contribution of electrons to the observed energy density of visible matter, ρ_b, because $m_e \ll m_p$.[4]

[4]Here I am using the astrophysical convention of classifying baryon energy density, ρ_b, as the total energy density of visibile

Prior to QCD phase transition, $T > T_{QCD} \approx 200$ MeV, when all the quarks were free, the number of quarks must exceed the number of antiquarks by

$$\sum_q \frac{n_q - \bar{n}_q}{s} = 3B. \tag{2.40}$$

As the quarks as relativistic prior to QCD phase transition, n_q and $\bar{n}_q$ is of the same order as the entropy density, $n_q \sim s \sim T_{SM}^3$. Which means the difference between their number densities is exceedingly small compared to their total number density. Consequently, their chemical potential also has to be very small, $\mu_q/T \ll 1$. One can then use eq. (2.21) to quantify the relationship between μ_q and $n_q - \bar{n}_q$. Doing so yields,

$$\frac{\mu_q}{T} \sim 10^{-8}, \tag{2.41}$$

thereby validating the assumption that chemical potentials are very small in the early universe.

After QCD phase transition, all the quarks and anti quarks are almost instantaneously bound into hadrons, of which only protons and neutrons survive until today. As $m_p, m_n \sim \text{GeV} \gg T_{QCD}$, the abundance of protons and neutrons is already Boltzmann suppressed by $\sim e^{-m/T} \sim 10^{-2}$ just after the QCD phase transition. As the universe cools further, nucleons and anti-nucleons undergo more annihilations and the Boltzmann suppression of their number densities continues such that

$$\frac{n_{p,eq} - \bar{n}_{p,eq} + n_{n,eq} - \bar{n}_{n,eq}}{s} \approx 2\sinh(\mu_N/T)\frac{4e^{-m_p/T}[m_pT/(2\pi)]^{3/2}}{2\pi^2 g_{*s}(T)T^3/45} = B, \tag{2.42}$$

where we have approximated $m_n = m_p$, and μ_N is the chemical potential of the nucleons in equilibrium. Once μ_N/T becomes order one, nucleon annihilations freeze-out because the anti-nucleons are not in sufficient concentration to significantly disrupt nucleon number density. Plugging $\mu_N/T = 1$ in the above equation gives us an approximate estimate of the freezeout temperature, $T \sim 40$ MeV. Note that anti-nucleons continue to undergo further annihilations below $T \sim 40$ MeV because nucleon density has stopped falling exponentially. Consequently, the freeze-out of anti-nucleon annihilation occurs later with a relic anti-nucleon density orders of magnitude smaller than the relic nucleon density.

2.3.3 Neutrino decoupling

So far we have focussed on periods when the interactions between SM particles are large enough to maintain thermal equilibrium between all the particles. However, as the SM particles cool in an expanding universe, the interaction rate eventually falls below the Hubble rate. Among all the stable SM particles, neutrinos have the weakest interaction with rest of the particles. Consequently, neutrinos are the first to thermally decouple at temperatures around $T_{SM} \sim 1.5$ MeV.

The neutrinos maintain equilibrium with other SM particles via scattering with electrons,

$$e^\pm + \nu_e \rightleftarrows e^\pm + \nu_e, \qquad\qquad e^\pm + \bar{\nu}_e \rightleftarrows e^\pm + \bar{\nu}_e. \tag{2.43}$$

Both the above interactions are mediated by W and Z bosons. At temperatures much smaller than W and Z

matter, which includes both nucleons and electrons.

boson mass, the interaction rate for the above processes is well approximated as

$$\Gamma_{e\nu} = n_e \sigma_{e\nu} \sim G_F^2 T_{SM}^5, \tag{2.44}$$

where $G_F = 1.16 \times 10^{-11}$ MeV^{-2} is the Fermi constant and the above rate assumes that electrons are relativistic, $n_e \propto T_{SM}^3$. Comparing this interaction rate with the Hubble rate in eq. (2.37), we find

$$\frac{\Gamma_{e\nu}}{H} \sim \left(\frac{T_{SM}}{1.5 \text{ MeV}}\right)^3. \tag{2.45}$$

Consequently, the neutrinos thermally decouple once T_{SM} falls below $\sim$ MeV. Beyond this point the neutrinos freestream without further scattering.

Note that even though the neutrinos are no longer in thermal equilibrium, they have a well-defined temperature because the neutrino thermal distribution remains unaffected after thermal decoupling. As the momentum of each neutrino redshifts as $p \propto 1/a$, the effective neutrino temperature also scales as $T_\nu \propto 1/a$ after neutrino decoupling. Consequently, right after neutrino decoupling, we have $T_\nu = T_\gamma$, where γ stands for photons, even if neutrinos are not in thermal contact with photons.

However, once T_γ falls below the mass of the electron, $m_e = 0.511$ MeV, the electrons and positrons annihilate and heat the photons compared to neutrinos. The corresponding temperature asymmetry between neutrinos and photons after electron-positron annihilation can be readily calculated using entropy conservation. Enforcing entropy conservation above and below the electron-positron annihilation in electron-photon fluid yields,

$$(2 + 4 \times 7/8)(aT_\gamma)_i^3 = 2(aT_{\gamma,f})^3, \tag{2.46}$$

where subscript i and f refer to some time just above and below electron positron annihilation. In contrast, the neutrinos just cool as $1/a$ and initially have the same temperature as photons,

$$(aT_\gamma)_i = (aT_\nu)_f. \tag{2.47}$$

Comparing the above two equations, we obtain

$$\left(\frac{T_\nu}{T_\gamma}\right)_f = \left(\frac{4}{11}\right)^{1/3}. \tag{2.48}$$

The presence of cosmic neutrinos can be indirectly detected in the fluctuations of the cosmic microwave background (CMB) spectrum because of the gravitational influence of free-streaming neutrino radiation density. In particular, the CMB is sensitive to the ratio ρ_ν/ρ_γ which is parameterized in terms of effective number of neutrino species, $N_{\rm eff}$,

$$\frac{\rho_\nu}{\rho_\gamma} = \frac{7/8 \times N_{\rm eff} \times 2 \times T_\nu^4}{2 \times T_\gamma^4} = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}. \tag{2.49}$$

The SM prediction of $N_{\rm eff}^{\rm SM}$ is 3.044 [31–35], which is slightly above three because the neutrinos have not completely decoupled from the photons when the electron-positron annihilations occur. The current measurement of $N_{\rm eff}$ from Planck mission is $N_{\rm eff} = 2.96 \pm 0.33$ [3], which agrees with the SM prediction.

2.3.4 Big Bang Nucleosynthesis

After the QCD phase transition, all the quarks are bounded into hadrons and mesons, of which the longest lived particles are protons and neutrons. Unbounded free neutrons are unstable because they decay into protons along with electrons and neutrinos via weak process. However, prior to all the neutrons can decay, the neutrons combine with protons to form nuclei such as helium, deuterium, lithium, etc. In this section we briefly review the formation of these nuclei in the early universe. In particular we highlight how their abundances are sensitive probes of $N_{\rm eff}$ and the baryon-to-photon ratio, $\eta_b = \rho_b/(m_p n_\gamma)$.

There are three primary processes that together determine the final relic abundance of nuclei:

1. At $T_\gamma \lesssim 1$ MeV, the reactions $n + \bar{\nu}_e \to p + e^-$ and $n + e^+ \to p + \nu_e$ start depleting neutrons. However, before any significant depletion occurs, the rate of the reaction falls below the Hubble rate and the depletion is halted.

2. At $T_\gamma \lesssim 0.1$ MeV, the age of the universe is around the half-life of the neutron decay process $n \to p + \nu_e + e^-$ and thus neutron depletion again starts to occur.

3. Fortunately, before the decays become significant, the neutrons get absorbed to form deuterium, $n + p \to D + \gamma$, at $T_\gamma \sim 0.07$ MeV. Once there is a significant abundance of deuterium, the deuterium quickly reacts with protons to form Helium and Lithium. Note that the neutron number is conserved as soon as neutrons are first bounded into deuterium.

We next look at equations governing each of the above steps in more detail.

The neutrons are slightly heavier than protons by $\Delta \equiv m_n - m_p = 1.293$ MeV. Consequently, at temperatures $T_\gamma \gg \Delta$, the same amount of forward and backward reaction of neutron-to-proton conversion occurs and the number densities of both neutrons and protons are equal, $n_n = n_p$. However, once T_γ falls below Δ, backward reactions becomes less kinetically feasible as electrons and neutrinos do not have enough kinetic energy to compensate for the mass difference. Hence, the abundance of neutrons start falling compared to that of protons.

The evolution of neutron and proton number abundances around $T_\gamma \sim \Delta$, is well described by the Boltzmann equations

$$\frac{dn_n}{dt} + 3Hn_n = -\mathcal{C}_{n\to p} \tag{2.50}$$

$$\frac{dn_p}{dt} + 3Hn_p = +\mathcal{C}_{n\to p}, \tag{2.51}$$

where $\mathcal{C}_{n\to p}$ is the collision term describing depletion of neutrons to protons via $n + \bar{\nu}_e \to p + e^-$ and $n + e^+ \to p + \nu_e$. In the limit $\mathcal{C}_{n\to p} = 0$, the above Boltzmann equations simply yield $n_n \propto 1/a^3$ and $n_p \propto 1/a^3$, as expected. The collision term, $\mathcal{C}_{n\to p}$, is given by

$$\mathcal{C} = \int \frac{d^3p_n}{2E_n(2\pi)^3}\frac{d^3p_{\bar{\nu}_e}}{2E_{\bar{\nu}_e}(2\pi)^3}\frac{d^3p_p}{2E_p(2\pi)^3}\frac{d^3p_{e^-}}{2E_{e^-}(2\pi)^3}(2\pi)^4\delta^3(\vec{p}_n + \vec{p}_{\bar{\nu}_e} - \vec{p}_p + \vec{p}_{e^-})\delta(E_n + E_{\bar{\nu}_e} - E_p - E_{e^-})|\mathcal{M}|^2$$
$$[f_n(p_n)f_{\bar{\nu}_e}(p_{\bar{\nu}_e})(1 \pm f_p(p_p))(1 \pm f_{e^-}(p_{e^-})) - f_p(p_p)f_{e^-}(p_{e^-})(1 \pm f_n(p_n))(1 \pm f_{\bar{\nu}_e}(p_{\bar{\nu}_e}))]. \tag{2.52}$$

As the temperature is close to MeV where neutrons and protons are non-relativistic, their distribution functions are well approximated as $f \sim e^{(-m+\mu-\vec{p}^2/2m)/T}$. The electrons and neutrinos are relativistic, but we

neglect the effect of quantum statistics (Pauli blocking) and assume $f \sim e^{-|\vec{p}|/T}$. Furthermore, the amplitude for the above processes is given by $|\mathcal{M}|^2 = 39.26\pi^3\tau_n^{-1}m_p^2 p_\nu p_e/m_e^5$, where $\tau_n = 886.7$ sec is the neutron lifetime. With the above replacements, the phase space integral of the collision term can be performed analytically to obtain

$$\mathcal{C} = \Gamma_{n\to p}(x)[n_n - n_p e^{-x}], \tag{2.53}$$

where $x = \Delta/T_\gamma$ and $\Gamma_{n\to p}$ is the rate at which the neutron to proton conversion reaction proceeds,

$$\Gamma_{n\to p}(x) = \frac{255}{\tau_n}\frac{1}{x^5}(x^2 + 12 + 6x). \tag{2.54}$$

Since we are primarily concerned with the neutron abundance relative to proton, it is convenient to express n_n in terms of neutron fraction, $X_n = n_n/(n_n + n_p)$, and total baryon abundance, $n_B = n_n + n_p$. Correspondingly, the equations given in eq. (2.50) simplify to yield $n_B \propto 1/a^3$ and

$$x\frac{dX_n}{dx} = -\frac{\Gamma_{n\to p}(x)}{H(x)}[X_n - (1 - X_n)e^{-x}]. \tag{2.55}$$

Note that as $H \propto T_{SM}^2 \propto 1/x^2$, the ratio $\Gamma_{n\to p}/H$ grows as $1/x$ for $x \to 0$. Consequently, at $T_\gamma \gg \Delta$, the above equation enforces the factors inside the bracket in the RHS to go to zero, which yields $X_n = 1/2$ as expected. Conversely, in the limit $x \to \infty$, the coefficient of RHS itself goes to zero and we obtain $X_n =$ constant. The constant asymptote is a consequence of the fact that neutrons cannot efficiently interact with electrons and neutrinos because the expanding universe is diluting their abundance faster than the rate at which the conversion occurs.

While one has to solve the above equation numerically to find the exact relic value of X_n, one can obtain an $\mathcal{O}(1)$ estimate using the instantaneous freeze-out approximation. The instantaneous freeze-out approximation assumes that the concerned process is in perfect equilibrium until $\Gamma = H$, after which the process is instantaneously shut off. Using the Hubble rate given in eq. (2.37) and using that $g_* = 2 + \frac{7}{8}(2N_{\rm eff} + 4) \equiv 10.75(1 + \delta_N)$ near $T_{SM} \sim 1$ MeV, we find $\Gamma_{n\to p}(x_f) = H(x_f)$ at

$$x_f^{-1} = \left(4\sqrt{1 + \delta_N} + \frac{1}{18}\right)^{1/3}\frac{1}{12^{1/3}} - \frac{1}{6}. \tag{2.56}$$

For $N_{\rm eff} = 3$ or $\delta_N = 0$, we find freeze-out of reaction occurs at $x_f \approx 1.89$ or $T_f \approx 0.7$ MeV. The final value of X_n is then approximated as $X_{n,\infty} \approx X_{n,eq}(x_f)$, where $X_{n,eq}$ is the X_n for which the RHS of eq. (2.55) is zero. This yields

$$X_{n,\infty} \approx \frac{1}{1 + e^{x_f}}. \tag{2.57}$$

The above analytical result gives $X_{n,\infty} = 0.13$ when $\delta_N = 0$, in comparison to the exact numerical result $X_{n,\infty} = 0.15$.

The remaining abundance of neutrons can convert into protons once neutron decays begin, $n \to p + \nu_e + e^-$, which occurs when age of the universe is around the neutron lifetime, τ_n. The corresponding Boltzmann

equation for this process is

$$\frac{dn_n}{dt} + 3Hn_n = -\tau_n^{-1} n_n. \tag{2.58}$$

Again we can rewrite the above process by replacing $n_n = X_n n_b$ and then integrate the differential equation to obtain

$$X_n(T_\gamma) = X_i \exp\left[-\frac{3}{2\pi}\frac{M_{\rm Pl}}{\tau_n}\left(\frac{10}{g_*}\right)^{1/2}\left(\frac{1}{T_\gamma^2} - \frac{1}{T_i^2}\right)\right], \tag{2.59}$$

where T_i is some initial temperature and we assumed g_* is constant between T_i and T_γ. The decays primarily take place after electron-positron annihilation. So we have $g_* = 3.36 + 7/8 \times 2 \times N_{\rm eff} \times (4/11)^{4/3} = 3.36(1 + 0.82\delta_N)$. Then taking $X_{n,i} = X_{n,\infty}$ and neglecting $1/T_{\gamma,i}^2$ yields,

$$X_n(T_\gamma) \approx \frac{1}{1+e^{x_f}} \exp\left[-\frac{0.3}{\sqrt{1+0.82\delta_N}}\left(\frac{0.07\ \text{MeV}}{T_\gamma}\right)^2\right]. \tag{2.60}$$

We can see that as T_γ continues to drop, the abundance of neutrons exponentially decreases.

The neutron decay stops once deuterium production starts and neutrons and protons are bounded into stable nuclei. To find the temperature when the deuterium production becomes feasible, we need to consider detailed balance for $n + p \to D + \gamma$, where D stands for deuterium. In thermal equilibrium, the chemical potential on both side of the reaction should match. Photons in thermal equilibrium do not have a chemical potential, so we obtain

$$e^{(\mu_n+\mu_p)/T_\gamma} = e^{\mu_D/T_\gamma} \tag{2.61}$$

$$\frac{n_n}{n_n^{eq}}\frac{n_p}{n_p^{eq}} = \frac{n_D}{n_D^{eq}}, \tag{2.62}$$

$$\frac{n_D}{n_p n_n} = \frac{n_D^{eq}}{n_p^{eq} n_n^{eq}} = \frac{(4\pi)^{3/2}}{2}\frac{1}{(m_p T_\gamma)^{3/2}} e^{B_D/T_\gamma} \tag{2.63}$$

$$\frac{n_D}{n_b} = X_n(1-X_n)\frac{(4\pi)^{3/2}}{\pi^2}\eta_b\Big(\frac{T_\gamma}{m_p}\Big)^{3/2} e^{B_D/T_\gamma} \sim \eta_b\Big(\frac{T_\gamma}{m_p}\Big)^{3/2} e^{B_D/T_\gamma}, \tag{2.64}$$

where $B_D = m_n + m_p - m_D$ is the binding energy for deuterium, and we used $n_n = X_n n_b$, $n_p = (1 - X_n)n_b$, and $\eta_b = n_b/n_\gamma$ to obtain the last line. The neutron decays almost stop instantaneously as soon as n_D/n_b becomes of order one. Solving for $T_\gamma = T_{\rm nuc}$ at which deuterium abundance reaches $\sim \mathcal{O}(1)$ yields,

$$\ln(\eta_b) + \frac{3}{2}\ln(T_{\rm nuc}/m_p) \sim -B_D/T_{\rm nuc}. \tag{2.65}$$

We can see that the determination of $T_{\rm nuc}$ is primarily set by the logarithm of η_b and is independent of the Hubble rate. For $\eta_b \sim 10^{-10}$, we find $T_{\rm nuc} \sim 0.07$ MeV. Note that naively, deuterium production would have happened much before at $T \sim B_D = 2.22$ MeV. However, due to very tiny amount of baryons in the plasma, this process is kinematically postponed until $T \sim 0.1$ MeV.

The final abundance of neutrons in the universe is obtained by setting $T_\gamma = T_{\rm nuc}$ in eq. (2.60). Note that even though the decay can potentially cause an exponential reduction in X_n, the fact that $T_{\rm nuc}$ occurs very close to the half life of neutrons halts the neutron decay rather quickly.

After $T_{\rm nuc}$, the number of neutrons remain conserved. While neutrons are first assembled in deuterium

nuclei, they are quickly transferred into He^4 nuclei because the binding energy of Helium nuclei is much larger than that of deuterium and other nuclei with small atomic numbers. Formation of much heavier nuclei with even larger binding energies does not occur because their formation typically require three particle interactions. Nuclear processes involving more than two particles occur very rarely in the early universe because baryon concentrations are very low.

Thus, virtually all the neutrons are swept up inside the helium nuclei after $T_{\rm nuc}$. Consequently, one can estimate the mass fraction of helium to hydrogen nuclei, $Y = m_{He} n_{He}/(m_p n_H)$, simply as

$$Y = 2X_n(T_{\rm nuc}) \approx 2\frac{1}{1+e^{x_f}} \exp\left[-\frac{0.3}{\sqrt{1+0.82\delta_N}}\right]. \tag{2.66}$$

For $\delta_N = 0$, the above yields $Y = 0.2$ which is slightly smaller than the numerical estimate of $Y = 0.22$. The present day abundance of helium to hydrogen mass fraction agrees with the primordial abundance calculated above. Consequently, the successful prediction of BBN provides stringent constraint on any deviation from Standard Cosmology at $T_\gamma \lesssim 1$ MeV.

Note that not all of the deuterium gets converted into Helium. This is because the rate of nuclear reactions converting deuterium into Helium eventually fall below the Hubble rate once the concentration of deuterium falls below a threshold. Compared to $T_{\rm nuc}$, the freeze-out of the nuclear reactions is much more sensitive to η_b. Consequently, the leftover trace of deuterium is a much more sensitive probe of η_b than the abundance of Helium.

2.4 Thermal origin for dark matter: WIMP

In this section we describe the mechanism by which dark matter can obtain the observed relic abundance by being part of the SM plasma in the early universe. Additionally, we emphasize a simple way to calculate the model parameters that yield the correct relic abundance.

Let us imagine that the particle comprising the dark matter today, χ was part of the SM plasma in the early universe. One can naively imagine a process of type: $\chi + \chi \to l + l$ kept dark matter in thermal equilibrium, where l is some particle in the SM that is much lighter than dark matter. Let us further consider χ to have no particle anti-particle asymmetry. Correspondingly, the relic abundance of χ is primarily set by when the annihilation reactions of χ drop out-of-equilibrium and the number density of χ is frozen in. In particular, we focus in the region of parameter space where the Boltzmann suppression of equlibrium dark matter abundance after $T < m_\chi$ causes the annihilation reaction to fall out-of-equilibrium.

One can find the relevant Boltzmann equations for dark matter number density in the limit of non-relativistic dark matter similar to how we derived eq. (2.25) earlier. Except, the equation is further simplified here because we consider $n_\chi = n_{\bar\chi}$,

$$a^{-3}\frac{d(a^3 n_{DM})}{dt} = \langle\sigma v\rangle(n_{DM,eq}^2 - n_{DM}^2), \tag{2.67}$$

where $n_{DM,eq}$ is the equilibrium number density of DM. Additionally, considering dark matter annihilation is a s-wave process, in the non-relativistic limit the annihilation cross-section has no temperature dependence, $\langle\sigma v\rangle =$ constant.

With the Hubble rate calculated considering a SM radiation-dominated universe, the above equation can be solved with the initial condition that $n_{DM} = n_{DM,eq}$. Because both entropy density and dark matter number

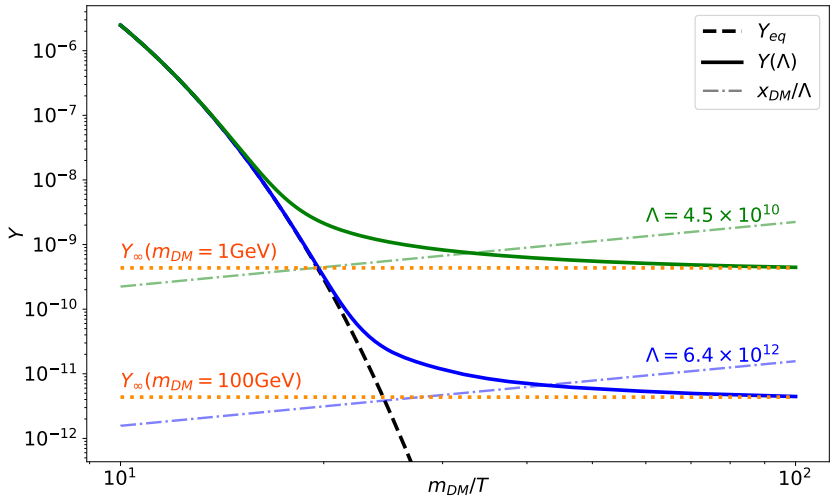


Figure 2.2: Evolution of $Y = n_{DM}/s$ as a function of $x = m_{DM}/T$. The black dashed line shows equilibrium evolution, the solid lines show evolution of Y for a given value of Λ, the dashed lines follows x_{DM}/Λ, and the dotted lines shows the value of Y_∞ for a given m_{DM}. The green lines have been plotted for $\Lambda = 4.5 \times 10^{10}$ while the blue lines have been plotted for $\Lambda = 6.4 \times 10^{12}$. We have set $g_* = g_{*s} = 100$ for the above plot for simplicity.

density evolve as $\propto 1/a^3$ after dark matter annihilations have stopped, we rewrite the above Boltzmann equation in terms of

$$Y = \frac{n_{DM}}{s}, \qquad x_{DM} = \frac{m_{DM}}{T}, \qquad \Lambda = \frac{s(T = m_{DM})\langle\sigma v\rangle}{H(T = m_{DM})}. \tag{2.68}$$

Doing so yields,

$$\frac{dY}{dx_{DM}} = -\left(1 - \frac{d\ln g_{*s}}{3d\ln x_{DM}}\right)\left(\frac{g_*(m_{DM})}{g_*(m_{DM}/x_{DM})}\right)^{0.5}\left(\frac{g_{*s}(m_{DM}/x_{DM})}{g_{*s}(m_{DM})}\right)\frac{\Lambda}{x_{DM}^2}(Y^2 - Y_{eq}^2). \tag{2.69}$$

Above the derivative of g_{*s} was obtained by calculating dx_{DM}/da using entropy conservation. For simplicity, in the following analysis we assume that the freeze-out of dark matter annihilation occurs away from mass thresholds, such that g_* and g_{*s} are constant.

For a given Λ, one can see that the above equation causes $Y = Y_{eq}$ in the early universe when $x_{DM} \ll 1$. As x becomes larger than one, Y_{eq} starts to fall exponentially as $e^{-x_{DM}}$. The value of Y follows the drop in Y_{eq} until,

$$Y_{eq}(x_f) = \frac{x_f}{\Lambda}. \tag{2.70}$$

Beyond this point, $x_{DM} > x_f$, Y_{eq} becomes too small and can be neglected. Note that as x_{DM} is increasing, the overall coefficient on the RHS is also decreasing. Thus soon after the contribution from the whole RHS becomes negligible and Y asymptotes to a constant. This asymptote sets the relic abundance of dark matter.

In figure 2.2, the solid lines show the evolution of Y as a function of x for two different values of Λ. The black dashed line shows the evolution of Y_{eq} and we can see that initially Y follows Y_{eq}. Once the value of Y_{eq} and Y drops near x_{DM}/Λ, Y departs from its equilibrium value and starts to asymptote to a constant.

Considering the asymptote value of Y to be Y_∞, we require that Y_∞ appropriately yields the present day dark matter density. Thus we have

$$Y_\infty = \frac{\rho_{DM,0}}{m_{DM} s(T_0)} = 4.3 \times 10^{-10} \left(\frac{\Omega_{DM} h^2}{0.12}\right) \left(\frac{\text{GeV}}{m_{DM}}\right). \tag{2.71}$$

where we have used the fact that $\rho_{DM,0} = \rho_{crit} \Omega_{DM} = 0.81 \times 10^{-46} \Omega_{DM} h^2 \text{GeV}^4$. Figure 2.2 also shows different values of Y_∞ that obtain the correct relic abundance for different dark matter masses.

From the figure it is apparent that a given value of m_{DM} fixes the value of Λ that yields the correct relic abundance of dark matter. Thus we obtain a one-to-one mapping between the required values of annihilation cross-section, $\langle \sigma v \rangle$, and the value of m_{DM} that gives the correct relic abundance. In practice we find that $\langle \sigma v \rangle$ is only weakly sensitive to m_{DM}. To understand why let us analytically solve the freeze-out abundance of dark matter.

One can analytically solve the freeze-out equation using instantaneous freeze-out approximation, where the following three equalities hold at x_f,

$$Y_{eq}(x_f) = \frac{x_f}{\Lambda} = Y_\infty. \tag{2.72}$$

From figure 2.2, we can see why the above is approximately valid.

The simplest way to solve for x_f is by equating the first and last term. In this way we avoid any superfluous dependence of x_f on Λ, which we would have obtained if matched the second and third term or first and second term. As Y_∞ is inversely proportional to m_{DM} and Y_{eq} is exponentially dependent on x_f, x_f only depends logarithmically on m_{DM}.

With x_f on hand we can directly evaluate Λ by matching the second and third term above. As Λ is proportional to m_{DM}, the factors on m_{DM} on both sides cancel and we obtain,

$$\langle \sigma v \rangle = 4.1 \times 10^{-10} \text{ GeV}^{-2} \left(\frac{x_f}{20}\right) \left(\frac{0.12}{\Omega_{DM} h^2}\right) \left(\frac{g_*(m_{DM})}{100}\right)^{0.5} \left(\frac{100}{g_{*s}(m_{DM})}\right) \tag{2.73}$$

The above cross-section has a weak dependence on m_{DM} through x_f.

A typical form of cross-section is given by

$$\langle \sigma v \rangle \sim \frac{\lambda^2}{16\pi m_{DM}^2}, \tag{2.74}$$

where λ is the coupling strength of dark matter with the light SM particles. Interestingly, one sees that for λ^2 being around weak coupling strength $\alpha_W^2 \sim 10^{-3}$ and for m_{DM} being around weak scale mass $m_W \sim 100$ GeV, we obtain the correct expected value of dark matter abundance. This coincidence is known as the WIMP miracle.

2.5 Out of equilibrium sectors and cosmological attractors

So far we have focussed on particles that were in equilibrium with the SM plasma in the early universe. The rest of thesis explores the dynamics of a hidden sector of particles that were always out-of-equilibrium with the SM plasma. In this section, we provide a general insight on how out-of-equilibrium sectors evolve if we consider there might be energy transfer from one sector to the other. Our primary interest here is in the

energy density contained in a thermal radiation bath, where notable examples of such attractor solutions include the $T \propto a^{-3/8}$ evolution of a radiation bath during (classical, perturbative) reheating [36] and the $T \propto a^{-3/4}$ behavior of a radiation bath fed by out-of-equilibrium renormalizeable scattering processes [37, 38]. Another important class of examples is realized by various models of freeze-in dark matter [15, 39], where the relevant quantity is the number density of DM. We show that the energy density of the dilute sector follow a quasi-static attractor solution in a sense that we make precise here.

Since we are considering two sectors that might have much more stronger internal interactions than interactions between them, we need to solve for energy momentum conservation in each of the sectors independently,

$$\frac{\mathrm{d}\rho_a}{\mathrm{d}t} + 3H\rho_a(1+w_a) = -\mathcal{C}_E \tag{2.75}$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + 3H\rho(1+w) = \mathcal{C}_E. \tag{2.76}$$

where the collision term encodes the amount of energy transfer from one sector to the other and the Hubble rate is determined by

$$H = \frac{1}{\sqrt{3}M_{\rm Pl}}\sqrt{\rho_a + \rho}. \tag{2.77}$$

Generically, one can consider one sector is much more dilute compared to the other, $\rho \ll \rho_a$. Consequently, the Hubble rate is primarily determined by ρ_a. Furthermore, one can neglect the effect of $\mathcal{C}$ on ρ_a as any energy transfer from b is insignificant for a. Under these conditions, ρ_a evolves independent of the other sector, while the evolution of the other sector is driven by ρ_a

With the above approximation, we can rewrite the equation for ρ in the form

$$a\frac{\mathrm{d}\rho}{\mathrm{d}a} + 3(1+w)\rho = \frac{\mathcal{C}_E}{H} \equiv F(\rho, a). \tag{2.78}$$

By defining new variables as

$$\lambda(a) = \frac{\rho(a)}{F(\rho(a), a)}, \qquad p(a) = \frac{\partial \ln F}{\partial \ln \rho}, \qquad q(a) = \frac{\partial \ln F}{\partial \ln a}, \tag{2.79}$$

and assuming w to be constant, we can further modify eq. (2.78) to yield

$$\frac{\mathrm{d}\ln\lambda}{\mathrm{d}\ln a} = (1-p)\frac{1}{\lambda} - 3(1+w)(1-p) - q. \tag{2.80}$$

This equation dictates how the ratio of $H\rho/\mathcal{C}_E$ evolves depending on the functional behavior of $F(\rho, a) = \mathcal{C}_E/H$ encoded in p, q. Now note that for $p < 1$ and $q > -3(1+w)(1-p)$,

$$\lambda = \frac{1}{3(1+w) + \frac{q}{1-p}} \tag{2.81}$$

is a stable attractor solution for this equation, provided that p and q slowly vary with a ($\frac{\partial \ln(p,q)}{\partial \ln a} \ll 1$). This solution is an attractor: radiation baths initially below this steady-state solution rise up very rapidly to meet it, while radiation baths initially above it redshift as $\rho \propto a^{-3(1+w)}$ until the attractor solution is attained. The attractor nature of the system of equations is a consequence of the fact that the process of thermalization

tends to remove the information about initial conditions.

The quasi-static behavior of ρ can be found by simply solving the equation

$$\rho(a) = \frac{1}{3(1+w) + \frac{q}{1-p}} \frac{\mathcal{C}_E}{H}. \tag{2.82}$$

The above result just highlights the fact that in a given e-fold, the energy density in the dilute sector is dominated by the amount of energy density transferred from the dominant sector in a Hubble time, $\mathcal{C}_E/H$. In cases of cosmological interest F very frequently has power law dependence on ρ and a, thus making λ a fixed and readily computable constant (usually of $O(1)$). In such cases, the relevant power law describing the temperature evolution can then be quickly obtained by solving $\rho \propto \mathcal{C}/H$.

When $q < -3(1+w)(1-p)$, there is no attractor solution (as λ is always positive) and λ increases uncontrollably. This corresponds to the cases when $\mathcal{C}_E$ is falling faster than the redshifting of the energy density, and the evolution of ρ is approximately adiabatic. On the other hand, when $p > 1$, the attractor solution (when it exists) is not stable. If $\mathcal{C}_E$ ever came to dominate in this scenario then it would lead to an indefinite explosive rise in ρ due to the positive feedback from $\mathcal{C}_E$. The subsequent solution can be obtained by simply solving $\dot{\rho} = \mathcal{C}_E$.

We can perform an analogous exercise for number density. The Boltzmann equation we start with here is

$$\frac{\mathrm{d}n}{\mathrm{d}t} + 3Hn = \mathcal{C}, \tag{2.83}$$

and, defining

$$\kappa(a) = \frac{n(a)}{F(n,a)}, \qquad p(a) = \frac{\partial \ln F}{\partial \ln n}, \qquad q(a) = \frac{\partial \ln F}{\partial \ln a}, \tag{2.84}$$

we may rewrite this equation as

$$\frac{\mathrm{d}\ln\kappa}{\mathrm{d}\ln a} = (1-p)\frac{1}{\kappa} - 3(1-p) - q. \tag{2.85}$$

Then the attractor solution is given by

$$\kappa = \frac{1}{3 + \frac{q}{1-p}}, \text{ or} \qquad n(a) = \frac{1}{3 + \frac{q}{1-p}} \frac{\mathcal{C}}{H}. \tag{2.86}$$

Chapter 3

Reheating in two-sector cosmology

3.1 Introduction

There is now strong evidence from observations of the fluctuations in the CMB that the thermal era was preceded by an epoch of early accelerated expansion—inflation. Inflation exponentially dilutes any pre-existing matter and radiation leaving the Universe cold and empty. The population of otherwise decoupled sectors cannot therefore be put in 'by hand' as an initial condition. Instead it must be generated dynamically in the post-inflationary evolution of the Universe. In the simplest scenarios, the accelerated expansion is driven by a single fundamental scalar degree of freedom, whose weak couplings to matter reheat the Universe via perturbative decays. One of the simplest mechanisms for populating hidden sectors is to couple them to the inflaton so that they are populated at reheating along with the visible sector. By arranging the couplings so that the hidden sector couples differently to the inflaton than the SM, reheating can be asymmetric, whereby the SM and the hidden sector are heated to different temperatures [40–43]. However, coupling both the SM and a hidden sector to the inflaton in the UV necessarily results in inflaton-mediated interactions between the two sectors. As demonstrated in reference [42], this irreducible inflaton-mediated scattering can thermalize the two sectors under fairly generic conditions.

This chapter is based on the work in Ref. [17], where we extend the analysis of reference [42] to explore the effects of out-of-equilibrium inflaton-mediated interactions on asymmetric reheating. Along the way, we develop and implement methods to numerically solve the Boltzmann equations describing the reheating of two otherwise-decoupled sectors from the perturbative decay of the inflaton. In particular, we develop accurate approximations (including the effects of quantum statistics) for the collision terms that describe the inflaton-mediated scattering between thermalized gases of fermionic and bosonic particles. We take an effective field theory approach and consider combinations of trilinear scalar, Yukawa, and pseudo-scalar couplings between fermions, bosons, and the inflaton. When inflaton couplings to matter become sufficiently large, both non-perturbative effects such as preheating and collective effects in the radiation baths such as Landau damping and thermal masses can provide important corrections to the inflaton decay rate and hence the evolution of the temperature asymmetry, particularly at very high radiation temperatures [44–46]. However, as we show here, both inflaton decays and inflaton-mediated scattering furnish cosmological attractor solutions during the perturbative phase of reheating, making the final temperature asymmetry largely sensitive to the dynamics of the system at and below the perturbative reheat temperature $T_{\rm rh}$. Thus the perturbative reheating process that we analyze in the present chapter will often serve as a good guide to

the final temperature asymmetry despite the presence of richer dynamics at early times.

This chapter is organized as follows. In section 3.2, we review standard perturbative reheating and extend the well-known single-sector results to include the effects of quantum statistics on the decay width of the inflaton. In section 3.3, we begin our study of reheating into two sectors and introduce the inflaton-mediated interactions between the two sectors (required by self-consistency). We show how inter-sector scattering dominates over any features from quantum statistics in most of the parameter space. We conclude in section 3.4.

We work in units where $\hbar = c = k_B = 1$, and denote by $M_{\rm Pl} = 2.435 \times 10^{18}$ GeV the reduced Planck mass.

3.2 Quantum statistics in single-sector reheating

In this section we revisit the perturbative reheating of a radiation bath. After reviewing the classic treatment, we demonstrate that at temperatures $T > M_\phi/4$, where M_ϕ is the inflaton mass, quantum effects such as Bose enhancement and Pauli blocking can significantly affect the evolution of the temperature of the radiation bath during reheating. We show that the effects of quantum statistics disappear once T drops below $M_\phi/4$, and thus alter the outcome of reheating only when $T_{\rm rh} \gtrsim M_\phi/4$. While we refer to the decaying particle as an inflaton and have post-inflationary reheating primarily in mind, our results apply also to other "reheatons" such as curvatons or moduli (see also [47]).

3.2.1 Perturbative reheating

A generic scenario of inflation [48–50] consists of one or more scalar fields ϕ_i slowly rolling on a sufficiently flat potential, $V(\phi_i)$ (see, for example, [51] and references within). Inflation ends when the slow-roll conditions are violated, and the fields ϕ_i roll quickly to the potential minima and start oscillating. For this work, we assume that only one field ϕ is relevant during the reheating process, and that its potential is analytic and can be expanded in a Taylor series about its minimum. We further assume that only the leading quadratic term in this Taylor series is needed.[1] The time-averaged equation of state of a field oscillating in a quadratic potential is that of a stationary massive particle, and thus the Universe undergoes a period of matter domination while the inflaton energy density dominates [54]. During this oscillating phase, the inflaton condensate starts to decay through its couplings to matter, initiating reheating. If these couplings are large enough, the first stage of reheating can proceed through a period of parametric resonance known as preheating [55, 56]. In the preheating regime, particle production is non-perturbative and typically requires numerical treatment (however, see [57]). As the amplitude of inflaton oscillations decreases, due to both Hubble friction and inflaton decay, preheating ceases and particle production can be treated perturbatively. Unless preheating is violent enough to drain an $\mathcal{O}(1)$ fraction of energy out of the inflaton condensate, this final epoch of perturbative reheating typically dominates the properties of the radiation bath produced by inflaton decays.

For this work, we thus consider perturbative reheating in a quadratic potential [58, 59]. We consider the generic case where all particle masses besides the inflaton mass are negligible at the energies we consider, and therefore treat all matter species as radiation. We further neglect inverse decays from radiation into inflaton quanta; this is a good approximation provided the number of species in the radiation bath is large, $g_* \gg 1$.

[1] We are explicitly ignoring anharmonic corrections to the inflaton potential that may be relevant during reheating. These anharmonic terms can be important for non-perturbative effects during reheating, such as the formation of oscillons, as recently reviewed in [52]. Conversely, the absence of a quadratic minima generically leads to a radiation equation-of-state very quickly following inflation [53].

With these approximations, the Boltzmann equations describing reheating read (see, for example, [36])

$$\frac{\mathrm{d}\rho_\phi}{\mathrm{d}t} + 3H\rho_\phi = -\Gamma\rho_\phi \tag{3.1}$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + 4H\rho = \Gamma\rho_\phi, \tag{3.2}$$

where the Hubble rate H is given by the Friedmann equation

$$H = \frac{1}{\sqrt{3}M_{\rm Pl}}\sqrt{\rho+\rho_\phi}\,. \tag{3.3}$$

The inflaton width is denoted by Γ, and ρ_ϕ and ρ are the inflaton and radiation energy densities, respectively. These equations are (approximately) valid from the end of inflation at some scale factor $a = a_I$, which we take as our initial point. The radiation sector is initially empty,[2] $\rho_{,I} = 0$, whereas the initial energy density of the inflaton is given in terms of the mean-square value of the inflaton field just after the end of inflation, $\langle\phi_I^2\rangle$, as $\rho_{\phi,I} = M_\phi^2\langle\phi_I^2\rangle/2$.

In figure 3.1, we plot the inflaton and radiation energy densities obtained by numerically solving eqns. (3.1) and (3.2) with a constant inflaton decay width, $\Gamma = \Gamma_0$. Initially, $\Gamma \ll H$ and therefore inflaton decays are inefficient. Thus the inflaton energy density during this phase can be well approximated as diluting only through redshifting, $\rho_\phi \approx \rho_{\phi,I}(a/a_I)^{-3}$. The evolution of the radiation sector, however, is dominated by the energy injection from inflaton decays. Initially, the radiation energy density grows rapidly until the rate at which energy is injected into the radiation bath by inflaton decays, governed by $\Gamma\rho_\phi$, matches the rate at which the radiation bath loses energy due to the expansion of the universe, governed by $4H\rho$. After this point, the evolution of the radiation sector follows an attractor solution, which realizes a quasi-static equilibrium between energy injection and dilution (see section 2.5),

$$4H\rho = \frac{4}{4+\frac{q}{1-p}}\Gamma\rho_\phi, \tag{3.4}$$

where

$$q(a) = \frac{\partial\ln(\Gamma\rho_\phi/H)}{\partial\ln a}, \qquad \text{and} \qquad p(a) = \frac{\partial\ln(\Gamma\rho_\phi/H)}{\partial\ln\rho}. \tag{3.5}$$

We call the above evolution imposed on the radiation sector by inflaton decays the *reheating attractor curve*. For a temperature-independent decay width, $\Gamma = \Gamma_0$, the factors q and p are readily determined to be constants, $q = -3/2$ and $p = 0$, yielding the relation $4H\rho = (8/5)\Gamma_0\rho_\phi$. On this attractor solution the radiation bath evolves as [36]

$$\rho = \left(\frac{2}{5}\Gamma_0 M_{\rm Pl}\sqrt{\rho_{\phi,I}}\right)\left(\frac{a}{a_I}\right)^{-3/2}, \tag{3.6}$$

when $a \gg a_I$. The attractor nature of this solution means that the evolution of the energy density of the radiation bath during reheating is relatively insensitive to its initial conditions. Radiation baths with energy

[2] This is generally a good approximation for models with tri-linear scalar couplings and Yukawa interactions with fermions, as in these cases the daughter fields get a large mass during inflation, shutting off inflaton decays. However, for a pseudo-scalar inflaton coupling to either fermions or gauge bosons, there can be significant energy density already in the radiation sector as inflation ends (see, for example, [60, 61]). However, as we demonstrate below, the specific initial conditions are largely irrelevant for the detailed outcome of reheating.

density initially below the attractor solution rise rapidly to meet the attractor. Meanwhile radiation densities initially above the attractor curve redshift as $\rho \propto a^{-4}$ until they meet the attractor, as can occur in (e.g.) scenarios where a modulus comes to dominate the post-reheating universe [47, 62, 63].

The attractor solution can also be obtained by solving the Boltzmann equations during reheating. Well before reheating, the inflaton condensate dominates the energy budget of the Universe; its comoving energy density is approximately constant and the Boltzmann equation describing the radiation bath can be simplified to yield [36]

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(\rho a^4\right) \quad = \quad \sqrt{3} M_{\rm Pl} a^3 \sqrt{\rho_\phi}\, \Gamma_0. \tag{3.7}$$

Solving eq. (3.7) with the initial condition $\rho(a_I) = 0$ also allows us to determine the maximum energy density attained by the radiation sector [36],

$$\rho_{\rm max} = 0.24 M_{\rm Pl} \Gamma_0 \sqrt{\rho_{\phi,I}}. \tag{3.8}$$

3.2.2 Quantum statistics during single-sector reheating

The preceding discussion neglected the possible effects of quantum statistics during reheating. Typically, the inflaton decays at rest, producing pairs of particles at a fixed energy $M_\phi/2$. To quantify the possible effects of Pauli blocking or Bose enhancement of the inflaton decay, we need to specify the phase space distribution in the radiation sector. For simplicity, we assume the radiation is in thermal equilibrium,

$$\rho = \frac{\pi^2 g_*}{30} T^4 \equiv \alpha T^4, \tag{3.9}$$

which amounts to assuming that the thermalization time scale for the radiation sector is much faster than any other time scale in the problem. This is in some sense a conservative assumption for the purpose of analyzing the scattering and reheating processes discussed in this chapter: a less equilibrated sector has a greater fraction of particles with energies concentrated near $M_\phi/2$, making both inflaton-mediated scattering and quantum statistics more important. However, as we demonstrate below, the post-reheating properties of the radiation baths are typically determined by the late-time behavior of the system, making the detailed approach to thermal equilibrium within each radiation bath largely immaterial for the final outcome of reheating. This separation of timescales generally makes prompt thermalization a robust assumption.

For temperatures $T \lesssim M_\phi/4$, a constant (zero-temperature) inflaton decay width is a good approximation. At these temperatures, the phase space where particles are injected by inflaton decays, $E \sim M_\phi/2$, is sparsely populated due to the fast thermalization of the injected particles, and thus the effects of Pauli blocking or Bose enhancement are negligible. However, at higher temperatures, $T \gtrsim M_\phi/4$, the equilibrium thermal distributions have significant support at $E \sim M_\phi/2$, and the inflaton decay rate can be significantly altered. The partial decay width of a parent scalar to pairs of particles in equilibrium at finite temperature is given by

$$\Gamma(T) = \Gamma_0 \frac{\exp(\frac{M_\phi}{2T}) \pm 1}{\exp(\frac{M_\phi}{2T}) \mp 1}, \tag{3.10}$$

where Γ_0 is the zero temperature decay width, and the upper (lower) sign holds for bosons (fermions) in the final state. At high temperatures, the decay width is enhanced (suppressed) for bosons (fermions) due to Bose enhancement (Pauli blocking). We now consider reheating to boson and fermionic radiation separately.

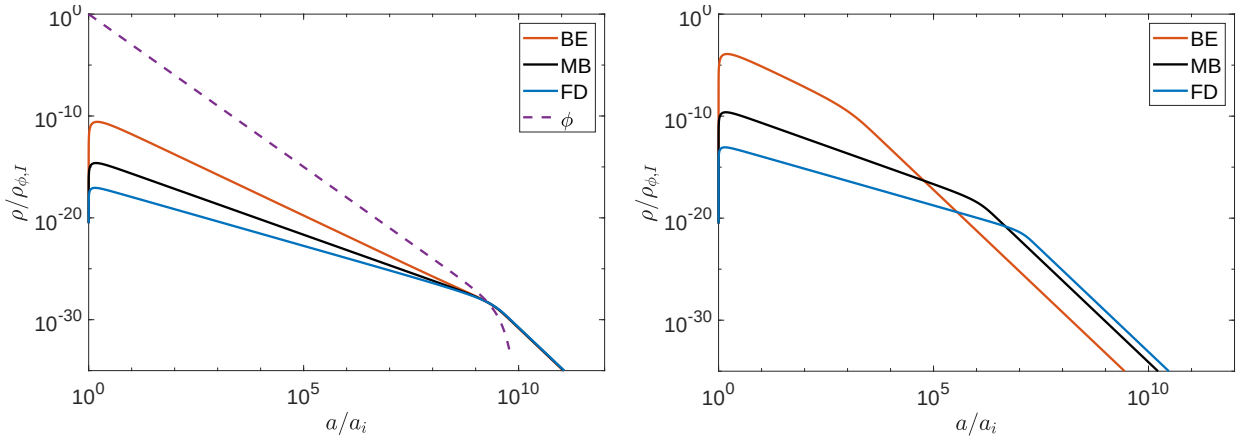


Figure 3.1: *Left panel:* Energy density as a function of scale factor during reheating for $\Gamma_0 = 10^{-14} M_\phi$, $\rho_{\phi,I} = M_{\rm Pl}^2 M_\phi^2$, and $M_\phi = 2.4 \times 10^5$ GeV. Solid lines show the energy density of a thermal radiation bath with different colors indicating different statistics: orange for Bose-Einstein (BE), black for Maxwell-Boltzman (MB) and blue for Fermi-Dirac (FD). The energy density in the inflaton field is shown by the purple dashed line. *Right panel:* Same as the left panel, with inflaton width given by $\Gamma_0 = 10^{-9} M_\phi$. For these parameters the reheat temperature is larger than the inflaton mass and hence different quantum statistics lead to different reheat temperatures.

Bosonic reheating: In the case of decays to bosons, for $T \gg M_\phi/4$ the inflaton decay width is approximately given by $\Gamma \approx 4T\Gamma_0/M_\phi$. Using this decay width in eq. (3.4) immediately yields a new quasi-static equilibrium solution for the radiation bath ($q = -3/2$ and $p = 1/4$), with power law evolution

$$T = M_\phi \left(\frac{2\sqrt{3}}{\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \right)^{1/3} \left(\frac{a}{a_I} \right)^{-1/2} . \tag{3.11}$$

This reheating attractor curve can again be found analytically by solving the approximate Boltzmann equation describing the radiation bath during reheating, analogous to the Maxwell-Boltzmann result. With initial condition $\rho(a_I) = 0$, the full temperature evolution is

$$T(a) = M_\phi \left(\frac{2\sqrt{3}}{\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \left[\left(\frac{a}{a_I} \right)^{-3/2} - \left(\frac{a}{a_I} \right)^{-3} \right] \right)^{1/3} . \tag{3.12}$$

The radiation bath attains its maximum temperature,

$$T_{\rm max} = 0.95 M_\phi \left(\frac{1}{\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \right)^{1/3} , \tag{3.13}$$

at $a = 1.6 a_I$. For bosons, the inflaton decay width decreases with temperature, making energy injection into the radiation sector less efficient as the temperature decreases. This results in the temperature dropping as $T \propto a^{-1/2}$, faster than the classical result $T \propto a^{-3/8}$ (eq. (3.6)).

Fermionic reheating: For an inflaton decaying to fermions at $T \gg M_\phi$, the decay width can be well approximated by $\Gamma \approx \Gamma_0 M_\phi/(4T)$, which gives $q = -3/2$ and $p = -1/4$. In this regime, the radiation sector

evolves as

$$T = M_\phi \Big(\frac{5\sqrt{3}}{56\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \Big)^{1/5} \left(\frac{a}{a_I} \right)^{-3/10} . \tag{3.14}$$

The full solution to the Boltzmann equations with initial condition $\rho(a_I) = 0$ is

$$T(a) = M_\phi \left(\frac{5\sqrt{3}}{56\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \left[\left(\frac{a}{a_I} \right)^{-3/2} - \left(\frac{a}{a_I} \right)^{-5} \right] \right)^{1/5} . \tag{3.15}$$

The maximum temperature attained by the radiation bath is

$$T_{\rm max} = 0.58 M_\phi \Big(\frac{1}{\alpha} \frac{M_{\rm Pl}}{M_\phi^4} \Gamma_0 \sqrt{\rho_{\phi,I}} \Big)^{1/5} , \tag{3.16}$$

attained at $a = 1.4 a_I$.

In the analytic treatment of the Boltzmann equations for reheating in the fermionic and bosonic cases above, we have taken $T(a_I) = 0$ as our initial condition. Strictly this is inconsistent with the high temperature expansion used for the inflaton width. A more complete analytic treatment would use the zero-temperature inflaton width to describe the early evolution of the radiation bath until its temperature rises to M_ϕ before implementing the high temperature expansion. However, such a procedure only alters the scale factor at which the maximum temperature is attained and not its value. Moreover, since the maximum temperature is attained very quickly compared to other timescales in our problem, the error due to this simplifying assumption is negligible. Perhaps the more consequential assumption in this region is that we have taken the radiation bath to attain internal thermal equilibrium nearly instantaneously. In the very early periods of reheating, the thermalization rate is likely to be smaller than the very rapid rate at which the energy density of the radiation bath grows. The simple solutions presented here for the decay width and the initial evolution of the energy densities are thus probably incorrect for describing these very early regions.

Once the temperature falls below the inflaton mass scale, the temperature dependence of the inflaton decay width in eq. (3.10) becomes unimportant as inflaton decays now populate sparsely occupied regions of phase space. Subsequently the radiation sector evolves as $T \propto a^{-3/8}$.

Reheating completes when the inflaton decays become efficient, $\Gamma \sim H$, and the inflaton energy density decreases exponentially. During this epoch the Universe transitions from the matter-dominated era of reheating to a radiation-dominated expansion, where the temperature of the radiation sector redshifts adiabatically as $T \propto a^{-1}$. If $\Gamma \sim H$ occurs while $T \gtrsim M_\phi/4$, then the temperature of the radiation sector directly transitions to $T \propto a^{-1}$ without going through the classical $T \propto a^{-3/8}$ regime. In this scenario, the resulting reheat temperature depends on the quantum statistics of the inflaton decay products. Estimating the reheat temperature by setting $H = \Gamma(T)$ and taking H to be dominated by the radiation bath, we find for $T_{\rm rh} \gg M_\phi$

$$T_{\rm rh} = \begin{cases} \dfrac{4\sqrt{3}}{\sqrt{\alpha}} \dfrac{M_{\rm Pl} \Gamma_0}{M_\phi} & \text{boson} \\ \left(\dfrac{\sqrt{3}}{4\sqrt{\alpha}} M_{\rm Pl} \Gamma_0 M_\phi \right)^{1/3} & \text{fermion} \end{cases} , \tag{3.17}$$

in contrast to the classical result

$$T_{\rm rh} = \left(\frac{\sqrt{3}}{\sqrt{\alpha}} M_{\rm Pl}\Gamma_0\right)^{1/2}, \tag{3.18}$$

which holds for $T_{\rm rh} \ll M_\phi$. We summarize these different power law behaviors of the radiation temperature in figure 3.1. In the left panel we show a case where the reheat temperature $T_{\rm rh}$ is below the inflaton mass. In this case, quantum statistics are unimportant for determining $T_{\rm rh}$, as all three scenarios converge onto the attractor solution governing classical perturbative reheating, eq. (3.4). In the right panel of figure 3.1 we show a case where $T_{\rm rh}$ is above the inflaton mass. As the inflaton decay width gets significant corrections from quantum statistics at these high temperatures, we observe the different reheat temperatures of eq. (3.17) expected for different quantum statistics at fixed zero-temperature decay width.

In the above scenario we have assumed that all particles coupled to the inflaton have the same quantum statistics (bosons or fermions). If the inflaton couples to both bosons and fermions then the energy density of the radiation sector as a whole evolves depending on the total inflaton decay width. In this scenario, the inflaton width is dominated by the Bose-enhanced partial widths at very high temperatures, and hence the radiation sector evolves according to the bosonic power law ($T \propto a^{-1/2}$). If the zero-temperature partial-width into fermions is larger than that to bosons, $\Gamma_0^{\rm fermion} > \Gamma_0^{\rm boson}$, then (assuming $T_{\rm rh} < M_\phi/4$) there is a temperature, T_*, for which $\Gamma^{\rm boson}(T_*) = \Gamma^{\rm fermion}(T_*)$, while $\Gamma^{\rm boson}(T) < \Gamma^{\rm fermion}(T)$ for $T < T_*$. For $M_\phi/4 < T < T_*$, the radiation bath transitions to the power law $T \propto a^{-3/10}$ (characteristic of high-temperature fermionic reheating) before ultimately transitioning to the classical $T \propto a^{-3/8}$ for $T < M_\phi/4$.

Despite model-dependent uncertainties associated with the initial evolution of the radiation baths, the attractor nature of these perturbative reheating solutions renders the later temperature evolution, and the resulting reheat temperatures, insensitive to variations to the initial conditions and early evolution provided that the attractor solution is obtained. Reaching the attractor solution requires that 1) the energy density of the oscillating inflaton dominates the Hubble rate for some time, during which inflaton decays become a cosmologically important source for the radiation bath, and 2) that the thermalization timescale is short compared to the duration of inflaton domination. As we demonstrate in the remainder of this chapter, similarly general results can be obtained for the more complicated scenarios that arise in two-sector reheating as well.

3.2.3 Preheating and the Bose power law

In previous section we showed that the evolution of radiation bath is significantly affected by quantum statistical distribution of its particles if the temperature of the radiation bath satisfies, $T > M_\phi/4$. To obtain such high temperatures, large values are required of the inflaton coupling with particles in the radiation bath. One might be concerned that such large couplings to matter place the inflaton in the regime where preheating dominates over perturbative reheating. In this section, we use a toy model to demonstrate that there is a region of parameter space where one can have perturbative production of particles to be significantly affected by quantum statistics.

We focus on a theory with an inflaton, ϕ, coupled to a scalar field, χ, via the trilinear coupling

$$\mathcal{L}_{\rm int} = \frac{1}{2}\mu\phi\chi^2. \tag{3.19}$$

This model can experience broad resonance preheating for sufficiently large values of μ and sufficiently large

inflaton oscillations $\Phi = \sqrt{\langle\phi^2\rangle}$, in which case energy is effectively drained from the inflaton condensate before perturbative reheating can occur. The condition that no broad resonances are present in the theory is

$$\mu\Phi < M_\phi^2. \tag{3.20}$$

When this condition is satisfied, preheating is inefficient and perturbative reheating dominates the properties of the radiation bath [55, 56, 64].

During perturbative reheating, for a given value of μ, the inflaton amplitude Φ uniquely specifies the temperature of the matter sector. Using the reheating attractor solution eq. (3.4) along with the quantum statistics correction to the inflaton decay width eq. (3.10), we obtain for the radiation bath

$$\alpha T^4 = \frac{1}{4 - \frac{3/2}{1-p}} \times \frac{\mu^2}{32\pi} \left(\frac{e^{\frac{M_\phi}{2T}} + 1}{e^{\frac{M_\phi}{2T}} - 1} \right) \sqrt{\frac{3}{2}} \Phi M_{\rm Pl}, \tag{3.21}$$

where

$$p = \frac{1}{2} \frac{e^{\frac{M_\phi}{2T}} \left(\frac{M_\phi}{2T} \right)}{e^{\frac{M_\phi}{T}} - 1}. \tag{3.22}$$

Eq. (3.21) can be solved to yield T as a function of Φ. At high and low temperatures, the above relation simplifies to

$$T = \begin{cases} \left(\frac{\sqrt{3}}{16\sqrt{2}\pi\alpha} \frac{\mu^2}{M_\phi} M_{\rm Pl} \Phi \right)^{1/3} & T \gg M_\phi/4 \\ \left(\frac{\sqrt{3}}{80\pi\alpha} \mu^2 M_{\rm Pl} \Phi \right)^{1/4} & T \ll M_\phi/4. \end{cases} \tag{3.23}$$

Eq. (3.21) is valid as long as $\Gamma(T) < H$, or

$$\frac{\mu^2}{32 M_\phi \pi} \left(\frac{e^{\frac{M_\phi}{2T}} + 1}{e^{\frac{M_\phi}{2T}} - 1} \right) < \frac{M_\phi \Phi}{\sqrt{6} M_{\rm Pl}}. \tag{3.24}$$

In figure 3.2 we show the resulting parameter space for perturbative reheating for three different values of M_ϕ. We show the equalities corresponding to broad resonance preheating (eq. (3.20)) and the end of perturbative reheating (eq. (3.24)) in red and blue respectively; the yellow shaded region represents the region where perturbative reheating dominates the evolution of the radiation bath. We further show contours of T, from eq. (3.21). Above $T = M_\phi/4$, the matter sector realizes the $T \propto a^{-1/2}$ power law. We thus observe that for all three mass points, there is some region of parameter space where reheating is dominated by perturbative processes and the radiation bath realizes the bosonic power law. Lower inflaton masses enable the radiation bath to reach higher temperatures during perturbative reheating.

For a given value of μ, the theory may avoid preheating if the inflaton amplitude at the end of inflation is below the red line in figure 3.2. As perturbative reheating occurs, the inflaton amplitude decreases due to redshifting in an expanding universe. This redshifting corresponds to traversing downward in the $\mu - \Phi$ parameter space. The temperature of the radiation bath decreases correspondingly along this trajectory. This downward trajectory continues until we reach the blue line and reheating occurs.

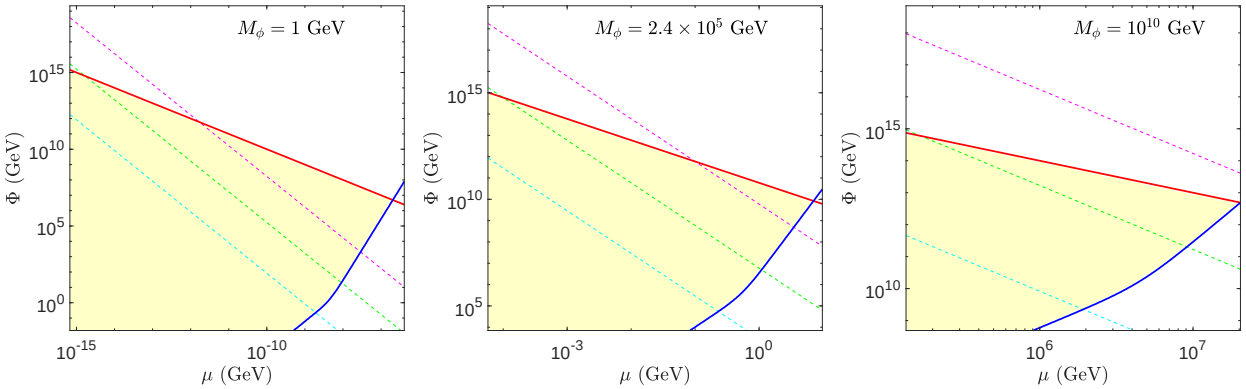


Figure 3.2: Reheating parameter space for $M_\phi = 1$ GeV (left), $M_\phi = 2.4 \times 10^5$ GeV (center) and $M_\phi = 10^{10}$ GeV (right). If the inflaton amplitude at the end on inflation is in the region above the red line, then the universe undergoes by broad resonance preheating, eq. (3.20). If the inflaton amplitude at the end of inflation is in the yellow shaded region, then the effects of preheating are sub-dominant and the Universe undergoes perturbative reheating. Perturbative reheating ends on the blue line, eq. (3.24). The dotted lines represent constant temperature contours during perturbative reheating calculated using eq. (3.21), $T = 10M_\phi$ (magenta), $T = M_\phi$ (green) and $T = 0.1M_\phi$ (cyan).

3.3 Two-sector reheating with inflaton-mediated interactions

In section 3.2, we demonstrated how quantum statistics alter the temperature evolution of the radiation sector prior to reheating. We now incorporate inflaton decays into two sectors and study the effect of inflaton-mediated scattering between the two sectors on the final temperature asymmetry. As we demonstrate in this section, inflaton-mediated energy transfer between sectors also yields an attractor solution for the temperature of the colder radiation bath, which allows us to make analytic predictions for the final temperature asymmetries in the regime where inflaton-mediated scattering is important.

We begin by establishing our notation. Introducing the scattering terms in the Boltzmann equations and ignoring any inflaton quanta, the Boltzmann equations in this limit read [42],

$$\frac{d\rho_\phi}{dt} + 3H\rho_\phi = -(\Gamma_a + \Gamma_b)\rho_\phi \tag{3.25}$$

$$\frac{\mathrm{d}\rho_a}{\mathrm{d}t} + 4H\rho_a = \Gamma_a\rho_\phi - \mathcal{C}_E \tag{3.26}$$

$$\frac{\mathrm{d}\rho_b}{\mathrm{d}t} + 4H\rho_b = \Gamma_b\rho_\phi + \mathcal{C}_E, \tag{3.27}$$

where $\Gamma_{a,b}$ are the (temperature-dependent) decay rates of ϕ to the respective sectors, H is the Hubble rate, which is given by the Friedmann equation

$$H = \frac{1}{\sqrt{3}M_{\rm Pl}}\sqrt{\rho_a + \rho_b + \rho_\phi}, \tag{3.28}$$

and $\mathcal{C}_E$ is the collision term describing the energy transfer from the hotter radiation sector to the colder

radiation sector via two-to-two scattering processes of the form $1+2 \to 3+4$,

$$\begin{aligned}\mathcal{C}_E = &\int \frac{d^3p_1}{2E_1(2\pi)^3}\frac{d^3p_2}{2E_2(2\pi)^3}\frac{d^3p_3}{2E_3(2\pi)^3}\frac{d^3p_4}{2E_4(2\pi)^3}(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)|\overline{\mathcal{M}}|^2 S(E_1+E_2)\\ &\times[f_1(p_1)f_2(p_2)(1\pm f_3(p_3))(1\pm f_4(p_4)) - f_3(p_3)f_4(p_4)(1\pm f_1(p_1))(1\pm f_2(p_2))]\\ \equiv &\mathcal{C}_E^f - \mathcal{C}_E^b. \end{aligned} \tag{3.29}$$

Here $\mathcal{C}_E^f$ and $\mathcal{C}_E^b$ are the collision terms for forward and backward reactions respectively, $|\overline{\mathcal{M}}|^2$ is the spin-summed scattering amplitude as determined by the particular inflaton-radiation interaction, and S is a symmetry factor accounting for possible identical particles in the initial and/or final state. We retain the full dependence on quantum statistics to accurately describe the energy transfer between two relativistic radiation baths [42], which makes the evaluation of the collision term more challenging. In appendix A, we show how the collision term eq. (3.29) can be simplified for two relativistic species at different temperatures.[3]

We use the subscript 'a' to denote the sector that attains the larger temperature at the end of reheating. Generically, this corresponds to the sector with the largest zero-temperature decay width.[4] 'Reheat temperature' in this context refers to the temperature of the hotter sector when the universe transitions from matter to radiation domination. We define the transition from matter domination to radiation domination at the point where energy density in the radiation becomes equal to the energy density in the inflaton, $\rho_a(a_{\rm rh}) + \rho_b(a_{\rm rh}) = \rho_\phi(a_{\rm rh})$. where

For the numerical results in the rest of the chapter we adopt a common reference set of numerical values for the inflaton mass and initial energy density as well as the number of degrees of freedom in each radiation bath,

$$\begin{gathered}\alpha_a = \frac{\pi^2 g_{*,a}}{30} = \alpha_b = \frac{\pi^2 g_{*,b}}{30} = 30, \quad M_\phi = 10^{-13} M_{\rm Pl} = 2.4\times 10^5 \text{ GeV}, \\ \rho_{\phi,I} = \frac{1}{2}M_\phi^2\langle\phi_I^2\rangle = M_{\rm Pl}^2 M_\phi^2.\end{gathered} \tag{3.30}$$

We assume for simplicity that α_a and α_b are constant over the range of temperatures we consider. While in what follows we have fixed the value of M_ϕ, our results are broadly independent of its precise value. As we demonstrate below, our results for the final temperature asymmetry depend on M_ϕ only through $T_{\rm rh}$ and the ratio $T_{\rm rh}/M_\phi$. The specific value of M_ϕ is generally only important insofar as smaller values of M_ϕ make it easier to obtain larger $T_{\rm rh}/M_\phi$.

We next demonstrate that the inflaton-mediated energy transfer yields a cosmological attractor solution for the colder radiation bath, using the model where the inflaton has trilinear couplings to scalar fields in both sectors as an illustrative example. We then analyze in detail how the interplay between this scattering attractor solution and the reheating attractor curve of the previous section determines the final temperature asymmetry. We then extend this analysis to other forms of the inflaton couplings to matter. In particular, we consider theories where the inflaton has: Yukawa couplings to fermions in both sectors; axion-like couplings to gauge bosons in both sectors; and a mixed scenario with a trilinear coupling to scalars in one sector and Yukawa coupling to fermions in the other. For the collision term, $\mathcal{C}_E$, we use the analytic approximations

[3]Note that there can also be energy exchange via t-channel scattering process mediated by inflaton. However the energy transfer via t-channel scattering process is orders of magnitude smaller compared to s-channel process because only the latter process is resonantly enhanced when inflaton is on-shell.

[4]This is not the case when the inflaton couples to bosons in one sector and fermions in the other and the resulting reheat temperature is large, $T_{\rm rh} \gtrsim M_\phi/4$. For this case, the effective decay width into bosons to be larger than the effective decay width into fermions even if the zero temperature decay width into bosons is smaller.

derived in appendix A.4. For simplicity we continue to assume that the inflaton only couples to a single species in each sector.

3.3.1 Scalar trilinear couplings

We begin by considering a theory where the inflaton is coupled to scalar fields in both sectors, χ_a and χ_b, via trilinear couplings

$$\mathcal{L}_{\rm int} = \frac{1}{2}\mu_a \phi \chi_a^2 + \frac{1}{2}\mu_b \phi \chi_b^2. \tag{3.31}$$

This interaction results in zero-temperature decay widths given by

$$\Gamma_{0a,b} = \frac{1}{32\pi}\frac{\mu_{a,b}^2}{M_\phi}\sqrt{1-\frac{4m_{a,b}^2}{M_\phi^2}} \approx \frac{\mu_{a,b}^2}{32\pi M_\phi}, \tag{3.32}$$

where $m_{a,b}$ denotes the mass of the fields $\chi_{a,b}$, which we have assumed to be much smaller than the inflaton mass, $m_{a,b} \ll M_\phi$. Our convention is that sector a is the hotter sector, and accordingly we take $\mu_a \geq \mu_b$ in what follows.

The collision term

The s-channel amplitude for $\chi_a\chi_a \leftrightarrow \chi_b\chi_b$ scattering mediated by inflaton exchange is given by

$$|\mathcal{M}(s)|^2 = \frac{\mu_a^2\mu_b^2}{(s-M_\phi^2)^2 + (\Gamma_{0a}+\Gamma_{0b})^2}. \tag{3.33}$$

In appendix A.4.1, we compute the collision term, $\mathcal{C}_E$, following from this amplitude.

The collision term in general is a function of both T_a and T_b. However, for large asymmetries $T_b \ll T_a$, the forward energy transfer term governing energy injection into the colder sector dwarfs the backward energy transfer term. Moreover, in this regime we can also ignore the final state Bose enhancement of C_E^f: while χ_b particles produced in the forward reaction typically have energies of order $\sim T_a$, for $T_b \ll T_a$ those energy levels are mostly unpopulated. In this simplified regime, the collision term thus depends only on T_a as

$$\mathcal{C}_E = \frac{1}{16\pi^3} \times \begin{cases} \dfrac{\mu_a^2\mu_b^2}{\mu_a^2+\mu_b^2} T_a^3 \left[1.6 \log\left(\frac{T_a}{M_\phi}\right) + 1.3\right] & T_a \gtrsim M_\phi \\ \dfrac{\mu_a^2\mu_b^2}{\mu_a^2+\mu_b^2} M_\phi^2 \dfrac{T_a}{4} K_2\left(\frac{M_\phi}{T_a}\right) & T_a \lesssim M_\phi \\ \dfrac{7.9}{32\pi^2}\dfrac{\mu_a^2\mu_b^2}{M_\phi^4} T_a^5 & m_{a,b} \ll T_a \ll M_\phi, \end{cases} \tag{3.34}$$

as derived in appendix A.4.1; see figure A.1. For $T_a \gtrsim M_\phi$, the collision term is substantially enhanced by the resonant exchange of inflaton particles. The divergence in the Bose-Einstein distributions at $E \to 0$ combined with the resonant peak in the scattering amplitude results in a logarithmic dependence on T_a/M_ϕ for $T_a \gg M_\phi$. As T_a drops below the inflaton mass, the scattering goes off resonance and $\mathcal{C}_E$ drops rapidly. Because the scattering is dominated by the energetic tail of the phase space distribution, this fall-off of

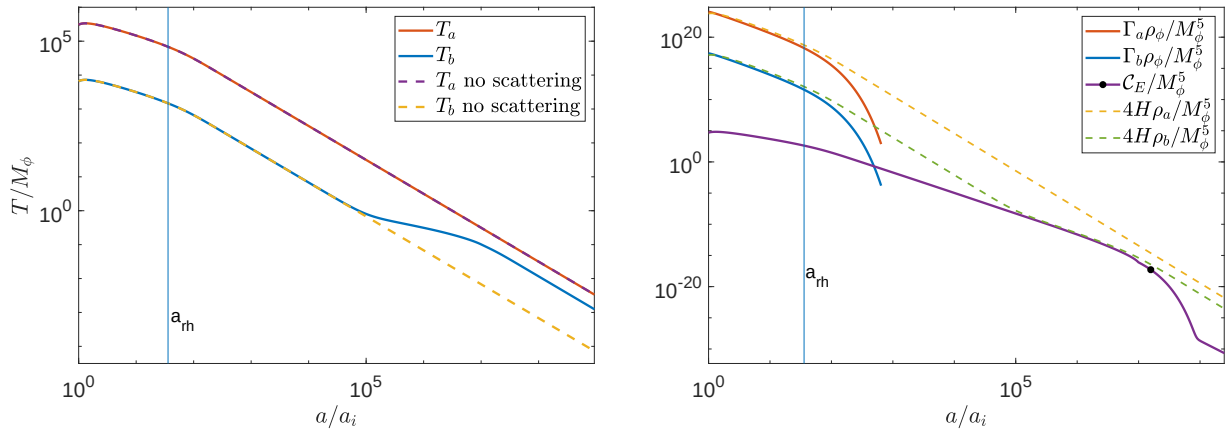


Figure 3.3: *Left panel:* Example temperature evolution of radiation sectors during and after reheating via scalar trilinear couplings. The solid (dashed) lines denote the temperatures obtained from numerically solving the Boltzmann equations including (neglecting) the inflaton-mediated scatterings between the sectors. Reheating is denoted by the vertical blue line at $a_{\rm rh}$. The figure has been plotted for $\mu_a = 10^{-3}M_\phi$, $\mu_b = 10^{-5.5}M_\phi$, $m_{a,b} = 10^{-3}M_\phi$ with other parameters as specified in eq. (3.30). *Right panel:* Comparison of the collision term to the redshifting of both sectors as well as inflaton decays for the same parameters as that in left panel. The black dot on the curve for the collision term indicates $T_a = M_\phi/4$, which approximately indicates the temperature below which the s-channel energy transfer rate becomes exponentially suppressed.

the energy transfer rate can be accurately described assuming Maxwell-Boltzmann statistics, thus yielding a Bessel function $K_2(M_\phi/T_a) \sim (M_\phi/T_a)^{3/2}\exp(-M_\phi/T_a)$. Note that, in the resonant regime, the energy transfer rate depends more strongly on the smaller coupling μ_b than the larger coupling μ_a, and in particular, when $\mu_b \ll \mu_a$, the rate is almost independent of μ_a. Below the resonance, the energy transfer rate drops rapidly until it reaches the low-temperature regime $T_a \ll M_\phi$. In this regime, the inflaton can be integrated out of the theory, leaving a constant scattering amplitude, $|\mathcal{M}(s)|^2 \approx \mu_a^2\mu_b^2/M_\phi^4$. Thus we obtain the $\mathcal{C}_E \propto T_a^5$ behavior in the last line of eq. (3.34). Finally, at temperatures low enough that one or both of the scattering species becomes non-relativistic, the energy transfer rate becomes Boltzmann-suppressed; we do not include this effect, as we find that generically the behavior of $\mathcal{C}_E$ below $T_a < M_\phi/4$ is inconsequential to determining the final temperature asymmetry.

Finally, we stress that the expression for $\mathcal{C}_E$ given in eq. (3.34) is a limiting version that neglects its dependence on T_b. Dependence on T_b can enter in two ways: first, via the backward energy transfer term, and second, from Bose enhancement of $\mathcal{C}_E^f$. The backward energy transfer term becomes important when $T_b \gtrsim 0.9T_a$, and as the two sectors approach equilibration the net energy transfer rate rapidly drops. The Bose enhancement of the forward energy transfer term is more involved to model. This Bose enhancement largely serves to increase $\mathcal{C}_E^f$ in the high and low temperature regimes in eq. (3.34) with increasing T_b. The middle regime in eq. (3.34), however, is insensitive to the possible Bose enhancement terms, as that regime is effectively described by Maxwell-Boltzmann statistics. As we show below, this last property enables us to obtain analytic predictions of the final temperature asymmetry without needing to keep track of the full behavior of the Bose enhancements.

The scattering attractor solution

Now we discuss the impact of the collision term on the temperature evolution of both sectors, and derive the corresponding scattering attractor curve for the temperature of the colder sector. We begin by considering

scenarios where $T_{\rm rh} \gg M_\phi$, where, as we demonstrate, scattering becomes important post-reheating. One such parameter point is shown in figure 3.3, which plots numerical solutions for the radiation temperatures obtained using the collision term as given in eq. (A.78). To highlight the importance of scattering, we also show the temperature evolution when the scattering has been turned off. In the right panel, we show the evolution of the collision term, $\mathcal{C}_E$, in comparison to the combinations $4H\rho_{a,b}$ and $\Gamma_{a,b}\rho_\phi$. We mark the point $T_a = M_\phi/4$ around which $\mathcal{C}_E$ begins to exhibit substantial Boltzmann suppression. As the scattering process affects the temperature evolution substantially post-reheating in this example, we can cleanly separate the effects of scattering from the contributions of reheating; in this discussion, reheating itself is only important insofar as it provides initial conditions for the subsequent post-reheating evolution of T_a and T_b.

As figure 3.3 shows, T_b begins to deviate from the no-scattering solution as soon as the fractional energy transfer rate into the colder sector becomes comparable to the Hubble rate, $\Gamma_{E,b} = \mathcal{C}_E/\rho_b \sim H$. When this happens we say that inflaton-mediated scattering becomes *effective*. In contrast, when the fractional energy transfer rate out of the hotter sector becomes comparable to the Hubble rate, $\Gamma_{E,a} = \mathcal{C}_E/\rho_a \sim H$, inflaton-mediated scattering becomes *efficient* and the two sectors attain thermal equilibrium. In the scenario shown in figure 3.3, inflaton-mediated scattering becomes effective but never efficient. The solution to the Boltzmann equation for T_b when scattering becomes effective is approximated by the quasi-static attractor solution (see section 2.5)

$$\rho_b(T_b) = \frac{1}{4 + \frac{q}{1-p}} \frac{\mathcal{C}_E(T_a, T_b)}{H(T_a)}, \tag{3.35}$$

where

$$p(a) = \frac{\partial \ln(\mathcal{C}_E/H)}{\partial \ln \rho_b}, \qquad \text{and} \qquad q(a) = \frac{\partial \ln(\mathcal{C}_E/H)}{\partial \ln a}. \tag{3.36}$$

We call this evolution of ρ_b the *scattering attractor curve*. In evaluating $q(a)$, the scale factor dependence in $\mathcal{C}_E/H$ comes through T_a, which in the present scenario is evolving adiabatically. For a given value of T_a, there is a single corresponding value of T_b that satisfies eq. (3.35). In general, solving eq. (3.35) for T_b is non-trivial given the dependence of $\mathcal{C}_E$ on T_b through Bose enhancement. The attractor curve exists as long as $4 + q/(1-p) > 0$, which translates to the condition that $\mathcal{C}_E$ falls off more slowly with scale factor than Ha^{-4}. At temperatures below $T \sim M_\phi/4$ the collision term falls off exponentially (eq. (3.34)), marking the end of the attractor evolution. Beyond that point, ρ_b evolves adiabatically as seen in figure 3.3. Thus the scattering attractor curve yields a final temperature asymmetry simply given by the asymmetry at $T_a \approx M_\phi/4$.

To further highlight the attractor nature of the collision term, figure 3.4 shows the post-reheating evolution of the temperature ratio, $x = (\alpha_a \rho_b/(\alpha_b \rho_a))^{1/4}$, for the parameter point of figure 3.3, but now considering a range of (post-reheating) initial conditions for ρ_b (or equivalently x). In the left panel, the solid blue line tracks the evolution of T_b/T_a following from figure 3.3, where the initial conditions are determined self-consistently from inflaton decays, $x_i = x_{\rm rh} = 0.02$. The purple dot-dashed line shows the evolution when the initial temperature ratio is instead zero, $x_{1,i} = 0$; again, initial densities below the attractor solution rise rapidly to attain the attractor. The yellow dot-dashed line shows the evolution with an initial temperature ratio $x_{2,i} = 0.1 > x_{\rm rh}$; this solution still attains the scattering attractor curve, eq. (3.35). The green dot-dashed line denotes evolution with an initial temperature ratio, $x_{3,i} = 0.5$, much above the final temperature ratio determined by the scattering attractor solution. In this case T_b remains mostly unaffected by inflaton-mediated interactions. Thus we see that the inflaton-mediated interactions impose a minimum

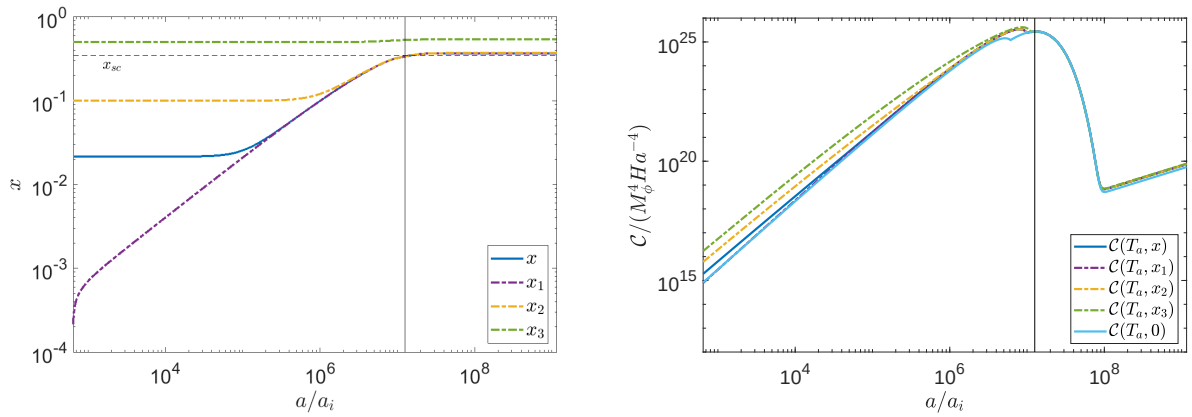


Figure 3.4: *Left panel:* Evolution of post-reheating temperature ratios for several different initial conditions. The solid blue line shows the evolution $x = T_b/T_a$ from figure 3.3. The dot-dashed lines (x_1, x_2 and x_3) indicate the temperature ratio evolution after artificially varying the initial condition for T_b ($x_{1,i} = 0$, $x_{2,i} = 0.1$ and $x_{3,i} = 0.5$). The horizontal black dashed line corresponds to the analytic estimation of the final temperature ratio derived in eq. (3.41). *Right panel:* The collision terms (normalized by $M_\phi^4 H a^{-4}$) experienced by the colder sectors plotted on the left panel. The solid light blue line represents the evolution of the collision term when the T_b-dependence of $\mathcal{C}_E$ is neglected. The vertical black line in both panels indicates the scale where $T_a = M_\phi/4$.

value for the final temperature ratio: any initial temperature ratio below this minimum value is increased to that value by the scattering attractor solution while initial temperature ratios above this minimum remain largely unaffected. This minimum final temperature ratio is simply determined by the behavior of $\mathcal{C}_E$ near $T_a \sim M_\phi/4$, as we elaborate below.

The right panel of figure 3.4 shows the ratio $\mathcal{C}_E(T_a, x)/(Ha^{-4})$ for all scenarios. The different behavior of the curves at early times shows the differing importance of Bose enhancement on the final state in the different scenarios. All curves converge onto a single common solution for $T_a \sim M_\phi$, when scattering is well described by Maxwell-Boltzmann statistics. As $H\rho_a$ evolves with scale factor as Ha^{-4}, the value of a where $\mathcal{C}_E/Ha^{-4}$ is maximized[5] coincides with the value of a where the scattering attractor solution for ρ_b ends.

To accurately determine the final temperature asymmetry predicted by the scattering attractor-curve, we need to integrate the Boltzmann equations around the region where $T_a \lesssim M_\phi$. Assuming no T_b dependence in $\mathcal{C}_E$ or H, the Boltzmann equation for ρ_b can be directly integrated,

$$\rho_b a^4 - (\rho_b a^4)_{a=a_1} = \int_{a_1}^{a} \frac{a'^3 \mathcal{C}_E(T_a)}{H} da' = a_1^4 \int_1^{a/a_1} \frac{z^5 \sqrt{3} M_{\rm Pl} \mathcal{C}_E(M_\phi/z)}{\alpha_a^{1/2} M_\phi^2} dz, \tag{3.37}$$

where we have defined a_1 as the scale factor at which $T_a(a_1) = M_\phi$, and $z = a'/a_1$. We evaluate $\mathcal{C}_E$ by taking

[5]The visible hiccup near the peak in $\mathcal{C}_E(T_a, 0)$ is a feature of the imperfect fitting used in our analytic approximation of $\mathcal{C}_E$, eq. (A.78).

the limit $x \to 0$ around $T_a \sim M_\phi$ in eq. (A.78) to obtain

$$\begin{aligned}\rho_b a^4 - (\rho_b a^4)_{a=a_1} &= a_1^4 \left(\frac{2}{\pi^2} M_\phi^2 \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}} \frac{\sqrt{3}M_{\rm Pl}}{\alpha_a^{1/2}} \right) \\ &\quad \times \int_1^{a/a_1} \max\left\{ z^2 \Big(1.3 - 1.6\log(z)\Big), \frac{1}{4} z^4 K_2(z) \right\} dz \\ &\xrightarrow{a/a_1 \text{large}} 6.3 a_1^4 \left(\frac{2}{\pi^2} M_\phi^2 \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}} \frac{\sqrt{3}M_{\rm Pl}}{\alpha_a^{1/2}} \right). \end{aligned} \tag{3.38}$$

We determine the initial energy density of the colder sector, $(\rho_b a^4)_{a=a_1}$, by assuming that the colder sector is already on the scattering attractor curve defined by evaluating eq. (3.35) using $\mathcal{C}_E = \mathcal{C}_E(T_a, 0)$. However, the Maxwell-Boltzmann behavior of the collision term in this temperature range helps to ensure that the final result is insensitive to the specific choice of $x = 0$. We find

$$\rho_b(a_1) = 0.71 \left(\frac{2}{\pi^2} M_\phi^2 \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}} \frac{\sqrt{3}M_{\rm Pl}}{\alpha_a^{1/2}} \right). \tag{3.39}$$

The final energy density of the colder sector is then given by

$$\rho_b(a) = 7.0 \left(\frac{a_1}{a} \right)^4 \left(\frac{2}{\pi^2} M_\phi^2 \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}} \frac{\sqrt{3}M_{\rm Pl}}{\alpha_a^{1/2}} \right). \tag{3.40}$$

The final temperature ratio between the two sectors predicted by inflaton-mediated scattering is then

$$x_{sc} \equiv \left(\frac{\alpha_a \rho_b}{\alpha_b \rho_a} \right)^{1/4}_{a>a_1} = 1.25 \left(\frac{1}{M_\phi^2} \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}} \frac{M_{\rm Pl}}{\sqrt{\alpha_a}\alpha_b} \right)^{1/4}, \tag{3.41}$$

where we have used $\rho_a(a) = \alpha_a M_\phi^4 (a_1/a)^4$. This is the value that the temperature ratios x, x_1 and x_2 asymptote to as shown by the horizontal black dashed line in figure 3.4. Eq. (3.41) holds as long as $x_{sc} \lesssim 0.9$. Once the temperature ratio approaches unity, backward energy transfer and the contribution of ρ_b to the Hubble parameter become important, and the attractor solution no longer captures the full behavior of the system. In these cases, where the two sectors approach thermalization, a more detailed numerical study is required.

Finally, it is worth emphasizing that the scattering attractor curve discussed here is dominated by the resonant behavior of the energy transfer rate, and depends on the properties of the radiation baths at $T_a \sim M_\phi$. In the trilinear scalar model, a second attractor phase appears at temperatures well below the resonance ($T_a \ll M_\phi$). This is evident from the late-time increase in $\mathcal{C}_E/(Ha^{-4})$ in the right panel of figure 3.4, after the resonant enhancement ends. This possibility of IR thermalization is a special feature of the trilinear scalar model, where integrating out the inflaton introduces a renormalizeable quartic interaction between the two sectors. In all other cases $\mathcal{C}_E$ falls off much faster at lower temperatures due to the higher (≥ 4) mass dimensions of the operators that couple the inflaton to the radiation baths. Once $T_{a,b} \sim m_{a,b}$, $\mathcal{C}_E$ becomes exponentially suppressed and scattering is cut off. Thus, thermalization in the IR depends on the mass scales in the matter sectors coupled to the inflaton, as well as the inflaton mass and $T_{\rm rh}$. Late-time equilibration through scalar portal interactions is studied in detail in [42, 65, 66] and we do not discuss it further here.

