EXOTIC DARK MATTER IN THE EARLY UNIVERSE

by

Gregory Ridgway

B.A., University of Maryland (2017)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author .....
Department of Physics
July 27, 2022

Certified by .....
Tracy R. Slatyer
Professor of Physics
Thesis Supervisor

Certified by .....
Iain Stewart
Professor of Physics
Thesis Supervisor

Accepted by .....
Deepto Chakrabarty
Associate Department Head of Physics

EXOTIC DARK MATTER IN THE EARLY UNIVERSE

by

Gregory Ridgway

Submitted to the Department of Physics
on July 27, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The fundamental nature of dark matter remains a mystery. As crucial evidence for its existence comes from its effects on early universe observables, perhaps clues to its fundamental nature also reside in its behavior in the early universe. In this thesis, I explore two scenarios in which dark matter is capable of exotic behavior in the early universe. In the first scenario, I consider particle dark matter that is able to decay or annihilate into standard model matter. I describe a code package, `DarkHistory`, that quickly and accurately calculates the effects such annihilations and decays have on the evolution of the ionization levels, matter temperature, and spectrum of photons in the early universe. I then use `DarkHistory` and measurements of the Lyα forest to place constraints on the decay lifetime and annihilation rates of dark matter. In the second scenario, I consider dark matter that consists of dark quarks and gluons. In the specific model I consider, the confinement phase transition is of first order, leading to the formation of bubbles. The dynamics of these bubbles and their interactions with the dark quarks dramatically modifies their present-day abundance.

Thesis Supervisor: Tracy R. Slatyer
Title: Professor of Physics

Thesis Supervisor: Iain Stewart
Title: Professor of Physics

Acknowledgments

It took a village to get me to this point, so I have many people to thank. My advisors Tracy Slatyer and Iain Stewart for their encouragement, mentorship, and incredible investment into my academic well-being. Hongwan Liu, for taking on an advisor role at such a crucial early stage, investing so much time and effort into me, and making perhaps the biggest difference in my grad school career. Krishna Rajagopal, for his sage-like guidance. Jesse Thaler, for being a perfect academic advisor. Scott Morley and Charles Suggs for consistently going above and beyond and making everyone's lives in the CTP better. My collaborators Wenzer Qin, Pouya Asadi, Siddharth Mishra-Sharma, Rebecca Leane, Ryan Weller, Yi Yin, Sanjay Raman, Anjie Gao, Ian Moult, Kim Boddy, Vivian Poulin, Ningqiang Song, Erik Kuflik, Erik David Kramer, Juri Smirnov, Tomer Volansky, Chen Sun, Sida Lu, and Nadav Outmezguine for their incredible work and inspiration. My other group mates, Patrick Fitzpatrick, Chih-Liang Wu, Nick Rodd, Stella Schindler, Gherardo Vita, Markus Ebert for the camaraderie. My office mates, Jasmine Brewer, Andrew Turner, Sarah Geller, Annie Wei, Sam Alipour-Fard, Elba Alonso-Monsalve, for putting up with my mess. Paulo Bedaque, Andrei Alexandru, Neill Warrington, Gokce Basar, for showing me how much fun physics research could be. Rashmish Mishra, for being my Rasputin, but in a good way. Cyrus Ishikawa, for lighting the spark and breathing new life into it every chance he gets. Alec Ridgway, for always being three steps ahead of me and sharing with me the wisdom that he gained. My friends, whose impact on my life is both profound and inarticulable, and should not read too far into the fact that they remain nameless here (there are too many of you to write down). And my family, for being my foundation through it all.

Contents

Chapter 1

Introduction

Dark matter is a great puzzle. First, it is real. It is responsible for important phenomena in our universe, including the formation of galaxies like our own and the statistical properties of radiation from the Big Bang. Second, it is mysterious. It could consist of undiscovered particles so light that their wavelengths extend astronomical distances, or conventional matter so heavy that it forms black holes. Therefore, when one puzzles over dark matter, they can feel satisfied that they are working towards understanding a substance real enough to actually exist in our universe, but mysterious enough to warrant full use of their imagination when trying to figure out what dark matter is.

Any solution to the dark matter puzzle would likely have important consequences. If dark matter consists of a new particle or field, or a collection of them, then the new particles or fields would have to be included in the Standard Model of Particle Physics. Dark matter could then provide solutions to outstanding problems like baryogenesis [1], the hierarchy problem (see Weakly Interacting Massive Particles [2]), or the Strong-CP problem (see the QCD axion [3]). If instead of a new particle or field, dark matter consists of primordial black holes [4, 5], then its formation mechanism could shed light on the inflationary epoch. Even falling short of identifying what dark matter is, but learning that it interacts non-gravitationally with itself, or interacts non-gravitationally with Standard Model matter, or has the ability to dissipate energy, could lead to a deeper understanding of structure formation and galactic dynamics.

1.1 Dark Matter Overview

Observations of light from space are key to dark matter research. Stars, gas, galaxies, and the hot plasma created after the Big Bang all emanate light that encodes information about dark matter. Through experiments that carefully collect this light and analyze it, physicists have pinned down important properties of dark matter, namely that it is gravitationally-interacting, cold, stable, abundant, and has no currently detected non-gravitational interactions. These properties taken together define the Cold Dark Matter (CDM) paradigm.

1.1.1 Evidence

To establish the CDM paradigm, the physics community performed a large variety of experiments and observations (for a more thorough review, see [6] or Chapter 27 of [7]). All of these experiments and observations find evidence of a non-luminous substance existing in and beyond galaxies, modifying the gravitational potential in a number of observable ways. These observations are so numerous, take place over such a wide range of length and time scales, and are so consistent with a single dark matter interpretation that they place the dark matter hypothesis on a very strong footing.

At the smallest scales, it has been observed within many galaxies that stellar rotation curves flatten well beyond the bulk of visible matter [8–11]. These observations point to the existence of large amounts of invisible matter that stretch well beyond the visible galaxy and tug on gravitationally bound stars. Not just the coherent rotational motion of stars, but also the random motion of stars in dwarf galaxies suggests that there is a lot more gravitational pull on these stars than can be accounted for by luminous matter, and can thus be attributed to dark matter [12]. In addition to stars, the distribution and properties of gas in large galaxies and galaxy clusters determine the gravitational potential of these galaxies, and indicate that they contain about 5 times more matter than is visible [13].

Dark matter does not just tug on matter, but also on light. For example, studies

of galaxy collisions like the Bullet Cluster (left panel of Fig. 1-1) show a large separation between X-ray emitting gas near the collision area and non-luminous matter well beyond the collision, the latter of which is only perceptible by its ability to gravitationally lens background light [14]. The interpretation of the Bullet Cluster is that after the collision of the galaxies, which both contain gas and dark matter, the gas slowed down due to its ability to dissipate energy, but the dark matter passed right through due to its inability to interact non-gravitationally. More general analyses of lensing within galaxies, galaxy clusters, and large-scale structure all indicate that there is a more abundant component of dark matter than luminous matter throughout the universe [15].

At the largest scales, dark matter plays an important role in the formation of gravitationally bound structures in our universe [16]. As opposed to standard model matter, for which electromagnetic interactions resist the tendency to clump under the influence of gravity, dark matter clumps efficiently very early on in the history of the universe. These clumps provide potential wells in which tiny initial matter overdensities are able to gradually grow into the galaxies and clusters we observe today. The statistical properties of these clumps then become imprinted on the statistical properties of present-day galaxies, allowing for an inference of the gravitational properties of dark matter [17]. These clumps also affect the statistical properties of the radiation bath, the Cosmic Microwave Background (CMB). By measuring the deviations of the CMB intensity from the mean and correlating different points along the sky at different angular scales, one can construct the right plot of Fig. 1-1. From this detailed measurement, we have learned that dark matter has been around for over 13 billion years, and accounts for about 84.4% of the mass contained in the universe [18].

These observations teach us valuable lessons about dark matter. We learn that dark matter accounts for most of the mass in the universe. We surmise that it is stable, since the fraction of dark matter remains fixed for so much of cosmic history. We also infer that dark matter interacts with matter and radiation gravitationally, and that any non-gravitational interactions are either non-existent or too feeble to be observed.

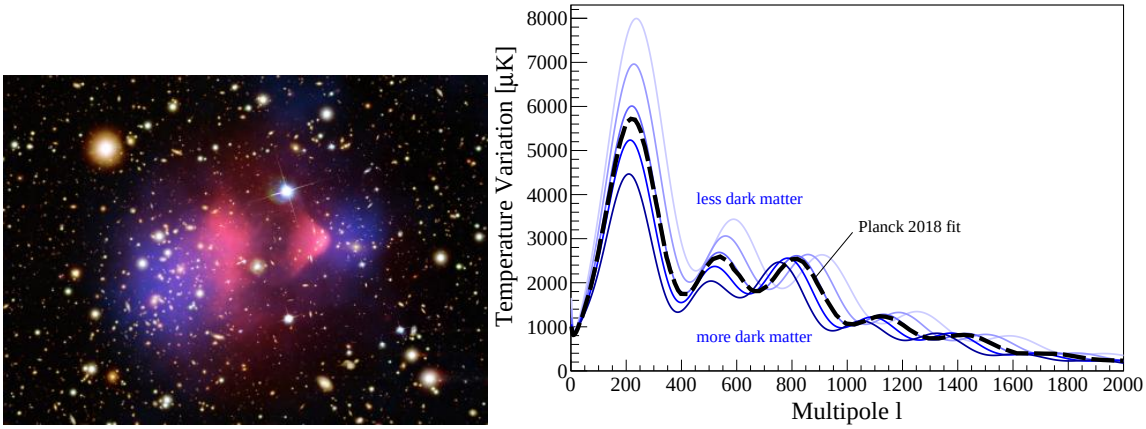


Figure 1-1: Two key pieces of evidence that establish the existence of dark matter. *Left*: The Bullet Cluster, 1E0657-56. (Figure credit: Left: Schumann et al.[19]; Right: NASA/CXC/CfA/ M.Markevitch et al.) *Right*: The CMB TT anisotropy power spectrum for dark matter densities $0.11 \leq \Omega_c \leq 0.43$, with all other cosmological parameters held constant. The best fit to the 2018 Planck data [18] is shown (black, dashed).

1.1.2 Deviations from CDM

If dark matter had no non-gravitational interactions, it would be difficult to learn what it actually is. We would not be able to capture it and send it through a magnetic field to learn about its intrinsic spin, as we can with electrons. We could not shine light on dark matter and search for deflection or absorption to learn about its characteristic size and electric dipole moment, as we can with hydrogen. Nor could we produce it in a collider to learn about its mass and interactions with other particles, as we can a W-boson.

On the other hand, non-gravitational interactions open the prospect of very informative signatures. If dark matter could collide with standard model matter, one could fill a container with a large density of targets and search for evidence of collisions between targets and dark matter traveling through our galaxy, motivating the field of direct detection. If dark matter could convert to standard model matter or vice versa, then one could try to produce it at a collider, or look for its products in space, motivating the field of indirect detection. In the first part of this thesis, I consider the indirect detection scenario by supposing that dark matter can convert

to standard model matter, and investigate the effects of its products in the early universe.

Non-gravitational interactions also open the possibility for an explanation of the dark matter abundance that we observe today. For a classic example, consider 2-to-2 thermal freeze-out [20]. Let us assume that dark matter is made of particles, denoted χ, and that there exists an interaction that can convert a pair of dark matter particles to a pair of standard model particles or vice versa, $\chi\,\chi \leftrightarrow SM\ SM$, with cross-section σ. Then by integrating the Boltzmann equation for dark matter over all phase space variables, we find that its number density evolves according to

$$\dot{n}_\chi = -3Hn_\chi - \langle\sigma v\rangle \left(n_\chi^2 - \left(n_\chi^{\rm eq}\right)^2\right) , \tag{1.1}$$

where v is the relative velocity between colliding particles in the 2-to-2 process and is averaged against the cross-section. The first term on the right accounts for the dilution of the number density due to the expansion of space, where the expansion occurs at the Hubble rate, H. The $\langle\sigma v\rangle n_\chi^2$ term is the dark matter particle loss rate per unit volume as the interaction converts χ particles to standard model particles, while the remaining term is the inverse rate that replenishes the dark matter number density, which has been calculated using detailed balance. At early times, densities are high, making $\langle\sigma v\rangle n_\chi^2$ large compared to the Hubble term. We can neglect the Hubble term, leaving us with an equation that drives n_χ to $n_\chi^{\rm eq}$ – when $n_\chi > n_\chi^{\rm eq}$ the right hand side is negative, causing n_χ to decrease down towards $n_\chi^{\rm eq}$; when $n_\chi < n_\chi^{\rm eq}$ the right hand side is positive, causing n_χ to increase up towards $n_\chi^{\rm eq}$. We thus find that at early enough times the number of dark matter particles is completely determined by the common temperature T that all particles in the early universe share. As the universe steadily becomes more dilute, the interaction rate becomes less than the Hubble rate, making the latter two terms negligible. The comoving number of dark matter particles stays essentially fixed, neglecting the residual amounts of annihilations that still occur, and is no longer determined by T. Dark matter is said to have "frozen out" at this point. By adjusting the interaction strength, $\langle\sigma v\rangle$, one can

arrange for the observed abundance of dark matter seen today. It is important to note that although freezeout means that the dark matter density is no longer significantly affected by annihilations, these annihilations are still important. As I will explain further below and in Chapters 2 and 3, for a wide range of interaction strengths, the residual amounts of annihilation that take place are still large enough to significantly heat or ionize matter in the universe.

2-to-2 thermal freezeout is far from the only way to achieve the correct dark matter abundance. There are spinoffs, like 3-to-2 thermal freeze-out [21, 22] or freeze-in [23], and completely different ideas, like gravitational production during inflation [24], the misalignment mechanism for axion dark matter [25], or primordial black hole (PBH) production for PBH dark matter [5]. In the second half of this thesis, I consider an extension of the thermal freeze-out scenario in which freeze-out initially produces too much dark matter, but then a first-order phase transition compresses the dark matter, effecting a second stage of annihilation that depletes the dark matter down to the correct abundance.

As a final note, typically within indirect detection there are two strategies, the model-independent and the model-dependent strategy. The model-independent strategy is concerned with generic types of interactions that dark matter may have, like self-interactions, energy dissipation, or collisions with nuclei. The model-dependent strategy is concerned with specific forms dark matter can take, like axion dark matter, WIMP dark matter, or sterile neutrino dark matter. Each strategy comes with its own advantages and disadvantages, and the two are complementary to one another. In the first half of this thesis I will take a more model-independent approach. The type of interaction I consider is annihilation or decay of dark matter to standard model matter. In the second half, I will take a more model-dependent approach. I will consider a model in which dark matter is comprised of dark quarks and gluons.

Figure 1-2: Types of interactions between dark matter and standard model matter. To the left is a 3-point interaction, which can lead to dark matter decay to standard model particles. To the right is a 4-point interaction, which can lead to dark matter annihilation to standard model particles. Both annihilation and decay give rise to dark matter energy in the early universe.

1.2 Annihilating and Decaying Dark Matter

In the first part of this thesis I explore the effects of dark matter conversion into standard model matter in the early universe. This type of interaction is a common feature of many models. To see why, let us assume that dark matter consists of particles, as I will throughout the rest of this thesis. Quantum Field theory then tells us that the interactions between dark matter and standard model particles can be described by n-point amplitudes, with at least one dark matter particle leg and at least one standard model particle leg. At the lowest non-trivial order is a three-point amplitude. If this amplitude has one dark matter leg and two standard model legs, as depicted in the left half of Fig. 1-2, then this interaction can give rise to dark matter decays. At the next order is a four-point amplitude. If it involves two dark matter particles and two standard model particles, as depicted in the right half of Fig. 1-2, then this type of interaction gives rise to annihilation. Different combinations can exist and have been explored (see [21, 26]), but in this thesis I restrict to decays and 2-to-2 annihilations.

1.2.1 Energy Injection

If dark matter has a decay rate of Γ, then its comoving number density evolves according to

$$\frac{d}{dt}\left(a^3 n_\chi\right) = -\Gamma\, n_\chi\,. \tag{1.2}$$

I have assumed that the universe is dilute enough that the inverse process in which the decay products combine to make a dark matter particle is negligibly small compared to the decay process. If we only consider regimes in which dark matter behaves like non-relativistic matter, then its energy density is given by $\rho_\chi = m_\chi n_\chi$. The amount of energy released by dark matter decay per volume and time is then given by

$$\left(\frac{dE}{dV\,dt}\right)_{\text{inj, decay}} = \frac{m_\chi n_\chi}{\tau}\,. \tag{1.3}$$

where I have defined the lifetime as $\tau = \Gamma^{-1}$.

If dark matter annihilates with a cross-section of σ, then the number of dark matter particles that annihilate per volume is $n_\chi^2 \langle\sigma v\rangle$. In this analysis, I assume that the two annihilating χ particles are indistinguishable, but the distinguishable case proceeds similarly. The energy injection rate is then

$$\left(\frac{dE}{dV\,dt}\right)_{\text{inj, ann}} = \rho_\chi^2 \frac{\langle\sigma v\rangle}{m_\chi},\,. \tag{1.4}$$

The velocity averaged cross-section, $\langle\sigma v\rangle$, can be simplified further. Again, by assuming that dark matter is non-relativistic, we can expand the velocity averaged cross-section in powers of the root-mean-square dark matter velocity, v_{rms} [27],

$$\langle\sigma v\rangle = b_0 + b_1 v_{rms}^2 + \ldots\,. \tag{1.5}$$

s-wave annihilation refers to the case when the first term dominates while p-wave annihilation refers to the case when the second dominates. In Chapter 3 I will constrain models of p-wave annihilating DM. Note that the above assumes a contact

interaction; with a long-range interaction we could have v^{-1} terms.

p-wave annihilation occurs for various specific dark matter models like, for example, scalar millicharged dark matter [28]. To write the p-wave velocity-averaged cross-section in a slightly more convenient way, evaluate it at a reference velocity, v_{ref}, which I will take to be $100\,\mathrm{km\,s^{-1}}$ throughout this thesis. We then have a reference value for the cross-section, $(\sigma v)_{ref} = b_1 v_{ref}^2$, which allows us to rewrite b_1. The p-wave cross-section at any velocity is then $\langle \sigma v \rangle = \frac{(\sigma v)_{ref}}{v_{ref}^2} v_{rms}^2$.

To calculate v_{rms} and ρ_{DM}, we must consider two regimes, the high-redshift regime and the low-redshift regime. In the high-redshift regime, matter over- and under-densities are small and the density of dark matter is homogeneous to a good approximation. After freezeout, the energy density evolves solely due to Hubble expansion, $\rho_{DM}(z) = \rho_{DM,0}(1+z)^3$, where $1+z$ is the redshift factor of the universe and $\rho_{DM,0}$ is the measured dark matter energy density today. Also in the post-freezeout, high-redshift regime, the dark matter velocity distribution can be taken to be Maxwell-Boltzmann, with a temperature that scales with two powers of redshift after kinetic decoupling as non-relativistic matter must in the expanding universe [29]. Therefore, $v_{rms}^2 = \frac{3}{2}\frac{T}{m_{DM}} \propto (1+z)^2$.

In the low-redshift regime, structure formation is important. The density of dark matter can no longer be taken to be homogeneous. I now show that the dark matter inhomogeneities lead to boost factors in the energy injection rates for annihilation, but not decay [30]. As time progresses, more and more dark matter falls into potential wells, wells that become deeper with time. p-wave annihilation, for example, gets boosted by two effects, the increase in ρ_χ^2 as gravitational attraction increases dark matter densities, and the increase in v_{rms} as gravitational potential energy converts to kinetic energy. s-wave annihilation also gets a boost due to the increase in ρ_χ^2. Hence, higher density regions become the dominant sources of energy injection. In reality, then, energy injection should be treated as an inhomogeneous process, as is done in Refs. [31, 32]. Instead, in this thesis I will simplify my calculations by assuming that energy injection can be considered homogeneous. This assumption could be justified if one showed that the injected particles had long mean free paths compared to the

typical distance between overdensities. Then, since the overdensities are distributed homogeneously, the particles injected from different overdensities would eventually mix and become homogeneous again.

This reasoning explains why there is no boost factor for decay. For every overdense region with increased decay there is an underdense region with decreased decay. If we assume that the injected particles quickly homogenize, then we can calculate the energy injection rate by placing the average dark matter density in Eq. 1.3, and forgetting about the inhomogeneities altogether. Otherwise put, my assumption that the injected particles quickly homogenize translates to an averaging procedure of the energy injection rate over space. Since Eq. 1.3 is linear in its fluctuating variable n_χ, the averaging procedure merely replaces n_χ by the average dark matter number density. On the other hand, the s-wave annihilation rate is quadratic in its fluctuating variable ρ_χ^2, and the p-wave is quartic in its fluctuating variables $\rho_\chi^2 v_{rms}^2$. The averaging procedure of Eq. 1.4 then produces boost factors relative to their averages, $\rho_\chi^2 \to \langle \rho_\chi^2 \rangle = \langle \rho_\chi \rangle^2 \mathcal{B}_s$ and $\rho_\chi^2 v_{rms}^2 \to \langle \rho_\chi^2 v_{rms}^2 \rangle = \langle \rho_\chi \rangle^2 \langle v_{rms} \rangle^2 \mathcal{B}_p$. We will find that the boost factors become important at redshifts $z \lesssim 50$, and that for unconstrained models, the p-wave annihilation rate is essentially negligible up until that point. In the rest of this thesis I will leave the averaging symbols $\langle \cdot \rangle$ for ρ and v_{rms} implicit.

In summary, the total homogeneous energy injection rate is

$$\left(\frac{dE}{dV\,dt}\right)^{\text{inj}} = \begin{cases} \rho_{\chi,0}(1+z)^3/\tau\,, & \text{decay}, \\ \rho_{\chi,0}^2(1+z)^6\langle\sigma v\rangle/m_\chi\,\mathcal{B}_s\,, & s-\text{wave annihilation}, \\ \rho_{\chi,0}^2(1+z)^8(\sigma v)_{ref}/m_\chi \frac{v_{rms}^2}{v_{ref}^2}\,\mathcal{B}_p\,, & p-\text{wave annihilation}, \end{cases} \tag{1.6}$$

where $\rho_{\chi,0}$ is the present day energy density of DM.

1.2.2 Energy Deposition

We have determined how much raw energy is injected into the universe in the form of standard model particles in either scenario, but this energy is quickly processed in

complicated ways. The energy can be injected in the form of any pair of particles – quark and anti-quark, electron and positron, neutrino and anti-neutrino, photons, or other gauge bosons. If any of these particles are unstable, as the heavy quarks and heavy leptons are, or if they hadronize, as the gluons and light quarks do, they resolve into a cascade of stable particles. The decay and hadronization timescales are always fast compared to other timescales, so we can assume that the conversion to stable particles happens instantaneously.

The stable particles that are produced are electrons, positrons, photons, neutrinos, protons, and other light nuclei. The neutrinos interact so weakly that we can ignore them, treating them solely as a form of energy loss. The nuclei larger than a proton are produced in negligible amounts, so we neglect them, too. The protons and anti-protons, are typically produced in smaller abundances [33] and interact more weakly than other electromagnetically interacting particles, but are not completely negligible [34]. As a first approximation, I ignore protons and anti-protons, taking the philosophy that the complications they add to my calculations do not justify the small corrections they add to my results. Moreover, the effect of adding protons and anti-protons is likely one-sided – adding them would just increase the amount of energy deposited. We are then left with electrons, positrons, and photons.

The injected electrons, positrons, and photons are typically initially at much higher energies than the thermal bath of photons and electrons present in the early universe. Therefore, these particles cool, depositing energy into a number of different channels as they thermalize. The fraction of energy deposited into each channel is summarized by deposition fraction functions,

$$f_c = \left(\frac{dE}{dV dt}\right)_{\text{dep, c}} \Bigg/ \left(\frac{dE}{dV dt}\right)_{\text{inj}} . \tag{1.7}$$

The channels, c, include ionization and excitation of atoms, heating of the cooler thermal baths, or a distortion to the photon blackbody spectrum. These fractions are very convenient, as they condense all the information that existed in the complicated cascade of injected particles into a set of simple functions.

1.2.3 Temperature and Ionization Evolution

With $\left(\frac{dE}{dV\,dt}\right)_{\rm inj}$ and f_c in hand, we can start exploring the effects that dark matter energy injection has on the evolution of the early universe. Given the channels – excitation, ionization, heating, and distortion – one may expect there to be four sets of evolution equations that should be solved, one for the excitation levels of atoms, another for ionization levels, another for temperatures of the different thermal baths, and another for the spectrum of photons. Fortunately, transition rates between different excited states are very fast compared to cosmological timescales, and the excitation levels quickly achieve a steady-state. The excitation levels can then be calculated via a set of steady-state conditions [35, 36]. Even these steady-state conditions usually do not need to be solved since all states but the ground state are negligibly populated.

The ionization equations, on the other hand, are important. Although the time scales for ionization and recombination of atoms are also fast compared to cosmological timescales, the *net* transition rate is comparable, and eventually slower. However, since all atomic transition rates are so fast, the ionization equations take a relatively simple form. Consider the hydrogen ionization equation [30, 37, 38],

$$\begin{aligned}\dot{x}_{\rm HII} =& -\mathcal{C}\left[n_{\rm H}x_e x_{\rm HII}\alpha_{\rm H} - 4(1-x_{\rm HII})\beta_{\rm H}e^{-E_\alpha/T_{\rm CMB}}\right] \\ &+ \left[\frac{f_{\rm H\ ion}}{\mathcal{R}n_{\rm H}} + \frac{(1-\mathcal{C})f_{\rm exc}}{E_\alpha n_{\rm H}}\right]\left(\frac{dE}{dV\,dt}\right)^{\rm inj} \\ &+ \dot{x}^{\rm re}_{\rm HII}\,, \end{aligned} \tag{1.8}$$

where $\mathcal{R} \approx 13.6\,\text{eV}$ is the Rydberg energy, $E_\alpha = 3\mathcal{R}/4$ is the Lyman-α excitation energy, $n_{\rm H}$ is the density of hydrogen atoms (bound or unbound), and $T_{\rm CMB}$ is the CMB temperature. $x_i = n_i/n_{\rm H}$ is a density normalized per hydrogen atom. Specifically, $x_e = n_e/n_{\rm H}$ where n_e is the density of unbound electrons and $x_{\rm HII} = n_{\rm HII}/n_{\rm H}$ where $n_{\rm HII}$ is the density of ionized hydrogen atoms. I use the convention where the roman numeral attached to an atomic symbol denotes the degree to which an atom is ionized: HI is neutral hydrogen, HII is ionized hydrogen, HeI is neutral helium, HeII is singly-ionized helium, and HeIII is doubly-ionized helium.

Due to the separation of atomic and cosmological timescales, all complications introduced by the atomic structure of the hydrogen atom, with its infinite tower of excited states, has been condensed into the Peebles-C factor, the recombination coefficient $\alpha_{\rm H}$, and the ionization coefficient, $\beta_{\rm H}$. The deposition functions appear in a simple way, with the hydrogen ionization and excitation coefficients each appearing as separate ionization terms. The f_{Hion} term appears as a direct ionization term, while the f_{exc} term appears with an extra factor of $(1 - \mathcal{C})$ to account for the net fraction of atoms excited by dark matter energy injection that end up fully ionizing. The last term accounts for the effects of reionization and will be the subject of Ch. 3.

The temperature evolution equation is also important. Naively, one may have expected multiple temperature evolution equations, one for each type of matter and a separate one for the radiation, but in reality there is only one non-trivial equation. Let us start with the CMB temperature. Recalling that for every one baryon there are about 2 billion photons [18], we see that the heat capacity of the CMB is much larger than the matter heat capacity. Therefore, even though electrons are able to act as a conduit through which heat can flow from the CMB to the matter bath, to a great approximation, T_{CMB} remains unaffected by any heat transfer. The evolution of T_{CMB} is then given by $T_{CMB}(z) = T_{CMB,0}(1 + z)$, where $T_{CMB,0}$ is the CMB temperature as measured today, as necessary for a radiation bath in an expanding universe. Now we move on to the dark matter temperature. If the annihilation or decay rate were comparable to the expansion rate, then heat exchange between dark matter and standard model matter would be important. Fortunately, we will find that constraints are stringent enough that annihilation and decay rates must be small. Since I will assume that there are no additional interactions between dark matter and standard model matter throughout this thesis, there is no thermal coupling between the two baths. Since the dark matter temperature is unobservable, we can then ignore it.

Finally, we turn to the matter temperature equation. It turns out that ions, electrons and neutral atoms exchange energy with each other and among themselves quicker than all other energy exchange rates in the problem [37]. All species of

matter can then be assumed to be at a common, well-defined temperature, T_m. This temperature evolves according to

$$\dot{T}_m = -2HT_m + \Gamma_C(T_{\rm CMB} - T_m) + \frac{2f_{\rm heat}}{3(1+\mathcal{F}_{\rm He}+x_e)n_{\rm H}}\left(\frac{dE}{dV\,dt}\right)^{\rm inj} + \dot{T}_m^{\rm re}\,, \qquad (1.9)$$

where Γ_c is the Compton cooling rate (defined in Chapter 2), $\mathcal{F}_{\rm He} = n_{\rm He}/n_{\rm H}$ is the fraction of helium atoms per hydrogen atoms. The first term cools matter due to the expansion of the universe. The second allows heat to exchange between the photon bath and the matter bath. The third is a source of heat due to dark matter energy injection. Finally, the fourth is a source of heat due to the sources that drive reionization.

The last evolution equation we must consider is that of the distortion to the CMB spectrum. Over a wide range of frequencies, the spectrum of CMB photons has been measured to be a perfect blackbody spectrum to within one part in 10^4 precision [39]. As injected electrons scatter against CMB photons and injected photons cool down and ultimately add to the CMB, this spectrum of photons can be distorted away from a perfect blackbody. While this distortion is an important effect and will be the subject of a future upgrade of `DarkHistory`, it will not be a focus of this thesis.

In Chapter 2 I will explain how `DarkHistory` effectively organizes all steps of this calculation. `DarkHistory` calculates the f_c functions for arbitrary energy injection scenarios, including decay and s- or p- wave annihilation to arbitrary spectra of photons or e^+e^- pairs. Using these f_c functions, it is able to quickly and accurately compute the ionization levels and temperature of the universe as a function of time. In Chapter 3 I will use `DarkHistory` to calculate how decay and p-wave annihilation modifies the temperature of matter at low redshifts. Using measurements of the matter temperature at low temperatures coming from observations of the Lyman-α forest, I constrain dark matter decay lifetimes and annihilation cross-sections.

1.3 Strongly Interacting Dark Matter

The standard model often serves as inspiration for dark matter models. The hope is that since nature has already set a precedent by allotting the Standard Model with certain types of particles and interactions, maybe it allotted similar particles and interactions to the dark sector. For example, authors have taken this philosophy and constructed models of dark photons [40], dark atoms [41], dark pions [42] – even an entire dark Standard Model [43]. In this thesis, I explore one such model inspired by Quantum Chromodynamics (QCD).

1.3.1 QCD-like Dark Matter

To begin, let us take inspiration from the Standard Model and guess that dark matter can be described by a gauge theory. Rather than the complicated gauge group of the Standard Model, $SU(3)\otimes SU(2)\otimes U(1)$, let us consider $SU(N_c)$ as a dark gauge group, where N_c will denote the number of dark quark colors. We have therefore specified that there are dark gauge bosons in our dark sector, which we will call dark gluons by analogy to QCD. At this point we could construct a valid theory of dark matter by ensuring that the stable glueballs in this theory constitute the dark matter [44], but instead let us continue our analogy with QCD and include dark quarks in our model. Let us assume that these dark quarks are in the fundamental representation of $SU(N_c)$. Unlike the standard model, we will assume that there is only one flavor of quark, and that it is stable. With the particle content we have specified so far, there are only two parameters that I need to fix, the interaction strength α and the quark mass m_q.

To fix α, I must first specify how it runs. At an energy scale μ, where $N_f(\mu)$ flavors contribute to the beta function, we have at leading order (see Chapter 9 of Ref. [7])

$$\alpha(\mu) = \frac{12\pi}{(11N_c - 2N_f(\mu))\log\frac{\mu^2}{\Lambda^2}} \tag{1.10}$$

α is then fully determined once Λ is specified. Importantly, we have chosen $N_f(\mu) \leq 1$,

meaning that $\alpha(\mu)$ gets weaker at higher energy scales. In other words, our dark sector is asymptotically free, just like QCD is. We can therefore interpret Λ as the confinement scale. The cosmological implication for asymptotic freedom is that when the dark sector is hot and dense enough in the early universe, with a temperature $T \gg \Lambda$, the quarks and gluons will have a much higher energy than Λ on average. Their interactions will not be strong enough to keep them in bound states, meaning that they will exist in an unbound Quark Gluon Plasma. At the other extreme, when the dark sector cools due to Hubble expansion and $T \ll \Lambda$, the quarks and gluons will have a much lower energy than Λ on average. Their interactions will be very strong causing them to confine, binding into color neutral bound states. At some intermediate temperature, there will then be a phase transition.

With α specified, we now move on to the quark mass, m_q. When specifying m_q, the important question is: how large is it compared to Λ? If we were to choose m_q comparable to or less than Λ, then many processes would be in the non-perturbative regime, and therefore too difficult to calculate without considerable aid from lattice simulation. Choosing $m_q \gg \Lambda$ instead not only fixes this problem, but also makes the cosmology of this model very interesting. The confining phase transition becomes a first-order phase transition [45], meaning that there will be an era of bubble formation and complicated bubble dynamics. In fact, my main motivation for considering $m_q \gg \Lambda$ will be to explore the interesting effects such a phase transition has on the evolution of this dark sector in the early universe.

1.3.2 The Heavy Quark Regime

Let us end this section by better appreciating how the heavy quark regime simplifies the analytic and numerical aspects of our problem. Many of the analytic simplifications stem from the fact that the heavy quarks can be taken to be in the non-relativistic limit. As a result, the entire 2-to-2 freezeout discussion leading up to Eq. 1.1, which assumed non-relativistic dark matter, can be applied to the quark, anti-quark to gluon annihilation process, $q\bar{q} \leftrightarrow gg$. Furthermore, we can use a non-relativistic scattering formalism to compute $\langle \sigma v \rangle$. Using the inter-quark potential

calculated on the lattice [46], we can calculate $\langle \sigma v \rangle$ and find an expression similar to that of Eq. 1.5, with additional corrections due to the long-range nature of the inter-quark force, and also due to the interactions with the surrounding plasma [47]. The entire process of quark freezeout is then calculable in the heavy quark limit. When we consider the phase transition, we will be able to neglect pair production of quarks. We will also be able to treat the quarks in an almost Newtonian manner, considering how its velocity (which is only well-defined in the non-relativistic limit) evolves due to the forces exerted on it by the bubbles.

In addition to analytic calculations, numerical calculations using lattice QCD are easier in the heavy-quark limit. In this limit, the most resource intensive part of the lattice calculation, the fermion determinant, can be neglected. As a result, many important numerical results are available from the lattice. First, the potential between two quarks and the potential between a quark and an anti-quark have been determined at any temperature [46]. This information is necessary for the calculation of $\langle \sigma v \rangle$. Second, the two most important non-perturbative quantities determining the bubble dynamics during the first-order phase transition have been calculated: the surface tension σ and the latent heat, l [48–52]. The surface tension determines the amount of energy per unit area it takes to form an interface between the two phases co-existing in the phase transition. The latent heat determines how much heat is released as the high energy phase converts to the lower energy phase.

