

COSMOLOGICAL ASPECTS OF NON-MINIMAL DARK SECTORS

by

Fei Huang

Copyright © Fei Huang 2019

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHYSICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY


In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Fei Huang titled *Cosmological Aspects of Non-Minimal Dark Sectors* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.



Keith R. Dienes

Date: April 5 2019



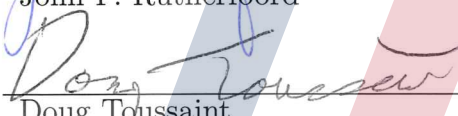
Shufang Su

Date: April 5 2019



John P. Rutherford

Date: April 5 2019



Doug Toussaint

Date: April 5 2019




Shufeng Zhang

Date: April 5 2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.




We hereby certify that we have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.



Keith R. Dienes
Professor
Physics

Date: April 5 2019



Shufang Su
Professor
Physics

Date: April 5 2019

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

Fei Huang

ACKNOWLEDGEMENTS

Pursuing a doctoral degree in physics is never easy. However, it can nevertheless be an enjoyable and rewarding experience. I feel very fortunate that I have been firmly supported by the people around me. Without them, my life as a PhD student would have been colorless. I would like to first thank my advisors KEITH DIENES and SHUFANG SU for their longstanding love and guidance. Keith has been a noble role model to me by showing his love of physics and his sense of responsibility for his students as well as the entire theoretical high-energy-physics community. I have enjoyed all of my conversations with him, on topics not only related to physics, but also those related to politics, culture, language, music and even personal life. The lessons that I have learnt from him span a huge spectrum. Shufang is the one who introduced me to the field of high-energy physics. She has taught me most of the fundamental techniques for my research and has been guiding me with her sharp instinct in physics. She is also a kind friend and a caring elder sister who has illuminated my career in physics with her encouragement, optimism and laughter. My sincere gratitude also goes to BROOKS THOMAS, who is a great collaborator, a heartfelt friend, and a teacher of highest virtue. I feel so lucky that he has literally been my third advisor. I would also like to thank DOUG TOUSSAINT for his wonderful lectures on quantum field theory, which essentially became the foundation of my research, JOHN RUTHERFOORD, for his insight which has inspired me to think more deeply about my research, and SHUFENG ZHANG for providing his interesting and valuable perspective as a knowledgeable physicist outside of high-energy physics. I cannot thank LUKA LESKOVIC enough for his numerous suggestions about coding, and YOUNGSOO PARK for his advice in solving Boltzmann equations. I have also tremendously benefited from the enlightening conversations with, and invaluable suggestions from, my officemates LUCIEN HEURTIER and DOOJIN KIM. My special thanks goes to JEFF KOST, who has been a perfect collaborator and also a mentor for me in the department, and HUANIAN ZHANG, who introduced me to the entire physics department and the city of Tucson when I first arrived six years ago, and has since then become my closest friend in the U.S. I would also like to thank my fellow graduate students ASHOORDIN ASHOORMARAM, PEDRO ESPINO, SOURATOSH KAHN, SHENGQIANG HUANG, SAMUEL LUK, HUAYANG SONG, OU ZHANG, SHULEI ZHANG, and YOU ZHOU. It is a privilege to have so many kind friends, and I hope our paths will cross again. I owe my deepest gratitude to my mother QIUBO HUANG for her eternal and unconditional love, and my late father ZIJIA ZENG who had been the spine of our family, who shared with me his wisdom, and taught me how to be a man. I deeply appreciate the trust of my father-in-law, ZHAOJUN LIU, and the dedication of my mother-in-law, HONGFANG LIU, who reached out to help and support my growing family. Finally, and most importantly, I would like to express my deep love and gratitude to my wife WEIWEI LIU, who sacrificed a lot to begin the journey with me and who encourages me when I need it the most, and my daughter JADE HUANG, who is

a gift to me from heaven and brings out the best in me.

DEDICATION

To the Memory of My Father Zijia Zeng

TABLE OF CONTENTS

LIST OF FIGURES	9
LIST OF TABLES	10
ABSTRACT	11
CHAPTER 1 INTRODUCTION	13
CHAPTER 2 COSMOLOGY BASICS	16
2.1 The Universe as We Observe Today	16
2.2 The Cosmic Timeline	20
2.3 Thermodynamics in Cosmology	25
CHAPTER 3 DARK MATTER BASICS	35
3.1 Observational Evidence	35
3.1.1 Why We Need Dark Matter	35
3.1.2 What Makes Dark Matter	38
3.2 Dark-Matter Production in the Early-Universe	41
3.2.1 Thermal Freeze-out	41
3.2.2 Misalignment Production	45
CHAPTER 4 DYNAMICAL DARK MATTER FROM STRONGLY-COUPLED DARK SECTORS	50
4.1 DDM ensembles of dark hadrons: Fundamental assumptions	55
4.1.1 The mass spectrum: Regge trajectories	56
4.1.2 Degeneracy of states: Exponential behavior	58
4.1.3 Physical interpretation of ensemble parameters	60
4.1.4 Constraints on parameters	65
4.2 Lifetimes and cosmological abundances for hadronic DDM ensembles	68
4.2.1 Cosmological abundances	69
4.2.2 Lifetimes and decays	75
4.3 Cosmological constraints on the dark-hadron ensemble	76
4.3.1 Total abundance, tower fraction, and effective equation of state	77
4.3.2 Cosmological constraints	80
4.4 Results	85
4.5 Conclusions	106

TABLE OF CONTENTS – *Continued*

CHAPTER 5	DECIPHERING THE ARCHAEOLOGICAL RECORD:	
	COSMOLOGICAL IMPRINTS OF NON-MINIMAL DARK SECTORS	112
5.1	From Intra-Ensemble Decays to Structure Formation	115
5.1.1	Packets to Packets, Dust to Dust: Evolution of Phase-Space Distributions	115
5.1.2	Structure Formation	131
5.1.2.1	Free-Streaming Horizon	135
5.1.2.2	Matter Power Spectrum	137
5.2	Toy Model	150
5.2.1	Parametrization of the Ensemble	150
5.2.2	Dynamical Quantities	158
5.2.3	Numerical Study: Initial Conditions	159
5.2.4	Numerical Study: Evolution Picture	163
5.2.5	Numerical Study: Matter Power Spectrum	172
5.3	Conclusion	177
CHAPTER 6	CONCLUSIONS	183
APPENDIX A	BOLTZMANN EQUATIONS	186
APPENDIX B	RELATIVISTIC KINEMATICS	189
B.1	Decay Term	189
B.2	Inverse Decay Term	193
B.3	Feed Term	195
REFERENCES	199

LIST OF FIGURES

FIGURE 2.1	Timeline in Standard Cosmology	24
FIGURE 2.2	Effective Number of Relativistic Degrees of Freedom	30
FIGURE 3.1	Thermal Freeze-Out	43
FIGURE 3.2	Misalignment Production	47
FIGURE 4.1	A Sketch for DDM	51
FIGURE 4.2	Internal consistency check	67
FIGURE 4.3	Exponentially Growing State Degeneracy	69
FIGURE 4.4	$r - s$ Plane Contours	87
FIGURE 4.5	Contours of $\eta(t_{\text{now}})$ and M_s	90
FIGURE 4.6	Contours of the Minimum Value of τ_0^{min}	92
FIGURE 4.7	$\Omega_{\text{tot}}(t)$, $w_{\text{eff}}(t)$ and $\eta(t)$	95
FIGURE 4.8	Level-by-Level Aggregate Cosmological Abundances	100
FIGURE 4.9	Aggregate Abundance Fractions $\hat{\Omega}_n(t_{\text{now}})/\Omega_{\text{tot}}(t_{\text{now}})$	101
FIGURE 4.10	Dark-Matter Pie Charts 1	104
FIGURE 4.11	Dark-Matter Pie Charts 2	105
FIGURE 5.1	Conveyor Belt: Deposits and Flow	119
FIGURE 5.2	Conveyor Belt: Producing Multi-Modal Distribution	122
FIGURE 5.3	Conveyor Belt: Deposit of Bricks	123
FIGURE 5.4	Examples: 3-State System	129
FIGURE 5.5	Free-Streaming Horizon	136
FIGURE 5.6	Effects on $P(k)$ from a single packet	140
FIGURE 5.7	Evolution of Different Perturbation Modes	141
FIGURE 5.8	Double Packets	143
FIGURE 5.9	“Single packet + CDM” vs “Double packets”	145
FIGURE 5.10	Multiple Packets	148
FIGURE 5.11	Total Decay Width	153
FIGURE 5.12	Partial Decay Widths	155
FIGURE 5.13	Decay Chains	156
FIGURE 5.14	Number Density	165
FIGURE 5.15	Energy Density	166
FIGURE 5.16	Tower Fraction	168
FIGURE 5.17	Equation of State Parameter	169
FIGURE 5.18	Evolution of Phase-Space Distribution	171
FIGURE 5.19	Present-Day Phase-Space Distribution	173
FIGURE 5.20	Matter Power Spectrum Ratios for Different Cases	175

LIST OF TABLES

TABLE 5.1	SM Temperatures during Intra-Ensemble Decays	174
-----------	--	-----

ABSTRACT

Many proposals for physics beyond the Standard Model do not give rise to only a single dark-matter candidate — they give rise to an entire dark sector consisting of many independent dark degrees of freedom. In this dissertation, we explore some of the cosmological implications of such non-minimal dark sectors. In the first part of this dissertation, we examine the phenomenology of dark sectors in which the density of dark states grows exponentially with mass. Ensembles of such states arise naturally as the “hadronic” resonances associated with the confining phase of a strongly-coupled dark sector; they also arise naturally as the gauge-neutral bulk states of Type I string theories. We study the dynamical properties of such ensembles, including their effective equations of state, and investigate some of the immediate model-independent observational (astrophysical and cosmological) constraints on such ensembles that follow. Remarkably, we find that these constraints allow such sectors to exhibit energy scales ranging from the GeV scale all the way to the Planck scale, but that the total present-day cosmological abundance of the dark sector must be spread across an increasing number of different states in the ensemble as these energy scales are dialed from the Planck scale down to the GeV scale. In the second part of this dissertation, by contrast, we examine the possibility of non-trivial dynamics within non-minimal dark sectors, focusing on processes in which heavier constituents within a non-minimal dark sector decay to lighter constituents. We begin by demonstrating that such decays can leave non-trivial imprints on the phase-space distribution of the resulting dark matter. Indeed, as a result of these effects, this phase-space distribution need not be thermal — it can even be multi-modal, with a non-trivial pattern of peaks and troughs as a function of momentum. We then proceed to study how these features can induce non-trivial changes in the shape of the resulting matter power spectrum. The results of this project therefore provide an interesting way of learning about (and

potentially even constraining) non-trivial dynamics in the early universe.

CHAPTER 1

INTRODUCTION

Particle physics, the study of the fundamental building blocks of the nature, has been a tremendous success in the past 100 years or so. The Standard Model (SM), whose last missing piece has been found by the discovery of the Higgs boson in 2012 [1, 2], is the “Rosetta Stone” of particle physics that elegantly summarizes our current understanding of all the fundamental interactions besides gravity and provides us a solid understanding of our world. However, there are still lots of unsolved fundamental problems. Some of them refer to finding a complete and consistent unification of all the theories, for example, the quest for a successful marriage between quantum mechanics and general relativity – quantum gravity. Some are related to the naturalness of the existing theories. For example, the hierarchy problem [3] questions why the quantum correction to the Higgs mass is quadratically sensitive to the ultraviolet cutoff while the Higgs mass is so small. Similarly, the strong CP problem concerns why the amount of CP violation in strong interactions is unexpectedly small as the experimental bound on the neutron electric dipole moment suggests the θ angle in the Lagrangian of the Quantum Chromodynamics (QCD), which controls the strength of the CP violation via strong interactions, is tiny, if not vanishing[4]. Others involve the unexplained experimental or observational evidence, *e.g.* the origin of the non-vanishing neutrino masses [5, 6], the nature of dark matter which provides gravity to form the structure we see today but does not seem to interact with light , and the nature of dark energy which drives the accelerated expansion of the universe in the current epoch [7, 8]. While many of these problems require new physics at microscopic scales, the limit of humans’ ability in constructing ground-based high-energy experiments severely constrains the search for physics beyond the Standard Model (BSM). Nevertheless, the universe, in which physics manifests itself at macroscopic scales, actually provides a rich playground for fundamental particle physics. Indeed,

the universe as we observe today is the result of the cosmological evolution from very early stage, and is in fact sensitive to the high-energy physical processes that have occurred in the early universe. Therefore, studying particle physics in the cosmological context allows us to decipher an enormous amount of information that is encoded in the history of the universe.

Indeed, various aspects of particle physics can be addressed in cosmology. For example, many cosmological parameters, such as the redshift of matter-radiation equality (MRE) and the present-day dark-matter energy density are sensitive to the sum of the neutrino masses, and therefore cosmological observations can be used to constrain the physics of neutrinos. In fact, the strongest constraint on the sum of the neutrino masses comes from cosmology [5]. Another example is the QCD axion [9–12], which is a field that dynamically drives the θ angle to zero and thus solves the strong CP problem. The QCD axion can be a natural candidate for dark matter, and thus can be constrained by the observed dark-matter relic abundance, as well as other cosmological and astrophysical observations, which, in turn, put stringent constraints on its parameter space as well as production mechanism. Likewise, many BSM theories which predicts one or more dark-matter candidate(s) receive constraints from dark-matter relic density, direct and indirect detections, or even structure formation.

Among a plethora of interesting topics that relate high-energy physics and cosmology, the focus of this dissertation is on dark matter. The existence of dark matter is strongly supported by astronomical observations, and it is widely believed that dark matter has a particle nature. Therefore, dark matter is one of the strongest evidence for new physics beyond the Standard Model. So far, there exist many models for dark-matter physics, and most of the traditional scenarios feature a single dark-matter candidate, *e.g.* a *weakly interacting massive particle* (WIMP). However, since the traditional paradigm is more and more disfavored by the null result from the direct and indirect experiments, we focus in this dissertation on non-traditional dark-matter scenarios. In particular, we are interested in scenarios involving non-minimal dark sectors, such as those that follow the framework of *dynamical-dark-matter* (DDM), in which the property of dark matter is not completely determined by any individual

particle alone.

The structure of the rest of the dissertation is as follows. In Chap. 2, we begin with a brief review on cosmology to establish a basic picture of the cosmic history, and the use thermodynamics in the universe. A general review on dark-matter physics is given in Chap. 3, where we discuss the evidence that supports the existence of dark matter and show some examples of dark-matter production mechanism. In Chap. 4, we present the first scenario, in which we construct DDM models from strongly-coupled dark sectors using the oscillator states in string theory or dark hadronic resonances in the confining phase of strongly-coupled gauge theories. We establish preliminary observational constraints, and show that a model like this is not only viable, but can also arise naturally to satisfy the DDM scaling relations. In chap. 5, we present another non-minimal scenario, in which the decays within the dark sector are able to produce non-trivial phase-space distribution for the lightest state in the dark sector. We show that the non-trivial phase-space distribution eventually leaves its imprints on the matter power spectrum at late time and provide a qualitative map between the features in the phase-space distribution and the matter power spectrum. In Chap. 6, we summarize the entire dissertation and discuss the outlook for future researches.

Note that, the work in Chap. 4 is already published in Refs. [13, 14], and the work in Chap. 5 is going to appear in Ref. [15].

CHAPTER 2

COSMOLOGY BASICS

In this chapter, we present a review on cosmology to introduce the background preliminaries and set up the stage for the later review on dark matter physics. We shall discuss some of the observational facts and the motivations for studying physics at the early epochs. Since it is a necessary tool, the application of thermodynamics in cosmology is also reviewed in detail.

2.1 The Universe as We Observe Today

In classical cosmology, the equation that governs the evolution of all the energy content in the universe is the Einstein's equation:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} , \quad (2.1)$$

where $R_{\mu\nu}$ and $R \equiv g^{\mu\nu}R_{\mu\nu}$ is the Ricci tensor and Ricci scalar, respectively, and $T_{\mu\nu}$ is the stress-energy tensor. From astronomical observations, we know that our universe today is homogeneous and isotropic on cosmological scales to a very good approximation [5]. The solution to the Einstein equation based on homogeneity and isotropy is the Friedmann-Robertson-Walker (FRW) metric:

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right] , \quad (2.2)$$

in which $a(t)$ is the *scale factor* that determines the physical distance between any two spatial coordinates *i.e.* $x_{\text{phys}} = a(t)x$, K is the spatial curvature constant which usually takes discrete values¹ $K = +1, 0, -1$ that corresponds to a closed, flat or open universe.

¹One can always rescale the radial coordinate r and the scale factor a to make K take these discrete values.

The curvature constant K is determined by the total energy density of the universe. To see this, let us assume that the stress-energy tensor takes the form of a perfect fluid with a time-dependent pressure and energy density $P(t)$ and $\rho(t)$:

$$T_{\mu}^{\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} . \quad (2.3)$$

Using this and the Einstein's equation, we can obtain two independent equations — the Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} , \quad (2.4)$$

where $H(t) \equiv (\dot{a}/a)$ is the Hubble parameter, and the Raychaudhuri equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) . \quad (2.5)$$

The continuity equation can be obtained from the two equations above:

$$\dot{\rho} + 3H(\rho + P) = 0 . \quad (2.6)$$

The continuity equation can also be recast in a form that shows the conservation of energy $d(a^3\rho) = -Pd(a^3)$ in a comoving volume a^3 . In fact, it can also be obtained from the conservation of energy momentum $\nabla_{\mu}T^{\mu}_{\nu} = 0$.

If we define the critical energy density

$$\rho_{\text{crit}} \equiv \frac{3H^2}{8\pi G} , \quad (2.7)$$

and the total abundance

$$\Omega(t) \equiv \rho(t)/\rho_{\text{crit}}(t) , \quad (2.8)$$

using Eq. (2.4), the curvature constant can be conveniently written as

$$\frac{K}{a^2} = H^2 (\Omega - 1) . \quad (2.9)$$

Therefore, the spatial geometry depends on whether or not the total energy density ρ is larger or smaller than, or equal to the critical energy density. If $\Omega > 1$, the universe is closed, while if $\Omega < 1$ the universe is open, and the universe is flat if $\Omega = 1$. Current observations show that our universe is flat to a very good precision with $\Omega = 1.0002 \pm 0.0026$ [6]. Therefore, for simplicity, we shall assume throughout this dissertation that our universe is precisely flat, *i.e.* $K = 0$.

Besides the spatial curvature, the evolution of the universe is also determined by the energy content in the universe. In a spatially flat expanding universe, if the energy density ρ scales as $a^{-\gamma}$ with $\gamma \geq 0$, Eq. (2.4) gives $\dot{a} \sim a^{1-\gamma/2}$. Solving this equation, one finds

$$a \sim \begin{cases} t^{2/\gamma} , & \text{if } \gamma > 0 ; \\ e^{Ht} , & \text{if } \gamma = 0 . \end{cases} \quad (2.10)$$

Therefore, the universe expands as a power law if $\gamma > 0$, and it expands exponentially if $\gamma = 0$. The way the energy density scales depends on the *equation of state parameter*

$$w \equiv \frac{P}{\rho} . \quad (2.11)$$

Inserting this into Eq. (2.6), one finds $\rho \sim a^{-3(1+w)}$, *i.e.* $\gamma = 3(1+w)$. The usual energy contents that we consider are dust, or in other words, matter ($w = 0$), radiation ($w = 1/3$) and vacuum energy ($w = -1$). It is obvious that a universe dominated by matter or radiation undergoes power-law expansion with $a \sim t^{2/3}$ or $a \sim t^{1/2}$, respectively. And if vacuum energy dominates, the universe would expand exponentially with $a \sim e^{Ht}$. Interested readers can also check the case where $\gamma < 0$ (or $w < -1$), in which the universe is filled with the so-called phantom energy for which the energy density increases as the universe expands. Phantom energy gives rise to an expansion rate even faster than the exponential expansion. In fact, the

scale factor a would diverge in finite time if phantom energy dominates, leading to the “Big Rip”, in which even the subatomic particles would be torn apart. Although interesting, phantom energy is beyond the scope of this dissertation, and therefore, we shall ignore it from now on. Note that, the Friedmann equation also admits contracting solutions. However, observations show that distant galaxies are receding from the Earth in a way that follows the Hubble’s law

$$v = H_0 D , \quad (2.12)$$

in which D is the physical distance between the Earth and a distant galaxy, $v = dD/dt$ is the receding velocity, and $H_0 > 0$ is the present-day Hubble parameter. This means that our universe is expanding, and therefore we shall not consider the contracting solutions. In addition to that, we also know that our universe is experiencing accelerated expansion, *i.e.* $\ddot{a} > 0$. From Eq. (2.5), we see that, to have accelerated expansion, one needs $w < -1/3$ — our universe is dominated by the type of energy content that produces negative pressure.

For concreteness, recent measurements [6] show that the universe today can be well described by the so-called Λ CDM model that basically consists of four types of energy: 1) dark energy, which can be either vacuum energy or simply a cosmological constant Λ in Einstein’s equation; 2) dark matter, which includes both cold dark matter (CDM) and non-relativistic neutrinos; 3) baryons, which include all the nuclei and electrons; 4) radiation, which consists of photons and relativistic neutrinos. The corresponding measurements on the present-day cosmological abundances, $\Omega_i \equiv \rho_i / \sum_j \rho_j$, that are pertinent to the types of energy listed above are [6]

$$\begin{aligned} \Omega_\Lambda &= 0.692 \pm 0.012 , \quad \Omega_{\text{CDM}} = 0.258 \pm 0.011 , \quad \Omega_b = 0.0484 \pm 0.0010 , \\ \Omega_\gamma &= (5.38 \pm 0.15) \times 10^{-5} , \quad 0.005 > \Omega_\nu \geq 0.0012 . \end{aligned} \quad (2.13)$$

Note that, instead of the total abundance of radiation, only the photon abundance is known to us. The neutrinos are pretty much completely non-relativistic today.

Though it is possible that the lightest flavor of neutrinos remains relativistic, its contribution to the total neutrino abundance would be negligible in this case. Surprisingly, our knowledge about the Standard Model only covers roughly 5% of the energy content (baryons, photons and neutrinos) in the universe while the other 95% is in forms of dark matter and dark energy which we know little of. Current observations find that the equation of state parameter for the dominant species, dark energy, is $w_{\text{DE}} = -1.01 \pm 0.04$ [16]. This is consistent with the simplest candidate, the cosmological constant Λ , which corresponds exactly to $w_{\Lambda} = -1$. Without further specification, we shall assume implicitly that dark energy is just the cosmological constant. For the present-day Hubble parameter, there is a recent tension between different measurements. Data from the *Cosmic Microwave Background* (CMB) measurements give $H_0 = 67.8 \pm 0.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ [5], whereas Type Ia Supernovae measurements suggest $H_0 = 73.24 \pm 1.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ [17]. Nevertheless, a discussion on potential causes of this tension is beyond the scope of this dissertation, and we shall take the value suggested by CMB measurements whenever needed.

2.2 The Cosmic Timeline

As the universe expands, different types of energy scale differently. Using the equation of state, Eq. (2.11), we see that energy density of matter scales as $\rho_m \sim a^{-3}$, the radiation energy density scales as $\rho_r \sim a^{-4}$, and vacuum energy has an energy density which is constant in time $\rho_{\Lambda} \sim \text{const.}$ The physical interpretation of these scaling behavior is straightforward. For matter, most of its energy is in its mass, therefore $\rho_m \approx nm$, where n is the particle number density and m is the mass of the corresponding particle. As the universe expands, the number density in a physical volume dilutes like a^{-3} , while the mass of the particle stays constant, which gives $\rho_m \sim a^{-3}$. For radiation, the number density scales in the same way as matter does, however, the energy of each particle is mostly in its momentum p , and the momentum redshifts as the universe expands. One easy trick is to use $p \sim 1/\lambda$, where λ is the particle's de Broglie wavelength. Since the de Broglie wavelength is a physical length, it is proportional to the scale factor, which gives $p \sim a^{-1}$. Bringing

all the things together, it is easy to see that radiation dilutes faster than matter, with $\rho_r \sim a^{-4}$. At last, for vacuum energy, the total vacuum energy inside a physical volume is proportional to the physical size of the volume, therefore the vacuum energy density stays a constant as the universe expands.

Clearly, the energy contents that dilutes faster as the universe expand will become less and less abundant in the future. However, if we evolve back in time, the energy contents that are not taking a significant share of the total abundance at the present epoch might play an important role or even dominate the universe in earlier epochs. For example, our universe is dominated by dark energy today, whereas matter is subdominant. However, in general, the ratio between the energy densities of matter (including both dark matter and baryons) and dark energy is

$$\frac{\rho_m(t)}{\rho_\Lambda(t)} = \frac{\rho_m(t_{\text{now}})}{\rho_\Lambda(t_{\text{now}})} \frac{a(t_{\text{now}})^3}{a(t)^3} . \quad (2.14)$$

Using $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ and the convention $a(t_{\text{now}}) = 1$, we can roughly estimate that, at $a \approx 0.75$, the energy densities of dark energy and matter are equal. That is, the universe only recently stepped into the dark-energy dominated phase, before that, it was dominated by matter. Similarly, using $\rho_m/\rho_r \sim a$, one can also estimate the value of the scale factor a when the energy density of radiation is equal to that of matter. This moment is usually referred to as *matter-radiation equality* (MRE), prior to which the universe was dominated by radiation. However, caution must be taken in the calculation since neutrinos which are non-relativistic today were relativistic and contribute to the radiation energy density at MRE. In fact, we know that $\rho_r \approx 1.68\rho_\gamma$ between neutrino decoupling and its transition from being relativistic to being non-relativistic. Putting things together, we have

$$1 = \frac{\rho_m(t_{\text{MRE}})}{\rho_r(t_{\text{MRE}})} \approx \frac{\rho_m(t_{\text{now}})}{\rho_\gamma(t_{\text{now}})} \frac{a_{\text{eq}}}{1.68} , \quad (2.15)$$

where $a_{\text{eq}} \equiv a(t_{\text{MRE}})$. A rough estimate with $\Omega_\gamma \approx 5 \times 10^{-5}$ and $\Omega_m \approx 0.3$, gives $a(t_{\text{MRE}}) \approx 2.8 \times 10^{-4} \sim 1/3500$.

When discussing the history of the universe, it is usually more convenient to

use the photon temperature T instead of the cosmological time t as the ticks on the timeline, since $T \sim a^{-1}$. The CMB temperature today is measured to be $T = 2.7255 \text{ K} \approx 0.234 \text{ meV}$ [5]. Before our universe is dominated by dark energy, there was a *matter-dominated* (MD) epoch which extends itself back to $T \sim \text{eV}$. The formation of galaxies and large scale structure all occurred in the MD epoch because the primordial density fluctuations were able to be amplified by gravity and start to grow in the MD epoch. Neutral atoms were also formed in the MD epoch, since when the temperature dropped below the eV scale, photons in the universe were no longer energetic enough to ionize the neutral atoms. By the time neutral atoms formed, photons decoupled from the rest of the cosmic plasma and became the CMB we observe today. Before the MD epoch, there was the *radiation-dominated* (RD) epoch. In the standard Big Bang picture, high-energy physical processes such as the Electro-Weak phase-transition ($T \sim \mathcal{O}(100) \text{ GeV}$), the QCD phase-transition ($T \sim \mathcal{O}(200) \text{ MeV}$), and the *Big Bang nucleosynthesis* (BBN, $T \sim \text{MeV}$) are all assumed to take place within the RD epoch. However, it is usually not emphasized that current observations only provide direct probe of the cosmic history up to the BBN epoch. In some non-standard scenarios, it is even possible to insert an additional *early matter-dominated* (EMD) epoch into the RD epoch before BBN [18–21]. In fact, any cosmological model that successfully reproduces all the physics since the BBN epoch is consistent with current observations.

Prior to the RD epoch, it is usually believed that there exists an epoch at an energy scale of $\mathcal{O}(10^{13}) \text{ GeV}$ in which the universe is completely dominated by vacuum energy and undergoes rapid exponential expansion. This epoch is called the cosmic *inflation*. Inflation is strongly motivated by the observed uniformity of CMB. On one hand, other than the observed almost perfect blackbody radiation with the 2.7255 K average temperature, the observed temperature fluctuations in CMB which is related to the density fluctuations in the early universe is as small as 1 part in 10^4 . On the other hand, the CMB photons that we see are coming from $\sim 10^4$ patches of space that were not causally connected at the time of recombination. It is a mystery that these regions which appear to have never talked to each other produce photons with

almost the same temperature. This is often referred to as the *horizon problem*. To solve this problem, the inflation mechanism is invoked so that those regions, that were not causally connected at the time of recombination, had been in causal contact at earlier times and therefore were able to smooth out quantum fluctuations before inflation sent them apart. After the end of inflation, the inflaton, which is the field that drives inflation, decays into particles which eventually thermalize and become the radiation bath of the universe. This epoch is usually referred to as the *reheating* phase. The primordial quantum fluctuations of the inflaton field thus become the density fluctuations in radiation, which eventually become the tiny temperature fluctuations in the CMB. Despite solving the horizon problem, inflation can also solve problems such as the *flatness problem* and *magnetic-monopole problem*, etc [22]. The observed almost scale-invariant primordial scalar power spectrum is a strong support for the existence of such an inflationary epoch.

One can further explore the history of the universe beyond the inflationary epoch. Naively, as the universe gets more and more dense, when evolving back in time, we will eventually hit the *big-bang singularity* at which the density diverges and the space-time ceases to exist. The big-bang singularity is usually thought as a sign that general relativity fails to describe the physics at such a high scale. The most common solution to the big-bang singularity problem is that, as we approach the Planck scale, quantum effects become important, and instead of using general relativity, one has to resort to *quantum gravity*. However, despite the pure theoretical efforts such as *string theory* [23–26] and *loop quantum gravity* [27–29], there is no observational evidence for any theory of quantum gravity.

To conclude this section, we sketch in Fig. 2.1 the timeline of the standard cosmic history and show the way the abundances of different species evolve.² Since we do not have a testable theory of quantum gravity yet, we start our storyline at the epoch of the cosmic inflation. After the end of the inflation, the inflaton starts to oscillate in the vacuum before it decays. The decay of the inflaton produces high-energy radiation which then thermalizes and reheats the universe, so the universe heads into the RD

²Of course, the “standard” picture can also vary between different physicists.

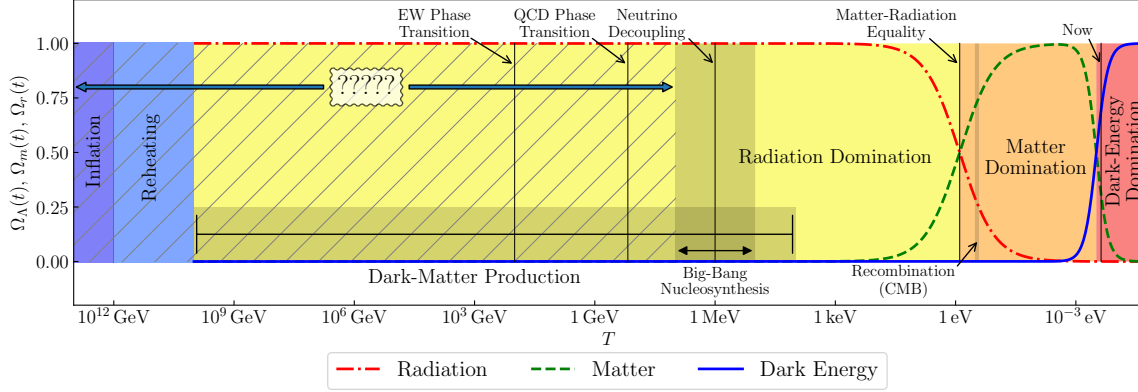


FIGURE 2.1: The cosmic timeline in the standard cosmology. The horizontal axis shows the temperature of the SM thermal bath after reheating, for which we assume that there is no additional degree of freedom beyond what exist already in the Standard Model. Before the end of reheating, the axis only suggests the energy scales. The red, green and blue curves represent the cosmological abundances of radiation, matter and dark energy, respectively. The purple, blue, yellow, orange and red shaded regions correspond to the epochs of inflation, reheating, radiation domination, matter domination and the dark-energy domination. Some of the important moments are labeled by the vertical lines. Events like BBN and recombination are shaded since they last for a small range of temperature. Since we do not yet know how dark matter is produced in the early universe, the wide horizontal region that is shaded darker suggests that dark matter can be produced at any scale in this region. The vertical regions that are shaded darker correspond to the time intervals in which the BBN and the recombination occur, since such events last for a non-negligible amount of time. Finally, the hatching and the question marks suggest that we do not have direct probes of such epochs, therefore, the existence of such epochs is hypothetical. Alternative cosmologies could modify this part of cosmic history as long as they successfully reproduce the rest of the history.

epoch. Since the energy scale of the RD epoch can be very high, it is an ideal test ground for many BSM models. For example, the dark-matter production is assumed to take place within the RD epoch. Depending on the scenarios in consideration, the dark-matter particles can be produced right after the decay of the inflation, or down to a few keV. This is indicated by the horizontal shaded region. As the temperature of the radiation drops, the SM physics eventually comes into play, and physical processes such as the electroweak phase transition, the QCD phase transition, the decoupling of neutrinos from the rest of cosmic plasma and the production of primordial light elements occur subsequently. Since radiation dilutes faster than matter, as the temperature drops further, the energy density of the universe starts to be dominated by matter. In the MD epoch, neutral atoms start to form, and

photons free-stream and become the CMB we observe today. The primordial density perturbations also start to grow within the MD epoch and form galaxies, clusters and the large-scale structure. Since both the matter and the radiation energy densities decrease as the universe expands, the dark energy whose energy density stays constant eventually dominates the universe and drives the accelerated expansion of the universe today. If the dark energy is indeed described by the cosmological constant, the universe would eventually be completely dominated by the dark energy and expand forever.

Once again, we emphasize that this storyline corresponds to the standard cosmic history, and therefore the existence of the epochs before the BBN is only hypothetical (the hatched regions). In general, a non-standard cosmic history can significantly alter the phenomenological predictions of many BSM models.

2.3 Thermodynamics in Cosmology

The previous subsections briefly reviewed some of the current observations on our universe and provide us with a rough sense of the basic facts as well as the evolution picture. To study the evolution of the universe in more detail, especially at early epochs when the universe was hot and dense, and was filled with a large number of different species, tools from thermodynamics and statistical mechanics are necessary.

Let us start with the phase-space distribution. When we study a particular particle species in the universe, we usually care about its number density, energy density as well as the pressure it produces. The information about these quantities is all contained in the distribution function $f(p, t)$ which gives the number of such particles occupying a phase-space volume element, $f(p, t) d^3x d^3p$. Through the phase-space distribution, the number density n , energy density ρ and pressure P can be defined

as follows:

$$n(t) \equiv g \int \frac{d^3p}{(2\pi)^3} f(p, t) , \quad (2.16)$$

$$\rho(t) \equiv g \int \frac{d^3p}{(2\pi)^3} E f(p, t) , \quad (2.17)$$

$$P(t) \equiv g \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} f(p, t) , \quad (2.18)$$

where $E = \sqrt{p^2 + m^2}$, and g is the number of internal degrees of freedom, such as spin and color, of the particle in consideration. Note that, the pressure defined in such a way is appropriate for particles. The integrand $p^2/(3E)f(p, t)$ properly counts the flow of momentum carried by particles through an infinitesimal surface.³ Therefore, it is not only true for collisional particles, but also valid for collisionless particles. Straightforwardly, the equation of state parameter takes the same old form $w(t) \equiv P(t)/\rho(t)$.

In the early universe, the whole universe was in a hot dense state, and lots of species in the primordial particle soup were in equilibrium. There are three different types of equilibrium — *kinetic*, *chemical* and *thermal* equilibrium, which refer to different conditions, respectively. For kinetic equilibrium, particles can quickly scatter and exchange energy momentum with the others. The phase-space distribution of any species in kinetic equilibrium (with other species or even with itself) takes the form of the Bose-Einstein or the Fermi-Dirac distribution:

$$f(p, t) = \frac{1}{\exp[(E - \mu)/T] \pm 1} , \quad (2.19)$$

where T is the temperature, μ is the chemical potential, and “ $-/+$ ” are for bosons and fermions respectively. Chemical equilibrium refers to a steady state of reactions. For a general process $a + b + \dots \leftrightarrow i + j + \dots$, the rates of the reaction in both directions are the same if the species involved are in chemical equilibrium. This

³Interestingly, the factor of 3 in the denominator comes from the number of spatial dimensions.

imposes a useful condition:⁴

$$\mu_a + \mu_b + \dots = \mu_i + \mu_j + \dots . \quad (2.20)$$

For example, for a process $a + b \leftrightarrow \gamma + \gamma$, since we know that the photon chemical potential $\mu_\gamma = 0$, chemical equilibrium implies $\mu_a + \mu_b = 2\mu_\gamma = 0$. Thermal equilibrium is a much stronger condition. It implies both kinetic and chemical equilibrium between and adds to them the condition that all the species share the same temperature T .

Since as the universe expands, particles redshift, we expect the temperature of the cosmic thermal bath to cool down as time goes. For the temperature of an entire radiation bath in equilibrium, its actual evolution is complicated by the change of degrees of freedom which we will discuss later. Here, let us look at what would happen if a species had been in thermal contact with the thermal bath, but has decoupled from the rest of the thermal bath at a certain point. Let us first choose a phase-space volume element $\Delta x^3 \Delta p^3$. As we have mentioned in Sec. 2.2, the physical coordinates $x \sim a$, while momentum $p \sim 1/a$ in an FRW universe, for which we can formally write

$$x(t) = \frac{a(t)}{a_i} x_i , \quad (2.21)$$

$$p(t) = \frac{a_i}{a(t)} p_i . \quad (2.22)$$

Here the subscript “ i ” stands for values at an arbitrary initial time t_i . Obviously the size of the phase-space volume element $\Delta x(t)^3 \Delta p(t)^3$ stays unchanged. The number of particles reside in this volume⁵, $f(p(t), t) \Delta x(t)^3 \Delta p(t)^3$, should also stay constant if the particles are in equilibrium. Therefore, we have $f(p(t), t) = f(p_i, t_i)$. The condition for kinetic equilibrium implies

$$\frac{E(t) - \mu(t)}{T(t)} = \frac{E_i - \mu_i}{T_i} . \quad (2.23)$$

⁴It is usually not mentioned that a and b , or i and j can be the same species.

⁵We are sloppy here in neglecting factors like $1/(2\pi)^3$

In the relativistic limit we can replace E with p . To keep the p/T part constant, we can define the effective temperature $T(t) = (a_i/a(t))T_i$. With this definition, the evolution of the chemical potential is forced to be

$$\mu(t) = \frac{T(t)}{T_i} \mu_i . \quad (2.24)$$

This definition is reasonable as we see the zero chemical potential for photon stays zero. For the non-relativistic limit, $E \simeq m + p^2/(2m)$, and thus we can similarly define $T(t) = (a_i/a(t))^2 T_i$. Eq. (2.23) therefore implies

$$\mu(t) = m - (m - \mu_i) \frac{T(t)}{T_i} . \quad (2.25)$$

To see why this is reasonable, one can check that in the limit $m - \mu \gg T$, Eq. (2.19) becomes the Maxwell-Boltzmann distribution

$$f(p, t) = \exp\left(-\frac{m - \mu(t)}{T(t)}\right) \exp\left(-\frac{p^2}{2mT(t)}\right) , \quad (2.26)$$

which is as expected. Therefore, we conclude here that the temperature of a decoupled species scales as a^{-1} if the species is relativistic, while it scales as a^{-2} if the species is non-relativistic.

Let us use the distributions in Eq. (2.19) to perform the integrals in Eqs. (5.1)–(5.4). We are interested in several limiting cases. For relativistic species ($T \gg m$), if $T \gg \mu$,

$$n \simeq \begin{cases} (\zeta(3)/\pi^2) gT^3 & \text{for boson} \\ (3/4) (\zeta(3)/\pi^2) gT^3 & \text{for fermion} \end{cases} , \quad (2.27)$$

$$\rho \simeq \begin{cases} (\pi^2/30) gT^4 & \text{for boson} \\ (7/8) (\pi^2/30) gT^4 & \text{for fermion} \end{cases} , \quad (2.28)$$

$$P \simeq \rho/3 , \quad (2.29)$$

where $\zeta(3) = 1.202\dots$ is the Riemann zeta function of 3. For non-relativistic species

$$(T \ll m),$$

$$n \simeq g \left(\frac{mT}{2\pi} \right)^{3/2} \exp[-(m - \mu)/T] , \quad (2.30)$$

$$\rho \simeq nm + \frac{3}{2}nT , \quad (2.31)$$

$$P \simeq nT . \quad (2.32)$$

We see that in this limit, the energy density is mostly from particles' rest mass, and pressure is negligible. The above equations confirm what we mentioned in the previous chapter that ultra-relativistic species has $w \approx 1/3$ and $\rho \propto a^{-4}$, while non-relativistic species has $w \approx 0$ and $\rho \propto a^{-3}$.

For the purpose of studying dark matter in the early universe, we are interested in the RD epoch in which the total energy density $\rho \sim \rho_r$. Using Eq. (2.28), we can sum all the energy density from radiation

$$\begin{aligned} \rho_r &= \sum_i \rho_{r,i} \\ &= \frac{\pi^2}{30} T^4 \times \left(\sum_i^{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_i^{\text{fermions}} g_i \left(\frac{T_i}{T} \right)^4 \right) \\ &= \frac{\pi^2}{30} g_*(T) T^4 , \end{aligned} \quad (2.33)$$

in which T_i is the temperature of each individual species, T is a reference temperature which is usually chosen to be the photon temperature, and $g_*(T)$ is given by the entire expression in the big brackets in the second line and is called the *effective number of relativistic degrees of freedom*. In principle, if we assume that g_* contains only contributions from the standard model, given the fact that we can calculate when the SM species are in thermal equilibrium and when they decouple, g_* is countable. A fit for g_* is shown in Fig. 2.2.

Using the Friedmann equation $3M_{\text{P}}^2 H^2 = \rho \simeq \rho_r$, where $M_{\text{P}} = 1/\sqrt{8\pi G}$ is the

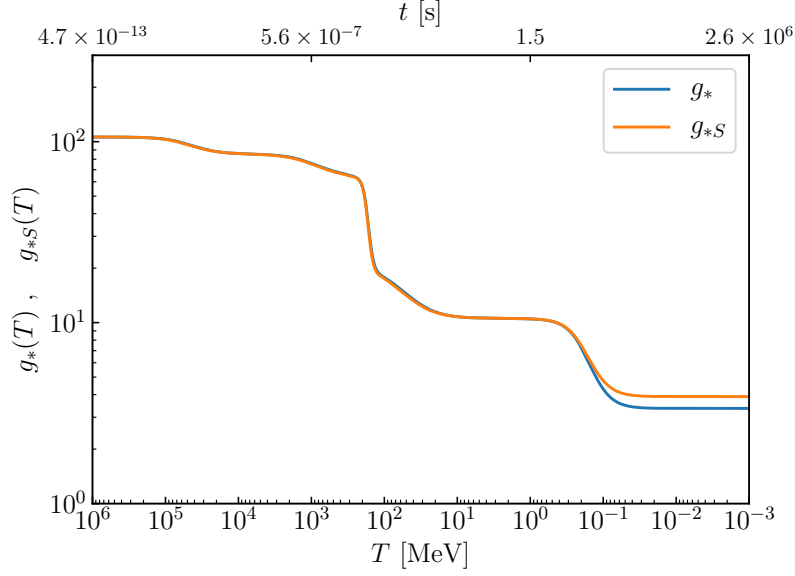


FIGURE 2.2: Effective number of relativistic degrees of freedom for energy density $g_*(T)$, and for entropy density $g_{*S}(T)$ using the analytical fit from Ref. [30]. The bottom axis shows temperature while the upper axis marks the cosmological time t at $T = 1$ keV, 1 MeV, 1 GeV, 1 TeV using the time-temperature relation Eq. (2.35).

reduced Planck mass, we have

$$H \simeq \frac{\pi}{\sqrt{90}} \frac{g_*^{1/2} T^2}{M_{\text{P}}} . \quad (2.34)$$

Since $a \sim t^{1/2}$ and thus $H \simeq 1/(2t)$, we obtain the useful *time-temperature relation*:

$$t \simeq \frac{\sqrt{90}}{2\pi} \frac{M_{\text{P}}}{g_*^{1/2} T^2} \simeq 4.85 g_*^{-1/2} \left(\frac{\text{MeV}}{T} \right)^2 \text{ s} . \quad (2.35)$$

To have a better understanding of this relation, we mark out the values of t for several typical temperature scales $T = 1$ keV, 1 MeV, 1 GeV and 1 TeV in the upper axis of Fig. 2.2, assuming all the relativistic degrees of freedom comes from the Standard Model.

As promised in previous paragraphs, we now discuss the evolution of temperature for the species in equilibrium. We begin with the second law of thermal dynamics in

the expanding universe:

$$TdS = d(\rho V) + PdV - \sum_i \mu_i d(n_i V) , \quad (2.36)$$

in which $V = a^3$ is the physical size of a unit comoving volume, ρ and p are the total energy density and pressure of the universe, and we have explicitly written down the contribution of chemical potential from all the species. If the entire universe is in thermal equilibrium or consists of decoupled sectors that each has its own thermal equilibrium, chemical equilibrium implies that the last term vanishes. The reason is simple. For a single decoupled species which equilibrates with itself, if all the reactions among its own particles conserve particle number, $d(n_i V) = 0$; while if particle-number non-conserving processes exist, $\mu_i = 0$. For species in thermal equilibrium, for example, through a process $a + b + \dots \leftrightarrow i + j + \dots$, the moment each species a, b, \dots on the left side of the reaction increase by a small amount dn , the species i, j, \dots on the other side of the reaction equation would decrease by the same amount dn , making the sum $\sum_i \mu_i d(n_i V) = 0$. In the following, we shall therefore ignore the last term in Eq. (2.36). Writing the factors of a explicitly, we have

$$\begin{aligned} TdS &= d(\rho a^3) + p d(a^3) \\ &= a^3 \left(\frac{d\rho}{dt} + 3 \frac{(\rho + P)}{a} \frac{da}{dt} \right) dt . \end{aligned} \quad (2.37)$$

From Eq. (2.6), the conservation of energy in an expanding universe, we see the terms in the big brackets add up to zero. Therefore,

$$\frac{dS}{dt} = 0 . \quad (2.38)$$

We thus arrive at the important conclusion: the entropy is conserved if the entire universe is in thermal equilibrium or if it is made up from decoupled sectors that is in their own thermal equilibrium.

It is often useful to define the entropy density

$$s \equiv \frac{S}{a^3} . \quad (2.39)$$

Entropy conservation thus implies that $s \sim a^{-3}$. To be able to actually calculate the entropy density, we first rewrite Eq. (2.36) as

$$\begin{aligned} dS &= \frac{\rho + P}{T} dV + V \frac{d\rho}{T} \\ &= d\left(\frac{\rho + P}{T} V\right) - \frac{dP}{T} + \frac{\rho + P}{T^2} dT . \end{aligned} \quad (2.40)$$

The next step is to find out the relation between dP and dT , and hope that the relevant terms cancel. Since the energy density is only a function of T in thermal equilibrium, we have

$$\left(\frac{\partial S}{\partial V}\right)_T = \frac{\rho + P}{T} , \quad \left(\frac{\partial S}{\partial T}\right)_V = \frac{V}{T} \frac{d\rho}{dT} . \quad (2.41)$$

Using the integrability condition

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T} , \quad (2.42)$$

we find

$$dP = - \frac{\rho + P}{T} dT , \quad (2.43)$$

which gives

$$dS = d\left(\frac{\rho + P}{T} V\right) . \quad (2.44)$$

With the definition in Eq. (2.39), we eventually arrive at

$$s = \frac{\rho + P}{T} , \quad (2.45)$$

in which we have ignored the integration constant. From Eqs. (2.27)-(2.32), we have seen that the energy density of any species in equilibrium is exponentially small if the

species is non-relativistic. Therefore, most of the contribution to the total entropy comes from relativistic species. Using Eq. (2.28) and Eq. (2.29), we find, to a very good approximation,

$$s = \frac{2\pi^2}{45} g_{*S} T^3, \quad (2.46)$$

in which the *effective number of degrees of freedom for entropy* is defined as

$$g_{*S} \equiv \sum_i^{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_i^{\text{fermions}} g_i \left(\frac{T_i}{T} \right)^3. \quad (2.47)$$

Like g_* , g_{*S} is also countable. A fit for g_{*S} is also shown in Fig. 2.2.

With entropy conservation $sa^3 = \text{const.}$, we see the temperature of the entire thermal bath goes as $T \sim g_{*S}^{-1/3} a^{-1}$. When g_{*S} stays constant, T drops regularly as a^{-1} . When g_{*S} drops as certain species become non-relativistic, the entropy of that species is transferred to the thermal bath in contact with it, therefore making the temperature T decrease slower than a^{-1} .

Besides the evolution of temperature, entropy conservation is often used in calculating the comoving number density. A frequently used quantity is called the *yield*:

$$Y \equiv \frac{n}{s}. \quad (2.48)$$

The yield is also sometimes referred to as the comoving number density since it is proportional to the normal comoving number density na^3 when entropy is conserved. Since the entropy can be directly calculated from temperature, it is very convenient to relate the yield in the early universe to its observed value in the late-time universe. A famous example is the calculation of dark-matter relic abundance in the thermal *freeze-out* scenario. A successful freeze-out which produces the correct dark-matter relic density must have the yield at freeze-out equal to the yield today

$$Y_{\text{fo}} = Y_{\text{now}}. \quad (2.49)$$

Since, Y_{now} can be inferred from the observed entropy density and dark-matter energy

density, the condition above immediately put constraints on the associated dark-matter model. We shall discuss the thermal freeze-out mechanism in more detail in Chap. 3.

CHAPTER 3

DARK MATTER BASICS

In this chapter, we shall review the basic observational evidence that supports the existence of dark matter. We shall see that the existence of a matter content on astrophysical and cosmological scales that has negligible interactions with light is perhaps one of the strongest evidence for physics beyond the Standard Model. Since dark matter might have been produced in the early universe, we shall discuss its production mechanism and see how processes in the early universe and the observations in the present-day universe are related. In particular, we shall present two typical scenarios, thermal freeze-out and misalignment production, which represent two major classes of dark-matter production mechanisms — thermal production and non-thermal production, respectively.

3.1 Observational Evidence

3.1.1 Why We Need Dark Matter

The observational evidence that supports the existence of dark matter comes from many different sides, such as the measurements of *mass-to-light ratios*, *galactic rotation curves*, *bullet cluster*, CMB and *large scale structure*. All of these pieces of evidence can essentially be explained by the existence of a matter component that interacts gravitationally but has a negligible electromagnetic interaction.

The earliest suggestion comes from the measurement of the radial velocity of galaxies in the *Coma cluster* by Zwicky in 1930s [31, 32]. Since the total mass can be estimated using the virial theorem, while visible mass can be inferred from the amount of light emitted, it was shown that the mass-to-light ratio is unexpectedly high, indicating that the system would not be gravitationally bound if all the mass in the system is visible.

The most commonly mentioned observational evidence for the presence of dark matter is perhaps from the measurements of galactic rotation curves, pioneered by Rubin in 1970s [33]. For simplicity, let us assume that a galaxy has a spherically symmetric density profile $\rho(r)$. Then the total mass within a radius r is $M(r) = \int_0^r dr \, 4\pi r^2 \rho(r)$. Using Newtonian mechanics, an object rotating around the center of the galaxy at a radial distance r should have a velocity that satisfies

$$\frac{v(r)^2}{r} = \frac{GM(r)}{r^2}. \quad (3.1)$$

Now let us assume that the visible part of the galaxy has a radius R . If the visible part contains most of the mass of the galaxy, beyond the radius R , $\rho(r) \approx 0$, and the total mass $M(r) \sim \text{const}$. Therefore, when looking at a galaxy, it is expected that the rotational velocity $v(r) \sim r^{-1/2}$ beyond the visible part. However, in astronomical observations, instead of seeing power-law falling tails, the observed rotation curves generally flatten out well beyond the visible galactic disks. This means, the objects in the outer region of a galaxy are moving as fast as those in the inner region. Since the mass of the visible part of a galaxy can be well measured, it is found that those objects would have already escaped if the total mass of the galaxy is mostly from the visible mass. Therefore, additional invisible mass has to exist to make the system gravitationally bound. One useful exercise is to find a density distribution that give rise to a flat rotation curve at large radius. From Eq. (3.1), we see if additional mass with $\rho(r) \sim 1/r^2$ exists for $r > R$, then $M(r) \sim r$, and the rotational velocity becomes a constant at large radius.

The galactic rotation curves seem to be an indisputable support for the existence of dark matter. However, there exist theories that are able to fit the rotation curves by modifying gravity rather than introducing dark matter. Those theories are called the *Modified Newtonian Dynamics* (MOND) [34, 35]. The basic idea of MOND is that Newton's second law $F = ma$ is not correct when the acceleration is small.

Instead, the gravitational acceleration of a test particle due to a mass M is

$$a = \begin{cases} a_N, & \text{for } a \gg a_c, \\ \sqrt{a_c a_N}, & \text{for } a \ll a_c, \end{cases} \quad (3.2)$$

where $a_N \equiv GM/r^2$ is the Newtonian acceleration, and the critical acceleration is interestingly related to the speed of light and the Hubble parameter $a_c \sim cH_0/(2\pi) \sim 1.2 \times 10^{-8} \text{ cm/s}^2$. Obviously, this is able to reproduce the flat rotation curve at large radius since $a \sim 1/r$ instead of $1/r^2$ when acceleration is small. In fact, MOND fits the galactic rotation curves very well.

To find a better theory between dark matter and MOND, additional complementary observations are necessary. The observation of Bullet Cluster provides evidence from a different perspective. The Bullet Cluster comprises two merging galaxy clusters. Naively, the region where two clusters are colliding is expected to have a higher density both in mass and in the amount of hot gas of baryons. The hot gas emits X-ray through bremsstrahlung, therefore we can observe and infer the mass distribution of the baryons in the system. Except for the mass of the hot gas, the total mass can be obtained through gravitational lensing. Surprisingly, observations on the Bullet Cluster show a clear separation between the regions of highest mass density and the regions of highest baryons density. MOND cannot easily explain the offset of these regions, if not entirely impossible. On the contrary, explaining it using dark-matter interpretation is straightforward — while the baryons collide with each other, the dark-matter part of the two clusters, which contains most of the mass, passes through each other without much interaction. With the dark-matter assumption, numerical simulations were able to set limits on the self-interaction cross section of dark matter per unit mass [36]:

$$\sigma_{\chi\chi}/m_\chi \lesssim 1 \text{ cm}^2 \text{ g}^{-1}, \quad (3.3)$$

where χ stands for dark matter.

Another piece of observational evidence that supports the dark-matter hypothe-

sis and poses difficulty for MOND is from the CMB measurements. The anisotropy power spectrum of CMB is very sensitive to the interplay between baryons, photon, and the gravitational potential which comes from the primordial density fluctuations. In particular, the oscillation of the photon-baryon plasma under the presence of gravitational potential gives rise to acoustic peaks in the CMB anisotropy power spectrum. Since dark matter is a source of gravity, and unlike baryons it is not tightly coupled with photons, its presence can significantly modify the shape of the CMB power spectrum, especially the height and position of the acoustic peaks. By measuring these acoustic peaks, the study of the CMB strongly supports the existence of a matter component that, at times before recombination, was already not coupled to the photon-baryon plasma and was interacting through gravity only. Indeed, the CMB provides probably the most precise measurement of the dark-matter abundance and the baryon abundance, which we have listed in Eq. (2.13).

From these observational facts, we conclude here that even though the modification of gravity is able to explain some of the observations, the overall picture from all the current observations strongly favors that dark matter is a matter content. Therefore we shall assume in the rest of this dissertation that dark matter is really one type of matter instead of a manifestation of MOND.

3.1.2 What Makes Dark Matter

Now the next natural question becomes what makes a good dark-matter candidate. Could dark matter be anything in the Standard Model of particle physics or be made of SM particles?