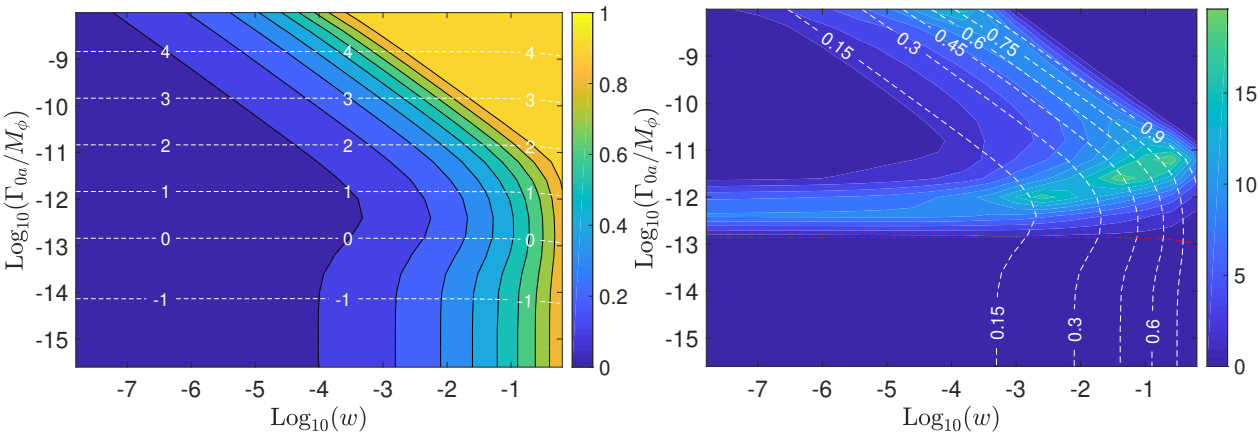


Figure 3.5: *Left panel*: Contours of the final temperature ratio obtained numerically, $x_{f,n} = T_b/T_a$, shown in color, for the case when the inflaton is coupled to scalars in both sectors. The white dashed lines show contours of $\log_{10}(T_{\rm rh}/M_\phi)$. *Right panel*: Contours of relative fractional discrepancy, $|x_{f,a} - x_{f,n}|/x_{f,n}$, where $x_{f,a}$ is the analytic estimate of eq. (3.42) with eq. (3.41), in percent. The white dashed contours depict $x_{f,n}$ and the red dashed line marks the region where $T_{\rm rh} = M_\phi$.

Final temperature ratios

Both reheating and scattering, considered independently, produce attractor solutions. In most of parameter space, one attractor solution dominates over the other and thus is primarily responsible for determining the final temperature asymmetry. As demonstrated above in section 3.3.1, when $T_{\rm rh} > M_\phi$, a good semi-analytic approximation to the final temperature asymmetry is therefore

$$x_f = \max[x_{\rm rh}, x_{sc}, 1], \tag{3.42}$$

where $x_{\rm rh} = (\alpha_a \Gamma_{0b}/\alpha_b \Gamma_{0a})^{1/4}$, is the temperature asymmetry obtained if we neglect inflaton mediated scattering. Both $x_{\rm rh}$ and x_{sc} can be straightforwardly computed from the Lagrangian parameters without any need to solve the full Boltzmann equations.

In the case $T_{\rm rh} \ll M_\phi/4$, the reheating attractor solution dominates, as we now show. As $\mathcal{C}_E$ redshifts more slowly than $\Gamma_b \rho_\phi$ during reheating, it suffices to show that $\mathcal{C}_E/\Gamma_b\rho_\phi < 1$ at $T_a \approx M_\phi/4$ when $\mathcal{C}_E$ is maximized. Using eq. (3.4) for ρ_ϕ at $T_a = M_\phi/4$ and eq. (A.75) for $\mathcal{C}_E(T_a = M_\phi/4)$ (at $x = 0$) we find

$$\left(\frac{\mathcal{C}_E}{\Gamma_{0b}\rho_\phi}\right)_{T_a = M_\phi/4} \approx \left(\frac{25\pi^2 \alpha_a}{3 \times 2^{15} K_2(4)}\right)^{1/4} \frac{T_{\rm rh}}{M_\phi} \approx \left(\frac{\alpha_a}{10}\right)^{1/4} \frac{T_{\rm rh}}{M_\phi}. \tag{3.43}$$

This ratio is small by assumption for SM-scale values of α_a. Hence for low reheat temperatures inflaton-mediated scattering is unimportant during reheating. After reheating, scattering cannot thermalize the two sectors as the resonant enhancement has already ended. Thus, for $T_{\rm rh} \ll M_\phi$ the final temperature asymmetry is simply given by $x_{\rm rh} = (\alpha_a \Gamma_{0b}/\alpha_b \Gamma_{0a})^{1/4}$. Although x_{sc}, as defined in eq. (3.41) as the temperature ratio obtained by the *post-reheating* scattering attractor curve, does not strictly pertain in this case, one can check that its value is always less than $x_{\rm rh}$ when $T_{\rm rh} \ll M_\phi/4$. Thereby we can extend eq. (3.42) to hold for $T_{\rm rh} \ll M_\phi/4$ as well.

We are now ready to consider the full numerical solution to the Boltzmann equations, eq. (3.25). The resulting numerical temperature asymmetry, $x_{f,n}$, is shown in the left panel of figure 3.5 as a function of the

ratio of the zero-temperature widths $w = \Gamma_{0,b}/\Gamma_{0,a}$.

As expected, for low reheat temperatures the reheating attractor curve dominates the evolution. The contours below $T_{\rm rh} \lesssim M_\phi$ show the same behavior as in the absence of inflaton-mediated scattering. Inflaton-mediated scattering becomes important roughly for $T_{\rm rh} \gtrsim M_\phi$, and dominates for $T_{\rm rh} \gtrsim 10 M_\phi$. In this high-temperature regime the contours are almost diagonal, reflecting the fact that x_{sc} is dominantly governed by the smaller decay width $\Gamma_{0,b}$. In the right panel of figure 3.5 we compare our analytic estimate of eq. (3.42) with the result obtained from numerically solving the Boltzmann equations. The analytic estimate agrees with the numerical results within 20%. The discrepancies are greatest exactly where the scattering and reheating attractor curves are no longer individually sufficient to capture the full behavior of the system: when both scattering and inflaton decays are important for determining the final asymmetry, around $T_{\rm rh} \sim$ few M_ϕ, and when the sectors are approaching (but not obtaining) thermalization, $x_f \sim 0.7 - 0.8$.

Finally, let us note the important point that our numerical results for $x_{f,n}$ are themselves based on analytic approximations to the collision term. Our analytic fits to the collision term deviate from the exact numerical values by almost 50% near $T \sim M_\phi$ (see figure A.1). As the final temperature ratio is predominantly determined by the behavior of the collision term near $T \sim M_\phi$, this error is unfortunately not negligible for our final results. However, this error is made less numerically consequential once we take the fourth root to find the temperature (eq. (3.41)), inducing uncertainties of up to $\sim 15\%$ in the numerical temperature ratio plots at high $T_{\rm rh}$, figure 3.5.

3.3.2 Final temperature asymmetry for other theories

The two key properties of $\mathcal{C}_E$—the exponential suppression at $T_a \lesssim M_\phi/4$ and the weak dependence on T_b in this range—that allowed us to analytically determine the final temperature asymmetry for the scalar trilinear case are generic features of resonant s-channel interactions. Much of our analysis in the previous section can thus be applied directly to other interaction structures. As we demonstrate, in models where the inflaton has renormalizeable couplings to matter, scattering is only important for determining the final temperature asymmetry when the endpoint of the scattering attractor curve occurs post-reheating. However, scattering during reheating can also be important when the inflaton is a pseudoscalar with dimension-five couplings to gauge fields in both sectors.

Yukawa couplings

We begin with a model where the inflaton has Yukawa couplings to fermions in both sectors,

$$\mathcal{L}_{\rm int} = y_a \phi \bar{\psi}_a \psi_a + y_b \phi \bar{\psi}_b \psi_b. \tag{3.44}$$

This interaction results in zero-temperature inflaton decay widths given by

$$\Gamma_{0a,b} = \frac{y_{a,b}^2}{8\pi} M_\phi \sqrt{1 - \frac{4m_{a,b}^2}{M_\phi^2}} \approx \frac{y_{a,b}^2}{8\pi} M_\phi, \tag{3.45}$$

where $m_{a,b} \ll M_\phi$ denotes the mass of fields $\psi_{a,b}$. The s-channel spin-summed scattering amplitude between the two species is

$$|\overline{\mathcal{M}}(s)|^2 = 4 y_a^2 y_b^2 \left(1 - \frac{4m_a^2}{s}\right)\left(1 - \frac{4m_b^2}{s}\right) \frac{s^2}{(s - M_\phi^2)^2 + (\Gamma_{0a} + \Gamma_{0b})^2}. \tag{3.46}$$

The total energy transfer collision term, $\mathcal{C}_E$, following from this amplitude is discussed in appendix A.4.2 and shown in figure A.2. Unlike the scalar case discussed in section 3.3.1, the collision term is almost insensitive to the temperature of the colder sector unless the two temperatures are very close and $\mathcal{C}_E^b$ becomes important. In the limit that the temperature ratio between the two sectors is very small, $x \ll 1$, $\mathcal{C}_E$ is approximately given by

$$\mathcal{C}_E = \frac{1}{4\pi^3} \times \begin{cases} \dfrac{3.0}{2\pi^2} y_a^2 y_b^2 T_a^5 & T_a \gg M_\phi \\ 0.29 \dfrac{y_a^2 y_b^2}{y_a^2 + y_b^2} M_\phi^2 T_a^3 & T_a \gtrsim M_\phi \\ \dfrac{y_a^2 y_b^2}{y_a^2 + y_b^2} M_\phi^4 \dfrac{T_a}{4} K_2\left(\frac{M_\phi}{T_a}\right) & T_a \lesssim M_\phi \\ \dfrac{1.4 \times 10^3}{2\pi^2} y_a^2 y_b^2 \dfrac{T_a^9}{M_\phi^4} & m_{a,b} \ll T_a \ll M_\phi. \end{cases} \tag{3.47}$$

At temperatures much larger than the inflaton mass, the inflaton mass can be neglected and the scattering amplitude is approximately constant, $|\overline{\mathcal{M}}(s)|^2 \approx y_a^2 y_b^2$, yielding the $\mathcal{C}_E \propto T^5$ behavior required from dimensional analysis. At temperatures closer to the inflaton mass, the energy transfer rate is resonantly enhanced, yielding $\mathcal{C}_E \propto T^3$ behavior. As the temperature drops below the inflaton mass, the energy transfer rate is dominated by resonant scattering in the Boltzmann-suppressed tails. Analogously to the scalar case, $\mathcal{C}_E$ can be well modeled in this region using Maxwell-Boltzmann statistics. In the low temperature regime the scattering amplitude can be approximated as $|\overline{\mathcal{M}}(s)|^2 \approx y_a^2 y_b^2 s^2 / M_\phi^4$, yielding the steep $\mathcal{C}_E \propto T^9$ behavior. Note that, like the scalar trilinear case, the energy transfer rate depends most strongly on the smaller coupling in the resonant regime.

We can again obtain an analytic expression for the final temperature asymmetry due to inflaton-mediated scattering, as we did for scalars in section 3.3.1. Using $\mathcal{C}_E$ from eq. (A.89) and taking $x \to 0$, we obtain

$$x_{sc} = 1.19 \left(\frac{1}{M_\phi^2} \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a} + \Gamma_{0,b}} \frac{M_{\rm Pl}}{\alpha_b \sqrt{\alpha_a}} \right)^{1/4}. \tag{3.48}$$

The final temperature asymmetry can then be estimated using eq. (3.42), i.e., by comparing the lower bounds from the scattering and reheating attractor solutions. In figure 3.6 we show numerical final temperature ratios in the left panel and in the right panel compare our analytic estimate to the numerical results. We again observe a transitional region around $T_{\rm rh} \sim$ few $\times M_\phi$ where both reheating and scattering are important for determining the final value of x_f. Note that the analytic estimate from the scattering attractor curve has better agreement with the numerical results in the region near thermalization, $x_f \to 1$, than we saw for the scalar case; this is because the Fermi blocking of C_E^f that occurs here is nowhere near as large an effect as the Bose enhancement we discussed in the previous subsection.

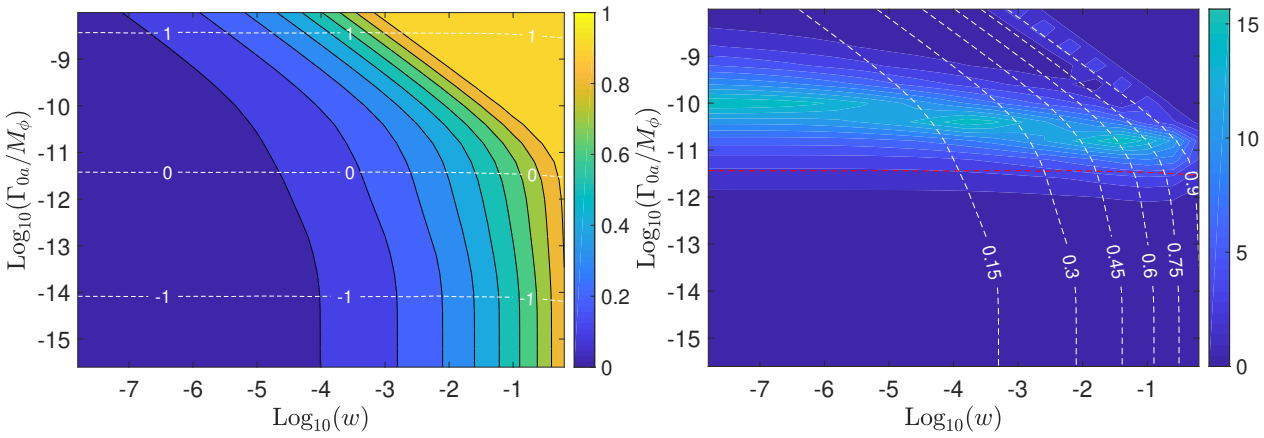


Figure 3.6: *Left panel*: Contours of the final temperature ratio obtained numerically, $x_{f,n} = T_b/T_a$, shown in color, for an inflaton with Yukawa couplings to fermions in both sectors. The white dashed lines show contours of $\log_{10}(T_{\rm rh}/M_\phi)$. *Right panel*: Contours of relative fractional discrepancy, $|x_{f,a} - x_{f,n}|/x_{f,n}$, where $x_{f,a}$ is the analytic estimate of eq. (3.42) with eq. (3.48), in percent. The white dashed contours depict $x_{f,n}$ and the red dashed line marks $T_{\rm rh} = M_\phi$.

Axionic couplings to gauge bosons

We next consider a theory where a pseudo-scalar inflaton couples to gauge bosons in both sectors,

$$\mathcal{L}_{\rm int} = -\frac{1}{4\Lambda_a}\phi F_a^{\mu\nu}\tilde{F}_{a,\mu\nu} - \frac{1}{4\Lambda_b}\phi F_b^{\mu\nu}\tilde{F}_{b,\mu\nu}. \tag{3.49}$$

This interaction results in zero-temperature decay widths given by

$$\Gamma_{0a,b} = \frac{1}{256\pi}\frac{M_\phi^3}{\Lambda_{a,b}^2}\sqrt{1-\frac{4m_{a,b}^2}{M_\phi^2}} \approx \frac{M_\phi^3}{256\pi\Lambda_{a,b}^2}, \tag{3.50}$$

where $m_{a,b} \ll M_\phi$ denotes the mass of the gauge fields, $A_{a,b}^\nu$. The s-channel spin-summed amplitude for $A_aA_a \leftrightarrow A_bA_b$ scattering mediated by inflaton exchange is

$$|\overline{\mathcal{M}}(s)|^2 = \frac{4}{128\Lambda_a^2\Lambda_b^2}\left(1-\frac{4m_a^2}{M_\phi^2}\right)\left(1-\frac{4m_b^2}{M_\phi^2}\right)\frac{s^4}{(s-M_\phi^2)^2+(\Gamma_{0a}+\Gamma_{0b})^2}. \tag{3.51}$$

In appendix A.4, we derive the total energy transfer rate, $\mathcal{C}_E$ for this amplitude; see figure A.3. When the temperature ratio between the two sectors is very small, $x \ll 1$, the temperature dependence of $\mathcal{C}_E$ is

approximately given by

$$\mathcal{C}_E = \frac{1}{64\pi^3} \times \begin{cases} \dfrac{14.0}{2\pi^2}\dfrac{1}{\Lambda_a^2\Lambda_b^2}T_a^9 & \Lambda_a > T_a \gg M_\phi \\ \dfrac{M_\phi^4}{\Lambda_a^2+\Lambda_b^2}T_a^3\left[1.6\log\left(\frac{T_a}{M_\phi}\right)+1.3\right] & T_a \gtrsim M_\phi \\ \dfrac{M_\phi^6}{\Lambda_a^2+\Lambda_b^2}\dfrac{T_a}{4}K_2\left(\frac{M_\phi}{T_a}\right) & T_a \lesssim M_\phi \\ \dfrac{7.1\times10^4}{2\pi^2}\dfrac{1}{\Lambda_a^2\Lambda_b^2}\dfrac{T_a^{13}}{M_\phi^4} & m_{a,b} \ll T_a \ll M_\phi. \end{cases} \tag{3.52}$$

The steep rise in the collision term ($\mathcal{C}_E \propto T_a^9$) at high temperatures is a consequence of the high mass-dimension of the operators mediating the interaction. This behavior will be modified when $T_a \gtrsim \Lambda_a$ and the effective field theory breaks down.

Repeating the calculation from section 3.3.1, using $\mathcal{C}_E$ from eq. (A.99) with $x \to 0$, we obtain the asymptotic temperature asymmetry resulting from the scattering attractor curve,

$$x_{sc} = 1.49\left(\frac{1}{M_\phi^2}\frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a}+\Gamma_{0,b}}\frac{M_{\rm Pl}}{\alpha_b\sqrt{\alpha_a}}\right)^{1/4}. \tag{3.53}$$

The final temperature asymmetry can then be estimated using eq. (3.42). In the left panel of figure 3.7 we show the final temperature ratio determined by numerically solving the Boltzmann equations. In this section, we take $\rho_{\phi,I} = 10^{-10}M_\phi^2 M_{\rm Pl}^2$ in order to keep $T_{\rm max} < \Lambda_a$ in all of our parameter space, thus ensuring that the effective field theory is valid throughout the entire evolution of the system. Due to the attractor nature of the Boltzmann equations describing reheating, larger values of $\rho_{\phi,I}$ do not change the final value of x that one would compute for a given set of Lagrangian parameters. However, changing $\rho_{\phi,I}$ does alter the maximum temperature attained (see eq. (3.13)), and therefore if we require $T_{\rm max} < \Lambda_a, \Lambda_b$ then we are restricted to parameters that satisfy

$$\frac{\Gamma_{0a}}{M_\phi} < \left(\frac{\alpha_a^2 M_\phi^6}{(256\pi)^3 M_{\rm Pl}^2\rho_{\phi,I}}\right)^{1/5}. \tag{3.54}$$

In the left panel of figure 3.7 the red dot-dashed lines indicate where $T_{\rm max} = \Lambda_a$ for different values of $\rho_{\phi,I}$. Above those lines $T_{\rm max} > \Lambda_a$, and thus the early evolution of the system probes the theory above the cutoff. In the right panel of figure 3.7 we compare our analytic estimate to the numerical result.

In the top left corner of the right panel of figure 3.7, large discrepancies between the analytic estimate and the numerical computation are becoming evident. In the same region in the left panel, the contours of fixed temperature asymmetry are beginning to extend more deeply into the region of small w than the previous examples. Both these features are the consequence of early (i.e. pre-reheating) thermalization, enabled by the UV-dominated energy transfer process ($\mathcal{C}_{E,UV} \propto T_a^9$) whose effects are not incorporated into the analytic estimate in eq. (3.42). At sufficiently high temperatures, T_a, $\mathcal{C}_{E,UV}$ dominates over the energy dumped from the inflaton. This UV behavior can be seen in figure 3.8, which shows the various contributions

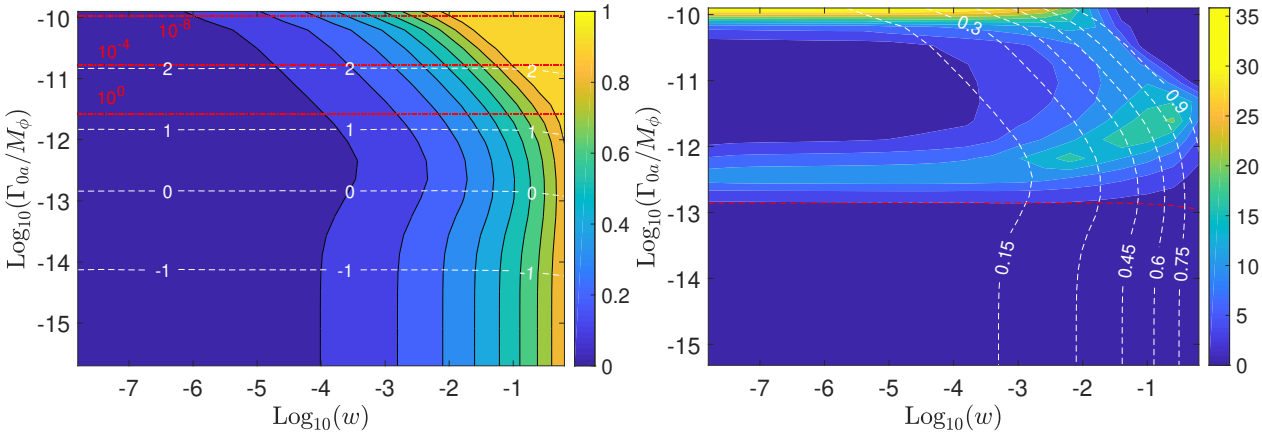


Figure 3.7: *Left panel*: Contours of the final temperature ratio obtained numerically, $x_{f,n} = T_b/T_a$, shown in color, for an inflaton with axionic couplings to gauge bosons in both sectors. The white dashed contours show $\log_{10}(T_{\rm rh}/M_\phi)$. The red dot-dashed lines mark the region where $T_{a,\max} = \Lambda_a$ for initial inflaton densities $\rho_{\phi,I} = 10^{-8} M_{\rm Pl}^2 M_\phi^2$, $10^{-4} M_{\rm Pl}^2 M_\phi^2$ and $10^{0} M_{\rm Pl}^2 M_\phi^2$. *Right panel*: Contours of relative fractional discrepancy, $|x_{f,a} - x_{f,n}|/x_{f,n}$, where $x_{f,a}$ is the analytic estimate of eq. (3.42) with eq. (3.53), in percent. The white dashed contours depict $x_{f,n}$ and the red dashed line indicates where $T_{\rm rh} = M_\phi$.

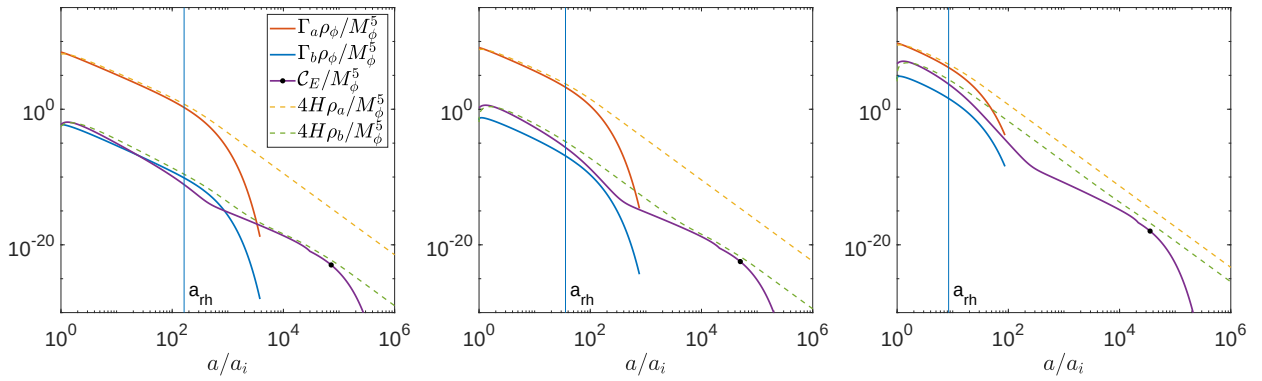


Figure 3.8: Comparison of the collision term with the inflaton decay terms into the two sectors. The plots are for parameters $\{\Gamma_{0a} = 10^{-11}, w = 10^{-8}\}$ (left), $\{\Gamma_{0a} = 10^{-10.5}, w = 10^{-8}\}$ (center) and $\{\Gamma_{0a} = 10^{-10}, w = 10^{-4}\}$ (right). The vertical blue line denotes the point where reheating occurs. In these plots $4H\rho_{a,b}$ serve as proxies for the temperatures of the two sectors. Since we have taken $\alpha_a = \alpha_b$, the temperature ratio is simply $x = (4H\rho_b/4H\rho_a)^{1/4}$.

to the evolution of the energy density in eq. (3.25) as a function of scale factor. Because $\mathcal{C}_{E,UV}$ redshifts faster than $\Gamma_{0a,b}\rho_\phi$, the energy injection due to inflaton decays can exceed $\mathcal{C}_{E,UV}$ before reheating terminates, $\mathcal{C}_{E,UV}(T_{\rm rh}) < \Gamma_{0,b}(T_{\rm rh})\rho_\phi$. When this occurs, (see, e.g. the left plot in figure 3.8), the temperature ratio at the end of reheating is the same as the one obtained due to the reheating attractor. Thus, asymmetric reheating overwhelms the collision term. However, when $\mathcal{C}_{E,UV}(T_{\rm rh}) > \Gamma_{0,b}(T_{\rm rh})\rho_\phi$, the temperature ratio at the end of reheating, $x_{\rm rh}$, is larger than the case without scattering, i.e. the result obtained from the reheating attractor. This deviation would not be reflected in the final temperature asymmetry if x_{sc} is larger than this modified $x_{\rm rh}$ (center plot in figure 3.8). It is only when the modified $x_{\rm rh}$ due to $\mathcal{C}_{E,UV}(T_{\rm rh})$ is larger than x_{sc} (right plot in figure 3.8), that the effects from $\mathcal{C}_{E,UV}$ impact the final temperature ratio as we see in the top left corner of figure 3.7. It is worth recalling that thermal effects beyond the scope of this chapter, in particular Landau damping and thermal blocking, can be important for determining the duration and dynamics of reheating in the high-$T_{\rm rh}$ regime where the effects from $\mathcal{C}_{E,UV}$ show up.

Mixed Yukawa and trilinear couplings

Finally, we consider a theory in which inflaton has trilinear couplings to scalars in sector a and Yukawa couplings to fermions in sector b,

$$\mathcal{L}_{\rm int} = \frac{1}{2}\mu_a \phi \chi_a \chi_a + y_b \phi \bar{\psi}_b \psi_b. \tag{3.55}$$

This interaction results in zero-temperature partial widths given by

$$\Gamma_{0,a} \approx \frac{\mu_a^2}{32\pi M_\phi}, \qquad \text{and} \qquad \Gamma_{0,b} \approx \frac{y_b^2}{8\pi} M_\phi. \tag{3.56}$$

The spin-summed s-channel scattering amplitude between the two sectors is

$$|\overline{\mathcal{M}}(s)|^2 = 2\mu_a^2 y_b^2 \left(1 - \frac{4m_b^2}{s}\right) \frac{s}{(s - M_\phi^2)^2 + (\Gamma_{0a} + \Gamma_{0b})^2}. \tag{3.57}$$

Using this scattering amplitude we derive the total energy transfer rate, $\mathcal{C}_E$, given in eq. (A.111); see figure A.4. The collision term is almost insensitive to T_b except when $T_b \approx T_a$. However, since the two sectors have different quantum statistics, the behavior of the collision term changes depending on which sector is hotter. When there is a large temperature asymmetry between the two sectors ($T_a \gg T_b$ or $T_b \gg T_a$), analogous to

the cases considered above, the temperature dependence of $\mathcal{C}_E$ is approximately

$$\mathcal{C}_E = \frac{1}{16\pi^3} \times \begin{cases} \dfrac{4\mu_a^2 y_b^2}{(\mu_a/M_\phi)^2 + 4y_b^2} T_a^3 \left[1.6 \log\left(\frac{T_a}{M_\phi}\right) + 1.1\right] & T_a > M_\phi,\ T_b \ll T_a \\ -0.30 \dfrac{4\mu_a^2 y_b^2}{(\mu_a/M_\phi)^2 + 4y_b^2} T_b^3 & T_b > M_\phi,\ T_b \gg T_a \\ \pm \dfrac{4\mu_a^2 y_b^2}{(\mu_a/M_\phi)^2 + 4y_b^2} M_\phi^2 \dfrac{T_{a,b}}{4} K_2\left(\frac{M_\phi}{T_{a,b}}\right) & T_{a,b} \lesssim M_\phi,\ T_{b,a} \gg T_{a,b} \\ \dfrac{31}{4\pi^2} \mu_a^2 y_b^2 \dfrac{T_a^7}{M_\phi^4} & m_{a,b} \ll T_a \ll M_\phi,\ T_b \ll T_a \\ -\dfrac{21}{4\pi^2} \mu_a^2 y_b^2 \dfrac{T_b^7}{M_\phi^4} & m_{a,b} \ll T_b \ll M_\phi,\ T_a \ll T_b, \end{cases} \tag{3.58}$$

where the minus signs appear when $T_b > T_a$, as consistent with our definition of the energy transfer term in eq. (3.25).

Determining the final temperature asymmetry due to inflaton-mediated scattering as in section 3.3.1, making use of $\mathcal{C}_E$ from eq. (A.111) with $x \to 0$, we find

$$x_{sc} = \left(\frac{1}{M_\phi^2} \frac{\Gamma_{0,a}\Gamma_{0,b}}{\Gamma_{0,a} + \Gamma_{0,b}} \frac{M_{\rm Pl}}{\alpha_{\rm cold}\sqrt{\alpha_{\rm hot}}}\right)^{1/4} \times \begin{cases} 1.24 & T_b \ll T_a \\ 1.19 & T_b \gg T_a, \end{cases} \tag{3.59}$$

where $\alpha_{\rm hot}$ ($\alpha_{\rm cold}$) denotes the value of $\alpha = \pi^2 g_*/30$ corresponding to the hotter (colder) sector. The final temperature asymmetry can then be estimated using eq. (3.42). In the left panel of figure 3.9 we show numerical results for the final temperature ratio. In the right panel of figure 3.9 we show the disagreement between our analytic estimate and the numerical result as a percentage of the numerical result.

3.4 Summary and conclusion

Asymmetric reheating is a minimal way to populate dark sectors that are otherwise completely decoupled from the SM following inflation. In this work, we have performed the first detailed analysis of perturbative asymmetric reheating. Specifically, by solving the Boltzmann equations describing the perturbative decay of the inflaton into two otherwise decoupled radiation sectors, we have studied in detail the resulting temperature asymmetries attained by the sectors. Scattering processes mediated by inflaton exchange couple the two sectors in the UV, and our self-consistent treatment takes into account the associated collision terms that transfer energy between the radiation sectors. Furthermore, we have carefully accounted for the effects of quantum statistics. At high temperatures (compared to the relevant mass scale in the problem, the inflaton mass) these quantum-statistical effects lead to important corrections in both the inflaton decay rate, as well as the inflaton-mediated scattering processes that transfer energy between the sectors.

The system of Boltzmann equations describing the evolution of the energy densities in the various sectors is a coupled set of three first-order non-linear differential equations, and a general analytic solution is not available. However, in this work we have demonstrated that the system can be accurately analyzed by making

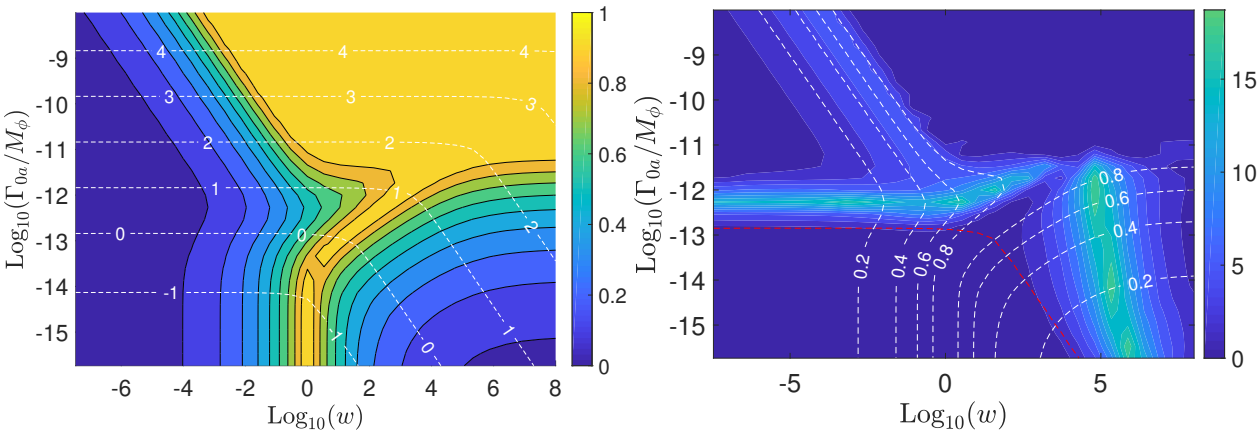


Figure 3.9: *Left panel*: Contours of the final temperature ratio obtained numerically, $x_{f,n} = T_b/T_a$, shown in color, for an inflaton coupled to scalars in sector a and fermions in sector b. The white dashed lines show contours of $\log_{10}(T_{\rm rh}/M_\phi)$. *Right panel*: Contours of relative fractional discrepancy, $|x_{f,a} - x_{f,n}|/x_{f,n}$, where $x_{f,a}$ is the analytic estimate of eq. (3.42) with eq. (3.59), in percent. The white dashed contours depict $x_{f,n}$ and the red dashed line marks $T_{\rm rh} = M_\phi$.

use of the attractor nature of its solutions. Broadly, we have identified two classes of quasi-static attractor solutions to which the energy density of the radiation bath evolves depending on the physical process that is dominating the evolution. In a broad range of parameter space and to a good approximation, at any given time the evolution is dominated by either 1) the energy injection from the decay of the inflaton, 2) the transfer of energy between the sectors through inflaton-mediated scattering, or 3) the adiabatic expansion of the Universe. Case 1) leads to a *reheating attractor curve*, case 2) yields a *scattering attractor curve*, while in case 3) the radiation density simply redshifts as $\rho \propto a^{-4}$. As we have demonstrated, the utility of these attractor solutions is that they allow for a very accurate semi-analytic determination of the resulting temperature asymmetry between the sectors; the asymmetry is simply determined by the process which dominates the evolution at the latest time.

Our results for the temperature asymmetries generated by asymmetric reheating are surprisingly universal across various coupling structures and particle types. The key property that determines the outcome of asymmetric reheating is the reheating temperature, $T_{\rm rh}$ relative to the inflaton mass-scale, M_ϕ, as follows:

- When $T_{\rm rh} \ll M_\phi/4$, the temperature asymmetry is solely determined by perturbative reheating process. More specifically, when $T_{\rm rh} < M_\phi/10$, the final temperature ratio is simply given by the ratio of the zero-temperature decay widths, $x = T_b/T_a = (\alpha_a \Gamma_{0b}/\alpha_b \Gamma_{0a})^{1/4}$. As the reheat temperature is increased (but still $< M_\phi$) quantum-statistical corrections to the inflaton decay width begin to significantly affect the final temperature asymmetry. In this region asymmetric reheating can be achieved by quantum statistical effects alone, with otherwise identical couplings.

- When $T_{\rm rh} \gg M_\phi/4$, the final temperature asymmetry is determined solely by inflaton-mediated scattering. Inflaton-mediated energy transfer between the sectors falls off exponentially when the temperature of the hotter sector falls below $T_a < M_\phi/4$ due to the s-channel scattering process going off-resonance. If the radiation sectors have not thermalized by this time, the colder sector is populated by a freeze-in like process where its final density (or equivalently the temperature ratio $x = T_b/T_a$) is

primarily determined by the collision term and the Hubble rate at $T_a = M_\phi/4$,

$$x_{sc} \sim \left(\frac{\alpha_a \mathcal{C}_E}{H \alpha_b \rho_a}\right)_{T_a = M_\phi/4}^{1/4}.$$

Because the collision term $\mathcal{C}_E(T_a = M_\phi/4)$ is largely insensitive to the inflaton coupling to the hotter sector as well as (at $T_a = M_\phi/4$) the quantum statistics of the interacting particles, the final temperature ratio is determined solely by the coupling strength of the colder sector irrespective of its particle identity.

In the region $T_{\rm rh} \sim M_\phi/4$, both reheating and scattering are important in determining the final temperature asymmetry. We find that the final temperature asymmetry, as a function of $T_{\rm rh}$ and the ratio of zero temperature partial widths w depends on the inflaton mass only through $T_{\rm rh}/M_\phi$. However, lower inflaton masses allow for the consistent realization of higher values of T/M_ϕ prior to reheating, which can be particularly important for models where the inflaton couples to the radiation baths through non-renormalizeable interactions (as in the axionic coupling to gauge bosons considered here).

The primary goal of this chapter was to analyze, in detail, the temperature evolution of two otherwise-decoupled radiation sectors during and after asymmetric reheating, but along the way we obtained a number of other novel results. We found novel power laws describing the evolution of radiation baths during reheating at temperatures larger than the inflaton mass scale, when quantum statistics are important. We developed methods to derive closed form (approximate) analytic expressions for energy transfer rates between two relativistic particles at different temperatures via s-channel interactions mediated by a massive scalar field. Finally, we derived reduced integral-expressions for energy-transfer rate between two relativistic sectors at different temperatures via t-channel interactions.

The analytic estimates of the final temperature ratio developed here for two-sector reheating can be straightforwardly extended to N-sector reheating scenarios [67]. In such cosmologies, for each of the sub-dominant sectors, the dominant energy injection from scattering is the collision term determined by the hottest sector. Provided the expansion rate is dominated by a single component (either the inflaton, or a single dominant radiation bath), to a very good approximation, the subdominant sectors are insensitive to each others presence.

In this work, we limited our analysis to perturbative reheating, ignoring the effects from 1) incomplete internal thermalization in the sectors during early reheating, 2) thermal modifications to the inflaton decays from collective effects, such as thermal blocking or Landau damping, 3) back-scatterings into inflaton quanta and 4) preheating. As long as these effects do not significantly alter the final reheating temperature obtained from perturbative reheating, our results for the final temperature asymmetry remain robust. Even in scenarios where such effects do significantly alter the reheat temperatures, the scattering attractor curve provides a strict upper bound to the temperature asymmetry between the sectors, $x \geq x_{sc}$ (see eq. (3.42)), as long as reheating occurs before inflaton-mediated scattering drops off resonance, $T_{\rm rh} > M_\phi/4$. We leave the further study of temperature asymmetries under these potentially disruptive effects to future work. Another possible extension of this work is to study scenarios that include large asymmetries in the number of degrees of freedom in the two sectors. In such a scenario, the sector with the higher temperature could have a sub-dominant energy density, a possibility we explicitly ignored in this work.

Chapter 4

Cannibal imprints on the matter power spectrum

4.1 Introduction

While standard cosmology posits that post-inflationary reheating is followed by uninterrupted radiation domination prior to matter-radiation equality, a variety of well-motivated particle physics scenarios predict departures from radiation domination in the poorly constrained epoch between the end of inflation and Big Bang nucleosynthesis (BBN) [68]. For instance, supersymmetric theories often predict moduli whose energy density can come to dominate the universe as they coherently oscillate in a quadratic potential, giving rise to an early matter-dominated era (EMDE) that ends when the modulus decays [62]. The semi-classical evolution of light spin-zero fields can also give rise to epochs of kination when the scalar field's kinetic energy dominates over its potential energy [69–71].

Early departures from radiation domination are also generic consequences of theories that contain an internally thermalized hidden sector that is thermally decoupled from the Standard Model (SM). Such decoupled self-interacting hidden sectors are readily obtained from straight-forward inflationary scenarios [17, 40–43], and can naturally provide a cosmological origin for the dark matter (DM) of our universe [6, 40, 72, 73], a possibility that becomes ever more compelling with the continued absence of direct detection signals to date. If the lightest state in the hidden sector is massive, then it can easily come to dominate the energy density of the universe after it becomes non-relativistic. If this particle is effectively pressureless when it dominates, it produces an EMDE [11–13, 74, 75].

Altered expansion histories prior to BBN can leave potentially observable footprints in dark matter perturbations on scales that experienced altered growth [18, 76–79]. Since subhorizon dark matter density perturbations grow linearly with the scale factor during matter domination, an EMDE generates a significantly enhanced population of sub-earth-mass dark matter halos if the dark matter particles are cold enough to form such structures [18, 76]. The masses and central densities of the smallest microhalos are determined by the small-scale cutoff in the matter power spectrum. The rapid growth of perturbations during the EMDE implies that the observational signatures of these microhalos, such as the dark matter annihilation rates within their dense cores, are extremely sensitive to the scale of this cutoff [18, 20, 80]. If dark matter does not interact with SM particles, the small-scale cutoff is most often determined by the microphysics of the species that produces the altered cosmic evolution, making the microhalo population a probe of the particle

physics of the early universe as well as its expansion history.

In many familiar theories, ranging from the simple and minimal example of a single scalar field to the exceptionally well-motivated scenario of a confining Yang-Mills sector, the lightest particle in the dark sector has number-changing "cannibal" self-interactions that remain in equilibrium even after the particle becomes non-relativistic [7, 10, 81–89]. Such "cannibal" interactions [7, 90, 91] are natural properties for the lightest particle in a hidden sector with a mass gap, and meta-stable cannibal species arise frequently in models of hidden sector DM [10, 11, 74, 75, 92–94]. If this cannibal particle comes to dominate the energy density of the universe prior to BBN, then the subsequent early cannibal-dominated era (ECDE) will leave its imprint on the small scale dark matter perturbations. As we show here, cannibal interactions alter the evolution of density perturbations during an early cannibal-dominated era (ECDE) compared to their evolution during an EMDE and typically determine the small-scale cutoff in the matter power spectrum.

In contrast to previous studies of structure formation with a cannibal species [89, 95, 96] or a related toy model [97], the cannibal here is *unstable*, decaying to SM particles during the Universe's first second. Therefore, it does not contribute to the present-day dark matter abundance, which we assume to be composed of a separate species. For simplicity we consider the DM abundance to have formed prior to the ECDE, as is easily realized, e.g., when DM is a thermal relic in the hidden sector. We are thus interested in the evolution of perturbations in a three-component universe, consisting of cold decoupled DM, the SM radiation bath, and the metastable cannibal species.

We find that an ECDE generates a peak in the power spectrum of DM density fluctuations on scales that enter the cosmological horizon during the ECDE. We show that the shape, amplitude, and scale of the ECDE peak is determined by the properties of the cannibal particle, with little sensitivity to DM particle properties. This enhancement to the matter power spectrum will generate DM microhalos long before structure formation would occur in the absence of an ECDE, and we relate the characteristic mass and formation time of these microhalos to the mass of the cannibal field, the strength of its number-changing interactions, its temperature relative to the Standard Model particles, and its lifetime. Our results reveal a new window into the thermal history of the pre-BBN Universe and further establish that hidden sectors can generate distinctive observational signatures even in the absence of interactions between DM and the SM.

The organization of this chapter is as follows. In section 4.2 we review the novel cosmic evolution of a cannibal fluid. Then in section 4.3, we embed the cannibal in the early universe along with dark matter and Strandard Model radiation, and discuss the homogeneous evolution of the resulting cosmologies. In section 4.4, we study perturbation growth in these cosmologies and highlight important length scales, showing that both the magnitude and scale of maximum DM perturbation growth are directly connected to cannibal particle properties. We discuss possible breakdowns of the perfect-fluid approximation in section 4.5. Implications of early cannibal-dominated eras for the earliest-forming microhalos are discussed in section 4.6, and we conclude in section 4.7. Several technical results used in this chapter are derived in appendix B.

4.2 Cannibal evolution

A representative particle model for cannibal involves a scalar field with quartic and cubic self couplings,

$$\mathcal{L}_{can} = \frac{1}{2}\partial^\mu\varphi\partial_\mu\varphi - \frac{1}{2}m^2\varphi^2 - \frac{g}{3!}\varphi^3 - \frac{\lambda}{4!}\varphi^4. \tag{4.1}$$

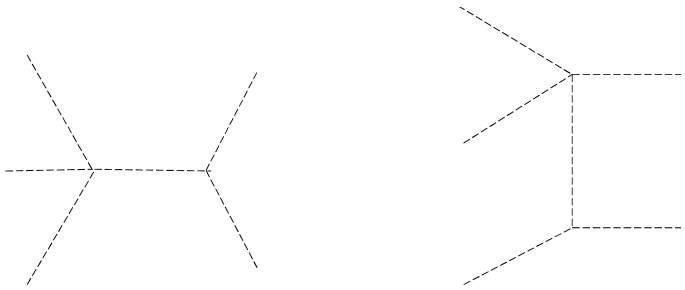


Figure 4.1: Representative Feynmann diagrams for cannibal three-to-two scattering process.

In this theory the Feynmann diagrams shown in figure 4.1 contribute to the three-to-two number changing interactions, which maintains the chemical equilibrium in the cannibal plasma. While eq. (4.1) describes a specific cannibal model, it also provides a useful toy model for a broad class of theories with cannibal interactions. For instance, the lightest glueballs in a pure SU(N) sector have cannibal interactions that can be described with an effective Lagrangian of the form in eq. (4.1) [83, 84, 86, 87, 89]. In the remainder of the section we solve the cosmic evolution of cannibal for the above model.

As there is no external energy injection or ejection in the cannibal, the cannibal density evolution follows the energy conservation equation in expanding spacetime,

$$\frac{d\rho_{can}}{dt} + 3H\rho_{can}(1 + w_c(a)) = 0, \tag{4.2}$$

where ρ_{can} is the energy density of the cannibal, w_c is the equation of state of thhe cannibal and H is the Hubble rate. Considering the cannibal to be in thermal equilibrium[1], we can express both its density and its equation of state in terms of temperature,

$$\rho_{can} = \rho_{can,eq} = \int \frac{d^3\vec{p}}{(2\pi)^3} E f_{eq}(E/T_c) = \frac{m^4}{2\pi^2}\int_1^\infty d\tilde{E}\tilde{E}^2\sqrt{\tilde{E}^2-1} f_{\rm eq}(\tilde{E}x) \equiv \frac{m^4}{2\pi^2}h(x), \tag{4.3}$$

$$w_c(a) = w_{c,eq}(a) = \frac{\mathcal{P}_{can,eq}}{\rho_{can,eq}} = \frac{\int_1^\infty d\tilde{E}(\tilde{E}^2-1)^{3/2} f_{\rm eq}(\tilde{E}x)}{3\int_1^\infty d\tilde{E}\tilde{E}^2\sqrt{\tilde{E}^2-1} f_{\rm eq}(\tilde{E}x)} \equiv \frac{g(x)}{h(x)}, \tag{4.4}$$

where $E = \sqrt{\vec{p}^2 + m^2}$, m is the mass of cannibal, $\tilde{E} \equiv E/m$, T_c is the temperature of cannibal, $\mathcal{P}_{can}$ is the pressure of cannibal, $\rho_{can,eq}$ and $\mathcal{P}_{can,eq}$ are equilibrium density and pressure respectively, and f_{eq} is the Bose-Einstein distribution in chemical equilibrium. At temperatures $T_c \gg m$, eq.. (4.4) yields $w_c = 1/3$. Consequently eq. (4.2) gives back the expected evolution of radiation bath. However, as cannibal becomes non-relativistic, the reverse $2 \to 3$ reactions are phase-space suppressed but annihilations via $3 \to 2$ processes remain active. This causes the comoving number density to deplete, and the rest mass energy is converted to thermal energy. This gives rise to the novel behaviour of cannibalism. Combining eq. (4.2) with eq. (4.3) and eq. (4.4), we can solve for the evolution of temperature with scale factor as the cannibal transitions from the relativistic to the non-relativistic regime. Note that— like the evolution behaviour for radiation and matter— the evolution of T_c and $\rho_{can,eq}$ with scale factor are independent of the form of Hubble rate.

[1]In this section by thermal equilibrium we mean that both chemical and kinetic equilibrium are maintained.

4.2.1 Equilibrium cannibal evolution

We first evaluate T_c as a function of scale factor by integrating eq. (4.2) through

$$-\ln(a/a_i) = \int_{0.1}^{x} \frac{h'(\tilde{x})}{3[h(\tilde{x}) + g(\tilde{x})]} d\tilde{x} \equiv F(x), \tag{4.5}$$

where we have used $x(a_i) = 0.1$ and primes denotes derivatives with respect to x. We evaluate $F(x)$ at several values of x and use the resulting table to define an interpolating function for x as a function of $F = -\ln(a/a_i)$. We find $\rho_{\rm can,eq}(a)$ by inserting the resulting $x(a)$ into eq. (4.3). For $a < a_{\rm fz}/3$ in figure 4.2, the blue and black curves shows the evolution of the equilibrium cannibal density and temperature obtained using this procedure. The orange dotted line shows the evolution if we completely neglect mass inside f_{eq} in eq. (4.3) and eq. (4.4). For $T_c > m$, the cannibal evolves like radiation as expected.

In the limits $x \gg 1$ and $x \ll 1$, we find simple analytical expressions for $F(x)$ using the asymptotic expansions

$$h(x) \approx \begin{cases} \dfrac{\pi^4}{15}\dfrac{1}{x^4} & x \ll 1 \\ \dfrac{1}{x^{3/2}}\left(1 + \dfrac{27}{8x} + \dfrac{705}{128x^2} + O(x^{-3})\right)\sqrt{\dfrac{\pi}{2}}e^{-x} & x \gg 1 \end{cases} \tag{4.6}$$

$$g(x) \approx \begin{cases} \dfrac{\pi^4}{45}\dfrac{1}{x^4} & x \ll 1 \\ \dfrac{1}{x^{3/2}}\left(\dfrac{1}{x} + \dfrac{15}{8x^2} + O(x^{-3})\right)\sqrt{\dfrac{\pi}{2}}e^{-x} & x \gg 1. \end{cases} \tag{4.7}$$

Using the $x \ll 1$ limits for $g(x)$ and $h(x)$ in eq. (4.5) gives the expected $T \propto 1/a$ scaling for relativistic particles. The $x \gg 1$ limits give us the evolution of the cannibal fluid during cannibalism. To connect the non-relativistic evolution of the cannibal fluid to its early relativistic evolution we need to integrate in the semi-relativistic regime ($x \sim 1$) where no simple analytical expressions are available. We handle the integration in the semi-relativistic regime by breaking up the integral in eq. (4.5) into two integrals: one in the region $0.1 < \tilde{x} < 10$, and one in the region $10 < \tilde{x} < x$. Then we use the large-x approximations for $h(x)$ and $g(x)$ in the second integral to obtain

$$F(10) - \frac{1}{3}\int_{10}^{x}\left(1 + \frac{1}{2\tilde{x}} + \frac{35}{8\tilde{x}^2} + \mathcal{O}(\tilde{x}^{-3})\right) d\tilde{x} \approx -\ln(a/a_i). \tag{4.8}$$

Taking $F(10) = -6.5$ in eq. (4.8) implies

$$x = 3\ln\left(\frac{a/a_i}{17.5x^{1/6}}\right) + \frac{35}{8x} + \mathcal{O}\left(x^{-2}\right). \tag{4.9}$$

To obtain a simpler relation between x and a, we neglect the $1/x$ term and set $x = 10$ in the logarithm, which is approximately true during cannibalism, as seen in figure 4.2. With these simplifications,

$$\frac{T_c}{m} = \frac{1}{x} \approx \frac{1}{3\ln\left(\frac{a}{25.6a_i}\right)}. \tag{4.10}$$

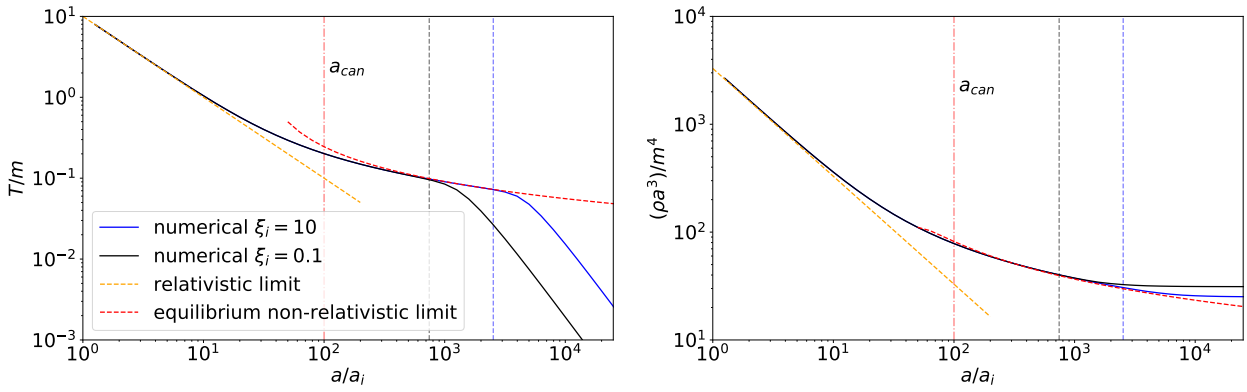


Figure 4.2: Temperature (left) and density (right) of the cannibal fluid as a function of scale factor. Solid blue (black) lines show numerical results for the cannibal when the universe is cannibal- (SM radiation-) dominated and correspond to $m = 35$ TeV, $\alpha_c = 1$, $T_{\rm rh} = 10$ MeV, and $\xi_i = 10$ ($\xi_i = 0.1$), where $T_{\rm rh}$ and ξ_i are as defined in eq. (4.34) and eq. (4.30), respectively. The orange and red dashed lines show analytical results for the equilibrium cannibal fluid applicable in the relativistic and non-relativistic limits, respectively (eqs. (4.10) and (4.12)). The vertical red dot-dashed line marks the onset of cannibalism. The remaining vertical dot-dashed lines indicate $a_{\rm fz}$ for the parameter point corresponding to the solid line of the same color.

In the left panel of figure 4.2 the red dashed curve shows this result for the temperature evolution, which accurately describes the evolution of the cannibal fluid once it becomes non-relativistic.

To determine the evolution of the cannibal density in the non-relativistic regime we first write its density in the large x limit by using eq. (4.6) in eq. (4.4). Since $h(x)$ has an exponential dependence on x, we use eq. (4.9) instead of eq. (4.10) in the exponential term. Expanding the resulting equation to order $1/x^2$ gives

$$\rho_{\rm can,eq} \approx m^4 \left[\sqrt{\frac{\pi}{2}}\frac{(17.5)^3}{2\pi^2}\right] \frac{1}{(a/a_i)^3 x}\left[1 - \frac{1}{x} + \mathcal{O}(x^{-2})\right]. \tag{4.11}$$

Since the temperature of cannibal particles remains of order $0.1m$ during cannibalism, as seen in left panel of figure 4.2, the next-to-leading order term in $1/x$ above provides a $\sim 10\%$ correction. As the above relation no longer depends exponentially on x, we express x in terms of a using eq. (4.10) to obtain

$$\rho_{\rm can,eq} \approx \frac{148 m^4}{(a/a_i)^3 \ln(a/[25.6 a_i])}\left[1 - \frac{1}{3\ln(a/[25.6 a_i])} + \mathcal{O}(x^{-2})\right]. \tag{4.12}$$

In the right panel of figure 4.2 we show the above estimate of $\rho_{\rm can,eq}$ as the red-dashed line. We define the beginning of cannibalism at

$$T_c(a_{can}) \equiv \frac{m}{5} \qquad\qquad a_{can} = 101 a_i, \tag{4.13}$$

which marks the point at which T_c and ρ_{can} can be well approximated by eq. (4.10) and eq. (4.12), respectively.

4.2.2 Cannibal freeze-out

The cannibalism phase continues until the number-changing interaction rate goes below the Hubble rate. After this point the cannibal no longer maintains chemical equilibrium and evolves like cold matter. Since two-to-two interactions are stronger than three-to-two interactions, the cannibal stays in kinetic equilibrium

throughout the freeze-out process. The freeze-out process for the cannibal is governed by the equation

$$\frac{dn_{can}}{dt} + 3Hn_{can} = \langle\sigma v^2\rangle_{\rm can} n_{can}^2 (n_{eq,can} - n_{can}), \tag{4.14}$$

where n_{can} is the number density of the cannibal, $n_{eq,can}$ is the number density of cannibal in thermal equilibrium and $\langle\sigma v^2\rangle_{\rm can}$ is the $3 \to 2$ cross-section of the cannibal. We are only interested in scenarios in which freeze-out happens when cannibal is non-relativistic. To leading order in T_c/m, the s-wave component of $\langle\sigma v^2\rangle_{\rm can}$ dominates and is given by (see appendix B.1)

$$\langle\sigma v^2\rangle_{\rm can} = \frac{25\sqrt{5}(g/m)^2[(g/m)^2 + 3\lambda]^2}{147456\pi m^5} \equiv \frac{25\sqrt{5}\pi^2\alpha_c^3}{5184 m^5}. \tag{4.15}$$

Here α_c parametrizes the combination of cannibal couplings that determines the strength of $3 \to 2$ reactions,[2]

$$(4\pi\alpha_c)^3 \equiv \frac{9}{4}(g/m)^2[(g/m)^2 - 3\lambda]^2. \tag{4.16}$$

We define the scale-factor at freeze-out of $3 \to 2$ reactions, $a_{\rm fz}$, through

$$\langle\sigma v^2\rangle_{\rm can} n_{\rm can}^2(a_{\rm fz}) = H(a_{\rm fz}). \tag{4.17}$$

At early times, $a \ll a_{fz}$, eq. (4.14) effectively sets $n_{can} = n_{can,eq}$, equivalently eq. (4.14) sets the chemical potential, $\mu = 0$. After $a \gg a_{fz}$, cannibal reactions stop and the cannibal particles evolve as cold matter, $\rho_{can} \propto 1/a^3$ and $T_c \propto 1/a^2$. However, to accurately find the cannibal density and temperature near a_{fz} we need to solve for both eq. (4.14) and eq. (4.2).

Numerically, we find that the cannibal fluid starts to depart from its chemical equilibrium around $a > a_{\rm fz}/3$. Since the cannibal particles are non-relativistic by the time freeze-out occurs, we use Maxwell-Boltzmann statistics to obtain

$$n_{\rm can} = e^{\mu/T} n_{\rm can,eq} = e^{\mu x/m} \frac{m^3}{2\pi^2} \frac{K_2(x)}{x}; \tag{4.18}$$

$$\rho_{\rm can} = e^{\mu/T} \rho_{\rm can,eq} = e^{\mu x/m} \frac{m^4}{2\pi^2} \left(\frac{xK_1(x) + 3K_2(x)}{x^2}\right); \tag{4.19}$$

$$w_c = \frac{\mathcal{P}_{\rm can}}{\rho_{\rm can}} = \frac{\mathcal{P}_{\rm can,eq}}{\rho_{\rm can,eq}} = \frac{K_2(x)}{xK_1(x) + 3K_2(x)}; \tag{4.20}$$

where $K_i(x)$ is the modified Bessel function of i^{th} order.

To find the evolution equation for x we begin by expressing $\rho_{\rm can}$ in terms of $n_{\rm can}$ and x using eqs. (4.18)-(4.20),

$$\rho_{\rm can} = m n_{\rm can} \frac{1}{x w_c(x)}. \tag{4.21}$$

Using the above relation to express $\rho_{\rm can}$ in terms of $n_{\rm can}$ in the energy conservation equation, eq. (4.2), yields

$$\frac{m}{x w_c}\left(a\frac{dn_{\rm can}}{da} + 3n_{\rm can}\right) - \frac{m n_{\rm can}}{x^2}\left(\frac{w_c + x w_c'(x)}{w_c^2} a\frac{dx}{da} - 3x\right) = 0. \tag{4.22}$$

[2]The factor of 9/4 in the front has been added to counter an earlier error made in calculating eq. (4.15).

We then simplify the first bracket using eq. (4.14) to obtain the evolution equation for x,

$$a\frac{dx}{da} - 3\frac{xw_c^2}{w_c + xw_c'(x)} = \frac{xw_c}{w_c + xw_c'(x)}\frac{\langle\sigma v^2\rangle}{H}n_{\rm can}(n_{\rm can,eq} - n_{\rm can}). \tag{4.23}$$

Eq. (4.14) and eq. (4.23) form coupled differential equations which are evaluated together to solve for $n_{\rm can}(a)$ and $x(a)$. In the left panel of figure 4.2 we plot the temperature (blue and black lines) for $a > a_{\rm fz}/3$ by numerically solving eq. (4.14) and eq. (4.23). In the right panel we plot the cannibal density for $a > a_{\rm fz}/3$ using the numerically evaluated x and $n_{\rm can}$ in eq. (4.21). Notice that the evolution of the cannibal thermodynamic quantities is very similar in the two cases where the universe is cannibal dominated (blue line) or SM radiation dominated (black line), with the only difference being the specific value of $a_{\rm fz}$. Here $a_{\rm fz}$ is evaluated by finding where equality in eq. (4.17) is satisfied for numerically obtained $n_{\rm can}$.

4.3 Homogeneous background evolution

We are interested in a universe comprised of three components: the cannibal species, DM, and the thermal SM radiation bath. For simplicity, we consider the DM relic abundance to already be in place at the beginning of our analysis. In the natural and minimal scenario where DM and the cannibal species are part of the same thermal bath in the early universe, we expect DM to be heavier than the cannibal: to experience cannibalism the cannibal species cannot be in equilibrium with any relativistic species while it is non-relativistic, and thus is generically the lightest state in that sector. The cannibal species must be thermally decoupled from the SM radiation bath, making the initial temperature ratio between the two sectors a free parameter. Once the cannibal particle becomes non-relativistic, its energy density dilutes more slowly than that of the SM radiation and will eventually come to dominate the universe provided the cannibal is sufficiently long-lived. We focus on the parameter space where the universe undergoes such an *early cannibal-dominated era* (ECDE) and caution that the cannibal may or may not be actively undergoing cannibalism during the ECDE. The cannibal eventually decays into SM particles, which must occur before neutrino decoupling to avoid spoiling the successful predictions of BBN [98–101] and altering the features of the CMB [102, 103].

The Boltzmann equations that describe the homogeneous evolution of the cannibal fluid in the early universe, together with DM and SM radiation, are

$$\frac{d\rho_{\rm can}}{dt} + 3H\rho_{\rm can}(1 + w_c(a)) = -\Gamma m n_{\rm can} \tag{4.24}$$

$$\frac{d\rho_r}{dt} + 4H\rho_r = \Gamma m n_{\rm can} \tag{4.25}$$

$$\frac{d\rho_{\rm DM}}{dt} + 3H\rho_{\rm DM} = 0 \tag{4.26}$$

$$\frac{dn_{\rm can}}{dt} + 3Hn_{\rm can} = \langle\sigma v^2\rangle_{\rm can} n_{\rm can}^2(n_{\rm can,eq} - n_{\rm can}) - \Gamma n_{\rm can}, \tag{4.27}$$

where the Hubble rate is given by

$$H = \frac{1}{\sqrt{3}M_{\rm Pl}}\sqrt{\rho_{\rm can} + \rho_r + \rho_{\rm DM}}, \tag{4.28}$$

ρ_r, $\rho_{\rm DM}$ and $\rho_{\rm can}$ are the energy densities of SM radiation, DM and the cannibals, respectively, w_c is the cannibal equation of state, $n_{\rm can}$ is the cannibal number density and $n_{\rm can,eq}$ its equilibrium value, Γ is the

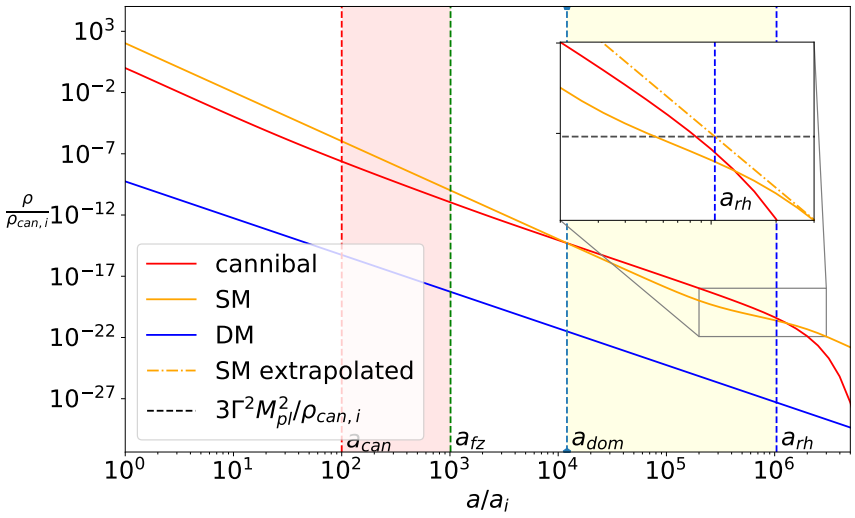


Figure 4.3: Evolution of the cannibal (red), SM radiation (solid orange) and DM (solid blue) energy densities as a function of scale factor. The inset figure highlights our definition of $a_{\rm rh}$ (see eq. (4.34)) as the scale factor when the SM energy density extrapolated adiabatically back in time (orange dot-dashed) becomes equal to the total energy density required for the Hubble rate to equal decay rate of the cannibal particles (black dashed). The shaded red region highlights the period of cannibalism. The yellow shaded region highlights the period of early matter domination produced by the cannibal particles. This figure takes $m = 190$ GeV, $T_{\rm rh} = 10$ MeV, $\alpha_c = 0.06$, and $\xi_i = 1$.

zero-temperature decay width of the cannibal particle to the SM, and $M_{\rm Pl} = 2.435 \times 10^{18}$ GeV is the reduced Planck mass. The collision operator describing cannibal decays that appears on the right-hand side of these equations is derived in appendix A.2.

Chemical equilibrium in the cannibal fluid is maintained as long as the $3 \to 2$ scattering rate is rapid compared to H; the freeze-out of this cannibal interaction is described by eq. (4.27). We assume that two-to-two cannibal scatterings are fast enough to maintain internal kinetic equilibrium. Thus all the thermal quantities for the cannibal fluid can be expressed in terms of its chemical potential, μ, and its temperature, T_c. Consequently the system of eqs. (4.24)-(4.27) can be solved for the four unknowns $\rho_r, \rho_{\rm DM}, \rho_{can}$, and T_c.

We set our initial conditions at an initial scale factor a_i, defined such that

$$T_c(a_i) = 10m. \tag{4.29}$$

The cannibal fluid is in chemical equilibrium initially, so that $\mu(a_i) = 0$. We find the initial DM density by scaling the observed relic density back in time. Since the cannibal fluid and the SM radiation bath are necessarily thermally decoupled, the initial SM temperature $T_r(a_i)$ must be separately specified; we parameterize it with the initial temperature ratio

$$\xi_i \equiv \frac{T_c(a_i)}{T_r(a_i)} = \frac{10m}{T_r(a_i)}. \tag{4.30}$$

Figure 4.3 shows the evolution of energy densities beginning from these initial conditions until the cannibal particles decay. For $\xi_i = 1$ as shown here, the cannibal is subdominant to SM radiation at a_i but eventually comes to dominate. This fluid system goes through four important transitions, which we will discuss in turn: 1) the onset of cannibalism within the hidden sector, 2) the freeze-out of the 3-to-2 cannibal reactions, 3) the

transition from SM radiation domination to cannibal domination, and 4) the decay of the cannibal particles into the SM.

At temperatures $T_c \gg m$, the cannibal behaves as radiation. As the cannibal particles become non-relativistic ($T_c < m$), $2 \to 3$ scattering processes become Boltzmann-suppressed while $3 \to 2$ processes remain active. Thus the comoving number density depletes, which self-heats the cannibal particles by converting rest mass to kinetic energy. In this "cannibal" phase of evolution, the cannibal temperature decreases as $T_c \propto 1/\log(a)$ while its energy density decreases as $\rho_{can} \propto 1/(a^3 \log(a))$, as seen in eq. (4.10) and eq. (4.12).

As long as the cannibal fluid is in chemical equilibrium, the evolution of $\rho_{\rm can}(a)$ with a is independent of the Hubble rate. However, the scale at which the cannibal fluid can no longer maintain chemical equilibrium depends on the Hubble rate and thus on the presence of other species. After cannibal freeze-out, the temperature of the cannibal cools as $T_c \propto 1/a^2$, as expected for massive non-interacting particles.

The universe is initially SM radiation-dominated when $\rho_r(a_i) > \rho_{\rm can}(a_i)$ or

$$g_*[T_r(a_i)]T_r^4(a_i) > (10m)^4, \tag{4.31}$$

where $g_*(T_r)$ is the effective number of degrees of freedom in the SM. For $g_*[T_r(a_i)] \sim 100$, SM radiation domination at a_i then requires

$$\xi_i \lesssim 3.2. \tag{4.32}$$

A universe that is SM radiation dominated at a_i will transition to cannibal domination at the scale factor $a_{\rm dom}$ where

$$\rho_{\rm can}(a_{\rm dom}) = \rho_r(a_{\rm dom}), \tag{4.33}$$

where we have implicitly assumed that the cannibal lifetime is long enough that it will come to dominate before it decays.

When Γ exceeds the Hubble rate, the cannibal particles decay into the SM radiation bath and the universe then evolves as in the standard ΛCDM cosmology. We define the reheat temperature, $T_{\rm rh}$, by equating the Hubble rate in a SM radiation-dominated universe with the cannibal decay rate,

$$\sqrt{\frac{\pi^2 g_*(T_{\rm rh})}{30}}\frac{T_{\rm rh}^2}{\sqrt{3}M_{\rm Pl}} \equiv \Gamma. \tag{4.34}$$

We define the scale factor at reheating, $a_{\rm rh}$, by isentropically extrapolating the temperature of the SM from $T_{\rm rh}$ to the present-day temperature T_0,

$$g_{*s}(T_{\rm rh})(a_{\rm rh}T_{\rm rh})^3 = g_{*s}(T_0)(a_0T_0)^3. \tag{4.35}$$

Here g_{*s} is the effective number of entropic degrees of freedom in the SM and a_0 is the present-day scale factor. Note that with the above definition of $a_{\rm rh}$, the temperature of the SM at $a_{\rm rh}$, $T_r(a_{\rm rh})$, is not equal to $T_{\rm rh}$. This can be seen in the inset panel of figure 4.3 where the SM energy density (solid orange line) at $a_{\rm rh}$ is smaller than the radiation density when $T = T_{\rm rh}$ (horizontal black dashed line). Figure 4.3 also shows that the SM radiation density evolves adiabatically until the energy injection rate from the cannibal fluid into the radiation becomes of order the Hubble rate ($\rho_{\rm can}\Gamma/\rho_r \sim H$). After this time, the radiation density is

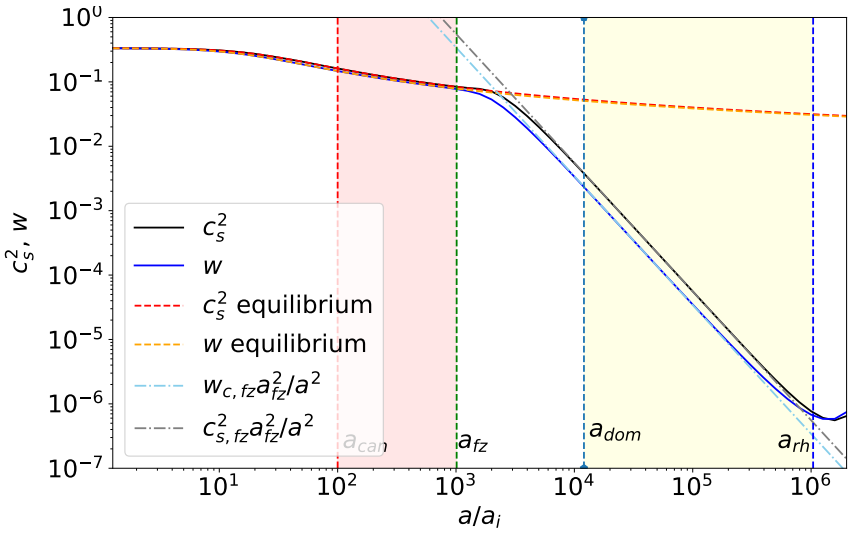


Figure 4.4: Evolution of the cannibal sound speed c_s^2 and equation of state w from initial cannibal temperature $T_c(a_i) = 10m$. The red (yellow) dashed line shows the evolution of c_s^2 (w) in thermal equilibrium, the gray (cyan) dot-dashed line when completely decoupled, and the solid black (blue) line shows the full numerical result. This figure uses the same parameter values as figure 4.3.

proportional to $\Gamma m n_{\rm can}/H$ until the cannibal energy density becomes subdominant.

In solving perturbation equations we need the cannibal equation of state w_c and the cannibal sound speed

$$c_s^2 = \frac{\partial \mathcal{P}_{\rm can}}{\partial \rho_{\rm can}} = w_c - \frac{a w_c'(a)}{3(1+w)}, \tag{4.36}$$

where $\mathcal{P}_{\rm can} = w_c \rho_{\rm can}$ is the pressure of the cannibal fluid. In figure 4.4 we plot the evolution of both w_c and c_s^2 as a function of scale factor. Both quantities begin to deviate from their equilibrium values around $2a_{\rm fz}$. For $a \gg a_{\rm fz}$ both quantities evolve as

$$c_s^2 = c_{s,\rm fz}^2 \frac{a_{\rm fz}^2}{a^2} \qquad\qquad w_c = w_{c,\rm fz} \frac{a_{\rm fz}^2}{a^2}, \tag{4.37}$$

where $c_{s,\rm fz}^2$ and $w_{c,\rm fz}$ are constants that give the correct asymptotic evolution (see dot-dashed line in figure 4.4). We find that $c_{s,\rm fz}$, to a good approximation, is given by $c_{s,\rm eq}(3a_{\rm fz})$ while $w_{c,\rm fz}$ is given by $w_{c,\rm eq}(2a_{\rm fz})$, where the subscript eq denotes that the variables are computed assuming the cannibal fluid to be in thermal equilibrium. Then using the fact that in the non-relativistic limit $c_{s,\rm eq}^2 \approx w_{c,\rm eq} \approx T_{c,\rm eq}/m$, we obtain

$$c_{s,\rm fz}^2 \approx \frac{1}{3\ln(3a_{\rm fz}/(c_2 a_i))} \qquad\qquad w_{c,\rm fz} \approx \frac{1}{3\ln(2a_{\rm fz}/(c_2 a_i))}. \tag{4.38}$$

In figure 4.4, both w_c and c_s^2 increase near $a_{\rm rh}$ because the cannibal particles with larger velocities decay later due to time dilation. Consequently, the temperature of the cannibal fluid increases as cannibal particles with slower speeds are removed first. However, the heating near $a_{\rm rh}$ is unimportant for the evolution of dark matter perturbations because we are interested in scenarios with $a_{\rm rh} \gg a_{\rm fz}$. Thus, the cannibal fluid is too cold at $a_{\rm rh}$ for the heating due to time dilation to have any impact.

4.3.1 Mapping between cannibal parameters and cosmological scales

Our early cannibal-dominated cosmology is governed by four free parameters: the initial temperature ratio ξ_i and the cannibal particle properties $m, T_{\rm rh}$, and α_c. These four parameters determine the three important scales $a_{\rm fz}/a_i$, $a_{\rm dom}/a_i$, and $a_{\rm rh}/a_i$ that will ultimately control the major features of the matter power spectrum. Due to the non-trivial evolution of the cannibal density, the exact relations between these cosmological scales and the underlying cannibal parameters are complicated, but useful approximate relations can be obtained by fixing $a/a_i = 10^3$ in the logarithm of the expression for $\rho_{\rm can}(a)$ given in eq. (4.12):

$$\rho_{\rm can} \sim \frac{40m^4}{(a/a_i)^3}. \tag{4.39}$$

This approximation is accurate to $\mathcal{O}(1)$ for scale factors between $a_{\rm can}$ and $a_{\rm rh}/10$ and will enable us to provide simple expressions for key quantities, albeit at the cost of obscuring logarithmic dependence on $a_{\rm fz}/a_i$.

We can express $a_{\rm rh}/a_i$ in terms of m and $T_{\rm rh}$ by setting the cannibal density at reheating equal to the radiation density. Using eq. (4.39) for the cannibal density then gives

$$\frac{a_{\rm rh}}{a_i} \sim 2.3 \left(\frac{m}{T_{\rm rh}}\right)^{4/3} \left(\frac{g_*(T_{\rm rh})}{10}\right)^{-1/3}. \tag{4.40}$$

Similarly, we can find $a_{\rm dom}$ in terms of ξ_i using eq. (4.39) for the cannibal density in the definition of $a_{\rm dom}$, eq. (4.33). With $\rho_r(a_{\rm dom}) = \rho_r(a_i)(a_i/a_{\rm dom})^4$, we then find

$$\frac{a_{\rm dom}}{a_i} \sim 80 g_*[T_r(a_i)] \frac{1}{\xi_i^4}. \tag{4.41}$$

To express $a_{\rm fz}$ in terms of cannibal parameters, we start with its definition in eq. (4.17). We then approximate $n_{\rm can}(a_{\rm fz}) \approx \rho_{\rm can}(a_{\rm fz})/m$ and express $\rho_{\rm can}$ and $\langle\sigma v^2\rangle_{\rm can}$ using eq. (4.39) and eq. (4.15) respectively. In the case where the Hubble rate is dominated by the cannibal density during freeze-out, i.e. $a_{\rm dom} < a_{\rm fz}$, we obtain

$$\frac{a_{\rm fz}}{a_i} \sim 3 \times 10^4 \alpha_c^{2/3} \left(\frac{\rm GeV}{m}\right)^{2/9}. \tag{4.42}$$

Similarly, in the case where the Hubble rate is dominated by the SM radiation during freeze-out, i.e. $a_{\rm dom} > a_{\rm fz}$, we use $\rho_r(a_{\rm fz}) = \rho_r(a_i)(a_i/a_{\rm fz})^4$ to obtain

$$\frac{a_{\rm fz}}{a_i} \sim 3.3 \times 10^4 \alpha_c^{3/4} \xi_i^{1/2} \left(\frac{\rm GeV}{m}\right)^{1/4} \left(\frac{g_*(10m/\xi)}{100}\right)^{-1/8}. \tag{4.43}$$

We see that $a_{\rm fz}/a_i$ decreases slowly as m increases because increasing m reduces the $3 \to 2$ cross-section for fixed α_c. When the universe is SM dominated at $a_{\rm fz}$, $a_{\rm fz}/a_i$ decreases as we decrease ξ_i because decreasing ξ_i increases ρ_r, which in turn increases the Hubble rate, causing freeze-out to occur earlier.

4.4 Evolution of perturbations

In this section we describe the evolution of cosmological perturbations during an ECDE with particular focus on the physics underlying the growth in DM density perturbations. We follow the conventions used in Ma

and Bertschinger [104]. We work in conformal Newtonian gauge with metric given by

$$ds^2 = -(1+2\psi)dt^2 + a^2(t)(1-2\phi)dx^2, \tag{4.44}$$

where ϕ and ψ are spatial and temporal metric perturbations respectively. We consider all fluids to be perfect fluids. Thus there is no anisotropic stress, which implies that

$$\psi = \phi. \tag{4.45}$$

In section 4.5 we revisit the perfect-fluid assumption for the DM and cannibal fluids.

Perturbations of perfect fluids can be described by two quantities: $\delta = [\rho(t, x^i) - \bar{\rho}(t)]/\bar{\rho}$, which is the density perturbation over the uniform background density $\bar{\rho}$, and $\theta = \partial_j v^j$, which is the comoving divergence of the physical fluid velocity, $v^j = a\, dx^j/dt$. Our cannibal perturbation equations are similar to those in [82, 89, 96] but we also include leading-order contributions from cannibal decays into radiation, which are derived in appendix B.3. Our suite of perturbation equations is then

$$\delta_c'(a) + (1+w_c)(\frac{\theta_c}{a^2 H} - 3\phi') + \frac{3}{a}\left(1 - \frac{\Gamma}{2H}\right)(c_s^2 - w_c)\delta_c = -\frac{\Gamma}{aH}\phi\left(1 - \frac{3}{2}w_c\right), \tag{4.46}$$

$$\theta_c{}'(a) + \frac{1}{a}(1-3w_c)\theta_c + \frac{w_c'}{1+w_c}\theta_c - \frac{c_s^2}{1+w_c}\frac{k^2}{a^2 H}\delta_c - \frac{k^2}{a^2 H}\phi = \frac{\Gamma}{aH}\theta_c c_s^2, \tag{4.47}$$

$$\delta_{\rm DM}'(a) + \frac{\theta_{\rm DM}}{a^2 H} - 3\phi' = 0, \tag{4.48}$$

$$\theta_{\rm DM}'(a) + \frac{1}{a}\theta_{\rm DM} - \frac{k^2}{a^2 H}\phi = 0, \tag{4.49}$$

$$\delta_r'(a) + \frac{4}{3}\frac{\theta_r}{a^2 H} - 4\phi' = \frac{\Gamma m n_{\rm can}}{aH\rho_r}\left[\phi + \delta_c - \delta_r + \frac{3}{2}\delta_c(w_c - c_s^2)\right], \tag{4.50}$$

$$\theta_r'(a) - \frac{1}{4}\frac{k^2}{a^2 H}\delta_r - \frac{k^2}{a^2 H}\phi = \frac{\Gamma m n_{\rm can}}{aH\rho_r}\left(\frac{3}{4}\theta_c - \theta_r\right), \tag{4.51}$$

$$k^2\phi + 3(aH)^2\left(a\phi' + \phi\right) = -\frac{1}{2}a^2\frac{1}{M_{\rm Pl}^2}(\rho_{\rm can}\delta_c + \rho_r\delta_r + \rho_{\rm DM}\delta_{\rm DM}). \tag{4.52}$$

Here the subscripts c, r and DM corresponds to perturbations of the cannibal, SM radiation and DM fluids respectively, and the prime denotes a derivative with respect to a. We have taken DM to be kinetically decoupled from both the cannibal and radiation fluids, so that its only interactions are gravitational; we will discuss the effects of adding kinetic couplings between cannibal and DM fluids below.

At a_i, when we begin our numerical calculations, the cannibal particles are still relativistic since $T_c(a_i) = 10m$. For adiabatic perturbations, the initial conditions for super horizon modes at a_i are:

$$\delta_r = \frac{4}{3}\delta_{\rm DM} = \delta_c = -2\phi_p \qquad\qquad \theta_r = \theta_{\rm DM} = \theta_c = \frac{1}{2}\frac{k^2}{aH}\phi_p, \tag{4.53}$$

where ϕ_p is the primordial metric perturbation. Adiabatic initial conditions for all fields are naturally obtained in the minimal cosmological scenario where the decays of a single inflaton field populate both the SM and a hidden sector containing the cannibals and DM.[3]

Our primary interest is the evolution of modes that enter the horizon prior to reheating and thus experience

[3]Strictly speaking, these adiabatic initial conditions are applicable to ρ_r as long as energy injection from cannibal decays is negligible at a_i. When instead $\Gamma\rho_{\rm can}/\rho_r \gg H$ at a_i, $\rho_r \propto a^{-2}$, and the initial conditions for the radiation perturbations become $\delta_r = \phi_p/2$ and $\theta_r = \theta_c$.

the epoch of modified cosmic expansion. Before numerically solving the perturbation equations given in eqs. (4.46)-(4.51), we first show how they simplify for modes deep inside the horizon ($k \gg aH$) during the ECDE to gain insight into the essential physics governing the growth of DM density perturbations. Starting with eq. (4.52) for the metric perturbation, we neglect the second term on the LHS of eq. (4.52). We can also ignore $\rho_{\rm DM}\delta_{\rm DM}$ on the RHS of eq. (4.52) because $\rho_{\rm DM}$ is at least seven orders of magnitude smaller than $\rho_{\rm can}$ and ρ_r prior to reheating (see figure 4.3), which must occur before BBN. Consequently, deep inside the horizon and prior to reheating we have

$$\phi = -\frac{3}{2}\left(\frac{aH}{k}\right)^2 \frac{\rho_{\rm can}\delta_c + \rho_r\delta_r}{\rho_{\rm can}+\rho_r}. \tag{4.54}$$

Next we consider the evolution of the cannibal perturbations because they determine the evolution of DM perturbations. We use eq. (4.47) to eliminate θ_c from eq. (4.46). In doing so, we make three approximations. First, we neglect terms proportional to $c_s^2 - w_c, w_c'(a)$ and $d(c_s^2(a))/da$, as w_c and c_s are slowly varying before $a_{\rm fz}$ and rapidly become negligible after $a_{\rm fz}$. Second, we neglect ϕ' in eq. (4.46) because the variation of the metric perturbation is negligible compared to $\theta_c/(aH)$ deep inside the horizon. Third, we neglect terms proportional to Γ/H: before $a_{\rm rh}$ we have $\Gamma/H \ll 1$, and after $a_{\rm rh}$, the cannibal fluid decays and becomes irrelevant. Around $a_{\rm rh}$, when $\Gamma/H \sim \mathcal{O}(1)$, the metric perturbation multiplying Γ in eq. (4.46) is negligible compared to δ_c for modes deep within the horizon, and the sound speed term multiplying Γ in eq. (4.47) is much smaller than one by the time of reheating (see figure 4.4). Finally we eliminate ϕ using eq. (C.1) to obtain

$$\delta_c''(a) + \left[\frac{(a^2H)'}{a^2H} + \frac{1}{a}(1-3w_c)\right]\delta_c' + \frac{1}{a^2}\left(\frac{c_s k}{aH}\right)^2 \delta_c = \frac{3}{2}\frac{(1+w_c)}{a^2}\frac{\rho_r\delta_r + \rho_{\rm can}\delta_c}{\rho_{\rm can}+\rho_r}. \tag{4.55}$$

Naively, eq. (4.55) implies that δ_r may affect δ_c during SM radiation domination, when $\rho_r \gg \rho_{\rm can}$. However, subhorizon radiation perturbation oscillate, and thus their gravitational influence on δ_c is negligible. Consequently one can set $\delta_r = 0$ and rewrite eq. (4.55) in the form

$$\frac{{\rm d}^2\delta_c(a)}{{\rm d}\ln^2(a)} - 3w_c\frac{{\rm d}\delta_c(a)}{{\rm d}\ln(a)} + \left[\left(\frac{c_s k}{aH}\right)^2 - \frac{3}{2}(1+w_c)\frac{\rho_{\rm can}}{\rho_{\rm can}+\rho_r}\right]\delta_c = 0. \tag{4.56}$$

The first term in the square brackets arises from thermal pressure in the cannibal fluid and induces oscillations in the cannibal density perturbation. The second term in the square brackets is inconsequential during SM radiation domination, but during cannibal domination, it induces growth in the cannibal perturbation due to the gravitational attraction between the cannibal particles. When $\rho_{\rm can} \gg \rho_r$, the terms in the square brackets thus determine a Jeans wavenumber, k_J, for the cannibal fluid,

$$k_J \equiv \sqrt{\frac{3}{2}(1+w_c)}\,\frac{aH}{c_s}. \tag{4.57}$$

The corresponding Jeans length scale k_J^{-1} determines when gravitational attraction overcomes the thermal pressure and leads to growth in δ_c.

Earlier we saw that the evolution of both c_s and w_c depend on when the cannibal self interactions freeze-out. Consequently, both the scales a_{fz} and a_{dom} determine the growth in δ_c and we find that the ordering of the two scale produce qualitatively different growth in δ_c. To better understand this qualitative

difference, in the next two subsections we describe the evolution of perturbations for two extreme cases: one where the universe is always cannibal dominated prior to reheating, and one where cannibal comes to dominate the universe much after the cannibal reactions freeze out, $a_{fz} \ll a_{dom} \ll a_{rh}$.

4.4.1 Cannibal freeze-out during cannibal domination

In this section we highlight the key features of perturbations in a scenario with $a_{dom} < a_{fz} < a_{rh}$. To reduce the number of free parameters we focus on the sub-case where universe is already cannibal-dominated at a_i, $\xi_i \gg 1$.

In the top panel of figure 4.5 we plot the evolution of comoving horizon size (solid blue line) and that of the cannibal Jeans scale (solid yellow line) and compare them with the length-scale of Fourier mode, k^{-1} (dashed black line). These scales will determine the evolution of perturbation modes that enter the horizon prior to cannibal freeze-out. In the bottom panel of figure 4.7 we show the evolution of density perturbations for one such mode, indicated in the top panel as the black-dashed line. These results are obtained by numerically solving eqs. (4.46)-(4.51) with initial conditions given by eq. (4.53). The particle parameters were chosen to obtain a large separation between transition scale factors, $a_{rh} \gg a_{fz} \gg a_{can} = 100 a_i$, for pedagogical purposes. Moreover, we show only the evolution of perturbations until shortly after reheating as beyond this time the usual ΛCDM evolution pertains.

The perturbations shown in the bottom panel start to evolve once the mode enters the horizon at $a_{\rm hor}$, defined through

$$a_{\rm hor} H(a_{\rm hor}) \equiv k. \tag{4.58}$$

All density perturbations grow by a factor of 10 to 100 shortly after they enter the horizon. Inside the horizon, thermal pressure causes both the cannibal and SM radiation density perturbations to oscillate, whereas the DM density perturbation approaches a constant value while the cannibal perturbation oscillates during ECDE.