In Chapter 4 I investigate this QCD-like dark matter model in more detail, focusing on its early universe evolution and its first-order phase transition in particular. I show that the phase transition leads to a second stage of dark matter annihilation, dramatically modifying the dark matter relic abundance calculation.

Chapter 2 was published as [53] in collaboration with Hongwan Liu and Tracy Slatyer. Chapter 3 was published as [54] in collaboration with Hongwan Liu, Wenzer Qin, and Tracy Slatyer. Chapter 4 was published as [55] with a companion paper [56] in collaboration with Pouya Asadi, Eric Kuflik, Eric David Kramer, Tracy Slatyer, and Juri Smirnov.

Chapter 2

`DarkHistory`: A Code Package for Calculating Modified Cosmic Ionization and Thermal Histories with Dark Matter and Other Exotic Energy Injections

2.1 Introduction

Dark matter annihilation or decay and other exotic sources of energy injection can significantly alter the ionization and temperature histories of the universe. In this chapter we describe a new public code package, `DarkHistory`, that allows fast and accurate computation of these possible effects of exotic energy injection on astrophysical and cosmological observables.

In particular, we will focus on interactions that allow dark matter (DM) to decay or annihilate into electromagnetically interacting Standard Model particles. This case has been studied extensively in the literature: stringent constraints on the dark matter annihilation cross section and decay lifetime have been derived from the way

these Standard Model products would distort the anisotropies of the cosmic microwave background (CMB) [57–60], or increase the temperature of the Inter-Galactic Medium (IGM), consequently affecting 21-cm and Lyman-α line emission [30, 61–63].

`DarkHistory` facilitates the calculation of these observables and the resulting constraints. In particular, `DarkHistory` makes the temperature constraint calculations significantly more streamlined, self-consistent, and accurate. It has a modular structure, allowing users to easily adjust individual inputs to the calculation – e.g. by changing the reionization model, or the spectrum of particles produced by dark matter annihilation/decay. Compared to past codes developed for such analyses [64], `DarkHistory` has a number of important new features:

- the first fully self-consistent treatment of exotic energy injection. Exotic energy injections can modify the evolution of the IGM temperature $T_{\rm IGM}$ and free electron fraction x_e, and previously this modification has been treated perturbatively, assuming the backreaction effect due to these modifications on the cooling of injected particles is negligible. This assumption can break down toward the end of the cosmic dark ages for models that are not yet excluded [30]. `DarkHistory` solves simultaneously for the temperature and ionization evolution and the cooling of the injected particles, avoiding this assumption;

- a self-contained treatment of astrophysical sources of heating and reionization, allowing the study of the interplay between exotic and conventional sources of energy injection;

- a large speed-up factor for computation of the full cooling cascade for high-energy injected particles (compared to the code employed in e.g. [30]), via precomputation of the relevant transfer functions as a function of particle energy, redshift and ionization level;

- support for treating helium ionization and recombination, including the effects of exotic energy injections; and

- a new and more correct treatment of inverse Compton scattering (ICS) for

mildly relativistic and non-relativistic electrons; previous work in the literature has relied on approximate rates which are not always accurate.

Due to these improvements, `DarkHistory` allows for rapid scans over many different prescriptions for reionization, either in the form of photoheating and photoionization rates, or a hard-coded background evolution for x_e. The epoch of reionization is currently rather poorly constrained, making it important to understand the observational signatures of different scenarios, and the degree to which exotic energy injections might be separable from uncertainties in the reionization model. Previous attempts to model the effects of DM annihilation and decay into the reionization epoch have typically either assumed a fixed ionization history [64] – requiring a slow re-computation of the cooling cascade if that history is changed [30] – or made an approximation for the effect of a modified ionization fraction on the cooling of high-energy particles [61, 63, 65–67].

Despite our emphasis on dark matter annihilation and decay, `DarkHistory` can be used to explore the effect of other forms of exotic particle injection. Other such possible sources include Hawking radiation from black holes [67, 68], radiation from accretion onto black holes [69], and processes from new physics such as de-excitation of dark matter or decay of meta-stable species [70].

In Section 2.2 we review the physics of the ionization and temperature evolution, in the context of the three-level-atom (TLA) approximation, including the possibility of exotic energy injections. In Section 2.3 we discuss the overall structure of `DarkHistory`, which self-consistently combines the TLA evolution of the ionization and gas temperature with the cooling of particles injected by exotic processes. This section also describes the implementation of various physical processes in the code, in particular the treatment of cooling and production of secondaries by electrons and photons. In Section 2.4 we relate these processes to the various modules of `DarkHistory`, before providing a number of worked examples in Section 2.5. We present our conclusions and discuss some future directions in Section 4.5. We discuss our improved treatment of ICS in detail in Appendix A.1, provide the photon spectra from positronium annihilation in Appendix A.2, discuss a series of cross checks

in Appendix A.3, and provide a table of definitions used throughout this chapter in Appendix A.4.

2.2 Ionization and Thermal Histories

`DarkHistory` computes the ionization and temperature evolution of the universe in the presence of an exotic source of energy injection, such as dark matter annihilation or decay, using a modified version of the three-level atom (TLA) model for both hydrogen and helium, based on `RECFAST` [38, 71]. The reader may refer to Ref. [72] for a detailed derivation of the unmodified TLA equations with hydrogen only, and Refs. [38, 71, 73] for the treatment of helium recombination in `RECFAST`. In this section, we will neglect the evolution of helium for simplicity, leaving a detailed discussion of our treatment of helium to Sec. 2.3.7.

In the absence of any source of energy injection, the TLA model, first derived in [37, 74], provides a pair of coupled differential equations for the matter temperature in the IGM and the hydrogen ionization fraction:

$$\begin{aligned} \dot{T}_m^{(0)} &= -2HT_m + \Gamma_C(T_{\rm CMB} - T_m)\,, \\ \dot{x}_{\rm HII}^{(0)} &= -\mathcal{C}\left[n_{\rm H} x_e x_{\rm HII}\alpha_{\rm H} - 4(1-x_{\rm HII})\beta_{\rm H} e^{-E_{21}/T_{\rm CMB}}\right]\,, \end{aligned} \tag{2.1}$$

where H is the Hubble parameter, $n_{\rm H}$ is the total number density of hydrogen (both neutral and ionized), $x_{\rm HII} \equiv n_{\rm HII}/n_{\rm H}$ where $n_{\rm HII}$ is the number density of free protons, $x_e \equiv n_e/n_{\rm H}$ is the free electron fraction with n_e being the free electron density, and $E_{21} = 10.2\,\text{eV}$ is the Lyman-α transition energy. T_m and $T_{\rm CMB}$ are the temperatures of the IGM and the CMB respectively.[1] $\alpha_{\rm H}$ and $\beta_{\rm H}$ are case-B recombination and photoionization coefficients for hydrogen respectively,[2] and $\mathcal{C}$ is the Peebles-C factor that represents the probability of a hydrogen atom in the $n = 2$ state decaying to the ground state before photoionization can occur [37, 72]. The photoionization coefficient

[1] We follow the standard astrophysical convention in which H and H^+ are denoted HI and HII, while He, He^+ and He^{2+} are denoted HeI, HeII and HeIII respectively.

[2] The value of $\beta_{\rm H}$ used in `DarkHistory` includes the constant and gaussian fudge factors used by version 1.5.2 of `RECFAST`.

is evaluated at the radiation temperature, $T_{\rm CMB}$, in agreement with Ref. [75]. Γ_C is the Compton scattering rate, given by

$$\Gamma_C = \frac{x_e}{1 + \mathcal{F}_{\rm He} + x_e} \frac{8\sigma_T a_r T_{\rm CMB}^4}{3 m_e} \,, \tag{2.2}$$

where σ_T is the Thomson cross section, a_r is the radiation constant, m_e is the electron mass, and $\mathcal{F}_{\rm He} \equiv n_{\rm He}/n_{\rm H}$ is the relative abundance of helium nuclei by number. In the absence of helium, note that $x_e = x_{\rm HII}$. The solutions to Eq. (2.1) — i.e. without any sources of energy injection — define what we will call the baseline temperature and ionization histories, $T_m^{(0)}(z)$ and $x_{\rm HII}^{(0)}(z)$. More accurate calculations of T_m and $x_{\rm HII}$ such as COSMOREC [76] and HYREC [72] are routinely used for CMB analyses, but such a high degree of accuracy is not currently needed when computing the impact of potential energy injection sources, since the uncertainties associated with these processes and the cooling of the injected particles are relatively large.

Exotic sources may inject additional energy into the universe, altering the thermal and ionization evolution shown in Eq. (2.1). For example, the rate of energy injection from DM annihilating with some velocity averaged cross section $\langle \sigma v \rangle$, or decaying with some lifetime τ much longer than the age of the universe, is given by

$$\left(\frac{dE}{dV\, dt} \right)^{\rm inj} = \begin{cases} \rho_{\chi,0}^2 (1+z)^6 \langle \sigma v \rangle / m_\chi \,, & \text{annihilation,} \\ \rho_{\chi,0} (1+z)^3 / \tau \,, & \text{decay,} \end{cases} \tag{2.3}$$

where $\rho_{\chi,0}$ is the mass density of DM today, and m_χ is the DM mass. This injected energy, however, does not in general manifest itself instantaneously as ionization, excitation, or heating of the gas. Instead, the primary particles injected into the universe may cool over timescales significantly larger than the Hubble time, producing secondary photons that may redshift significantly before depositing their energy into the gas.

Although the primary particles injected into the universe may be any type of Standard Model particle, we will only need to consider the cooling of photons and electron/positron pairs [57]. This simplification occurs because either the primaries

are stable particles like photons, electrons and positrons, neutrinos, protons and antiprotons, and heavier nuclei, or are unstable particles that resolve into these particles on time scales much shorter than the cosmological time scales under consideration. For typical sources of energy injection we can neglect heavier nuclei because they are produced in negligible amounts, and neutrinos because they are very ineffective at depositing their energy. Protons and antiprotons generally form a subdominant component of stable electromagnetic particles across all possible Standard Model primaries [33], and deposit energy less effectively than electrons, positrons, and photons (although their effects are not completely negligible [34]). We therefore only decompose the injection of any primary into an effective injection of photons, electrons, and positrons, in accordance with Ref. [57] and subsequent works. Adding the contribution from protons and antiprotons may strengthen these constraints by a small amount.

A significant amount of work has been done on computing the cooling of high energy photons, electrons, and positrons [57, 58, 77–84]. Once the cooling of injected primary particles is determined, the energy deposited into channel c (hydrogen ionization, excitation, or heating) can be simply parametrized as

$$\left(\frac{dE}{dV\,dt}\right)_c^{\rm dep} = f_c(z, \mathbf{x}) \left(\frac{dE}{dV\,dt}\right)^{\rm inj}, \tag{2.4}$$

with all of the complicated physics condensed into a single numerical factor that is dependent on the redshift and the ionization fractions of all of the relevant species in the gas, which we denote $\mathbf{x} \equiv (x_{\rm HII}, x_{\rm HeII}, x_{\rm HeIII})$. When helium is neglected, the ionization dependence of these f_c functions simplifies to a dependence on $x_{\rm HII} = x_e$. These f_c functions also depend on the energies and species of the injected particles, but for simplicity of notation we will not write these arguments explicitly.

The effect of energy injection on the thermal and ionization history can now be

captured by additional source terms,

$$\dot{T}_m^{\rm inj} = \frac{2 f_{\rm heat}(z, \mathbf{x})}{3(1 + \mathcal{F}_{\rm He} + x_e) n_{\rm H}} \left(\frac{dE}{dV\, dt} \right)^{\rm inj},$$
$$\dot{x}_{\rm HII}^{\rm inj} = \left[\frac{f_{\rm H\ ion}(z, \mathbf{x})}{\mathcal{R} n_{\rm H}} + \frac{(1 - \mathcal{C}) f_{\rm exc}(z, \mathbf{x})}{0.75 \mathcal{R} n_{\rm H}} \right] \left(\frac{dE}{dV\, dt} \right)^{\rm inj}, \tag{2.5}$$

where $\mathcal{R} = 13.6\,\text{eV}$ is the ionization potential of hydrogen.

Prior to this work, $f_c(z, \mathbf{x})$ has largely been computed assuming the standard ionization history computed by recombination codes $\mathbf{x}_{\rm std}(z)$, essentially making z the only independent variable of f_c as a function. These calculations are therefore applicable only so long as any perturbations to the assumed ionization history (e.g. by additional sources of energy injection) are sufficiently small. This is generally a good approximation near recombination: at these redshifts, the ionization history is well-constrained by CMB power spectrum measurements, and therefore large perturbations to x_e are highly disfavored. For $z \lesssim 100$, however, ionization levels that exceed the standard value of $x_e \sim 2 \times 10^{-4}$ by several orders of magnitude are experimentally allowed [30]. Moreover, star formation during the process of reionization rapidly ionizes and heats the universe at $z \lesssim 20$, causing the ionization and thermal history to diverge from the baseline histories.

The primary effect of an increase in ionization levels is to decrease the number of neutral hydrogen and helium atoms available to ionize, decreasing the fraction of injected power that goes into ionization of these species; on the other hand, increasing x_e increases the number of charged particles available for low-energy electrons to scatter off and heat the IGM, increasing the fraction of power going into heating. Since energy injection processes generally increase x_e with time, the power into heating increases at an accelerated rate at late times, making a proper calculation of $f_c(z, \mathbf{x})$ crucial for an accurate computation of the temperature history.

Computing the full $\mathbf{x}$-dependence of $f_c(z, \mathbf{x})$ also allows us to perform, for the first time, a consistent calculation of the temperature and ionization histories with both exotic energy injection processes and reionization. At the onset of reionization, stars begin to form, and the ionizing radiation emitted by these objects injects a

large amount of energy into the IGM. There remains a large degree of uncertainty regarding how reionization proceeds, but given some model for the photoionization and photoheating rates, and including other important energy transfer processes such as collisional ionization and excitation, additional terms $\dot{T}_m^{\rm re}$ and $\dot{x}_{\rm HII}^{\rm re}$ (as well as the corresponding terms for helium) can be included in Eq. (2.1) to model reionization. These terms are discussed in much greater detail in Sec. 2.3.7.

To summarize, `DarkHistory` computes the ionization and thermal history in the presence of exotic sources of energy injection, with the evolution equations in the absence of helium given by

$$\begin{aligned} \dot{T}_m &= \dot{T}_m^{(0)} + \dot{T}_m^{\rm inj} + \dot{T}_m^{\rm re}\,, \\ \dot{x}_{\rm HII} &= \dot{x}_{\rm HII}^{(0)} + \dot{x}_{\rm HII}^{\rm inj} + \dot{x}_{\rm HII}^{\rm re}\,. \end{aligned} \tag{2.6}$$

In the rest of the chapter, we will describe how we calculate the inputs required to integrate these equations, i.e. $f_c(z, \mathbf{x})$, $\dot{T}_m^{\rm re}$, $\dot{x}_{\rm HII}^{\rm re}$ and the modifications necessary to include helium.

2.3 Code Structure and Content

In this section we discuss the structure and physics content of the `DarkHistory` package.

2.3.1 Overview

Fig. 2-1 shows a flowchart depicting the overall structure of `DarkHistory`. The overall goal of the code is to take in some injected spectrum of photons and electron/positron pairs at a given redshift, and partition the energy into several categories as they lose their energy over a small redshift step:

1. *High-energy deposition.* This is the total amount of energy deposited into ionization, excitation and heating by any high-energy (above 3 keV) electron generated

during any of the cooling processes;

2. *Low-energy electrons.* These are electrons that have kinetic energy below 3 keV where atomic cooling processes typically dominate over ICS after recombination. These electrons are separated out at each step in order to treat their energy deposition (which occurs in a timescale much shorter than the time step) more carefully;

3. *Low-energy photons.* These are photons with energies below 3 keV that either photoionize within the redshift step, or lie below 13.6 eV. Such photons either lose all their energy within the redshift step, or cool only through redshifting, and thus can be treated in a simplified manner; and

4. *Propagating photons.* These are photons that are present at the end of the redshift step and are not included in the low-energy photons category.

Throughout the chapter, we use the word "electrons" to refer to both electrons and positrons. Although the interactions of electrons and positrons with the gas differ, the ICS cross-sections are identical, and ICS dominates the energy losses down to energy scales where the positron is nonrelativistic [85]. For nonrelativistic positrons, their mass energy is converted into photons through annihilation with electrons. Since the positron mass is much larger than the kinetic energy in this regime, neglecting differences in kinetic energy loss between electrons and positrons is unlikely to be important. In a future version of `DarkHistory` we plan to include a more sophisticated treatment of low energy electrons and positrons.

The outputs in the first three categories are used to compute the evolution of the ionization and temperature history at this redshift step, before the code moves on to the next step and performs the same calculation again. A brief description of a step in this loop is as follows:

1. *Input.* Before the code begins, the user specifies a DM energy injection model or some other redshift-dependent energy injection rate, as well as the photon and e^+e^- spectra produced per energy injection event. By default, `DarkHistory`

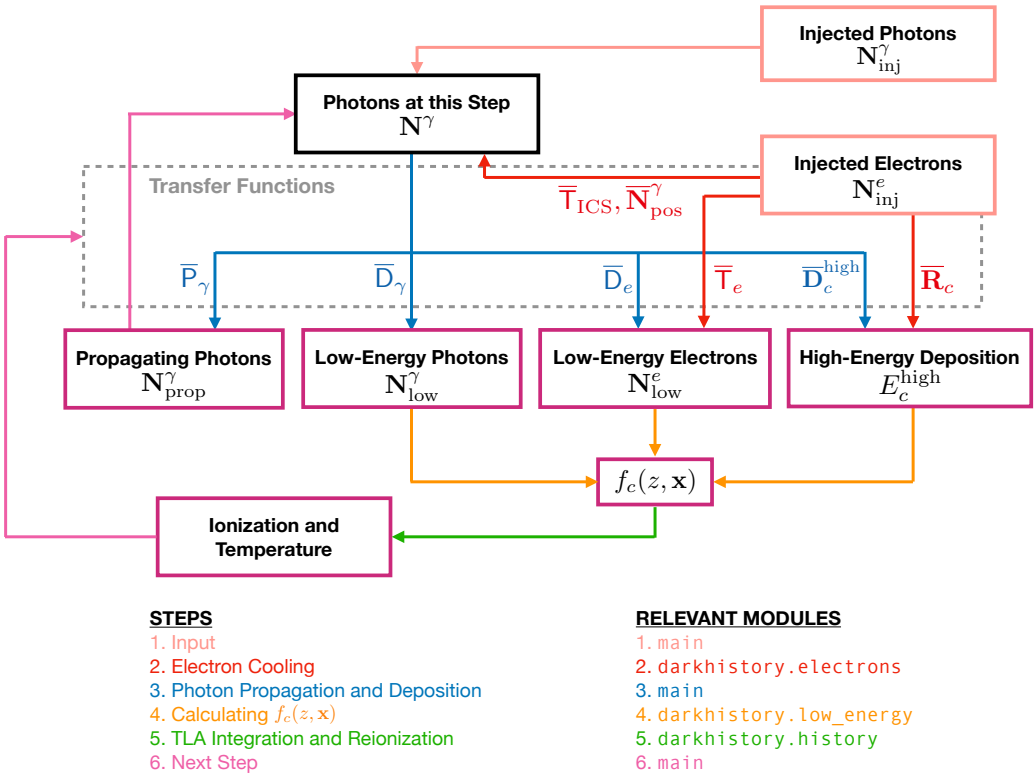


Figure 2-1: Flowchart showing schematically how the calculation of ionization and thermal histories in `DarkHistory` proceeds. Solid boxes represent input spectra (light pink), intermediate spectra used in calculations (black) and output spectra and quantities (purple), while arrows indicate numerical calculations that take place within the corresponding color-coded modules. The dashed grey box encloses all of the transfer functions for electron cooling (blue) and photon propagation and deposition (red), which are defined in Sec. 2.3.4 and 2.3.5 respectively. The calculation of $f_c(z)$ (orange) and the integration of the TLA (green) are explained in Sec. 2.3.6 and 2.3.7 respectively. Propagating photons and ionization/temperature values, which are used in calculating the transfer functions, are used as inputs for the next step (purple). All notation used here are defined in the text, and a summary table with their definitions can be found in Appendix A.4. Each step is outlined in Sec. 4.2, and then explained in more detail in subsequent subsections within Sec. 2.3. The modules shown here will also be outlined in Sec. 2.4.

starts from an initial redshift of $1 + z = 3000$, ensuring that the spectra of particles present at and after recombination (at $z \sim 1000$) are accurate. Details are provided in Sec. 2.3.3. Inputs to the code are provided to the function `evolve()` found in the module `main`; some tools for obtaining spectra from an arbitrary injection of Standard Model particles can be found in the `pppc` module;

2. *Injected electron cooling.* Injected electrons (and positrons) cool through a combination of atomic processes and ICS. Transfer functions that map these injected electrons to high-energy deposition, secondary photons from ICS and positron annihilation, and low-energy electrons are computed and applied to the injected electrons. A discussion of these calculations can be found in Sec. 2.3.4 and in the `electrons` module of the code.

   The sum of the secondary photons produced by electron cooling, photons injected on this timestep, and propagating photons from the previous timestep are used as input to the photon cooling transfer functions, which we describe next;

3. *Photon propagation and energy deposition.* At this stage, we have a spectrum of photons that can undergo a range of cooling processes to lose their energy over this redshift step. The effect of these cooling processes on the photon spectrum can be reduced to three transfer functions that we will describe in detail in Sec. 2.3.5. These transfer functions have been pre-computed separately and can be downloaded at https://doi.org/10.7910/DVN/DUOUWA, together with all the other data required to run the code. These transfer functions determine how photons in this redshift step turn into propagating photons that continue on to the next redshift step, and low-energy photons and low-energy electrons that undergo further processing. All of these computations occur in the `main` module;

4. *Calculating* $f_c(z, \mathbf{x})$. The low-energy photons and low-energy electrons from this redshift step deposit their energy into ionization, heating and excitation of atoms, and the value of $f_c(z, \mathbf{x})$ at this step is computed by comparing the energy deposited in each channel to the energy injection rate for this timestep. Details of this computation are given in Sec. 2.3.6, and can be found in the `low_energy` module;

5. *TLA integration and reionization.* With $f_c(z, \mathbf{x})$ at this step, we can now integrate the TLA across this redshift step. We can also include a reionization

model, or track helium ionization, both of which add more terms to the TLA, as detailed in Sec. 2.3.7. We now know the $\mathbf{x}$ and T_m that are reached at the end of this step. These calculations are done in the `history` module; and

6. *Next step.* The $\mathbf{x}$ and T_m values computed above are passed to the next redshift step, so that all transfer functions at the next step can be computed at the appropriate ionization level. The propagating photons found above are also passed to the next step, and the loop repeats.

Because $f_c(z, \mathbf{x})$ is computed by integrating the TLA at each step, and all transfer functions are evaluated at the value of $\mathbf{x}$ in the step, the backreaction of increased ionization levels is now fully accounted for.

In the next several sections, we will describe both the physics and numerical methods that go into the loop.

2.3.2 Discretization

Before describing in detail each part of `DarkHistory`, we will first describe how discretization occurs in our code, and the notation we will use throughout this chapter. Typically, we will deal with some smooth spectrum of particles $dN/dE(E, A, B, \cdots)$, which is a function of the energy abscissa E, and several other variables that we denote here as $A, B, \cdots$. Smooth functions that are derivatives will always use 'd' to denote differentiation, and parentheses to denote functional dependence. We shall always discretize such spectra as

$$\frac{dN}{dE}(E_i, A_j, B_k, \cdots) \approx \mathsf{S}[E_i, A_j, B_k, \cdots]\,. \tag{2.7}$$

The discretized spectrum S is a matrix of dimension equal to the number of variables it depends on, where $i, j, k, ...$ index discrete values of these variables. Throughout this chapter, we will denote vectors (quantities which depend on a single variable) by a bold typeface and matrices (quantities that depend on multiple variables) by a sans-serif typeface. Discrete steps or changes are denoted by 'Δ', and discrete functional

dependencies are written in square brackets.

S times the bin width should always be regarded as a matrix of number of particles inside some bin, all with energy given by E_i. This matrix is mathematically defined as

$$\mathsf{N}[E_i, A_j, B_k, \cdots] \quad \equiv \quad \mathsf{S}[E_i, A_j, B_k, \cdots] \quad \times \quad E_i \Delta \log E_i \,, \tag{2.8}$$

where $\Delta \log E_i$ is the log-energy bin width. We will always take $E_i \Delta \log E_i$ to be the bin width by convention. In `DarkHistory`, spectra are binned into energy values that are evenly log-spaced. E_i should be regarded as the bin center, with the bin boundaries occurring at the geometric mean of adjacent energy values, and the boundaries of the first and last bin are taken to be symmetric (in log-space) about the bin centers.

2.3.3 Input

To initialize the loop described above, the user must specify the discretized photon and electron spectra produced per injection event, which we denote $\overline{\mathbf{N}}^{\gamma}_{\text{inj}}[E'_j]$ and $\overline{\mathbf{N}}^{e}_{\text{inj}}[E'_j]$. Bars denote spectra or transfer functions that have been normalized by some process or quantity, while spectra without any markings denote a number of particles per baryon from here on, unless otherwise specified.

Given the redshift-dependent rate of injection events per volume $(dN/dV\,dt)^{\text{inj}}$ we can determine the spectrum of particles $\mathbf{N}^{\alpha}_{\text{inj}}$ injected within a log-redshift step of width $\Delta \log(1+z)$ per baryon by

$$\mathbf{N}^{\alpha}_{\text{inj}}[E'_i, z] = \overline{\mathbf{N}}^{\alpha}_{\text{inj}}[E'_i] \left(\frac{dN}{dV\,dt}\right)^{\text{inj}} G(z) \,, \tag{2.9}$$

where α take on values γ or e, and

$$G(z) \equiv \frac{\Delta \log(1+z)}{n_B(z) H(z)} \,, \tag{2.10}$$

where n_B is the number density of baryons. $G(z)$ converts between the rate of injection

events per volume and the number of injection events per baryon in the log-redshift step.

In the following sections, we will be mostly concerned with log-redshift steps, and so it is convenient to define

$$y \equiv \log(1+z)\,, \tag{2.11}$$

and likewise $\Delta y \equiv \Delta \log(1+z)$.

2.3.4 Injected Electron Cooling

After specifying the injected spectra, the next step of the code is to resolve the injected electron/positron pairs, $\mathbf{N}^e_{\text{inj}}$. High-energy electrons and positrons cool through atomic processes (collisional ionization, collisional excitation and Coulomb heating), as well as ICS off CMB photons. After losing their kinetic energy to these processes, positrons ultimately annihilate with free electrons in the IGM, producing high-energy photons. All of these processes occur within a timescale much shorter than the timesteps considered in `DarkHistory`. Because of this, the code converts all input high-energy electrons into energy deposited into ionization, excitation, heating, scattered photons from ICS, and low-energy electrons (below 3 keV), which we treat separately. The photons produced from ICS are added to those that are injected promptly from the DM energy injection process, as well as propagating photons from the previous step.

We will first briefly discuss our calculation of the scattered photon and electron spectra from ICS, and then move on to describe the numerical method used to compute electron cooling.

Inverse Compton Scattering

ICS off CMB photons is an important energy loss mechanism for electrons/positrons over a large range of energies and redshifts. The efficiency of ICS as a cooling mechanism relative to atomic cooling processes has been the subject of some confusion in the

literature, with some earlier studies [78, 79] underestimating the cooling rate of the electrons. ICS becomes more important relative to atomic processes as the electron energy increases, but a correct treatment shows that even nonrelativistic electrons can have ICS as the main cooling mechanism in the early universe; at $z \sim 600$, for example, it is the primary energy loss mechanism for electrons with kinetic energy $\gtrsim 10\,\text{keV}$ [80, 82]. Existing work on electron cooling has focused on the highly nonrelativistic regime (electron kinetic energy below 3 keV) [86], where ICS is unimportant compared to atomic cooling processes, or on the relativistic regime [78, 79, 87].

Earlier work by one of the authors [58, 82] already incorporates ICS cooling for electrons across both the Thomson and the relativistic regimes. `DarkHistory` improves the accuracy of the calculation in the Thomson regime by using the full expression for the spectrum of scattered photons, with no further approximation. As a result, the code is able to accurately calculate the scattered photon spectrum and the energy loss spectrum of electrons. This means that we fully cover all relevant regimes for ICS for electrons of arbitrary energy scattering off the CMB at all redshifts $z \sim 10^9$ and below.[3] These calculations are fast and numerically stable even for nonrelativistic electrons, where conventional numerical integration can be unreliable due to the presence of catastrophic cancellations between large terms.

We leave a full discussion of how `DarkHistory` treats ICS to Appendix A.1. In summary, the code is able to compute the scattered photon and electron spectra that are produced per unit time due to ICS off the CMB across all relevant kinematic regimes. These spectra are then taken as inputs for the numerical computation of how an electron cools taking into account all processes, which is described below.

Numerical Method

Consider an injected electron (or positron) with kinetic energy E' (all quantities associated with injected particles throughout this chapter will be denoted with $'$). Let $R_c(E')$ be the energy eventually deposited into some channel c by this electron,

[3] Above this redshift, photons have energies comparable to the electron mass m_e, and Klein-Nishina scattering can occur between photons and non-relativistic electrons, which falls outside of the two regimes considered here.

once it has lost all of its initial energy. Within a short time interval Δt (taken to be 1 s in our calculation), the electron undergoes all possible cooling processes with some probability, producing the (averaged) secondary electron spectrum dN/dE. Within this same interval Δt, some portion of the energy $P_c(E')$ is also deposited promptly into the channel under consideration. The secondary electron spectrum then deposits its energy according to R_c for energies lower than E'. We can thus write the following recursive equation:

$$R_c(E') = \int dE\, R_c(E) \frac{dN}{dE} + P_c(E')\,. \tag{2.12}$$

Note that $R_c(E')$ does not include deposition to the channel c via secondary photons from ICS or positron annihilation; because the cooling times of secondary photons can be much longer than a timestep, they must be treated separately. $R_c(E')$ as defined here is the "high-energy deposition" from electrons within the timestep, as described in Section 4.2. The relevant channels are $c = \{\text{'ion'}, \text{'exc'}, \text{'heat'}\}$ for deposition into collisional ionization, collisional excitation and heating respectively. The 'ion' and 'exc' channels include ionization and excitation off all species.

As long as the time step Δt is much shorter than the characteristic interaction timescale of all of the interactions, dN/dE is simply the sum of all of the scattered electron spectra due to each process within Δt, normalized to a single injected electron. A detailed accounting of the relevant cross sections and secondary spectra is provided in Ref. [57], and these results can be used to calculate dN/dE and P_c. We will denote the discretized version of the normalized scattered electron spectra by $\overline{\mathsf{N}}$, since it is normalized to one electron.

Numerically, we would like to compute $\overline{\mathbf{R}}_c$, a vector containing the energy deposited into channel c, with each entry corresponding to a single electron with initial kinetic energy E'. The overline notation serves as a reminder that the quantity is normalized to one injected electron. The discretized version of Eq. (2.12) reads

$$\overline{\mathbf{R}}_c[E'_i] = \sum_j \overline{\mathsf{N}}[E'_i, E_j]\overline{\mathbf{R}}_c[E_j] + \overline{\mathbf{P}}_c[E'_i]\,, \tag{2.13}$$

where $\overline{\mathbf{P}}_c$ is the vector of the prompt energy deposition in channel c per electron. This is a linear system of equations, and we can solve for each $\overline{\mathbf{R}}_c$ given $\overline{\mathsf{N}}$ and $\overline{\mathbf{P}}_c$.

A similar procedure also works for finding the ICS photon spectrum after an electron completely cools. Let the discretized spectrum be $\overline{\mathsf{T}}_{\mathrm{ICS},0}[E'_{e,i}, E_{\gamma,j}]$, where E'_e is the initial electron kinetic energy, and E_γ is the photon energy. Then the ICS photon spectrum produced after complete cooling of a single electron satisfies

$$\begin{aligned}\overline{\mathsf{T}}_{\mathrm{ICS},0}[E'_{e,i}, E_{\gamma,j}] = &\sum_k \overline{\mathsf{N}}[E'_{e,i}, E_{e,k}]\overline{\mathsf{T}}_{\mathrm{ICS},0}[E_{e,k}, E_{\gamma,j}] \\ &+ \overline{\mathsf{N}}_{\mathrm{ICS}}[E'_{e,i}, E_{\gamma,j}]\,,\end{aligned} \tag{2.14}$$

with $\overline{\mathsf{N}}_{\mathrm{ICS}}$ being the discretized version of the scattered photon spectrum defined in Eq. (A.1.1) within Δt, and indices e and γ have been inserted to clarify the difference between electron and photon energies. This spectrum consists of CMB photons that are upscattered by the injected electron; in order to be able to track energy conservation, we also need to keep track of the initial energy of the upscattered photons. We therefore also need to solve

$$\overline{\mathbf{R}}_{\mathrm{CMB}}[E'_i] = \sum_j \overline{\mathsf{N}}[E'_i, E_j]\overline{\mathbf{R}}_{\mathrm{CMB}}[E_j] + \overline{\mathbf{P}}_{\mathrm{CMB}}[E'_i]\,, \tag{2.15}$$

where $\overline{\mathbf{P}}_{\mathrm{CMB}}$ is the total initial energy of photons upscattered in Δt.[4] At this point, we now define $\overline{T}_{\mathrm{ICS}}$ to be the ICS photon spectrum with the upscattered CMB spectrum subtracted out, so that $\overline{T}_{\mathrm{ICS}}$ now represents a *distortion* to the CMB spectrum:

$$\begin{aligned}\overline{\mathsf{T}}_{\mathrm{ICS}}[E'_{e,i}, E_{\gamma,j}] = &\overline{\mathsf{T}}_{\mathrm{ICS},0}[E'_{e,i}, E_{\gamma,j}] \\ &- \overline{\mathbf{R}}_{\mathrm{CMB}}[E'_{e,i}]\overline{\mathbf{N}}_{\mathrm{CMB}}[E_{\gamma,j}]\,,\end{aligned} \tag{2.16}$$

where $\overline{\mathbf{N}}_{\mathrm{CMB}}$ is the CMB spectrum normalized to unit total energy. The total of

[4] We do not have to track the photon spectrum, since the initial CMB photon energy is only significant for nonrelativistic injected electrons, which are always in the Thomson regime and hence scatter in a frequency-independent manner. For relativistic electrons, the initial CMB photon energy is neglected, as the photon is overwhelmingly upscattered to a much higher final energy.

energy of $\overline{T}_{\rm ICS}$ for each $E'_{e,i}$ therefore gives the energy lost by the incoming electron through ICS.