To answer these questions, let us examine through a list of candidates with the observational evidence we have. First of all, we shall exclude all the massless particles *i.e.* photon and gluon. Second, the fact that dark matter is invisible suggests its electric charge has to be neutral. Therefore, charged particles such as electron and proton are out. Third, since we can see the observed gravitational effects of dark matter at the present epoch, it has to be overall stable on cosmological time scale.¹ In

¹We shall see in later chapters how this requirement can be relaxed.

fact the lifetime of dark matter is severely constrained. If the decay products of dark matter are SM particles except for neutrinos, they would generate visible signals, and thus can be easily detected. Even if the decay products are dark radiation, they can be constrained by their gravitational effects. Therefore electric neutral particles like the Higgs boson, Z and neutron, are also immediately ruled out. The only massive, neutral and stable particle candidates in the Standard Model are neutrinos. Neutrinos only participate in weak and gravitational interactions, and therefore can perfectly hide themselves inside galaxies and clusters if they are non-relativistic. Indeed, at least most of the neutrinos are non-relativistic today. However, the problem concerns the time at which they became non-relativistic, since we know that, deep within the RD epoch, when the temperature is above a few MeV, neutrinos were relativistic and in equilibrium with the rest of the SM thermal bath. Because the neutrino masses are so small, it turns out that they are still relativistic at the time when structure formation begins, which is around the MRE. This means neutrinos are able to free-stream over large cosmological length scales. If neutrinos were to take a major share of dark-matter abundance, the density perturbations at scales below the neutrino free-streaming scale would not have enough time to grow and form galaxies, clusters and the large-scale structure that we observe today. Therefore neutrinos cannot be suitable dark-matter candidates.

To create massive, neutral and stable objects out of the Standard Model, one might consider macroscopic objects such as Jupiter-like planets and brown dwarfs which are made of baryons. Those objects are usually referred to as *massive compact halo objects* (MACHOs). However, observations on the CMB also suggests that dark matter is not coupled to the photon-baryon plasma at the times important for generating its anisotropy. Since MACHOs are only formed at later time, the baryons that made them were indeed coupled to the plasma. Thus they cannot make a significant fraction of dark matter today, though in principle they can contribute a small fraction to the measurement of mass-to-light ratio, galactic rotation curves and Bullet Cluster.

The only possibility for SM particles to become dark matter is through *primordial black holes* (PBHs). PHBs are generated from the primordial density perturbation

$\delta\rho$ in the RD epoch. Their energy density scales like matter and could behave as if they are non-baryonic at times important for the CMB. The mass range of the PBHs is huge. PBHs formed at Planck time would have mass $M_{\text{PBH}} \sim 10^{-5}$ g, while those formed at 1 s would have $M_{\text{PBH}} \sim 10^5 M_{\odot}$. However, the density of PBHs are constrained by the Hawking radiation: PBHs with $M_{\text{PBH}} \sim 10^{15}$ g would be evaporating at the present epoch, producing photons at ~ 100 MeV, and thus are constrained by the observed γ -ray background. It is shown that the energy density of PBH at this mass scale is lower than 10^{-8} of the critical energy density. PBHs with masses smaller than 10^{15} g would evaporate at earlier epochs. The photons they inject in the early universe could potentially modify the predictions of the BBN and distort the CMB anisotropy. PBHs with masses larger than 10^{15} g could survive today, and be constrained by its gravitational effects such as lensing and gravitational waves. Although in principle PBHs with 10^{17} g $\lesssim M_{\text{PBH}} \lesssim 10^{24}$ g are still viable as dark-matter candidate, the fact that its formation relies a lot on the primordial density-perturbation spectrum and its mass spans such a huge range makes it highly model-dependent and not particularly predictive. We shall therefore not consider PBHs as the major dark-matter candidate. Interested readers are encouraged to look at Ref. [37] for constraints on PBHs.

Based on all the previous discussions, we conclude here that dark matter has to be primarily made of particles beyond the Standard Model. The gross features that a BSM dark-matter candidate needs to have are listed below:

- *Massive*: It has a mass.
- *Cold*: It has to be non-relativistic before the MRE.
- *Dark*: Its electromagnetic interaction has to be negligible at least from BBN to the present day.
- *Stable*: It has to be overall stable on cosmological timescales.

3.2 Dark-Matter Production in the Early-Universe

The previous section lists the major pieces of evidence for the existence of dark matter in the part of cosmic history that we are able to probe. Most of the observations, such as the mass-to-light ratio, the galactic rotation curves and the Bullet Cluster concerns the universe at late time — deep within the MD epoch or even later than that. However, the study of dark matter in general covers a wide range of temperature scales in the history of the universe. In fact, most dark-matter scenarios rely heavily on the physics in the early universe which we refer to as deeply within the RD epoch.

In this section, we shall discuss the production of dark matter in the early universe. The ways through which dark matter is produced have already been thoroughly studied in the literature, and lots of different mechanisms have been proposed. Though the details of these production mechanisms and the ways they are related to the late-time detection signals vary-among different models, in general, there are two major classes of dark matter production mechanisms — thermal production and non-thermal production. The former assumes that dark matter had been in thermal equilibrium in the early universe until it decoupled when the processes that had kept it in equilibrium became too slow. The latter allows the possibility that dark matter was never in thermal equilibrium with the rest of the cosmic plasma, and in this case dark matter can usually be generated from misalignment production or out-of-equilibrium decay, etc. In what follows, we shall briefly introduce two typical examples for the two classes of production mechanisms — thermal freeze-out and vacuum misalignment.

3.2.1 Thermal Freeze-out

The thermal freeze-out is one of the most popular dark-matter production mechanisms. It relies on the assumption that dark matter had been in thermal equilibrium with the rest of the thermal bath at high temperature in the RD epoch. As the universe expands and cools down, at a certain point the processes that had kept dark matter in equilibrium became inefficient, and thus dark matter decoupled from the thermal bath with its comoving number density fixed.

To fully appreciate the beauty of the thermal freeze-out mechanism, let us take a simple example which only concerns the dark-matter annihilation: $\chi\chi \leftrightarrow \phi\phi$. Here ϕ can be any SM particle or BSM particle that couples to the SM particles, but for simplicity, let us just assume it is a SM particle. The equation that governs the evolution of dark-matter number density n_χ is the Boltzmann equation:

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -n_\chi^2 \langle \sigma v \rangle \left[1 - \left(\frac{n_\chi^{\text{eq}}}{n_\chi} \right)^2 \right], \quad (3.4)$$

in which n_χ^{eq} is the number density if dark matter is in equilibrium, and $\langle \sigma v \rangle$ is the *thermally averaged cross-section* of the annihilation process.

Let us first analyze the general behavior of this type of equation. We first notice there is a competition between H and $n_\chi \langle \sigma v \rangle$ — the Hubble expansion rate and the reaction rate. If $H \gg n_\chi \langle \sigma v \rangle$, the Hubble expansion dominates as the dark-matter particles can hardly find each other to annihilate. Therefore n_χ simply dilutes as a^{-3} , as if there is no interaction at all. If $H \ll n_\chi \langle \sigma v \rangle$, the annihilation is very efficient such that particles can annihilate in a time scale much shorter than the Hubble time H^{-1} . In this case, if $n_\chi > n_\chi^{\text{eq}}$, the right-hand side of Eq. (3.4) is negative, and therefore n_χ decreases faster than a^{-3} — dark matter annihilates more into the SM particles. On the contrary, if $n_\chi < n_\chi^{\text{eq}}$, the right-hand side of Eq. (3.4) is positive, which means n_χ scales slower than a^{-3} — SM particles annihilate more to produce dark-matter particles. Therefore, we see the scattering term on the right-hand side always tends to keep maintain the species in equilibrium.

Now let us look back at Eq. (2.30). Since dark matter had already become non-relativistic in the RD epoch, if it had been kept in equilibrium until today, its relic density would be exponentially small. Therefore, dark matter that was in equilibrium with the thermal bath has to break away from the thermal equilibrium sufficiently early in order to have a large relic abundance as we have observed today. To guide the eye, an plot for dark-matter freeze-out is shown in Fig. 3.1

Let us do some order of magnitude estimates. First, suppose that dark-matter freezes out when relativistic, $T_{\text{fo}} \gg m_\chi$, and its temperature right at freeze-out

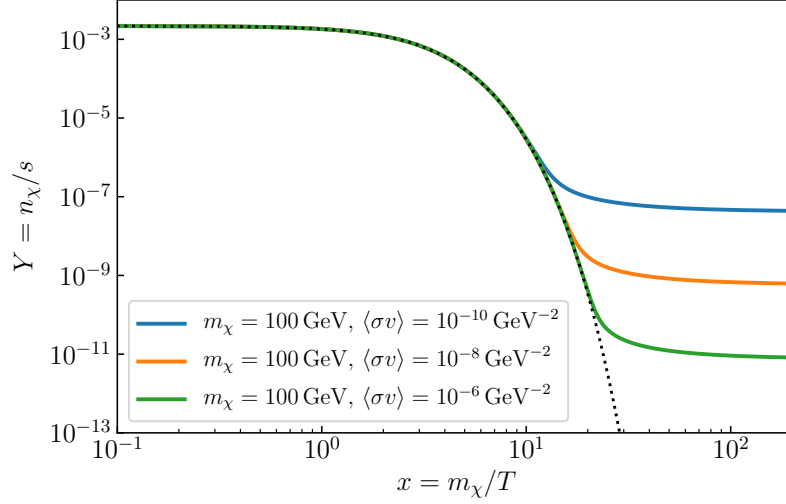


FIGURE 3.1: Thermal freeze-out of dark matter with mass $m_\chi = 100$ GeV. The solid curves show three different cases in which $\langle\sigma v\rangle = 10^{-10}, 10^{-8}, 10^{-6}$ GeV $^{-2}$ at freeze-out. The black dotted curve represents the equilibrium case.

is still the same as the temperature of the rest of the thermal bath. The yield at freeze-out is equal to the yield today due to entropy conservation:

$$Y_{\chi,\text{fo}} = \frac{n_\chi(T_{\text{fo}})}{s(T_{\text{fo}})} = Y_{\chi,\text{now}} = \frac{n_\chi(T_{\text{now}})}{s(T_{\text{now}})} = \frac{\rho_\chi(T_{\text{now}})}{m_\chi s(T_{\text{now}})}. \quad (3.5)$$

Let us ignore the difference between g_* and $g_{*,S}$ as we have seen that they are very similar during the the history of the universe. Substituting in the numbers, we see $Y_{\chi,\text{fo}} \approx 1/(3g_*(T_{\text{fo}}))$. This leads to the relic abundance calculation

$$\Omega_\chi = \frac{\rho_\chi(T_{\text{now}})}{\rho_{\text{crit}}} = Y_{\chi,\text{fo}} \frac{m_\chi s(T_{\text{now}})}{\rho_{\text{crit}}} \approx \frac{1}{3g_*(T_{\text{fo}})} \frac{m_\chi s(T_{\text{now}})}{\rho_{\text{crit}}}. \quad (3.6)$$

Using $T_{\text{now}} = 0.234$ meV and $\rho_\gamma(T_{\text{now}}) \sim (3/4)s(T_{\text{now}})T_{\text{now}} \sim 5 \times 10^{-5}\rho_{\text{crit}}$, we find $s(T_{\text{now}})/\rho_{\text{crit}} \sim 30$ eV $^{-1}$. Therefore a rough estimate gives:

$$\Omega_\chi \sim \frac{1}{g_*(T_{\text{fo}})} \frac{m_\chi}{10 \text{ eV}}. \quad (3.7)$$

This suggests a dark-matter mass at the eV scale. Just like neutrinos, such dark-matter particles would be too hot for structure formation. This is why thermal dark

matter that freezes out while being relativistic are typically categorized as the *hot dark matter*.

Let us now look at the case in which dark matter freezes out non-relativistically $T_{\text{fo}} \ll m_\chi$. This time, we have $Y_{\chi,\text{fo}} \approx x_{\text{fo}}^{3/2} e^{-x_{\text{fo}}} / (6g_*(T_{\text{fo}}))$, in which the convention $x \equiv m_\chi/T$ is used. Unlike freezing out relativistically which depends on the freeze-out temperature only through g_* , the non-relativistic case has a stronger dependence through x_{fo} . The temperature at which dark matter freezes out can be estimated using the condition $H(T_{\text{fo}}) \sim n_\chi(T_{\text{fo}}) \langle \sigma v \rangle$. Together with Eq. (2.34), this condition leads to

$$n_\chi(T_{\text{fo}}) \sim \frac{\pi}{\sqrt{90}} \frac{g_*(T_{\text{fo}})^{1/2} T_{\text{fo}}^2}{M_{\text{P}} \langle \sigma v \rangle}. \quad (3.8)$$

After some algebra, one can write it in a very suggestive way:

$$\Omega_\chi \sim \left(\frac{x_{\text{fo}}}{10} \right) \left(\frac{100}{g_*(T_{\text{fo}})} \right)^{1/2} \left(\frac{10^{-8} \text{ GeV}^{-2}}{\langle \sigma v \rangle} \right). \quad (3.9)$$

The fact that a dark-matter candidate with $m_\chi \sim \mathcal{O}(100) \text{ GeV}$ and a weak scale cross-section that freezes out non-relativistically at $x_{\text{fo}} \sim \mathcal{O}(10)$ is able to give rise to a relic abundance $\Omega_\chi \sim \mathcal{O}(1)$ strongly motivates the traditional WIMP scenarios. This is usually referred to as the “*WIMP miracle*”.

However, this does not mean that the thermal freeze-out mechanism favors a weak-scale mass. Do notice that in deriving Eq. (3.9), no assumption for x_{fo} has been made. By playing with the way the cross-section scales with mass and couplings, the dark-matter mass can be in principle as small as 0.1 eV [38]. Although for dark-matter candidates that are produced thermally, similar to neutrinos, constraints from structure formation have set the lower limit for dark-matter mass to be above a few keV, the point emphasized here is that there is a lot of room to play in dark-matter model building instead of hanging upon the WIMP miracle.

In the end, we briefly comment on the fact that light dark-matter candidates ($m_\chi \lesssim \text{keV}$) produced from thermal freeze-out receive severe constraint from structure formation. The reason behind this is very simple. On one hand, around freeze-out, dark matter has the same temperature with the rest of the thermal bath. After

the decoupling of dark matter, its temperature simply scales $T_\chi \sim a^{-1}$ if it is relativistic, whereas $T_\chi \sim a^{-2}$ if it is non-relativistic. At the same time, entropy conservation tells us that the temperature of the thermal bath scales as $T \sim a^{-1}$ up to small heating effects from the change of $g_{*S}^{-1/3}$ which only gives $\mathcal{O}(1)$ corrections. Therefore, the temperature of dark matter in general follows the temperature of the thermal bath until dark matter is non-relativistic. On the other hand, most of the contribution to the dark-matter free-streaming length comes from the time when dark matter is relativistic. Now recall that structure formation starts around the MRE, at which $T \sim \text{eV}$. For thermal dark matter with a mass below a few keV, no matter when it decouples, it would always become non-relativistic at $T \sim \text{keV}$, which is close enough to the MRE. Therefore, dark matter would almost always acquire a free-streaming length large enough to leave a trace in the large-scale structure. In the worst case, the time of the MRE will be shifted if the mass is too small. Thus, to produce very light dark matter, non-thermal mechanisms are usually employed.

3.2.2 Misalignment Production

The misalignment production is one of the typical non-thermal production mechanisms. It is widely used in the production light scalars² such as axions or axion-like particles (ALPs) whose mass scale ranges from 10^{-3} eV to 10^{-22} eV . In these scenarios, dark matter is usually described as a classical scalar field $\phi(t)$, whose energy density and pressure are

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) , \quad (3.10)$$

$$P_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) . \quad (3.11)$$

Note that the definition of pressure here for a classical field is different from that for particles (see Eq. (2.18)). Depending on the evolution of the field, the equation of state parameter $w = P_\phi/\rho_\phi$ can vary from -1 to 1 . For simplicity, let us use a

²Fermions have to have a mass above $\sim \text{keV}$ due to their finite phase-space density in a gravitational bounded system [39].

quadratic potential $V(\phi) = m^2\phi^2/2$ as an example. The equation of motion of the field ϕ follows from the continuity equation, Eq. (2.6):

$$\ddot{\phi} + 3H\dot{\phi} + m^2\phi = 0, \quad (3.12)$$

in which we have used $\partial V/\partial\phi = m^2\phi$. The above differential equation takes the form of a damped oscillator, in which the $3H\dot{\phi}$ term acts as a friction term. Let us assume that H can be treated as a constant in the regimes we are interested in. Inserting $\phi(t) \sim e^{i\omega t}$ and solving for ω , it is easy to find

$$\omega = i\frac{3}{2}H \pm \sqrt{m^2 - \frac{9}{4}H^2}. \quad (3.13)$$

Now we can investigate the behaviors in different limits. At early times, $H \gg m$, we see $\omega = 0, 3iH$, which means ϕ evolves quickly towards a constant

$$\phi(t) \simeq \phi_0 + \phi_1 e^{-3Ht} \xrightarrow{3Ht \gg 1} \phi_0. \quad (3.14)$$

The energy density therefore also quickly evolves towards an asymptotic constant value $\rho_\phi \simeq V = m^2\phi_0^2/2$ — the field behaves like dark energy. At later times, $H \ll m$, thus $\omega = 3iH/2 \pm m$ and the field fast oscillates about its minimum:

$$\phi(t) = \phi_2 e^{\pm imt} e^{-\frac{3H}{2}t}. \quad (3.15)$$

Using the virial theorem $m^2\phi^2 = \dot{\phi}^2$, which is valid since the oscillation time scale $1/m$ is much smaller than the expansion time scale $1/H$, we see that on average $\rho_\phi \simeq m^2\phi_2^2 e^{-3Ht}/2$. Notice that, with the assumption that H can be treated as a constant, $e^{-3Ht} \sim a^{-3}$. Therefore, the energy density $\rho_\phi \sim a^{-3}$, that is, the field behaves like matter in this limit.

The transition point between these two regime is $m \simeq 3H/2$ when the real part of ω becomes available which allows the field to oscillate. One can use this to define a production time by $H(t_{\text{prod}}) = 2m/3$. The production time defined by this relation

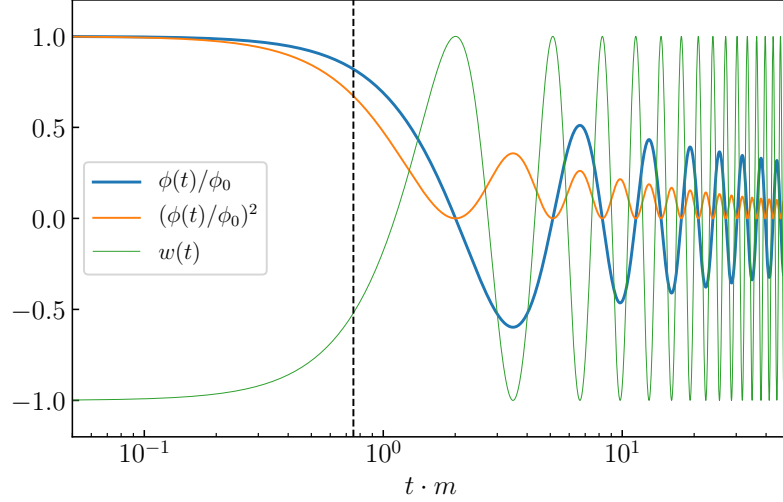


FIGURE 3.2: Evolution of the classical field ϕ . The dashed black line indicates the production time defined by the relation $H(t_{\text{prod}}) \simeq 2m/3$. Before t_{prod} , the field and its energy density ($\propto \phi^2$) stay as a constant, and $w \simeq -1$. After t_{prod} , the field oscillates rapidly. This makes $\langle w \rangle \simeq 0$ and $\rho_\phi \sim a^{-3}$.

is shown in Fig. 3.2 as the dashed black vertical line. Clearly, there is a transition from the phase in which the field is a constant into the fast oscillating phase after the production time. Moreover, we see the average of the equation of state parameter $\langle w \rangle = 0$ in the oscillating phase, which means the average energy density of the field is indeed behaving like matter. Needless to say, our qualitative estimate agrees well with the numerical calculation.

We still need to make sure that the mass scales which we are interested in correspond to a production time t_{prod} before the MRE. To prove that, we notice that Eq. (2.34) gives $H(t_{\text{MRE}}) \sim 10^{-28}$ eV. Therefore, even for the masses as small as $m \sim 10^{-22}$ eV, the production time is ensured to be before MRE.

In the end, in order to relate the dark-matter relic abundance with the initial field value, let us glue the two limits together at t_{prod} . First of all, the conservation of energy at the transition time t_{prod} gives $\phi_0 = \phi_2 = \phi(t_{\text{prod}})$.³ Requiring ρ_ϕ to be the

³We see here ϕ_0 shows how much the initial field value is misaligned with the minimum of the potential which it oscillates about later, and hence the name misalignment production.

energy density of dark matter, we have

$$\begin{aligned}
0.25 \approx \Omega_\phi &= \frac{\rho_\phi(t_{\text{now}})}{\rho_{\text{crit}}} \sim \frac{m^2 \phi(t_{\text{now}})^2}{(2.5 \times 10^{-3} \text{ eV})^4} \\
&= \frac{m^2 \phi(t_{\text{prod}})^2}{(2.5 \times 10^{-3} \text{ eV})^4} \left(\frac{a(t_{\text{prod}})}{a(t_{\text{now}})} \right)^3 \\
&= \frac{m^2 \phi(t_{\text{prod}})^2}{(2.5 \times 10^{-3} \text{ eV})^4} \frac{g_{*S}(T_{\text{now}})}{g_{*S}(T_{\text{prod}})} \left(\frac{T_{\text{now}}}{T_{\text{prod}}} \right)^3. \quad (3.16)
\end{aligned}$$

After a few algebra, the estimate shows

$$m\phi(t_{\text{prod}}) \sim (mM_{\text{P}})^{3/4} \text{eV}^{1/2}. \quad (3.17)$$

To conclude this chapter, we have seen that there is ample evidence showing that dark matter has to exist, and there are also compelling reasons for dark matter to have a particle nature. Although we have not yet observed such particles, we know that it should have some gross features — massive, cold, dark and overall stable. Particle species like this do not exist in the Standard Model, and the production of those particles has to take place in the early universe, deeply within the RD epoch. There are two major classes of dark-matter production mechanisms, thermal production and non-thermal production. We have shown one typical example for each class — dark matter production from thermal freeze-out and from vacuum misalignment. We have also demonstrated the way the relic abundance calculation is related to the physical observables relevant in the early universe. Of course, there exist many other production mechanisms as well as extensions to these two basic mechanisms, and the details of dark matter production is highly model-dependent. Nevertheless, the two production mechanisms that we have shown here are often used as necessary ingredients.

We emphasize that the study of dark matter involves physics from many aspects, including particle physics, astrophysics, cosmology and statistical physics, *etc.* Therefore, when performing the calculations, it is important to check the implicit assumptions made from different sides. For example, the relic abundance calculation that

we have shown has the following implicit assumptions: i) the history of the universe is standard; ii) dark matter only has a single component and its phenomenology depends only on the properties of itself; iii) dark matter is stable. Indeed, these implicit assumptions prevail in most of the traditional scenarios. However, these assumptions can be relaxed. First, as we have mentioned before, since we have no direct probe of the physics before BBN, any cosmological model that successfully reproduces all the physics from BBN to the present epoch cannot be ruled out by current observations. Therefore, in models within non-standard cosmological history, the relic abundance calculation can be easily modified [18, 19, 21]. Second, other than being single-component, it is not unreasonable to consider dark sectors with non-minimal construction such that the dark-matter phenomenology depends on the properties of not only one single particle, but the entire dark sector. After all, the Standard Model has exhibited complex and delicate structure while dark matter is much more abundant than the SM particles. If dark matter consists of more than one components, absolute stability would not be a necessary requirement. In general, dark-matter components whose cosmological abundance is small does not have to be stable at all. Even if dark matter has only one component today, the early-universe dynamics in the dark sector might also leave non-trivial imprints in the late-time cosmology.

Through out this dissertation, we shall assume a standard cosmology. However, we shall relax the other two assumptions that dark matter physics is minimal and that dark matter is absolutely stable. In the following chapters, we shall present two types of non-minimal scenarios — those in which dark matter consists of multiple components which decay predominantly into the SM final states, and those in which dark matter is single-component today, but decays that occur entirely within the non-minimal dark sector give rise to non-trivial cosmological consequences.

CHAPTER 4

DYNAMICAL DARK MATTER FROM STRONGLY-COUPLED DARK SECTORS

In the previous chapter, we have discussed two typical traditional DM scenarios in which the dark matter has only a single component. In this chapter, we shall study a new framework for dark-matter physics which is called “Dynamical Dark Matter” (DDM) [40–42].

The basic idea behind DDM is relatively simple. Rather than focus on one or more stable dark-matter particles, we instead consider a multi-component framework in which the dark matter of the universe comprises a vast ensemble of interacting fields with a variety of different masses, mixings, and cosmological abundances. Moreover, rather than impose stability for each field individually, we instead ensure the phenomenological viability of such a scenario by requiring that those states with larger masses and SM decay widths have correspondingly smaller relic abundances, and vice versa. In other words, dark-matter stability is not an absolute requirement in the DDM framework, but is replaced by a *balancing* of lifetimes against cosmological abundances across the entire ensemble. This leads to highly dynamical scenario in which cosmological quantities such as Ω_{CDM} experience non-trivial time-dependences beyond those normally associated with cosmological expansion. It turns out that many extensions to the SM — including string theory as well as theories with large extra dimensions or large hidden sectors — give rise to large ensembles of dark states in which such a balancing naturally occurs. Moreover, because the dark-matter “candidate” in this framework consists of a carefully balanced DDM ensemble which cannot be characterized in terms of a single well-defined mass, decay width, or interaction cross section, this framework generically gives rise to many striking signatures at colliders as well as at direct- and indirect-detection dark-matter experiments — signatures which transcend those usually associated with dark matter and which ultimately reflect the collective behavior of the entire DDM ensemble. An artistic sketch

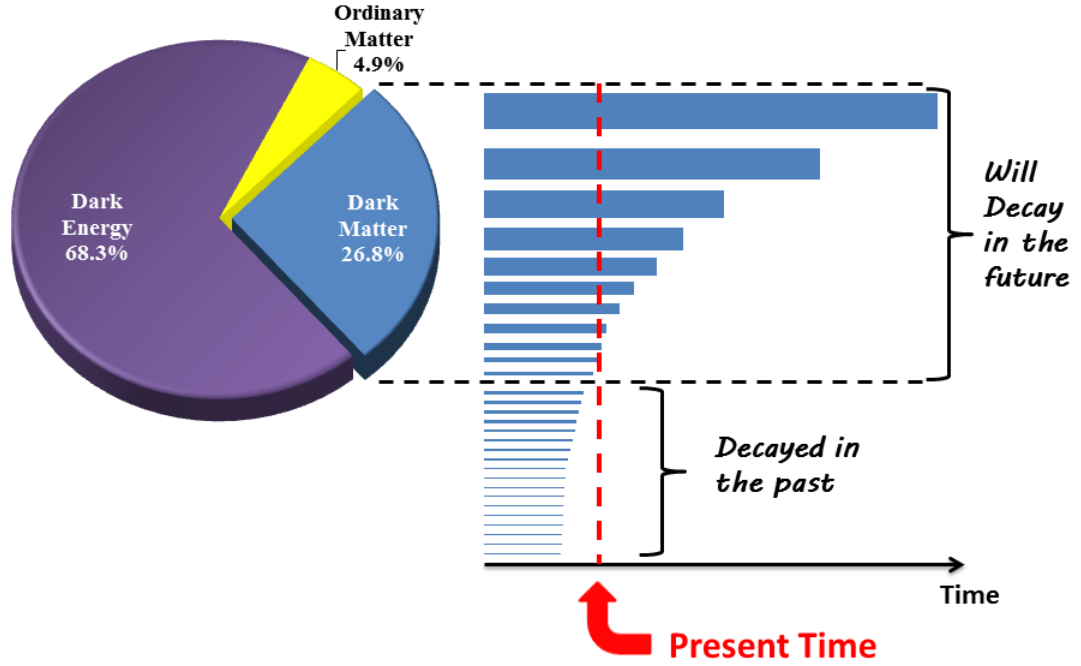


FIGURE 4.1: A sketch for DDM. In this plot, the blue stripes stand for different dark-matter components, and their lengths and widths represent the corresponding lifetimes and abundances, respectively. Clearly, the dark matter that we observe today only consists of the “surviving” components. There are components that have already decayed in the past, and, eventually, all of the dark-matter components will decay in the future.

for the DDM idea is shown in Fig. 4.1 in which the dark-matter components are represented by the blue stripes of which the lifetimes are indicated by their lengths, while the abundances are indicated by their widths, and what we observe today as the dark matter are merely the components that survive to the present day.

The DDM framework was originally introduced by K. R. Dienes and B. Thomas in Ref. [40], in which its theoretical aspects have been surveyed. The “existence proof” of the phenomenological viability of the DDM framework is shown in Refs. [43–45] in which an explicit DDM model was constructed, and it is demonstrated that this model satisfies all known collider, astrophysical, and cosmological constraints. Various signatures by which this framework might be experimentally tested and con-

strained have also been studied. These include unique signatures at direct-detection experiments [46], at indirect-detection experiments [47–50], and at colliders [51–55].

Of course, many of the constraints on such DDM ensembles depend on model-specific details associated with the ensemble in question, such as the specific particle nature of the individual dark constituent fields and the precise form of their decays into SM states. By contrast, other phenomenological properties of (and constraints on) these DDM ensembles depend simply on the manner in which the lifetimes and abundances of ensemble constituents scale with respect to each other, and thus have a greater degree of model-independence. For example, the effective equations of state for these ensembles are governed in large part solely by these scaling relations. As a result, all phenomenological/observational constraints on the equations of state of the dark sector are essentially constraints on the types of balancing relations that DDM ensembles may exhibit. These are thus model-independent constraints which can be placed on such ensembles simply as a result of their inherent scaling relations.

One general class of DDM ensembles consisting of large numbers of dark particle species exhibiting suitable scaling relations between lifetimes and cosmological abundances are those whose constituents are the Kaluza-Klein (KK) modes of a gauge-neutral bulk field in a theory with extra spacetime dimensions in which cosmological abundances are established through misalignment production [40]. Indeed, explicit realizations of DDM ensembles of this type have been constructed [43, 44]. Although many aspects of these ensembles depend on the details of the particular fields under study, certain general properties are common across all such ensembles in this class. One of these is that the cosmological abundance of each component scales as a power of the lifetime of that component. Likewise, the density of states within such ensembles is either insensitive to mass or scales roughly as a polynomial function of mass across the ensemble. For these reasons, most phenomenological studies of the DDM framework have focused on ensembles exhibiting polynomial scaling relationships.

Polynomial scaling relations also emerge in other (purely four-dimensional) contexts as well. For example, under certain circumstances, thermal freeze-out mechanisms for abundance generation can also lead to appropriate polynomial inverse

scaling relations between lifetimes and abundances [56]. In fact, such inverse scaling relations can even emerge *statistically* in contexts in which the dynamics underlying the dark sector is essentially random [57].

There are, however, other well-motivated theoretical constructions which do not give rise to dark sectors with polynomial scaling relations. One example is a dark sector consisting of a set of fermions (dark “quarks”) charged under a non-Abelian gauge group G which becomes confining below some critical temperature T_c . At temperatures $T \lesssim T_c$, when the theory is in the confining phase, the physical degrees of freedom are composite states (dark “hadrons”). Another well-motivated type of DDM ensemble consists of the bulk (*i.e.*, closed-string) states in Type I string theories. Such bulk states are typically neutral with respect to all brane gauge symmetries, and interact with those brane states only gravitationally. As such, from the perspective of brane-localized observers, these bulk states too are dark matter.

At first glance, these two latter types of ensembles may seem to have little in common with each other. Indeed, many aspects of the detailed phenomenologies associated with these ensembles will be completely different. However, they nevertheless exhibit certain underlying model-independent commonalities which are relevant for their viability as DDM ensembles. Indeed, these features are identical to those which characterize the “visible” sector of ordinary hadrons, namely

- mass distributions which follow linear Regge trajectories (*i.e.*, $\alpha' M_n^2 \sim n$ where α' is a corresponding Regge slope), and
- exponentially growing (“Hagedorn-like”) degeneracies of states (*i.e.*, $g_n \sim e^{\sqrt{n}} \sim e^{\sqrt{\alpha'} M_n}$).

These features — especially the appearance of an *exponential* scaling of the state degeneracies with mass — represent a behavior which is markedly different from that exhibited by DDM ensembles with polynomial scaling relations. For example, as a result of their exponentially growing densities of states, such ensembles have a critical temperature [58] beyond which their partition functions diverge.

In this chapter, we shall study the generic properties of DDM ensembles which

exhibit the two features itemized above. We shall calculate the effective equations of state $w_{\text{eff}}(t)$ for such ensembles, and subject these ensembles to those immediate model-independent observational constraints that follow directly from these equations of state. We shall therefore be able to place zeroth-order model-independent bounds on some of the quantities that parametrize these features, such as the effective Regge slope as well as the rate of exponential growth in the state degeneracies. Our primary motivation is to understand the phenomenology that might apply to strongly-coupled dark sectors in their confined (“hadronic”) phase, imagining nothing more than that our DDM ensemble resembles the visible hadronic sector in the two respects itemized above. However, the results of such analyses might also be useful in constraining the bulk sector of various classes of string theories, since these bulk sectors also give rise to ensembles of dark-matter states which share these two grossest features. We shall therefore aim to keep our discussion as model-independent as possible, subject to our assumption of the above two properties itemized above. In this way, our analysis and the constraints we obtain can serve as useful phenomenological guides in eventually building realistic dark-matter models of this type.

This chapter is organized as follows. In Sect. 4.1, we begin by reviewing the properties that we shall assume for the mass spectrum and density of states of our DDM dark “hadron” ensemble. We shall also discuss the physical interpretations of these properties in terms of a variety of underlying flux-tube models and string theories. This section will also serve to establish our conventions and notation. Then, in Sect. 4.2, we discuss how the required balancing between lifetimes and abundances naturally arises for such DDM ensembles. In particular, we examine the mechanism through which primordial abundances for these hadron resonances are generated, and we determine how these abundances scale across the ensemble as a function of the hadron mass. We also discuss the scaling behavior of the decay widths that characterize the decays of the hadronic ensemble constituents to SM states, as well as the assumptions that enter into such calculations. In Sect. 4.3, we then derive expressions for the total abundance $\Omega_{\text{tot}}(t)$, the tower fraction $\eta(t)$, and the effective equation-of-state parameter $w_{\text{eff}}(t)$ for these DDM ensemble as functions of time. As

discussed in Refs. [40, 43] and reviewed in Sect. 4.3, these three functions characterize the time-evolution of DDM ensembles and allow us to place a variety of general, model-independent constraints on such ensembles. In Sect. 4.4, we then present the results of our analysis of the phenomenological viability of such DDM ensembles, identifying those regions of the corresponding parameter space which lead to the most promising ensembles and uncovering generic phenomenological behaviors and correlations across this space. One of our key findings is that these DDM ensembles can satisfy our constraints across a broad range of energy scales ranging from the GeV scale all the way to the Planck scale, but that the present-day cosmological abundance of the dark sector must be distributed across an increasing number of different states in the ensemble as the fundamental mass scales associated with the ensemble are dialed from the Planck scale down to the GeV scale. Finally, in Sect. 4.5, we summarize our results and discuss possible avenues for future work.

4.1 DDM ensembles of dark hadrons: Fundamental assumptions

As discussed in above, in this chapter we are primarily concerned with the properties of DDM ensembles whose constituents are the “hadronic” composite states or resonances of a strongly-coupled dark sector. As has been well known since the 1960’s, many of the attributes of such an ensemble can be successfully modeled by strings. These attributes include linear Regge trajectories, linear confinement, an exponential rise in hadron-state degeneracies, and s - and t -channel duality. It is not a complete surprise that there is a deep connection between hadronic spectroscopy and the spectra of string theory. Hadronic resonances (particularly mesons) can be viewed as configurations of dark “quarks” linked together by flux tubes. The spectrum of excitations in such a theory therefore corresponds to the spectrum of fluctuations of these flux tubes. However, it is well known that these flux tubes can be modeled as non-critical strings. Thus string theory can provide insight into the properties of such collections of composite states.

In what follows, we shall use this analogy between hadronic physics and string theory to motivate our parametrization for the mass spectrum and for the density of

states of our dark-“hadronic” DDM ensembles. We shall also make recourse to modern string technology, when needed, for refinements of our basic picture. Throughout, however, we shall attempt to keep our parametrizations as general as possible so that they might apply to the widest possible set of DDM ensembles sharing these properties. As discussed in the Introduction, this will allow our analysis and eventual constraints to serve as useful guides in future attempts to build realistic models exhibiting these features.

4.1.1 The mass spectrum: Regge trajectories

The first feature that we shall assume of our hadronic dark sector is a mass spectrum consistent with the existence of Regge trajectories. The existence of such trajectories follows directly from nothing more than our assumption that our dark-sector bound states can be modeled by dark quarks connected by the confining flux tube associated with a strong, attractive, dark-sector interaction. Taking meson-like configurations as our guide and temporarily assuming massless quarks, it can easily be shown that the mass M_n associated with a relativistic rotating flux tube scales with the corresponding total angular momentum n as $n \sim \alpha' M_n^2$, where α' is the so-called Regge slope. In the *visible* sector, this successfully describes the so-called leading Regge trajectory of the observed mesons, with $\alpha' \sim 1 \text{ (GeV)}^{-2}$ appropriate for QCD. Moreover, there also exist subleading (parallel) Regge trajectories of observed mesons which have the same Regge slope but different intercepts: $n \sim \alpha' M_n^2 + \alpha_0$.

Regge trajectories of this form, both leading and subleading, also emerge in string theory. For example, the perturbative states of a quantized open bosonic string have masses M and spins $J = 0, 1, \dots, J_{\max}$ which satisfy $J_{\max} = \alpha' M^2 + 1$ where α' is now the Regge slope associated with string theory [typically assumed to be $\sim (M_{\text{Planck}})^{-2}$]. The states with $J = J_{\max}$ thus sit along the leading Regge trajectory, while those with smaller values of J sit along the subleading Regge trajectories. Similar results also hold for superstrings and heterotic strings.

Given these observations, in this chapter we shall assume that the states of our

dark “hadronic” DDM ensemble have discrete positive masses M_n of the general form

$$M_n^2 = nM_s^2 + M_0^2 . \quad (4.1)$$

where n is an index labeling our states in order of increasing mass. Here $M_s \equiv 1/\sqrt{\alpha'}$ is the corresponding “string scale”, while M_0 represents the mass of the lightest “hadronic” constituent in the DDM ensemble. Indeed, since we do not expect to have any tachyonic states in our DDM ensemble, we shall assume throughout this chapter that $M_0^2 \geq 0$. We shall avoid making any further assumptions about the nature of the dark sector by treating both M_s and M_0 as free parameters to be eventually constrained by cosmological data.

Our choice of sign for M_0^2 perhaps deserves further comment. For the visible sector, most hadrons lie along Regge trajectories with $M_0^2 \geq 0$. While there do exist Regge trajectories with $M_0^2 < 0$, the lowest states in such trajectories are of course absent. In string theory, by contrast, all Regge trajectories have $M_0^2 < 0$. However, just as in the hadronic case, all tachyonic states which might result for small n are ultimately removed from the string spectrum by certain “projections” which are ultimately required for the self-consistency of the string. In other words, for Regge trajectories with $M_0^2 < 0$, one could equivalently relabel our remaining states by shifting $n \rightarrow n - 1$ and thereby obtain an “effective” $M_0^2 \geq 0$. This is not normally done in string theory because in string theory the index n is correlated with other physical quantities such as the spin of the state. However we are making no such assumption for the states of our dark sector, and are treating the index n as a mere labelling parameter. Our assumption of a tachyon-free dark sector then leads us to take $M_0^2 \geq 0$.

There is also another motivation for taking $M_0^2 \geq 0$. All of the above results which treat n as an angular momentum assume massless quarks at the endpoints of the flux tube. However, while such an approximation holds well for the lightest states in the visible sector, we do not wish to make such an approximation for our unknown dark sector. We shall therefore assume $M_0^2 \geq 0$ in what follows, recognizing

that this parameter may in principle also implicitly include the positive contributions from dark quark masses as well.

4.1.2 Degeneracy of states: Exponential behavior

The second generic feature associated with hadronic spectroscopy is the well-known exponential rise in the degeneracies of hadrons as a function of mass: $g_n \sim e^{\sqrt{n}}$. This behavior was first predicted and observed for hadrons (both mesons and baryons) in Ref. [58], and also holds as a generic feature for both bosonic and fermionic states in string theory [23–26].

In general, we can understand this behavior as follows. If we model our hadrons as quarks connected by flux tubes, the degeneracy g_n of hadronic states at any mass level n can be written as the product of two contributions: one factor κ representing a multiplicity of states due to the degrees of freedom associated with the quarks (such as the different possible configurations of quantities like spin and flavor), and a second factor \hat{g}_n representing the multiplicity of states due to the degrees of freedom associated with the flux tube. We thus have

$$g_n \approx \kappa \hat{g}_n . \quad (4.2)$$

While κ is a constant which is independent of the particular mass level n , the remaining degeneracy factor \hat{g}_n counts the rapidly increasing number of ways in which a state of given total energy n can be realized as a combination of the vibrational, rotational, and internal excitations of the different harmonic oscillators which together comprise a quantized string. It is this quantity which grows exponentially with mass, and in string theory the leading behavior of \hat{g}_n for large n generally takes the form [23–26]

$$\hat{g}_n \approx A n^{-B} e^{C\sqrt{n}} \quad \text{as } n \rightarrow \infty , \quad (4.3)$$

where A, B, C are all positive quantities which depend on the particular type of string model under study. Indeed, for any B and C , it turns out that the proper

normalization for \hat{g}_n in string theory is given by

$$A = \frac{1}{\sqrt{2}} \left(\frac{C}{4\pi} \right)^{2B-1}. \quad (4.4)$$

Thus our asymptotic degeneracy of states is parametrized by two independent quantities B and C , and we shall assume that this continues to be true in our dark sector as well.

The most salient property of the expression in Eq. (4.3) is that it rises exponentially with \sqrt{n} , or equivalently with the mass M_n of the corresponding state. This represents a crucial difference relative to the KK-inspired DDM ensembles previously considered in Refs. [40, 43, 44] (or even the purely four-dimensional DDM ensembles considered in Refs. [56, 57]). For example, the KK states corresponding to a single flat extra spacetime dimension have degeneracies \hat{g}_n which are constant, or which become so above the $n = 0$ level. The key difference here is that the degrees of freedom associated with our flux tube consist of not only KK excitations (if the flux tube happens to be situated within a spacetime with a compactified dimension), but also so-called *oscillator* excitations representing the internal fluctuations of the flux tube itself. It is these oscillator excitations which give rise to the exponentially growing degeneracies and which are a direct consequence of the non-zero spatial extent of the flux tube. As such, they are intrinsically stringy and would not arise in theories involving fundamental point particles.

Unfortunately, the asymptotic form in Eq. (4.3) is not sufficient for our purposes. Although we are interested in the behavior of all states across the DDM ensemble, it is the lighter states rather than the heavier states which are most likely to have longer lifetimes and therefore greater cosmological abundances. Thus, even though we want to keep track of all of the states in our ensemble, we need to be particularly sensitive to the degeneracies of the lighter states, *i.e.*, the states with smaller values of n . This poses a problem because the asymptotic expression in Eq. (4.3) is fairly accurate in the large- n limit but is not especially accurate in the small- n limit.

Fortunately, for values of B and C which correspond to self-consistent strings (to

be discussed below), the tools of modern string technology (specifically conformal field theory and modular invariance) furnish us with a more precise approximation for \hat{g}_n which remains accurate even for very small values of n . This expression is given by [59–62]

$$\hat{g}_n \approx 2\pi \left(\frac{16\pi^2 n}{C^2} - 1 \right)^{\frac{1}{4}-B} I_{|2B-\frac{1}{2}|} \left(C \sqrt{n - \frac{C^2}{16\pi^2}} \right), \quad (4.5)$$

where $I_\nu(z)$ denotes the modified Bessel function of the first kind of order ν . Use of the approximation $I_\nu(z) \approx e^z/\sqrt{2\pi z}$ for $z \gg 1$ then reproduces the result in Eq. (4.3). However, the expression in Eq. (4.5) remains valid to within only a few percent all the way down to $n = 1$, assuming $C \leq 4\pi$ (so that the argument of the Bessel function remains real even for $n = 1$).

In what follows, we therefore shall adopt the expression in Eq. (4.5) as our general parametrization for the degeneracy of states \hat{g}_n for arbitrary values of B and $C \leq 4\pi$ and for all $n \geq 1$. For values of B and C corresponding to bona-fide string theories, this expression yields results for the state degeneracies which, though not necessarily integral, are highly accurate for all values of $n \geq 1$. An explicit example of this will be provided below. More generally, however, this expression is smooth and well-behaved for all values of the B and C parameters, and in all cases exhibits the exponential Hagedorn-like behavior whose primary effects we seek to analyze in this chapter. For $n = 0$, by contrast, we shall define $\hat{g}_0 \equiv 1$, representing the unique ground state of our flux tube.

4.1.3 Physical interpretation of ensemble parameters

Thus far we have introduced four parameters to describe our dark “hadron” DDM ensemble: M_s , M_0 , B , and C . The first two parameters have immediate interpretations: M_0 is the mass of the lightest state in the DDM ensemble, while M_s parametrizes the splitting between the states. We would now like to develop analogous physical interpretations of B and C .

Clearly B and C describe the dynamics of the flux tube. However, in the case of the ordinary strong interaction, many possible theories governing this dynamics have been

proposed. These range from early examples such as the scalar (Nambu) string [63], the Ramond string [64], and the Neveu-Schwarz (NS) string [65] to more modern examples such as Polyakov’s “rigid string” [66], Green’s “Dirichlet string” [67], and the Polchinski-Strominger “effective string” [68]. Many other possibilities and variants have also been proposed.

All of these theories begin by imagining a one-dimensional line of flux energy (*i.e.*, a string) which sweeps out a two-dimensional flux-sheet (or worldsheet) as it sweeps through an external D -dimensional spacetime. Here D is the number of spacetime dimensions which are effectively uncompactified with respect to the fundamental energy scale M_s associated with the flux tube. As such, as it propagates, our string/flux tube is free to fluctuate into any of the $D_\perp \equiv D - 2$ spatial dimensions transverse to the string. We can describe such fluctuations by specifying D_\perp embedding functions $X^i(\sigma_1, \sigma_2)$, $i = 1, \dots, D_\perp$, which are nothing but the transverse spacetime locations of any point on the flux-tube worldsheet with coordinates (σ_1, σ_2) . As such, these embedding functions may be regarded as fields on the two-dimensional flux-tube worldsheet. The dynamics of this system is then governed by the Polyakov action

$$S \sim M_s^2 \int d^2\sigma \sum_{i=1}^{D_\perp} \left(\frac{\partial}{\partial \sigma^\alpha} X^i \right) \left(\frac{\partial}{\partial \sigma_\alpha} X^i \right) . \quad (4.6)$$

Minimizing this action is classically equivalent to minimizing the area of the flux-tube worldsheet.

By itself, the expression in Eq. (4.6) describes the action of the so-called D_\perp -dimensional “scalar” string. In some sense this theory provides the simplest possible description of a strongly-interacting flux tube, with the term in Eq. (4.6) representing the bare minimum that must always be present for any flux-tube description. The various possible refinements of this basic theory then differ in the extra terms that might be added to this action. Some of these theories mentioned above introduce extra terms which correspond to additional, purely internal degrees of freedom [*e.g.*, additional fields analogous to $X^i(\sigma_1, \sigma_2)$ but without interpretations as the coordinates of uncompactified spacetime dimensions] on the flux-tube worldsheet. By

contrast, other theories introduce extra interaction terms for the X^i -fields which alter their short-distance behavior.

The action in Eq. (4.6) can be interpreted as that of a two-dimensional (2D) field theory (where the two dimensions are those of the flux-tube worldsheet), and we immediately see that it is endowed with a 2D conformal symmetry. There are good reasons to expect that the long-distance limit of any self-consistent flux-tube theory should exhibit such a symmetry, since we expect the physics of this system to be invariant under reparametrizations of our flux-tube worldsheet coordinates. As a result, those flux-tube theories that augment the scalar string by introducing extra purely internal degrees of freedom on the flux-tube worldsheet must not break this conformal symmetry; this requirement constrains what kinds of terms can be added. By contrast, the theories that introduce extra interaction terms for the X^i fields do break this conformal symmetry, but they do so only in the short-distance limit. The 2D conformal symmetry of the long-distance limit is then preserved as an effective symmetry.

In any 2D conformal field theory, either exact or effective, the total number of degrees of freedom is encoded within the so-called *central charge* c . Each X^i field contributes a central charge $c = 1$, and thus the minimal scalar-string action in Eq. (4.6) describes a theory with central charge $c = D_\perp$. However the introduction of additional degrees of freedom on the flux-tube worldsheet will necessarily increase the central charge, producing a theory with $c > D_\perp$.

Given a particular action for our flux-tube dynamics, it is straightforward to quantize the fields in question. In this way, we can determine the corresponding spectrum of the theory at all mass levels. These calculations are standard in string theory (see, *e.g.*, Ref. [23–26]), and ultimately one obtains [60–62] asymptotic state degeneracies \hat{g}_n of the forms given in Eq. (4.3) or Eq. (4.5). Remarkably, one finds a relatively straightforward connection between the parameters (B, C) appearing in our state degeneracies and the parameters (D_\perp, c) of our underlying flux-tube theory [59–

62]:

$$\begin{cases} B = \frac{1}{4}(3 + D_{\perp}) \\ C = \pi\sqrt{2c/3} . \end{cases} \quad (4.7)$$

Indeed, for any value of B and C , we may regard the total central charge c as having two contributions: one contribution $c_{\text{fluc}} = D_{\perp}$ associated with the degrees of freedom associated with the transverse uncompactified spacetime fluctuations of the flux tube, and a remaining contribution

$$c_{\text{int}} \equiv c - D_{\perp} \equiv \frac{3C^2}{2\pi^2} - 4B + 3 \quad (4.8)$$

associated with those additional, purely internal degrees of freedom which might also exist within the full flux-tube theory (including those associated with any *compactified* spacetime dimensions which may also exist).

At first glance, it might seem that our dark sector must have $D_{\perp} = 2$, just as does our visible sector. This would certainly be true if our dark-sector flux tube were to experience the same spacetime geometry as does the visible sector. However, we emphasize that in a string-theoretic or “braneworld” context, the dark sector could correspond to physics in the “bulk” — *i.e.*, physics perpendicular to the brane on which the visible-sector resides. The degrees of freedom in the bulk would then be able to interact with those on the brane at most gravitationally, and would thus constitute dark matter by construction. However, the geometric properties of the bulk will generally differ from those of the brane — the bulk might contain not only extra spacetime dimensions which are effectively large (*i.e.*, uncompactified) with respect to the fundamental string scale, but also extra spacetime dimensions which are small (*i.e.*, compactified). The bulk may also be populated by additional fields with no spacetime interpretations at all. It is for this reason that we make no assumptions about the values of c or D_{\perp} associated with the dark sector.

Once our flux-tube theory is specified and the corresponding values of B and C determined, we may calculate the corresponding effective static-quark potential $V(R)$

between two quarks a distance R apart. We find [59]

$$\begin{aligned} V(R) &= \left(\frac{M_s}{2\pi} \right) \sqrt{(M_s R)^2 - (C/2)^2} \\ &\approx \frac{M_s^2 R}{2\pi} - \frac{C^2}{16\pi} \frac{1}{R} + \dots \quad \text{for } R \gg M_s^{-1}. \end{aligned} \quad (4.9)$$

The first term in the final expression indicates a linear confinement potential, as expected; this is nothing but the classical energy in the flux tube. By contrast, the second term resembles a Coulomb term but is actually an attractive universal quantum correction (or Casimir energy) which arises due to the transverse zero-point vibrations of the flux tube.

For visible-sector hadrons, it is natural to take $D = 4$. As a result, the $D_\perp = 2$ scalar string with $c_{\text{int}} = 0$ (corresponding to $B = 5/4$ and $C = 2\pi/\sqrt{3} \approx 3.63$) is the “minimal” string that we expect to underlie all descriptions of the actual visible-sector QCD flux tube. In fact, it has been shown in Ref. [59] that this minimal $D_\perp = 2$ scalar string with $\kappa = 36$ provides an excellent fit to hadronic data, both for low energies (which are sensitive to the Casimir energy within the confinement potential) as well as high energies (which are governed by the asymptotic degeneracy of hadronic states and the corresponding Hagedorn temperature). As discussed in Ref. [59], this success — coupled with the appearance of the same quantity C in both places — provides a highly non-trivial test of the classical conformal invariance of the QCD string.

In this chapter, we shall imagine that our DDM ensemble of dark-sector hadrons mimics that of the visible-sector hadrons to the extent that it corresponds to a set of masses M_n and state degeneracies \hat{g}_n parametrized by the functional forms given in Eqs. (4.1) and (4.5). However, we shall not insist on an actual string interpretation governing our dark-sector confinement dynamics, and as discussed above we shall therefore regard B and C as free parameters which may be adjusted at will (subject to certain constraints to be discussed below). Nevertheless it is only when B and C correspond to appropriate values of D_\perp and c via the relations in Eq. (4.7) that we may describe our resulting spectrum as corresponding to that of a classically

self-consistent string moving in a specific geometry. Moreover, motivated by our experience with visible-sector hadrons, we shall continue to regard the special scalar-string case with $B = 5/4$ and $C = 2\pi/\sqrt{3}$ as our “minimal” theory, corresponding to the action in Eq. (4.6) with $D_\perp = 2$. Adjusting the value of B above or below $5/4$ can then be interpreted as changing the effective number of uncompactified spacetime dimensions felt by our dark-sector flux tube (*i.e.*, the number of uncompactified spacetime dimensions into which it can experience fluctuations), while increasing the value of C beyond $2\pi/\sqrt{3}$ corresponds to introducing additional purely internal degrees of freedom with central charge c_{int} into our flux-tube theory.

Note, in this regard, that the degrees of freedom associated with fluctuations into extra *compactified* spacetime dimensions count towards c_{int} rather than D_\perp . Thus, in terms of its effects on the dark sector, the act of compactifying a spacetime dimension to a radius below the associated string scale preserves the central charge c (and thus the coefficient C) and merely shifts the associated degrees of freedom from D_\perp to c_{int} . The resulting change in the asymptotic state degeneracies \hat{g}_n due to the change in B then reflects the appearance of new Kaluza-Klein resonances in the total flux-tube spectrum.

4.1.4 Constraints on parameters

Even though M_s , M_0 , B , and C are henceforth to be viewed as unrestricted quantities parametrizing our hadron-like DDM ensemble, they are nevertheless subject to certain self-consistency constraints.

First, we note that while the asymptotic form for \hat{g}_n in Eq. (4.5) is remarkably accurate within those regions of (B, C) parameter space for which actual string realizations exist, there are other regions of (B, C) parameter space within which this approximation provides unphysical results. For example, given that the expression for \hat{g}_n in Eq. (4.5) multiplies a growing Bessel function against a falling monomial, for any given value of B it is in principle possible for there to exist a critical value of C below which \hat{g}_n is not always monotonically increasing for all $n \geq 0$. Such a situation is clearly unphysical, implying that the number of accessible flux-tube states fails to

grow with the total energy in the flux tube. We therefore demand that

$$\hat{g}_{n+1} > \hat{g}_n \quad \text{for all } n \geq 0 . \quad (4.10)$$

Given that we have taken $\hat{g}_0 = 1$, it turns out throughout the parameter range of interest that this requirement is tantamount to demanding

$$\hat{g}_1 > 1 . \quad (4.11)$$

If we further wish to demand that our ensemble of dark “hadrons” admit a string-theoretic description, then certain additional consistency conditions on the parameters B and C must be satisfied as well. For example, since $D_\perp \in \mathbb{Z} > 0$ in any self-consistent string construction, we must have

$$B \in \mathbb{Z}/4 > 3/4 . \quad (4.12)$$

Likewise, as discussed above, any self-consistent string theory will also have $c \geq D_\perp$ (or $c_{\text{int}} \geq 0$), which in turn implies

$$C^2 \geq \frac{2\pi^2}{3}(4B - 3) . \quad (4.13)$$

There are, of course, further string-derived constraints that might be imposed. For example, the allowed set of worldsheet central charges c that can be realized in such non-critical string theories depends crucially on the types of string models under study and the types of conformal field theories used in their constructions. However, the constraints in Eqs. (4.12) and (4.13) can be taken as a minimal model-independent set of constraints that must be satisfied as a prerequisite to any possible string interpretation.