The oscillations in δ_c is due to the non-negligible sound speed of the cannibal (see figure 4.4). In the top panel of figure 4.5, as long as k^{-1} (black dashed line) is inside the Jeans length (yellow shaded region), the cannibal perturbations oscillate in the bottom panel. The Jeans length peaks at roughly $\sim 2a_{fz}$ because cannibal freeze-out is a slow process which starts at a_{fz}, as seen earlier in figures 4.2 and 4.4. After $\sim 2a_{fz}$, the sound speed falls as $c_s^2 \propto 1/a^2$ and the comoving horizon grows as $(aH)^{-1} \propto a^{1/2}$, yielding $k_J^{-1} \propto a^{-1/2}$. Because of the slow decrease of the Jeans length, the cannibal density perturbation continues to oscillate long after a_{fz}. Eventually, once the Jeans length becomes smaller than the scale of the Fourier mode, $k < k_J$, the cannibal density starts to grow linearly as expected.

Using the WKB approximation (details in appendix C) in eq. (4.56) we find the evolution of amplitude, D_c, of δ_c when the mode is inside the Jeans length as

$$D_c(a) \propto \frac{1}{\sqrt{c_s}} a^{-(1-3w_c)/2}. \tag{4.59}$$

We can see that the amplitude is inversely proportional to the frequency of oscillations determined by c_s^2. The dependence on w_c in the exponent of a is associated with the damping caused by the $(1-3w_c)$ factor multiplying δ_c' in eq. (4.56). In the relativistic limit, $w_c = c_s^2 = 1/3$, eq (4.59) recovers the standard result of radiation perturbations oscillating with constant amplitude. In figure 4.5, after entering the Jeans damping

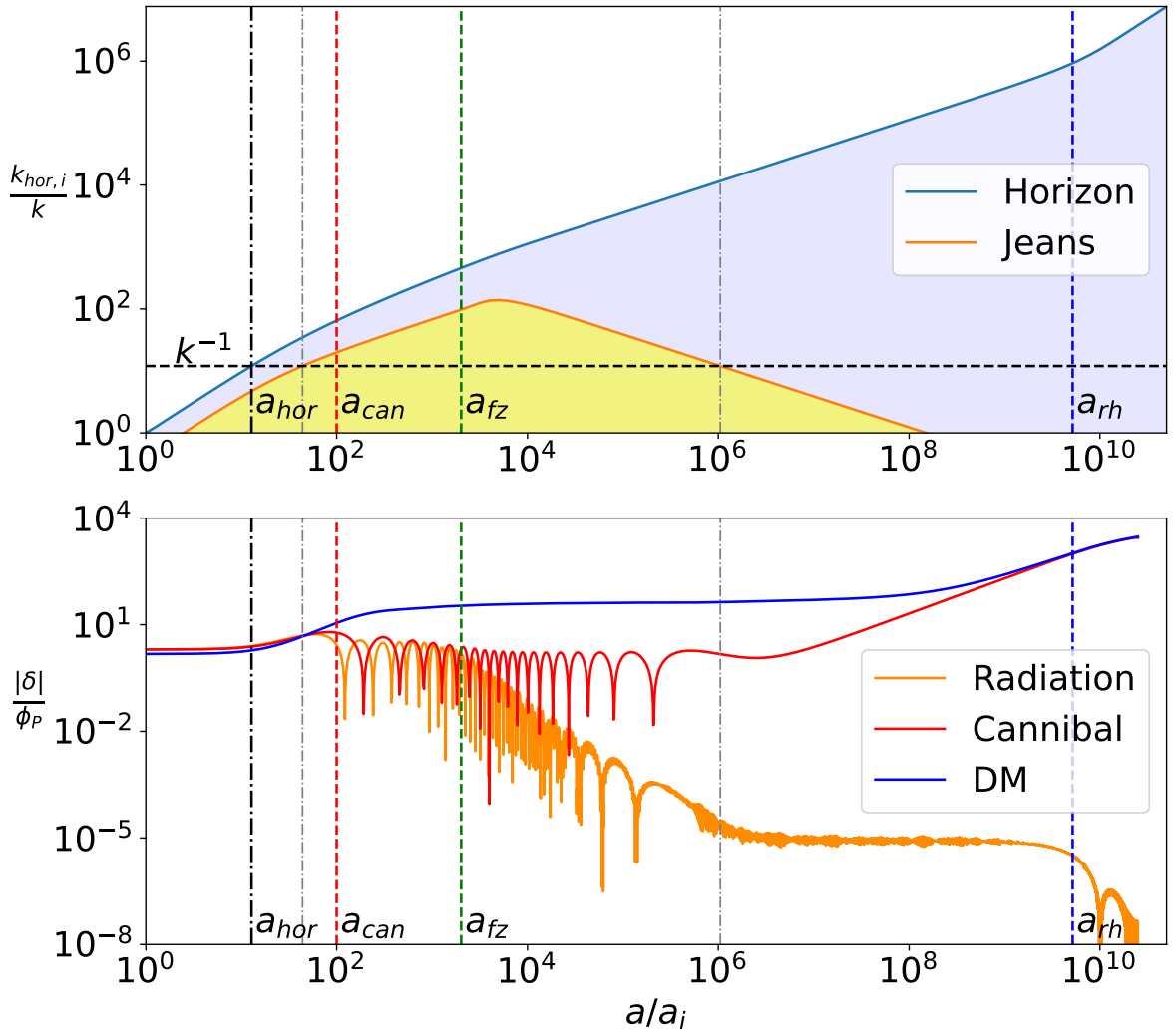


Figure 4.5: **Top**: Evolution of comoving horizon (solid blue line) and cannibal's Jean's damping horizon (solid yellow line) relative to the horizon size at the beginning, $k_{hor,i}^{-1} \simeq 9$ pc. The horizontal black dashed line marks the size of the Fourier mode, $k = k_{hor,i}/12$, for whom perturbations are plotted in the bottom panel. The blue and shaded yellow region highlights when a mode is inside the horizon and the Jeans length, respectively. **Bottom:** Evolution of SM radiation, cannibal and DM density perturbation relative to the primordial metric perturbation, ϕ_p. The vertical dashed red, green and blue lines marks the scale at a_{can}, a_{fz} and a_{rh} respectively. The vertical dot dashed black line marks the scale when mode k enters the horizon at a_{hor}. The vertical dashed grey lines mark the scale factor when the mode enters and exits the Jeans damping scale. The figure is plotted for a scenario with parameters $\xi_i = 500$, $m \simeq 58$ TeV, $T_{rh} = 5$ MeV and $\alpha_c \simeq 0.88$, chosen to give $a_{fz}/a_i \simeq 2 \times 10^3$ and $a_{rh}/a_i = 5 \times 10^9$.

scale but before a_{fz}, we can see that D_c decreases slowly. This slow decrease is due to the logarithmic decay of sound speed and the small value of w_c, which does not completely counter the $a^{-1/2}$ decay in eq (4.59). After a_{fz}, we can neglect w_c in the exponent and the sound speed evolves as $c_s \propto 1/a$, leading to a constant amplitude of oscillations as observed.

The evolution of δ_r seen in the bottom panel of figure 4.7, although interesting, has no significant impact on $\delta_{\rm DM}$. Radiation perturbations have an important gravitational impact on $\delta_{\rm DM}$ only during SM radiation domination. However, during SM radiation domination δ_r oscillates, and hence its gravitational feedback on both $\delta_{\rm DM}$ and δ_c is negligible. The only time δ_r has a significant influence on the other perturbations is near horizon entry ($a \lesssim 10 a_{\rm hor}$) before δ_r starts oscillating.

Finally, we analyze DM density perturbations which is the quantity we are primarily interested in. In the bottom panel of figure 4.5 we see that δ_{DM} (blue line) stagnates after entering the horizon even though the energy density dominating the universe is diluting as matter (approximately). The stagnation of δ_{DM} ca be understood by simplifying the DM perturbation equations in the same way we simplified δ_c in eq. (4.56). Correspondingly we obtain

$$\delta''_{DM}(a) + \Big[\frac{(a^2H)'}{a^2H} + \frac{1}{a}\Big]\delta'_{DM} = \frac{3}{2a^2}\frac{\rho_{can}}{\rho_{can}+\rho_r}\delta_c. \tag{4.60}$$

The term on RHS comes directly from the metric perturbation term in eq. (4.49). Note that when the mode is inside the Jeans length, δ_c oscillates (as does ϕ) and δ_c's time average feedback in the above equation is effectively zero. Assuming the cannibal density dilutes as $1/a^3$, the coefficient of δ'_{DM} equals to $3/(2a)$. As a result, for modes inside Jeans length, δ_{DM} has two solutions: $\delta_{DM} \propto a^{-1/2}$ and $\delta_{DM} \propto a^0$. Consequently we see δ_{DM} stagnating to a constant value in figure 4.5 when δ_c is oscillating. Once the mode escapes the Jeans length, δ_c starts to grow and consequently, DM starts falling into the gravitational potential well formed by δ_c. This condition leads to the growth in δ_{DM}. Once δ_c becomes comparable to δ_{DM}, the DM perturbation grows with $\delta_{DM} = \delta_c \propto a$ as seen in figure 4.5.

Above, we have described in detail the behaviour of perturbations for a mode entering the horizon prior to freeze out of cannibal $3 \to 2$ reactions. Modes that enter the horizon after $\sim a_{fz}$ do not enter the Jeans damping scale and thus δ_c does not experience oscillations. For these modes, δ_c behaves like cold matter perturbation, $w, c_s^2 = 0$, and both δ_{DM} and δ_c grow linearly with a.

Transfer function

We now compare the present-day linear matter power spectrum following an ECDE to the matter power spectrum in standard cosmology. Here, by "standard cosmology" we mean that the universe experienced uninterrupted radiation domination between inflationary reheating and matter-radiation equality, $a_{\rm eq}$. With $\delta_{\rm DM,s}$ denoting the DM density perturbation in the standard cosmology, we define the transfer function

$$T(k) \equiv \frac{\delta_{\rm DM}(k,a)}{\delta_{\rm DM,s}(k,a)}, \tag{4.61}$$

which is evaluated after matter-radiation equality.

After a perturbation mode enters the horizon, $\delta_{\rm DM,s}$ grows logarithmically with scale factor while the

universe is radiation dominated,

$$\delta_{\rm DM,s}(a) = -A_s \phi_p(k) \ln\left(\frac{B_s a}{a_{\rm hor,s}}\right) \tag{4.62}$$

where $A_s = 9.11$ and $B_s = 0.594$ are numerical fitting factors [105], and $a_{\rm hor,s}$ is the scale factor when the mode k enters the horizon in standard cosmology. With H_s denoting the Hubble rate in the standard cosmology, we define $a_{\rm hor,s}$ through

$$a_{\rm hor,s} H_s(a_{\rm hor,s}) \equiv k. \tag{4.63}$$

The logarithmic evolution of $\delta_{\rm DM}$ continues until matter-radiation equality, after which $\delta_{\rm DM}$ grows linearly with scale factor. This evolution is described by the growing solution of the Meszaros equation [105, 106] with initial conditions provided by eq. (4.69),

$$\delta_{\rm DM,s}(a) = -\frac{3A_s\phi_p(k)}{2} \ln\left(\frac{4B_s e^{-3} a_{\rm eq}}{a_{\rm hor,s}}\right) (1 + a/a_{\rm eq})^{0.9} \qquad a > a_{\rm eq}. \tag{4.64}$$

Here the exponent of 0.9 results from the fact that the scales affected by an ECDE are much smaller than the baryon Jeans length [107]. Consequently, $\sim$15% of the matter density does not participate in the gravitational growth, causing the dark matter overdensities to undergo slower than linear growth. The argument of the logarithmic term in eq. (4.64) also obtains $\mathcal{O}(1)$ baryonic corrections as described in [105]. However, we have ignored these corrections because they have an insignificant effect on the final transfer function. The value of ϕ_p in eq. (4.64) is the same as in eq. (4.53) because the universe is radiation dominated at a_i regardless of whether the cannibal fluid or SM radiation is dominant at a_i.

During the radiation-dominated era that follows an ECDE, $\delta_{\rm DM}$ also grows logarithmically with scale factor for subhorizon modes. Consequently, the post-reheating evolution of $\delta_{\rm DM}$ can be described by eq. (4.62), but with A_s and B_s replaced by k-dependent functions $A(k)$ and $B(k)$, which encode the evolution history of $\delta_{\rm DM}$ prior to reheating. After matter-radiation equality, the evolution of $\delta_{\rm DM}$ can similarly be described using eq. (4.64) but with A_s and B_s again replaced by $A(k)$ and $B(k)$. For $k < k_{\rm rh}$, we recover $A(k) = A_s$ and $B(k) = B_s$. It follows that

$$T(k) \equiv \frac{\delta_{\rm DM}(k,a)}{\delta_{\rm DM,s}(k,a)} \approx \frac{A(k)}{A_s} \frac{\ln[4B(k)e^{-3}a_{\rm eq}/a_{\rm hor}(k)]}{\ln[4B_s e^{-3} a_{\rm eq}/a_{\rm hor,s}(k)]}, \tag{4.65}$$

where in the latter equality we have neglected baryonic effects in the logarithm.

In the bottom right panel of figure 4.6 we plot the transfer function for a scenario where $a_{fz}/a_i = 5 \times 10^3$ and $a_{rh}/a_i = 10^8$. To relate the transfer function at a given wavenumber with the background experienced by that Fourier mode as shown in the top left panel of figure 4.6, we plot an inverted transfer function in the top right panel. To increase computational speed while evaluating the transfer function, we ignore radiation perturbations deep inside the horizon, as the feedback of δ_r on $\delta_{\rm DM}$ and δ_c is negligible. We also neglect the heating of the cannibal fluid caused by its decay (see figure 4.4), which has no noticeable impact on $\delta_{\rm DM}$ in this regime.

The transfer function is unity for modes that enter the horizon after reheating because $\delta_{\rm DM}$ and $\delta_{\rm DM,s}$ undergo the same evolution for these modes. As we increase k (i.e. go down in y-axis in the top right panel), the transfer function increases. This is because those modes (see k_1 in top left panel of figure 4.6) enter

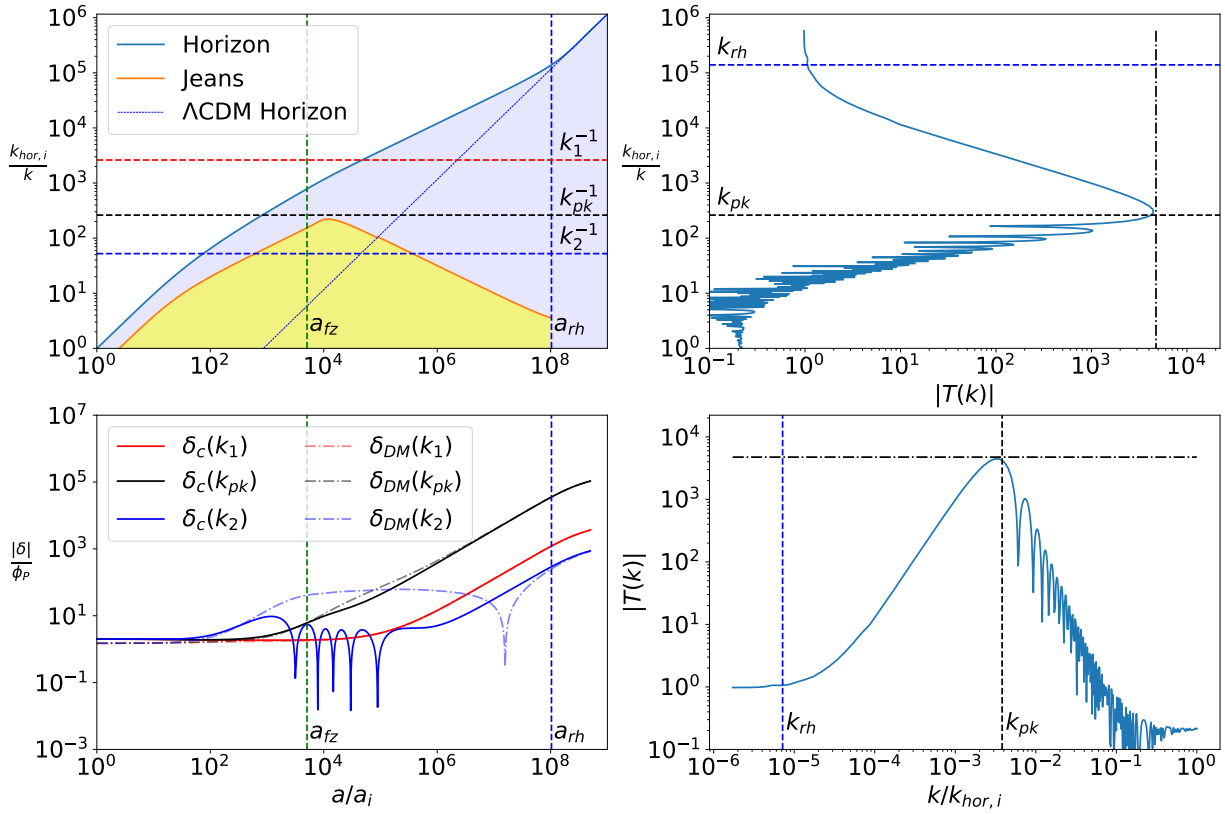


Figure 4.6: **Top left**: Evolution of comoving horizon (solid blue line), cannibal's Jeans length (solid yellow line) and comoving horizon in standard cosmology (dotted blue line) relative to $k_{hor,i}^{-1} \simeq 0.3$ pc. The horizontal dashed line marks the wavenumbers for which we have plotted the evolution of δ_c and δ_{DM} in the bottom panel. We have set $k_1 = k_{pk}/10$ and $k_2 = 5k_{pk}$. **Top right**: The absolute value of the transfer function, eq. (4.65), on the x-axis for every inverse wavenumber k^{-1} shown on the y-axis. The horizontal blue dashed line marks the Fourier mode that enters the horizon at reheating. The horizontal black dashed line marks our analytical estimate of the inverse wavenumber corresponding to the peak of transfer function, eq. (4.66). The vertical black dot-dashed line is our analytical estimate of the size of the peak of transfer function, eq. (4.84). **Bottom left**: Evolution of δ_c (solid line) and δ_{DM} (semi-transparent dashed lines) corresponding to the mode with largest transfer function, k_{pk}, and two other modes. **Bottom right**: Same as top right panel just with axes inverted to show the transfer function on the y-axis and k on the x-axis. This figure is plotted for parameters $\xi_i = 10^2$, $m \simeq 3.2$ TeV, $T_{rh} = 5$ MeV and $\alpha_c \simeq 1.4$.

the horizon when cannibal behaves as cold matter and, correspondingly, the density perturbations grow as $\delta_{DM} = \delta_c \propto a$. This linear growth can be seen in the bottom left panel of figure 4.6 where we plot the evolution of δ_c in solid red and δ_{DM} (which always equals to δ_c in this case) in light dashed-red for the k_1 mode. Since the horizon size grows as $(aH)^{-1} \propto a^{1/2}$ between $a_{\rm dom}$ and $a_{\rm rh}$, $a_{\rm hor} \propto k^{-2}$. Consequently, $\delta_{\rm DM}(a_{\rm rh})$ scales as $\delta(a_{\rm rh}) \propto a_{\rm rh}/a_{\rm hor} \propto k^2$. In the case of $\delta_{\rm DM,s}$, the modes entering the horizon earlier undergo more logarithmic growth between $a_{\rm hor,s}$ and $a_{\rm eq}$ but the same linear growth after $a_{\rm eq}$. Since the horizon size grows as $(aH_s)^{-1} \propto a$ during radiation domination, an increase in k results in a linear decrease in $a_{\rm hor,s}$, which implies that $\delta_{\rm DM,s} \propto \ln[k/(8k_{\rm eq})]$ [105], where $k_{\rm eq} \equiv (aH)_{\rm eq}$. Thus in this regime the transfer function goes as $k^2/\ln[k/(8k_{\rm eq})]$.

As we further increase k, the transfer function continues to grow until the wavenumbers start intersecting the yellow shaded region in the top left panel. The cannibal perturbations for modes entering the Jeans length (k_2 mode in bottom left panel of figure 4.6) oscillates until the mode exits the Jeans length. Consequently, these modes have less time to grow linearly after escaping the Jeans length but before reheating. The oscillation in $T(k)$ is caused by the change in the phase of cannibal perturbation from when the mode enters the horizon to when reheating occurs. Since δ_{DM} eventually falls into the potential well formed by δ_c after the mode escapes the Jeans length (see k_2 mode in bottom left panel), the oscillation of cannibal perturbation is imprinted onto $T(k)$.

Large values of the transfer function imply that the DM perturbations after an ECDE reach the nonlinear regime much earlier than they would in a standard cosmology. Once $\delta_{\rm DM} \gtrsim 1$, overdense fluctuations collapse to form halos [108]. In section 4.6 we discuss how $k_{\rm pk}$ determines the mass of the earliest-forming microhalos, while $T(k_{\rm pk})$ determines the redshift of their formation. Due to the importance of $k_{\rm pk}$ and $T(k_{\rm pk})$ in controlling microhalo formation, in the following subsection we find analytical estimates for both quantities and highlight their connection to cannibal parameters.

The mode corresponding to the peak of the transfer function, k_{pk}, just misses entering the Jeans length as seen in top left panel of figure 4.6. As k_J^{-1} peaks near $2a_{fz}$, the peak wavenumber is given by

$$k_{pk}^{-1} = 1.4(k_J(2a_{fz}))^{-1}, \tag{4.66}$$

where the factor 1.4 we find empirically. This estimate is accurate to within 30% for $a_{\rm rh} = 5a_{\rm fz}$ and reaches 3% accuracy for larger $a_{\rm rh}/a_{\rm fz}$.

In the above scenario we focussed on the sub-case where universe is already cannibal-dominated at a_i. More generally, if the universe was SM-radiation-dominated at a_i but becomes cannibal-dominated when cannibal reaction freezes out, $a_{dom} < a_{fz}$, the qualitative feature of the transfer function is not drastically affected. In particular, $T(k)$ for $k < (aH)_{dom}$ remains completely unaffected as those modes enter the horizon after a_{dom} and are not affected by the period of SM-radiation domination. Since $k_{pk} < (aH)_{dom}$ for $a_{dom} < a_{fz}$, the value of k_{pk} is explained by the same physics as discussed in this sub-section.

4.4.2 Cannibal freeze-out during SM radiation domination

In this section we highlight the key features of perturbations in a scenario where the universe was SM radiation-dominated during the freeze-out of cannibal $3 \to 2$ reactions, $a_{fz} < a_{dom} < a_{rh}$. In the previous subsection we saw that k_{pk} depends sensitively on when the cannibal reactions freeze-out. Consequently, the SM radiation-dominated universe during the freeze out of cannibal reactions significantly changes the behaviour of k_{pk} and $T(k_{pk})$, compared to the results obtained in the previous scenario where $a_{dom} < a_{fz}$.

We can understand the essential behavior of the perturbations in cosmologies with an initially-subdominant cannibal density by separately considering the evolution of perturbation modes that enter the horizon prior to cannibal freeze-out and modes that enter the horizon between cannibal freeze-out and cannibal decay.

Modes that enter the horizon prior to cannibal freeze-out

In the top panel of figure 4.7 we show the comoving horizon, $(aH)^{-1}$ (solid blue), the cannibal Jeans length (solid orange), and the comoving cannibal sound horizon,

$$r_s \equiv \int^t c_s \frac{d\tilde{t}}{\tilde{a}} = \int^a \frac{c_s}{\tilde{a}H} \mathrm{d}\ln(\tilde{a}). \tag{4.67}$$

These scales will determine the evolution of perturbation modes that enter the horizon prior to cannibal freeze-out. In the bottom panel of figure 4.7 we show the evolution of density perturbations for one such mode, indicated in the top panel as the black-dashed line. These results are obtained by numerically solving eqs. (4.46)-(4.51) with initial conditions given by eq. (4.53).

When the wavenumber k is much larger than aH/c_s, the time scale of cannibal oscillations is much smaller than the time scale over which the instantaneous frequency and the anti-damping terms in eq. (4.56) evolve. Thus, one can use a WKB approximation to obtain

$$\delta_c \approx \frac{C_1}{\sqrt{c_s}} \exp\left(-\int_{a_{\mathrm{s,hor}}}^a \frac{1-3w_c}{2} \mathrm{d}\ln(\tilde{a})\right) \sin\left[kr_s + C_2\right], \tag{4.68}$$

as detailed in appendix C. Here $a_{\mathrm{s,hor}}$ is the scale factor for which $c_s k/(aH) = 1$, and C_1 and C_2 are constants determined by $\delta_c(a_{\mathrm{s,hor}})$ and $\delta_c'(a_{\mathrm{s,hor}})$. Note that the instantaneous frequency of δ_c oscillations is set by the cannibal sound horizon.

In the bottom panel of figure 4.7, we see that the amplitude of δ_c oscillations decreases slowly for $a < 2a_{\mathrm{fz}}$ and reaches a constant value for $a > 2a_{\mathrm{fz}}$. The slow decay prior to $2a_{\mathrm{fz}}$ results from the logarithmic decay of c_s partially compensating for the exponential in eq. (4.68). While $a \gtrsim 2a_{\mathrm{fz}}$, c_s and w_c decay as $c_s = c_{s,\mathrm{fz}} a_{\mathrm{fz}}/a$ and $w_c = w_{c,\mathrm{fz}} a_{\mathrm{fz}}^2/a^2$ (see figure 4.4). Inserting this evolution in eq. (4.68), one can check that the amplitude of δ_c remains constant after $2a_{\mathrm{fz}}$.

Once cannibal domination begins, the Jeans length is again the scale that controls the oscillations in δ_c. For instance, the δ_c oscillations in the bottom panel of figure 4.7 end when the mode exits the Jeans horizon in the top panel. The linear growth of δ_c after the mode exits the Jeans horizon can be seen analytically by solving eq. (4.55) while neglecting c_s, w_c, and ρ_r and using the fact that Hubble rate evolves as $H \propto a^{-3/2}$.

Like in previous subsection, the evolution of δ_r seen in the bottom panel of figure 4.7, has no significant impact on δ_{DM} because they oscillate too rapidly within the horizon.

In the bottom panel of figure 4.7 we see that δ_{DM} (blue line) grows logarithmically for $a_{\mathrm{hor}} < a < a_{\mathrm{dom}}$. This is the expected evolution for δ_{DM} in a radiation-dominated universe and is given by

$$\delta_{\mathrm{DM}}(a) = -A_s \phi_p(k) \ln\left(\frac{B_s a}{a_{\mathrm{hor}}}\right) \qquad a_{\mathrm{hor}} < a < a_{\mathrm{dom}}, \tag{4.69}$$

where $A_s = 9.11$ and $B_s = 0.594$ are numerical fitting factors [105]. After the universe becomes cannibal dominated, δ_{DM} is constant until δ_c grows to be of order δ_{DM}, after which δ_{DM} grows linearly as $\delta_{\mathrm{DM}} = \delta_c \propto a$. After reheating, δ_{DM} again grows logarithmically. Consequently, the growth experienced by DM perturbations

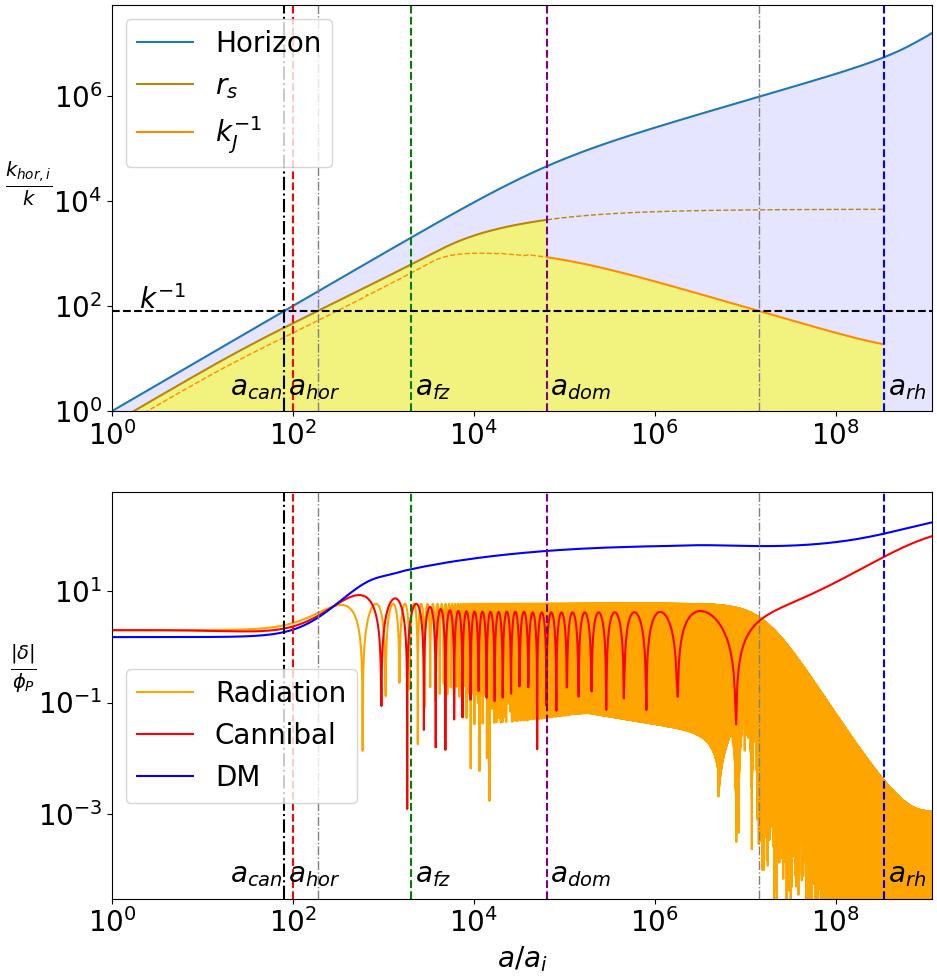


Figure 4.7: **Top**: Comoving horizon (solid blue line), cannibal Jeans length (solid yellow line) and cannibal sound horizon (solid brown line) relative to the horizon size at a_i, with $k_{\rm hor,i}^{-1} \equiv [a_i H(a_i)]^{-1}$. The dashed yellow (brown) line represents when k_J^{-1} (r_s) is plotted in regimes outside of its validity. The horizontal black dashed line marks the Fourier mode, $k = 12.5 \times 10^{-3} k_{\rm hor,i}$, of the perturbations shown in the bottom panel. The shaded yellow region highlights the region within which δ_c oscillates. **Bottom:** Evolution of SM radiation, cannibal and DM density perturbations relative to the primordial metric perturbation, ϕ_p. The vertical dashed red, green, purple, and blue lines marks the scale at $a_{\rm can}$, $a_{\rm fz}$, $a_{\rm dom}$, and $a_{\rm rh}$ respectively. The vertical dot-dashed black line marks the scale $a_{\rm hor}$ when the mode enters the horizon. The vertical dashed grey lines mark the scale factors when the mode enters the sound horizon and exits the Jeans horizon. The figure has been plotted for parameters $\xi_i = 0.68$, $m = 15$ TeV, $T_{\rm rh} = 10$ MeV and $\alpha_c = 1$, for which we obtain $k_{\rm hor,i}^{-1} = 2 \times 10^{-6}$ pc.

during an ECDE is determined by the growth of the cannibal perturbation.

Modes that enter the horizon after cannibal freeze-out

The homogeneous cannibal fluid behaves like pressureless matter after $2a_{\rm fz}$ because $w_c \ll 1$ in this regime. Thus, for modes that enter the horizon after $2a_{\rm fz}$, one might expect δ_c to simply evolve as expected for pressureless matter, i.e., with δ_c growing logarithmically between horizon entry and the end of radiation domination. However, lingering thermal pressure affects the evolution of the cannibal perturbations that enter the horizon shortly after cannibal freeze-out. The sound horizon grows logarithmically for $a_{\rm fz} < a < a_{\rm dom}$, as can be seen analytically by substituting $H \propto 1/a^2$ and $c_s = c_{s,\rm fz} a_{\rm fz}/a$ in eq. (4.67):

$$r_s \sim \frac{c_{s,\rm fz}}{(aH)_{\rm fz}} \ln\left(\frac{a}{2a_{\rm fz}}\right) \qquad\qquad 2a_{\rm fz} < a < a_{\rm dom}. \tag{4.70}$$

This logarithmic growth of r_s is also evident in figure 4.7 and in the top right panel of figure 4.8. Since the sound horizon continues to grow while the universe is radiation dominated, modes that enter the horizon after cannibal freeze-out may still oscillate. For example, in the bottom left panel of figure 4.8 we can see that $\delta_c(k_2)$ undergoes oscillations once k_2^{-1} enters the sound horizon in the top left panel.

To better understand the evolution of δ_c for modes entering the horizon between $2a_{\rm fz}$ and $a_{\rm dom}$, we solve eq. (4.56). For $a > a_{\rm fz}$, the anti-damping term in eq. (4.56) rapidly decays while the frequency remains constant. Moreover, since w_c is already much less than one by $a_{\rm fz}$, the anti-damping term never has a significant impact, as we have verified numerically. Consequently, for $a > a_{\rm fz}$, eq. (4.56) simplifies to a simple harmonic oscillator equation in $\ln(a)$. Using $c_s = c_{s,\rm fz} a_{\rm fz}/a$ and $H = H(a_{\rm fz}) a_{\rm fz}^2/a^2$ for $a > a_{\rm fz}$, we can exactly solve this simple harmonic oscillator equation to obtain

$$\delta_c = \tilde{C}_1 \sin\left(\frac{c_{s,\rm fz} k}{a_{\rm fz} H(a_{\rm fz})} \ln\left(\tilde{C}_2 \frac{a}{a_{\rm hor}}\right)\right) \qquad\qquad a_{\rm fz}, a_{\rm hor} < a < a_{\rm dom}, \tag{4.71}$$

where $\tilde{C}_1$ and $\tilde{C}_2$ are constants.

Eq. (4.71) suggests that the cannibal perturbation evolves logarithmically with a for a short time after horizon entry. Since $c_{s,\rm fz} k/(aH)_{\rm fz} \ll 1$ for modes entering the horizon after $a_{\rm fz}$, eq. (4.71) simplifies to

$$\delta_c \approx \tilde{C}_1 \frac{c_{s,\rm fz} k}{(aH)_{\rm fz}} \ln\left(\tilde{C}_2 \frac{a}{a_{\rm hor}}\right), \tag{4.72}$$

while a does not greatly exceed $a_{\rm hor}$. Hence the naive expectation that δ_c should evolve logarithmically for modes entering the horizon after cannibal freeze-out does hold for a brief period after horizon entry. During this period of logarithmic growth, the influence of thermal pressure is initially negligible. However, the influence of thermal pressure keeps growing logarithmically until it becomes large enough that the argument of the sine becomes $\mathcal{O}(1)$ and δ_c begins to oscillate.

We can find $\tilde{C}_1$ and $\tilde{C}_2$ for modes entering the horizon after $2a_{\rm fz}$ by using the fact that the super horizon initial condition is the same for both the cannibal and DM perturbations in this regime because the cannibal particles are non-relativistic. Consequently, the early-time solution in eq. (4.72) should match the standard logarithmic growth of DM during radiation domination in eq. (4.69). By matching eq. (4.72) to eq. (4.69) we find the constants to be $\tilde{C}_1 = -A_s \phi_p (aH)_{\rm fz}/(c_{s,\rm fz} k)$ and $\tilde{C}_2 = B_s$. It follows that the cannibal perturbation

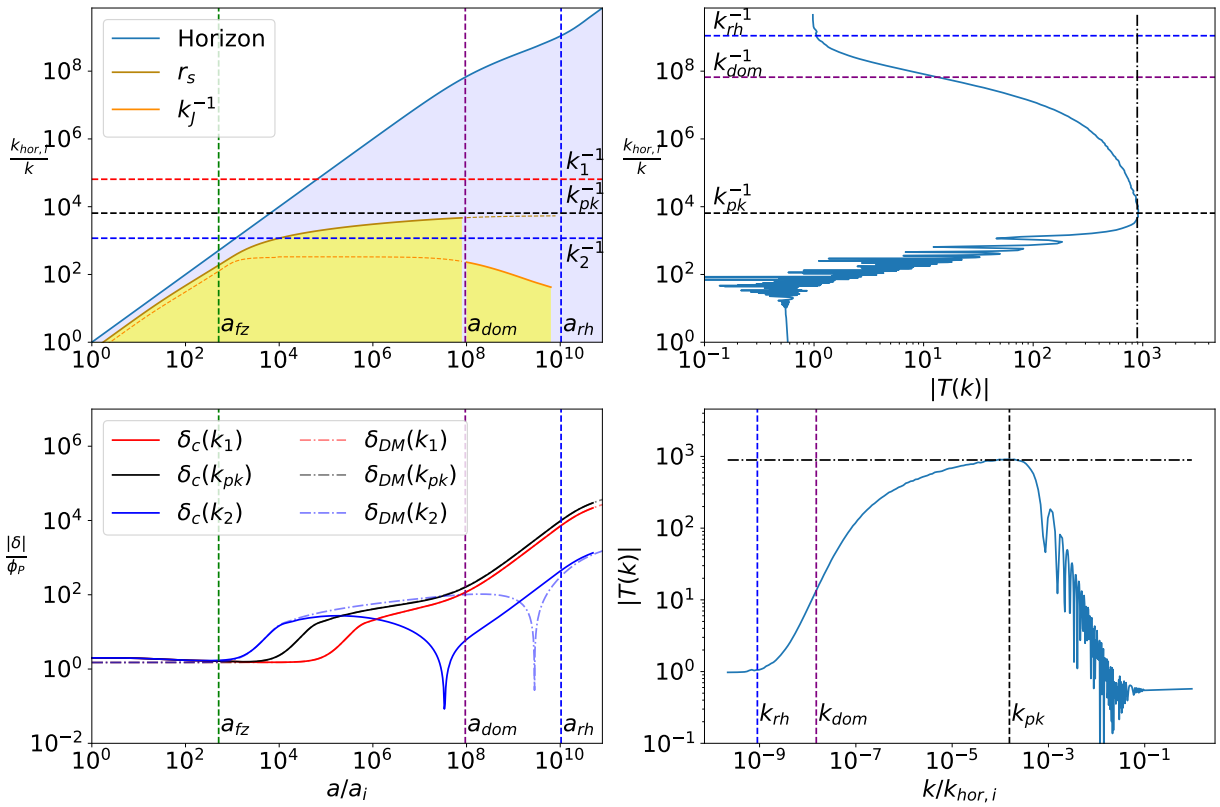


Figure 4.8: **Top left**: Comoving horizon (blue solid line), sound horizon (brown solid line) and Jeans length (yellow solid line) relative to $k_{\rm hor,i}^{-1} \equiv [a_i H(a_i)]^{-1}$. Horizontal dashed lines mark the wavenumbers corresponding to the perturbations in the bottom panel. **Top right**: Absolute value of the transfer function, eq. (4.65), on the x-axis, for every inverse wavenumber k^{-1} shown on the y-axis. The horizontal blue (purple) dashed line marks the wavenumber that enters the horizon at $a_{\rm rh}$ ($a_{\rm dom}$). The horizontal black dashed line marks the analytical estimate of the wavenumber $k_{\rm pk}$ that maximizes $\delta_{\rm DM}/\phi_P$, eq. (4.78). The vertical black dot-dashed line is our analytical estimate of the the transfer function at horizontal black dashed line, eq. (4.88). **Bottom left**: Evolution of δ_c (solid) and $\delta_{\rm DM}$ (light dot-dashed) for $k_{\rm pk}$ and two other modes $k_1 = k_{\rm pk}/10$ and $k_2 = 4k_{\rm pk}$. **Bottom right**: Transfer function as in the top right panel but rotated by ninety degrees. In this figure we take $\xi_i = 0.1$, $m = 180$ TeV, $T_{\rm rh} = 10$ MeV and $\alpha_c = 1$.

evolution within the horizon is given by

$$\delta_c(a) = -A_s \phi_p \frac{(aH)_{\rm fz}}{c_{s,\rm fz} k} \sin\left(\frac{c_{s,\rm fz} k}{(aH)_{\rm fz}} \ln\left(B_s \frac{a}{a_{\rm hor}}\right)\right) \qquad a_{\rm hor} < a < a_{\rm dom}. \tag{4.73}$$

The fact that the cannibal and DM perturbations follow the same initial evolution after horizon entry can be seen in the bottom left panel of figure 4.8.

Note that the argument of the sine in eq. (4.73) is similar to kr_s. By using the approximate evolution of r_s given by eq. (4.70), we find the difference between kr_s and the argument of the sine in eq. (4.73) to be $\sim c_{s,\rm fz} k/(aH)_{\rm fz} \ln(a_{\rm hor}/2a_{\rm fz})$. This difference is much less than one for $a_{\rm hor} > 2a_{\rm fz}$ because $c_{s,\rm fz} \ll 1$ and $k/(aH)_{\rm fz} = a_{\rm fz}/a_{\rm hor} < 1$, which follows from the fact that $H \propto a^{-2}$ between $a_{\rm fz}$ and $a_{\rm hor}$. Thus δ_c deviates from the logarithmic growth experienced by $\delta_{\rm DM}$ approximately when the mode enters the sound horizon. This can also be seen in the left panels of figure 4.8, where the intersection of k_2^{-1} and r_s in the top panel coincides with $\delta_c(k_2)$ beginning to deviate from $\delta_{\rm DM}(k_2)$ in the bottom panel.

The dark matter density perturbation evolves in the same manner as in the previous subsection: it grows logarithmically between $a_{\rm hor}$ and $a_{\rm dom}$, after which it remains constant until δ_c grows to be of order $\delta_{\rm DM}$, and then it grows linearly along with the cannibal perturbation, $\delta_{\rm DM} = \delta_c \propto a$. Note that this linear growth occurs independently of whether the mode enters the horizon before or after $a_{\rm fz}$, as long as the mode is outside the cannibal Jeans horizon during cannibal domination.

So far we have discussed modes that enter the horizon prior to cannibal domination. For modes that enter the horizon after $a_{\rm dom}$, δ_c evolves as cold matter, since $k^{-1} \gg r_s(a_{\rm dom})$. Thus, the evolution of perturbations for modes that enter the horizon after $a_{\rm dom}$ is the same as those studied in early matter-dominated eras [18, 24, 76–78].

The linear matter power spectrum after an ECDE

We now compare the present-day linear matter power spectrum following an ECDE to the matter power spectrum in standard cosmology. In the bottom right panel of figure 4.8 we plot the transfer function for a scenario with $a_{\rm dom} \gg a_{\rm fz}$.

The transfer function is unity for modes that enter the horizon after reheating because $\delta_{\rm DM}$ and $\delta_{\rm DM,s}$ undergo the same evolution for these modes. As we increase k, the transfer function increases as approximately $k^2/\ln[k/(8k_{\rm eq})]$ until $k \sim k_{\rm dom} \equiv (aH)_{\rm dom}$ because these modes experience evolution similar to modes in an early matter dominated universe.

For $k > k_{\rm dom}$, modes enter the horizon during SM radiation domination. Modes with larger k now see $\delta_{\rm DM}$ undergo a larger logarithmic growth between $a_{\rm hor}$ and $a_{\rm dom}$ but the same linear growth between $a_{\rm dom}$ and $a_{\rm rh}$ (seen, for example, in the differing growth of $\delta_{\rm DM}(k_1)$ and $\delta_{\rm DM}(k_{\rm pk})$ in the bottom left panel of fig. 4.8). Consequently, an increase in k results in a $\sim \ln[k/(8k_{\rm dom})]$ increase in the growth experienced by $\delta_{\rm DM}$. Thus the transfer function increases as approximately $\ln[k/(8k_{\rm dom})]/\ln[k/(8k_{\rm eq})]$, and as $k_{\rm dom} \gg k_{\rm eq}$, the k-dependence of the transfer function is primarily driven by the logarithm in the numerator.

As we further increase k beyond $k_{\rm dom}$ in the top right panel of figure 4.8, the transfer function continues to grow until k^{-1} intersects the cannibal sound horizon in the top left panel. For these modes, the cannibal perturbations undergo oscillations while their wavelength is contained within the yellow shaded region in the top left panel, which inhibits the growth of DM perturbations. Modes with larger k spend more time within the Jeans horizon, and thus have less time to grow prior to reheating. Since the Jeans length decays as $a^{-1/2}$ for $a > a_{\rm fz}$, the envelope of the transfer function for $k > k_{\rm pk}$ falls as $k^{-2}/\ln[k/(8k_{\rm eq})]$. The oscillations

in $T(k)$ are caused by changes in the phase of the cannibal perturbation between horizon entry and Jeans horizon exit; the dark matter inherits these oscillations because it falls into the gravitational wells generated by the cannibal perturbations after they stop oscillating [26]. When δ_c keeps oscillating until $a_{\rm rh}$, however, the cannibal mode does not have a net gravitational impact on $\delta_{\rm DM}$, and thus for $k > k_J(a_{\rm rh})$ the oscillations in $T(k)$ stop.

4.4.3 Analytical estimate of the peak of the matter power spectrum

In this section we find analytical relationship between the key features in the small scale matter power spectrum, $\{k_{\rm pk}, T(k_{\rm pk})\}$ and the key cosmological transitions scales in ECDE, $\{a_{fz}, a_{dom}, a_{rh}\}$.

Estimating k_{pk}

Earlier in section 4.4.1 we found that peak in the transfer function occurs at $k_{\rm pk}^{-1} = 1.4 k_J^{-1}(2a_{\rm fz})$ when the cannibal reactions freeze-out during cannibal domination, $a_{\rm dom} < 2a_{\rm fz}$. The k_{pk} mode here corresponded to the smallest-scale mode for which the cannibal perturbation did not oscillate. Similarly, if instead cannibal reactions freeze out while the universe is SM dominated, then the matter power spectrum peaks near the smallest-scale mode whose cannibal perturbation do not oscillate. This occurs for a mode that is larger than the cannibal sound horizon, i.e.,

$$k_{\rm pk} \approx r_s^{-1}(a_{\rm dom}). \tag{4.74}$$

We can more accurately determine $k_{\rm pk}$ by using the fact that $\delta_{\rm DM}(a_{\rm rh}) = \delta_c(a_{\rm rh})$ for wavenumbers in the vicinity of $k_{\rm pk}$. For these wavenumbers, δ_c and $\delta_{\rm DM}$ undergo the same amount of linear growth between $a_{\rm dom}$ and $a_{\rm rh}$, as illustrated in the bottom left panel of figure 4.8. Consequently, we can find $k_{\rm pk}$ by finding the wavenumber that maximizes $\delta_c(a_{\rm dom})$.

As the top left panel of figure 4.8 demonstrates, wavenumbers with $k \sim r_s^{-1}(a_{\rm dom})$ enter the horizon between $a_{\rm fz}$ and $a_{\rm dom}$. We can estimate $\delta_c(a_{\rm dom})$ for these modes using the analytical solution for δ_c given in eq. (4.73). Strictly speaking, the approximations yielding eq. (4.73) do not include the gradual transition to cannibal domination around $a_{\rm dom}$, but as the impact of this transition is similar for all modes with k between $k_{\rm dom}$ and $k_{\rm fz}$, these neglected terms will not affect the determination of $k_{\rm pk}$.

In eq. (4.73), some of the k-dependence is hidden inside $a_{\rm hor}$. This dependence can be made explicit by using the fact that the modes near $k_{\rm pk}$ enter the horizon during radiation domination, yielding $k/(aH)_{\rm fz} = (aH)_{\rm hor}/(aH)_{\rm fz} = a_{\rm fz}/a_{\rm hor}$. Expressing k in terms of $a_{\rm hor}$ and defining $\gamma \equiv a_{\rm hor}/a_{\rm fz}$, eq. (4.73) becomes

$$\delta_c(a_{\rm dom}) \approx -A_s \phi_p \frac{\gamma}{c_{s,\rm fz}} \sin\left(\frac{c_{s,\rm fz}}{\gamma} \ln\left(\frac{B_s}{\gamma}\frac{a_{\rm dom}}{a_{\rm fz}}\right)\right). \tag{4.75}$$

Apart from the weak k-dependence in ϕ_p ($\phi_p \propto k^{-0.02}$) [3], the rest of the k dependence is now encoded inside γ. We maximize $\delta_c(a_{\rm dom})$ by taking ϕ_p to be constant and setting the derivative of $\delta_c(a_{\rm dom})$ with respect to γ to zero. Doing so yields

$$\tan\left[\frac{c_{s,\rm fz}}{\gamma} \ln\left(\frac{B_s}{\gamma}\frac{a_{\rm dom}}{a_{\rm fz}}\right)\right] = \frac{c_{s,\rm fz}}{\gamma} + \frac{c_{s,\rm fz}}{\gamma} \ln\left(\frac{B_s}{c_{s,\rm fz}\gamma}\frac{a_{\rm dom}}{a_{\rm fz}}\right). \tag{4.76}$$

This equation has multiple solutions. We want the largest γ for which the above equation is satisfied, because the largest solution corresponds to the value of $a_{\rm hor}$ for which δ_c does not oscillate. Since oscillations suppress

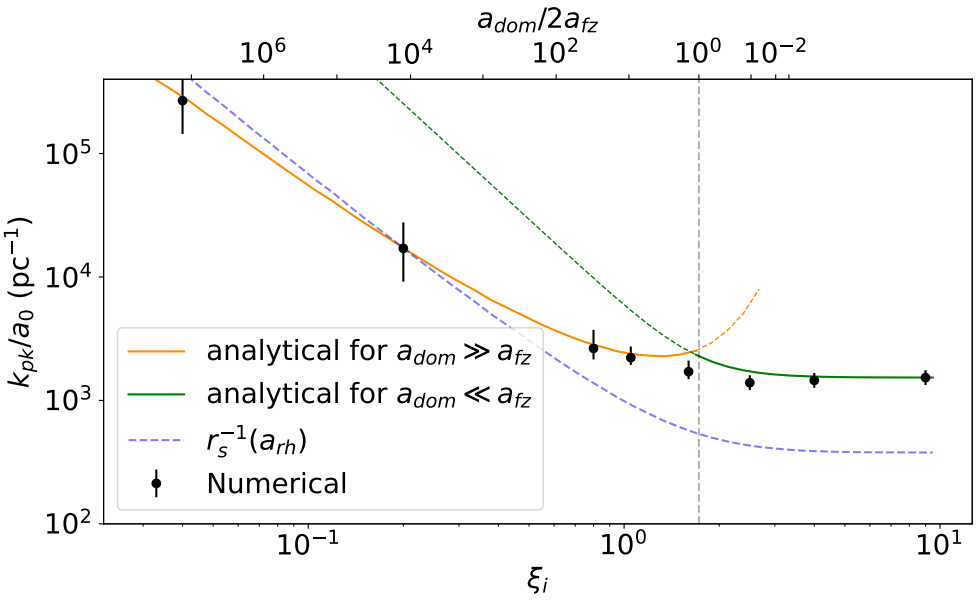


Figure 4.9: Wavenumber corresponding to the peak of the matter power spectrum, $k_{\rm pk}$, as a function of initial temperature ratio between cannibals and SM radiation, $\xi_i = T_c(a_i)/T_r(a_i)$. The yellow (green) solid line shows the estimate of $k_{\rm pk}$ given in eq. (4.78) (eq. (4.66)); the dashed yellow and green dashed lines show extrapolations of these analytical estimates outside their regime of validity. Black dots are the numerically obtained values of $k_{\rm pk}$. The bars on the black dots show the range of k around $k_{\rm pk}$ for which $\delta_{\rm DM}(k, a_{\rm eq})$ is within 5% of $\delta_{\rm DM}(k_{\rm pk}, a_{\rm eq})$. The blue dashed line is the inverse of the cannibal sound horizon (see eq. (4.67)) at $a_{\rm rh}$. The vertical gray dashed line marks the point at which $a_{\rm dom} = 2a_{\rm fz}$. In the secondary x-axis on top, we show the values of $a_{\rm dom}/(2a_{\rm fz})$. For $\xi_i \gtrsim 3.2$, the SM radiation density is subdominant to the cannibal density until reheating, and $a_{\rm dom}$ is not defined. In this figure we take $m = 300$ TeV, $\alpha_c = 0.2$ and $T_{\rm rh} = 8$ MeV. We have restricted the plot to $\xi_i > 0.03$ because for smaller ξ_i the cannibal density never exceeds SM density.

$\delta_c(a_{\rm dom})$, this mode is the global maximum of $\delta_c(a_{\rm dom})$. For the largest γ satisfying the above equation, the tangent is well-approximated by a second-order Taylor expansion. After simplifying the resulting equation we obtain

$$\gamma_{\rm pk} = \frac{3c_{s,\rm fz}}{2\sqrt{2}} W^{3/2}\left[2\left(\frac{B_s}{3c_{s,\rm fz}}\frac{a_{\rm dom}}{a_{\rm fz}}\right)^{2/3}\right], \tag{4.77}$$

where W is the Lambert function. Using $\gamma = a_{\rm hor}/a_{\rm fz} = (aH)_{\rm fz}/k$, we obtain

$$k_{\rm pk}^{-1} = \frac{\gamma_{\rm pk}}{(aH)_{\rm fz}}. \tag{4.78}$$

In the right panels of figure 4.8, this estimate of $k_{\rm pk}$ is shown with a black dashed line.

In figure 4.9, we plot the estimates of $k_{\rm pk}$ given in eqs. (4.78) and (4.66) as functions of ξ_i. We also show the values of k that maximize $\delta_{\rm DM}(k, a_{\rm eq})/\phi_P$ as black dots. Since ϕ_P is only weakly scale-dependent, the peak location in $\delta_{\rm DM}(k, a_{\rm eq})/\phi_P$ is an excellent estimate of the peak location in the matter power spectrum. The bars around the black dots indicate the range of k for which $\delta_{\rm DM}(k, a_{\rm eq})/\phi_P$ is within 5% of its maximum value. This range is larger for $a_{\rm dom} < 2a_{\rm fz}$ because the logarithmic evolution of δ_c prior to $a_{\rm dom}$ causes the transfer function to be flatter near the peak (see figure 4.8). Our analytical estimates are accurate

approximations to the numerical results except in the transition region where $a_{\rm dom} \sim 2a_{\rm fz}$. In figure 4.9, we also plot the cannibal sound horizon at reheating, $r_s(a_h)$. We plot $r_s(a_{\rm rh})$ and not $r_s(a_{\rm dom})$ because $r_s(a_{\rm dom})$ provides an approximate estimate of $k_{\rm pk}$ only for $a_{\rm dom} > 2a_{\rm fz}$, while $r_s(a_{\rm rh})$ can provide an order-of-magnitude estimate of $k_{\rm pk}$ for all values of ξ_i.

To understand how $k_{\rm pk}$ depends on ξ_i, we need to express $k_{\rm pk}$ in terms of cannibal parameters: m, α_c, $T_{\rm rh}$, and ξ_i. For $a_{dom} < 2a_{fz}$, we can find a rough dependence of k_{pk} on particle parameters by using eq. (4.57) to estimate $k_J(2a_{fz})$ and then performing the following steps to simplify eq. (4.57): 1) approximate $c_s^2 \approx T_c/m$ and neglect w_c, 2) use eq. (4.39) to substitute ρ_{can} residing inside H, and 3) use (4.10) to substitute T_c. Performing the above steps gives us the relation $k_{pk} \sim 0.3\sqrt{a_i/a_{fz}}k_{hor,i}$, where $k_{hor,i} = a_i H(a_i)$. We can get an estimate of $k_{\rm hor,i}$ as a function of our cannibal parameters by splitting $a_i/a_0 = a_i/a_{\rm rh} \times a_{\rm rh}/a_0$, estimating $a_{\rm rh}/a_0$ using eq. (4.35) and $a_{\rm rh}/a_i$ using eq. (4.40), and using the initial densities of SM radiation and cannibals in the Hubble rate to obtain

$$\frac{k_{\rm hor,i}}{a_0} \sim 34.5\ {\rm pc}^{-1} \times \left(\frac{T_{\rm rh}}{10\,{\rm MeV}}\right)^{1/3} \left(\frac{m}{\rm GeV}\right)^{2/3} \sqrt{[1 + g_*(10m/\xi_i)\xi_i^{-4}]}. \tag{4.79}$$

Finally, by replacing a_{fz} using eq. (4.42) we obtain

$$k_{pk} \sim 0.05 \times \alpha_c^{-1/3} \left(\frac{m}{\rm GeV}\right)^{7/9} \left(\frac{T_{rh}}{10{\rm MeV}}\right)^{1/3} {\rm pc}^{-1}. \tag{4.80}$$

As we can see k_{pk} has no ξ_i dependence for $a_{dom} < 2a_{fz}$ case.

In the case where $2a_{fz} < a_{dom}$, we first use $\gamma_{\rm pk} \approx \ln^{3/2}(a_{\rm dom}/a_{\rm fz})/3.5$ in eq. (4.78). This approximation for $\gamma_{\rm pk}$ is an empirical relation that we found to be accurate to within 20% for $10 < a_{\rm dom}/a_{\rm fz} < 10^5$ and $10^2 < a_{\rm fz}/a_i < 10^4$. Furthermore, we use $(aH)_{\rm fz} = k_{\rm hor,i}a_i/a_{\rm fz}$ because the universe is radiation dominated between a_i and $a_{\rm fz}$. Finally, expressing $a_{\rm fz}$ and $a_{\rm dom}$ in terms of cannibal parameters using eq. (4.43) and eq. (4.41) respectively, we obtain for $a_{\rm dom} > 2a_{\rm fz}$

$$\frac{k_{\rm pk}}{a_0} \sim 3.7 \times 10^{-2}\ {\rm pc}^{-1}\alpha_c^{-3/4}\xi_i^{-5/2} \left(\frac{m}{\rm GeV}\right)^{11/12} \left(\frac{T_{\rm rh}}{10\,{\rm MeV}}\right)^{1/3} \left(\frac{g_*(10m/\xi_i)}{100}\right)^{5/8} \times \ln^{-3/2}\left(\frac{1}{4}\alpha_c^{-3/4}\xi_i^{-9/2} \left(\frac{m}{\rm GeV}\right)^{1/4} \left(\frac{g_*(10m/\xi_i)}{100}\right)^{9/8}\right) {\rm pc}^{-1}. \tag{4.81}$$

The location of the peak in the transfer function thus depends on all four cannibal parameters, but it is most sensitive to ξ_i, which is in contrast to eq. (4.80).

While our analytic estimates of $k_{\rm pk}$ in the scenarios with $a_{\rm dom} < 2a_{\rm fz}$ and $a_{\rm dom} > 2a_{\rm fz}$ are given by different functions, the underlying scale determining $k_{\rm pk}$ in both cases is the cannibal sound horizon, up to an order of magnitude. For $a_{\rm dom} > 2a_{\rm fz}$ we have already shown that $k_{\rm pk}^{-1} \approx r_s(a_{\rm dom})$. In the case of $a_{\rm dom} < 2a_{\rm fz}$ one can show that $r_s(2a_{\rm fz})$ is within an $\mathcal{O}(1)$ factor of $k_J^{-1}(2a_{\rm fz}) \approx k_{\rm pk}^{-1}$. Moreover, after cannibals freeze out and dominate the universe, r_s asymptotes to a constant value (see figure 4.7 and 4.8) as seen by inserting $H \propto a^{-3/2}$ and $c_s \propto 1/a$ in the definition of r_s, eq. (4.67). Consequently, the total comoving distance traveled by the sound waves in the cannibal fluid, $r_s(a_{\rm rh})$, provides an order-of-magnitude estimate of $k_{\rm pk}$ for all scenarios. In figure 4.9, we compare the dependence of $r_s^{-1}(a_{\rm rh})$ on ξ_i to that of $k_{\rm pk}$. The value of $r_s^{-1}(a_{\rm rh})$ always falls within a factor of 5 from $k_{\rm pk}$ in our parameter space.

Estimating $T(k_{pk})$

We now semi-analytically estimate the value of $T(k_{pk})$.[4] Since DM particles are falling into the gravitational potential formed by cannibal, the value of the DM perturbation at the end of reheating is approximately the same as that of the cannibal's, $\delta_{DM}(a_{rh}) \approx \delta_c(a_{rh})$. Consequently, we can estimate the peak of the transfer function by first estimating the value of cannibal perturbation for k_{pk}.

First we solve for the case where $a_{dom} < 2a_{fz}$. Since for this scenario, the cannibal density perturbations with $k < k_J$ grow linearly after horizon entry, and the peak wavenumber $k_{\rm pk}$ enters the horizon at a scale factor $a_{\rm hor,pk} \propto a_{\rm fz}$, we expect $\delta_{\rm c}(k_{\rm pk}, a_{\rm rh}) \propto \phi_p(k_{\rm pk}) a_{\rm rh}/a_{\rm fz}$. This relation, particularly the assumed linear scaling of $a_{\rm hor,pk} = k_{\rm pk}/H(a_{\rm hor,pk})$ with $a_{\rm fz}$, does receive logarithmic corrections owing to the cannibal interactions, but the logarithmic nature of those corrections implies that the linear scaling provides a useful estimate. After reheating, we expect the DM perturbation to grow logarithmically. Thus we expect that the behavior of $\delta_{\rm DM}(k_{\rm pk})$ during the post-reheating epoch of radiation domination can be parametrized as

$$\delta_{\rm DM}(k_{\rm pk}, a > a_{\rm rh}) = -b_1 \frac{a_{\rm rh}}{a_{\rm fz}} \phi_p(k_{\rm pk}) \ln\left(b_2 \frac{a}{a_{\rm rh}}\right), \tag{4.82}$$

where b_1 and b_2 reflect, respectively, the deviation of $\delta_{\rm DM}(k_{\rm pk})$ from perfect linear growth between $a_{\rm rh}$ and $a_{\rm fz}$, and the transition of $\delta_{\rm DM}$ from linear to logarithmic growth around $a_{\rm rh}$. Empirically we find that both b_1 and b_2 are $\mathcal{O}(1)$; for instance, in the parameter point shown in Fig. 4.6, $b_1 = 2.5$ and $b_2 = 1.7$. Both b_1 and b_2 are insensitive to variations in $a_{\rm rh}/a_{\rm fz}$ as long as $\delta_{\rm DM}$ achieves linear growth by the time of reheating, or equivalently $a_{\rm rh}/a_{\rm fz} \gtrsim 200$. However, while b_2 is insensitive to variations in $a_{\rm fz}/a_i$, b_1 has a logarithmic dependence on $a_{\rm fz}/a_i$ arising from the logarithmic corrections to the assumed proportionality $a_{\rm hor,pk} \propto a_{\rm fz}$.

Comparing Eq. (4.82) with Eq. (4.69) lets us identify

$$A(k_{\rm pk}) = b_1 \frac{a_{\rm rh}}{a_{\rm fz}} \qquad\qquad B(k_{\rm pk}) = b_2 \frac{a_{\rm hor,pk}}{a_{\rm rh}}. \tag{4.83}$$

Using these relations of $A(k_{\rm pk})$ and $B(k_{\rm pk})$ in the transfer function, Eq. (4.65), yields

$$T(k_{\rm pk}) = \frac{b_1}{A_{\rm s}} \frac{a_{\rm rh}}{a_{\rm fz}} \left[1 - \frac{\ln(B_{\rm s} b_2^{-1} a_{\rm rh}/a_{\rm hor,s})}{\ln(4B_{\rm s} e^{-3} a_{\rm eq}/a_{\rm hor,s})}\right] \sim \frac{1}{5} \frac{a_{\rm rh}}{a_{\rm fz}}, \tag{4.84}$$

where in the second relation we dropped the logarithmic factors (since $a_{\rm rh} \ll a_{\rm eq}$) and estimated $b_1 \sim 2$. Thus the peak of the transfer function is roughly proportional to the decades between the freeze out of $3 \to 2$ reactions and reheating.

The semi-analytical estimate of the peak of the transfer function given in the first equality is found to be accurate with the numerical value (less than a 1% deviation) for different values of a_{rh} as long as $a_{rh} \gg a_{fz}$. However, the semi-analytical estimate has a small variation of order $\sim 20\%$ with respect to its numerical counterpart as we change a_{fz} by a factor of order 10. This semi-analytical estimate is also shown in the right panels of figure 4.6 as black dot-dashed lines.

Next we find an analytic estimate of $T(k_{\rm pk})$ for scenarios with $a_{\rm dom} > 2a_{\rm fz}$. Again, we start by first estimating the value of $\delta_c(a_{\rm rh})$. Since the mode with wavenumber $k_{\rm pk}$ typically remains outside of the cannibal sound horizon, $\delta_c(k_{\rm pk})$ evolves similarly to a cold matter perturbation. That is, δ_c evolves logarithmically from horizon entry at $a_{\rm hor} = \gamma_{\rm pk} a_{\rm fz}$ until $a_{\rm dom}$. The linear growth of $\delta_c(k_{\rm pk})$ after cannibal domination is

[4] Due to the presence of $\delta_{\rm DM,s}$ in the denominator of the transfer function (eq (4.65)), the location of the peak in the transfer function is slightly different from $k_{\rm pk}$. However, this difference is negligible, as can be seen in figure 4.8.

well described by the growing solution of the Meszaros equation:

$$\delta_c(k_{\rm pk}, a) \approx \frac{3A_s\phi_p}{2} \ln\left[\frac{4B_s e^{-3} a_{\rm dom}}{\gamma_{\rm pk} a_{\rm fz}}\right] \frac{a}{a_{\rm dom}}. \tag{4.85}$$

As $\delta_{\rm DM}(k_{\rm pk}, a_{\rm rh}) = \delta_c(k_{\rm pk}, a_{\rm rh})$, the logarithmic growth of $\delta_{\rm DM}$ after reheating will then be of the form

$$\delta_{\rm DM}(k_{\rm pk}, a) = \frac{3A_s\phi_p}{2} \ln\left[\frac{4B_s e^{-3} a_{\rm dom}}{\gamma_{\rm pk} a_{\rm fz}}\right] \frac{a_{\rm rh}}{a_{\rm dom}} \tilde{b}_1 \ln\left(\tilde{b}_2 \frac{a}{a_{\rm rh}}\right). \tag{4.86}$$

Here $\tilde{b}_1$ and $\tilde{b}_2$ parameterize the transition from linear to logarithmic growth through reheating. Numerically we find $\tilde{b}_1 = 1.29$ and $\tilde{b}_2 = 1.66$.

Comparing eq. (4.86) with the standard logarithmic growth of $\delta_{\rm DM}$ during radiation domination, eq. (4.69), we find

$$A(k_{\rm pk}) = \frac{3A_s}{2} \ln\left[\frac{4B_s e^{-3} a_{\rm dom}}{\gamma_{\rm pk} a_{\rm fz}}\right] \frac{a_{\rm rh}}{a_{\rm dom}} \tilde{b}_1 \qquad\qquad B(k_{\rm pk}) = \tilde{b}_2 \frac{a_{\rm hor}}{a_{\rm rh}}. \tag{4.87}$$

Using the above relations in the definition of the transfer function, eq. (4.65), gives

$$T(k_{\rm pk}) \approx \frac{3}{2}\tilde{b}_1 \ln\left(\frac{4B_s e^{-3} a_{\rm dom}}{\gamma_{\rm pk} a_{\rm fz}}\right) \frac{a_{\rm rh}}{a_{\rm dom}} \left[1 - \frac{\ln(B_s \tilde{b}_2^{-1} a_{\rm rh}/a_{\rm hor,s})}{\ln(4B_s e^{-3} a_{\rm eq}/a_{\rm hor,s})}\right]. \tag{4.88}$$

This estimate is accurate to within 5% as long as $a_{\rm dom} > 100 a_{\rm fz}$ and $a_{\rm rh} > 10 a_{\rm dom}$, and is shown as a black dot-dashed line in the right panels of figure 4.8).

We obtain a simple approximation for $T(k_{\rm pk})$ by neglecting the logarithmic factors in the square bracket in eq. (4.88) as they provide only an $\mathcal{O}(1)$ correction and using the fact that $\gamma_{\rm pk}$ is typically of $\mathcal{O}(1)$, yielding

$$T(k_{\rm pk}) \sim 2 \ln\left(\frac{a_{\rm dom}}{10 a_{\rm fz}}\right) \frac{a_{\rm rh}}{a_{\rm dom}}. \tag{4.89}$$

Notice that in both $a_{dom} < 2a_{fz}$ and $a_{dom} > 2a_{fz}$ scenario, the peak of the transfer function is primarily determined by the duration of cannibal domination after the freeze-out of cannibal reactions. In the scenarios with $2a_{\rm fz} < a_{\rm dom}$ one also gets an additional logarithmic enhancement due to the logarithmic growth of $\delta_c(k_{\rm pk})$ prior to $a_{\rm dom}$.

4.4.4 The effects of DM-cannibal interactions

Until now we have focused on scenarios where the DM only interacts gravitationally with the other constituents of the universe. However, if DM and the cannibal particle are part of the same hidden sector, then it is natural for the two species to have non-gravitational interactions as well. In this section we show that the presence of non-gravitational DM-cannibal interactions does not change the key features of the transfer function and does not change $k_{\rm pk}$ and $T(k_{\rm pk})$.

Scattering between the DM and cannibal particles can cause the DM to be kinetically coupled to the cannibal fluid. The scenario we consider involves an interaction between two non-relativistic particles with a mass hierarchy ($m_{\rm DM} > m$) and hence is similar to the baryon-DM interactions studied in Ref. [109]. Using the results of Ref. [109], the momentum transfer rate ($\frac{1}{p}\frac{{\rm d}p}{{\rm d}t}$) experienced by a DM particle due to its s-wave scattering interactions with the cannibal bath is given by $n_{\rm can}(m/m_{\rm DM})\langle\sigma_{\rm DM,c} v_c\rangle$, where

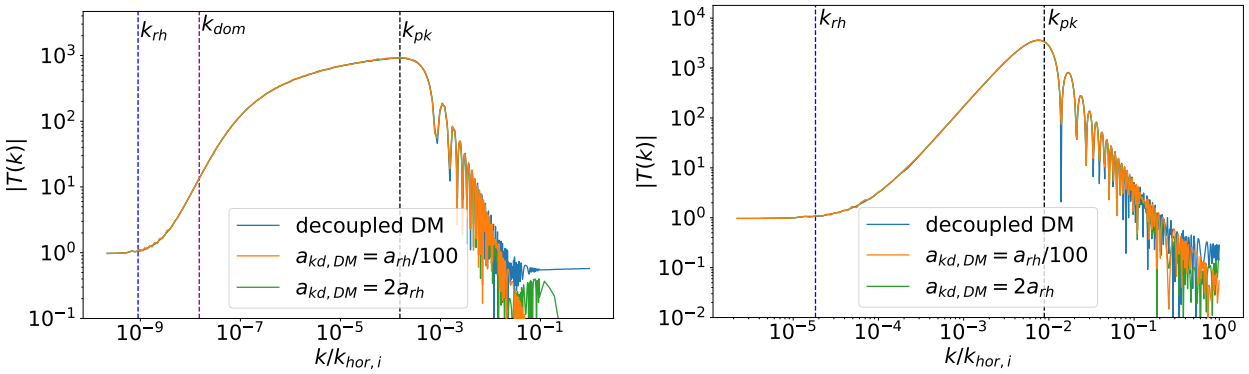


Figure 4.10: Transfer functions for scenarios with different cannibal-DM interaction cross-sections. The left panel shows an initially subdominant cannibal scenario with the same parameters as in figure 4.8. The right panel shows a scenario with an initially dominant cannibal density with $m = 1.8$ TeV, $T_{\rm rh} = 10$ MeV, $\alpha_c = 0.1$, and $\xi_i = 10$,. The blue line corresponds to no kinetic coupling between DM and the cannibals. The orange line corresponds to a scenario where the DM-cannibal scattering rate falls below the Hubble rate at $a_{\rm kd,DM} = a_{\rm rh}/100$, while the green line is for a scenario with $a_{\rm kd,DM} = 2a_{\rm rh}$. The transfer function is unaffected by cannibal-DM interaction for modes where $T(k) > 1$.

$\langle\sigma_{\rm DM,c}v_c\rangle \propto \sqrt{T_c/m}$ is the velocity averaged cross-section for cannibal-DM scattering. When we include these interactions, eq. (4.49)) for $\theta_{\rm DM}$ becomes

$$\theta'_{\rm DM}(a) + \frac{1}{a}\theta_{\rm DM} - \frac{k^2}{a^2 H}\phi = \frac{m}{m_{\rm DM}}\frac{n_{\rm can}\langle\sigma_{\rm DM,c}v_c\rangle}{aH}(\theta_c - \theta_{\rm DM}). \tag{4.90}$$

The momentum transfer rate for a cannibal particle interacting with the DM fluid is given by $n_{\rm DM}\langle\sigma_{\rm DM,c}v_c\rangle$. Consequently, a term similar to the RHS of eq. (4.90) would also appear in the θ_c equation but with an additional factor of $-\rho_{\rm DM}/\rho_{\rm can}$. As $\rho_{\rm DM} \ll \rho_{\rm can}$ prior to reheating, the effect of DM-cannibal scattering on θ_c is much smaller than its effect on $\theta_{\rm DM}$. Consequently, the DM perturbations track the cannibal perturbations while providing negligible feedback on the evolution of the cannibal perturbations. Hence we ignore the impact of DM interactions on the cannibal fluid.

In figure 4.10 we compare the transfer functions resulting from scenarios with different values of $\langle\sigma_{\rm DM,c}v_c\rangle$. We see that DM-cannibal interactions have no impact on the peak of the transfer function. In the limit of strong kinetic coupling, i.e. $n_{\rm can}\langle\sigma_{\rm DM,c}v_c\rangle \gg H$, the DM-cannibal interactions cause $\delta_{\rm DM}$ to track δ_c. However, regardless of the strength of the DM-cannibal kinetic coupling, the metric perturbation will always drive $\delta_{\rm DM}$ toward δ_c once the mode escapes the cannibal Jeans horizon. Consequently, the value of $\delta_{\rm DM}$ at reheating is insensitive to DM-cannibal scattering for modes that escape the cannibal Jeans horizon.

DM-cannibal interactions do affect the DM transfer function on very small scales, corresponding to modes that do not escape the cannibal Jeans horizon before reheating. For these modes, δ_c oscillates until reheating and so never generates a coherent gravitational pull on the DM perturbation. In scenarios with only gravitational interactions, shown by the blue line in figure 4.10, $\delta_{\rm DM}$ is larger than δ_c at reheating for these small-scale modes because δ_c oscillates while $\delta_{\rm DM}$ grows logarithmically until $a_{\rm dom}$. However, in the opposite limit where DM remains kinetically coupled to the cannibals until $2a_{\rm rh}$, shown with the green line, $\delta_{\rm DM}$ has the same value as δ_c at $a_{\rm rh}$.

The intermediate case shown by the orange line in figure 4.10 is more suppressed on very small scales than the tightly coupled case shown in green, even though the intermediate case has a smaller value of $\langle\sigma_{\rm DM,c}v_c\rangle$.

This relative suppression results from diffusion damping of the dark matter perturbations. Diffusion damping occurs when the cannibal perturbations oscillate faster than the DM-cannibal scattering rate prior to kinetic decoupling, i.e. $c_s k > n_{\rm can}\langle\sigma_{{\rm DM},c} v_c\rangle > H$. During this period, the DM perturbations oscillate with the same frequency as the oscillations in the cannibal perturbations, but the amplitude of their oscillation is highly damped. This damping is similar to the Silk damping of baryon density perturbations [110]. In figure 4.10, the orange line is more suppressed than the green line because the diffusion damping scale, $k_D^{-1} \sim c_s/(n_{\rm can}\langle\sigma_{{\rm DM},c} v_c\rangle)$, is larger for smaller values of $\langle\sigma_{{\rm DM},c} v_c\rangle$. Consequently, the modes experience damping at smaller k values when DM decouples shortly before reheating compared to the tightly coupled case.

4.5 Beyond the perfect-fluid approximation

So far we have assumed that the cannibals and the DM fluids are perfect fluids. The perfect-fluid approximation will break down on scales where the random motion of the particles comprising the fluid cannot be neglected, which can occur in a variety of regimes. Even while the homogeneous cannibal fluid is in kinetic equilibrium, the cannibals still have a finite diffusion length. For perturbations on scales smaller than the diffusion length, a perfect-fluid description is not sufficient. Once kinetic equilibrium is lost, the random thermal motion of particles becomes important on scales quantified by either the free-streaming length or the collisionless Jeans length, depending on the gravitational forces experienced by the particles. Again, for perturbations with wavelengths smaller than these scales, the perfect-fluid description breaks down.

Momentum exchange among cannibal particles is dominated by elastic 2-to-2 scatterings, with a rate given by $n_{\rm can}\langle\sigma_c v_c\rangle$. In appendix B.2 we derive the two-to-two scattering rate for the φ^4 theory described by eq. (4.1). We find that the $s-$wave contribution in the non-relativistic limit is

$$\langle\sigma_c v_c\rangle = \frac{1}{64\pi^{3/2} m^2}\left(\lambda - \frac{5}{3}\frac{g^2}{m^2}\right)^2 \times \sqrt{\frac{T_c}{m}} \equiv \sigma_{\rm eff}\sqrt{\frac{T_c}{m}}, \tag{4.91}$$

where λ and g are the coupling constants of the cubic and quartic interactions in eq. (4.1), respectively, and in the second equality we pulled out a factor of $v_c = \sqrt{T_c/m}$ to define an effective scattering cross-section. Note that the parameter α_c that controls the cannibal number-changing interactions does not uniquely determine σ_c because α_c and σ_c depend on different combinations of g and λ.

In a Hubble time, a cannibal particle will undergo $N = (n_{\rm can}\langle\sigma_c v_c\rangle)/H$ scatterings. The average distance travelled by a cannibal particle between two collisions is $\ell_{\rm mfp} \sim 1/(n_{\rm can}\sigma_{\rm eff})$. Consequently, the comoving diffusion length is given by

$$\lambda_{\rm diff} = \frac{1}{a}\sqrt{N} \times \ell_{\rm mfp} = \frac{1}{a}\sqrt{\frac{v_c}{n_{\rm can}\sigma_{\rm eff} H}}. \tag{4.92}$$

For modes with wavelengths shorter than the comoving diffusion length, the higher moments of the Boltzmann hierarchy can no longer be neglected, and will suppress δ_c [111]. The perfect-fluid approximation also breaks down for modes that oscillate faster than the 2-to-2 scattering rate, i.e. if $c_s k > n_{\rm can}\langle\sigma_c v_c\rangle$. Since $c_s \sim v_c$, requiring the oscillation frequency to be slower than the scattering rate is equivalent to requiring $k^{-1} > \ell_{\rm mfp}$. Since the cannibal diffusion length is larger than the mean free path prior to kinetic decoupling (as $N > 1$), modes will be damped by diffusion before the scattering rate falls below the oscillation frequency.

The diffusion length is relevant as long as the cannibal fluid maintains internal kinetic equilibrium, i.e.

$n_{\rm can}\langle\sigma_c v_c\rangle > H$. We define the scale factor, $a_{\rm kd}$, at which the cannibal fluid falls out of its kinetic equilibrium through the relation

$$n_{\rm can}(a_{\rm kd})\langle\sigma_c v_c(a_{\rm kd})\rangle = H(a_{\rm kd}). \tag{4.93}$$

After kinetic decoupling, the cannibal fluid is effectively collisionless. Cannibal number-changing interactions, which involve three particles in the initial state, freeze out substantially before the cannibal fluid loses internal kinetic equilibrium, so after kinetic decoupling the cannibal fluid evolves as pressureless matter.[5]

While the universe is radiation dominated, the cannibals experience no coherent gravitational force and have a comoving free-streaming length given by

$$\lambda_{\rm fs}(a) = \lambda_{\rm diff}(a_{\rm kd}) + \int_{a_{\rm kd}}^{a} \frac{v_c}{a^2 H} da \qquad\qquad a_{\rm kd} < a < a_{\rm dom}. \tag{4.94}$$

Here we have imposed that the cannibal diffusion length is equal to the free-streaming length at kinetic decoupling. When the cannibal comes to dominate the universe, metric perturbations can begin to pull particles toward overdense regions. In this regime, departures from perfect-fluid behavior are governed by the collisionless Jeans length. Analogous to the collisional Jeans length described in the previous section, the collisionless Jeans length determines the scale above which gravitational attraction is sufficient to overcome the random motion of particles.

To find the collisionless Jeans length we need to include the anisotropic stress, $\sigma_{\rm can}$, in eq. (4.47), which governs the evolution of θ_c. Before kinetic decoupling, elastic cannibal scattering ensures that $\sigma_{\rm can}$ is only relevant for modes within the diffusion length. After kinetic decoupling, the anisotropic stress is determined by the free-streaming velocity of the cannibals. Ref. [112] finds the anisotropic stress for a collisionless fluid to be given by $\sigma = -\frac{5}{3}\langle v^2\rangle\delta$, which follows from the assumption that the phase-space density of the particles remains unchanged while particles fall into gravitational potential wells. After kinetic decoupling, the sound speed term in eq. (4.47) is ill-defined and no longer appears in that equation. Consequently, the θ_c equation after cannibal kinetic decoupling is

$$\theta_c{}'(a) = -\frac{1}{a}(1-3w_c)\theta_c - \frac{w_c'}{1+w_c}\theta_c + \frac{k^2}{a^2 H}\phi + \frac{k^2}{a^2 H}\frac{5}{3}\langle v^2\rangle\delta. \tag{4.95}$$

We find the collisionless Jeans length from eq. (4.95) by following the same steps as we performed for calculating the collisional Jeans length: find the simple harmonic oscillator equation for δ_c analogous to eq. (4.56) and then find the wavenumber k for which the frequency becomes imaginary. Doing so, we find the collisionless Jeans length, $k_{J,c}^{-1}$, to be the same as the collisional Jeans length in eq. (4.57), except with $c_s^2/(1+w_c)$ replaced by $5\langle v_c^2\rangle/3$:

$$k_{J,c} = \sqrt{\frac{9}{10\langle v_c^2\rangle}}\, aH. \tag{4.96}$$

After cannibal freeze-out and before cannibal kinetic decoupling, the sound speed is given by $c_s^2 = \frac{5}{3}\frac{T_c}{m} = \frac{5}{3}\langle v_c^2\rangle$

[5]The s-wave component of the two-to-two scattering cross-section vanishes for $\lambda = (5/3)g^2/m^2$; for couplings in the neighborhood of such values, the p-wave component will dominate the elastic scattering cross-section in the non-relativistic regime. As the three-to-two cannibal interactions are phase-space suppressed as well as higher order in the couplings, they will still generically decouple earlier than the elastic scattering interactions, but to examine this specific sliver of parameter space in detail requires the retention of p-wave contributions beyond eq. 4.91, and is beyond the scope of this work.

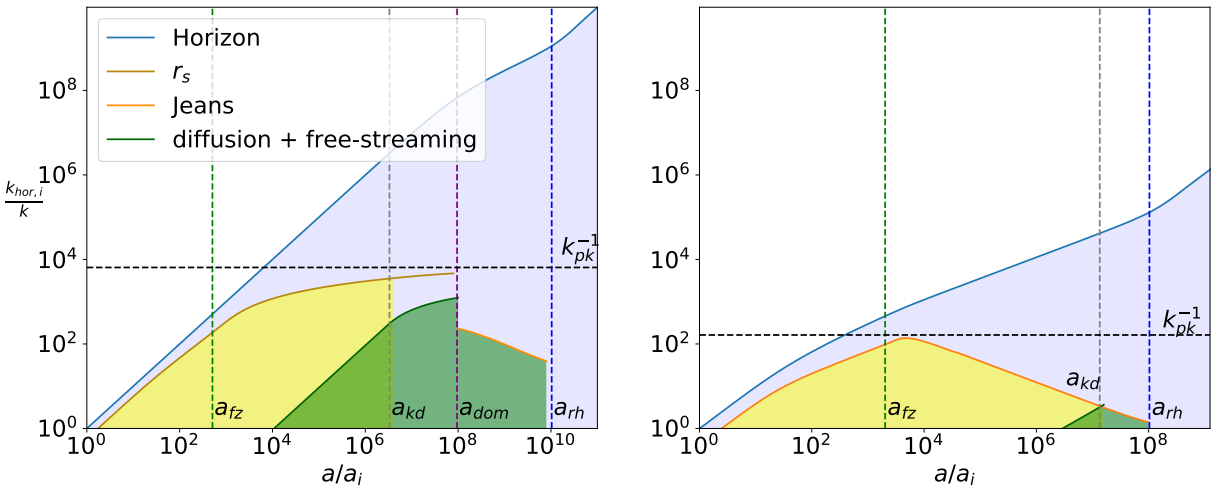


Figure 4.11: Length scales where the perfect-fluid approximation for the cannibals breaks down. The left panel uses the same parameters as in figure 4.8 and shows a scenario where both cannibal freeze-out and kinetic decoupling occur during SM radiation domination. The right panel takes $m = 6.1$ TeV, $T_{\rm rh} = 10$ MeV, $\alpha_c = 0.42$, and $\xi_i = 10$ and shows a scenario where cannibal freeze-out occurs during cannibal domination. The green line prior to $a_{\rm kd}$ shows the comoving cannibal diffusion length as given in eq. (4.92), and in the left panel between $a_{\rm kd}$ and $a_{\rm dom}$, the comoving free-streaming length of the cannibals as given in eq. (4.94). The orange solid line is the collisional Jeans length, eq. (4.57), for $a < a_{\rm kd}$ and the collisionless Jeans length, eq. (4.96), for $a > a_{\rm kd}$. The brown line indicates the cannibal sound horizon, eq. (4.67). The green shaded region marks the regimes where the perfect-fluid approximation for the cannibals breaks down. Here we show the diffusion length taking $\lambda = 0$ in eq. (4.91).

and $w_c \ll 1$. Therefore, the collisionless Jeans length has the same value as the collisional Jeans length would have had in the absence of kinetic decoupling.

We show the evolution of all three length scales (diffusion, free-streaming, and collisional Jeans lengths) in fig. 4.11. In the left panel we show a scenario where kinetic decoupling occurs before the cannibal density comes to dominate, and in the right panel we show a scenario where the cannibal density is always dominant. In both panels, the solid green line shows the growth of the diffusion length up until $a_{\rm kd}$, which, in the left panel, transitions smoothly into the free-streaming length given in eq. (4.94) in the region $a_{\rm kd} < a < a_{\rm dom}$. In the left panel, the solid orange line after $a_{\rm dom}$ shows the collisionless Jeans length. In the right panel, the orange line shows the collisional Jeans length before $a_{\rm kd}$ and the collisionless Jeans length after $a_{\rm kd}$. The green shaded region indicates the scales where the cannibal particles no longer behave as a perfect fluid. For modes that enter the green shaded region in figure 4.11, we expect $\delta_c(a_{\rm rh})$ to experience a suppression that is not captured in our suite of perturbation equations. The black dashed line shows the peak of the matter power spectrum in the perfect-fluid approximation, and thus indicates the location of the cutoff that follows from oscillations in the cannibal fluid. We see that, in the cases shown, the modes that are affected by the breakdown of the perfect-fluid approximation are already suppressed by the cannibal oscillations.

Cutoffs arising from imperfect-fluid behavior can be important for the transfer function when they occur on scales larger than the small-scale cutoff provided by cannibal oscillations. In scenarios where the cannibals freeze out during cannibal domination, the cutoff provided by cannibal interactions sets $k_{\rm pk} \approx k_J(2a_{\rm fz})/1.4$, which ensures that any deviations from perfect-fluid behavior occur at scales substantially below $k_{\rm pk}^{-1}$ when $a_{\rm fz} \ll a_{\rm kd}$. In scenarios where the cannibals freeze out during radiation domination, however, the situation is

a little more subtle. In this case, the perfect-fluid calculation of the cannibal cutoff gives $k_{\rm pk} \sim r_s(a_{\rm dom})$. As both the sound horizon and the free-streaming length grow logarithmically during the period of radiation domination following $a_{\rm kd}$, the sound horizon will generically stay outside the free-streaming length, and therefore the cannibal oscillation cutoff $k_{\rm pk}^{-1}$ as given by eq. (4.78) will occur at larger scales than the free-streaming length. However, the derivation of eq. (4.78) assumes a collisional fluid. Thus while we expect the cannibal oscillation cutoff to be the relevant small-scale cutoff for $a_{\rm kd} > a_{\rm dom}$, this conclusion does not necessarily hold if $a_{\rm kd} < a_{\rm dom}$. The regime with $a_{\rm kd} < a_{\rm dom}$ can be realized in a small region of parameter space, as we show in section 4.6. A full calculation of the small-scale cutoff in this regime would require incorporating higher moments of the Boltzmann hierarchy and is beyond the scope of this work; see [113] for related calculations in a similar model. However in general we can expect this cutoff to lie somewhere in the vicinity of $r_s(a_{\rm kd})$ and $\lambda_{\rm fs}(a_{\rm dom})$. These two scales are relatively similar: the sound horizon is governed by the distance traveled by sound waves in the cannibal fluid, while the free-streaming horizon is governed by the distance traveled by the non-relativistic cannibal particles in this epoch. Both the sound speed and the cannibal particle speed are determined by the cannibal temperature, which changes only logarithmically between $a_{\rm kd}$ and $a_{\rm dom}$.