Finally, the low-energy electron spectrum produced is similarly given by

$$\begin{aligned}\overline{\mathsf{T}}_e[E'_{e,i}, E_{e,j}] = &\sum_k \overline{\mathsf{N}}_{\rm high}[E'_{e,i}, E_{e,k}]\overline{\mathsf{T}}_e[E_{e,k}, E_{e,j}] \\ &+ \overline{\mathsf{N}}_{\rm low}[E'_{e,i}, E_{e,j}]\,,\end{aligned} \tag{2.17}$$

where $\overline{\mathsf{N}}_{\rm high}$ ($\overline{\mathsf{N}}_{\rm low}$) is $\overline{\mathsf{N}}$ with only high-energy (low-energy) $E_{e,k}$ included.

In `DarkHistory`, we choose a square matrix $\overline{\mathsf{N}}$ with the same abscissa for both injected and scattered electron energies. As a result, $\overline{\mathsf{N}}$ has diagonal values that are very close to 1, since most particles do not scatter within Δt. Because of this, we find that it is numerically more stable to solve the equivalent equation

$$\frac{\widetilde{E}[E'_i]}{E'_i}\overline{\mathbf{R}}_c[E'_i] = \sum_j \widetilde{\mathsf{N}}[E'_i, E_j]\overline{\mathbf{R}}_c[E_j] + \overline{\mathbf{P}}_c[E'_i]\,, \tag{2.18}$$

where

$$\widetilde{\mathsf{N}}[E'_i, E_j] \equiv \begin{cases} \overline{\mathsf{N}}[E'_i, E_j]\,, & E'_i < E_j\,, \\ 0\,, & \text{otherwise,} \end{cases} \tag{2.19}$$

$$\begin{aligned}\widetilde{E}[E'_i] \equiv &\sum_j \widetilde{\mathsf{N}}[E'_i, E_j]E_j + \sum_c \overline{\mathbf{R}}_c[E'_i] \\ &+ \sum_j \overline{\mathsf{T}}_{\rm ICS}[E'_i, E_{\gamma,j}]E_{\gamma,j}\,.\end{aligned} \tag{2.20}$$

The variables $\widetilde{\mathsf{N}}$ and $\widetilde{E}$ are simply the number of electrons and total energy excluding electrons that remained in the same energy bin after Δt. Eqs. (2.16) and (2.17) can be similarly transformed in the same way as Eq. (2.18) and solved. Since $\widetilde{N}$ is a triangular matrix, the SciPy function `solve_triangular()` is used for maximum speed.[5]

[5]The upscattering of electrons during ICS is negligible: see Appendix A.1 for more details.

Having calculated $\overline{\mathbf{R}}_c$, $\overline{\mathsf{T}}_{\rm ICS}$ and $\overline{\mathsf{T}}_e$, all normalized to a single electron, the final result when an arbitrary electron spectrum $\mathbf{N}^e_{\rm inj}[E'_{e,i}]$ completely cools is simply given by contracting these quantities with $\mathbf{N}^e_{\rm inj}$. Note that all of these quantities are also dependent on redshift: we have simply suppressed this dependence for notational simplicity in this section.

Finally, after positrons have lost all of their kinetic energy, they are assumed to form positronium and annihilate promptly, producing a gamma ray spectrum that also gets added to the propagating photon spectrum. The positronium spectrum is given simply by

$$\mathbf{N}^\gamma_{\rm pos}[E_i] = \frac{1}{2}\overline{\mathbf{N}}^\gamma_{\rm pos}[E_i] \sum_j \mathbf{N}^e_{\rm inj}[E'_j]\,, \tag{2.21}$$

where $\overline{\mathbf{N}}^\gamma_{\rm pos}$ is the positronium annihilation spectrum normalized to a single positron, shown in Appendix A.2. The factor of 1/2 accounts for the fact that $\mathbf{N}^e_{\rm inj}$ contains both electrons and positrons in equal number.

Since all calculated quantities depend on z and $\mathbf{x}$, all quantities discussed in this section have to be computed at each redshift step. This allows us to properly capture the effect of changing ionization levels on the energy deposition process.

2.3.5 Photon Propagation and Energy Deposition

After resolving the injected electrons and obtaining the photons produced from their cooling, the spectrum of photons that have been newly injected per baryon per log-redshift can be discretized as

$$\frac{dN^\gamma_{\rm new}}{dE'_j\,dy}(E'_j) \times E'_j \log \Delta E'_j \times \Delta y \approx \mathbf{N}^\gamma_{\rm new}[E'_j]\,, \tag{2.22}$$

where $\mathbf{N}^{\gamma}_{\text{new}}$ is the sum of photons injected directly by the injection event, and photons produced by the cooling of injected electrons, i.e.

$$\begin{aligned}\mathbf{N}^{\gamma}_{\text{new}}[E'_j] = \mathbf{N}^{\gamma}_{\text{inj}}[E'_j] + \mathbf{N}^{\gamma}_{\text{pos}}[E'_j] \\ + \sum_i \overline{\mathsf{T}}_{\text{ICS}}[E'_{e,i}, E'_j]\mathbf{N}^{e}_{\text{inj}}[E'_{e,i}]\,.\end{aligned} \tag{2.23}$$

These photons can cool through a number of processes, including redshifting, pair production, Compton scattering and photoionization. Within a particular log-redshift step, low-energy photons and low-energy electrons are produced, and some high-energy deposition from high-energy electrons produced by $\mathbf{N}^{\gamma}_{\text{new}}$ occur. On the other hand, some part of the photon spectrum lies above 13.6 eV and does not photoionize within the log-redshift step; instead, these photons propagate forward to the next step.

The resulting deposition into low-energy photons and electrons was used to compute f_c in Ref. [82], assuming the fixed baseline ionization history. In order to capture the dependence on ionization history, however, we need to be able to calculate the propagation and deposition processes at any ionization level, redshift and injected particle energy.

One of the main ideas of `DarkHistory` is to capture the photon cooling processes as precomputed transfer functions with injection energy, redshift and ionization levels as the dependent variables. These transfer functions then act on some incoming spectrum and produce a spectrum of propagating particles, a spectrum of deposited particles or some amount of deposited energy within a log-redshift step. These transfer functions can be evaluated at various points in injection energy, redshift, and ionization levels, and interpolated at other points. With a given injection model, we can then string together these transfer functions to work out the propagation of photons and the deposition of energy, over an extended redshift range, given any exotic source of energy injection.

Propagating Photons

Consider a spectrum of photons per baryon denoted dN^γ/dE' that is present in the universe at some log-redshift y. As these photons propagate, various cooling processes result in these photons being scattered into energies below 13.6 eV, or they may photoionize on an atom in the gas. Those particles that do not undergo either process within a redshift step are called "propagating photons", and continue to propagate into the next redshift step.

We define the transfer function for propagating photons $\overline{P}^\gamma(E', E, y', y)$ through the following relation:

$$\left.\frac{dN^\gamma_{\rm prop}}{dE}\right|_y = \int dE'\, \overline{P}^\gamma(E', E, y', y)\, \left.\frac{dN^\gamma}{dE'}\right|_{y'} . \tag{2.24}$$

$\overline{P}^\gamma$ takes a spectrum of photons that are present at y' and propagates them forward to a spectrum of propagating photons at y. $\overline{P}^\gamma(E', E, y', y)$ is exactly the number of propagating photons per unit energy that results from a single photon injected at log-redshift y' with energy E' cooling until log-redshift y. The $\overline{P}^\gamma$ functions are calculated separately using the code described in Ref. [57, 82].

We distinguish between two different sources of photons between two redshifts y' and y (with $y' > y$): propagating photons at y', $dN^\gamma_{\rm prop}/dE'$, and the newly injected photons between the redshifts y' and y, defined in discretized form in Eq. (2.23). With these sources, we can write the spectrum of propagating photons at y as

$$\begin{aligned}\left.\frac{dN^\gamma_{\rm prop}}{dE}\right|_y = &\int dE'\, \overline{P}^\gamma(E', E, y', y)\, \left.\frac{dN^\gamma_{\rm prop}}{dE'}\right|_{y'} \\ &+ \int dE' \int_y^{y'} d\eta\, \overline{P}^\gamma(E', E, \eta, y)\, \left.\frac{dN^\gamma_{\rm new}}{dE'\, d\eta}\right|_\eta . \end{aligned} \tag{2.25}$$

We discretize this expression by defining the following discrete quantities according

to the conventions set down in Eqs. (2.7) and (2.8):

$$\overline{\mathsf{P}}^{\gamma}[E_i', E_j, y', \Delta y] E_i' \Delta \log E_i' \approx \overline{P}^{\gamma}(E_i', E_j, y', y' - \Delta y)\,,$$
$$\mathbf{N}^{\gamma}_{\text{prop}}[E_i', y'] \approx \left.\frac{dN^{\gamma}_{\text{prop}}}{dE'}\right|_{y'} E_i' \,\Delta \log E_i'\,, \tag{2.26}$$

where we have chosen some fixed value of Δy, so that the final redshift is $y = y' - \Delta y$. In `DarkHistory`, the default value is $\Delta y = 10^{-3}$, although this can be adjusted by the process of coarsening, described in Sec. 2.3.5. Dropping the dependence on Δy for simplicity, the discretized version of Eq. (2.25) reads

$$\mathbf{N}^{\gamma}_{\text{prop}}[E_j, y] = \sum_i \overline{\mathsf{P}}^{\gamma}[E_i', E_j, y'] \mathbf{N}^{\gamma}[E_i', y']\,, \tag{2.27}$$

where we have defined

$$\mathbf{N}^{\gamma}[E_i', y] \equiv \mathbf{N}^{\gamma}_{\text{prop}}[E_i', y] + \mathbf{N}^{\gamma}_{\text{new}}[E_i', y]\,. \tag{2.28}$$

Energy Deposition

Aside from $\overline{\mathsf{P}}^{\gamma}$, we also have three deposition transfer functions describing the energy losses of $\mathbf{N}^{\gamma}$ into high-energy deposition, low-energy electrons and low-energy photons, as defined in Sec. 4.2. These transfer functions are defined by their action on the discretized photon spectrum, $\mathbf{N}^{\gamma}$, and are discretized in a similar manner.

The low-energy electron deposition transfer matrix, $\overline{\mathsf{D}}^{e}$, yields the low-energy electrons produced via cooling of $\mathbf{N}^{\gamma}$. Adding the low-energy electrons produced directly from the injected electrons $\mathbf{N}^{e}_{\text{inj}}$, we obtain the full low-energy electron spectrum $\mathbf{N}^{e}_{\text{low}}[E_j, y]$ at a particular redshift step:

$$\mathbf{N}^{e}_{\text{low}}[E_{e,j}, y] = \sum_i \overline{\mathsf{D}}^{e}[E'_{\gamma,i}, E_{e,j}, y'] \mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$$
$$+ \mathbf{N}^{e}_{\text{low,inj}}[E_{e,j}, y]\,, \tag{2.29}$$

where

$$\mathbf{N}^{e}_{\text{low,inj}}[E_{e,j}, y] = \sum_i \overline{\mathsf{T}}^{e}[E'_{e,i}, E_{e,j}, y]\mathbf{N}^{e}_{\text{inj}}[E'_{e,i}, y]\,, \tag{2.30}$$

while the deposition transfer matrix $\overline{\mathsf{D}}^{\gamma}$ yields the low-energy photons,

$$\mathbf{N}^{\gamma}_{\text{low}}[E_j, y] = \sum_i \overline{\mathsf{D}}^{\gamma}[E'_i, E_j, y']\mathbf{N}^{\gamma}[E'_i, y']\,. \tag{2.31}$$

$\mathbf{N}^{\gamma}_{\text{low}}$ is computed as a *distortion* to the CMB spectrum, with $\overline{\mathsf{D}}^{\gamma}$ computed with the initial spectrum of upscattered CMB photons subtracted, in the same way as $\overline{T}_{\text{ICS}}$, as shown in Eq. (2.16).

As the propagating photons cool over a single log-redshift step, they generate high-energy electrons along the way. These are handled in a similar manner to injected high-energy electrons as described in Sec. 2.3.4, but instead of performing the calculation at each step, we simply provide transfer functions $\overline{\mathbf{D}}^{\text{high}}_{\text{c}}$ that act on propagating photons and return the high-energy deposition into the channels $c =$‘ion’, ‘exc’, ‘heat’}.[6] We can then combine this with the result from electron cooling to obtain the high-energy deposition per baryon within a log-redshift step into each channel c:

$$\begin{aligned} E^{\text{high}}_{c}[y] = &\sum_i \overline{\mathbf{D}}^{\text{high}}_{\text{c}}[E'_{\gamma,i}, y']\mathbf{N}^{\gamma}[E'_{\gamma,i}, y'] \\ &+ \sum_i \overline{\mathbf{R}}_{c}[E'_{e,i}, y']\mathbf{N}^{e}_{\text{inj}}[E'_{e,i}, y']\,. \end{aligned} \tag{2.32}$$

To summarize, we have defined the following transfer functions: $\overline{\mathsf{P}}^{\gamma}$ for propagating photons, and $\overline{\mathsf{D}}^{\gamma}$, $\overline{\mathsf{D}}^{e}$ and $\overline{\mathbf{D}}^{\text{high}}_{c}$ for deposition into low-energy photons, low-energy electrons and high-energy deposition channels respectively. These transfer functions act on the spectrum of photons $\mathbf{N}^{\gamma}$ (from both the injection source and the cooling of injected electrons). Together with the transfer functions for the cooling of injected

[6] For legacy reasons, `DarkHistory` actually computes the transfer function that returns the high-energy deposition per second; this is just a difference in convention.

electrons, we have all the information needed to propagate injected particles and compute their energy deposition as a function of redshift.

Coarsening

The propagating photons transfer function $\overline{\mathsf{P}}^{\gamma}$ can always be evaluated with the same input and output energy abscissa, so that the 2D transfer matrix at each y is square. If the transfer function $\overline{\mathsf{P}}^{\gamma}$ does not vary significantly over redshift, then in the interest of computational speed, we can make the following approximation of Eq. (2.27) for propagation transfer matrices:

$$\mathbf{N}^{\gamma}_{\text{prop}}[E_j, y - n\Delta y] \approx \left(\overline{\mathsf{P}}^{\gamma}_{1/2}\right)^{n}_{ji} \mathbf{N}^{\gamma}_{i}[y]\,, \tag{2.33}$$

where repeated indices are summed. i and j index input and output energies, and $\overline{\mathsf{P}}^{\gamma}_{1/2}$ is $\overline{\mathsf{P}}^{\gamma}$ evaluated at log-redshift $y - n\Delta y/2$ to minimize interpolation error. When making this approximation, we also have to ensure that we redefine

$$\mathbf{N}^{\alpha}_{\text{inj}}[E'_i, y] \to n\mathbf{N}^{\alpha}_{\text{inj}}[E'_i, y] \tag{2.34}$$

for both channels $\alpha = e$ and γ, so that we (approximately) include all of the particles injected between y and $y - n\Delta y$.

Likewise, if both the deposition and propagation matrices do not vary significantly over redshift, we can approximate Eq. (2.31) as

$$\mathbf{N}^{\gamma}_{\text{low}}[E_j, y - n\Delta y] \approx \left(\overline{\mathsf{D}}^{\gamma}_{1/2}\right)_{jk} \sum_{m} \left(\overline{\mathsf{P}}^{\gamma}_{1/2}\right)^{m}_{ki} \mathbf{N}^{\gamma}_{i}[y]\,, \tag{2.35}$$

with repeated indices once again being summed over. $\overline{\mathsf{D}}^{\gamma}_{1/2}$ is defined in the same manner as $\overline{\mathsf{P}}^{\gamma}_{1/2}$. This equation essentially applies the deposition transfer matrix at $y - n\Delta y/2$ to all n steps of the propagation of the spectrum $\mathbf{N}^{\gamma}$ from y to $y - \Delta y$, which itself is approximated by $\overline{\mathsf{P}}^{\gamma}_{1/2}$. In our code, we call these approximations "coarsening", and the number n in both Eqs. (2.33) and (2.35) the "coarsening factor".

Different Redshift Regimes

In `DarkHistory` we separate our transfer matrices into three redshift regimes: redshifts encompassing reionization ($z < 50$), redshifts encompassing the times between recombination and reionization ($50 \leq z \leq 1600$), and redshifts well before recombination ($z > 1600$). During the redshifts encompassing reionization, we allow our transfer functions to be functions of x_{HII} and x_{HeII}, enabling the use of reionization models that evolve hydrogen and helium ionization levels separately. We only consider singly-ionized helium in the current version of `DarkHistory` since we expect x_{HeIII} not to play an important role until $z \sim 6$. We compute the transfer functions on a grid of z^k, x_{HII}^m, and x_{HeII}^n, and linearly interpolate over the grid of pre-computed transfer functions.

Between recombination and reionization, the helium ionization level lies at or below the hydrogen ionization level, since helium has a larger ionization potential at 24.6 eV. After recombination, current experimental constraints typically forbid a large ionization fraction, i.e. we expect $x_{\text{HII}} \lesssim 0.1$ [30]. As such, setting $x_{\text{HeII}} = 0$ is a good approximation for the photon propagation and deposition functions: since $\mathcal{F}_{\text{He}} \sim 8\%$, neglecting helium ionization only results in $\lesssim 8\%$ error to x_e, and $\lesssim 10\%$ error in the density of neutral helium. We therefore follow the same procedure as before, except we now calculate and interpolate the transfer functions over a grid of z^k and x_{HII}^m values while holding the helium ionization level fixed to zero.

Finally, well before recombination, we expect the universe to be close to 100% ionized and tightly coupled thermally to the CMB. Any extra source of exotic energy injection that is consistent with current experimental constraints will likely have a negligible effect on the ionization and thermal histories. We thus calculate and interpolate our transfer functions over a grid of z^k values while holding the hydrogen and helium ionization levels to the baseline values provided by `RECFAST` [73].

The actual grid values z^k, x_{HII}^m, and x_{HeII}^n in each of these regimes can be found in the code, and have been chosen so that interpolation errors are at the sub-10% level when $f_c(z)$ is calculated using the same method detailed in Ref. [82]. Our results for

$f_c(z)$ without taking into account backreaction, including some improvements over Ref. [82], can be found in Appendix A.3.

2.3.6 Calculating $f_c(z)$

The low-energy photons $\mathbf{N}^{\gamma}_{\text{low}}[E_i, z]$ and low-energy electrons $\mathbf{N}^{e}_{\text{low}}[E_i, z]$, defined in Sec 4.2, transfer their energy into ionization and excitation of atoms, heating of the IGM, and free-streaming photons to be added to the CMB continuum. In `DarkHistory` we keep track of how much energy low energy photons and electrons deposit into channels c $\in$ {'H_{ion}', 'He_{ion}', 'exc', 'heat', 'cont'}, which represent hydrogen ionization, helium ionization, hydrogen excitation, heating of the IGM, and sub-10.2 eV continuum photons respectively. The energy deposition fractions $f_c(z)$ are then found by normalizing the total energy deposited into channel c within a redshift step by the total energy injected within that step according to Eq (2.4). We closely follow the method for computing $f_c(z)$ described in Ref. [82].

Before calculating $f_c(z)$ for each channel, it is instructive to see how to calculate the total amount of energy deposited per unit time and volume, $(dE/dV\,dt)^{\text{dep}}$. The low-energy photon and electron spectra $\mathbf{N}^{\gamma}_{\text{low}}[E_i]$ and $\mathbf{N}^{e}_{\text{low}}[E_i]$ as defined above contain a number of particles per baryon deposited within each log-redshift bin (the z-dependence has been suppressed since all calculations in this section occur at the same redshift step). We can convert between these and spectra containing the number of particles produced per unit volume and unit time using the conversion factor $G(z)$ introduced in Eq. (2.10). For example, to obtain the total amount of energy deposited at a given redshift per unit time and volume, one simply sums over low-energy particle type and applies the conversion factor,

$$\left(\frac{dE}{dV\,dt}\right)^{\text{dep}}_{\text{low}} = \frac{1}{G(z)} \sum_{\alpha} \sum_{i} E'_i\, \mathbf{N}^{\alpha}_{\text{low}}[E'_i]\,. \tag{2.36}$$

To calculate the total amount of energy deposited we must also add the amount

deposited by high energy electrons and photons, which we computed in Eq. (2.32):

$$\left(\frac{dE}{dV\,dt}\right)^{\rm dep}_{\rm high} = \frac{1}{G(z)}\sum_c E_c^{\rm high}\,. \tag{2.37}$$

Then the total deposited energy summed over all channels is given by

$$\left(\frac{dE}{dV\,dt}\right)^{\rm dep} = \left(\frac{dE}{dV\,dt}\right)^{\rm dep}_{\rm low} + \left(\frac{dE}{dV\,dt}\right)^{\rm dep}_{\rm high}\,. \tag{2.38}$$

With this example in mind, we are now ready to understand how to split the energy deposition into the different channels.

Photons

We first compute $f_c(z)$ for low-energy photons, starting with energy deposition into continuum photons. These are photons with energy below $3\mathcal{R}/4 = 10.2\,\text{eV}$ that are unable to effectively transfer their energy to free electrons or atoms, so they just free stream. The energy of these photons constitutes deposition into the continuum channel, i.e.

$$\left(\frac{dE^\gamma}{dV\,dt}\right)^{\rm dep}_{\rm cont} = \frac{1}{G(z)}\sum_{E_i=0}^{3\mathcal{R}/4} E_i\,\mathbf{N}^\gamma_{\rm low}[E_i]\,. \tag{2.39}$$

To calculate the total amount of energy deposited into hydrogen excitation, we make the approximation that all photons with energies between $3\mathcal{R}/4 = 10.2\,\text{eV}$ and $\mathcal{R} = 13.6\,\text{eV}$ deposit their energy instantaneously into hydrogen Lyman-α excitation, following [82]:

$$\left(\frac{dE^\gamma}{dV\,dt}\right)^{\rm dep}_{\rm exc} = \frac{1}{G(z)}\sum_{E_i=3\mathcal{R}/4}^{\mathcal{R}} E_i\,\mathbf{N}^\gamma_{\rm low}[E_i]\,. \tag{2.40}$$

A more complete treatment of excitation would involve keeping track of sub-13.6 eV energy photons as they redshift into the Lyman-α transition region at 10.2 eV, and

should also include two-photon excitation into the $2s$ state.[7] Finally, helium excitation has been neglected, since the de-excitation of helium atoms, which occurs quickly, produces photons that can eventually photoionize hydrogen. We therefore expect almost no net deposition of energy into helium excitation. Energy injection through helium excitation would mainly affect the process of helium recombination, when the probability of ionization after excitation to a higher state is significant due to the photon bath. However, we do not track this small effect, since the change to x_e would be very small. We leave a more careful treatment of excitation that can correctly take into account all of these effects to future work.

We now move on to ionization. All photons above $\mathcal{R} = 13.6\,\text{eV}$ that are included in N^{γ}_{low} have photoionized one of the atomic species (HI, HeI and HeII). However, after photoionizing a helium atom, the resulting ion may quickly recombine with an ambient free electron, producing an $\mathcal{R}_{\text{He}} = 24.6\,\text{eV}$ or $4\mathcal{R} = 54.4\,\text{eV}$ photon, which may then go on to photoionize hydrogen instead.[8]

We can handle low-energy photons with energy E_γ that photoionize neutral helium in one of the following three ways:

1. if helium is completely ignored, the photon is assumed to photoionize hydrogen, producing a low-energy electron with energy $E_\gamma - \mathcal{R}$ from photoionization and depositing $\mathcal{R}$ into hydrogen ionization. This is the approach used in previous calculations of $f_c(z)$ [82], but leaves us unable to self-consistently track x_{HeII} if desired;

2. the photon produces a low-energy electron with energy $E_\gamma - \mathcal{R}_{\text{He}}$ from photoionization, depositing $\mathcal{R}$ into hydrogen ionization from the recombination photon (with energy $\mathcal{R}_{\text{He}}$) and producing an electron with energy $\mathcal{R}_{\text{He}} - \mathcal{R}$, which ultimately deposits energy into hydrogen excitation, heating and sub-10.2 eV photons. This approach was previously discussed in Ref. [80], and found to

[7] Two-photon $1s \to 2s$ transitions are in fact as important as Lyman-α transitions near recombination in determining the ionization history, due to the fact that the Lyman-α line is optically thick at this time.

[8] The photoionization rate on neutral hydrogen is much faster than the Hubble rate for $x_{\text{HII}} \lesssim 0.9999$ for $z > 3$.

result in very little difference when compared to method (1); or

3. the photon produces a low-energy electron with energy $E_\gamma - \mathcal{R}_{\rm He}$ from photoionization and deposits $\mathcal{R}_{\rm He}$ into helium ionization.

The most accurate accounting of helium ionization lies somewhere between methods (2) and (3); however, either method will likely lead to very similar results in terms of x_e and T_m, since the bulk of the energy is deposited by the electron from the initial photoionization for photon energies $E_\gamma \gg \mathcal{R}_{\rm He}$, and the remaining energy always leads to one ionization event overall. `DarkHistory` offers the choice of these three options for implementing helium ionization.

We have checked that all three methods lead to similar ionization and temperature histories for DM models over a large range of masses decaying to both e^+e^- and $\gamma\gamma$; these checks are shown in Appendix A.3. We recommend simply using method (1) with helium turned off if the user is interested in ionization and temperature histories well before reionization, and using both method (2) and (3) with helium turned on to bracket the uncertainties associated with energy deposition on helium if the user is interested in the epoch of reionization.

To summarize, the amount of deposited energy into hydrogen per unit time and volume is given by

$$\left(\frac{dE^\gamma}{dV\,dt}\right)^{\rm dep}_{\rm H_{ion}} = \frac{\mathcal{R}}{G(z)} \sum_{E_i > \mathcal{R}} q^\gamma_{\rm H}[E_i] \mathbf{N}^\gamma_{\rm low}[E_i]\,, \tag{2.41}$$

and into helium ionization by:

$$\left(\frac{dE^\gamma}{dV\,dt}\right)^{\rm dep}_{\rm He_{ion}} = \frac{\mathcal{R}_{\rm He}}{G(z)} \sum_{E_i > \mathcal{R}_{\rm He}} q^\gamma_{\rm He}[E_i] \mathbf{N}^\gamma_{\rm low}[E_i]\,, \tag{2.42}$$

Method	$q^{\gamma}_{\rm H}$	$q^{e}_{\rm H}$	$q^{\gamma}_{\rm He}$	$q^{e}_{\rm He,a}$	$q^{e}_{\rm He,b}$
1	1	1	0	0	0
2	1	q	0	$1-q$	$1-q$
3	q	q	$1-q$	$1-q$	0

Table 2.1: List of q-coefficients for use in Eqs. (2.41)–(2.43). The variable q is defined in Eq. (2.44).

producing a low-energy electron spectrum after photoionization of

$$\begin{aligned}\mathbf{N}^{e}_{\rm ion}[E_i] = {} & q^{e}_{\rm H}(E_i+\mathcal{R})\mathbf{N}^{\gamma}_{\rm low}[E_i+\mathcal{R}] \\ & + q^{e}_{\rm He,a}(E_i+\mathcal{R}_{\rm He})\mathbf{N}^{\gamma}_{\rm low}[E_i+\mathcal{R}_{\rm He}] \\ & + \delta[E_i-\mathcal{R}_{\rm He}+\mathcal{R}]\sum_j q^{e}_{\rm He,b}(E_j)\mathbf{N}^{\gamma}_{\rm low}[E_j]\,, \end{aligned} \tag{2.43}$$

where $\delta[E_i-\mathcal{R}_{\rm He}+\mathcal{R}]$ is one when the bin boundaries span the energy $\mathcal{R}_{\rm He}-\mathcal{R}$ and is zero otherwise, and

$$q(E_i) \equiv \begin{cases} \frac{n_{\rm HI}\sigma_{\rm HI}(E_i)}{n_{\rm HI}\sigma_{\rm HI}(E_i)+n_{\rm HeI}\sigma_{\rm HeI}(E_i)}, & E_i > \mathcal{R}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.44}$$

with the σ's denoting the photoionization cross section of the appropriate species. $\mathbf{N}^{e}_{\rm ion}$ is added to the low-energy electron spectrum, $\mathbf{N}^{e}_{\rm low}$, which is then treated in the next section. The values of the q-coefficients depend on the method, and are shown in Table 2.1.

Electrons

To compute how low-energy electrons deposit their energy into the different channels, we use the results obtained by the MEDEA code [78, 79], following a similar treatment to Ref. [80]. Although `DarkHistory` also includes a calculation of electron energy deposition, which we discussed in Sec. 2.3.4, the MEDEA results are more accurate in the sub-3 keV electron energy range, including a more detailed accounting of all

possible atomic processes (such as $2s \to 1s$ deexcitations) and with more up-to-date cross sections. However, at mildly nonrelativistic to mildly relativistic regimes, our calculation of ICS is more accurate, as argued in Sec. 2.3.4. Furthermore, the MEDEA results assume that hydrogen and helium are at similar ionization levels, which is not always a good assumption. In future versions of `DarkHistory`, an improved treatment of electrons may be a useful addition to the code.

The MEDEA code uses a Monte Carlo method to track high-energy electrons as they are injected into the IGM, and determines the fraction of the initial electron energy deposited into ionization, Lyman-α excitation, heating of the gas and sub-10.2 eV photons. We use a table of these energy deposition fractions $p_c(E_i, x_{e,j})$ [80], where $c \in \{\text{'H}_{\text{ion}}\text{'}, \text{'He}_{\text{ion}}\text{'}, \text{'exc'}, \text{'heat'}, \text{'cont'}\}$ as before, $x_{e,j}$ ranges between 0 and 1, and E_i ranges between 14 eV and 3 keV, and perform an interpolation over these values. The energy deposition from electrons is then simply given by

$$\left(\frac{dE^e}{dV\,dt}\right)^{\text{dep}}_{\text{c}} = \frac{1}{G(z)} \sum_i p_c(E_i, x_e) E_i \, \mathbf{N}^e_{\text{low}}[E_i] \,, \tag{2.45}$$

keeping in mind that $\mathbf{N}^e_{\text{ion}}$ has already been added to $\mathbf{N}^e_{\text{low}}$. Between energies of 10.2 eV and 13.6 eV, where collisional excitations of hydrogen are possible but not ionization, we use the result at 14 eV, but setting the component into hydrogen ionization to zero and normalizing to unit probability. Below 10.2 eV, electrons can only deposit energy through Coulomb heating.

High-Energy Deposition

Finally, the high-energy deposition component of the total energy deposited is given by:

$$\left(\frac{dE^{\text{high}}}{dV\,dt}\right)^{\text{dep}}_{c} = \frac{1}{G(z)} E_c^{\text{high}} \,, \tag{2.46}$$

where $c \in \{$ 'ion', 'exc', 'heat' $\}$. Here, we add the high-energy excitation and ionization component to Lyman-α excitation and hydrogen ionization for simplicity, even

though the high-energy deposition is computed for all atomic species. A more accurate computation of this together with a more consistent treatment of helium ionization will be a potential improvement in a future version of `DarkHistory`.

With the rate of energy deposition through both low-energy photons and low-energy electrons computed, the total energy deposition rate is then straightforwardly given by

$$\left(\frac{dE}{dV\,dt}\right)_c^{\text{dep}} = \sum_\alpha \left(\frac{dE^\alpha}{dV\,dt}\right)_c^{\text{dep}}, \tag{2.47}$$

where $\alpha \in \{\gamma, e, \text{high}\}$.

2.3.7 TLA Integration and Reionization

`DarkHistory` offers several options for which set of assumptions should be used when integrating the ionization and thermal histories. In the simplest case, the user may integrate Eq. (2.6) at each redshift step based on the $f_c(z, \mathbf{x})$ calculated above, with the reionization terms switched off. As we have discussed, including this backreaction is already a significantly better treatment compared to calculations which assume a standard recombination history, i.e. using $f_c(z, \mathbf{x}_{\text{std}}(z))$ (although backreaction can also be switched off within `DarkHistory`).

The next significant improvement that is implemented within `DarkHistory` is the tracking of the neutral helium ionization fraction. Well before reionization, neglecting helium is a good approximation, since the number density of helium nuclei is only $\mathcal{F}_{\text{He}} \simeq 0.08$ of hydrogen, and we should expect only at most an 8% correction to x_e if we include helium.

However, tracking helium allows us to accomplish a self-consistent modeling of exotic energy injection and the reionization of hydrogen and neutral helium. `DarkHistory` allows users to input a model of reionization, for the first time extending the validity of these energy injection calculations into a regime where hydrogen is fully ionized and helium is singly ionized.

Helium

The `DarkHistory` evolution equation governing helium without any energy injection is identical to the `RECFAST` model, and is given by [73]

$$\begin{aligned}\dot{x}^{(0)}_{\text{HeII}} = \mathcal{C}^s_{\text{HeI}}\big(&x_{\text{HeII}} x_e n_{\text{H}} \alpha^s_{\text{HeI}} \\ &- \beta^s_{\text{HeI}}(\mathcal{F}_{\text{He}} - x_{\text{HeII}}) e^{-E^{s,\text{He}}_{21}/T_{\text{CMB}}}\big) \\ + \mathcal{C}^t_{\text{HeI}}\big(&x_{\text{HeII}} x_e n_{\text{H}} \alpha^t_{\text{HeI}} \\ &- 3\beta^t_{\text{HeI}}(\mathcal{F}_{\text{He}} - x_{\text{HeII}}) e^{-E^{t,\text{He}}_{21}/T_{\text{CMB}}}\big)\,. \end{aligned} \tag{2.48}$$

The singlet and triplet ground states of helium must be treated separately, and terms relevant to the singlet or triplet state are represented with a superscript s or t respectively. Here, α_{HeI} and β_{HeI} are the recombination and photoionization for HeI, E^{He}_{21} represents the energy difference between the corresponding $n = 1$ and $n = 2$ states, and finally $\mathcal{C}_{\text{HeI}}$ is the analog to the Peebles-C coefficient found in Eq. (2.1), representing the probability of a helium atom in the $n = 2$ state decaying to either the singlet or triplet ground state before photoionization can occur. The reader should refer to Refs. [73, 88, 89] for details on the numerical values of the coefficients, as well as how to compute $\mathcal{C}_{\text{HeI}}$.

We emphasize that although we have implemented all of the modifications to the standard TLA in Eq. (2.1), our code should not be used for high-precision cosmology, given that it has not been tested extensively, e.g. with different cosmological parameters from the central values used in `DarkHistory`. We find that our code agrees to within 3% of the `RECFAST` x_e values for the cosmological parameters used here, which is sufficient for computing the effects of exotic energy injection at this stage.

In the presence of exotic sources of energy injection, low-energy photons and electrons can also change the helium ionization level. Once again, we express the energy injection source term as

$$\dot{x}^{\text{inj}}_{\text{HeII}} = \frac{f_{\text{He ion}}(z, \mathbf{x})}{\mathcal{R}_{\text{He}} n_{\text{H}}} \left(\frac{dE}{dV\, dt}\right)^{\text{inj}}, \tag{2.49}$$

where $\mathcal{R}_{\rm He} = 24.6\,\text{eV}$ is the ionization potential of neutral helium. As we discussed in Sec. 2.3.6, there are three different methods available to evaluate $f_{\text{He}_{\text{ion}}}$ which bracket the uncertainties involved in helium ionization.