In Fig. 4.2, we indicate the region of (B, C) parameter space which is consistent with the constraints in Eqs. (4.11), (4.12), and (4.13). We emphasize that the first of these constraints must always be satisfied as a matter of internal self-consistency. By

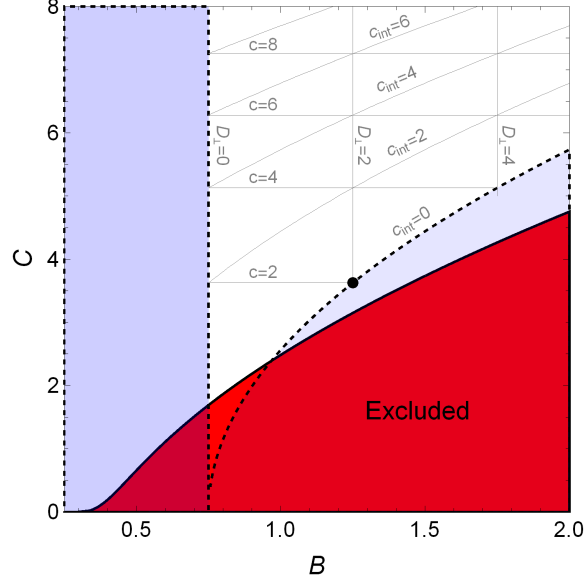


FIGURE 4.2: The region of (B, C) parameter space of interest for a DDM ensemble of dark “hadrons.” The red shaded region is excluded by the theoretical self-consistency condition $\hat{g}_1 \geq 1$. By contrast, the blue shaded regions are excluded by the constraint $B > 3/4$ as well as by the constraint in Eq. (4.13), and thus correspond to regions in which it would not be possible to interpret the ensemble constituents as the states of a quantized string. Note that locations for which $B \notin \mathbb{Z}/4$ would also suffer from this difficulty. Within the (unshaded) string-allowed region, we have indicated contours of D_\perp , c , and c_{int} , as defined in Eqs. (4.7) and (4.8). The black dot indicates the point in parameter space corresponding to the minimal $D_\perp = 2$ scalar string with $c_{\text{int}} = 0$. As demonstrated in Ref. [59], this model provides the best fit to the visible hadron spectrum.

contrast, as discussed above, the latter two conditions need to be satisfied only if one imposes the additional stipulation that our ensemble of dark “hadrons” admit a string-theory description. We observe in this connection that the first constraint is always weaker than the remaining string-motivated constraints. In other words, a string-based description with $B \in \mathbb{Z}/4 \geq 1$ is always guaranteed to have monotonically growing degeneracies \hat{g}_n . In Fig. 4.2 we also highlight the point $(B, C) = (5/4, 2\pi/\sqrt{3})$ corresponding to the “minimal” $D_\perp = 2$ scalar string. While this theory need not necessarily provide the best-fit description for our dark hadrons (as it does for the visible hadrons), its minimality nevertheless provides a useful benchmark for exploring the parameter space of our DDM model. Finally, we observe from Fig. 4.2 that our

combined constraints imply that

$$C \gtrsim 1.693 . \quad (4.14)$$

Indeed, this is the allowed range in C for which $\hat{g}_1 > 1$ when $B = 3/4$.

As an illustration of the results of this section, let us focus further on this “minimal” $D_\perp = 2$ scalar string. As noted above, the action for this string is given in Eq. (4.6). Quantizing this theory then gives rise to a discrete spectrum of states whose exact degeneracies are¹ $\hat{g}_{n \geq 0} = \{1, 2, 5, 10, 20, 36, 65, 110, 185, \dots\}$. Indeed it is only because of the existence of a quantized string formulation that we are even able to calculate the degeneracies of the corresponding ensemble from first principles. However, as we have asserted, these degeneracies are extremely well approximated by the expression in Eq. (4.5) with $(B, C) = (5/4, 2\pi/\sqrt{3})$. This is shown in Fig. 4.3, where we plot both the discrete exact degeneracies \hat{g}_n and the approximate functional form in Eq. (4.5). As evident from Fig. 4.3, our functional form matches these discrete values of \hat{g}_n extremely well for all values of $n \geq 0$ — even though the degeneracies \hat{g}_n are necessarily integers and even though our functional form was originally designed to be accurate only in the asymptotic $n \rightarrow \infty$ limit! Indeed, as claimed above, this functional form is accurate to within two percent over the entire range of n . This demonstrates the power of the functional form we have adopted, as well as the utility of an underlying string formulation for our flux tube.

4.2 Lifetimes and cosmological abundances for hadronic DDM ensembles

In the previous section, we discussed the spectra of our dark “hadronic” DDM ensembles. Our next step, then, is to consider the lifetimes and cosmological abundances of the individual states within these ensembles.

¹These degeneracies \hat{g}_n may be extracted as the coefficients of q^n in a small- q power-series expansion of the infinite product $\prod_n (1 - q^n)^{-2}$. With only minor modifications and a proper physical definition for q , this infinite product turns out to be the partition function of the $D_\perp = 2$ scalar string theory in Eq. (4.6).

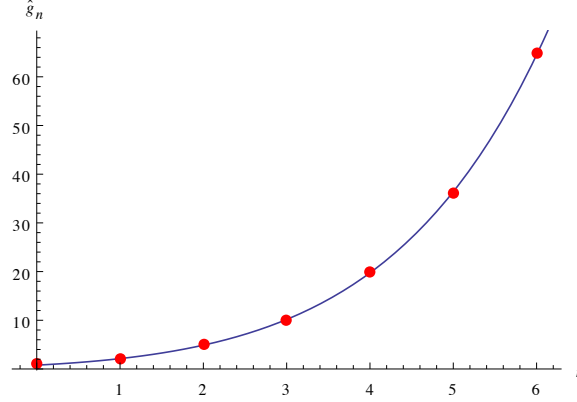


FIGURE 4.3: State degeneracies \hat{g}_n for the $D_\perp = 2$ scalar-string flux-tube model of Eq. (4.6) (red circles), with the asymptotic functional form in Eq. (4.5) superimposed (blue line). It is clear that our asymptotic functional form succeeds in modelling the state degeneracies extremely accurately all the way down to the ground state, as we shall require for our analysis.

4.2.1 Cosmological abundances

As we have seen, the degeneracy of states g_n for our ensemble of dark “hadrons” grows exponentially with the mass of the state, with asymptotic behavior $g_n \sim e^{\sqrt{n}} \sim e^{M_n/M_s}$. This exponential rise in the state degeneracies places severe constraints on the possible, physically consistent cosmological production mechanisms by which the corresponding abundances Ω_n might be established. Indeed, unless the corresponding abundances Ω_n fall sufficiently rapidly with n , our ensemble is likely to encounter severe phenomenological difficulties.

Fortunately, our interpretation of the individual components of such an ensemble as dark hadrons suggests a natural mechanism through which the corresponding abundances Ω_n are generated with an exponential suppression factor capable of overcoming this exponential rise in g_n . As we have discussed, we have been imagining that these dark “hadrons” emerge as the result of a dark-sector confining phase transition triggered by the strong interactions of some dark-sector gauge group G . This phase transition occurs when the temperature T in the dark sector drops below the critical temperature T_c associated with this phase transition. This event marks the time t_c at which the primordial abundances of our individual hadrons are established. Moreover, it is reasonable to assume that residual G interactions establish thermal

equilibrium among these hadrons at $T \sim T_c$. Thus, the primordial abundances Ω_n of our hadrons can be assumed to follow a Boltzmann distribution at $t = t_c$:

$$\Omega_n(t_c) \equiv \frac{\rho_n(t_c)}{\rho_{\text{crit}}(t_c)} = \frac{1}{3\widetilde{M}_P^2 H(t_c)^2} \int \frac{d^3\mathbf{p}}{(2\pi)^3} E_{\mathbf{p}} e^{-E_{\mathbf{p}}/T_c} \quad (4.15)$$

where $E_{\mathbf{p}} \equiv \sqrt{\mathbf{p} \cdot \mathbf{p} + M_n^2}$ and $\rho_{\text{crit}}(t) \equiv 3\widetilde{M}_P^2 H(t)^2$ where $\widetilde{M}_P \equiv M_P/\sqrt{8\pi} = 1/\sqrt{8\pi G_N}$ is the reduced Planck mass and $H(t)$ the Hubble parameter. Indeed, we may equivalently regard these abundances as emerging from an infinitely rapid succession of thermal freeze-outs. Evaluating Eq. (4.15) explicitly, we find

$$\begin{aligned} \Omega_n(t_c) = & X \left\{ (M_n T_c)^2 K_2(M_n/T_c) \right. \\ & \left. + \frac{1}{2} M_n^3 T_c \left[K_1(M_n/T_c) + K_3(M_n/T_c) \right] \right\} \end{aligned} \quad (4.16)$$

where $K_\nu(z)$ are modified Bessel functions of the second kind and where $X \equiv [6\pi^2 \widetilde{M}_P^2 H(t_c)^2]^{-1}$ is a common overall multiplicative factor.

In general, a given state with mass M produced at temperature T_c will be non-relativistic (behaving like massive matter) if $T_c \lesssim M$ and relativistic (behaving like radiation) otherwise. In such limiting cases, the abundances in Eqs. (4.15) and (4.16) take the simplified forms

$$\Omega_n(t_c) \approx \begin{cases} \sqrt{\pi/2} X M_n (M_n T_c)^{3/2} e^{-M_n/T_c} & \text{non-rel} \\ 6X T_c^4 & \text{rel.} \end{cases} \quad (4.17)$$

At first glance, it may seem that any value for T_c might be phenomenologically permissible. However, this production mechanism can only be self-consistent if it injects a finite total energy density into our system. In other words, as a bare minimum, we must require that

$$\Omega_{\text{tot}}(t_c) \equiv \sum_{n=0}^{\infty} g_n \Omega_n(t_c) < \infty. \quad (4.18)$$

However, this condition is sensitive to the behavior of the abundances $\Omega_n(t_c)$ for extremely large n , corresponding to states which are non-relativistic. For such states,

we see from Eq. (4.17) that $\Omega_n(t_c) \sim e^{-M_n/T_c}$. With $g_n \sim n^{-B} e^{C\sqrt{n}}$ as $n \rightarrow \infty$, we find using Eqs. (4.1) and (4.15) that Eq. (4.18) can only hold if

$$\frac{T_c}{M_s} \leq \frac{1}{C}. \quad (4.19)$$

This then becomes a hard bound on the allowed values of T_c , one which ensures that the Boltzmann exponential suppression factor in Eq. (4.15) ultimately overcomes the exponential rise in the degeneracy of states g_n . Indeed, Eq. (4.19) reflects nothing more than the statement that $T_c \leq T_H$, where $T_H \equiv M_s/C$ is the Hagedorn temperature of our dark ensemble. For the *visible* hadronic sector, one often assumes that T_c and T_H are related to each other parametrically, with T_c either directly identified as T_H or positioned not too far below T_H . We shall implicitly make the same assumption for the dynamics of our dark sector as well.

The next question is to determine which of our ensemble components are produced relativistically or non-relativistically at $T = T_c$. To do this, we shall henceforth assume that $T_c, M_s, M_0 > T_{\text{MRE}}$ where t_{MRE} and T_{MRE} are the time and temperature associated with matter-radiation equality. This assumption, which parallels what occurs for the hadrons of the visible sector, ensures that our abundances $\Omega_n(t)$ are established during the radiation-dominated era prior to matter-radiation equality and that all ensemble constituents have become effectively non-relativistic by t_{MRE} . Note that the assumption that $T_c > T_{\text{MRE}}$ follows from our expectation that our dark degrees of freedom prior to t_c (*i.e.*, prior to “hadronization” in the dark sector) are likely to be relativistic, thereby reinforcing the radiation-dominated nature of the era prior to T_{MRE} and making matter-radiation equality impossible to achieve using only visible-sector matter, as would have been required had we taken $T_c < T_{\text{MRE}}$. Similarly, the assertion that $M_s > T_c$ follows directly from our assumption that $T_c > T_{\text{MRE}}$, given the constraints in Eqs. (4.14) and (4.19). Finally, although it is not impossible to imagine self-consistent scenarios in which $M_0 < T_{\text{MRE}}$, taking $M_0 > T_{\text{MRE}}$ also helps to preserve t_{MRE} at its standard cosmological value. We shall nevertheless make no assertion regarding the relative sizes of M_0 and T_c .

The above assumptions enable us to determine which of the components of our ensemble are relativistic or non-relativistic at $T = T_c$. To do this, we simply compare T_c against the ensemble masses M_n given in Eq. (4.1). Given the constraint in Eq. (4.19), it is straightforward to demonstrate that

$$T_c \leq \frac{M_s}{C} \leq \frac{M_1}{C} . \quad (4.20)$$

Since $C > 1$ [as follows from Eq. (4.14)], we conclude that *all of our ensemble components with $n \geq 1$ are necessarily non-relativistic at $t = t_c$. By contrast, the $n = 0$ component will be relativistic at $t = t_c$ if $T_c \gtrsim M_0$, and non-relativistic otherwise.*

Eq. (4.15) describes the abundances of our dark-sector hadrons at the time t_c when these hadrons come into existence as the result of a dark-sector confining transition. However, once established, these abundances then evolve non-trivially with time as a result of two effects. The first of these is Hubble expansion; the second is particle decay. We shall treat each of these effects separately.

In order to evaluate the effect of Hubble expansion on the abundances $\Omega_n(t)$, we shall assume a standard cosmological history in which the universe remains radiation-dominated (RD) from very early times up to the time t_{MRE} of matter-radiation equality. We shall also approximate the universe as matter-dominated (MD) throughout the subsequent epoch. In general, we recall that the abundance $\Omega(t)$ of non-relativistic matter scales as $t^{1/2}$ during an RD epoch but remains constant in an MD epoch; by contrast, the abundance of relativistic matter remains constant during an RD epoch but scales as $t^{-2/3}$ during an MD epoch. Likewise, we recall that the temperature T of the universe scales as $T \sim t^{-1/2}$ during RD but $T \sim t^{-2/3}$ during MD. Thus any ensemble component of mass M which is “born” relativistic at $T = T_c \gg M$ will eventually transition to non-relativistic behavior as the temperature ultimately drops below $T \sim M$.

Collecting these observations, we then find that the net effect of Hubble expansion is to rescale the original abundance of given state of mass M by a factor which depends on whether that state was non-relativistic or relativistic at the time t_c of its

production:

$$\Omega(t) = \Omega(t_c) \times \begin{cases} \sqrt{t_{\text{MRE}}/t_c} & \text{non-rel} \\ \sqrt{t_{\text{MRE}}/t_M} & \text{rel} \end{cases} \quad (4.21)$$

where t_M denotes the time at which $T = M$. Note that this result is valid for any time $t \geq t_{\text{MRE}}$. Since it follows from our assumptions that $t_c, t_M < t_{\text{MRE}}$, we see that the abundances of all of our ensemble states are necessarily *enhanced* before reaching the current MD era. However, as evident from Eq. (4.21), these abundances are not enhanced equally: the abundance of a non-relativistic component is enhanced more greatly than that of any relativistic component of mass M a factor $\sqrt{t_M/t_c}$.

We have already seen that the states with $n \geq 1$ are all non-relativistic, while the $n = 0$ ground state is either relativistic or non-relativistic depending on the value of M_0/T_c . Thus, putting all of the pieces together, we find for all $n \geq 1$ that

$$\begin{aligned} \Omega_n(t) &= \sqrt{\frac{\pi}{2}} X M_n (M_n T_c)^{3/2} e^{-M_n/T_c} \sqrt{\frac{t_{\text{MRE}}}{t_c}} \\ &= \sqrt{\frac{\pi}{2}} X \left(\frac{g_c}{g_{\text{MRE}}} \right)^{1/4} \frac{(M_n T_c)^{5/2}}{T_{\text{MRE}}} e^{-M_n/T_c} \\ &= \sqrt{\frac{\pi}{72}} \frac{1}{g_c^{3/4} g_{\text{MRE}}^{1/4}} \frac{M_n^{5/2}}{T_c^{3/2} T_{\text{MRE}}} e^{-M_n/T_c} . \end{aligned} \quad (4.22)$$

Note that in passing to the second line we have exploited the standard time/temperature relationship suitable for an RD epoch, specifically²

$$t = \sqrt{\frac{\pi}{32}} g_*(T)^{-1/2} \frac{M_P}{T^2} , \quad (4.23)$$

where $g_*(T)$ tallies the number of effectively relativistic degrees of freedom driving the Hubble expansion at any temperature T , with $g_\alpha \equiv g_*(T_\alpha)$. Likewise, in passing to the final line of Eq. (4.22) we have recognized that $H = 1/(2t)$ for an RD epoch, from which it follows that $X = 1/(6g_c T_c^4)$.

² Note that the factor of $\sqrt{\pi/32}$ in Eq. (4.23) is consistent with our adoption of Boltzmann statistics in Eq. (4.15); for Bose-Einstein statistics this would instead become $\sqrt{45/16\pi^3}$.

For $n = 0$, however, the corresponding cosmological abundance is given by

$$\Omega_0(t) = \begin{cases} \sqrt{\frac{\pi}{72}} \frac{1}{g_c^{3/4} g_{\text{MRE}}^{1/4}} \frac{M_0^{5/2}}{T_c^{3/2} T_{\text{MRE}}} e^{-M_0/T_c} & T_c \lesssim M_0 \\ \frac{1}{g_c} \left(\frac{g_{M_0}}{g_{\text{MRE}}} \right)^{1/4} \left(\frac{M_0}{T_{\text{MRE}}} \right) & T_c \gtrsim M_0 . \end{cases} \quad (4.24)$$

As expected, the cosmological abundances in Eqs. (4.22) and (4.24) depend non-trivially on the three mass scales which parametrize our dark-hadron mass spectrum, namely M_0 , T_c , and M_s (the latter appearing implicitly through M_n). They also depend on the fixed mass scale T_{MRE} . However, if we disregard the numerical g -factors which appear in these results and which only serve to parametrize the external time/temperature relationship, we see that the *ratios* between these abundances depend only on the *ratios* between our input mass scales. In particular, such abundance ratios are no longer anchored to a fixed external mass scale such as T_{MRE} . To make this point explicit, let us define the dimensionless quantities

$$r \equiv \frac{M_0}{M_s} \quad \text{and} \quad s \equiv \frac{T_c}{M_s} \quad (4.25)$$

and imagine that $g_*(T)^{1/4}$ does not change significantly between T_c and M_0 . (Note, indeed, that $g^{1/4}$ varies much more slowly than g .) We then find from Eqs. (4.22) and (4.24) that

$$\frac{\Omega_{n \geq 1}(t)}{\Omega_0(t)} = \begin{cases} \frac{(n + r^2)^{5/4}}{r^{5/2}} e^{-(\sqrt{n+r^2}-r)/s} & s \lesssim r \\ \sqrt{\frac{\pi}{72}} \frac{(n + r^2)^{5/4}}{r s^{3/2}} e^{-\sqrt{n+r^2}/s} & s \gtrsim r . \end{cases} \quad (4.26)$$

Thus, up to an overall rescaling factor Ω_0 , we see that all of our abundances Ω_n depend purely on the dimensionless ratios r and s . It then follows that the cosmological abundance of each state in our dark-hadron ensemble is determined once Ω_0 is anchored to a particular numerical value and specific values of r and s are chosen.

This observation will be important in what follows.

4.2.2 Lifetimes and decays

As indicated above, our derivation of the dark-sector cosmological abundances $\Omega_n(t)$ has thus far disregarded the effects of particle decays. In other words, we have implicitly assumed that each ensemble component is absolutely stable once produced at T_c . As our final step, we shall therefore now incorporate the effects of such decays into our analysis. In doing so, we shall make several simplifying assumptions. First, we shall assume that the net injection of energy density in the form of radiation from these decays has a negligible effect on the total radiation-energy density of the universe. Hence, this effect decouples from the effect of Hubble expansion. Second, we shall further assume that the contribution to the total decay width Γ_n of each ensemble constituent from intra-ensemble decays is negligible. In other words, we shall assume that Γ_n is dominated by decays to visible-sector final states which do not include lighter ensemble constituents. We shall discuss the consequences of relaxing this assumption in Sect. 4.5. Third, we shall assume that all states at a given mass level n share a common decay width Γ_n , and that this width scales with n across our dark-hadron ensemble according to

$$\Gamma_n = \Gamma_0 \left(\frac{M_n}{M_0} \right)^\xi \quad (4.27)$$

where M_n are the dark-hadron masses in Eq. (4.1) and where Γ_0 (or, equivalently, the corresponding lifetime τ_0) and the scaling exponent $\xi > 0$ are taken to be additional free parameters of our model. Thus each state in our dark-sector ensemble has a lifetime $\tau_n \equiv 1/\Gamma_n$ given by

$$\tau_n = \tau_0 \left(\frac{n}{r^2} + 1 \right)^{-\xi/2}. \quad (4.28)$$

Finally, for simplicity, we shall imagine that all states with lifetimes τ_n indeed actually decay at $t = \tau_n$.

Under these assumptions, the abundance $\Omega_n(t)$ of any ensemble constituent at any time $t \geq t_c$ is given by the expressions quoted above, but now multiplied by an additional decay factor

$$e^{-(t-t_c)/\tau_n} \approx e^{-(\sqrt{n+r^2}/r)^\xi t/\tau_0} \quad (4.29)$$

where we have approximated $t \gg t_c$. For $s \lesssim r$, we thus have

$$\Omega_{n \geq 1}(t) = \Omega_0^{(\text{NR})}(t) \frac{(n+r^2)^{5/4}}{r^{5/2}} \mathcal{E}_n^{(\text{NR})}(t) \quad (4.30)$$

where

$$\mathcal{E}_n^{(\text{NR})}(t) \equiv e^{-(\sqrt{n+r^2}-r)/s - [(\sqrt{n+r^2}/r)^\xi - 1]t/\tau_0} \quad (4.31)$$

and where

$$\Omega_0^{(\text{NR})}(t) = \sqrt{\frac{\pi}{72}} \frac{1}{g_c^{3/4} g_{\text{MRE}}^{1/4}} \left(\frac{r}{s}\right)^{3/2} \left(\frac{M_0}{T_{\text{MRE}}}\right) e^{-r/s - t/\tau_0} . \quad (4.32)$$

By contrast, for $s \gtrsim r$, we have

$$\Omega_{n \geq 1}(t) = \sqrt{\frac{\pi}{72}} \Omega_0^{(\text{R})}(t) \frac{(n+r^2)^{5/4}}{r s^{3/2}} \mathcal{E}_n^{(\text{R})}(t) \quad (4.33)$$

where

$$\mathcal{E}_n^{(\text{R})}(t) \equiv e^{-\sqrt{n+r^2}/s - [(\sqrt{n+r^2}/r)^\xi - 1]t/\tau_0} \quad (4.34)$$

and where

$$\Omega_0^{(\text{R})}(t) = \frac{1}{g_c} \left(\frac{g_{M_0}}{g_{\text{MRE}}}\right)^{1/4} \left(\frac{M_0}{T_{\text{MRE}}}\right) e^{-t/\tau_0} . \quad (4.35)$$

4.3 Cosmological constraints on the dark-hadron ensemble

Having determined the abundances and lifetimes of each of the individual components of our dark-hadron DDM ensemble, we now proceed to study the overall properties of our ensemble and its behavior as a function of time. However, as we shall see, many of the phenomenological properties and constraints that apply to such an ensemble do not rest upon the properties of the individual ensemble components *per se*, but

rather upon various aggregate quantities that collectively describe the ensemble as a whole. Accordingly, in this section we shall begin by describing three aggregate quantities which ultimately play the most important roles in characterizing and constraining such dark-hadron DDM ensembles. We shall then discuss some of the most immediate cosmological constraints that can be placed upon these quantities.

4.3.1 Total abundance, tower fraction, and effective equation of state

Perhaps not surprisingly, the first aggregate property of a given dark-hadron DDM ensemble that shall concern us is its total abundance

$$\Omega_{\text{tot}}(t) \equiv \sum_{n=0}^{\infty} g_n \Omega_n(t) = \kappa \sum_{n=0}^{\infty} \hat{g}_n \Omega_n(t). \quad (4.36)$$

Given our results in Eqs. (4.30) and (4.33), this total abundance takes the form

$$\Omega_{\text{tot}}(t) = \begin{cases} \kappa \Omega_0^{(\text{NR})}(t) \left[1 + \sum_{n=1}^{\infty} \hat{g}_n \frac{(n+r^2)^{5/4}}{r^{5/2}} \mathcal{E}_n^{(\text{NR})}(t) \right] & s \lesssim r \\ \kappa \Omega_0^{(\text{R})}(t) \left[1 + \sqrt{\frac{\pi}{72}} \sum_{n=1}^{\infty} \hat{g}_n \frac{(n+r^2)^{5/4}}{r s^{3/2}} \mathcal{E}_n^{(\text{R})}(t) \right] & s \gtrsim r \end{cases} \quad (4.37)$$

where $\Omega_0^{(\text{NR,R})}(t)$ are given in Eqs. (4.32) and (4.35). Indeed, we further note from Eqs. (4.32) and (4.35) that

$$\Omega_0^{(\text{NR,R})}(t) = e^{-(t-t_{\text{now}})/\tau_0} \Omega_0^{(\text{NR,R})}(t_{\text{now}}) \quad (4.38)$$

where $t_{\text{now}} \approx 4 \times 10^{17}$ s denotes the current age of the universe. We thus see from Eqs. (4.37) and (4.38) that the overall magnitude of $\Omega_{\text{tot}}^{(\text{NR,R})}(t)$ can be viewed as being set by the single number $\Omega_0^{(\text{NR,R})}(t_{\text{now}})$.

In characterizing the properties of our DDM ensemble and how they evolve with time, we are certainly interested in tracking $\Omega_{\text{tot}}(t)$. However, we are also interested

in tracking the *distribution* of this total abundance among the individual ensemble constituents. One quantity of particular interest that provides essential information about this distribution is the so-called “tower fraction” $0 \leq \eta(t) \leq 1$ originally introduced in Ref. [40]. This quantity is typically defined in the DDM literature as the fraction of the abundance carried by all ensemble components *other* than the dominant component, where the dominant component is the one making the largest individual contribution to $\Omega_{\text{tot}}(t)$. As such, the quantity η tracks the degree to which a single component carries the bulk of the total abundance. When η is close to zero, our ensemble effectively resembles a traditional single-component dark-matter setup. By contrast, when η differs significantly from zero, our ensemble is more truly “DDM-like”, with many of the ensemble constituents playing a non-trivial role in together shaping the properties of the dark sector.

Such a definition for η is appropriate in cases in which each ensemble constituent has a unique mass and lifetime. Indeed, this has often been the case for the types of DDM ensembles previously studied. However, for the dark-hadron DDM ensembles on which we are focusing here, the states at a given Regge level n have been assumed to have essentially equal masses and lifetimes. Thus, in this chapter, we shall adopt a modified definition for $\eta(t)$ in which the comparison is made between the aggregate abundance contributions that accrue *level by level* rather than state by state. Specifically, we define

$$\hat{\Omega}_n(t) \equiv g_n \Omega_n(t) \quad (4.39)$$

as the aggregate cosmological abundance arising from all states at a particular oscillator level n . In terms of these aggregate abundances, we then define

$$\eta(t) \equiv 1 - \frac{\max_n \{\hat{\Omega}_n(t)\}}{\Omega_{\text{tot}}(t)} . \quad (4.40)$$

Thus we continue to have $0 \leq \eta(t) \leq 1$, with $\eta \approx 0$ signifying a dark sector resembling traditional single-component dark matter and $\eta > 0$ indicating (and quantifying) a DDM-like departure from this traditional scenario.

At first glance, one might assume that the $n = 0$ ground state(s) must always yield

the largest aggregate abundance $\widehat{\Omega}_n(t)$ because the primordial abundances $\Omega_n(t)$ for the states at all higher levels $n > 0$ are exponentially suppressed by the corresponding Boltzmann factor in Eq. (4.15). However, for the DDM ensembles of dark hadrons studied here, it often turns out that the Hagedorn-like exponential growth of the degeneracies g_n as a function of n can more than compensate for the Boltzmann suppression for small values of n . Indeed, this is true even for combinations of the ensemble parameters B, C, r , and s which satisfy the consistency conditions discussed in Sect. 4.1 and which yield a finite value of $\Omega_{\text{tot}}(t_c)$. As a result of this net balancing between these two competing exponential effects, the level carrying the greatest aggregate cosmological abundance $\widehat{\Omega}_n(t)$ need not always be the $n = 0$ ground state. It need not even be fixed as a function of time. This possibility must therefore be taken into account when evaluating $\eta(t)$.

Finally, another important quantity which can be taken to characterize our dark sector is the so-called equation-of-state parameter w . For a single-component dark sector, this quantity is nothing but the ratio between the pressure p and energy density ρ of the dark component: $p = w\rho$. However, we are dealing here with a multi-component dark sector in which each component has its own individual lifetime and abundance. As a result, the total energy density and pressure associated with our dark sector will generally experience a rather non-trivial time dependence which causes our ensemble as a whole to behave collectively as if it had a non-trivial w — even if each individual component is taken to be pure matter with $w = 0$.

To describe these collective effects, we therefore define [40] an *effective* equation-of-state parameter $w_{\text{eff}}(t)$ which describes the behavior of our ensemble as a single collective entity:

$$w_{\text{eff}}(t) \equiv - \left(\frac{1}{3H} \frac{d \log \rho_{\text{tot}}}{dt} + 1 \right) . \quad (4.41)$$

Here H is the Hubble parameter and $\rho_{\text{tot}} = 3\widetilde{M}_P H^2 \Omega_{\text{tot}}$ is the total energy density of the ensemble. Note that the definition in Eq. (4.41) is nothing but the usual definition of w prior to any assumptions of dark-sector minimality. As discussed above, we are primarily concerned with the evolution of the ensemble during the present matter-

dominated epoch, within which $H(t) \approx 2/(3t)$. Thus, the effective equation-of-state parameter for our DDM ensemble within this epoch is given by

$$w_{\text{eff}}(t) = - \frac{t}{2\Omega_{\text{tot}}} \frac{d\Omega_{\text{tot}}(t)}{dt} . \quad (4.42)$$

As discussed in Sect. 4.2, the only explicit dependence of $\Omega_{\text{tot}}(t)$ on t within a matter-dominated epoch is due to the exponential decay factor (4.29) within each individual abundance $\Omega_n(t)$. We thus find that

$$w_{\text{eff}}(t) = \frac{t}{2\tau_0\Omega_{\text{tot}}(t)} \sum_{n=0}^{\infty} g_n \left(\frac{\sqrt{n+r^2}}{r} \right)^{\xi} \Omega_n(t) . \quad (4.43)$$

Note that even though each of the individual components of our ensemble has been taken to be matter-like (with $w = 0$), the collective equation-of-state parameter $w_{\text{eff}}(t)$ for our ensemble as a whole is *positive*, reflecting the fact that the ensemble as a whole is continually losing abundance as its individual components decay. Indeed, it is only in the $\tau_0 \rightarrow \infty$ limit that $w_{\text{eff}}(t) \rightarrow 0$. As we shall see in Sect. 4.3.2, $w_{\text{eff}}(t)$ plays an important role in constraining the parameter space of these DDM ensembles.

4.3.2 Cosmological constraints

Given our time-dependent aggregate quantities $\Omega_{\text{tot}}(t)$, $\eta(t)$, and $w_{\text{eff}}(t)$, we now turn to the cosmological constraints that bound these functions. In this way, we shall ultimately be placing non-trivial constraints on the parameter space underlying these hadronic DDM ensembles.

In this connection, we again stress that our aim in this chapter is not to perform a detailed analysis of the astrophysical and/or cosmological constraints on this parameter space. Such a detailed analysis would clearly be an important but extensive task which is beyond the scope of this chapter. Moreover, such an analysis would require a host of further assumptions concerning the particular nature of our ensemble, the specific decay modes of its constituents into SM states, and so forth. Rather, in this chapter, our goal is to simply to obtain a rough initial sense of those regions of

parameter space in which a DDM ensemble of dark “hadrons” might have at least the *potential* of phenomenological viability. Accordingly, in what follows, we shall put forth a set of requirements which directly constrain the fundamental quantities $\Omega_{\text{tot}}(t)$, $\eta(t)$, and $w_{\text{eff}}(t)$ we have defined above, but which do not require any further information concerning these hadronic ensembles beyond those properties already discussed. In some sense, then, these might be viewed as the immediate “zeroth-order” model-independent constraints that any DDM ensemble of this sort must satisfy.

Our first constraint is an obvious one: despite the presence of an infinite tower of dark-hadronic resonances, each with its own cosmological abundance and lifetime, we shall demand that

$$\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26 . \quad (4.44)$$

This requirement is clearly predicated on the assumption that our dark-hadronic ensemble represents the totality of the dark sector; for other cases we would simply require that $\Omega_{\text{tot}}(t_{\text{now}}) \lesssim 0.26$. As we shall see, in either situation this is a severe and unavoidable constraint which ultimately “anchors” our entire construction in terms of actual numbers and mass scales.

Second, we may also consider the *time-variation* of $\Omega_{\text{tot}}(t)$. The time-variation of this total abundance is constrained by experimental probes which yield information about the dark-matter abundance during different cosmological epochs. For example, CMB data [5] provides information about the dark-matter abundance around the time of last scattering — *i.e.*, at a redshift $z \approx 1100$, or equivalently a time of roughly $2.7 \times 10^{-5} t_{\text{now}}$. On the other hand, observational data on baryon acoustic oscillations [69] and the relationship between luminosity and redshift for Type Ia supernovae [70] provide information about $H(t)$ and the dark-energy abundance Ω_{Λ} at subsequent times, down to redshifts of around $z \approx 0.5$. Within the context of the Λ CDM cosmology, the agreement between these different measurements implies that the dark-matter abundance has not changed dramatically since the time of last scattering.

In order to be consistent with this result, we shall therefore demand that the total

abundance of our DDM ensemble not vary by more than 5% between an early “look-back” time t_{LB} and today:

$$\frac{\Omega(t) - \Omega(t_{\text{now}})}{\Omega(t_{\text{now}})} \leq 0.05 \quad \text{for all } t_{\text{LB}} \leq t \leq t_{\text{now}} . \quad (4.45)$$

In what follows, we shall choose a look-back time $t_{\text{LB}} = 10^{-6} t_{\text{now}}$, which lies comfortably before the recombination epoch.

In addition to these constraints on the time-variation of the dark-matter abundance, there are further considerations which constrain the decays of the DDM-ensemble constituents more directly. These constraints depend on the decay properties of the dark-sector particles and are thus ultimately model-dependent. However, for those rather general cases in which the ensemble constituents can decay to final states involving visible-sector particles, one must ensure that these decay products not disrupt big-bang nucleosynthesis [71], not produce observable distortions in the CMB [72, 73], not reionize the universe [74], and not violate current limits on the fluxes of photons or other cosmic-ray particles [75, 76]. Indeed, even if the ensemble constituents decay exclusively into other, lighter dark-sector particles, such decays can nevertheless leave observable imprints on small-scale structure [77, 78], alter the scale- and redshift-dependence of the cosmological gravitational-lensing power spectrum [79], and affect the luminosity-redshift relation for Type Ia supernovae [80, 81]. Since these effects all arise from the decays of ensemble constituents, non-observation of these effects also leads to constraints on the time-variation of Ω_{tot} .

Some of these latter constraints admittedly depend on model-dependent aspects of the decay kinematics of the dark-ensemble constituents. However the strongest and most general of these constraints effectively amount to limits on the variation of $\Omega_{\text{tot}}(t)$ within the recent past — *i.e.*, for redshifts $0 \lesssim z \lesssim 3$. Therefore, in addition to our look-back-time constraint in Eq. (4.45), we shall also impose an additional constraint on our effective equation-of-state parameter:

$$w_{\text{eff}}(t_{\text{now}}) \leq 0.05 . \quad (4.46)$$

Through Eq. (4.42), this thus becomes a constraint on the present-day *time-derivative* of $\Omega_{\text{tot}}(t)$. It is important to stress that this constraint is independent of that in Eq. (4.45): while Eq. (4.45) constrains accumulated changes in $\Omega_{\text{tot}}(t)$ over a relatively long interval, Eq. (4.46) constrains the time-variation of $\Omega_{\text{tot}}(t)$ near the present time.

Other considerations will also guide our interest in certain regions of parameter space. For example, from a DDM-inspired standpoint, we are particularly interested in scenarios for which

$$\eta(t_{\text{now}}) \sim \mathcal{O}(1) , \quad (4.47)$$

i.e., scenarios in which the present-day value of η is significantly different from zero. This ensures that a sizable number of ensemble constituents continue to survive and contribute meaningfully to Ω_{tot} at the present time, with dark-matter decays occurring *throughout* the present epoch and not just in the distant past or future. Although Eq. (4.47) is not a strict requirement for phenomenological consistency, this condition guides the degree to which we may regard our ensemble as being fully DDM-like, with a significant portion of the ensemble playing a non-trivial role in the phenomenology of the dark sector. For example, this condition rules out regions of parameter space in which $\tau_n \ll \tau_0$ for all $n \geq 1$, with $\tau_1 \ll t_{\text{LB}}$. In such regions of parameter space, all excited dark-hadronic states have decayed prior to our look-back time, leaving us with a single dark-hadronic ground state in the present epoch. Such a scenario trivially satisfies all of our phenomenological constraints on the time-variations of the total dark-sector abundance, but is effectively no different from that of a traditional, single-component dark sector. It is thus less interesting from a DDM perspective.

There are two further phenomenological constraints which will be useful for us to consider in the following. First, we shall demand that $\tau_0 \gg t_{\text{now}}$. Although we do not necessarily require $\tau_0 \approx 10^9 t_{\text{now}}$ as in traditional single-component dark sectors, we generally expect that τ_0 must exceed t_{now} by at least several orders of magnitude in order to satisfy look-back and w_{eff} constraints. This assumption will be discussed further in Sect. 4.4. Likewise, although we have thus far assumed $M_0 \geq T_{\text{MRE}}$ throughout our analysis, we actually must impose the somewhat stronger bound

$M_0 \gtrsim \mathcal{O}(10^3)T_{\text{MRE}} \approx \mathcal{O}(\text{keV})$ in order to satisfy BBN and structure-formation constraints. This last requirement implicitly assumes that our lightest ensemble component carries the largest cosmological abundance (or at least a sizable fraction of the total cosmological abundance), but we shall see in Sect. 4.4 that this turns out to be true for the vast majority of phenomenologically interesting cases.

Finally, we shall also make certain simplifying assumptions. First, for concreteness, we shall restrict our attention to situations with $\xi = 3$. In other words, we shall assume that the dominant contributions to the decay lifetimes τ_n of our DDM constituents ϕ_n scale as $\tau_n \sim 1/M_n^3$ across the DDM ensemble. Decay widths of the form $\Gamma_n \sim M_n^3/\Lambda^2$ emerge naturally from operators such as $\phi_n F_{\mu\nu} F^{\mu\nu}/\Lambda$ where Λ parametrizes the energy scale associated with such couplings and where $F^{\mu\nu}$ denotes a field-strength tensor associated with either the visible-sector (SM) photon or a dark-radiation photon associated with an additional Abelian gauge group under which the ensemble constituents are not charged. The contributions from such operators will dominate the decays of our DDM constituents in scenarios in which our DDM ensemble is uncharged with respect to all SM symmetries, and in which intra-ensemble decays can be neglected. Likewise, we shall also make the simplifying assumption that $\kappa = 1$ in Eq. (4.2). This restricts us to the bare “minimal” case in which we do not ascribe non-trivial degrees of freedom to our dark-sector quarks, and thereby focus exclusively on the ensemble of states generated by our infinite tower of hadronic resonances. Finally, throughout our analysis, we shall continue to impose the self-consistency constraints listed in Eqs. (4.10) [or equivalently (4.11)], (4.12), (4.13), and (4.19).

Thus, going forward, the free parameters governing our dark-hadron DDM ensemble may be tallied as follows. First, there are the two parameters $\{B, C\}$ which govern the individual state degeneracies \hat{g}_n according to Eq. (4.5). Second, there are the four parameters $\{r, s, M_0, \tau_0\}$ which govern the individual abundances $\Omega_n(t)$ in Eqs. (4.29) through (4.35). However, imposing Eq. (4.44) as an overall normalization condition allows us to remove M_0 as a free parameter. Thus, for the rest of this chapter, we shall consider our DDM ensembles as functions of their locations within

the five-dimensional parameter space corresponding to the variables $\{B, C, r, s, \tau_0\}$ where $B \geq 1$, $C^2 \geq 2\pi^2(4B - 3)/3$, and $s \leq 1/C$.

4.4 Results

In general, we seek to determine which values of our defining parameters $\{B, C, r, s, \tau_0\}$ lead to self-consistent and potentially viable dark sectors — *i.e.*, sectors which satisfy our abundance, look-back, and w_{eff} constraints in Eqs. (4.44), (4.45), and (4.46) respectively, along with our $M_0 > \mathcal{O}(\text{keV})$ constraint. For each such set, we also seek to determine the corresponding values of relevant mass scales such as the string scale M_s . We also seek to determine the extent to which the corresponding ensemble is truly DDM-like, with a relatively large number of component states playing a significant role in the phenomenology of the dark sector and contributing to Ω_{tot} at the present time. In general, the larger the value of $\eta(t_{\text{now}})$, the more DDM-like the corresponding ensemble.

At first glance, it might seem rather daunting to orient ourselves within the five-dimensional $\{B, C, r, s, \tau_0\}$ parameter space. However, there are really two separate parts to our analysis — one part which depends only on *relative* mass scales, and one part which makes explicit reference to *absolute* mass scales. It is clear from Eqs. (4.37) and (4.38) that once we know $\{B, C, r, s, \tau_0\}$, we can determine the function $\Omega_{\text{tot}}^{(\text{R,NR})}(t)$ up to an overall multiplicative constant $\Omega_0^{(\text{R,NR})}(t_{\text{now}})$. Setting $\Omega_{\text{tot}}^{(\text{R,NR})}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$ therefore immediately determines a required numerical value of $\Omega_0^{(\text{R,NR})}(t_{\text{now}})$. This also determines the corresponding values of $\eta(t_{\text{now}})$ and $w_{\text{eff}}(t_{\text{now}})$. Up to this point, we have not yet anchored our results in terms of absolute mass scales. However, this can also easily be done: we simply set our required numerical value of $\Omega_0^{(\text{R,NR})}(t_{\text{now}})$ to the expression in either Eq. (4.32) or Eq. (4.35). This then determines an absolute value for the mass scale M_0 , whereupon we find that $M_s = rM_0$ and $T_c = (s/r)M_0$. Thus, in this way, we can extract the values for M_s and $\eta(t_{\text{now}})$ corresponding to every point in the $\{B, C, r, s, \tau_0\}$ parameter space.

Certain observations can be made rather rapidly. For example, given Eq. (4.44), it immediately follows that $\Omega_0(t_{\text{now}}) \lesssim 0.26$ — a bound which can be saturated only

when $\eta(t_{\text{now}}) = 0$. More generally and more schematically, we might write this constraint in the rough order-of-magnitude form

$$\Omega_0(t_{\text{now}}) \lesssim \mathcal{O}(0.1) . \quad (4.48)$$

However, let us now consider the expression in Eq. (4.35) for $\Omega_0(t)$ in the relativistic case. Since τ_0 must significantly exceed t_{now} by at least several orders of magnitude, as discussed in Sect. 4.3, we see that the exponential factor e^{-t/τ_0} is essentially 1. Likewise we recall that $M_0/T_{\text{MRE}} \geq \mathcal{O}(10^3)$, as also discussed in Sect. 4.3. Let us assume that this bound is saturated, so that $M_0/T_{\text{MRE}} = \mathcal{O}(10^3)$. We therefore find that Eq. (4.48) can be satisfied only if $g_c \sim 10^4$. This would in turn require a mass scale T_c which at the very minimum exceeds the TeV scale (thereby introducing a hierarchy between T_c and M_0 which is at least a factor of 10^6) and which actually must be so high that there are at least ten times as many effectively relativistic degrees of freedom below this scale than are known to exist below the TeV scale — a rather unlikely proposition resting entirely on currently unknown physics. Considering greater values of M_0/T_{MRE} only worsens this situation and requires even greater values of g_c . Therefore, although there might exist finely tuned slivers of parameter space in which one might manage to achieve a balancing between g_c and M_0/T_{MRE} sufficient to satisfy Eq. (4.48), we shall abandon any further consideration of the relativistic case in what follows.

This situation changes dramatically when we turn to the non-relativistic case in Eq. (4.32). In this case, we continue to find that $e^{-t_{\text{now}}/\tau_0} \approx 1$. However, the presence of the factor $(r/s)^{3/2}e^{-r/s}$ allows us greater freedom in satisfying the constraint in Eq. (4.48). Indeed, the first thing we learn is that our system is going to be very sensitive to the ratio r/s — not surprising, given that this was already the ratio that determined the extent to which our lightest mode was relativistic or non-relativistic. However, we now see that r/s is also going to play a large role in governing the allowed values of the overall mass scales in our system, with greater (lesser) values of r/s generally corresponding to higher (lower) absolute mass scales for our ensemble.

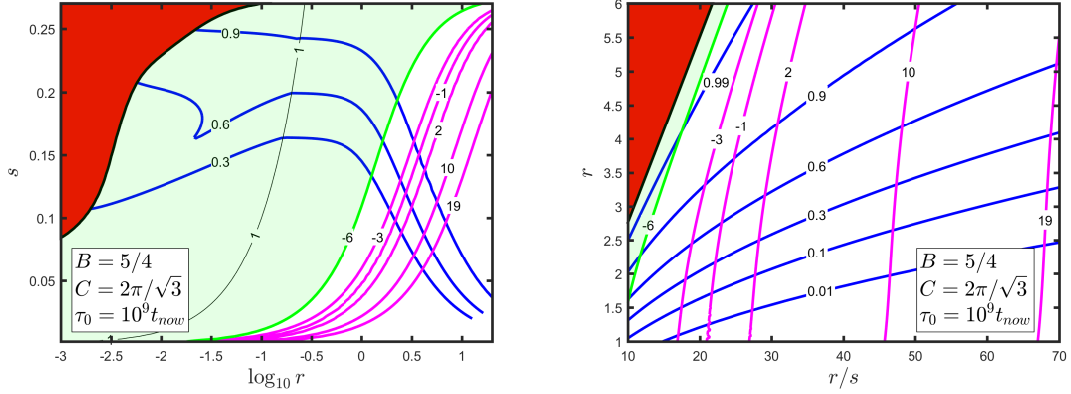


FIGURE 4.4: A survey of physics in the (r, s) plane, with B , C , and τ_0 set to the “benchmark” values shown. *Left panel:* The thin black line labeled ‘1’ indicates the contour with $r/s = 1$; this is thus the dividing line between the region in which the lightest state is relativistic (left of this line) versus non-relativistic (right of this line). The blue curves indicate contours of $\eta(t_{\text{now}})$, while the magenta lines indicate contours of M_s and are labelled by the value of $\log_{10}(M_s/\text{GeV})$. The red region is excluded by look-back and w_{eff} constraints, while the pale green region is excluded by the constraint $M_0 \gtrsim \mathcal{O}(\text{keV})$ which is saturated along the single green contour. Increasing (decreasing) the value of τ_0 does not affect the M_s or $\eta(t_{\text{now}})$ contours, and simply shifts the red exclusion region to the left (right). *Right panel:* Same as left panel, but with features plotted relative to the variables r and r/s . The entire region shown in this panel corresponds to the non-relativistic case.

We shall therefore proceed through our parameter space as outlined above, paying special attention to the values of r and s and in particular to the ratio r/s . Specifically, for each value of $\{B, C, r, s, \tau_0\}$, we shall determine whether our internal consistency constraints $B \geq 1$, $C^2 \geq 2\pi^2(4B - 3)/3$, and $s \leq 1/C$ are satisfied and whether the phenomenological consistency constraints in Eqs. (4.45) and (4.46) are satisfied. If so, we shall then determine the corresponding values of M_s and $\eta(t_{\text{now}})$, with the overall goal of understanding which regions of parameter space potentially lead to viable ensembles and which subregions correspond to ensembles which are particularly DDM-like.

Because of the somewhat natural and intuitive role played by the $D_\perp = 2$ scalar flux tube, as discussed in Sect. 4.1, we shall adopt the values

$$B = 5/4, \quad C = 2\pi/\sqrt{3} \approx 3.63 \quad (4.49)$$

as “benchmark” values and begin our exploration within (r, s) space. Taking $\tau_0 =$

$10^9 t_{\text{now}}$, we find the results shown in Fig. 4.4.

Let us first concentrate on the left panel of Fig. 4.4. The red region indicates those values of (r, s) which are excluded by look-back and w_{eff} constraints, while the pale green region is excluded by the requirement that $M_0 \gtrsim \mathcal{O}(\text{keV})$. The blue curves indicate contours of $\eta(t_{\text{now}})$ and the magenta curves indicate contours of M_s , labelled by values of $\log_{10}(M_s/\text{GeV})$. The single green curve indicates the contour with $M_0 = 1 \text{ keV}$. The thin black curve indicates the contour with $r/s = 1$, and thus serves as the nominal dividing line between the regions in which the lowest ensemble state is relativistic (above and to the left) or non-relativistic (below and to the right).

Several things are immediately apparent from this figure. First, we see that the portion of the parameter space corresponding to the relativistic case is excluded by our constraint on M_0 . This is entirely in keeping with our conclusions already reached above. Nevertheless, we also see that beyond this region there exists an entire area of parameter space in which all of our constraints are satisfied. Moreover, within this region we see that M_s varies from the keV/MeV-range all the way to the Planck scale. Likewise, $\eta(t_{\text{now}})$ varies through all of its possible values. This is therefore not only an allowed region, but one which is likely to be exceedingly rich in phenomenology. Indeed, given the contours plotted in this figure, we see that the “sweet spot” within the (r, s) parameter space lies roughly within the range

$$\begin{cases} 1 \lesssim r \lesssim 6 \\ 0.05 \lesssim s \lesssim 0.18 . \end{cases} \quad (4.50)$$

This is the region of (r, s) parameter space where the plotted blue and magenta contours intersect each other and form a “cross-hatched” region, as illustrated in the left panel of Fig. 4.4. This sweet spot is therefore the region that will be of maximum interest to us. Indeed, within this region, we observe from the left panel of Fig. 4.4 that $\eta(t_{\text{now}})$ increases if either r or s is increased, while M_s increases in the former case but decreases in the latter.

The right panel of Fig. 4.4 focuses on this sweet-spot region and shows the same

M_s and η contours, only now plotted with respect to the variables r/s and s using a linear rather than logarithmic axis. The fact that the M_s contours are approximately vertical in this region indicates that M_s is dominantly determined by the ratio r/s , exactly as anticipated above, with increasing values of r/s corresponding to increasing values of M_s . Indeed, we see from the right panel of Fig. 4.4 that M_s increases extremely rapidly as a function of r/s , in keeping with the exponential dependence in Eq. (4.32). Likewise, increasing the value of r/s while holding r fixed tends to *decrease* the value of $\eta(t_{\text{now}})$. Thus, for fixed r , we find that M_s and $\eta(t_{\text{now}})$ tend to vary inversely with respect to each other as functions of r/s , with our ensembles becoming less DDM-like at higher mass scales and more DDM-like at lower mass scales. Likewise, for fixed r/s , we find that increasing r tends to increase $\eta(t_{\text{now}})$, as already evident from the left panel of Fig. 4.4.

It is easy to understand these results physically. For fixed r , increasing r/s corresponds to decreasing s . This lowers the critical temperature T_c at which our initial cosmological abundances are established, which has the effect of decreasing the abundances of the heavier states relative to the lighter states. This therefore decreases the value of $\eta(t_{\text{now}})$. By contrast, holding r/s fixed and increasing r corresponds to increasing s as well. The increase in r renders all of the ensemble states more massive but provides a smaller proportional mass increase for the heavier states than for the lighter states. Thus the mass *ratios* between heavier and lighter states decreases, which tends to increase the value of $\eta(t_{\text{now}})$. Likewise, as discussed above, increasing s also tends to increase the value of $\eta(t_{\text{now}})$. These two effects then tend to reinforce each other, as evident in Fig. 4.4.

Having identified our sweet-spot region in (r, s) parameter space, we now investigate how these values of M_s and $\eta(t_{\text{now}})$ vary as our other parameters B , C , and τ_0 are varied. To do this, we study variations in these parameters relative to an (r, s) “benchmark”

$$r = 3.5 \ , \quad r/s = 30 \ , \quad (4.51)$$

which we henceforth take as representative of our sweet-spot region in the (r, s) plane.

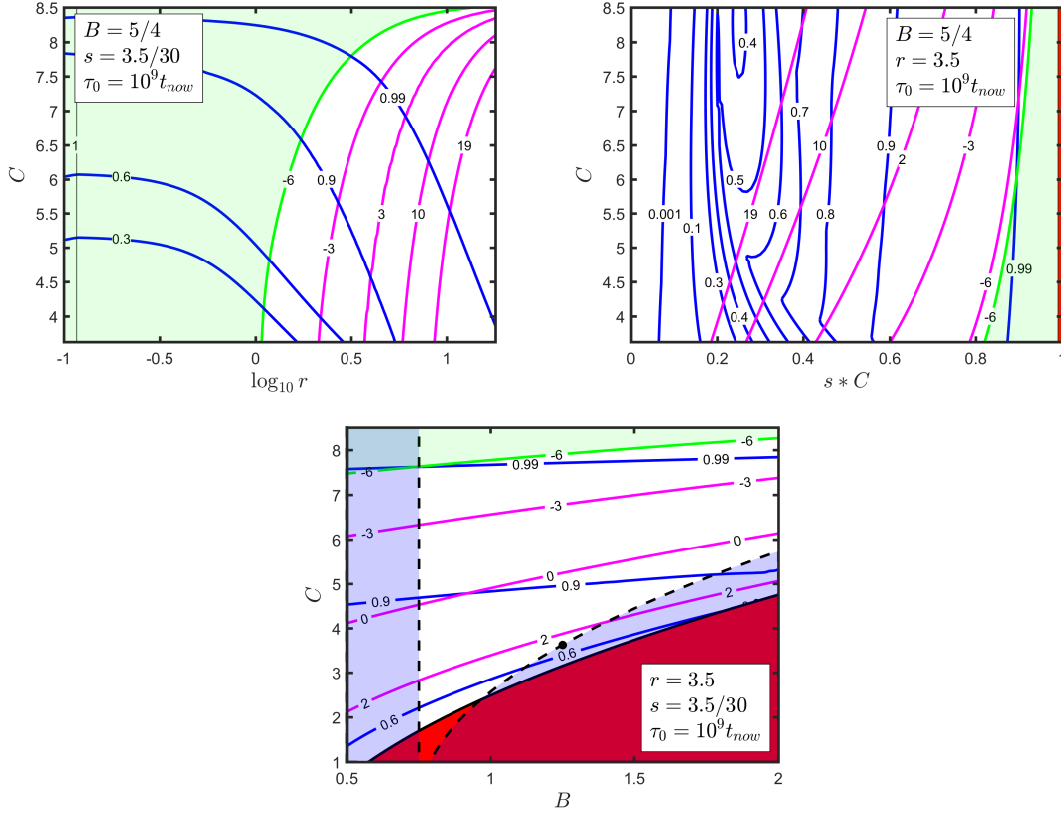


FIGURE 4.5: Contours of $\eta(t_{\text{now}})$ (blue curves) and M_s (magenta curves), labelled as in Fig. 4.4 and plotted in three different planar “slices” through the (B, C, r, s) parameter space. The top two panels show these contours plotted in the (r, C) and (s, C) planes, respectively, while the bottom panel shows these contours plotted in the (B, C) plane. In all panels, colored shaded regions are excluded by either string consistency constraints (blue shaded regions), internal consistency constraints (red region in lower panel), or phenomenological look-back, w_{eff} , or $M_0 \gtrsim \mathcal{O}(\text{keV})$ constraints (pale green regions as well as the red region along the right edge of the upper right panel). As in Fig. 4.4, the thin black vertical $r/s = 1$ contour (visible at the extreme left of the upper left panel) continues to represent the boundary between the regions in which the lightest state is either relativistic (left of the line) or non-relativistic (right of the line).