Finally, we quantify the relationship between $a_{\rm fz}$ and $a_{\rm kd}$ in our cannibal model. Since the $2 \to 2$ and $3 \to 2$ scattering cross-sections depend on different combinations of the quartic coupling λ and cubic coupling g/m, we can obtain a range of possible $a_{\rm kd}$ for a fixed $a_{\rm fz}$. To evaluate $a_{\rm kd}/a_{\rm fz}$ in terms of the Lagrangian couplings, we first divide eq. (4.93) by eq. (4.17):

$$\frac{n_{\rm can}(a_{\rm kd})\langle\sigma_c v_c(a_{\rm kd})\rangle}{n_{\rm can}^2(a_{\rm fz})\langle\sigma v^2\rangle_{\rm can}} = \frac{H(a_{\rm kd})}{H(a_{\rm fz})}. \tag{4.97}$$

We then express $n_{\rm can}(a_{\rm kd})$ and $T_c(a_{\rm kd})$ in terms of their values at $a_{\rm fz}$ by using $n_{\rm can} \propto 1/a^3$ and $T_c \propto 1/a^2$ for $a > a_{\rm fz}$. Next, we approximate $m n_{\rm can}(a_{\rm fz}) \approx \rho_{\rm can,eq}(a_{\rm fz})$ and $T_c(a_{\rm fz}) \approx T_{c,\rm eq}(a_{\rm fz})$ and use eq. (4.12) and eq. (4.10) to express $\rho_{\rm can,eq}$ and $T_{c,\rm eq}$ in terms of the scale factor. Finally, we set $a_{\rm fz}/a_i = 10^3$ inside the logarithms to obtain

$$\frac{a_{\rm kd}^4}{a_{\rm fz}^4}\frac{H(a_{\rm kd})}{H(a_{\rm fz})} \sim 10^{-2}\left(\frac{a_{\rm fz}}{a_i}\right)^3 \frac{\sigma_{\rm eff}}{m^3\langle\sigma v^2\rangle_{\rm can}} = 10^{-2}\left(\frac{a_{\rm fz}}{a_i}\right)^3 \frac{20736}{225\sqrt{5\pi}} \frac{\left[\lambda - 5g^2/(3m^2)\right]^2}{(g/m)^2[(g/m)^2 - 3\lambda]^2}. \tag{4.98}$$

In the last equality above we used eq. (4.91) and eq. (4.15) for $\sigma_{\rm eff}$ and $\langle\sigma v^2\rangle_{\rm can}$ respectively. When the universe is cannibal dominated between $a_{\rm fz}$ and $a_{\rm kd}$, then $H(a_{\rm fz})/H(a_{\rm kd}) = (a_{\rm kd}/a_{\rm fz})^{3/2}$ and $a_{\rm fz}$ is given by eq. (4.42). Defining q as the ratio of the quartic and the cubic coupling, $q \equiv \lambda/(g/m)^2$, and expressing $(g/m)^2$ using the definition of α_c in eq. (4.98) yields

$$\frac{a_{\rm kd}}{a_{\rm fz}} \sim 10^5 \alpha_c^{2/5}\left(\frac{\rm GeV}{m}\right)^{4/15}\left|\frac{q - 5/3}{(3q-1)^{2/3}}\right|^{4/5}. \tag{4.99}$$

Similarly, if the universe is radiation dominated between $a_{\rm fz}$ and $a_{\rm kd}$, $H(a_{\rm fz})/H(a_{\rm kd}) = (a_{\rm kd}/a_{\rm fz})^2$ and $a_{\rm fz}$ is given by eq. (4.43). Eq. (4.98) then implies that

$$\frac{a_{\rm kd}}{a_{\rm fz}} \sim 2 \times 10^6 \alpha_c^{5/8}\xi_i^{3/4}\left(\frac{\rm GeV}{m}\right)^{3/8}\left|\frac{q - 5/3}{(3q-1)^{2/3}}\right|. \tag{4.100}$$

In most of the parameter space that realizes an ECDE, varying λ while keeping α_c fixed results in a variation

in $a_{\rm kd}/a_{\rm fz}$ of up to an order of magnitude. Figure 4.11 shows results with $\lambda = q = 0$; for the value of α_c shown in figure 4.11 increasing λ to non-perturbative values results in a correction of less than 30% to the values of $a_{\rm kd}$ shown in the figure.

Departures of the DM from perfect-fluid behavior can also be important for determining the transfer function. Prior to reheating, the DM free-streaming, diffusion, and collisionless Jeans lengths are always smaller than the perfect-fluid result for $k_{\rm pk}^{-1}$, as the DM speed $v_{\rm DM} = \sqrt{T_{\rm DM}/m_{\rm DM}}$ is always smaller than the cannibal sound speed, $c_s \sim \sqrt{T/m}$, which controls the scale of $k_{\rm pk}$. However, DM free streaming after reheating can affect the peak of the DM transfer function in some regions of parameter space. Ref. [26] evaluated the post-reheating free streaming of DM in the case where DM kinetically decouples from the cannibal fluid after $a_{\rm fz}$. In this case we have $T_{\rm DM}(a_{\rm rh}) = T_c(a_{\rm rh})$, which gives for the DM free-streaming length

$$\lambda_{\rm DM,fs} \approx \int_{t_{\rm rh}}^{t_0} v_{\rm DM} \frac{dt}{a} = \sqrt{\frac{T_c(a_{\rm rh})}{m_{\rm DM}}} \frac{1}{(aH)_{\rm rh}} \int_{a_{\rm rh}}^{a_0} \frac{da}{a^3} \left[G(a) \left(\frac{1}{a}\right)^4 + G(a_{\rm eq}) \left(\frac{1}{a^3 a_{\rm eq}}\right) \right]^{-1/2}, \tag{4.101}$$

where we defined $G(a) \equiv g_*(a) g_*^{1/3}(a_{\rm rh})/g_{*s}^{4/3}(a)$ and dropped negligible contributions from dark energy at late times. This result is applicable regardless of whether the universe was radiation dominated or SM radiation dominated at a_i. While the DM free-streaming length in any given model will depend in detail on the DM coupling to the cannibal species, eq. (4.101) provides an upper bound on $\lambda_{DM,fs}$: DM that decouples from the cannibals prior to $a_{\rm fz}$ will have a reduced free-streaming length as the temperature of the DM at reheating will be colder than the cannibal temperature.

If the DM free-streaming length is larger than the small-scale cutoff coming from cannibal self-interactions, then the DM transfer function will be maximized on a scale $\sim \lambda_{\rm DM,fs}^{-1}$, which depends on DM as well as cannibal microphysics. For $m_{\rm DM} \gtrsim 10m$, we find DM free-streaming can provide the small-scale cutoff in the transfer function in a small portion of the parameter space, as we discuss in the following section.

4.6 Implications for microhalo formation

In this section, we first discuss how the key features of the linear transfer function, namely $k_{\rm pk}$ and $T(k_{\rm pk})$, relate to the properties of the earliest-forming microhalos. We then express $k_{\rm pk}$ and $T(k_{\rm pk})$ as a function of the cannibal parameters $m, T_{\rm rh}, \alpha_c$, and ξ_i. Finally, we briefly discuss the microhalos' observational signatures and how these observations probe cannibalism in the early Universe.

After an ECDE, the DM perturbations with wavenumber $k_{\rm pk}$ have experienced the most growth. Although the stochastic nature of the primordial perturbations prevents us from knowing exactly which mode has the largest amplitude, the near scale-invariance of the primordial power spectrum implies that perturbations on scales near $k_{\rm pk}$ are the first to collapse and form gravitationally bound structures. Since perturbations that enter the horizon prior to BBN form halos that are too small to capture baryons [114], the characteristic mass of the earliest-forming halos is given by the amount of DM in a sphere of comoving radius $k_{\rm pk}^{-1}$:

$$M_{\rm pk} \equiv \frac{4\pi}{3} k_{\rm pk}^{-3} \rho_{\rm DM,0}, \tag{4.102}$$

where $\rho_{\rm DM,0}$ is the dark matter density today, which we take to be $\rho_{\rm DM,0} = 9.7 \times 10^{-48}$ GeV^4 [3]. When cannibals freeze out while they dominate the energy density of the universe ($a_{\rm dom} < 2a_{\rm fz}$), we calculate $M_{\rm pk}$

from the expression for $k_{\rm pk}$ given in eq. (4.80):

$$M_{\rm pk} \sim 10^{-11} M_\odot \left(\frac{\alpha_c}{0.1}\right) \left(\frac{10\,{\rm MeV}}{T_{\rm rh}}\right) \left(\frac{\rm TeV}{m}\right)^{7/3} . \tag{4.103}$$

For $a_{\rm dom} > 2a_{\rm fz}$, we calculate $M_{\rm pk}$ from the expression for $k_{\rm pk}$ given in eq. (4.81):

$$M_{\rm pk} \sim 3 \times 10^{-13} M_\odot \left(\frac{\alpha_c}{0.1}\right)^{9/4} \left(\frac{\xi_i}{0.4}\right)^{15/2} \left(\frac{10\,{\rm MeV}}{T_{\rm rh}}\right) \left(\frac{\rm TeV}{m}\right)^{11/4} \left(\frac{100}{g_*(10m/\xi_i)}\right)^{15/8} \times \left(\frac{1}{6} \ln\left[500 \left(\frac{0.4}{\xi_i}\right)^{9/2} \left(\frac{0.1}{\alpha_c}\right)^{3/4} \left(\frac{m}{\rm TeV}\right)^{1/4} \left(\frac{g_*(10m/\xi_i)}{100}\right)^{9/8}\right]\right)^{9/2} . \tag{4.104}$$

We remind the reader that the expression for $k_{\rm pk}$ given in eq. (4.81) is a good approximation for $10 < a_{\rm dom}/a_{\rm fz} < 10^5$ and $10^2 < a_{\rm fz}/a_i < 10^4$. Since the peak halo mass is typically much smaller than one Earth mass, the earliest-forming halos are microhalos.

In both cases, $M_{\rm pk}$ increases as either $T_{\rm rh}$ or m decreases because $M_{\rm pk}$ is determined by the sound horizon at reheating, $r_s(a_{\rm rh})$. Decreasing $T_{\rm rh}$ delays reheating and hence increases $r_s(a_{\rm rh})$. Decreasing m increases $r_s(a_{\rm rh})$ by delaying the freeze-out of cannibal reactions, which increases the cannibal temperature. Since increasing α_c also delays the freeze-out of cannibal reactions, we see a positive correlation between $M_{\rm pk}$ and α_c. The peak halo mass has a stronger dependence on α_c when the cannibals freeze out while the universe is SM radiation dominated because the Hubble rate falls faster in a radiation-dominated universe compared to a cannibal-dominated universe.

An ECDE enhances the amplitude of all perturbations with $k < k_{\rm pk}$ that enter the horizon during the ECDE. Therefore, the largest halos that are affected by the ECDE have masses equal to the amount of DM within the horizon at reheating, $M_{\rm rh}$, which is given by eq. (4.102) but with $k_{\rm pk}$ replaced by $k_{\rm rh} = a_{\rm rh} H(a_{\rm rh})$. We find $M_{\rm rh}$ in terms of cannibal parameters by taking $H(a_{\rm rh}) \sim \Gamma$ and then expressing Γ in terms of $T_{\rm rh}$.[6] We then express $a_{\rm rh}/a_0$ in terms of SM temperatures using entropy conservation to obtain

$$M_{\rm rh} \sim 10^{-4} M_\odot \left(\frac{10\,{\rm MeV}}{T_{\rm rh}}\right)^3 \left(\frac{10}{g_*(T_{\rm rh})}\right)^{1/2} . \tag{4.105}$$

While deriving the above relation we set $g_{*s}(T_{\rm rh}) = g_*(T_{\rm rh})$. An ECDE increases the abundance of all halos with masses between $M_{\rm pk}$ and $M_{\rm rh}$, and these halos form earlier than they would in a standard cosmology.

Halos form when $\delta_{\rm DM}$ becomes of order unity. In a standard cosmology, the amplitude of small-scale perturbations increases only logarithmically with k, so microhalos with masses within several orders of magnitude of an earth mass form near a redshift of 60 [115, 116]. Since baryons do not participate in structure formation for modes that enter the horizon during an ECDE, $\delta_{\rm DM} \propto (1+z)^{-0.9}$ for $z < z_{\rm eq}$ on these scales [105, 107]. Consequently, the collapse redshift of the microhalos corresponding to overdensities with wavenumber k increases by a factor of $\sim [T(k)]^{1.11}$ compared to that in the standard cosmology as long as the collapse occurs after matter-radiation equality, i.e. for $T(k) < 30$. For $T(k) > 30$, the formation of the microhalos occurs prior to matter-radiation equality, and the exact increase in the collapse redshift depends non-trivially on $T(k)$ [19].

The central density of a dark matter halo scales with the homogeneous matter density at the time of its

[6] The Hubble rate at $a_{\rm rh}$ does not equal Γ because $a_{\rm rh}$ is defined as the scale factor when the Hubble rate equals Γ in a standard cosmology. However, since $\rho_{\rm can} \sim \rho_r$ at $a_{\rm rh}$, $H(a_{\rm rh})$ is some $\mathcal{O}(1)$ factor times Γ.

formation [117–119], so the microhalos that form after an ECDE have central densities that are significantly larger than those in standard cosmology [120]. These central densities are large enough for the microhalos to survive within galaxies, although they experience significant mass loss [20, 25, 116, 121–123].

If $T(k_{\rm pk})$ is significantly large, then the cannibals and DM particles assemble into microhalos before reheating. For modes in the vicinity of the peak in the matter power spectrum, eq. (4.87) implies that the DM overdensity at reheating is related to the primordial metric fluctuation via $\delta_{\rm DM}(k, a_{\rm rh}) \approx A(k)\phi_P$. For a nearly scale-invariant spectrum, we expect $\phi_P \sim 10^{-5}$ [3], and thus density perturbations on all scales remain perturbative until reheating provided $A(k_{\rm pk}) \lesssim 10^5$. As $T(k_{\rm pk}) \approx A(k_{\rm pk})/A_s$, where $A_s = 9.11$, microhalos will form prior to reheating if $T(k_{\rm pk})$ exceeds 10^4. These microhalos are destroyed once reheating occurs because they are primarily composed of cannibal particles. When the cannibal particles decay, DM particles free stream out of the microhalos with typical speeds given by the virial speed within the microhalos, which is of order 10^{-2} [19]. The subsequent free streaming of DM particles acts to erase the structure within the comoving horizon at the time of reheating, thus washing out much of the enhanced structure resulting from the ECDE.

The peak amplitude of the transfer function, and thus the formation time of the first microhalos, can be directly related to the properties of the cannibal field. In the case where cannibal freeze-out occurs during cannibal domination ($a_{\rm dom} < 2a_{\rm fz}$), we use the expression for $T(k_{\rm pk})$ given in eq. (4.84) and express $a_{\rm fz}$ and $a_{\rm rh}$ using eq. (4.42) and eq. (4.40) respectively to obtain

$$T(k_{\rm pk}) \sim 2 \times 10^3 \left(\frac{0.1}{\alpha_c}\right)^{2/3} \left(\frac{m}{\rm TeV}\right)^{14/9} \left(\frac{10\,{\rm MeV}}{T_{\rm rh}}\right)^{4/3} \left(\frac{10}{g_*(T_{\rm rh})}\right)^{1/3} . \tag{4.106}$$

For $a_{\rm dom} > 2a_{\rm fz}$, we use eq. (4.89) for $T(k_{\rm pk})$ and express $a_{\rm fz}$, $a_{\rm rh}$, and $a_{\rm dom}$ using eq. (4.43), eq. (4.40), and eq. (4.41) respectively to obtain

$$\begin{aligned} T(k_{\rm pk}) \sim 2 \times 10^2 \left(\frac{\xi_i}{0.4}\right)^4 \left(\frac{m}{\rm TeV}\right)^{4/3} \left(\frac{10\,{\rm MeV}}{T_{\rm rh}}\right)^{4/3} \left(\frac{10}{g_*(T_{\rm rh})}\right)^{1/3} \left(\frac{100}{g_*(10m/\xi_i)}\right) \\ \times \frac{1}{6} \ln\left[50 \left(\frac{0.4}{\xi_i}\right)^{9/2} \left(\frac{0.1}{\alpha_c}\right)^{3/4} \left(\frac{m}{\rm TeV}\right)^{1/4} \left(\frac{g_*(10m/\xi_i)}{100}\right)^{9/8}\right] . \end{aligned} \tag{4.107}$$

In both cases, $T(k_{\rm pk})$ is approximately proportional to $m/T_{\rm rh}$ because for a given α_c and ξ_i this ratio determines the post-freeze-out duration of the ECDE. Since $\delta_{\rm DM}(k_{\rm pk})$ grows faster during this period than at any other time prior to matter-radiation equality, increasing this duration increases $T(k_{\rm pk})$. The amplitude of the transfer function at $k_{\rm pk}$ has a power-law dependence on α_c when the cannibals freeze out in a cannibal-dominated universe, while it only depends logarithmically on α_c when the cannibals freeze out in a SM radiation-dominated universe. This difference in sensitivity to α_c reflects the linear growth of $\delta_{\rm DM}(k_{\rm pk})$ between $a_{\rm fz}$ and $a_{\rm rh}$ for $a_{\rm dom} < 2a_{\rm fz}$, as opposed to its logarithmic growth between $a_{\rm fz}$ and $a_{\rm dom}$ for $a_{\rm dom} > 2a_{\rm fz}$. These two growth histories for $\delta_{\rm DM}(k_{\rm pk})$ also explain why $T(k_{\rm pk})$ is independent of ξ_i when $a_{\rm dom} < 2a_{\rm fz}$, but is strongly dependent on ξ_i when $a_{\rm dom} > 2a_{\rm fz}$: since ξ_i determines $a_{\rm dom}/a_i$, it sets the transition from logarithmic to linear growth when cannibals freeze out prior to the start of the ECDE.

In figure 4.12, white-dashed contours show $M_{\rm pk}$ as a function of m and $T_{\rm rh}$ for fixed values of ξ_i and α_c. The $M_{\rm pk}$ contours were calculated from the expression for $k_{\rm pk}$ given in eq. (4.66) if the cannibals freeze out in a cannibal-dominated universe and eq. (4.78) if the cannibals freeze out in a SM radiation-dominated universe. The colored contours show $T(k_{\rm pk})$, which is evaluated by numerically solving the cosmological

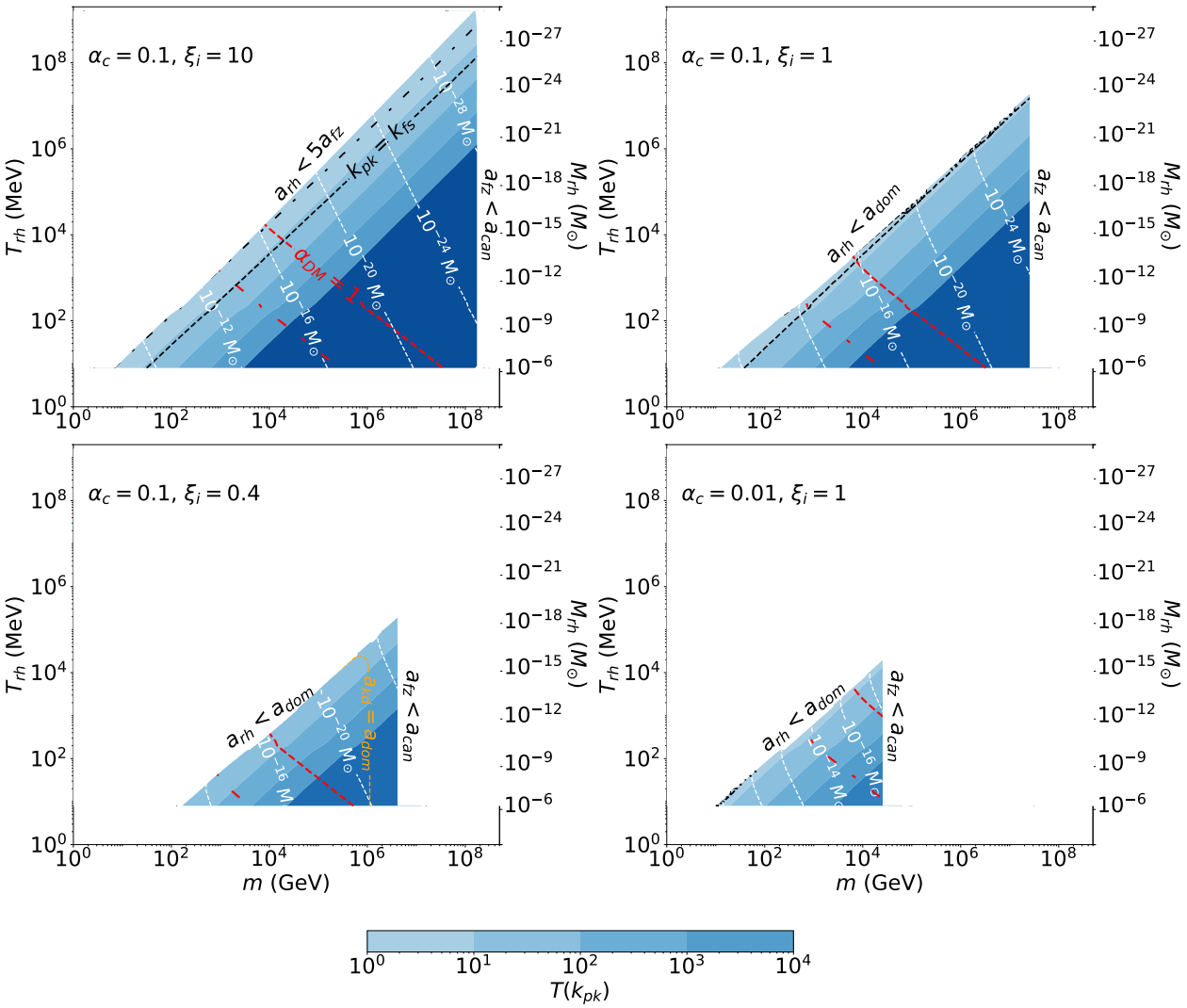


Figure 4.12: Colored contours show the value of the DM transfer function at the scale that maximizes the matter power spectrum, $T(k_{\rm pk})$, as a function of m and $T_{\rm rh}$ for fixed α_c and ξ_i. In the top left panel, the cannibal density exceeds the SM radiation density up until reheating. In all other panels, the cannibals freeze out in a SM radiation-dominated universe. The white dashed contours show the mass scale of the first microhalos, eq. (4.102). The secondary y-axis shows the microhalo mass scale corresponding to modes entering the horizon at reheating. In the white space on the top left, reheating either occurs prior to cannibal freeze-out (top left panel) or the cannibal density does not dominate the universe prior to the decay of the cannibals (all other panels). In the white space on the right, cannibalism does not occur. Above the red dashed (dot-dashed) line, thermal freezeout cannot generate the observed DM abundance for $m_{\rm DM} \geq 10m$ ($m_{\rm DM} \geq 100m$). To the left of the black dashed (dot-dashed) line, the post-reheating free streaming of DM modifies $k_{\rm pk}$ if DM kinetically decouples from the cannibals after $a_{\rm fz}$ and $m_{\rm DM} = 10m$ ($m_{\rm DM} = 100m$). In the bottom left panel, the cannibal fluid becomes collisionless prior to $a_{\rm dom}$ to the right of the orange-dashed line, and consequently cannibal free streaming modifies $k_{\rm pk}$.

perturbation equations for $k_{\rm pk}$. The secondary y-axis on the right shows the values of $M_{\rm rh}$ obtained from eq. (4.105). The parameter space with $5 \lesssim T(k_{\rm pk}) \lesssim 10^4$ is the region that generates a significantly enhanced abundance of microhalos with masses between $M_{\rm pk}$ and $M_{\rm rh}$. As $T(k_{\rm pk})$ is roughly proportional to $m/T_{\rm rh}$, there is an enhanced abundance of microhalos for a band of $m/T_{\rm rh}$ values.

The parameter space shown in figure 4.12 is bounded on all sides by three conditions. First, the reheat temperature defined by eq. (4.34) must exceed 8.1 MeV to be compatible with the constraints from BBN and the CMB [102, 103]. Second, significantly increasing the microhalo abundance requires a period of cannibal domination following cannibal freeze-out, i.e. $\max(a_{\rm fz}, a_{\rm dom}) < a_{\rm rh}$. Finally, as here we are specifically interested in the impact of cannibal interactions on perturbation growth, we require an epoch of cannibalism to occur, i.e., $a_{\rm can} < a_{\rm fz}$.

A period of cannibalism only occurs if the cannibals remain in chemical equilibrium after they become non-relativistic. For a fixed value of α_c, this condition imposes an upper bound on m because $a_{\rm fz}/a_i$ decreases as m increases. Using eq. (4.42) and eq. (4.43) for $a_{\rm fz}/a_i$ and the fact that $a_{\rm can} \approx 100 a_i$, the $a_{\rm fz} > a_{\rm can}$ condition can be rewritten as:

$$\left(\frac{m}{\rm GeV}\right) < \begin{cases} 10^8 \left(\dfrac{\alpha_c}{0.1}\right)^3 & a_{\rm dom} < 2a_{\rm fz}, \\ 2\times 10^6 \left(\dfrac{\xi_i}{0.4}\right)^2 \left(\dfrac{\alpha_c}{0.1}\right)^3 \left(\dfrac{100}{g_*(10m/\xi)}\right)^{1/2} & a_{\rm dom} > 2a_{\rm fz}. \end{cases} \tag{4.108}$$

If m exceeds the bound in eq. (4.108), then the number-changing self-interactions decouple while the cannibal particles are relativistic, and any subsequent ECDE is simply a matter-dominated era. The evolution of perturbations in such scenarios (without cannibal interactions) have already been studied in the context of decoupled hidden sector theories [19, 74].

To obtain a substantial amount of growth, reheating must occur well after the cannibal reactions freeze out ($a_{\rm rh} > 5a_{\rm fz}$) and in a cannibal-dominated universe ($a_{\rm rh} > a_{\rm dom}$).[7] This requirement imposes a lower bound on m for a given $T_{\rm rh}$, following from the expressions for $a_{\rm rh}$, $a_{\rm fz}$, and $a_{\rm dom}$ given in eq. (4.40), eq. (4.42), and eq. (4.41):

$$\left(\frac{m}{\rm GeV}\right) > \begin{cases} 10 \left(\dfrac{\alpha_c}{0.1}\right)^{3/7} \left(\dfrac{T_{\rm rh}}{10\,\rm MeV}\right)^{6/7} \left(\dfrac{g_*(T_{\rm rh})}{10}\right)^{3/14} & a_{\rm dom} < 2a_{\rm fz}, \\ 70 \left(\dfrac{0.4}{\xi_i}\right)^3 \left(\dfrac{T_{\rm rh}}{10\,\rm MeV}\right) \left(\dfrac{g_*(10m/\xi_i)}{100}\right)^{3/4} \left(\dfrac{g_*(T_{\rm rh})}{10}\right)^{1/4} & a_{\rm dom} > 2a_{\rm fz}. \end{cases} \tag{4.109}$$

If $a_{\rm rh} < 5a_{\rm fz}$, the modes that enter the horizon during the ECDE do not escape the cannibal Jeans horizon much prior to reheating. Consequently, there is no significant enhancement of DM perturbations, and the $a_{\rm rh} < 5a_{\rm fz}$ section of parameter space does not provide interesting cosmological signatures.

As the initial density of the cannibals decreases relative to the SM radiation density, the upper and lower bounds on m given by eqs. (4.108) and (4.109) become more restrictive, as seen in figure 4.12. For smaller ξ_i, larger values of $m/T_{\rm rh}$ are needed to give sufficient time for the cannibal density to overcome the SM radiation density prior to reheating. Decreasing ξ_i also increases the Hubble rate at a given cannibal temperature, so

[7] Our numerical calculations neglect cannibal decays during cannibal freezeout. For the parameter space of interest for enhanced structure formation, this is an excellent approximation, but for $a_{\rm rh} < 5a_{\rm fz}$, the impact of cannibal decays can be nonnegligible during freezeout. Thus we only show numerical results for $a_{\rm rh} > 5a_{\rm fz}$.

smaller values of m/α_c^3 are required to keep the cannibals in equilibrium after they become non-relativistic.

Figure 4.12 also illustrates how decreasing α_c shrinks the region of cannibal parameter space that yields a substantially enhanced microhalo population. If cannibal freeze-out occurs while the universe is cannibal dominated, then decreasing α_c makes it possible for lighter cannibals to freeze out before reheating (for fixed $T_{\rm rh}$). However, decreasing α_c also reduces the parameter space where the cannibal particles will freeze out while non-relativistic, so the net effect of lowering α_c is to reduce the range of m values that can realize $a_{\rm can} < a_{\rm fz} < 5a_{\rm rh}$. If cannibal freeze-out occurs during SM radiation domination, the lower bound on m is set by the requirement that $a_{\rm dom} < a_{\rm rh}$, which is independent of α_c. Consequently, in the right panels in figure 4.12, only the upper bound on m moves as α_c is changed.

If either ξ_i or α_c becomes too small, then it is not possible for particles that undergo an epoch of cannibalism to significantly affect the growth of structure because they do not dominate the universe after freezing out. Since decreasing $T_{\rm rh}$ reduces the lower bound on m given by eq. (4.109), we compare the lower and upper bounds on m at the smallest reheat temperatures allowed by BBN and the CMB constraints. With $T_{\rm rh} = 10$ MeV, it is possible to satisfy both upper and lower bounds on m if

$$\alpha_c > \begin{cases} 2\times 10^{-4}\left(\dfrac{g_*(T_{\rm rh})}{10}\right)^{1/12} & a_{\rm dom} \ll 2a_{\rm fz} \\ 10^{-3}\xi_i^{-5/3}\left(\dfrac{g_*(10m/\xi_i)}{100}\right)^{5/12}\left(\dfrac{g_*(T_{\rm rh})}{10}\right)^{1/12} & a_{\rm dom} \gg 2a_{\rm fz}. \end{cases} \tag{4.110}$$

Not all regions of cannibal parameter space that realize an ECDE are compatible with DM production through thermal freeze-out. If we suppose that the DM relic abundance is fixed by annihilations to hidden sector species (e.g., the cannibal itself), then for a given set of cannibal parameters, $\{m, T_{\rm rh}, \alpha_c, \xi_i\}$, we can solve for the value of the DM annihilation cross-section, $\langle\sigma v\rangle_{\rm DM}$, that generates the observed DM abundance. If the DM annihilation cross-section takes the form

$$\langle\sigma v\rangle_{\rm DM} \equiv \frac{\pi\alpha_{\rm DM}^2}{m_{\rm DM}^2}, \tag{4.111}$$

requiring $\alpha_{\rm DM} < 1$ for perturbativity then implies an upper bound on $m_{\rm DM}$.

We can estimate this upper bound using a sudden freezeout approximation for the DM abundance,

$$n_{\rm DM}(a_{\rm f,DM}) \equiv \frac{H(a_{\rm f,DM})}{\langle\sigma v\rangle_{\rm DM}}, \tag{4.112}$$

which defines the scale factor at DM freeze-out, $a_{\rm f,DM}$. The DM number density today is

$$n_{\rm DM}(a_0) \approx n_{\rm DM}(a_{\rm f,DM})\left(\frac{a_{\rm f,DM}}{a_0}\right)^3 = \frac{H(a_{\rm f,DM})}{\langle\sigma v\rangle_{\rm DM}}\left(\frac{a_{\rm f,DM}}{a_i}\times\frac{a_i}{a_{\rm rh}}\times\frac{a_{\rm rh}}{a_0}\right)^3. \tag{4.113}$$

As the cannibals evolve like radiation prior to a_i, we have

$$\frac{a_{\rm f,DM}}{a_i} = \frac{T_c(a_i)}{T_c(a_{\rm f,DM})} = \frac{10m}{m_{\rm DM}}x_{\rm DM}, \tag{4.114}$$

where $x_{\rm DM} \equiv m_{\rm DM}/T_c(a_{\rm f,DM})$. We use eq. (4.35) and eq. (4.40) to express $a_{\rm rh}/a_0$ and $a_{\rm rh}/a_i$ in terms of $T_{\rm rh}$

and m. Given that

$$H(a_{\rm f,DM}) = H(a_i)\frac{a_i^2}{a_{\rm f,DM}^2} = H(a_i)\frac{m_{\rm DM}^2}{(10m)^2}x_{\rm DM}^{-2}, \tag{4.115}$$

eq. (4.113) implies that

$$\langle\sigma v\rangle_{\rm DM} \approx \frac{10^{-12}}{\rm GeV^2}\sqrt{[1+g_*(10m/\xi_i)/\xi_i^4]}\left(\frac{x_{\rm DM}}{10}\right)\left(\frac{\Omega_{\rm dm}h^2}{0.12}\right)^{-1}\left(\frac{T_{\rm rh}}{10\,\rm MeV}\right)\left(\frac{\rm TeV}{m}\right), \tag{4.116}$$

where we assume $g_{*s}(T_{\rm rh}) = g_*(T_{\rm rh})$. Keeping $\alpha_{\rm DM} < 1$ then demands that

$$\left(\frac{x_{\rm DM}}{10}\right)\left(\frac{m_{\rm DM}/m}{10}\right)^2\left(\frac{\Omega_{\rm dm}h^2}{0.12}\right)^{-1}\left(\frac{T_{\rm rh}}{10\,\rm MeV}\right)\left(\frac{m}{\rm TeV}\right)\sqrt{[1+g_*(10m/\xi_i)/\xi_i^4]} < 10^4. \tag{4.117}$$

The above bound still depends on $x_{\rm DM}$. While $x_{\rm DM}$ is typically determined by inserting the equilibrium number density into Eq. (4.112), this procedure makes $x_{\rm DM}$ logarithmically dependent on $\langle\sigma v\rangle_{\rm DM}$. To avoid this dependency, we instead determine $x_{\rm DM}$ through

$$n_{\rm DM,eq}(x_{\rm DM})\frac{a_{\rm f,DM}^3}{a_0^3} = n_{\rm DM}(a_0). \tag{4.118}$$

Thus, given $m_{\rm DM}/m$, eq. (4.117) provides an upper bound on m. The red dashed line in figure 4.12 shows the upper bound on m resulting from the condition $\alpha_{\rm DM} < 1$ for $m_{\rm DM} = 10m$.[8] For larger values of m, alternative production mechanisms such as freeze-in can still generate the observed DM density [15].

As discussed in section 4.5, departures from perfect-fluid behavior for either the cannibals or DM can be important in some regions of parameter space. The impact of DM free streaming depends on its kinetic coupling to the cannibal fluid and is model-dependent. In figure 4.12 the black dashed lines show where the free-streaming horizon $\lambda_{\rm DM,fs}$, given in eq. (4.101), equals $k_{\rm pk}$ for $m_{\rm DM} = 10m$, in the case where DM kinetically decouples from the cannibal fluid after cannibal freeze-out. Above and to the left of this line, DM free-streaming, rather than cannibal interactions, can determine the peak of the transfer function.

To better illustrate when DM free-streaming can be relevant, we simplify eq. (4.101) by neglecting the temperature dependence of g_* and the DM density inside the Hubble rate. This yields

$$\lambda_{\rm DM,fs} \sim \sqrt{\frac{T_c(a_{\rm rh})}{m_{\rm DM}}}(aH)_{\rm rh}^{-1}\log(a_{\rm eq}/a_{\rm rh}) \sim \sqrt{\frac{m}{m_{\rm DM}}}k_J^{-1}(a_{\rm rh})\log(a_{\rm eq}/a_{\rm rh}). \tag{4.119}$$

In the last relation we have used the definition of the Jeans length, eq. (4.57), and the fact that $c_s^2 = 5T_c/(3m)$ and $w_c = T_c/m \ll 1$ for $a \gg a_{\rm fz}$. For scenarios where cannibals freeze out in a cannibal-dominated universe, we have $k_J^{-1}(a_{\rm rh})/k_{\rm pk}^{-1} \approx k_J(2a_{\rm fz})/k_J(a_{\rm rh})$. Consequently, DM free streaming becomes relevant when $a_{\rm rh}$ is close to $a_{\rm fz}$ and the ratio $m/m_{\rm DM}$ is not too small. In contrast, when cannibals freeze out in a SM-dominated universe, we have $k_J^{-1}(a_{\rm rh})/k_{\rm pk}^{-1} \approx k_J^{-1}(a_{\rm rh})/r_s(a_{\rm dom})$. Here, the logarithmic growth of the sound horizon until $a_{\rm dom}$ increases the gap between $k_J^{-1}(a_{\rm rh})$ and $k_{\rm pk}^{-1}$ (as seen in the top left panel of figure 4.8). In the bottom left panel of figure 4.12, this gap is large enough that the DM free-streaming horizon remains less than $k_{\rm pk}^{-1}$ for $m_{\rm DM} \geq 10m$.

[8] While decreasing $m_{\rm DM}$ relative to m relaxes the upper bound on m, our analysis assumes the DM and cannibal species to be chemically decoupled by $T_c(a_i) = 10m$, and thus we consider $m_{\rm DM} \gtrsim 10m$.

In the bottom left panel of figure 4.12, the cannibal fluid becomes collisionless prior to $a_{\rm dom}$, i.e. $a_{\rm kd} < a_{\rm dom}$, in the region right of the orange dashed line. In this regions, the peak of the transfer function is determined by the cannibal free-streaming horizon instead of the cannibal sound horizon. In computing the $a_{\rm kd} = a_{\rm dom}$ boundary shown in figure 4.12 we set $\lambda = 0$ when evaluating $\langle \sigma_c v_c \rangle$ through eq. (4.91). To see how the timing of cannibal kinetic decoupling depends on cannibal parameters more generally, we begin with the inequality $a_{\rm kd} > a_{\rm dom}$, write $a_{\rm kd} = a_{\rm kd}/a_{\rm fz} \times a_{\rm fz}$, and subsequently use eqs. (4.100), (4.43), and (4.41) for $a_{\rm kd}/a_{\rm fz}$, $a_{\rm fz}$, and $a_{\rm dom}$, respectively. The condition $a_{\rm kd} > a_{\rm dom}$ then becomes

$$m < 6 \times 10^5 \left(\frac{\alpha_c}{0.1}\right)^{11/5} \left(\frac{\xi_i}{0.4}\right)^{42/5} \left|\frac{q - 5/3}{(3q-1)^{2/3}}\right|^{8/5} \text{GeV}, \tag{4.120}$$

where $q \equiv \lambda/(g/m)^2$. This restriction on m is only relevant if it is more constraining than eq. (4.108). Consequently, the restriction on m in eq. (4.120) becomes relevant when

$$\xi_i < 0.68 \left(\frac{\alpha_c}{0.1}\right)^{1/8} \left|\frac{q - 5/3}{(3q-1)^{2/3}}\right|^{-1/4}. \tag{4.121}$$

For values of ξ_i larger than the RHS of eq. (4.121), $a_{\rm kd}$ is always greater than $a_{\rm dom}$, and cannibal free streaming does not affect $k_{\rm pk}$ in the parameter space where the cannibals significantly enhance microhalo abundance and also undergo cannibalism. Thus, there is no $a_{\rm kd} = a_{\rm dom}$ boundary in the panels with $\xi_i = 1$ or 10 in figure 4.12. As we decrease ξ_i below the RHS of eq. (4.121), a larger fraction of the parameter space has $a_{\rm kd} < a_{\rm dom}$. However, the parameter space where the cannibals significantly enhance microhalo abundance also shrinks as we decrease ξ_i, until there is no allowed parameter space for $\xi_i < 0.07(\alpha_c/0.1)^{-3/5}$ (see eq. (4.110)). Consequently, cannibal free streaming may affect $k_{\rm pk}$ only in the narrow parameter space between $0.07(\alpha_c/0.1)^{-3/5} < \xi_i < 0.68(\alpha_c/0.1)^{1/8} \left|\frac{q-5/3}{(3q-1)^{2/3}}\right|^{-1/4}$. Furthermore, as discussed in section 4.5, we expect only a marginal correction to the perfect-fluid result for $k_{\rm pk}$ in the parameter space where $a_{\rm kd} < a_{\rm dom}$.

The early-forming microhalos generated by an ECDE have large enough central densities to survive accretion into galaxies [20, 122, 123]. While sub-Earth-mass halos are too diffuse to be detected by photometric microlensing searches [124] and too small to be detected via astrometric microlensing [124–126], they can be detected by pulsar timing arrays [21, 22] and by their impact on stellar microlensing within galaxy clusters [23–25]. Furthermore, if the dark matter is a thermal relic, early-forming halos significantly boost the dark matter annihilation rate regardless of their masses, and the isotropic gamma-ray background places powerful constraints on the microhalo population [19, 20]. If dark matter annihilation is eventually detected in dwarf spheroidal galaxies, the emission profile could distinguish annihilation within microhalos from both decaying dark matter and dark matter annihilation outside of microhalos [20].

A full analysis of the observational constraints on cannibalism within a hidden sector lies beyond the scope of this article, but we can use constraints on EMDE cosmologies to forecast which regions of cannibal parameter space are likely to be probed by current and future observations. Constraints on EMDE cosmologies are often expressed in terms of a generic exponential cutoff scale: $P(k) \propto \exp[-k^2/k_{\rm cut}^2]$. Ref. [127] showed that weekly observations of 500 pulsars over 20 years with an rms timing residual of 10 ns would detect microhalos arising from an EMDE with $k_{\rm cut}/k_{\rm rh} > 20$ and $T_{\rm rh} \leq 30$ MeV. Increasing the observational period to 40 years extends the reach of pulsar timing arrays to reheat temperatures up to 100 MeV with 200 pulsars; see also Ref. [128]. The EMDE transfer function [76] implies that $T(k_{\rm pk}) \simeq 25$ for $k_{\rm cut}/k_{\rm rh} = 20$, nearly independently of the reheat temperature. If cannibal reactions freeze out during the ECDE, then

the power spectrum on scales $k \lesssim k_{\rm pk}$ is the same in EMDE and ECDE cosmologies, and ECDE scenarios with $25 \lesssim T(k_{\rm pk}) \lesssim 10^4$ would generate pulsar timing signals that are at least as strong as those produced by an EMDE with $k_{\rm cut}/k_{\rm rh} \simeq 20$. If cannibal reactions freeze out prior to cannibal domination, then the ECDE power spectrum differs more substantially from the EMDE power spectrum analyzed by Ref. [127] for $k \gtrsim k_{\rm dom}$, but we can still predict which ECDE scenarios are likely to be accessible by pulsar timing arrays. If $k_{\rm cut}/k_{\rm rh} = 20$, then the EMDE power spectrum peaks at $k_{\rm pk} \simeq 27 k_{\rm rh}$. Therefore, ECDE scenarios with $T(27k_{\rm rh}) \gtrsim 25$ and $T(k_{\rm pk}) \lesssim 10^4$ will generate a microhalo population that is at least as detectable as the microhalos that result from an EMDE with $k_{\rm cut}/k_{\rm rh} = 20$. For $\alpha_c = 0.1$, obtaining $T(27k_{\rm rh}) \gtrsim 25$ requires $T(k_{\rm pk}) \gtrsim 100$ for $\xi_i = 1$ and $T(k_{\rm pk}) \gtrsim 500$ for $\xi_i = 0.4$. Estimates of potential sensitivity from observations of cluster caustic microlensing are at a far more preliminary stage, but suggest broadly similar reach for $T_{\rm rh}$ and $T(k_{\rm pk})$ individually [25].

The best current constraints on EMDE cosmologies with thermal relic dark matter come from the isotropic gamma-ray background [19, 20]. Since the dark matter annihilation rate within microhalos does not change after the microhalos form and the microhalos track the dark matter density, dark matter annihilations within early-forming microhalos generate the same constant emission per dark matter mass as decaying dark matter. It is therefore possible to define an effective dark matter decay lifetime for these scenarios [19]:

$$\tau_{\rm eff} = \left(\frac{10^{-10}\,{\rm GeV}^{-2}}{\langle \sigma v \rangle_{\rm DM}}\right)\left(\frac{m_{\rm DM}}{10^6\,{\rm GeV}}\right)\left(\frac{7 \times 10^{38}\,{\rm seconds}}{B_0}\right), \tag{4.122}$$

where $B_0 \equiv \langle \rho_{\rm DM}^2 \rangle / \bar{\rho}_{\rm DM}^2$ is the structure boost factor generated by the microhalos. This effective lifetime should be compared to bounds on dark matter lifetime for particles with twice the mass. When accounting for emission from astrophysical sources, Ref. [129] found that Fermi-LAT observations of the IGRB [130] demand that $\tau_{\rm eff} \gtrsim 10^{28}$ seconds for $m_{\rm DM}$ between 10 GeV and 10^9 GeV and a wide range of annihilation channels.

The microhalo boost factor that arises from an EMDE has been calculated for scenarios in which all modes with $k_{\rm rh} < k < k_{\rm cut}$ enter the horizon during the EMDE [18, 20] and for scenarios that include a radiation-dominated era prior to the EMDE with $k_{\rm cut} > k_{\rm dom}$ [19]. The former case generates a sharp peak in the matter power spectrum that is qualitatively similar to the peak generated when cannibal reactions freeze out during the ECDE, while the latter generates the same plateau feature as an ECDE that starts after cannibals freeze out. However, for the limited range of $T(k_{\rm pk})$ values that were considered in both analyses, the two scenarios have values of B_0 that differ by less than a factor of 10, and much of that variation can be attributed to differing assumptions regarding the microhalo density profiles [20]. The fact that B_0 is largely insensitive to changes in reheat temperature for fixed $T(k_{\rm pk})$ further supports the conclusion that the shape of the peak in the power spectrum does not significantly affect the dark matter annihilation rate: it does not matter how the microhalos are distributed in mass as long as they have the same formation time and contain the same fraction of the dark matter, both of which are determined by $T(k_{\rm pk})$.

The ECDE scenarios shown in Figure 4.12 generally require $\langle \sigma v \rangle_{\rm DM} \gtrsim 10^{-12}$ GeV^{-2} to generate the observed DM abundance through thermal freeze-out, which implies that $m_{\rm DM} \lesssim 2 \times 10^6$ GeV is required to satisfy the unitarity bound. For these parameters, $B_0 \lesssim 10^{13}$ is required to keep $\tau_{\rm eff} > 10^{28}$ seconds if the annihilations are predominantly s-wave so that $\langle \sigma v \rangle_{\rm DM}$ is independent of the DM velocity. Refs. [18, 20] did not consider boost factors this large because they restricted their analyses to microhalos that form after matter-radiation equality, but Ref. [19] included microhalos that form during radiation domination and found that $B_0 \gtrsim 10^{13}$ for $T(k_{\rm pk}) \gtrsim 80$. However, if we only consider ECDE scenarios with $T(k_{\rm pk}) \lesssim 80$, then

$\langle\sigma v\rangle_{\rm DM} \gtrsim 10^{-10}$ GeV^{-2} and $m_{\rm DM} \lesssim 2\times10^5$ GeV. For these parameters, the IGRB bound on $\tau_{\rm eff}$ demands that $B_0 \lesssim 10^{10}$, which corresponds to $T(k_{\rm pk}) \lesssim 20$. It therefore seems likely that all of the ECDE parameter space in Figure 4.12 that contains dark matter that thermally froze out (via s-wave annihilations) prior to the ECDE is already ruled out by observations of the IGRB.

4.7 Summary and conclusions

We have shown that an early cannibal-dominated era (ECDE) leaves a distinctive peak in the matter power spectrum. Perturbation modes that enter the horizon after the freeze-out of cannibal reactions but before the end of the ECDE are enhanced. On smaller scales, the pressure generated by the self-heating of the cannibal particles suppresses the growth of dark matter perturbations. Consequently, the properties of the cannibal field generally establish the minimum halo mass following an ECDE. We show how the properties of the cannibal field establish the minimum halo mass even if the cannibal reactions freeze out prior to cannibal domination.

Cannibals are generically predicted in theories with thermally decoupled hidden sectors that have a mass gap and a number-changing self-interactions. If the lightest particle in such a hidden sector remains in chemical equilibrium after it becomes non-relativistic, it undergoes a period of cannibalism. During cannibalism, the cannibal number-changing self-interactions convert the particles' rest-mass energy into kinetic energy to maintain chemical equilibrium while conserving entropy within the cannibal fluid. The period of cannibalism ends when the rate of number-changing self-interaction falls below the Hubble rate. Such a cannibal fluid can easily come to dominate the energy density of the universe even if the hidden sector was initially colder than the SM bath. The ECDE ends when the cannibal particles decay into relativistic SM particles prior to the onset of BBN.

During the ECDE, we find that sub-horizon cannibal density perturbations grow linearly with the scale factor on scales that are larger than the cannibal Jeans length. The DM perturbations follow the cannibal density perturbations because the DM particles fall into the gravitational potential wells formed by the cannibals. Consequently, the enhancement of the DM perturbations after an ECDE relative to those in the standard cosmology reflects the cannibal perturbation spectrum and contains information about the cannibal self-interactions. This enhancement of the DM perturbations due to an ECDE is unaffected by possible scattering between the DM and the cannibals.

Enhanced small-scale DM perturbations collapse earlier than they otherwise would and hence lead to an enhanced population of halos at high redshift. Since an ECDE only affects perturbations on scales that enter the horizon during the ECDE, perturbations on these scales form microhalos with masses far less than the mass of the Sun. The characteristic mass of the earliest-forming microhalos, $M_{\rm pk}$, is determined by the scale with the largest enhancement in DM perturbations ($k_{\rm pk}$) whereas the formation time of these microhalos is determined by the amplitude of the enhancement, which is given by $T(k_{\rm pk})$.

The location of the peak of the DM power spectrum is determined by the process that counteracts gravitationally induced growth and prevents structure formation on small scales. In earlier works that have studied microhalo formation due to an early matter-dominated era (EMDE), this cutoff in the matter power spectrum was assumed to be generated by DM free streaming [18, 76–78] or axion DM oscillations [24]. Consequently, the peak scale is determined by DM microphysics. If the DM belongs to a hidden sector whose lightest particle causes the EMDE, then the DM particle may be cold enough that the relativistic pressure of the lightest hidden-sector particle sets the small-scale cutoff [19, 74]. We showed here that the cutoff in the

matter power spectrum following an ECDE is typically generated by the thermal pressure in the cannibal fluid and is independent of DM microphysics when there is no period of SM radiation domination prior to the ECDE. We then extended our analysis to scenarios in which the cannibals freeze-out while cannibal density is subdominant to SM density and showed that the cannibal thermal pressure still determines the cutoff. We find the cutoff scale to be given by the cannibal sound horizon at reheating, up to an order of magnitude, irrespective of the initial temperature ratio between the cannibal fluid and SM plasma and the properties of the DM particles. The only exceptions occur in narrow bands of parameter space where the DM free-streaming horizon overcomes the cannibal sound horizon or where the cannibal fluid becomes collisionless prior to cannibal domination.

While the cannibal sound horizon provides a rough estimate of the wavenumber at which the power spectrum is maximized, $k_{\rm pk}$, we have also found a more accurate (within $\sim 10\%$) expression for $k_{\rm pk}$. In the case where the freeze-out of cannibal reactions occurs during cannibal domination we find $k_{\rm pk} \approx k_J(2a_{\rm fz})/1.4$, where k_J^{-1} is the cannibal Jeans length, and $a_{\rm fz}$ is the scale factor when cannibal reactions freeze out. If the freeze-out of cannibal reactions occurs during SM radiation domination, then we find $k_{\rm pk}$ to be given by eq. (4.78). These analytical estimates allow us to provide a map between the key microhalo properties, $M_{\rm pk}$ and $T(k_{\rm pk})$, and the cannibal particle properties.

The peak amplitude of enhancement in dark matter perturbations due to an ECDE, $T(k_{\rm pk})$, is determined by how long cannibals dominate the universe after they freeze out, which depends on the cannibal mass, m, its $3 \to 2$ reaction coupling, α_c, its initial temperature relative to the SM, ξ_i, and the reheat temperature, $T_{\rm rh}$. A longer period of post-freeze-out cannibal domination leads to larger values of $T(k_{\rm pk})$ and earlier microhalo formation. If $T(k_{\rm pk}) \gtrsim 10^4$, the cannibals and DM particles assemble into halos prior to reheating. After the cannibals decay, the DM particles are released from these halos with sufficient velocity that their subsequent motion erases nearly all the perturbations that are within the horizon at reheating [19]. Consequently, an ECDE will generate a significantly enhanced abundance of microhalos for $5 \lesssim T(k_{\rm pk}) \lesssim 10^4$. Since $T(k_{\rm pk})$ is roughly proportional to the ratio $m/T_{\rm rh}$, a band of $m/T_{\rm rh}$ values is expected to yield an enhanced microhalo population. The upper and lower limits of this band are fixed by α_c and ξ_i. The range of possible values for m, $T_{\rm rh}$, α_c, and ξ_i, is further constrained by the requirement that cannibals undergo cannibalism and that reheating occurs early enough to avoid altering the neutrino abundance, which would spoil the success of BBN [98–101] and alter the anisotropies in the CMB [102, 103]. Thus, we have identified a bounded region in the cannibal parameter space that produces an enhanced abundance of microhalos due to an ECDE. Within this parameter space, we provide estimates for the masses of the earliest-forming halos and their formation times in terms of the properties of the cannibal field.

Finally, we briefly discussed potential observational sensitivity to this enhanced microhalo population. We expect the microhalos generated by ECDEs with reheat temperatures up to $T_{\rm rh} \simeq 100$ MeV with $T(27k_{\rm rh}) \gtrsim 25$ and $T(k_{\rm pk}) \lesssim 10^4$ to be detectable in the future pulsar timing arrays analyzed in Refs. [127, 128], while the results of Refs. [19, 20, 129, 130] imply that the observed IGRB likely excludes s-wave thermal relic DM in almost all ECDE scenarios. Cluster caustic microlensing is a promising alternative gravitational means of detecting the ECDE-enhanced population of microhalos in the low-redshift universe, but projections for such observations are not developed enough to allow for similarly definitive statements.

It is important to remember, however, that all of these observational probes are sensitive to the internal structure of the microhalos. While it is possible to predict the density profiles of the first microhalos from the matter power spectrum [120], it is unknown how subsequent mergers between microhalos and their further evolution within galactic halos affect their internal structure. Analyses that employ different assumptions

regarding the microhalos' density profiles, substructure, and survival rate give similar but not identical bounds on EMDE cosmologies. There is also a great deal of uncertainty regarding how the gravitational heating of the dark matter following structure formation during the EMDE or ECDE affects the subsequent formation of microhalos [19], and it has been suggested that microhalo remnants could persist through reheating [131]. Therefore, we cannot yet establish robust observational constraints on cannibalism within a hidden sector. Nevertheless, we have identified which regions of cannibal parameter space enhance the microhalo abundance, which demonstrates how observations of small-scale structure provide a window into the evolution and particle content of the early Universe.

Chapter 5

$N_{\rm eff}$ constraints on portal couplings with hidden sectors

5.1 Introduction

Next-generation cosmic microwave background (CMB) experiments, such as CMB-Stage 4 [132], will measure the details of the acoustic peaks in the microwave background with unprecedented accuracy. These measurements will result in subpercent-level determinations of the contents and geometry of the Universe. In particular, the fidelity with which the locations of the acoustic peaks are forecast to be determined will improve the measurement of the energy density in free-streaming radiation, parametrized by the effective number of neutrino species, $N_{\rm eff}$, by almost an order of magnitude. Future CMB experiments, beyond Stage-4, aim to reach a threshold of $\sigma_{N_{\rm eff}} < 0.027$, where any new relativistic beyond-the-Standard Model (BSM) particle must be always out of equilibrium with the Standard Model (SM) in the early Universe [133] if the measured central value agrees with the SM prediction of $N_{\rm eff}^{\rm SM} = 3.044$ [34, 35, 134–136]. A measurement of $N_{\rm eff}$ that deviates from $N_{\rm eff}^{\rm SM}$ would be compelling evidence of physics beyond the standard model. Conversely, models that require additional light states must be coupled to the SM in such a way that does not violate bounds on $N_{\rm eff}$.

Constraints on new relativistic degrees of freedom through $N_{\rm eff}$ are often restated as a constraint on the decoupling temperature at which any BSM relativistic particle must lose thermal contact with the SM plasma in the early universe (see, e.g., Refs. [137–140]). For an out-of-equilibrium relativistic particle, measurements of $N_{\rm eff}$ can be used to constrain the total energy transferred between BSM relativistic particles and the SM plasma in the early Universe, and can provide a powerful probe of the interactions of the SM with light, feebly-interacting particles.

Our primary interest in this work is the case where a SM singlet mediator particle has renormalizable couplings to both the SM and the dark radiation species. This scenario is ultraviolet (UV) insensitive insofar as it yields interaction rates that grow more rapidly than the Hubble rate as the universe expands, provided that the SM temperature remains larger than the mediator mass. This UV insensitivity means that the asymptotic dark radiation density predicted in these models does not depend on the unknown early thermal history of our universe provided the reheating temperature is above the mediator mass. In this work, we focus on mediator masses $m > 0.1$ MeV where thermal production in the early universe provides one of the leading avenues to test these models. Constraints from stellar cooling are typically stronger than cosmological

constraints for masses $m < 0.1$ MeV [140, 141].

Similar UV-insensitive and out-of-equilibrium dark radiation production has been explored earlier in the context of specific models. For instance, in the case of axions, freeze-in production can receive important contributions from both heavy states in the UV completion [142] and fermion annihilation, which proceeds through infrared-dominated processes below the scale of electroweak symmetry breaking [143–146]. BSM neutrino model-building can also yield sizeable out-of-equilibrium dark radiation production [147, 148]. Meanwhile LHC searches can provide a complementary window onto the freeze-in of dark radiation in scenarios where a weak-scale mediator carries SM charge [149].

In this chapter, we study the production of dark radiation in minimal BSM models that consist of a massive ($m > 0.1$ MeV) SM gauge singlet mediator coupled to new light degrees of freedom. We begin by considering two well-motivated extensions to the SM: a millicharged particle (MCP) model [150], and a model where the SM baryon-number-minus-lepton number ($B - L$) symmetry is gauged [151, 152]. In the MCP model, a dark photon that kinetically mixes with SM hypercharge is the dark radiation and the MCP is the mediator. In the gauged $B - L$ model the three right-handed neutrinos required to cancel gauge anomalies are the dark radiation, while the new $B - L$ gauge boson is the massive mediator. By developing and solving the relevant Boltzmann equations, we use the production of dark radiation in these models to place constraints on the strength of their interactions with the SM. We update constraints on the minimal MCP model given in Refs. [153, 154] and present forecasts for future CMB observatories. We further demonstrate that future CMB experiments will be able to rule out (or discover evidence for) the extended model proposed by Ref. [155] to explain the EDGES anomaly. For the $B - L$ model, we improve on the analysis of Ref. [156] by incorporating two further effects that lead to more stringent constraints in the unequilibriated regime. In particular, we take into account the out-of-equilibrium production of right-handed neutrinos, and further show that the out-of-equilibrium decays of the $B - L$ gauge bosons lead to a more powerful constraint on the $B - L$ coupling in the relevant regions of parameter space.

In the process of deriving these results, we develop a number of approximations which allow us to analytically solve the Boltzmann equations in the regions of parameter space where the new light degrees of freedom are out of equilibrium with the SM. We use these solutions to argue, on general grounds, that a conservative lower bound on the dark radiation density can be quickly obtained for a generic class of hidden sectors containing light degrees of freedom that interact with the SM via a heavier SM gauge singlet mediator. The lower bound is governed by the properties of the mediator and is insensitive to the details of the hidden sector, such as the number of degrees of freedom and their internal interactions, and relies solely on the assumption that the mediator preferentially transfers its energy into the HS rather than the SM. This amounts to assuming that the mediator interacts more strongly with the HS than the SM.

This chapter is organized as follows. In sections 5.2 and 5.3, we study dark radiation production in the MCP and gauged $B - L$ models, respectively. We develop and solve the relevant Boltzmann equations to find the allowed regions of parameter space given current and projected CMB constraints on $N_{\rm eff}$. In both models, we develop approximations that allow us to analytically solve the Boltzmann equation in various regimes. In section 5.4, we consider the applicability of dark radiation constraints to generic classes of hidden sectors containing relativistic particles. We conclude in section 5.5. The details of many of our computations are relegated to appendices. In appendix A.5 we describe various processes transferring energy between the SM and the dark photons in the MCP model, and similarly in appendix A.6 we describe processes transferring energy from the SM into right-handed neutrinos in the gauged $B - L$ model. Finally, in appendix A, we simplify the phase space integral of the energy transfer collision terms for generic annihilations, decays, and

elastic scatterings, while taking into account the quantum statistical distributions of relevant particles.

5.2 Millicharged particle model

In this section we derive constraints on the allowed parameter space of a MCP model from CMB measurements of $N_{\rm eff}$. In this model, a massless dark photon kinetically mixes with the SM hypercharge gauge boson, while the MCP is a massive Dirac fermion charged under the dark $U(1)$.

MCP models have recently been explored in detail as potential explanations of the anomalously small spin temperature of the hydrogen atoms inferred from the 21 cm signal measured by the EDGES experiment [155, 157–160]. This anomaly can be resolved if the baryons were cooled by scattering with DM particles. In the scenario where the MCP comprises some of the dark matter, the millicharge interactions can cool the baryons to explain the EDGES anomaly. However, the required values of the millicharge, Q, are ruled out by a combination of bounds from the CMB and e^+e^- colliders [159].[1] Recently, an extension of the minimal MCP model was proposed with multiple millicharged fermions that could resolve the EDGES anomaly while evading current constraints [155]. In this section we both update the current CMB constraints on the minimal MCP model and show that measurements of $N_{\rm eff}$ from future CMB experiments will provide a stringent test of these extended MCP models.

This section is organized as follows. We begin by describing the MCP model in section 5.2.1. In section 5.2.2, we describe the relevant Boltzmann equations and solve them to find the region of parameter space that saturates the $N_{\rm eff}$ bounds from current and upcoming CMB experiments, updating the results of [153]. Next, in section 5.2.3 we go into more detail about the physics responsible for the production of dark radiation, and the relevant features of the resulting parameter space constraints from $N_{\rm eff}$ measurements. Finally, in section 5.2.4, we show how these constraints can be extended to models with multiple MCPs in a detail-insensitive way. We then apply these constraints to the MCP model proposed by Ref. [155] and show that measurements of $N_{\rm eff}$ at the level of accuracy forecast by CMB-S4 can potentially rule out this explanation of the EDGES anomaly.

5.2.1 The millicharged particle model

The MCP model is an extension of the SM that contains a massless dark photon, A'_μ, and an additional Dirac fermion, ψ, with mass m. The dark photon kinetically mixes with the SM hypercharge gauge boson, A_μ, and the Dirac fermion has charge e' under the dark $U(1)$. The relevant interactions for our study are

$$\mathcal{L}_{\rm int} = -\frac{\epsilon}{2} B^{\mu\nu} F'_{\mu\nu} + e J^\mu_{\rm EM} A_\mu + e J^\mu_Z Z_\mu + e' \bar{\psi} \gamma^\mu \psi A'_\mu, \tag{5.1}$$

where $B_{\mu\nu}$ is the hypercharge field strength, Z_μ is the Z boson, $J^\mu_{\rm EM}$ is the electromagnetic current, and J^μ_Z is the weak neutral current.

We work in the basis where the gauge boson kinetic terms are diagonal and where $J^\mu_{\rm EM}$ and J^μ_Z do not couple to the dark photon. Thus the dark photon remains 'dark'. After performing the relevant redefinitions

[1] If the baryons are cooled by a millicharged dark fermion that is not coupled to dark radiation, then one can explain the EDGES result if the dark fermions compose a 0.4% fraction of dark matter [161–164]. However, Ref. [165] found that this solution is incompatible with the constraints on the millicharge and dark fermion mass imposed by its production history in the early universe.

of the A and A' fields and considering the limit of weak kinetic mixing, $\epsilon \ll 1$, the interaction Lagrangian is

$$\mathcal{L}_{\rm int} \approx e\left(J^{\mu}_{\rm EM} - Q\bar{\psi}\gamma^{\mu}\psi\right)A_{\mu} + e'\bar{\psi}\gamma^{\mu}\psi A'_{\mu} + Qe\tan\theta_W\bar{\psi}\gamma^{\mu}\psi Z_{\mu}, \tag{5.2}$$

where θ_W is the weak mixing angle, and the dark fermion has obtained a millicharge, Q, given by

$$Q \equiv \epsilon\frac{e'}{e}\cos\theta_W. \tag{5.3}$$

While the dark photon does not directly couple to SM degrees of freedom, dark photons are produced by annihilations of millicharged fermions, which themselves are produced by interactions with the SM plasma in the early Universe. In this work, we consider the regime where the fermion mass is $m > 0.1$ MeV; stellar cooling observations provide the dominant constraint for smaller masses [153].

5.2.2 Evaluation of the dark radiation density and the constraints on the model

Dark photons contribute to the energy budget of the Universe as radiation, and their presence in the early Universe is constrained by measurements of the effective number of (free-streaming) relativistic species, $N_{\rm eff}$. Specifically, dark photons shift the value of $N_{\rm eff}$ away from its SM value of $N^{\rm SM}_{\rm eff} = 3.044$, by

$$\Delta N_{\rm eff} \equiv N_{\rm eff} - N^{\rm SM}_{\rm eff} = \frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\frac{\rho_{A'}}{\rho_{\gamma}}, \tag{5.4}$$

where $\rho_{A'}$ and ρ_{γ} are the energy densities of the dark photon and the SM photon, respectively. The dark photon energy density $\rho_{A'}$ during recombination is controlled by Q, m, and e', and thus measurements of $N_{\rm eff}$ can be translated into constraints on the parameter space of the model.

We demonstrate below that, for the regions of parameter space that lead to dark radiation densities that saturate the bounds on $\Delta N_{\rm eff}$ from upcoming experiments, the dark charge e' must be large enough to enable almost all the MCPs to efficiently annihilate. In this limit, the final dark photon abundance is insensitive to the value of e'. Moreover, due to the tight coupling of the MCPs to the dark photons, the hidden sector (HS) thermal bath comprising the MCP and the dark photon is well-approximated by a fluid in chemical equilibrium. Thus, instead of solving for the individual MCP and dark photon abundances, we can solve for the combined HS energy density through the Boltzmann equations

$$\begin{aligned}\frac{d\rho_{\rm SM}}{dt} + 3H(1+w_{\rm SM})\rho_{\rm SM} &= -\mathcal{C}\\ \frac{d\rho_{\rm HS}}{dt} + 3H(1+w_{\rm HS})\rho_{\rm HS} &= \mathcal{C},\end{aligned} \tag{5.5}$$

where $\mathcal{C}$ is the energy transfer collision term due to millicharge interactions, ρ is the energy density, $H = \sqrt{\rho_{\rm HS}+\rho_{\rm SM}}/[\sqrt{3}M_{\rm Pl}]$, $w = \mathcal{P}/\rho$ is the equation of state, $\mathcal{P}$ is the pressure, and $M_{\rm Pl} = 2.435\times10^{18}$ GeV is the reduced Planck mass. After the MCPs become non-relativistic and annihilate into dark photons, $\rho_{\rm HS} \approx \rho_{A'}$.

Both $\rho_{\rm HS}$ and $w_{\rm HS}$ are determined in terms of $T_{\rm HS}$ by

$$\rho_{\rm HS} = \frac{\pi^2}{30}g_{\rm HS}T^4_{\rm HS}, \qquad w_{\rm HS} = \frac{g_{{\rm HS},p}}{3g_{\rm HS}}, \tag{5.6}$$

where

$$g_{\rm HS} = 2 + \frac{30}{\pi^2 T_{\rm HS}^4} \times 4 \int_0^\infty \frac{d^3p}{(2\pi)^3} E \frac{1}{\exp(E/T_{\rm HS}) + 1}, \tag{5.7}$$

$$g_{{\rm HS},p} = 2 + \frac{90}{\pi^2 T_{\rm HS}^4} \times 4 \int_0^\infty \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} \frac{1}{\exp(E/T_{\rm HS}) + 1}. \tag{5.8}$$

Similarly, $w_{\rm SM}$ is related to $T_{\rm SM}$ via

$$w_{\rm SM} = \frac{g_{*p}(T_{\rm SM})}{3 g_*(T_{\rm SM})}, \tag{5.9}$$

where g_* and g_{*p} count the effective degrees of freedom in the SM energy density and the SM pressure, respectively. We model the QCD phase transition using the g_* tables from Ref. [166] for $T_{\rm SM} > 100$ MeV.

The collision term in eq. (5.5) includes all processes that transfer energy from the SM plasma into the HS bath due to the millicharge interactions. There are four important processes contributing to energy transfer: (1) SM fermion annihilation into MCPs; (2) Z-boson decays into MCPs; (3) plasmon decays into MCPs; and (4) Coulomb scattering of SM fermions with MCPs.[2] We include the quantum statistical distributions of SM particles while deriving the collision term for each of these processes, relegating the details to appendix A.5. The use of quantum statistics instead of Maxwell-Boltzmann distributions provides a $\sim 20\%$ correction to the net energy transfer. Among the three s-channel processes (numbers 1-3 above), we find that energy transfer via fermion annihilation dominates over the other two in the bulk of parameter space. For instance, the energy transferred by fermion annihilations dominates over that from Z-boson decays except for the region of parameter space where 1 GeV $\lesssim m \lesssim$ 40 GeV. The energy transferred via plasmon decays is typically around $\sim 20\%$ of that transferred by fermion annihilations.

Finally, we find that the energy transferred by Coulomb scattering dominates over that from fermion annihilations for the values of Q and m that saturate the bounds on $\Delta N_{\rm eff}$ from both current and upcoming experiments. Naively, one might expect energy transfer via Coulomb scattering to be subdominant in the out-of-equilibrium regime because these processes are suppressed by the small MCP abundance in the initial state. However, due to the forward-scattering singularity, the energy transfer via Coulomb scattering dominates over that via SM fermion annihilations for $T_{\rm HS}$ as low as $0.35 T_{\rm SM}$ (for more detail, see appendix A.5). Temperature ratios of $T_{\rm HS}/T_{\rm SM} > 0.35$ during recombination produce enough dark radiation to shift $\Delta N_{\rm eff} > 0.06$, which can be detected in the upcoming CMB-S4 experiments [132]. Hence, Coulomb scattering processes are key for evaluating the dark radiation densities relevant for the values of $\Delta N_{\rm eff}$ that can be tested in upcoming as well as current experiments.

To determine the relic dark radiation density, and the resulting parameter space allowed by $N_{\rm eff}$ measurements, we solve the Boltzmann equations in eq. (5.5) from an initial SM temperature $T_{\rm SM} \gg m$ until the energy injection into the HS ends, $T_{\rm SM} \ll m$. We assume the HS is initially empty, which provides a conservative constraint; any significant initial HS density only increases the final dark radiation density and thus $\Delta N_{\rm eff}$.

The various shaded regions in figure 5.1 show the regions of parameter space where the resulting energy density in dark radiation exceeds various current and future experimental sensitivities to shifts in $N_{\rm eff}$. The constraint contours saturate the current one-sided 2σ upper bound from Planck [3] $\Delta N_{\rm eff} = 0.3$ (black solid), the projected 2σ sensitivity for CMB-S4 [132] $\Delta N_{\rm eff} = 0.06$ (gray solid), and the threshold goal for future

[2] Energy transfer from Compton-like scattering, $A + \psi \to \psi + A'$, can be more important than the processes mentioned here for large values of the dark coupling constant, $e' > 0.9$. We neglect this process for simplicity and genericity.

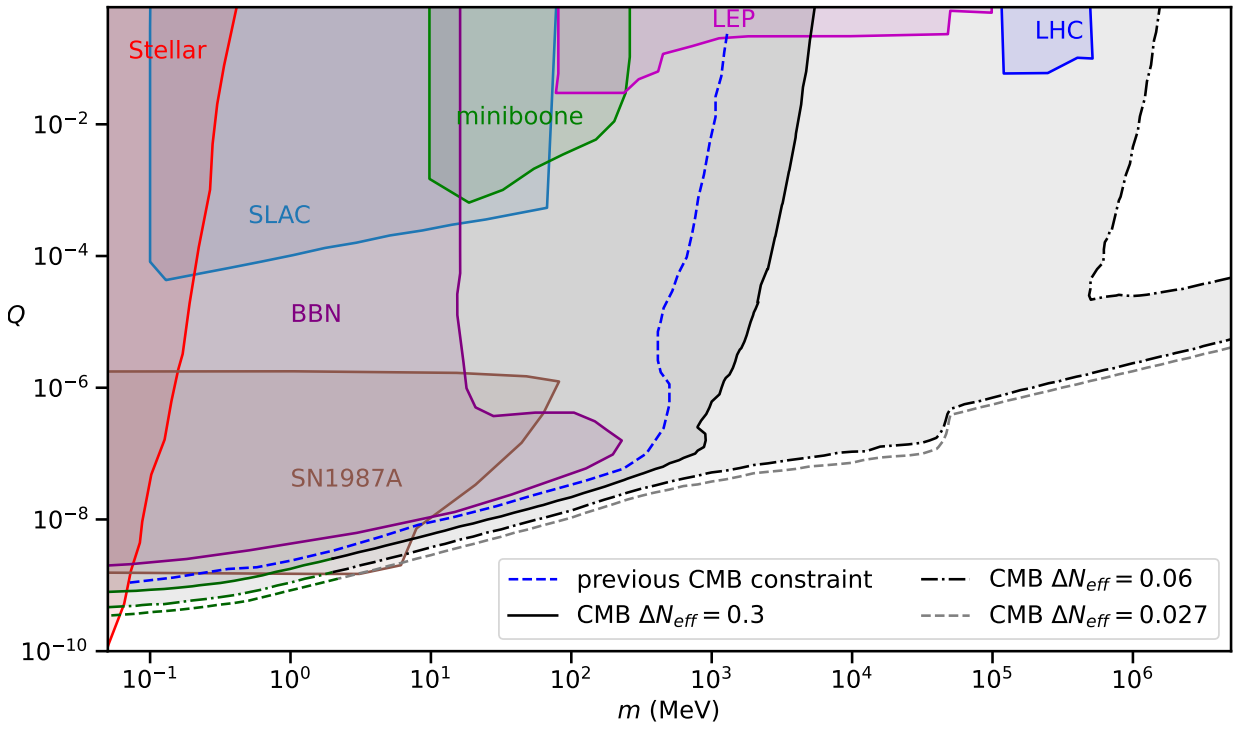


Figure 5.1: Constraints on the mass and millicharge of the millicharged particle. The black solid, gray solid, and gray dashed contours indicate the parameter space that yields $\Delta N_{\rm eff} = 0.3$, $\Delta N_{\rm eff} = 0.06$, and $\Delta N_{\rm eff} = 0.027$, respectively. The blue dashed contour is the CMB constraint derived in Ref. [153] for $\Delta N_{\rm eff} = 0.8$. The green color on our $\Delta N_{\rm eff}$ contours marks the region where we expect the millicharge interactions to cool the electron-photon bath relative to neutrinos and strengthen our constraints by an $\mathcal{O}(1)$ factor. Also shown are constraints from SLAC [167], MiniBooNE [168], LEP [169], LHC [170], BBN [153], supernova 1987A [171] and stellar observations [153].

CMB experiments $\Delta N_{\rm eff} = 0.027$ (gray dashed). The threshold of $\Delta N_{\rm eff} = 0.027$ physically corresponds to the shift in $N_{\rm eff}$ due to the energy density at recombination in a relativistic spin-zero particle that was in thermal equilibrium with the SM in the early Universe and decoupled while all SM species were relativistic.

For comparison, in figure 5.1 we also display the results of Ref. [153] as the blue dashed line, which shows the parameter points that lead to $\Delta N_{\rm eff} = 0.8$. We have verified that our results agree with Ref. [153] within $\mathcal{O}(1)$ when we assume Maxwell-Boltzmann statistics for all particles. The use of Maxwell-Boltzmann statistics as opposed to Fermi-Dirac statistics overestimates the dark radiation density by around 20%. As the energy density in dark radiation depends on Q^2, using Fermi-Dirac statistics for SM fermions weakens the $N_{\rm eff}$ constraint on Q by around 10%.

In figure 5.1 we also show the constraints on the MCP from collider experiments, stellar evolution, and supernova observations (see Ref. [172] for a review). We omit limits from direct detection experiments because those constraints are dependent on the interaction of MCPs with the magnetic fields in the galaxy [173, 174]. Among the displayed constraints, the current Planck limit is already the dominant bound in a substantial portion of parameter space, while upcoming CMB observations will provide the strongest constraint for the entire region with $m \gtrsim 0.1$ MeV, assuming no deviation is observed from the SM value of $N_{\rm eff}^{\rm SM} = 3.044$.

The curves of constant $\Delta N_{\rm eff}$ in the MCP model parameter space shown in figure 5.1 have four key features. First, at low masses, the contours of constant dark radiation density at recombination (and therefore constant $\Delta N_{\rm eff}$) relate the millicharge, Q, to the MCP mass via $Q \propto \sqrt{m \Delta N_{\rm eff}}$. In this region the HS is out-of-equilibrium with the SM. Second, as one moves along the contour of constant $\Delta N_{\rm eff}$ toward increasing m, one reaches a threshold mass $m_{\rm th}$ where the millicharge Q becomes large enough that the HS thermalizes with the SM. When the HS is thermalized with the SM plasma, the net energy transfer between sectors becomes insensitive to the specific value of Q since forward and backward processes balance each other. In this regime the asymptotic dark radiation density, and therefore the constraint from $\Delta N_{\rm eff}$, depends primarily on m and only logarithmically on Q, as seen in the figure.

Third, the contour corresponding to $\Delta N_{\rm eff} = 0.06$ has a narrow exclusion region (where $\Delta N_{\rm eff} > 0.06$) extending from $m \approx m_{\rm th}$ up to arbitrarily large m, while no such excluded strip exists for either the $\Delta N_{\rm eff} = 0.3$ or the $\Delta N_{\rm eff} = 0.027$ contours. The existence (or non-existence) of this strip beyond the threshold mass is related to the fact that $\Delta N_{\rm eff} < 0.06$ still allows the dark photon itself to have been in equilibrium with the SM plasma for temperatures above the TeV scale, but is not compatible with the MCP also having entered equilibrium, which would increase the hidden sector relativistic degrees of freedom to an unacceptably large value at early times. Finally, the $\Delta N_{\rm eff} < 0.06$ and $\Delta N_{\rm eff} < 0.027$ constraints have a bump below $m \sim m_Z/2$ which is due to energy injection from on-shell Z-boson decays. In the following subsection we elaborate on this discussion by analytically solving the Boltzmann equations in the relevant regimes.

For $m \lesssim 2$ MeV, energy transfer into the HS occurs predominantly after neutrino decoupling. In this part of parameter space, the production of dark photons as well as the relative cooling of the electron-photon bath compared to neutrinos contributes to $\Delta N_{\rm eff}$ during recombination, while our analysis only considers the contribution from dark photons. Taking into account the relative cooling of photons should further strengthen the $N_{\rm eff}$ constraints calculated in this study by an $\mathcal{O}(1)$ factor. We indicate this region in figure 5.1 by coloring the $N_{\rm eff}$ contours green. A full treatment of early universe constraints on the MCP model below $m \lesssim 2$ MeV requires a detailed treatment of neutrino decoupling as well as light element formation during BBN, and is beyond the scope of this work.

5.2.3 Dark radiation production in different regimes

The parameter space that saturates the bounds on $N_{\rm eff}$ can be separated into two distinct regions: a region where the HS remains out of equilibrium with the SM plasma and a region where the HS thermalizes with the SM. In this subsection we focus on the evolution of the HS energy density, $\rho_{\rm HS}$, in these two regions of parameter space. By studying the Boltzmann equations, we develop approximate analytic descriptions that enable a deeper understanding of the shapes of the curves in figure 5.1.

Collisions, redshifting, and the evolution of $\rho_{\rm HS}$: The evolution of the energy density in dark radiation is controlled by two factors. The first is the (net) rate at which energy is injected into the HS, $\mathcal{C} = \mathcal{C}_{\rm f} - \mathcal{C}_{\rm b}$, where $\mathcal{C}_{\rm f}$, and $\mathcal{C}_{\rm b}$ are the forward and backward collision terms describing energy transfer from the SM into the HS. The second factor is the rate at which the energy density is redshifting, $H\rho_{\rm HS}$. The ratio $\mathcal{C}_{\rm f}/H$, then, indicates the energy density transferred to the HS within a Hubble time. When $\rho_{\rm HS}$ is out of equilibrium with the SM, $\mathcal{C}_{\rm f}/H$ serves as a useful indicator of whether energy injection is important ($\mathcal{C}_{\rm f}/H > \rho_{\rm HS}$) or not ($\mathcal{C}_{\rm f}/H < \rho_{\rm HS}$) in governing its evolution. When $\rho_{\rm HS}$ is in equilibrium with the SM, $\rho_{\rm HS,eq} = [\pi^2 g_{\rm HS}(T_{\rm SM})/30](T_{\rm SM})^4$ and $\mathcal{C} = \mathcal{C}_{\rm f} - \mathcal{C}_{\rm b} = 0$, as both forward and backward rates become large. The HS remains in equilibrium with the SM plasma as long as the fractional energy injection rate, $\Gamma_E \equiv \mathcal{C}_{\rm f}/\rho_{\rm HS,eq}$, is larger than H.

To develop some intuition about the evolution of these rates, and their impact on the resulting dark radiation density, in figure 5.2 we show the evolution of $\rho_{\rm HS}a^4$ (black line) along with $\mathcal{C}_{\rm f}a^4/H$ (blue dot-dashed line) after numerically solving the Boltzmann equations given in eq. (5.5). The red-dashed line shows the evolution of $\rho_{\rm HS,eq}a^4 = [\pi^2 g_{\rm HS}(T_{\rm SM})/30](aT_{\rm SM})^4$.[3] The ratio between the black and red lines is proportional to $(T_{\rm HS}/T_{\rm SM})^4$ and thus indicates how far away the HS is from equilibrating with the SM plasma. Two parameter choices are shown to illustrate the two different regimes for computing the resulting dark radiation density. The left panel shows a parameter point where $\mathcal{C}_{\rm f}/H$ is always smaller than $\rho_{\rm HS,eq}$, and consequently the HS remains out-of-equilibrium with the SM plasma. The right panel shows a second choice of parameters where the HS comes into thermal equilibrium with the SM for some period of time, indicated by the overlapping red and black lines. In both panels, the initial hidden sector energy density is small compared to the energy injection from the SM, $\rho_{\rm HS} < \mathcal{C}_{\rm f}/H$, and the evolution of $\rho_{\rm HS}$ is driven by the energy injection, giving the initial increase in $\rho_{\rm HS}a^4$.

In the left panel of figure 5.2, energy injection into the HS ceases to be important after $\mathcal{C}_{\rm f}$ becomes Boltzmann-suppressed and $\mathcal{C}_{\rm f}/H$ falls below $\rho_{\rm HS}$. In particular, $\mathcal{C}_{\rm f}a^4/H$ attains its maximum around $T_{\rm SM} = m/2$, but it is not until $T_{\rm SM} = m/4$ (yellow dashed line) that energy injection into the HS effectively ends. For this choice of parameters, the HS does not come into thermal equilibrium with the SM, and consequently the final value of $\rho_{\rm HS}a^4$ can be estimated from the maximum value of $\mathcal{C}_{\rm f}a^4/H$. As $\Delta N_{\rm eff}$ parametrizes the energy density of dark photons, it constrains the maximum value of $\mathcal{C}_{\rm f}a^4/H$, which is proportional to $Q^2 M_{\rm Pl}/m$.

In the right panel, $\mathcal{C}_{\rm f}/H$ grows until it exceeds $\rho_{\rm HS,eq}$ and subsequently the HS thermalizes with the SM plasma. The two sectors remain in equilibrium until $\mathcal{C}_{\rm f}/H$ falls below $\rho_{\rm HS,eq}$. The final value of $\rho_{\rm HS}a^4$ is given by $\rho_{\rm HS,eq}a^4$ evaluated at T_d, where T_d is the temperature below which HS thermally decouples from

[3] We use the same scale factor for both the red dashed and black solid lines, which is obtained after numerically solving for $\rho_{\rm HS}$ indicated by the black line. The red line should not be confused with the solution for the comoving energy density for a HS always in thermal equilibrium. The bump in the red line near $T_{\rm SM} = 200$ MeV is due to the sudden decrease in g_* below the QCD phase transition. The red line decreases for $T_{\rm SM} \lesssim m/4$ because the degrees of freedom in the HS decreases when MCPs become non-relativistic.

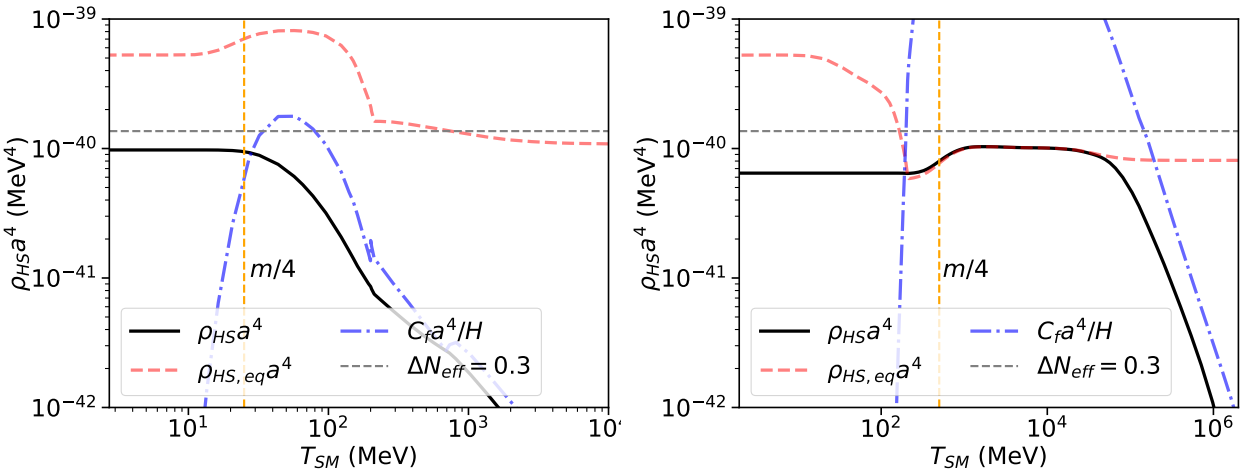


Figure 5.2: Evolution of the comoving HS energy density (solid black) as a function of SM temperature for MCP mass and charges $\{m, Q\} = \{10^2 \text{ MeV}, 2 \times 10^{-8}\}$ (left) and $\{m, Q\} = \{10^4 \text{ MeV}, 8 \times 10^{-5}\}$ (right). The red dashed line shows $(\pi^2/30) g_{\rm HS}(T_{\rm SM}) T_{\rm SM}^4 a^4$, and the blue dot-dashed line is $\mathcal{C}_{\rm f} a^4/H$. The vertical orange dashed line marks $T_{\rm SM} = m/4$. The horizontal black dashed line marks the dark radiation density that produces $\Delta N_{\rm eff} = 0.3$. For an out-of-equilibrium HS, the final dark radiation density depends on the maximum value of $\mathcal{C}_{\rm f} a^4/H$, while for a HS that thermalizes with the SM plasma the dark radiation density depends on the decoupling temperature.

the SM plasma, $H(T_d) = \Gamma_E$. Consequently, if the HS thermalizes with the SM plasma, measurements of $N_{\rm eff}$ probe T_d, which is only logarithmically sensitive to Q.

We now separately study the regimes where the HS remains out of equilibrium with the SM plasma and where it equilibrates.

Dark radiation production in the out-of-equilibrium regime: For scenarios where the HS remains out of equilibrium with the SM plasma, the Boltzmann equations simplify because for $T_{\rm HS} \ll T_{\rm SM}$, the cooling of the SM plasma due to millicharge interactions is negligible, and the Hubble rate is dominated by $\rho_{\rm SM}$. Consequently, the SM plasma evolves adiabatically and we only need to integrate the Boltzmann equation for $\rho_{\rm HS}$.

Integrating the remaining Boltzmann equation for $\rho_{\rm HS}$ is non-trivial because the collision term depends on both $T_{\rm SM}$ and $T_{\rm HS}$. In particular, for $T_{\rm HS} \ll T_{\rm SM}$, $\mathcal{C} \approx \mathcal{C}_{\rm f}$, and while $\mathcal{C}_{\rm f} = \mathcal{C}_{\rm f}(T_{\rm SM})$ for s-channel processes, for Coulomb scattering processes $\mathcal{C}_{\rm f} = \mathcal{C}_{\rm f}(T_{\rm HS}, T_{\rm SM})$. The energy transfer from Coulomb scattering process dominates over that from s-channel processes for the regions of parameters space that saturate $\Delta N_{\rm eff} = 0.3$ (see section 5.2.2). However, if $\Delta N_{\rm eff}$ is constrained to smaller values by future experiments, the HS will be constrained to regions of parameter space with lower temperatures, and consequently, the contribution from Coulomb scattering processes will become less important compared to the contribution from s-channel processes.