To summarize, the user may opt to track the change in helium ionization levels. This means that in addition to Eq. (2.6), we also include

$$\dot{x}_{\rm HeII} = \dot{x}^{(0)}_{\rm HeII} + \dot{x}^{\rm inj}_{\rm HeII} + \dot{x}^{\rm re}_{\rm HeII}\,, \tag{2.50}$$

where $\dot{x}^{\rm re}_{\rm HeII}$ is the contribution from processes that are active during reionization.

Reionization

The evolution equations shown in Eqs. (2.6) and (2.50) can be integrated with all reionization terms switched off if the user is primarily interested in temperatures or ionization levels well before reionization starts at $z \sim 20$. In this regime, turning off helium is also a reasonable approximation.

With reionization however, the helium ionization level should be solved as well for complete consistency. We solve the TLA differential equations shown in Eqs. (2.6) and (2.50) in two separate redshift regimes. Prior to some user-defined reionization redshift $1+z_{\rm re}$ ($z_{\rm re} \leq 50$), we set $\dot{T}^{\rm re}_m$, $\dot{x}^{\rm re}_{\rm HII}$ and $\dot{x}^{\rm re}_{\rm HeII}$ to zero. Once reionization begins, we set $\dot{x}^{(0)}_{\rm HII}$ and $\dot{x}^{(0)}_{\rm HeII}$ to zero for $z < z_{\rm re}$ instead, switching over to the specified reionization model with its own photoionization and recombination rates.[9] We also begin tracking doubly-ionized helium $x_{\rm HeIII}$, which is always assumed to be zero before reionization.

The $\dot{T}^{\rm re}_m$, $\dot{x}^{\rm re}_{\rm HII}$ and $\dot{x}^{\rm re}_{\rm HeII}$ terms depend on the details of how reionization proceeds, which is still relatively uncertain. However, choosing a model for the formation of stars and active galactic nuclei (AGNs) and the associated photoionization and photoheating rates, these terms can be evaluated. `DarkHistory` by default includes the Puchwein+ model of Ref. [90]. We also demonstrate how to implement the older

[9] We do not set $\dot{T}^{(0)}_m = 0$, since both adiabatic cooling and Compton scattering off the CMB remain active during reionization.

Madau and Haardt model [91] in Example 8. Both models provide a photoionization rate $\Gamma^{\rm ion}_{\gamma X}(z)$ and a photoheating rate $\mathcal{H}^{\rm ion}_{\gamma X}(z)$ as a function of redshift and species X.

Along with these energy injection rates, we must also include other relevant processes that alter the ionization fraction of each species. Since these processes generally convert kinetic energy to atomic binding energy, cooling or heating of the gas due to these processes must also be included in $\dot{T}^{\rm re}_m$. The processes we include are:

1. collisional ionization, occuring at a rate $\Gamma^{\rm ion}_{eX}$ for each species X, and an associated cooling rate $-\mathcal{H}^{\rm ion}_{eX}$;

2. case-A recombination, described by a rate coefficient $\alpha_{A,X}$ for each species X, and an associated cooling rate $-\mathcal{H}^{\rm rec}_X$;

3. collisional excitation cooling, with a rate $-\mathcal{H}^{\rm exc}_{eX}$; and

4. bremsstrahlung cooling, with a rate $-\mathcal{H}^{\rm br}$.

The cooling rates here have been defined with a negative sign so that all quantities denoted by $\mathcal{H}$ contribute positively to any temperature change. Expressions for all of these rates can be found in Ref. [92]. They are explicitly dependent on the ionization fraction of all three of the relevant species, namely $x_{\rm HI}$, $x_{\rm HeI}$ and $x_{\rm HeII}$. The full expressions for the evolution of each of these fractions are as follows:

$$\begin{aligned}
\dot{x}^{\rm re}_{\rm HII} &= x_{\rm HI}\left(\Gamma^{\rm ion}_{\gamma\rm HI} + n_e\Gamma^{\rm ion}_{e\rm HI}\right) - x_{\rm HII}n_e\alpha_{A,\rm HI}\,,\\
\dot{x}^{\rm re}_{\rm HeII} &= x_{\rm HeI}\left(\Gamma^{\rm ion}_{\gamma\rm HeI} + n_e\Gamma^{\rm ion}_{e\rm HeI}\right) + x_{\rm HeIII}n_e\alpha_{A,\rm HeIII}\\
&\quad - x_{\rm HeII}\left(\Gamma^{\rm ion}_{\gamma\rm HeII} + n_e\Gamma^{\rm ion}_{e\rm HeII} + n_e\alpha_{A,\rm HeII}\right)\,,\\
\dot{x}^{\rm re}_{\rm HeIII} &= x_{\rm HeII}\left(\Gamma^{\rm ion}_{\gamma\rm HeII} + n_e\Gamma^{\rm ion}_{e\rm HeII} - x_{\rm HeIII}n_e\alpha_{A,\rm HeIII}\right)\,,
\end{aligned} \tag{2.51}$$

with the temperature evolution given by

$$\begin{aligned}
\dot{T}^{\rm re}_m &= \frac{2}{3(1+\mathcal{F}_{\rm He}+x_e)n_{\rm H}}\\
&\quad\times\sum_X\left(\mathcal{H}^{\rm ion}_{eX} + \mathcal{H}^{\rm rec}_X + \mathcal{H}^{\rm exc}_{eX} + \mathcal{H}^{\rm br}\right)\,.
\end{aligned} \tag{2.52}$$

Instead of specifying a full reionization model, the user may also choose the simpler alternative of fixing the value of x_{HII} and x_{HeII} as a function of redshift once reionization begins, and integrate the temperature evolution alone instead. We note that this approach is not self-consistent, since fixing the ionization levels forces us to neglect any additional contribution to ionization from exotic energy injection sources. However, if the contribution to ionization is known to be small, this can serve as a useful approximation.

Numerical Integration

To ensure that ionization fractions always remain appropriately bounded during integration, we introduce the variable

$$\zeta_i \equiv \text{arctanh}\left[\frac{2}{\chi_i}\left(\frac{n_i}{n_{\text{H}}} - \frac{\chi_i}{2}\right)\right], \tag{2.53}$$

where $\chi_i = 1$ for HI and $\chi_i = \mathcal{F}_{\text{He}}$ for HeI and HeII. This transformed equation is then integrated using the standard `odeint` integrator provided by SciPy.

At early times, the equations we are integrating are very stiff, and solving them directly with numerical integration can often run into difficulties. We therefore assume that when $x_{\text{HII}} > 0.99$ or $x_{\text{HeII}} > 0.99\mathcal{F}_{\text{He}}$, either variable follows their Saha equilibrium values.

In Sec. 2.5.4, we will show several thermal and ionization histories that showcase `DarkHistory`'s capabilities in tracking the helium ionization level, exotic energy injection and reionization all at the same time.

2.4 Modules

In this section we summarize the main modules in `DarkHistory`. We will pay particular attention to the modules shown in the flow chart in Fig. 2-1, and as far as possible provide links between the code and the text. Keep in mind that this is not a complete list and that it is subject to change in future versions of the code. There is

more thorough documentation in `DarkHistory` itself that will be periodically updated at https://darkhistory.readthedocs.io, and will contain a more complete explanation of the code. In the interest of space, we only provide the full path of each module in the code when it is mentioned for the first time.

2.4.1 Data

First, the user must download the data files found at https://doi.org/10.7910/DVN/DUOUWA. These files contain the photon propagation transfer function $\overline{\mathsf{P}}_\gamma$ and deposition transfer functions $\overline{\mathsf{D}}_\gamma$, $\overline{\mathsf{D}}_e$ and $\overline{\mathbf{D}}_c^{\text{high}}$, which have all been precomputed as discussed above. They also contain transfer functions for ICS calculations discussed in Appendix A.1, structure formation annihilation boost factors computed in Ref. [30], the baseline thermal and ionization histories, data from PPPC4DMID [33, 93] and $f_c(z)$ computed without backreaction for DM annihilation and decay, where photons and e^+e^- are injected at a fixed set of energies.

2.4.2 `config`

The `config` module contains the code required to access the downloaded data, and to store them in memory for use. Users should ensure that the variable `data_path` points to the directory containing the data files.

2.4.3 `main`

The `main` module contains the function that implements the loop shown in Fig. 2-1, `evolve()`. The usage of this function will be discussed in great detail in Sec. 2.5.

2.4.4 `darkhistory.physics`

This module contains physical constants and useful functions found in cosmology, particle physics and atomic physics. We use units of cm for length, s for time and eV for energy, mass and temperature. Some examples of functions that are included in this

module include the Hubble parameter as a function of redshift, `physics.hubble()`, and the Peebles-C factor $\mathcal{C}$ found in Eq. (2.1), `physics.peebles_C()`. Constants provided in this module are taken from central values of the Planck 2018 TT,TE,EE+lowE results [18] and the Particle Data Group review of particle physics [94].

2.4.5 `darkhistory.electrons`

The `electrons` module contains all of the functions necessary to perform the electron cooling calculation. The `positronium` submodule contains functions that return the spectrum of photons obtained during positronium annihilation, which we denoted as $\overline{\mathbf{N}}_\gamma^{\text{pos}}$ in Eq. (2.21); Example 7 demonstrates how to use this module. The `ics` submodule contains all of the machinery necessary to compute the ICS scattered photon and electron spectra; for more details on how to use this submodule, refer to Example 4 in the code.

`elec_cooling` contains the code necessary to compute the transfer functions $\overline{\mathbf{R}}_c$, $\overline{\mathsf{T}}_{\text{ICS}}$ and $\overline{\mathsf{T}}_e$, as defined in Eqs. (2.13), (2.16) and (2.17) respectively; Example 6 shows how this module is used.

2.4.6 `darkhistory.history`

This module contains our implementation of the TLA and reionization. The submodule `tla` corresponds to Sec. 2.2 where the function `get_history` implements the TLA, including all of the terms discussed in Eqs. (2.6) and Eqs. (2.50)–(2.52). The submodule `reionization` contains the Puchwein+ reionization model, and contains all of the coefficients found in Eqs. (2.51) and (2.52).

2.4.7 `darkhistory.low_energy`

This module calculates $f_c(z)$. The `lowE_photons` and `lowE_electrons` submodules correspond to Sec. 2.3.6 and Sec. 2.3.6, respectively, implementing Eqs. (2.39)–(2.43) and Eq. (2.45) respectively. The `lowE_deposition` submodule then combines the

energy deposited by photons, electrons (including photoionized electrons) and high-energy deposition to make $f_c(z, \mathbf{x})$.

2.4.8 `darkhistory.spec`

This module contains functions for handling and generating spectra and transfer functions. All one dimensional spectra in the code can be handled using the class `Spectrum`, which stores not just the data of the spectrum, but also the abscissa, and other relevant information like redshift or the injection energy of the particle that produced the spectrum. This class includes many convenience functions, such as the ability to rebin the spectrum into a new binning while conserving total number and energy, or the ability to quickly obtain the total number of particles within some energy range. Example 1 in our code gives a quick introduction to this class.

The user may also want to store closely related spectra in one object. This may be desirable for spectra of the same particle type over different redshifts, or if they correspond to spectra from the same injected particle but at different injection energies. The class `Spectra` has been written to do exactly this. Example 2 provides a good overview of what this class can do.

2.4.9 `darkhistory.spec.pppc`

Within the `spec` module, a dedicated submodule `pppc` has been written to calculate the electron and photon spectra from the injection of any arbitrary Standard Model particle, based on the PPPC4DMID results. The function `pppc.get_pppc_spec()` is the main function to use for this end. See Example 4 for more information on how to use this function.

2.5 Using the Code

We will now apply `DarkHistory` to perform a variety of calculations in order to highlight the key functionalities of the code. Each of the subsections corresponds to

an example Jupyter notebook that has been provided as part of the code; the user should refer to these examples for a deeper look at the full capability of the code, as well as to the online documentation. In this chapter, we will simply highlight capabilities and interesting physics results.

Within the code and in this section, the word "redshift" and variables that represent redshift (usually called `rs` in the code) refer to the quantity $1 + z$, since this is the physically relevant quantity in many cosmological calculations.

2.5.1 A Simple Model: $\chi\chi \to b\bar{b}$

As a first example, we will demonstrate how to compute the ionization and thermal history of a simple annihilation model. Consider a 50 GeV Majorana fermion DM particle that undergoes s-wave annihilation to a pair of $b\bar{b}$ quarks, with an annihilation cross section $\langle\sigma v\rangle = 2 \times 10^{-26}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$, close to the required thermal freezeout cross section for the correct relic abundance. Similar models have been considered as a possible dark matter explanation for the galactic center excess [95] and the AMS-02 antiproton excess [96, 97]. We perform the calculation in a relatively simplified setting: with no reionization, no backreaction included, but with a boost to the annihilation rate from structure formation. For more details, see Example 9 in the code.

The function that we use to compute histories is `main.evolve()`. There are many keyword parameters that can be used with this function, and the user should refer to the example notebooks and the online documentation for more information. To find the thermal history for this model, `evolve()` can be called in the following fashion:

```
import main
import darkhistory.physics as phys

bbbar_noBR = main.evolve(
    DM_process='swave', mDM=50e9,
    sigmav=2e-26,
```

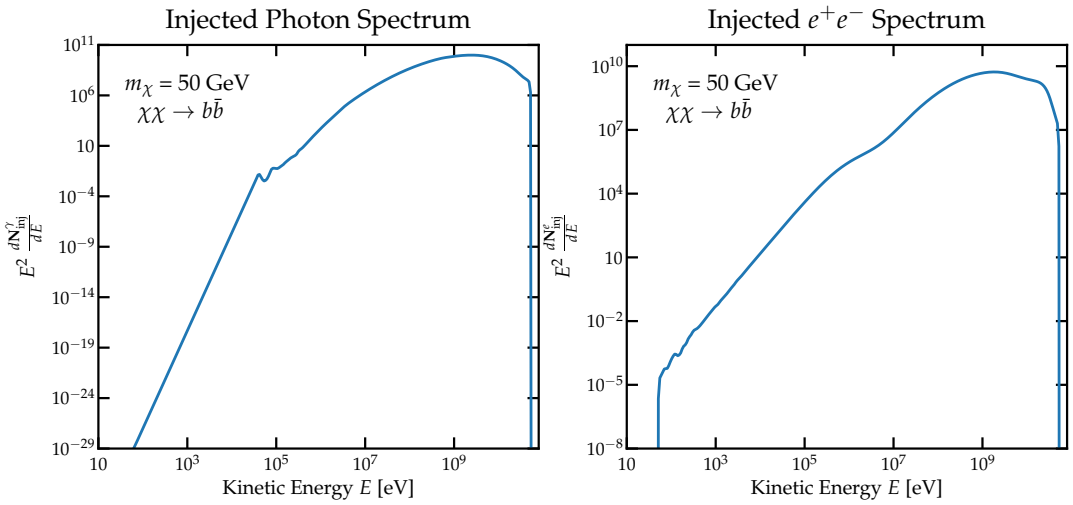


Figure 2-2: Photon (left) and e^+e^- (right) spectra produced by a single annihilation event, $\chi\chi \to b\bar{b}$, with $m_\chi = 50\,\text{GeV}$. These spectra are based on the raw data provided by PPPC4DMID.

```
    primary='b', start_rs=3000.,
    coarsen_factor=32, backreaction=False,
    struct_boost=phys.struct_boost_func()
)
```

The keyword parameters are as follows:

1. `DM_process='swave'` – specifies the DM process of interest. Currently, `DarkHistory` can handle s-wave annihilating and decaying DM models (`DM_process='decay'`) with this keyword;

2. `mDM=50e9` – specifies the DM mass, in eV;

3. `sigmav=2e-26` – specifies the velocity averaged annihilation cross section, in $\text{cm}^3\,\text{s}^{-1}$;

4. `primary='b'` – specifies the annihilation channel. The options include all of those offered by PPPC4DMID [33, 93], and the spectra are extracted from the raw data provided by the cookbook. The e^+e^- and photon spectra from the showering of a single $b\bar{b}$ pair are shown in Fig. 2-2. These are proportional to

the injection spectra $\mathbf{N}_{\text{inj}}^{\alpha}$ defined in Sec. 2.3.3, and can be generated using the function `pppc.get_pppc_spec()`;

5. `start_rs=3000` – the redshift at which to start the evaluation. $1 + z = 3000$ is the highest redshift at which we have produced the photon cooling transfer functions, and represents the highest redshift that should be specified here. In this example, `start_rs` fixes the initial conditions of the TLA in Eq. (2.1) at the baseline ionization and temperature values at this redshift;

6. `coarsen_factor=32` – the coarsening factor, defined in Sec. 2.3.5. For a comparison between solutions with different coarsening factors, see Appendix A.3;

7. `backreaction=False` – this turns backreaction on and off; and

8. `struct_boost=phys.struct_boost_func()` – the structure formation prescription to use. Once dark matter halos start to collapse, the annihilation rate gets enhanced by the factor

$$1 + \mathcal{B}(z) \equiv \frac{\langle \rho_\chi^2 \rangle}{\langle \rho_\chi \rangle^2} \tag{2.54}$$

compared to the smooth annihilation rate shown in Eq. (2.3). Here, `struct_boost` is a function that takes redshift as the argument, and returns $1 + \mathcal{B}(z)$. The user can make use of the structure formation boosts that are saved by default in `DarkHistory` in the `physics` module, which include the boost factors computed in Ref. [30], and is used as the default boost factor by `struct_boost_func()`.

By default, the solver integrates the equations down to $1 + z = 4$, and will not evolve the helium ionization levels. These choices can of course be changed with other keyword parameters. Note that the function is not limited to DM processes or PPPC4DMID spectra; other keyword parameters allow the user to specify their own injection rates as a function of redshift (see the documentation for the keyword parameters `rate_func_N` and `rate_func_eng`), along with the spectra of photons

and e^+e^- injected (see the documentation for the keyword parameters `in_spec_elec` and `in_spec_phot`).

The output of `evolve()`, stored in `bbbar_noBR`, is a dictionary containing the redshift abscissa of the solutions, the ionization and temperature solutions, the propagating photon, low-energy photon and low-energy electron spectra, and the computed value of $f_c(z)$. To access the redshift, ionization and temperature, we can simply do:

```
# Redshift abscissa.
rs_vec  = bbbar_noBR['rs']
# Matter temperature in eV.
Tm_vec  = bbbar_noBR['Tm']
# Ionization fraction.
# Stored as 1+z by {xHII, xHeII, xHeIII}.
xHII_vec = bbbar_noBR['x'][:,0]
```

In Fig. 2-3 we plot T_m and $x_{\rm HII}$ as a function of redshift for the $\chi\chi \to b\bar{b}$ model. For DM masses above $\gtrsim 10\,\text{GeV}$, values of $\langle\sigma v\rangle$ required for thermal freezeout are unconstrained by the CMB anisotropy power spectrum energy injection constraints: the ionization fraction, which changes by approximately 25% only at high redshifts, does not change enough to affect the power spectrum significantly. The sudden increase in ionization and temperature at $z \sim 30$ corresponds to an increase in the boost factor used (halos with an Einasto profile with halo substructure boost included [30], found in `physics.struct_boost_func()`).

We also show in Fig. 2-3 for completeness the effect of turning on backreaction, i.e. including the effect of the increased ionization level on the evolution of the ionization and thermal histories. This is conveniently done by setting `backreaction=True`. In this particular example, the effect of backreaction is small, but we will show more scenarios where backreaction has large effect on T_m, and explain why this can be significant in the next example.

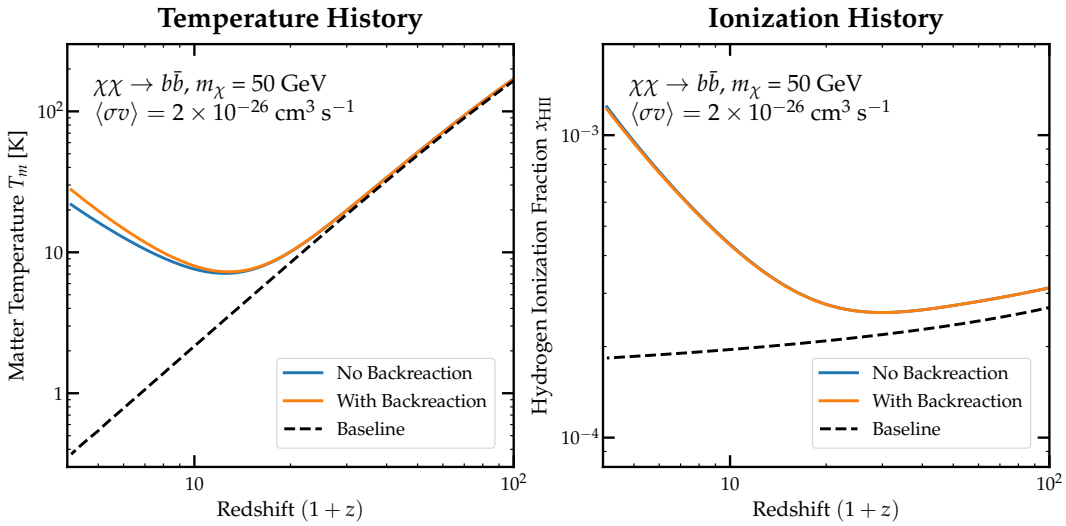


Figure 2-3: Matter temperature T_m (left) and hydrogen ionization fraction $x_{\rm HII}$ (right) solved in the presence of dark matter annihilation into $b\bar{b}$ pairs using `DarkHistory` . Eq. 2.6 is solved without dark matter energy injection to produce the baseline histories (black, dashed), with energy injection but without backreaction (blue), and with dark matter annihilation and backreaction (orange). We assume a dark matter mass of 50 GeV and a velocity averaged annihilation cross section of 2×10^{-26} cm^3 s^{-1}.

2.5.2 Backreaction

Let us explore the effects of backreaction a bit more using some of the code found in Example 10 of `DarkHistory`. As was described in Sec 2.2, one of `DarkHistory`'s main improvements to ionization and temperature history calculations is its ability to include the effects of back-reaction. To see its importance, consider the example of 100 MeV dark matter decaying to a pair of e^+e^-, with a lifetime of $\tau = 3 \times 10^{25}$ s, a value that is close to the minimum lifetime allowed by constraints from the CMB power spectrum [59]. The ionization and thermal histories can be evaluated in this way:

```
decay_BR = main.evolve(
  DM_process='decay', mDM=1e8, lifetime=3e25,
  primary='elec_delta', start_rs=3000.,
  coarsen_factor=16, backreaction=True
)
```

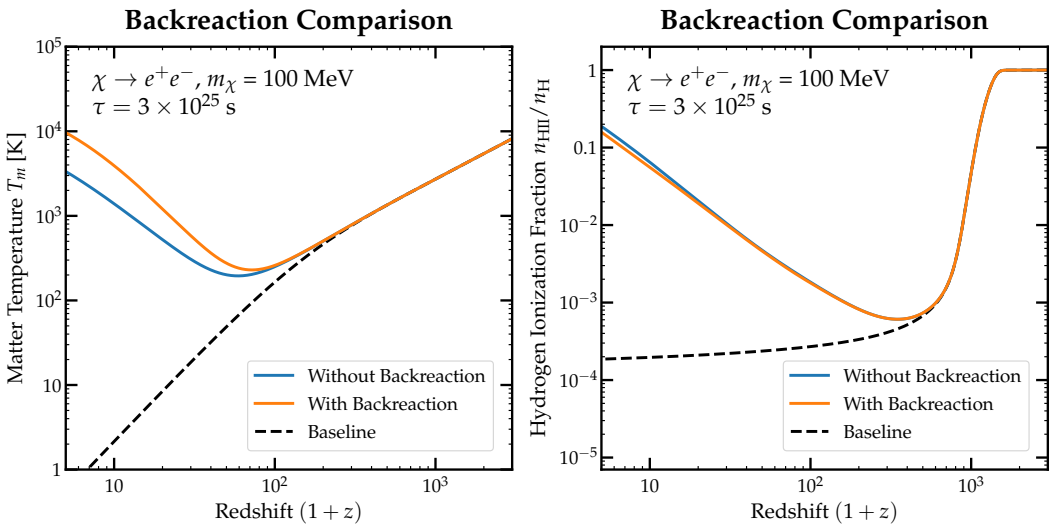


Figure 2-4: Temperature (left) and ionization (right) histories including the effects of dark matter decay to electrons and positrons. We choose a lifetime of 3×10^{25} s, which is consistent with the CMB constraints from Ref. [59]. We plot the baseline histories (black, dashed), the histories including dark matter energy injection but not backreaction (blue), and the histories including energy injection and backreaction (orange). These plots are a single vertical slice of the contour plots in Fig. 2-5. Additionally, these plots constitute a cross-check on `DarkHistory`, as they agree well with similar results obtained in Ref. [30].

The new keywords here are:

1. `DM_process='decay'` – specifies the DM process of interest to be decays;
2. `lifetime=3e25` – specifies the decay lifetime in seconds; and
3. `primary='elec_delta'` – the `primary` channel options `'elec_delta'` and `'phot_delta'` can be used to inject an e^+e^- and $\gamma\gamma$ pair respectively, with no electroweak corrections applied.

To do the calculation without backreaction, we can simply set `backreaction=False`. However, with `primary='elec_delta'` or `'phot_delta'`, `DarkHistory` can instead rely on tabulated results of $f_c(z)$ for these two channels, using the same method based on results from Ref. [82], to calculate the ionization and thermal histories without evolving the input spectrum, leading to a significant speed-up. This can be done using the function `tla.get_history()`:

```
import numpy as np
from darkhistory.tla import get_history
# get_history takes a redshift vector:
rs_vec = np.flipud(np.arange(5, 3000, 0.1))

result = get_history(
   rs_vec, baseline_f=True, mDM=1e8,
   lifetime=3e25, DM_process='decay',
   inj_particle='elec_delta'
)
```

with the following parameters:

1. `rs_vec` – the redshift vector, ordered from high to low, over which the temperature and ionization histories are to be evaluated;
2. `baseline_f=True` – this tells the code to use the baseline $f_c(z)$ computed by `DarkHistory` without backreaction. As we discussed in Sec. 2.3.6, these $f_c(z)$ agree with those computed in Ref. [82] to within 10%, and

3. `inj_particle='elec_delta'` – used to specify one of two options `'elec_delta'` or `'phot_delta'`.

The output `result` is an array of shape `(len(rs_vec), 4)`, with the second dimension indexing $\{T_m, x_{\text{HII}}, x_{\text{HeII}}, x_{\text{HeIII}}\}$. The temperature (in eV) can be accessed through `T_m = results[-1,0]`.

Although only the $f_c(z)$ values for the injection for an e^+e^- and $\gamma\gamma$ pair have been saved for use with `DarkHistory`, the $f_c(z)$ for any arbitrary channel can be computed from a weighted average of the electron and photon results [82]. We stress once again, however, that this can only be done assuming no backreaction.

The histories are shown in Fig 2-4, with and without backreaction turned on. First, even though the ionization level at $z \sim 10$ is three orders of magnitude larger than the baseline, such a scenario is actually still consistent with the CMB power spectrum constraints, owing to the fact that the ionization build-up occurs relatively late: the CMB constraints are sensitive to changes in x_e near recombination, and become less sensitive at later times.

Comparing the temperature histories with and without backreaction, we see that the main effect of this increase in x_e on the energy deposition processes is to increase energy deposition into heating. Ionization and excitation rates depend on the neutral fraction, which is still close to 100% even with energy deposition from DM. However, the energy rate into Coulomb heating is proportional to x_e, so taking into account the significantly elevated x_e values leads to higher temperature levels. By about $z \sim 10$, T_m with backreaction is larger than without backreaction by a factor of ~ 4, with the difference continuing to grow. Neglecting backreaction therefore leads to a severe underestimate of T_m, and including this effect consistently will certainly be important in understanding what measurements of T_m at $z \simeq 20$ through the 21-cm signal or the Lyman-α power spectrum can tell us about exotic sources of energy injection.

We can perform the calculation over a range of DM masses by looping over values of `mDM`. For each value of m_χ, we select the minimum lifetime τ which is consistent with the CMB power spectrum constraints, and compare the difference between the temperature history with backreaction ($T_{m,\text{BR}}$) and without ($T_{m,0}$) by computing the

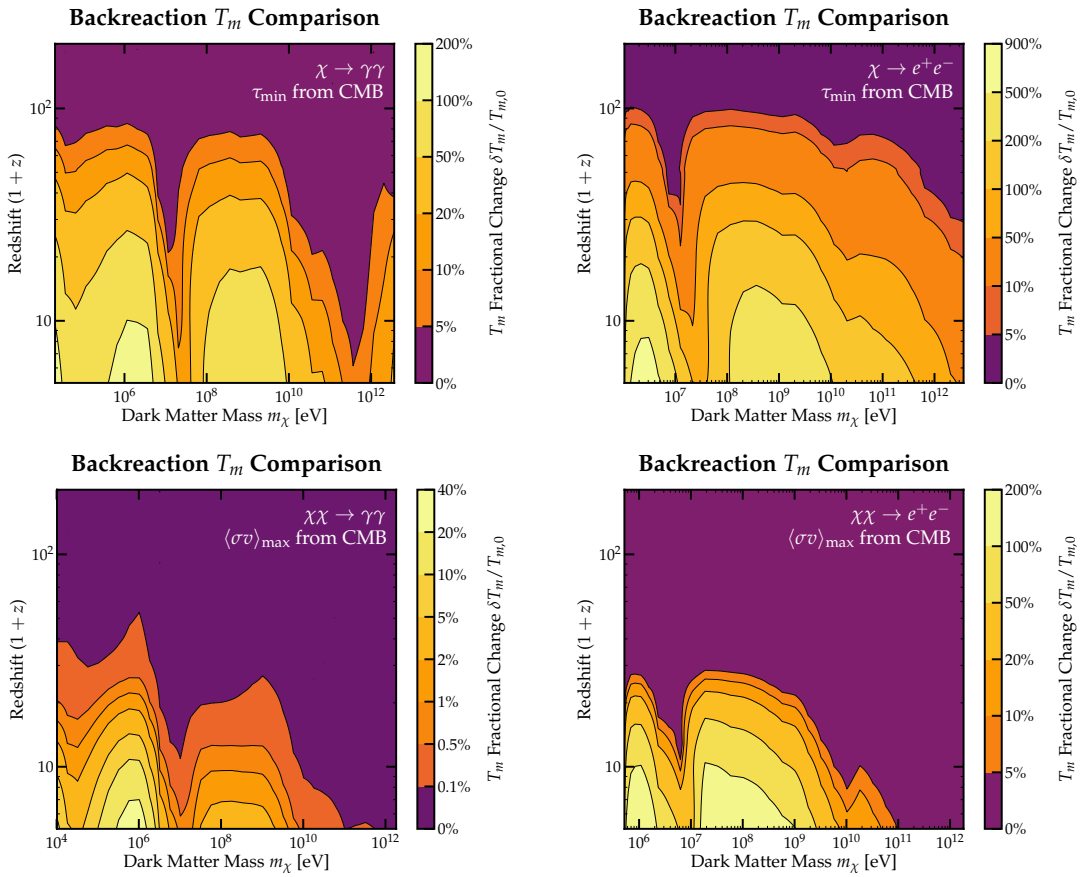


Figure 2-5: Contour plots of the fractional change in temperature $\delta T_m/T_{m,0}$ caused by including the effects of backreaction, as a function of dark matter mass and redshift (See Eq. (2.55)). For each dark matter mass, we choose the minimum τ or maximum $\langle\sigma v\rangle$ allowed by current CMB power spectrum constraints [59, 98].

fractional change in temperature,

$$\frac{\delta T_m}{T_{m,0}}(m_\chi, z) = \frac{T_{m,\rm BR}(m_\chi, z) - T_{m,0}(m_\chi, z)}{T_{m,0}(m_\chi, z)}\,. \tag{2.55}$$

In Fig 2-5 we plot this variable over a range of redshifts and dark matter masses for this particular channel ($\chi \to e^+e^-$), but also for decay and annihilation into e^+e^- and $\gamma\gamma$, taking the maximum $\langle\sigma v\rangle$ again allowed by the CMB power spectrum constraints. At a redshift of $z \sim 17$ near the end of the cosmic dark ages, $\delta T_m/T_{m,0} \sim 100\%$ (i.e. T_m with backreaction is a factor of 2 larger than without) or more can easily be obtained. Even larger deviations are possible at lower redshifts, depending on the channel under consideration.

2.5.3 21-cm Sensitivity

The global 21-cm signal is a measurement of the sky-averaged differential brightness temperature T_{21} with respect to the background radiation. Measurements of this signal would open a window into the ionization and temperature histories of the universe at the cosmic dawn (see e.g. Ref. [99] for a review of 21-cm cosmology). A first claim of such a measurement has already been made by the EDGES collaboration [100]. The brightness temperature of the 21-cm hydrogen absorption line relative to the background radiation temperature is given by [99]:

$$T_{21} \approx x_{\rm HI}(z) \left(\frac{0.15}{\Omega_m}\right)^{1/2} \left(\frac{\Omega_b h}{0.02}\right) \times \left(\frac{1+z}{10}\right)^{1/2} \left[1 - \frac{T_R(z)}{T_S(z)}\right] 23\,{\rm mK}\,, \tag{2.56}$$

where Ω_b is the baryon energy density today as a fraction of the critical density, h is the Hubble parameter today in $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, T_R is the background radiation temperature (typically assumed to be the CMB temperature) and T_S is the spin temperature of neutral hydrogen as a function of redshift, which determines the relative population of neutral hydrogen in the two hyperfine states. Due to the presence of an intense Lyman-α radiation field once stars begin to form, it is expected that $T_S \approx T_m$ at the cosmic dawn. This fact allows us to turn the 21-cm global signal into a limit on T_m itself, assuming that $T_R = T_{\rm CMB}$.

We will focus on $1 + z \approx 18$, roughly the central value of the absorption trough measured by EDGES [100]. At this redshift, almost all hydrogen is neutral, i.e. $x_{\rm HI} \approx 1$, and we can invert Eq. (2.56) to find T_S as a function of T_{21}. Since $T_m < T_S$, this yields the bound

$$T_m(z = 17) < \left(1 - \frac{T_{21}}{35\,{\rm mK}}\right)^{-1} 49\,{\rm K}\,. \tag{2.57}$$

This temperature bound in turn puts a limit on the DM decay lifetime or cross-section because too much dark matter decay/annihilation would heat up T_m past this point.