In Fig. 4.5 we illustrate the effects of variations in B and C relative to this benchmark, plotting contours of M_s and $\eta(t_{\text{now}})$ in the (r, C) plane (upper left panel), the (s, C) plane (upper right panel), and the (B, C) plane (lower panel). Note that since we must always have $s \leq 1/C$, it is actually the normalized product $s \cdot C$ which captures the dependence on s in situations where C might also be varied. In the upper right panel we therefore plot our contours relative to $s \cdot C$ rather than s alone. Likewise, in the lower panel of Fig. 4.5 we have continued to indicate our allowed regions of B

and C as in Fig. 4.2, where the dot continues to represent the $D_\perp = 2$ scalar-string benchmark values in Eq. (4.49).

Together, the three panels of Fig. 4.5 tell a consistent story. First, with r and s held fixed, we see from the upper left and lower panels of Fig. 4.5 that increasing C generally tends to increase $\eta(t_{\text{now}})$. This result makes sense: increasing C corresponds to increasing the *degeneracies* of the heavier states relative to the lighter states. However, with s held constant, each of these heavier states continues to accrue the same abundance as before. Thus increasing C increases the total abundance carried by the heavier states relative to that carried by the lighter states, thereby increasing $\eta(t_{\text{now}})$. Second, we see from the lower panel of Fig. 4.5 that while our values of M_s and $\eta(t_{\text{now}})$ are quite sensitive to C , they are far less sensitive to B . This too makes sense, since C governs the exponential rate of growth in the state degeneracies while B governs only the subleading polynomial behavior. Third, in each of the above two cases, we also note that increasing C while holding r or B fixed also corresponds to decreasing M_s . Thus, once again, we see that M_s and $\eta(t_{\text{now}})$ tend to vary inversely with each other, giving rise to more DDM-like ensembles at lower energy scales and more traditional ensembles at higher energy scales.

Finally, we see from the upper right panel of Fig. 4.5 that our values of $\eta(t_{\text{now}})$ are largely *insensitive* to variations in C as long as $s \cdot C$ is held fixed. However, this too is easy to understand. Increasing C while holding $s \cdot C$ fixed corresponds to decreasing s as we increase C . Increasing C induces an exponential increase in the degeneracy of each massive state, while decreasing s decreases the critical temperature T_c , thereby inducing a corresponding exponential decrease in the abundance associated with each such state. Thus, to first approximation, these two effects tend to mitigate each other: they produce more states, but also cause each state to carry a correspondingly smaller abundance.

Thus far we have not discussed the effects of varying our remaining free parameter τ_0 . Varying τ_0 does not affect the degeneracies of states or their cosmological abundances. Indeed, variations in τ_0 affect only the *lifetimes* of these states. In principle, this has the potential to affect the values of quantities such as $\eta(t_{\text{now}})$ since the deter-

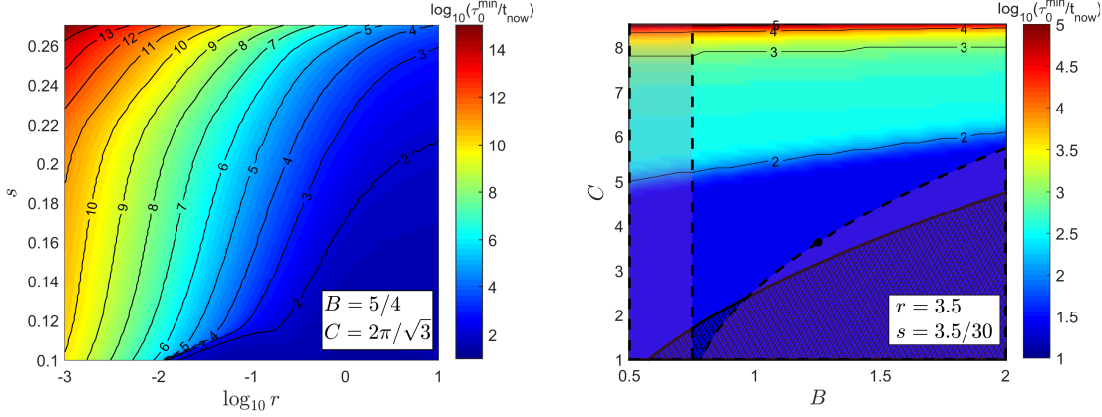


FIGURE 4.6: Contours of the minimum value of τ_0^{\min} consistent with the look-back and w_{eff} constraints discussed in the text, plotted in the (r, s) plane (left panel) and in the (B, C) plane (right panel).

mination of $\eta(t_{\text{now}})$ requires totalling the abundances of only those states which have not yet decayed at the present time. However, under the assumption that $\tau_0 \gg t_{\text{now}}$ (or under the equivalent assumption that our scenario already satisfies the look-back and w_{eff} constraints), we know that $\Omega_{\text{tot}}(t_{\text{now}})$ is not changing rapidly at the present time. In other words, the total abundances of those states which are decaying at the present time is relatively small. In such cases, the M_s and $\eta(t_{\text{now}})$ contours are therefore largely insensitive to τ_0 . Indeed, in Fig. 4.4, the sole effect of varying τ_0 is therefore merely to “slide” the red exclusion regions in Fig. 4.4 horizontally relative to the rest of the plot: these exclusion regions move to the right (and therefore become more threatening to our sweet-spot region) if τ_0/t_{now} is decreased, and move to the left (and therefore become even less of a concern) if τ_0/t_{now} is increased.

While this is entirely as expected, the natural question then arises: for any values of $\{B, C, r, s\}$, what is the minimum value of τ_0 that can be tolerated before violating our look-back and w_{eff} constraints? Contours indicating the resulting minimum values τ_0^{\min} are plotted in Fig. 4.6 in both the (r, s) and (B, C) planes, taking our “benchmark” values in Eqs. (4.51) and (4.49) respectively. In general, we see from Fig. 4.6 that a wide variety of values of τ_0^{\min} are possible, depending on our specific location in parameter space, with larger values of τ_0^{\min} corresponding to very small values of r or relatively large values of s or C . However, for our sweet-spot bench-

mark values in Eqs. (4.49) and (4.51), we see from Fig. 4.6 that τ_0^{\min} can be as small as approximately $10^2 t_{\text{now}}$.

This, too, is not entirely a surprise. After all, a bound on the lifetime of the longest-lived DDM constituent on the order $\tau_0/t_{\text{now}} \sim \mathcal{O}(100)$ is roughly on the same order as the most conservative bounds on the lifetime τ_χ of a traditional single-component dark-matter candidate which decays into other purely dark-sector states. Indeed, model-independent bounds on decaying dark matter in traditional single-component models in which the dark-matter particle carries essentially all of the observed dark-matter abundance and decays into dark radiation have been derived by a number of groups (see, *e.g.*, Refs. [81–84]). Depending on the assumptions inherent in the various analyses and on the breadth of cosmological data incorporated, such studies place a bound on the lifetime of such a dark-matter candidate on the order of $\tau_\chi/t_{\text{now}} \gtrsim \mathcal{O}(10 - 100)$. Thus, a bound on τ_0 in this range is *a priori* reasonable — especially since our analysis in Fig. 4.6 determines the value of τ_0^{\min} based only on cosmological look-back and w_{eff} constraints. Of course, if the ensemble constituents decay into *visible*-sector particles with a non-negligible branching fraction, the constraints on τ_0 are expected to increase significantly. Indeed, the most stringent bounds on a single dark-matter particle χ which decays primarily into visible-sector radiation require that this particle be hyperstable, with $\tau_\chi \sim 10^9 t_{\text{now}}$.

Despite the possibilities for lowering τ_0 afforded by the results in Fig. 4.6, we shall continue to retain our benchmark value $\tau_0 = 10^9 t_{\text{now}}$. We do this in order to be consistent with the most conservative decay scenarios possible. Although this value for τ_0 is quite large, we emphasize that this is only the lifetime of the lightest ensemble constituent, and that a significant fraction of the ensemble constituents will generally have lifetimes much less than τ_0 . Moreover, even in cases for which the majority of the ensemble is long-lived, DDM ensembles can nevertheless yield striking astrophysical signatures [47, 49, 50] which differ from those of traditional dark-matter candidates. Thus, even with such values of τ_0 , the phenomenology of the resulting ensemble can differ significantly from that of traditional dark-matter candidates.

Having explored the relevant $\{B, C, r, s, \tau_0\}$ parameter space of our ensemble and

identified our sweet-spot region, we now examine the characteristics of the corresponding ensembles in more detail. In particular, we seek to understand what these ensembles look like, and how their overall structure evolves with time. As discussed in Sect. 4.3.1, the most relevant aggregate properties of any dark-sector ensemble are its total cosmological abundance $\Omega_{\text{tot}}(t)$, its effective equation-of-state parameter $w_{\text{eff}}(t)$, and its tower fraction $\eta(t)$, each of which is generally time-dependent. We therefore begin by examining how each of these quantities evolves with time for ensembles in and near our sweet spot.

This information is shown in Fig. 4.7. In this figure, we consider a “benchmark” ensemble with $B = 5/4$, $C = 2\pi/\sqrt{3}$, $r = 3.5$, $s = 3.5/30$, and $\tau_0 = 10^9 t_{\text{now}}$, as well as nearby ensembles in which τ_0 is varied (top row), r is varied (second row), s is varied (third row), C is varied (fourth row), and B is varied (fifth row). In each case, we plot the corresponding total cosmological abundance Ω_{tot} (left column), equation-of-state parameter w_{eff} (middle column), and tower fraction η (right column) as functions of time. Note that in each case the overall abundance is normalized through an appropriate choice of M_s such that $\Omega(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$, as required.

In each panel of Fig. 4.7 (except for those along the bottom row), the blue curve corresponds to our “benchmark” point. We therefore begin by focussing on these benchmark curves. The curve for $\Omega_{\text{tot}}(t)$ appears nearly constant at $\Omega_{\text{CDM}} \approx 0.26$ for all of the cosmological history plotted (which we assume to have been matter-dominated), including the present time t_{now} . Indeed, this behavior continues all the way into the future until $t \approx 10^9 t_{\text{now}}$, at which point $\Omega_{\text{tot}}(t)$ begins to decline gently to $\Omega_{\text{tot}} = 0$. This behavior is matched by $w_{\text{eff}}(t)$, which remains near zero for most its cosmological evolution before gently rising to $w_{\text{eff}} > 0$ at $t \approx 10^9 t_{\text{now}}$. This makes sense, since Eq. (4.42) tells us that $w_{\text{eff}}(t)$ is proportional to the time-derivative of $\Omega_{\text{tot}}(t)$. Finally, we see that $\eta(t)$ remains more or less fixed at approximately $\eta \approx 0.72$ during most of its cosmological history before smoothly dropping to $\eta = 0$.

This behavior is easy to understand. If this has been a traditional ensemble with a single dark-matter component whose decay we could model as essentially instantaneous (just as we are assuming for the individual components of our dark-matter

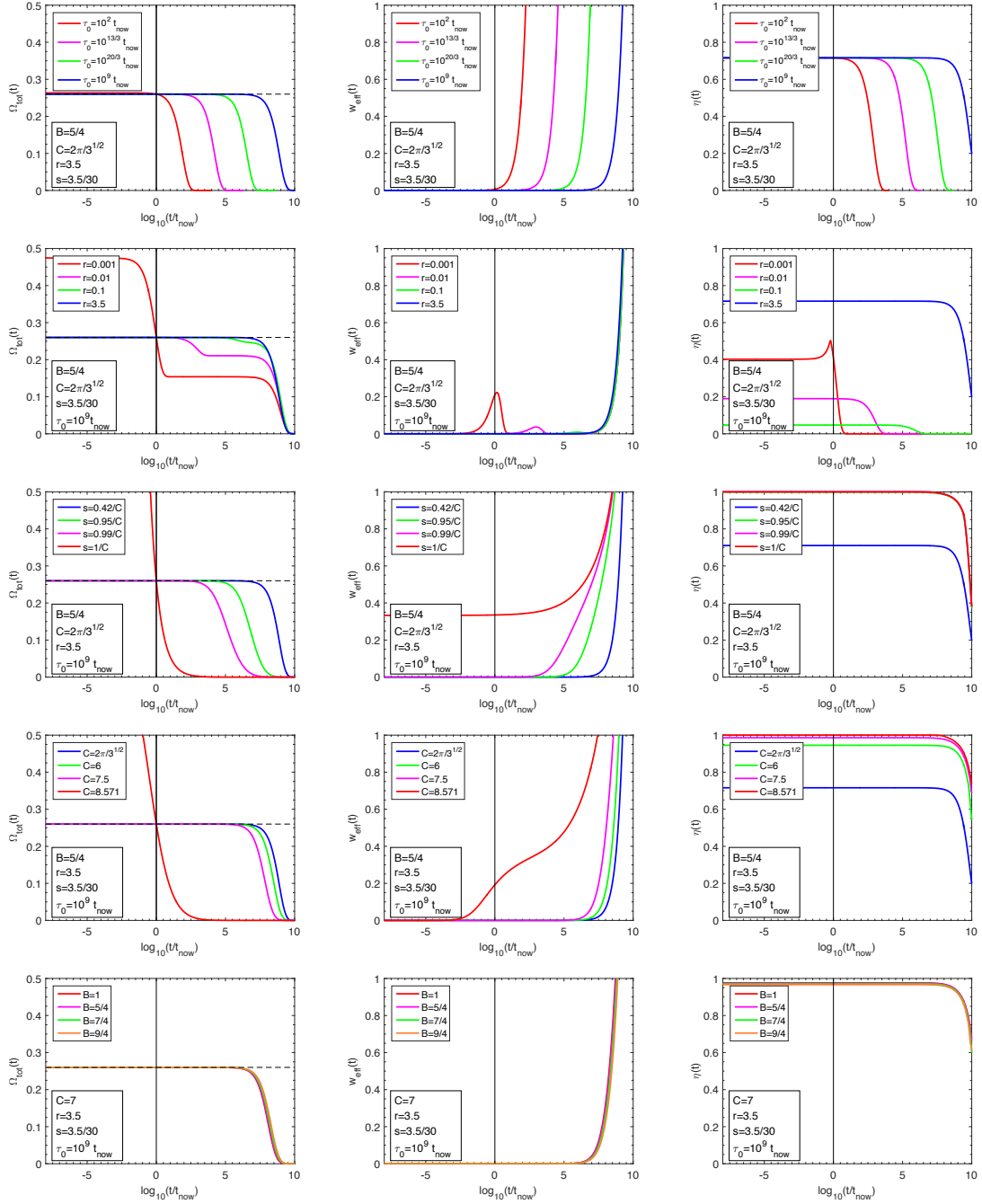


FIGURE 4.7: Total cosmological abundances Ω_{tot} (left column), equation-of-state parameters w_{eff} (middle column), and tower fractions η (right column) for our DDM ensembles, plotted as functions of time when all input variables are held fixed at their “benchmark” values except for τ_0 (top row), r (second row), s (third row), C (fourth row), and B (bottom row). In all panels the blue curve corresponds to our “benchmark” point with $B = 5/4$, $C = 2\pi/\sqrt{3}$, $r = 3.5$, $s = 3.5/30$, and $\tau_0 = 10^9 t_{\text{now}}$, while the curves of other colors indicate departures away from this point. For reasons discussed in the text, the bottom row illustrates variations in B along a line that does not include the benchmark point. Note that, as expected, some variations away from the benchmark point violate our look-back, w_{eff} , or M_0 constraints. However, our internal self-consistency constraints are always satisfied, with $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$ in all cases.

ensembles), our curve for $\Omega_{\text{tot}}(t)$ would have been fixed precisely at its present value $\Omega_{\text{CDM}} \approx 0.26$ over the entire range shown until suddenly dropping (essentially discontinuously) to $\Omega_{\text{tot}} = 0$ when the single dark-matter particle decays at $t \approx 10^9 t_{\text{now}}$. Likewise, $w_{\text{eff}}(t)$ would have been strictly fixed at $w_{\text{eff}} = 0$ during the cosmological evolution, while $\eta(t)$ would have been fixed at zero all along. However, this is not a traditional dark-matter setup: this is a DDM ensemble in which the present-day cosmological abundance $\Omega_{\text{tot}}(t_{\text{now}}) \approx 0.26$ is spread across a relatively large number of individual components with different masses and different lifetimes. It is thus the continued, ordered, sequential decays of these different components which produce the softer, gentler drop in $\Omega_{\text{tot}}(t)$ as t approaches $t \approx 10^9 t_{\text{now}}$. In fact, $\Omega_{\text{tot}}(t)$ is actually falling slightly *throughout* the cosmological evolution shown; this behavior is not visible in Fig. 4.7 only because at early times prior to $t \approx 10^9 t_{\text{now}}$ the states which are decaying are extremely heavy and thus carry extremely small abundances. By contrast, at late times approaching $t \approx 10^9 t_{\text{now}}$, the states which are decaying are relatively low-lying and carry more significant abundances. This is also evident in our curve for $\eta(t)$: for most of the cosmological history, the value $\eta \approx 0.72$ tells us that only approximately 28% of the total dark-sector cosmological abundance is carried by the dominant (lightest) state in the ensemble, even at early times, while the remaining 72% of the abundance is carried by the more massive states — particularly those which, though more massive, are nevertheless relatively low-lying. As a result of the sequential decays of such states, $\eta(t)$ — like $\Omega_{\text{tot}}(t)$ — is also actually falling slightly *throughout* the cosmological evolution shown. It is only due to the decays of the relatively low-lying states near $t \approx 10^9 t_{\text{now}}$ that $\eta(t)$ ultimately falls gently but noticeably to zero.

At first glance, it may seem surprising that all three of our primary quantities Ω_{tot} , w_{eff} , and η are nearly constant at $t \approx t_{\text{now}}$. However, this is ultimately the direct consequence of our benchmark choice $\tau_0 = 10^9 t_{\text{now}}$: with this choice, those states within the ensemble which are decaying today are all extremely massive and thus carry very little abundance. The DDM nature of such an ensemble is nevertheless clear from its η -value, which is as high as 0.72 even at the present time. In this

connection, we again emphasize that taking $\tau_0 = 10^9 t_{\text{now}}$ was merely a conservative choice which is not by itself intrinsic to the DDM framework; indeed we learned from Fig. 4.6 that we could easily have chosen τ_0 as small as $\tau_0 \approx 10^2 t_{\text{now}}$ without running afoul of our look-back and w_{eff} constraints. Indeed, without further details concerning the precise nature of these ensembles (including, most critically, the ultimate decay products of their constituents), such small values for τ_0 would have been equally viable.

This observation is illustrated along the top row of Fig. 4.7, where we show the evolution of our blue “benchmark” curves as we vary τ_0 between our conservative value $\tau_0 \approx 10^9 t_{\text{now}}$ and the more extreme value $\tau_0 \approx 10^2 t_{\text{now}}$. In general, changing τ_0 does not affect the internal structure of the ensemble — it merely affects the lifetimes of the individual ensemble constituents, rescaling them all up or down together. Since it is these lifetimes which produce the non-trivial time-dependence for Ω_{tot} , w_{eff} , and η , we expect that changing τ_0 should preserve the general shapes of these curves and merely translate these curves along the time axis. This behavior is verified in the panels along the top row of Fig. 4.7. Indeed, we can even see from these panels why $\tau_0 \approx 10^2 t_{\text{now}}$ is the minimum value of τ_0 that may be chosen for our benchmark point: choosing τ_0 any smaller would shift our curves even further towards earlier times, whereupon $\Omega_{\text{tot}}(t)$ would begin to experience significant variations within the interval $10^{-6} t_{\text{now}} \lesssim t \lesssim t_{\text{now}}$ and $w_{\text{eff}}(t_{\text{now}})$ would begin to deviate significantly from zero. Such behavior would then violate our look-back and w_{eff} constraints, respectively.

Let us now turn to the behavior of our Ω_{tot} , w_{eff} , and η curves as we vary r , as shown in the panels along the second row of Fig. 4.7. Two observations underlie the behavior shown. First, we note that changing r changes the lifetimes of the states at each mass level according to Eq. (4.28), with $\tau_n/\tau_0 \rightarrow 0$ as $r \rightarrow 0$. This result is simple to understand: as $r \rightarrow 0$, the $n = 0$ states become hierarchically lighter than the $n \geq 1$ states and thus the $n > 0$ states have hierarchically shorter lifetimes. Second, we note that changing r also changes the relative abundances which are

generated at t_c according to

$$\frac{\Omega_n(t_c)}{\Omega_0(t_c)} = \frac{(n + r^2)^{5/4}}{r^{5/2}} \exp\left(-\frac{\sqrt{n + r^2} - r}{s}\right). \quad (4.52)$$

This quantity is non-monotonic as a function of r , first dropping as r is reduced from large values and ultimately hitting a minimum before increasing again and diverging as $r \rightarrow 0$. Indeed, for $n = 1$ and s set to its benchmark value $s = 3.5/30 \approx 0.117$, this minimum occurs at $r \approx 0.4$.

These two effects are responsible for the behaviors shown in the second row of Fig. 4.7. As r decreases from its benchmark value with τ_0 held fixed, the excited states with $n > 0$ start decaying earlier and earlier. Rescaling our overall abundances in order to keep $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}}$ produces the effects shown in the left panel. Indeed, we see from this panel that the case with $r = 0.001$ actually violates our look-back and w_{eff} constraints, as already evident from Fig. 4.4. Even the $\Omega_{\text{tot}}(t)$ curve with $r = 0.01$ is tightly constrained: shifting τ_0 towards any smaller values below $10^9 t_{\text{now}}$ (*i.e.*, shifting this curve further towards the left) also leads to violations of our look-back and w_{eff} constraints, as already anticipated in the left panel of Fig. 4.6. Likewise, as a result of the observations below Eq. (4.52), the relative sizes of the abundances Ω_n associated with the excited $n > 0$ states relative to the abundance Ω_0 associated with the $n = 0$ ground state vary non-monotonically with r , shrinking as r drops from 3.5 to approximately 0.4, and then growing again as r drops still further. This then explains the non-monotonic behavior for $\eta(t)$ as a function of r , as shown in the right panel.

By contrast, the effects of varying s and C are shown along the third and fourth rows of Fig. 4.7, respectively. While the quantity s governs the exponential rate at which the Boltzmann suppression of the abundances of the ensemble constituents *decreases* with n , the quantity C governs the exponential rate at which the degeneracy of states for the ensemble *grows* with n . As a result, the effects of decreasing s or increasing C are largely similar to each other as far as $\Omega_{\text{tot}}(t)$ is concerned, as evident in Fig. 4.7: both tend to increase the primordial aggregate abundances $\hat{\Omega}_n$

of the heavier states in the ensemble. This effect causes $\Omega_{\text{tot}}(t)$ to begin to decline earlier and earlier as these heavier states are the first to decay. By contrast, it is important to note that increasing C and decreasing s nevertheless have *opposite* effects on the value of $\eta(t_{\text{now}})$: the former increases $\eta(t_{\text{now}})$, as anticipated in Fig. 4.5, while the latter decreases $\eta(t_{\text{now}})$, as anticipated in Fig. 4.4. This difference occurs because increasing C merely increases the state degeneracies \hat{g}_n of the heavy states, thereby injecting more abundance into the heavy states relative to the light states, while decreasing s has the effect of increasing the abundances of *all* of our states, including the abundance of the dominant abundance-carrier at $n = 0$. This causes the total abundance of the ensemble to grow more rapidly than the abundances of the excited $n > 0$ states alone, thereby decreasing $\eta(t_{\text{now}})$.

One important feature to note from these plots is the appearance of a Hagedorn instability as $s \rightarrow 1/C$ (or equivalently as $C \rightarrow 1/s$). In these limiting cases, the total energy density Ω_{tot} injected into the system through our confining phase transition at $t = t_c$ diverges, violating the constraint in Eq. (4.18). Such cases therefore violate our look-back and w_{eff} constraints, as evident in Fig. 4.7. Indeed, the Hagedorn instability is a critical feature of theories with exponentially growing degeneracies of states [58].

Finally, we turn to the fifth and final row of Fig. 4.7. Note that in order to remain within the self-consistency bound in Eq. (4.13), it is not possible to increase B above our benchmark value $5/4$ when $C = 2\pi/\sqrt{3}$. For this reason, we have chosen to hold C fixed at a greater value, specifically $C = 7$, when exploring the effects of varying B . Unfortunately, we see that variations in B are barely distinguishable in these plots, even when B is varied all the way from $B = 1$ (corresponding to $D_{\perp} = 1$) to $B = 9/4$ (corresponding to $D_{\perp} = 6$). This tells us that the sorts of abundance-based or equation-of-state-based analyses we are doing here are relatively insensitive to the number of uncompactified transverse spacetime directions into which our dark-sector flux tube can vibrate, as long as C (related to the total central charge of the degrees of freedom on the flux-tube worldsheet) is held fixed. Of course, in a realistic setting, there are likely to be many other more specific probes of D_{\perp} , including probes that are based on specific properties of the dark-sector dynamics. Our result here merely

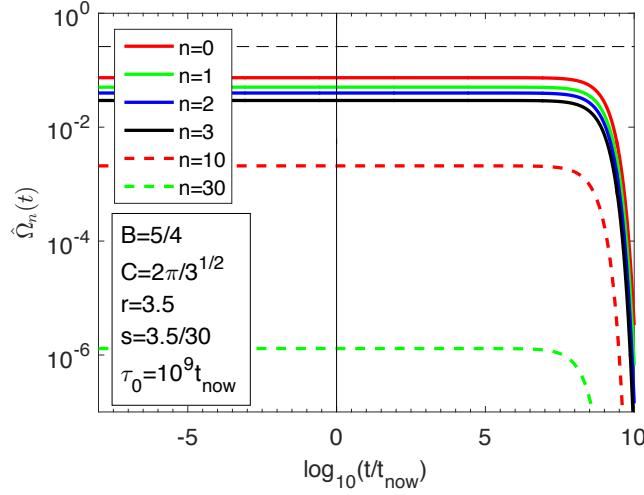


FIGURE 4.8: The level-by-level aggregate cosmological abundances $\hat{\Omega}_n \equiv g_n \Omega_n$ of our benchmark DDM model, plotted as functions of time for a series of low-lying mass levels n . We see that the lightest states decay later and carry the largest cosmological abundances, while the heavier states decay earlier and carry smaller cosmological abundances — a key feature of the DDM framework. As required, the sum of all abundance contributions at $t = t_{\text{now}}$ is $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$.

indicates that studies based on cosmological abundances alone are not likely to be the most useful in this regard.

We have seen in Fig. 4.7 how the total abundances Ω_{tot} of our DDM ensembles vary as a function of time. However, it is also interesting to understand how the *individual* aggregate abundances $\hat{\Omega}_n(t)$ at each mass level n contribute to this behavior. The result is shown in Fig. 4.8 for our benchmark DDM model. As we see from Fig. 4.8, there are many mass levels n whose states contribute to $\Omega_{\text{tot}}(t_{\text{now}})$: states with smaller values of n carry larger abundances and have longer lifetimes, persisting into later times before decaying, while those with larger values of n carry smaller abundances and have shorter lifetimes, decaying earlier. Indeed, this balancing between lifetimes and abundances is a fundamental hallmark of the DDM framework. Although the sum of these abundances at $t = t_{\text{now}}$ is fixed at $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$, we see that even states with relatively large values of n have lifetimes τ_n exceeding t_{now} and thus contribute non-trivially to $\Omega_{\text{tot}}(t_{\text{now}})$. Indeed, for our benchmark model, we find that there are no fewer than seven distinct mass levels contributing more than 0.01 to $\Omega_{\text{tot}}(t_{\text{now}})$ and no fewer than ten distinct mass levels contributing more than 1%

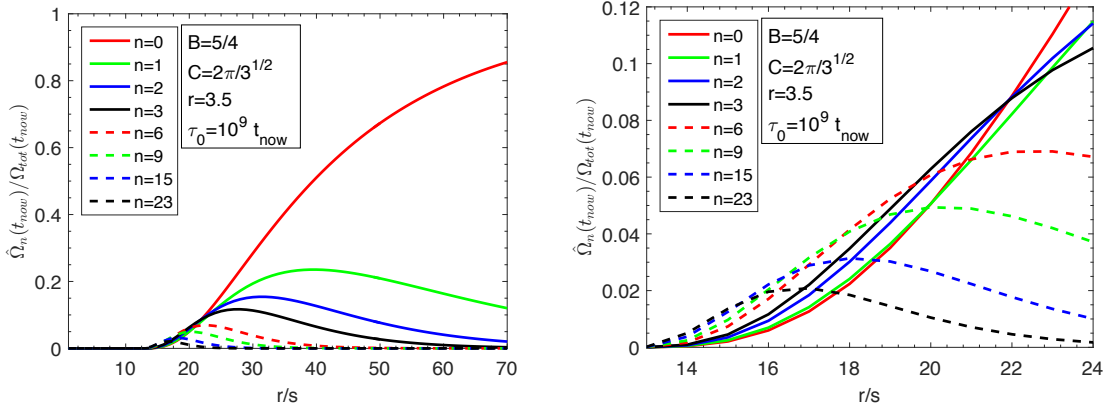


FIGURE 4.9: *Left panel:* Present-time aggregate abundance fractions $\hat{\Omega}_n(t_{\text{now}})/\Omega_{\text{tot}}(t_{\text{now}})$, plotted as functions of r/s . As r/s increases, the $n = 0$ state carries an increasingly large fraction of the total abundance, resulting in scenarios which have smaller values of η and which are therefore less DDM-like. By contrast, for smaller r/s , we see that the lightest state carries a smaller proportional fraction of the total abundance and in fact may not even be the dominant state for sufficiently small r/s . *Right panel:* A zoom-in of the left panel, illustrating how the level n of the states carrying the largest collective abundance $\Omega_n(t_{\text{now}})$ shifts as a function of r/s . For example, for $r/s = 15$, all states carry relatively small abundances and it is actually the $n = 23$ states which collectively carry the largest collective abundance at the present time. Such scenarios are therefore extremely DDM-like.

of $\Omega_{\text{tot}}(t_{\text{now}})$.

It is also interesting to examine how these results vary as a function of the ratio r/s which, as we have seen, governs the overall mass scales associated with these DDM ensembles. The results are shown in Fig. 4.9, where we plot the aggregate fractions $\hat{\Omega}_n(t_{\text{now}})/\Omega_{\text{tot}}(t_{\text{now}})$ for a variety of different mass levels n as a function of r/s . As evident in Fig. 4.9, the lightest state carries a larger and larger fraction of the total abundance as r/s increases, resulting in scenarios which have smaller values of η and which are therefore less DDM-like. By contrast, the lightest state carries a smaller proportional fraction of the total abundance as r/s decreases, and in fact may not even be the dominant state for sufficiently small r/s . Indeed, for $r/s = 15$, we find that all states carry relatively small abundances, and it is actually the states at the $n = 23$ mass level which collectively carry the largest individual abundance at the present time. Such scenarios are therefore extremely DDM-like.

Putting all the pieces together, we can summarize our results as in Figs. 4.10 and 4.11. Fig. 4.10 consists of a sequence of dark-matter pie charts showing the

relative contributions to $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$ from the lowest-lying states for $r = 3.5$ (top row) and $r = 4$ (bottom row), with $r/s = \{25, 30, 50, 65\}$ across each row. Within each pie, we illustrate the corresponding collective abundances $\hat{\Omega}_n(t_{\text{now}})$ as separate slices, one for each value of n , while the numbers listed within each slice indicate the number of individual states \hat{g}_n contributing at that mass level. For each pie chart we have also shown the corresponding values of M_0 , T_c , and M_s . For these calculations we have used the input values $T_{\text{MRE}} = 0.7756$ eV, $g_{\text{MRE}} = 3.36$, and $g_c = \{10.75, 61.75, 106.75, 106.75\}$, respectively, for $r/s = \{25, 30, 50, 65\}$. We have also assumed our standard benchmark values $B = 5/4$, $C = 2\pi/\sqrt{3}$, and $\tau_0 = 10^9 t_{\text{now}}$.

Let us begin by focusing on the “benchmark” pie chart within Fig. 4.10 corresponding to $r = 3.5$ and $r/s = 30$. For this pie chart, we see that the largest pie slice corresponds to the abundance contribution from the $n = 0$ mass level, while the successively smaller pie slices progressing in a clockwise fashion within the pie chart correspond to the abundance contributions from successively higher mass levels. For this pie chart, we find that $M_0 \approx 532$ GeV, $T_c \approx 18$ GeV, and $M_s \approx 152$ GeV. Note that this value for M_s is in agreement with the M_s contours shown in Fig. 4.4. We also see geometrically from this pie chart that $\eta \approx 0.72$, in agreement with the results shown in Figs. 4.4, 4.5, and 4.7.

Given this, we can now investigate how this benchmark pie chart deforms as a function of r/s and r . Results are illustrated in the other pie charts shown in Fig. 4.10. We see in general that increasing r from 3.5 to 4.0 (*i.e.*, passing from the top row of pie charts in Fig. 4.10 to the bottom row) has the net effect of shifting cosmological abundance away from the ground state, thereby increasing η and generally making each pie slice smaller while simultaneously lowering the corresponding mass scales. This is in complete accord with the results shown in Fig. 4.4. Likewise, decreasing or increasing r/s (*i.e.*, moving left or right along either row) has the effect of increasing or decreasing η while decreasing or increasing our corresponding mass scales. Indeed, we see that the variable r/s allows us to interpolate between two extremes: traditional ensembles with high mass scales at large r/s versus DDM-like ensembles with smaller mass scales at small r/s . We further observe that for sufficiently small r/s , the largest

pie slice is no longer the $n = 0$ slice (labelled ‘1’ in each pie chart) — as r/s decreases, this honor gradually shifts towards the pie slices corresponding to higher mass levels. This is in accordance with the results in Fig. 4.9.

Fig. 4.11 is similar to the top row of Fig. 4.10, except that we have now increased our values of C and B to $\sqrt{2}\pi$ and $B = 3/2$, respectively. These new values maintain $c_{\text{int}} = 0$ and correspond to the $D_{\perp} = 3$ scalar string. These changes in C and B increase the degeneracies \hat{g}_n of states at each mass level, with the new values indicated within the corresponding pie slices. Although the cosmological abundances per state are not affected by the changes in C and B , these increased degeneracies result in ensembles which are even more DDM-like and which have correspondingly smaller mass scales than those along the top row of Fig. 4.10. These results are consistent with those shown in Fig. 4.5.

We see, then, that a tremendous variety of DDM ensembles exist which have the two fundamental features outlined in the Introduction — Regge trajectories and exponentially rising degeneracies of states. These ensembles are consistent with our look-back and w_{eff} constraints, and thus satisfy the zeroth-order constraints that may be imposed on such ensembles on the basis of their total energy densities and equations of state alone. We also observe an important feature, a inverse correlation between the tower fraction η (which governs the extent to which our ensemble is truly DDM-like) and the magnitude of its underlying mass scales. Indeed, we have seen that while traditional ensembles typically have high corresponding mass scales, our ensembles become increasingly DDM-like for lower mass scales — all while remaining consistent with our look-back and w_{eff} constraints. These observations will likely be an important guide and ingredient in any future attempts to build realistic dark-matter models of this type.

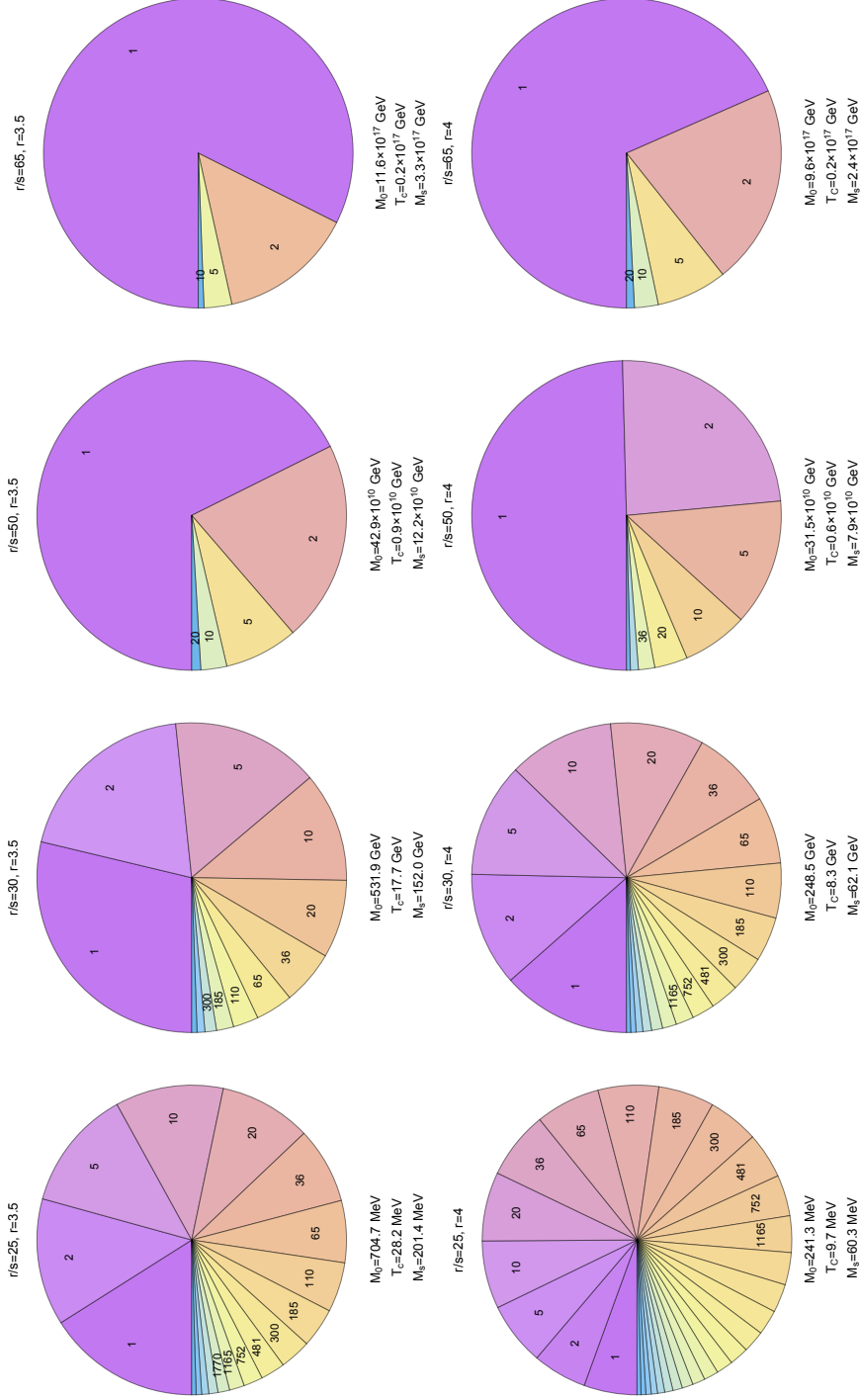


FIGURE 4.10: Pie for everyone! Dark-matter pie charts showing the relative contributions to $\Omega_{\text{tot}}(t_{\text{now}}) = \Omega_{\text{CDM}} \approx 0.26$ from the lowest-lying states for $r = 3.5$ (top row) and $r = 4$ (bottom row), with $r/s = \{25, 30, 50, 65\}$ across each row and with $B = 5/4$ and $C = 2\pi/\sqrt{3}$ held fixed. In the majority of cases (but not all cases), the largest pie slice corresponds to the abundance contribution from the $n = 0$ mass level, and the successively smaller pie slices (progressing in a clockwise fashion within the pie chart) correspond to the aggregate abundance contributions from successively higher mass levels. Within each pie slice we have also indicated the degeneracy \hat{g}_n of individual states whose cosmological abundances comprise the contribution from that slice. Note that in each case, the pie shown only corresponds to the dark-matter slice $\Omega_{\text{CDM}} \approx 0.26$ of the bigger “cosmic pie” which also includes contributions from dark energy and visible matter.

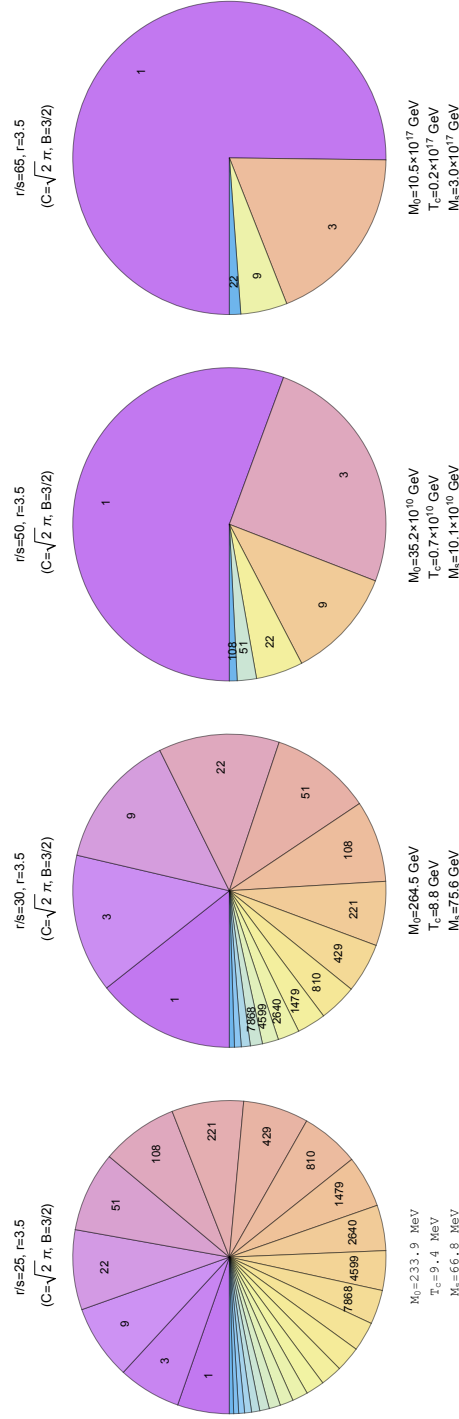


FIGURE 4.11: Seconds on pie! Same as the top row of Fig. 4.10 except that we have now taken $C = \sqrt{2}\pi$ and $B = 3/2$, corresponding to the $D_\perp = 3$ scalar string. These changes in C and B change the degeneracies \hat{g}_n of states at each mass level, as indicated within the corresponding pie slices, and result in ensembles which are even more DDM-like and which have correspondingly smaller mass scales than those along the top row of Fig. 4.10.

4.5 Conclusions

In this chapter, we have investigated the properties of a hitherto-unexplored class of DDM ensembles whose constituents are the composite states which emerge in the confining phase of a strongly-coupled dark sector. In ensembles of this sort, the masses of the constituent particles lie along well-defined Regge trajectories and the density of states within the ensemble grows exponentially as a function of the constituent-particle mass. This exponential growth is ultimately compensated by a Boltzmann suppression factor in the primordial abundances of the individual constituents, resulting in a finite total energy density $\Omega_{\text{tot}}(t)$. We also showed that such ensembles can naturally exhibit a balancing between lifetimes and cosmological abundances of the sort required by the DDM framework.

For each such ensemble, we calculated the corresponding effective equation-of-state parameter $w_{\text{eff}}(t)$ as well as the tower fraction $\eta(t)$. We also imposed a number of zeroth-order model-independent phenomenological constraints which follow directly from knowledge of $\Omega_{\text{tot}}(t)$, $w_{\text{eff}}(t)$, and $\eta(t)$. In general, we found that the imposition of such constraints tends to introduce *correlations* between the different underlying variables which parametrize our DDM ensembles, so that an increase in one variable (such as, *e.g.*, the exponential rate of growth in the state degeneracies) requires a corresponding shift in another variable (in this case, an increase in the lifetime of the lightest state in the ensemble, as indicated in the right panel of Fig. 4.6). Perhaps one of our most important results is the existence of an inverse correlation between the tower fraction $\eta(t)$ associated with a given DDM ensemble and its corresponding fundamental mass scales, so that the present-day cosmological abundance of the dark sector must be distributed across an increasing number of different states in the ensemble as these fundamental mass scales are dialed from the Planck scale down to the GeV scale.

We are certainly not the first to consider dark-matter scenarios in which the dark matter is composite. Indeed, within the context of traditional dark-matter models, it has been appreciated for some time that the dark-matter particle could be a com-

posite state. For example, the lightest technibaryon in technicolor theories was long ago identified as a promising dark-matter candidate [85, 86], and mechanisms [87] were advanced by which this particle could be rendered sufficiently light so as to be phenomenologically viable. Indeed, several explicit models [88] have been developed along these lines. Other more exotic baryon-like composites have also been advanced as potential dark-matter candidates [89]. Lattice studies of baryon-like states in the confining phases of both $SU(3)$ and $SU(4)$ gauge theories have also been performed [90–92].

A variety of scenarios in which a long-lived meson-like state which appears in the confining phase of a strongly-coupled hidden sector have been developed as well (for a review see, *e.g.*, Ref. [93]). These include scenarios in which the dark-matter particle is a pseudo-Nambu-Goldstone boson (PNGB) stabilized by a dark-sector analogue of flavor symmetry [94–98] or G -parity [99], or alternatively by some other symmetry of the theory with no SM analogue [100–103]. Complementary lattice studies of strongly-coupled dark-sector scenarios in which the dark-matter candidate is a PNGB have been performed as well [104, 105]. Scenarios in which the dark-matter candidate is not a PNGB, but rather a bound state of one heavy quark and one light quark, have also received recent attention [106–108], primarily due to the non-standard direct-detection phenomenology to which they give rise, as have scenarios in which the dark-matter candidate is a bound state of heavy quarks alone [109]. More general studies of composite hidden-sector theories which give rise to meson-like or baryon-like dark-matter candidates within different regions of parameter space have also been performed [110, 111].

Composite hidden-sector states consisting of non-Abelian gauge fields alone (so-called “glueball” states) have also long been recognized as promising dark-matter candidates [112, 113] — a possibility which has received renewed attention [114, 115] as well. Indeed, hidden sectors involving cosmologically stable dark glueball states arise naturally in a variety string constructions [116, 117], as well as in certain anomaly-mediated supersymmetry-breaking scenarios [118].

In addition, the possibility that composite states in the dark sector could them-

selves form bound states (so-called “dark nuclei”) has also been studied [119, 120], as has the possibility that these nuclei themselves could combine to form dark “atoms” or even dark “molecules” [121, 122]. Indeed, lattice studies [119, 123] corroborate the existence of stable dark nuclei states even within simple, two-flavor models with $SU(2)$ as the confining gauge group. In such models, a dark-sector equivalent of BBN serves as the mechanism for abundance generation. Such models can have interesting phenomenological consequences, especially in the regime in which a significant fraction of the dark-matter abundance is contributed by nuclei with large nucleon numbers [124, 125].

Composite dark-matter models are interesting from a phenomenological perspective as well. For example, the states of a strongly-coupled hidden sector provide a natural context [126] for strongly-interacting massive particle (SIMP) dark matter [127, 128] models, in which $3 \rightarrow 2$ processes rather than $2 \rightarrow 2$ processes play a dominant role in determining the dark-matter abundance. Indeed, a number of explicit models along these lines have been constructed [129–132]. One of the most interesting ramifications of SIMP models is that they naturally give rise to dark-matter self-interactions with cross-sections sufficiently large that dark-matter scattering can have an observable impact on structure formation [133]. Such composite dark-matter models can have other phenomenological consequences as well, both at indirect-detection experiments [134, 135] and at colliders [136–139]. Finally, the presence of additional non-Abelian gauge sectors, each with their own analogue of the QCD Θ -angle, could have potential implications for the physics of axions and axion-like particles [140].

While all of these represent theoretically viable possibilities for the dark sector, the dark ensemble we have considered in this chapter is unique for several important reasons. In traditional composite dark-matter models, it is usually a single bound state (usually the lightest bound state) which serves as the primary dark-matter candidate and which therefore carries the full dark-matter abundance Ω_{CDM} . While there may be several other dark states to which this bound state couples — and which may play a role in determining the abundance of the dark-matter candidate — it is neverthe-

less true that only one (or a few) composite states carry the dark-matter abundance Ω_{CDM} and thereby play a significant role in dark-sector phenomenology. By contrast, within the DDM framework, the dark-matter abundance is potentially spread across a relatively large set of composite states with various masses and lifetimes. Thus the usual required stability of the traditional dark-matter candidate is not a required feature of the DDM ensemble, thereby allowing the associated dark-matter abundance $\Omega_{\text{CDM}}(t)$ and dark-matter equation-of-state parameter $w_{\text{eff}}(t)$ to vary with time — even during the current, matter-dominated era.

Moreover, because the DDM framework requires an enlarged viewpoint in which the entire spectrum of composite states are potentially relevant for determining the properties of the dark sector, features that describe the entire composite spectrum suddenly become relevant for determining dark-sector phenomenology — features which would not have been relevant for previous studies within more traditional frameworks. These features include the fact that the masses of such bound states actually lie along Regge trajectories, and that the densities of such bound states experience a Hagedorn-like exponential growth as a function of mass. Indeed, these features do not play a role within traditional studies of composite dark states, but they have been the cornerstones of the analysis we have presented here. In this context, we note that a similar approach was also adopted in Ref. [124] with regard to ensembles of dark nuclei whose abundances are generated via a dark-sector analogue of BBN. This is indeed another context in which the full ensemble of dark-sector states plays an important role in dark-matter phenomenology.

Given the initial steps presented here, there are many avenues for future research. For example, in this chapter we have primarily focused on the phenomenology associated with the “sweet-spot” region in Eq. (4.50), as this region gives rise to a rich spectrum of associated mass scales and DDM-like behaviors. However, other regions may also be relevant for different situations, including the case of dark ensembles emerging from the bulk sectors of actual critical Type I string theories. Indeed, such theories typically have significantly larger central charges and values of D_{\perp} than those corresponding to the $D_{\perp} = 2$ flux tube, and thus correspond to values of (B, C)

which are very far from the “benchmark” values in Eq. (4.49). Such strings also likely correspond to values of (r, s) which are far from those in Eq. (4.50). Likewise, in our analysis we have taken $\kappa = 1$ and $\xi = 3$. Although these simple choices were well-motivated and conservative, it would certainly be interesting to explore the consequences of alternative choices. It would also be of interest to explore the ramifications of relaxing some of the approximations we have made in our analysis. These include the “instantaneous freeze-out” approximation that underpins the Boltzmann suppression factor in Eq. (4.15), as well as our implicit assumption that the Hubble expansion within which our calculations have taken place is unaffected by potential gravitational backreaction from our continually evolving dark sector. While these approximations may certainly be justified to first order, a more refined calculation is still capable of altering our results numerically if not qualitatively.

It would also be interesting to subject the DDM ensembles we have studied here to more detailed phenomenological constraints. The constraints we have studied here, such as our look-back and w_{eff} constraints, are those that follow directly (and in a completely model-independent manner) from knowledge of $\Omega_{\text{tot}}(t)$ and $w_{\text{eff}}(t)$ alone, and as such we have seen that they are sufficient to rule out vast regions of parameter space. It is nevertheless true that a plethora of additional constraints could be formulated once a particular scenario with a particular particle content is specified, and that imposing such additional constraints could potentially narrow our viable parameter space still further.

Finally, and perhaps most importantly, in this chapter we have assumed that the effects of *intra-ensemble* decays on the decay widths of the ensemble constituents are negligible. Such an assumption is certainly consistent with our other assumptions about the structure of the theory. In general, following our string-based approach to understanding the dynamics of these bound-state flux tubes, we may regard the strength of the interactions among the different dark hadrons in our DDM model as being governed by an additional parameter, a so-called “string coupling” g_s , which we have not yet specified but which does not impact any of the results we have presented thus far. In general, g_s can be different from the coupling which governs

the decays of our ensemble states to SM states and which is thus embedded within τ_0 . In an actual string construction, the value of g_s is determined by the vacuum expectation value (VEV) of the dilaton field, but the dynamics that determines this VEV is not well understood. In general, however, intra-ensemble decays will provide an additional contribution to the total decay widths Γ_n , especially for the heavier ensemble constituents, and the decays of these heavier constituents can serve as an additional source for the abundances of the lighter constituents. The effects of such intra-ensemble decays will be discussed in more detail in Ref. [\[15\]](#).

CHAPTER 5

DECIPHERING THE ARCHAEOLOGICAL RECORD:
COSMOLOGICAL IMPRINTS OF NON-MINIMAL DARK SECTORS

In the previous chapter, we have seen that instead of having a single WIMP, dark matter can in general have multiple different components with different masses, lifetimes and cosmological abundances. We have also seen that dark matter models which belong to the framework of DDM rely on a balance between the lifetimes and abundances across the entire dark sector. Although it seems less elegant to introduce more particles, many theories beyond the Standard Model do naturally offer a vast number of different particles with appropriate scaling relations between the masses and the decay widths [13, 40, 43, 44, 56, 57]. With minimal assumptions, those models can be every bit as predictive as the traditional WIMP scenarios [44, 46, 47, 49–52].

The cosmological constraints of these *multi-component* dark-matter models are usually obtained by studying the abundances of each state. In previous DDM models, it is often assumed that the decays of each dark-matter component produce predominantly the SM particles, or other states that live outside the dark ensemble, such as dark radiation. Therefore, the constituents in the ensemble can be simply organized by balancing their cosmological abundances against their SM decay widths. While such assumption is valid in all the previously considered DDM models, in principle, decays from the heavier states to the lighter states within the dark ensemble might also occur. In some scenarios, similar to the decays of QCD hadrons in the Standard Model, such *intra-ensemble decays* might even be dominant.

Similar to those scenarios in which decays of dark-matter components produces solely the SM final states, intra-ensemble decays would not only change the abundance distribution across the dark ensemble, but would also change the evolution of the total abundance of the ensemble. The difference is that, this is not done by leaking energy

into other sectors, but via converting the mass energy of the dark ensemble into its own kinetic energy. Therefore, even though the energy is conserved during each decay, the way the total energy density scales is changed. For example, when non-relativistic particles decay and produce relativistic decay products, the way the energy density scales changes from a^{-3} to a^{-4} until the expansion of the universe makes the decay products non-relativistic.

Beyond the consideration on the total dark-matter abundance and the distribution of it across all the dark-matter components, intra-ensemble decays could also lead to unique imprints in late-time phenomenology even when there is only a single ensemble constituent left in the dark sector today. This may seem bizarre at first sight. However, the phase-space distribution of the lightest constituent resulting from intra-ensemble decays would in general be non-thermal and could even be highly non-trivial. The non-trivial phase-space distribution would subsequently exert its influence during structure formation and could possibly leave some imprints on the large-scale structure of the universe [141, 142]. In general, the way such decays leave their imprints in the late-time universe comprises two steps:

1. States with larger masses decay into states with smaller masses and give them momenta which are determined by the decay kinematics. Since the injection of momenta depends specifically on the decay channels, and at the same time the momenta are redshifted due to the expansion of the universe, in general, the phase-space distributions of the lighter states would be different from the thermal phase-space distributions.
2. In the second step, the kinetic energy obtained from the decays allows dark-matter particles to free-stream, and thus potentially weakening the gravitational potential as regions of higher density tend to have a larger out-going particle flux. When structure starts to form, such weakening of gravitational potential would slow down the growth of density perturbation, preventing structures below certain scales to form. The phase-space distributions contain information about particles' velocities, and hence the distances those particles could travel

and the scales they could suppress. Therefore, the phase-space distribution will give rise to observable features in the large-scale structure. Such features can be detected by measuring the *matter power spectrum*.

Essentially, studying the relation between the dark-matter phase-space distribution and the large scale structure would enable us to learn about dark matter from its gravitational interaction only. It also provides a way to learn about the dark sector even when the dark sector is not talking to the SM sector at all, other than via gravity!

Therefore, in this chapter, we are interested in the whole physical processes from the details of the decays to the resulting phase-space distribution, and eventually to the observational signatures on the matter power spectrum. For concreteness, we shall for simplicity assume all the intra-ensemble-decay processes take place within the RD epoch and only the lightest state is left to play the role of the dark matter that we observe today. In addition, we shall assume that, when intra-ensemble decays start, all the constituents of the dark ensemble have already kinetically decoupled from the rest of the cosmic plasma (or they have never been in equilibrium) and, for consistency, the decay products cannot re-couple with the thermal bath. In principle, it is possible for particles to establish kinetic equilibrium through scatterings with themselves and obtain a thermal distribution. However, such a scenario prevents us from connecting late-time observation with the early-universe processes and would only produce observable signatures that are indistinguishable from those of the traditional thermal dark-matter scenarios. Therefore, we shall also require for consistency that the particle scatterings among the ensemble constituents are not sufficient enough to smear the phase-space distribution established after the intra-ensemble decays. We shall see that the characteristic features in the phase-space distribution indeed produce interesting patterns in the matter power spectrum. In the end, to avoid studying hot dark-matter scenarios which are obviously ruled out by current observations, we shall require that the lightest constituent becomes non-relativistic before matter-radiation equality. The ultimate goal of this chapter is to build a map between the particle-physics model which governs the dynamics of the intra-ensemble decay process and the present-day observational signatures on the large-scale struc-

ture via the phase-space distribution. We shall see that though the forward mapping is straightforward by numerically solving the Boltzmann equations, the attempt to invert the mapping is extremely difficult due to fact that distinct features in phase-space distribution are usually blurred in the matter power spectrum. Nevertheless, we shall also see that qualitative features can be obtained which allow us to partially decipher the information about the early-universe processes.