In order to obtain a simple expression for a conservative lower bound on the asymptotic dark radiation density, we neglect the Coulomb scattering processes. This allows us to take $\mathcal{C}_{\rm f} = \mathcal{C}_{\rm f}(T_{\rm SM})$. Then with the additional simplifying assumption that $w_{\rm SM} = w_{\rm HS} = 1/3$, we can integrate the Boltzmann equation for $\rho_{\rm HS}$

to obtain

$$\left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_F - \left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_I \approx \int_{T_{{\rm SM},F}}^{T_{{\rm SM},I}} \frac{dT_{\rm SM}}{T_{\rm SM}} \frac{\mathcal{C}_{\rm f}}{H\rho_{\rm SM}} = \frac{\sqrt{3}M_{\rm Pl}}{(g_*\pi^2/30)^{3/2}} \int_{T_{{\rm SM},F}}^{T_{{\rm SM},I}} \frac{dT_{\rm SM}}{T_{\rm SM}^7}\mathcal{C}_{\rm f}. \tag{5.10}$$

Because the annihilations of SM fermions into MCPs typically dominates the s-channel energy transfer processes, we focus on its contribution to the production of dark photons. In appendix A, we compute the collision term describing the forward energy transfer for these annihilation processes. The corresponding collision term for the forward energy transfer in SM fermion annihilations into MCPs is given by[4]

$$\mathcal{C}_{\rm f}^{\rm an} = \sum_f \frac{1}{32\pi^4} \int_{4\max(m_f,m)^2}^{\infty} ds(s-4m_f^2)s\sigma_{ff\to\psi\bar{\psi}}(s)T_{\rm SM}G_{\zeta_f}(\sqrt{s}/T_{\rm SM}), \tag{5.11}$$

where $\sigma_{ff\to\psi\bar{\psi}}$ is the spin-summed center-of-mass (CM) frame cross-section (see eq. A.114) and the summation runs over all SM fermions. The dimensionless function $G_{\zeta_f}(z)$, given by eq. (A.18), is determined by the quantum statistical distribution $f(p) = [e^{-E/T} + \zeta]^{-1}$, where $\zeta = 1$ for fermions and $\zeta = -1$ for bosons. In the limit when SM fermions can be approximated to have a Maxwell-Boltzmann distribution ($\zeta_f \to 0$), $G_{\zeta_f}(z)$ asymptotes to the second-order modified Bessel function of the second kind, $K_2(z)$, and eq. (5.11) then matches with the well-known result of Ref. [175].

The integral on the RHS of eq. (5.10) can be simplified for the collision term of eq. (5.11) by first rewriting the integral as

$$\int_{T_{{\rm SM},F}}^{T_{{\rm SM},I}} \frac{dT_{\rm SM}}{T_{\rm SM}^7}\mathcal{C}_{\rm f}^{\rm an} = \sum_f \frac{1}{32\pi^4} \int_{4\max(m,m_f)^2}^{\infty} ds \frac{(s-4m_f^2)}{s\sqrt{s}}\sigma_{ff\to\psi\psi} \int_{x_F}^{x_I} \frac{dx}{x^6} G_{\zeta_f}\left(\frac{1}{x}\right), \tag{5.12}$$

where $x = T_{\rm SM}/\sqrt{s}$. One can show, to an excellent approximation, that the integration limits for x can be replaced by 0 and infinity for $T_{{\rm SM},F} \ll \max(m, m_e) \ll T_{{\rm SM},I}$.[5] With this approximation the integral over x yields a factor of $15\pi\kappa_{\zeta_f}/2$, where $\kappa_1 = 0.80$, $\kappa_0 = 1$ and $\kappa_{-1} = 1.5$. Taking into account quantum statistics in the phase-space distribution of fermions therefore leads to a correction of about 20% to the final dark photon density. This result is indicative of the size of quantum-statistical effects in all energy transfer processes we considered (including Coulomb scattering).

Numerically, we find that the integral over the collision term obtains its asymptotic value at $T_{\rm SM} \sim \max(m, m_e)/4$. The dependence on m_e is a consequence of the fact that for $T_{\rm SM} \ll m_e$ the abundance of all electromagnetically-charged SM fermions is Boltzmann-suppressed.

Using the simplified collision integral, we find the fraction of energy transferred from the SM plasma into the HS is given by

$$\left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_{\rm leak} \equiv \left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_{T_{\rm SM}=\Lambda} - \left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_I \approx \frac{15\sqrt{3}}{64\pi^3[g_*(4\Lambda)\pi^2/30]^{3/2}} \frac{M_{\rm Pl}}{\Lambda} \times L, \tag{5.13}$$

[4]While deriving eq. (5.11) we make two key approximations. First, we neglect the Pauli-blocking effect from MCPs; second, we assume $T_{\rm SM} \gg m_f$. The first approximation is valid in the parameter space where MCPs are produced out-of-equilibrium with $T_{\rm HS} \ll T_{\rm SM}$. The second approximation has negligible impact on the production of dark radiation for $m \gg m_f$ because MCP production is Boltzmann-suppressed by the time $T_{\rm SM} \sim m_f$, while for $m < m_f$ the energy injection is dominated by lighter fermions that are relativistic during $T_{\rm SM} \sim m$.

[5]This approximation is possible for two reasons. First, the terms outside of the x integral peak at energy scale $\sqrt{s} \sim \max(m, m_e)$. Second, the integrand of the x integral goes to 0 as $x \to \infty$ and as $x \to 0$. Thus, as long as $T_{{\rm SM},F} \ll \max(m, m_e) \ll T_{{\rm SM},I}$, the total integral is insensitive to the initial and final temperatures.

where

$$\Lambda \equiv \frac{\max(m, m_e)}{4}, \qquad L = \Lambda\kappa_1 \sum_f \int_{4\max(m,m_f)^2}^{\infty} ds \frac{(s-4m_f^2)}{s\sqrt{s}} \sigma_{f\bar{f}\to\psi\psi}(s). \tag{5.14}$$

The energy injection decoupling temperature, Λ, determines the SM temperature below which energy injection ceases to be important, and the leak factor, L, parameterizes the leakage of energy from the SM plasma due to BSM interactions. While the above approximation assumes a constant g_* until $T_{\rm SM} < \Lambda$, numerically we find that $(\rho_{\rm HS}/\rho_{\rm SM})_{T_{\rm SM}=\Lambda}$ is primarily sensitive to g_* at $T_{\rm SM} = 4\Lambda$. Finally, this result is only valid as long as the HS does not thermalize with the SM, or equivalently, if $(\rho_{\rm HS}/\rho_{\rm SM})_{T_{\rm SM}=\Lambda} < (g_{\rm HS}/g_*)_{T_{\rm SM}=\Lambda}$.

The leak factor L has a simple analytical form when the MCP mass m is not close to any of the SM fermion masses. Using the cross-section for fermion annihilation to MCP pairs given in eq. (A.114), and neglecting Z-mediated contributions, we obtain

$$L \approx \kappa_1 \sum_f^{m_f<4\Lambda} \frac{3\pi^2\alpha^2 Q^2 Q_f^2 N_c(f)}{8}, \tag{5.15}$$

where α is the fine structure constant, Q_f and $N_c(f)$ are the charge and color factor of the SM fermions, respectively, and the sum over f runs over all SM fermions that are relativistic at $T_{\rm SM} \sim 4\Lambda$.

To evaluate the final dark photon energy density we adiabatically evolve $\rho_{\rm HS}$ from the end of energy injection until recombination, $(\rho_{A'}a^4)_{\rm CMB} = (\rho_{\rm HS}/\rho_{\rm SM})_{\rm leak}\,(\rho_{\rm SM}a^4)_{T_{\rm SM}=\Lambda}$. The dark photons act as free-streaming dark radiation and cause a shift in $N_{\rm eff}$ given in eq. (5.4). Requiring that the resulting $\Delta N_{\rm eff}$ remains below the $(\Delta N_{\rm eff})_{\rm max}$ upper bound set by CMB measurements limits the value of the charge to

$$Q^2 \quad < \quad \frac{10^{-14}}{\kappa_1}\left(\frac{1/137}{\alpha}\right)^2\left(\frac{g_*(\Lambda)}{10}\right)^{1/3}\left(\frac{g_*(4\Lambda)}{10}\right)^{3/2}\left(\frac{4}{\sum_f Q_f^2 N_c(f)}\right)\left(\frac{(\Delta N_{\rm eff})_{\rm max}}{0.3}\right)\left(\frac{\Lambda}{\rm GeV}\right). \tag{5.16}$$

In deriving eq. (5.16), we set $g_{*s}(T_{\rm CMB}) = 3.94$ and approximated $g_{*s}(\Lambda) = g_*(\Lambda)$, where g_{*s} counts the effective entropic degrees of freedom in the SM and $T_{\rm CMB} = 0.25$ eV is the temperature of photons near recombination. The constraint on Q for $m > m_e$ is roughly proportional to $\sqrt{m}$, with the proportionality constant determined by $\sqrt{(\Delta N_{\rm eff})_{\rm max}}$. Note that taking into account the Fermi-Dirac statistics of SM fermions weakens the constraint on Q by $\sim 10\%$.

In the left panels of figure 5.3, the dark blue dashed lines show the values of the parameters that saturate various $N_{\rm eff}$ thresholds. These points are evaluated by numerically solving the Boltzmann equations after including all s-channel energy transfer processes but not the t-channel Coulomb scattering processes. We compare this s-channel result with the full result, which includes Coulomb scattering processes, given by the black solid lines. Note that the agreement between the full and the s-channel-only results improves as $\Delta N_{\rm eff}$ is restricted to smaller values. The light blue dot-dashed lines show the analytical result given by eq. (5.16). Our analytical result does not include the contribution from Z-boson decays and hence underestimates the dark radiation density in the range 1 GeV$\lesssim m \lesssim$ 40 GeV in the bottom left panel of figure 5.3.[6]

Dark radiation production in the equilibrium regime: The analysis in the previous section is only valid when the HS remains out of equilibrium with the SM plasma. However, starting in the out-of-equilibrium

[6]One can straightforwardly incorporate Z-boson decays into the approximate analytical treatment by substituting the corresponding collision term, given in eq. (A.115), into eq. (5.10). We omit this calculation for brevity.

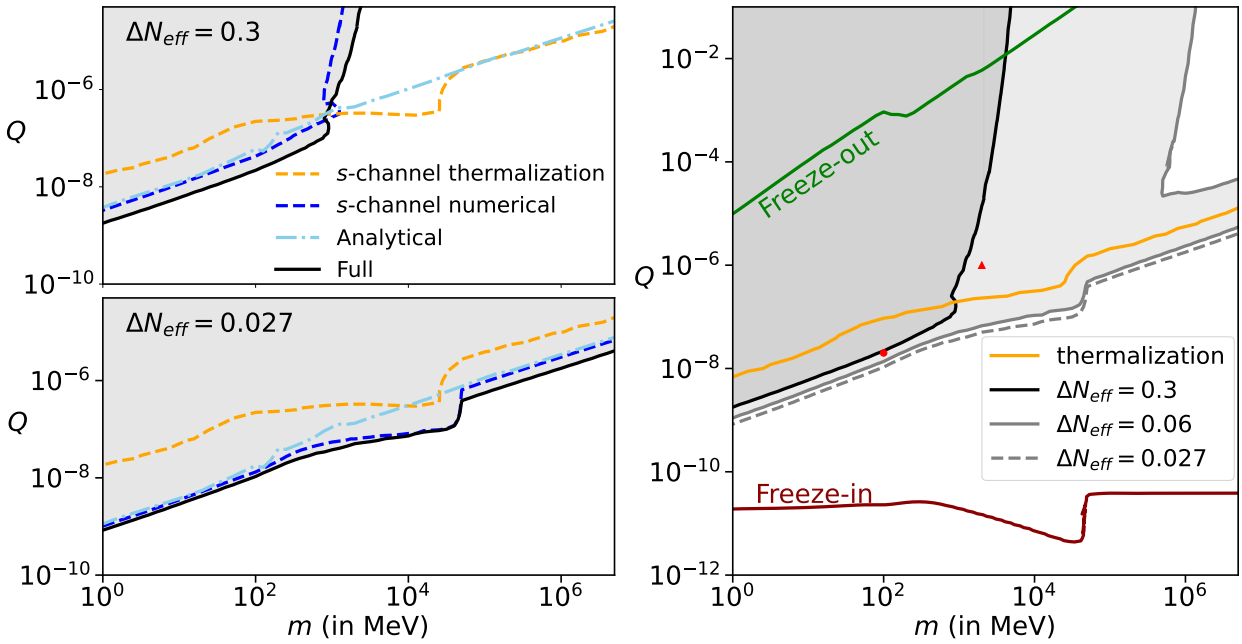


Figure 5.3: **Left**: Solid black lines mark the parameter space for the MCP model that yields $\Delta N_{\rm eff} = 0.3$ (top) and $\Delta N_{\rm eff} = 0.027$ (bottom). The blue dashed lines show the numerical solution after neglecting contributions from Coulomb scattering processes. The light blue dot-dashed line is our analytical approximation to the blue dashed line, as given in eq. (5.16). The orange dashed line marks the values of Q at which the HS thermalizes with the SM plasma after neglecting Coulomb scattering processes. The exact $N_{\rm eff}$ constraint is well described by the constraint calculated with only s-channel processes as $N_{\rm eff}$ measurements are improved. **Right**: Solid black, solid gray and dashed gray lines mark the MCP parameter space that yields $\Delta N_{\rm eff} = 0.3$, $\Delta N_{\rm eff} = 0.06$, and $\Delta N_{\rm eff} = 0.027$, respectively, and are the same as those in figure 5.1. The orange line marks the parameter space above which the hidden sector thermalizes with the SM plasma. The green and maroon lines mark the parameter space where the MCP relic density matches the observed dark matter density via freeze-out [165] and freeze-in [37], respectively. For values of Q relevant for $N_{\rm eff}$ constraints, almost all MCPs produced in the early universe must annihilate into dark photons to avoid overclosure of the universe.

regime, as one follows a contour of constant $\Delta N_{\rm eff}$ by increasing m, the value of Q increases. At some point the coupling can become large enough that the HS thermalizes with the SM. Once the sectors are thermalized, the dark radiation density is no longer sensitive to the maximum of the forward energy transfer $\mathcal{C}_{\rm f}a^4/H$. Instead, the final dark radiation density is determined by the temperature T_d at which the HS decouples from the SM. This decoupling temperature is determined by the Boltzmann suppression of the collision term, and is principally determined by the mass of the MCP, while remaining only weakly dependent on the coupling Q. This is illustrated in the right panel of figure 5.2, which shows the evolution of the densities for a parameter point where the HS and SM thermalize. Here, decoupling occurs with the Boltzmann suppression of the collision term at $T \sim m/4$.

The orange lines in figure 5.3 mark the values of Q above which the HS thermalizes with the SM plasma for a given mass m. In the left panel, the orange lines are plotted after considering only s-channel energy transfer processes while in the right panel they are plotted after including all processes. In the left panel, the s-channel result that saturates $\Delta N_{\rm eff} = 0.3$ becomes largely insensitive to the coupling Q once the curve crosses above the s-channel thermalization contour; similar weakening occurs in the right panel for the full result.

We can determine the thermalization threshold, the mass scale beyond which we can no longer use the out-of-equilibrium result in eq. (5.16), as follows. On the one hand, a given relic dark radiation density, or value of $\Delta N_{\rm eff}$, can be translated to a decoupling temperature, T_d, by assuming that entropy is separately conserved in the HS and SM sectors after T_d. This leads to the implicit relation

$$\frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\left(\frac{g_{*s}(T_{\rm CMB})}{g_{*s}(T_d)}\right)^{4/3}\frac{g_{\rm HS}}{2} = \Delta N_{\rm eff}, \tag{5.17}$$

which can be solved to determine $T_d(\Delta N_{\rm eff}, g_{\rm HS})$.[7] This expression for $T_d(\Delta N_{\rm eff}, g_{\rm HS})$ is independent of the masses and couplings in the hidden sector, depending only on the effective number of degrees of freedom.

On the other hand, given a model, in this case the MCP model, we can compute the decoupling temperature directly from the collision term by setting the energy transfer rate $\Gamma_E(T) \equiv \mathcal{C}_{\rm f}(T)/\rho_{\rm HS,eq}(T)$ equal to the Hubble rate at T_d. This condition determines the decoupling temperature in terms of the model parameters Q and m, $T_d(Q, m)$. Consequently, when the HS is thermalized with the SM plasma, the contour in MCP parameter space that yields a given value of $\Delta N_{\rm eff}$ is found by setting

$$T_d(Q, m) = T_d(\Delta N_{\rm eff}, g_{\rm HS}). \tag{5.18}$$

The energy transfer rate Γ_E increases compared to the Hubble rate until $T_{\rm SM} \sim m/2$, after which it starts decreasing. Consequently, the decoupling temperature has to be smaller than $m/2$. Thus the lowest value of m for which the HS can be in equilibrium with the SM plasma for a given $(\Delta N_{\rm eff})_{\rm max}$ is determined by $T_d(\Delta N_{\rm eff}, g_{\rm HS})$. Empirically we find that the Boltzmann suppression of Γ_E becomes prohibitive for $T_{\rm SM} \lesssim m/4$, and thus the precise location of the decoupling temperature becomes logarithmically sensitive to the value of Q for $T_d < m/4$. Therefore, the value of m above which the the dark radiation constraint on Q become exponentially weak occurs at

$$m_{\rm th} \equiv 4T_d[(\Delta N_{\rm eff})_{\rm max}, g_{\rm HS}]. \tag{5.19}$$

[7] Note that there is a many-to-one map from T_d to $\Delta N_{\rm eff}$ because $g_{*s}(T_d)$ is constant away from mass thresholds. For $\Delta N_{\rm eff}$ values that exactly coincide with regions where $g_{*s}(T_d)$ is constant, we calculate $T_d(\Delta N_{\rm eff}, g_{\rm HS})$ by finding the minimum T_d that satisfies eq. (5.17).

Notice that the evaluation of $m_{\rm th}$ is independent of the strength of energy transfer processes and only depends on the sensitivity of the $N_{\rm eff}$ measurement and the degrees of freedom in the HS. Consequently, eq. (5.19) does not depend on the detailed calculation of $\mathcal{C}$, and in particular whether we do or do not include contributions from Coulomb scattering.

If future CMB missions continue to see an agreement with the SM value of $N_{\rm eff}$, the thermalization threshold $m_{\rm th}$ will be pushed to larger values. The gray solid and dashed lines in the right panel of figure 5.3 show the values of the parameters that lead to $\Delta N_{\rm eff} = 0.06$ and $\Delta N_{\rm eff} = 0.027$, respectively. The excluded regions extend to much larger values of m because more of the parameter space is required to have the HS remain out of equilibrium with the SM plasma. For $\Delta N_{\rm eff} < 0.027$, there is no allowed thermalization threshold.

The exponential behavior of the constant $\Delta N_{\rm eff}$ contours for $m > m_{\rm th}$ eventually stops at sufficiently large values of Q, when direct energy transfer from SM into dark photons through off-shell MCPs become larger than the Boltzmann-suppressed energy transfer into on-shell dark fermions. These off-shell processes depend on additional model parameters, in particular the dark gauge coupling constant, and are beyond the scope of the study.

Requirement of chemical equilibrium: Our analysis assumes that the HS energy density can be treated as a whole, including both the MCP and the dark photon, instead of tracking their energy densities separately. This assumption is strictly valid when the HS is in internal chemical equilibrium throughout the period of energy transfer, which is not necessarily true everywhere throughout our parameter space. However, this assumption of internal chemical equilibrium is only critical to our final result for the dark radiation abundance in the regions near and above the thermalization threshold(s) for the MCP, where it does hold (as we discuss below). Below the thermalization threshold, where the MCPs remain out-of-equilibrium with the SM, the assumption of internal chemical equilibrium remains an excellent approximation as long as (i) the HS energy density is dominated by radiation throughout the period of energy transfer, and (ii) we can treat all the entropy carried by the MCPs as deposited into dark radiation, rather than the SM, after it becomes non-relativistic. Given these two conditions, the detailed evolution of the MCP number density itself is unimportant to the final dark radiation abundance. In fact condition (ii) follows from condition (i) when the MCPs are out of equilibrium with the SM, as requiring the HS to be dominated by radiation means that almost all the produced MCPs must rapidly annihilate, and if the MCP is out of equilibrium with the SM, then necessarily $n_\psi^2 \langle \sigma v \rangle_{\psi\psi\to\gamma\gamma} < H$. Thus the MCP must dominantly annihilate into dark photons.

The condition that almost all the produced MCPs efficiently annihilate into dark radiation *is* met in the regions of our parameter space relevant for current and forecast out-of-equilibrium constraints, given the mild constraint on the dark gauge coupling e' that follows from requiring that the relic MCP abundance does not overclose the universe, as we now argue. The green line in figure 5.3 indicates where the freezeout of SM annihilations into pairs of MCPs would produce the observed DM relic density in the absence of dark photons, i.e., if the MCP's only annihilation channel is to SM fermions [165]. Meanwhile the maroon line indicates where the freezein production of MCPs from the SM produces the observed DM relic density, again turning off the MCP annihilations into dark photons [37]. As current and future $N_{\rm eff}$ constraints lie between these two lines (except for a small region above the thermalization threshold in the case of current constraints), in the region of parameter space relevant for evaluating these constraints, SM processes alone overproduce MCPs by multiple orders of magnitude. Thus the dark gauge coupling constant must be large enough to enable the vast majority of MCPs to annihilate efficiently into dark photons. If this condition is not met, the model is

excluded simply by overclosure; our $N_{\rm eff}$ analysis applies to the surviving model parameter space where e' is large enough to avoid overclosure, and otherwise does not depend on the detailed value of e'. Thus avoiding overclosure alone establishes the requirement for out-of-equilibrium case discussed above, which suffices as long as the final result for dark radiation density does not depend on the evolution of $g_{\rm HS}$ with temperature.

To accurately determine the production of dark radiation when the hidden sector is close to the thermalization threshold, we need to track the evolution of $g_{\rm HS}$ with temperature, and thus the evolution of the MCP number abundance with temperature, up until $T_{\rm HS} \lesssim m/3$. Once $T_{\rm HS} < m/3$, the hidden sector equation of state is given by $w_{\rm HS} = 1/3$ to an excellent approximation. In particular this is necessary to accurately determine the location of the excluded strip running up to high masses in fig. 5.1. Using the results of Ref. [176], we have checked that internal chemical equilibrium for $T_{\rm HS} > m/3$ is indeed necessary if the relic MCPs are not to overclose the universe.

5.2.4 Dark radiation production in extended MCP models and implications for EDGES

We have so far considered a minimal MCP model where only one fermion is charged under the dark $U(1)$ gauge symmetry. More generally, the hidden sector may contain multiple particles with dark charges. A full evaluation of the resulting dark radiation density in these models depends on the detailed spectrum of the hidden sector, including properties such as the number of particles and the values of their various dark charges. However, in this section, we show that a conservative lower bound on the dark photon density at recombination can be estimated that is insensitive to such details. This conservative lower bound can then be used to place general constraints on the allowed parameter space of these models.

The conservative lower bound on the dark photon density is obtained by considering only s-channel energy transfer processes (annihilations or decays) and considering only one MCP and one dark photon in the HS bath. On the one hand, if the HS equilibrates with the SM plasma, the final dark radiation density is largely insensitive to the specifics of the energy transfer processes but remains proportional to the degrees of freedom in the HS bath, $g_{\rm HS}$. Thus, minimizing the particles in the HS also minimizes the final value obtained for the dark radiation density. On the other hand, if the HS remains out-of-equilibrium with the SM plasma, the dark radiation density is determined by the energy transfer from the SM. While the energy transferred by t-channel scattering processes decreases as we increase $g_{\rm HS}$ (holding the total $\rho_{\rm HS}$ fixed), the energy transferred by s-channel processes is insensitive to $g_{\rm HS}$ as long as it is dominated by a single mediating species. Consequently, the dark radiation density cannot be smaller than that following from s-channel processes alone for an out-of-equilibrium hidden sector.

This conservative lower bound on the dark radiation density can be translated directly into a lower bound on $\Delta N_{\rm eff}$. This lower bound has an immediate application to the MCP model proposed by Ref. [155] to explain the anomalously small hydrogen spin temperature as measured by the EDGES experiment [158]. Their model consists of two fermions that are charged under a dark $U(1)$ gauge symmetry. One fermion is the main component of dark matter, χ_1, and the other fermion, χ_2, constitutes a small fraction of dark matter. The particle χ_2 is responsible for cooling hydrogen atoms via millicharge interactions and then transfers that heat to the dark matter bath via dark long range interactions. The dark photon mediating the long-range interaction is a light relativistic relic that contributes to $N_{\rm eff}$.

In figure 5.4 we show the parameter space in the model of Ref. [155] that is consistent with various current and projected CMB measurements of $N_{\rm eff}$. The orange lines in figure 5.4 show the values of the millicharge, Q_2, and mass, m_2, of the χ_2 particle required to resolve the EDGES anomaly, as calculated in Ref. [155].

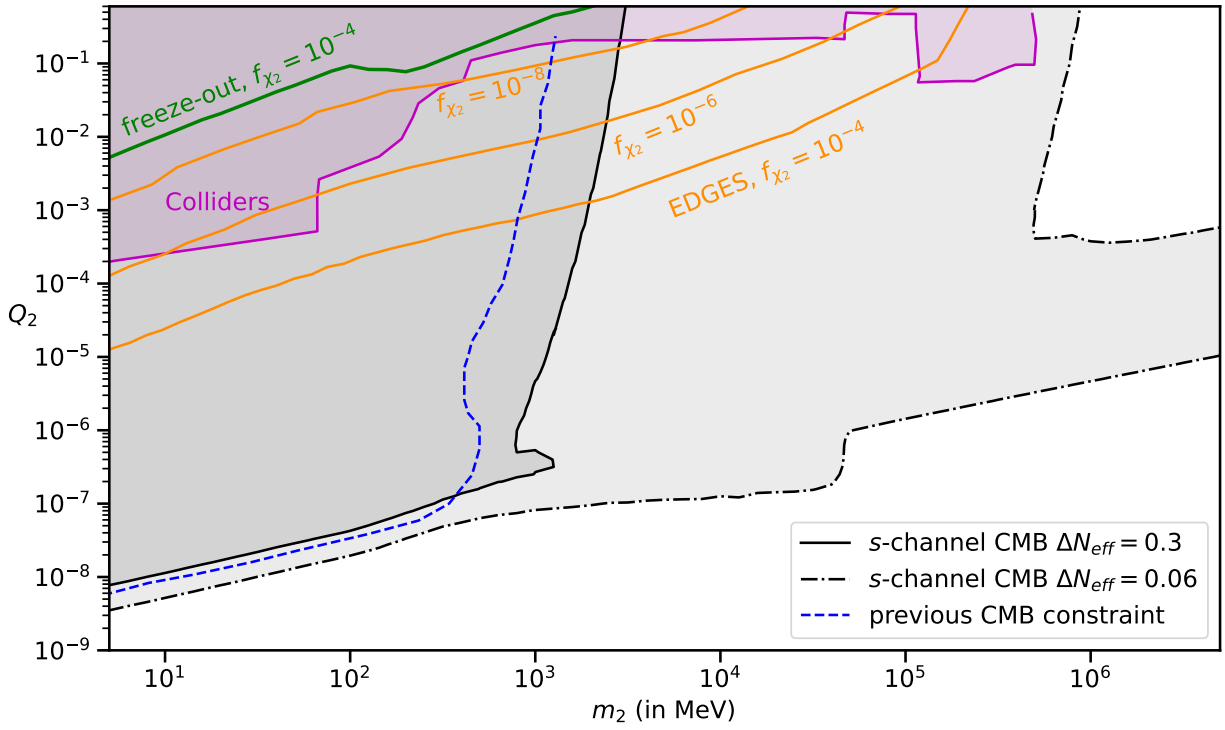


Figure 5.4: Constraints on the mass and millicharge of the millicharged particle in the context of extended models. The orange lines mark the values of charge and mass of the MCP χ_2 for which the model given in Ref. [155] resolves the EDGES anomaly. The orange lines have been plotted after fixing the dark matter mass to 10 MeV and setting the fraction of χ_2 density relative to dark matter to $f_{\chi_2} = 10^{-4}$, 10^{-6}, and 10^{-8} as indicated. The green line marks the values of Q_2 for which χ_2 would obtain $f_{\chi_2} = 10^{-4}$ in the absence of dark annihilation channels. The black solid and dot-dashed contours mark the parameter space that yields $\Delta N_{\rm eff} = 0.3$ and $\Delta N_{\rm eff} = 0.06$, respectively, after neglecting energy transfer from Coulomb scattering processes and assuming one millicharged particle in the hidden sector bath. The blue dashed contour is the CMB constraint derived in Ref. [153] for $\Delta N_{\rm eff} = 0.8$. The pink shaded regions marks the parameter space ruled out by SLAC [167], MiniBooNE [168], LEP [169] and LHC [170].

The lines are plotted for fixed values of dark charges and χ_1 masses chosen such that the cooling of hydrogen atoms is maximized while remaining consistent with cosmological bounds from the CMB and BBN. The black solid and dashed lines show the values of $N_{\rm eff}$ computed using the conservative method described above that saturate the Planck and projected CMB-S4 2σ bounds, respectively. The contours below the thermalization threshold are well described by eq. (5.16). Current measurements of $N_{\rm eff}$ already limit $m_2 > 2$ GeV, while future CMB experiments can completely rule out the MCP model proposed by Ref. [155]. Since the dark radiation constraints we show here are largely insensitive to the details of the specific extended MCP model, they offer a powerful way to constrain model-building in this direction to explain the EDGES anomaly.

Naively one might imagine that the dark radiation constraints on Q_2 can be circumvented if χ_2 predominantly annihilates into SM particles rather than dark photons. However, for χ_2 to resolve the EDGES anomaly, it must have significant couplings with a lighter dark particle in order to avoid being overproduced in the early universe. For instance, the solid green line in figure 5.4 marks the values of Q_2 that produce χ_2 constituting a fraction $f_{\chi_2} = 10^{-4}$ of dark matter density today if χ_2 only has annihilation channels to SM fermions. Assuming SM-only freezeout, the relic abundance of χ_2 increases below the green line by a factor of $1/Q_2^2$. Consequently, the values of Q_2 required to resolve the EDGES anomaly result in a χ_2 relic abundance multiple orders of magnitude larger than what is required unless χ_2 has an additional annihilation channel. The minimal possibility is that χ_2 dominantly annihilates into the dark mediator that sources the requisite long-ranged interaction between χ_1 and χ_2.[8] The produced dark mediator is then constrained by the $N_{\rm eff}$ measurements, which consequently restricts Q_2 as shown in figure 5.4.

Finally, applying the CMB $N_{\rm eff}$ constraint to any MCP model assumes that the dark photon is free-streaming during recombination. If the dark photon and the MCPs have sufficiently large self-interactions during recombination, they can instead form a fluid, and the dark photon would accordingly contribute to $N_{\rm fluid}$ instead of producing a neutrino-like signal. The ability to form a fluid depends on the MCP relic abundance as well as the interaction between the MCP and the dark photon, both of which are determined by the dark coupling constant, e'. A more detailed analysis would be required to find the relevant values of e' that can produce a self-interacting radiation bath without violating either unitarity or cosmological bounds. For such values of e', one would instead have to look to $N_{\rm fluid}$ measurements, which are factors of $2-3$ less sensitive than measurements of $N_{\rm eff}$ [177]. A future CMB-S4 constraint of $\Delta N_{\rm fluid} \lesssim 0.16$ would yield a thermalization threshold of $m_{2,\rm th} \sim$ GeV.

5.3 $B - L$ right-handed neutrinos

In this section we derive dark radiation constraints on the scenario where the global SM symmetry of baryon number minus lepton number $(B-L)$ is promoted to a gauge symmetry. This promotion requires the addition of three right-handed neutrinos to cancel gauge anomalies. When these three additional neutrinos are light, they contribute to the energy budget of the Universe as dark radiation. Consequently, their energy density and the parameter space of the model are constrained by measurements of $N_{\rm eff}$.

The gauged $B-L$ model is also constrained by fifth-force searches [178], stellar evolution [179], supernova 1987A [180], and collider experiments [181–188]. Constraints on this model from $N_{\rm eff}$ measurements have been studied previously in Ref. [189] and updated in Ref. [156]. Here we improve over previous studies by taking into account the out-of-equilibrium production of right-handed neutrinos.

[8] An alternative non-minimal method to dilute the χ_2 abundance is to have an unstable field preferentially reheat the SM plasma at some temperature $T_{\rm rh} < m_2$.

This section is organized as follows. We begin in section 5.3.1 by introducing the model and our conventions. In section 5.3.2 we describe the relevant Boltzmann equations, detailing the approximations within which we work. We then solve the Boltzmann equations to find the model parameter space that saturates the $N_{\rm eff}$ bounds from current and upcoming CMB experiments. Next, in section 5.3.3 we analyze the evolution of the energy density in right handed neutrinos, ρ_{ν_R}, and show that its final value is qualitatively changed depending on the lifetime of the Z' boson. Finally in section 5.3.4, we provide an analytical explanation of the features of the dark radiation constraint on the model parameter space.

5.3.1 The model

The Lagrangian describing the interactions of the SM with the $B-L$ gauge boson Z' and the right-handed neutrinos is given by

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F'_{\mu\nu}F^{\mu\nu\prime} + \frac{1}{2}M^2_{Z'}Z'_\mu Z'^\mu + g' Z'_\mu \sum_i \left[\frac{1}{3}(\bar{u}_i\gamma^\mu u_i + \bar{d}_i\gamma^\mu d_i) - \bar{e}_i\gamma^\mu e_i - \bar{\nu}_{L,i}\gamma^\mu \nu_{L,i}\right] \\ &- g' Z'_\mu \sum_i \bar{\nu}_{R,i}\gamma^\mu \nu_{R,i}. \end{aligned} \tag{5.20}$$

Here, the index i runs over the three generations of SM fermions, while u, d, e, ν_L and ν_R denote the up quark, down quark, electron, left-handed neutrino and right-handed neutrino counterparts of each generation. Above we have explicitly separated the interaction of the Z' gauge boson with the ν_R from its interactions with the known SM fermions. We consider the minimal version of the model where the three right-handed neutrinos form Dirac particles with the left-handed neutrinos after electroweak symmetry breaking. Because the neutrinos are always relativistic during and prior to recombination, we ignore neutrino masses in the subsequent analysis and treat ν_L and ν_R as distinct Weyl fermions. The Z' gauge boson has mass $M_{Z'}$, which can come from a Stueckelberg or a Higgs mechanism. To remain as model-independent as possible, we ignore potential contributions to the dark radiation density arising from possible Higgs fields associated with $B-L$ breaking and focus on the irreducible contribution from the Z' itself.[9]

Right-handed neutrinos are produced in this model as a result of the $B-L$ interactions with the Standard Model in the early Universe. Because they are approximately massless and sterile at late times, after the Z' freezes out, these right-handed neutrinos are dark radiation and contribute to $N_{\rm eff}$. Furthermore, for values of g' allowed by current $N_{\rm eff}$ constraints, the $B-L$ interactions with ν_L are significantly weaker than the weak interactions with ν_L prior to neutrino decoupling. We focus on the region of parameter space where dark radiation is produced prior to BBN, and thus before the weak interactions freeze out and the ν_L leave equilibrium. In this region of parameter space, the production of ν_R provides the major contribution to $\Delta N_{\rm eff}$,

$$\Delta N_{\rm eff} = \frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\frac{\rho_{\nu_R}}{\rho_\gamma}. \tag{5.21}$$

5.3.2 Boltzmann equations and constraints for the $B-L$ model

Right-handed neutrinos in this model are dominantly produced by Z'-mediated SM fermion annihilation. In part of the relevant parameter space, the Z' bosons are long-lived, i.e., they do not decay within a Hubble

[9] This is an excellent approximation when a $B-L$ Higgs is more massive than the Z', and conservative in the case when it is not; this treatment is also applicable to the technically natural scenario where the Z''s only interactions are the Stueckelberg mass and the coupling to the SM $B-L$ current as given in eq. 5.20.

time. Consequently, the energy transferred into ν_R can depend on the cosmic evolution of the on-shell Z' density. The relevant Boltzmann equations for this system need to track the evolution of both Z' and ν_R, and read

$$\frac{d\rho_{\rm SM}}{dt} + 3H(1+w_{\rm SM})\rho_{\rm SM} = -\mathcal{C}_{ff\to Z'} - \mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}, \tag{5.22}$$

$$\frac{d\rho_{Z'}}{dt} + 3H(1+w_{Z'})\rho_{Z'} = \mathcal{C}_{ff\to Z'} - \mathcal{C}_{Z'\to\nu_R\nu_R}, \tag{5.23}$$

$$\frac{d\rho_{\nu_R}}{dt} + 4H\rho_{\nu_R} = \mathcal{C}_{Z'\to\nu_R\nu_R} + \mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}. \tag{5.24}$$

Here the Hubble rate is $H = \sqrt{\rho_{\nu_R}+\rho_{\rm SM}+\rho_{Z'}}/[\sqrt{3}M_{\rm Pl}]$, and the various $\mathcal{C}_i$ denote energy transfer collision terms from three processes: $\mathcal{C}_{ff\to Z'}$, for the inverse decay of SM fermions into Z's; $\mathcal{C}_{Z'\to\nu_R\nu_R}$, describing the decay of Z's into right handed neutrinos; and $\mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}$, which describes contact interactions between SM fermions and ν_R, mediated by off-shell Z's (see also Refs. [190, 191]). We include the quantum phase space distributions for initial state particles but ignore final state quantum effects in the evaluation of the collision terms.[10] The detailed expressions for the collision terms are given in appendix A.6.

While evaluating the backward collision terms describing $\nu_R\nu_R \to ff$ and $\nu_R\nu_R \to Z'$, we assume that right-handed neutrinos are internally thermalized with a temperature $T_{\nu_R} \equiv [\rho_{\nu_R}/(g_{\nu_R}\pi^2/30)]^{1/4}$, where $g_{\nu_R} = 2\times 3\times 7/8$. On the one hand, this assumption is unimportant if the two sectors do not thermalize as the backward collision term is negligibly small in comparison to the forward collision term for $T_{\nu_R} \ll T_{\rm SM}$. On the other hand, if they do thermalize then the assumption is automatically satisfied. The transition regime, where the backward collision term can be important, is relevant for the current $N_{\rm eff}$ constraints in the mass range 1 GeV $< M_{Z'} <$ 2 GeV; in this range a differential treatment of the phase space distribution of the right-handed neutrinos would be required to improve on our treatment. For $\Delta N_{\rm eff} < 0.06$ we expect to be well into the out-of-equilibrium regime where the backward collision term is unimportant.

When right-handed neutrinos are in equilibrium with the SM plasma, the decoupling temperature is determined by either Z' decays or contact interactions. When the right-handed neutrinos are out-of-equilibrium with the SM plasma, the energy transferred through Z' decays and inverse decays is orders of magnitude larger than that via contact interactions. The forward energy transfer collision terms for $Z' \to ff$ and $Z' \to \nu_R\nu_R$ are given by

$$\mathcal{C}^{\rm f}_{Z'\to ii} = M_{Z'}\Gamma_{Z'\to i}n_{Z'}, \tag{5.25}$$

where $n_{Z'}$ is the number density of the Z' and $\Gamma_{Z'\to i}$ is the Z' decay width into particle species i. While computing $\Gamma_{Z'\to SM}$ we neglect the decays of Z' bosons into hadrons for $M_{Z'} < 2T_{\rm QCD}$, where we set the QCD transition scale at $T_{\rm QCD} = 200$ MeV. For $M_{Z'} > 2T_{\rm QCD}$, we include Z' decays into free quarks in $\Gamma_{Z'\to SM}$.

We approximate $\rho_{Z'} = M_{Z'}n_{Z'}$ and $w_{Z'} = 0$ in the Boltzmann equations, as appropriate for non-relativistic Z' bosons. Most of the energy injection into ν_R occurs when the Z' bosons are non-relativistic, as demonstrated explicitly below, and therefore this approximation has a minimal effect on the final dark radiation density and the ensuing constraints.

[10] Ignoring final-state quantum effects is an excellent approximation as long as the Z' is out of equilibrium with both the ν_R and SM plasma. On the other hand, if the $B-L$ interactions are strong enough to thermalize the Z' and the ν_R with the SM, the precise value of the collision term has only a marginal impact on the final densities of Z' and ν_R and hence final-state quantum effects are not quantitatively important.

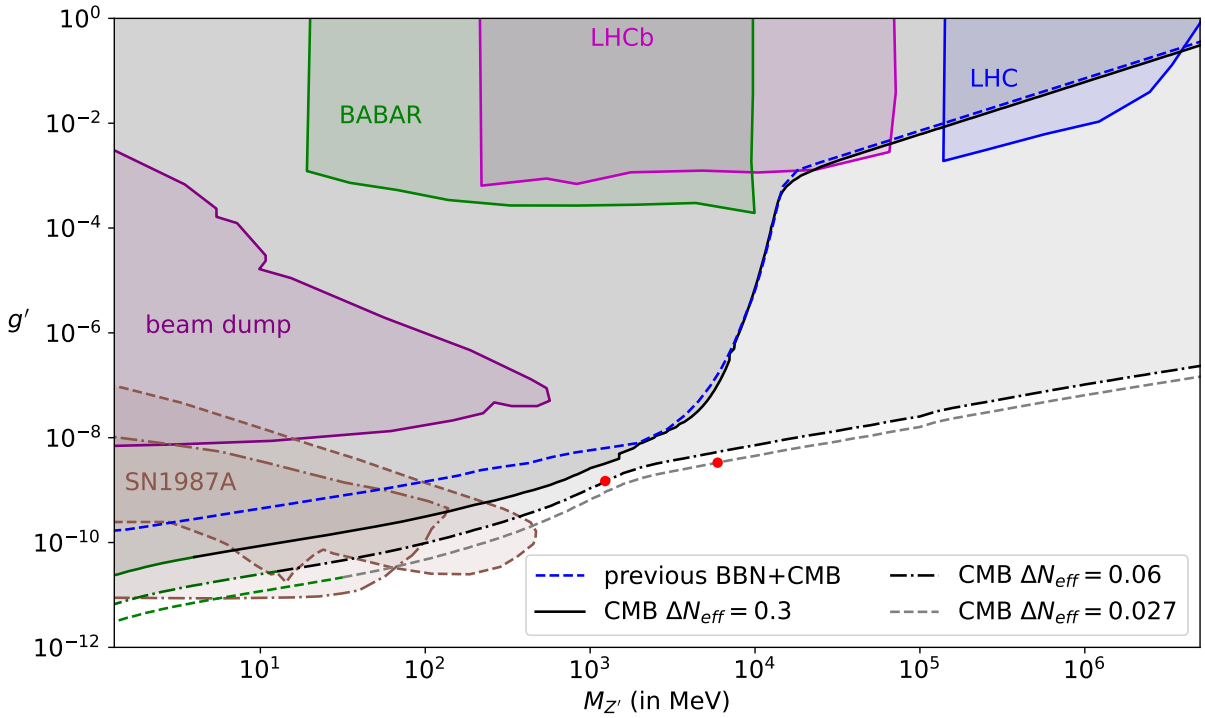


Figure 5.5: Constraints on $B - L$ gauge coupling and gauge boson mass. The black solid, black dot-dashed, and gray dashed contours mark the parameter space that yields $\Delta N_{\rm eff} = 0.3$, $\Delta N_{\rm eff} = 0.06$, and $\Delta N_{\rm eff} = 0.027$ respectively. These bounds correspond to the 2σ upper limit for Planck [3], 2σ upper limit for CMB-S4 [132], and the sensitivity goal for future CMB experiments, and update the BBN+CMB constraints derived in Ref [156], which are shown with the blue dashed contour. The red dots mark the points on the constant $N_{\rm eff}$ curves below and to the left of which $\Gamma_{Z'}$ is smaller than the Hubble rate at $T_{\rm SM} = M_{Z'}/2$. The green color on our $\Delta N_{\rm eff}$ contours marks the region where we expect Z' decays into ν_L to contribute to $\Delta N_{\rm eff}$ and alter our results by an $\mathcal{O}(1)$ factor. Brown lines show constraints from supernova 1987A from Refs. [180] (dashed) and [192] (dot-dashed). We also show constraints from BABAR [181], LHCb [182], LHC [183, 184], and beam dump experiments [185–188].

We are now ready to compute the final dark radiation density in ν_R by solving the Boltzmann equations given in eq. (5.22)-(5.24). We begin the evolution at an initial SM temperature $T_{\rm SM} \gg M_{Z'}$, setting $\rho_{Z'} = \rho_{\nu_R} = 0$, and evolve forward until the end of energy injection. In figure 5.5, we show the contours of g' as a function of $M_{Z'}$ that saturate the current one-tailed 2σ upper limit from Planck [3], $\Delta N_{\rm eff} = 0.3$ (black solid); the projected 2σ upper limit from CMB-S4 [132], $\Delta N_{\rm eff} = 0.06$ (black dot-dashed); and the threshold goal for future CMB experiments $\Delta N_{\rm eff} = 0.027$ (gray dashed).

The curves of constant $\Delta N_{\rm eff}$ in figure 5.5 have a number of key features. As in the MCP model, these curves have a thermalization threshold beyond which they are only logarithmically sensitive to g'. For the $\Delta N_{\rm eff} = 0.3$ curve, the threshold is at $M_{Z'} \sim 1.7$ GeV, while for other contours displayed, there is no threshold. This is because $\Delta N_{\rm eff} = 0.3$ allows three BSM Weyl fermions to decouple from the SM plasma before the QCD phase transition, but the smaller values $\Delta N_{\rm eff} = 0.06$ and $\Delta N_{\rm eff} = 0.027$ cannot accommodate so many new degrees of freedom ever thermalizing with the SM. For $\Delta N_{\rm eff} = 0.3$, the logarithmic sensitivity to g' becomes a power law again above $g' \lesssim M_{Z'}/(16\text{ TeV})$ (see also [193]), as the decoupling temperature goes from being determined by Z' decays and inverse decays to being determined by contact interactions, described by $\mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}$.

The curves corresponding to $\Delta N_{\rm eff} = 0.06$ and lower (as well as the curve for $\Delta N_{\rm eff} = 0.3$ below the thermalization threshold) are controlled by the out-of-equilibrium production of right-handed neutrinos. As we describe below, there are two qualitatively different out-of-equilibrium production mechanisms depending on the ratio $\Gamma_{Z'}/H$ at $T_{\rm SM} \sim M_{Z'}/2$, where $\Gamma_{Z'}$ is the total decay width of Z' bosons. The red dots on the curves indicate where $\Gamma_{Z'}$ is equal to the Hubble rate at $T_{\rm SM} = M_{Z'}/2$. Along the contours below and to the left of the red dots, the Z' bosons become long-lived and we need to track their number density to evaluate dark radiation production. This key result, together with the usual out-of-equilibrium production of ν_R, accounts for the difference between the results in this work and those previously obtained in Ref. [156], shown in figure 5.5 as the blue dashed curve.

Constraints on the $B-L$ gauge boson can also be derived by considering the production of ν_R in colliders or in supernova. In figure 5.5 we also show the regions of parameter space that are excluded by measurements from these other sources. Current CMB constraints are already the leading probe of this hidden sector across much of parameter space, with LHC constraints taking over for masses above 100 GeV. The $N_{\rm eff}$ measurements from future CMB experiments along with existing supernova measurements will provide the strongest constraint on g' for all masses $M_{Z'} \gtrsim 1$ MeV.

For $g' \lesssim 10^{-10}\sqrt{\text{MeV}/M_{Z'}}$, the Z' bosons decay after neutrino decoupling. In this part of parameter space, decays to both ν_L and ν_R contribute to $\Delta N_{\rm eff}$ during recombination, while our analysis only considers the contribution from ν_R. We estimate that the additional production of ν_L provides no more than an $\mathcal{O}(1)$ correction to the $N_{\rm eff}$ constraints calculated in this study. We indicate this region in figure 5.5 by coloring the $N_{\rm eff}$ contours green. Furthermore, for $M_{Z'} < 2m_e$, the dominant energy transfer occurs between ν_L and ν_R, while our Boltzmann equations assume energy injection from a thermal SM plasma with all species at the photon temperature. Thus below the MeV scale, our analysis no longer applies, and hence we restrict our attention here to $M_{Z'} > 2m_e$. Meanwhile, stellar cooling places powerful constraints on this theory for $M_{Z'} < 0.1$ MeV [140, 194]. A full treatment of early universe constraints on the $B-L$ model in the mass range between $0.1\text{ MeV} < M_{Z'} < (10^{-20}/g'^2)$ MeV requires a detailed treatment of neutrino decoupling as well as light element formation during BBN, and is beyond the scope of this work.

5.3.3 Dark radiation density in the out-of-equilibrium regime

In the out-of-equilibrium (OOE) regime, the final energy deposited into ν_R depends on whether or not the total decay width of the Z', $\Gamma_{Z'}$, is less than the Hubble rate at SM temperatures around $T_{\rm SM} \sim M_{Z'}$, where the production rate of Z's is maximized. In the case where $\Gamma_{Z'}/H \gg 1$ at $T_{\rm SM} \sim M_{Z'}$, the large population of on-shell Z' bosons produced at resonance decay almost immediately into ν_R. However when $\Gamma_{Z'}/H \ll 1$ at $T_{\rm SM} \sim M_{Z'}$, the on-shell Z' bosons produced at resonance are cosmologically long-lived and, because they are non-relativistic at production, their energy density redshifts like matter. The right-handed neutrinos are then dominantly produced at some SM temperature $T_{\rm decay} \ll M_{Z'}$ when the population of massive Z' bosons decays, $\Gamma_{Z'} = H(T_{\rm decay})$. Numerically, we find that setting $(\Gamma_{Z'}/H)_{T_{\rm SM}=M_{Z'}/2} = 1$ is a convenient criterion to separate the long- and short-lived regimes.

We illustrate these two regimes with two representative parameter points in figure 5.6. Here in both panels the black line shows the comoving energy density of ν_R, while the red dashed line indicates the energy density of ν_R after setting $T_{\nu_R} = T_{\rm SM}$ (similarly to the red line in figure 5.2). The ν_R do not thermalize with the SM for either the parameter points shown, and correspondingly the black line remains below the red line in both panels. The blue dot-dashed line shows the evolution of $M_{Z'}\Gamma_{Z'\to\nu_R} n_{Z'} a^4/H$, which indicates the amount of comoving energy injected into ν_R in a Hubble time from the decay of on-shell Z' bosons. The energy injected by SM fermions annihilating to ν_R through off-shell Z' bosons, given by $\mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R} a^4/H$, is below the range covered in figure 5.6 and is not shown. The vertical orange dashed line marks when $T_{\rm SM} = M_{Z'}/8$, after which temperature we find empirically that the production of Z' bosons from the SM plasma is negligible.

The left panel in figure 5.6 corresponds to a parameter point where $\Gamma_{Z'}$ exceeds the Hubble rate at some $T_{\rm SM} > M_{Z'}/2$. The Z' bosons produced after $\Gamma_{Z'} = H$ are short-lived and decay within a Hubble time. The SM plasma keeps producing Z' bosons until $T_{\rm SM} \sim M_{Z'}/8$, and thus the energy injection into ν_R ends once $T_{\rm SM} < M_{Z'}/8$. The right panel of figure 5.6 corresponds to a parameter point where $(\Gamma_{Z'}/H)_{T_{\rm SM}=M_{Z'}/2} \ll 1$. In this scenario, the SM plasma first produces Z' bosons via inverse decays. The production of Z' bosons ends once $T_{\rm SM} < M_{Z'}/8$. Subsequently, $n_{Z'}$ evolves adiabatically until $\Gamma_{Z'}$ becomes of the order of H, after which Z' decays into SM particles as well as ν_R.

We now develop analytic approximations to the final value of ρ_{ν_R} for the short- and long-lived Z' cases separately.

Dark radiation production for short-lived Z' bosons: In the regime where the Z's are cosmologically short-lived, $(\Gamma_{Z'}/H)_{T_{\rm SM}=M_{Z'}/2} > 1$, the Boltzmann equations can be simplified by noticing that after $\Gamma_{Z'} = H$ the abundance of Z' bosons follows a quasi-static equilibrium where the production rate of Z' bosons balances its decay rate. Setting the RHS of eq. (5.23) to zero and replacing $\mathcal{C}_{ff\to Z'}$ and $\mathcal{C}_{Z'\to\nu_R\nu_R}$ using eq. (**??**) gives the quasi-static equilibrium abundance of Z' bosons,

$$n^{qs}_{Z'} = \frac{\Gamma_{Z'\to SM}}{\Gamma_{Z'}}\tilde{n}_\zeta(T_{\rm SM}) + \frac{\Gamma_{Z'\to\nu_R}}{\Gamma_{Z'}}\tilde{n}_\zeta(T_{\nu_R}), \tag{5.26}$$

where $\tilde{n}_\zeta$ is defined in eq. (**??**). Substituting this quasi-static abundance $n^{qs}_{Z'}$ into eq. (5.24), we obtain an effective collision term describing energy injection into ν_R given by

$$\mathcal{C}_{ff\to\nu_R\nu_R} = \frac{3M^3_{Z'}}{2\pi^2}\left[\frac{\Gamma(Z'\to\nu_R)\Gamma(Z'\to SM)}{\Gamma_{Z'}}\right]\left(T_{\rm SM}G_1\left(\frac{M_{Z'}}{T_{\rm SM}}\right) - T_{\nu_R}G_1\left(\frac{M_{Z'}}{T_{\nu_R}}\right)\right) + \mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}, \tag{5.27}$$

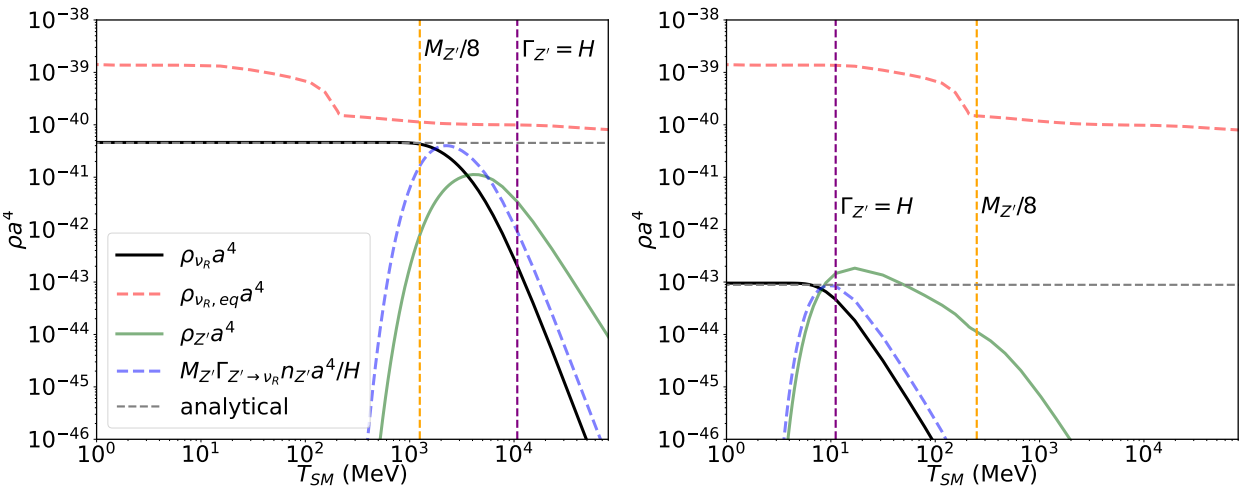


Figure 5.6: Evolution of comoving ν_R energy density (black line) and comoving Z' energy density (light green line) as a function of SM temperature for $\{M_{Z'}, g'\} = \{10 \text{ GeV}, 10^{-8}\}$ and $\{M_{Z'}, g'\} = \{2\times \text{ GeV}, 2\times 10^{-11}\}$ in the left and right panels, respectively. The red line shows the evolution of $(\pi^2/30) g_{\nu_R} T_{\rm SM}^4 a^4$, and the blue dot-dashed line is $M_{Z'} \Gamma_{Z' \to \nu_R} n_{Z'} a^4 / H$. The vertical orange dashed line marks the point where $T_{\rm SM} = M_{Z'}/8$. The vertical purple dashed line marks the point where the Z' decay rate equals the Hubble rate. The gray dashed line shows the analytical estimate of the asymptotic value of $\rho_{\nu_R} a^4$, which is calculated using eq. (5.28) in the left panel and eq. (5.33) in the right panel.

where G_1 is a dimensionless function given by eq. (A.18). In this regime, the collision term in eq. (5.27) reproduces the collision term calculated using the complete $ff \to \nu_R \nu_R$ cross-section, including the on-shell Z' bosons, as we show explicitly in appendix A.6. As this collision term no longer depends on $\rho_{Z'}$, we need only solve for ρ_{ν_R} and $\rho_{\rm SM}$ to find the contribution of ν_R to $\Delta N_{\rm eff}$. Thus, in the short-lived Z' limit, the resulting system of Boltzmann equations is similar to that for the MCP model, eq. (5.5).

We can determine the asymptotic value of ρ_{ν_R} by following steps similar to those in section 5.2.3 to obtain eq. (5.13). We can neglect the contribution from $\mathcal{C}^{\rm off}_{ff \to \nu_R \nu_R}$ because the net energy transferred to out-of-equilibrium ν_R through contact interactions is much smaller than the resonantly-enhanced contribution from the on-shell collision term. The fraction of SM energy transferred into ν_R is then given by

$$\left(\frac{\rho_{\nu_R}}{\rho_{\rm SM}}\right)_{\rm leak} = \frac{15\sqrt{3}}{64\pi^3 [g_*(4\Lambda)\pi^2/30]^{3/2}} \frac{M_{\rm Pl}}{\Lambda} \times L, \tag{5.28}$$

where $\Lambda = M_{Z'}/8$, and

$$L = 6\pi^2 \kappa_1 \left[\frac{\Gamma_{Z' \to \nu_R} \Gamma_{Z' \to SM}}{\Gamma_{Z'}}\right] = \frac{3\pi\kappa_1}{4} g'^2 \left[\frac{\Gamma_{Z' \to SM}}{\Gamma_{Z'}}\right]. \tag{5.29}$$

This limiting result for the comoving density of ν_R, $\rho_{\nu_R} a^4 = (\rho_{\nu_R}/\rho_{\rm SM})_{\rm leak} (a^4 \rho_{\rm SM})_{T_{\rm SM} = M_{Z'}/8}$, is shown by the gray dashed line in the left panel of figure 5.6, which demonstrates its agreement with the numerically evaluated asymptote of $\rho_{\nu_R} a^4$. Eq. (5.28) is only valid as long as the ν_R do not thermalize with the SM plasma. Numerically we find that for $(\Gamma_{Z'}/H)_{T_{\rm SM} = M_{Z'}/2} \gtrsim 30$, the ν_R thermalize with the SM plasma and the final density ratio is simply given by $(\rho_{\nu_R}/\rho_{\rm SM})_f = g_{\nu_R}/g_*$.

Out-of-equilibrium dark radiation production from long-lived Z' bosons: To solve for the dark radiation density in ν_R in the case where the Z' bosons are long-lived, we first need to calculate the freeze-in abundance of Z'. To proceed, we make two simplifications. First, since the Z' bosons are long-lived and, until they decay, the ν_R abundance is negligible, we can neglect the decays of Z' as well as the inverse decays of ν_R into Z' when calculating the freeze-in Z' abundance. Second, because $\rho_{\rm SM} \gg \rho_{Z'}$, ρ_{ν_R} we neglect the contributions of the Z' and ν_R in determining the Hubble rate. Assuming the SM degrees of freedom remain constant until the production of Z' ends at $T_{\rm SM} \lesssim M_{Z'}/8$, we can then simply integrate eq. (5.23) for $\rho_{Z'}$. With $\rho_{Z'} = M_{Z'} n_{Z'}$ and $\mathcal{C}_{ff\to Z'}$ given by eq. (**??**), the frozen-in abundance of Z' bosons is then

$$a^3 n_{Z'}^{\rm fz-in} = (aT_{\rm SM})^3_{T_{\rm SM}=M_{Z'}/8} \times \frac{3\lambda}{8\pi^2}\left[\frac{\Gamma_{Z'\to \rm SM}}{H(T_{\rm SM}=M_{Z'}/2)}\right], \tag{5.30}$$

where $\lambda = \int G_1(1/x)x^{-5}dx \approx 5.93$.

The frozen-in population of Z' boson eventually decays, and accordingly the comoving number density evolves as

$$n_{Z'} = n_{Z'}^{\rm fz-in}\, e^{-\Gamma_{Z'}t}. \tag{5.31}$$

Note that the final number density of ν_R is not affected by whether the Z' bosons decay before or after achieving their freeze-in abundance. This is because the number density of ν_R is set by the branching ratio of Z' decays into ν_R and the number of Z' bosons produced by the SM plasma, neither of which depend on when the Z' bosons decay. In contrast, the energy density of the ν_R does depend on the timing of the Z' boson decay because the ν_R are produced with a fixed energy of $M_{Z'}/2$, which subsequently redshifts as $1/a$. Consequently, Z' bosons that decay later result in more energetic ν_R at recombination, and thus a larger contribution to $N_{\rm eff}$.

The asymptotic value of ρ_{ν_R} is found by substituting the evolution of the massive Z's, eq. (5.31), into the Boltzmann equation for ν_R, eq. (5.24). Once again, both inverse decays of ν_R into Z' and off-shell contributions to SM fermion annihilation can be ignored in comparison to the contribution from Z' decays. The resulting ρ_{ν_R} is given by

$$\left(\frac{\rho_{\nu_R}}{\rho_{\rm SM}}\right)_{\rm decay} = \frac{M_{Z'}\Gamma_{Z'\to\nu_R}}{a^4\rho_{\rm SM}}\int_0^a \frac{\tilde{a}^3 n_{Z'}^{\rm fz-in} e^{-\Gamma_{Z'}t}}{H}d\tilde{a} \tag{5.32}$$

$$\approx \frac{2\sqrt{2}\lambda}{15\sqrt{\pi}\kappa_1}\left(\frac{\rho_{\nu_R}}{\rho_{\rm SM}}\right)_{\rm leak}\left[\frac{H(T_{\rm SM}=M_{Z'}/2)}{\Gamma_{Z'}}\right]^{1/2}\left[\frac{g_*^3(M_{Z'}/2)g_*(T_{\rm decay})}{g_*^4(M_{Z'}/8)}\right]^{1/4}, \tag{5.33}$$

where $T_{\rm decay}$ is the SM temperature at which $H(T_{\rm decay}) = \Gamma_{Z'}$ and $(\rho_{\nu_R}/\rho_{\rm SM})_{\rm leak}$ is defined in eq. (5.28). In the second line we approximated g_* to be constant around $T_{\rm decay}$ and set $g_{*s} = g_*$. The numerical coefficient in eq. (5.33) and the ratio of g_* factors in the square brackets are both $\mathcal{O}(1)$. Consequently, ρ_{ν_R} is enhanced by a factor of $(\sqrt{H/\Gamma_{Z'}})_{T_{\rm SM}=M_{Z'}/2}$ if the Z' bosons are long-lived compared to the cases where the Z' bosons decay instantaneously. The right panel of figure 5.6, shows the analytical estimate of the comoving density of ν_R, given by $\rho_{\nu_R}a^4 = (\rho_{\nu_R}/\rho_{\rm SM})_{\rm decay}\,(a^4\rho_{\rm SM})_{T_{\rm SM}=T_{\rm decay}}$, as the gray dashed line. At late times, this analytical estimate is in close agreement with the numerically evaluated $\rho_{\nu_R}a^4$, as shown by the black solid line.

5.3.4 Dark radiation production and analytical approximations to the $N_{\rm eff}$ constraint

In this section we provide analytical expressions for the curves of constant $\Delta N_{\rm eff}$ in the $B-L$ parameter space. We consider the out-of-equilibrium and equilibrated regions of parameter space separately.

Out-of-equilibrium dark radiation production In the case when ν_R remains out-of-equilibrium with the SM plasma, the dependence of the final dark radiation density, ρ_{ν_R}, on the $B-L$ coupling, g', depends on whether $\Gamma_{Z'}$ is larger or smaller than the Hubble rate at $T_{\rm SM}=M_{Z'}/2$. We find that the g' and $M_{Z'}$ values on the $\Delta N_{\rm eff}=0.3$ constraint contour typically satisfy $(\Gamma_{Z'}/H)_{T_{\rm SM}=M_{Z'}/2}\gtrsim 1$. Consequently, we use eq. (5.28) to evaluate the constraint on g' and $M_{Z'}$ for $\Delta N_{\rm eff}<0.3$. In particular, we adiabatically evolve ρ_{ν_R} given in eq. (5.28) from the end of energy injection at $T_{\rm SM}=M_{Z'}/8$ to recombination and restrict the $\Delta N_{\rm eff}$ shift given in eq. (5.21) to remain below the $(\Delta N_{\rm eff})_{\rm max}$ upper bound set by CMB measurements. Doing so yields

$$g'^2 \quad < \quad 5.8\ \times\ 10^{-19}\left(\frac{g_*(M_{Z'}/2)}{10}\right)^{3/2}\left(\frac{g_*(M_{Z'}/8)}{10}\right)^{1/3}\left(\frac{\Gamma_{Z'}}{\Gamma_{Z'\to {\rm SM}}}\right)\left(\frac{(\Delta N_{\rm eff})_{\rm max}}{0.3}\right)\left(\frac{M_{Z'}}{\rm GeV}\right). \quad (5.34)$$

The ratio of decay widths here is typically an $\mathcal{O}(1)$ number depending on the value of $M_{Z'}$.

For the $\Delta N_{\rm eff}=0.06$ and $\Delta N_{\rm eff}=0.027$ constraint contours, the condition $(\Gamma_{Z'}/H)_{T_{\rm SM}=M_{Z'}/2}>1$ is satisfied above and to the right of the red dot in figure 5.5. Consequently, the analytical result for short-lived Z's in eq. (5.34) also applies to the $\Delta N_{\rm eff}=0.06$ and $\Delta N_{\rm eff}=0.027$ contours in this region. To find an analytical result applicable below and to the left of the red dot, we start from the expression for ρ_{ν_R} given in eq. (5.33). We then evolve ρ_{ν_R} adiabatically from the end of Z' decays at $T_{\rm decay}$ to recombination. The corresponding constraint on g' is then given by

$$g'^2 < 8.1\times 10^{-21}\left(\frac{g_*(M_{Z'}/2)}{10}\right)\left(\frac{g_*(M_{Z'}/8)}{10}\right)^2\left(\frac{g_*(T_{\rm decay})}{10}\right)^{1/6}\left(\frac{\Gamma_{Z'}}{\Gamma_{Z'\to\nu_R}}\right)\left(\frac{\Gamma_{Z'}}{\Gamma_{Z'\to {\rm SM}}}\right)^2 \times\left(\frac{(\Delta N_{\rm eff})_{\rm max}}{0.06}\right)^2\left(\frac{M_{Z'}}{\rm GeV}\right). \quad (5.35)$$

Note that the constraint on g' for short-lived Z's, given in eq. (5.34), is proportional to $(\Delta N_{\rm eff})_{\rm max}$ while the long-lived Z' result in eq. (5.35) is proportional to $(\Delta N_{\rm eff})^2_{\rm max}$. The delayed Z' decays parametrically enhance the ultimate dark radiation density and hence the sensitivity of $N_{\rm eff}$ measurements to the model parameters.

Dark radiation production in the equilibrium regime: If the right-handed neutrinos thermalize with the SM, then the final comoving energy density in ν_R depends on the decoupling temperature, T_d, which is only logarithmically sensitive to g'. The thermalization threshold for the $B-L$ model can be calculated in a similar manner to the MCP model in section 5.2.3 above, see eq. (5.19). Since the ratio of the energy injection rate to the Hubble rate, $\Gamma_E/H=\mathcal{C}_{\rm f}/(\rho_{\nu_R,eq}H)$, is negligible for temperatures below $T_{\rm SM}\sim M_{Z'}/8$ in the $B-L$ model (as compared to to $T_{\rm SM}\sim m/4$ in the MCP model), the thermalization threshold in eq. (5.19) is for the $B-L$ model

$$M_{Z',{\rm th}}=8T_d[(\Delta N_{\rm eff})_{\rm max},g_{\nu_R}], \quad (5.36)$$

where T_d is defined through eq. (5.17). In figure 5.5, this thermalization threshold for $\Delta N_{\rm eff} = 0.3$ occurs around $M_{Z',\rm th} \sim 1.7 \times 10^3$. For $\Delta N_{\rm eff} < 0.14$, the thermalization threshold is pushed to arbitrarily large values of $M_{Z',\rm th}$ because restricting $\Delta N_{\rm eff} < 0.14$ rules out ever having three right-handed neutrinos in thermal equilibrium with the SM plasma, assuming no new degrees of freedom in the SM.

The exponential weakening of the $\Delta N_{\rm eff} = 0.3$ curve in figure 5.5 stops around $M_{Z'} \sim 15$ GeV, after which the constraint follows $g' \lesssim M_{Z'}/(16\ \text{TeV})$ (see also Ref. [156, 193]). At these large masses the contact-operator-mediated annihilations, described by $\mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}$, dominate over on-shell Z' production, $M_{Z'}\Gamma_{Z'\to\nu_R} n_Z$, in determining decoupling temperatures that are much smaller than $M_{Z'}$.

Note that, unlike the MCP model, the $B-L$ model does not have an excluded strip in parameter space extending up to high masses. For such a strip to exist, the mediator between the SM particles and the BSM relativistic particles must have stronger couplings to the BSM relativistic particles than it does to the SM. In the $B-L$ model, the decay width of Z' into SM particles is larger than its decay width into ν_R, in contrast to the MCP model, where the mediating MCP has much stronger interactions with the dark photons than with the SM photon.

5.4 Dark radiation constraints on classes of hidden sectors

We have so far considered dark radiation constraints on specific, minimal BSM models where a particle ϕ with mass $m_\phi \gtrsim$ MeV has renormalizable couplings to both the SM and new relativistic particles. While in general the heavy particle could be the SM Higgs boson (or indeed the Z boson), in this work we focus on the case where the heavy particle is a new SM gauge singlet particle. Additionally we focus on mass scales $m_\phi \gtrsim$ MeV because for lighter masses the constraints from stellar cooling observations generically become important. We have seen in two specific examples that such models will be stringently tested by upcoming CMB experiments that promise to measure $N_{\rm eff}$ to an accuracy of $(\Delta N_{\rm eff})_{\rm max} = 0.06$ at 95% confidence. In particular, we have demonstrated that, while detailed constraints on the parameter space require numerical evaluation of a coupled system of Boltzmann equations, a conservative, semi-analytic estimate of the allowed parameter space can be made by making a number of simplifying assumptions.

In this section we highlight the general methodology and assumptions required to estimate this conservative constraint and argue that the constraint holds even when the new relativistic particles are part of a much larger hidden sector (HS). We then explore the restrictions on HS model building that will be placed by upcoming CMB measurements of $N_{\rm eff}$.

5.4.1 Hidden sector models

We consider classes of HS models that contain light degrees of freedom that are relativistic during recombination. These degrees of freedom may be required by symmetries (as the ν_R were above), or they may be required to sequester entropy to facilitate the freezeout of HS dark matter. Stellar cooling observations strongly constrain direct renormalizable couplings of these light degrees of freedom to stable SM particles, and we assume they couple to the SM through new, heavy, SM gauge-singlet particles, ϕ, with masses which we take to be $m_\phi >$ MeV. The existence of this portal coupling enables the production of the mediator ϕ particles in the early universe via annihilation or decay of SM particles. These mediator particles then lead to the production of the light degrees of freedom in the HS, whose energy density is constrained by measurements of $N_{\rm eff}$. These $N_{\rm eff}$ constraints are applicable as long as these light degrees of freedom remain relativistic during

recombination. For masses larger than 0.1 eV, the combined constraints from the large scale structure and the CMB measurements are more stringent [195].

We seek to derive conservative constraints on the couplings of such sectors to the SM by estimating the production of dark radiation. The precise computation of the dark radiation density depends on the details of the HS, such as the number of degrees of freedom, masses of the particles, and interactions between them. However, we argue that a lower bound on the dark radiation density can be estimated from the production of mediator particles by the SM plasma, provided the energy in the mediator is preferentially transferred to the HS degrees of freedom. This amounts to assuming that the mediator interacts more strongly with the HS than with the SM. Any HS energy density that subsequently becomes freestreaming dark radiation at the decoupling of the CMB is minimized if all the energy dilutes as radiation as soon as it is produced. Therefore, assuming that all the energy that is transferred to the mediator particles by the SM plasma is rapidly deposited into light degrees of freedom in the HS provides a lower bound on the resulting dark radiation density. This lower bound provides a conservative estimate on the shift in $\Delta N_{\rm eff}$. Below we quantify this conservative estimate for different types of couplings between the mediator and the Standard Model. As above, we separate the estimates into the regions where the HS thermalizes, and those where it remains out of equilibrium.

Out-of-equilibrium dark radiation production

Practically, in the out-of-equilibrium regime, our conservative estimate of $\Delta N_{\rm eff}$ is obtained by assuming a hidden sector consisting of a single massless particle together with a massive mediator ϕ that couples to the SM. The HS equation of state is taken to be that of radiation, $w_{\rm HS} = 1/3$. The dark radiation density is determined by solving the Boltzmann equation given by eq. (5.5), where the collision term is determined by all the energy transfer processes from the SM plasma into the HS bath enabled by the portal coupling. While taking into account all the energy transfer process depends on the specifics of a particular HS model under consideration, energy transfer through the production of ϕ is common in the vast majority of HS models. Consequently, to obtain a conservative estimate of the asymptotic energy density in dark radiation, we evaluate the collision term only for processes involving on-shell production of ϕ. For renormalizable interactions with the SM, these processes are annihilations of SM particles, $aa \to \phi\phi$; decays of SM particles $a \to \phi\phi$; or inverse decays of SM particles $aa \to \phi$. In all cases a denotes a SM particle coupled to ϕ via a renormalizable portal coupling. We further restrict our attention to s-channel processes, which are independent of the properties of the HS radiation bath as long as the interaction proceeds well out of equilibrium.

In the out-of-equilibrium regime, we can analytically find the energy transferred into the HS by taking the SM temperature to evolve as $T \propto 1/a$ (a good approximation away from mass thresholds). The calculation is analogous to that in section 5.2.3 leading to eq. (5.10), and we obtain

$$\left(\frac{\rho_{\rm HS}}{\rho_{\rm SM}}\right)_{T_{\rm SM}=\Lambda} \approx \frac{\sqrt{3} M_{\rm Pl}}{[g_*(4\Lambda)\pi^2/30]^{3/2}} \int_0^\infty \frac{dT_{\rm SM}}{T_{\rm SM}^7} \mathcal{C}_{\rm f}, \tag{5.37}$$

where Λ is the energy injection decoupling temperature, below which the production of ϕ ends and $\mathcal{C}_{\rm f}$ is the forward energy transfer collision term for production of ϕ.

The integral over the forward collision term can be carried out given a specific model for the cross-section, allowing us to express the energy density injected during out-of-equilibrium scattering in terms of a leak

factor L,

$$L = \frac{64\pi^3}{15}\Lambda \int_0^\infty \frac{dT_{\rm SM}}{T_{\rm SM}^7}\mathcal{C}_{\rm f}. \tag{5.38}$$

Annihilation of SM particles into the HS. For annihilations of SM particles into the HS, $aa \to \phi\phi$, the forward collision term is

$$\mathcal{C}_{\rm f} = \sum_f \frac{1}{32\pi^4}\int_{4\max(m_a,m_\phi)^2}^\infty ds(s-4m_a^2)s\sigma_{aa\to\phi\phi}(s)T_{\rm SM}G_{\zeta_f}(\sqrt{s}/T_{\rm SM}). \tag{5.39}$$

This production process occurs in the millicharged particle model when the SM fermions annihilate into millicharged particles. The corresponding energy injection decoupling temperature and the leak factor are given by

$$\Lambda = \frac{1}{4}\max(m_a, m_\phi), \qquad L = \kappa_{\zeta_a}\Lambda\int_{64\Lambda^2}^\infty ds\frac{(s-4m_a^2)}{s\sqrt{s}}\sigma_{aa\to\phi\phi}(s). \tag{5.40}$$

Here $\sigma_{aa\to\phi\phi}(s)$ is the spin-summed CM frame cross-section and κ_{ζ_a} is determined by the quantum statistical distribution of a, as described below eq. (5.12).

Inverse decay of SM particles into the HS. For the inverse decay process, $aa \to \phi$, the collision term is of the form

$$\mathcal{C}_{\rm f} = m_\phi\Gamma_{\phi\to a}\tilde{n}_{\zeta_a}(T_{\rm SM}), \tag{5.41}$$

where $\Gamma_{\phi\to a}$ is the decay width of ϕ to a, and $\tilde{n}_{\zeta_a}$ is given by eq. (**??**). We encountered this production process for the gauged $B-L$ model in section 5.3.2. After integrating the RHS of eq. (5.37) for the process $aa \to \phi$, the final result can can be written in the form of eq. (5.38) with

$$\Lambda = \frac{m_\phi}{8}, \qquad L = 2\pi^2 g_\phi \kappa_{\zeta_a}\frac{\Gamma_{\phi\to a}}{m_\phi}, \tag{5.42}$$

where g_ϕ is the number of spin degrees of ϕ.

The final HS energy density calculated using eq. (5.42) is different from the one we obtained in the case of the gauged $B-L$ model for two reasons. First, the decay width of Z' into ν_R is smaller than its total decay width into SM particles for $M_{Z'} > 2m_e$. Thus most of the energy transferred into Z' bosons does not end up in ν_R but is rather returned to the SM plasma. If we consider the Z' bosons to couple much more strongly with additional HS particles, then the above calculation would accurately reflect the minimum energy transferred into the HS. Second, the final energy density in ν_R is enhanced when the Z' bosons are long lived.

Decays of SM particles into the HS. Finally for decays of SM particles into the HS, $a \to \phi\phi$, the collision term is of the form

$$\mathcal{C}_{\rm f} = m_a\Gamma_{a\to\phi}n_{{\rm eq},a}(T_{\rm SM}), \tag{5.43}$$

where

$$n_{\mathrm{eq},a} = g_a \int \frac{d^3p}{(2\pi)^3} \frac{1}{[e^{E/T_{\mathrm{SM}}} + \zeta_a]} \tag{5.44}$$

is the equilibrium number density of particle a. This production process occurs in the millicharged particle model when the Z bosons decay into millicharged particles.[11] Again the energy transferred into the HS can be expressed by eq. (5.38) with

$$\Lambda = \frac{m_a}{8}, \qquad L = \pi^2 g_a \tilde{\kappa}_{\zeta_a} \frac{\Gamma_{a\to\phi}}{m_a}, \tag{5.45}$$

where $\tilde{\kappa}_1 = 31\pi^6/30240$, $\tilde{\kappa}_0 = 1$, and $\tilde{\kappa}_{-1} = \pi^6/945$.

When the HS remains out of equilibrium with the SM plasma, we can find the contribution to ΔN_{eff} by starting from eq. (5.38) and then adiabatically evolving ρ_{HS} as radiation from the end of energy injection until recombination. Requiring ΔN_{eff} to be less than the CMB sensitivity, $(\Delta N_{\mathrm{eff}})_{\mathrm{max}}$, yields

$$L < g_*^{3/2}(4\Lambda) g_*^{1/3}(\Lambda) (\Delta N_{\mathrm{eff}})_{\mathrm{max}} \frac{\Lambda}{M_{\mathrm{Pl}}}. \tag{5.46}$$

The above calculations assume that all produced ϕ particles decay rapidly into relativistic HS particles. This assumption holds if ϕ has sufficiently strong couplings with HS particles. This is a conservative assumption because a long-lived ϕ would result in a larger density in the HS, and a larger shift in ΔN_{eff}.

Equilibrium dark radiation production

If the HS thermalizes with the SM plasma, then the final energy density in the HS depends on the decoupling temperature, T_d, which is only logarithmically sensitive to the strength of the portal coupling. Consequently, the N_{eff} constraint on the portal coupling become exponentially weak once the HS thermalizes. Similar to the case of the $B-L$ and millicharged particle models, the weakening of constraints occur for values of the energy injection decoupling temperature, Λ, larger than

$$\Lambda_{\mathrm{th}} \equiv T_d[(\Delta N_{\mathrm{eff}})_{\mathrm{max}}, g_{\mathrm{HS}}], \tag{5.47}$$

where T_d is given by eq. (5.17). Thus, the N_{eff} constraint given by eq. (5.46) is only valid for $\Lambda < \Lambda_{\mathrm{th}}$. Note that a larger g_{HS} would push the thermalization threshold given in eq. (5.47) to larger Λ_{th}. Thus, eq. (5.46) together with a thermalization threshold scale Λ_{th} calculated assuming $g_{\mathrm{HS}} = 1$ provides a conservative constraint on the portal coupling that is independent of details within the hidden sector.

5.4.2 Implications for HS model building

In this section, we have argued that under fairly generic conditions, a conservative lower bound on the energy density in dark radiation in a generic HS may be estimated. This lower bound can in turn be used to place bounds on the couplings between a mediator that couples the dark sector to the SM. As future measurements of N_{eff} become more and more precise, increasing pressure will be placed on models of BSM physics that contain light states contributing to dark radiation.

[11] This production process can also be realized in the case of a SM singlet scalar coupling through the Higgs portal, in which case the Higgs boson can decay into pairs of scalar fields; for a specific recent application producing dark radiation through this coupling, see [148].

From a different perspective, this analysis also points at ways such models may be brought into agreement with future data. There are a number of possibilities. In particular, one may simply be able to arrange the couplings so that the mediator interacts more strongly with the SM than the HS, and thereby energy is transferred back into the SM from the mediator. Another possibility is to have new degrees of freedom in equilibrium with the SM plasma that become non-relativistic after $T_{\rm SM} < \Lambda$. Consequently, the annihilation of the new degrees of freedom heat the SM plasma relative to the HS, diluting the dark radiation today. Similarly, if a massive field comes to dominate the universe and subsequently decays predominantly into the SM at some temperature $T_{\rm rh} < \Lambda$, the resulting reheating of the SM relaxes constraints from $N_{\rm eff}$; this mechanism was invoked, for instance, to ameliorate dark radiation constraints on Twin Higgs models [196, 197]. The entropy of the SM plasma can also increase if the SM comes into equilibrium with a new light species at a temperature $T_{\rm eq} < \Lambda$ that later becomes nonrelativistic and deposits its entropy into the SM. This mechanism generally requires a light BSM field with couplings to the SM that become cosmologically important at late times. While stellar cooling constraints are typically prohibitive for models that realize equilibration *after* SM neutrino decoupling [198], the BSM dark radiation considered here has a thermal decoupling scale $\Lambda >$ MeV and thus suppressing $N_{\rm eff}$ using this mechanism can be much simpler. Finally, $N_{\rm eff}$ also decreases if one or more of the states contributing to the dark radiation at $T \sim \Lambda$ can decay back into the SM prior to recombination; in this case the decay can produce visible signatures in light element abundances and/or CMB spectral distortions, depending on the details of the decay.

If one considers a minimal extension of the SM, where the SM has renormalizable interactions with a single massive particle in the HS and the cosmological evolution of the SM plasma is not otherwise altered, then one cannot completely evade the bounds set by $N_{\rm eff}$ measurements. However, the constraints can be somewhat ameliorated if the relic energy density in the HS does not always evolve as free-streaming dark radiation. For instance, if the relativistic HS particles have strong self-interactions, such that they behave as an ideal fluid during recombination, then they would instead contribute to $N_{\rm fluid}$, the constraints on which are are weaker by factors of 2-3 compared to $N_{\rm eff}$ [177]. Examples of this scenario include interacting neutrino models, recently surveyed in [199]. Alternatively, while a single hot HS relic that subsequently becomes nonrelativistic is more stringently constrained than if it remains relativistic [195], the combination of $N_{\rm eff}$ and large-scale structure constraints may be mitigated in a system with more than one hot relic if one HS species becomes non-relativistic before recombination while at least one other species remains relativistic. In principle, one can obtain a conservative constraint on portal interactions between the SM and a HS containing light degrees of freedom that can accommodate such variations in the spectrum of the HS by combining both CMB and large-scale structure measurements. We leave this to future work.

5.5 Summary and discussion

In this work, we have studied the production of dark radiation in scenarios where the SM has renormalizable interactions with a heavy ($m_\phi >$ MeV) gauge singlet mediator that annihilates or decays into dark radiation prior to BBN. We have focused on two specific minimal models: (i) a MCP model with a massless dark photon, and (ii) a gauged $B - L$ model with light right-handed neutrinos. By numerically solving the relevant Boltzmann equations, we have computed the resulting dark radiation abundance and determined the corresponding shifts in $N_{\rm eff}$ in the regions of parameter space relevant for upcoming CMB experiments. We present updated CMB constraints for the MCP model, and have shown that future CMB measurements will be sensitive enough to either rule out or discover the extended MCP model invoked to explain the

EDGES anomaly [155]. In the case of the gauged $B-L$ model, our computations extend and improve previous analyses by taking into account all relevant out-of-equilibrium processes, including the potentially out-of-equilibrium decays of the $B-L$ gauge boson. As a result, our projected constraints on the allowed parameter space of the $B-L$ model are stronger than previous studies. In both models we take into account the quantum statistical phase space distribution for Standard Model particles, which was not done in previous studies. We find that quantum statistics provide a correction of about 10% to the predicted shift in $N_{\rm eff}$.

The relation between dark radiation production and the model parameters depends crucially on whether or not the HS comes into thermal equilibrium with the SM. We have provided simple semi-analytical recipes to obtain the predicted shift in $N_{\rm eff}$ in both cases. When the HS remains out of equilibrium with the SM, we have demonstrated that the resulting dark radiation density is determined by the energy transfer rate from the SM into the HS at temperatures of order the mediator mass, $(\Gamma_E/H)_{T_{\rm SM}\sim m_\phi}$. The energy transfer rate typically goes like $\Gamma_E(T_{\rm SM}\sim m_\phi)\propto g_\phi^2 m_\phi$, where g_ϕ is the Standard Model coupling with the heavy mediator particle with mass m_ϕ. Consequently, the contour of constant $\Delta N_{\rm eff}$ relates $g_\phi\propto\sqrt{m_\phi/M_{\rm Pl}}$, which accounts approximately for the shape of the contours in the regions where the sectors are out-of equilibrium in figures 5.1 and 5.5. We provide a simple formula for evaluating the resulting $N_{\rm eff}$ constraint, given an input cross-section. CMB $\Delta N_{\rm eff}$ constraints are already the leading limit on both models in most of the out-of-equilibrium parameter space, along with constraints from SN1987A; these astrophysical and cosmological constraints far exceed terrestrial accelerator constraints in the sub-GeV regime.

As one increases m_ϕ at a fixed value of the dark radiation density, the coupling g_ϕ can increase to a point where the HS comes into thermal equilibrium with the SM. When the HS thermalizes with the SM, the resulting dark radiation density is determined by the temperature at which the HS and SM decouple. This decoupling temperature is primarily determined by Boltzmann suppression of the collision term. Consequently, the decoupling temperature is mainly set by the mass of the mediator, m_ϕ, and only depends logarithmically on the coupling g_ϕ; once the sectors are in thermal equilibrium, increasing the coupling only marginally decreases the resulting decoupling temperature, and thus marginally increases the resulting dark radiation density. Because of the weak sensitivity to g_ϕ, the constraint imposed by $N_{\rm eff}$ measurements on g_ϕ is exponentially weakened if the HS thermalizes with the SM. This effect gives rise to a thermalization mass threshold, $m_{\rm th}$, beyond which the constraint curves in figures 5.1 and 5.5 are exponentially weakened.

The example models discussed above consider a minimal hidden sector that is coupled to the SM via a heavy mediator. More generally, one can consider the mediator to communicate with a hidden sector that may have a nonminimal internal spectrum. While the exact evaluation of dark radiation production in extended models would require a numerical computation of the Boltzmann equations that take into account all internal hidden sector interactions, we have shown how to obtain a simple analytical lower bound on the relic dark radiation that depends only on the mass and coupling of the mediator, and is independent of the number of particles in the hidden sector or their internal interactions. This minimum dark radiation abundance is obtained by considering that energy transfer into the HS occurs through the production of heavy mediators by the SM plasma, and assuming that any energy transferred to the mediator is promptly deposited in the relativistic HS degrees of freedom. This amounts to assuming that the mediator is more strongly coupled to the HS than to the SM. In the regime where CMB constrains the HS to remain out of equilibrium with the SM in the early universe, this is a very mild requirement on the mediator coupling. This model-insensitive lower bound on $\Delta N_{\rm eff}$ assumes there are no BSM contributions to the entropy of the SM plasma, and that the relic dark radiation remains a free-streaming relativistic relic throughout the formation of the CMB. Relaxing these assumptions can evade our lower bound.

We have shown that future CMB measurements of $N_{\rm eff}$ have the potential to constrain portal couplings to values which typically are orders of magnitude weaker than those probed by collider experiments, and provided simple semi-analytic recipes to evaluate their reach. If future CMB observations do not find any deviation from the Standard Model prediction for $N_{\rm eff}$, hidden sector models with light species will also be out of reach for accelerator experiments, unless there are departures from the standard cosmology. This work highlights the potential of future CMB missions to significantly narrow down the space of observationally relevant BSM theories.