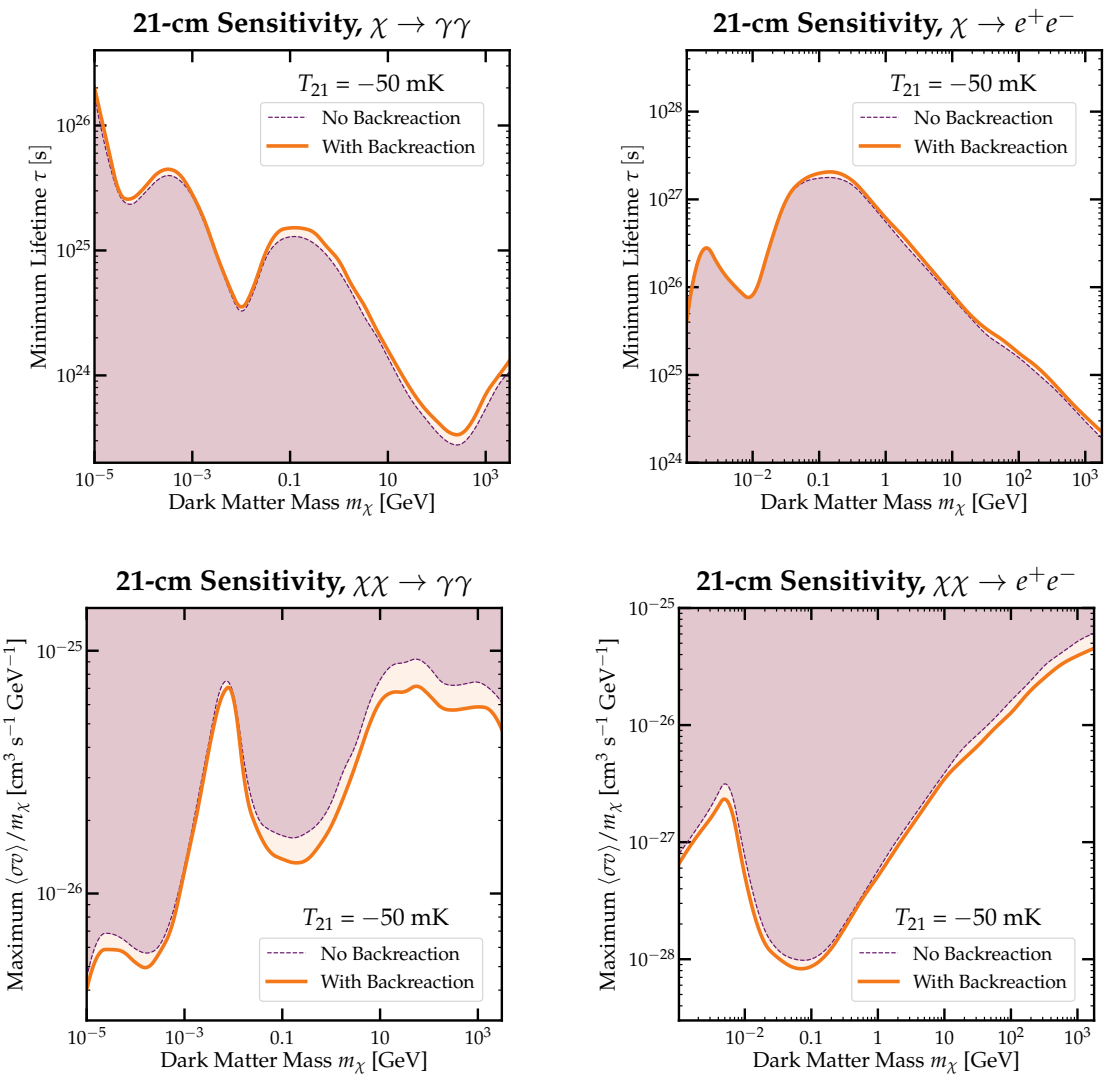


Figure 2-6: The minimum dark matter decay lifetime (top row) and maximum annihilation cross section (bottom row) bounds, derived from the global 21-cm signal. We assume a differential 21-cm brightness temperature of $T_{21} = -50\,\mathrm{mK}$, corresponding to a maximum T_m of about 20.3 K at $z \sim 17$. We consider decay and annihilation into $\gamma\gamma$ (left column) and e^+e^- (right column) and compute the bounds with (orange, solid) and without (purple, dashed) backreaction.

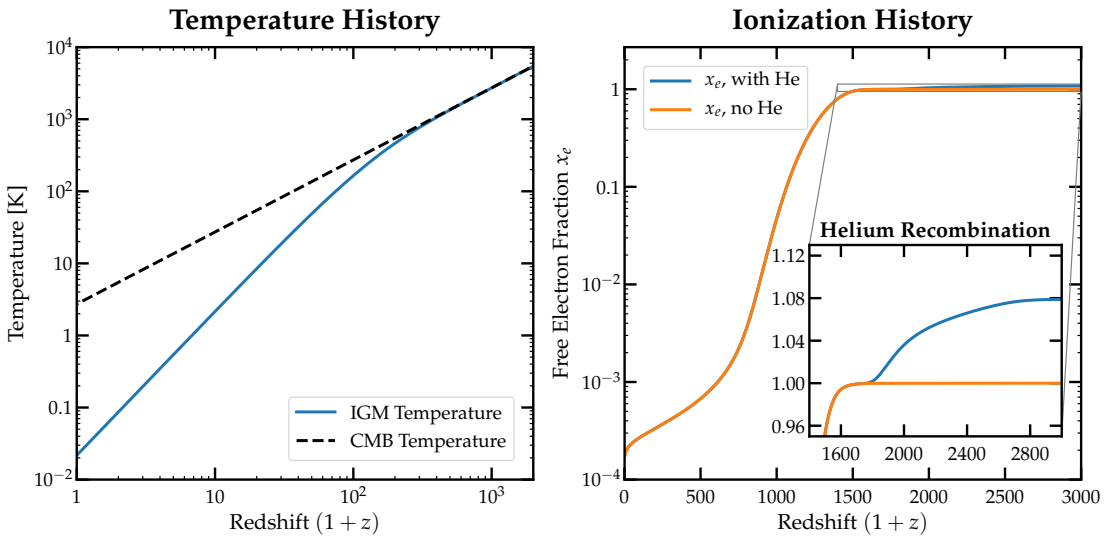


Figure 2-7: Temperature (left) and free electron fraction x_e (right) as a function of redshift. x_e is solved in `DarkHistory` with (blue) and without (orange) helium; both options lead to a similar temperature history (blue). With helium, helium recombination is correctly computed (inset). The CMB temperature is shown (black, dashed) for reference.

In contrast to the CMB power spectrum energy injection bounds, which is most sensitive to changes in x_e around the time of recombination, the 21-cm global signal constraints are more sensitive to energy injection processes that are more active at late times, and are dependent primarily on T_m instead. Since T_m is significantly impacted by including the effects of backreaction, the calculation performed by `DarkHistory` becomes important for setting accurate constraints using the 21-cm global signal.

To illustrate this, we perform a simple sensitivity study by obtaining the constraints for a measured T_{21} of $-50\,\text{mK}$, and compare the constraints with and without backreaction taken into account. Although this value of T_{21} is inconsistent with the EDGES experiment, it is impossible to interpret the EDGES result without proposing new physics that may be at play during the cosmic dark ages [62], which is a more complicated task and less relevant to helping users understand the code. The following analysis is worked out in more detail within the code in Example 11.

$T_{21} = -50\,\text{mK}$ means that we require $T_m < 20.3\,\text{K}$ according to Eq. (2.57). We once again scan over a grid of dark matter masses and lifetimes/cross-sections decaying/annihilating into e^+e^- and $\gamma\gamma$, using `get_history()` for the case with no

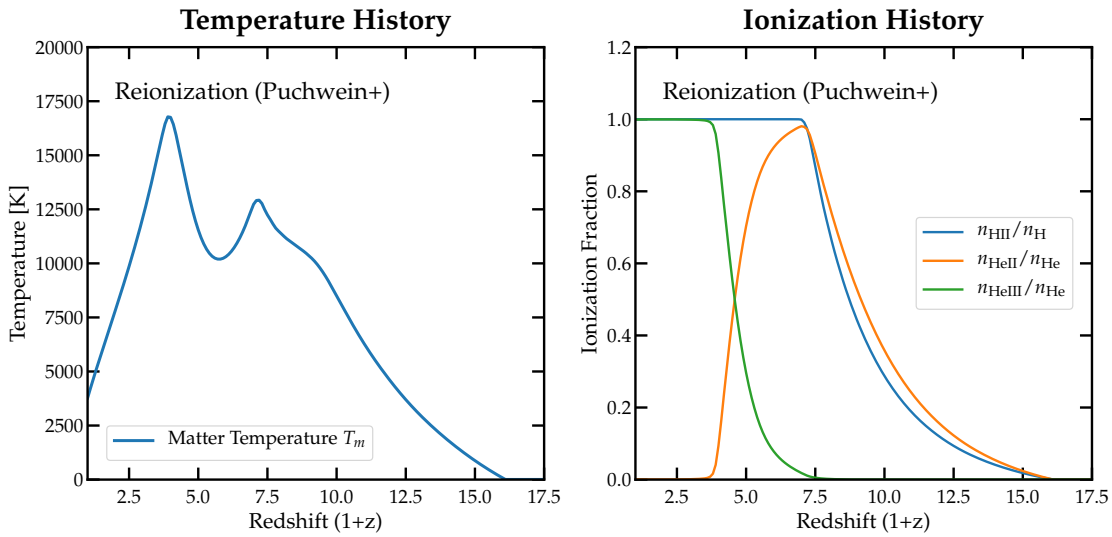


Figure 2-8: Temperature (left) and free electron fraction (right) as a function of redshift, solved in `DarkHistory` with the default Puchwein+ reionization model [90]. The IGM temperature (blue) is shown on the left, while the ionization fractions $n_{\rm HII}/n_{\rm H}$ (blue), $n_{\rm HeII}/n_{\rm He}$ (orange) and $n_{\rm HeIII}/n_{\rm He}$ (green) are shown as well. These results agree very well with the same plots shown in Ref. [90].

backreaction and `evolve()` for the case with backreaction, as explained in the previous section, to find where in parameter space dark matter energy injection leads to a violation of Eq. (2.57).

The resulting exclusion plots are shown in Fig. 2-6. We see that in each case the calculation with backreaction can be between 10%-50% stronger than without backreaction, which we would expect because backreaction leads to larger temperatures. We emphasize that this is the result for just one chosen value of T_{21}; for larger (less negative) T_{21}, we expect that the importance of backreaction will increase, since the energy injection is less constrained, allowing for larger values of x_e.

2.5.4 Helium, Dark Matter and Reionization

Finally, we will take a closer look at the different options one can use within the code to evaluate temperature and ionization histories. Throughout this section, we will demonstrate these different options mostly using `get_history()`, but similar options are also available in `evolve()`, which calls `get_history()` with all of the relevant

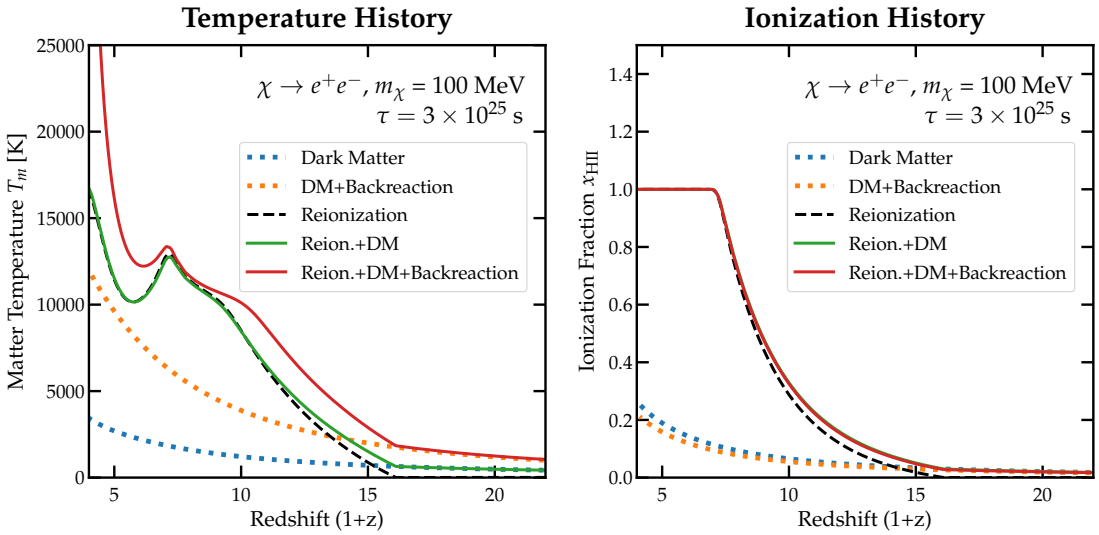


Figure 2-9: Temperature (left) and hydrogen ionization (right) history of the universe with DM decay and the default reionization model. The DM has a mass of $m_\chi = 100\,\text{MeV}$ and decays to e^+e^- with a lifetime of $3 \times 10^{25}\,\text{s}$. The temperature and ionization with DM decay alone is shown without (blue, dotted) and with (orange, dotted) backreaction included. The combined effect of DM decay and reionization without (green) and with (red) backreaction can be compared to the reference reionization model (black, dashed).

options provided. We refer the reader to the online documentation and to Example 8 in the code for more details.

Without any exotic energy injection or any reionization, the function `get_history()` accepts a redshift vector, and simply returns the baseline ionization and temperature histories, obtained by solving Eq. (2.6):

```
import numpy as np
from darkhistory.tla import get_history

# Redshift vector in decreasing order.
rs_vec = np.flipud(np.arange(1., 3000., 0.1))
soln_baseline = get_history(rs_vec)
```

Turning on helium evolution within `get_history()` is controlled by the flag `helium_TLA`, i.e.

```
soln_He = get_history(rs_vec,helium_TLA=True)
```

Fig. 2-7 shows the solution to Eq. (2.1) with just the "(0)" terms, i.e. without any energy injection or reionization, and compares that solution to one with Eq. (2.50) added as well. This is simply the standard ionization history with helium recombination ($z \sim 1800$) and hydrogen recombination ($z \sim 1100$), eventually leading to the residual ionization fraction at redshifts well below hydrogen recombination of about $x_e \sim 2 \times 10^{-4}$. The inset of Fig. 2-7 shows that `DarkHistory` is able to correctly reproduce helium recombination; the entire ionization history agrees with `RECFAST` results at the central cosmological parameters used by `DarkHistory` to within $\sim 3\%$. We recommend that helium ionization levels are tracked when used in combination with reionization.

The next important option is whether to include the effects of reionization. This option is controlled by the flag `reion_switch`:

```
soln_default_reion = get_history(
 rs_vec, helium_TLA=True, reion_switch=True
)
```

With no other options set, setting `reion_switch` to `True` causes `DarkHistory` to use the standard reionization model, which is based on the photoionization and photoheating rates provided in [90]. Fig. 2-8 shows the IGM temperature as well as the ionization levels of the different atomic species as a function of redshift. Both of these results agree well with the same result shown in Ref. [90]. Reionization of hydrogen and neutral helium is complete by about $z \sim 6$; soon after, HeII starts to become doubly ionized, leading to a decrease in n_{HeII} and a corresponding increase in n_{HeIII}. Dips in T_m correspond to a decrease in photoheating rates once a species becomes completely ionized and the production of high-energy electrons from photoionization off these species ceases.

Aside from the default reionization model, the user may also supply their own reionization models in two different ways: by either providing their own photoionization and photoheating rates on each atomic species (e.g. based on a model that is different from the default, e.g. [91]), or by fixing the ionization history below a certain

redshift, e.g. with a tanh model [101, 102]. We leave a discussion of how to use these options to Example 8 in the code.

With the ability to include both helium and reionization, we can now add a new source of energy injection and compute the effects on ionization and temperature levels. We remind the reader that this means we are solving Eq. (2.6) together with Eq. (2.50). This is accomplished in the code with both `reion_switch` and `helium_TLA` set to `True`, and supplying the same keyword parameters used to inject energy from DM shown in Sec. 2.5.2. We can add decaying DM with mass 100 MeV into an e^+e^- pair with a lifetime of 3×10^{25} s, like so (using `evolve()` in this example):

```
main.evolve(
  DM_process='decay', mDM=1e8,
  lifetime=3e25, primary='elec_delta',
  start_rs=3000., coarsen_factor=1,
  backreaction=True, helium_TLA=True,
  reion_switch=True
)
```

By turning on and off the various flags `backreaction`, `helium_TLA` and `reion_switch`, we can produce histories including or excluding these various effects.

The results from different combinations of these switches are summarized in Fig. 2-9. The dashed lines shows the histories with DM decay only, and illustrates the significant difference that can arise after taking into account backreaction, which we have already seen in Fig. 2-4. Combining the DM energy injection with the reionization model gives the solid lines in Fig. 2-9. These curves should be compared to the default reionization model temperature and ionization histories, shown in the black, dashed lines. When computing the DM energy deposition without taking into account backreaction, we find that the amount of energy deposited into heating from DM is much smaller than heating from reionization processes once they begin in earnest, and so adding the DM decays on top of reionization produces only a small perturbation in the temperature history relative to T_m for just the reionization model

alone. In some cases, the addition of DM actually decreases T_m: this can happen due to reionization proceeding at a faster rate, leaving fewer atoms to photoionize and thus suppressing photoheating.

It is clear, however, that neglecting backreaction leads to a severe underestimation of the energy deposition into heating. Performing the full calculation with DM, reionization and backreaction correctly accounted for produces the line in red, which shows that the addition of DM significantly increases T_m compared to both the reionization model and the case where DM energy deposition is added without backreaction. Reionization greatly enhances the energy deposition rate into heating of the IGM by increasing the number of free charged particles available for Coulomb heating, and properly accounting for backreaction using `DarkHistory` is critical to predicting the IGM temperature growth due to energy injection once reionization begins.

2.6 Future Directions

An important future application for `DarkHistory` will be the computation of spectral distortions. The effects of DM energy injection have already been explored at very high redshifts in the form of μ-type distortions for $5 \times 10^4 \lesssim z \lesssim 2 \times 10^6$ and y-type distortions for lower redshifts [103], and also another class of so-called 'non-thermal relativistic' distortions at redshifts $1100 \lesssim z \lesssim 2 \times 10^5$ [104]. `DarkHistory` thus will be able to complement these earlier studies and calculate the spectral distortions coming from redshifts below recombination, $z \lesssim 1100$. As a demonstration of the type of calculation `DarkHistory` will perform, we show the component of this late-time spectral distortion that `DarkHistory` is currently able to compute in Fig. 2-10. We consider the case of DM annihilation into photon pairs, with 5 different masses and the maximum $\langle \sigma v \rangle$ allowed by CMB anisotropy measurements. We emphasize that this computed spectral distortion is incomplete. `DarkHistory` does not resolve the spectrum of photons produced during the cooling of sub-3 keV electrons, and it also does not account for the distortions induced by absorption or emission of photons due to atomic transitions. We leave the complete calculation of late-time spectral

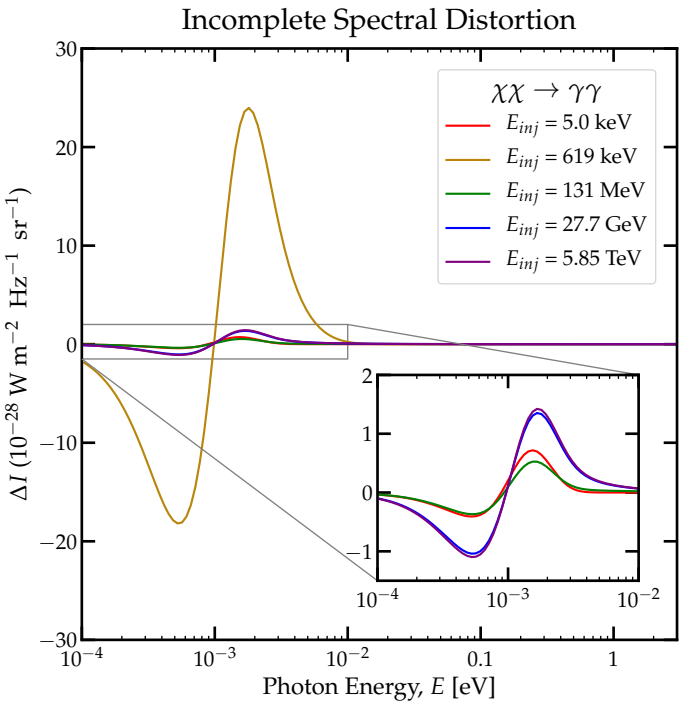


Figure 2-10: Calculation of the spectral distortion caused by DM annihilation into photon pairs during redshifts $4 < 1 + z < 3000$. We use 5 different DM masses and use the maximum $\langle \sigma v \rangle$ allowed by CMB anisotropy measurements [82]. Since `DarkHistory` currently does not resolve the spectrum of photons produced during the cooling of sub-3keV electrons, the distortion is incomplete and may be missing the dominant source of distortions.

distortions to future work.

Another future application of `DarkHistory` will be to compute the effects of DM energy injection on stars, galaxies, and other aspects of structure formation. Since these are features of an inhomogeneous universe, we will need to generalize `DarkHistory` to relax the assumption of homogeneity that is currently built into it. Once generalized, `DarkHistory` can be used to explore the effects of DM energy injection on the formation of H_2, and hence star formation, as in Ref. [105], on the growth of DM halos as in Ref. [106], on star fragmentation as in Ref. [107], and the effects on gas within halos as in Ref. [108].

2.7 Conclusion

We have developed and made public a new code package for mapping out the effects of arbitrary exotic energy injections — including dark matter annihilation and decay to arbitrary Standard Model final states — on the temperature and ionization history of the early universe. `DarkHistory` is capable of self-consistently including the effects of conventional astrophysical sources of ionization and heating, and of including feedback effects that can significantly enhance the degree of heating. Additionally, the ICS module can be employed independently of the rest of the code, as an accurate and efficient numerical calculator of ICS across a very wide range of electron and photon energies. We have outlined here a number of worked examples, and provide more examples with the online code at https://github.com/hongwanliu/DarkHistory.

`DarkHistory` has a modular framework and can in the future be improved in several different directions, while keeping the same underlying structure. In this first version we have focused on the homogeneous signal, and neglected the possible effect of new radiation backgrounds and/or gas inhomogeneities on the cascade of secondaries produced by injected high-energy particles. Such effects may become important in the late cosmic dark ages and the epoch of reionization. The spectrum of low-energy photons produced by energy injection, and the resulting distortion to the spectrum of the CMB, is a possible observable in its own right; the current version of `DarkHistory` provides only a partial calculation of this spectral distortion, due to our approximate treatment of low-energy electrons, but we intend to improve this aspect in future work. The effects of other new physics on the temperature/ionization evolution – in particular, scattering between baryons and DM – can be incorporated within the same framework. We also intend to explore the possibility of interfacing DarkHistory with existing public codes for computing the recombination history, perturbations to the CMB, and 21cm signals.

The tools we have developed in this work can be used to understand the visible imprints of exotic energy injections that could appear in the CMB and the 21cm line of neutral hydrogen, and hence to place precise constraints on dark matter annihilation

and decay. We hope they will help pave the way for a comprehensive description of the ways in which dark matter interactions, and other physics beyond the Standard Model, could reshape the early history of our cosmos.

Chapter 3

Lyman-α Constraints on Cosmic Heating from Dark Matter Annihilation and Decay

3.1 Introduction

Dark matter (DM) interactions such as annihilation or decay can inject a significant amount of energy into the early Universe, producing observable changes in both its ionization and temperature histories. Changes in the free electron fraction, for example, can alter the cosmic microwave background (CMB) anisotropy power spectrum [109–111], allowing constraints on the annihilation cross section [57, 58, 60, 80, 98, 112–118] and the decay lifetime of DM [59, 67, 119, 120] to be set using Planck data [121]. Constraints based on modifications to the temperature history focus on two redshift ranges where measurement data is or will potentially be available: *(i)* before hydrogen reionization at $z \sim 20$, and *(ii)* during the reionization epoch at $2 \lesssim z \lesssim 6$. In the former redshift range, the 21-cm global signal [62, 67, 68, 122–124] and power spectrum [61, 81] have been shown to be powerful probes of DM energy injection, and have the potential to be the leading constraint on the decay lifetime of sub-GeV DM [62]. In the latter range, measurements of the intergalactic

medium (IGM) temperature derived from Lyman-α flux power spectra [125, 126] and Lyman-α absorption features in quasar spectra [127, 128] have been used to constrain the s-wave annihilation cross section [129], the p-wave annihilation cross section, and the decay lifetime of DM [63, 67, 130]. The IGM temperature can also be used to set limits on the kinetic mixing parameter for ultralight dark photon DM [131–133], the strength of DM-baryon interactions [134], and the mass of primordial black hole DM [135].

In this chapter, we revisit the constraints on p-wave annihilating and decaying dark matter from the IGM temperature measurements during reionization. This work is timely for two reasons. First and foremost, the development of `DarkHistory` [53] allows us to improve on the results of Refs. [63, 129, 130] considerably. We can now self-consistently take into account the positive feedback that increased ionization levels have on the IGM heating efficiency of DM energy injection processes. This effect can give rise to large corrections in the predicted IGM temperature [53] during reionization. Furthermore, `DarkHistory` can solve for the temperature evolution of the IGM in the presence of both astrophysical reionization sources and dark matter energy injection; previous work only set constraints assuming no reionization [129] or a rudimentary treatment of reionization and the energy deposition efficiency [63, 130]. Second, experimental results published since Refs. [63, 129, 130] have considerably improved our knowledge of the Universe during and after reionization. These include:

1. *Planck constraints on reionization.* The low multipole moments of the Planck power spectrum provide information on the process of reionization [121]. In particular, Planck provides 68th and 95th percentiles for the ionization fraction in the range $6 \lesssim z \lesssim 30$ using three different models [136, 137], arriving at qualitatively similar results.

2. *New determinations of the IGM temperature.* By comparing mock Lyman-α power spectra produced by a large grid of hydrodynamical simulations to power spectra calculated [138] based on quasar spectra measured by BOSS [139], HIRES [140, 141], MIKE [142], and XQ-100 [143], Ref. [144] (hereafter Walther+)

determined the IGM temperature at mean density in the range $1.8 < z < 5.4$, overcoming a degeneracy between gas density and deduced temperature that hampered previous analyses [126, 145]. More recently, Ref. [146] (hereafter Gaikwad+) fit the observed width distribution of the Lyα transmission spikes to simulation results, enabling a determination of the IGM temperature at mean density in the $5.4 < z < 5.8$ redshift range, again with only a weak dependence on the temperature-density relation.

These improvements to both the understanding of energy deposition and the ionization/temperature histories are combined in our analysis into robust constraints on DM p-wave annihilation rates and decay lifetimes. These constraints are competitive in the light DM mass regime ($\lesssim 10\,\mathrm{GeV}$) with existing limits on DM decay from the CMB anisotropy power spectrum [59] and are complementary to indirect detection limits [147–150], being less sensitive to systematics associated with the galactic halo profile and interstellar cosmic ray propagation.

In the rest of this chapter, we introduce the IGM ionization and temperature evolution equations, discuss the data and statistical tests used, and finally present our new constraints. We also include Supplemental Materials that provide additional details to support our main text.

3.2 Ionization and temperature histories

In this section, we write down the equations governing the evolution of the IGM temperature, T_m, and the IGM hydrogen ionization level, $x_\mathrm{HII} \equiv n_\mathrm{HII}/n_\mathrm{H}$, where n_H is the number density of both neutral and ionized hydrogen. The ionization evolution equation is:

$$\dot{x}_\mathrm{HII} = \dot{x}_\mathrm{HII}^\mathrm{atom} + \dot{x}_\mathrm{HII}^\mathrm{DM} + \dot{x}_\mathrm{HII}^\star \,. \tag{3.1}$$

Here, $\dot{x}_\mathrm{HII}^\mathrm{atom}$ corresponds to atomic processes, i.e. recombination [38, 71, 72, 76] and collisional ionization, which depend in a straightforward way on the ionization and

temperature of the IGM, while $\dot{x}_{\rm HII}^{\rm DM}$ is the contribution to ionization from DM energy injection. These terms are discussed in detail in Ref. [53], and are given in full in the Supplemental Materials, as well as a completely analogous HeII evolution equation. The remaining term, $\dot{x}_{\rm HII}^{\star}$, corresponds to the contribution to photoionization from astrophysical sources of reionization. This term will inevitably source photoheating, which will be important for the IGM temperature evolution equation (discussed below). $\dot{x}_{\rm HII}^{\star}$ can in principle be determined given a model of astrophysical sources of reionization, but there are large uncertainties associated with these sources. For example, the fraction of ionizing photons that escape into the IGM from their galactic sites of production is highly uncertain, ranging from essentially 0 to 1 depending on the model [151].

Instead, we rely on the Planck constraints on the process of reionization to fix the form of $\dot{x}_{\rm e}$, allowing us to fix $\dot{x}_{\rm HII}^{\star}$ while remaining agnostic about astrophysical sources of reionization. Specifically, we begin by choosing a late time ionization history, $x_{\rm e}^{\rm Pl}(z)$ for $z < 30$, within the 95% confidence region determined using either the "Tanh" or "FlexKnot" model adopted by Planck [121]. We then make the common assumption that during hydrogen reionization HI and HeI have identical ionization fractions due to their similar ionizing potentials, but that helium remains only singly ionized due to HeII's deeper ionization potential [152]. These assumptions allow us to set $x_{\rm HII}^{\rm Pl} = x_{\rm e}^{\rm Pl}/(1+\chi)$, where $\chi \equiv n_{\rm He}/n_{\rm H}$ is the primordial ratio of helium atoms to hydrogen atoms. Given a choice of $x_{\rm e}^{\rm Pl}(z)$ we can then rearrange Eq. (3.1) to set

$$\dot{x}_{\rm HII}^{\star} = \left(\frac{\dot{x}_{\rm e}^{\rm Pl}}{1+\chi} - \dot{x}_{\rm HII}^{\rm atom} - \dot{x}_{\rm HII}^{\rm DM} \right) \theta(z^{\star} - z)\,, \tag{3.2}$$

where θ is a step function that enforces $\dot{x}_{\rm HII}^{\star} = 0$ at sufficiently early redshifts when astrophysical reionization sources do not exist yet. To fix $z^{\star}$, notice that at early times when $\dot{x}_{\rm HII}^{\star}$ is turned off, ionization due to DM energy injection produces $x_{\rm e}(z) \geq x_{\rm e}^{\rm Pl}(z)$. Since DM cannot significantly reionize the universe [130], there will exist a redshift past which $x_{\rm e}(z) < x_{\rm e}^{\rm Pl}(z)$ if we do not turn on $\dot{x}_{\rm HII}^{\star}$. We define $z^{\star}$ to be this cross-over redshift where $x_{\rm e}(z^{\star}) = x_{\rm e}^{\rm Pl}(z^{\star})$.

Thus, for any given DM model and $x_e^{\rm Pl}$ we can use Eq. (3.2) to construct ionization histories that self-consistently include the effects of DM energy injection and reionization simultaneously. We do not require the astrophysics that produces $\dot{x}_{\rm HII}^\star$ to obey any constraint other than $\dot{x}_{\rm HII}^\star \geq 0$, which maximizes freedom in the reionization model and leads to more conservative DM constraints.

The IGM temperature history can similarly be described by a differential equation:

$$\dot{T}_{\rm m} = \dot{T}_{\rm adia} + \dot{T}_{\rm C} + \dot{T}_{\rm DM} + \dot{T}_{\rm atom} + \dot{T}^\star \,, \tag{3.3}$$

where $\dot{T}_{\rm adia}$ is the adiabatic cooling term, $\dot{T}_{\rm C}$ is the heating/cooling term from Compton scattering with the CMB, $\dot{T}_{\rm DM}$ is the heating contribution from DM energy injection, and $\dot{T}_{\rm atom}$ comprises all relevant atomic cooling processes. These terms are also fully described in Ref. [53], and included in the Supplemental Materials for completeness. We stress that $\dot{T}_{\rm DM}$ is computed, using `DarkHistory` [53], as a function of both redshift and ionization fraction $x_{\rm e}$, self-consistently taking into account the strong dependence of $\dot{T}_{\rm DM}$ on $x_{\rm e}$, and strengthening the constraints we derive.

The remaining term, $\dot{T}^\star$, accounts for photoheating that accompanies the process of photoionization, as described in Eq. (3.2). We adopt two different prescriptions for treating the photoheating rate, which we name 'conservative' and 'photoheated'. In the 'conservative' treatment, we simply set $\dot{T}^\star = 0$. This treatment produces highly robust constraints on DM energy injection since the uncertainties of the reionization source modeling do not appear in our calculation. Any non-trivial model would only serve to increase the temperature of the IGM, strengthening our constraints.

In the 'photoheated' treatment, we implement a two-stage reionization model. In the first stage — prior to the completion of HI/HeI reionization — we follow a simple parametrization adopted in e.g. Refs. [144, 152, 153] and take $\dot{T}^\star = \dot{x}_{\rm HII}^\star (1+\chi)\Delta T$ for some constant ΔT. This parameter is expected to be within the range $2\times10^4\,{\rm K}$–$3\times10^4\,{\rm K}$ based on analytic arguments [154] and simulations [155, 156]. We will either restrict $\Delta T \geq 0$ or impose a physical prior of $\Delta T \geq 2\times10^4\,{\rm K}$ in what we call our 'photoheated-I' or 'photoheated-II' constraints, respectively.

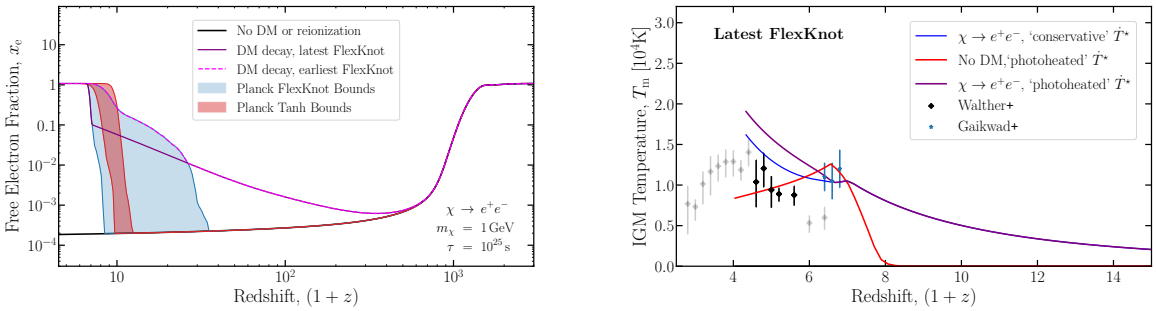


Figure 3-1: The ionization history *(Left)* and IGM temperature history *(Right)* as functions of redshift. The left plot shows the ionization history in the absence of DM energy injection and reionization sources (solid black), the 95% confidence region for Planck's FlexKnot (shaded blue) and Tanh (shaded red) reionization histories, and the ionization history in the presence of both DM energy injection and reionization sources that produce Planck's latest (solid purple) and earliest (dashed magenta) FlexKnot histories at late times. The right plot shows the temperature history assuming *(i)* DM decay and the 'conservative' treatment of $\dot{T}^\star$ (blue), *(ii)* the 'photoheated' treatment and no DM energy injection (red), and *(iii)* the 'photoheated' treatment with DM decay (purple). *(i)* and *(iii)* assume a DM mass of 1 GeV and decay to e^+e^- pairs with a lifetime of 10^{25} s while *(ii)* and *(iii)* assume the latest FlexKnot reionization history and use parameter values $(\Delta T, \alpha_{\rm bk}) = (24\,665\,\mathrm{K}, 0.57)$ and $(0\,\mathrm{K}, 1.5)$, respectively. Also included are the data from Ref. [144] (black diamonds) and Ref. [146] (blue stars), where the solid data constitute our fiducial data set.