This chapter is structured as follows. In Sec. 5.1, we first develop a theoretical understanding on the way non-trivial phase-space distributions are generated via the interplay between the decay branching ratio, the kinematics and the expansion of the universe. Then, we study the structure formation using free-streaming analysis and construct a map between the phase-space distribution and the shape of the corresponding the matter power spectrum. In Sec. 5.2, we present the study of a toy model. We obtain the phase-space distributions by solving a large set of Boltzmann equations, and use them to obtain the matter power spectra. We then apply the techniques developed in the previous section to demonstrate the mapping. In Sec. 5.3, we conclude.

5.1 From Intra-Ensemble Decays to Structure Formation

As is mentioned at the beginning of this chapter, the physics of intra-ensemble decays consists of two steps. The first step concerns the production of non-thermal distribution via the decays, while the second step concerns the relation between the phase-space distribution and the matter power spectrum. In the following subsections, we shall study the two steps one by one.

5.1.1 Packets to Packets, Dust to Dust: Evolution of Phase-Space Distributions

For a non-minimal dark sector, the phase-space distribution function $f_i(\vec{x}, \vec{p}, t)$ contains all information pertinent to the cosmological evolution of a particular constituent i in the dark sector. To zeroth order, our universe is homogeneous and isotropic, and

therefore the spatial and directional dependence are usually dropped from the distribution function, i.e. $f_i(\vec{x}, \vec{p}, t) \approx f_i(p, t)$, where $p \equiv |\vec{p}|$. The zeroth-order number density n_i , energy density ρ_i , pressure P_i and the equation of state (EoS) parameter w_i for any species i are defined as follows:

$$n_i(t) \equiv g_i \int \frac{d^3p}{(2\pi)^3} f_i(p, t) , \quad (5.1)$$

$$\rho_i(t) \equiv g_i \int \frac{d^3p}{(2\pi)^3} E f_i(p, t) , \quad (5.2)$$

$$P_i(t) \equiv g_i \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} f_i(p, t) , \quad (5.3)$$

$$w_i(t) \equiv \frac{P_i(t)}{\rho_i(t)} , \quad (5.4)$$

where $E = \sqrt{p^2 + m_i^2}$, and g_i stands for the internal degrees of freedom. Obviously, for ultra-relativistic species ($E \sim p$), we have $w_i \approx 1/3$, whereas, for non-relativistic species ($E \sim m_i \gg p$), we have $w_i \approx 0$. If a species in consideration is in kinetic equilibrium, $f_i(p, t)$ would take the form of a Bose-Einstein or Fermi-Dirac distribution $f_i(p, t) = 1/(\exp(E/T) \pm 1)$ with “−” for bosons and “+” for fermions.

It is useful to study the way the distribution function evolves with time. According to the FRW metric (see Eq. (2.2)), as the universe expands, the physical distance $x(t)$ between two otherwise stationary points at time t is related to that at time t' by $x(t) = x(t')a(t)/a(t')$, where a is the scale factor. To maintain the Poisson bracket $\{x, p\} = 1$, the momentum p therefore has to evolve inversely with respect to the coordinate, i.e. $p(t) = p(t')a(t')/a(t)$. This describes the redshift of momentum due to the expansion of the universe. We thus have

$$\begin{aligned} \frac{dp(t)}{dt} &= \frac{d}{dt} \left(p(t') \frac{a(t')}{a(t)} \right) = -H(t)p \\ \Rightarrow \frac{d \log p}{dt} &= -H(t) , \end{aligned} \quad (5.5)$$

where $H \equiv \dot{a}/a$ is the Hubble parameter. Thus time evolution corresponds to additive shifts in the value of $\log p$. This makes $\log p$ a particularly convenient variable

for studying time evolution. Assuming spatial homogeneity and isotropy of the phase-space distribution, we can therefore write

$$n(t) \sim \int d^3p f(p, t) \sim \int dp p^2 f(p, t) \sim \int d \log p p^3 f(p, t) . \quad (5.6)$$

Thus, the proper phase-space distribution with respect to $\log p$ is $p^3 f$. Given the manner in which $x(t)$ rescales under time evolution, we know that n must scale under time evolution as a^{-3} . Moreover, since $\log p$ merely accrues an additive shift under time evolution, $d \log p$ is invariant. Therefore, $p^3 f$ scales as a^{-3} . For this reason, it proves convenient to define the invariant quantity $g(p, t) \equiv a(t)^3 p^3 f(p, t)$ which therefore satisfies¹

$$g(p(t), t) = g(p(t'), t') . \quad (5.7)$$

Thus, if we plot $g(p)$ versus $\log p$, the total area under the curve is proportional to the dark-matter particle number $N \sim n a^3$, and under time evolution and barring any further dark-matter creation, the curve for $g(p)$ merely slides towards smaller values of $\log p$ without distortion, as if carried along on a “conveyor belt” moving with velocity $H(t)$.

In this paper, we shall be concerned with the manner in which a dark-matter distribution $g(p)$ is created as a result of physical processes that occur during cosmological evolution. In general, various processes can create dark matter and thereby make different contributions to $g(p)$ as the universe evolves. Our goal is ultimately to understand what kinds of cosmological history can give rise to non-trivial features in the resulting $g(p)$ distribution, and perhaps even to approach the inverse problem of determining the extent to which a determination of the late-time dark-matter distribution $g(p)$ can be used to constrain the cosmological history that give rise to it.

To help simplify and conceptualize the processes in play, we shall begin by considering an actual physical conveyor belt that moves with constant velocity v in the $-\hat{x}$ direction. The coordinate $x(t)$ of an object on the conveyor belt at time t is therefore

¹We are neglecting effects from particle interactions here.

related to its coordinate at an earlier time t' by

$$x(t) = x(t') - v(t - t') . \quad (5.8)$$

At time t , one can measure the number distribution of objects $N(x)$ as a function of location x . In general, there are many ways in which a given object distribution $N(x)$ at time t might be generated as the result of objects having been deposited onto the conveyor belt at different locations $y \geq x$ and at different times $t' \leq t$. Indeed, we can formally write $N(x)$ at any time t as an integral over time which sums over all the deposits:

$$N(x) = \int^t dt' \Delta(x + v(t - t'), t') , \quad (5.9)$$

where $\Delta(x, t')$ indicates the deposit profile at time t' which gives the rate at which objects are being deposited at any specific location and time, and where the integral is understood to extend from an early “primordial” time prior to all deposits. If the deposits occur at discrete times t_i prior to t , the rate function can be expressed as

$$\Delta(x, t') = \sum_i \Delta_i(x) \delta(t' - t_i) , \quad (5.10)$$

in which $\Delta_i(x)$ is the spatial distribution of the deposit at t_i . Substituting this into Eq. (5.9), we see that

$$N(x) = \sum_i \Delta_i(x + v(t - t_i)) \quad (5.11)$$

which clearly shows the distribution of objects along the entire conveyor belt at any moment t is an accumulative result of all the deposits at previous times.

An example of this is illustrated in Fig. 5.1, in which packets are dropped on a left-moving conveyor belt. The packets shaded in black at positions A/B and E are related to the packets in red at C/D and G at t_2 through the horizontal flow of the conveyor belt. In turn, the packets in red consists of packet D which is translation of the packet F at t_1 as well as packets C and G which are deposited right at t_2 . Using

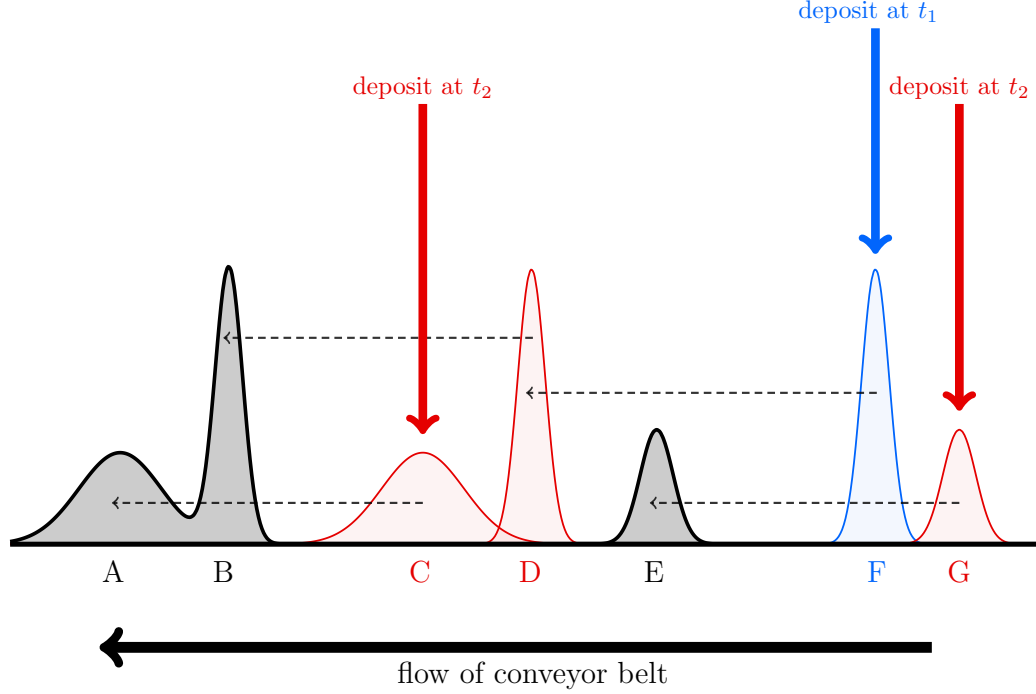


FIGURE 5.1: Sketch of the cosmological “conveyor belt” at three different times: an early time t_1 (blue) at which wavepacket F is deposited, a later time t_2 (red) at which wavepackets C and G are deposited, and a final time (black). The horizontal axis is $\log p$, and the area of the packets are associated with the comoving number density $g(p, t)$. During the time interval between t_1 and t_2 , wavepacket F has redshifted to position D , where it overlaps with the newly deposited wavepacket C . By contrast, after time t_2 , both the combined wavepacket C/D and the wavepacket G continue to redshift into their final positions A/B and E . In the absence of further information, knowledge of the final wavepackets A/B and E allows us to resurrect only a portion of this cosmological history.

the formulas that we have just developed above, the rate function is

$$\Delta(x, t') = \Delta_F(x)\delta(t' - t_1) + [\Delta_C(x) + \Delta_G(x)]\delta(t' - t_2) , \quad (5.12)$$

and the final distribution at time t can be expressed as

$$N(x) = \Delta_F(x + v(t - t_1)) + \Delta_C(x + v(t - t_2)) + \Delta_G(x + v(t - t_2)) . \quad (5.13)$$

Note that the distribution “observed” at any moment t depends on the entire history of the deposits before t . It would be ideal if one can recover the previous history of deposits by looking at a snapshot of the distribution at a certain moment. However,

without further information, only a portion of this history can be resurrected. For example, even if we know packets A, B and E are translated from C, D and G, we cannot tell if packet D is deposited right at t_2 or it is produced at position F at t_1 . In fact, the two different processes can produce exactly the same final phase-space distribution. So, when one measures the distribution, it is only the summation over the history that is measurable — what happened at a specific location on the conveyor belt prior to the measurement have already been integrated out.

Let us map our conveyor-belt analogy back to the phase-space evolution in cosmology. We can interpret the packets in Fig. 5.1 as the phase-space distribution of the dark-matter particles. The deposits thus correspond to early-universe processes that produce those dark-matter particles. And the flow of the conveyor belt is just the redshift effect which manifests as a translation of the phase-space distribution in the direction of lower momentum (to the left). Therefore, we have the following dictionary:

$$\begin{aligned} x &\rightarrow p , \\ x + v(t - t') &\rightarrow p \frac{a(t)}{a(t')} , \\ N(x) &\rightarrow g(p) . \end{aligned} \tag{5.14}$$

Using this map in Eq. (5.9), and let $\Delta(p, t)$ indicate the profile of the dark-matter deposit rate at time t , we have

$$g(p) = \int^t dt' \Delta \left(p \frac{a(t)}{a(t')}, t' \right) . \tag{5.15}$$

Likewise, if deposits occur at discrete times t_i , *i.e.*

$$\Delta(p, t') = \sum_i \Delta_i(p) \delta(t' - t_i) , \tag{5.16}$$

then

$$g(p) = \sum_i \Delta_i \left(p \frac{a(t)}{a(t_i)} \right) . \tag{5.17}$$

Therefore, all the information about the early-universe processes are encoded in the deposit profile $\Delta(p, t')$. However, it is only the phase-space distribution, which is the integral of $\Delta(p, t')$, that can be measured.

For concreteness, let us study a scenario in which the deposits occur via decays. Assuming the existence of some parent packets labelled by ℓ which decay instantaneously at moments T_ℓ and produce packets in the phase-space distribution of dark matter, then the deposit profile can be expressed as

$$\Delta(p, t') = \sum_{\text{parent } \ell} \int d \log p_\ell g_\ell(p_\ell) P_{\ell \rightarrow 0}(p_\ell, p) \delta(t' - T_\ell) , \quad (5.18)$$

in which $g_\ell(p_\ell)$ is the phase-space distribution of the packet ℓ when it decays at time T_ℓ , and $P_{\ell \rightarrow 0}(p_\ell, p)$ is the probability that the decay of the momentum slice at p_ℓ would produce a dark-matter particle with momentum p normalized by the multiplicity of the decay. Substituting this expression into Eq. (5.15), we then have

$$g(p) = \sum_{\text{parent } \ell} \int d \log p_\ell g_\ell(p_\ell) P_{\ell \rightarrow 0} \left(p_\ell, p \frac{a(t)}{a(T_\ell)} \right) . \quad (5.19)$$

In general, the parent packets ℓ could be descendants from some grandparent packets k , which in turn could also be descendants from some grand-grandparent packets and so on. Therefore, we can recursively add all the previous decays which eventually gives

$$\begin{aligned} g(p) = & \sum_{\{a, b, \dots, k, \ell\}} \int d \log p_\ell \int d \log p_k \cdots \int d \log p_b \int d \log p_a g_a(p_a) \\ & \times P_{a \rightarrow b} \left(p_a, p_b \frac{a(T_b)}{a(T_a)} \right) P_{b \rightarrow c} \left(p_b, p_c \frac{a(T_c)}{a(T_b)} \right) \cdots P_{k \rightarrow \ell} \left(p_k, p_\ell \frac{a(T_\ell)}{a(T_k)} \right) \\ & \times P_{\ell \rightarrow 0} \left(p_\ell, p \frac{a(t)}{a(T_\ell)} \right) , \end{aligned} \quad (5.20)$$

where $\{a, b, \dots, k, \ell\}$ denotes the summation over all the decay chains of the packets that produce dark matter, $P_{i \rightarrow j}$ is defined in the same way as $P_{\ell \rightarrow 0}(p_\ell, p)$, T_i is the instant when the packet i decay, and it is defined to be $T_i \equiv T_0^{(a)} + \tau_a + \tau_b + \cdots + \tau_i$

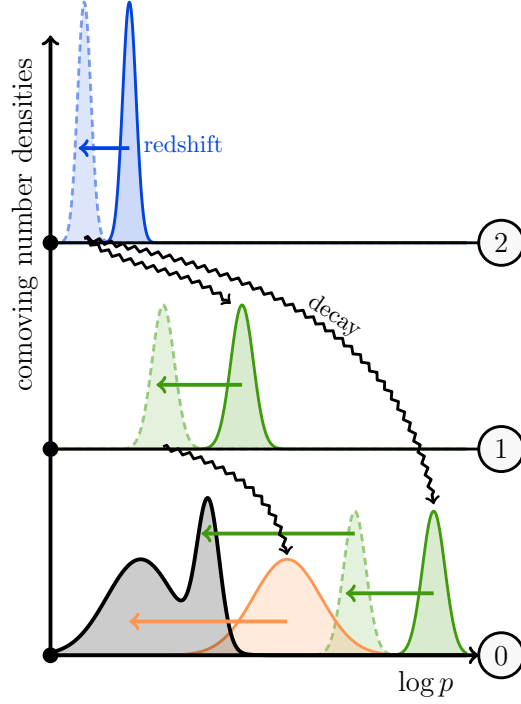


FIGURE 5.2: One possible scenario leading to a non-trivial distribution $a^3 p^3 f_0(p)$ at late times through a sequence of successive “deposits” of the sort sketched in Fig. 5.1. An excited state is created with a thermal distribution at level $\ell = 2$ (blue) and redshifts towards smaller momenta before undergoing a two-body decay $2 \rightarrow 1 + 0$. Each of these distributions then redshifts until the daughter at level $\ell = 1$ undergoes its own decay $1 \rightarrow 0 + 0$. The resulting distributions combine and continue to redshift, ultimately producing a final distribution $a^3 p^3 f_0(p)$ (black).

in the instantaneous decay limit in which τ_i is the lifetime of the packet i , and $T_0^{(a)}$ is the time of production of the earliest ancestor a .

Fig. 5.2 is an example of having two decay chains in a three-state system. At the very beginning, an excited state is created with a thermal distribution at level $\ell = 2$ (blue) and redshifts towards smaller momenta before undergoing a two-body decay with daughter states at levels $\ell = 1$ and $\ell = 0$ (green). Each of these distributions then redshifts until the daughter at level $\ell = 1$ undergoes its own decay into two kinematically identical grand-daughters at $\ell = 0$ (orange). The new-born orange packet happens to have a non-negligible overlap with the redshifted green packet (dashed green). Therefore, they superpose, and the resulting distributions continue to redshift, ultimately producing a final distribution $a^3 p^3 f_0(p)$ (black) which resembles wavepackets A/B in Fig. 5.2.

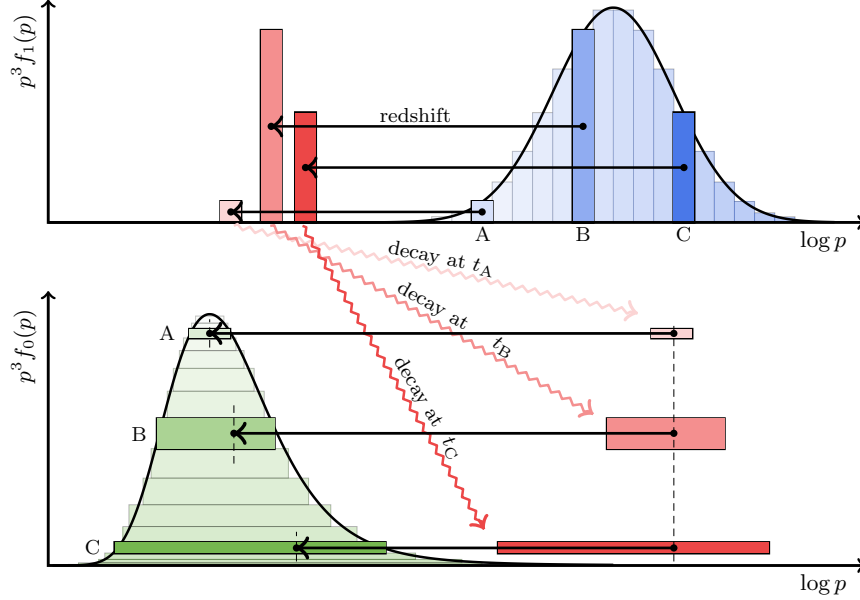


FIGURE 5.3: The process through which a parent packet (blue) decays into a daughter packet (green) in an expanding FRW cosmology, with intermediate steps indicated (red/pink). Each momentum slice of the parent packet experiences a different time-dilation factor which not only extends the lifetime associated with that slice but also increases the total accumulated cosmological redshift experienced by that momentum slice prior to decay. The sequential decays of these redshifted momentum slices of the parent then make sequential contributions to the daughter packet, with each contribution extending over an increasingly broad range of momenta and experiencing its own cosmological redshift immediately upon production. This process ends with the decay of the last momentum slice of the parent, thereby completely depleting the parent packet (blue) and yielding the total daughter packet (green).

Note that, although the original parent has a thermal distribution with relatively small width Δp , each decay to successive generations (blue, then green, then orange) produces offspring whose distributions have not only greater central momenta but also greater widths. Likewise, the superposition of results from the two different decay channels $2 \rightarrow 0$ and $2 \rightarrow 1 \rightarrow 0$ produces the final distribution $a^3 p^3 f_0(p)$ which is non-thermal and in fact multi-modal.

The process by which a parent packet decays into a daughter packet is surprisingly subtle within an expanding cosmology. These subtleties arise due to the relativistic effects which emerge if the momenta involved are large compared to the mass of the parent, and are sketched in Fig. 5.3. Let us assume that the parent species has mass m and is produced at some time $t = t_0$ with a phase-space distribution $a^3 p^3 f_0(p)$

as measured in the lab frame. Let us also assume that the parent species has decay width Γ , and for simplicity we shall further assume that the decays of the parent occur promptly at $\tau \equiv 1/\Gamma$, as measured in the parent rest frame. Because the parent packet $a^3 p^3 f(p)$ stretches across a non-zero range of momenta, each momentum slice of the parent packet will experience a different time-dilation factor. Then then increases the effective lifetime of the slice prior to decay, causing slices with larger momenta to decay later than those with smaller momenta. However, during the time interval prior to decay, each momentum slice also experiences a cosmological redshift. This redshift decreases the momentum of each slice and thereby partially mitigates the time-dilation effect. Indeed, combining both effects, we see that each momentum slice with initial momentum p_0 at $t = t_0$ will decay at

$$t_{\text{decay}}(p_0) = \int_{t_0}^{t_0+\tau} dt' \frac{\sqrt{p_0^2 (t_0/t')^{2\kappa/3} + m^2}}{m}, \quad (5.21)$$

where we have assumed an FRW cosmology with the Hubble factor scaling as $H \sim \kappa/3t$. Likewise, the momentum of this slice at the time of its decay is given by

$$p_{\text{decay}}(p_0) = p_0 \left[\frac{t_0}{t_{\text{decay}}(p_0)} \right]^{\kappa/3}, \quad (5.22)$$

where $t_{\text{decay}}(p_0)$ is given in Eq. (5.21).

As we move upwards within the parent packet towards momentum slices with increasing values of p_0 , the time-dilation factor increases but this also provides a longer time interval during which cosmological redshifting occurs. It is therefore important to understand the extent to which this extra redshifting might compensate for the higher original momentum of the momentum slice. Indeed, it is even possible that momentum slices with greater initial momenta p_0 might have *smaller* final momenta p_{decay} when they actually decay. However, it is straightforward to verify that $dp_{\text{decay}}(p_0)/dp_0 > 0$. Thus, momentum slices with greater initial momenta p_0 continue to have greater momenta when they each decay. However, it is also straightforward to verify that $dp_{\text{decay}}(p_0)/dp_0 < 1$. Thus, any two momentum slices whose original momenta p_0 differ by an amount Δp_0 will have decay momenta p_{decay} differing by

an amount $\Delta p_{\text{decay}} < \Delta p_0$. This “momentum compression” is illustrated along the top portion of Fig. 5.3, where the relative horizontal spacings between the original (blue) momentum slices labelled A , B , and C are larger than the relative horizontal spacings between the corresponding redshifted (red/pink) momentum slices which are sketched at the (different) times of their decays.

The decay of each momentum slice of the parent then yields a contribution to the emerging phase-space distribution $a^3 p^3 f_0(p)$ of the daughter. In general, the shapes of these contributions depend on the details of the decay process. However, two features tend to be rather general and will be assumed here. To explain these features, we first note that because the combined masses of the daughters are generally less than the mass of the parent, each daughter will always be produced with a non-zero momentum in the rest frame of the parent. Likewise, because the direction of the daughter momentum in the parent rest frame is uncorrelated with the direction of the momentum of the parent in the lab frame, the magnitude of the momentum of each daughter — a quantity which is uniquely determined in the rest frame of the parent — is broadened by the boost of the parent into a *range* of momentum magnitudes as measured in the lab frame.

These observations tell us that the magnitudes of the daughter momenta emerging from the decay of each parent momentum slice will generally exhibit two characteristics: they will fall within a range of momenta which grows as a function of p_{decay} , and this range of momenta will be “centered” around a non-zero value which we may identify as the daughter momentum that would emerge even if p_{decay} were zero. Thus, the contribution to the daughter packet which comes from the decay of a momentum slice of the parent with momentum p_{decay} will extend horizontally across different momenta, spanning a range that grows with p_{decay} and therefore with p_0 . We further note, of course, that the total area associated with each such contribution will generally be proportional to the total area of the original decaying momentum slice of the parent, with the proportionality constant signifying the number of daughters produced through the decay of each parent.

These features can easily be understood in the case of a simple two-body decay

of the form $X \rightarrow YY$ with parent and daughter masses m_X and m_Y respectively, where $m_Y < m_X/2$. The energy of each daughter in the rest frame of the parent is simply $m_X/2$, and likewise the magnitude of the momentum of each daughter in the rest frame of the parent is given by $\sqrt{(m_X/2)^2 - m_Y^2}$. Indeed, this is the non-zero momentum that is imparted to the daughters in the lab frame when the parent is at rest. However, if the parent has momentum p_X in the lab frame at the time of its decay, then the energy of each daughter in the lab frame is given by

$$E_Y = \frac{1}{2}\sqrt{m_X^2 + p_X^2} + \frac{p_X}{2}\sqrt{1 - 4m_Y^2/m_X^2} \cos \theta \quad (5.23)$$

where θ is the angle between the daughter momentum in the parent rest frame and the parent momentum. Given that the angle θ is unfixed by the kinematics of the decay, these daughter energies E_Y will therefore vary within a total range of magnitude

$$\Delta E_Y = E_Y^{(\max)} - E_Y^{(\min)} = p_X \sqrt{1 - 4m_Y^2/m_X^2}. \quad (5.24)$$

Likewise, for any daughter with energy E_Y within this range, the corresponding daughter momentum in the lab frame is simply given by $p_Y = \sqrt{E_Y^2 - m_Y^2}$; thus the existence of a range of possible energies for the daughters implies the existence a corresponding range of daughter momenta. Moreover, we see that daughter energy range ΔE_Y — and indeed the corresponding daughter momentum range Δp_Y — both increase as functions of p_X , as claimed.

These features are illustrated in the lower (red/pink) portion of Fig. 5.3. For visual simplicity, we have sketched each contribution to the daughter packet as a “brick” whose height does not vary with momentum, but of course this precise shape will generally depend on the detailed kinematics associated with the decay. However, we observe as a general result that our bricks are deposited in order of increasing width along the momentum axis. Moreover, because our original parent packet $a^3 p^3 f_1(p)$ has a profile which first rises and then falls as a function of momentum, the total areas of the bricks that are deposited first grow and then shrink as the decay process proceeds. Finally, in making this sketch we have assumed that the mass of the parent greatly

exceeds the combined masses of the daughters. It is for this reason that we have sketched our bricks as having momenta which greatly exceed p_{decay} . Likewise, while we have accurately indicated the *relative* areas of each brick in making this sketch, in the interests of generality we have not assumed a particular daughter multiplicity associated with this decay. Indeed, it is only through such a multiplicity that we would be able to assign a relative normalization between the area of each brick and that of the corresponding momentum slice of the parent prior to decay.

The final stage of the evolution from parent to daughter packet once again involves cosmological redshifting. Because the narrowest “bricks” are deposited first, they begin redshifting towards smaller momenta before the subsequent, wider “bricks” have even been deposited. Thus the narrowest “bricks” experience the largest redshifts, a fact which tends to cause the emerging daughter packet to experience a “tilt” towards smaller momenta — an effect which is particularly pronounced when the daughter particles are originally relativistic when produced.

Ultimately, the final shape of the daughter packet — *i.e.*, the final phase-space distribution of the daughter — is then given by the sum of these different bricks at the time when the final brick is deposited. The resulting packet then simply continues to redshift intact along the momentum conveyor belt towards lower momenta without any further changes to its shape.

Looking over the entire process shown in Fig. 5.3, we see that each *vertical* momentum slice in the original parent packet ultimately gives rise to a *horizontal* slice in the daughter packet. Thus, while the resulting shape of the daughter packet is ultimately sensitive to many kinematic details associated with the decay process, certain general conclusions can be drawn. For example, the *widths* of the daughter packet are correlated with the redshifted momenta of the individual momentum slices of the parent at the times when they decay, with the maximum width along the “base” of the daughter packet corresponding to the right-most of these momentum slices within the parent packet and the narrowest width (which governs the shape of the “peak” of the daughter packet) corresponding to the left-most of the parent momentum slices. In a similar fashion, as discussed above, the ascending and descending

slopes of the daughter packet are correlated with the vertical heights of the corresponding deposited bricks, yet these are correlated with the *area* of each brick which is in turn correlated with the area of the originally decaying momentum slice of the parent and thus with the corresponding value of $a^3 p^3 f_1(p)$. In this way, many gross features concerning the shape of the daughter packet can be directly connected to the shape of the parent packet.

Notice that, even though our example in Fig. 5.3 involves only two packets, the dynamics of decay is already very complicated, not to say that we have made the assumptions that all the momentum slices of the parent packet were created at the same time, and that decays occur instantaneously. In reality, different parts of the parent packet might be produced through different processes at different moments, and decays occur continuously. Therefore, even the particles belonging to the same momentum slice should not be expected to decay at the same time. Given this complexity, it is important to study examples numerically by exactly solving the Boltzmann equations of some toy models. In Fig. 5.4 we show the phase-space distributions of the ground state from a 3-component system right at the moment when intra-ensemble decays end (*i.e.* when the comoving number density of the ground state reaches 99.5% of its final asymptotic value). The masses of the states are $m_2 = 7m_0$ and $m_1 = 3m_0$, where m_0 is the mass of the ground state. The initial condition is that the heaviest state is the only state populated and has a Boltzmann-like distribution $f_2 \sim \exp(-E/T_0)$ (black, upper panel), with $T_0 = m_0/20$. We present five different cases which produce five different decay histories and thus five different ground-state phase-space distributions:

- $\Gamma_{200} = 1, \Gamma_{211} = 0, \Gamma_{100} = 0$ (blue):

The heaviest state decays into the ground state directly. This is effectively a 2-component system. By setting $\Gamma_{200} = 1$, all the other decay widths are normalized by this case.

- $\Gamma_{200} = 0, \Gamma_{211} = 1, \Gamma_{100} = 0.01$ (jade):

The direct decay into the ground state is forbidden. The decays have to go

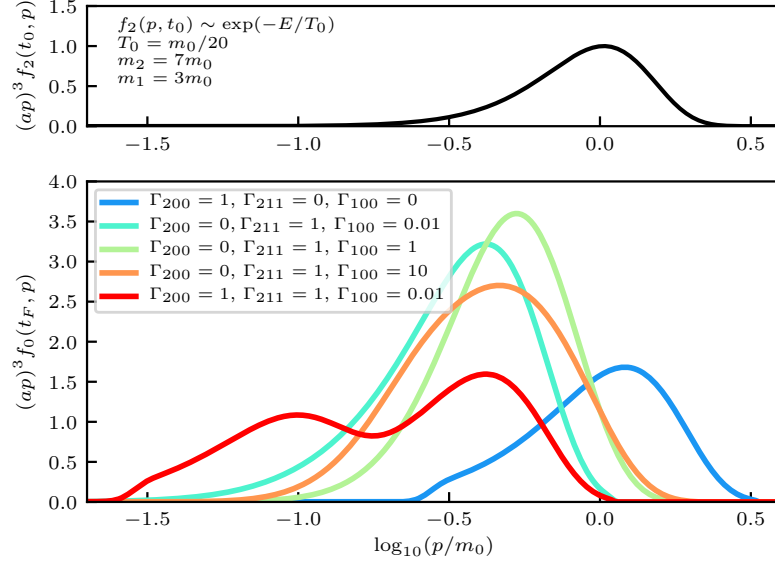


FIGURE 5.4: Phase-space distributions of the ground state right after intra-ensemble decays complete. Five cases are shown with different partial decay widths indicated in the legend. The vertical axis is normalized randomly, but the relative area of each packet is directly proportional to the corresponding comoving particle number. The decay widths are made dimensionless and normalized such that $\Gamma_{200} \sim H(t_0)/10$. The final time t_F is chosen at the moment when the comoving number density of the ground state reaches 99.5% of its final asymptotic value. Since the total decay width varies in each case, these cases do not necessarily end at the same time.

through two steps to reach the ground state — $2 \rightarrow 1+1$, and then $1 \rightarrow 0+0$. However, as compared with the first step, the decay rate of the second step is much slower. Therefore, the energy from the state 2 can be stored in the state 1 before the state 1 eventually decays into the ground state.

- $\Gamma_{200} = 0, \Gamma_{211} = 1, \Gamma_{100} = 1$ (green)

Similar to the last case, but the decay rate from 1 to 0 is the same with the decay rate from 2 to 1. Therefore, when the energy of state 2 is transferred to the state 1, it is also gradually transferred to the state 0.

- $\Gamma_{200} = 0, \Gamma_{211} = 1, \Gamma_{100} = 10$ (orange)

Similar to the previous two cases, but with a much faster second step. Therefore, as soon as the energy is transferred from state 2 to state 1, it is almost immediately dumped into the ground state.

- $\Gamma_{200} = 1, \Gamma_{211} = 1, \Gamma_{100} = 0.01$ (red)

Two decay chains to the ground state. It can be seen as the (weighted) combination of the blue and the jade case cases. Therefore, packets are deposited to the ground state at two different time scales.

We see that, though the initial distribution of the parent state is the same, the phase-space distribution of the ground state after going through different decay histories can be quite different. Let us first compare the number of particles produced in each case by looking at the decay topology and the area under each curve. Denoting the area under the black curve A_{black} , we see $A_{\text{blue}} = 2A_{\text{black}}$, $A_{\text{jade}} = A_{\text{green}} = A_{\text{orange}} = 4A_{\text{black}}$ and $A_{\text{red}} = 3A_{\text{black}}$. This is consistent with decay topology, since one direct decay produces two ground-state particles ($2 \rightarrow 0 + 0$), whereas a two-step decay produce four ground-state particles ($2 \rightarrow 1 + 1 \rightarrow 0 + 0 + 0 + 0$).

The shape of the ground-state distributions can be understood using what we have learnt from Figs. 5.1 - 5.3. Let us first focus on the jade, green and orange cases, since they only differ in the decay width of the last step. The overall width of the orange packet is the largest among the three. This is not surprising. Indeed, in the orange case, particles in the state 1 can decay as soon as they are produced, which means they have larger momenta when decaying. Therefore, their products obtain a larger range of momenta, and make the width of the packet larger, and the height smaller. However, this argument does not quite apply when comparing the green case with the jade case, as the jade packet appears to have a smaller height. To understand this, one needs to keep in mind that 1) the decay is not a instantaneous process, 2) these snapshots of phase-space distributions are not taken at the same time — they are taken right at the moment when intra-ensemble decays just finish. Comparing with the green case, the jade case has a very slow second step, and thus the snapshot is taken at a much later time. Its earliest deposits in the ground state therefore experience a larger amount of redshift, making the left edge very extended. This explains why the bottom part of the jade packet is wider than the green packet, and why its height is smaller.

In the end, the red case has two available decay chains for the heaviest state to decay into the ground state which make it a “superposition” of the blue and the jade case. Indeed, the resulting phase-space distribution is multi-modal. It is a weighted sum of the blue and the jade packets with packet at lower momentum corresponding to the blue packet, and the packet at higher momentum corresponding to the jade packet. Besides, since the value of Γ_{100} is the same in both the jade and the red cases, the decays in both cases complete at the same time. Therefore, we see the second packet in the red case aligns perfectly with packet in the jade case.

To conclude, we have seen that intra-ensemble decays can in general produce non-thermal and sometimes multi-modal phase-space distributions. The shape of the final phase-space distribution of the ground state depends on many factors, such as the decay kinematics, the phase-space distribution of the parent, the available decay chains, and the branching ratios, *etc.* Since there might be different ways to produce the same ground-state phase-space distribution, knowing its shape only allows us to partially resurrect the decay history. The process of a simple direct decay like the one depicted in Fig. 5.3 is surprisingly subtle. And we have seen in Fig. 5.4 that adding just one intermediate state is enough to make the analysis much more complicated. In later sections, when dealing with a large number of constituents in the dark ensemble, we shall therefore rely on numerical simulation to obtain the full evolution of phase-space distributions.

5.1.2 Structure Formation

The previous subsection enables us to understand how decays in a non-minimal dark sector affect the phase-space distribution of the ground state. We have seen that, highly non-trivial multi-modal phase-space distributions can easily emerge even from a simple 3-component system. In this subsection we try to develop tools useful for analyzing the observational effects from those general non-trivial distribution functions.

So far, the phase-space distributions we have been studying are assumed to be homogeneous and isotropic. However, the structures that we have observed across

various scales today imply the existence of perturbations to the homogeneous background in the early universe. Indeed, in the early universe, the primordial perturbations emerge as overdense or underdense regions. The overdense regions have stronger gravitational pull and tend to attract more and more matter. As a result, those initially small perturbations can be amplified by gravity and eventually form complex structures.

The perturbation to the energy density is usually written in the form below

$$\delta(\vec{x}, t) \equiv \frac{\delta\rho(\vec{x}, t)}{\bar{\rho}(t)} = \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \quad (5.25)$$

where $\bar{\rho}$ is the zeroth-order, unperturbed energy density which depends on time only, and $\rho(\vec{x}, t) \equiv (1/(2\pi)^3) \int d^3p f_i(\vec{x}, \vec{p}, t)$ is the actual energy density that varies in space. It is often convenient to switch to the Fourier modes using

$$\delta(\vec{x}, t) = \frac{1}{(2\pi)^3} \int d^3k \delta_k(t) \exp\left[-\frac{i\vec{k} \cdot \vec{x}}{a(t)}\right], \quad (5.26)$$

where \vec{k} is the comoving wavenumber. The formation of structures is essentially related to the evolution of the density perturbation modes δ_k .² A complete study of this requires solving the Einstein's equation, which is very complicated. We shall do that numerically using existing packages later. Normally, it is easy to read off some characteristic physical behaviors using the *Jeans'* equation, which is appropriate for perturbation modes whose spatial scales are smaller than the size of the Hubble horizon H^{-1} :

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k + \left(\frac{\sigma^2 k^2}{a^2} - 4\pi G\bar{\rho}\right)\delta_k = 0, \quad (5.27)$$

where the quantity σ is the adiabatic sound speed ($\sigma \equiv \sqrt{(\partial P/\partial\rho)_{\text{ad}}}$) that describes the reaction of pressure due to the change of energy density for collisional fluid, such as photons and baryons, or the velocity dispersion ($\sigma \equiv \sqrt{(\langle v^2 \rangle - \langle v \rangle^2)/3}$) that describes particle free-streaming for collisionless fluid. For dark-matter particles that

²Of course there are also perturbations to the metric, velocity, *etc.* These are all necessary ingredients in solving the Einstein's equation, but we are ignoring the details here.

have already decoupled from the rest of the thermal bath, it is more appropriate to consider them as collisionless. The usual Jeans' analysis is done by studying the expression in the brackets. The first term is always positive, which describes the resistance to gravitational collapse from the velocity of the particles. The second term is the gravitational term — the minus sign in front of it suggests that gravity is always driving the growth of perturbations. Therefore, by setting the expression in the brackets equal to zero, one can obtain a critical wavenumber, the *Jeans' scale*, that separates the gravitationally stable and unstable modes:

$$k_J \equiv \sqrt{\frac{4\pi G \bar{\rho} a^2}{\sigma^2}} = \frac{3}{2} \frac{aH}{\sigma} . \quad (5.28)$$

For modes with $k \gg k_J$, the solutions are oscillatory, which means the growth of perturbations is impeded, whereas, for those with $k \ll k_J$, growing solutions exist, and perturbations can grow.

Note that, the Jeans' scale evolves. The velocity dispersion (or sound speed) σ does not redshift in the ultra-relativistic regime, whereas it falls as $\sigma \sim 1/a$ once the particles become non-relativistic. The scaling of Hubble parameter depends on its energy content since $H \sim \sqrt{\bar{\rho}}$. In the RD epoch, $H \sim a^{-2}$, while in the MD epoch, $H \sim a^{-3/2}$ (we neglect the recent Λ -dominated epoch since this has little effect on our results). Therefore, for viable dark-matter candidates which become non-relativistic during the RD epoch, k_J takes its minimum value around the matter-radiation equality. Thus, $k_J(t_{\text{eq}})$ often serves as a rough estimate for the largest suppression scale at the present day.

Despite being time dependent, a single quantity k_J is sufficient for most traditional scenarios where the dark-matter phase-space distribution $a^3 p^3 f(p, t)$ is thermal and unimodal, *i.e.*, there is a single local maximum at a given time. In these cases, the velocity dispersion is a useful indicator of the actual speed of particles in the distribution. This is important, since, in the traditional free-streaming analysis, all information regarding $a^3 p^3 f(p, t)$ is reduced down and carried by the velocity dispersion. However, the phase-space distribution resulting from intra-ensemble decays

can easily be non-trivial and even be multi-modal. For these situations, the velocity dispersion can only capture the gross feature of the phase-space distribution but fail to retain all the information.

To keep the full information from the phase-space distribution, one might attempt to treat each momentum slice in the phase space as an “effective” dark-matter species and make use of the multi-species Jeans’ equations [22]:

$$\ddot{\delta}_{\alpha,k} + 2\frac{\dot{a}}{a}\dot{\delta}_{\alpha,k} + \left(\frac{\sigma_\alpha^2 k^2}{a^2} \delta_{\alpha,k} - 4\pi G \bar{\rho} \sum_{\beta} \epsilon_{\beta} \delta_{\beta,k} \right) = 0. \quad (5.29)$$

In the above expression, subscripts α and β are used to denote different species (momentum slices), $\epsilon_{\alpha} \equiv \bar{\rho}_{\alpha}/\bar{\rho}$ is the fraction of energy density contributed by the species α , and $\delta_{\alpha,k}$ is redefined to be the fractional density perturbation of the species itself $\delta_{\alpha,k} \equiv (\delta\rho_{\alpha}/\bar{\rho}_{\alpha})_k$. The summation $\sum_{\beta} \epsilon_{\beta} \delta_{\beta,k}$ therefore reflects the fact that the gravitational potential is built up by density perturbations from all the species. Following the analysis of Eq. (5.28), we can also set the expression in the brackets in Eq. (5.29) to zero and obtain a Jeans’ scale for each species:

$$k_{J,\alpha} \equiv \sqrt{\frac{4\pi G \bar{\rho} a^2 \sum_{\beta} \epsilon_{\beta} \delta_{\beta,k}}{\sigma_{\alpha}^2 \delta_{\alpha,k}}} = k_{\alpha} \sqrt{\epsilon_{\alpha} + \sum_{\beta \neq \alpha} \epsilon_{\beta} \frac{\delta_{\beta,k}}{\delta_{\alpha,k}}}, \quad (5.30)$$

where $k_{\alpha} \equiv \sqrt{4\pi G \bar{\rho} a^2 / \sigma_{\alpha}^2}$ is the would-be Jeans scale in a single-species scenario. Clearly, the evolution of the Jeans’ scale for any individual species depends on an intricate interplay between all the species through gravity. If the species in consideration dominates the energy density as well as the perturbation to it, we see $k_{J,\alpha} \approx k_{\alpha}$. However, if the species is not dominating, but have $\delta_{\alpha,k} \gg \delta_{\beta,k}$, then $k_{J,\alpha} < k_{\alpha}$ — the perturbation in this species is dampened by the other species. On the contrary, if $\delta_{\alpha,k} \ll \delta_{\beta,k}$, $k_{J,\alpha} > k_{\alpha}$ — the perturbations in other species are driving the perturbation in the species α . Thus, we see, the analysis relies highly on the knowledge of each $\delta_{\alpha,k}$, which is not clear if there is no numerical results, and could be highly non-trivial, for example, when $\delta_{\alpha,k}$ is close to zero or even becomes negative. It is also not clear what the velocity dispersion is for a single momentum slice. Besides,

the entire discussion above assumes that dark matter dominates the universe. In reality, other species such as photons and baryons should also be included in Eq. (5.29), which further complicates the analysis. Therefore, it is not advantageous to study the Jeans' scales for scenarios involving non-trivial phase-space distributions.

5.1.2.1 Free-Streaming Horizon

It is therefore very important to find a quantity that can be attributed to each packet or even each momentum slice in the phase-space distribution and at the same time do not have non-trivial time evolution. Indeed, there exists a quantity that has these advantages, *i.e.*, the *free-streaming horizon* [141]:

$$d_{\text{FSH}}(v_{\text{now}}) \equiv a_{\text{now}} \int_{t_{\text{P}}}^{t_{\text{now}}} v(t) \frac{dt}{a} = a_{\text{now}} \int_{a_{\text{P}}}^{a_{\text{now}}} \frac{v}{aH} \frac{da}{a} . \quad (5.31)$$

The expression above gives the maximum distance that a particle with speed v can traverse from the time t_{P} when it was produced to the present time t_{now} , therefore it is straightforward to map it uniquely into the phase space. Since the horizon is an integrated quantity with a positive integrand, it grows in time monotonically. The horizon grows differently depending on the type of energy which dominates the epoch, so the integral is split up as

$$\begin{aligned} \frac{d_{\text{FSH}}}{a_{\text{now}}} &= \overbrace{\int_{a_{\text{P}}}^{a_{\text{eq}}} \frac{v}{a^2 H} da}^{\text{radiation epoch}} + \overbrace{\int_{a_{\text{eq}}}^{a_{\text{now}}} \frac{v}{a^2 H} da}^{\text{matter epoch}} \\ &\approx \int_{a_{\text{P}}}^{a_{\text{eq}}} \frac{v da}{a_{\text{eq}}^2 H_{\text{eq}}} + \int_{a_{\text{eq}}}^{a_{\text{now}}} \sqrt{\frac{a_{\text{eq}}}{a}} \frac{v da}{a_{\text{eq}}^2 H_{\text{eq}}} , \end{aligned} \quad (5.32)$$

in which one can set the present-day scale factor $a_{\text{now}} = 1$, and hence the scale factor and the Hubble parameter at the MRE take the values $a_{\text{eq}} \approx 2.84 \times 10^{-4}$, $H_{\text{eq}} \approx H_0 (a_{\text{now}}/a_{\text{eq}})^{\frac{3}{2}} \approx 69.70 h/\text{Mpc}$. Additionally, we have used that $H \propto 1/a^2$ during radiation domination and $H \propto 1/a^{3/2}$ during matter domination. Assuming that dark-matter particles become non-relativistic before matter-radiation equality, the

horizon is then approximately

$$\frac{d_{\text{FSH}}}{a_{\text{now}}} \approx \int_{a_{\text{P}}}^{a_{\text{eq}}} \frac{p_{\text{now}}}{\sqrt{p_{\text{now}}^2 + (ma)^2}} \frac{da}{a_{\text{eq}}^2 H_{\text{eq}}} + \int_{a_{\text{eq}}}^{a_{\text{now}}} \frac{(p_{\text{now}}/m)da}{(aa_{\text{eq}})^{\frac{3}{2}} H_{\text{eq}}}, \quad (5.33)$$

where p_{now} is the present-day momentum. Performing the integrals, and using $a_{\text{now}} \gg a_{\text{eq}}$, we find our expression for the present-day free-streaming horizon:

$$d_{\text{FSH}}/a_{\text{now}} \approx \frac{p_{\text{now}}/m}{a_{\text{eq}}^2 H_{\text{eq}}} \left[2 + \operatorname{arctanh} \frac{m}{\sqrt{m^2 + p_{\text{eq}}^2}} - \operatorname{arctanh} \frac{m}{\sqrt{m^2 + p_{\text{P}}^2}} \right], \quad (5.34)$$

in which we have used $p_{\text{now}}/a_{\text{eq}} = p_{\text{eq}}$ and $p_{\text{now}}/a_{\text{P}} = p_{\text{P}}$ which are the momenta at matter-radiation equality and the production time. The *free-streaming wavenumber* k_{FSH} associated with the free-streaming horizon is defined as [143]

$$k_{\text{FSH}}(v_{\text{now}}) \equiv \frac{a_{\text{now}}}{d_{\text{FSH}}(v_{\text{now}})}. \quad (5.35)$$

With this, one can calculate the k_{FSH} for each momentum slice in the phase-space, which gives the largest scale that particles with certain momentum are able to reach. We shall see later that, at a particular scale k , particles with present-day velocity $v_{\text{now}} \equiv p_{\text{now}}/E_{\text{now}}$ are able to suppress the corresponding perturbation mode if $k_{\text{FSH}} \lesssim k$ (or $d_{\text{FSH}}/a_{\text{now}} \gtrsim 1/k$), otherwise they should leave no impact.

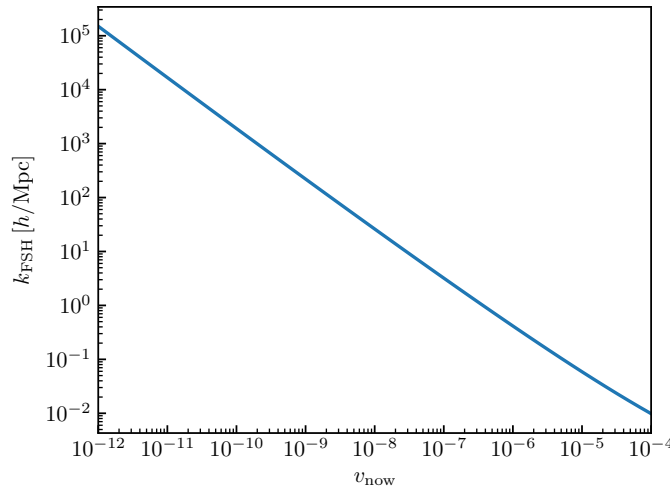


Figure 5.5: Present-day velocity and the associated free-streaming wavenumber k_{FSH} .

To see why the free-streaming wavenumber captures the physics, consider a plane wave with wavenumber k whose physical wavelength at time t is $\lambda(t) \equiv 2\pi a(t)/k$. On this scale, the overdense region and underdense regions are centered around the maxima and minima of the wave which are separated by a spatial distance $\lambda/2 = \pi a/k$. To be able to suppress the perturbation, particles in overdense regions must be able to travel to the underdense regions. This gives a physical length scale of the order a/k . Therefore, particles that are able to leave a trace at a comoving scale k today have to have $d_{\text{FSH}} = a_{\text{now}}/k_{\text{FSH}} \gtrsim a_{\text{now}}/k$.

A map between k_{FSH} and the present-day velocity v_{now} is shown in Fig. 5.5. We shall make use of the free-streaming wavenumbers in the subsequent discussion on matter power spectrum.

5.1.2.2 Matter Power Spectrum

The growth of structure is eventually reflected in the *matter power spectrum*. The matter power spectrum can be defined through the average of the density perturbation:

$$\langle \delta(x)^2 \rangle = \frac{1}{V} \frac{1}{(2\pi)^3} \int d^3k \mathcal{P}(k) , \quad (5.36)$$

in which $\langle \dots \rangle$ indicates the average over a space with physical volume V , which is usually taken to be our entire universe. After some algebra, it is easy to show that

$$\mathcal{P}(k) \equiv |\delta_k^2| . \quad (5.37)$$

We shall demonstrate the connection of the phase-space distribution today $p_{\text{now}}^3 f(p_{\text{now}})$ to the matter power spectrum $\mathcal{P}(k)$.³ In order to do this, let us construct some simple artificial phase-space distributions that are easy to identify. This will have the advantage of not only controlling the spectrum of k_{FSH} , but also getting an intuitive grasp for how the $p_{\text{now}}^3 f(p_{\text{now}}) \longrightarrow \mathcal{P}(k)$ mapping works, beyond critical wavenumbers.

Let us first define a form for the phase-space distributions. In non-minimal sce-

³To simplify the notation, we make use of the convention $a_{\text{now}} = 1$.

narios, it is common to have a large number of decay channels with wildly different decay widths. Therefore, the injection of particles could occur at time scales that are widely separated. Since particles redshift, the packets generated at logarithmically different time scales are thus also naturally separated logarithmically in the momentum space. Based on these considerations, for our artificial distribution functions, we choose the form such that quantity $p_{\text{now}}^3 f(p_{\text{now}})$ is constructed by a combination of Gaussians on a logarithmic scale. We define this as a log-normal distribution:

$$p_{\text{now}}^3 f(p_{\text{now}}) \equiv \frac{A}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \left[\log \left(\frac{p_{\text{now}}}{\langle p_{\text{now}} \rangle} \right) + \frac{1}{2}\sigma^2 \right]^2 \right\}, \quad (5.38)$$

for which σ is the width of the distribution. Note that, even on log-scale, the central value of such a distribution is not located at $\log p_{\text{now}} = \log \langle p_{\text{now}} \rangle$. Instead, the distribution is symmetric about $\log \langle p_{\text{now}} \rangle - \sigma^2/2$. Besides, the distribution function is normalized such that $\int p_{\text{now}}^2 f(p_{\text{now}}) dp_{\text{now}} = A$. Therefore, we can easily vary the width and average momentum of the distribution while retaining a fixed dark-matter abundance.

There exist several public numerical packages for computing the matter power spectrum. Since it is our purpose to study matter power spectrum from various non-trivial phase-space distribution, we choose to use the famous **CLASS**⁴ code, which allows us to input phase-space distribution of our dark-matter candidate.

In the following, we shall use the artificial distribution functions to learn about the relation between phase-space distribution and matter power spectrum. For simplicity, the mass of the dark-matter particle is set to $m = 10$ keV which is a typical mass scale for warm dark matter.

A Single Packet

To understand the cosmological consequences of having a non-trivial phase-space distribution, let us start by looking at the effects from a single packet. We shall let that the packet carries a certain fraction of the total dark-matter abundance,

⁴The ‘‘Cosmic Linear Anisotropy Solving System’’, see [144–147]

while the rest of dark matter are perfectly cold with zero velocity. The perfectly cold dark-matter component will be referred to as the CDM component.

In the upper left panel of Fig. 5.6, we show several cases in which we vary the height of each packet while keeping the width fixed. The fraction of the dark-matter abundance carried by each packet, which is proportional to the area under the packet, is therefore also varied. In the upper right panel, we plot the ratio $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$, in which $\mathcal{P}(k)$ is the matter power spectrum obtained from each phase-space distribution in the left panel, and $\mathcal{P}_{\text{CDM}}(k)$ is the standard case which is obtained by assuming the entirety of dark matter is perfectly cold.⁵ The effects of dark matter being not perfectly cold, can be seen from the deviation of the ratio from 1. In the same panel, we also plot the phase-space distribution $p_{\text{now}}^3 f(p_{\text{now}})$ against the free-streaming wavenumber k_{FSH} using Eq. (5.35). The benefit is that, at any k , the amount of particles having $k_{\text{FSH}} = k$ can be easily read from the plot.

We first notice in the right panel that the locations of the peaks on the k -axis, which correspond to the free-streaming wavenumber of the average velocity, accurately predict the point at which the matter power spectra start to deviate from the standard case. Indeed, in these cases that we show, the phase-space distributions are sufficiently sharp such that the average velocity has nicely captured the features in the distribution functions.

Beyond this, the amount of drop is related to the abundance carried by the packet. Having more abundance in the packet means more particles can free-streaming out of the overdense regions and less particles are aggregating to build up the gravitational potential. Therefore, a larger area under the packet leads to a stronger suppression. Interestingly, for all the cases, we see some wiggles in the matter power spectrum and this effect seems to become more distinct as we increase the abundance carried by the packet. This is in fact due to the acoustic oscillation of the packet component which we will explain later.

Besides the area under the packet, we are also interested in how the shape of the packet affects matter power spectrum. In the lower panels of Fig. 5.6, we fix

⁵This ratio is often referred to as the *squared transfer function*.