Chapter 6

BBN constraints on dark radiation isocurvature

6.1 Introduction

Upcoming stage 4 cosmic microwave background (CMB) experiments will make exquisite measurements of the energy content of the Universe [200]. These measurements will improve the constraint on the contribution of free-streaming radiation (through the effective number of relativistic species $N_{\rm eff}$) by an order of magnitude over current constraints. A measurement of $N_{\rm eff}$ consistent with the standard model (SM) prediction of $N_{\rm eff} = 3.044$ [31–35] will place extremely strong constraints on particle content beyond the SM [201].

Alternatively, these measurement could reveal the existence of additional light-particles (dark radiation) beyond the SM by measuring $N_{\rm eff} \neq 3.044$ at high significance. If these additional particles were ever in thermal equilibrium with the standard model, they will exhibit the usual adiabatic fluctuations in their density (see, e.g. [202]), and their effects on cosmology would be indistinguishable from additional neutrinos. However, this dark radiation may be completely decoupled from the standard model sector. In this decoupled scenario, fluctuations in the dark radiation density may be independent of the density fluctuations in the visible sector—there may be a dark-radiation isocurvature mode. Isocurvature modes are generically predicted by cosmological theories that have a second clock beyond the SM temperature field [203]. In this work we remain agnostic to the exact origin of such an initial condition and leave detailed model-building to future work.

The presence of dark radiation isocurvature affects the cosmic microwave background (CMB). Earlier work [204] constrained neutrino + dark radiation isocurvature to be less than $\sim 10^{-5}$ at scales around 500 Mpc using data from WMAP and ACT (see also Ref. [205] for non-Gaussian dark radiation isocurvature constraints). Planck is sensitive to dark-radiation isocurvature for scales ≥ 10 Mpc [16]. However, the inability to observe CMB fluctuations on angular sizes smaller than ~ 5 arcmins prohibits the estimation of isocurvature constraints on smaller scales [206]. In this work we probe dark radiation isocurvature down to ~ 1 Mpc scales through its impact on Big Bang nucleosynthesis (BBN).

BBN is a period in the early universe when the SM plasma became cold enough for the free protons and neutrons to combine and form the first nuclei. This process primarily produces Helium and deuterium (along with trace amounts of tritium and Lithium). Adiabatic fluctuations during BBN do not lead to spatial variations in the outcome of BBN. This result follows directly from the separate universe picture—different

patches of the Universe with differing density fluctuations simply appear to be a little older or younger as viewed by local observers. Since the local physics is identical, the outcome is identical. The presence of isocurvature during BBN can change the story by changing the physical conditions locally in a way that is distinguishable from a local shift in the clock. In this way, isocurvature leads to spatial variation in primordial elemental yields. Spatial variations in the yields of ^{4}He/H and D/H during BBN would then lead to corresponding differences in abundances in widely separated locations. To date, baryon isocurvature modes as a source of inhomogeneous BBN have been extensively studied in the literature [207–211]. In this work we consider the effect of dark radiation isocurvature on the primordial elemental abundances from BBN.

Our results can be summarized as follows. We demonstrate that the presence of dark-radiation isocurvature leads to spatially varying elemental abundances. As a result, galactic ^{4}He/H and D/H ratios are sensitive to dark-radiation isocurvature on galactic scales, ~ 1 Mpc. We use data on Helium abundances in nearby galaxies [212] and deuterium abundances in high-redshift Lyman-α absorption systems [213] to place constraints on the existence of dark radiation isocurvature. We constrain the variance of average isocurvature fluctuations in galaxies, to be less than $0.13/\Delta\bar{N}_{\rm eff}$ at 2σ confidence for scales around ~ 1 Mpc. Here $\Delta\bar{N}_{\rm eff}$ is the spatially averaged increase in $N_{\rm eff}$ due to the additional dark radiation component. In the absence of any dark-radiation, i.e. $\Delta\bar{N}_{\rm eff} = 0$, our constraints are relaxed as expected.

This chapter is organized as follows. In section 6.2, we show how dark-radiation isocurvature leads to spatially varying BBN yields and demonstrate that this leads to differences in the primordial abundances of light elements in different galaxies. In section 6.3, we use excess variance in existing ^{4}He/H and D/H data to place constraints on dark-radiation isocurvature. We conclude in section 6.4. Finally, in appendix D we use the separate universe approach to demonstrate how dark-radiation isocurvature leads to spatially varying $\Delta N_{\rm eff}$.

6.2 Inhomogenous Big Bang Nucleosynthesis through dark radiation isocurvature

In this section we demonstrate the impact of dark-radiation isocurvature on BBN. We first show how the effect of dark-radiation isocurvature on BBN is distinct from the more studied baryon-isocurvature case [207–211]. We then demonstrate that dark-radiation isocurvature leads to spatially varying $N_{\rm eff}$ that in turn causes spatial variation in primordial abundances of hydrogen and helium.

The elemental abundances produced by BBN are primarily determined by two processes: the weak processes which convert neutrons to protons, and by the deuterium formation process that forms deuterium from all the remaining neutrons. The temperature at which deuterium formation begins, $T_{\rm nuc}$, is insensitive to the Hubble rate and is primarily determined by the baryon-to-photon ratio. Baryon-isocurvature modes cause the baryon-to-photon ratio to vary spatially. This in turn leads to a spatially varying $T_{\rm nuc}$, and to spatial variations of the resulting elemental abundances. In contrast, a dark-radiation isocurvature mode leads to a spatially varying $N_{\rm eff}$, as we show below. The abundance of neutrons at $T_{\rm nuc}$ is sensitive to the Hubble rate at $T_{\rm nuc}$, and since $N_{\rm eff}$ affects the Hubble rate through the Friedmann equation, an inhomogeneous $N_{\rm eff}$ leads to an inhomogeneous abundance of neutrons at $T_{\rm nuc}$.

To show how dark-radiation isocurvature leads to spatial variation in $N_{\rm eff}$, we first highlight the relation between dark-radiation energy density and $N_{\rm eff}$. At $T \sim 1$ MeV, before BBN begins, neutrinos have chemically decoupled from SM plasma and thus evolve adiabatically like dark radiation. We can therefore absorb the density of dark radiation, ρ_{DR}, into an extra neutrino component [3],

$$\Delta N_{\rm eff} = \left[\frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\frac{\rho_{DR}}{\rho_\gamma}\right]_{\rm today}. \tag{6.1}$$

Thus $\Delta N_{\rm eff}$ depends on the ratio of homogeneous densities of dark-radiation to SM.

Now consider a spherical volume of radius $r = \lambda_{\rm gal}/2$, the matter within which later collapses to form a galaxy. In appendix D we show that an isocurvature fluctuation, $\delta\rho_i'(k)$, of a superhorizon-sized mode can be absorbed into the homogeneous density. For the spherical patch we are considering, using the separate universe principle, we can absorb the net $\delta\rho_i'(\vec{x})$ inside the volume into the homogeneous density (using eq. (D.17) and eq. (D.14)),

$$\tilde{\rho}_{SM} = \rho_{SM} + \Delta\rho'_{DR} \qquad\qquad \tilde{\rho}_{DR} = \rho_{DR} - \Delta\rho'_{DR}, \tag{6.2}$$

where $\Delta\rho'_{DR} = -\Delta\rho'_{SM}$ is the average isocurvature fluctuation in the dark radiation inside a spherical volume of radius r,

$$\Delta\rho'_{DR} = \int_0^\infty \frac{d^3x}{V_r}\delta\rho'_{DR}(\vec{x})W_r(\vec{x}). \tag{6.3}$$

Here W_r is a window function which weights the integral to be within r radius from origin and V_r is the volume swept by the window function, $V_r = \int d^3x W_r(\vec{x})$. The spherical volume effectively has $\tilde{\rho}_i$ as its homogeneous density. Note that while $\Delta\rho_2'$ involves contributions from all Fourier modes, the contribution from modes $k^{-1} \ll r$ is suppressed. The suppression is because the small wavelength modes have around the same number of over-densities and under-densities in a patch much larger than the mode's wavelength. Consequently, the super-horizon assumption in eq. (6.2) approximately holds as long as r is super-horizon sized.

Due to the presence of isocurvature, $N_{\rm eff}$ inside the spherical volume is also modified

$$\Delta N_{\rm eff} \propto \frac{\tilde{\rho}_{DR}}{\tilde{\rho}_{SM}} = \frac{\rho_{DR} + \Delta\rho'_{DR}}{\rho_{SM} - \Delta\rho'_{DR}}, \tag{6.4}$$

where we have used the fact that $\rho_{SM} \propto \rho_\gamma$.

Isocurvature between the dark-radiation and the SM plasma is defined as

$$S_{\rm DR} = \frac{3}{4}\left(\frac{\delta\rho_{DR}}{\rho_{DR}} - \frac{\delta\rho_{SM}}{\rho_{SM}}\right) = \frac{3}{4}\frac{\rho_{SM}+\rho_{DR}}{\rho_{SM}\rho_{DR}}\delta\rho'_{DR}, \tag{6.5}$$

where in the second equality we have expressed S_{DR} in the uniform density gauge where $\delta\rho'_{DR} = -\delta\rho'_{SM}$. Consequently, the average isocurvature in the spherical volume is given by

$$\Delta S_{DR} = \int_0^\infty \frac{d^3x}{V_r} S_{DR}(\vec{x})W_r(\vec{x}) = \frac{3}{4}\frac{\rho_{SM}+\rho_{DR}}{\rho_{SM}\rho_{DR}}\Delta\rho'_{DR}. \tag{6.6}$$

Replacing above back in eq. (6.4), we can rewrite $\Delta N_{\rm eff}$ as

$$\Delta N_{\rm eff} \propto \frac{\rho_{DR}}{\rho_{SM}}\frac{1 + \frac{4}{3}\frac{1}{1+\rho_{DR}/\rho_{SM}}\Delta S_{DR}}{1 - \frac{4}{3}\frac{\rho_{DR}/\rho_{SM}}{1+\rho_{DR}/\rho_{SM}}\Delta S_{DR}} \xrightarrow{(\rho_{DR}/\rho_{SM})\Delta S_{DR}\ll 1} \frac{\rho_{DR}}{\rho_{SM}}\left(1 + \frac{4}{3}\Delta S_{DR}\right). \tag{6.7}$$

Because ΔS_{DR} takes different values in different regions, dark-radiation isocurvature leads to spatial variations in $\Delta N_{\rm eff}$.

While we have derived Eq. (6.7) in uniform-density slicing, this equation is gauge invariant. Uniform density slicing makes transparent the relation between $\Delta N_{\rm eff}$ in a super-horizon patch and $\mathcal{S}_{DR}$, both of which are gauge-invariant. Although the Hubble rate has the same value everywhere in uniform density slicing, the presence of dark-radiation isocurvature causes the individual SM and dark radiation densities to be inhomogeneous. Consequently, in this slicing, the photon temperature is inhomogeneous. Because the photon temperature is the relevant clock during BBN, the effect of dark radiation isocurvature on BBN is most easily understood in the slicing where the temperature is uniform. In uniform SM temperature (density) slicing, the presence of dark radiation isocurvature causes the Hubble rate to be inhomogeneous.

In the presence of a dark-radiation isocurvature mode, regions of the Universe that were causally disconnected during BBN have different primordial abundances of light elements due to their different values of $\Delta N_{\rm eff}$. For example, the D/H ratio, D, is primarily a function of $\Delta N_{\rm eff}$ and the baryon fraction $\Omega_b h^2$. Assuming small fluctuations in $\Delta N_{\rm eff}$, the fluctuation in D is given by

$$D \approx \bar{D} + \left.\frac{\partial D}{\partial \Delta N_{\rm eff}}\right|_{\Delta \bar{N}_{\rm eff}} (\Delta N_{\rm eff} - \Delta \bar{N}_{\rm eff}), \tag{6.8}$$

where $\bar{D} = D(\Delta \bar{N}_{\rm eff}, \Omega_b h^2)$. This gives us a direct relation between the variance in D, given by σ_d, and the variance in $\Delta N_{\rm eff}$ fluctuations

$$\sigma_d = \left.\frac{\partial D}{\partial \Delta N_{\rm eff}}\right|_{\Delta \bar{N}_{\rm eff}} \sigma_{N_{\rm eff}}. \tag{6.9}$$

In practice, the derivatives, $\partial D/\partial \Delta N_{\rm eff}$, can be obtained numerically from publicly available codes. In this work, we use Parthenope [214].

Immediately following BBN, the primordial abundances in the patches are conserved. As the Universe expands, and overdensities collapse to form galaxies, variations on scales smaller than those scales that collapse to form galaxies get mixed.[1] Measurements from different galaxies are therefore sensitive to isocurvature fluctuations down to galactic scales. Consequently, the scale r entering in eq. (6.6) is the comoving size of a patch, $\lambda_{\rm gal}/2$, which collapses to form the galaxies we observe

$$\lambda_{\rm gal} = 3.7 \left(\frac{M_{\rm gal}}{10^{12} M_\odot}\right)^{1/3} \left(\frac{\Omega_{\rm m} h^2}{0.14}\right)^{-1/3} {\rm Mpc}, \tag{6.10}$$

where $M_{\rm gal}$ is the mass of the galaxy. The scale $\lambda_{\rm gal}$ is larger than the horizon during BBN, $\sim$kpc, which implies that our analysis built on eq. (6.2) is self-consistent.

The value of $\Delta N_{\rm eff}$ experienced by a galaxy is sampled from a distribution with mean $\Delta \bar{N}_{\rm eff}$ and variance $\sigma_{N_{\rm eff}}$. Moreover, $\sigma_{N_{\rm eff}}$ is related to the power spectrum of isocurvature fluctuation, P_S, as

$$\sigma^2_{N_{\rm eff}} = \frac{16}{9} \Delta \bar{N}^2_{\rm eff} \langle \Delta S^2_{DR} \rangle = \frac{16}{9} \frac{\Delta \bar{N}^2_{\rm eff}}{\left[\int d^3x\, W_{\lambda_{\rm gal}/2}(\vec{x})\right]^2} \int_0^\infty \frac{dk}{k} |W_{\lambda_{\rm gal}/2}(k)|^2 \frac{k^3 P_S(k)}{2\pi^2}. \tag{6.11}$$

[1] The ratio of BBN yields to hydrogen can increase slightly during the collapse of structures [215]. However, we ignore this effect as the increase is well below the current sensitivities of measurements. Moreover, a post-BBN diffusion of elements [216] will erase differences in ^{4}He/H or D/H ratios inside the diffusion volume. Our analysis is unaffected by this post-BBN diffusion as long as the diffusion length scales are smaller than the galactic-scales.

where $W_r(k)$ and $P_S(k)$ are the Fourier transforms of $W_r(\vec{x})$ and $\langle S_{DR}(\vec{x})S_{DR}(\vec{x}\,')\rangle$ respectively. Since the details of galaxies providing ^{4}He/H or D/H are usually not observable, the accurate estimation of $W_{\lambda_{\rm gal}/2}$ is not feasible. Consequently, we cannot exactly relate the variance in the average isocurvature experienced by a galaxy, $\langle \Delta S^2_{DR}\rangle$, to the isocurvature power spectrum. However, we can obtain an approximate relation between $\langle \Delta S^2_{DR}\rangle$ and P_S. Assuming a blue-tilted isocurvature power spectrum

$$\Delta^2_S \equiv \frac{k^3 P_S(k)}{(2\pi^2)} \propto k^n, \tag{6.12}$$

with $n > 0$, and assuming a spherical Gaussian window function, $W_{\lambda/2}(k) = \exp(-k^2\lambda^2/8)$, eq. (6.11) yields

$$\langle \Delta S^2_{DR}\rangle = \frac{\Gamma(n/2)}{2}\Delta^2_S(2\lambda^{-1}_{\rm gal}). \tag{6.13}$$

Here $\Gamma(x)$ is the Euler Gamma function. As $\langle \Delta S^2_{DR}\rangle$ determines $\sigma_{N_{\rm eff}}$ which in turn informs the variance in D/H (or ^{4}He/H) ratios, the intrinsic variance in the D/H (or ^{4}He/H) ratio in a given galaxy is determined by dark-radiation isocurvature at scales $\sim \lambda_{\rm gal}/2$.

6.3 Constraints from ^{4}He/H and D/H data

In this section we use data from observations of the ratios of ^{4}He/H and D/H to place constraints on dark-radiation isocurvature. We first describe the datasets which we use for our analysis and then describe our methodology for D/H and ^{4}He/H data separately.

6.3.1 Datasets

D/H measurements are taken from gas clouds that are seen in absorption against the light of an unrelated background quasar [217]. Correspondingly, by looking at the frequency distribution of the light from the quasar, one can estimate the redshift of the gas cloud as well as the column densities of neutral hydrogen and deuterium atoms.

For our analysis we use the D/H measurements provided in Ref. [212]. The data uses measurements from seven damped Lyman-α systems[2] around redshifts $z \sim 2-3$, that satisfy the strict selection criteria of precision stated in Ref. [219]. To estimate the comoving scale in the early universe from which the gas cloud formed, we require the mass of the gas cloud. While the masses of individual damped Lyman-α systems are not known, their masses have been estimated to be in the range $10^{11} - 10^{12} M_\odot$ [220, 221].

The ^{4}He abundance is derived from observations of the helium and hydrogen emission lines from H II regions in low-metallicity blue compact dwarf galaxies that have undergone little chemical evolution [222]. Regions with minimal chemical evolution are selected so as to minimize ^{4}He enrichment by stellar processes. However, there still remains some contamination that leads to an increase in the ^{4}He/H ratio over its primordial value.

In this study we use ^{4}He/H data provided in Ref. [213]. The data consists of 15 measurements of He II regions from 14 different galaxies. For our analysis, we assume that each galaxy has a uniform value of the primordial ^{4}He/H ratio. Correspondingly, we combine the two measurements of the same galaxy into

[2]The damped Lyman-α systems are distinct from the Lyman-α forest systems which provide matter structure constraints around $\gtrsim 1$ Mpc. They are differentiated on the basis of the amount of neutral hydrogen column densities, N(H I). Lyman-α forest systems are those with N(H I) $< 10^{17}$cm^{-1} and damped Lyman-α systems are those with 2×10^{20}cm$^{-1} < N$(H I) [218].

a single data point using a weighted average. Unlike in the case of deuterium measurements, the galaxies providing Helium measurements have low redshifts $z < 0.05$. Out of the 14 galaxies used in measuring ^{4}He/H abundance we find the masses of three[3] of them in the SPARC database [223]. Their masses are in the range $10^{10.2} - 10^{10.6} M_\odot$.

6.3.2 Constraints from D/H data

The gas in damped Lyman-α systems is assumed not to have produced or destroyed significant amounts of deuterium. Correspondingly, the measurement from a gas cloud samples the primordial value of D which is assumed to be drawn from a distribution with mean and variance given by $\{\bar{D}, \sigma_d\}$. The probability of getting a measurement of D_i from gas cloud i is then given by

$$P(D_i|\{\bar{D},\sigma_d\}) = \frac{1}{\sqrt{2\pi(\sigma_{n,i}^2+\sigma_d^2)}} \exp\left(-\frac{(D_i-\bar{D})^2}{2(\sigma_{n,i}^2+\sigma_d^2)}\right), \tag{6.14}$$

where $\sigma_{n,i}$ is the estimated noise in the measurement of D. We have assumed that each measurement has the same intrinsic variance in D. We do so because the damped Lyman-α systems typically have masses in the relatively narrow range $10^{11} - 10^{12} M_\odot$ [220, 221]. Correspondingly, the gas clouds in our data have roughly the same $\lambda_{\rm gal}$ (see eq. (6.10)) and thus the same variance in D (see eqs. (6.13) and (6.9)). Moreover, we have neglected covariance between different measurements. This approximation is valid because isocurvature on the scales of separation between different galaxies in our data (usually > 100 Mpc) is constrained by CMB measurements [16] to be much smaller than the variance to which our analysis is sensitive.

The constraints from D_i measurements are degenerate in $\Delta N_{\rm eff}$ and $\Omega_b h^2$. To remove this degeneracy we fix the value of $\Omega_b h^2$ using the Planck data,[4] $\Omega_b h^2 = 0.02239 \pm 0.00018 \equiv \bar{\Omega}_b h^2 \pm \sigma_{\Omega_b}$ [3], where σ_{Ω_b} is the uncertainty on the baryon density, which is assumed to be spatially homogeneous. The corresponding likelihood function is then given by

$$\mathcal{L}_0(\Delta N_{\rm eff}, \sigma_{\Delta N_{\rm eff}}) = \int_0^\infty \left[\prod_i P(D_i|\{\bar{D},\sigma_d\})\right]_{\Delta N_{\rm eff}, \Omega_b h^2, \sigma_{N_{\rm eff}}} \frac{\exp\left(-\frac{(\Omega_b h^2 - \bar{\Omega}_b h^2)^2}{2\sigma_{\Omega_b}^2}\right)}{\sqrt{2\pi\sigma_{\Omega_b}^2}} d(\Omega_b h^2). \tag{6.15}$$

We numerically marginalize over $\Omega_b h^2$ to obtain our likelihood estimate.

Using Parthenope [214] to estimate $D(\Delta N_{\rm eff}, \Omega_b h^2)$ and $\frac{\partial D}{\partial \Delta N_{\rm eff}}$, we find the 1σ and 2σ limits on $\{\Delta\bar{N}_{\rm eff}, \sigma_{N_{\rm eff}}\}$ shown as orange contours in left panel of figure 6.1.

6.3.3 Constraints from ^{4}He/H data

The methodology used in analysing D/H data is also applicable for ^{4}He/H data after accounting for the ^{4}He produced by stellar processes. To estimate the amount of primordial ^{4}He/H ratio, Y_p, in a given galaxy we assume a linear relation between the oxygen to hydrogen ratio (O/H) and the ^{4}He/H ratio produced by

[3]The galaxies of whose masses we found are aliased as Mrk 209, Mrk 71 and SBS 1415+437 in [213]. While their aliases used in SPARC database are UGCA 281, NGA 2366, and PGC51017, respectively.

[4]The Planck constraints on $\Omega_b h^2$ are slightly degenerate with $N_{\rm eff}$. Correspondingly we take Planck constraints on $\Omega_b h^2$ values marignalised over $N_{\rm eff}$ from TT+TE+EE+lowl+lowlE+BAO data.

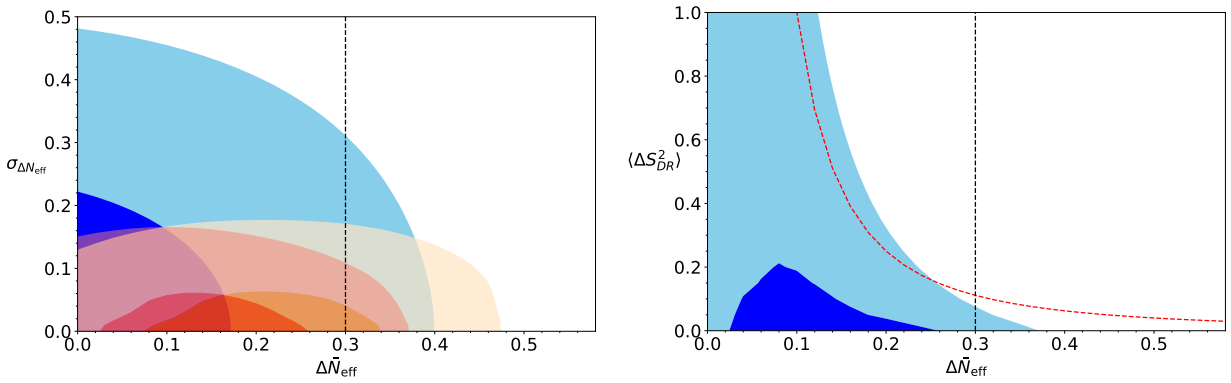


Figure 6.1: *Left*: 1σ and 2σ constraints on $\{\Delta\bar{N}_{\rm eff}, \sigma_{\Delta N_{\rm eff}}\}$ from ^{4}He/H data (blue contours), D/H data (orange contours) and from combined data (red contours). The vertical dashed line denotes the upper limit on $\Delta N_{\rm eff}$ from Planck data with 95% confidence [3]. *Right*: Constraints on $\{\Delta\bar{N}_{\rm eff}, \langle\Delta S^2_{\rm DR}\rangle\}$ from combined data. Here $\Delta S_{\rm DR}$ is the average dark radiation in a galaxy. For blue-tilted isocurvature mode $\langle\Delta S^2_{\rm DR}\rangle$ is approximately same as the normalized isocurvature power spectrum at 1 Mpc, $\Delta^2_S(1{\rm Mpc}^{-1})$ (see eq. (6.13)). The red-dashed line marks the parameters space at which $\Delta\bar{N}_{\rm eff}\Delta S_{\rm DR} = 0.1$. Correspondingly, the small $\Delta\bar{N}_{\rm eff}\Delta S_{\rm DR}$ approximation made in eq. (6.7) holds for most of our parameter space.

stellar processes. Thus the net ^{4}He/H ratio, Y, found in a galaxy is given by

$$Y = Y_p + m \times ({\rm O/H}), \tag{6.16}$$

where m is the proportionality between O/H production and ^{4}He/H production through stellar processes. Above, Y_p fluctuates with $\Delta N_{\rm eff}$ in a similar manner as in eq. (6.8), except with D replaced by Y_p. Similarly, the variance in Y_p, given by σ_y, and the variance in $\Delta N_{\rm eff}$ fluctuations are related by eq. (6.9), except with D replaced by Y_p.

Taking into account the linear relation between O/H and Y, the probability of getting a measurement of Y_i from galaxy i is given by

$$\tilde{P}(Y_i|\{\bar{Y}_p, \sigma_y, m\}) = \frac{1}{\sqrt{2\pi(\sigma^2_{n,i} + \sigma^2_y + m^2\sigma^2_{{\rm O}n,i})}} \exp\Big(-\frac{(Y_i - (\bar{Y}_p + m({\rm O/H})_i))^2}{2(\sigma^2_{n,i} + \sigma^2_y + m^2\sigma^2_{{\rm O}n,i})}\Big), \tag{6.17}$$

where $\sigma_{{\rm O}n,i}$ is the noise in O/H measurement. Just like in the case of deuterium, we have considered all galaxies to have the same variance in $\Delta N_{\rm eff}$ and neglected covariance between different galaxies. We marginalize over $\Omega_b h^2$ as we did for deuterium in eq. (6.15). Additionally, as the precise value of m is unknown, we explicitly marginalize over m assuming a uniform prior,

$$\begin{aligned}\mathcal{L}(\Delta N_{\rm eff}, \sigma_{N_{\rm eff}}) = \int_0^\infty \int_0^\infty \Big[\prod_i \tilde{P}(Y_i|\{\bar{Y}, \sigma_y, m\})\Big]_{\Delta N_{\rm eff}, \Omega_b h^2, \sigma_{N_{\rm eff}}} \\ \times \frac{\exp\Big(-\frac{(\Omega_b h^2 - \bar{\Omega}_b h^2)^2}{2\sigma^2_{\Omega_b}}\Big)}{\sqrt{2\pi\sigma^2_{\Omega_b}}} dm\ d(\Omega_b h^2).\end{aligned} \tag{6.18}$$

Since the ^{4}He/H and O/H data prefers values of $m \sim O(10^2)$ [213], we have $m\sigma_{{\rm O}n,i} \sim 10^{-2}\sigma_{n,i}$. Thus we neglect the contribution from $m^2\sigma^2_{{\rm O}n,i}$ terms in our likelihood function. Consequently, using the definition of

$\mathcal{L}_0$ in eq. (6.15) but with D replaced by Y, the above integral simplifies to,

$$\begin{aligned}\mathcal{L}(\Delta N_{\rm eff}, \sigma_{N_{\rm eff}}) \approx & \mathcal{L}_0(\Delta N_{\rm eff}, \sigma_{N_{\rm eff}}) \\ & \times \int_0^\infty \exp\left(-\frac{1}{2}\left[\sum_i \frac{({\rm O/H})_i^2}{(\sigma_{n,i}^2+\sigma_y^2)}\right] m^2 + \left[\sum_i \frac{(Y_i-\bar{Y})({\rm O/H})_i}{(\sigma_{n,i}^2+\sigma_y^2)}\right] m\right) dm \\ = & \mathcal{L}_0(\Delta N_{\rm eff}, \sigma_{N_{\rm eff}}) \frac{\sqrt{\pi} e^{b^2/(2a)}\left({\rm erf}\left(\frac{b}{\sqrt{2a}}\right)+1\right)}{\sqrt{a}},\end{aligned} \tag{6.19}$$

where a and b are the inverse variance weighted sum of $({\rm O/H})_i^2$ and $(Y_i-\bar{Y})({\rm O/H})_i$ respectively (terms in the square bracket in the first line). Using Parthenope [214] to estimate $Y(\Delta N_{\rm eff}, \Omega_b h^2)$ and $\frac{\partial Y}{\partial \Delta N_{\rm eff}}$ we find the 1σ and 2σ limits on $\{\Delta\bar{N}_{\rm eff}, \sigma_{N_{\rm eff}}\}$, shown as blue contours in left panel of figure 6.1.

The red contours in the left panel of figure 6.1 show the combined constraints from Helium and deuterium data, which restricts the variance in $\Delta N_{\rm eff}$ to be $\sigma_{N_{\rm eff}} \leq 0.17$ at 95% confidence. In the right panel of figure 6.1 we convert the constraints on $\sigma_{N_{\rm eff}}$ to constraints on $\langle \Delta S_{\rm DR}^2 \rangle$ using eq. (6.11). Since the masses of the galaxies used in our measurements lie in the range $10^{10} - 10^{12} M_\odot$, we consider all our measurements to have $\lambda_{\rm gal} \sim 2$ Mpc (see eq. (6.10)). Correspondingly, the variance in average dark radiation isocurvature in a galaxy, $\langle \Delta S_{\rm DR}^2 \rangle$ approximately measures isocurvature on scales around $\lambda_{\rm gal}/2 \sim 1$ Mpc, i.e. $\langle \Delta S_{\rm DR}^2 \rangle \sim \Delta_S^2(k \sim 1{\rm Mpc}^{-1})$ (see eq. (6.13)). The constraints on isocurvature become significantly weaker for smaller values of $\Delta\bar{N}_{\rm eff}$. This is indicative of the fact that smaller dark radiation densities make it harder for the isocurvature component to gravitationally affect BBN.

6.4 Discussion and conclusions

In a universe with a dark radiation field that is populated independently of the SM sector following inflation, an isocurvature mode can naturally occur between the two sectors. In this work, we have demonstrated that such an isocurvature mode leads to spatially varying BBN yields. Correspondingly, we have derived constraints on the existence of an isocurvature mode between SM plasma and putative dark radiation by looking at spatial variations in ^{4}He/H and D/H abundances.

A lack of excess variance in observed ^{4}He/H and D/H data limits the amount of isocurvature present during BBN. Assuming each galaxy has internally uniform ^{4}He/H and D/H ratios, a single galaxy probes dark radiation isocurvature at scales $\sim \lambda_{\rm gal}/2$. Here $\lambda_{\rm gal}$ corresponds to the comoving size of the overdensity which eventually collapses to form the galaxy in question. Since the structures which provide ^{4}He/H or D/H measurements typically have masses around 10^{11} $M_\odot$, our analysis is sensitive to dark radiation isocurvature at scales ~ 1 Mpc. Subsequently, using ^{4}He/H data from measurements of nearby galaxies [212] and D/H data from measurements of high-redshift Lyman-α absorption systems [213], we constrained the variance of average dark radiation isocurvature fluctuations, to be $\sqrt{\langle \Delta S_{\rm DR}^2 \rangle} < 0.13/\Delta\bar{N}_{\rm eff}$ (see right panel of figure 6.1) at 95% confidence. The quantity $\langle \Delta S_{\rm DR}^2 \rangle$ is approximately the same as the normalized isocurvature power spectrum at 1 Mpc. The exact relation between $\langle \Delta S_{\rm DR}^2 \rangle$ and the isocurvature power spectrum requires an accurate estimation of the primordial overdensities that collapse to form the galaxies in our data. Finally, we also constrain the variance in $\Delta N_{\rm eff}$ to be $\sigma_{N_{\rm eff}} < 0.17$ at 95% confidence.

By translating the neutrino isocurvature constraints by Planck [16] to dark radiation isocurvature,[5] we

[5]CMB measurements cannot distinguish between the effects from dark-radiation and neutrinos. Consequently the isocurvature, $\mathcal{S}$ constrained by CMB would have contributions from both dark-radiation and neutrinos, $\mathcal{S} = \frac{3}{4}\left(\frac{\delta\rho_{DR}+\delta\rho_\nu}{\rho_{DR}+\rho_\nu} - \frac{\delta\rho_\gamma}{\rho_\gamma}\right) =$

find that dark radiation isocurvature fluctuations are constrained to be less than $\sim 10^{-5} \times N_{\rm eff}/\Delta\bar{N}_{\rm eff}$ on scales larger than ~ 10 Mpc. Although the CMB is much more sensitive probe than the ^{4}He/H and D/H data, the latter is able to probe isocurvature at scales that are inaccessible to CMB measurements.

If dark radiation and dark matter fluctuations are correlated—which could occur, for example, in theories of dark freeze-out in the presence of isocurvature—then constraints from measurements of clustering in the Lyman-α forest on dark matter isocurvature can be translated to dark radiation. The Lyman-α forest data constraints DM isocurvature to be less than 10^{-4} at 1 Mpc [224], by putting limits on the excess power over the adiabatic matter power spectrum extrapolated from CMB measurements. In contrast, the analysis in this study is sensitive directly to the isocurvature mode between dark radiation and the SM sector, and is unaffected by the adiabatic fluctuations.

The extension of constraints from Lyman-α forest data to smaller scales is limited by solving non-linear structure formation—a complication that does not affect our analysis. In contrast, the techniques used in this study can theoretically be used to extend the constraints down to scales ~ 0.01 Mpc, i.e. scales slightly larger than the comoving horizon during BBN. To achieve the constraints at such small scales one would require measurements of ^{4}He/H or D/H from structures with masses of order $\sim 10^6 - 10^7 M_\odot$. Potential future measurements of ^{4}He/H in halos of masses $\sim 10^{9.5} M_\odot$ [225] would extend the constraints down to ~ 0.3 Mpc.

$\frac{\rho_{DR}}{\rho_{DR}+\rho_\nu}\mathcal{S}_{DR} + \frac{\rho_\nu}{\rho_{DR}+\rho_\nu}\mathcal{S}_\nu$. The subscript ν refer to quantities for neutrinos.

Chapter 7

Summary and conclusion

This thesis explores various novel early universe cosmologies that are possible with hidden sectors and that can potentially lead to observable signatures in the cosmic microwave background spectrum or in the distribution of matter today.

The thesis particularly focuses on hidden sectors that are out-of equilibrium with the Standard Model plasma. A simple mechanism to populate such hidden sectors is by making the inflaton to couple with both Standard Model and hidden sector particles with different couplings. Consequently, after the end of inflation, inflaton decays can populate the two sectors at different temperatures. However, inflaton mediated interaction between the hidden sector and the Standard Model can alter the temperature ratio between the two sectors from the naive expectation of the temperature ratio being determined by the ratio of inflaton couplings to the two sectors. Chapter 3 highlights a simple criteria that determined when the naive expectation is violated: when the inflaton decays occur at temperatures larger than quarter of inflaton mass, $T_{\rm rh} > M_\phi/4$. The simple criteria is a consequence of the fact that the temperature ratio is primarily determined by processes that occur later in time. As the energy transfer via inflaton mediated interactions dominantly occur at $T \sim M_\phi/4$ because of on-shell resonance, these interactions determine the final temperature ratio for $T_{\rm rh} > M_\phi/4$.

A hidden sector that is out-of-equilibrium with the Standard Model plasma can alter the early universe cosmology in a variety of ways. Chapter 4 discusses one such way where the hidden sector causes an early cannibal dominated era (ECDE) prior to Big Bang nucleosynthesis (BBN). If the lightest particle in the hidden sector has strong number changing interactions, then as the particle becomes non-relativistic its interactions can maintain chemical equilibrium via self-annihilations. When such a cannibal particle comes to dominate the universe, it can lead to a growth in dark matter perturbations with the peak enhancement in the perturbations being determined by the thermal pressure in the cannibal fluid due to cannibal annihilations. These enhanced perturbations collapse to form dark matter microhalos long before the collapse that occurs in standard cosmology. As the mass and central density of the microhalos are determined by the peak enhancement in the dark matter perturbations, we show a direct map between the observable properties of the microhalos and the properties of the cannibal particle.

If the lightest particle in the hidden sector is massless, then the hidden sector can leave imprints on the cosmic microwave background (CMB) spectrum because the CMB spectrum is sensitive to dark radiation. Conventionally, the energy density in dark radiation is parameterized by its contribution to the effective number of neutrino species, $N_{\rm eff}$. As upcoming CMB measurement of $N_{\rm eff}$ become more precise, they will constraint the lightest hidden sector particle to always be out-of-equilibrium with the Standard Model plasma

in the early universe if no deviation is observed from the Standard Model value of $N_{\rm eff}$. In chapter 5 we show that the $N_{\rm eff}$ measurements can directly constrain the energy transfer rate between an out-of-equilibrium hidden sector and the Standard Model without being dependent on the number of particles in the hidden sector or internal hidden sector interactions. As the energy transfer rate in the early universe is determined by beyond Standard Model interactions, $N_{\rm eff}$ measurements can be used to directly constrain beyond Standard Model interactions at energies $E >$ MeV. The $N_{\rm eff}$ constraints on beyond Standard Model interactions are found to be orders of magnitude stronger than the sensitivity of collider experiments.

Apart from populating the hidden sector via inflaton decays or directly from interactions with Standard Model particles, another natural possibility is that the hidden sector was populated by a curvaton field. In such scenarios, if the hidden sector remains thermally decoupled from the Standard Model plasma, then the density perturbations of the two sectors are out of sync, i.e. they have an isocurvature mode between them. In chapter 6, we show that such an isocurvature causes a variation in the Hubble rate in different patches of the universe during BBN. This inhomogeneous Hubble rate in turn leads to an inhomogeneous production of the Helium to hydrogen ratio (He/H) as well as the deuterium to hydrogen ratio (D/H). Consequently, by looking at the variation in D/H data collected from different Lyman-alpha clouds and the He/H data collected from different galaxies, the variance of average dark radiation isocurvature fluctuations was constrained to $\sqrt{\langle \Delta S_{\rm DR}^2 \rangle} < 0.13/\Delta \bar{N}_{\rm eff}$ at 95% confidence, where $\Delta \bar{N}_{\rm eff}$ is the shift in $N_{\rm eff}$ due to the dark radiation.

Taken together, the analyses in this thesis points towards interesting implications for hidden sector model building relevant for upcoming experiments. For instance, we have argued that if future CMB experiments do not detect dark radiation then portal couplings between Standard Model and hidden sectors containing relativistic particles will typically be restricted to values that are orders of magnitude smaller than the sensitivities of collider experiments. Then to have a chance of detecting hidden sectors in future collider experiments, one would either have to consider the Standard Model plasma to be reheated at temperatures below the mass scale of the portal interaction or consider hidden sectors without dark radiation. On one hand, reheating at low temperatures implies an early matter dominated era (EMDE) and thus a possibly enhanced abundance of dark matter microhalos today. On the other hand, hidden sector with a massive lightest particle typically overclose the universe if the lightest particle is also stable. And if the lightest massive hidden sector particle is unstable, then we open the possibility of hidden sector causing an early matter dominated era (EMDE) or an early cannibal dominated era (ECDE), which would also enhance the abundance of dark matter microhalos today. Thus, for hidden sectors theories that can be observed in colliders, we argue that one should either expect a detection of dark radiation in future CMB experiments or an increased possibility of large abundance of dark matter microhalos.

References

[1] P. Zyla et al. (Particle Data Group), "Review of Particle Physics", PTEP **2020**, 083C01 (2020).

[2] P. A. R. Ade et al. (Planck), "Planck 2015 results. XIII. Cosmological parameters", Astron. Astrophys. **594**, A13 (2016).

[3] N. Aghanim et al. (Planck), "Planck 2018 results. VI. Cosmological parameters", (2018).

[4] M. Escudero, A. Berlin, D. Hooper, and M.-X. Lin, "Toward (Finally!) Ruling Out Z and Higgs Mediated Dark Matter Models", JCAP **1612**, 029 (2016).

[5] J. Alexander et al., "Dark Sectors 2016 Workshop: Community Report", in (2016).

[6] E. W. Kolb, D. Seckel, and M. S. Turner, "The Shadow World", Nature **314**, 415–419 (1985).

[7] E. D. Carlson, M. E. Machacek, and L. J. Hall, "Self-interacting dark matter", Astrophys. J. **398**, 43–52 (1992).

[8] K. Sigurdson, "Hidden Hot Dark Matter as Cold Dark Matter", (2009).

[9] C. Cheung, G. Elor, L. J. Hall, and P. Kumar, "Origins of Hidden Sector Dark Matter I: Cosmology", JHEP **03**, 042 (2011).

[10] D. Pappadopulo, J. T. Ruderman, and G. Trevisan, "Dark matter freeze-out in a nonrelativistic sector", Phys. Rev. **D94**, 035005 (2016).

[11] J. A. Dror, E. Kuflik, and W. H. Ng, "Codecaying Dark Matter", Phys. Rev. Lett. **117**, 211801 (2016).

[12] A. Berlin, D. Hooper, and G. Krnjaic, "Thermal Dark Matter From A Highly Decoupled Sector", Phys. Rev. **D94**, 095019 (2016).

[13] J. A. Dror, E. Kuflik, B. Melcher, and S. Watson, "Concentrated dark matter: Enhanced small-scale structure from codecaying dark matter", Phys. Rev. **D97**, 063524 (2018).

[14] J. Georg, B. Melcher, and S. Watson, "Primordial Black Holes and Co-Decaying Dark Matter", (2019).

[15] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, "Freeze-In Production of FIMP Dark Matter", JHEP **03**, 080 (2010).

[16] Y. Akrami et al. (Planck), "Planck 2018 results. X. Constraints on inflation", Astron. Astrophys. **641**, A10 (2020).

[17] P. Adshead, P. Ralegankar, and J. Shelton, "Reheating in two-sector cosmology", JHEP **08**, 151 (2019).

[18] A. L. Erickcek, "The Dark Matter Annihilation Boost from Low-Temperature Reheating", Phys. Rev. **D92**, 103505 (2015).

[19] C. Blanco, M. S. Delos, A. L. Erickcek, and D. Hooper, "Annihilation Signatures of Hidden Sector Dark Matter Within Early-Forming Microhalos", Phys. Rev. **D100**, 103010 (2019).

[20]M. Sten Delos, T. Linden, and A. L. Erickcek, "Breaking a dark degeneracy: The gamma-ray signature of early matter domination", Phys. Rev. D **100**, 123546 (2019).

[21]J. A. Dror, H. Ramani, T. Trickle, and K. M. Zurek, "Pulsar timing probes of primordial black holes and subhalos", Phys. Rev. D **100**, 023003, 023003 (2019).

[22]H. Ramani, T. Trickle, and K. M. Zurek, "Observability of Dark Matter Substructure with Pulsar Timing Correlations", (2020).

[23]L. Dai and J. Miralda-Escudé, "Gravitational Lensing Signatures of Axion Dark Matter Minihalos in Highly Magnified Stars", Astron. J. **159**, 49 (2020).

[24]N. Blinov, M. J. Dolan, and P. Draper, "Imprints of the Early Universe on Axion Dark Matter Substructure", Phys. Rev. D **101**, 035002 (2020).

[25]N. Blinov, M. J. Dolan, P. Draper, and J. Shelton, "Dark Matter Microhalos From Simplified Models", Phys. Rev. D **103**, 103514 (2021).

[26]A. L. Erickcek, P. Ralegankar, and J. Shelton, "Cannibal domination and the matter power spectrum", (2020).

[27]A. L. Erickcek, P. Ralegankar, and J. Shelton, "Cannibalism's lingering imprint on the matter power spectrum", JCAP **01**, 017 (2022).

[28]P. Adshead, P. Ralegankar, and J. Shelton, "Neff constraints on portal interactions with hidden sector", In preparation.

[29]P. Adshead, G. Holder, and P. Ralegankar, "BBN constraints on dark radiation isocurvature", JCAP **09**, 016 (2020).

[30]M. D'Onofrio and K. Rummukainen, "Standard model cross-over on the lattice", Phys. Rev. D **93**, 025003 (2016).

[31]G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, and P. D. Serpico, "Relic neutrino decoupling including flavour oscillations", Nuclear Physics B **729**, 221–234 (2005).

[32]E. Grohs, G. M. Fuller, C. T. Kishimoto, M. W. Paris, and A. Vlasenko, "Neutrino energy transport in weak decoupling and big bang nucleosynthesis", Physical Review D **93**, 10.1103/physrevd.93.083522 (2016).

[33]P. F. de Salas and S. Pastor, "Relic neutrino decoupling with flavour oscillations revisited", JCAP **1607**, 051 (2016).

[34]K. Akita and M. Yamaguchi, "A precision calculation of relic neutrino decoupling", (2020).

[35]M. E. Abenza, "Precision early universe thermodynamics made simple: neff and neutrino decoupling in the standard model and beyond", Journal of Cosmology and Astroparticle Physics **2020**, 048–048 (2020).

[36]D. J. H. Chung, E. W. Kolb, and A. Riotto, "Production of massive particles during reheating", Phys. Rev. **D60**, 063504 (1999).

[37]X. Chu, T. Hambye, and M. H. G. Tytgat, "The Four Basic Ways of Creating Dark Matter Through a Portal", JCAP **05**, 034 (2012).

[38]J. E. Evans, C. Gaidau, and J. Shelton, "Leak-in Dark Matter and Hidden Sectors Below the Equilibration Floor", to appear.

[39]J. McDonald, "Thermally generated gauge singlet scalars as selfinteracting dark matter", Phys. Rev. Lett. **88**, 091304 (2002).

[40]H. M. Hodges, "Mirror baryons as the dark matter", Phys. Rev. **D47**, 456–459 (1993).

[41]Z. G. Berezhiani, A. D. Dolgov, and R. N. Mohapatra, "Asymmetric inflationary reheating and the nature of mirror universe", Phys. Lett. B **375**, 26–36 (1996).

[42]P. Adshead, Y. Cui, and J. Shelton, "Chilly Dark Sectors and Asymmetric Reheating", JHEP **06**, 016 (2016).

[43]J. Halverson, C. Long, B. Nelson, and G. Salinas, "Axion reheating in the string landscape", Phys. Rev. D **99**, 086014 (2019).

[44]E. W. Kolb, A. Notari, and A. Riotto, "On the reheating stage after inflation", Phys. Rev. **D68**, 123505 (2003).

[45]M. Drewes and J. U. Kang, "The Kinematics of Cosmic Reheating", Nucl. Phys. **B875**, [Erratum: Nucl. Phys.B888,284(2014)], 315–350 (2013).

[46]E. Hardy and J. Unwin, "Symmetric and Asymmetric Reheating", JHEP **09**, 113 (2017).

[47]M. Reece and T. Roxlo, "Nonthermal production of dark radiation and dark matter", JHEP **09**, 096 (2016).

[48]A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems", Phys. Rev. **D23**, [Adv. Ser. Astrophys. Cosmol.3,139(1987)], 347–356 (1981).

[49]A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems", Phys. Lett. **108B**, [Adv. Ser. Astrophys. Cosmol.3,149(1987)], 389–393 (1982).

[50]A. Albrecht and P. J. Steinhardt, "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking", Phys. Rev. Lett. **48**, [Adv. Ser. Astrophys. Cosmol.3,158(1987)], 1220–1223 (1982).

[51]J. Martin, C. Ringeval, and V. Vennin, "Encyclopedia Inflationaris", Phys. Dark Univ. **5-6**, 75–235 (2014).

[52]M. A. Amin, M. P. Hertzberg, D. I. Kaiser, and J. Karouby, "Nonperturbative Dynamics Of Reheating After Inflation: A Review", Int. J. Mod. Phys. **D24**, 1530003 (2014).

[53]K. D. Lozanov and M. A. Amin, "Equation of State and Duration to Radiation Domination after Inflation", Phys. Rev. Lett. **119**, 061301 (2017).

[54]M. S. Turner, "Coherent Scalar Field Oscillations in an Expanding Universe", Phys. Rev. **D28**, 1243 (1983).

[55]J. H. Traschen and R. H. Brandenberger, "Particle Production During Out-of-equilibrium Phase Transitions", Phys. Rev. **D42**, 2491–2504 (1990).

[56]L. Kofman, A. D. Linde, and A. A. Starobinsky, "Reheating after inflation", Phys. Rev. Lett. **73**, 3195–3198 (1994).

[57]W. T. Emond, P. Millington, and P. M. Saffin, "Boltzmann equations for preheating", (2018).

[58]L. F. Abbott, E. Farhi, and M. B. Wise, "Particle Production in the New Inflationary Cosmology", Phys. Lett. **117B**, 29 (1982).

[59]A. Albrecht, P. J. Steinhardt, M. S. Turner, and F. Wilczek, "Reheating an Inflationary Universe", Phys. Rev. Lett. **48**, 1437 (1982).

[60]P. Adshead, J. T. Giblin, T. R. Scully, and E. I. Sfakianakis, "Gauge-preheating and the end of axion inflation", JCAP **1512**, 034 (2015).

[61]P. Adshead and E. I. Sfakianakis, "Fermion production during and after axion inflation", JCAP **1511**, 021 (2015).

[62]G. Kane, K. Sinha, and S. Watson, "Cosmological Moduli and the Post-Inflationary Universe: A Critical Review", Int. J. Mod. Phys. **D24**, 1530022 (2015).

[63]R. T. Co, F. D'Eramo, L. J. Hall, and D. Pappadopulo, "Freeze-In Dark Matter with Displaced Signatures at Colliders", JCAP **1512**, 024 (2015).

[64]J. F. Dufaux, G. N. Felder, L. Kofman, M. Peloso, and D. Podolsky, "Preheating with trilinear interactions: Tachyonic resonance", JCAP **0607**, 006 (2006).

[65]G. Krnjaic, "Probing Light Thermal Dark-Matter With a Higgs Portal Mediator", Phys. Rev. **D94**, 073009 (2016).

[66]J. A. Evans, S. Gori, and J. Shelton, "Looking for the WIMP Next Door", JHEP **02**, 100 (2018).

[67]N. Arkani-Hamed, T. Cohen, R. T. D'Agnolo, A. Hook, H. D. Kim, and D. Pinner, "Solving the Hierarchy Problem at Reheating with a Large Number of Degrees of Freedom", Phys. Rev. Lett. **117**, 251801 (2016).

[68]R. Allahverdi, M. A. Amin, A. Berlin, N. Bernal, C. T. Byrnes, M. S. Delos, A. L. Erickcek, M. Escudero, D. G. Figueroa, K. Freese, and et al., "The first three seconds: a review of possible expansion histories of the early universe", The Open Journal of Astrophysics **4**, 10.21105/astro.2006.16182 (2021).

[69]B. Spokoiny, "Deflationary universe scenario", Phys. Lett. B **315**, 40–45 (1993).

[70]M. Joyce, "Electroweak Baryogenesis and the Expansion Rate of the Universe", Phys. Rev. D **55**, 1875–1878 (1997).

[71]P. G. Ferreira and M. Joyce, "Cosmology with a primordial scaling field", Phys. Rev. D **58**, 023503 (1998).

[72]X. Chen and S. .-. H. Tye, "Heating in brane inflation and hidden dark matter", JCAP **06**, 011 (2006).

[73]J. L. Feng, H. Tu, and H.-B. Yu, "Thermal Relics in Hidden Sectors", JCAP **0810**, 043 (2008).

[74]Y. Zhang, "Long-lived Light Mediator to Dark Matter and Primordial Small Scale Spectrum", JCAP **05**, 008 (2015).

[75]A. Berlin, D. Hooper, and G. Krnjaic, "PeV-Scale Dark Matter as a Thermal Relic of a Decoupled Sector", Phys. Lett. **B760**, 106–111 (2016).

[76]A. L. Erickcek and K. Sigurdson, "Reheating Effects in the Matter Power Spectrum and Implications for Substructure", Phys. Rev. D **84**, 083503 (2011).

[77]G. Barenboim and J. Rasero, "Structure formation during an early period of matter domination", Journal of High Energy Physics **2014**, 138, 138 (2014).

[78]J. Fan, O. Özsoy, and S. Watson, "Nonthermal histories and implications for structure formation", Phys. Rev. D **90**, 043536, 043536 (2014).

[79]K. Redmond, A. Trezza, and A. L. Erickcek, "Growth of Dark Matter Perturbations during Kination", Phys. Rev. D **98**, 063504 (2018).

[80]A. L. Erickcek, K. Sinha, and S. Watson, "Bringing Isolated Dark Matter Out of Isolation: Late-time Reheating and Indirect Detection", Phys. Rev. D **94**, 063502 (2016).

[81]S. Heeba, F. Kahlhoefer, and P. Stöcker, "Freeze-in production of decaying dark matter in five steps", JCAP **11**, 048 (2018).

[82]S. Heimersheim, N. Schöneberg, D. C. Hooper, and J. Lesgourgues, "Cannibalism hinders growth: Cannibal Dark Matter and the S_8 tension", JCAP **12**, 016 (2020).

[83]K. K. Boddy, J. L. Feng, M. Kaplinghat, and T. M. P. Tait, "Self-Interacting Dark Matter from a Non-Abelian Hidden Sector", Phys. Rev. **D89**, 115017 (2014).

[84]K. K. Boddy, J. L. Feng, M. Kaplinghat, Y. Shadmi, and T. M. P. Tait, "Strongly interacting dark matter: Self-interactions and keV lines", Phys. Rev. **D90**, 095016 (2014).

[85]N. Bernal, X. Chu, C. Garcia-Cely, T. Hambye, and B. Zaldivar, "Production Regimes for Self-Interacting Dark Matter", JCAP **03**, 018 (2016).

[86]A. Soni and Y. Zhang, "Hidden SU(N) Glueball Dark Matter", Phys. Rev. **D93**, 115025 (2016).

[87]L. Forestell, D. E. Morrissey, and K. Sigurdson, "Non-Abelian Dark Forces and the Relic Densities of Dark Glueballs", Phys. Rev. D **95**, 015032 (2017).

[88]M. Farina, D. Pappadopulo, J. T. Ruderman, and G. Trevisan, "Phases of Cannibal Dark Matter", JHEP **12**, 039 (2016).

[89]M. A. Buen-Abad, R. Emami, and M. Schmaltz, "Cannibal Dark Matter and Large Scale Structure", Phys. Rev. D **98**, 083517 (2018).

[90]A. Dolgov, "On Concentration of Relict Theta Particles (in Russian)", Yad. Fiz. **31**, 1522–1528 (1980).

[91]A. Dolgov, "New Old Mechanism of Dark Matter Burning", (2017).

[92]T. Tenkanen and V. Vaskonen, "Reheating the Standard Model from a hidden sector", Phys. Rev. **D94**, 083516 (2016).

[93]J. Kopp, J. Liu, T. R. Slatyer, X.-P. Wang, and W. Xue, "Impeded Dark Matter", JHEP **12**, 033 (2016).

[94]G. Krnjaic, "Freezing In, Heating Up, and Freezing Out: Predictive Nonthermal Dark Matter and Low-Mass Direct Detection", JHEP **10**, 136 (2018).

[95]M. E. Machacek, "Growth of adiabatic perturbations in selfinteracting dark matter", Astrophys. J. **431**, 41–51 (1994).

[96]A. A. de Laix, R. J. Scherrer, and R. K. Schaefer, "Constraints of selfinteracting dark matter", Astrophys. J. **452**, 495 (1995).

[97]A. Das, B. Dasgupta, and R. Khatri, "Ballistic Dark Matter oscillates above ΛCDM", JCAP **04**, 018 (2019).

[98]M. Kawasaki, K. Kohri, and N. Sugiyama, "Cosmological constraints on late time entropy production", Phys. Rev. Lett. **82**, 4168 (1999).

[99]M. Kawasaki, K. Kohri, and N. Sugiyama, "MeV scale reheating temperature and thermalization of neutrino background", Phys. Rev. **D62**, 023506 (2000).

[100]S. Hannestad, "What is the lowest possible reheating temperature?", Phys. Rev. **D70**, 043506 (2004).

[101]K. Ichikawa, M. Kawasaki, and F. Takahashi, "The Oscillation effects on thermalization of the neutrinos in the Universe with low reheating temperature", Phys. Rev. **D72**, 043522 (2005).

[102]P. de Salas, M. Lattanzi, G. Mangano, G. Miele, S. Pastor, and O. Pisanti, "Bounds on very low reheating scenarios after Planck", Phys. Rev. D **92**, 123534 (2015).

[103]T. Hasegawa, N. Hiroshima, K. Kohri, R. S. Hansen, T. Tram, and S. Hannestad, "MeV-scale reheating temperature and thermalization of oscillating neutrinos by radiative and hadronic decays of massive particles", JCAP **12**, 012 (2019).

[104]C.-P. Ma and E. Bertschinger, "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges", Astrophys. J. **455**, 7–25 (1995).

[105]W. Hu and N. Sugiyama, "Small scale cosmological perturbations: An Analytic approach", Astrophys. J. **471**, 542–570 (1996).

[106]P. Meszaros, "The behaviour of point masses in an expanding cosmological substratum.", Astronomy and Astrophysics **37**, 225–228 (1974).

[107]E. Bertschinger, "The Effects of Cold Dark Matter Decoupling and Pair Annihilation on Cosmological Perturbations", Phys. Rev. D **74**, 063509 (2006).

[108]P. J. E. Peebles, *The large-scale structure of the universe* (1980).

[109]C. Dvorkin, K. Blum, and M. Kamionkowski, "Constraining Dark Matter-Baryon Scattering with Linear Cosmology", Phys. Rev. D **89**, 023519 (2014).

[110]J. Silk, "Cosmic Black-Body Radiation and Galaxy Formation", Astrophys. J. **151**, 459 (1968).

[111]W. Hu and M. J. White, "The Damping tail of CMB anisotropies", Astrophys. J. **479**, 568 (1997).

[112]O. F. Piattella, D. C. Rodrigues, J. C. Fabris, and J. A. de Freitas Pacheco, "Evolution of the phase-space density and the Jeans scale for dark matter derived from the Vlasov-Einstein equation", JCAP **11**, 002 (2013).

[113]D. Egana-Ugrinovic, R. Essig, D. Gift, and M. LoVerde, "The Cosmological Evolution of Self-interacting Dark Matter", JCAP **05**, 013 (2021).

[114]D. Tseliakhovich and C. Hirata, "Relative velocity of dark matter and baryonic fluids and the formation of the first structures", Phys. Rev. D **82**, 083520 (2010).

[115]J. Diemand, B. Moore, and J. Stadel, "Earth-mass dark-matter haloes as the first structures in the early Universe", Nature **433**, 389–391 (2005).

[116]T. Ishiyama, J. Makino, and T. Ebisuzaki, "Gamma-ray Signal from Earth-mass Dark Matter Microhalos", Astrophys. J. Lett. **723**, L195 (2010).

[117]J. F. Navarro, C. S. Frenk, and S. D. M. White, "A Universal Density Profile from Hierarchical Clustering", Astrophys. J. **490**, 493–508 (1997).

[118]J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel, "Profiles of dark haloes. Evolution, scatter, and environment", Mon. Not. Roy. Astron. Soc. **321**, 559–575 (2001).

[119]R. H. Wechsler, J. S. Bullock, J. R. Primack, A. V. Kravtsov, and A. Dekel, "Concentrations of dark halos from their assembly histories", Astrophys. J. **568**, 52–70 (2002).

[120]M. S. Delos, M. Bruff, and A. L. Erickcek, "Predicting the density profiles of the first halos", Phys. Rev. D **100**, 023523 (2019).

[121]V. Berezinsky, V. Dokuchaev, Y. Eroshenko, M. Kachelrieß, and M. A. Solberg, "Superdense cosmological dark matter clumps", Phys. Rev. D **81**, 103529, 103529 (2010).

[122]M. S. Delos, "Tidal evolution of dark matter annihilation rates in subhalos", Phys. Rev. D **100**, 063505 (2019).

[123]M. S. Delos, "Evolution of dark matter microhalos through stellar encounters", Phys. Rev. D **100**, 083529 (2019).

[124]F. Li, A. L. Erickcek, and N. M. Law, "A new probe of the small-scale primordial power spectrum: astrometric microlensing by ultracompact minihalos", Phys. Rev. D **86**, 043519 (2012).

[125]A. L. Erickcek and N. M. Law, "Astrometric Microlensing by Local Dark Matter Subhalos", Astrophys. J. **729**, 49 (2011).

[126]K. Van Tilburg, A.-M. Taki, and N. Weiner, "Halometry from Astrometry", JCAP **07**, 041 (2018).

[127]M. S. Delos and T. Linden, "Dark Matter Microhalos in the Solar Neighborhood: Pulsar Timing Signatures of Early Matter Domination", (2021).

[128]V. S. H. Lee, S. R. Taylor, T. Trickle, and K. M. Zurek, "Bayesian Forecasts for Dark Matter Substructure Searches with Mock Pulsar Timing Data", JCAP **08**, 025 (2021).

[129]C. Blanco and D. Hooper, "Constraints on Decaying Dark Matter from the Isotropic Gamma-Ray Background", JCAP **03**, 019 (2019).

[130]M. Ackermann et al. (Fermi-LAT), "The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV", Astrophys. J. **799**, 86 (2015).

[131]G. Barenboim, N. Blinov, and A. Stebbins, "Smallest remnants of early matter domination", JCAP **12**, 026 (2021).

[132]K. Abazajian et al., "CMB-S4 Science Case, Reference Design, and Project Plan", (2019).

[133]D. Green et al., "Messengers from the Early Universe: Cosmic Neutrinos and Other Light Relics", Bull. Am. Astron. Soc. **51**, 159 (2019).

[134]G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, and P. D. Serpico, "Relic neutrino decoupling including flavour oscillations", Nuclear Physics B **729**, 221â€"234 (2005).

[135]E. Grohs, G. Fuller, C. Kishimoto, M. Paris, and A. Vlasenko, "Neutrino energy transport in weak decoupling and big bang nucleosynthesis", Physical Review D **93**, 10.1103/physrevd.93.083522 (2016).

[136]P. F. de Salas and S. Pastor, "Relic neutrino decoupling with flavour oscillations revisited", JCAP **1607**, 051 (2016).

[137]K. A. Olive, D. N. Schramm, and G. Steigman, "Limits on New Superweakly Interacting Particles from Primordial Nucleosynthesis", Nucl. Phys. B **180**, 497–515 (1981).

[138]C. Brust, D. E. Kaplan, and M. T. Walters, "New Light Species and the CMB", JHEP **12**, 058 (2013).

[139]S. Weinberg, "Goldstone Bosons as Fractional Cosmic Neutrinos", Phys. Rev. Lett. **110**, 241301 (2013).

[140]S. Knapen, T. Lin, and K. M. Zurek, "Light Dark Matter: Models and Constraints", Phys. Rev. D **96**, 115021 (2017).

[141]H. An, M. Pospelov, and J. Pradler, "New stellar constraints on dark photons", Phys. Lett. B **725**, 190–195 (2013).

[142]F. D'Eramo, F. Hajkarim, and S. Yun, "Thermal QCD Axions across Thresholds", JHEP **10**, 224 (2021).

[143]R. Z. Ferreira and A. Notari, "Observable Windows for the QCD Axion Through the Number of Relativistic Species", Phys. Rev. Lett. **120**, 191301 (2018).

[144]F. D'Eramo, R. Z. Ferreira, A. Notari, and J. L. Bernal, "Hot Axions and the H_0 tension", JCAP **11**, 014 (2018).

[145]F. Arias-Aragón, F. D'Eramo, R. Z. Ferreira, L. Merlo, and A. Notari, "Production of Thermal Axions across the ElectroWeak Phase Transition", JCAP **03**, 090 (2021).

[146]D. Green, Y. Guo, and B. Wallisch, "Cosmological implications of axion-matter couplings", JCAP **02**, 019 (2022).

[147]X. Luo, W. Rodejohann, and X.-J. Xu, "Dirac neutrinos and N_{eff}. Part II. The freeze-in case", JCAP **03**, 082 (2021).

[148]A. Biswas, D. Borah, N. Das, and D. Nanda, "Freeze-in Dark Matter and $\Delta N_{\rm eff}$ via Light Dirac Neutrino Portal", (2022).

[149]E. Bernreuther, F. Kahlhoefer, M. Lucente, and A. Morandini, "Searching for dark radiation at the LHC", (2022).

[150]B. Holdom, "Two U(1)'s and Epsilon Charge Shifts", Phys. Lett. B **166**, 196–198 (1986).

[151]E. D. Carlson, "LIMITS ON A NEW U(1) COUPLING", Nucl. Phys. B **286**, 378–398 (1987).

[152]D. Feldman, P. Fileviez Perez, and P. Nath, "R-parity Conservation via the Stueckelberg Mechanism: LHC and Dark Matter Signals", JHEP **01**, 038 (2012).

[153]H. Vogel and J. Redondo, "Dark Radiation constraints on minicharged particles in models with a hidden photon", JCAP **02**, 029 (2014).

[154]R. Foot and S. Vagnozzi, "Dissipative hidden sector dark matter", Phys. Rev. D **91**, 023512 (2015).

[155]H. Liu, N. J. Outmezguine, D. Redigolo, and T. Volansky, "Reviving Millicharged Dark Matter for 21-cm Cosmology", Phys. Rev. D **100**, 123011 (2019).

[156]K. N. Abazajian and J. Heeck, "Observing Dirac neutrinos in the cosmic microwave background", Phys. Rev. D **100**, 075027 (2019).

[157]J. B. Muñoz and A. Loeb, "A small amount of mini-charged dark matter could cool the baryons in the early Universe", Nature **557**, 684 (2018).

[158]J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, "An absorption profile centred at 78 megahertz in the sky-averaged spectrum", Nature **555**, 67–70 (2018).

[159]R. Barkana, N. J. Outmezguine, D. Redigolo, and T. Volansky, "Strong constraints on light dark matter interpretation of the EDGES signal", Phys. Rev. D **98**, 103005 (2018).

[160]A. Mathur, S. Rajendran, and H. Ramani, "A composite solution to the EDGES anomaly", (2021).

[161]A. Berlin, D. Hooper, G. Krnjaic, and S. D. McDermott, "Severely Constraining Dark Matter Interpretations of the 21-cm Anomaly", Phys. Rev. Lett. **121**, 011102 (2018).

[162]E. D. Kovetz, V. Poulin, V. Gluscevic, K. K. Boddy, R. Barkana, and M. Kamionkowski, "Tighter limits on dark matter explanations of the anomalous EDGES 21 cm signal", Phys. Rev. D **98**, 103529 (2018).

[163]T. R. Slatyer and C.-L. Wu, "Early-Universe constraints on dark matter-baryon scattering and their implications for a global 21 cm signal", Phys. Rev. D **98**, 023013 (2018).

[164]R. de Putter, O. Doré, J. Gleyzes, D. Green, and J. Meyers, "Dark Matter Interactions, Helium, and the Cosmic Microwave Background", Phys. Rev. Lett. **122**, 041301 (2019).

[165]C. Creque-Sarbinowski, L. Ji, E. D. Kovetz, and M. Kamionkowski, "Direct millicharged dark matter cannot explain the EDGES signal", Phys. Rev. D **100**, 023528 (2019).

[166]L. Husdal, "On Effective Degrees of Freedom in the Early Universe", Galaxies **4**, 78 (2016).

[167]A. A. Prinz et al., "Search for millicharged particles at SLAC", Phys. Rev. Lett. **81**, 1175–1178 (1998).

[168]G. Magill, R. Plestid, M. Pospelov, and Y.-D. Tsai, "Millicharged particles in neutrino experiments", Phys. Rev. Lett. **122**, 071801 (2019).

[169]S. Davidson, S. Hannestad, and G. Raffelt, "Updated bounds on millicharged particles", JHEP **05**, 003 (2000).

[170]J. Jaeckel, M. Jankowiak, and M. Spannowsky, "LHC probes the hidden sector", Phys. Dark Univ. **2**, 111–117 (2013).

[171]J. H. Chang, R. Essig, and S. D. McDermott, "Supernova 1987A Constraints on Sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle", JHEP **09**, 051 (2018).

[172]M. Fabbrichesi, E. Gabrielli, and G. Lanfranchi, "The Dark Photon", 10.1007/978-3-030-62519-1 (2020).

[173]L. Chuzhoy and E. W. Kolb, "Reopening the window on charged dark matter", JCAP **07**, 014 (2009).

[174]A. Stebbins and G. Krnjaic, "New Limits on Charged Dark Matter from Large-Scale Coherent Magnetic Fields", JCAP **12**, 003 (2019).

[175]P. Gondolo and G. Gelmini, "Cosmic abundances of stable particles: Improved analysis", Nucl. Phys. B **360**, 145–179 (1991).

[176]N. Fernandez, Y. Kahn, and J. Shelton, "Freeze-in, glaciation, and UV sensitivity from light mediators", (2021).

[177]D. Baumann, D. Green, J. Meyers, and B. Wallisch, "Phases of New Physics in the CMB", JCAP **01**, 007 (2016).

[178]T. A. Wagner, S. Schlamminger, J. H. Gundlach, and E. G. Adelberger, "Torsion-balance tests of the weak equivalence principle", Class. Quant. Grav. **29**, 184002 (2012).

[179]J. Redondo and G. Raffelt, "Solar constraints on hidden photons re-visited", JCAP **08**, 034 (2013).

[180]D. Croon, G. Elor, R. K. Leane, and S. D. McDermott, "Supernova Muons: New Constraints on Z' Bosons, Axions and ALPs", JHEP **01**, 107 (2021).

[181]J. P. Lees et al. (BaBar), "Search for a Dark Photon in e^+e^- Collisions at BaBar", Phys. Rev. Lett. **113**, 201801 (2014).

[182]R. Aaij et al. (LHCb), "Search for Dark Photons Produced in 13 TeV pp Collisions", Phys. Rev. Lett. **120**, 061801 (2018).

[183]M. Aaboud et al. (ATLAS), "Search for new high-mass phenomena in the dilepton final state using 36 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", JHEP **10**, 182 (2017).

[184]M. Escudero, S. J. Witte, and N. Rius, "The dispirited case of gauged U(1)$_{B-L}$ dark matter", JHEP **08**, 190 (2018).

[185]M. Bauer, P. Foldenauer, and J. Jaeckel, "Hunting All the Hidden Photons", JHEP **07**, 094 (2018).

[186]J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, "New Fixed-Target Experiments to Search for Dark Gauge Forces", Phys. Rev. D **80**, 075018 (2009).

[187]S. Andreas, C. Niebuhr, and A. Ringwald, "New Limits on Hidden Photons from Past Electron Beam Dumps", Phys. Rev. D **86**, 095019 (2012).

[188]J. Blümlein and J. Brunner, "New Exclusion Limits on Dark Gauge Forces from Proton Bremsstrahlung in Beam-Dump Data", Phys. Lett. B **731**, 320–326 (2014).

[189]J. Heeck, "Unbroken B – L symmetry", Phys. Lett. B **739**, 256–262 (2014).

[190]A. Pilaftsis and T. E. J. Underwood, "Resonant leptogenesis", Nucl. Phys. B **692**, 303–345 (2004).

[191]G. F. Giudice, A. Notari, M. Raidal, A. Riotto, and A. Strumia, "Towards a complete theory of thermal leptogenesis in the SM and MSSM", Nucl. Phys. B **685**, 89–149 (2004).

[192]C. S. Shin and S. Yun, "Dark gauge boson production from neutron stars via nucleon-nucleon bremsstrahlung", JHEP **02**, 133 (2022).

[193]P. Fileviez Pérez, C. Murgui, and A. D. Plascencia, "Neutrino-Dark Matter Connections in Gauge Theories", Phys. Rev. D **100**, 035041 (2019).

[194]D. K. Hong, C. S. Shin, and S. Yun, "Cooling of young neutron stars and dark gauge bosons", Phys. Rev. D **103**, 123031 (2021).

[195]N. DePorzio, W. L. Xu, J. B. Muñoz, and C. Dvorkin, "Finding eV-scale light relics with cosmological observables", Phys. Rev. D **103**, 023504 (2021).

[196]Z. Chacko, N. Craig, P. J. Fox, and R. Harnik, "Cosmology in Mirror Twin Higgs and Neutrino Masses", JHEP **07**, 023 (2017).

[197]N. Craig, S. Koren, and T. Trott, "Cosmological Signals of a Mirror Twin Higgs", JHEP **05**, 038 (2017).

[198]A. Berlin, N. Blinov, and S. W. Li, "Dark Sector Equilibration During Nucleosynthesis", Phys. Rev. D **100**, 015038 (2019).

[199]J. M. Berryman et al., "Neutrino Self-Interactions: A White Paper", in 2022 Snowmass Summer Study (Mar. 2022).

[200]K. N. Abazajian et al. (CMB-S4), "CMB-S4 Science Book, First Edition", (2016).

[201]J. Errard, S. M. Feeney, H. V. Peiris, and A. H. Jaffe, "Robust forecasts on fundamental physics from the foreground-obscured, gravitationally-lensed CMB polarization", JCAP **1603**, 052 (2016).

[202]S. Weinberg, "Must cosmological perturbations remain non-adiabatic after multi-field inflation?", Phys. Rev. **D70**, 083522 (2004).

[203]D. H. Lyth, C. Ungarelli, and D. Wands, "The Primordial density perturbation in the curvaton scenario", Phys. Rev. **D67**, 023503 (2003).

[204]M. Kawasaki, K. Miyamoto, K. Nakayama, and T. Sekiguchi, "Isocurvature perturbations in extra radiation", JCAP **1202**, 022 (2012).

[205]E. Kawakami, M. Kawasaki, K. Miyamoto, K. Nakayama, and T. Sekiguchi, "Non-Gaussian isocurvature perturbations in dark radiation", JCAP **1207**, 037 (2012).

[206]Y. Akrami et al. (Planck), "Planck 2018 results. I. Overview and the cosmological legacy of Planck", (2018).

[207]G. P. Holder, K. M. Nollett, and A. van Engelen, "On Possible Variation in the Cosmological Baryon Fraction", **716**, 907–913 (2010).

[208]K. Inomata, M. Kawasaki, A. Kusenko, and L. Yang, "Big Bang Nucleosynthesis Constraint on Baryonic Isocurvature Perturbations", JCAP **1812**, 003 (2018).

[209]C. J. Copi, K. A. Olive, and D. N. Schramm, "Implications of a primordial origin for the dispersion in d/h in quasar absorption systems", Submitted to: Astrophys. J. (1996).

[210]R. V. Wagoner, "Big-Bang Nucleosynthesis Revisited", ApJ **179**, 343–360 (1973).

[211]J. D. Barrow and R. J. Scherrer, "Constraining density fluctuations with big bang nucleosynthesis in the era of precision cosmology", Physical Review D **98**, 10.1103/physrevd.98.043534 (2018).

[212]R. J. Cooke, M. Pettini, and C. C. Steidel, "One Percent Determination of the Primordial Deuterium Abundance", Astrophys. J. **855**, 102 (2018).

[213]E. Aver, K. A. Olive, and E. D. Skillman, "The effects of He I λ10830 on helium abundance determinations", JCAP **1507**, 011 (2015).

[214]R. Consiglio, P. F. de Salas, G. Mangano, G. Miele, S. Pastor, and O. Pisanti, "PArthENoPE reloaded", Comput. Phys. Commun. **233**, 237–242 (2018).

[215]D. Medvigy and A. Loeb, *Element diffusion during cosmological structure formation*, 2001.

[216]M. Pospelov and N. Afshordi, *Lithium diffusion in the post-recombination universe and spatial variation of [li/h]*, 2012.

[217]T. F. Adams, "The detectability of deuterium Lyman alpha in QSOs.", A& A **50**, 461–462 (1976).

[218]A. M. Wolfe, E. Gawiser, and J. X. Prochaska, "Damped lyα systems", Annual Review of Astronomy and Astrophysics **43**, 861–918 (2005).

[219]R. Cooke, M. Pettini, R. A. Jorgenson, M. T. Murphy, and C. C. Steidel, "Precision measures of the primordial abundance of deuterium", Astrophys. J. **781**, 31 (2014).

[220]R. Mackenzie, M. Fumagalli, T. Theuns, D. J. Hatton, T. Garel, S. Cantalupo, L. Christensen, J. P. U. Fynbo, N. Kanekar, P. Møller, J. O'Meara, J. X. Prochaska, M. Rafelski, T. Shanks, and J. Trayford, "Linking gas and galaxies at high redshift: MUSE surveys the environments of six damped Lyα systems at z $\approx$ 3", MNRAS, 1435 (2019).

[221]A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme, I. Pâris, P. Petitjean, J. Rich, E. Rollinde, N. P. Ross, D. P. Schneider, M. White, and D. G. York, "The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS", JCAP **2012**, 059, 059 (2012).

[222]M. Peimbert and S. Torres-Peimbert, "Chemical composition of H II regions in the Large Magellanic Cloud and its cosmological implications.", Apj **193**, 327–333 (1974).

[223]P. Li, F. Lelli, S. McGaugh, M. S. Pawlowski, M. A. Zwaan, and J. Schombert, "The halo mass function of late-type galaxies from HI kinematics", Astrophys. J. **886**, L11 (2019).

[224]M. Beltran, J. Garcia-Bellido, J. Lesgourgues, and M. Viel, "Squeezing the window on isocurvature modes with the lyman-alpha forest", Phys. Rev. **D72**, 103515 (2005).

[225]C. Sykes, M. Fumagalli, R. Cooke, and T. Theuns, "Determining the primordial helium abundance and UV background using fluorescent emission in star-free dark matter haloes", (2019).

[226]J. Birrell, C.-T. Yang, and J. Rafelski, "Relic Neutrino Freeze-out: Dependence on Natural Constants", Nucl. Phys. **B890**, 481–517 (2014).

[227]G. G. Raffelt, *Stars as laboratories for fundamental physics: The astrophysics of neutrinos, axions, and other weakly interacting particles* (May 1996).

[228]C. Dvorkin, T. Lin, and K. Schutz, "Making dark matter out of light: freeze-in from plasma effects", Phys. Rev. D **99**, 115009 (2019).

[229]S. Weinberg, "Adiabatic modes in cosmology", Phys. Rev. **D67**, 123504 (2003).

[230]D. H. Lyth and D. Wands, "The CDM isocurvature perturbation in the curvaton scenario", Phys. Rev. **D68**, 103516 (2003).

Appendix A

Energy transfer collision term with quantum statistics

In this section we calculate the collision term describing energy transfer between two baths at different temperatures via annihilation, decays, and scattering. Instead of approximating the thermal distribution of all involved particles as Maxwell-Boltzmann distribution, which is the norm, we take into account the quantum statistical thermal distributions (Bose-Einstein or Fermi-Dirac) in our calculations.

We then calculate these collision terms for the specific models used in the main body. In particular, in appendix A.4 we calculate the collision term for energy transfer mediated by inflaton, in appendix A.5 we calculate the collision terms for various energy transfer processes from Standard Model plasma into millicharged particle, and finally in appendix A.6 we calculate the collision term for energy transfer from Standard Model plasma into right handed neutrinos via massive $B-L$ gauge boson.

A.1 Annihilation

In this section we simplify the forward energy transfer collision term for particle a annihilating into particle b,

$$1(a) + 2(\bar{a}) \to 3(b) + 4(\bar{b}). \tag{A.1}$$

We start the forward collision term given by

$$\begin{aligned} \mathcal{C}_{\rm f} = \int \left[\prod_{i=1}^{4} \frac{d^4p_i}{(2\pi)^3}\delta(p_i^2 - m_i^2)\Theta(p_i^0)\right] (2\pi)^4\delta^4(p_1+p_2-p_3-p_4) S|\mathcal{M}_f|^2(p_1^0+p_2^0) \\ \times \left[f_a(p_1)f_a(p_2)(1 \pm f_b(p_3))(1 \pm f_b(p_4))\right], \end{aligned} \tag{A.2}$$

where Θ is the Heaviside function. Above $|\mathcal{M}_f|^2$ is the spin-summed matrix element and S is the symmetry factor.

In the limit density of b is more dilute compared to density of a, the final state effects (Bose-enhancement/Pauli-blocking) can be neglected regardless of the exact distribution of b, f_b. For instance, if b has large enough self-interactions to thermalize then f_b will be peaked at momentas $p \sim T_b \ll T_a$, where T is the temperature. As the b particles produced by annihilation of a will typically be at momenta $p \sim T_a$, the values of f_b probed

by the collision integral will be much smaller than one even if $f_b \sim \mathcal{O}(1)$ at its peak. In the case particle b does not thermalize, f will be peaked at $p \sim T_a$ but its value will be much less than one because density of b is more dilute compared to density of a.

By neglecting the final state effects, the phase space integration can be significantly simplified because one can directly integrate the matrix element over the phase space of outgoing particles, particles 3 and 4, as we show in section. If on the other hand we want to incorporate the contribution from final state effect, then we need to consider that the matrix element is only a function of energy in center of mass frame to simplify the integral, as we show in section.

A.1.1 Negligible final state effects

By neglecting the final state effect, the phase-space integration over p_3 and p_4 can simply be absorbed into the definition of Lorentz invariant cross-section, σ (see [175]),

$$4F\sigma_{aa\to bb} \equiv \int \frac{d^3p_3}{(2\pi)^3 2E_3}\frac{d^3p_4}{(2\pi)^3 2E_4}(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)S|\mathcal{M}_f|^2, \tag{A.3}$$

where $F = \sqrt{(p_1\cdot p_2)^2 - m_1^2 m_2^2}$. If the mass of particles 1 and 2 are the same, which is typically true for the processes we consider in this chapter, one can show that $[(2E_1)(2E_2)|\vec{v}_3 - \vec{v}_4|]$ calculated in the CM frame is equal to F. Thus for $m_1 = m_2$, σ is identical to the spin-summed center-of-mass (CM) frame cross-section. With the above mentioned simplifications, the collision term becomes

$$\mathcal{C}_{\rm f} = \int \left[\prod_{i=1}^{2}\frac{d^4p_i}{(2\pi)^3}\delta(p_i^2 - m_a^2)\Theta(p_i^0)\right] 4F(p_1^0+p_2^0)\sigma_{aa\to bb,CM} f_a(p_1) f_a(p_2). \tag{A.4}$$

The integral can be further simplified if we make the following change of variables

$$p = p_1 + p_2, \quad q = p_1 - p_2. \tag{A.5}$$

Note that in this basis the Mandelstam parameter s is simply $s = p^2$. Performing the above change of variables in the collision term integral yields

$$\begin{aligned}\mathcal{C}_{\rm f}(T_f) = &\int \frac{1}{2^4}\frac{d^4p}{(2\pi)^2}\left[\frac{d^4q}{(2\pi)^4}\delta((p+q)^2/4 - m_a^2)\delta((p-q)^2/4-m_a^2)\Theta(p^0-|q^0|)\right] \\ &\times 4Fp^0\sigma_{aa\to bb,CM} f_a((p^0+q^0)/2) f_a((p^0-q^0)/2) \qquad \text{(A.6)} \\ \equiv &\int \frac{1}{2^4}\frac{d^4p}{(2\pi)^2} dI_q \times 4Fp^0\sigma_{aa\to bb,CM} f_a((p^0+q^0)/2) f_a((p^0-q^0)/2), \qquad \text{(A.7)}\end{aligned}$$

where the phase space element dI_q is given by the square brackets in the first line.

Next we simplify dI_q. The delta functions in dI_q together impose the following constraints:

$$q_3 = \frac{q^0 p^0}{|\vec{p}|} \qquad\qquad |\vec{q_{12}}|^2 = p^2\left(1 - \frac{(q^0)^2}{|\vec{p}|^2}\right) - 4m_a^2, \tag{A.8}$$

where q_3 is the component of $\vec{q}$ along $\vec{p}$, while $\vec{q}_{12}$ is the component of $\vec{q}$ perpendicular to $\vec{p}$. Consequently, we

can perform the integral over q_3 and $|\vec{q_{12}}|$ in dI_q to integrate over the delta functions, yielding

$$dI_q = \left[\delta((p+q)^2/4 - m_a^2)\delta((q-p)^2/4 - m_a^2)\Theta(p^0 - |q^0|)dq^3\ |\vec{q_{12}}|d|\vec{q_{12}}|\right]d\theta_{12}\frac{dq^0}{(2\pi)^4} \tag{A.9}$$

$$= \frac{2}{|\vec{p}|}\Theta(p^0 - |q^0|)\Theta\left(p^2\left[1 - \frac{(q^0)^2}{|\vec{p}|^2}\right] - 4m_a^2\right)d\theta_{12}\frac{dq^0}{(2\pi)^4}, \tag{A.10}$$

where θ_{12} is the Azimuthal angle made by $\vec{q}_{12}$ in the plane perpendicular to $\vec{p}$. The second Θ function in the last line imposes the requirement that $|\vec{q_{12}}| > 0$. The two Θ function together rule out the possibility of p^2 to be less than $4m_a^2$, which is expected. Thus the arguments of the Θ functions can be rewritten as

$$dI_q = \frac{2}{|\vec{p}|}\Theta(p^2 - 4m_a^2)\Theta(|\vec{p}|\beta_a - |q^0|)d\theta_{12}\frac{dq^0}{(2\pi)^4}, \tag{A.11}$$

where

$$\beta_a = \sqrt{1 - \frac{4m_a^2}{s}}. \tag{A.12}$$

Substituting the simplified dI_q back in the collision term, we obtain

$$\mathcal{C}_{\rm f} = \int \frac{2\pi}{2^4}\frac{d^4p}{(2\pi)^6}\frac{2}{|\vec{p}|}\Theta(p^2 - 4m_a^2) \times 4Fp^0\sigma_{aa\to bb,CM}\int_{-|\vec{p}|\beta_a}^{|\vec{p}|\beta_a} dq^0 f_a((p^0+q^0)/2)f_a((p^0-q^0)/2). \tag{A.13}$$

To integrate over the phase-space distribution, we consider the particles a to be in thermal equilibrium such that

$$f_a(p) = \frac{1}{e^{p/T_a} + \zeta_a}, \tag{A.14}$$

where T_a is the temperature of particles a and $\zeta_a = 1$ (-1) if a is a fermion (boson).

For a thermal phase-space distribution, the integral over q^0 can be analytically performed to yield

$$\mathcal{C}_{\rm f} = \int \frac{2\pi}{2^4}\frac{d^4p}{(2\pi)^6}\frac{2}{|\vec{p}|}\Theta(p^2 - 4m_a^2) \times 4Fp^0\sigma_{aa\to bb,CM} \times \frac{4T_a}{e^{p^0/T_a} - \zeta_a^2}\ln\left(\frac{\exp\frac{p^0+|\vec{p}|\beta_a}{2T_a} + \zeta_a}{\exp\frac{p^0}{2T_a} + \zeta_a\exp\frac{|\vec{p}|\beta_a}{2T_a}}\right). \tag{A.15}$$

Rewriting the integration variable $p^\mu = (p^0, \vec{p})$ in terms of Mandelstam s and $y = |\vec{p}|/\sqrt{s}$, and using $F = \sqrt{s(s-4m_a^2)}/2$, we obtain

$$\mathcal{C}_{\rm f} = \frac{T_a}{32\pi^4}\int_{4\max(m_a^2,m_b^2)}^{\infty} ds s\sqrt{s(s-4m_a^2)}\sigma_{aa\to bb,CM} \times \left[2\int_0^\infty dy y\frac{1}{\left[\exp\left(\frac{\sqrt{y^2+1}}{T_a/\sqrt{s}}\right) - \zeta_a^2\right]}\ln\left(\frac{\exp\left(\frac{\sqrt{y^2+1}+\beta_a y}{2T_a/\sqrt{s}}\right) + \zeta_a}{\exp\left(\frac{\sqrt{y^2+1}}{2T_a/\sqrt{s}}\right) + \zeta_a\exp\left(\frac{\beta_a y}{2T_a/\sqrt{s}}\right)}\right)\right]. \tag{A.16}$$

In the limit the thermal distribution of particle a can be approximated as Maxwell-Boltzmann, i.e. $\zeta_a \to 0$, the integral in the square brackets simplifies to $\beta_a K_2(\sqrt{s}/T_a)$, where K_n is the modified Bessel function of

second kind. Correspondingly, the collision term becomes

$$\mathcal{C}_{\rm f}(T_a) = \frac{T_a}{32\pi^4}\int_{4\max(m_a^2,m_b^2)}^{\infty} ds s(s-4m_a^2)\sigma_{aa\to bb}K_2(\sqrt{s}/T_a), \tag{A.17}$$

which agrees with the result in Ref. [175].