In the second stage — after reionization is complete — the IGM becomes optically thin. In this regime, reionization-only models find that the IGM is, to a good approximation, in photoionization equilibrium [157]. The photoheating rate in this limit is specified completely by the spectral index $\alpha_{\rm bk}$ of the average specific intensity J_ν [with units $\mathrm{eV\,s^{-1}\,Hz^{-1}\,sr^{-1}\,cm^{-2}}$] of the ionizing background near the HI ionization threshold, i.e. $J_\nu \propto \nu^{-\alpha_{\rm bk}}$ [153, 156]. By considering a range of reionization source models and using measurements of the column-density distribution of intergalactic hydrogen absorbers, the authors of Ref. [156] bracketed the range of $\alpha_{\rm bk}$ to be within $-0.5 < \alpha_{\rm bk} < 1.5$, which we will use in our analysis.

In summary, the 'photoheated' prescription is

$$\dot{T}^\star = \begin{cases} \dot{x}^\star_{\rm HII}(1+\chi)\Delta T\,, & x_{\rm HII} < 0.99\,, \\ \displaystyle\sum_{i\in\{\rm H,He\}} \frac{E_{i\rm I} x_i}{3(\gamma_{i\rm I} - 1 + \alpha_{\rm bk})} \alpha_{{\rm A},i\rm I} n_{\rm H}\,, & x_{\rm HII} \geq 0.99\,, \end{cases} \tag{3.4}$$

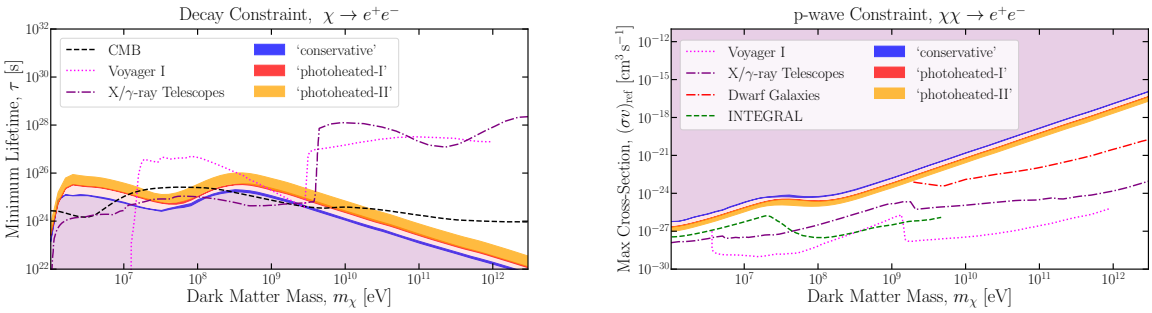


Figure 3-2: Constraints for decay (left) or p-wave annihilation (right) to e^+e^- pairs with $v_{\rm ref} = 100\,\mathrm{km\,s^{-1}}$. We show our constraints using the 'conservative' (blue band), 'photoheated-I' (red band), or 'photoheated-II' (orange band) treatment. The darkly shaded bands show the variation of our constraints as we vary through the 95% confidence regions of Planck's Tanh and FlexKnot reionization models. We also include constraints from the CMB [59] (dashed-black), X/γ-ray telescopes [147, 148, 159] (dot-dashed purple), INTEGRAL [160] where we have assumed $\langle v^2 \rangle = 220\,\mathrm{km\,s^{-1}}$ in the Milky Way, Voyager I [149, 150] (dotted pink), and gamma-ray observations of dwarf galaxies [161] (dot-dashed red).

where i runs over H and He (thus x_i =1, $n_{\rm He}/n_{\rm H}$), and for species i, $E_{i\rm I}$ is the ionization potential, $\gamma_{i\rm I}$ denotes the power-law index for the photoionization cross-section at threshold, and $\alpha_{A,i\rm I}$ is the case-A recombination coefficient [156]. The 'photoheated' model is therefore fully specified by two parameters, ΔT and $\alpha_{\rm bk}$. Additionally, once HI/HeI reionization is complete, we set $1 - x_{\rm e} = 4 \times 10^{-5}$, which is approximately its measured value [158]. This small fraction of neutral HI and HeI atoms dramatically decreases the photoionization rate relative to its pre-reionization value for photons of energy $13.6\,\mathrm{eV} < E_\gamma < 54.4\,\mathrm{eV}$ injected by DM. Consequently, there is a non-negligible unabsorbed fraction of photons in each timestep, $\exp\left(-\sum_{i\in\{\rm HI,HeI\}} n_i \sigma_i^{\rm ion}(E_\gamma)\Delta t\right)$, where $\sigma_i^{\rm ion}(E_\gamma)$ is the photoionization cross-section for species i at photon energy E_γ. We modify `DarkHistory` to propagate these photons to the next timestep.

To demonstrate the effects of DM energy injection and our reionization modeling, we show in Fig. 3-1 example histories obtained by integrating Eqs. (3.1) and (3.3) for both the 'conservative' and 'photoheated' treatments, with and without DM decay. The left plot shows how our method can produce ionization histories that both take into account the extra ionization caused by DM energy injection and also vary over Planck's 95% confidence region for the late-time ionization levels. In the right panel,

we assume the Planck FlexKnot curve with the latest reionization, and show in red our best fit temperature history assuming no DM energy injection, with the 'photoheated' treatment. This history is a good fit to the fiducial data, with a total χ^2 of about 5. Additionally, once DM is added we show a model that is just consistent with our (95% confidence) 'conservative' constraints but ruled out by the 'photoheated' constraints.

3.3 Comparison with data

We compare our computed temperature histories with IGM temperature data obtained from Walther+ [144] within the range $1.8 < z < 5.4$ and Gaikwad+ [146] within $5.4 < z < 5.8$. To construct our fiducial IGM temperature dataset, we only consider data points with redshifts $z > 3.6$ (see Fig. 3-1, solid data points) since these redshifts are well separated from the redshift of full HeII reionization [126], allowing us to safely use the transfer functions that `DarkHistory` currently uses, which assume $x_{\text{HeIII}} = 0$. By neglecting HeII reionization and its significant heating of the IGM [151] we derive more conservative constraints. Additionally, the two Walther+ data points above $z = 4.6$ are in tension with the Gaikwad+ result; we discard them in favor of the higher T_{m} values reported by Gaikwad+, since this results in less stringent limits.

To assess the agreement between a computed temperature history and our fiducial temperature dataset using our 'conservative' method, we perform a modified χ^2 test. Specifically, our test statistic only penalizes DM models that overheat the IGM relative to the data, which accounts for the fact that any non-trivial photoheating model would only result in less agreement with the data, whereas DM models that underheat the IGM could be brought into agreement with the data given a specific photoheating model. We define the following test statistic for the ith IGM temperature bin:

$$\text{TS}_i = \begin{cases} 0\,, & T_{i,\text{pred}} < T_{i,\text{data}}\,, \\ \left(\dfrac{T_{i,\text{pred}} - T_{i,\text{data}}}{\sigma_{i,\text{data}}}\right)^2\,, & T_{i,\text{pred}} \geq T_{i,\text{data}}\,, \end{cases} \tag{3.5}$$

where $T_{i,\text{data}}$ is the fiducial IGM temperature measurement, $T_{i,\text{pred}}$ is the predicted

IGM temperature given a DM model and photoheating prescription, and $\sigma_{i,\rm data}$ is the 1σ upper error bar from the fiducial IGM temperature data. We then construct a global test statistic for all of the bins, simply given by $\mathrm{TS} = \sum_i \mathrm{TS}_i$. Assuming the data points $\{T_{i,\rm data}\}$ are each independent, Gaussian random variables with standard deviation given by $\sigma_{i,\rm data}$, the probability density function of TS given some model $\{T_{i,\rm pred}\}$ is given by

$$f(\mathrm{TS}|\{T_{i,\rm pred}\}) \quad = \quad \frac{1}{2^N}\sum_{n=0}^{N}\frac{N!}{n!(N-n)!}f_{\chi^2}(\mathrm{TS};n)\,. \quad (3.6)$$

N is the total number of temperature bins and $f_{\chi^2}(x;n)$ is the χ^2-distribution with argument x and number of degrees-of-freedom n, where the $n = 0$ case is defined to be a Dirac delta function, $f_{\chi^2}(x;0) \equiv \delta(x)$. The hypothesis that the data $\{T_{i,\rm data}\}$ is consistent with the $\{T_{i,\rm pred}\}$ can then be accepted or rejected at the 95% confidence level based on Eq. (3.6). See the Supplemental Materials for more details.

For our 'photoheated' constraints, we perform a standard χ^2 goodness-of-fit test. For any given DM model we marginalize over the photoheating model parameters by finding the ΔT and $\alpha_{\rm bk}$ values that minimize the total χ^2 subject to the constraints $\Delta T \geq 0$ ('photoheated-I') or $2 \times 10^4\,\mathrm{K}$ ('photoheated-II') and $-0.5 < \alpha_{\rm bk} < 1.5$. We then accept or reject DM models at the 95% confidence level using a χ^2 test with 6 degrees of freedom (8 data points - 2 model parameters).

Fig. 3-2 shows constraints for two classes of DM models: DM that decays or p-wave annihilates to e^+e^-. Our p-wave annihilation cross-section is defined by $\sigma v = (\sigma v)_{\rm ref} \times (v/v_{\rm ref})^2$ with $v_{\rm ref} = 100\,\mathrm{km\,s^{-1}}$. We also use the NFW boost factor for p-wave annihilation calculated in Ref. [130]. Although we only show constraints for e^+e^- final states, our method applies to any other final state (see the Supplemental Materials). The blue, red, and orange regions are excluded by our 'conservative,' 'photoheated-I,' and 'photoheated-II' constraints, respectively. The 'photoheated' limits are generally a factor of $2-8$ times stronger than the 'conservative' constraints.

The thickness of the darkly shaded bands correspond to the variation in the constraints when we vary $x_{\rm e}^{\rm Pl}$ in Eq. (3.2) over the 95% confidence region of Planck's Flex-

Knot and Tanh late-time ionization curves. The 'conservative' and 'photoheated-I' bands are narrow, demonstrating that the uncertainty in the late-time ionization curve is not an important uncertainty for these treatments. However, the 'photoheated-II' treatment shows a larger spread, since the larger values of ΔT imposed by the prior significantly increase the rate of heating at $z \sim 6$, making the earliest temperature data points more constraining, and increasing the sensitivity to the ionization history at $z \simeq 6$. A better understanding of the process of reionization could therefore enhance our constraints significantly.

Our 'conservative' constraints for decay to e^+e^- are the strongest constraints in the DM mass range $\sim 1\,\mathrm{MeV} - 10\,\mathrm{MeV}$ and competitive at around 1 GeV while our p-wave constraints are competitive in the range $\sim 1\,\mathrm{MeV} - 10\,\mathrm{MeV}$. For higher masses, constraints from Voyager I observations of interstellar cosmic rays are orders of magnitude stronger for both p-wave [150] and decay [149]. Constraints from X/γ-ray telescopes [147, 148, 159, 161] are stronger than ours for $m_\chi > 1\,\mathrm{GeV}$ and comparable for $m_\chi < 1\,\mathrm{GeV}$.

Importantly, all three types of constraints are affected by different systematics. The telescope constraints are affected by uncertainties in our galactic halo profile while Voyager's are affected by uncertainties in cosmic ray propagation. The p-wave boost factor is relatively insensitive to many details of structure formation, since it is dominated by the largest DM halos, which are well resolved in simulations (see the Supplemental Materials). A more important systematic comes from our assumption of homogeneity. We assume that energy injected into the IGM spreads quickly and is deposited homogeneously, when in reality injected particles may be unable to efficiently escape their sites of production within halos [31, 32]. We leave a detailed exploration of these inhomogeneity effects for future work.

3.4 Conclusion

We have described a method to self-consistently construct ionization and IGM temperature histories in the presence of reionization sources and DM energy injection

by utilizing Planck's measurement of the late-time ionization level of the IGM. We construct two types of constraints for models of DM decay and p-wave annihilation. For the first 'conservative' type of constraint, we assume that reionization sources can ionize the IGM but not heat it, resulting in constraints that are robust to the uncertainties of reionization. For the second 'photoheated' type of constraint, we use a simple but well-motivated photoheating model that gives stronger limits than the 'conservative' constraints by roughly a factor of $2 - 8$. We expect that as the uncertainties on the IGM temperature measurements shrink, and as reionization and photoheating models become more constrained, these 'photoheated' constraints will strengthen considerably.

Chapter 4

Thermal Squeezeout of Dark Matter

4.1 Introduction

The cosmic abundance of dark matter (DM) is comparable to the abundance of Standard Model (SM) particles up to an $\mathcal{O}(1)$ factor [18]. The similarity of these two ostensibly unrelated abundances raises the suspicion that the two sectors may have been in chemical equilibrium at some point in their history, implying some sort of interaction portal between the SM and DM. Numerous experimental efforts to look for such a portal are under way. Nonetheless, the particle nature of DM and any potential portals to the SM remain unknown at present.

In addition to probing interactions between DM and the SM, many experimental and theoretical efforts aim to probe possible dynamics within the dark sector itself. Often simplified dark sectors with only a single DM particle are considered. Yet, the rich gauge structure of the SM offers no particular reason to believe that the dark sector will be significantly simpler. A wide range of more involved dark sectors have been studied, especially scenarios with a new confining force in the dark sector; see for instance Refs. [22, 162–194]. Depending on the details of the sector, different hadronic states can be the DM candidate in different theories and the stabilizing symmetry and the DM mass scale can vary widely [189]. Such a sector can further give rise to rich dynamics that can potentially solve other problems in the SM as well, e.g. see Refs. [166, 195] where the observed baryon asymmetry is tied to the DM abundance.

An interesting class of confining dark sector models is the scenario where all the

dark quarks are substantially heavier than the dark confinement scale, Λ. These models and their experimental signals have been studied extensively, see for instance Refs. [166, 183–185, 187]. For sufficiently heavy quarks, lattice calculations have shown that the phase transition in such a sector is of first-order for $SU(N)$ with $N = 3$ [46, 196–199] or $N > 3$ [48, 200]. There has been a recent surge of interest in the study of the potential effects of first-order phase transitions in other DM models, e.g. see Refs. [201–204][1], but the effects of the phase transition on the relic abundance of dark matter in confining dark sector models are mostly unexplored, with the exception of a recent study of dark sectors with only light quarks ($m_q \leqslant \Lambda$) [208].

In this work, we consider the simplest such confining model – an $SU(3)$ gauge theory with one heavy quark in the fundamental representation – and focus on the effects of the first-order phase transition on the DM relic abundance calculation. Similar to the arguments put forward in Ref. [209], we will argue that toward the end of the phase transition we will be left with pockets of the high temperature, i.e. deconfined, phase submerged in a sea of low temperature, i.e. confined, phase. We will argue that the heavy quarks are all initially trapped inside these contracting pockets.[2] To determine important properties of these pockets such as their initial characteristic size and contraction rate, we will develop a simplified model to numerically simulate the phase transition.

As a pocket contracts, the dark quarks within it are compressed, allowing them to recouple and go through a second stage of annihilation. We calculate the fraction of the quarks that survive this new annihilation stage and escape the pockets in the form of dark baryons. We refer to this process as "thermal squeezeout", as the dark quarks are squeezed within the pockets and eventually leak out, in contrast to the standard "thermal freezeout". We find a dramatic suppression in the final abundance thanks to this phenomenon, which points to much heavier dark matter parameter space than was previously thought. In particular, the fact that the local DM density is much

[1] See also Refs. [205–207] for another mechanism affecting DM abundance in the presence of significant supercooling during a phase transition.

[2] This depends on the representation of the quarks under the dark confining gauge group. For instance, if the quarks were in the adjoint representation (similar to the model in Ref. [186]), they could combine with the surrounding gluons, form color-neutral hadrons, and move into the confined phase. Similarly, if in addition to the heavy quarks the spectrum were to include light quarks in the fundamental representation as well, the heavy quarks would not remain trapped within the pocket as they could make a color-neutral bound state with one of many light quarks surrounding them to escape the pocket.

larger than the globally averaged DM density during this second stage of annihilation invalidates the homogeneity assumption made in the unitarity bound argument of Ref. [210].[3] As a result, this model allows the DM candidate to be heavier than the perturbative unitarity bound on Weakly-Interacting Massive Particles (WIMPs), despite being thermal.

Since this second stage of annihilation is controlled by the dynamics within the dark sector and not the interaction between the dark sector and SM, our results are largely independent of the portal to the SM. In fact, we do not constrain ourselves to any specific portal in this chapter. The only assumptions we make about such a portal is that (i) it exists, (ii) it keeps the SM and the dark sector in thermal contact during the phase transition, and (iii) it respects the dark baryon number that stabilizes the dark baryons. These assumptions streamline our calculations significantly. However, it is worth considering the possibility of models in which we can relax one or more of these assumptions; we leave this for future work.

The current work is merely the start of a broader program of studying such models in more details. The phenomenology of all such models should be revisited in light of the dramatic change in the relic abundance calculation. Depending on the gauge group under study, the quarks' representation, and the portal, different models (with vastly different phenomenology) can be constructed.

Our study indicates a natural window of DM masses between 1-100 PeV for such a setup. While conventional searches may lose sensitivity for such a high DM mass, the stochastic gravitational wave background due to the first-order phase transition in this scenario can be detected by planned future facilities. (See Ref. [213] for the projected reach of such facilities.) While this signal depends on the UV parameters that control the thermodynamics of the phase transition, it does not depend on the nature of the portal.

The rest of this chapter is organized as follows. In Sec. 4.2 we provide an overview of the cosmology of our dark sector. In Sec. 4.3 we write down and solve the Boltzmann equations that determine the relic abundance of the dark matter. In Sec. 4.4 we provide an overview of the possible phenomenological implications of our dark sector before concluding in Sec. 4.5. We also provide three appendices to supply more details. In App. C.1 we provide more details on the thermodynamics of first-order

[3] See also Refs. [211, 212] for more recent studies of the unitarity bound on thermal DM models.

phase transitions in an expanding universe and detail a simulation we performed to fix the phase transition parameters that enter the relic abundance calculation. In Apps. C.2 and C.3 we review some results in the literature for the cross sections and binding energies of heavy quarks and their bound states.

4.2 A Qualitative Overview of the Cosmology

We consider a dark sector with a non-abelian $SU(3)$ gauge group and a single flavor of heavy quarks q in the fundamental representation

$$\mathcal{L} \supset -\frac{1}{4} G^{\mu\nu} G_{\mu\nu} + \bar{q} \left(i \gamma_\mu D^\mu - m_q \right) q \,, \tag{4.2.1}$$

where $G^{\mu\nu}$ is the dark gluon field strength and m_q is the dark quark mass with $m_q \gg \Lambda$ (in practice we consider $m_q \geqslant 100\Lambda$), where Λ is the dark confinement scale at which a phase transition takes place.

Given that $m_q \gg \Lambda$ we expect that such heavy quarks decouple from the thermal bath well before the phase transition, so that the phase diagram of this model is very close to that of pure Yang-Mills for $T \ll m_q$. Since the heavy quark regime can be well-approximated by the pure-gauge regime, the phase transition behavior is almost independent of the number of heavy quark flavors [45]. The only constraining condition on the number of quark flavors is then asymptotic freedom, or in special cases asymptotic safety [214]. Various lattice gauge theory studies have established that the $SU(3)$ phase transition takes place at a critical temperature very near the confinement scale, $T_c \approx \Lambda$, and is first-order [46, 197–199]. We will therefore assume that this dark sector features a first-order phase transition exactly at $T_c = \Lambda$.

The effects of this phase transition on the relic abundance of DM have been relatively unexplored in previous studies of such a confining gauge sector, e.g. [184]. We will discuss in this section how this phase transition dramatically changes the relic abundance calculation by causing a second stage of significant DM annihilation.

We remain agnostic about how the dark quark mass is generated as it will not affect our study. We also do not commit to any specific portal between the dark sector and the SM. We merely assume such a portal exists and keeps the dark sector in kinetic equilibrium with the SM. The portal enables the decay of the dark glueballs and mesons to the SM while respecting the dark baryon number symmetry, thus

stabilizing the dark baryons. These baryons, which are three-quark bound states, are the DM candidate in this setup. We will also assume a symmetric initial condition, $n_q = n_{\bar{q}}$.

In this section we provide an overview of the cosmology of such a sector, focusing primarily on the effect that the phase transition has on the DM relic abundance. The goal is to provide the reader with a broad picture of the various moving parts in this study, while leaving some of the more detailed calculations for later sections and App. C.1.

4.2.1 Pre-confinement epoch

For high enough temperatures, $T > T_c$, the dark sector exists in a deconfined thermal state in which quarks move freely within a gluon bath. Naively this setup seems at odds with confinement, which requires that colored objects not propagate freely over distances greater than the confinement length, Λ^{-1}. However, qualitatively, these colored quarks can move freely because they are connected to a network of thermal gluons [215]. These gluons screen a quark's color charge so that the quark effectively behaves like a color neutral object, in a process analogous to Debye shielding in plasmas. More quantitatively, lattice simulations have shown that when $T \geq T_c$, the potential between two heavy quarks flattens when they are separated by a distance of roughly more than Λ^{-1} [46]. In other words, distant quarks in a gluon bath do not influence one another.

In the deconfined phase, the quark relic abundance calculation proceeds analogously to a standard WIMP relic abundance calculation. For large enough dark quark masses, $m_q \gtrsim 20\Lambda$, which will be satisfied for all the parameter space we consider in our analysis, the dominant number changing process $q\bar{q} \leftrightarrow gg$ freezes out before confinement. The Boltzmann equation governing this freeze-out is simply

$$\dot{n}_q + 3H(T)n_q = -\langle\sigma v\rangle \left(n_q^2 - \left(n_q^{\rm eq}\right)^2\right), \tag{4.2.2}$$

where $H(T)$ is the Hubble constant at temperature T, $\langle\sigma v\rangle$ is the thermally averaged annihilation rate for $q\bar{q} \leftrightarrow gg$, n_q is the quark number density, and $n_q^{\rm eq}$ is its value in thermal equilibrium with a thermal bath of temperature T and with zero chemical

potential. Since $T \gg \Lambda$ in this epoch, $\langle \sigma v \rangle$ can be calculated perturbatively [47],

$$\langle \sigma v \rangle = \zeta \, \pi \frac{\alpha^2(m_q)}{m_q^2} \,, \tag{4.2.3}$$

where $\alpha(m_q)$ is the dark, strong coupling constant evaluated at the dark quark mass scale and ζ is a prefactor encapsulating plasma effects and non-relativistic enhancements (with numerical values presented in Fig. C-5). We will find that the exact size of this cross section does not qualitatively change the main results of this chapter. For further elaboration about this cross section and others, see App. C.2.[4] For the running coupling constant we use [7]

$$\alpha(m_q) = \frac{12\pi}{(11N_c - 2N_f(m_q)) \log \frac{m_q^2}{\Lambda^2}} \,, \tag{4.2.4}$$

where $N_f(\mu)$ is the number of light flavors contributing to the beta function at mass scale μ. We set $N_c = 3$ and $N_f(m_q) = 1$.

In the left panel of Fig. 4-1 we show the resulting quark number density evolution for specific choices of quark mass and confinement scale. A generic obstacle in the study of strong sectors is the uncertainty in determining cross sections. To characterize this uncertainty, we vary the cross section within an order of magnitude around the central value in Eq. (4.2.3), which produces the green bands.

Importantly, we find that heavy quarks are well-separated just before the phase transition begins. To characterize their separation, we define the typical inter-quark spacing in units of the confinement length,

$$\xi(t_n) \equiv \frac{\Lambda}{(n_q(t_n))^{1/3}} \,, \tag{4.2.5}$$

where t_n is the time at which extensive bubble nucleation starts, i.e. the onset of the phase transition, and $n_q(t_n)$ is the number density of the quarks at this time. This quantity measures, in units of Λ^{-1}, the typical distance between quarks at the onset of the phase transition. When $\xi(t)$ is large, quarks are separated by much more than a confinement length. In the right panel of Fig. 4-1 we show $\xi(t_n)$ as a function of m_q and Λ. Indeed, quarks are generically further from each other than

[4] In addition to $\bar{q}q$ annihilation, quarks are able to bind into diquarks via the attractive anti-triplet channel [47] and diquarks can bind with quarks to form baryons. We have checked that in the pre-confinement epoch and for the parameter space we are considering, this bound state production is negligible (see Sec. 4.3).

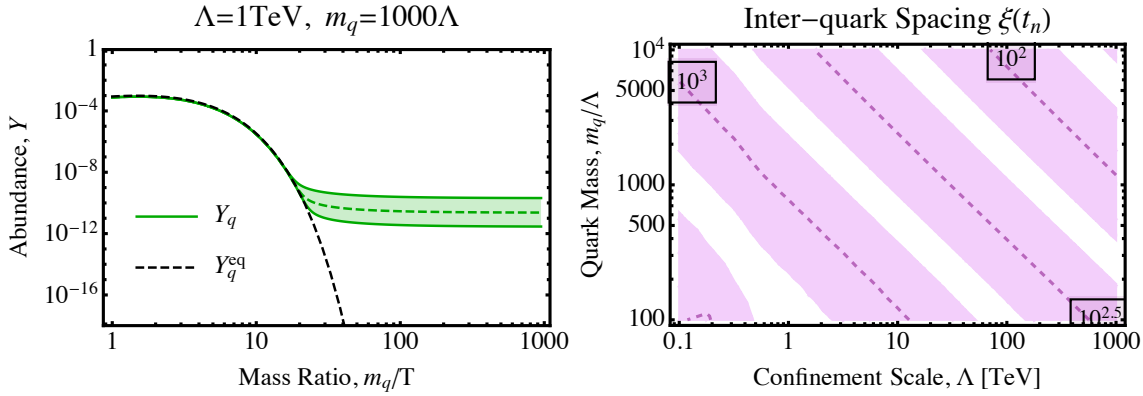


Figure 4-1: **Left:** The yield (number density normalized by the entropy density of the universe), $Y = \frac{n_q}{s}$, of the quarks for a quark mass of 10^3 TeV and confinement scale of 1 TeV. **Right:** The average separation of quarks as defined in Eq. (4.2.5) at the onset of the phase transition for various values of Λ and m_q/Λ. In each plot we vary the cross section by a factor of 10 above or below the central value (dashed lines) in Eq. (4.2.3) to produce the shaded regions.

a confinement length just as the phase transition begins. Were quarks not so well separated, the details of the phase transition would have a less dramatic effect on the DM relic abundance calculation and we would be able to use the combinatorial method of Ref. [184].

4.2.2 Bubble dynamics

Once the universe cools down to the critical temperature, $T_c = \Lambda$, a first-order phase transition begins. Phase conversion cannot occur right at the critical temperature as both phases have the same free energy, so the temperature of the deconfined phase initially cools slightly below T_c. As the deconfined phase supercools further into a metastable state, bubbles of the confined phase begin to nucleate and expand at a non-negligible rate.

As the deconfined phase is converted to confined phase, latent heat is released. In contrast to weakly coupled phase transitions, there is no perturbative parameter suppressing the latent heat, meaning that phase conversion will serve as a significant heat source in the temperature evolution of the universe. As a result, the plasma heats back up to a temperature very close to T_c quickly after bubble nucleation becomes efficient. Since the nucleation rate is exponentially sensitive to the degree of supercooling, $(T - T_c)/T_c$, subsequent nucleation of bubbles is completely suppressed.

For the phase transition to continue, at least some of the bubbles from the brief period of efficient nucleation must continue to grow. To determine the bubble growth

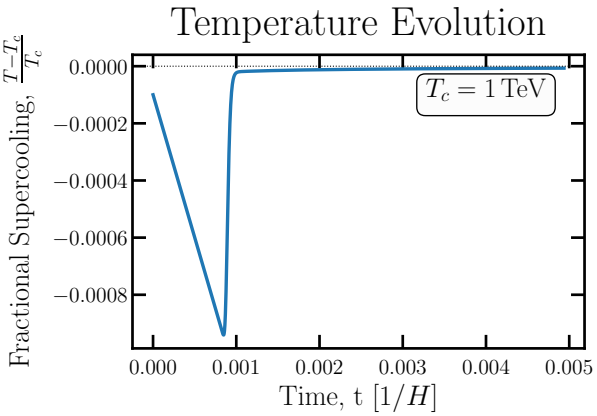


Figure 4-2: The degree of supercooling prior to percolation. The temperature supercools until bubbles nucleate efficiently. The nucleated bubbles quickly expand, depositing latent heat that drives T back up to T_c and eliminates further bubble nucleation. The temperature then stays roughly constant as latent heat deposited from bubble growth nearly cancels Hubble cooling. This plot ends at percolation, when half of the universe is in each phase. Notice that the timescale of the phase transition is much shorter than the Hubble timescale.

rate, we borrow an argument from [209]. As bubbles grow, the local temperature at the bubble walls increases towards T_c, diminishing the free energy difference between the two phases that drives the expansion. The expansion rate is then limited by the rate at which the wall can cool. The cooling rate is controlled by the temperature gradient between the wall and surrounding fluid – if there were no temperature difference, heat would not flow. Since the wall temperature cannot exceed T_c without reversing direction, we assume that this temperature difference is bounded above by the small degree of supercooling $(T_c - T)/T_c$. By modeling the heat dynamics near the wall in App. C.1, we estimate that the wall velocity is also bounded above by the degree of supercooling, $v_w \leq (T_c - T)/T_c$. For simplicity, we assume that v_w saturates this bound.

In Fig. 4-2 we plot the degree of supercooling as a function of time during the bubble expansion stage of the phase transition. This result comes from a simple simulation that we develop to track the nucleation and growth of bubbles during this epoch. Further details about this simulation can be found in App. C.1. The stages of the phase transition discussed above are visible in this plot. The universe initially supercools through Hubble expansion until bubble nucleation becomes efficient, leading to quick bubble growth and latent heat injection that reheats the universe. The heating and cooling rates then roughly balance one another, leaving the temperature at a value very near T_c, which suppresses further bubble nucleation as explained above.

Eventually, half of the universe converts to the confined phase. At around this so-called percolation time, most bubbles are in contact with one another and start coalescing. Soon after we are left with isolated "pockets" of the deconfined phase submerged in a sea of the confined phase. To properly compute the spectrum of shapes and sizes of these pockets would require a full numerical 3D bubble simulation. Instead, to simplify our analysis, we will assume that soon after percolation there is a characteristic size of a typical pocket, that pockets can be approximated as spherical, and that the details of the spectrum of pocket shapes and sizes will give only sub-dominant corrections to our results.

To determine this characteristic pocket size, we first determine the characteristic size of bubbles just before percolation. Using our simulation from App. C.1, we find that at percolation the spectrum of bubble radii peaks strongly at

$$R_0 \approx 10^{-6} \times \left(\frac{\Lambda}{M_{pl}}\right)^{-0.9} \frac{1}{\Lambda}\,, \tag{4.2.6}$$

where $M_{pl} = 2.4 \times 10^{18}$ GeV is the reduced Planck mass. We now borrow another argument from [209] to argue that for most values of Λ, these bubbles coalesce quickly until they reach a larger characteristic size, denoted as R_1.

The central idea is that small bubbles coalesce and merge quickly into bigger bubbles, and that the time scale for two bubbles in contact to merge becomes longer as bubble sizes grow. Intuitively, the larger the coalescing bubbles, the more matter has to be moved via the bubbles' surface tension, which takes more time. Thus there is a special bubble size, R_1, above which bubbles merge slower than the timescale over which the phase transition takes place. We find this critical size to be

$$R_1 \approx \left(\frac{M_{\rm pl}}{10^4\,\Lambda}\right)^{2/3} \frac{1}{\Lambda}\,. \tag{4.2.7}$$

Figure 4.2.2 shows that for $\Lambda \gtrsim 1$ TeV the typical size of bubbles just before percolation (R_0) is always smaller than R_1. Thus, we assume that, for this range of Λ, at percolation all bubbles quickly coalesce until they reach a size of R_1. For smaller Λs we assume that all bubbles will have radius R_0 instead. We then make the simplifying assumption that the characteristic size of pockets just after percolation is the same as the characteristic size of bubbles just before percolation, i.e.

$$R_i = \max\left(R_0, R_1\right)\,, \tag{4.2.8}$$

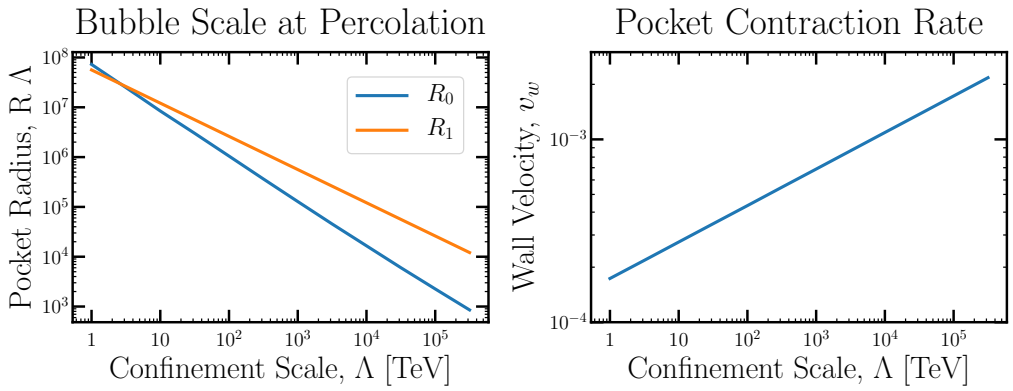


Figure 4-3: **Left:** The typical radius of bubbles just before percolation (blue line) and the characteristic coalescence radius R_1 (orange line). For any Λ with $R_0 \leq R_1$, we assume that bubbles quickly coalesce and grow to radius R_1 at percolation. **Right:** The asymptotic velocity of the pocket wall during its contraction as a function of the confinement scale when quark pressure is ignored. In the more realistic case where internal quark pressure is allowed to resist the contraction of the pocket, we expect v_w to be much smaller, and to not necessarily asymptote to a constant value at small radii.

where R_i is the characteristic initial pocket radius after percolation.

It is more complicated to determine the wall velocity of the contracting pocket. The main complication is that quarks are trapped within pockets, which we will show in the next section, and will generically slow down the wall. For now we will assume that we can neglect the effect of the enclosed quarks, which will lead to an overestimate of v_w. In Sec. 4.2.4 we will revisit the effect these quarks have on v_w.

In App. C.1 we find that at radii much smaller than R_i, the pocket contraction rate asymptotes to a constant value, which is shown in the right panel of Fig. 4.2.2 and can be fit by

$$v_w(\Lambda) \approx 0.2 \times \left(\frac{\Lambda}{M_{pl}}\right)^{0.2}. \tag{4.2.9}$$

It will turn out in Sec. 4.3 that the relic abundance of DM is set while $R \ll R_i$, so we can neglect the initial stages when v_w varies and treat it as a constant. The pockets' radii therefore shrink as a function of time according to

$$R(t') = R_i - v_w t' \tag{4.2.10}$$

where t' is the time after percolation.

To the best of our knowledge, the problem of characteristic bubble properties, e.g. the wall velocity and charactristic size at percolation, is not completely settled

for first-order phase transitions even in weakly interacting theories (see [216] and the references therein for recent discussions on calculating the wall velocity). In our numerical calculations in Sec. 4.3, we characterize these uncertainties by varying both v_w and R_i within one order of magnitude of the results shown in Fig. 4.2.2.

We note here that since the quark temperature is fixed near T_c throughout the phase transition, the typical quark velocity is

$$v_q \sim \sqrt{\Lambda/m_q}\,. \tag{4.2.11}$$

For the range of parameters that we will be interested in, we find $v_q \gg v_w$. This inequality will become important in the next section when we analyze the effects that the walls have on the quarks.