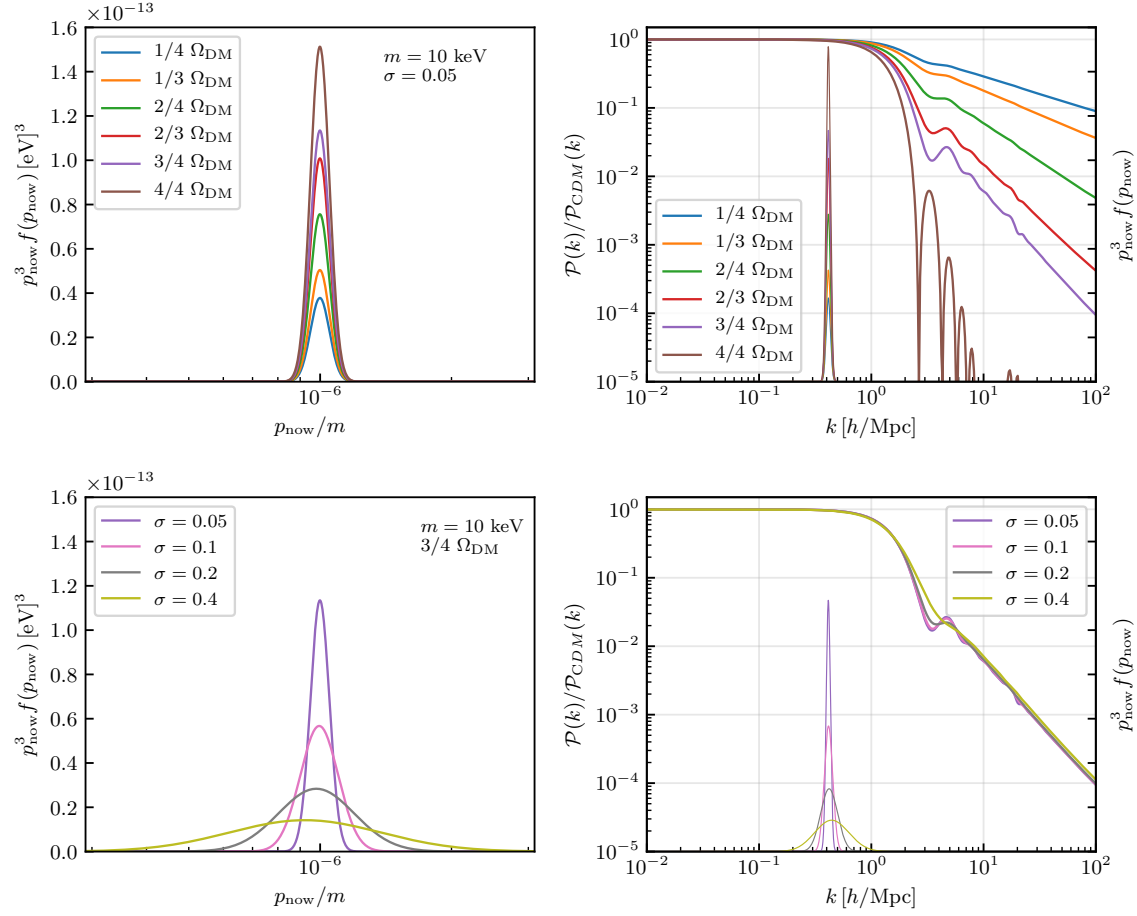


FIGURE 5.6: Different ways of varying a single Gaussian packet with the same average velocity fixed at $\langle v_{\text{now}} \rangle = 10^{-6}$. **Upper panels:** The width $\sigma = 0.05$ is fixed, but the height of the packet varies, which changes the dark matter abundance carried by each packet. The total dark-matter abundance is fixed by adding a complementary CDM component, *i.e.* $\Omega_{\text{packet}} + \Omega_{\text{CDM}} = \Omega_{\text{DM}} = 0.26$. The ratio between the matter power spectrum associated with each case and the standard CDM matter power spectrum is shown in the right panel. We also plot in the same panel the phase-space distribution against k_{FSH} using Eq. (5.35). **Lower panels:** The area of each packet is fixed, so that the abundance carried by each packet is set to $3\Omega_{\text{DM}}/4$. The distributions differ in their widths and heights.

the abundance of the packet to be $3/4$ of the total DM abundance, and vary the width of the packet. In the lower right panel, once again, the location of each peak still matches well with the point where the ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ start to deviate from 1. Beyond this point, the general trend is very similar in each case. However, small differences do exist. For wider distributions, since they have more particles at larger velocities, the deviation actually starts at slightly smaller k . At the same time, since wider distribution functions also means more particles at lower velocities, their

power is slightly larger at larger k . Besides, just like before, wiggles show up around $k \sim 2 - 6 \ h/\text{Mpc}$. However, this effect is weaker as the width gets larger, and the power at larger k seems to be insensitive to the variation in width.

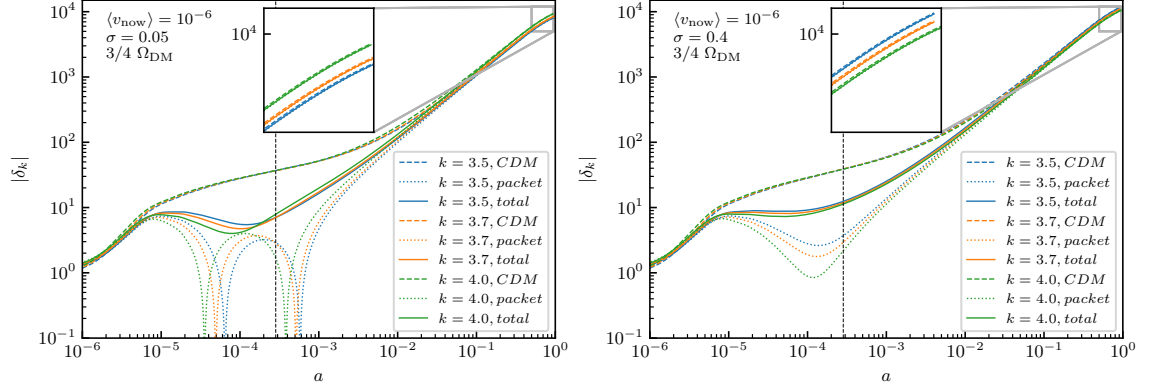


FIGURE 5.7: Evolution of perturbation modes $k = 3.5, 3.7$ and $4.0 \ h/\text{Mpc}$. For each mode, the dashed and the dotted curves stand for the perturbation in the CDM component and the packet component, respectively. The total dark-matter perturbation is represented by the solid curves. The left panel shows the case with $\sigma = 0.05$, while the right panel is associated with $\sigma = 0.4$.

To understand the emergence of the wiggles, we pick out several perturbation modes $k = 3.5, 3.7$ and $4.0 \ h/\text{Mpc}$ in the two cases with $\sigma = 0.05$ and 0.4 , and look at their evolution in time. In Fig. 5.7, we separate the total dark-matter perturbation into the perturbation of the packet component and the CDM component.⁶ Naively, in this regime, one would expect the present-day perturbation amplitude to decrease monotonically as k increases. However, while this is true in the right panel, it is not true in the left panel as we see $\delta_{k=4.0} > \delta_{k=3.7} > \delta_{k=3.5}$ in the small window. Looking back in time, we find that this is because these two cases have different initial conditions at the beginning of MD epoch which is labelled by the vertical dashed lines in the panels. In the left panel, the modes of the packet component have experienced suppression and have suffered from acoustic oscillation in the RD epoch after they enter horizon. During the oscillation, these modes actually have changed sign once which means overdense regions can change into underdense regions and vice versa. Since the CDM component dominates gravity and its perturbation modes

⁶The total dark-matter perturbation is a weighted sum over different components: $\delta_{\text{DM}} \equiv \sum_j \epsilon_j \delta_j$, where $\epsilon_j \equiv \bar{\rho}_j / \bar{\rho}_{\text{DM}}$.

never change sign, the packet component have to flip the sign again before its final growing phase in the MD epoch. It turns out that, in this small regime, the larger modes of the packet component which experience suppression earlier also have larger frequencies such that they cross zero and start to grow at an earlier time in the MD epoch. This gives the larger modes an opportunity to have a larger amplitude than the shorter modes today which explains the emergence of wiggles.

The situation is a bit different in the right panel. The packet component has also experienced suppression. However, the suppression is not strong enough to flip the sign of the perturbation modes. In other words, the acoustic oscillation never has the chance to complete one full cycle. Therefore, larger modes cannot have the advantage to grow first in the MD epoch.

The observation that a sharper packet tends to cause a stronger suppression than a wider packet does lead us to an very intuitive explanation. In this small regime of k , almost all the particles in the sharper packet have their free-streaming horizon larger than $1/k$ in the RD epoch, whereas the particles from the lower-momentum part of the wider packet are not able to free-stream a distance large enough to cause suppression. These slower particles therefore protect the gravitational potential and prevent the packet component from being significantly suppressed. This picture is also consistent with the behavior at both smaller and larger k values. For smaller k , particles in either packet are not fast enough to reach these scales, whereas for larger k , particles in both packets are able to free-stream through them.

Finally, let us summarize the points that we have learnt from the single-packet case:

- The largest length scale (smallest k) that a packet suppress agrees with the free-streaming wavenumber k_{FSH} associated with the packet.
- The “strength” of the suppression depends on the area (the abundance) of the packet. A larger area corresponds to a stronger suppression.
- The width of the packet also matters. A sharper packet tends to create wiggles in the matter power spectrum while a wider packet gives rise to a more smooth

spectrum. The wiggles are more distinct when the area of the packet is larger.

Double Packets

Beyond a single packet, the next more complicated cases are distribution functions with two packets. We shall let the two packets carry the entire dark-matter abundance together, but at the same time, vary the fraction taken by each individual packet. We show these cases in the left panel of Fig. 5.8. We shall call the packet at larger momentum the warm packet, and the one at lower momentum the cold packet. The fraction of dark-matter abundance carried the cold and the warm packets are $\{5\%, 95\%\}$, $\{50\%, 50\%\}$ and $\{95\%, 5\%\}$, respectively. For each distribution function, both packets have the same width. The solid curves stands for sharp distributions with $\sigma = 0.05$, while the dashed curves represent wide distributions with $\sigma = 0.7$.

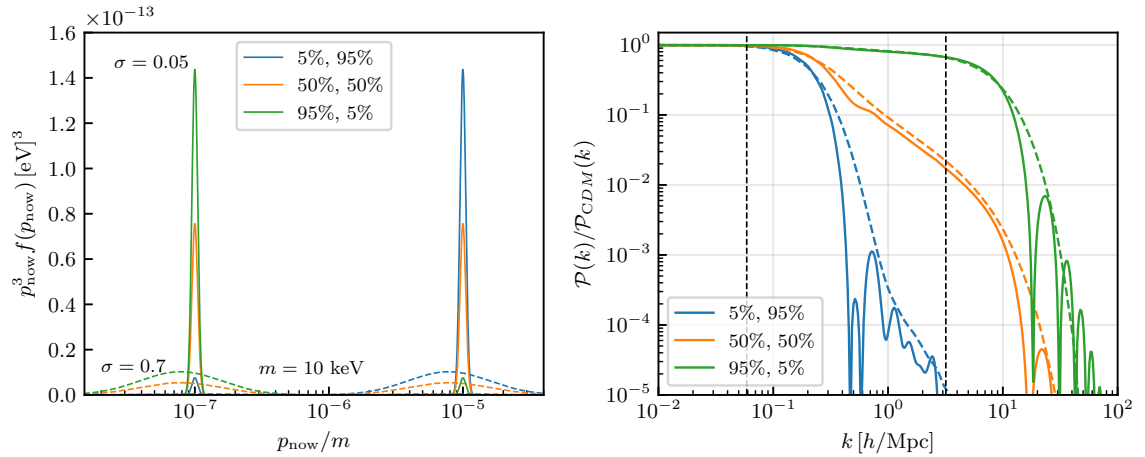


FIGURE 5.8: Various abundance arrangements for the double-packet distribution. All the distributions in the left panel have two peaks with the same width whose average are $\langle v_{\text{now}} \rangle = 10^{-5}$ and $\langle v_{\text{now}} \rangle = 10^{-7}$, respectively. There three types of partitioning of the abundance between the two packets — $\{5\%, 95\%\}$, $\{50\%, 50\%\}$ and $\{95\%, 5\%\}$, which are represented by different colors. The width of the distribution is $\sigma = 0.05$ for the sharp packets (solid curves), and $\sigma = 0.7$ for the wide packets (dashed curves). The matter power spectrum associated with each distribution is shown in the right panel. The vertical dashed lines indicate the values of k_{FSH} which correspond to the average velocity of each packet.

In the right panel, we see from the ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ that the first vertical line which corresponds to the free-streaming wavenumber associated with the warm packet nicely

predicts the beginning point of deviation. The size of the first packet determines the amount of initial drop after passing by the first free-streaming wavenumber, which is consistent with what we have learnt from Fig. 5.6. The second drop in $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ happens after passing by the second vertical line which indicates the free-streaming wavenumber associated with the average velocity of the cold packet. Likewise, the second drop is bigger when the cold packet is larger. In the blue cases, since the warm packet takes a large portion of abundance, the ratios take a big drop right after passing the first vertical line, which makes the second drop difficult to see. However, in the orange and green cases, the two-step drop is quite distinct as we see the ratios slowly decrease between the two vertical lines and only drop significantly after passing the second vertical line. Moreover, just as in Fig. 5.6, when the warm packet is very sharp and carries a sizable abundance, wiggles caused by acoustic oscillations could show up. In the blue and orange cases (solid curves), this happens even before reaching the second vertical line. Such effect is completely smoothed out when packets are wide. This is because the tails of the packets are able to extend to very small momentum and protect the gravitational potential. Despite the wiggles, we also notice that, before reaching the second vertical line, the general trend in both the case with wide packets and the case with sharp packets is quite similar. The curves are almost overlapping when the warm packet is small, and their difference gets larger as the area of the warm packet increases. Even in the blue case, the solid curve starts to follow the dashed curve before they approach the second vertical line. Such behavior is consistent with our physical picture — after passing the first vertical line, the contribution to the gravitational potential from the warm packet gets smaller and smaller, while the amount of cold particles that protect the gravitational potential becomes similar in both cases.

To understand the role played by each packet and the interplay between them in more detail, we once again look at the evolution of the perturbations of different components in the phase-space distribution at several scales. In Fig. 5.9, the orange case is the same with the solid orange curve in Fig. 5.8 which has two identical packets with $\langle v_{\text{now}} \rangle = 10^{-5}$ and $\langle v_{\text{now}} \rangle = 10^{-7}$. We compare it with the blue case which

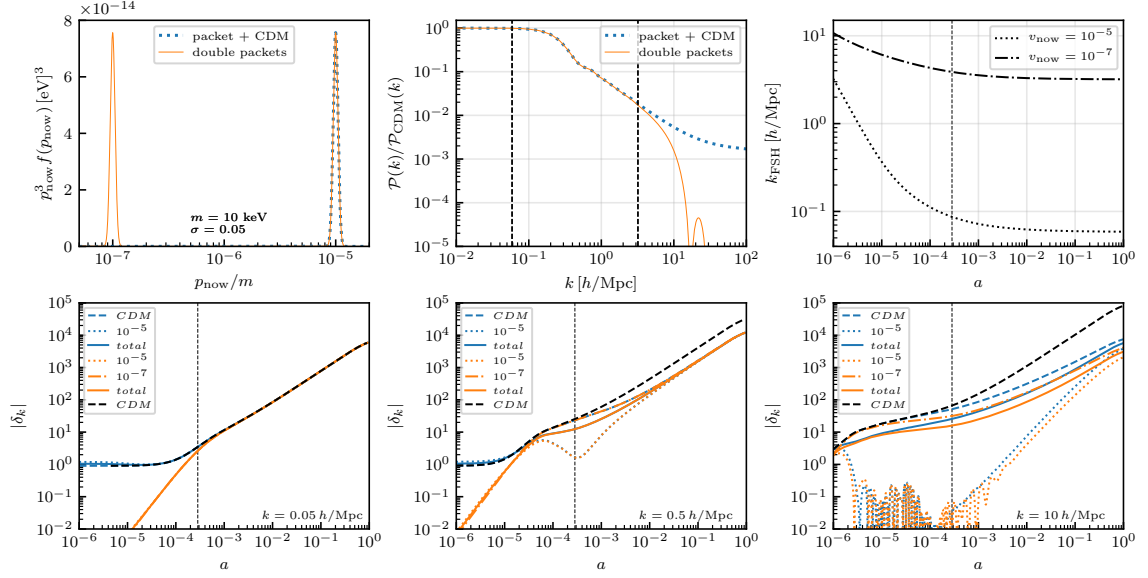


FIGURE 5.9: ‘Single packet + CDM’ vs ‘double packets’. In the top left panel, the blue curve has one sharp packet with $\langle v_{\text{now}} \rangle = 10^{-5}$ which carries half of the dark-matter abundance while the other half is in a CDM component. The orange curve has the same packet at $v_{\text{now}} = 10^{-5}$ while another packet which has the same shape has $\langle v_{\text{now}} \rangle = 10^{-7}$ and carries the other half of the dark-matter abundance. The ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ is shown in the top middle panel where the two vertical lines indicates the free-streaming wavenumber k_{FSH} associated with the two characteristic velocities. The top right panel shows the free-streaming wavenumbers that particles with $v_{\text{now}} = 10^{-7}$ or $v_{\text{now}} = 10^{-5}$ can reach at different epochs in the history of the universe. The dotted vertical lines in this panel and the lower panels marks the location of a_{eq} . The lower panels show the evolution of $|\delta_k|$ at 3 different scales, $k = 0.05, 0.5$, and 10 h/Mpc . The dotted curve stands for the warm packet, the dash-dotted curve represents the cold packet, and the dashed curve represents the CDM component. The evolution in the standard CDM case is plotted with the dashed black curve.

has exactly the same warm packet with $\langle v_{\text{now}} \rangle = 10^{-5}$, but with the cold packet replaced by a CDM component with the same abundance. The power-spectrum ratio in the top middle panel exhibit very interesting behavior. For both cases, the first drop takes place at exactly the same place, which is precisely predicted by the free-streaming wavenumber associated with the warm packet (the left vertical line) which we refer to as $k_{\text{FSH}}^{\text{warm}}$. After that, the power at larger k is exactly the same in those two different cases until the free-streaming wavenumber associated with the cold packet (we shall refer to it as $k_{\text{FSH}}^{\text{cold}}$) is reached. The two cases only start to deviate after passing by $k_{\text{FSH}}^{\text{cold}}$.

Such behavior can be explained by looking at the evolution of density perturbations of different components (*i.e.* the packet components and the CDM component) as well as the entire dark matter. In the lower panels, we show the evolution at three different scales.

For $k = 0.05 \ h/\text{Mpc}$, all the perturbations have almost exactly the same growth with the standard CDM case in the MD epoch. Indeed, the wavenumber is smaller than $k_{\text{FSH}}^{\text{warm}}$. And we also see from the top right panel that particles with $v_{\text{now}} = 10^{-5}$ can never free-stream to this scale. Therefore, at this scale, both the packet at $v_{\text{now}} = 10^{-7}$ and the packet at $v_{\text{now}} = 10^{-5}$ behave like a CDM component, and the power is not suppressed as compared with the standard CDM case (black dashed curve).

The bottom middle panel shows the mode $k = 0.5 \ h/\text{Mpc}$ which is just between $k_{\text{FSH}}^{\text{warm}}$ and $k_{\text{FSH}}^{\text{cold}}$. We notice that, despite having a slower growth compared with the standard CDM case, the total perturbations in both the orange and the blue case are exactly the same. This is not surprising as we see the perturbations of the warm packets in both cases receive the same amount of suppression, and at the same time, the perturbation of the cold packet in the orange case has the same growth as the CDM component in the blue case. Indeed, as one can check again in the top right panel, while particles from the warm packet are able to free-stream to this scale within the RD epoch, particles from the cold packet can never travel such a distance.

In the bottom right panel, since $k = 10 \ h/\text{Mpc} > k_{\text{FSH}}^{\text{cold}}$, the total perturbation in the blue and the orange cases start to bifurcate. The perturbations of the warm packets in both cases are damped completely and start acoustic oscillation during the RD epoch. The difference comes from the cold packet in the orange case and the CDM component in the blue case. While the perturbation of CDM component is only suppressed by free-streaming effects from the warm packet, the perturbation of the cold packet suffers additionally from the free-streaming of itself. This makes total perturbation in the orange case smaller and explains the bifurcation of the orange and blue curves beyond $k \sim 3$ in the top middle panel.

Both Fig. 5.8 and Fig. 5.9 tell us that, when two packets are present, the matter

power spectrum can be separated into three different regimes by the free-streaming wavenumbers associated with each peak: 1) $k \lesssim k_{\text{FSH}}^{\text{warm}}$, 2) $k_{\text{FSH}}^{\text{warm}} \lesssim k \lesssim k_{\text{FSH}}^{\text{cold}}$, 3) $k \gtrsim k_{\text{FSH}}^{\text{cold}}$. The first regime corresponds to length scales beyond the reach of all the components. Therefore all the components behave like cold dark matter, and the total perturbation is not suppressed. In the second regime, the perturbation is damped due to the free-streaming of the particles associated with the warm packet while the free-streaming effects from the cold packet can not affect this scale. This gives rise to a initial deviation from the standard CDM case, and the amount of deviation depends on the size and shape of the warm packet. In the third regime, the perturbation is suppressed by the free-streaming of the particles from both packets which leads to a further drop beyond $k_{\text{FSH}}^{\text{cold}}$.

We thus conclude the underlying physics that we have learnt from this set of examples here:

- The suppression of power at a particular scale k is determined by the part of phase-space distribution that has particles with free-streaming horizon $d_{\text{FSH}} \gtrsim 1/k$ (or $k_{\text{FSH}} \lesssim k$).
- The part of phase-space distribution that corresponds to particles whose $k_{\text{FSH}} \gtrsim k$ has negligible effect on scales larger than $1/k$. On these scales those particles behave exactly like perfectly cold dark matter. Therefore, the shape of the distribution of this “cold part” does not matter.
- The matter power spectrum is fully suppressed only at scales where a large fraction of dark-matter particles are able to free-stream through.

Multiple Packets

So far we have only looked at the effects from a single packet or widely separated isolated packets. However, in some scenarios, the distribution generated from decays can be more like a continuum. To gradually develop our understanding from the discrete limit to the continuum limit, in Fig. 5.10, we start with the double packet scenario that we have studied before, in which two identical packets are located at

$v_{\text{now}} = 10^{-7}$ and $v_{\text{now}} = 10^{-5}$, and each carries half of the entire dark-matter abundance. We gradually “flatten” this distribution by adding packets with the same widths in between. The new distribution functions after adding more packets have multiple identical packets evenly distributed on a logarithmic scale, while the total area, *i.e.* the total dark-matter abundance is kept fixed. The examples of having $n_{\text{pac}} = 2, 3, 4, 5$ packets are shown in the left panel, and eventually we show the case with $n_{\text{pac}} = 100$ which is a very good approximation of the $n_{\text{pac}} \rightarrow \infty$ continuum limit.

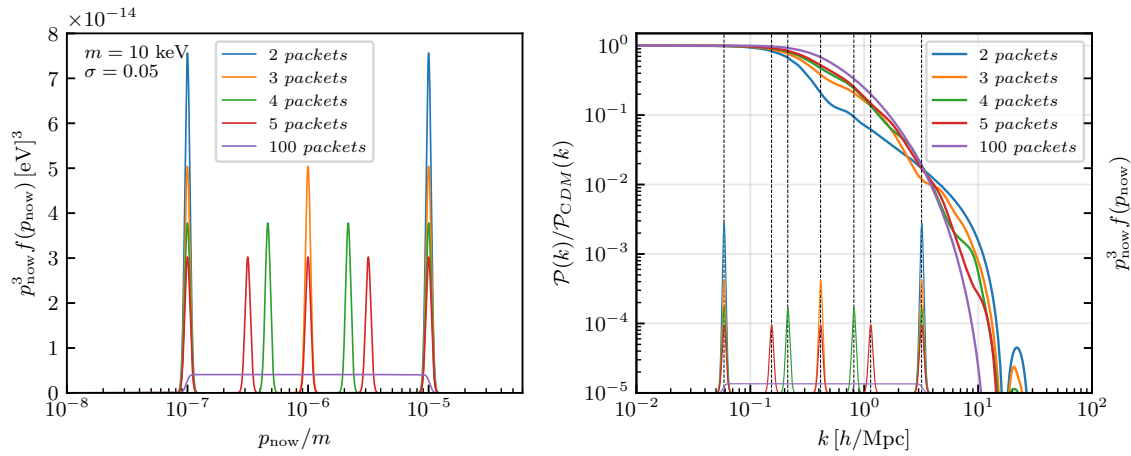


FIGURE 5.10: Picket Fence vs Wall: In the left panel, narrow packets with $\sigma = 0.05$ are distributed evenly on logarithmic scale between $v_{\text{now}} = 10^{-7}$ and $v_{\text{now}} = 10^{-5}$. The height of the packets are adjusted so that the entire dark-matter abundance is evenly distributed among the packets. The ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ associated with each case is shown on the right panel, where the dashed vertical lines mark k_{FSH} associated with the peaks in the $n_{\text{pac}} = 2, 3, 4, 5$ cases. The phase-space distributions are also plotted against k in the right panel. Their heights are lowered to avoid intersecting the matter-power-spectrum curves.

The ratio $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$ associated with each case is shown in the right panel. Just like before, all the cases start to deviate from the standard CDM case after reaching the k_{FSH} associated with the warmest packet. However, the amount of initial drop is decreased after adding more packets, since the height of the warmest packet is reduced. Therefore, we see the order of the curves right after the initial deviation goes like $n_{\text{pac}} = 100, 5, 4, 3, 2$ from the highest to the lowest. As we pass by the subsequent k_{FSH} ’s associated with the intermediate packets, which are marked out by the vertical lines, we see this flattening tend to reshuffle the order of the curves.

For example, we see the red curve is approaching and eventually crossing the green curve, and the green curve is doing the same to the orange curve. Comparing with the other curves, the purple curve, which is associated with the continuum limit, starts with the largest power, smoothly turns down, and eventually ends with the smallest power. This is due to the fact that the fraction of the cold part ($k_{\text{FSH}} \gtrsim k$) and the warm part ($k_{\text{FSH}} \lesssim k$) of the purple distribution successively changes as k increases — it starts with the largest fraction of cold particles at small k , but end with the largest fraction of warm particles at large k . In the end, shortly after passing by the last vertical line, the order of the curves are completely inverted. The blue curve is on top because its phase-space packet at $v_{\text{now}} = 10^{-7}$ is the largest, while the purple curve is at the bottom since it has the least cold particles. Therefore, we see that ‘flattening’ a distribution tend to change the amount of suppression in different regimes. It makes the decrease of power as k increases more gradual and successive. Besides, the wiggles created by acoustic oscillations which can be seen when the distribution function consists of a small number of isolated sharp packets become less distinct as we insert more packets. They are eventually smoothed out as we approach the continuum limit.

All the above observations tell us an important message: *As we track along the matter power spectrum $\mathcal{P}(k)$ (or the ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$) in the direction of increasing k , we are also effectively scanning over the phase-space distribution from larger momentum to smaller momentum.* This motivate us to define a quantity which scans the fraction of dark matter that can be effectively treated as a CDM component at a particular scale k :

$$F(k) \equiv \frac{\int_0^{p_{\text{now}}(k)} p_{\text{now}}^2 f(p_{\text{now}}) dp_{\text{now}}}{\int_0^\infty p_{\text{now}}^2 f(p_{\text{now}}) dp_{\text{now}}} , \quad (5.39)$$

where $p_{\text{now}}(k)$ is the momentum that maps to the free-streaming wavenumber k . One can obtain it by inverting Eq. (5.35). We shall make use of this fraction function to analyze the results in the next section.

To conclude this subsection, we emphasize that there is an infinite number of way to modify the shape of a phase-space distribution. We have shown cases in which

we vary the height and width of a single packet, as well as cases where there is more than one packet. Although there is no way to be all-inclusive, the fact that the free-streaming wavenumber k_{FSH} can be attributed to each individual momentum slice allows us to use these fundamental examples to at least qualitatively study the matter power spectrum from any kind of phase-space distribution. We shall demonstrate this in the next section in which we are going to study the matter power spectra obtained from “real” distributions shaped by intra-ensemble decays.

5.2 Toy Model

5.2.1 Parametrization of the Ensemble

Thus far, we have only discussed how phase-space distribution can be affected by decays within a multi-component dark sector, as well as the relation between phase-space distribution and the suppression pattern on matter power spectrum. We have not yet specified anything about the dark sector itself. In general, we can imagine an dark ensemble consisting of all kinds of particles with different spin, masses and couplings, just like the Standard Model. However, the main point of this chapter is to study the dynamics of intra-ensemble decays in terms of the flow, redistribution and redshift of energy densities across the ensemble, and the cosmological implications of a non-trivial relic phase-space distribution. Studying a system that is too complicated would only obscure our understanding of physics, not to say it might also be extremely computationally expensive.

As a toy model, we assume the dark sector consists of an ensemble of $N + 1$ real scalar fields ϕ_j neutral to the Standard Model gauge interactions, where the index $j = 0, 1, \dots, N$ labels the constituents in order of increasing mass m_j . The number of constituents in the ensemble depends on the mathematical structure of the underlying theory — it can be a large finite integer, like those from a large gauge group [57], or even infinity, such as those from extra-dimension theories [40, 43] and strongly-coupled dark sectors [13]. Here, instead of taking a top-down perspective, we shall simply take some reasonable finite integer N without looking at the fundamental

theory. For concreteness, the m_j is assumed to scale across the ensemble according to the relation

$$m_j = m_0 + j^\delta \Delta m , \quad (5.40)$$

where the mass m_0 of the lightest ensemble constituent, the scaling exponent δ , and the mass-splitting parameter Δm are taken to be free parameters of the model. We shall also assume that these fields couple to each other through a trilinear Lagrangian coupling which leads to decays of the form $\ell \rightarrow i + j$. For simplicity, we assume that the partial widths of all other decay channels are negligible. Thus, the effective Lagrangian relevant for the period of intra-ensemble decays can be taken to be

$$\mathcal{L} = \sum_{\ell=0}^N \left(\frac{1}{2} \partial_\mu \phi_\ell \partial^\mu \phi_\ell - \frac{1}{2} m_\ell^2 \phi_\ell^2 - \sum_{i=0}^{\ell} \sum_{j=0}^i c_{\ell ij} \phi_\ell \phi_i \phi_j \right) + \dots . \quad (5.41)$$

In the above expression, “...” includes all the other terms, such as quardrilinear terms and other higher-order terms which help stabilize the vacuum. Those terms might be important at earlier times if the ensemble constituents were in equilibrium with each other. However, since we assume intra-ensemble decays occur while all the constituents have already decoupled, those terms are not relevant during the period of intra-ensemble decays. The coupling coefficients $c_{\ell ij}$ (with dimensions of mass) follow the scaling relation

$$c_{\ell ij} = c_0 \mu R_{\ell ij} \left(\frac{m_\ell - m_i - m_j}{\Delta m} \right)^r \left(1 + \frac{m_i - m_j}{\Delta m} \right)^{-s} . \quad (5.42)$$

Here, μ is the energy scale corresponding to the Lagrangian, and c_0 , r , and s are taken to be dimensionless free parameters, and

$$R_{\ell ij} \equiv \begin{cases} 6 & \text{all indices different} \\ 3 & \text{only two indices equal} \\ 1 & \text{all indices equal} . \end{cases} \quad (5.43)$$

The expressions in Eqs. (5.40) and (5.42) are sufficient to specify the total decay

widths of the ϕ_ℓ and branching fractions $\text{BR}(\ell \rightarrow i+j)$ for individual decay channels.

The two factors r and s in Eq. (5.42), can be explained as follows. The first factor r represents the dependence of the coupling on the fraction of the energy released in ϕ_ℓ decay that goes into the *kinetic* energies, rather than the mass energies, of the daughter particles ϕ_i and ϕ_j . Note that the sums in Eq. (5.41) are defined such that $m_\ell \geq m_i \geq m_j$. Thus, the greater the exponent r , the more decays are dominated by processes involving final states with proportionally smaller masses. Indeed, as the value of r is increased, decay processes in which the decay products behave more like “radiation” and less like “matter” are increasingly preferred. By contrast, the second factor s in some sense reflects the degree of “symmetry” between the daughter-particle masses. In particular, it depends on how symmetrically the total mass energy of the daughter particles is partitioned between them. Note that, the negative sign in front of s in Eq. (5.42) implies that, for $s > 0$, the decays of ϕ_ℓ are dominated by processes in which the daughter-particle masses are similar, while, for $s < 0$, these decays are dominated by those in which the daughter-particle masses are significantly different.

For a two-body decay process $\ell \rightarrow i+j$, the center-of-mass momentum of the decay products is

$$|\vec{p}| = \frac{\sqrt{[m_\ell^2 - (m_i + m_j)^2] [m_\ell^2 - (m_i - m_j)^2]}}{2m_\ell}. \quad (5.44)$$

The associated width is

$$\Gamma_{\ell \rightarrow i+j} = |\mathcal{M}_{\ell \rightarrow i+j}|^2 \frac{|\vec{p}|}{8\pi m_\ell^2} \Theta(m_\ell - m_i - m_j). \quad (5.45)$$

where $|\vec{p}|$ is given by Eq. (5.44), $|\mathcal{M}_{\ell \rightarrow i+j}|^2$ refers to the amplitude, and $\Theta(m_\ell - m_i - m_j)$ is a unit step-function that ensures the kinematic accessibility. For simplicity, we shall only show the results from a linear mass spectrum in the following.

Fig. 5.11 shows the total intra-ensemble decay width of each state with $j > 0$ for various combinations of r and s . The curves with the same r have the same color, whereas those with the same s have the same line style. We have chosen a linear

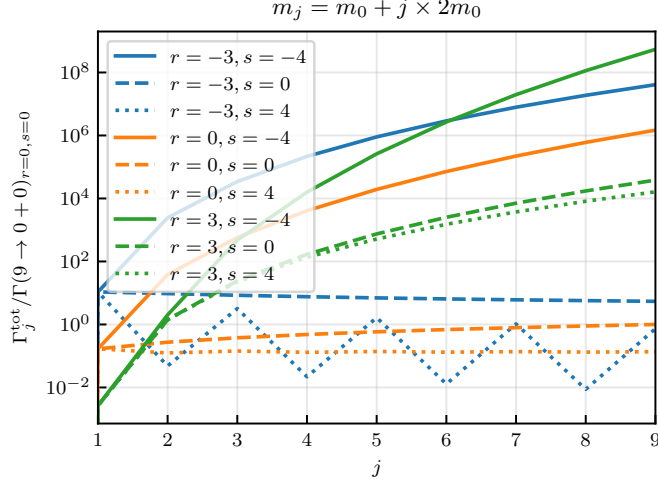


FIGURE 5.11: Total decay width at each tower level for various combinations of r and s . The decay widths are normalized by the width of the $9 \rightarrow 0 + 0$ channel in the $r = 0, s = 0$ case. The curves with the same r have the same color, while curves with the same s have the same line style.

mass spectrum, in which $\delta = 1$ and $\Delta m = 2m_0$. Note that, ϕ_1 always has the same decay width for the same r , since the only decay channel available is symmetric $1 \rightarrow 0 + 0$. This channel seems to decrease as r increases. However, this is not general. As we can see from Eq. (5.42), this is simply due to our choice of this mass spectrum such that $m_1 - 2m_0 = m_0 < \Delta m = 2m_0$.

For fixed r , the total decay width decreases as s increases. This is because the couplings in Eq. (5.42) are invariant under the change of s for the symmetric decay channels. Therefore, varying the value of s only changes the decay widths of the asymmetric channels. Consequently, as symmetry is more and more favored when s is increased, asymmetric channels close up, which results in a net decrease in the total decay width.

The dependence on r is more subtle. Eq. (5.42) and Eq. (5.45) tell us that the decay widths depend not only on the couplings but also the kinematics. Besides, varying r changes the branching ratios across the accessible decay channels — it opens certain decay channels while closing the others. For example, whereas the decay channels that have a larger mass difference between the parent state and the daughter states are enhanced when increasing r , those with a smaller mass difference

are weakened. Therefore, it is not straightforward to determine how the total decay widths when adjusting r .

Besides, not all the decay widths increases with the tower level j . For the curves with $s \geq 0$ and $r \leq 0$, the total width can even go down at higher levels. The blue dotted curve shows the extreme case when radiation is disfavored and symmetry strongly enforced. In that case, only the odd levels can produce a symmetric and kinematically accessible channel that minimize the radiation, therefore it exhibits oscillating behavior between the odd and the even levels.

Fig. 5.12 uses ϕ_9 as an example to show how r and s affect the branching ratio as well as the overall size of each partial decay width. In each panel, the color of each pixel indicates the decay width of one particular channel $9 \rightarrow i + j$, which is normalized by the $\Gamma(9 \rightarrow 0 + 0)$ in the $r = 0, s = -4$ case. As expected, varying r from negative to positive amounts to changing the preference over various decay channels from marginal to energetic. Therefore, when s is fixed, we see the hot region moves from the northwest-southeast diagonal to the lower left corner. Moreover, when symmetry is preferred ($s > 0$), the decays along the southwest-northeast diagonal (which is symmetric in i and j) are enhanced; whereas when symmetry is disfavored ($s < 0$), the channels far away from this diagonal are enhanced. The $r = 0, s = 0$ neutral case shows a very democratic distribution where all the partial widths are around the same order of magnitude. It is also clear that the decay width of a single channel as well as the total decay width also vary with r and s .

Besides the total decay width of each state and the partial width of a particular decay channel, we are also interested in the “decay chains”, *i.e.* the paths that a heavy state takes to decay into the lightest state. In Fig. 5.13, we present the major decay chains of the heaviest state from a system with $N + 1 = 10$ states. The vertical axis shows the tower level, while the horizontal axis indicates the number of ‘hops’ that the initial state has taken to reach a particular tower level. The line segments illustrate the hops from one parent state to one daughter state. However, for a particular hop, if a final state is produced less than 5% of all the products from the parent, the corresponding chain is ignored such that we can keep clear track of

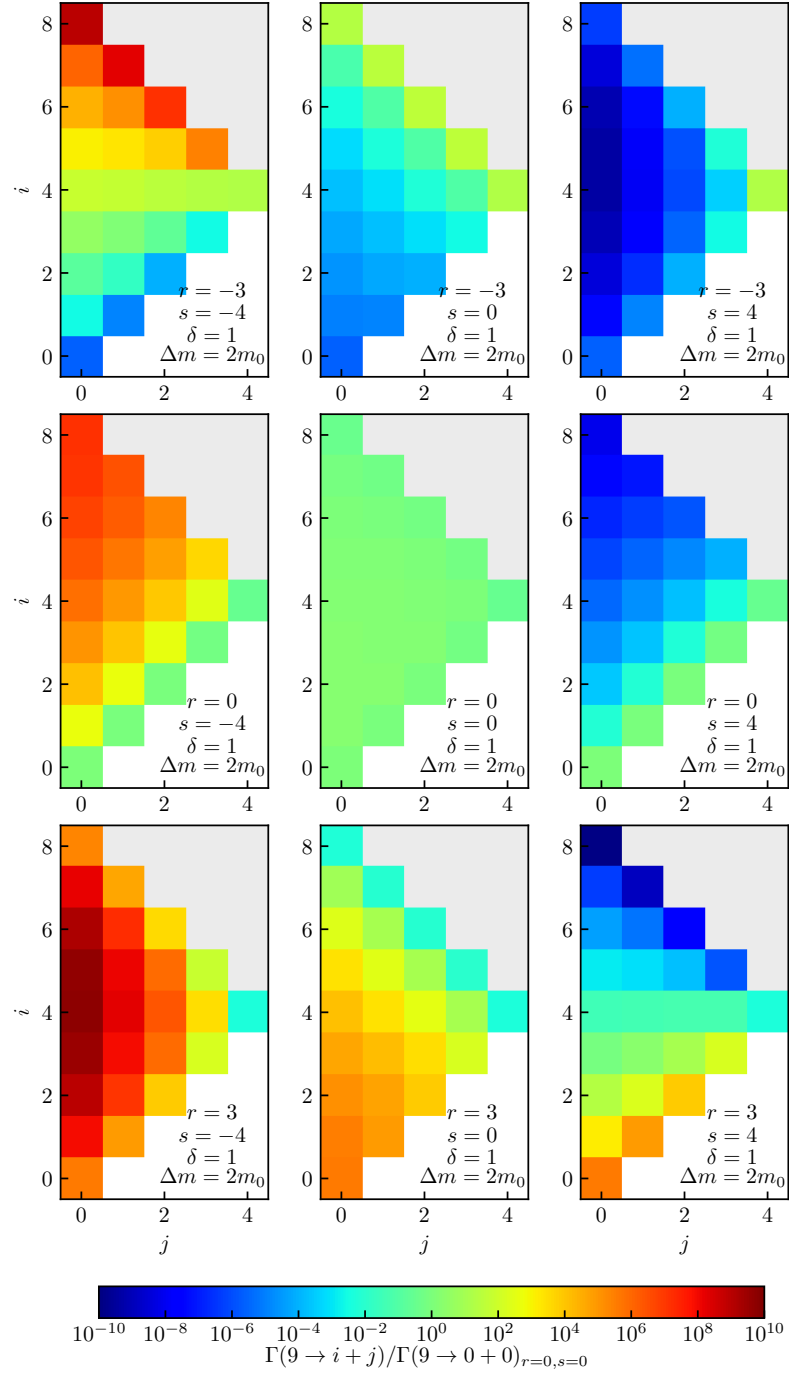


FIGURE 5.12: Partial decay widths of ϕ_9 . The widths are normalized by the $9 \rightarrow 0 + 0$ channel in the $r = 0, s = 0$ case. Each panel corresponds to one combination of r and s , while other parameters are kept fixed. The horizontal and vertical axes indicate the final products. In each panel, the blank space is due to the redundancy in permuting i and j . The gray shaded region corresponds to channels that are kinematically inaccessible. Note that adjusting these r and s not only changes the branching ratios, but also changes the overall decay width of each channel.

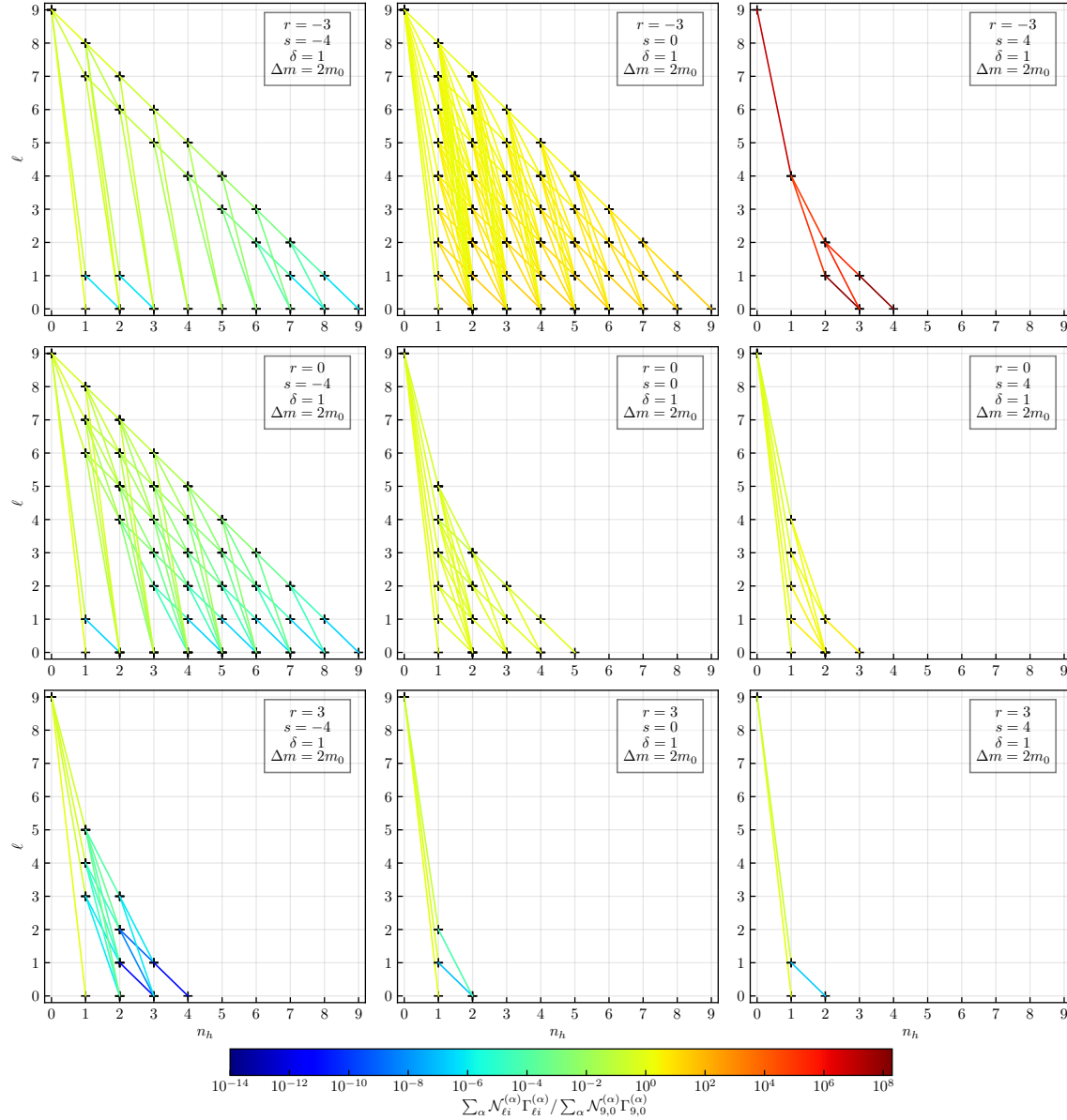


FIGURE 5.13: Major decay chains of a system with 10 states. The vertical axis represents the tower level ℓ of a parent particle. The horizontal axis is the number of “hops” that the heaviest state takes to jump to the current level. For example, the channel $9 \rightarrow 0 + \dots$ takes one hop to the ground state, whereas decay chain $9 \rightarrow 8 + \dots$, $8 \rightarrow 4 + \dots$, $4 \rightarrow 0 + \dots$ takes 3 hops. The color of each line indicates the “production rate” of a certain final state normalized by the production rate of the lightest state from the direct decays of the heaviest state in each case. The normalized production rate is defined as $Pr(\ell \rightarrow i) \equiv \sum_{\alpha} \mathcal{N}_{\ell i}^{(\alpha)} \Gamma_{\ell i}^{(\alpha)} / \sum_{\alpha} \mathcal{N}_{9,0}^{(\alpha)} \Gamma_{9,0}^{(\alpha)}$. In the above panels, only major decay chains are shown. At each hop, if a particular daughter state is produced less than 5% of what its parent can produce in total, the corresponding decay chain is ignored.

the major decay chains. The color coding shows how fast a final state is produced from a parent, which we define as the “production rate”:

$$Pr(\ell \rightarrow i) \equiv \frac{\sum_{\alpha} \mathcal{N}_{\ell i}^{(\alpha)} \Gamma_{\ell i}^{(\alpha)}}{\sum_{\alpha} \mathcal{N}_{9,0}^{(\alpha)} \Gamma_{9,0}^{(\alpha)}}. \quad (5.46)$$

For our toy model, the summation over various decay channels $\ell \rightarrow i + j$ is just a summation over the decay partner j , and thus we have $\mathcal{N}_{\ell i}^{(\alpha)} = 1$ if $i \neq j$, and $\mathcal{N}_{\ell i}^{(\alpha)} = 2$ if $i = j$. Note that we have normalized the production rate in Eq. (5.46) using the the production rate of the lightest state from the direct decays of the heaviest state in each case. At a first glance, this normalization may seem bizarre. But it actually enables us to compare the production rate between direct decays and cascades. The time scale of each decay chain can be inferred by adding the inverse of the production rates. Since the color-coding is on log-scale, we can easily tell the time scales of each decay chain by simply inverting the “slowest” color.

For the four panels around the upper right corner, the ground state is produced around one single time scale since the lines going to the ground state all have similar colors. For the other five panels, we see that the hops to the ground state are colored by yellow, green and/or blue which suggests that they have two or more different time scales for the decay chains to the ground state. According to the previous discussion using the analogy of a conveyor belt, we can predict that the former cases tend to produce distribution functions with simple structure, while the latter cases tend to produce more complicated multi-modal distributions. We will see later that the results in Fig. 5.19 agree with our prediction. The color and line structure also give out the information about the dynamics of the intermediate steps. For example, in the upper right panel, the heaviest state tends to go fast through intermediate states with $\ell = 4, 2, 1$ before eventually reaching the ground state. This is easy to understand, since with $r = -3$, $s = 4$, decays tend to go in a symmetric way and minimize the kinetic energy. Therefore, decay chains like $9 \rightarrow 4 + 4$, $4 \rightarrow 2 + 2$, $2 \rightarrow 1 + 1$ and $1 \rightarrow 0 + 0$ are strongly favored. For the bottom panels and the panel on the left, a significant part of the decays goes directly into the ground state, while a

non-negligible portion of the energy can be stored in some intermediate states which decay into the ground state later. In some of the cases, the decays go through many different intermediate steps while others just have a few channels effectively working. The upper middle panel shows the case in which the heaviest state essentially goes through all the possible decay chains to the ground state. We shall see later that the structure of the major decay chains here allows us to somewhat predict the results from solving the Boltzmann equations.

5.2.2 Dynamical Quantities

For a generic multi-component system, the quantities of cosmological interest, such as those defined in Eqs. (5.1)-(5.4), can be applied to each individual constituent as well as the entire ensemble. The total number density, energy density and pressure is simply a sum of the individual ones:

$$n_{\text{tot}}(t) \equiv \sum_{j=0}^N n_j(t), \quad \rho_{\text{tot}}(t) \equiv \sum_{j=0}^N \rho_j(t), \quad P_{\text{tot}}(t) \equiv \sum_{j=0}^N P_j(t). \quad (5.47)$$

The equation of state parameter of the entire ensemble, on one hand, is defined to be the ratio between the total pressure and the total energy density,

$$w_{\text{tot}}(t) \equiv \frac{P_{\text{tot}}(t)}{\rho_{\text{tot}}(t)} = \sum_{j=0}^N \frac{\rho_j(t)}{\rho_{\text{tot}}(t)} w_j(t), \quad (5.48)$$

which can also be expressed as a weighted sum of each individual $w_j(t)$. On the other hand, from the conservation of energy in a comoving volume $d(a^3 \rho_{\text{tot}}) = -pd(a^3)$, we see [40]

$$w_{\text{tot}}(t) = - \left(\frac{1}{3H} \frac{d \log \rho_{\text{tot}}}{dt} + 1 \right). \quad (5.49)$$

Therefore, w_{tot} really indicates how the total energy density of the entire ensemble scales with time. Essentially, since the kinetic energy dilutes faster than the mass energy in an expanding universe, w_{tot} can be seen as a measure of the kinetic energy in the ensemble.

Following Ref. [40], we define the *tower fraction* that characterizes the non-minimal nature of the dark sector:

$$\eta(t) \equiv 1 - \frac{\max\{\rho_j(t)\}}{\rho_{\text{tot}}(t)} . \quad (5.50)$$

The tower fraction η shows how much energy density is *NOT* taken by the most dominant constituent in the ensemble. If $\eta \rightarrow 1$, the total abundance is spread across the entire ensemble, and even the dominating constituent is not taking a significant share. However, if $\eta \rightarrow 0$, the ensemble is dominated by a single constituent as the abundances of the other constituents are much smaller than the dominant one. To easily visualize how η is built up, we also define the *partial tower fraction*

$$\eta_j(t) \equiv 1 - \frac{\rho_j(t)}{\rho_{\text{tot}}(t)}, \quad (5.51)$$

so that $\eta(t) = \min\{\eta_j(t)\}$.

5.2.3 Numerical Study: Initial Conditions

To simulate the evolution, we must first specify the initial conditions for the ensemble constituents, both in terms of their cosmological abundances and in terms of their phase-space distributions. The starting point of the simulation is assumed to be at a moment deep within the RD epoch when the only efficient processes left in the dark sector are just decays that occur entirely within the dark ensemble. In other words, we assume at the beginning of the simulation, all the constituents of the dark ensemble have already decoupled from both the SM thermal bath and the species in the dark sector. It is also possible that they have never been in equilibrium with any other species.

Depending on the dark-matter production mechanisms, the initial conditions for the simulation can vary a lot. For example, for dark matter produced from thermal freeze-out, the constituents were all in thermal equilibrium before freeze-out, and therefore the distribution functions right at that moment of decoupling would

be Fermi-Dirac or Bose-Einstein: $f_j(p) = 1/(\exp(E/T_j) \pm 1)$. After all the constituents decouple, the distribution functions just redshift. Therefore, the shape of the distribution function is kept fixed in the comoving momentum space, and one can specify the initial condition for each constituent by giving each of them an effective temperature T_j . For dark matter produced from vacuum misalignment, such as the scenarios considered in Ref. [40, 43] in which the dark sector consists of an infinite tower of Kaluza-Klein states, the initial abundance of each constituent depends on the mixing between the KK eigenstates and mass eigenstates. The initial phase-space distribution of each constituent in such a scenario is a δ -function: $f_j(\vec{p}) \sim \delta^{(3)}(\vec{p})$, and one just need to specify an overall constant for each constituent that gives the correct number density.

While each scenario is interesting and important by itself, it is not the purpose of this dissertation to study all the possible initial conditions. Without loss of generality, we assume that the initial abundance of the dark ensemble is completely dominated by the heaviest state, and the initial shape of the distribution function of the heaviest state is Boltzmann-like. Therefore,

$$\begin{aligned} f_N(p, t_{\text{ic}}) &= A \exp(-E_j/T_0) , \\ f_j(p, t_{\text{ic}}) &= 0, \text{ if } j \neq N , \end{aligned} \tag{5.52}$$

where t_{ic} is the cosmological time at which the initial condition is taken, A is an overall factor that determines the initial cosmological abundance, and $T_0 \ll m_0$ can be viewed as the initial temperature of the dark sector.

Note that, as long as the initial condition is sufficiently “cold” and the condition $f_j \ll 1$ is satisfied for all the states throughout the simulation, the exact initial shape of the distribution function will not significantly affect our general results. This is because of the following facts: 1) When $f_j \ll 1$, the factors from the phase-space distribution will not significantly change the decay widths. Therefore, the timescales at which packets are deposited in the phase space of the ground state, *i.e.* the “conveyor belt”, are not significantly changed. 2) When the initial condition

is cold, all the parent particles are approximately at rest at the beginning of the intra-ensemble decay process. Thus, no matter how the phase-space distribution of the parent particles changes, the first decay products will be created with roughly the same momenta around which similar narrow packets would form in the phase space of the daughter particles. This means, the phase-space distribution of the daughter particles are not very sensitive to the change of the initial phase-space distribution of their parent. As the decays go on further, the phase-space distributions of the subsequent decay products will also lose the sensitivity to the initial phase-space distribution.

The above discussion tells us that we can freely choose any T_0 as long as it is much smaller than the lightest mass in the ensemble. The factor A can also be left as a free parameter as long as $f_j \ll 1$ can be satisfied throughout the simulation. Indeed, as we can see from Eq. (A.0.11), as long as $f_j \ll 1$ is satisfied all the time, the Boltzmann equation system is effectively invariant under an overall rescaling of f_j .