In the case where $m_a < m_b$, the energy injection into b particles is mostly dominated by annihilations of a particles when a is relativistic. In the relativistic limit, we can approximate $\beta_a = 1$ making the integral in square brackets in eq. (A.16) only a function of $\sqrt{s}/T_a$. Defining

$$G_\zeta(x) = 2\int_0^\infty dt t \frac{1}{e^{x\sqrt{t^2+1}}-\zeta^2}\ln\left(\frac{e^{x(\sqrt{t^2+1}+t)/2}+\zeta}{e^{x\sqrt{t^2+1}/2}+\zeta e^{xt/2}}\right), \tag{A.18}$$

and approximating the term in square brackets in eq. (A.16) as $\beta_a G_{\zeta_a}(\sqrt{s}/T_a)$ we obtain

$$\mathcal{C}_{\rm f}(T_a) = \frac{T_a}{32\pi^4}\int_{4\max(m_a^2,m_b^2)}^{\infty} ds s(s-4m_a^2)\sigma_{aa\to bb}G_{\zeta_a}(\sqrt{s}/T_a). \tag{A.19}$$

The above collision term matches with eq. (A.16) in the limit $T_a \gg m_a$. We use this simplified form of collision term while calculating the energy injection from particles a into particles b in chapter 5. The error induced from using the simplified collision term in the calculation of energy injection is maximum when $m_a \gg m_b$. This maximum error is about 2% if particle a is a fermion and 8% if it is a boson. Even this error is typically inconsequential because energy injection into b is dominated by annihilations of particles lighter than b.

The function G_ζ can be computed analytically in the limit $x \gg 1$ and $x \ll 1$. In the large x limit, G_ζ asymptotes to K_2 as expected. In the small x limit, we find that

$$G_\zeta(x) \xrightarrow{x\ll 1} \begin{cases} \frac{\pi^2}{6x^2}\ln(2) & \zeta = 1, \\ \frac{2}{x^2} & \zeta = 0, \\ \frac{\pi^2}{3x^2}\ln(\frac{8\pi e A^{-12}}{x^2}) & \zeta = -1, \end{cases} \tag{A.20}$$

where A is the Glaisher-Kinkelin constant. At high temperatures, $T_a \gg m_a$ and m_b, most of the integral in eq. (A.19) is weighted at $\sqrt{s} \ll T_a$. Thus the collision term computed with $\zeta_a = 1$ (Fermi-Dirac statistics) is suppressed by a factor of ~ 2 compared to the collision term calculated using $\zeta_a = 0$ (Maxwell-Boltzmann statistics). While the collision term for $\zeta = -1$ (Bose-Einstein statistics) gets a non-trivial logarithmic enhancement compared to $\zeta = 0$ case.

The total collision term is well approximated by $\mathcal{C}_{\rm f}$ as long as particles b remain more dilute than particle a. However, close to thermalization, the phase-space distribution of particle b can no longer be neglected in the computation of the total collision term. In scenarios where we are only interested in computing the final energy density of particle b, we can approximate the total collision term using

$$\mathcal{C} = \mathcal{C}_{\rm f}(T_a) - \mathcal{C}_{\rm f}(T_b), \tag{A.21}$$

where T_b is the temperature of particle b if it is in internal thermal equilibrium. The above collision term is not accurate close to thermalization because we have ignored the contribution of final state effects while calculating $\mathcal{C}_{\rm f}$. However, the exact value of $\mathcal{C}$ is not important for the evaluation of energy density of particle

b once b and a have thermalized. Consequently, the approximation will deviate from the true answer only in the narrow parameter space where b is slightly away from thermalization with a.

One can include the contribution from final state effects if one makes some simplifying assumptions about the matrix element, as we do in the next section.

A.1.2 With final state effects

If we want to include the contribution from final state effects, then we find that the computation of the collision term is simplified if we shift to the following variables

$$\begin{aligned} p =& p_1 + p_2, \quad p' = p_3 + p_4, \\ q =& p_1 - p_2, \quad q' = p_3 - p_4. \end{aligned} \tag{A.22}$$

In terms of these variables, the Mandelstam variables are $s = p^2$, $t = (q - q')^2/4$ and $u = (q + q')^2/4$. In the CM frame, $p = (\sqrt{s}, 0, 0, 0)$ and consequently $U = \frac{1}{\sqrt{s}}(\sqrt{|\vec{p}|^2 + s}, 0, 0, |\vec{p}|)$, where $|\vec{p}|$ is the spatial component of p in the frame $U = (1, 0, 0, 0)$. Reference [226] shows that the 12-dimensional phase-space integral of eq. (A.2) can be reduced to

$$\begin{aligned} \mathcal{C}_E^f =& \frac{1}{256(2\pi)^8} \int 4\pi |\vec{p}|^2 d|\vec{p}| \int ds \frac{4rr'}{s} S \\ & \times \int \left[\int |\overline{\mathcal{M}}|^2 (1 \pm f_3(U \cdot p_3))(1 \pm f_4(U \cdot p_4)) d\theta' dy' \right] f_1(U \cdot p_1) f_2(U \cdot p_2) d\theta dy, \end{aligned} \tag{A.23}$$

where r is the magnitude of $\vec{q}$ and $y = \cos\phi, \theta$ give the direction of $\vec{q}$ with respect to $\vec{p}$, while r', y', and θ' denote the corresponding quantities for q'. The spatial and temporal magnitudes of q and q' are given by

$$r = \frac{1}{\sqrt{s}} \sqrt{(s - (m_1 + m_2)^2)(s - (m_1 - m_2)^2)}, \quad q^0 = \frac{m_1^2 - m_2^2}{\sqrt{s}}, \tag{A.24}$$

$$r' = \frac{1}{\sqrt{s}} \sqrt{(s - (m_3 + m_4)^2)(s - (m_3 - m_4)^2)}, \quad q'^0 = \frac{m_3^2 - m_4^2}{\sqrt{s}}. \tag{A.25}$$

For scenarios where the scattering amplitude is a function only of s, eq. (A.23) further simplifies to

$$\begin{aligned} \mathcal{C}_E^f =& \frac{S}{64(2\pi)^5} \int_0^\infty d|\vec{p}||\vec{p}|^2 \int_{s_0}^\infty ds |\overline{\mathcal{M}}(s)|^2 \frac{rr'}{s} \left[\int_{-1}^1 (1 \pm f_3(U \cdot p_3))(1 \pm f_4(U \cdot p_4)) dy' \right] \\ & \times \left[\int_{-1}^1 f_1(U \cdot p_1) f_2(U \cdot p_2) dy \right] \end{aligned} \tag{A.26}$$

where $s_0 = \max((m_1 + m_2)^2, (m_3 + m_4)^2)$. To evaluate these final integrals, we need to specify the distribution functions. We take particles 1 and 2 to be of species a at a temperature T_a and particles 3 and 4 of species b

at temperature T_b. Inserting the corresponding equilibrium distribution functions, we obtain

$$\mathcal{C}_E^f = \frac{T_a T_b S}{4(2\pi)^5}\int_0^\infty\int_{s_0}^\infty d|\vec{p}|\; ds\; |\overline{\mathcal{M}}(s)|^2 \frac{\exp\left(\frac{\sqrt{|\vec{p}|^2+s}}{T_b}\right)}{\left[\exp\left(\frac{\sqrt{|\vec{p}|^2+s}}{T_a}\right)-1\right]\left[\exp\left(\frac{\sqrt{|\vec{p}|^2+s}}{T_b}\right)-1\right]}$$
$$\times\log\left(\frac{\exp\left(\frac{\sqrt{|\vec{p}|^2+s}+\beta_a|\vec{p}|}{2T_a}\right)+\zeta_a}{\exp\left(\frac{\sqrt{|\vec{p}|^2+s}}{2T_a}\right)+\zeta_a\exp\left(\frac{\beta_a|\vec{p}|}{2T_a}\right)}\right)\log\left(\frac{\exp\left(\frac{\sqrt{|\vec{p}|^2+s}+\beta_b|\vec{p}|}{2T_b}\right)+\zeta_b}{\exp\left(\frac{\sqrt{|\vec{p}|^2+s}}{2T_b}\right)+\zeta_b\exp\left(\frac{\beta_b|\vec{p}|}{2T_b}\right)}\right). \tag{A.27}$$

Here $\zeta_{a,b}=\pm1$ depending on whether the respective particles are fermions/bosons and

$$\beta_{a,b}=\sqrt{1-\frac{4m_{a,b}^2}{s}}. \tag{A.28}$$

Next, we scale out the temperature of the hotter sector, T_a, by defining

$$\tilde{s}=\frac{s}{T_a^2},\qquad \tilde{p}=\frac{|\vec{p}|}{T_a},\qquad \tilde{m}_{a,b}=\frac{m_{a,b}}{T_a},\quad\text{and}\quad x=\frac{T_b}{T_a}\le 1. \tag{A.29}$$

This isolates the temperature dependence in the integral, which becomes

$$\mathcal{C}_E^f = \frac{S}{4(2\pi)^5}xT_a^5\int_0^\infty\int_{s_0/T_a^2}^\infty d\tilde{p}\; d\tilde{s}\; |\overline{\mathcal{M}}(\tilde{s})|^2 \frac{\exp(\frac{1}{x}\sqrt{\tilde{p}^2+\tilde{s}})}{\left[\exp(\frac{1}{x}\sqrt{\tilde{p}^2+\tilde{s}})-1\right]\left[\exp(\sqrt{\tilde{p}^2+\tilde{s}})-1\right]}$$
$$\times\log\left(\frac{\exp\left(\frac{1}{2}(\sqrt{\tilde{p}^2+\tilde{s}}+\beta_a\tilde{p})\right)+\zeta_a}{\exp\left(\frac{1}{2}\sqrt{\tilde{p}^2+\tilde{s}}\right)+\zeta_a\exp\left(\frac{1}{2}\beta_a\tilde{p}\right)}\right)\log\left(\frac{\exp\left(\frac{1}{2x}(\sqrt{\tilde{p}^2+\tilde{s}}+\beta_b\tilde{p})\right)+\zeta_b}{\exp\left(\frac{1}{2x}\sqrt{\tilde{p}^2+\tilde{s}}\right)+\zeta_b\exp\left(\frac{1}{2x}\beta_b\tilde{p}\right)}\right). \tag{A.30}$$

The temperature T_a enters the integrand only through $\mathcal{M}(T_a^2\tilde{s})$ and $\tilde{m}_{a,b}$.

The total collision term describing net energy transfer is

$$\mathcal{C}_E = S'xT_a^5\int_0^\infty\int_{s_0/T_a^2}^\infty d\tilde{p}\; d\tilde{s}\; |\overline{\mathcal{M}}(\tilde{s})|^2 \frac{\exp(\frac{1}{x}\sqrt{\tilde{p}^2+\tilde{s}})-\exp(\sqrt{\tilde{p}^2+\tilde{s}})}{\left[\exp(\frac{1}{x}\sqrt{\tilde{p}^2+\tilde{s}})-1\right]\left[\exp(\sqrt{\tilde{p}^2+\tilde{s}})-1\right]} \tag{A.31}$$
$$\times\log\left(\frac{\exp\left(\frac{1}{2}(\sqrt{\tilde{p}^2+\tilde{s}}+\beta_a\tilde{p})\right)+\zeta_a}{\exp\left(\frac{1}{2}\sqrt{\tilde{p}^2+\tilde{s}}\right)+\zeta_a\exp\left(\frac{1}{2}\beta_a\tilde{p}\right)}\right)\log\left(\frac{\exp\left(\frac{1}{2x}(\sqrt{\tilde{p}^2+\tilde{s}}+\beta_b\tilde{p})\right)+\zeta_b}{\exp\left(\frac{1}{2x}\sqrt{\tilde{p}^2+\tilde{s}}\right)+\zeta_b\exp\left(\frac{1}{2x}\beta_b\tilde{p}\right)}\right)$$
$$\equiv S'xT_a^5\int_0^\infty\int_{s_0/T_a^2}^\infty d\tilde{p}\; d\tilde{s}\; |\overline{\mathcal{M}}(\tilde{s})|^2 D(\tilde{s},\tilde{p},x,\tilde{m}_{a,b}), \tag{A.32}$$

where $S'=S/(4(2\pi)^5)$.

Even if we are interested in the regime where all external particles are relativistic, $\tilde{m}_{a,b}\ll 1$, retaining finite masses can be important for regulating collision terms with bosons because Bose-Einstein distribution diverges when $p\to 0$. This divergence for bosons ($\zeta=-1$) causes the integrand in the above equation to diverges as $\tilde{s},\tilde{p}\to 0$ when $\tilde{m}=0$. However, if $|\overline{\mathcal{M}}(\tilde{s})|^2$ is finite at $\tilde{s}\to 0$ then the divergence vanishes after

integration over $\tilde{s}, \tilde{p}$ and the collision term remains finite as $\tilde{m}_{a,b} \to 0$.[1] Thus, for our calculations we can freely work in the limit $\tilde{m}_{a,b} \to 0$.

A.2 Decays

In this section we simplify the collision term describing energy transfer via particle a decaying into particle b,

$$a \to b + b.$$

The collision term describing the forward energy transferred from a to b is given by

$$\mathcal{C}^f_{a\to b+b} = \int d\Pi d\Pi_1 d\Pi_2 \, (2\pi)^4 \delta^4(p - p_1 - p_2) \, S|\mathcal{M}_\Gamma|^2 f_a(p)(1 \pm f_b(p_1))(1 \pm f_b(p_2))E, \tag{A.33}$$

where f_a is the distribution function for a, $d\Pi_k = d^3p_k/[(2\pi)^3 2E_k]$, $|\mathcal{M}_\Gamma|^2$ is the spin-summed matrix element corresponding to the decay process, S is the symmetry factor, and variables with subscripts 1 and 2 correspond to the daughter particles while those with no subscripts correspond to a.

In the following calculation we neglect the final state effects (Bose-enhancement/Pauli-blocking) from particle b. As discussed in the previous subsection, this approximation is valid as long as density of particle b is much more dilute than the density of particle a.

By neglecting f_b, we can perform the phase space integration of the daughter particles in the rest frame of particle a by using the definition of the rest frame decay width,

$$\Gamma \equiv \frac{1}{2mg_a} \int d\Pi_1 d\Pi_2 \, (2\pi)^4 \delta^4(p - p_1 - p_2) \, S|\mathcal{M}_\Gamma|^2 = \frac{S|\mathcal{M}_\Gamma|^2}{4\pi} \tilde{\beta}_b \frac{1}{2mg_a}, \tag{A.34}$$

where m is the mass of particle a, g_a are the spin degrees of freedom of a, $\tilde{\beta}_b = \sqrt{1 - 4m^2/m_b^2}$ and m_b is the mass of particle b. Doing so simplifies the collision term to

$$\mathcal{C}^f_{a\to b+b} = m\Gamma g_a \int \frac{d^3p}{(2\pi)^3} f_a(p) = m n_a \Gamma. \tag{A.35}$$

Next, we simplify the collision term describing energy transferred by inverse decays of particle b into particle a. We start with,

$$\mathcal{C}^f_{b+b\to a} = \int d\Pi d\Pi_1 d\Pi_2 (2\pi)^4 \delta^4(p_1 + p_2 - p)E \; |\mathcal{M}_\Gamma|^2 S f_b(E_1) f_b(E_2). \tag{A.36}$$

Again we have neglected the final state effect from particle a by assuming particle a is much more dilute than particle b.

For simplicity, we perform the calculation in the rest frame of particle a. Considering $U = (1, 0, 0, 0)$ to denote the original isotropic frame, after changing frames such that, $(\sqrt{\vec{p}^2 + m^2}, \vec{p}) \to (m, \vec{0})$, we obtain

[1] In the collision term for rate of annihilations, i.e. when the collision term is not weighted with $E_1 + E_2$, the cancellation of the divergence depends on summing both forward and backward processes; the forward collision term alone retains a logarithmic dependence on $\tilde{m}$. In general one expects thermal self-energies to regulate this behavior when $T_{a,b} \gg m_{a,b}$; see also [66].

$U = \frac{1}{m}(\sqrt{m^2+\vec{p}^2}, -\vec{p})$. Consequently, the above collision term becomes

$$\mathcal{C}^f_{b+b\to a} = \int \frac{d^3p}{2(2\pi)^3}\left[\int d\Pi_1 d\Pi_2 (2\pi)^4\delta^3(p_1+p_2)\delta(2|\vec{p}_1| - m\tilde{\beta}_b)\ |\mathcal{M}_\Gamma|^2 S f_b(p_1\cdot U) f_b(p_2\cdot U)\right]. \tag{A.37}$$

Note that after shifting the frame, $\vec{p}$ no longer stands for momentum of a particle but instead is the label of boosted frame. Integrating over $\vec{p}_2$ and $|\vec{p}_1|$, yields

$$\mathcal{C}^f_{b+b\to a} = \frac{\tilde{\beta}_b^2}{8(2\pi)^5}\int d^3p \int d\Omega_{p_1} |\mathcal{M}_\Gamma|^2 S f_b(p_1\cdot U) f_b(p_2\cdot U). \tag{A.38}$$

Now note that the spin-summed matrix element of a decay process is isotropic as well as independent of the momentum of a. Consequently, we can pull $|\mathcal{M}_\Gamma|^2$ outside of the integral. Considering b particles to have a thermal distribution of form

$$f_b(p) = \frac{1}{e^{p/T_b}+\zeta_b}, \tag{A.39}$$

we can perform the angular integral over the distribution functions to yield,

$$\mathcal{C}^f_{b+b\to a} = T_b m^2 \tilde{\beta}_b \frac{|\mathcal{M}_\Gamma|^2 \hat{S}}{8\pi^3}\int_0^\infty dt \frac{t}{\exp(x\sqrt{t^2+1}) - \zeta_b^2}\log\left(\frac{\exp(\frac{x}{2}(\sqrt{1+t^2}+t\tilde{\beta}_b))+\zeta_b}{\exp(\frac{x\sqrt{1+t^2}}{2}) + \zeta_b\exp(\frac{tx\tilde{\beta}_b}{2})}\right), \tag{A.40}$$

where $x = m/T_b$ and $t = |\vec{p}|/m$. In the limit $m_b \ll m$, we can approximate $\tilde{\beta}_b = 1$ inside the integral, yielding

$$\mathcal{C}^f_{b+b\to a} \approx m\Gamma \times \left[m^2\frac{g_{Z'}}{2\pi^2}T_b G_{\zeta_b}(m/T_b)\right] \equiv m\Gamma \times \tilde{n}_{\zeta_b}(T_b), \tag{A.41}$$

where G is as defined in eq. (A.18). In the limit $\zeta_b = 0$, we have $G_{\zeta_b} = K_2$, where K_n is the modified Bessel function of the second kind. Consequently, $\tilde{n}_0(T_b)$ is the equilibrium number density of particles with Maxwell-Boltzmann distribution as expected.

A.3 Scattering

In this section we simplify the phase space integral for the collision term which describes the energy transfer from species a to b via t-channel process:

$$1(a) \ + \ 2(b) \ \to \ 3(a) \ + \ 4(b). \tag{A.42}$$

We consider both particles to be at different temperatures and additionally do not approximate their distribution as Maxwell-Boltzmann.

The relevant energy transfer collision term for the process given in eq. (A.42) is

$$\begin{aligned}\mathcal{C} = &\int\left[2\prod_i\left(\frac{d^4p_i}{(2\pi)^3}\delta(p_i^2-m_i^2)\Theta(p_i^0)\right)(2\pi)^4\delta^4(p_1+p_2-p_3+p_4)\right] S|\mathcal{M}|^2(p_1^0-p_3^0) \\ &\times [f_a(p_1)f_b(p_2)(1\pm f_a(p_3))(1\pm f_b(p_4))] \qquad \text{(A.43)} \\ \equiv &\int dP\ S|\mathcal{M}|^2(p_1^0-p_3^0)[f_a(p_1)f_b(p_2)(1\pm f_a(p_3))(1\pm f_b(p_4))], \qquad \text{(A.44)}\end{aligned}$$

where f_i are the distributions for the i^{th} particle, $\mathcal{M}$ is the amplitude of the process, a and b correspond to two species and $\hat{S}$ includes the symmetry factors from the process along with each particles degree of freedom. The first factor of two inside the integral is because the contribution from the backward scattering is the same as the forward process and we have summed over both backward and forward process above. The phase space element dP is given by the square bracket in the first line.

Because of the energy transfer term $(p_1^0 - p_3^0)$ in the integrand, we find that the computation of the collision term is simplified if we shift to the following variables

$$p = p_1 - p_3 \qquad p' = p_2 - p_4 \tag{A.45}$$

$$q = p_1 + p_3 \qquad q' = p_2 + p_4 \tag{A.46}$$

Correspondingly the Mandelstam variables are given by $s = (q+q')^2/4$, $t = p^2$ and $u = (q-q')^2/4$.

After performing the above shift in variables and integrating out p' using the momentum conserving dirac delta, we obtain

$$\begin{aligned} dP =& \frac{(2\pi)^4}{2^7}\frac{d^4p}{(2\pi)^4}\left[\frac{d^4q}{(2\pi)^4}\delta((p+q)^2/4 - m_a^2)\delta((q-p)^2/4 - m_a^2)\Theta(q^0 - |p^0|)\right] \\ & \times \left[\frac{d^4q'}{(2\pi)^4}\delta((-p+q')^2/4 - m_b^2)\delta((q'+p)^2/4 - m_b^2)\Theta(q'^0 - |p^0|)\right] \end{aligned} \tag{A.47}$$

$$dP \equiv \frac{d^4p}{2^7}\, dI_q\, dI_{q'}, \tag{A.48}$$

where dI_q and $dI_{q'}$ are given by the first and second square brackets, respectively. Note that dI_q and $dI_{q'}$ are identical integral elements except for the masses. Moreover, except for the argument in the Θ function, dI_q defined above has the same form as dI_q defined in eq. (A.7). Thus by performing the same steps as we did before to obtain eq. (A.10), we find

$$dI_q = \frac{2}{|\vec{p}|}\Theta(q^0 - |p^0|)\Theta\left(p^2\left[1 - \frac{(q^0)^2}{|\vec{p}|^2}\right] - 4m_a^2\right)d\theta_{12}\frac{dq^0}{(2\pi)^4}, \tag{A.49}$$

where θ_{12} is the Azimuthal angle made by $\vec{q}$ in the plane perpendicular to $\vec{p}$. The second Θ function in the last line imposes the requirement that $|\vec{q_{12}}| > 0$. The two Θ function together rule out the possibility of $p^2 > 0$, which is expected as $t = p^2 < 0$. Thus the arguments of the Θ functions can be rewritten as

$$dI_q = \frac{2}{|\vec{p}|}\Theta(-p^2)\Theta(q^0 - |\vec{p}|\beta_a)d\theta_{12}\frac{dq^0}{(2\pi)^4}, \tag{A.50}$$

where

$$\beta_a = \sqrt{1 - \frac{4m_a^2}{t}}. \tag{A.51}$$

The phase space element $dI_{q'}$ has the exact same form as dI_q with just m_a interchanged with m_b. Thus the phase space element dP simplifies to

$$dP = \frac{16\pi}{2^7(2\pi)^8}\Theta(-p^2)\left[\Theta(q^0 - |\vec{p}|\beta_a)d\theta_{12}dq^0\right]\left[\Theta(q'^0 - |\vec{p}|\beta_b)d\theta'_{12}dq'^0\right]d|\vec{p}|dp^0\frac{d\Omega_p}{4\pi}. \tag{A.52}$$

Replacing the above phase space element back in the collision term, and using the fact that the integrand is

independent of the orientation of $\vec{p}$ as well as the overall phase $\theta_{12}+\theta'_{12}$, we obtain

$$\begin{aligned}\mathcal{C} =&\frac{32\pi^2}{2^7(2\pi)^8}\int\left[\Theta(q^0-|\vec{p}|\beta_a)f_a(p_1)(1\pm f_a(p_3))dq^0\right]\left[\Theta(q'^0-|\vec{p}|\beta_b)f_b(p_2)(1\pm f_b(p_4))dq'^0\right]\\&\times\left(\int|\mathcal{M}|^2d\theta\right)p^0d|\vec{p}|dp^0\Theta(-p^2),\end{aligned}\tag{A.53}$$

where $\theta=\theta_{12}-\theta'_{12}$. Note that none of the square brackets depend on θ because the Boltzmann distributions are only functions of p^0 and q^0. Thus only the matrix element can have possible θ dependence.

Now the matrix element is a function of both $t=p^2=(p^0)^2-|\vec{p}|^2$ and $s=(q+q')^2/4$. So its dependence on q,q' forbids an independent integration of the square brackets in eq. (A.53). A t-channel matrix element can generically be written as

$$|\mathcal{M}|^2=\frac{\sum_{vw}c_{vw}s^vt^w}{(t-m_\phi^2)^2},\tag{A.54}$$

where the m_ϕ is the mediator mass. The Mandelstam t is simply equal to p^2 while s has a complicated dependence on q, q', and θ given by

$$s=\frac{1}{4}(q^2+q'^2+2q^0q'^0-2q_3q'_3-2q_{12}q'_{12}\cos\theta)\tag{A.55}$$

$$\begin{aligned}=&\frac{1}{4}\Bigg[4m_a^2+4m_b^2-2p^2-2q^0q'^0\frac{p^2}{|\vec{p}|^2}\\&-2\left(p^2\left(1-\frac{(q^0)^2}{|\vec{p}|^2}\right)-4m_a^2\right)^{0.5}\left(p^2\left(1-\frac{(q'^0)^2}{|\vec{p}|^2}\right)-4m_b^2\right)^{0.5}\cos\theta\Bigg].\end{aligned}\tag{A.56}$$

Note that after integration of matrix element over θ, all terms with odd powers of $\cos\theta$ vanish. Hence the integrated matrix element is simply given by a polynomial of form

$$\int|\mathcal{M}|^2d\theta=\frac{1}{(p^2-m_\phi^2)^2}\sum_{nm\lambda}c_{nm\lambda}(q^0)^n(q'^0)^m\frac{p^{2\lambda}}{|\vec{p}|^{n+m}}.\tag{A.57}$$

The values of of exponents above are restricted to $n,m,\lambda\in\{0,1,2\}$ because we require $v+w\leq 2$ in eq. (A.54) for the matrix element to be unitary. Furthermore, since $\int d\theta s^v$ only depends on even combinations of q^0 and q'^0, $n+m$ is always even.

Substituting the above matrix element in the collision term we obtain

$$\begin{aligned}\mathcal{C}=&\sum_{nm\lambda}\frac{32\pi^2c_{nm\lambda}}{2^7(2\pi)^8}\int\frac{p^0}{(p^2-m_\phi^2)^2}p^{2\lambda}\left[\int\Theta(q^0-|\vec{p}|\beta_a)f_a(p_1)(1\pm f_a(p_3))\frac{(q^0)^n}{|\vec{p}|^n}dq^0\right]\\&\times\left[\int\Theta(q'^0-|\vec{p}|\beta_b)f_b(p_2)(1\pm f_b(p_4))\frac{(q'^0)^m}{|\vec{p}|^m}dq'^0\right]d|\vec{p}|dp^0\Theta(-p^2)\end{aligned}\tag{A.58}$$

$$\equiv\sum_{nm\lambda}\frac{32\pi^2c_{nm\lambda}}{2^7(2\pi)^8}\int\frac{p^0}{(p^2-m_\phi^2)^2}p^{2\lambda}\times I_{n,\zeta_a}I'_{m,\zeta_b}d|\vec{p}|dp^0\Theta(-p^2),\tag{A.59}$$

where the terms I_{n,ζ_a} and I'_{n,ζ_b} are given by the first and second square brackets in the first equation above.

Considering a thermal distribution for particle a as given in eq. (A.14), the term I can be analytically

computed to yield

$$I_{n,\zeta_a} = \frac{2T_a}{e^{p^0/T_a}-1}\left(\frac{2T_a}{|\vec{p}|}\right)^n L_{n,\zeta_a}\left(\frac{|\vec{p}|\beta_a}{2T_a}, \frac{p^0}{2T_a}\right), \tag{A.60}$$

where

$$L_{n,\zeta}(a,b) = \sum_{r=0}^{n} \frac{n!}{(n-r)!} a^{n-r}[-\zeta \mathrm{Li}_{r+1}(-\zeta e^{-a+b}) + \zeta \mathrm{Li}_{r+1}(-\zeta e^{-a-b})], \tag{A.61}$$

and Li is the Polylogarithmic function. Recall that $\zeta = 1$ if particle a is fermion and $\zeta = -1$ if particle a is boson. In the Maxwell-Boltzmann limit, the square brackets in L simplify to e^{-a+b} for all r.

I'_m is same as I_n except with p^0 replaced by $-p^0$ and subscript a replaced by subscript b. Putting the simplified I_n and I'_m back in the collision term yields

$$\mathcal{C} = \frac{32\pi^2}{2^7(2\pi)^8} 4T_aT_b \sum_{nm\lambda} c_{nm\lambda} \int_{-\infty}^{\infty} dp^0 p^0 \int_{|p^0|}^{\infty} d|\vec{p}| \frac{p^{2\lambda}}{(p^2-m_\phi^2)^2} \left(\frac{2T_a}{|\vec{p}|}\right)^n \left(\frac{2T_b}{|\vec{p}|}\right)^m \times \frac{L_{n,\zeta_a}\left(\frac{|\vec{p}|\beta_a}{2T_a}, \frac{p^0}{2T_a}\right)}{e^{p^0/T_a}-1} \frac{L_{m,\zeta_b}\left(\frac{|\vec{p}|\beta_b}{2T_b}, -\frac{p^0}{2T_b}\right)}{e^{-p^0/T_b}-1}. \tag{A.62}$$

Note that $p^0 > 0$ indicates forward energy transfer from a to b, while $p^0 < 0$ indicates backward energy transfer. Consequently, the forward energy transfer collision term is given by the above integral, except with integral limits of p^0 changed to 0 to ∞.

Using the fact that $L_{m,\zeta}(a,-b) = -L_{m,\zeta}(a,b)$ we convert the integral over negative values of p^0 to positive values, yielding

$$\begin{aligned}\mathcal{C} =& \frac{32\pi^2}{2^7(2\pi)^8} 4T_aT_b \int_0^{\infty} dp^0 p^0 \left[\frac{1}{(e^{p^0/T_a}-1)(1-e^{-p^0/T_b})} - \frac{1}{(e^{p^0/T_b}-1)(1-e^{-p^0/T_a})}\right] \\ &\times \sum_{nm\lambda} c_{nm\lambda} \int_{|p^0|}^{\infty} d|\vec{p}| \frac{p^{2\lambda}}{(p^2-m_\phi^2)^2} \left(\frac{2T_a}{|\vec{p}|}\right)^n \left(\frac{2T_b}{|\vec{p}|}\right)^m L_{n,\zeta_a}\left(\frac{|\vec{p}|\beta_a}{2T_a}, \frac{p^0}{2T_a}\right) L_{m,\zeta_b}\left(\frac{|\vec{p}|\beta_b}{2T_b}, \frac{p^0}{2T_b}\right).\end{aligned} \tag{A.63}$$

Notice that the first square-bracket vanishes in the limit $T_a = T_b$ as expected.

In the limit both particles have Maxwell-Boltzmann distribution, the collision term simplifies to

$$\begin{aligned}\mathcal{C} =& \frac{32\pi^2}{2^7(2\pi)^8} 4T_aT_b \int_0^{\infty} dp^0 p^0 \left[e^{-p^0/2T_a} e^{p^0/2T_b} - e^{p^0/2T_a} e^{-p^0/2T_b}\right] \\ &\times \sum_{nm\lambda} c_{nm\lambda} \int_{p^0}^{\infty} d|\vec{p}| \frac{p^{2\lambda}}{(p^2-m_\phi^2)^2} \left(\frac{2T_a}{|\vec{p}|}\right)^n \left(\frac{2T_b}{|\vec{p}|}\right)^m \left(\sum_{r=0}^{n} \frac{n!}{(n-r)!} \left(\frac{|\vec{p}|\beta_a}{2T_a}\right)^{n-r}\right) \\ &\times \left(\sum_{r=0}^{m} \frac{m!}{(m-r)!} \left(\frac{|\vec{p}|\beta_b}{2T_b}\right)^{n-r}\right) e^{-\frac{\beta_a|\vec{p}|}{2T_a}} e^{-\frac{\beta_b|\vec{p}|}{2T_b}}.\end{aligned} \tag{A.64}$$

A.4 Collision terms for inflaton mediated annihilations

In this section we calculate the collision terms that are used in chapter 3 using the results from appendix A.1.2.

A.4.1 Trilinear scalar couplings

We consider two scalar species, χ_a and χ_b, interacting via

$$\mathcal{L}_{\rm int} = \frac{1}{2}\mu_a \phi \chi_a^2 + \frac{1}{2}\mu_b \phi \chi_b^2. \tag{A.65}$$

Here ϕ is a massive scalar (inflaton) mediator with mass M_ϕ. As both the coupled fields are scalars (and hence bosons), we take $\zeta_{a,b} = -1$ and $S' = 1/(16(2\pi)^5)$.

The scattering amplitude for the s-channel process in this theory, for $m_{a,b} \ll M_\phi$, is given by

$$|\overline{\mathcal{M}}(\tilde{s})|^2 = \frac{\mu_a^2 \mu_b^2}{(s - M_\phi^2)^2 + \Big(\Gamma_{0a} + \Gamma_{0b}\Big)^2}, \quad \Gamma_{0a,b} = \frac{\mu_{a,b}^2}{32\pi M_\phi}. \tag{A.66}$$

For $\mu_{a,b} \ll M_\phi$ we can approximate the scattering amplitude as [42]

$$|\overline{\mathcal{M}}(\tilde{s})|^2 \approx 32\pi^2 \frac{w\mu_a^2}{w+1}\frac{1}{T_a^2}\delta(\tilde{s} - \tilde{M}_\phi^2) + \Theta(\tilde{M}_\phi^2 - \tilde{s})\frac{\mu_a^4 w}{M_\phi^4}. \tag{A.67}$$

where $w = \Gamma_{0b}/\Gamma_{0a} = \mu_b^2/\mu_a^2$. To analytically estimate the behavior of $\mathcal{C}_E$, we combine the simplified form of the scattering amplitude given in eq. (A.67) along with approximations $\tilde{M}_\phi \ll 1$ and $\tilde{M}_\phi \gg 1$ at high and low temperatures respectively.

High temperature limit, $T_a \gg M_\phi$. In the high-temperature limit $\tilde{M}_\phi \to 0$, the contribution to the integral in eq. (A.32) from the Θ function term in eq. A.67 is dwarfed by the contribution from the Dirac delta term. Subsequently, in the high temperature limit we can to good approximation retain only the Dirac delta portion, giving

$$\mathcal{C}_{\text{high-T}} = S' x T_a^5 \int_0^\infty d\tilde{p}\; 32\pi^2 \frac{w\mu_a^2}{w+1}\frac{1}{T_a^2} D(\tilde{M}_\phi^2, \tilde{p}, x, 0). \tag{A.68}$$

To evaluate the above integral we separate it into two domains: $\tilde{p} < 0.1$ and $\tilde{p} > 0.1$. In the latter region we approximate $\tilde{p} \gg \tilde{M}_\phi$ to give

$$\begin{aligned}
\int_{0.1}^\infty d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0)\Big|_{\tilde{p}\gg\tilde{M}_\phi} &\approx \int_{0.1}^\infty d\tilde{p}\; \frac{\exp(\tilde{p}/x) - \exp(\tilde{p})}{[\exp(\tilde{p}) - 1][\exp(\tilde{p}/x) - 1]}\Big[\log^2\left(1/\tilde{M}_\phi^2\right) \\
&\quad + \log\Big(64\tilde{p}^2 \sinh(\tilde{p}/2)\sinh(\tilde{p}/(2x))x\Big)\log\left(1/\tilde{M}_\phi^2\right) \\
&\quad + \log\Big(8\tilde{p}\sinh(\tilde{p}/2)\Big)\log\Big(8\tilde{p}\sinh(\tilde{p}/(2x))x\Big)\Big] \\
&= \frac{1}{x}\Big(Y_1(x)\log^2\left(\frac{T_a}{M_\phi}\right) + Y_2(x)\log\left(\frac{T_a}{M_\phi}\right) + Y_3(x)\Big)
\end{aligned} \tag{A.69}$$

where

$$
\begin{aligned}
Y_1(x) &= 4x\int_{0.1}^{\infty} d\tilde{p}\frac{\exp(\tilde{p}/x)-\exp(\tilde{p})}{[\exp(\tilde{p})-1][\exp(\tilde{p}/x)-1]} = 9.01x + 4x^2\log(e^{0.1/x}-1) \\
&\xrightarrow{x<0.1} \approx 9.4x \qquad (A.70)\\
Y_2(x) &= 2x\int_{0.1}^{\infty} d\tilde{p}\frac{\exp(\tilde{p}/x)-\exp(\tilde{p})}{[\exp(\tilde{p})-1][\exp(\tilde{p}/x)-1]}\log\left(64\tilde{p}^2\sinh(\tilde{p}/2)\sinh(\tilde{p}/(2x))x\right) \\
&\xrightarrow{x<0.1} \approx 0.71x + 1.6 + 4.7x\log(x) \qquad (A.71)\\
Y_3(x) &= x\int_{0.1}^{\infty} d\tilde{p}\frac{\exp(\tilde{p}/x)-\exp(\tilde{p})}{[\exp(\tilde{p})-1][\exp(\tilde{p}/x)-1]}\log\left(8\tilde{p}\sinh(\tilde{p}/2)\right)\log\left(8\tilde{p}\sinh(\tilde{p}/(2x))x\right) \\
&\xrightarrow{x<0.1} \approx 3.2x - 0.82x\log(x) + 1.3. \qquad (A.72)
\end{aligned}
$$

To evaluate the integral in the region $\tilde{p} < 0.1$ we first consider the case where $T_a, T_b \gg M_\phi$, allowing us to approximate $\tilde{M}_\phi \ll x \ll 1$. Next, note that the integrand $D(\tilde{M}_\phi^2, \tilde{p}, x, 0, 0)$ is peaked near $\tilde{p} \sim \tilde{M}_\phi$. Near this peak we can use the approximation $\tilde{p} \ll x$. Assuming the contribution from the peak dominates the integral, we extend the approximation $\tilde{p} \ll x$ to the entire integration range $\tilde{p} \in (0, 0.1)$, yielding

$$
\begin{aligned}
\int_0^{0.1} d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0)\Big|_{\tilde{p},\tilde{M}_\phi \ll x} &\approx \int_0^{0.1} d\tilde{p}\frac{(1-x)}{\sqrt{\tilde{p}^2+\tilde{M}_\phi^2}}\log^2\left(\frac{\sqrt{\tilde{p}^2+\tilde{M}_\phi^2}+\tilde{p}}{\sqrt{\tilde{p}^2+\tilde{M}_\phi^2}-\tilde{p}}\right) \\
&\approx \frac{4}{3}(1-x)\log^3\left(0.2\frac{T_a}{M_\phi}\right). \qquad (A.73)
\end{aligned}
$$

In the case $T_b \ll M_\phi \ll T_a$ the assumptions we used above no longer hold. One can instead use the approximations $\tilde{p}, \tilde{M}_\phi \ll 1$ along with $e^{\tilde{M}_\phi/x} \gg 1$ to simplify the integral and show that its contribution is always dwarfed by the contribution from $\tilde{p} > 0.1$. For brevity we do not show the calculations here. Thus we can neglect contributions from $\tilde{p} < 0.1$ in eq. (A.68) when $T_b < M_\phi$[2]. We find empirically that using eq. (A.73) for all T_b helps improve the agreement between the analytic estimate and the full numerical calculation for T_a as low as $T_a \sim M_\phi$. Thus we approximate the full collision term at high temperatures as

$$
\begin{aligned}
\mathcal{C}_{\text{high-T}} \approx & S' 32\pi^2\frac{\mu_a^2 w}{w+1}T_a^3\left[\frac{4}{3}(1-x)x\log^3\left(0.2\frac{T_a}{M_\phi}\right) + Y_1(x)\log^2\left(\frac{T_a}{M_\phi}\right)\right. \\
& \left. + Y_2(x)\log\left(\frac{T_a}{M_\phi}\right) + Y_3(x)\right]. \qquad (A.74)
\end{aligned}
$$

From the asymptotic behavior at small x we see that $\mathcal{C}_{\text{high-T}}$ is largely insensitive to T_b. At extremely small x, the logarithmic term is dominant. However, at large temperatures $T \gtrsim 10^2 M_\phi$, as x increases to $x \sim 0.5$, the higher powers of the logarithm take over and enhance the collision term by roughly two orders of magnitude. This enhancement is due to the Bose enhancement of the forward energy transfer. As x further increases towards unity, the backward collision term starts catching up to forward collision term, eventually completely cancelling it at $x = 1$.

In the left panel of figure A.1, we compare our high temperature estimate with the numerically evaluated collision term. The Bose-Einstein enhancement over the classical Maxwell-Boltzmann result is clearly visible at high temperatures.

[2]In fact, even when $\tilde{M}_\phi \ll x$, the contributions from the $p < 0.1$ integral remains sub-dominant until extremely large temperatures, $T_a \geq 10^5 M_\phi$.

Intermediate temperatures, $T_a \leq M_\phi$. As $\tilde{M}_\phi$ begins to exceed unity, the Dirac delta contribution to the matrix element ensures that the integral of eq. (A.32) has support dominantly at $\tilde{s} = \tilde{M}_\phi > 1$. However, here the phase space distribution functions, contributing through the factor D, are exponentially suppressed. This Boltzmann suppression causes the collision term to fall sharply. In other words, in the intermediate temperature regime the integral receives its dominant contribution from an energy scale much larger than either temperature, which means that to excellent approximation the scattering here can be described using classical statistics.

Using classical statistics, the overall integral over $\tilde{p}$ (eq. (A.68)) can be performed exactly,

$$\mathcal{C}_{\rm MB} = S' 32\pi^2 M_\phi^2 \frac{\mu_a^2 w}{(w+1)} \frac{T_a}{4} \Big(K_2\Big(\frac{M_\phi}{T_a}\Big) - x K_2\Big(\frac{M_\phi}{x T_a}\Big)\Big), \tag{A.75}$$

where K_2 is the modified Bessel function of second kind. Again, we can see that at small x the collision term becomes insensitive to variations in the colder sector. As the temperatures fall further below the inflaton mass, the collision term becomes Boltzmann-suppressed.

Low temperature limit, $m_{a,b} \ll T \ll M_\phi$. In the low-temperature regime, the integral is dominated by off-shell inflaton scattering, described by the Heaviside term in eq. (A.67). Thus at low temperatures we need to evaluate

$$\mathcal{C}_{\text{low-T}} \approx S' \frac{x T_a^5}{M_\phi^4} \mu_a^4 w \int_0^\infty \int_0^{\tilde{M}_\phi^2} d\tilde{p}\; d\tilde{s} D(\tilde{s}, \tilde{p}, x, 0). \tag{A.76}$$

As D is exponentially suppressed at large values of $\tilde{s}$, we can take the upper limit of the $\tilde{s}$ integral to infinity with negligible errors. Both the integrand and the limits of integration thus become independent of temperature, giving

$$\begin{aligned} \mathcal{C}_{\text{low-T}} &\approx S' \frac{T_a^5}{M_\phi^4} \mu_a^4 w \Big[x \int_0^\infty \int_0^\infty d\tilde{p}\; d\tilde{s} D(\tilde{s}, \tilde{p}, x, 0) \Big], \\ &\equiv S' \frac{T_a^5}{M_\phi^4} \mu_a^4 w f(x) \xrightarrow{x<0.1} 7.9 S' \frac{T_a^5}{M_\phi^4} \mu_a^4 w. \end{aligned} \tag{A.77}$$

Again, as required, we find that the energy transfer function becomes insensitive to the colder sector as $x \to 0$.

Total collision term. To get a complete analytic estimate of $\mathcal{C}_E$ over all temperature ranges we combine the analytic estimates as

$$\begin{aligned} \mathcal{C}_E(T_a) = \mathcal{C}_{\text{low-T}} \Theta(M_\phi - T_a) + \mathcal{C}_{\text{high-T}} \Theta(T_a - M_\phi) + \mathcal{C}_{\rm MB} \Theta(0.2 M_\phi - T_a) \\ + \max(\mathcal{C}_{\rm MB}, \mathcal{C}_{\text{high-T}}) \Theta(M_\phi - T_a) \Theta(T_a - 0.2 M_\phi), \end{aligned} \tag{A.78}$$

where $\mathcal{C}_{\text{high-T}}, \mathcal{C}_{\rm MB}$ and $\mathcal{C}_{\text{low-T}}$ are given in eq. (A.74), (A.75) and (A.77). The Heaviside functions ensure that each function contributes only in its region of validity. This function describes the collision term for all temperature ranges as long as the scattering particles remain relativistic ($T_{a,b} \gg m_{a,b}$).

In the left panel of figure A.1, we compare the analytic approximations to the energy transfer collision term derived in high-, low-, and intermediate temperature regions with the exact numerical value. For illustrative purposes, each approximation is shown over a range larger than that taken in eq. (A.78). Together

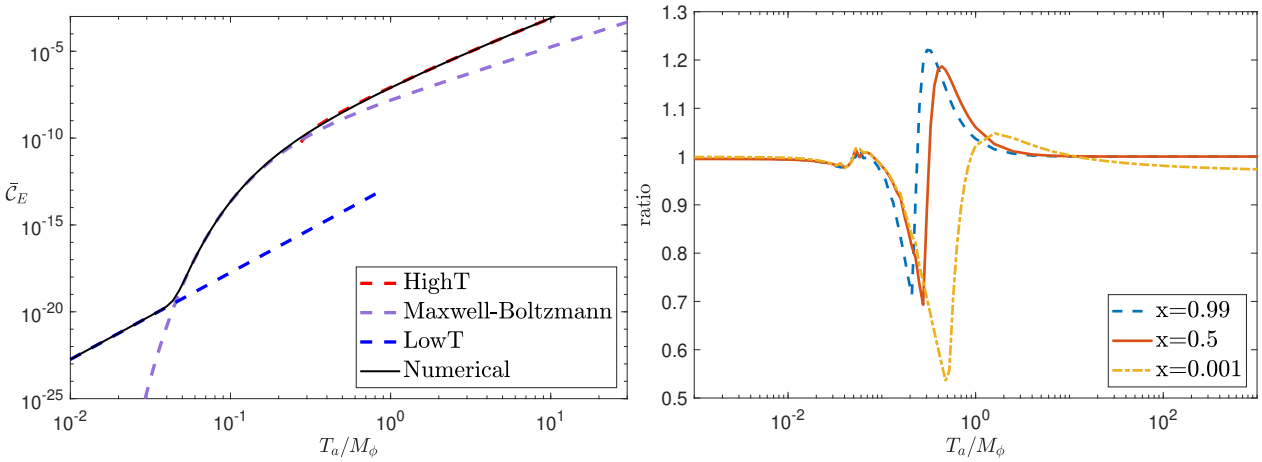


Figure A.1: *Left Panel:* Total energy transfer collision term $\bar{\mathcal{C}}_E = \mathcal{C}_E/M_\phi^5$ as a function of T_a/M_ϕ at fixed temperature ratio $x = T_b/T_a = 0.5$. The black solid line corresponds to $\mathcal{C}_E$ numerically evaluated using eq. (A.32) with matrix element given in eq. (A.66), red dashed line corresponds to $\mathcal{C}_{\text{high-T}}$ (eq. (A.74)), purple dashed line corresponds to $\mathcal{C}_{\text{MB}}$ (eq. (A.75)) and the blue dashed line corresponds to $\mathcal{C}_{\text{low-T}}$ (eq. (A.77)). *Right Panel:* The ratio of the analytic approximation to $\mathcal{C}_E$ (eq. (A.78)) with the full numerical value as a function of T_a/M_ϕ for $x = 0.99, 0.5, 0.001$. Results are shown for $\mu_a = 0.01M_\phi$, $\mu_b/\mu_a = 0.5$, and $m_{a,b} = 10^{-8}M_\phi$.

these approximations accuratelly describe the behavior of $\mathcal{C}$ in their respective regions. In the right panel of figure A.1, we compare the ratio of our analytic approximation, eq. (A.78), to the numerically evaluated collision term using the full matrix element of eq. (A.66). The largest deviation occurs during the transition from $\mathcal{C}_{\text{high-T}}$ to $\mathcal{C}_{\text{MB}}$ between $M_\phi/4 < T_a < M_\phi$ and is of the order $\sim 50\%$.

A.4.2 Yukawa couplings

We next consider two Dirac fermions, ψ_a and ψ_b, interacting with a scalar inflaton ϕ via

$$\mathcal{L}_{\text{int}} = y_a\phi\bar{\psi}_a\psi_a + y_b\phi\bar{\psi}_b\psi_b. \tag{A.79}$$

In this case we have $\zeta_{a,b} = 1$ and $S' = 1/(4(2\pi)^5)$.

The s-channel scattering amplitude in this theory, for $m_{a,b} \ll M_\phi$, is given by

$$|\overline{\mathcal{M}}(s)|^2 = 4y_a^2y_b^2\left(1-\frac{4m_a^2}{s}\right)\left(1-\frac{4m_b^2}{s}\right)\frac{s^2}{(s-M_\phi^2)^2+\left(\Gamma_{0a}+\Gamma_{0b}\right)^2}, \tag{A.80}$$

where

$$\Gamma_{0a,b} = \frac{y_{a,b}^2M_\phi}{8\pi}. \tag{A.81}$$

For small $y_{a,b}$ the scattering amplitude can be approximated as [42]

$$|\overline{\mathcal{M}}(\tilde{s})|^2 \approx 8\pi^2\frac{4y_a^2w}{w+1}\tilde{M}_\phi^2\delta(\tilde{s}-\tilde{M}_\phi^2) + \Theta(\tilde{M}_\phi^2-\tilde{s})\left(1-\frac{4\tilde{m}_a^2}{\tilde{s}}\right)\left(1-\frac{4\tilde{m}_b^2}{\tilde{s}}\right)4y_a^4w\frac{\tilde{s}^2}{\tilde{M}_\phi^4}$$
$$+\Theta(\tilde{s}-\tilde{M}_\phi^2)4y_a^4w. \tag{A.82}$$

where $w = \Gamma_{0,b}/\Gamma_{0,a} = y_b^2/y_a^2$.

To estimate $\mathcal{C}_E$ analytically, we combine the simplified form of the scattering amplitude given in eq. (A.82) along with the approximations $\tilde{M}_\phi \ll 1$ and $\tilde{M}_\phi \gg 1$ at high and low temperatures respectively. Moreover, since for fermions the contribution of the distribution functions to the integrand, D, is regular at $\tilde{s}, \tilde{p} = 0$, the limit $m_{a,b} = 0$ does not need any special attention.

High temperature limit, $T_a \gg M_\phi$. In the high temperature limit only the Dirac delta term and the second Heaviside theta term in eq. (A.82) contribute to the integral. The collision term then becomes

$$\mathcal{C}_{\text{high-T}} \approx S' x T_a^5 \left[8\pi^2 \frac{4y_a^2 w}{w+1} \tilde{M}_\phi^2 \int_0^\infty d\tilde{p}\ D(\tilde{M}_\phi^2, \tilde{p}, x, 0) + 4y_a^4 w \int_0^\infty d\tilde{p} \int_{\tilde{M}_\phi^2}^\infty d\tilde{s} D(\tilde{s}, \tilde{p}, x, 0) \right]. \tag{A.83}$$

Note that as $\tilde{s} \to 0$ the integrand $D(\tilde{s}, \tilde{p}, x, 0)$ asymptotes to a finite value over all $\tilde{p}$. Thus, we can safely approximate $\tilde{M}_\phi = 0$ in the integrand, making the integrals independent of T_a,

$$\mathcal{C}_{\text{high-T}} = S' \left[8\pi^2 M_\phi^2 \frac{4y_a^2 w}{w+1} V_1(x) T_a^3 + 4y_a^4 w V_2(x) T_a^5 \right]. \tag{A.84}$$

where,

$$V_1(x) = x \int_0^\infty d\tilde{p}\ D(0, \tilde{p}, x, 0) \xrightarrow{x<0.1} 0.29 \tag{A.85}$$

$$V_2(x) = x \int_0^\infty d\tilde{p} \int_0^\infty d\tilde{s} D(\tilde{s}, \tilde{p}, x, 0) \xrightarrow{x<0.1} 3.0 \tag{A.86}$$

We can check that at small x we see that $\mathcal{C}_{\text{high-T}}$ is in this limit insensitive to T_b, and at $x = 1$ all these functions go to zero as backward energy transfer exactly balances the forward energy transfer. The collision term at very high temperatures in this case is not sensitive to the inflaton mass.

Intermediate regime, $T_a \lesssim M_\phi$. For $T_a \sim M_\phi$ the Dirac delta part of the scattering amplitude will dominate the collision term. As discussed in section A.4.1 above for scalars, the collision term can be well approximated using Maxwell-Boltzmann statistics as the temperature drops below the inflaton mass scale, $T_a < M_\phi$. Thus, the collision term can be simply written as

$$\mathcal{C}_{\text{MB}} = S' 8\pi^2 M_\phi^4 \frac{4y_a^2 w}{w+1} \frac{T_a}{4} \left(K_2\left(\frac{M_\phi}{T_a}\right) - x K_2\left(\frac{M_\phi}{x T_a}\right) \right), \tag{A.87}$$

where K_2 is the modified Bessel function of the second kind.

Low temperature regime, $m_{a,b} \ll T \ll M_\phi$. In the low temperature regime, the integral is dominated by the $\Theta(\tilde{M}_\phi^2 - \tilde{s})$ term. Just as for scalars, we can to a good approximation replace $\tilde{M}_\phi \to \infty$ in the limit of integration. This yields

$$\begin{aligned} \mathcal{C}_{\text{low-T}} &\approx S' \frac{4T_a^9}{M_\phi^4} y_a^4 w \left[x \int_0^\infty \int_0^\infty d\tilde{p}\ d\tilde{s}\ \tilde{s}^2 D(\tilde{s}, \tilde{p}, x, 0) \right], \\ &\xrightarrow{x<0.1} 1.4 \times 10^3 S' y_a^4 w 4 \frac{T_a^9}{M_\phi^4}. \end{aligned} \tag{A.88}$$

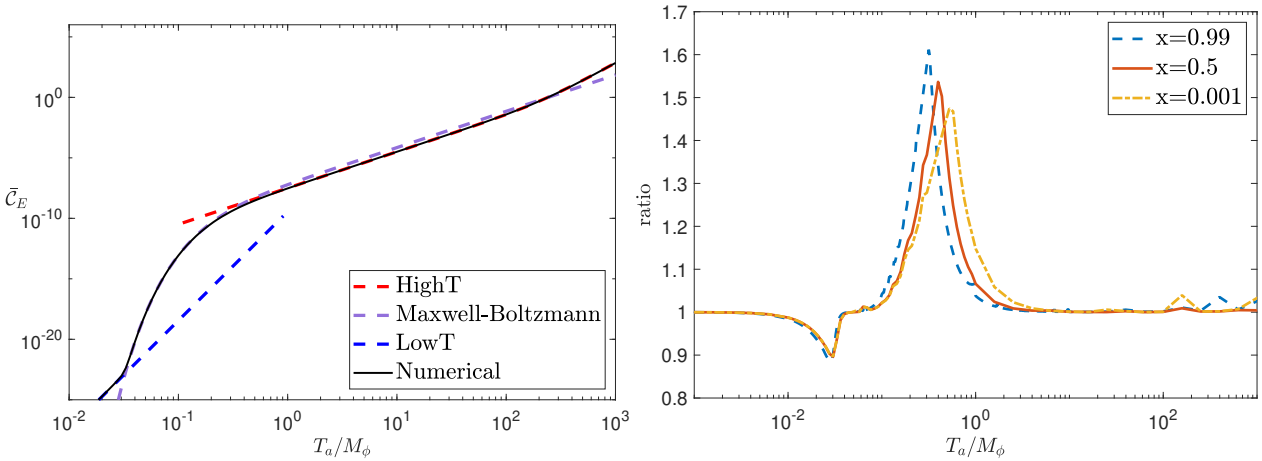


Figure A.2: *Left Panel:* Energy transfer collision term $\mathcal{C}_E/M_\phi^5$ as a function of T_a/M_ϕ for fixed $x = T_b/T_a = 0.5$. The black solid line corresponds to $\mathcal{C}_E$ numerically evaluated using eq. (A.32) and scattering amplitude as given in eq. (A.80), the red dashed line corresponds to $\mathcal{C}_{\text{high-T}}$ (eq. (A.84)), the purple dashed line corresponds to $\mathcal{C}_{\text{MB}}$ (eq. (A.87)) and the blue dashed line corresponds to $\mathcal{C}_{\text{low-T}}$ (eq. (A.88)). *Right Panel:* Ratio of the analytic approximation to the total collision term as given in eq. (A.89) to the numerical value as a function of T_a/M_ϕ for fixed temperature ratios $T_b/T_a = x = 0.99, 0.5, 0.001$. Results are shown for $y_a = 0.01$, $y_b/y_a = 0.5$, and $m_{a,b} = 10^{-8}M_\phi$.

Total collision term. We combine the analytic estimates derived above to approximate the collision term for all temperatures as

$$\begin{aligned}\mathcal{C}_E =& \mathcal{C}_{\text{low-T}}\Theta(M_\phi - T_a) + \mathcal{C}_{\text{MB}}\Theta(0.2M_\phi - T_a) + \min(\mathcal{C}_{\text{MB}}, \mathcal{C}_{\text{high-T}})\Theta(M_\phi - T_a)\Theta(T_a - 0.2M_\phi) \\ &+ \mathcal{C}_{\text{high-T}}\Theta(T_a - M_\phi).\end{aligned} \tag{A.89}$$

where $\mathcal{C}_{\text{high-T}}, \mathcal{C}_{\text{MB}}$, and $\mathcal{C}_{\text{low-T}}$ are as described in eq. (A.84),(A.87) and (A.88). Eq. (A.89) predicts the collision term for all temperature ranges as long as the particles remain relativistic ($T_{a,b} \gg m_{a,b}$).

In left panel of figure A.2, we compare our analytic estimate of the energy transfer collision term with the exact numerical value. Together these approximations accurately model the behavior of $\mathcal{C}_E$ in their respective regions. In the right panel of figure A.2, we compare the ratio of our analytic approximation, eq. (A.89), to the collision term numerically evaluated (eq. (A.32)) with the scattering amplitude in eq. (A.80). The largest deviation occurs during the transition from $\mathcal{C}_{\text{high-T}}$ to $\mathcal{C}_{\text{MB}}$ between $M_\phi/4 < T_a < M_\phi$ and is of the order $\sim 50\%$.

A.4.3 Axionic couplings to gauge bosons

Next, we consider two (Abelian) gauge bosons interacting with a (pseudoscalar) inflaton ϕ via

$$\mathcal{L}_{\text{int}} = -\frac{1}{4\Lambda_a}\phi F_a^{\mu\nu}\tilde{F}_{a,\mu\nu} - \frac{1}{4\Lambda_b}\phi F_b^{\mu\nu}\tilde{F}_{b,\mu\nu}. \tag{A.90}$$

In this case we have $\zeta_{a,b} = -1$ and $S' = 1/(16(2\pi)^5)$.

The s-channel scattering amplitude in this theory, for $m_{a,b} \ll M_\phi$, is given by

$$|\overline{\mathcal{M}}(s)|^2 = \frac{4}{128\Lambda_a^2\Lambda_b^2} \frac{s^4}{(s-M_\phi^2)^2 + \left(\Gamma_{0a}+\Gamma_{0b}\right)^2}, \tag{A.91}$$

where

$$\Gamma_{0a,b} = \frac{M_\phi^3}{256\pi\Lambda_{a,b}^2}. \tag{A.92}$$

For $\Lambda_{a,b} \gg 1$, we approximate the scattering amplitude as [42]

$$|\overline{\mathcal{M}}(\tilde{s})|^2 \approx \frac{2\pi^2}{\Lambda_a^2}\frac{4w}{w+1}\frac{M_\phi^4}{T_a^2}\delta(\tilde{s}-\tilde{M}_\phi^2) + \frac{4wT_a^8}{128\Lambda_a^4 M_\phi^4}\Theta(\tilde{M}_\phi^2-\tilde{s})\tilde{s}^4 + \frac{4wT_a^4}{128\Lambda_a^4}\Theta(\tilde{s}-\tilde{M}_\phi^2)\tilde{s}^2, \tag{A.93}$$

where $w = \Gamma_{0b}/\Gamma_{0a} = \Lambda_a^2/\Lambda_b^2$.

To estimate $\mathcal{C}_E$ analytically, we combine the simplified scattering amplitude given in eq. (A.93) along with high and low-temperature approximations in the limits $\tilde{M}_\phi \ll 1$ and $\tilde{M}_\phi \gg 1$ respectively.

High temperature limit, $T_a \gg M_\phi$. At high temperatures both the Dirac delta contribution and the Heaviside theta term $\propto \tilde{s}^2$ contribute importantly to the integral. Subsequently we approximate the scattering amplitude as

$$\mathcal{C}_{\text{high-T}} = S' x T_a^5 \left[\frac{2\pi^2}{\Lambda_a^2}\frac{4w}{w+1}\frac{M_\phi^4}{T_a^2}\int_0^\infty d\tilde{p}\; D(\tilde{M}_\phi^2,\tilde{p},x,0) + \frac{4wT_a^4}{128\Lambda_a^4}\int_0^\infty d\tilde{p}\int_{\tilde{M}_\phi^2}^\infty d\tilde{s}\;\tilde{s}^2 D(\tilde{s},\tilde{p},x,0)\right]. \tag{A.94}$$

In the above equation we have already assumed $\tilde{m}_{a,b} = 0$. The first integral on the RHS is exactly the same as that evaluated for scalars at high temperatures. In the second integral, the integrand vanishes as $s \to 0$, allowing us to freely take $\tilde{M}_\phi \approx 0$. This yields

$$\begin{aligned}\mathcal{C}_{\text{high-T}} = S'\Bigg[&\frac{2\pi^2 M_\phi^4}{\Lambda_a^2}\frac{4w}{w+1}T_a^3\Bigg(\frac{4}{3}(1-x)x\log^3\left(\frac{T_a}{M_\phi}\right) + Y_1(x)\log^2\left(\frac{T_a}{M_\phi}\right) + Y_2(x)\log\left(\frac{T_a}{M_\phi}\right) \\ &+ Y_3(x)\Bigg) + \frac{4w}{\Lambda_a^4}Z(x)T_a^9\Bigg],\end{aligned} \tag{A.95}$$

where

$$Z(x) = \frac{x}{128}\int_0^\infty d\tilde{p}\int_0^\infty d\tilde{s}\;\tilde{s}^2 D(\tilde{s},\tilde{p},x,0) \xrightarrow{x<0.1} \approx 14. \tag{A.96}$$

Here the Y_i are defined in eq. (A.70), (A.71) and (A.72).

In the left panel of figure A.3 we compare this high temperature estimate with the numerically evaluated collision term.

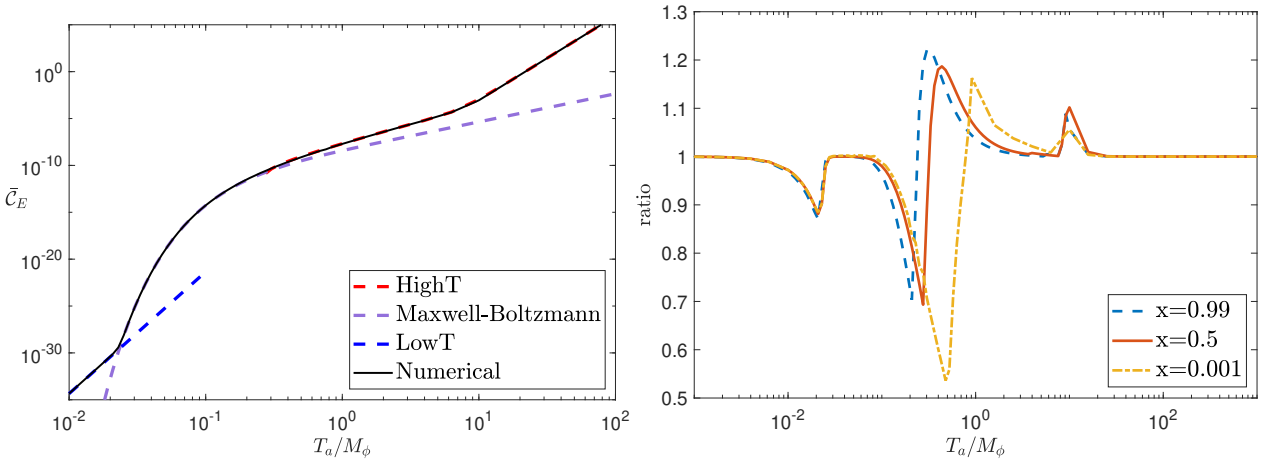


Figure A.3: *Left Panel:* Total energy transfer collision term $\mathcal{C}_E/M_\phi^5$ as a function of T_a/M_ϕ at fixed $x = T_b/T_a = 0.5$. The black solid line corresponds to $\mathcal{C}_E$ numerically evaluated using eq. (A.32) with scattering amplitude given by eq. (A.91), the red dashed line corresponds to $\mathcal{C}_{\text{high-T}}$ (eq. (A.95)), the purple dashed line corresponds to $\mathcal{C}_{\text{MB}}$ (eq. (A.97)) and the blue dashed line corresponds to $\mathcal{C}_{\text{low-T}}$ (eq. (A.98)). *Right Panel:* Ratio of our analytic estimate as given in eq. (A.99) to the numerically evaluated collision term as a function of T_a/M_ϕ for fixed temperature ratios $T_b/T_a = x = 0.99, 0.5, 0.001$. Here we fix $\Lambda_a = 100 M_\phi$, $\frac{1/\Lambda_b}{1/\Lambda_a} = 0.5$ and $m_{a,b} = 10^{-8} M_\phi$.

Intermediate temperatures, $T_a \leq M_\phi$. In the intermediate regime near $T_a \sim M_\phi$, integral can again be well approximated with Maxwell-Boltzmann distribution. Thus the collision term can be simply written as

$$\mathcal{C}_{\text{MB}} = S' \frac{2\pi^2 M_\phi^6}{\Lambda_a^2} \frac{4w}{w+1} \frac{T_a}{4} \Big(K_2\Big(\frac{M_\phi}{T_a}\Big) - x K_2\Big(\frac{M_\phi}{x T_a}\Big)\Big), \tag{A.97}$$

where K_2 is the modified Bessel function of the second kind.

Low temperature limit $m_{a,b} \ll T \ll M_\phi$ In this regime, the $\Theta(\tilde{M}_\phi^2 - \tilde{s})$ term dominates in eq. (A.32). We can again take $\tilde{M}_\phi \to \infty$ in the limit of integration. This yields

$$\begin{aligned} \mathcal{C}_{\text{low-T}} &\approx S' T_a^{13} \frac{4w}{M_\phi^4 \Lambda_a^4} \left[\frac{x}{128} \int_0^\infty \int_0^\infty d\tilde{p}\; d\tilde{s}\; \tilde{s}^4 D(\tilde{s}, \tilde{p}, x, 0) \right], \\ &\xrightarrow{x<0.1} 7.1 \times 10^4 S' \frac{4w}{M_\phi^4 \Lambda_a^4} T_a^{13}. \end{aligned} \tag{A.98}$$

Total collision term. We combine the analytic estimates in the following manner,

$$\mathcal{C}_E = \mathcal{C}_{\text{low-T}} \Theta(M_\phi - T_a) + \max(\mathcal{C}_{\text{MB}}, \mathcal{C}_{\text{high-T}}) \Theta(M_\phi - T_a) + \mathcal{C}_{\text{high-T}} \Theta(T_a - M_\phi). \tag{A.99}$$

where $\mathcal{C}_{\text{high-T}}, \mathcal{C}_{\text{MB}}$, and $\mathcal{C}_{\text{low-T}}$ are as described in eq. (A.95),(A.97) and (A.98). This function describes the collision term for all temperature ranges as long as the scattering particles remain relativistic ($T_{a,b} \gg m_{a,b}$).

In the left panel of figure A.3, we compare our analytic estimate derived in the three regions with the numerical value. In the right panel of figure A.3, we compare the ratio of our analytic estimate, eq. (A.99), to the numerically evaluated collision term (eq. (A.32)) with scattering amplitude given in eq. (A.91). The largest deviation occurs during transition from $\mathcal{C}_{\text{high-T}}$ to $\mathcal{C}_{\text{MB}}$ between $M_\phi/4 < T_a < M_\phi$ and is of the order

$\sim 50\%$.

A.4.4 Mixed Yukawa and scalar trilinear couplings

In this case we consider a Dirac fermion ψ and a scalar field χ that interact with the inflaton ϕ via

$$\mathcal{L}_{\rm int} = \frac{1}{2}\mu_a \phi \chi_a \chi_a + y_b \phi \bar{\psi}_b \psi_b. \tag{A.100}$$

Note that sector a is not necessarily hotter in this scenario. For this theory $\zeta_a = -1$, $\zeta_b = +1$, and $S' = 1/(8(2\pi)^5)$.

The spin-summed s-channel scattering amplitude in this theory, for $m_{a,b} \ll M_\phi$, is given by

$$|\overline{\mathcal{M}}(s)|^2 = 2\mu_a^2 y_a^2 \left(1 - \frac{4m_b^2}{s}\right) \frac{s}{(s - M_\phi^2)^2 + (\Gamma_{0a} + \Gamma_{0b})^2}, \tag{A.101}$$

where

$$\Gamma_{0a} = \frac{\mu_a^2}{32\pi M_\phi}, \qquad \Gamma_{0b} = \frac{y_b^2 M_\phi}{8\pi}. \tag{A.102}$$

For $y, \mu \ll 1$ we approximate the scattering amplitude as [42]

$$|\overline{\mathcal{M}}(\tilde{s})|^2 \approx 16\pi^2 \mu_a^2 \frac{w}{w+1} \frac{1}{T_a^2} \delta(\tilde{s} - \tilde{M}_\phi^2) + \frac{1}{2}\left(1 - \frac{4m_b^2}{s}\right) \Theta(\tilde{M}_\phi^2 - \tilde{s}) \frac{\mu_a^4 w}{M_\phi^6} T_a^2 \tilde{s}, \tag{A.103}$$

where $w = \Gamma_{0b}/\Gamma_{0a} = 4y_b^2 M_\phi^2/\mu_a^2$.

To estimate the behavior of $\mathcal{C}_E$ analytically, we combine the simplified form of the scattering amplitude given in eq. (A.103) along with the high- and low-temperature approximations $\tilde{M}_\phi \ll 1$ and $\tilde{M}_\phi \gg 1$.

High temperature limit, $T \gg M_\phi$. When the temperature of the hotter sector, $T = \max(T_a, T_b)$, is larger than the inflaton mass, the Dirac-delta term of the scattering amplitude dominates in the integral in eq. (A.32), giving

$$\mathcal{C}_{\text{high-T}} = S' x T_a^5 \left[16\pi^2 \mu_a^2 \frac{w}{w+1} \frac{1}{T_a^2} \int_0^\infty d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0)\right]. \tag{A.104}$$

As the scalars χ_a follow BE distribution, $D(\tilde{M}_\phi^2, \tilde{p}, x, 0)$ has a pole as $\tilde{M}_\phi^2 \to 0$. We cannot simply approximate $\tilde{M}_\phi = 0$, and we proceed analogously to the case of section A.4.1 and split the integral as,

$$\int_0^\infty d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0) = \int_0^{\tilde{M}_\phi} d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0)\Big|_{\tilde{p}, \tilde{M}_\phi \ll 1} + \int_{\tilde{M}_\phi}^\infty d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0)\Big|_{\tilde{p} \gg \tilde{M}_\phi}. \tag{A.105}$$

The integrand in the first term on RHS vanishes for small $\tilde{p}$. Hence, for $\tilde{M}_\phi \ll 1$ the first integral can be ignored. Expanding in $\tilde{p} \gg \tilde{M}_\phi$ limit, the second integral yields

$$\int_0^\infty d\tilde{p}\; D(\tilde{M}_\phi^2, \tilde{p}, x, 0) \approx \int_{\tilde{M}_\phi}^\infty d\tilde{p}\; \frac{\exp(\tilde{p}/x) - \exp(\tilde{p})}{[\exp(\tilde{p}) - 1][\exp(\tilde{p}/x) - 1]} \Big[\log(8\tilde{p}\sinh(\tilde{p}/2)) \log(\cosh(\tilde{p}/2/x)) + 2\log\Big(\frac{1}{\tilde{M}_\phi}\Big) \log(\cosh(\tilde{p}/2/x)) \Big]. \tag{A.106}$$

As the integrand above diverges only logarithmically as $\tilde{p} \to 0$, the integral is insensitive to its lower limit, which can be replaced with 0 with negligible error. The collision term at high temperatures can then be simply written as

$$\mathcal{C}_{\text{high-T}} = S' 16\pi^2 \frac{\mu_a^2 w}{w+1} T_a^3 \Big(W_1(x) \log\Big(\frac{T_a}{M_\phi}\Big) + W_2(x) \Big), \tag{A.107}$$

where

$$W_1(x) = 2x \int_0^\infty d\tilde{p} \frac{\exp(\tilde{p}/x) - \exp(\tilde{p})}{[\exp(\tilde{p}) - 1][\exp(\tilde{p}/x) - 1]} \log(\cosh(\tilde{p}/2/x)) \longrightarrow \begin{cases} 1.6 & x < 0.1 \\ -0.48x^2 & x > 10 \end{cases}$$

$$W_2(x) = x \int_0^\infty d\tilde{p} \frac{\exp(\tilde{p}/x) - \exp(\tilde{p})}{[\exp(\tilde{p}) - 1][\exp(\tilde{p}/x) - 1]} \log(8\tilde{p}\sinh(\tilde{p}/2)) \log(\cosh(\tilde{p}/2/x)) \longrightarrow \begin{cases} 1.1 & x < 0.1 \\ -0.30x^3 & x > 10 \end{cases} \tag{A.108}$$

Note that the high temperature collision term is approximately insensitive to the colder sector in this case. The collision term's logarithmic sensitivity on the inflaton mass depends on whether or not the scalars are hotter than the fermions. However, in both the cases the collision term is IR-sensitive due to the dependence on inflaton mass.

In left panels of figure A.4 we compare our high temperature estimate with the numerically evaluated collision term.

Intermediate temperatures, $T \leq M_\phi$. For temperatures near $T \sim M_\phi$ the Dirac delta part of the scattering amplitude dominates the behavior of collision term. As discussed above, in this region, the distribution functions are well approximated by Maxwell-Boltzmann distributions as temperature drops below the inflaton mass scale, $T < M_\phi$. The collision term can therefore be simply computed as,

$$\mathcal{C}_{\text{MB}} = S' 16\pi^2 \frac{\mu_a^2 w}{(w+1)} \frac{M_\phi^2}{4} \Big(T_a K_2\Big(\frac{M_\phi}{T_a}\Big) - T_b K_2\Big(\frac{M_\phi}{T_b}\Big) \Big), \tag{A.109}$$

where K_2 is the modified Bessel function of second kind. Again, we can see that at small/large x the collision term becomes insensitive to variations in the colder sector.

Low temperature limit, $m_{\chi,\psi} \ll T \ll M_\phi$. In the low-temperature regime, the contribution from the Dirac delta part of the matrix element falls below the one from $\Theta(\tilde{M}_\phi^2 - \tilde{s})$ term in the integral in eq. (A.32).

Just like in scalar case, to a good approximation we can replace $\tilde{M}_\phi \to \infty$ in the integral limit. This yields

$$\mathcal{C}_{\text{low-T}} \approx S' \frac{T_a^7}{M_\phi^6} \mu_a^4 w \left[\frac{x}{2} \int_0^\infty \int_0^\infty d\tilde{p}\; d\tilde{s}\; \tilde{s} D(\tilde{s}, \tilde{p}, x, 0) \right] \longrightarrow \begin{cases} 31\, S' \frac{\mu_a^4 w}{M_\phi^6} T_a^7 & x < 0.1 \\ -21\, S' \frac{\mu_a^4 w}{M_\phi^6} T_b^7 & x > 10. \end{cases} \tag{A.110}$$

Similar to our high temperature estimate, we find that the energy transfer function becomes insensitive to the colder sector as $x \to 0, \infty$.

Total collision term. To get an analytic estimate of $\mathcal{C}_E$ over all temperature ranges we combine the analytic estimates as

$$\begin{aligned} \mathcal{C}_E =& \mathcal{C}_{\text{low-T}} \Theta(M_\phi - T) + \mathcal{C}_{\text{MB}} \Theta(0.2 M_\phi - T) + \max(\mathcal{C}_{\text{MB}}, \mathcal{C}_{\text{high-T}}) \Theta(M_\phi - T) \Theta(T - 0.2 M_\phi) \\ &+ \mathcal{C}_{\text{high-T}} \Theta(T - M_\phi), \end{aligned} \tag{A.111}$$

where $T = \max(T_a, T_b)$ and $\mathcal{C}_{\text{high-T}}, \mathcal{C}_{\text{MB}}$ and $\mathcal{C}_{\text{low-T}}$ are as described in eq. (A.107), (A.109) and (A.110).

In the left panels of figure A.4, we compare our analytic estimate with the exact numerical value. In the right panels of figure A.4, we compare the ratio of our analytic fit, eq. (A.111), with the numerical evaluation of the full expression (eq. (A.32)) with scattering amplitude given in eq. (A.103). The largest deviation occurs during the transition from $\mathcal{C}_{\text{high-T}}$ to $\mathcal{C}_{\text{MB}}$ between $M_\phi/4 < T < M_\phi$ and is of the order $\sim 50\%$.

A.5 Collision terms for millicharge particle model

The Standard Model particles can inject energy into the hidden sector bath formed by millicharged particle (MCP) and dark photon via annihilations, decays and scattering. The calculation of the collision term for these processes involves phase space integration of the matrix element along with appropriate particle distribution function. In appendix A, we show the how the multi-dimensional phase space integral can be simplified for energy injection via generic annihilation or decay processes after considering a quantum statistical thermal distribution of particles. Similarly in appendix A.3 we simplify the phase space integral for scattering processes. In this section, we describe the particle physics processes that determine the matrix element in the collision term for MCP model. We highlight the simplifications we employed in calculation of these processes through electroweak and QCD phase transitions.

In what follows we first discuss energy injection via SM fermion annihilations into MCPs in section A.5.1. In section A.5.2 and section A.5.3, we describe energy injection into HS due to Z-boson and plasmon decays into MCPs, respectively. Finally, in section A.5.4 we describe energy transfer via Coulomb scattering between SM fermions and MCPs. The Coulomb scattering processes has a forward singularity which is mitigated by plasmon mass.

There are also additional processes involved in energy transfer into the HS bath, such as SM boson annihilations into MCPs, Compton scattering of MCPs with photons and dark photons, and photon dark photon fusion into MCPs. We have verified that energy transfer through SM boson annihilations and fusion processes are around 2 orders of magnitude weaker than the processes we mentioned in previous paragraphs. The Compton scattering is different because it depends on both the dark coupling constant, e', and the millicharge, Q, while all other processes only depend on Q. We have checked that the collision term for Coulomb scattering is always larger than the one for Compton scattering when $e' < 0.9$. Consequently in

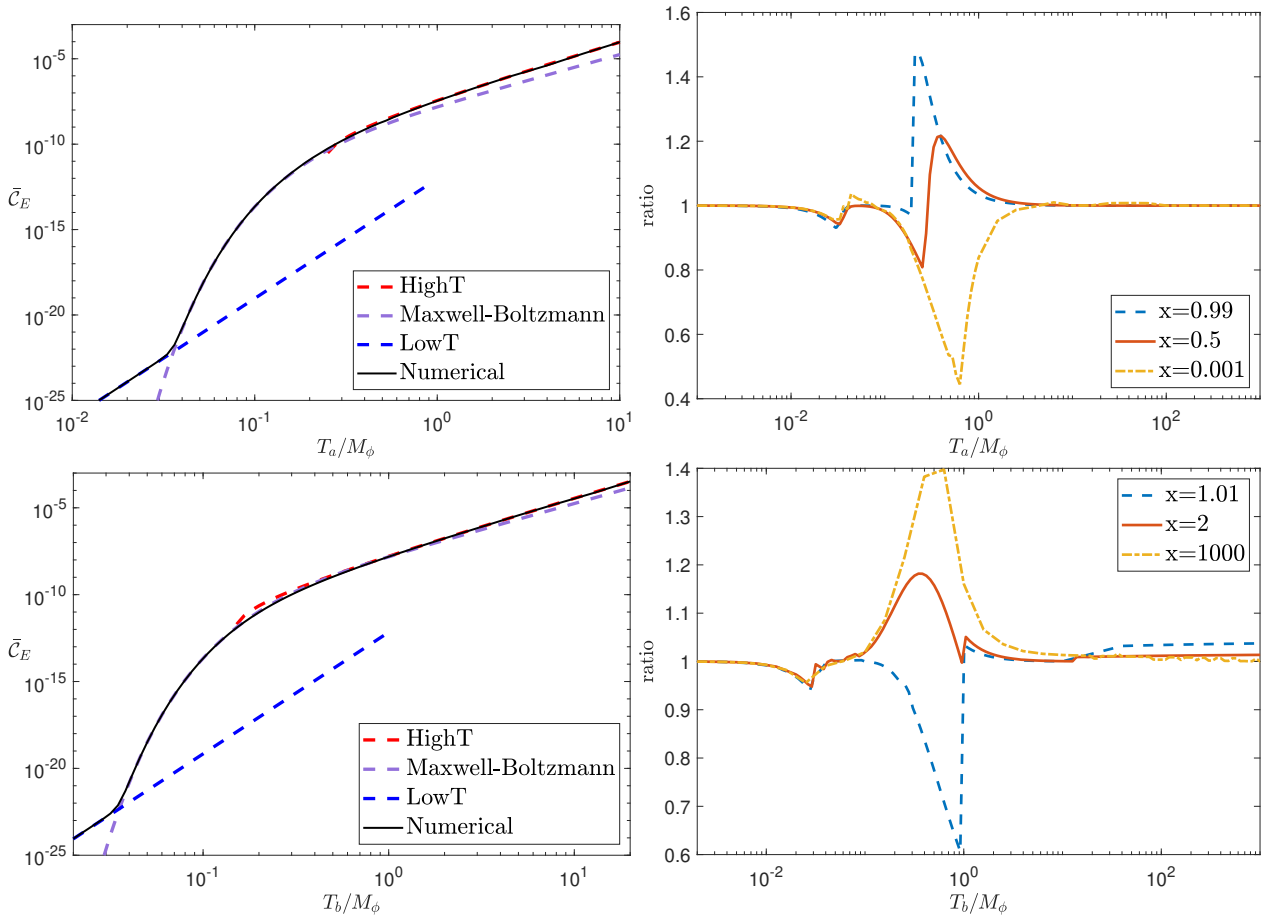


Figure A.4: *Left Panel:* Magnitude of total energy transfer collision term $\bar{\mathcal{C}}_E = \text{abs}(\mathcal{C}_E)/M_\phi^5$ as a function of $T = \max(T_a, T_b)$ at fixed temperature ratio $x = T_b/T_a = 0.5$ (top left) and $x = 2$ (bottom left). The collision term is always positive for the top left panel and always negative for the bottom left. The black solid line corresponds to $\mathcal{C}$ numerically evaluated using eq. (A.32), the red dashed line corresponds to $\mathcal{C}_{\text{high-T}}$ (eq. (A.107)), the purple dashed line corresponds to $\mathcal{C}_{\text{MB}}$ (eq. (A.109)) and the blue dashed line corresponds to $\mathcal{C}_{\text{low-T}}$ (eq. (A.110)). *Right Panel:* Ratio of the analytic approximation to $\mathcal{C}_E$ (eq. (A.111)) to the full numerical value as a function of $T = \max(T_a, T_b)$ for fixed temperature ratios $T_b/T_a = x = 0.99, 0.5, 0.001$ (top right) and $x = 1.01, 1, 1000$ (bottom right). Results are shown for $\mu_a = 0.01 M_\phi$, $2y_b M_\phi/\mu_a = 0.5$, and $m_{a,b} = 10^{-8} M_\phi$.

our study we neglect the contribution of all the processes mentioned in this paragraph, and we omit their detailed calculation here for brevity.

A.5.1 Fermion annihilations

The center-of-momentum (CM) frame spin-summed cross-section for SM fermion annihilation into MCPs is given by

$$\begin{aligned}\sigma_{ff\to\psi\psi} =& \frac{4\pi Q^2 N_c(f)\alpha^2}{s^3}\frac{\sqrt{s-4m^2}}{\sqrt{s-4m_f^2}} \\ &\times\left\{\frac{4}{3}(2m^2+s)(2m_f^2+s)\left[Q_f^2-\frac{Q_f C_V}{\cos^2\theta_W}\frac{s(s-m_Z^2)}{(s-m_Z^2)^2+m_Z^2\Gamma_Z^2}\right]\right. \\ &\left.+\frac{1}{4\cos^4\theta_W}\frac{s^2(s+2m^2)}{(s-m_Z^2)^2+m_Z^2\Gamma_Z^2}\left[\frac{4}{3}(C_V^2+C_A^2)(s-m_f^2)+4(C_V^2-C_A^2)m_f^2\right]\right\},\end{aligned} \tag{A.112}$$

where C_V and C_A are the vector and axial couplings of the SM fermion f to the Z boson, respectively, given by $C_V = T_f^3 - 2Q_f \sin^2\theta_W$ and $C_A = T_f^3$. Here the term proportional to Q_f^2 comes from the photon-mediated interaction. The terms proportional to C_V^2 and C_A^2 comes from the Z-mediated interaction while the term proportional to $C_V Q_f$ comes from the interference between photon and Z-mediated terms.

The cross-section in eq. (A.112) has a pole at $s = M_{Z'}^2$, which can be seen explicitly in the narrow width limit,

$$\frac{1}{(s-M_Z^2)^2+\Gamma_Z^2 M_Z^2} \approx \frac{1}{M_Z^4}\Theta(M_Z^2-s)+\frac{\pi\delta(s-M_Z^2)}{M_Z\Gamma_Z}+\frac{1}{s^2}\Theta(s-M_Z^2), \tag{A.113}$$

where Θ is the Heaviside function. The contribution to the collision integral from the Dirac delta term gives an identical contribution to the collision term due to Z-boson decays (see appendix A.6 or Refs. [190, 191]), discussed in the next subsection. To avoid double-counting we subtract the Dirac delta piece. Additionally, we also neglect the terms proportional to $\Theta(M_Z^2 - s)$ as the contribution from those terms is heavily suppressed compared to others. This yields the effective off-shell cross-section

$$\begin{aligned}\sigma^{\text{off}}_{ff\to\psi\psi} =& \frac{4\pi Q^2 N_c(f)\alpha^2}{s^3}\frac{\sqrt{s-4m^2}}{\sqrt{s-4m_f^2}} \\ &\times\left\{\frac{4}{3}(2m^2+s)(2m_f^2+s)\left[Q_f^2+\Theta(s-M_Z^2)\left(\frac{(C_V^2+C_A^2)}{4\cos^4\theta_W}-\frac{C_V Q_f}{\cos^2\theta_W}\right)\right]\right. \\ &\left.-\Theta(s-M_Z^2)\frac{(C_V^2+3C_A^2)m_f^2}{2\cos^4\theta_W}(s+2m^2)\right\}.\end{aligned} \tag{A.114}$$

We find that the cross-section from photon contributions alone (i.e., retaining only terms proportional to Q_f^2) to be at least an order of magnitude larger than the contribution from the remaining terms that involve at least one coupling to the Z. Thus, in the analytical calculation of the leak factor in eq. (5.15) we neglect the Z-mediated contributions for simplicity.

The forward energy transfer collision term, $\mathcal{C}_{\text{f}}$, corresponding to fermion annihilations into MCPs is calculated by using the cross-section in eq. (A.114) inside the generic collision term derived in eq. (A.19) and summing over all SM fermions. The total energy transfer collision term is then evaluated through $\mathcal{C} = \mathcal{C}_{\text{f}}(T_{\text{SM}}) - \mathcal{C}_{\text{f}}(T_{\text{HS}})$. We include quarks, treated as free fermions, for $T_{\text{SM}} > T_{\text{QCD}}$, where we take

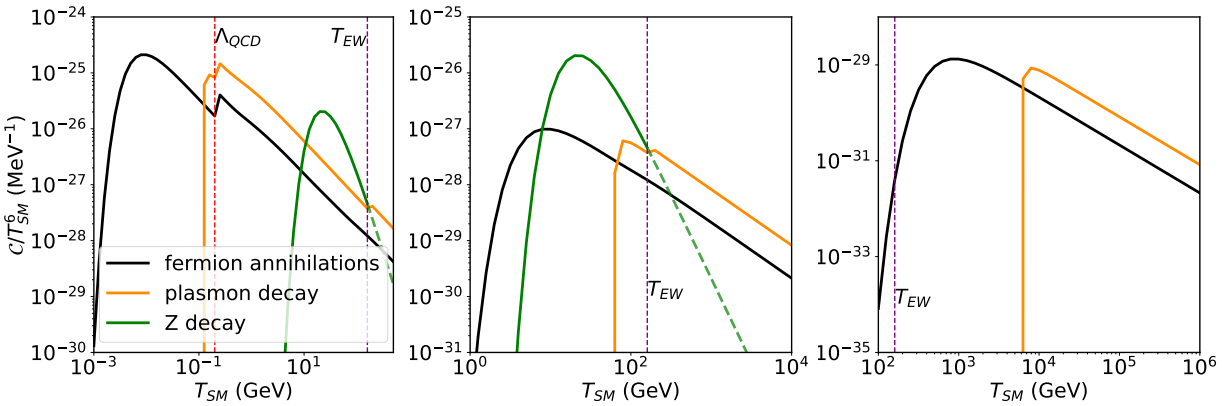


Figure A.5: Collision terms from the three s-channel processes (fermion annihilations, Z decays, and plasmon decays) for $Q = 10^{-9}$ and $m = 10$ MeV (left), $m = 10$ GeV (center), and $m = 1$ TeV (right). We plot $\mathcal{C}/T_{\rm SM}^6$, the maximum value of which provides the dominant contribution to the energy injected into the hidden sector. The black line corresponds to the collision term from SM fermion annihilations, the green line to Z decays (eq. (A.115)), and the orange line to plasmon decays (eq. (A.123)). Vertical red dashed and purple dashed lines mark the temperatures used for the QCD and electroweak phase transitions, respectively. Fermion annihilations into MCPs provide the dominant s-channel energy transfer process except in the mass range 0.3 GeV $\lesssim m \lesssim$ 40 GeV where Z-boson decays dominate.