To summarize, the phase transition begins with an initial, complicated stage of bubble nucleation and growth until bubbles come into contact with one another. It then enters an even more complicated bubble coalescence stage. The space between bubbles is made of pockets of the deconfined phase with the same characteristic size, R_i. These pockets become isolated and eventually spherical, then contract initially with a velocity that is determined by the local heat diffusion rate. The contraction rate gradually slows down due to the pressure of the enclosed quarks, and the pockets eventually vanish. The phase transition has completed at this point, and the universe can proceed with its standard expansion history.

Further details of this phase transition, as well as an overview of the relevant thermodynamics can be found in App. C.1. Our study of the phase transition's effect on the DM relic abundance is insensitive to many of the details of the phase transition; we merely need an expression for the characteristic initial radius of pockets and their wall velocity, which are respectively provided in Eqs. (4.2.8) and (4.2.9). We emphasize that this latter expression for v_w, which neglects the effect of quark pressure, overestimates the wall velocity during the contraction phase .

4.2.3 Heavy quarks during the phase transition

During the entire process of bubble nucleation and expansion described in the previous section, bubble walls run into quarks and anti-quarks. In this section we study these encounters in detail, and argue that the walls are impermeable to quarks, but permeable to color-neutral bound states. While we will focus on the interaction be-

tween walls and quarks, our conclusions hold for anti-quarks as well.

When a wall encounters a quark, the quark can push against it and deform it locally. Whereas in electroweak-like phase transitions a particle is able to penetrate through the wall at the cost of only a finite mass difference [201, 202], the energy cost for an isolated quark to enter the confined phase is unbounded [46], preventing it from traveling far into the confined phase. Therefore, a quark can enter a bubble only if it either forms a color-neutral bound state before it enters the bubble, or deforms the wall so that it remains immersed in the color-screening gluon bath (see Fig. 4-4).

There are two ways in which the quark could form a bound state. First, $\bar{q}q$ pairs could be spontaneously created, binding with the quark as it passes through the wall. We can imagine a scenario as in Fig. 4-4 in which the quark pushes into the bubble and is connected to a gluon string [215] starting from its initial point of contact with the wall. If the quark were light enough, at some point this stretched string could break into a $\bar{q}q$ pair and the $\bar{q}$ could bind with the incoming quark to form a color-singlet bound state that enters the bubble (see [208] for an example in which this process is efficient). However, for a heavy quark, the string breaking rate is extremely suppressed; this rate can be approximated [217] using the Schwinger mechanism [218],

$$(t_{\text{string}})^{-1} \sim \frac{m_q}{4\pi^3} e^{-m_q^2/\Lambda^2} \,. \tag{4.2.12}$$

The exponential of the square of the large ratio m_q/Λ makes this string breaking timescale much larger than all other timescales, completely shutting off this process. The inefficiency of string breaking and the quark's inability to pass through bubble walls is the main feature distinguishing our model from those that involve light quarks.

The second way a quark could form a bound state is by encountering an anti-quark or two quarks somewhere within the deconfined phase, binding, then escaping into the confined phase before the bound state dissociates. These processes are important and will be analyzed in the next section via Boltzmann equations.

If a quark has not managed to bind into a color singlet state by the time it reaches a bubble wall, it deforms the wall to avoid entering the confined phase. As the wall deforms, its surface area increases, which increases the energy of the system. The surface tension therefore creates a force that opposes this deformation. If we estimate this force to be of order Λ^2 on dimensional grounds, then we find that the timescale for the surface tension to restore the shape of the wall and reverse the quark's velocity

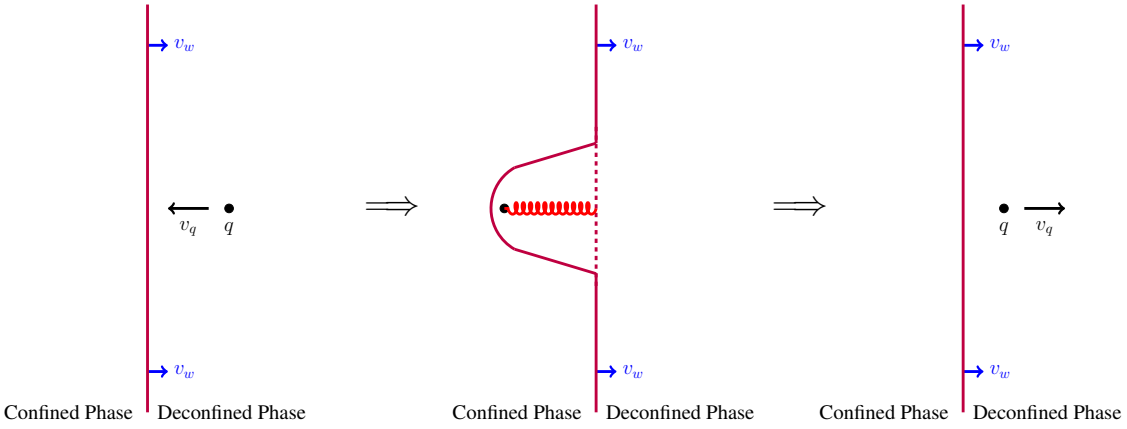


Figure 4-4: A depiction of a quark interacting with a phase boundary. As bubbles of confined phase grow (the wall moves to the right in the figure), their walls run in to quarks that move with typical velocity, v_q, which is much larger than the wall velocity, v_w (**left** configuration). The quark locally deforms the wall (**center** configuration), introducing an opposing force via surface tension. One can think of the quark as connected to the deconfined phase through a gluon string [215]. Since string breaking is shut off, i.e. the quark-gluon configuration does not have sufficient energy to pull a heavy $\bar{q}q$ pair out of the thermal background, eventually the quark comes to a halt and then rebounds back into the deconfined phase with its initial speed (the **right** configuration).

is

$$t_{\text{rebound}} \sim \frac{v_q}{\dot{v}_q} \sim \frac{v_q}{\Lambda^2/m_q} = \sqrt{\frac{m_q}{\Lambda}}\frac{1}{\Lambda}\,. \tag{4.2.13}$$

This rebound timescale is much shorter than the string breaking timescale. It is also orders of magnitude smaller than the phase transition timescale, which we find in App. C.1 to be $t_{\text{PT}} \sim 10^{-2}H^{-1} \sim 10^{-2}M_{\text{pl}}/\Lambda^2$ (also, see Fig. 4-2). Finally, the pocket contraction timescale is of order $t_{\text{contract}} \sim R/v_w$. Since $R > \Lambda^{-1}$ and we find that $v_w \lesssim 10^{-3}$, we have $t_{\text{contract}} > 10^3/\Lambda$. Since we only consider quark masses that satisfy $\sqrt{m_q/\Lambda} \leq 10^2$ in this chapter, we have $t_{\text{rebound}} \ll t_{\text{contract}}$. Since this rebound timescale is the shortest timescale in the problem, quarks rebound off walls very quickly before any other process can take place. Therefore, the bubble walls act like very stiff surfaces that quickly reflect quarks that come into contact with them.

As these bubbles grow, the walls sweep quarks and anti-quarks into the ever-shrinking deconfined regions, increasing the quark density over time. Moreover, since $v_q \gg v_w$, quarks that are swept in can quickly travel through the shrinking deconfined region and maintain homogeneity, meaning that n_q is independent of position in the pocket throughout the phase transition. Eventually, these particles end up inside the isolated pockets formed toward the final stage of the phase transition (see Fig. 4-5).

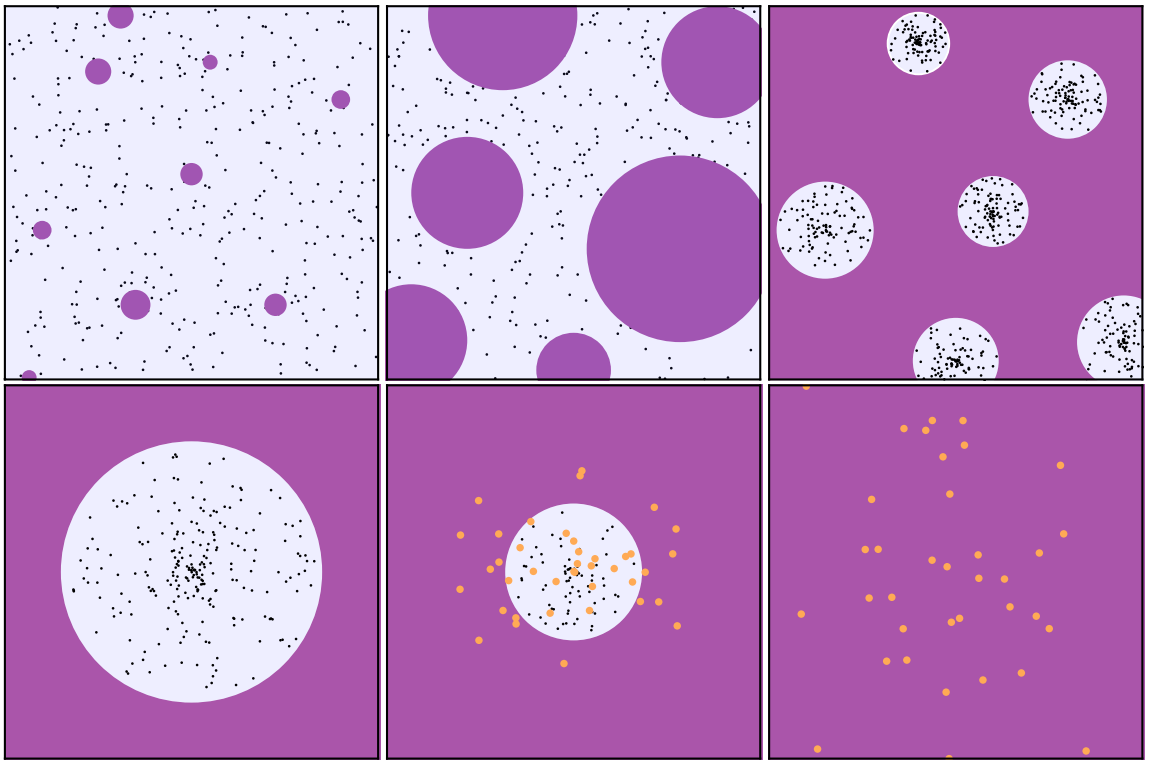


Figure 4-5: Schematic illustration of different stages of the phase transition and its effect on the DM abundance. Violet indicates the confined phase and light blue the deconfined phase. **Top Left:** Once the temperature drops slightly below $T_c = \Lambda$, bubbles of confined phase begin nucleating everywhere. The nucleated bubbles start growing and push quarks (black dots) around. **Top Middle:** The bubbles grow to a point that a $\mathcal{O}(1)$ fraction of the universe has converted into the confined phase. At this point bubbles start coalescing and quickly grow larger. **Top Right:** As the bubbles keep growing and combining, eventually we are left with isolated pockets of the deconfined phase submerged in a sea of the confined phase. **Bottom Left:** A single isolated pocket with quarks trapped inside it. Each pocket contracts as the phase transition continues. **Bottom Middle:** The particles in the pocket are compressed and their interactions recouple. During this phase the particles can either annihilate or bind with one another. Non-singlet states can not enter the confined phase, but once they form color-neutral bound states (orange dots) they can escape into the confined phase. **Bottom Right:** In the end, the pockets vanish and only the fraction of quarks that ended up inside color-neutral baryons survive. These particles diffuse away from the original pocket's position due to their local overdensity.

We note that in models with additional light (though not massless) quarks, there would be no first-order phase transition [219], and even if such a transition did exist, quarks would easily pass through bubble walls and would most likely be unaffected by the phase transition.

Within any fixed volume of the universe, including the isolated pockets, the baryon number is a fluctuating random variable. Although the baryon number averaged over all pockets must be zero due to our symmetric initial condition, any given pocket is expected to have an overabundance or underabundance of quarks relative to anti-quarks, which we call the pocket asymmetry, η. We find that the initial total number of quarks in a pocket, N_q^{initial}, is large, so that by the central limit theorem the standard deviation of fluctuations above and below the mean is $\sqrt{N_q^{\text{initial}}}$. Therefore, no matter how efficient $\bar{q}q$ annihilation processes are in these contracting pockets, on average, at least a fraction

$$\eta_{\text{rms}} \equiv \sqrt{\langle \eta^2 \rangle} = \sqrt{N_q^{\text{initial}}}/N_q^{\text{initial}} = 1/\sqrt{N_q^{\text{initial}}} \tag{4.2.14}$$

of the initial quarks (or antiquarks) in a pocket will survive. This observation will have important consequences for our relic abundance calculation in the next section.

Once isolated pockets have formed and their asymmetries have been set, they will contract and compress quarks and anti-quarks until formerly frozen-out interactions turn back on. These interactions include $\bar{q}q$ annihilation as well as $q + q$ binding via the attractive anti-triplet channel [47]. As these diquarks build up their occupation number, they can eventually bind with quarks to form color-singlet baryons that can quickly fly out of the pocket.[5] These escaping stable baryons constitute the DM candidate of our model, while the rest of the particles eventually dump their energy into the SM sector through an unspecified portal interaction.

We define a survival factor as the fraction of quarks and antiquarks that escape the pocket within baryons and antibaryons,

$$\mathcal{S} \equiv \frac{N_q^{\text{survived}}}{N_q^{\text{initial}}}. \tag{4.2.15}$$

In the next section we write down the Boltzmann equations governing the quark dynamics within contracting pockets and calculate this survival factor. As remarked

[5] Notice that the formation of baryons through an intermediate diquark is more efficient than the formation of baryons via direct 3-body recombination, which we ignore.

above, $\mathcal{S}$ is bounded below by the asymmetry of a given pocket, and the expectation value of this lower bound is

$$\mathcal{S} \geqslant \eta_{\rm rms} = \frac{1}{\sqrt{N_q^{\rm initial}}} \,. \tag{4.2.16}$$

After these surviving baryons escape the pockets and after the phase transition eventually completes, these baryons continue to diffuse away until they re-establish homogeneity in the universe. If the asymmetry bound is not saturated, baryons and anti-baryons can continue to annihilate as they diffuse outside of the pocket. A more detailed study of this final annihilation stage requires integrating inhomogeneous Boltzmann equations, which we leave for future work.

In summary, the dark matter undergoes a short squeeze, where collapsing bubbles during the phase transition induce a second stage of rapid annihilation that drastically depletes the universe's pre-existing stock of dark matter. This extra annihilation after freeze-out opens up parameter space that had previously been ruled out due to overproduction of DM. We will show that this effect allows for this thermal DM to be heavier than the conventional unitarity bound of ~ 300TeV [210].

4.2.4 The quark pressure on the wall

Although we have considered the effect that the wall has on the trapped quarks, we have ignored the effect that the trapped quarks have on the wall. In this section, we will argue that the trapped quarks generically slow down the contraction rate of the wall.[6]

Much like a piston, a pocket wall can contract only if it does work on the enclosed gas of heavy quarks. Since we have assumed that this gas is thermally coupled to the rest of the SM bath, which has a much larger heat capacity than that of the dark sector, the quarks contract at constant temperature. Using an ideal gas equation of state, we can write the pressure of this quark gas as $p_q = n_q T$. The work that the wall does on the gas when it contracts by an amount dR is therefore $p_q dV = 4\pi R^2 n_q T dR$.

The forces that are responsible for this change of pocket radius are the surface tension and net gluonic pressure, the latter of which is directed inward whenever $T < T_c$. We will show in the next section that during the earliest stages of pocket

[6] We thank Filippo Sala for pointing out this effect.

contraction, quark interactions are inefficient and the total number of quarks in the pocket is initially conserved. As a result, when the pocket shrinks the work required to contract the pocket grows like R^{-1}. At the same time, the work that the surface tension and net gluonic pressure do when contracting the pocket by dR is proportional to the change in area and volume, respectively, and so shrink like R and R^2. Altogether, as R contracts, the forces pushing out grow while the forces pushing in shrink. We therefore expect that v_w decreases with decreasing R.

As this physics involves non-equilibrium, strong dynamics, we cannot reliably compute v_w as a function of R. Instead, in the remainder of this section we will argue that the effect of quark pressure is to slow down v_w by orders of magnitude relative to the upper bound of Eq. (4.2.9) we computed when we neglected the quark pressure. For more details relevant to the following discussion, we refer the reader to App. C.1.4.

When we simulate pocket contraction while keeping track of the quark density within a pocket (see next section), we find that there always comes a point when the quark pressure has grown to such an extent that, were we to suddenly include it, the quark pressure would exactly oppose all inward-pointing forces. This point of mechanical equilibrium is defined by $\sigma\, dA + (\sum p)\, dV = 0$, where σ is the surface tension, dA the change in surface area, dV the change in volume, and $\sum p$ the sum of pressures acting on the wall. (Inward-pointing pressures are defined to be positive while outward facing pressures are negative.) If we were to suddenly include the effects of quark pressure, the motion of the wall would suddenly become calculable, since the state of the wall would be determined by equilibrium physics. The pocket would suddenly slow down and proceed to adiabatically shrink while maintaining mechanical equilibrium. Number changing processes would deplete n_q, diminishing the quarks' outward-pointing pressure, and the universe would supercool further, increasing the net gluonic inward-pointing pressure. We find that in this scenario, v_w suddenly drops by orders of magnitude when mechanical equilibrium is achieved, and v_w steadily decreases many more orders of magnitude as the pocket contracts.

The discontinuous drop in v_w signals a breakdown of our assumption that quark pressure was negligible before mechanical equilibrium was achieved. This simulation merely demonstrates that it is inconsistent to neglect quark pressure, and that it can potentially slow down the pocket contraction rate by orders of magnitude. We

therefore expect that a more realistic simulation that correctly includes the effects of quark pressure from the very beginning will lead to a v_w that gradually decreases from our upper bound of Eq. (4.2.9), which eventually overestimates v_w by orders of magnitude.

We will use the results of our pocket evolution simulations to calculate a few parameters that enter the Boltzmann Equations that govern the abundances of various bound states in the pocket. While the expression for the pocket radius, Eq. (4.2.8), is robust to the uncertainties introduced by quark pressure, we argued that the wall velocity v_w is sensitive to this uncertainty. In the next section, we will study the evolution of the bound state abundances in the pocket in two extreme cases: (i) when the effect of quark pressure on v_w is completely ignored, or (ii) when its effect dramatically reduces v_w.

4.3 Boltzmann Equations During Compression

As described above, toward the end of the phase transition, the deconfined regions form isolated pockets that contain all of the dark quarks. In this section, we describe the dynamics of the dark quarks and their bound states as the contracting pockets compress them. The Boltzmann equations that we solve keep track of the many processes by which quarks either ultimately annihilate into gluons or form baryons that escape the pockets and become dark matter. We will solve the Boltzmann equations for a typical pocket with initial characteristic radius R_i and pocket asymmetry set to its root-mean-square value, $\sqrt{N_q^{\text{initial}}}$. We assume that the $\mathcal{S}$ of this typical pocket is approximately equal to $\mathcal{S}$ averaged over the full distribution of initial pocket radii and pocket asymmetries. The total number of DM particles that survive until today will then equal the total number of DM particles entering the phase transition times $\mathcal{S}$.

4.3.1 Ingredients of the Boltzmann equations

We begin by listing the degrees of freedom that we will include in our Boltzmann equations, which have been tabulated in Tab. 4.1. We have neglected a host of exotic hadronic bound states like tetra- and penta-quark states because we assume that they are unstable and promptly decay to the states listed in Tab. 4.1. We also do not consider excited states of any of the bound states. To simplify the notation, we label

State	Dark Quark Number	Color Rep.
Gluons	0	**8**
Quark	1	**3**
Diquark	2	$\bar{\mathbf{3}}$
Baryon	3	**1**

Table 4.1: Different degrees of freedom entering the Boltzmann equations of the contracting pockets. We use the quark number of each state to refer them throughout the text. The existence of anti-particles, with negative dark quark numbers and conjugate representations under $SU(3)$, are implied.

states by their quark number throughout the text (for example, a baryon state is a 3 state while an anti-diquark is a -2 state).

We also neglect the mesons $\bar{q}q$ in our analysis. This can be justified by comparing their decay rate to the fastest annihilation rate that we will encounter (see App. C.2)

$$\begin{aligned} \Gamma_{\bar{q}q} &\sim \alpha^5 m_q, \\ \langle \sigma v \rangle_{\max} n_q &\sim \frac{1}{\alpha^3}\frac{\alpha^2}{m_q^2}\left(\frac{\Lambda}{\xi(t)}\right)^3, \\ \Longrightarrow \frac{\langle \sigma v \rangle_{\max} n_q}{\Gamma_{\bar{q}q}} &\sim \left(\frac{1}{\xi(t)\alpha^2}\frac{\Lambda}{m_q}\right)^3 \ll 1, \end{aligned} \tag{4.3.1}$$

where the last inequality is obtained because we have heavy quarks ($m_q/\Lambda \geqslant 100$) and the interquark spacing in units of Λ satisfies $\xi(t) \geq 1$. Such a fast meson decay rate ensures that these states are kept in equilibrium so that their number density is negligibly small. We also have verified numerically that including the mesons in our Boltzmann equations below has a negligible effect on our results.

Let us now look into the Boltzmann equation for the particles in Tab. 4.1 as they are compressed by the contracting pockets. We start with the Liouville operators. For the colored particles, i.e. ± 1 (quarks/anti-quarks) and ± 2 (diquarks/anti-diquarks), we have

$$L[i] = \dot{n}_i - 3\frac{v_w}{R} n_i, \quad i = 1, 2, \tag{4.3.2}$$

where the second term captures the effect of pocket compression. Notice that we have not included the usual factor of $+3Hn_i$ for the dilution of space due to Hubble expansion. As argued above, $t_{\rm PT} \ll H^{-1}$. Therefore, the Hubble dilution rate is negligible during the phase transition and can be ignored.

For the color-neutral particles, i.e. baryons and anti-baryons, the compression

term will be absent. Unlike the colored particles, the baryons are not constrained by confinement to remain in the deconfined pockets. The baryons formed in the pocket can then be thought of as a gas created in a container without walls. The gas of baryons will thus escape with a rate governed by its internal pressure, or equivalently by the thermal velocities of the baryons.[7]

Once the baryons escape the pocket they are no longer tracked by the Boltzmann equations. So we must include baryon escape as a sink term in our Boltzmann equations, which we do by modifying the Liouville operator,

$$L[3] = \dot{n}_3 + 3\frac{v_q}{R}n_3. \tag{4.3.3}$$

To derive this escape rate, consider a small time step, dt. In each time step the pocket radius contracts by $v_w dt$. The typical baryon moves a distance of about $v_q dt$, where we ignore the distinction between the baryon and quark velocities. We then overestimate the escape rate by an $\mathcal{O}(1)$ factor by assuming that all baryons at the edge of the bubble move radially outward, giving a total number of escaped baryons of

$$dN_3^{\rm esc} = 4\pi R^2 n_3(R)(v_q + v_w)dt\,. \tag{4.3.4}$$

Combining this with the rate of change for pockets volume gives the density loss rate due to baryon escape used in Eq. (4.3.3).

It will be convenient to track the evolution of the total number of particles in a pocket as opposed to number densities. Define the pocket volume,

$$V(R) = \frac{4\pi}{3}R^3\,. \tag{4.3.5}$$

Multiplying the number density of species i by the volume of a pocket then gives the total number of species i in the pocket,

$$N_i \equiv V n_i\,. \tag{4.3.6}$$

It will also be convenient to replace the time coordinate with R using Eq. (4.2.10).

[7]Notice that the justification for why baryons in the pocket are homogeneously distributed is different than that of the quarks and diquarks. Gradients in the baryon density naturally arise as the baryons flow from their high density points of creation to the low density exterior of the pockets. However, a homogeneous component of baryons is constantly being produced within a pocket due to the binding of (homogeneously distributed) quarks and diquarks. We find that the rate of production is faster than the escape rate, so the baryon density in the pocket remains homogeneous to a good approximation.

We can then rewrite the Liouville operators as

$$L[i] = -\frac{v_w}{V} N_i', \quad i = 1, 2, \tag{4.3.7}$$

$$L[3] = -\frac{v_w}{V}\left(N_3' - \frac{3}{R}\frac{v_q + v_w}{v_w} N_3\right), \tag{4.3.8}$$

where $N' \equiv dN/dR$ and we have used $\dot{R} = -v_w$.

Now that we have dealt with the Liouville operators we write down the collision operators. We will only be concerned with 2-to-2 processes since n-to-2 processes are Boltzmann suppressed while 2-to-n processes are suppressed by extra factors of $\alpha(m_q)$ and phase space factors. We denote each of these terms with the following notation,

$$\begin{aligned}\Big\langle (a,b) \to (\alpha,\beta) \Big\rangle &= \langle \sigma v \rangle_{ab\to\alpha\beta} \left(n_a n_b - n_\alpha n_\beta \frac{n_a^{eq} n_b^{eq}}{n_\alpha^{eq} n_\beta^{eq}} \right) \\ &= \frac{\langle \sigma v \rangle_{ab\to\alpha\beta}}{V^2} \left(N_a N_b - N_\alpha N_\beta f_{ab,\alpha\beta} \right) ,\end{aligned} \tag{4.3.9}$$

with $a, b, \alpha, \beta = 0, \pm 1, \pm 2, \pm 3$, and $f_{ab,\alpha\beta} \equiv \frac{N_a^{eq} N_b^{eq}}{N_\alpha^{eq} N_\beta^{eq}}$. For gluons we have $n_0 = n_0^{(eq)}$, i.e. the gluons are always in equilibrium.

Once we have identified all the important interactions to be included in our Boltzmann equations, we can write down the complete system of differential equations for $N_i(R)$. We supply these equations with the initial conditions, which were derived in Sec. 4.2. The initial pocket radius is R_i while the initial quark number in the pocket, N_1, is found by multiplying the number density result of the pre-confinement freeze-out calculation in Eq. (4.2.2) by $\frac{4\pi}{3} R_i^3$. We find that the initial conditions for N_2 and N_3 are irrelevant, as they quickly approach an equilibrium value independent of whatever values we initially choose (so long as $N_2, N_3 \ll N_1$ initially). All that is left is to write down these equations and solve them.

4.3.2 Complete set of Boltzmann equations

The complete set of Boltzmann equations is:

$$L[i] = C[i], \quad i = 1, 2, 3. \tag{4.3.10}$$

$$\begin{aligned} C[1] &= -\Big\langle (-3,1) \to (-1,-1) \Big\rangle - \Big\langle (-3,1) \to (-2,0) \Big\rangle + 2\Big\langle (3,-1) \to (1,1) \Big\rangle \\ &+ \Big\langle (3,-2) \to (1,0) \Big\rangle - \Big\langle (1,-1) \to (0,0) \Big\rangle + \Big\langle (2,2) \to (3,1) \Big\rangle - 2\Big\langle (1,1) \to (2,0) \Big\rangle \end{aligned}$$

$$
\begin{aligned}
&+ \left\langle (-3,2) \to (-2,1) \right\rangle + \left\langle (2,-2) \to (1,-1) \right\rangle + \left\langle (2,-1) \to (1,0) \right\rangle \\
&- \left\langle (2,1) \to (3,0) \right\rangle - \left\langle (-2,1) \to (-1,0) \right\rangle + \left\langle (3,-3) \to (1,-1) \right\rangle \quad , \\
C[2] \;&= \left\langle (1,1) \to (2,0) \right\rangle - \left\langle (-3,2) \to (-1,0) \right\rangle + \left\langle (3,-1) \to (2,0) \right\rangle \\
&- \left\langle (2,-2) \to (0,0) \right\rangle + \left\langle (3,-2) \to (2,-1) \right\rangle + \left\langle (3,-3) \to (2,-2) \right\rangle \\
&- \left\langle (2,-1) \to (1,0) \right\rangle - 2\left\langle (2,2) \to (3,1) \right\rangle - \left\langle (2,1) \to (3,0) \right\rangle \\
&- \left\langle (-3,2) \to (-2,1) \right\rangle - \left\langle (2,-2) \to (1,-1) \right\rangle \quad , \\
C[3] \;&= \left\langle (2,1) \to (3,0) \right\rangle + \left\langle (2,2) \to (3,1) \right\rangle - \left\langle (3,-3) \to (0,0) \right\rangle - \left\langle (3,-1) \to (2,0) \right\rangle \\
&- \left\langle (3,-1) \to (1,1) \right\rangle - \left\langle (3,-3) \to (1,-1) \right\rangle - \left\langle (3,-3) \to (2,-2) \right\rangle \\
&- \left\langle (3,-2) \to (2,-1) \right\rangle - \left\langle (3,-2) \to (1,0) \right\rangle
\end{aligned}
$$

where $\left\langle (\cdot\,,\cdot) \to (\cdot\,,\cdot) \right\rangle$ is defined in Eq. (4.3.9). The right-hand side consists of all interactions that are consistent with quark number conservation. We also make the approximation that

$$N_i = N_{-i}. \tag{4.3.11}$$

While this equality is not strictly satisfied due to the pocket asymmetry, we are able to make it because only one of three scenarios can occur: either (i) the symmetric component is never depleted to the point that the asymmetry is important, (ii) it is completely depleted and the accidental asymmetric abundance is all that survives, or (iii) the symmetric and the asymmetric components are comparable and our answer is off by an $\mathcal{O}(1)$ factor. As argued before, this asymmetry introduces a lower bound on $\mathcal{S}$ (Eq. (4.2.16)).

Despite Eq. (4.3.10) having numerous terms, solving these equations numerically is rather straightforward. For convenience we list the important parameters entering into these equations and their reference values in Tab. 4.2. We remind the reader that Eq. (4.2.9) overestimates v_w since it neglects the quark pressure's ability to oppose pocket contraction. As we will discuss below, we can bracket the effect that a slower v_w would have on the final DM relic abundance quite robustly, see Sec. 4.3.5 for further details. We also reemphasize that we have used simple approximations for some of other quantities – particularly the bubble radius – and a rigorous determination of them is only possible through more extensive numerical calculations.

In Fig. 4-6, we show the solution of the Boltzmann equations in Eq. (4.3.10) for

Quantity	v_w	$R_i(\Lambda)$	$\xi(t_n)$	v_q	σv	Binding energies
Central Value	See main text.	Eq. (4.2.8)	Eq. (4.2.5)	Eq. (4.2.11)	See App. C.2	See App. C.3

Table 4.2: The relevant quantities in the Boltzmann equations and our expression for each. More discussion on how we treat v_w is included in the main text.

a specific quark mass and confinement scale when we neglect quark pressure and use Eq. (4.2.9) for the pocket wall velocity. There are a number of important observations to be made about this figure. First, the fractions of diquarks and baryons are initially very small, justifying why we did not include them in our calculations prior to pocket formation. Next, as the pockets contract, the number of bound states initially grows while the number of free quarks decreases due to binding or annihilation to gluons. As the number of free quarks decreases, the annihilation or escape of bound states become more important than their production, so their occupation numbers reach a maximum and monotonically decrease from there. Finally, we see that each step in the chain of bound state formation ($q+q+q \to qq+q \to qqq$) results in a suppression, i.e. the total number of diquarks is suppressed compared to the total number of free quarks, while the total number of baryons is suppressed compared to the diquarks. We anticipate that had we started with a larger $SU(N)$ gauge group ($N \geqslant 4$), the bound states with higher quark numbers would have been further suppressed and the final DM survival factor would be lower. We leave a more detailed analysis of this scenario to future work.

Finally, to calculate the survival factor, we simply integrate Eq. (4.3.4) to calculate the total number of baryons that escaped during the contraction of the pocket and normalize to the initial quark number in the pocket. Rewriting Eq. (4.3.4) in terms of N and R, we find

$$\mathcal{S}_{\text{symm.}} = \frac{3\int dN_3^{\text{esc}}}{N_q^{\text{initial}}} = \frac{9}{N_1(R_i)}\int dR \frac{v_q + v_w}{v_w R} N_3(R), \tag{4.3.12}$$

where we have used Eqs. (4.2.10) and (4.3.6) to change variables, and the subscript in $\mathcal{S}_{\text{symm.}}$ is to indicate that this is the survival factor of the symmetric component of the dark quarks. The factor of 3 in the first equality accounts for the fact that three quarks exists within every one baryon that escapes.

Note that in deriving this result we assumed no asymmetry exists in the pocket. Combining this result with the lower bound on $\mathcal{S}$ from the asymmetry component,

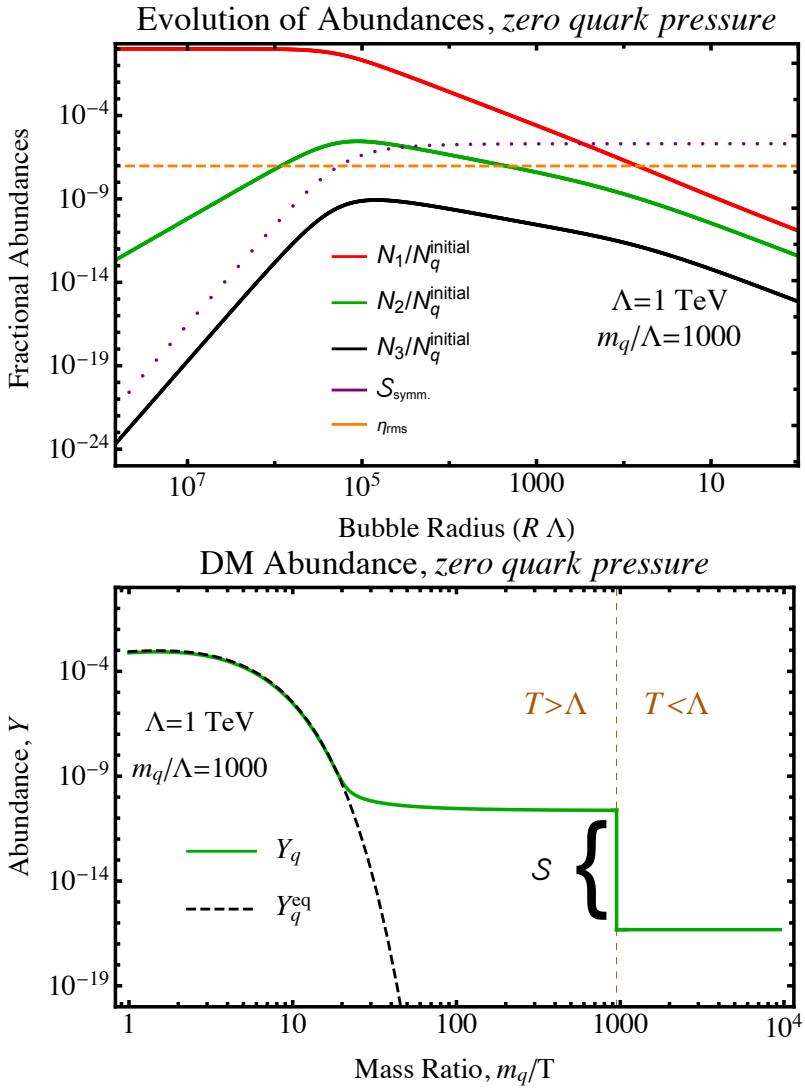


Figure 4-6: **Top:** Evolution of the fractional number of free quarks (red), diquarks (green), and baryons (black) inside the pocket, normalized to the initial quark number. We neglect the effect of quark pressure on v_w in solving the Boltzmann equations for this plot, meaning that we use Eq. (4.2.9) for v_w. As the phase transition proceeds, the pocket radius decreases. Initially, almost all the quarks are unbound. As the pocket contracts, more bound states are formed and fewer quarks are found as free particles. As their numbers increase, the various states' annihilation rates increase as well. At some point their production and annihilation rates are comparable and the number of bound states inside the pocket reaches its maximum. The accumulative surviving fraction assuming zero pocket asymmetry predicted by Eq. (4.3.12) is denoted by the dotted purple line. The asymptotic value of this line is equal to $\mathcal{S}_{\text{symm.}}$. We denote the asymmetry lower bound on $\mathcal{S}$ from Eq. (4.2.16) by the orange dashed line. **Bottom:** The DM abundance evolution for this mass and confinement scale. The $T > \Lambda$ region is similar to Fig. 4-1; the confinement takes place at $T = \Lambda$ and gives rise to the abundance suppression predicted by Eq. (4.3.13).