The initial time t_{ic} corresponds to an initial temperature T_{ic} of the SM thermal bath. They can be related using the standard time-temperature relation in the RD epoch:

$$t = \sqrt{\frac{45}{2\pi^2}} g_*(T)^{-1/2} \frac{M_{\text{P}}}{T^2}, \quad (5.53)$$

where $g_*(T)$ is the effective relativistic degrees of freedom, and $M_{\text{P}} = 1/\sqrt{8\pi G}$ is the reduced Planck mass. However, when actually solving the Boltzmann equation, instead of using the cosmological time t , we use a dimensionless time variable $\tilde{t} \in [\tilde{t}_{\text{ic}}, \tilde{t}_{\text{f}}]$, where \tilde{t}_{ic} and \tilde{t}_{f} correspond to the beginning time and the end time of the simulation. To map the dimensionless time variable \tilde{t} into cosmological time t , we first notice that, since $H = 1/(2t)$ in the RD epoch, the Boltzmann equation system Eq. (A.0.11) is also invariant under a rescaling of the time variable. Therefore, essentially, \tilde{t} and the real cosmological time t only differ by a multiplicative factor. To find this multiplicative factor, we actually fix T_{ic} first. Then the cosmological time

at the beginning of simulation is

$$t_{\text{ic}} = \gamma \cdot \frac{M_P}{T_{\text{ic}}^2} , \quad (5.54)$$

where $\gamma = \sqrt{45/(2\pi^2)}g_*(T_{\text{ic}})^{-1/2}$ depends only on the initial temperature of the SM thermal bath. By matching t_{ic} to the starting time of the simulation \tilde{t}_{ic} , we find their ratio is

$$\zeta \equiv \frac{\tilde{t}_{\text{ic}}}{t_{\text{ic}}} = \tilde{t}_{\text{ic}} \cdot \frac{T_{\text{ic}}^2}{\gamma M_P} . \quad (5.55)$$

This ratio is fixed between any t and \tilde{t} , so, the conversion is

$$t = \frac{\tilde{t}}{\zeta} = \tilde{t} \times \frac{\gamma M_P}{\tilde{t}_{\text{ic}} T_{\text{ic}}^2} . \quad (5.56)$$

One can also view $\gamma M_P/(T_{\text{ic}}^2 \tilde{t}_{\text{ic}})$ as the unit of \tilde{t} when mapping the timeline of the simulation into the real history of the universe. There is a caveat here. Since by choosing different model parameters, decays would proceed with different rates, the entire intra-ensemble decay process in general does not all end uniformly at \tilde{t}_{f} . Just like Sec. 5.1.1, we can define a \tilde{t}_{end} at which all the decays within the ensemble have effectively ended. Therefore, when mapping the time variable \tilde{t} in the simulation to the real cosmological time t , it is only necessary to map the “physical” part of it, namely, from \tilde{t}_{ic} to \tilde{t}_{end} .

With the freedom in choosing the initial temperature T_{ic} and rescaling f , our simulation can be mapped to a wide range of possible scenarios. We explain our procedure in the following. We first pick up a set of dimensionless model parameters $\{\Delta m/m_0, \delta, r, s\}$ together with the initial shapes of the distributions, and then evolve the Boltzmann equation system until only the lightest constituent exists. From here we can obtain the average momentum in the units of the lightest mass at the end of the intra-ensemble decay process, *i.e.*, $\langle p \rangle_{\text{end}}/m_0$ at \tilde{t}_{end} . We then match it with a scenario in which there is a dark matter mass m_0 , a present-day relic density $\rho_{DM}(t_{\text{now}})$ as well as an average momentum $\langle p \rangle_{\text{now}}$. Since the zeroth order distribution function simply redshifts after the intra-ensemble decay process, matching the present-day average

momentum amounts to a simple shift in the momentum space. So, the first thing we do is redshifting $\langle p \rangle_{\text{end}}$ to match $\langle p \rangle_{\text{now}}$, and then we rescale f_0 to match the relic density. As long as $f_j \ll 1$ is still satisfied after the rescaling, this procedure is self-consistent.

We conclude this subsection by mentioning some additional consistency checks. First of all, although t_{ic} can be any moment after the decoupling of dark matter, it should be significantly smaller than the lifetime of the most short-lived constituent. Otherwise, some constituents in the ensemble would have already been decaying when our simulation starts. Moreover, since it is our assumption that intra-ensemble decays occur within the RD epoch, consistency requires $a_{\text{end}} < a_{\text{eq}}$. Using $p \propto 1/a$, we have

$$\frac{a_{\text{now}}}{a_{\text{end}}} = \frac{\langle p \rangle_{\text{end}}}{\langle p \rangle_{\text{now}}} . \quad (5.57)$$

Therefore, the condition translates into $\langle p \rangle_{\text{end}} / \langle p \rangle_{\text{now}} > a_{\text{now}} / a_{\text{eq}} \sim 3400$. Since $a_{\text{now}} / a_{\text{end}}$ can be easily determined while matching the average momentum, it is straightforward to determine the temperature T_{end} and T_{ic} numerically, using the entropy conservation and the time-temperature relation in Eq. (5.53).

5.2.4 Numerical Study: Evolution Picture

After specifying the initial conditions, we also need to specify the choice of parameters. In principle, the parametrization of the ensemble has a huge effect on the dynamics. For example, the mass spectrum determines the decay kinematics, and the choice of r and s controls the branching ratio as well as the overall decay widths. However, it is not the purpose of this chapter to explore the entire parameter space and establish various bounds. Our goal is to show examples in which various non-trivial phase-space distributions can be generated, and study how these distributions leave their imprints on matter power spectrum. Therefore, we shall take only one simple linear mass spectrum, with $\delta = 1$, $\Delta m = 2m_0$. For r and s , we study all the combinations of $r \in \{-3, 0, 3\}$ and $s \in \{-4, 0, 4\}$, which covers most of the characteristic scenarios. For the number of constituents, we choose $N + 1 = 10$, which is large enough to show

complex dynamics, and, at the same time, computationally affordable. The range of the time variable \tilde{t} is chosen to be $[10^{-6}, 10^6]$, and it is sliced into 200 log-spaced time blocks. The momentum space is discretized logarithmically with 300 momentum bins. Since there are $N + 1 = 10$ states, the system consists of $300 \times 10 = 3000$ mutually coupled differential equations.

Number Density

Fig. 5.14 shows the evolution of comoving number density of each state. The vertical axis is normalized by the initial total comoving number density, with a_{ic} being the scale factor at the beginning of the simulation. We see that, as soon as the heaviest state starts to decay, the lighter states start to populate. As we have seen before, different cases end up with different number of ground-state particles. In general, the cases with larger r tend to have smaller particle number in the end because more mass energy is converted into kinetic energy and get redshifted away.

In the left column, we have $s = -4$, and the decay rate of the heaviest state is relatively fast. Since negative s prefers decay products with different mass, it allows intermediate states to populate one after another. When $r = -3$, decay chains tend to minimize kinetic energy, thus allowing the states right below the heaviest state to dominate. Therefore we see a series of peaks popping out from $i = 8, 7 \dots$ to $i = 1$ in the upper left panel. As r increases, products with smaller masses become favored, thus the first peak becomes $i = 7$ when $r = 0$, and $i = 4$ when $r = 3$. Before proceed to the next column, we remind the readers here that the cases in the upper and bottom left panel violate the consistency requirement since the heaviest state has already started decaying at the initial time of the simulation. We nevertheless study these cases by pretending the associated initial conditions exist.

In the middle column, the heaviest state lives relatively longer, and all the lower states start to populate at similar times as we see the peaks are packed up together. Just like before, we see heavier states, such as $i = 5, 6, 7, 8$ are able to have a non-vanishing population when $r = -3, 0$. But when $r = 3$, only states with smaller masses are able to populate.

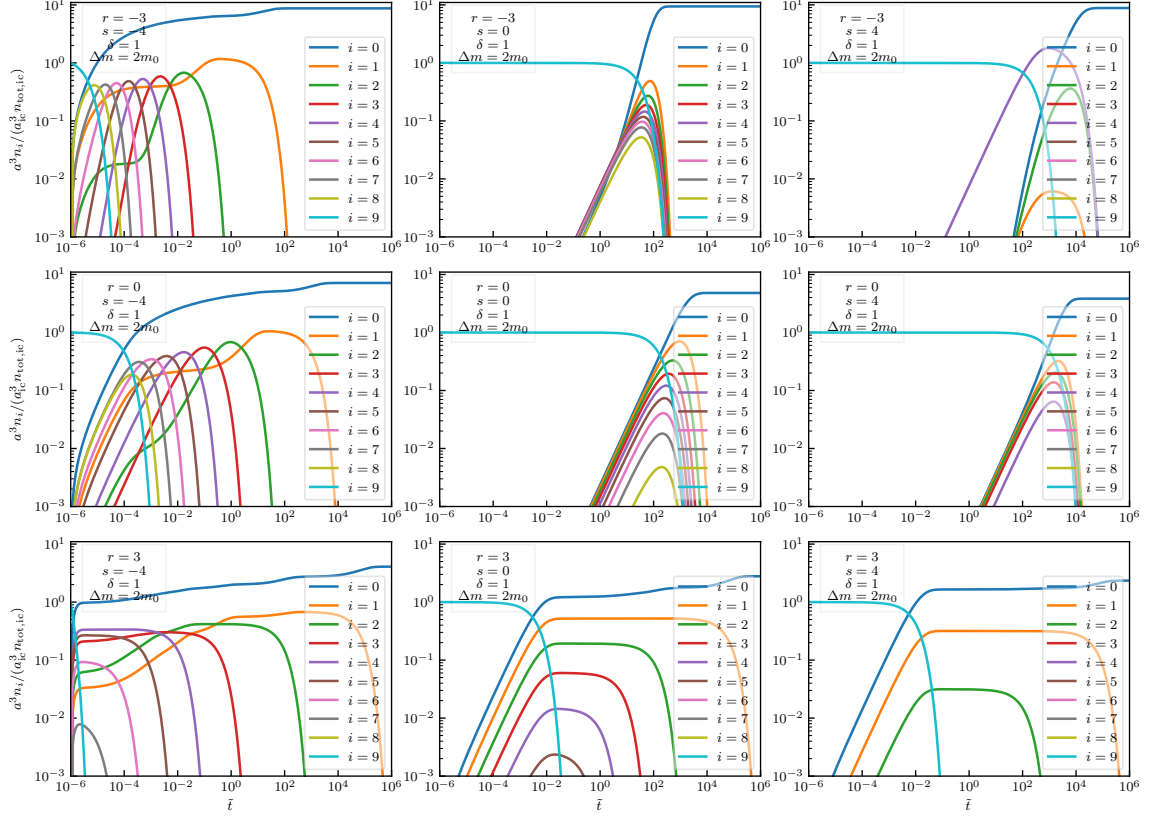


FIGURE 5.14: The evolution of the comoving number density $n_i a^3$ for each state. The curves are normalized by the total initial comoving number density.

In the right column, since $s = 4$, strong preference in symmetry severely reduces the number of decay products whose mass is larger than half of the mother particle. Among relatively lighter states, a smaller r favors the heavier ones, while a larger r favors the lighter ones.

In general, by comparing with Fig. 5.13, we see that the cases with simpler decay-chain structure tend to have less states populating, *i.e.* the decay products tend to be produced with widely different number densities. On the contrary, in the cases where heavier states can cascade through various different decay chains, the evolution history becomes very rich, in which multiple different state are able to populate before decaying.

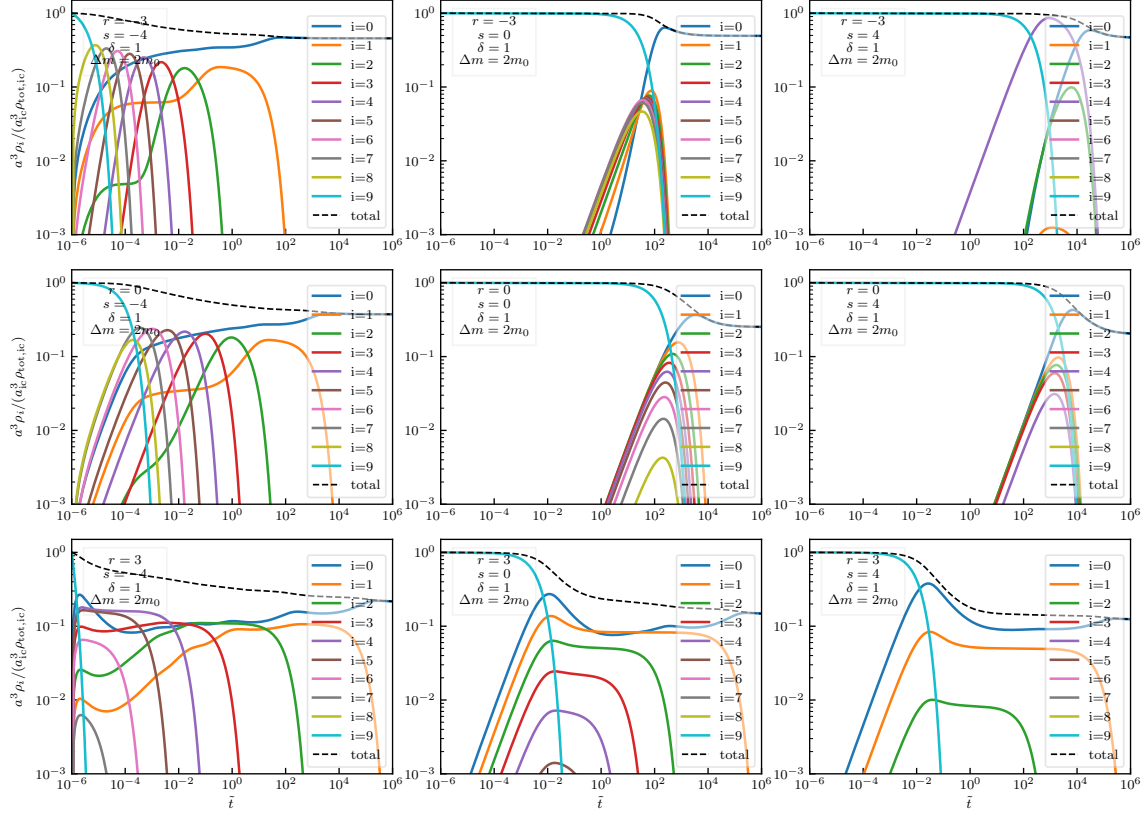


FIGURE 5.15: Comoving energy density of each state as well as the total energy density normalized by the initial total comoving energy density.

Energy Density

Fig. 5.15 shows the evolution of comoving energy density $\rho_i a(\tilde{t})^3$ for each state. The total comoving energy density $\rho_{\text{tot}} a(\tilde{t})^3$ of the entire ensemble is represented by the dashed black curve. The general behavior for each state is similar to Fig. 5.14. However, the energy density takes into account not only the mass energy but also the kinetic energy. Therefore, when a constituent is very energetic, it is able to have a large energy density with a small number density.

For the total energy density, we know that $\rho \sim a^{-3}$ if the entire ensemble is overall cold, while $\rho \sim a^{-4}$ if the ensemble is relativistic. Since the initial distribution is cold, the energy is initially stored almost entirely within the heaviest particle's mass. If there is no intra-ensemble decay at all, the total energy within a comoving volume a^3 would stay constant. However, decays convert mass energy into kinetic energy and

change the way the total energy scales. When decays are actively occurring, we see the comoving energy density $\rho_{\text{tot}} a^3$ decreases, which means the energy density of the entire ensemble is diluting faster than a^{-3} as kinetic energy is redshifted away. Therefore, the evolution of the total comoving energy density reveals how much “radiation” are being produced during the decay history — the comoving energy density decreases faster when there is more radiation, and it stays constant when the mass energy is dominant.

Obviously, the cases with larger r always end up with smaller comoving energy since a larger r indicates a larger portion of the initial mass energy can be converted into radiation via decays. Eventually, as the kinetic energy redshifts away, the total energy density is stored completely within the mass of the lightest state. Having a larger r therefore also means having a smaller number of ground state particles and vice versa. This is consistent with what we have seen in Fig. 5.14.

Tower Fraction

The evolution of tower fractions $\eta_i(\tilde{t})$ and $\eta(\tilde{t})$ are shown in Fig. 5.16. As we have said before, η is a measure of the non-minimal nature of the dark sector. A large η means the dynamics is controlled by multiple states in the ensemble, while a vanishing tower fraction means there exists a single state which takes most of the total abundance.

In our initial condition, we have set the heaviest constituent to dominate at the beginning. As the heaviest state starts to decay, lower states start to populate. Therefore, we see η always increases right from the start. In the $r = -3, 0, s = -4$ cases, the dark sector is consecutively dominated by the intermediate states. This makes η show some oscillatory behavior. For the $r = -3, 0, s = 0$ and $r = 0, s = 4$ cases, the heaviest state dominates for a long time until the ground state takes over. For the intermediate states, though none of them has ever dominated the dark sector, their population is still large enough to keep η sufficiently far away from zero before they decay. When $r = -3, s = 4$, since radiation is disfavored and symmetry is enhanced, the dark sector is shortly taken over by the state $i = 4$ before it is eventually dominated by the ground state. For the $r = 3, s = 0, 4$ cases, η can even grow after

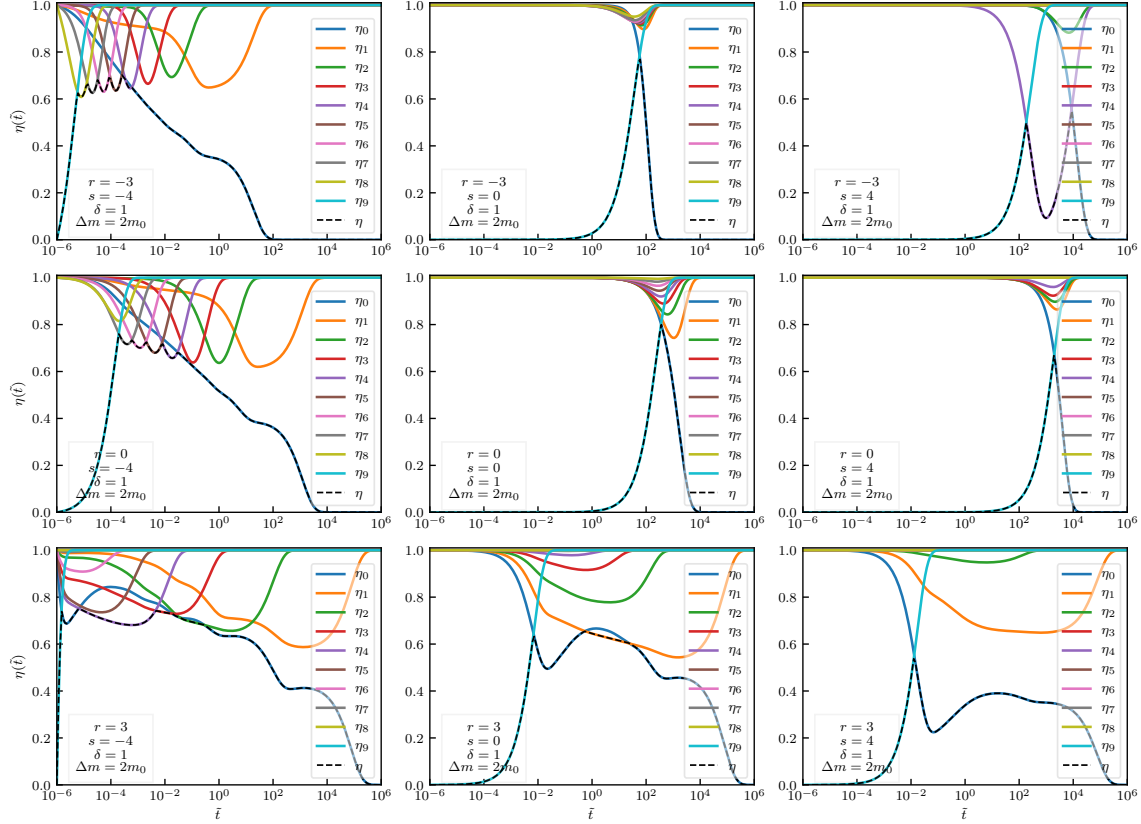


FIGURE 5.16: Evolution of the tower fraction η (dashed black curve) and the partial tower fraction η_i (solid curves).

the ground state has dominated the dark sector. This is because, 1) at the beginning, there is a lot of kinetic energy in the ground state, and therefore they lose their energy rapidly through redshift; 2) at the same time, there exists a relatively long-lived competing state — the $i = 1$ state in the $r = 3, s = 0, 4$ cases, which has less kinetic energy than the ground state and is able to hold its energy for a long time before decaying. Therefore, before the competing state decays, the energy density of the ground state decreases faster than the competing state, making the fraction of abundance carried by the ground state decrease. In the $r = 3, s = 0$ case, the $i = 1$ state even shortly dominates the dark sector.

In general, during intra-ensemble decays the evolution of η can be highly non-trivial. However, in this work, such non-trivial behaviors only occur within the RD epoch. Once intra-ensemble stops, only the lightest state will be propagating, and

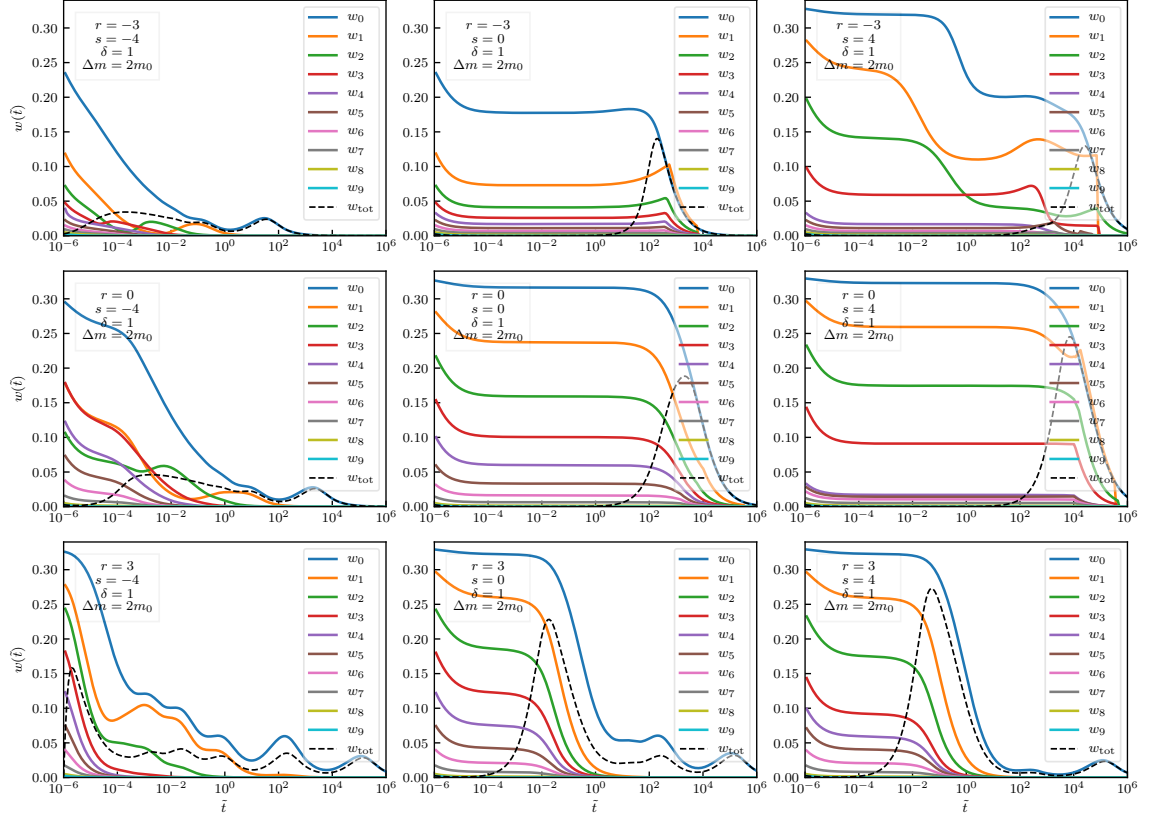


FIGURE 5.17: The equation of state parameter of each constituent and of the whole ensemble.

therefore η will eventually decrease to zero.

Equation of State Parameter

The evolution of the equation of state parameter of each state as well as the entire ensemble is shown in Fig. 5.17. As is easily seen from Eq. (5.4) and Eq. (5.48), the equation of state reveals the “warmness” of a certain energy component in the universe. For the heaviest constituent, since it starts cold and its kinetic energy can only decrease, w_9 stays close to zero all the time. On the contrary, the lower states start empty. Therefore, as soon as the heaviest state decays, the lower states acquire kinetic energy, which results in a positive w_i for $i < N$. The value of w_i depends on how relativistic the corresponding particles are when they are produced. Therefore, we see in most of the cases, the w_i ’s are roughly ordered by their masses at the beginning.

For the left column, since the heaviest state has a relatively short lifetime, a lot of kinetic energy are released early, leaving enough time for the lighter states to redshift and cool down, thus w_i 's go down in general.

For the middle and right columns, the decay of the heaviest state persists for a longer time, thus all the lower states are able to maintain almost constant w_i for quite a while.

For a particular state i , after the heaviest state decays, if there is no more decay from the intermediate states, the w_i simply decreases due to redshift. However, if some of the states are able to retain a non-negligible amount of energy and release later, peaks or wiggles might emerge in the equation of state of the products. For example, in the lower left panel, multiple wiggles and peaks show up in w_1 and w_2 because of the later decays of intermediate states.

The total equation of state parameter w_{tot} is a weighted sum of each individual w_i . It is thus sensitive to decays as long as the state decaying has a non-negligible fraction of energy density. Therefore, the peaks and wiggles in w_{tot} reveals the moments when decays take place. In most of the cases, w_{tot} peaks at the moment when the heaviest state decays. The only exception is the case in the upper right panel in which only the decay of certain intermediate states like $i = 4, 2$ are able to produce a significant amount of radiation.

In the end, notice that in the upper middle, upper right, middle center and middle right panels, there are curves are not smooth when approaching the end of the simulation. These are numeric artifacts, and they only emerge when the corresponding state takes a vanishing fraction of the total abundance. Therefore, we see the smoothness of w_{tot} is not affected.

Phase-Space Evolution

Finally, we show the evolution of the phase-space distribution $a^3 p^3 f(p, \tilde{t})$ of the ground state in Fig. 5.18. We first specify \tilde{t}_{end} — the time at which intra-ensemble decays end, by looking at the comoving number number density of the lightest particle. We set \tilde{t}_{end} to be the moment at which the comoving number density reaches 99.9% of its

asymptotic value. Among all the cases we are considering, \tilde{t}_{end} ranges from $\mathcal{O}(100)$ to $\mathcal{O}(10^5)$. We pick out five snapshots which are evenly spaced on logarithmic time scale between \tilde{t}_{ic} and \tilde{t}_{end} to exhibit the evolution.

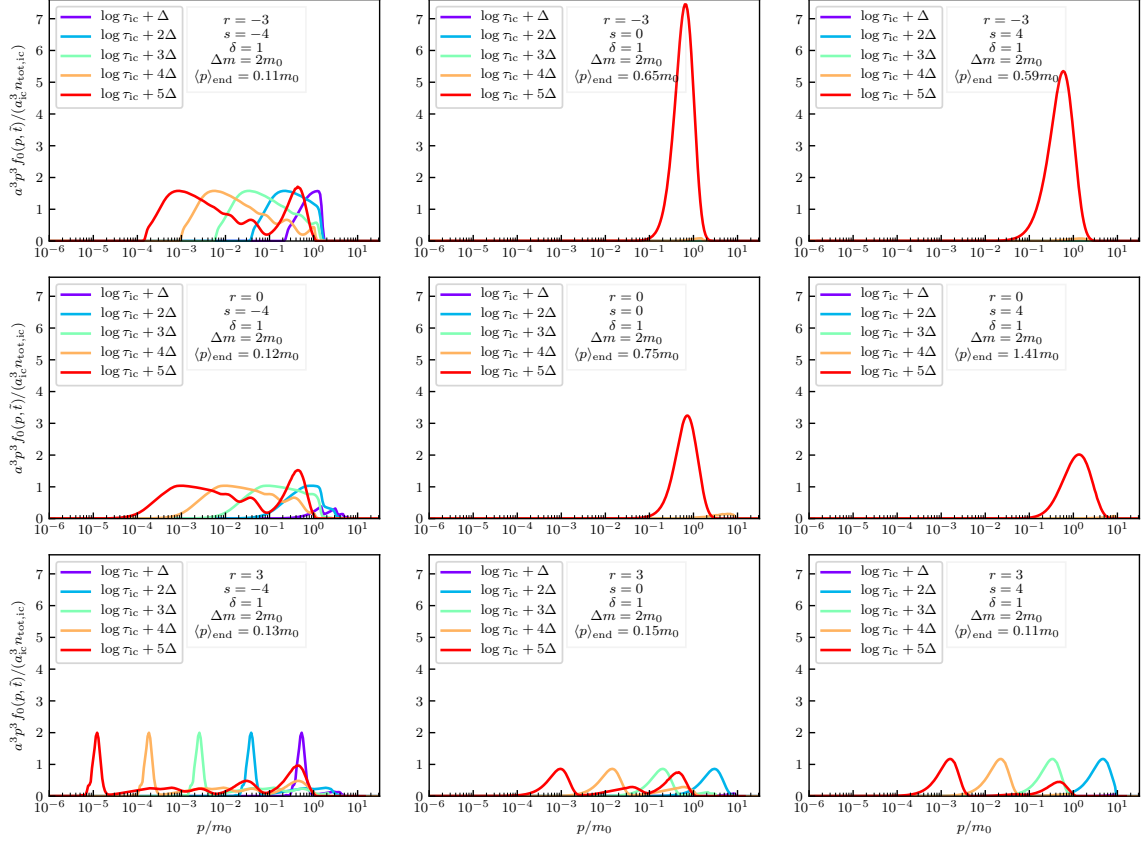


FIGURE 5.18: Snapshots of phase-space distribution during intra-ensemble decays. For each panel, five time slices that are evenly spaced in log-scale between \tilde{t}_{ic} and \tilde{t}_{end} are chosen to show the evolution of phase-space distribution. Here, $\Delta = (1/5) \log_{10}(\tilde{t}_{\text{end}}/\tilde{t}_{\text{ic}})$. Therefore the corresponding time slices are $\tilde{t}_l = \tilde{t}_{\text{ic}} \times 10^{l\Delta}$, where $l = 1, \dots, 5$. Different colors are used to distinguish different times slices. The vertical and horizontal axes are chosen such that the area under each curve is proportional to the comoving number density. The average momentum at the end of the intra-ensemble decay process is also shown in each case.

For the four panels around the upper right corner, the distribution functions take a simple unimodal form. Despite various decay-chain structures that we have seen in Fig. 5.13, this is simply due to the fact that the majority of decays to the ground state all occur around the same time scale, and the expansion of the universe are not able to separate different “deposits”. The other panels, however, have decay chains to the ground state that differ by several orders of magnitude in time. Therefore,

the universe can expand and redshift old deposits to lower momenta before new deposits come in. As a result, these cases tend to give rise to multi-modal phase-space distributions.

The phase-space distributions allow us to calculate the average momentum $\langle p \rangle_{\text{end}}$ at the end of the intra-ensemble decay process. It seems that the unimodal cases in general have a relatively higher average momentum right after the intra-ensemble decay process. Indeed, since decays in these cases tend to occur around the same time scale, at \tilde{t}_{end} , the ground state particles are relatively “fresh” — they are newly produced, and their kinetic energy has not been redshifted away much.

5.2.5 Numerical Study: Matter Power Spectrum

In this section, we would like to look at the matter power spectrum resulting from the phase-space distribution that we have obtained in the toy model. We have mentioned in the previous subsection that, by choosing different parameters, the toy model will end up with different particle number. This means, if we start our model with certain common initial condition, not all of our cases under study would end up generating the desired relic density. While this is indeed true, our freedom in rescaling the distribution function f_j and the time variable \tilde{t} allows us to obtain the correct dark matter abundance in all the cases, up to the self-consistency requirements. Therefore, in this subsection, we shall make use of this freedom. Instead of fixing the initial condition and looking for regions in the parameter space that satisfies all sorts of phenomenological constraints, we shall alter our perspective and ask “*What if the dark matter we observe today had a phase space distribution like those produced after intra-ensemble decays? What can we see in the matter power spectrum that relates to the features in the phase-space distribution?*”

Following the procedure laid out in Sec. 5.2.3, we first pick up a mass for the lightest constituent by setting $m_0 = 10$ keV, a typical mass scale for warm dark matter. Then we redshift the distribution functions in all the cases so that they have the desired average momentum today, which we set to be $\langle p_{\text{now}} \rangle = 100 T_{\text{CMB}}$, where $T_{\text{CMB}} \approx 0.23 \times 10^{-3}$ eV is the present-day CMB temperature. This also

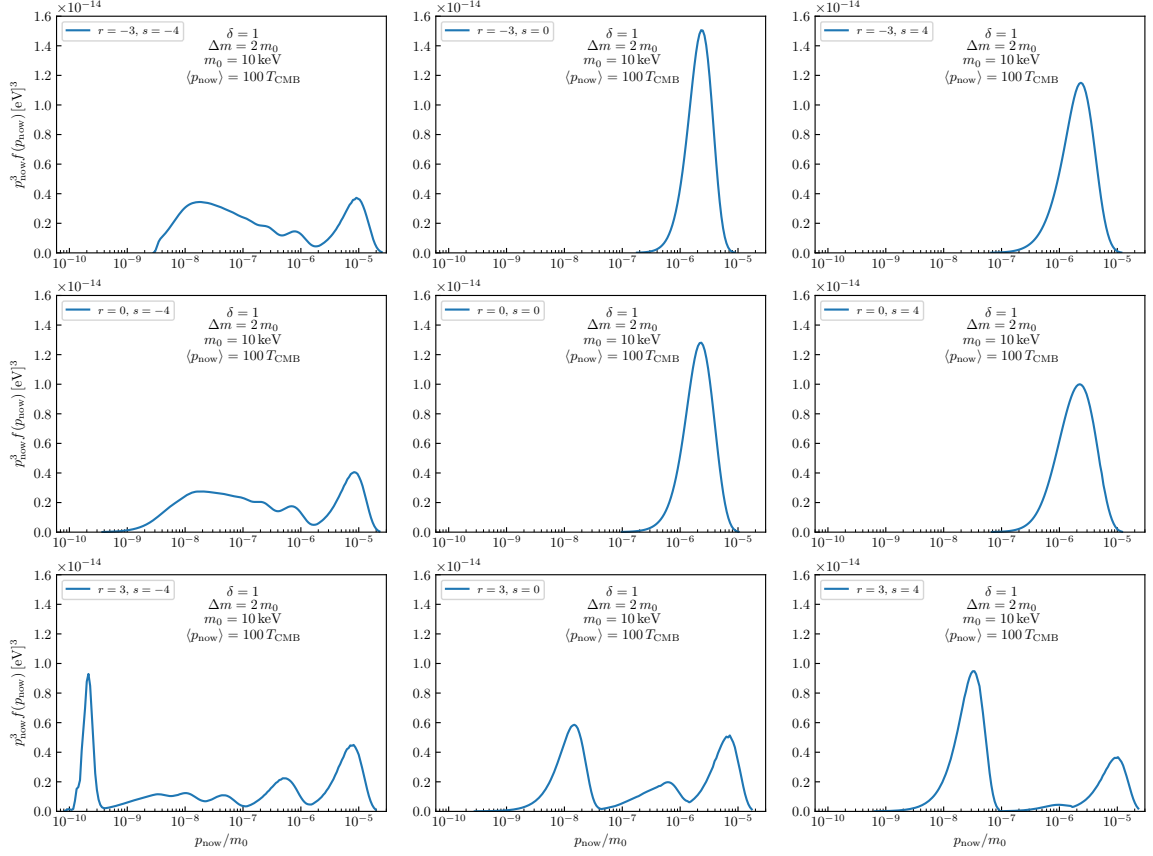


FIGURE 5.19: Phase-space distribution $p_{\text{now}}^3 f_0(p_{\text{now}})$ for all the cases normalized to have the correct present-day DM energy density. The distributions are also redshifted so that the average momentum at the present epoch is 100 times the CMB temperature. The mass of the ground state is taken to be $m_0 = 10$ keV. Each panel corresponds to one choice of r and s .

means the present-day average velocity is $\langle v \rangle_{\text{now}} \sim 2.3 \times 10^{-6}$. In the end, we rescale $f_0(p_{\text{now}}, t_{\text{now}})$ to match the present day dark matter energy density $\rho_{\text{DM}} \sim 0.27 \times 10^{-5} h^2 \text{ GeV cm}^{-3}$. After doing so, we obtain the phase-space distributions in Fig. 5.19. One might have noticed that the blue curve in the lower panel once again violates a consistency requirement as it has to $f_0 > 1$ in the lower-momentum region. However, this can be easily alleviated by slightly increasing both m_0 and $\langle p_{\text{now}} \rangle$ while keeping the average velocity today $\langle v_{\text{now}} \rangle$ unchanged. Such change has no visible effect on the matter power spectrum. Therefore, we shall nevertheless show the matter power spectrum generated from this distribution function.

Since we now know both $\langle p \rangle_{\text{end}}$ and $\langle p_{\text{now}} \rangle$, we are able to calculate the temperature

$\begin{smallmatrix} s \\ r \end{smallmatrix}$	-4	0	4
-3	$T_{\text{ic}} = 0.1 \text{ MeV}$ $T_{\text{end}} = 11.2 \text{ eV}$	$T_{\text{ic}} = 1.0 \text{ MeV}$ $T_{\text{end}} = 65.2 \text{ eV}$	$T_{\text{ic}} = 11.0 \text{ MeV}$ $T_{\text{end}} = 59.0 \text{ eV}$
0	$T_{\text{ic}} = 0.8 \text{ MeV}$ $T_{\text{end}} = 12.3 \text{ eV}$	$T_{\text{ic}} = 5.3 \text{ MeV}$ $T_{\text{end}} = 75.1 \text{ eV}$	$T_{\text{ic}} = 13.1 \text{ MeV}$ $T_{\text{end}} = 140.9 \text{ eV}$
3	$T_{\text{ic}} = 6.5 \text{ MeV}$ $T_{\text{end}} = 13.4 \text{ eV}$	$T_{\text{ic}} = 7.5 \text{ MeV}$ $T_{\text{end}} = 15.2 \text{ eV}$	$T_{\text{ic}} = 5.2 \text{ MeV}$ $T_{\text{end}} = 11.3 \text{ eV}$

TABLE 5.1: SM temperature at the beginning and the end of the intra-ensemble decay process, after shifting the phase-space distributions to $\langle p_{\text{now}} \rangle = 100 T_{\text{CMB}}$.

of the SM sector at the beginning and the end of the intra-ensemble decay process using entropy conservation. The results are collected in Tab. 5.1. Roughly speaking, by choosing $m_0 = 10 \text{ keV}$ and $\langle p_{\text{now}} \rangle = 100 T_{\text{CMB}}$, intra-ensemble decays start at $T_{\text{ic}} \sim \mathcal{O}(1 - 10) \text{ MeV}$ and end at $T_{\text{end}} \sim \mathcal{O}(10 - 100) \text{ eV}$. Since $\langle p_{\text{end}} \rangle / m_0$ is fixed no matter how we rescale the time parameter (see Fig. 5.18), it is essentially the present-day velocity $\langle v_{\text{now}} \rangle \approx \langle p_{\text{now}} \rangle / m_0$ that determines the number of expansions the universe has gone through after the intra-ensemble decay process. Therefore, if a larger mass m_0 or a smaller $\langle p_{\text{now}} \rangle$ is chosen, $\langle v_{\text{now}} \rangle$ can be made smaller, which allows a larger number of expansions between today and the end of the intra-ensemble decay process. In this way, both T_{ic} and T_{end} would be pushed further back in the history of the universe.

The matter-power-spectrum ratios $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$ obtained from the phase-space distributions in Fig. 5.19 are presented in Fig. 5.20. As we have discussed in Sec. 5.1.2.2, traversing along the matter power spectrum in the direction of increasing k amounts to “scanning” the phase-space distribution in the direction of decreasing momentum/velocity. Though the panels in Fig. 5.20 has one-to-one correspondence to the panels in Fig. 5.19, in order to show how this scanning works, we also plot in each panel the present-day phase-space distribution as a function of k . Therefore, at each k , one can easily tell which momentum slice is just reaching this scale. In order to guide our eyes in this “scanning”, we also use a color code based on Eq. (5.39). One can refer to the colorbar to tell the fraction of dark matter that can be considered effectively as CDM component at each scale. The four dotted verti-

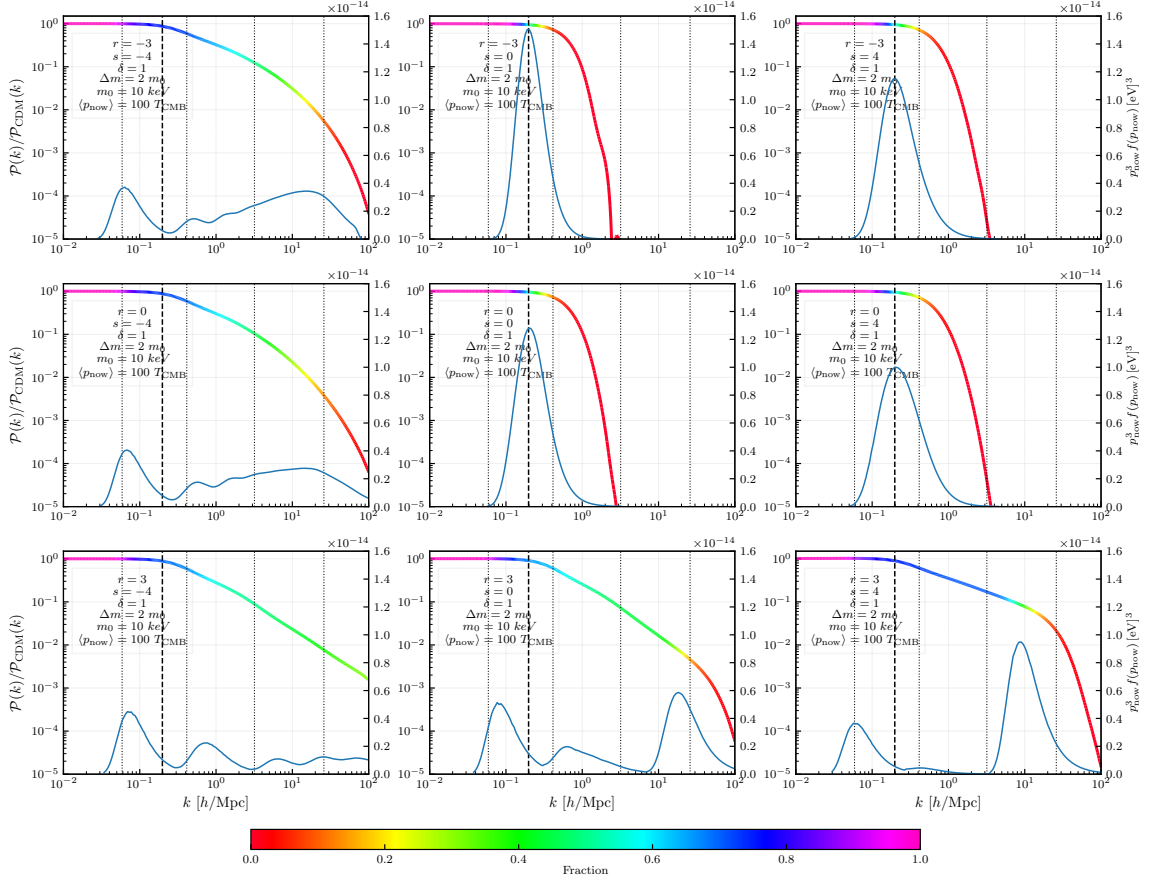


FIGURE 5.20: The matter-power-spectrum ratios $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$ as a scan over the phase-space distribution. The color coding follows Eq. (5.39), which shows the fraction of particles that have $k_{\text{FSH}} > k$. The phase-space distribution $p_{\text{now}}^3 f(p_{\text{now}})$ of each case is plotted against the free-streaming wavenumber by converting p to the corresponding k_{FSH} using Eq. (5.35). The dashed black vertical line corresponds to the free-streaming wavenumber associated with the average momentum today. The four dotted vertical lines from left to right indicate the free-streaming wavenumbers associated with $v_{\text{now}} = 10^{-5}, 10^{-6}, 10^{-7}$, and 10^{-8} , respectively.

cal lines from left to right indicates the free-streaming wavenumbers associated with $v_{\text{now}} = 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}$, respectively. The dashed vertical line marks the free-streaming wavenumber corresponding to the present-day average momentum $\langle p_{\text{now}} \rangle$. We expect the power starts to be suppressed around there. Indeed, it gives a good estimate. However, notice that many of the phase-space distributions actually extend themselves beyond $v_{\text{now}} \sim 10^{-5}$. We can see that the transition of color from purple to blue which suggests a non-vanishing fraction of particles is able to free-streamed to that scale actually gives a even better prediction on the starting point of deviation.

This is more obvious if the vertical axis is on linear scale.

The non-trivial phase-space distribution also gives more information beyond a simple deviation from CDM. For example, the $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ curves in the upper center, upper right, middle center and middle right all exhibit a sharp drop. This is simply due to the single narrow packet in their phase-space distribution which carries the entire DM abundance. As we can see from the scan over the phase-space in Fig. 5.20, in all the cases, when $k \gtrsim 1$, there is no much CDM fraction left to provide gravitational potential, which explains the sharp drop. The lower center and lower right panels also show interesting behaviors. Instead of being strongly suppressed, the ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ seem to approach a constant slope as we scan over the first packet. This is consistent with our previous findings, since the CDM fraction is not changing a lot as both distribution functions have a large packet located in the lower momentum region around $v_{\text{now}} \sim 10^{-7} - 10^{-8}$. Consequently, they both exhibit a sharp turn after scanning the last packet. On the contrary, the cases in the upper left and middle left panels show a more smooth turning because their distribution functions are very spread-out instead of consisting of isolated peaks. The color coding also tells us the change of CDM fraction as k increases in these two cases is more gradual compared with the previous two cases. In the end, the the $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ curve in lower left panel seems to have a constant slope after it first deviate from 1. As we see in the scan, the distribution function is very extended. After passing the first two peaks, there is still $\sim 30 - 40\%$ of dark matter that has not been scanned and acts as an effective CDM component. This non-negligible CDM component protects the gravitational potential and prevents further suppression in the matter power spectrum, therefore giving rise to a nearly constant slope.

Besides identifying the features in the phase-space distribution with the features in the matter power spectrum, one can also sort of invert this mapping to make qualitative predictions about the phase-space distribution based on an observed matter power spectrum. For example, if the power-spectrum ratio $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ sharply decreases right after the first deviation, we can expect a narrow packet in the phase-space distribution that holds a large fraction of dark-matter abundance. If the ratio seems

to reach a constant slope, then beside the packet that causes the first drop, a non-negligible part of the distribution must be located at regions of sufficiently small momentum which can be considered as an effective CDM component. For the power-spectrum ratio that gradually turns, the phase-space distribution is likely to extend over a large range of momentum scales. Moreover, if the ratio reaches a constant slope and suddenly turns down, then the associated distribution function might consist of at least two isolated major packets. One can then use the scales at which these features emerge to predict the locations of the packets in the phase-space. A potentially possible way of reconstructing the phase-space distribution is to make use of the “packet + CDM” cases that we have shown in Sec. 5.1.2.2. One can simulate the actual packet by trying Gaussian packets with different widths and heights, and check if features in the matter power spectrum can be reproduced.

In the end, we emphasize that, while certain predictions about the phase-space distribution can be made by analyzing the matter power spectrum, an actual inverted map is extremely difficult if not entirely impossible. It is not clear whether there exists an one-to-one map. And even if it does exist, the detailed features on the matter power spectrum can be very subtle. For instance, the $\mathcal{P}/\mathcal{P}_{\text{CDM}}$ curves in the upper left and middle left panels are almost indistinguishable. The only predictions we can make are just the average velocity of the first packet, and that the entire phase-space distribution function is not sharp but extended. It is very difficult to obtain additional information beyond these. The same argument also applies to the upper middle, upper right, middle center and middle right panels. In reality, the observed matter power spectrum is also blurred by observational uncertainties. Therefore, a systematic exploration of how phase-space distribution can be reconstructed from matter power spectrum is beyond the scope of this dissertation.

5.3 Conclusion

In this chapter, we have studied the cosmological imprints from non-minimal dark sectors. In particular, we have presented a thorough study on the phase-space distribution generated from the decays within a non-minimal dark sector. As a first

step, we analyzed how the decay processes affect the phase-space distribution of the decay products. In general, this dynamical process in the phase space of a particular constituent consists of three parts, 1) packets redshift due to the expansion of the universe; 2) packets disappearing due to the decay of itself; 3) packets growing due to the decay from higher states. We have found that the shape of a newly-formed packet is related to the shape of the decaying packet of the parent. When multiple decay channels are available, each packet works like a single “brick” that drops on the “conveyor belt” at different times and builds up the distribution function of the daughter particles. The whole intra-ensemble decay process eventually left its imprints as a set of packets stacking up together in the phase space of the lightest state. Depending on the parameters that controls the decay widths and branching ratios, the resulting phase-space distribution could be wildly different. We have found that, for scenarios in which major decay chains from the heaviest state are separated by different timescales, the phase-space distribution of the ground state will naturally be multi-modal. However, if a single decay chain dominates, or if all the major decay chains reach the ground state at the same timescale, the shape of the phase-space distribution will be relatively simple. When the intra-ensemble decay process is over, the phase-space distribution of the relic simply redshifts. Therefore, the nontrivial shape of the momentum distribution molded by the decays will not change in the comoving momentum space until the structure formation starts.

We further studied the relation between the phase-space distribution and the structure formation by looking at the suppression patterns in matter power spectrum. Traditionally, the Jeans’ scale which is obtained from the velocity dispersion or the sound speed is used to give an estimate on the scale at which the structure formation is suppressed. However, the story it tells is far from complete. First of all, the Jeans’ scale is time dependent and reaches its maximum not at today, but around the epoch of matter-radiation equality. Second, it only vaguely separates modes that are oscillatory from the modes that are growing. The growing modes near the Jeans’ scale can also be suppressed if free-streaming effects are not completely negligible. Besides, the velocity dispersion can well characterize the phase-space distribution only

when the distribution is unimodal. However, the sort of phase-space distributions derived from decays can easily be multi-modal and even feature a combination of isolated packets and extended continuum. Therefore in those cases a single velocity dispersion is not enough to capture all the features. We found it very useful to use the free-streaming horizon to analyze the effects from phase-space distribution. The free-streaming horizon is an integrated quantity which gives the largest distance that particles with certain velocity have traveled today. Thus, it naturally gives the largest scale those particles can impact. Besides, since particles can have more than one characteristic velocities if the distribution function is non-trivial, using free-streaming horizon allows us to associate particles with any velocity to a corresponding length scale and identify the possible effects from them.

We have subsequently studied the matter power spectra resulting from different types of distribution functions and compared them with the standard CDM scenario. To be specific, we looked at the ratio $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$ and studied the way deviation from the standard scenario occurs as well as all the features beyond the deviation point. We learnt that, as we move along the matter-power-spectrum ratio in the direction of increasing k , we are also effectively “scanning” the phase-space distribution in the direction of decreasing particle velocity. The matter power spectrum starts to deviate from the standard CDM scenario at the location where there is a non-negligible fraction of dark-matter particles able to free-stream to this scale. Beyond this initial deviation, the way the power decreases as k increases depends on the details in the distribution function. In general, at a particular k , we found it useful to separate the entire dark matter into two parts – an effective warm component, who has its $k_{\text{FSH}} > k$, and an effective cold component whose $k_{\text{FSH}} < k$. We found the suppression patterns are sensitive to the phase-space distribution of the warm component, while only sensitive to the abundance of the cold component. Since the exact shape of its phase-space distribution does not matter, the cold component can be well approximated by a CDM component with the same abundance. The way the warm component exert its influence is by smoothing out its inhomogeneities in the RD epoch, and thus weakening the gravitational potential. On the contrary, the cold

component always protects the gravitational potential and resists the dampening from the warm component. The evolution of the total matter perturbation thus depends on the interplay between the two. We then qualitatively identified the features in the phase-space distribution $p_{\text{now}}^3 f(p_{\text{now}})$ with the features in the matter-power-spectrum ratio $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$ using several sets of characteristic examples.

To show the application of our analysis, we chose to study a DDM-motivated toy model with $N + 1 = 10$ constituents. We presented the parametrization of the mass spectrum and the couplings through which we can easily vary the total decay width, as well as the branching ratios. We have also showed the way the decay chains change while varying the ensemble parameters. We then presented the results from solving Boltzmann equations at the level of the phase-space distribution $f_i(p, \tilde{t})$. With all the information obtained from the phase-space distribution, we presented the non-trivial evolution of the number density, the energy density, the tower fraction and the equation of state. We saw that, by going through different decay chains, the final yield of ground-state particles can be quite different. This is because kinetic energy scales differently from the mass energy, and the amount of kinetic energy that is produced during decays varies in different cases. In the same way, the total energy density also scales differently in different cases. When decays are not occurring and particles have cooled down, $\rho_{\text{tot}} \sim a^{-3}$. However, when particles are actively decaying and producing energetic daughter particles, ρ_{tot} dilutes faster. In general, when r is bigger, more radiation is produced and more energy is lost in the end due to redshift. The tower fraction which indicates the non-minimality of the dark ensemble also evolves in a highly non-trivial way when decays occur. In some cases, intermediate states are able to shortly dominate the dark sector before they decay, while in others, the heaviest state can quickly transfer its energy to the ground state. In all the cases, η are able to stay significantly away from zero, which means the dark sector is in general far away from the single-particle picture. The equation of state parameter for the entire dark sector, which is built up from that of each constituent, also works as a indicator of the amount of kinetic energy in the dark sector. When decays occur, the equation of state parameter will increase because of the injection of

momentum. When decays end, it will decrease as the injected momentum redshifts away. Therefore, the fluctuations in the equation of state parameter also reveal the time when decays occur.

We further showed the evolution of phase-space distribution from the beginning to the end of the intra-ensemble decay process. Just as expected, the phase-space distribution can have multi-modal structure when there exist multiple decay chains with different timescales. However, unimodal structure can also be obtained when decays to the ground state tend to occur around the same time. We made use of the freedom in rescaling the time variable and the distribution function to make the phase-space distributions have the same average momentum and to make them have the correct dark-matter relic abundance. By fixing the present-day average momentum to $\langle p_{\text{now}} \rangle = 100 T_{\text{CMB}}$, we were able to calculate the temperature scale of the universe when intra-ensemble decays begin and complete. Roughly, the cases that we studied have $T_{\text{ic}} \sim \mathcal{O}(1 - 10)$ MeV and $T_{\text{end}} \sim \mathcal{O}(10 - 100)$ eV. But this can be modified by choosing other parameters.

In the end, we showed the matter power spectra generated from CLASS by inputting the “real” phase-space distributions. We were able to use the conclusions from Sec. 5.1.2.2 to qualitatively map certain features in the distribution functions to the patterns in the matter-power-spectrum ratio $\mathcal{P}(k)/\mathcal{P}_{\text{CDM}}(k)$. However, it is very difficult if not completely impossible to invert the mapping. Nevertheless, limited but perhaps valuable information about the phase-space distribution is able to be obtained by studying the matter-power-spectrum ratio.

Finally, let us discuss some of the explicit assumptions that have been made in this work. First, we have assumed that the dark matter obtains non-thermal phase-space distribution only from decays that occur completely within the dark sector. Other decay scenarios can also be considered. For instance, decays from heavier dark-matter states to lighter dark-matter states can be accompanied by one or more SM final states. In these cases, in addition to the suppression of structure, if such decays occur at late time, they might give rise to visible signals that explain the PAMELA positron excess [148] or give rise to interesting direct detection signals due to the existence of

the excited states [149]. Besides being produced from dark-sector particles or other particles in the BSM sector, dark matter can also obtain non-thermal phase-space distribution from decays of the SM particles. For example, Ref. [150] explored the scenario in which the SM photons obtain an in-medium plasma mass and decay into the dark-matter particles. In general, there are various early-universe processes that are able to produce non-thermal phase-space distributions, and it is possible that multiple processes need to be taken into account in the simulation. It would be useful to summarize all the possible patterns associated with each specific type of process using techniques similar to the analysis of decay chains.

Second, in the numerical study, we have required that our phase-space density $f_j \ll 1$. This is due to the fact that the decay term and the feed term in the Boltzmann equations receive strong Bose enhancement once $f_j \gtrsim 1$ (see Eq. (A.0.7) and Eq. (A.0.9)). For axion-like particles whose masses are extremely small, effects from such enhancement could be huge and would easily dominate the decay-width calculations which lead to severe stiffness problem in numerics. However, in these cases, effects from inverse decays also need to be taken into account and might be the key to a computationally affordable code. We shall leave it for future work.

Last but not the least, we have assumed that intra-ensemble decays complete before the MRE. This is a very conservative choice since we do not want to inject too much kinetic energy during the structure formation in the MD epoch. Nevertheless, we have seen in Refs. [151–155] that dark matter decaying into dark radiation between the recombination and the present epoch could help alleviate the Hubble tension [156] and the σ_8 tension [157]. Since intra-ensemble decays essentially transform the dark-matter mass energy into the radiation energy and change the expansion history, it might provide an alternative approach in alleviating these tensions. We shall also leave this for future work.