$T_{\rm QCD} = 200$ MeV; for $T_{\rm SM} < T_{\rm QCD}$ we neglect hadronic contributions, as they are generally Boltzmann-suppressed.

In section 5.2.2, we found that the maximum of $\mathcal{C}_{\rm f} a^4/H$ roughly determines the final comoving energy density of dark radiation, where $\mathcal{C}_{\rm f}$ is the forward collision term and H is the Hubble rate. As $\mathcal{C}_{\rm f} a^4/H \propto \mathcal{C}_{\rm f}/T_{\rm SM}^6$, in figure A.5 we plot $\mathcal{C}_{\rm f}/T_{\rm SM}^6$ for fermion annihilation into MCPs (black line) for different values of m. The collision term $\mathcal{C}_{\rm f}/T_{\rm SM}^6$ reaches its maximum around $T_{\rm SM} \sim m/2$ below which it becomes Boltzmann suppressed. (We focus on the regime with $m > m_e$.)

Above the electroweak phase transition, the dark photon mixes with the hypercharge gauge boson. In the $s \gg M_Z^2$ limit, the cross-section in eq. (A.112) reduces to the cross-section describing annihilation through a hypercharge boson. For simplicity, we neglect the temperature dependence of the Higgs vev, and thus the (tree-level) Z mass, through the electroweak phase transition, as Z contributions are subleading below the transition and negligible above it.

A.5.2 Z-boson decay

Next, we discuss energy transfer from Z decays into MCPs. The collision term due to Z decays is given by

$$\mathcal{C}_{Z\to\psi\psi} = \Gamma_{Z\to\psi\psi} m_Z (n_Z - n_{eq,Z}(T_{\rm HS})), \tag{A.115}$$

where

$$\Gamma_{Z\to\psi\psi} = \frac{Q^2\alpha\tan^2\theta_W}{3} m_Z \sqrt{1-\frac{4m^2}{m_Z^2}}\left(1+\frac{2m^2}{m_Z^2}\right)\Theta(m_Z - m/2), \tag{A.116}$$

n_Z is the Z number density, $n_{Z,\rm eq}$ is the equilibrium number density of bosons with three degrees of freedom, and Θ is the Heaviside function. As Z bosons are always in equilibrium with the SM plasma, we have

$n_Z = n_{Z,\mathrm{eq}}(T_{\mathrm{SM}})$.

At temperatures above $T_{\mathrm{EW}} = 160$ GeV [30], electroweak symmetry is unbroken and thus the contribution from Z decays is absent. We use a simple model of the electroweak phase transition, where we set $\mathcal{C}_{Z\to\psi\psi} = 0$ for $T > T_{\mathrm{EW}}$ but neglect any temperature variation in the Z mass for $T < T_{\mathrm{EW}}$. As the main contribution from Z decays arises at temperatures significantly smaller than T_{EW}, this is a sufficient approximation for our purposes. In figure A.5, we plot the resulting collision term, normalized by T_{SM}^6, as the green line. The net energy injected into the HS is dominated by the peak values of $\mathcal{C}/T_{\mathrm{SM}}^6$, which occurs around $T_{\mathrm{SM}} \sim M_Z/4 \ll T_{\mathrm{EW}}$.

The contribution from Z decays to the total energy injected into the HS can dominate over the contribution from photon-mediated fermion annihilation when $0.3\ \mathrm{GeV} \lesssim m \lesssim 40\ \mathrm{GeV}$. Elsewhere, Z decays provide a sub-leading contribution.

A.5.3 Plasmon decay

Below the electroweak phase transition, $T_{\mathrm{SM}} < T_{\mathrm{EW}}$, the thermal effects in the plasma cause photons to acquire an in-medium plasma mass. The corresponding plasmons can decay into MCPs with the collision term given by [153, 227, 228]

$$\mathcal{C}_{\gamma\to\psi\psi} = \sum_{\mathrm{pol}} \int \frac{d^3k}{(2\pi)^3} \left(\frac{1}{e^{\omega/T_{\mathrm{SM}}} - 1} - \frac{1}{e^{\omega/T_{\mathrm{HS}}} - 1} \right) \omega \Gamma_{\gamma\to\psi\psi}, \tag{A.117}$$

where

$$\Gamma_{\gamma\to\psi\psi} = \frac{\alpha Q^2}{3\omega} Z (m_\gamma^2 + 2m^2) \sqrt{1 - \frac{4m^2}{m_\gamma^2}}. \tag{A.118}$$

Here Z is a wave-function renormalization factor and m_γ is the plasmon mass, both of which differ for transverse and longitudinal polarizations. For a relativistic photon, where $\omega(k) \equiv \sqrt{m_\gamma^2 + k^2} \gg m_\gamma$, the decays from the longitudinal polarization are negligible compared to the decays from the transverse polarizations [227, 228]. Moreover, for the transverse polarization at relativistic energies we have $Z \approx 1$ and

$$m_\gamma^2 = \sum_f g_f Q_f^2 \frac{4\alpha}{\pi} \int_0^\infty dp\ p\, f_f(p), \tag{A.119}$$

where f_f is the phase space distribution of the SM fermion f and the summation runs over all fermions; g_f counts the spin degrees of freedom of each fermion. Since $m_\gamma \ll T_{\mathrm{SM}}$, approximating $\omega \gg m_\gamma$ in eq. (A.117) is valid as the integrand is dominated by momenta with $\omega \sim T_{\mathrm{SM}}$. Thus the collision term simplifies to

$$\mathcal{C}_{\gamma\to\psi\psi} = \frac{2\alpha Q^2}{3} (m_\gamma^2 + 2m^2) \sqrt{1 - \frac{4m^2}{m_\gamma^2}} \times \left(n_\gamma(T_{\mathrm{SM}}) - n_\gamma(T_{\mathrm{HS}}) \right), \tag{A.120}$$

where n_γ is the equilibrium number density of photons. Energy transfer from plasmon decay is prohibited when $m_\gamma < 2m$. Since $m_\gamma \sim 0.1 T_{\mathrm{SM}}$, energy injection via plasmon is only efficient at high temperatures where $T_{\mathrm{SM}} > 10m$.

Above the electroweak phase transition, $T_{\mathrm{SM}} > T_{\mathrm{EW}}$, we need to evaluate the decay of hypercharge bosons into MCPs. The collision term for this process is similar to that for photon decay, up to the replacement

of the electric charge e by the hypercharge coupling $e/\cos\theta_W$ and the fermion electric charges Q_f by their hypercharges Q_Y. Consequently, we obtain

$$\mathcal{C}_{B\to\psi\psi} = \frac{2\alpha Q^2}{3\cos^4\theta_W}(m_B^2+2m^2)\sqrt{1-\frac{4m^2}{m_B^2}} \times (n_B(T_{\rm SM}) - n_B(T_{\rm HS}))\,, \tag{A.121}$$

where n_B is the equilibrium number density of hypercharge gauge bosons and m_B is the thermal mass, given by

$$m_B^2 = \frac{11\alpha\pi}{3\cos^2\theta_W} T_{\rm SM}^2, \tag{A.122}$$

for large temperatures. We take the plasmon decay contribution to be given by

$$\mathcal{C}_{\rm plasmon} = \begin{cases} \mathcal{C}_{\gamma\to\psi\psi} & T_{\rm SM} < T_{\rm EW} \\ \mathcal{C}_{B\to\psi\psi} & T_{\rm SM} > T_{\rm EW}. \end{cases} \tag{A.123}$$

In figure A.5 we compare the resulting collision term $\mathcal{C}_{\rm f}/T_{\rm SM}^6$ (orange line) to the collision term describing photon-mediated SM fermion annihilations. The collision term $\mathcal{C}_{\rm f}/T_{\rm SM}^6$ from fermion annihilation is maximized around $T_{\rm SM} = m/4$, while that from plasmon decay is maximized around $T_{\rm SM} = m/10$. Since the maximum value of $\mathcal{C}/T_{\rm SM}^6$ controls the final energy injected into the HS, the energy injected into the HS via plasmon decay is subdominant to the energy injected via fermion annihilations, even though at high temperatures the collision term for plasmon decay is larger than the collision term for fermion annihilation. Thus, the approximations used in $\mathcal{C}_{\rm plasmon}$ near the electroweak and QCD phase transitions are of marginal consequence in evaluating the resulting dark radiation density.

A.5.4 Coulomb scattering

SM particles can also inject energy into the HS through the Coulomb scattering of MCPs with SM particles, $\psi + f \to \psi + f$. The cross-section for Coulomb scattering has a forward singularity, which we regulate by adding a plasmon mass in the propagator [153].

Below the electroweak scale, the Coulomb scattering is mediated by photons, with the plasmon mass given by eq. (A.119). The relevant spin-summed matrix element for SM fermion scattering with MCPs is given by

$$|\mathcal{M}|^2_{f\psi\to f\psi} = \frac{8Q^2N_c(f)Q_f^2e^4}{(t-m_\gamma^2)^2}\left(2(s-m_f^2-m^2)^2+2st+t^2\right), \tag{A.124}$$

where Q, Q_f, m and m_f are the charge and mass of the MCP and the SM fermion, respectively, $N_c(f)$ is the color factor of the SM fermion, m_γ is the plasmon mass given by eq. (A.119), and s and t are the Mandelstam variables. The collision term for the above process, including quantum statistics, is given by eq. (A.63) with

$m_\phi \to m_\gamma$ and the coefficients of non-zero $c_{nm\lambda}$, defined in eq. (A.57), given by

$$\frac{c_{222}}{16\pi Q^2 Q_f^2 e^4} = \frac{3}{4}, \quad \frac{c_{202}}{16\pi Q^2 Q_f^2 e^4} = -\frac{1}{4}, \quad \frac{c_{022}}{16\pi Q^2 Q_f^2 e^4} = -\frac{1}{4}, \quad \frac{c_{002}}{16\pi Q^2 Q_f^2 e^4} = \frac{3}{4},$$
$$\frac{c_{001}}{16\pi Q^2 Q_f^2 e^4} = (m_f^2 + m^2), \quad \frac{c_{201}}{16\pi Q^2 Q_f^2 e^4} = m^2, \quad \frac{c_{021}}{16\pi Q^2 Q_f^2 e^4} = m_f^2,$$
$$\frac{c_{000}}{16\pi Q^2 Q_f^2 e^4} = 4 m_f^2 m^2. \tag{A.125}$$

Additionally, we multiply the resulting collision term by a factor of four to account for all combinations of particles and antiparticles. While solving the Boltzmann equations in section 5.2, we sum over the contribution from all SM fermions. Again, we include quarks for $T_{\rm SM} > T_{\rm QCD}$, and neglect hadron contributions for $T_{\rm SM} < T_{\rm QCD}$.

Above the electroweak scale the Coulomb scattering is mediated by the hypercharge boson. For a (Weyl) SM fermion scattering with MCPs, the spin-summed matrix element is

$$|\mathcal{M}|^2_{f\psi\to f\psi} = \frac{4Q^2 N_c(f) Q_Y(f)^2 e^4}{\cos^4\theta_W (t - m_B^2)^2} \left(2(s - m_f^2 - m^2)^2 + 2st + t^2\right), \tag{A.126}$$

where θ_W is the weak mixing angle, $Q_Y(f)$ is the hypercharge of the fermion, and m_B is the thermal mass of hypercharge gauge boson (eq. (A.122)). The coefficients $c_{nm\lambda}$ for the above matrix element are the same as those given in eq. (A.125) up to an overall rescaling by the factor $Q_Y^2/(2Q_f^2\cos^4\theta_W)$. Additionally, the Higgs doublet can also scatter with MCPs, with the corresponding matrix element being

$$|\mathcal{M}|^2_{H\psi\to H\psi} = 2\frac{Q^2 e^4}{4\cos^4\theta_W} \frac{1}{(t - m_B^2)^2} \times 8[s^2 + st - m^2(t + 2s) + m^4]. \tag{A.127}$$

The corresponding coefficients $c_{nm\lambda}$ are

$$\frac{c_{222}}{4\pi Q^2 e^4/\cos^4\theta_W} = \frac{3}{4}, \quad \frac{c_{202}}{4\pi Q^2 e^4/\cos^4\theta_W} = -\frac{1}{4}, \quad \frac{c_{022}}{4\pi Q^2 e^4/\cos^4\theta_W} = -\frac{1}{4},$$
$$\frac{c_{002}}{4\pi Q^2 e^4/\cos^4\theta_W} = -\frac{1}{4}, \quad \frac{c_{201}}{4\pi Q^2 e^4/\cos^4\theta_W} = m^2, \quad \frac{c_{001}}{4\pi Q^2 e^4/\cos^4\theta_W} = -m^2. \tag{A.128}$$

Unlike the s-channel processes, for Coulomb scattering the forward collision term describing energy transfer into the HS is sensitive to the distributions of both HS and SM particles. Moreover, the backward collision term for Coulomb scattering is of the same order of magnitude as the forward collision term for $T_{\rm HS} > 0.1 T_{\rm SM}$, while the backward term for s-channel processes is almost negligible compared to the forward term for $T_{\rm HS} < 0.9 T_{\rm SM}$.

In figure A.6 we compare the total collision term for Coulomb scattering between MCPs and SM particles with the forward collision term for SM fermion annihilation into MCPs. The collision term for Coulomb scattering decreases for smaller $T_{\rm HS}$ as the number density of HS particles in the initial state drops. The Coulomb scattering collision becomes the dominant process for $T_{\rm HS}/T_{\rm SM} > 0.35$.

A.6 Collision terms in $B - L$ model

The dominant energy injection from the Standard Model plasma into right-handed neutrinos occurs through the annihilations of the Standard Model fermions into right-handed neutrinos: $f + \bar{f} \to Z' \to \bar{\nu}_R + \nu_R$. In

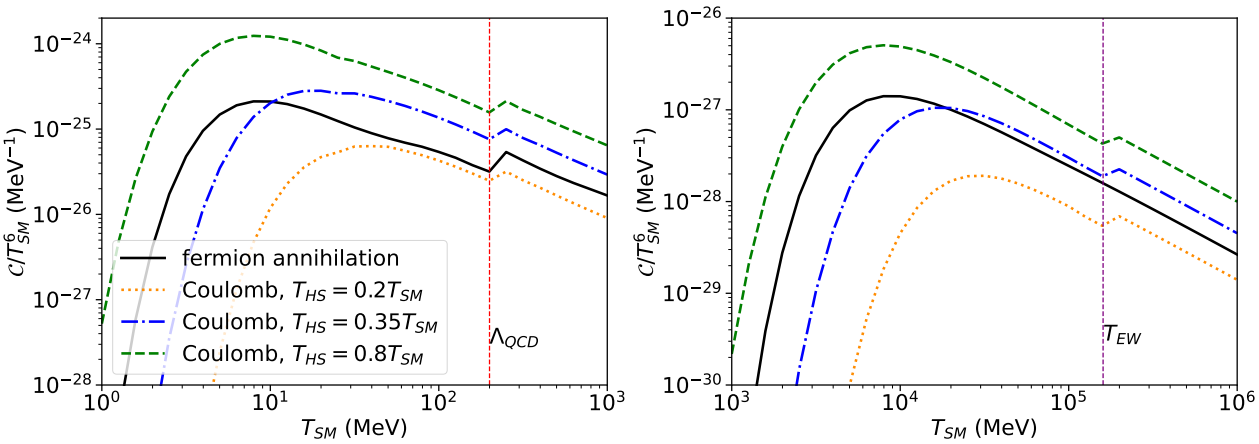


Figure A.6: Comparison between Coulomb scattering and fermion annihilation for $Q = 10^{-9}$ and MCP masses $m = 10$ MeV (left) and $m = 10$ GeV (right). Green, blue and orange lines show the collision term from Coulomb scattering, normalized by $T_{\rm SM}^6$, at different values of $T_{\rm HS}/T_{\rm SM}$ as indicated in the legend, while the black line shows the forward collision term due to SM fermion annihilation. The energy transferred via Coulomb scatterings dominates over that via annihilation for $T_{\rm HS} > 0.35T_{\rm SM}$.

appendix A we compute the collision term describing the forward energy transfer for the above process and include the quantum statistical phase space distribution for SM fermions. The corresponding collision term for the forward reaction is given by

$$\mathcal{C}_{\rm f} = \frac{1}{32\pi^4}\int_{4m_f^2}^{\infty} ds(s-4m_f^2)s\sigma_{f\bar{f}\to\nu_R\bar{\nu}_R}(s)T_{\rm SM}G_{\zeta_f}(\sqrt{s}/T_{\rm SM}), \tag{A.129}$$

where $\sigma_{f\bar{f}\to\nu_R\bar{\nu}_R}$ is the spin-summed center-of-mass (CM) frame cross-section, and G_ζ is a dimensionless function given by eq. (A.18) and is determined by the quantum statistical distribution $f(p) = [e^{-E/T} + \zeta]^{-1}$, where $\zeta = 1$ for fermions and $\zeta = -1$ for bosons. Since the f particles are fermions, we take $\zeta_f = 1$. In the limit fermion f can be approximated to have a Maxwell-Boltzmann distribution ($\zeta_f \to 0$), G asymptotes to the second order modified Bessel function of second kind, K_2, and eq. (A.129) then matches with the well-known result in Ref. [175].

The formula for the cross-section is

$$\sigma_{f\bar{f}\to\nu_R\bar{\nu}_R} = \frac{1}{2\pi}Q_f^2 g'^4 N_c(f)S(f)\frac{s}{(s-M_{Z'}^2)^2+\Gamma_{Z'}^2M_{Z'}^2}\sqrt{\frac{s}{s-4m_f^2}}\left(1+\frac{2m_f^2}{s}\right), \tag{A.130}$$

where Q_f is the $B-L$ charge, $N_c(f)$ is color factor that is equal to 3 for quarks and 1 for leptons, m_f is the mass of the fermion, and $\Gamma_{Z'}$ is the total decay width of the Z' boson,

$$\Gamma_{Z'} = \frac{g'^2}{12\pi}M_{Z'}\left[3+\sum_f^{2m_f<M_{Z'}} Q_f^2 N_c(f)\left(1+\frac{2m_f^2}{M_{Z'}^2}\right)\sqrt{1-\frac{4m_f^2}{M_{Z'}^2}}\right]. \tag{A.131}$$

Above the first factor of 3 comes from decays into ν_L and ν_R and the summation is over all SM fermions, except ν_L, with mass $m_f < M_{Z'}/2$.

In the narrow width limit, $\Gamma_{Z'} \ll M_{Z'}$, we can approximate

$$\frac{s}{(s-M_{Z'}^2)^2+\Gamma_{Z'}^2 M_{Z'}^2} \approx \frac{s}{M_{Z'}^4}\Theta(M_{Z'}^2-s) + \frac{\pi M_{Z'}\delta(s-M_{Z'}^2)}{\Gamma_{Z'}} + \frac{1}{s}\Theta(s-M_{Z'}^2), \tag{A.132}$$

where Θ is the Heaviside function. The resonant part of the cross-section, which correspond to the term with Dirac delta, is proportional to g'^2 due to the division by the total decay width, while the cross-section due to contact interactions, which correspond to terms with Heaviside functions, are proportional to g'^4. As g' is typically expected to be much smaller than 1, the energy injection is primarily determined by the resonant cross-section.

The contact interactions have off-shell Z' bosons. Consequently, the mediator Z' boson is a virtual particle and is unaffected by the expanding universe. The collision term corresponding to contact interaction, $\mathcal{C}^{\rm off}_{ff\to\nu_R\nu_R}$, is then simply given by eq. (A.129) with the following cross-section,

$$\sigma^{\rm off}_{f\bar{f}\to\nu_R\bar{\nu}_R} = \frac{1}{2\pi}Q_f^2 g'^4 N_c(f) S(f)\sqrt{\frac{s}{s-4m_f^2}}\left(1+\frac{2m_f^2}{s}\right)\left[\frac{s}{M_{Z'}^4}\Theta(M_{Z'}^2-s)+\frac{1}{s}\Theta(s-M_{Z'}^2)\right]. \tag{A.133}$$

In the limit $\Gamma_{Z'}$ is much larger than the Hubble rate, the on-shell Z' bosons do not experience any significant evolution in the expanding universe. Consequently, we can calculate the forward collision term due to the contribution from the Dirac-delta term, $\mathcal{C}^{\rm on}_{ff\to\nu_R\nu_R}$, using eq. (A.129). After summing over the annihilations from all SM fermions, we obtain

$$\mathcal{C}^{\rm on}_{ff\to\nu_R\nu_R} = \frac{3M_{Z'}^3}{2\pi^2}\left[\frac{\Gamma(Z'\to\nu_R)\Gamma(Z'\to SM)}{\Gamma_{Z'}}\right] T_{\rm SM} G_{\zeta_f}(M_{Z'}/T_{\rm SM}), \tag{A.134}$$

where $\Gamma(Z'\to\nu_R)$ and $\Gamma(Z'\to SM)$ are decay widths into right-handed neutrinos and SM fermions, respectively. Note that the above collision matches with the effective collision term we derived using Boltzmann equations in eq. (5.27).

In the limit $\Gamma_{Z'}$ is much smaller than the Hubble rate, the on-shell Z' bosons experience significant evolution in an expanding universe and hence we can no longer use eq. (A.134). Instead, we need to independently solve for the evolution of abundance of Z' using Boltzmann equations given in eqs. (5.22)-(5.24).

Appendix B

Collision operators relevant for cannibals

In this appendix we derive the various collision terms that are used in chapter 4. We first calculate the rate of $3 \to 2$ interactions that drive the cannibalistic evolution. We then calculate the $2 \to 2$ scattering rate between cannibals that are required to maintain internal kinetic equilibrium. Finally, we derive the perturbation equations for cannibal decaying into relativistic particles.

B.1 Three to two cannibal interaction

We begin by considering a number changing process of form

$$1 + 2 + 3 \to 4 + 5.$$

The corresponding scattering collision term is given by

$$\mathcal{C} = \int \frac{d^3p_1}{2E_1(2\pi)^3}\frac{d^3p_2}{2E_2(2\pi)^3}\frac{d^3p_3}{2E_3(2\pi)^3}\frac{d^3p_4}{2E_4(2\pi)^3}\frac{d^3p_5}{2E_5(2\pi)^3}(2\pi)^4\delta^4(p_1+p_2+p_3-p_4-p_5)S|\mathcal{M}|^2 \tag{B.1}$$
$$[f_1(p_1)f_2(p_2)f_3(p_3)(1 \pm f_4(p_4))(1 \pm f_5(p_5)) - f_4(p_4)f_5(p_5)(1 \pm f_1(p_1))(1 \pm f_2(p_2))(1 \pm f_3(p_3))].$$

Here $|\mathcal{M}|^2$ is the spin and polarizaition summed matrix element and S include the identical particle factors of initial and final states. Define Lorentz invariant phase space volume element as,

$$d\Pi_k = \prod_{i=1}^{k}\frac{d^3p_i}{2E_i(2\pi)^3}. \tag{B.2}$$

We can neglect the contribution from final state effects because we are in interested in regimes where $T \ll m$, where Maxwell-Boltzmann distribution is appropriate. Consequently, the collision term simplifies to

$$\mathcal{C} = \int d\Pi_5(2\pi)^4\delta^4(p_1+p_2+p_3-p_4-p_5)S|\mathcal{M}|^2[f_1(p_1)f_2(p_2)f_3(p_3) - f_4(p_4)f_5(p_5)]. \tag{B.3}$$

For such a collision term, principle of detailed balance tells us that the equilibrium distribution f is of the form:

$$f_{eq}(p) = e^{-p^0/T}. \tag{B.4}$$

In the absence of chemical equilibrium but assuming kinetic equilibrium (due to some strong $2 \to 2$ scattering process), the distribution function will attain a chemical potential,

$$f(p) = e^{\mu/T} e^{-p^0/T} = \frac{n}{n_{eq}} e^{-p^0/T}, \tag{B.5}$$

where $n = 2m \int d\Pi_1 f(p)$ is the number density and n_{eq} is the equilibrium number density. Substituting this in the collision term we get,

$$\begin{aligned}
\mathcal{C} =& \int d\Pi_5 (2\pi)^4 \delta^4(p_1 + p_2 + p_3 - p_4 - p_5) S|\mathcal{M}|^2 [e^{3\mu/T} e^{-(p_1^0+p_2^0+p_3^0)/T} - e^{2\mu/T} e^{-(p_4^0+p_5^0)/T}] \\
=& [e^{3\mu/T} - e^{2\mu/T}] \int d\Pi_5 (2\pi)^4 \delta^4(p_1 + p_2 + p_3 - p_4 - p_5) S|\mathcal{M}|^2 e^{-(p_1^0+p_2^0+p_3^0)/T} \\
=& n^2 [n - n_{eq}] \left[\frac{1}{n_{eq}^3} \int d\Pi_5 (2\pi)^4 \delta^4(p_1 + p_2 + p_3 - p_4 - p_5) S|\mathcal{M}|^2 e^{-(p_1^0+p_2^0+p_3^0)/T} \right] \\
\equiv& \langle \sigma v^2 \rangle n^2 [n - n_{eq}].
\end{aligned} \tag{B.6}$$

Here in the second equality we made use of energy conservation to pull out the factor $e^{-(p_1^0+p_2^0+p_3^0)/T}$.

Now let us concentrate on the form of $\langle \sigma v^2 \rangle$:

$$\langle \sigma v^2 \rangle = \frac{1}{n_{eq}^3} \int d\Pi_5 (2\pi)^4 \delta^4(p_1 + p_2 + p_3 - p_4 - p_5) S|\mathcal{M}|^2 e^{-(p_1^0+p_2^0+p_3^0)/T}. \tag{B.7}$$

In the limit the ingoing particles are non-relativistic, the s-wave component of the matrix element dominates over other component and consequently the matrix element can be approximated to be momentum independent. The remaining integral only has dependence on m (mass of the interacting particles) and T. We would like to disentangle dependence on them so that we just have an integral over dimensionless number.

We are primarily interested in the scenario with $T \ll m$ where the cannibal dominated era begins. In this limit, we can assume that particles $1, 2$ and 3 are non-relativistic,

$$p_i^0 \approx m + \frac{\vec{p}_i^2}{2m} \qquad\qquad i = 1, 2, 3.$$

However, particles 4 and 5 produced from such $3 \to 2$ process will necessarily be relativistic due to energy conservation,

$$p_j^0 = m\sqrt{1 + \frac{\vec{p}_j^2}{m^2}} \qquad |\vec{p}_j| \gg |\vec{p}_i| \qquad i = 1, 2, 3 \quad j = 4, 5.$$

With the above approximations we can rewrite $\langle \sigma v^2 \rangle$ as,

$$\langle \sigma v^2 \rangle \approx S|\mathcal{M}|^2 \frac{1}{n_{eq}^3} \int \frac{d^3p_1}{2m(2\pi)^3} \frac{d^3p_2}{2m(2\pi)^3} \frac{d^3p_3}{2m(2\pi)^3} \frac{d^3p_4}{2\sqrt{m^2+\vec{p}_4^2}(2\pi)^3} \frac{d^3p_5}{2\sqrt{m^2+\vec{p}_5^2}(2\pi)^3} e^{-3m/T} e^{-(\vec{p}_1^2+\vec{p}_2^2+\vec{p}_3^2)/(2mT)}$$
$$\times (2\pi)^4 \delta^3(-\vec{p}_4 - \vec{p}_5) \delta(3m - \sqrt{m^2+\vec{p}_4^2} - \sqrt{m^2+\vec{p}_5^2}). \tag{B.8}$$

Looking at above equation, we get obvious choice of making $\vec{p}_i$ dimensionless with

$$\vec{x}_i = \frac{\vec{p}_i}{\sqrt{mT}} \qquad \vec{x}_j = \frac{\vec{p}_j}{m} \qquad i = 1,2,3 \quad j = 4,5.$$

This yields,

$$\langle \sigma v^2 \rangle \approx S|\mathcal{M}|^2 \frac{1}{n_{eq}^3} \int \frac{(\sqrt{mT})^3 d^3x_1}{2m(2\pi)^3} \frac{(\sqrt{mT})^3 d^3x_2}{2m(2\pi)^3} \frac{(\sqrt{mT})^3 d^3x_3}{2m(2\pi)^3} \frac{m^3 d^3x_4}{2m\sqrt{1+\vec{x}_4^2}(2\pi)^3} \frac{m^3 d^3x_5}{2m\sqrt{1+\vec{x}_5^2}(2\pi)^3}$$
$$\times e^{-3m/T} e^{-(\vec{x}_1^2+\vec{x}_2^2+\vec{x}_3^2)/2} (2\pi)^4 \frac{1}{m^3} \delta^3(-\vec{x}_4 - \vec{x}_5) \frac{1}{m} \delta(3 - \sqrt{1+\vec{x}_4^2} - \sqrt{1+\vec{x}_5^2}). \tag{B.9}$$
$$= S|\mathcal{M}|^2 \frac{1}{n_{eq}^3} (\sqrt{mT})^9 \frac{1}{m^3} e^{-3m/T} \left[\int \frac{d^3x_1}{2(2\pi)^3} \frac{d^3x_2}{2(2\pi)^3} \frac{d^3x_3}{2(2\pi)^3} \frac{d^3x_4}{2\sqrt{1+\vec{x}_4^2}(2\pi)^3} \frac{d^3x_5}{2\sqrt{1+\vec{x}_5^2}(2\pi)^3} e^{-(\vec{x}_1^2+\vec{x}_2^2+\vec{x}_3^2)/2} \right.$$
$$\left. \times (2\pi)^4 \delta^3(-\vec{x}_4 - \vec{x}_5) \delta(3 - \sqrt{1+\vec{x}_4^2} - \sqrt{1+\vec{x}_5^2}) \right]. \tag{B.10}$$
$$\equiv S|\mathcal{M}|^2 \frac{1}{n_{eq}^3} (\sqrt{mT})^9 \frac{1}{m^3} e^{-3m/T} \times \xi \tag{B.11}$$

where ξ is some dimensionless number. Using equilibrium number density formula $n_{eq}(T) = (\frac{mT}{2\pi})^{3/2} e^{-m/T}$, the above becomes

$$\langle \sigma v^2 \rangle = \xi S|\mathcal{M}|^2 \frac{(2\pi)^{9/2}}{(\sqrt{mT})^9} e^{3m/T} (\sqrt{mT})^9 \frac{1}{m^3} e^{-3m/T}$$
$$= \frac{\xi'}{m^3} S|\mathcal{M}|^2 \tag{B.12}$$

where $\xi' = (2\pi)^{9/2}\xi$.

The quantity ξ' can be evaluated analytically,

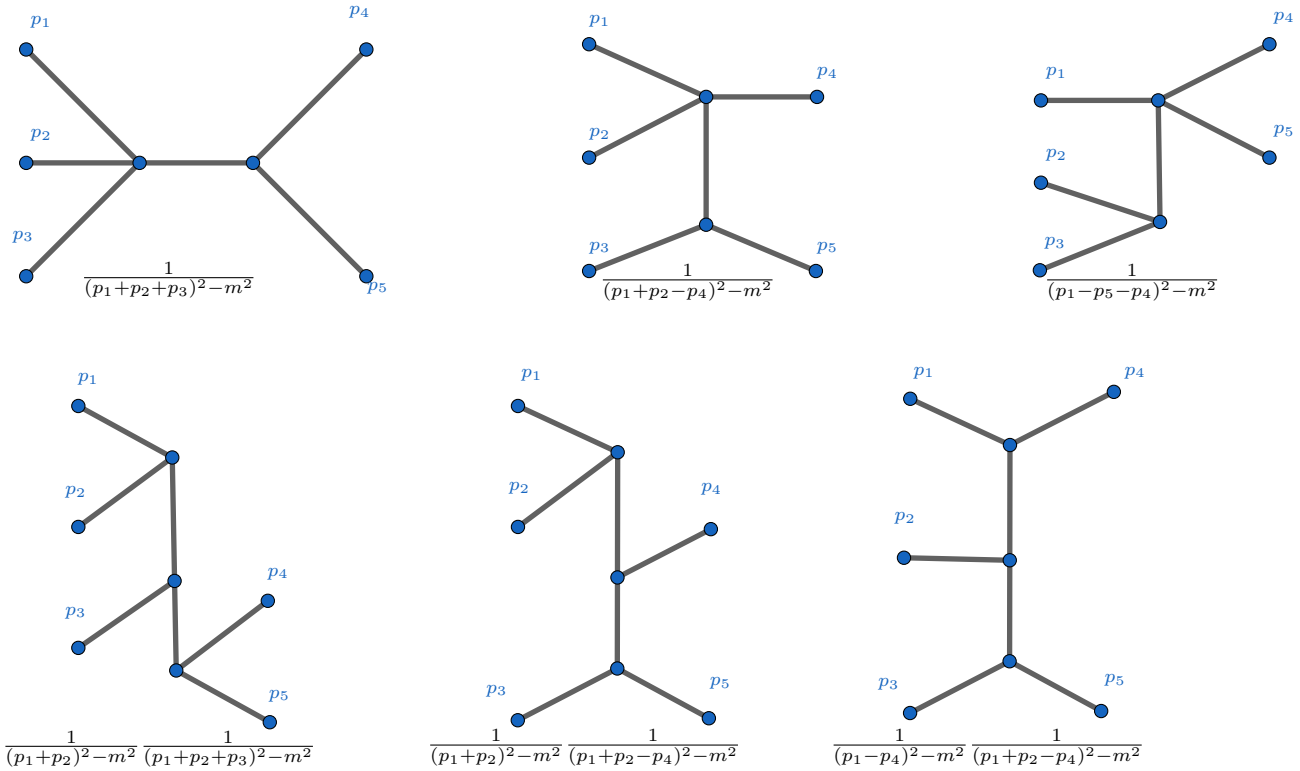
$$\xi' = (2\pi)^{9/2} \left(\int \frac{d^3x}{2(2\pi)^3} e^{-\vec{x}^2/2} \right)^3 \int \frac{d^3x_4}{2\sqrt{1+\vec{x}_4^2}(2\pi)^3} \frac{d^3x_5}{2\sqrt{1+\vec{x}_5^2}(2\pi)^3} (2\pi)^4 \delta^3(-\vec{x}_4 - \vec{x}_5) \delta(3 - \sqrt{1+\vec{x}_4^2} - \sqrt{1+\vec{x}_5^2})$$
$$= (2\pi)^{9/2} \frac{1}{2^5(2\pi)^{15}} (2\pi)^4 \left(\int dx e^{-x^2/2} \right)^9 \int d^3x_5 \frac{1}{1+x_5^2} \delta(2\sqrt{1+\vec{x}_5^2} - 3) \tag{B.13}$$
$$= \frac{\sqrt{5}}{192\pi}. \tag{B.14}$$

B.1.1 Calculating matrix element for real scalar field

Consider theory with

$$V(\phi) = \frac{\lambda}{4!}\phi^4 + \frac{A}{3!}\phi^3 + \frac{m^2}{2}\phi^2 = \frac{\lambda}{4!}\phi^4 + \frac{am}{3!}\phi^3 + \frac{m^2}{2}\phi^2.$$

The topologically distinct diagrams which contribute to 3 to 2 process are



The third and the fourth diagrams each have two more arrangements with different permutations of p_1, p_2 and p_3. The second, fifth and the sixth diagram has 5 more terms similarly. Thus we can write the matrix element as

$$\begin{aligned} i\mathcal{M} =& A\lambda\Big(\frac{1}{(p_1+p_2+p_3)^2-m^2}+\frac{6}{(p_1+p_2-p_4)^2-m^2}+\frac{3}{(p_1-p_4-p_5)^2-m^2}\Big) \\ & - A^3\Big(\frac{3}{(p_1+p_2)^2-m^2}\frac{1}{(p_1+p_2+p_3)^2-m^2}+\frac{6}{(p_1+p_2)^2-m^2}\frac{1}{(p_1+p_2-p_4)^2-m^2} \\ & +\frac{6}{(p_1-p_4)^2-m^2}\frac{1}{(p_1+p_2-p_4)^2-m^2}\Big) \end{aligned} \tag{B.15}$$

Assuming p_1, p_2, p_3 as non-relativistic and p_4, p_5 as semi relativistic; the momentum energy conservation gives

$$|\vec{p}_4| = |\vec{p}_5| = \frac{\sqrt{5}}{2}m \qquad\qquad \vec{p}_4 = -\vec{p}_5. \tag{B.16}$$

The above approximations simplifies the matrix element to

$$\begin{aligned} i\mathcal{M} =& \frac{5a}{8m}(-3\lambda+a^2) \\ S|\mathcal{M}|^2 =& \frac{1}{3!\times 2!}\frac{25a^2}{64m^2}(3\lambda-a^2)^2 \end{aligned} \tag{B.17}$$

The 3-to-2 cross section is then

$$\langle \sigma v^2 \rangle = \frac{\xi'}{m^3} S|\mathcal{M}|^2 = \frac{25\sqrt{5}a^2}{147456\pi m^5}(3\lambda - a^2)^2 \tag{B.18}$$

B.2 Cannibal 2-to-2 scattering rate

The 2-to-2 scattering rate can be computed from the forward piece of the collision term, $\Gamma_{\rm sc} = \hat{C}/n_{\rm can}$, where $\hat{C}$ is given by

$$\hat{C} = \int \frac{d^3p_1}{2E_1(2\pi)^3}\frac{d^3p_2}{2E_2(2\pi)^3}\frac{d^3p_3}{2E_3(2\pi)^3}\frac{d^3p_4}{2E_4(2\pi)^3}(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)S|\mathcal{M}|^2 f(p_1)f(p_2). \tag{B.19}$$

Here $|\mathcal{M}|^2$ is the matrix element, $S = 1/4$ includes the identical particle factors of initial and final states and f is the phase-space distribution. Since we are primarily concerned with the scattering rate when the cannibal particles are non-relativistic, we have dropped the final state phase space distributions.

When the cannibal fluid is in kinetic equilibrium, f is given by the Maxwell-Boltzmann distribution,

$$f(p) = e^{(\mu - E)/T_c}. \tag{B.20}$$

In equilibrium, the collision term can be written as

$$\hat{C} = e^{2\mu/T_c}\frac{T_c}{64\pi^4}\frac{1}{2}\int_{4m^2}^{\infty} ds \mathcal{A}(s)\sqrt{s-4m^2}K_1(\sqrt{s}/T_c) \tag{B.21}$$

where K_1 is the modified Bessel function of the second kind, s is the Mandelstam variable, and $\mathcal{A}(s)$ is the integral of the squared matrix element over final state phase space,

$$\mathcal{A}(s) = \frac{1}{2}\frac{1}{8\pi}\frac{\sqrt{s-4m^2}}{\sqrt{s}}\int \frac{d\Omega}{4\pi}|\mathcal{M}|^2. \tag{B.22}$$

We have included factors of 1/2 in both eq. (B.21) and eq. (B.22) to account for identical particles appearing in both the final and initial states.

For the cannibal Lagrangian given in eq. (4.1), we find the matrix element describing scattering is, in the non-relativistic limit,

$$|\mathcal{M}|^2 = \left(\lambda - \frac{5}{3}\frac{g^2}{m^2}\right)^2. \tag{B.23}$$

Inserting the above matrix element in eq. (B.21) and expanding the integrand in T/m gives the leading contribution to the collision term in the non-relativistic limit:

$$\hat{C} \approx e^{2(\mu-m)/T_c}\left(\frac{mT_c}{2\pi}\right)^3\sqrt{\frac{T_c}{m}}\frac{\left(\lambda - \frac{5}{3}\frac{g^2}{m^2}\right)^2}{64\pi^{3/2}m^2}. \tag{B.24}$$

Expressing the chemical potential in terms of number density and the temperature of the cannibal fluid

using the relation

$$n_{\rm can} = e^{\mu/T_c} n_{\rm can,eq} = e^{(\mu-m)/T_c} \left(\frac{mT_c}{2\pi}\right)^{3/2} \tag{B.25}$$

yields our desired result

$$\Gamma_{\rm sc} = \frac{\hat{C}}{n_{\rm can}} = n_{\rm can} \langle \sigma_c v_c \rangle, \tag{B.26}$$

where

$$\langle \sigma_c v_c \rangle = \frac{1}{64\pi^{3/2} m^2} \left(\lambda - \frac{5}{3}\frac{g^2}{m^2}\right)^2 \sqrt{\frac{T_c}{m}}. \tag{B.27}$$

B.3 Perturbed collision operators for a decaying semi-relativistic particle

In this section we derive the contribution from cannibal decays to the cosmological perturbation equations. We include the decay terms up to first order in w_c and c_s^2, or equivalently up to first order in T_c/m.

We begin by writing the Boltzmann equations for a generic particle in a perturbed FRW universe, whose metric given by

$$ds^2 = -[1+2\psi]dt^2 + a^2(t)[1-2\phi](dx^2+dy^2+dz^2). \tag{B.28}$$

Expressing the particle's phase space distribution in the form, $f(\vec{p},\vec{x},t) = \bar{f}(p,t) + \delta f(\vec{p},\vec{x},t)$, where $\bar{f}$ and δf are unperturbed homogeneous and perturbed inhomogeneous pieces, respectively, the Fourier transform of the Boltzmann equation is given, to first order in perturbations, by

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + i\frac{\vec{k}\cdot\vec{p}}{aE}\delta f - \left[H - \frac{d\phi}{dt}\right]\frac{p^2}{E}\frac{\partial f}{\partial E} - i\frac{\vec{k}\cdot\vec{p}}{a}\psi\frac{\partial \bar{f}}{\partial E} = \frac{1+\psi}{E}\hat{C}[f]. \tag{B.29}$$

Here $\vec{k}$ is the comoving Fourier wavenumber and $\hat{C}$ is the collision operator.

We are interested in the collision operator that describes the cannibal particle decaying into pairs of relativistic SM particles. The corresponding collision operators for the cannibal and radiation distributions are then given by

$$\hat{C}_\Gamma[f_c(p)] = -\frac{1}{2}\int d\Pi_1 d\Pi_2 (2\pi)^4 \delta(E-E_1-E_2)\delta^3(\vec{p}-\vec{p}_1-\vec{p}_2) S|\mathcal{M}_\Gamma|^2 f_c(p) \tag{B.30}$$

$$\hat{C}_\Gamma[f_r(p_1)] = \int d\Pi d\Pi_2 (2\pi)^4 \delta(E-E_1-E_2)\delta^3(\vec{p}-\vec{p}_1-\vec{p}_2)|\mathcal{M}_\Gamma|^2 f_c(p), \tag{B.31}$$

where f_c and f_r are distribution functions for the cannibals and relativistic SM particles, respectively, $d\Pi_k = d^3k/[(2\pi)^3 2E_k]$, $|\mathcal{M}_\Gamma|^2$ is the matrix element corresponding to cannibal decays into radiation, and S is the identical particle factor. The collision term for SM radiation does not include a factor of $1/2$ because two SM particles are produced in the decay. We have neglected the contribution from final state effects as well as those from inverse decays because the cannibal decays become important ($\Gamma \sim H$) when $T_c \ll m$.

To obtain the evolution equations for density and velocity perturbations we take the energy-weighted phase space integral ($\int \frac{d^3p}{(2\pi)^3} E$) and the phase space integral of the first moment ($\int \frac{d^3p}{(2\pi)^3}[\vec{k}\cdot\vec{p}]$) of eq. (B.29).

B.3.1 Cannibal Perturbation Equations

First we use the definition of the rest-frame decay width, Γ, in eq. (A.34) to evaluate eq. (B.30) for the cannibal collision operator:

$$\hat{C}_\Gamma[f_c] = -m\Gamma f_c, \tag{B.32}$$

where m is the mass of the cannibal particle. The Boltzmann equation for cannibals (eq. (B.29)) will also include collision operators originating from cannibal self-interactions. However, these collision operators do not contribute to the perturbation equations for energy density or velocity as the self-interactions do not affect the energy and momentum of the fluid. Consequently, only the contribution from the decay collision operator remains after integrating the Boltzmann equation for cannibals over $\int \frac{d^3p}{(2\pi)^3} E$:

$$\frac{\partial \rho_{\rm can}}{\partial t} + \frac{1}{a}(\bar{\rho}_{\rm can} + \bar{\mathcal{P}}_{\rm can})\theta_{\rm can} + 3\left[H - \frac{d\phi}{dt}\right](\rho_{\rm can} + \mathcal{P}_{\rm can}) = -m\Gamma(1+\psi)n_{\rm can}. \tag{B.33}$$

To obtain the above result we used the definitions of energy density (ρ), number density (n), and pressure ($\mathcal{P}$) in terms of f. We also used the definition of the divergence of fluid velocity:

$$\theta = \frac{i}{\bar{\rho}(1+w)} \int \frac{d^3p}{(2\pi)^3} (\vec{k}\cdot\vec{p})\delta f. \tag{B.34}$$

Writing ρ and $\mathcal{P}$ in terms of homogeneous and perturbed pieces,

$$\rho = \bar{\rho}(1+\delta) \qquad\qquad \mathcal{P} = w\bar{\rho} + c_s^2\bar{\rho}\delta, \tag{B.35}$$

in eq. (B.33) and using the energy conservation equation of the cannibal fluid (eq. (4.24)) to evaluate $\mathrm{d}\bar{\rho}_{\rm can}/\mathrm{d}t$, we obtain

$$\dot{\delta}_c + \frac{1}{a}(1+w_c)\theta_c - 3\dot{\phi}(1+w_c) + 3H(c_s^2 - w_c)\delta_c = -\Gamma\frac{m\bar{n}_{\rm can}}{\bar{\rho}_{\rm can}}\left[\psi + \frac{\delta n_{\rm can}}{\bar{n}_{\rm can}} - \delta_c\right], \tag{B.36}$$

where dot denotes differentiation with respect to t.

While $T_c \ll m$, we can further simplify the RHS by relating $n_{\rm can}$ to $\rho_{\rm can}$ and $\mathcal{P}_{\rm can}$ using

$$\rho_{\rm can} \approx m\int \frac{d^3p}{(2\pi)^3}\left(1 + \frac{p^2}{2m^2}\right) f_c = mn_{\rm can} + \frac{3}{2}\mathcal{P}_{\rm can}. \tag{B.37}$$

It follows that

$$\frac{\delta n_{\rm can}}{\bar{n}_{\rm can}} \approx \delta_c \frac{1 - \frac{3}{2}c_s^2}{1 - \frac{3}{2}w_c} \approx \delta_c\left[1 + \frac{3}{2}(w_c - c_s^2) + \mathcal{O}(w_c^2)\right]. \tag{B.38}$$

Using the above result to evaluate $\delta n_{\rm can}/n_{\rm can}$ in eq. (B.36), we obtain the perturbation equation for δ_c by expanding the terms proportional to Γ to first order in w_c and c_s^2,

$$\dot{\delta}_c + \frac{1}{a}(1+w_c)\theta_c - 3\dot{\phi}(1+w_c) + 3\left(H - \frac{\Gamma}{2}\right)(c_s^2 - w_c)\delta_c = -\Gamma\psi\left(1 - \frac{3}{2}w_c\right). \tag{B.39}$$

Next we calculate the perturbation equation for the divergence of fluid velocity, θ, by evaluating the phase-

space integration of the first moment ($\int \frac{d^3p}{(2\pi)^3}\left(\vec{k}\cdot\vec{p}\right)$) of the Boltzmann equation for cannibals (eq. (B.29)). Note that all integrals that are odd in $\hat{p}$ will be proportional to δf, so any products of such integrals with metric perturbations can be neglected. Most of the remaining integrals can be evaluated using the definitions of ρ, n, $\mathcal{P}$, and θ. The only integral not covered by these definitions contains $(\hat{k}\cdot\hat{p})^2$ in the integrand, which is contained in the definition of anisotropic stress:

$$\sigma \equiv -\frac{1}{\bar{\rho}(1+w)}\int \frac{d^3p}{(2\pi)^3}\frac{p^2}{E}\left[\left(\hat{k}\cdot\hat{p}\right)^2-\frac{1}{3}\right]\delta f. \tag{B.40}$$

Finally, we use $\dot{\bar{\mathcal{P}}} = c_s^2\dot{\bar{\rho}}_{\rm can}$ and evaluate $\dot{\bar{\rho}}_{\rm can}$ using the energy conservation equation (eq. (4.24)) to obtain

$$\begin{aligned}\dot{\theta}_c + H(1-3c_s^2)\theta_c - \frac{k^2}{a}\psi + \frac{k^2}{a}\sigma_{\rm can} - \frac{c_s^2k^2}{a(1+w_c)}\delta_c \\ = m\Gamma\left[\frac{\bar{n}_{\rm can}(1+c_s^2)}{\bar{\rho}_{\rm can}(1+w_c)}\theta_c - \frac{i}{\bar{\rho}_{\rm can}(1+w_c)}\int\frac{d^3p}{(2\pi)^3}\left(\vec{k}\cdot\vec{p}\right)\frac{\delta f}{E}\right].\end{aligned} \tag{B.41}$$

We further simplify the square bracket on the RHS by approximating $E \approx m + p^2/(2m)$ and using eq. (B.37). Simplifying the resulting expression by keeping only leading order terms in w_c and c_s^2 and then using the definition of θ, we obtain

$$\begin{aligned}\dot{\theta}_c + H(1-3c_s^2)\theta_c - \frac{k^2}{a}\psi + \frac{k^2}{a}\sigma_{\rm can} - \frac{c_s^2k^2}{a(1+w_c)}\delta_c \\ = \Gamma\left[\left(-\frac{5}{2}w_c + c_s^2\right)\theta_c + \frac{i}{\bar{\rho}_{\rm can}(1+w_c)}\int\frac{d^3p}{(2\pi)^3}\frac{p^2}{2m^2}\left(\vec{k}\cdot\vec{p}\right)\delta f\right].\end{aligned} \tag{B.42}$$

To simplify the integral on the RHS, we note that the distribution function for a non-relativistic perfect fluid can be written as

$$f = e^{(\mu+\delta\mu)/T}e^{-(E-\vec{p}\cdot\vec{v})/T} \approx \bar{f}(E) - \bar{f}'(E)(\delta\mu + \vec{p}\cdot\vec{v}), \tag{B.43}$$

where $\delta\mu$ and $\vec{v}$ encode the density and velocity perturbations in the fluid. Using the above expression of f in the θ definition (eq. (B.34)) we obtain $\theta = i\vec{k}\cdot\vec{v}$. Consequently, the integral on the RHS of eq. (B.42) simplifies to yield

$$\int\frac{d^3p}{(2\pi)^3}\frac{p^2}{2m^2}\left(\vec{k}\cdot\vec{p}\right)\delta f \approx -\int\frac{d^3p}{(2\pi)^3}\frac{p^2}{2m^2}\left(\vec{k}\cdot\vec{p}\right)\bar{f}'(E)(\vec{p}\cdot\vec{v}) = -i\bar{\rho}_{\rm can}\theta_c\left[\frac{5}{2}w_c + \mathcal{O}(w_c^2)\right]. \tag{B.44}$$

Therefore, the perturbation equation for θ_c is given by

$$\dot{\theta}_c + H(1-3c_s^2)\theta_c - \frac{k^2}{a}\psi + \frac{k^2}{a}\sigma_{\rm can} - \frac{c_s^2k^2}{a(1+w_c)}\delta_c = \Gamma c_s^2\theta_c \tag{B.45}$$

to leading order in w_c and c_s^2 for terms proportional to Γ.

B.3.2 Radiation Perturbation Equations

We find the equation for radiation density perturbations by taking the energy-weighted phase space integral ($\int \frac{d^3p}{(2\pi)^3}E$) of the Boltzmann equation for radiation (eq. (B.29)). The resulting integral of the collision term

on the RHS is same as that encountered for the cannibal except with an opposite sign,

$$(1+\psi)\int \frac{d^3p_1}{(2\pi)^3 E_1} S E_1 \hat{C}_\Gamma[f_r(p_1)] = -(1+\psi)\int \frac{d^3p}{(2\pi)^3 E} E \hat{C}_\Gamma[f_c(p)] = (1+\psi) m n_{\rm can}\Gamma, \tag{B.46}$$

where $\hat{C}_\Gamma[f_c(p)]$ and $\hat{C}_\Gamma[f_r(p_1)]$ are given in eq. (B.30) and eq. (B.31). The above equality is a direct consequence of energy conservation, which sets the energy of the daughter particle equal to half of the energy of the cannibal particle, $E_1 = E/2$. Since the expression on the LHS now features integration over the phase space of both radiation particles, the symmetry factor S appears.

Similar to the cannibal case, we simplify the LHS of the energy-weighted phase space integral of the Boltzmann equation by using the definitions in eq. (B.35) and using the energy conservation equation for ρ_r (eq. (4.25)) to yield

$$\dot{\delta}_r + \frac{4}{3}\frac{\theta_r}{a} - 4\dot{\phi} = \frac{m\Gamma\bar{n}_{\rm can}}{\bar{\rho}_r}\left[\psi + \frac{\delta n_{\rm can}}{\bar{n}_{\rm can}} - \delta_r\right]. \tag{B.47}$$

Above we have made use of the fact that $w = c_s^2 = 1/3$ for radiation. The $\delta n_{\rm can}$ in the RHS can be further simplified using eq. (B.38) to give

$$\dot{\delta}_r + \frac{4}{3}\frac{\theta_r}{a} - 4\dot{\phi} = \frac{m\Gamma\bar{n}_{\rm can}}{\bar{\rho}_r}\left[\psi + \delta_c - \delta_r + \frac{3}{2}\delta_c(w_c - c_s^2)\right]. \tag{B.48}$$

Next, we find the perturbation equations for the divergence of the radiation fluid velocity by evaluating the phase-space integration of the first moment ($\int \frac{d^3p}{(2\pi)^3}[\vec{k}\cdot\vec{p}]$) of the Boltzmann equation for radiation (eq. (B.29)). The RHS of the resulting equation is of the form

$$\begin{aligned}(1+\psi)\int \frac{d^3p_1}{(2\pi)^3 E_1}(\vec{p}_1\cdot\vec{k})\hat{C}_\Gamma[f_r(p_1)] \\ = 2(1+\psi)\int d\Pi d\Pi_1 d\Pi_2 (2\pi)^4 (\vec{p}_1\cdot\vec{k})\delta(E-E_1-E_2)\delta^3(\vec{p}-\vec{p}_1-\vec{p}_2) S|\mathcal{M}_\Gamma|^2 f_c(p).\end{aligned} \tag{B.49}$$

In the above integral we replace $\vec{p}_1\cdot\vec{k} \to (\vec{p}_1+\vec{p}_2)\cdot\vec{k}/2$ as the labels 1 and 2 are interchangeable. Moreover, by momentum conservation we have $\vec{p}_1+\vec{p}_2 = \vec{p}$, which yields

$$\begin{aligned}(1+\psi)\int \frac{d^3p_1}{(2\pi)^3 E_1}(\vec{p}_1\cdot\vec{k})\hat{C}_\Gamma[f_r(p_1)] \\ =&(1+\psi)\int d\Pi\left[\int d\Pi_1 d\Pi_2 (2\pi)^4\delta^4(p-p_1-p_2) S|\mathcal{M}_\Gamma|^2\right](\vec{p}\cdot\vec{k}) f_c(p) \qquad \text{(B.50)} \\ =&\Gamma\int \frac{d^3p}{(2\pi)^3}\left(\vec{k}\cdot\vec{p}\right)\frac{\delta f}{\sqrt{1+p^2/m^2}} \approx -i\Gamma\bar{\rho}_{\rm can}\theta_c\left[1-\frac{3}{2}w_c\right]. \qquad \text{(B.51)}\end{aligned}$$

Here in the second line we first expanded the denominator to first order in p^2/m^2 and then used the definition of θ (eq. (B.34)) along with the result given in eq. (B.44) to obtain the final answer.

We simplify the phase space integration of the first moment of the LHS of Boltzmann equation in the same way as we did for cannibal perturbations. Expressing the cannibal energy density in terms of the cannibal

number density using eq. (B.37) gives

$$\dot{\theta}_r - \frac{k^2}{4a}\delta_r - \frac{k^2}{a}\psi + \frac{k^2}{a}\sigma_r = \Gamma \frac{m\bar{n}_{\rm can}}{\bar{\rho}_r}\left[\frac{3}{4}\theta_c - \theta_r\right]. \tag{B.52}$$

Appendix C

Steady-state solutions in cosmological perturbations

In this section we show the procedure to analytically calculate the evolution of amplitude for cosmological perturbations that are coupled to an external source, using WKB approximation deep inside the horizon. In particular we explain the behaviour of cannibal and radiation perturbation amplitudes as seen in figures 4.5 and 4.7. We also explain why the feedback of radiation perturbation on other density perturbations can be neglected deep inside the horizon.

We first simplify the system of perturbation equations for cannibal and radiation, eqs. (4.46)-(4.52), for modes deep inside the horizon. For $k \gg aH$, the equation for metric perturbation, eq. (4.52), simplifies to give

$$\phi = -\frac{3}{2}\left(\frac{aH}{k}\right)^2 \frac{\rho_{can}\delta_c + \rho_r\delta_r}{\rho_{can} + \rho_r}. \tag{C.1}$$

We can see that $\phi \ll \delta_c, \delta_r$ because $k \gg aH$. Using the smallness of ϕ we can then simplify radiation and cannibal density perturbation. In particular, we can ignore ϕ' and ϕ terms in δ_c (eq. (4.46)) and δ_r (eq. (4.50)) equations. Moreover for cannibal perturbation we also neglect the terms $c_s^2 - w_c, w_c'(a)$ and $d(c_s^2(a))/da$ as w_c is slowly varying before a_{fz} and rapidly become negligible after a_{fz}. Next we replace ϕ in the θ equations for both species (eq. (4.51) and in eq. (4.47)) using eq. (C.1). Finally, we combine the θ and δ equations: eq. (4.51) with eq. (4.50) for radiation and eq. (4.47) with eq. (4.46) for cannibal. While combining eq. (4.51) with eq. (4.50) we neglect derivatives of

$$\zeta = \Gamma\rho_{can}/(H\rho_r). \tag{C.2}$$

This is because ζ' is at order ~ 1 values only for a very short while when decays just become important in SM radiation. Before the decays become important, ζ' is much less than one and afterwards $\zeta = 5/2$ (see eq. 3.4

and section 2.5) and so $\zeta' = 0$. The simplified cannibal and radiation perturbation equations are of form,

$$\delta_c''(a) + \Big(\frac{(a^2H)'}{a^2H} + \frac{1}{a}(1-3w_c)\Big)\delta_c' + \frac{(1+w_c)}{a^2}\Big(\frac{c_s^2}{1+w_c}\frac{k^2}{(aH)^2} - \frac{3}{2}\frac{\rho_{can}}{\rho_{can}+\rho_r}\Big)\delta_c = \frac{3}{2}\frac{(1+w_c)}{a^2}\frac{\rho_r\delta_r}{\rho_{can}+\rho_r} \tag{C.3}$$

$$\delta_r''(a) + \Big(\frac{(a^2H)'}{a^2H} + \frac{2\zeta}{a}\Big)\delta_r'(a) + \frac{1}{3}\frac{\tilde{k}^2}{(a^2H)^2}\delta_r = \Big(\frac{(a^2H)'}{a^2H}\frac{\zeta}{a} + \frac{\zeta^2}{a^2} - \frac{\zeta}{a^2} + \frac{2}{a^2}\frac{\rho_{can}}{\rho_{can}+\rho_r}\Big)\delta_c + \frac{2+w_c}{1+w_c}\frac{\zeta}{a}\delta_c'(a). \tag{C.4}$$

We checked numerically that above equations accurately describe the evolution of cannibal and radiation perturbations for modes deep inside the horizon.

The terms with $\rho_{can} + \rho_r$ in denominator come from the contribution of metric perturbation and causes the density perturbation to grow. For radiation perturbation, the decays from cannibal act both like a damping term (ζ term in the coefficient of δ_r') as well as provide an external source to radiation perturbation (ζ terms in the RHS). When the decays from cannibal have not become dominant for SM radiation bath, $\zeta \ll 1$, cannibal density affects radiation density only gravitationally.

We are primarily interested in understanding cannibal perturbations when the mode is within Jeans length scale, and radiation perturbations when mode is within comoving horizon. Both these scenarios are well described by the generic equation of form,

$$\delta''(a) + \left(\frac{(a^2H)'}{a^2H} + \frac{n}{a}\right)\delta'(a) + \frac{\omega^2(a)}{(a^2H)^2}\delta = S(a)e^{i\int\Omega_s(a)da}, \tag{C.5}$$

where n/a and ω^2 represents damping term and oscillation frequency and S represents a source which is oscillating with frequency Ω_s. The source function, S, we shall find to be independent of δ in the scenarios of our interest (see discussion in section C.4). Consequently the equation described above has two forms of solutions: a transient (homogeneous) solution that describes the evolution of perturbations in the absence of external source, and a steady-state (inhomogeneous) solution that is driven purely by S.

In the following subsections we shall first apply WKB approximation to find the transient and steady-state solution for the generic equation given in eq. (C.5). The application of WKB approximation varies significantly for the steady-state solution depending on whether S is oscillating or evolving slowly ($\Omega_s = 0$). Consequently, we solve for the inhomogeneous solution separately depending on evolution of S. Finally, we describe the behaviour of cannibal and radiation perturbation amplitudes as seen in figures 4.5 and 4.7 using the transient and steady-state solution.

C.1 Transient (homogeneous) solution

In this sub-section we solve for the transient solution of eq. (C.5) using WKB approximation.

The transient solution will be of form

$$\delta_t = D_t(a)e^{i\int\Omega(a)da}, \tag{C.6}$$

where D_t and Ω are the amplitude and frequency of oscillations respectively. Substituting this solution back in eq. (C.5) and setting $S = 0$ gives,

$$\left[D_t'' + \left(\frac{(a^2H)'}{a^2H} + \frac{n}{a}\right) D_t' + D_t\left(\frac{\omega^2(a)}{(a^2H)^2} - \Omega^2\right)\right] + i\left[2D_t'\Omega + D_t\Omega' + \left(\frac{(a^2H)'}{a^2H} + \frac{n}{a}\right) D_t\Omega\right] = 0. \tag{C.7}$$

Both imaginary part and the real part above have to be zero identically. The imaginary part can be solved exactly to give

$$D_t^2\ \Omega(a^2H)e^{\int_{a_*}^{a} n(\tilde{a})d\ln(\tilde{a})} = \text{constant}, \tag{C.8}$$

where a_* is some scale factor which provides the initial condition. To solve for the real part we use the WKB approximation, $\omega^2/(a^2H)^2 \gg D_t''/D_t, D_t'/(aD_t)$, i.e. the oscillations are happening much faster than the rate at which amplitude is changing. With this approximation the real part of the equation effectively gives us the condition

$$\Omega(a) = \frac{\omega(a)}{a^2H}. \tag{C.9}$$

The amplitude D_t can be solved by substituting the above relation back in eq. (C.8) to give the full transient solution as

$$\delta_t(a) = \delta_t(a_*)\sqrt{\frac{\omega(a_*)}{\omega(a)}}e^{-\int_{a_*}^{a} n(\tilde{a})d\ln(\tilde{a})/2}e^{i\int_{a_*}^{a}\omega(\tilde{a})/(\tilde{a}^2H)d\tilde{a}}. \tag{C.10}$$

Note that the damping due to the Hubble term present in the coefficient of δ' in eq. (C.5) is exactly compensated by the Hubble term appearing in the frequency. Hence expanding universe does not lead to damping of perturbations as one might naively think by looking at eq. (C.5). In fact by rewriting eq. (C.5) in terms of conformal time, $d\eta = da/(a^2H)$, one can see that the Hubble damping term completely disappears.

C.2 Steady-state (inhomogeneous) solution for slowly evolving source

In this sub-section we solve for the steady-state solution of eq. (C.5) for a slowly varying $S(a)$ with $\Omega_S = 0$, using WKB approximation.

Since the source is slowly varying, we can expect the steady-state solution, $\delta_{st}(a)$, sourced by S will also be slowly varying compared to ω. Consequently assuming $\omega^2/(a^2H)^2 \gg a\delta_{st}'/\delta_{st}, a^2\delta_{st}''/\delta_{st}$ in eq. (C.5) we obtain

$$\delta_{st} = \frac{S(a)}{\omega^2(a)}(a^2H)^2. \tag{C.11}$$

The condition on S for the above solution to hold is simply $aS'/S \ll \omega^2/(a^2H)^2$.

C.3 Steady-state (inhomogeneous) solution for rapidly oscillating source

In this sub-section we solve for the steady-state solution of eq. (C.5) for a rapidly oscillating source using WKB approximation.

The steady-state solution, δ_{st}, being driven by external source, should primarily oscillate with frequency Ω_s. Correspondingly we assume δ_{st} of form

$$\delta_{st} = D_{st}(a) e^{i(\phi(a) + \int \Omega_s(a) da)}, \tag{C.12}$$

where ϕ is some phase difference which is varying slowly compared to oscillation frequency[1], $\phi' \ll \Omega_s$. Replacing the above form of δ_{st} back in eq. (C.5), assuming
$\Omega_s \gg D''_{st}/D_{st}, D'_{st}/(aD_{st}), \phi'$ and separating out the real and imaginary parts of the equation we obtain,

$$D_{st}\left(\frac{\omega^2(a)}{(a^2H)^2} - \Omega_s^2\right) = S(a)\cos(\phi) \tag{C.13}$$

$$\frac{1}{D_{st}(a^2H)^2 e^{\int n(a) d\ln(a)}} (D_{st}^2 \Omega_s (a^2H)^2 e^{\int n(a) d\ln(a)})' = -S(a)\sin(\phi). \tag{C.14}$$

The above two equations can further be simplified if $\phi \ll 1$. We can check that indeed $\phi \ll 1$ by dividing the two equations,

$$\tan(\phi) = -\frac{\Omega_s^2}{\omega^2/(a^2H)^2 - \Omega_s^2}\left[\frac{\left\{D_{st}^2(a^2H)^2 e^{\int n(a) d\ln(a)}\right\}' / \left\{D_{st}^2(a^2H)^2 e^{\int n(a) d\ln(a)}\right\}}{\Omega_s} + \frac{\Omega'_s/\Omega_s}{\Omega_s}\right]. \tag{C.15}$$

The quantities in the square bracket on LHS are very small under WKB approximation. Thus we have $\phi \ll 1$ as long as $\Omega_s \neq \omega/(a^2H)$.Using the above equation one can also verify our earlier assumption of $\phi' \ll \Omega_s$ is consistent.

Consequently, setting $\phi = 0$ in eq. (C.13) and eq. (C.12), we obtain

$$\delta_{st}(a) = \frac{S(a)}{\omega^2/(a^2H)^2 - \Omega_s^2} e^{i\int \Omega_s(a) da}. \tag{C.16}$$

In the limit of $\Omega_s \ll \omega/(a^2H)$, we recover the steady-state solution obtained for slowly varying S in eq. (C.11).

Since we do not have two species with same oscillating frequency, we do not solve the steady-state solution for the resonant case ($\Omega_s = \omega/(a^2H)$).

C.4 Application to cannibal and radiation perturbations

In this section we apply the formalism of steady-state and transient solutions developed in previous subsections, to explain the evolution of cannibal and radiation density perturbations.

The cannibal and radiation perturbations, described by eq. (C.3) and eq. (C.4) respectively, do not simultaneously feedback onto each other. For instance, when universe is SM radiation-dominated, terms with both ζ and $\rho_{can}/(\rho_{can} + \rho_r)$ are much smaller than one in eq. (C.4). Consequently, the feedback of δ_c on δ_r

[1]In reality the condition is $\phi'/\phi \ll \Omega_s$ but since $\phi < 2\pi$, the slowly varying condition is equivalent to $\phi' \ll \Omega_s$.

is suppressed. However, the feedback of δ_r to δ_c cannot be naively neglected (although we shall later prove it is negligible). Thus, in SM radiation-dominated universe, δ_r acts as an external source for δ_c. Similarly, in cannibal-dominated universe, δ_c acts as an external source for δ_r. Hence, we can describe the evolution of both cannibal and radiation perturbation using the generic equation given in eq. (C.5). During radiation (cannibal) domination we shall take $S = 0$ in radiation (cannibal) perturbation equation.

Given an equation of form eq. (C.5), the final solution will be determined by the steady-state solution only if the transient solution falls below the steady-state solution *and* if transient solution is more damped than steady-state solution. The latter condition is because at every moment of evolution of $\delta(a)$, a transient solution is generated with $a_* = a$ in eq. (C.10). For δ to follow steady-state solution, the transient solution must be suppressed compared to steady-state solution. In the scenarios where δ follows the steady-state solution, the evolution of δ is insensitive to the initial conditions. This insensitivity is because the steady-state solution only depends on the value of S at that instant and thus acts like an attractor solution.

We shall first focus on the transient solutions of δ_c and δ_r as they are independent of whether the universe is cannibal or radiation dominated. For radiation perturbations, comparing eq. (C.4) with eq. (C.5), we have $n = 2\zeta$ and $\omega = k/(3a^2H)$. Using eq. (C.10), we obtain the evolution of the amplitude of the transient solution of the radiation perturbation as

$$D_{r,t} = D_{r,t}(a_*)e^{-\int_{a_*}^{a}\zeta(a)d\ln(\tilde{a})}. \tag{C.17}$$

When universe is radiation-dominated and subsequently $\zeta = \Gamma\rho_{can}/H\rho_r \ll 1$, we recover the constant amplitude of oscillations of δ_r.

For cannibal perturbations we are primarily interested in the period when the mode is inside the Jeans scale where δ_c (eq. (C.3)) is well described by eq. (C.5) with $\omega = c_s k$ and $n = 1 - 3w_c$. Substituting this back in eq. (C.10), we find amplitude of the transient solution of cannibal perturbations as

$$D_{c,t} = D_{c,t}(a_*)\sqrt{\frac{c_s(a_*)}{c_s(a)}}e^{-\int_{a_*}^{a}(1-3w_c(\tilde{a}))d\ln(\tilde{a})/2}. \tag{C.18}$$

For slowly varying w_c, the above relation can be approximated as $D_{c,t} \propto c_s^{-1/2}a^{(1-3w_c)/2}$. Additionally, putting $\omega = c_s k$ in eq. (C.10), we can see that the phase of oscillation is given by kr_s, where r_s is the cannibal sound horizon (see eq. (4.67)). We use eq. (C.18) along with the phase being kr_s to describe the evolution of δ_c during SM radiation domination in eq. (4.68).

When analyzing δ_r in cannibal-dominated universe and vice-versa for δ_c, the source in the RHS of (C.3) (or eq. (C.4)) becomes important. Consequently, the final solution is determined either by transient or steady-state solution depending on which is larger and less damped.

In the case of δ_c in SM-radiation-dominated universe, radiation perturbation acts as an external source to the cannibal perturbations with $S = 1.5(1 + w_c)D_{r,t}/a^2$ and $\Omega_s = k/(3a^2H)$ in eq. (C.5). Consequently, using eq. (C.16) we find the amplitude of the steady-state solution of δ_c as

$$D_{c,st} = \frac{3}{2}\frac{(1+w_c)D_{r,t}(a)}{c_s^2 - 1/3}\frac{(aH)^2}{k^2} \propto a^{-2}. \tag{C.19}$$

Above we obtain the proportionality relation by approximating $w_c, c_s^2 \ll 1$ and using the fact that $D_{r,t}$ and a^2H are constant during radiation domination. One can easily see that the steady-state solution sourced by the radiation perturbations is suppressed by $(aH/k)^2$ and also more damped than the transient solution

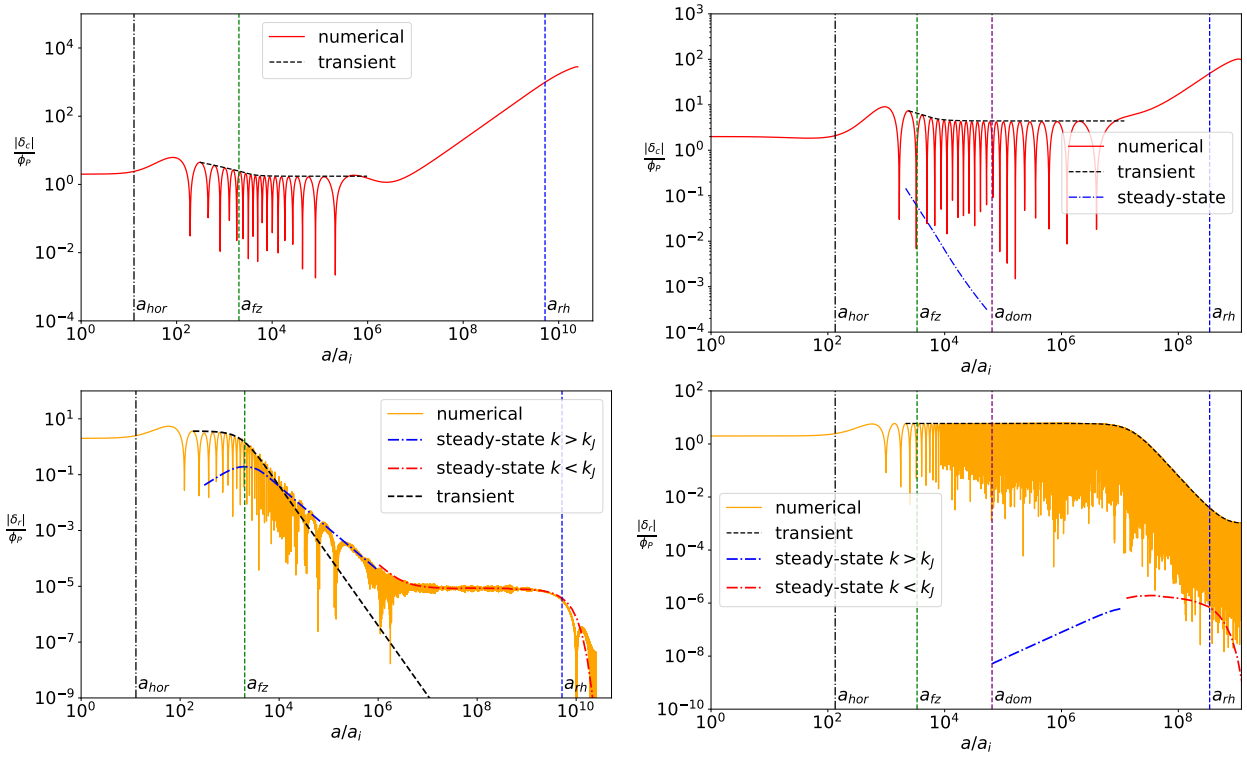


Figure C.1: Comparing the transient (black-dashed lines) and steady-state solutions (dot-dashed lines) obtained using WKB approximation with the numerical (solid lines) δ_c (top panels) and δ_r (bottom panels) evolution. The left (right) panels are for scenario shown in figure 4.5 (figure 4.7) where the universe is always cannibal-dominated (initially SM radiation-dominated). The amplitude of the transient solution are described by eq. (C.18) for cannibal and eq. (C.17) for radiation. The steady-state solution for cannibal is only valid when universe is radiation-dominated, $a < a_{dom}$, and its amplitude is described by eq. (C.19). Similarly the steady-state solution for radiation is only valid when universe is cannibal-dominated. The steady-state solution is broken into two parts: inside Jeans scale we plot its amplitude as described by eq. (C.20) (blue dot-dashed line) and outside Jeans scale we plot δ_r as described by eq. (C.21) (red dot-dashed line). The steady-state solution describes the density perturbation only when steady-state solution is larger than the transient solution and when the transient solution decays faster than the steady-state one.

in eq. (C.18). Thus deep inside the horizon, δ_r can never appreciably affect δ_c, irrespective of whether the universe is radiation or cannibal dominated. This is seen explicitly in the top panels of figure. C.1 where δ_c always follows the transient solution. In the top right panel, the steady-state solution (blue dot-dashed)— determined by δ_r during SM-radiation domination— is heavily damped and thus cannot overcome transient solution.

We break down the steady-state solution of δ_r into two components. The first one corresponds to the steady-state solution sourced by oscillating δ_c when the mode is inside the Jeans length, and the second one corresponds to the steady-state solution sourced by monotonically evolving δ_c when the mode escapes the Jeans length. For the first case, since δ_c is oscillating rapidly frequency $c_s k/a^2 H$), we have $a\delta_c' \approx (c_s k/aH)\delta_c \gg \delta_c$. Consequently the δ_c' term in the RHS of eq. (C.4) is the dominant source for δ_r. Correspondingly, the amplitude of δ_r can be described using eq. (C.16) with $S = (2+w_c)\zeta\Omega_s D_{c,t}/[(1+w_c)a]$ and $\Omega_s = c_s k/(a^2 H)$ giving,

$$D_{r,st} = \frac{1}{1/3 - c_s^2}\frac{(aH)}{k}\frac{2+w_c}{1+w_c}\zeta c_s D_{c,t} \propto \zeta\frac{c_s D_{c,t}}{a^{1/2}} \qquad k > k_J. \tag{C.20}$$

Above we obtain the proportionality relation by approximating $w_c, c_s^2 \ll 1$ and $H \propto a^{-3/2}$. In the bottom panels of figure C.1 we compare the numerical δ_r with $D_{r,st}$ (blue dot-dashed line) described above. For the case in left panel, ζ is small initially and correspondingly $D_{r,t}$ has constant amplitude. Once the decays in SM radiation bath become important, ζ becomes 5/2 and correspondingly $D_{r,t}$ decays. After $D_{r,t}$ falls below $D_{r,st}$, $D_{r,st}$ accurately describes the behaviour of δ_r while δ_c is oscillating. In the right panel, ζ is initially very small but a growing quantity. Due to this, $D_{r,st}$ is orders of magnitude smaller than the transient solution but is growing. Thus the evolution of δ_r is given by $D_{r,t}$ even after cannibal domination. The transient solution gets damped once cannibal decays in SM-radiation bath become important. The fact that escape from Jeans scale matches with the point where cannibal decays become important in SM-radiation bath (in the right plot) is due to pure coincidence of parameters. In both the panels we stop plotting $D_{r,st}$ once the mode exits the Jeans scale.

Next we analyse the steady-state solution of δ_r once the mode escapes the Jeans scale. In this case $a\delta_c' = \delta_c \propto a$. Correspondingly the source is determined by both δ_c' and δ_c terms in the RHS of eq. (C.4). We estimate S in eq. (C.11) by neglecting w_c and setting $a\delta_c' = \delta_c$ in RHS term of eq. (C.4). Doing so results in

$$\delta_{r,st} = \frac{(aH)^2}{k^2/3}\Big[\frac{\mathrm{d}\ln(a^2 H)}{\mathrm{d}\ln(a)}\zeta + \zeta^2 + \zeta + \frac{2\rho_{can}}{\rho_{can}+\rho_r}\Big]\delta_c \qquad k < k_J. \tag{C.21}$$

Note that $\delta_{r,st}$ is not an oscillating solution unlike the solutions we discussed earlier. Using the fact that $H \propto a^{-3/2}$, $\delta_c \propto a$ and $\zeta = 5/2$ before reheating, we can see that $\delta_{r,st}$ remains constant. Interestingly, δ_c sources δ_r via both the decays (ζ terms) and the metric ($\rho_{can}/(\rho_{can}+\rho_r)$ term). In the bottom left panel of figure C.1 we see that the behaviour of δ_r outside of Jeans scale is accurately explained by $\delta_{r,st}$ described above (red dot-dashed line). After reheating both ζ and ρ_{can} drop exponentially thereby leading to the decay of $\delta_{r,st}$ just after reheating. Once the decay of $\delta_{r,st}$ becomes larger than the decay of $D_{r,t}$, the transient solution takes over. Correspondingly δ_r oscillates with constant amplitude after reheating. In the bottom right panel, we find the $\delta_{r,st}$ to be orders of magnitude smaller than the homogeneous solution. However, had there been sufficient time between the decays becoming important in radiation bath and reheating, then $D_{r,t}$ would have eventually decayed below $\delta_{r,st}$. After which the steady-state solution would have governed the evolution of δ_r.

Finally, note that the effects of oscillating density perturbations on any other density perturbations, through gravitational coupling, can always be ignored deep inside the horizon. In the above discussion, we showed the negligible feedback of δ_r for cannibal perturbations using eq. (C.20). One can show that the steady-state solution imposed by oscillating δ_c on δ_{DM} will also be given by eq. (C.20), except with w_c and c_s^2 set to 0, the factor of $1/3$ replaced with c_s^2, and $D_{r,t}$ replaced by $D_{c,t}$. Again we can see that the steady-state solution imposed by cannibal on DM is suppressed by factors of $(aH)/k^2$. Had there been no cannibal decays into SM-radiation, $\zeta = 0$, the feedback of oscillating δ_c on δ_r would have also been negligible.

Appendix D

Separate universe principle and total density fluctuations

The separate universe principle (see, e.g., Refs. [229, 230]) posits that each super-horizon sized patch can be treated as an isolated, independent, Friedman-Robertson-Walker (FRW) universe. Any fluctuations on scales larger than the horizon simply become a part of the background variables of that island universe.

We begin by demonstrating that zero-mode fluctuations (Fourier mode $k = 0$) can be absorbed into the background quantities (density, pressure, etc.). The reason is straightforward. After fixing our coordinate system (fixing the gauge), there remains a residual coordinate transformation which can be used to absorb the fluctuations into a redefinition of the background. This can be seen explicitly as follows. We write the perturbed FRW metric as

$$[g_{\mu\nu}] = \begin{bmatrix} -(1+2\psi) & -a\partial_i\beta \\ -a\partial_i\beta & a^2[(1+2\phi)\delta_{ij} + 2\partial_i\partial_j\gamma] \end{bmatrix}. \tag{D.1}$$

Under an infinitesimal coordinate transformation (gauge transformation) $x^{\mu\prime} = x^\mu + \xi^\mu$, with $\xi_i = \partial_i\xi$, the scalar parts of the metric perturbations transform as

$$\begin{aligned} \psi' &= \psi - \dot{\xi}^0, & \beta' &= \beta - \frac{1}{a}\xi^0 + a\dot{\xi}, \\ \phi' &= \phi - H\xi^0, & \gamma' &= \gamma - \xi. \end{aligned} \tag{D.2}$$

The two scalar parts of the coordinate transformation, ξ^0 and ξ, can be used to set two of the four scalar perturbations in eq. (D.1) to zero. In conformal Newton gauge, ξ^0 and ξ are chosen to set $\beta = \gamma = 0$ and make the metric diagonal. However, because only gradients of β and γ appear in the metric, spatially uniform changes of coordinates leave the metric diagonal. Specifically, consider a transformation of the conformal Newton metric given by

$$\xi^0 = \epsilon(t) \qquad \xi = \omega x^i x^j \delta_{ij}, \tag{D.3}$$

where ω is a constant. The metric perturbations transform as

$$\psi' = \psi - \dot{\epsilon} \qquad \beta' = -\frac{1}{a}\epsilon(t), \tag{D.4}$$

$$\phi' = \phi - H\epsilon(t) \qquad \gamma' = -\omega x^i x^j \delta_{ij}. \tag{D.5}$$

Because only spatial derivatives of β and γ appear in the metric, the transformation in eq. (D.3) leaves the metric diagonal. The diagonal term from γ' can be absorbed into ϕ to give

$$\psi' = \psi - \dot{\epsilon}, \tag{D.6}$$

$$\phi' = \phi - H\epsilon(t) + \partial_i\partial_i\gamma'/3 = \phi - H\epsilon(t) - 2\omega. \tag{D.7}$$

The transformation as described in eq. (D.3) keeps the metric diagonal and is a residual gauge freedom for conformal Newton gauge. This residual gauge freedom in eqs. (D.6) and (D.7) can be used to remove the spatially homogeneous fluctuations in ϕ and ψ, and set the $k = 0$ fourier mode to zero, $\phi'_{k=0} = \psi'_{k=0} = 0$, giving the relations

$$H\epsilon + 2\omega = \phi_{k=0}, \qquad \dot{\epsilon} = \psi_{k=0}. \tag{D.8}$$

In the absence of metric perturbations, the Einstein equations imply that the total density perturbation vanishes. We can demonstrate this explicitly by noting that, under the residual gauge transformation, eq. (D.3), the density perturbation transforms as

$$\delta\rho' = \delta\rho - \dot{\rho}\epsilon = \frac{1}{4\pi G}(4\pi G\delta\rho - 3H\dot{H}\epsilon), \tag{D.9}$$

where in the second equality we used Friedmann equation for the background, $H^2 = 8\pi G\rho/3$. Further, because $\omega = \text{const.}$, eq. (D.8) gives us the relation

$$\dot{H}\epsilon = \dot{\phi}_{k=0} - H\psi_{k=0}. \tag{D.10}$$

Substituting this result into the gauge transformation given in eq. (D.9) for $k = 0$ mode, we obtain

$$\delta\rho'_{k=0} = \frac{1}{4\pi G}(4\pi G\delta\rho_{k=0} - 3H\dot{\phi}_{k=0} + 3H^2\psi_{k=0}). \tag{D.11}$$

The above vanishes identically after using the Einstein equation for $\delta\rho_{k=0}$. The residual gauge transformation in eq. (D.3) self-consistently removes all zero mode perturbations in the total density. The shift ϵ that is required to gauge away the perturbations is simply a uniform shift in coordinate time.

While the above analysis holds exactly for $k = 0$ mode, it can be extended to superhorizon modes up to corrections of order $O(k^2/(aH)^2)$. To see this, consider a Universe with a single superhorizon mode fluctuation, ϕ_k with $k \ll aH$. On a patch of the Universe with a scale sufficiently small compared to k^{-1}, but still large compared to the horizon, the fluctuation ϕ_k appears almost constant. If we consider two such patches separated by a distance comparable to k^{-1}, each patch samples a different approximately uniform value of ϕ_k. Consequently, in each patch, we may use the residual gauge freedom to remove this approximately constant ϕ_k. The approximately uniform shift in coordinate needed to make each patch uniform is different in each patch. Due to the equivalence principle, the only observable effects of such a shift enter at order

$O(k^2/(aH)^2)$.

D.1 Adiabatic vs isocurvature perturbations and residual gauge shifts

Above we showed that the total density fluctuations can be removed on super-horizon scales. However, if the Universe is filled with a multi-component fluid, the perturbations in each fluid species need not necessarily vanish after this procedure. In fact, only adiabatic perturbations in each species necessarily vanish, as we now demonstrate. Adiabatic perturbations between different species are related by

$$\frac{\delta\rho_1}{\dot{\rho}_1} = \frac{\delta\rho_2}{\dot{\rho}_2} = \frac{\delta\rho_3}{\dot{\rho}_3} = \ldots = \epsilon \tag{D.12}$$

where ϵ is the quantity obtained by solving eq. (D.8). We also show that isocurvature perturbations do not in general vanish, but can be absorbed into the background quantites within each patch. In a Universe with isocurvature perturbations, separate Universes evolve with both shifted clocks due to the background adiabatic fluctuations, but also spatially varying background densities, as we now demonstrate.

Consider now a universe comprised of two non-interacting perfect fluids, with densities ρ_1 and ρ_2. We suppose that on large scales $k^{-1} \gg (aH)^{-1}$, there exists density fluctuations in both species. We then consider a patch of the Universe small compared to k^{-1} but large compared to the horizon. In this patch, after the residual gauge shift with magnitude specified by eq. (D.8), the total density perturbation in this patch vanishes giving

$$\delta\rho'_{tot} = \delta\rho'_1 + \delta\rho'_2 = 0. \tag{D.13}$$

However, note that this only constrains the total density to vanish; it is not necessary for individual $\delta\rho'_i$ to also vanish. When they do, then using eq. (D.9), we see that the species satisfy eq. (D.12).

For an isocurvature mode, the density perturbations in individual species do not vanish after the gauge shift

$$\delta\rho'_1 = -\delta\rho'_2 \neq 0. \tag{D.14}$$

However, these isocurvature perturbations can be absorbed into the background variables. To see this explicitly we consider the zero-mode density-perturbation equation,

$$\delta\dot{\rho}_i + 3H(1+w_i)\delta\rho_i + 3\rho_i\dot{\phi} = 0, \tag{D.15}$$

where $i = 1, 2$ and w_i is the equation of state of the i-th perfect fluid. After gauge shift this becomes

$$\delta\dot{\rho}'_i + 3H(1+w_i)\delta\rho'_i = 0, \tag{D.16}$$

which can be trivially absorbed into the a redefinition of the background density

$$\tilde{\rho}_i = \rho_i + \delta\rho'_i. \tag{D.17}$$

This redefinition will not affect the metric or other Einstein variables as they only depend on total density in the universe, which remains unchanged after the absorption of isocurvature perturbations.