CHAPTER 6

CONCLUSIONS

In this dissertation, we have studied some cosmological aspects of non-minimal dark sectors. We have seen that the phenomenology of non-minimal dark sectors is just not relying on the properties of one or a few constituents — it depends on the properties of all the constituents within the dark sector.

In scenarios like the DDM, the dark sector often consists of a large number or even an infinite number of dark-matter components. Therefore, it is the *collective* behavior of those components that governs the cosmological evolution of dark matter. In the first part of this dissertation, we have explored the possibility of constructing DDM models using an ensemble of dark hadrons or gauge-neutral bulk states of Type I string theories. In both constructions, the mass spectrum of the dark-matter components follows the linear Regge trajectory, and the density of states in the dark ensemble grows exponentially with mass, which is called the Hagedorn behavior. Therefore both scenarios share the same parametrization, and we have studied them together. We started by examine the internal consistency of the models, and investigated the ways to properly organize the dark-matter components so that one can obtain the correct total relic abundance and have the desired balance between the SM decay widths and the cosmological abundances of the dark-matter components. We subsequently studied the evolution of the total dark-matter abundance, the effective equation of state as well as the tower fraction. We have seen that their evolution picture is very dynamical as the decays of the dark-matter components result in additional time-dependence beyond those normally associated with the expansion of the universe. We further demonstrated the phenomenological viability of such models using the constraints that follow directly from the evolution picture. Surprisingly, we found these preliminary constraints not only allow the dark sector to remain non-minimal today, but also tend to introduce correlations between various properties of these

DDM ensembles such as their associated mass scales, lifetimes, and abundance distributions. For example, the fundamental mass scales and the diversity of the dark sector today are inversely correlated — while the energy scales of the DDM ensemble are allowed to range from the GeV scale all the way to the Planck scale, the total present-day cosmological abundance of the dark sector must be spread across an increasing number of different states in the ensemble as these energy scales are dialed from the Planck scale down to the GeV scale.

In other scenarios, the dynamics in non-minimal dark sectors could leave behind only a single dark-matter candidate. Therefore, the phenomenology of those scenarios depends heavily on the processes that have taken place in the early universe. We focused on processes that leave non-trivial imprints on the phase-space distribution of the resulting dark-matter particles. In particular, we found that decays from the heavier constituents to the lighter constituents in a non-minimal dark sector could give rise to non-thermal or even multi-modal phase-space distributions. In general, such the non-trivial multi-modal distributions can be easily obtained by having different decay chains from heavier states that reach the ground state at highly separated timescales. We then investigated the cosmological consequences of having such non-trivial phase-space distribution by examining the suppression patterns in the shape of the resulting matter power spectrum. We found it particularly useful to analyze the matter power spectrum with the free-streaming horizon — particles can only suppress density perturbations at scales smaller than their free-streaming horizon, otherwise, they behave like perfectly cold dark matter. We then established a qualitative map from the features in phase-space distribution to the changes in the shape of the matter power spectrum. Finally, we also demonstrated the application of those techniques by studying a toy model with $N+1$ ensemble constituents. We have shown that, while the forward map from the phase-space distribution to the matter power spectrum works very well, the inverse map is much more difficult. However, in some circumstances, qualitative information can be learnt and might actually reveal valuable information about the early-universe processes.

In the end, we emphasize that the scenarios that we have considered in this dis-

sertation are at the certain extremes of a general non-minimal dark-sector scenario. In principle, both the decays that produce final states external to the dark ensemble and the intra-ensemble decays may play a non-negligible role in the evolution of the dark sector. There could also be decay channels that produce both the intra-ensemble states and the states outside the dark ensemble. In the most general context, processes like scatterings, annihilations and coannihilations also need to be taken into account, and they might smear the phase-space distribution produced via the decays, or even reestablish the equilibrium either within the dark sector or with the other sectors. Such effects are highly model-dependent. Instead of taking realistic models and establish various phenomenological bounds, the goals of the two works are to provide a proof-of-principle study, to show a viable framework in which specific models can be constructed and studied in future work.

APPENDIX A

BOLTZMANN EQUATIONS

In this appendix, we present a full derivation of the Boltzmann equations for the distribution functions when decays entirely within the ensemble are taken into account. In general, the evolution of phase-space distribution is subject to three effects – the dilution due to the expansion of the universe, the loss due to the decay of the particle itself and the gain from the decays of other heavier particles. Let us begin the derivation by considering only one species and ignoring any interactions. The conservation of comoving particle number density $d(na^3)/dt = 0$ gives

$$\frac{dn}{dt} = -3Hn. \quad (\text{A.0.1})$$

As universe expands, particles redshift. In a FRW universe, the momentum redshifts as

$$\frac{dp}{dt} = -Hp, \quad (\text{A.0.2})$$

which means the particles with momentum p at time t will have momentum p' at a later time t' which satisfies

$$\begin{aligned} p' &= p \cdot e^{-\int_t^{t'} H dt} \\ &= p \cdot \frac{a(t)}{a(t')}. \end{aligned} \quad (\text{A.0.3})$$

In a comoving volume a^3 , the redshift effect can be seen as a continuous flow of particles in the momentum space – particles that were in the momentum interval $[p, dp]$ at time t , flow into the interval $[p', p' + dp']$ at a later time t' with the speed $dp/dt = -Hp$. Therefore, the continuity equation applies:

$$\frac{\partial}{\partial t} (a^3 p^2 f) + \frac{\partial}{\partial p} (-Hp \cdot a^3 p^2 f) = 0, \quad (\text{A.0.4})$$

in which we have properly taken in to account the expansion of the universe by multiplying the factor a^3 in each term. After some simplification, we find

$$\frac{\partial f}{\partial t} = Hp \frac{\partial f}{\partial p} , \quad (\text{A.0.5})$$

which is the Boltzmann equation for the distribution functions in the absence of interactions. To check the result, one can integrate both sides and obtain

$$\begin{aligned} \int dp \frac{p^2}{2\pi^2} \frac{\partial f}{\partial t} &= H \int dp \frac{p^3}{2\pi^2} \frac{\partial f}{\partial p} \\ \frac{\partial}{\partial t} \int dp \frac{p^2}{2\pi^2} f &= -3H \int dp \frac{p^2}{2\pi^2} f \\ \left(\frac{dn}{dt} \right)_{\text{exp}} &= -3Hn , \end{aligned} \quad (\text{A.0.6})$$

where, in the second line, integration by parts is used, and the subscript ‘exp’ means the change of particle number density due to the expansion of the universe only. The result is consistent with Eq. (A.0.1).

Now let us consider an ensemble with multiple constituents, and assume that decays can occur completely within the ensemble. For the i th ($i = 0, 1, 2, \dots$) ensemble constituent, its decay width for a general decay channel $\alpha : i \rightarrow j + k + \dots$ is

$$\Gamma_i^{(\alpha)}(p_i, t) = \frac{1}{2E_i} \int [d\pi_j (1 \pm f_j)] [d\pi_k (1 \pm f_k)] \dots |\mathcal{M}|^2 , \quad (\text{A.0.7})$$

where $d\pi_i = d^3p_i / [(2\pi)^2 2E_i]$, $|\mathcal{M}|$ is the corresponding amplitude, and $+/-$ is chosen for bosons/fermions, respectively. The change of number density due to the decays is,

$$\left(\frac{dn_i}{dt} \right)_{\text{dec}} = - \int \frac{dp_i}{2\pi^2} p_i^2 f_i \sum_{\alpha} \Gamma_i^{(\alpha)} . \quad (\text{A.0.8})$$

While one constituent is decaying, heavier constituents can also decay into the con-

stituent in consideration. The rate of change due to this feeding effect is

$$\left(\frac{dn_i}{dt}\right)_{\text{feed}} = \sum_{\ell, \alpha} \mathcal{N}_{\ell i}^{(\alpha)} \int [d\pi_\ell f_\ell] [d\pi_i (1 \pm f_i)]^{N_{\ell i}^{(\alpha)}} \left\{ \prod_{m \neq i} [d\pi_m (1 \pm f_m)]^{N_{\ell m}^{(\alpha)}} \right\} |\mathcal{M}|^2 . \quad (\text{A.0.9})$$

Here, $\mathcal{N}_{\ell i}^{(\alpha)}$ is the multiplicity of the particle i from the decay of ℓ in the channel $\alpha : \ell \rightarrow i + \dots$

Putting all the relevant sources together, the change of particle number density when intra-ensemble decays are present is

$$\left(\frac{dn_i}{dt}\right) = \left(\frac{dn_i}{dt}\right)_{\text{exp}} + \left(\frac{dn_i}{dt}\right)_{\text{dec}} + \left(\frac{dn_i}{dt}\right)_{\text{feed}} . \quad (\text{A.0.10})$$

We can then write everything in terms of the phase-space distribution $f_i(p, t)$ and reach the master equation:

$$\begin{aligned} \frac{\partial f_i}{\partial t} &= H p_i \frac{\partial f_i}{\partial p_i} - f_i \sum_{\alpha} \Gamma_i^{(\alpha)} \\ &\quad + \frac{(1 \pm f_i)}{2E_i} \sum_{\ell, \alpha} \mathcal{N}_{\ell i}^{(\alpha)} \int [d\pi_\ell f_\ell] [d\pi_i (1 \pm f_i)]^{N_{\ell i}^{(\alpha)} - 1} \\ &\quad \times \left\{ \prod_{m \neq i} [d\pi_m (1 \pm f_m)]^{N_{\ell m}^{(\alpha)}} \right\} |\mathcal{M}|^2 . \end{aligned} \quad (\text{A.0.11})$$

Therefore, for each constituent i , and for each point p_i in the momentum space, there is a Boltzmann equation that governs the evolution of $f_i(p_i, t)$. And this evolution depends on the full information of all the distribution functions in the ensemble. In other words, the evolution of each individual distribution function is tightly coupled to the evolution of the entire ensemble.

APPENDIX B

RELATIVISTIC KINEMATICS

For Boltzmann equations in the form of Eq. (A.0.11), all the information about the kinematics is encoded in the second term and the third term on the right-hand side of the equation, which we shall refer to as the decay term and the feed term, respectively. In this appendix, we shall show detailed calculations for these two terms separately.

B.1 Decay Term

Consider the case in which particle i with momentum p_i decays into particle j and k with momentum p_j and p_k at moment t . The explicit expression for decay width is

$$\begin{aligned} \Gamma_{(jk)}^i(t, \vec{p}_i) &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{2E_i} \int \frac{d^3 p_j}{(2\pi)^3 2E_j} \frac{d^3 p_k}{(2\pi)^3 2E_k} \\ &\times (1 \pm f_j(\vec{p}_j, t))(1 \pm f_k(\vec{p}_k, t)) \\ &\times (2\pi)^4 \delta^{(4)}(p_i - p_j - p_k) , \end{aligned} \quad (\text{B.1.1})$$

where we use “+” for bosons and “−” for fermions. The reason for the fermionic case is such: were the particle a fermion, the probability that the momentum eigenstate $|\vec{p}_i\rangle$ is not occupied is $1 - f_i(\vec{p}_i, t)$ (recall that the occupation number can only be 0 or 1). For bosons, $f_i(\vec{p}_i, t)$ is the occupation number of the state $|f_i(\vec{p}_i, t)\rangle$ with momentum \vec{p}_i . It is straightforward that

$$a_{\vec{p}_i}^\dagger |f_i(\vec{p}_i, t)\rangle = \sqrt{1 + f_i(\vec{p}_i, t)} |1 + f_i(\vec{p}_i, t)\rangle . \quad (\text{B.1.2})$$

Therefore the transition rate into the state $|1 + f_i(\vec{p}_i, t)\rangle$ is enhanced by the occupation number of this state as compared with that into $|0\rangle$. We shall take “+” signs from now on since we are dealing with bosons.

Integrating over one of the momentum, say \vec{p}_k , we get

$$\begin{aligned}\Gamma_{(jk)}^i(t, p_i) &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{32\pi^2 E_i} \int \frac{d^3 p_j}{E_j E_k} (1 + f_j(|\vec{p}_j|, t)) \\ &\times (1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \delta(E_i - E_j - E_k) .\end{aligned}\quad (\text{B.1.3})$$

We shall use p_i for $|\vec{p}_j|$ from here on whenever there is no confusion. Since $E_k = \sqrt{p_i^2 + p_j^2 - 2p_i p_j \cos \theta_j + m_k^2}$, the delta function is now a function of p_j and $\cos \theta_j$, where θ_j is the angle between \vec{p}_i and \vec{p}_j . This means there are two different ways to integrate the delta function – integrate with the angle fixed or with the momentum fixed.

Angle fixed

Let's first assume θ_j is fixed and rewrite the delta function in terms of momentum p_j

$$\begin{aligned}\delta(E_i - E_j - E_k) &= \sum_l \delta(p_j - p_j^l) \left| \frac{\partial(E_i - E_j - E_k)}{\partial p_j} \right|_{p_j^l}^{-1} \\ &= \sum_l \delta(p_j - p_j^l) \frac{E_j^l E_k^l}{|p_j^l E_i - p_i E_j^l \cos \theta_j|} ,\end{aligned}\quad (\text{B.1.4})$$

where l denotes different solutions. We will see below that the solutions depend on the angle and there might be one or two solutions at each angle in different situations. Substituting this in the previous formula, we get

$$\begin{aligned}\Gamma_{(jk)}^i(t, p_i) &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{32\pi^2 E_i} \sum_l \int d^3 p_j (1 + f_j(p_j, t))(1 + f_k(p_k, t)) \frac{\delta(p_j - p_j^l)}{|p_j^l E_i - p_i E_j^l \cos \theta_j|} \\ &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi E_i} \sum_l \int d \cos \theta_j (1 + f_j(p_j^l, t))(1 + f_k(p_k^l, t)) \\ &\times \frac{p_j^{l^2}}{|p_j^l E_i - p_i E_j^l \cos \theta_j|}\end{aligned}\quad (\text{B.1.5})$$

Let us calculate the solution p_j^l . Let us first set up a frame where \vec{p}_i is in \hat{z} direction, and \vec{p}_j, \vec{p}_k are in the polar angles θ_j, θ_k , respectively. According to the conservation of energy $E_i = E_j + E_k$, we can solve for p_j :

$$p_j = \frac{(m_i^2 + m_j^2 - m_k^2)p_i \cos \theta_j \pm 2E_i \sqrt{m_i^2 p^{*2} - m_j^2 p_i^2 \sin^2 \theta_j}}{2(m_i^2 + p_i^2 \sin^2 \theta_j)}, \quad (\text{B.1.6})$$

where p^* is the solution in the rest frame, namely, the expression in Eq. (5.44). p_k can be solved easily by

$$p_k = \sqrt{p_i^2 + p_j^2 - 2p_i p_j \cos \theta_j}. \quad (\text{B.1.7})$$

So, in each angle θ_j , there could be two solutions for p_j . However, we need to carefully classify different cases and pick up the correct solution.

- $m_i p^* \leq m_j p_i$

This is the case where the parent has a relatively large momentum. To make sure this solution is physical, we need the part in the square root to be positive, i.e. $\sin \theta_j \leq m_i p^* / (m_j p_i) \leq 1$. So, there is a maximum angle $\theta_j \leq \pi/2$ for the daughter j , and both solutions must be kept. We can easily understand this by rewriting the condition as $p_i/m_i > p^*/m_j$ – this means the parent has a larger velocity than what the daughter j would have if the decay happens in a rest frame. So, all the decay products will go in the forward direction, i.e. we need to boost all the solutions in the backward direction in the rest frame to the forward direction, and thus there have to be two solutions at each possible angle. In such a situation the integration with respect to $\cos \theta_j$ goes from $\sqrt{1 - (m_i p^* / m_j p_i)^2}$ to 1.

- $m_i p^* > m_j p_i$

In this case, θ_j is allowed to take value from $[0, \pi]$, and therefore, at each angle, there exists only one solution. We shall therefore take the + sign in this case, and the integration bound is $[-1, 1]$.

This approach is clear physically. However, in numerical computation, there exists a pole in the first case. One can easily show that, when $\cos \theta_j = \sqrt{1 - (m_i p^* / m_j p_i)^2}$,

i.e. right on the boundary of integration, the denominator in Eq. (B.1.5) vanishes. Thus, we shall resort to the next approach whenever necessary.

Momentum fixed

The second approach is to fix the magnitude of momentum p_j first and look for solutions in $\cos \theta_j$ in the equation $E_i - E_j - E_k = 0$:

$$\cos \theta_j = \frac{2E_i E_j - (m_i^2 + m_j^2 - m_k^2)}{2p_i p_j}, \quad (\text{B.1.8})$$

$$\frac{\partial(E_i - E_j - E_k)}{\partial \cos \theta_j} = -\frac{E_k}{\partial \cos \theta_j} = \frac{p_i p_j}{E_k}, \quad (\text{B.1.9})$$

$$\delta(E_i - E_j - E_k) = \frac{E'_k}{p_i p_j} \delta(\cos \theta_j - \cos \theta'_j), \quad (\text{B.1.10})$$

where the “prime” stands for the existing solution for certain momentum magnitude. So the decay rate can be rewritten as:

$$\begin{aligned} \Gamma_{(jk)}^i(t, p_i) &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi E_i} \int dp_j d\cos \theta_j \frac{p_j}{p_i E_j} (1 + f_j(p_j, t))(1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \\ &\quad \times \delta(\cos \theta_j - \cos \theta'_j) \\ &= \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi E_i} \int_{p_j^{min}}^{p_j^{max}} dp_j \frac{p_j}{p_i E_j} (1 + f_j(|\vec{p}_j|, t))(1 + f_k(p'_k, t)), \end{aligned} \quad (\text{B.1.11})$$

where p_j^{min} and p_j^{max} are the lower and upper bounds of integration, respectively. This is because p_i determines how much the system is boosted and hence the maximum and minimum momentum of the daughter j . On one hand, p_j^{max} is obtained by setting $\theta_j = 0$ in the “+” solution of Eq. (B.1.6):

$$p_j^{max} = \frac{(m_i^2 + m_j^2 - m_k^2)p_i + 2E_i m_i p^*}{2m_i^2}. \quad (\text{B.1.12})$$

On the other hand, the expression for p_j^{min} depends on different situations. In the case where $m_i p^* / m_j p_i > 1$, p_j^{min} corresponds to $\theta_j = \pi$:

$$p_j^{min} = \frac{-(m_i^2 + m_j^2 - m_k^2)p_i + 2E_i m_i p^*}{2m_i^2}; \quad (\text{B.1.13})$$

whereas, in the case where $m_i p^*/m_j p_i \leq 1$, p_j^{min} corresponds to setting $\theta_j = 0$ in the “−” solution:

$$p_j^{min} = \frac{(m_i^2 + m_j^2 - m_k^2)p_i - 2E_i m_i p^*}{2m_i^2}. \quad (\text{B.1.14})$$

Obviously, the second approach can't be used in numerical calculation when $p_i = 0$. Therefore, one needs to switch between two approaches whenever necessary.

B.2 Inverse Decay Term

When decays exist, in principle, inverse decays should also exist. Here we shall prove that inverse decays can be safely neglected. Consider two beams of particles j and k collide and produce particles i , *i.e.* $j + k \rightarrow i$. The volume of interaction in a unit time is $V = vA$, where A is the cross-sectional area of the two beams, and v is the relative velocity. Let us first take the rest frame of particle k . The possibility that particles j and k will collide is equal to $N_k \sigma / A$, where σ is the cross section, and N_k is the number of particles k in the interaction volume. The number of production events per unit volume then is

$$\begin{aligned} \frac{N_j N_k \sigma}{VA} &= \frac{n_j n_k \sigma V}{A} \\ &= n_j n_k \sigma v \\ &= n_j n_k \sigma |\vec{v}_j - \vec{v}_k|, \end{aligned} \quad (\text{B.2.1})$$

where n_j, n_k are the number density of each particle, and, in the third line, we have switched to a general frame, so the relative velocity becomes the general expression $|\vec{v}_j - \vec{v}_k|$.

The above analysis applies to the case where the two bunches of particle have well defined unique momenta. However, in our cosmic collider, particles travel in all the different directions. Therefore, we have to generalize Eq. (B.2.1). In our study of the Boltzmann equations of distribution functions, we are interested in the production rate of the particle i with a specific momentum p_i , so that we can calculate the time

evolution of $f_i(p_i, t)$. So, in the above equation, the cross section σ should be the differential cross section of particles j, k with momenta \vec{p}_j, \vec{p}_k producing the particle i with the momentum \vec{p}_i , which is

$$d\sigma(\vec{p}_j, \vec{p}_k \rightarrow \vec{p}_i) = \frac{d^3p_i}{(2\pi)^3 2E_i} \frac{|\mathcal{M}|^2}{(2E_j)(2E_k)|\vec{v}_j - \vec{v}_k|} \times (2\pi)^4 \delta^{(4)}(p_j + p_k - p_i) . \quad (\text{B.2.2})$$

Besides, n_j and n_k should really be the number density of these two particles with momenta \vec{p}_j and \vec{p}_k , which are

$$\frac{d^3p_j}{(2\pi)^3} f_j(\vec{p}_j, t), \frac{d^3p_k}{(2\pi)^3} f_k(\vec{p}_k, t) , \quad (\text{B.2.3})$$

respectively. So, the production rate of particle i due to this particular inverse-decay channel is

$$\begin{aligned} \left. \frac{\partial n_i}{\partial t} \right|_{\text{inv.dec}} &= \int \frac{d^3p_i}{(2\pi)^3} (1 + f_i(p_i, t)) \frac{|\mathcal{M}_{j+k \rightarrow i}|^2}{2E_i} \\ &\times \int \frac{d^3p_j}{(2\pi)^3 (2E_j)} \int \frac{d^3p_k}{(2\pi)^3 (2E_k)} f_j(p_j, t) f_k(p_k, t) \\ &\times (2\pi)^4 \delta^{(4)}(p_j + p_k - p_i) , \end{aligned} \quad (\text{B.2.4})$$

where we have dropped the vector by assuming isotropy. Also note that we add a $1 + f$ term in the integration due to the Bose enhancement effect. In Boltzmann equations of the phase-space distribution, we can simply put this in after stripping down the integration $\int d^3p_i / (2\pi)^3$. This is very similar with Eq. (B.1.1) with the $(1 + f_j)(1 + f_k)$ term replaced by $f_j f_k$. As long as $f_\ell \ll 1, \forall \ell$, this production rate can be safely ignored from the Boltzmann equation for f_i during the period of intra-ensemble decays.

Other than contributing to the particle i , the inverse decays give rise to a particle loss term in the Boltzmann equations of f_j (and f_k). To study this term, we can

rearrange the integration is Eq. (B.2.4):

$$\begin{aligned} \left. \frac{\partial n_j}{\partial t} \right|_{\text{inv. dec}} &= \int \frac{d^3 p_j}{(2\pi)^3} f_j(p_j, t) \frac{|\mathcal{M}_{j+k \rightarrow i}|^2}{2E_j} \int \frac{d^3 p_i}{(2\pi)^3 (2E_i)} (1 + f_i(p_i, t)) \\ &\times \int \frac{d^3 p_k}{(2\pi)^3 (2E_k)} f_k(p_k, t) (2\pi)^4 \delta^{(4)}(p_j + p_k - p_i) . \end{aligned} \quad (\text{B.2.5})$$

Therefore, by similar argument, as long as $f_\ell \ll 1$, $\forall \ell$, this term is much smaller than the feed term which are discussed in what follows. This means we can also ignore it.

Moreover, the effects from the scattering processes ($i + j \rightarrow k + \ell$) induced by the trilinear coupling in Eq. (5.41) can also be neglected due to the same argument – they are all suppressed by the phase-space factors. In addition, they are also suppressed by an additional coupling constant. Therefore, during intra-ensemble decays, the scattering rates among the ensemble constituents through the trilinear term would be much smaller than decay rates, and thus it is safe to ignore the scatterings if they can only proceed through the trilinear couplings.

B.3 Feed Term

Let us now consider the contribution from the process $i \rightarrow j + k$ to $f_j(p_j, t)$. The exact form of the feed term is

$$\begin{aligned} &\int \frac{d^3 p_i}{(2\pi)^3 2E_i} f_i(p_i, t) |\mathcal{M}_{i \rightarrow j+k}|^2 \int \frac{d^3 p_j}{(2\pi)^3 2E_j} \frac{d^3 p_k}{(2\pi)^3 2E_k} (1 + f_j(p_j, t))(1 + f_k(p_k, t)) \\ &\times (2\pi)^4 \delta^{(4)}(p_i - p_j - p_k) \\ &= \int \frac{d^3 p_j}{(2\pi)^3} \frac{1 + f_j(p_j, t)}{2E_j} \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi^2} \int \frac{d^3 p_i}{E_i E_k} f_i(p_i, t) (1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \\ &\times \delta(E_i - E_j - E_k) , \end{aligned} \quad (\text{B.3.1})$$

before removing the integration with respect to p_j . Similar to the decay term, we need to develop two equivalent ways to calculate the integral since $E_i - E_j - E_k$ is a function of both p_i and the angle θ_i between \vec{p}_i and \vec{p}_j .

Angle fixed

Let us rewrite the δ -function by fixing θ_i first:

$$\delta(E_i - E_j - E_k) = \sum_l \frac{\delta(p_i - p_i^l) E_i^l E_k^l}{|p_i^l E_j - p_j E_i^l \cos \theta_i|} . \quad (\text{B.3.2})$$

Substituting this in Eq. (B.3.1) and ripping off $\int d^3 p_j / (2\pi)^3$, we have

$$\begin{aligned} & \frac{1 + f_j(p_j, t)}{2E_j} \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi^2} \int \frac{d^3 p_i}{E_i E_k} f_i(p_i, t) (1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \delta(E_i - E_j - E_k) \\ = & \frac{1 + f_j(p_j, t)}{16\pi E_j} |\mathcal{M}_{i \rightarrow j+k}|^2 \sum_l \int d \cos \theta_i d p_i f_i(p_i, t) (1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \\ & \times \frac{p_i^2 \delta(p_i - p_i^l)}{|p_i^l E_j - p_j E_i^l \cos \theta_i|} \\ = & \frac{1 + f_j(p_j, t)}{16\pi E_j} |\mathcal{M}_{i \rightarrow j+k}|^2 \sum_l \int d \cos \theta_i f_i(p_i^l, t) (1 + f_k(p_k^l, t)) \frac{p_i^{l^2}}{|p_i^l E_j - p_j E_i^l \cos \theta_i|} . \end{aligned} \quad (\text{B.3.3})$$

Let us assume the momentum \vec{p}_j , is in the \hat{z} direction, while \vec{p}_i and \vec{p}_k form angles θ_i and θ_k with respect to it, respectively. We can use the conservation of energy to solve for p_i :

$$p_i = \frac{(m_i^2 + m_j^2 - m_k^2) p_j \cos \theta_i \pm 2m_i E_j \sqrt{p^{*2} - p_j^2 \sin^2 \theta_i}}{2(E_j^2 - p_j^2 \cos^2 \theta_i)} . \quad (\text{B.3.4})$$

We can now study different cases:

- $p^* < p_j$

Once again, in order for the solution to be physical, we need $\sin \theta_i < p^*/p_j$. This makes sense because you can only obtain $p_j > p^*$ in the forward direction (forward with respect to the parent particle). Both solutions must be kept in this case – one from boosting a forward direction decay in the rest frame, the other from boosting a decay in backward direction. In this case, the integration of $\cos \theta_i$ goes from $\sqrt{1 - (p^*/p_j)^2}$ to 1.

- $p^* \geq p_j$

This is the case in which the daughter j is relatively slow. There is no constraint on the θ_j . We take the upper sign in this case, since the lower sign leads to unphysical solution when $\cos \theta_i$ becomes negative. The solution in backward direction ($\theta_i < \pi/2$) corresponds to a “slow” parent producing a daughter particle backwards, while a forward direction solution ($\theta_i \geq \pi/2$) comes from boosting a backward solution in the rest frame with a large velocity. In this case, the integration bound is $[-1, 1]$.

Just like before, whenever there are two solutions at the same angle, there exists a pole in E.q.(B.3.2) when $\cos \theta_i = \sqrt{1 - (p^*/p_j)^2}$. Therefore, the other approach with momentum fixed first is still necessary.

Momentum fixed

Let us repeat the previous calculations by fixing p_j first and solve for $\cos \theta_j$ in the equation $E_i - E_j - E_k = 0$:

$$\cos \theta_i = \frac{2E_i E_j - (m_i^2 + m_j^2 - m_k^2)}{2p_i p_j} \quad (\text{B.3.5})$$

$$\frac{\partial(E_i - E_j - E_k)}{\partial \cos \theta_i} = -\frac{\partial E_k}{\partial \cos \theta_i} = \frac{p_i p_j}{E_k} \quad (\text{B.3.6})$$

$$\delta(E_i - E_j - E_k) = \frac{E'_k}{p_i p_j} \delta(\cos \theta_i - \cos \theta'_i) \quad (\text{B.3.7})$$

Inserting these in Eq. (B.3.1), the feed term becomes:

$$\begin{aligned} & \frac{1 + f_j(p_j, t)}{2E_j} \frac{|\mathcal{M}_{i \rightarrow j+k}|^2}{16\pi^2} \int \frac{d^3 p_i}{E_i E_k} f_i(p_i, t) (1 + f_k(|\vec{p}_i - \vec{p}_j|, t)) \frac{E'_k}{p_i p_j} \delta(\cos \theta_i - \cos \theta'_i) \\ &= \frac{1 + f_j(p_j, t)}{16\pi E_j} |\mathcal{M}_{i \rightarrow j+k}|^2 \int_{p_i^{min}}^{p_i^{max}} dp_i \frac{p_i}{p_j E_i} f_i(p_i, t) (1 + f_k(p'_k, t)) . \end{aligned} \quad (\text{B.3.8})$$

For different cases, the bounds for integration are taken differently:

- $p^* < p_j$

$$p_i^{min} = \frac{(m_i^2 + m_j^2 - m_k^2)p_j - 2m_i E_j p^*}{2m_j^2}, \quad (\text{B.3.9})$$

$$p_i^{max} = \frac{(m_i^2 + m_j^2 - m_k^2)p_j + 2m_i E_j p^*}{2m_j^2}, \quad (\text{B.3.10})$$

- $p^* \geq p_j$

$$p_i^{min} = \frac{-(m_i^2 + m_j^2 - m_k^2)p_j + 2m_i E_j p^*}{2m_j^2}, \quad (\text{B.3.11})$$

$$p_i^{max} = \frac{(m_i^2 + m_j^2 - m_k^2)p_j + 2m_i E_j p^*}{2m_j^2}. \quad (\text{B.3.12})$$

Those are the formulas that we explicitly use in the code.

REFERENCES

- [1] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett.* **B716**, 1 (2012), [arXiv:1207.7214 \[hep-ex\]](#) .
- [2] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Lett.* **B716**, 30 (2012), [arXiv:1207.7235 \[hep-ex\]](#) .
- [3] S. Weinberg, *Phys. Rev.* **D13**, 974 (1976).
- [4] C. A. Baker *et al.*, *Phys. Rev. Lett.* **97**, 131801 (2006), [arXiv:hep-ex/0602020 \[hep-ex\]](#) .
- [5] P. A. R. Ade *et al.* (Planck), *Astron. Astrophys.* **594**, A13 (2016), [arXiv:1502.01589 \[astro-ph.CO\]](#) .
- [6] M. Tanabashi *et al.* (Particle Data Group), *Phys. Rev.* **D98**, 030001 (2018).
- [7] A. G. Riess *et al.* (Supernova Search Team), *Astron. J.* **116**, 1009 (1998), [arXiv:astro-ph/9805201 \[astro-ph\]](#) .
- [8] S. Perlmutter *et al.* (Supernova Cosmology Project), *Astrophys. J.* **517**, 565 (1999), [arXiv:astro-ph/9812133 \[astro-ph\]](#) .
- [9] R. D. Peccei and H. R. Quinn, *Phys. Rev. Lett.* **38**, 1440 (1977).
- [10] R. D. Peccei and H. R. Quinn, *Phys. Rev.* **D16**, 1791 (1977).
- [11] S. Weinberg, *Phys. Rev. Lett.* **40**, 223 (1978).
- [12] F. Wilczek, *Phys. Rev. Lett.* **40**, 279 (1978).
- [13] K. R. Dienes, F. Huang, S. Su, and B. Thomas, *Phys. Rev.* **D95**, 043526 (2017), [arXiv:1610.04112 \[hep-ph\]](#) .
- [14] K. R. Dienes, F. Huang, S. Su, and B. Thomas, *Proceedings, Workshop on Neutrino Physics: Session of CETUP* 2016: Lead/Deadwood, South Dakota, USA, Jule 6-July 17, 2016*, *AIP Conf. Proc.* **1900**, 040003 (2017).
- [15] J. K. S. S. K. R. Dienes, F. Huang and B. Thomas, .
- [16] S. Alam *et al.* (BOSS), *Mon. Not. Roy. Astron. Soc.* **470**, 2617 (2017), [arXiv:1607.03155 \[astro-ph.CO\]](#) .
- [17] A. G. Riess *et al.*, *Astrophys. J.* **826**, 56 (2016), [arXiv:1604.01424 \[astro-ph.CO\]](#) .

- [18] A. Berlin, D. Hooper, and G. Krnjaic, *Phys. Rev.* **D94**, 095019 (2016), [arXiv:1609.02555 \[hep-ph\]](#) .
- [19] A. Berlin, D. Hooper, and G. Krnjaic, *Phys. Lett.* **B760**, 106 (2016), [arXiv:1602.08490 \[hep-ph\]](#) .
- [20] J. T. Giblin, G. Kane, E. Nesbit, S. Watson, and Y. Zhao, *Phys. Rev.* **D96**, 043525 (2017), [arXiv:1706.08536 \[hep-th\]](#) .
- [21] L. Heurtier and F. Huang, .
- [22] E. W. Kolb and M. S. Turner, *Front. Phys.* **69**, 1 (1990).
- [23] M. B. Green, J. H. Schwarz, and E. Witten, *SUPERSTRING THEORY. VOL. 1: INTRODUCTION*, Cambridge Monographs on Mathematical Physics (1988).
- [24] M. B. Green, J. H. Schwarz, and E. Witten, *SUPERSTRING THEORY. VOL. 2: LOOP AMPLITUDES, ANOMALIES AND PHENOMENOLOGY* (1988).
- [25] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2007).
- [26] J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2007).
- [27] C. Rovelli, *Living Rev. Rel.* **1**, 1 (1998), [arXiv:gr-qc/9710008 \[gr-qc\]](#) .
- [28] A. Ashtekar, M. Bojowald, and J. Lewandowski, *Adv. Theor. Math. Phys.* **7**, 233 (2003), [arXiv:gr-qc/0304074 \[gr-qc\]](#) .
- [29] A. Ashtekar and P. Singh, *Class. Quant. Grav.* **28**, 213001 (2011), [arXiv:1108.0893 \[gr-qc\]](#) .
- [30] O. Wantz and E. P. S. Shellard, *Phys. Rev.* **D82**, 123508 (2010), [arXiv:0910.1066 \[astro-ph.CO\]](#) .
- [31] F. Zwicky, *Helv. Phys. Acta* **6**, 110 (1933), [Gen. Rel. Grav.41,207(2009)].
- [32] F. Zwicky, *Astrophys. J.* **86**, 217 (1937).
- [33] V. C. Rubin and W. K. Ford, Jr., *Astrophys. J.* **159**, 379 (1970).
- [34] M. Milgrom, *Astrophys. J.* **270**, 365 (1983).
- [35] M. Milgrom, *Astrophys. J.* **270**, 371 (1983).
- [36] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, *Astrophys. J.* **679**, 1173 (2008), [arXiv:0704.0261 \[astro-ph\]](#) .

- [37] B. J. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *Phys. Rev.* **D81**, 104019 (2010), [arXiv:0912.5297 \[astro-ph.CO\]](#) .
- [38] S. Profumo, *An Introduction to Particle Dark Matter* (World Scientific, 2017).
- [39] A. Boyarsky, O. Ruchayskiy, and D. Iakubovskyi, *JCAP* **0903**, 005 (2009), [arXiv:0808.3902 \[hep-ph\]](#) .
- [40] K. R. Dienes and B. Thomas, *Phys. Rev.* **D85**, 083523 (2012), [arXiv:1106.4546 \[hep-ph\]](#) .
- [41] K. R. Dienes and B. Thomas, *Proceedings, Workshop on Dark Matter, Unification and Neutrino Physics (CETUP* 2012): Lead/Deadwood, South Dakota, USA, July 10-August 1, 2012*, *AIP Conf. Proc.* **1534**, 57 (2012).
- [42] K. R. Dienes and B. Thomas, *Proceedings, 36th International Conference on High Energy Physics (ICHEP2012): Melbourne, Australia, July 4-11, 2012*, *PoS ICHEP2012*, 452 (2013).
- [43] K. R. Dienes and B. Thomas, *Phys. Rev.* **D85**, 083524 (2012), [arXiv:1107.0721 \[hep-ph\]](#) .
- [44] K. R. Dienes and B. Thomas, *Phys. Rev.* **D86**, 055013 (2012), [arXiv:1203.1923 \[hep-ph\]](#) .
- [45] K. R. Dienes and B. Thomas, *Proceedings, 36th International Conference on High Energy Physics (ICHEP2012): Melbourne, Australia, July 4-11, 2012*, *PoS ICHEP2012*, 460 (2013).
- [46] K. R. Dienes, J. Kumar, and B. Thomas, *Phys. Rev.* **D86**, 055016 (2012), [arXiv:1208.0336 \[hep-ph\]](#) .
- [47] K. R. Dienes, J. Kumar, and B. Thomas, *Phys. Rev.* **D88**, 103509 (2013), [arXiv:1306.2959 \[hep-ph\]](#) .
- [48] K. R. Dienes, J. Kumar, and B. Thomas, *Proceedings, Workshop on Dark Matter, Neutrino Physics and Astrophysics CETUP* 2013: 7th International Conference on Interconnection between Particle Physics and Cosmology (PPC 2013): Lead/Deadwood, South Dakota, USA, July, 8-13, 2013*, *AIP Conf. Proc.* **1604**, 22 (2014).
- [49] K. K. Boddy, K. R. Dienes, D. Kim, J. Kumar, J.-C. Park, and B. Thomas, *Phys. Rev.* **D94**, 095027 (2016), [arXiv:1606.07440 \[hep-ph\]](#) .
- [50] K. K. Boddy, K. R. Dienes, D. Kim, J. Kumar, J.-C. Park, and B. Thomas, *Phys. Rev.* **D95**, 055024 (2017), [arXiv:1609.09104 \[hep-ph\]](#) .
- [51] K. R. Dienes, S. Su, and B. Thomas, *Phys. Rev.* **D86**, 054008 (2012), [arXiv:1204.4183 \[hep-ph\]](#) .

- [52] K. R. Dienes, S. Su, and B. Thomas, *Phys. Rev.* **D91**, 054002 (2015), [arXiv:1407.2606 \[hep-ph\]](#) .
- [53] K. R. Dienes, S. Su, and B. Thomas, *Proceedings, Workshop on Neutrino Physics : Session of CETUP* 2015 and 9th International Conference on Interconnections between Particle Physics and Cosmology (PPC2015): Lead/Deadwood, South Dakota, USA, July 6-17, 2015*, *AIP Conf. Proc.* **1743**, 020013 (2016).
- [54] D. Curtin *et al.*, (2018), [arXiv:1806.07396 \[hep-ph\]](#) .
- [55] D. Curtin, K. R. Dienes, and B. Thomas, *Phys. Rev.* **D98**, 115005 (2018), [arXiv:1809.11021 \[hep-ph\]](#) .
- [56] K. R. Dienes, J. Fennick, J. Kumar, and B. Thomas, *Phys. Rev.* **D97**, 063522 (2018), [arXiv:1712.09919 \[hep-ph\]](#) .
- [57] K. R. Dienes, J. Fennick, J. Kumar, and B. Thomas, *Phys. Rev.* **D93**, 083506 (2016), [arXiv:1601.05094 \[hep-ph\]](#) .
- [58] R. Hagedorn, *Nuovo Cim. Suppl.* **3**, 147 (1965).
- [59] K. R. Dienes and J.-R. Cudell, *Phys. Rev. Lett.* **72**, 187 (1994), [arXiv:hep-th/9309126 \[hep-th\]](#) .
- [60] G. H. Hardy and S. Ramanujan, *Proc. London Math. Soc.* **17**, 75 (1918).
- [61] I. Kani and C. Vafa, *Commun. Math. Phys.* **130**, 529 (1990).
- [62] K. R. Dienes, *Nucl. Phys.* **B429**, 533 (1994), [arXiv:hep-th/9402006 \[hep-th\]](#) .
- [63] Y. Nambu, unpublished (1970).
- [64] P. Ramond, *Phys. Rev.* **D3**, 2415 (1971).
- [65] A. Neveu and J. H. Schwarz, *Nucl. Phys.* **B31**, 86 (1971).
- [66] A. M. Polyakov, *Nucl. Phys.* **B268**, 406 (1986).
- [67] M. B. Green, *Phys. Lett.* **B266**, 325 (1991).
- [68] J. Polchinski and A. Strominger, *Phys. Rev. Lett.* **67**, 1681 (1991).
- [69] L. Anderson *et al.* (BOSS), *Mon. Not. Roy. Astron. Soc.* **441**, 24 (2014), [arXiv:1312.4877 \[astro-ph.CO\]](#) .
- [70] N. Suzuki *et al.* (Supernova Cosmology Project), *Astrophys. J.* **746**, 85 (2012), [arXiv:1105.3470 \[astro-ph.CO\]](#) .
- [71] R. H. Cyburt, J. Ellis, B. D. Fields, F. Luo, K. A. Olive, and V. C. Spanos, *JCAP* **0910**, 021 (2009), [arXiv:0907.5003 \[astro-ph.CO\]](#) .

- [72] W. Hu and J. Silk, *Phys. Rev.* **D48**, 485 (1993).
- [73] W. Hu and J. Silk, *Phys. Rev. Lett.* **70**, 2661 (1993).
- [74] T. R. Slatyer, *Phys. Rev.* **D87**, 123513 (2013), [arXiv:1211.0283 \[astro-ph.CO\]](#) .
- [75] L. Accardo *et al.* (AMS), *Phys. Rev. Lett.* **113**, 121101 (2014).
- [76] A.-. Collaboration (AMS), (April 15-17, 2015).
- [77] A. H. G. Peter, C. E. Moody, and M. Kamionkowski, *Phys. Rev.* **D81**, 103501 (2010), [arXiv:1003.0419 \[astro-ph.CO\]](#) .
- [78] M.-Y. Wang, A. H. G. Peter, L. E. Strigari, A. R. Zentner, B. Arant, S. Garrison-Kimmel, and M. Rocha, *Mon. Not. Roy. Astron. Soc.* **445**, 614 (2014), [arXiv:1406.0527 \[astro-ph.CO\]](#) .
- [79] M.-Y. Wang and A. R. Zentner, *Phys. Rev.* **D82**, 123507 (2010), [arXiv:1011.2774 \[astro-ph.CO\]](#) .
- [80] Y. Gong and X. Chen, *Phys. Rev.* **D77**, 103511 (2008), [arXiv:0802.2296 \[astro-ph\]](#) .
- [81] G. Blackadder and S. M. Koushiappas, *Phys. Rev.* **D93**, 023510 (2016), [arXiv:1510.06026 \[astro-ph.CO\]](#) .
- [82] S. De Lope Amigo, W. M.-Y. Cheung, Z. Huang, and S.-P. Ng, *JCAP* **0906**, 005 (2009), [arXiv:0812.4016 \[hep-ph\]](#) .
- [83] B. Audren, J. Lesgourgues, G. Mangano, P. D. Serpico, and T. Tram, *JCAP* **1412**, 028 (2014), [arXiv:1407.2418 \[astro-ph.CO\]](#) .
- [84] E. Aubourg *et al.*, *Phys. Rev.* **D92**, 123516 (2015), [arXiv:1411.1074 \[astro-ph.CO\]](#) .
- [85] S. Nussinov, *Phys. Lett.* **165B**, 55 (1985).
- [86] S. M. Barr, R. S. Chivukula, and E. Farhi, *Phys. Lett.* **B241**, 387 (1990).
- [87] S. B. Gudnason, C. Kouvaris, and F. Sannino, *Phys. Rev.* **D73**, 115003 (2006), [arXiv:hep-ph/0603014 \[hep-ph\]](#) .
- [88] T. A. Ryttov and F. Sannino, *Phys. Rev.* **D78**, 115010 (2008), [arXiv:0809.0713 \[hep-ph\]](#) .
- [89] K. Harigaya, T. Lin, and H. K. Lou, *JHEP* **09**, 014 (2016), [arXiv:1606.00923 \[hep-ph\]](#) .
- [90] T. Appelquist *et al.* (Lattice Strong Dynamics (LSD)), *Phys. Rev.* **D89**, 094508 (2014), [arXiv:1402.6656 \[hep-lat\]](#) .

- [91] T. Appelquist *et al.*, *Phys. Rev.* **D92**, 075030 (2015), [arXiv:1503.04203 \[hep-ph\]](#) .
- [92] T. Appelquist *et al.*, *Phys. Rev. Lett.* **115**, 171803 (2015), [arXiv:1503.04205 \[hep-ph\]](#) .
- [93] G. D. Kribs and E. T. Neil, *Int. J. Mod. Phys.* **A31**, 1643004 (2016), [arXiv:1604.04627 \[hep-ph\]](#) .
- [94] C. Kilic, T. Okui, and R. Sundrum, *JHEP* **02**, 018 (2010), [arXiv:0906.0577 \[hep-ph\]](#) .
- [95] T. Hur and P. Ko, *Phys. Rev. Lett.* **106**, 141802 (2011), [arXiv:1103.2571 \[hep-ph\]](#) .
- [96] M. Holthausen, J. Kubo, K. S. Lim, and M. Lindner, *JHEP* **12**, 076 (2013), [arXiv:1310.4423 \[hep-ph\]](#) .
- [97] H. Hatanaka, D.-W. Jung, and P. Ko, *JHEP* **08**, 094 (2016), [arXiv:1606.02969 \[hep-ph\]](#) .
- [98] Y. Ametani, M. Aoki, H. Goto, and J. Kubo, *Phys. Rev.* **D91**, 115007 (2015), [arXiv:1505.00128 \[hep-ph\]](#) .
- [99] Y. Bai and R. J. Hill, *Phys. Rev.* **D82**, 111701 (2010), [arXiv:1005.0008 \[hep-ph\]](#) .
- [100] M. R. Buckley and E. T. Neil, *Phys. Rev.* **D87**, 043510 (2013), [arXiv:1209.6054 \[hep-ph\]](#) .
- [101] S. Bhattacharya, B. Melić, and J. Wudka, *JHEP* **02**, 115 (2014), [arXiv:1307.2647 \[hep-ph\]](#) .
- [102] A. Carmona and M. Chala, *JHEP* **06**, 105 (2015), [arXiv:1504.00332 \[hep-ph\]](#) .
- [103] M. Frigerio, A. Pomarol, F. Riva, and A. Urbano, *JHEP* **07**, 015 (2012), [arXiv:1204.2808 \[hep-ph\]](#) .
- [104] R. Lewis, C. Pica, and F. Sannino, *Phys. Rev.* **D85**, 014504 (2012), [arXiv:1109.3513 \[hep-ph\]](#) .
- [105] A. Hietanen, C. Pica, F. Sannino, and U. I. Sondergaard, *Phys. Rev.* **D87**, 034508 (2013), [arXiv:1211.5021 \[hep-lat\]](#) .
- [106] D. S. M. Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, *Phys. Lett.* **B692**, 323 (2010), [arXiv:0903.3945 \[hep-ph\]](#) .
- [107] M. Lisanti and J. G. Wacker, *Phys. Rev.* **D82**, 055023 (2010), [arXiv:0911.4483 \[hep-ph\]](#) .

- [108] D. Spier Moreira Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, *JHEP* **06**, 113 (2010), [arXiv:1003.4729 \[hep-ph\]](#) .
- [109] G. D. Kribs, T. S. Roy, J. Terning, and K. M. Zurek, *Phys. Rev.* **D81**, 095001 (2010), [arXiv:0909.2034 \[hep-ph\]](#) .
- [110] O. Antipin, M. Redi, and A. Strumia, *JHEP* **01**, 157 (2015), [arXiv:1410.1817 \[hep-ph\]](#) .
- [111] O. Antipin, M. Redi, A. Strumia, and E. Vigiani, *JHEP* **07**, 039 (2015), [arXiv:1503.08749 \[hep-ph\]](#) .
- [112] L. B. Okun, *JETP Lett.* **31**, 144 (1980), [*Pisma Zh. Eksp. Teor. Fiz.*31,156(1979)].
- [113] L. B. Okun, *Nucl. Phys.* **B173**, 1 (1980).
- [114] A. Soni and Y. Zhang, *Phys. Rev.* **D93**, 115025 (2016), [arXiv:1602.00714 \[hep-ph\]](#) .
- [115] L. Forestell, D. E. Morrissey, and K. Sigurdson, *Phys. Rev.* **D95**, 015032 (2017), [arXiv:1605.08048 \[hep-ph\]](#) .
- [116] A. E. Faraggi and M. Pospelov, *Astropart. Phys.* **16**, 451 (2002), [arXiv:hep-ph/0008223 \[hep-ph\]](#) .
- [117] J. Halverson, B. D. Nelson, and F. Ruehle, *Phys. Rev.* **D95**, 043527 (2017), [arXiv:1609.02151 \[hep-ph\]](#) .
- [118] J. L. Feng and Y. Shadmi, *Phys. Rev.* **D83**, 095011 (2011), [arXiv:1102.0282 \[hep-ph\]](#) .
- [119] W. Detmold, M. McCullough, and A. Pochinsky, *Phys. Rev.* **D90**, 115013 (2014), [arXiv:1406.2276 \[hep-ph\]](#) .
- [120] G. Krnjaic and K. Sigurdson, *Phys. Lett.* **B751**, 464 (2015), [arXiv:1406.1171 \[hep-ph\]](#) .
- [121] J. M. Cline, Z. Liu, G. Moore, and W. Xue, *Phys. Rev.* **D90**, 015023 (2014), [arXiv:1312.3325 \[hep-ph\]](#) .
- [122] K. K. Boddy, M. Kaplinghat, A. Kwa, and A. H. G. Peter, *Phys. Rev.* **D94**, 123017 (2016), [arXiv:1609.03592 \[hep-ph\]](#) .
- [123] W. Detmold, M. McCullough, and A. Pochinsky, *Phys. Rev.* **D90**, 114506 (2014), [arXiv:1406.4116 \[hep-lat\]](#) .
- [124] E. Hardy, R. Lasenby, J. March-Russell, and S. M. West, *JHEP* **06**, 011 (2015), [arXiv:1411.3739 \[hep-ph\]](#) .

- [125] E. Hardy, R. Lasenby, J. March-Russell, and S. M. West, *JHEP* **07**, 133 (2015), [arXiv:1504.05419 \[hep-ph\]](#) .
- [126] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, *Phys. Rev. Lett.* **113**, 171301 (2014), [arXiv:1402.5143 \[hep-ph\]](#) .
- [127] E. D. Carlson, M. E. Machacek, and L. J. Hall, *Astrophys. J.* **398**, 43 (1992).
- [128] A. A. de Laix, R. J. Scherrer, and R. K. Schaefer, *Astrophys. J.* **452**, 495 (1995), [arXiv:astro-ph/9502087 \[astro-ph\]](#) .
- [129] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, *Phys. Rev. Lett.* **115**, 021301 (2015), [arXiv:1411.3727 \[hep-ph\]](#) .
- [130] M. Hansen, K. Langæble, and F. Sannino, *Phys. Rev.* **D92**, 075036 (2015), [arXiv:1507.01590 \[hep-ph\]](#) .
- [131] N. Bernal and X. Chu, *JCAP* **1601**, 006 (2016), [arXiv:1510.08527 \[hep-ph\]](#) .
- [132] A. Kamada, M. Yamada, T. T. Yanagida, and K. Yonekura, *Phys. Rev.* **D94**, 055035 (2016), [arXiv:1606.01628 \[hep-ph\]](#) .
- [133] K. K. Boddy, J. L. Feng, M. Kaplinghat, and T. M. P. Tait, *Phys. Rev.* **D89**, 115017 (2014), [arXiv:1402.3629 \[hep-ph\]](#) .
- [134] K. K. Boddy, J. L. Feng, M. Kaplinghat, Y. Shadmi, and T. M. P. Tait, *Phys. Rev.* **D90**, 095016 (2014), [arXiv:1408.6532 \[hep-ph\]](#) .
- [135] M. Freytsis, D. J. Robinson, and Y. Tsai, *Phys. Rev.* **D91**, 035028 (2015), [arXiv:1410.3818 \[hep-ph\]](#) .
- [136] H. M. Lee and M.-S. Seo, *Phys. Lett.* **B748**, 316 (2015), [arXiv:1504.00745 \[hep-ph\]](#) .
- [137] Y. Hochberg, E. Kuflik, and H. Murayama, *JHEP* **05**, 090 (2016), [arXiv:1512.07917 \[hep-ph\]](#) .
- [138] C. Englert, K. Nordström, and M. Spannowsky, *Phys. Rev.* **D94**, 055028 (2016), [arXiv:1606.05359 \[hep-ph\]](#) .
- [139] S. Bruggisser, F. Riva, and A. Urbano, *SciPost Phys.* **3**, 017 (2017), [arXiv:1607.02474 \[hep-ph\]](#) .
- [140] P. Di Vecchia and F. Sannino, *Eur. Phys. J. Plus* **129**, 262 (2014), [arXiv:1310.0954 \[hep-ph\]](#) .
- [141] A. Merle and M. Tatzauer, *JCAP* **1506**, 011 (2015), [arXiv:1502.01011 \[hep-ph\]](#) .
- [142] R. Murgia, A. Merle, M. Viel, M. Tatzauer, and A. Schneider, *JCAP* **1711**, 046 (2017), [arXiv:1704.07838 \[astro-ph.CO\]](#) .

- [143] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel, *JCAP* **0905**, 012 (2009), [arXiv:0812.0010 \[astro-ph\]](#) .
- [144] J. Lesgourgues, (2011), [arXiv:1104.2932 \[astro-ph.IM\]](#) .
- [145] D. Blas, J. Lesgourgues, and T. Tram, *JCAP* **1107**, 034 (2011), [arXiv:1104.2933 \[astro-ph.CO\]](#) .
- [146] J. Lesgourgues, (2011), [arXiv:1104.2934 \[astro-ph.CO\]](#) .
- [147] J. Lesgourgues and T. Tram, *JCAP* **1109**, 032 (2011), [arXiv:1104.2935 \[astro-ph.CO\]](#) .
- [148] M. Fairbairn and J. Zupan, *JCAP* **0907**, 001 (2009), [arXiv:0810.4147 \[hep-ph\]](#) .
- [149] B. Batell, M. Pospelov, and A. Ritz, *Phys. Rev.* **D79**, 115019 (2009), [arXiv:0903.3396 \[hep-ph\]](#) .
- [150] C. Dvorkin, T. Lin, and K. Schutz, (2019), [arXiv:1902.08623 \[hep-ph\]](#) .
- [151] K. Enqvist, S. Nadathur, T. Sekiguchi, and T. Takahashi, *JCAP* **1509**, 067 (2015), [arXiv:1505.05511 \[astro-ph.CO\]](#) .
- [152] Z. Berezhiani, A. D. Dolgov, and I. I. Tkachev, *Phys. Rev.* **D92**, 061303 (2015), [arXiv:1505.03644 \[astro-ph.CO\]](#) .
- [153] A. Chudaykin, D. Gorbunov, and I. Tkachev, *Phys. Rev.* **D94**, 023528 (2016), [arXiv:1602.08121 \[astro-ph.CO\]](#) .
- [154] K. Hamaguchi, K. Nakayama, and Y. Tang, *Phys. Lett.* **B772**, 415 (2017), [arXiv:1705.04521 \[hep-ph\]](#) .
- [155] K. L. Pandey, T. Karwal, and S. Das, (2019), [arXiv:1902.10636 \[astro-ph.CO\]](#) .
- [156] W. L. Freedman, *Nat. Astron.* **1**, 0121 (2017), [arXiv:1706.02739 \[astro-ph.CO\]](#) .
- [157] R. A. Battye, T. Charnock, and A. Moss, *Phys. Rev.* **D91**, 103508 (2015), [arXiv:1409.2769 \[astro-ph.CO\]](#) .