Eq. (4.2.16), we have

$$\mathcal{S} = \max\left\{\mathcal{S}_{\text{symm.}}, \eta_{\text{rms}}\right\}. \tag{4.3.13}$$

In Fig. 4-6 we show $\mathcal{S}_{\text{symm.}}$ and η_{rms} as well. We find that for the chosen Λ and m_q, while neglecting the effect of quark pressure on the wall velocity, $\mathcal{S}_{\text{symm.}} \geqslant \eta_{\text{rms}}$, i.e. the local pocket asymmetry is not saturated during the contraction, but $\mathcal{S}$ is within ~ 1 order of magnitude of this asymmetry bound. In fact, we find this is true for all the points in the parameter space that we study. In the upcoming section we will describe how we can leverage this observation to bracket the range of parameter space that gives rise to the correct DM relic abundance.

4.3.3 Analytic approximation

While the Boltzmann equations in Eq. (4.3.10) can be solved numerically, the large number of terms involved can muddle one's intuition. In this section, we develop a simple analytic approximation for solving these equations and determining $\mathcal{S}_{\text{symm.}}$.

From the full set of interactions included in Eq. (4.3.10), we identify and neglect all but the most relevant processes that provide a closed set of equations with an analytic, asymptotic solution that shows good qualitative agreement with the numerical treatment. The subset of processes that we include are the formation of diquarks, and the subsequent capture of quarks that lead to the formation of baryons. The reduced set of Boltzmann equations is then

$$\begin{aligned}
-\frac{v_w}{V} N_1' &= -\Big\langle (1,-1) \to (0,0) \Big\rangle - 2\Big\langle (1,1) \to (2,0) \Big\rangle - \Big\langle (2,1) \to (3,0) \Big\rangle \\
&\quad + \Big\langle (2,-1) \to (1,0) \Big\rangle, \\
-\frac{v_w}{V} N_2' &= -\Big\langle (2,-2) \to (0,0) \Big\rangle - \Big\langle (2,1) \to (3,0) \Big\rangle + 2\Big\langle (1,1) \to (2,0) \Big\rangle, \\
-\frac{v_w}{V} N_3' &= -\Big\langle (3,-3) \to (0,0) \Big\rangle - \Big\langle (3,-1) \to (2,0) \Big\rangle + \Big\langle (2,1) \to (3,0) \Big\rangle - \frac{dN_3^{\text{esc}}}{dR}.
\end{aligned} \tag{4.3.14}$$

The analytic solution for this set of equations is obtained relying on several assumptions.

- The initial dark quark abundance N_q^{initial} is determined by the pre-confinement freezeout of the elementary constituents.

- As long as the annihilation rate and the baryon escape rate in the contracting pocket is slower than the pocket contraction rate v_w/R, the total quark number

is conserved. Once those rates are of the same order, the annihilation process "recouples", and the free quark abundance drops toward zero. The condition $\Gamma_{\rm ann} \approx N_q^{\rm initial}\langle\sigma v\rangle_{1(-1)\to 00}/V = v_w/R$ defines the recoupling pocket radius

$$R_{\rm rec} = \sqrt{\frac{3N_q^{\rm initial}\langle\sigma v\rangle_{1(-1)\to 00}}{(4\pi v_w)}}. \tag{4.3.15}$$

- The initial number of bound states, N_X $(X = 2, 3)$, is negligible. As the pocket contracts, bound states start forming. Thus, we can write $N_X \sim R^{-n}$, with $n > 0$, which implies $N_X' \sim N_X/R$. Inserting this into the Boltzmann equation shows that there is a small parameter controlling the rate of change, which is proportional to $\delta = N_X\langle\sigma v\rangle/(R^2 v_w) \propto \langle\sigma v\rangle/R_1^2 \sim \langle\sigma v\rangle\Lambda^{10/3}/M_{\rm pl}^{4/3} \ll 1$. Thus expanding in δ the leading order result is obtained by setting $N_X' \approx 0$, which is the equilibrium condition before the recoupling due to pocket contraction.

Given the above assumptions, before the annihilation process recouples, we have the quark number conservation $N_1 = N_q^{\rm initial} - 2N_2 - 3N_3$. Applying the equilibrium condition before recoupling and neglecting the escape and annihilation terms for the bound states at that point gives

$$\begin{aligned} 2\,\langle\sigma v\rangle_{11\to 20}\left(N_1^2 - \tilde{f}_1 N_2 V\right) &= \langle\sigma v\rangle_{21\to 30}\left(N_2 N_1 - \tilde{f}_2 N_3 V\right), \\ \langle\sigma v\rangle_{21\to 30}\left(N_1 N_2 - \tilde{f}_2 N_3 V\right) &= \langle\sigma v\rangle_{3(-1)\to 20}\left(N_3 N_1 - \tilde{f}_3 N_2 V\right), \end{aligned} \tag{4.3.16}$$

where

$$\begin{aligned} \tilde{f}_1 &= \frac{\left(n_1^{eq}\right)^2}{n_2^{eq}} \propto \exp\left(-\Delta E_1/T_c\right), \\ \tilde{f}_2 &= \frac{n_2^{eq} n_1^{eq}}{n_3^{eq}} \propto \exp\left(-\Delta E_2/T_c\right), \\ \tilde{f}_3 &= \frac{n_3^{eq} n_1^{eq}}{n_2^{eq}} \propto \exp\left(-\Delta E_3/T_c\right), \end{aligned} \tag{4.3.17}$$

and ΔE_i denote the heat released during the above processes. The solution to the above algebraic set of equations provide the abundances of quarks and bound states in the contracting pocket before recoupling as a function of the pocket radius R. The total abundance of produced color-singlet baryons is given by the total baryon abundance N_3 evaluated at the recoupling radius $R_{\rm rec}$. Notice that $e^{-\Delta E_{1,2}/T_c} \sim e^{-\alpha^2 m_q/T_c} \gg e^{-\Delta E_3/T_c} \sim e^{-m_q/T_c}$, where α is evaluated at the bound state's Bohr

radius. Thus, we identify a strong hierarchy $\tilde{f}_1 \ , \ \tilde{f}_2 \gg \tilde{f}_3$.

As a result, a simple analytic expression for the baryon fraction that survives the phase transition relative to the initial quark abundance N_q^{initial} can be found. In the limit of inefficient bound state breaking reactions $\tilde{f}_{1,2}\, V \ll 1$ it is

$$\frac{N_3}{N_q^{\text{initial}}} = \frac{2\langle\sigma v\rangle_{21\to 30}\langle\sigma v\rangle_{11\to 20}}{\langle\sigma v\rangle_{3(-1)\to 20}\langle\sigma v\rangle_{21\to 30} + 4\langle\sigma v\rangle_{3(-1)\to 20}\langle\sigma v\rangle_{11\to 20}} \,. \tag{4.3.18}$$

Thus, assuming all the cross sections are of the same order of magnitude, we see that the baryon survival factor is of order one, if deeply bound states dominate the system. This is the case if the scale hierarchy $m_q \gg \Lambda$ is taken to be extremely large.

In the regime of efficient bound state breaking $\tilde{f}_{1,2} V \gg 1$, we find stronger DM abundance suppression. To simplify things even further, we assume the terms with $\tilde{f}$ dominate and that $\tilde{f}_1 \sim \tilde{f}_2$.[8] With these assumptions, we find that at the recoupling radius we have

$$\frac{N_3}{N_q^{\text{initial}}} = \frac{4\pi v_w^3}{3\tilde{f}_1^2 N_q^{\text{initial}} \langle\sigma v\rangle^3_{1(-1)\to 00}} \,. \tag{4.3.19}$$

Now if we assume in Eq. (4.3.12) the integral is dominated by the contribution around the recoupling point where the bound states total numbers peak, we find

$$\mathcal{S}_{\text{symm.}} \approx 9\frac{v_q}{v_w}\frac{4\pi v_w^3}{3\tilde{f}_1^2 N_q^{\text{initial}} \langle\sigma v\rangle^3_{1(-1)\to 00}} \,. \tag{4.3.20}$$

We can better understand from this equation the effects that various parameters have on the survival factor. Increasing the quark velocity v_q enhances their escape rate (see Eq. (4.3.4)) thus increasing $\mathcal{S}_{\text{symm.}}$. We also see that by increasing $\langle\sigma v\rangle_{1(-1)\to 00}$, the survival factor decreases, which was expected since by increasing this cross section quarks annihilate more against each other instead of binding in bound states. For shallower bound states the binding processes are less favored, thus we expect the survival factor should decrease. This is exactly what Eq. (4.3.20) suggests: for shallower bound states, the Boltzmann suppression in $\tilde{f}_1$ becomes less severe and $\tilde{f}_1$ increases, thus $\mathcal{S}_{\text{symm.}}$ decreases.

The initial density of quarks in a pocket is determined via a pre-confinement, perturbative freezeout calculation. Yet, N_q^{initial} in Eq. (4.3.20) depends on the initial

[8] Both $\tilde{f}_1$ and $\tilde{f}_2$ depend on the ratio m_q/Λ. For $m_q/\Lambda \lesssim 1000$ they are within an orders of magnitude of each other, justifying our assumption. Neglecting this difference allows us to find a simple analytic formula that sheds light on the effect of various quantities on the survival factor.

pocket radius too. Thus, through N_q^{initial} we find that $\mathcal{S}_{\text{symm.}} \sim R_i^{-3}$.

Finally, by decreasing v_w, according to Eq. (4.3.15), the recoupling radius increases, which gives less time for the baryon abundance in the pocket to build up before the interactions become efficient again, see Fig. 4-6. A larger recoupling radius means a smaller peak value for the N_3 abundance, like the one seen at $R\,\Lambda \sim 10^5$ in Fig. 4-6, which in turn decreases the survival factor $\mathcal{S}_{\text{symm.}}$. This behavior is exactly what we see in Eq. (4.3.20).

4.3.4 The effect of quark pressure and summary of assumptions

The v_w scaling of Eq. (4.3.20) helps us better understand how our determination of $\mathcal{S}$ would change had we included the effect of quark pressure on v_w. This equation suggests that by using Eq. (4.2.9) for v_w and ignoring the fact that quark pressure can oppose pocket contraction, we are actually calculating an upper bound on the survival factor, since we are certainly overestimating v_w. Also, this v_w scaling, combined with the proximity of $\mathcal{S}_{\text{symm.}}$ to the asymmetry bound η_{rms} across our parameter space when we use Eq. (4.2.9), motivates us to believe that when the quark pressure is properly taken into account, we should expect that we saturate the the asymmetry bound $\mathcal{S} = \eta_{\text{rms}}$ for every point in the parameter space that we study (see App. C.1.4 for more empirical evidence of this claim). In the upcoming section we use these two limits to bracket the parameter space of the model that reproduces the observed DM relic abundance. We refer to these two limiting scenarios as the *zero quark pressure* and the *asymmetry* scenarios.

Before jumping to the result of solving the Boltzmann equations, it is useful to review all the parameters affecting our calculation of $\mathcal{S}$ and the final DM abundance. The UV model has a very limited set of parameters: the confinement scale, Λ, and the quark mass, m_q. These parameters feed into the calculation of a few secondary quantities that directly affect the calculation of $\mathcal{S}$ and are listed in Tab. 4.2. A precise calculation of these secondary quantities requires various non-perturbative studies. These quantities can be divided into two broad categories: *macroscopic* and *microscopic*.

The *macroscopic* quantities are those concerning the dynamics of the bubbles and pockets, i.e. their initial radius R_i and their wall velocity v_w. While our expressions

for these quantities in Eqs. (4.2.8) and (4.2.9) were based on a simplified simulation of the phase transition (see App. C.1), there is extensive literature concerned with the detailed calculation of these quantities. Unfortunately, this literature has not yet settled on a single, definitive calculation of these quantities, which is why we content ourselves with simple order of magnitude estimates. (See, for example, Refs. [216, 220–222] and references within for various calculations of the wall velocity.)

The *microscopic* quantities include various cross sections and binding energies. They also determine the dimensionless inter-quark spacing, ξ, which directly affects our final results as well. We use the results from [47] for the cross sections and the binding energies. We summarize the relevant quantities in Apps. C.2-C.3.

It is also worth reiterating a few important assumptions that significantly streamlined our analysis. Recall that in Sec. 4.2.2 we argued that the wall velocity is controlled by the amount of supercooling and quark pressure during the phase transition. Following that assumption, we found that the typical velocity of quarks in Eq. (4.2.11) is much faster than the wall velocity even when the quark pressure effect is neglected in Eq. (4.2.9). Therefore, any density gradient within a pocket caused by the compression of the walls can be quickly smoothed out by the thermal motions of the quarks. As a result, we assume that the particles within the pockets are homogeneously distributed, which simplifies our analysis significantly.

We also neglect the abundance of the bound states before the phase transition. Furthermore, as suggested in Fig. 4-1, we assume the quarks are initially well-separated inside the pockets, and that they rebound off the wall surface promptly. As we will argue later, all of our assumptions determine the parts of the parameter space where our analysis is valid.

In principle the formed baryons could further interact to form dark nuclei at the end of the pocket contraction phase. The formation of such multi baryon states in dark sectors has been discussed in Ref. [223]. To estimate the effect in our scenario, we can rescale the predicted interaction rate by the compression factor inside the pockets, which quadratically enhances the formation rate. The contraction factor is defined as $c_f = (R_1/R_{\rm rec})^3$, where $R_{\rm rec}$ is the radius where the interactions recouple. An estimate of this value based on our analytic solution gives $c_f \approx 10^4 (m_q/\Lambda)^{3/2}$. In addition the number density of baryons is rescaled by the survival factor of $1/\sqrt{N_q^{\rm initial}}$, which counteracts the compression effect. Evaluating this in our scenarios, where

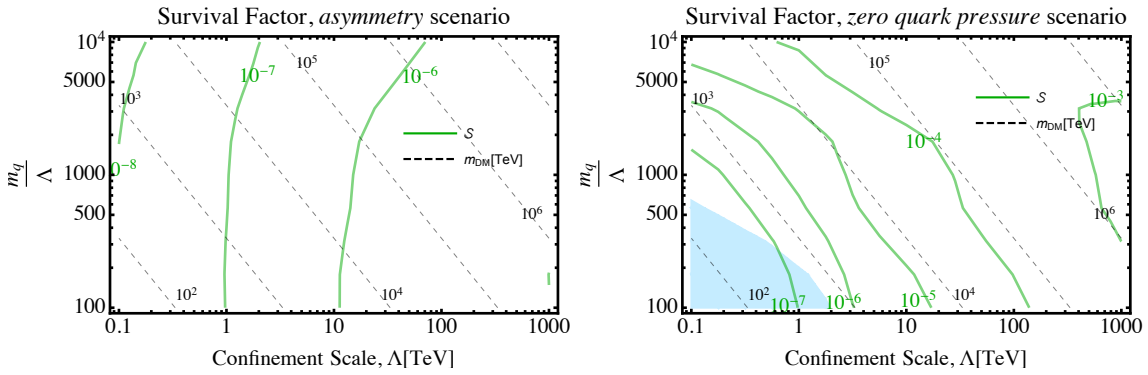


Figure 4-7: Contours of constant survival factor $\mathcal{S}$ (green contours) in the two limiting scenarios that we consider: (i) assuming the asymmetry bound on $\mathcal{S}$ is saturated (on the **left**), or (ii) neglecting the quark pressure effect on the pocket wall velocity (on the **right**). The contours of constant DM mass in TeV are shown as well (black dashed). To obtain the **right** plot, we solve the full set of Boltzmann equations in Eq. (4.3.10) using the values for the initial pocket radius R_i and the pocket wall velocity v_w based on our simulation results discussed in Sec. 4.2 and App. C.1. In the light blue region on the **right** plot we find that the suppression is so severe that only the accidental asymmetric abundance of quarks in each pockets survives after the phase transition. For the **left** plot we simply assume $\mathcal{S} = \eta_{\rm rms}$ for every point in the parameter space. We observe orders of magnitude suppression in the DM abundance due to the second stage of annihilation during the phase transition in either scenarios. Note that the small difference in the 10^{-7} contours in the overlap region is a plotting artifact.

the constituent quark mass is large, and thus the number density at freezeout is significantly reduced, leads to the conclusion that in the parameter space region we consider the di-baryon fraction is at most of the order of 0.5 %. Formation of larger multi-baryon clusters is also severely suppressed due to this di-baryon production bottleneck.

4.3.5 Results and discussion

We now turn to the central results of this chapter. We scan over a range of Λ and m_q/Λ values, solving the Boltzmann equations at each point to calculate the survival factor $\mathcal{S}$. As mentioned above, we use Eq. (4.2.9) for the wall velocity when solving the equations and finding the viable part of the model's parameter space that produces the correct present-day abundance of DM. We argued that this *zero quark pressure* scenario and the *asymmetry* scenario, in which we assume $\mathcal{S} = \eta_{\rm rms}$, are the two limiting cases that bracket the uncertainties in our DM relic abundance calculation. We will find that these two scenarios only give rise to an $\mathcal{O}(1)$ difference in the DM mass range that can explain the observed relic abundance.

In Fig. 4-7 we show contours of constant survival factor for both these scenarios

and for different values of Λ and m_q/Λ. The *asymmetry* scenario plot shows the smallest survival factor possible while the *zero quark pressure* scenario gives an upper bound on the survival factor for every point in the parameter space, see the discussion in Sec. 4.3.3. In the *asymmetry* scenario the only sources of uncertainty are those affecting the pre-confinement calculation and the initial pocket size, while in the *zero quark pressure* scenario the uncertainty in determining the wall velocity v_w should also be included.

The available parameter space in the *asymmetry* limit scenario is shown in Fig. 4-8. Equation (4.2.7) shows that as Λ increases, R_1, and so the number of trapped quarks inside the pocket, decreases. Thus, as expected from Eq. (4.2.16), we find that the larger the initial radius, the smaller the survival factor.

We should keep in mind that many simplifying approximations were made about the dynamics of the phase transition in App. C.1 in order to obtain Eq. (4.2.8) for the bubble radius. This, inevitably, introduces some uncertainty in our calculation. To characterize this uncertainty, in Fig. 4-8 we introduce a fudge factor for the bubble radius denoted by f_R, to be multiplied against the values from Eq. (4.2.8). The observed relic abundance line moves within the light purple band as we vary f_R between 0.1 and 10. Any point above and to the right of the relic abundance line, including the entire red region, is ruled out.

Since the *asymmetry* limit scenario was the lowest attainable $\mathcal{S}$ in our setup, the relic abundance line in this scenario is an upper bound on the possible masses in our model.

In the other limit, the *zero quark pressure* scenario provides us with a lower bound on the range of DM masses in this setup that can explain the observed DM abundance. In Fig. 4-9 we show the available parameter space in this scenario. The calculation can now be affected by a change in both the initial pocket radius R_1 and its wall velocity v_w. To characterize this uncertainty, in Fig. 4-9 we introduce a fudge factor for both the bubble radius and the wall velocity, denoted by f_R and f_v, respectively. As expected, for any fixed Λ the observed DM relic abundance in this scenario is obtained by smaller DM masses than that of the *asymmetry* scenario in Fig. 4-8.

In this figure we also show the relic abundance line when we use the analytic approximation of Eq. (4.3.20) to calculate the survival factor. We find reasonable agreement between our analytic approximation (the orange curve) and the full nu-

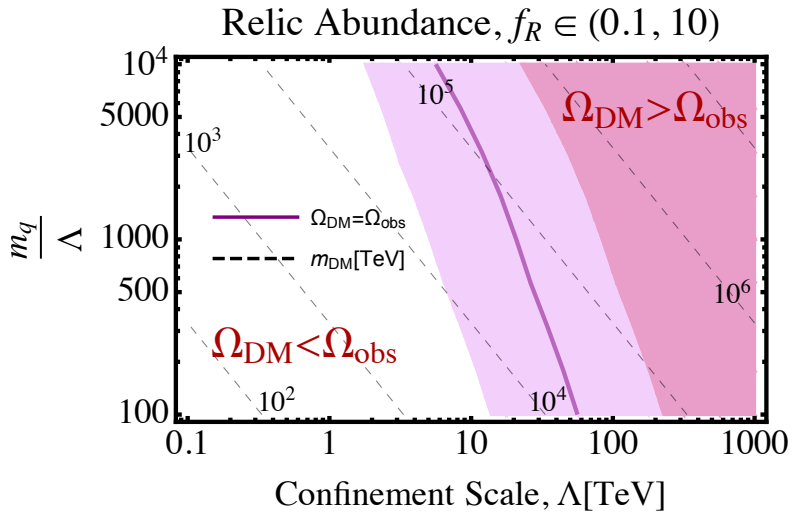


Figure 4-8: The produced abundance of dark baryons, the DM candidate, in the *asymmetry* abundance scenario. The black dashed lines are contours of constant DM mass in TeV. The relic abundance line (with the initial radius fixed to its central value, i.e. $f_R = 1$) is plotted (purple line) along with its uncertainty (light purple shades) corresponding to an order of magnitude variation of the initial radius. The shaded red region is excluded, as it produces too much DM, while the unshaded region produces too little DM. The baryons can therefore constitute a sub-component of the DM within the unshaded regions of parameter space. The survival factor is determined by the accidental asymmetry of the pocket, which is independent of the wall velocity as long as the asymmetry bound is saturated. Thus, the major source of uncertainty in the location of the relic abundance line is the initial pocket radius. We also find that the uncertainty from *microscopic* quantities is sub-dominant to those of initial pocket radius. This figure clearly shows that the baryon masses accounting for the observed DM abundance can be much heavier than the unitarity bound [210].

merical result (the purple curve).

In both these limiting scenarios studied in Figs. 4-8-4-9, we find a similar range of DM masses and Λ that can account for the present-day DM abundance. We expect that these two scenarios bracket the true location of the relic abundance line when the effect of the quark pressure on the pocket wall velocity is appropriately included. The figures indicate that, depending on the *macroscopic* parameters, the region of parameter space that produces the observed DM abundance predicts $m_{\rm DM} \sim \mathcal{O}(1) - \mathcal{O}(100)$ PeV, well above the thermal relic unitarity bound of $m_{\rm DM} \lesssim 300$ TeV [210]. Even with various sources of uncertainty, our results predict a confinement scale roughly in the $\mathcal{O}(1) - \mathcal{O}(100)$ TeV range, in contrast to [184], which predicts a much wider range of confinement scales in such models. The parameter space above this range is ruled out, while the remaining parameter space is allowed, producing a sub-component of DM.

The values of the cross sections and the binding energies entering the Boltzmann

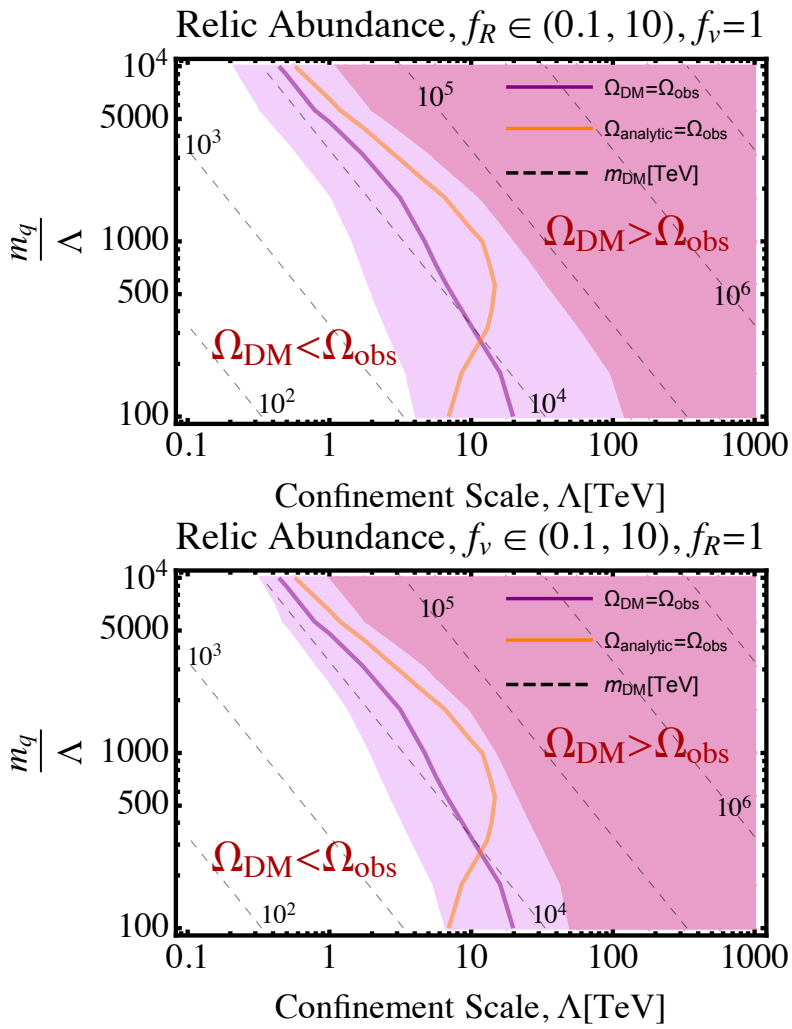


Figure 4-9: Similar to Fig. 4-8 but now with *zero quark pressure*. We vary the pocket initial radius (**top**) or the wall velocity (**bottom**) within one order of magnitude of the central values in Eqs. (4.2.8) and (4.2.9) to characterize the uncertainty in the final relic abundance calculation stemming from these quantities. The relic abundance line using the analytic approximation of Eq. (4.3.20) for the survival factor is denoted by the orange curve too. Even with the quark pressure neglected, we still find a substantial suppression in the DM abundance during the phase transition. We still find that the baryon masses accounting for the observed DM abundance can be much heavier than the unitarity bound [210].

equations can be found in appendices C.2 and C.3, respectively. We find that the uncertainty in our results due to these *microscopic* quantities is sub-dominant to the uncertainty from the *macroscopic* bubble dynamics parameters discussed above. For further details about these parameters and the uncertainties in determining them see the aforementioned appendices and the references therein.

Determining the exact position of the relic abundance line requires more precise calculations of both *macroscopic* and *microscopic* quantities. Nonetheless, such calculations will not change our qualitative conclusions: that the phase transition gives

rise to a new stage of annihilation that reduces the relic abundance by orders of magnitude and shifts the DM mass to well above the unitarity bound.

We can also understand the expected results for the parts of parameter space not plotted. For larger Λs than were plotted, Figs. 4-8-4-9 suggests that this model always overproduces DM and is ruled out. For smaller Λs than were plotted, our assumption that the pre-confinement abundances of bound states are negligible breaks down. Since so many baryons are produced before the start of the phase transition, the survival factor becomes comparable to 1. For low enough Λ, we should use the combinatoric calculation of the relic abundance described in Ref. [184]. An additional aspect is that in this region of parameter space the glueball lifetime can be substantial, which leads to entropy injection and makes the abundance calculation more model dependent, as discussed explicitly for several models in Ref. [184]. Further investigation of this region is left for future works.

As we go to larger values of m_q/Λ, our assumption that $v_q \gg v_w$ breaks down. In this case, local inhomogeneities appear in the distribution of particles in the pockets and the entire homogeneous system of equations in Eq. (4.3.10) must be modified. Furthermore, we find that, for higher m_q/Λ than is shown in Figs. 4-8-4-9, the quark separations during the contraction epoch can become as low as $\sim 1/\Lambda$ (due to the small cross sections allowing for a greater degree of compression). In this case, the picture of well-separated quarks that rebound off the stiff bubble wall (before they run into other colored particles) must be modified. Non-perturbative effects become more relevant in this case. It is also possible that at such high densities quarks bind into more stable and massive dark nuclear states such as nuggets, see [224] for a study of dark quark nuggets in the light dark quark limit.

Furthermore, a recent study has investigated the creation of multi-baryon bound states in a very similar framework [225]. It was found that at $m_q/\Lambda > 10^8$ the fermi-pressure of the quark constituents can lead to the formation of compact objects. However, the framework of the mentioned investigation significantly differs from the setup used in our work. In particular the DM number density is fixed by a preexisting asymmetry and the temperature of the decoupled dark sector is chosen as an initial condition.

Finally, as we go to lower values of m_q/Λ, eventually the first-order phase transition turns into a second order one and then a cross over, see e.g. [199]. In this regime

there will be no bubble walls to compress quarks into a second stage of annihilation. Even for lower values of m_q/Λ for which there still exists a first-order phase transition we run the risk of breaking our assumption that the string breaking rate is negligible, so that quarks and diquarks can escape from the pocket before significant annihilation takes place.

All in all, outside of the window shown in Figs. 4-8-4-9, either the parameter space is already ruled out, or at least one of the simplifying assumptions we made fails and our analysis becomes unreliable.

4.3.6 Extensions of our analysis

So far we have focused on a confining $SU(3)$ gauge group with a single generation of heavy fermions in the fundamental representation. Nonetheless, it is conceptually straightforward to repeat our analysis for slightly different setups. In this section we comment on the differences that we expect would have arisen had we varied the number of colors, N_c, or the quark representation.

Had we chosen a gauge group with a larger number of colors, $SU(N_c \geqslant 4)$, we expect that we would have found a smaller $\mathcal{S}_{\text{symm.}}$ since the stable DM candidate in such a theory (the analogue of the baryon) requires more constituent quarks to bind together in more steps. (Notice that in Fig. 4-6 as the quark number of a state increases, its abundance decreases within a pocket.) However, if even with $N_c = 3$ we find $\mathcal{S}_{\text{symm.}} \leqslant \eta_{\text{rms}}$, we expect to saturate the asymmetry bound for larger gauge groups as well. The additional $\mathcal{S}_{\text{symm.}}$ suppression would not change the final survival factor $\mathcal{S}$.

The quark representation under the dark gauge group has a slightly more complicated effect on our results. For any quark representation, one must first identify the list of all possible bound states and then write down the Boltzmann equations with all possible interactions. As explained in Sec. 4.3.3, the binding energies of these bound states can also have a significant effect on the solutions of the Boltzmann equations. As an example, consider the case in which quarks are in the adjoint representation of the group. These quarks can bind with gluons to form color-neutral gluequarks [186]. Since the gluons can be found abundantly, we expect that quarks can easily pass through pocket walls by binding with a nearby gluon. Thus, pocket walls will not compress the quarks, and there will be no second stage of annihilation due to the

phase transition.

Besides changing the model under consideration, our work would also benefit from improving our order of magnitude estimates and simplifying assumptions. Dedicated numerical simulations that more carefully model the bubble dynamics and non-perturbative physics could reduce the uncertainties in both *macroscopic* and *microscopic* quantities listed in Tab. 4.2, narrowing down the uncertainty on the relic abundance line in Figs. 4-8-4-9.

4.4 Potential Experimental Signals

Our study so far only relied on fairly general properties of a dark sector. We only assumed the dark sector under study is a confining $SU(3)$ gauge theory with a single generation of heavy fermions; we also assumed a portal exists between the sectors that keeps them in kinetic equilibrium and allows the glueballs and mesons to decay to the SM. All the conclusions drawn in the previous sections were independent of further details of the portal and the origin of the heavy dark quark mass.

A detailed study of all the phenomenological signals of such a sector has to be carried out in a model-dependent way with a specified portal. As a result, here we merely list the signals and constraints that should be expected from this broad class of models.

- The main feature of our setup is a first-order phase transition in the early universe. Such a phase transition can also give rise to a stochastic gravitational wave (GW) background that can be detected in a host of different future experiments, e.g. see [226] for a recent study of the GW signals of confining dark sectors. The characteristics of the resulting GW, such as the frequency and the strength, depend on a handful of thermodynamical parameters, see [227] for a brief review. This GW signal is independent of the portal to the SM. A naive estimate[9] shows that different parts of our parameter space could potentially be probed in future experiments like DECIGO [229, 230] and BBO [231]. Early universe phase transitions can also give rise to anisotropies in the GW spectrum, which can potentially be detected at future facilities (see e.g. [232]). Given the extremely high mass range of the DM candidates in our model, the GW signals

[9] We use the formulas in [228] to estimate the GW signal produced during the phase transition. We use the interface introduced in [213] to compare the result to the reach of various experiments.

could have the highest discovery potential in such sectors. We leave the further study of GW signals in this class of models for future work.

- The glueballs and the mesons are unstable due to the portal to the SM. Stringent bounds from BBN require that these relics have a short lifetime. See for instance [233–235] for recent studies. As a rule of thumb, one can avoid various constraints by assuming all these bound states decay before the BBN, i.e. their lifetime is $\tau \leqslant 1\text{s}$. This bound on the lifetime introduces a lower bound on the strength of the portal. This lower bound can vary substantially depending on the details of the portal. Our requirement that both sectors are in kinetic equilibrium also imposes a lower bound, though we expect the BBN bound to be more stringent.

- The portal to the SM introduces possible direct and indirect detection signals. However, the DM number density in the universe and in our galaxy is very suppressed due to this model's heavy DM mass. A naive estimation suggests that our model's indirect signal from DM annihilation within the Milky Way is severely suppressed and undetectable. The direct detection signal, however, depends on the details of the portal and should be studied model-dependently. We note that in this heavy mass range even very large DM-SM elastic cross sections are allowed, but within the reach of upcoming and ongoing experiments [236].

- A separate indirect detection signal comes from the observation that our composite DM model admits excited states. De-excitations from these excited states might lead to radiation that could be detected. Excitations could be produced in the early universe or via interactions with matter today.

- Yet another indirect signal could come from the capture of DM in celestial bodies see for example Refs. [237, 238].[10] As DM accumulates at the bottom of these potential wells, it can begin to annihilate at a significant rate, possibly affecting the evolution of these celestial bodies in an observable way or enhancing a potential annihilation signal [241, 242].

- For the $\mathcal{O}(\text{PeV})$ and above DM masses predicted in our model, direct production of DM at collider facilities is not possible in the foreseeable future. Yet, if

[10] See also [239] and [240] for studies of lighter DM capture in gravitational basins or exoplanets, respectively.

the portal is substantially lighter, it can be directly observed at collider experiments. While the dark quarks are too heavy to produce at collider facilities, the glueballs of the new dark sector, whose mass is $\mathcal{O}(10\Lambda)$, could potentially be produced at future colliders.

- Various studies suggest an upper bound on the DM self-scattering [243–249]. As a rough estimate

$$\sigma_{\rm SI}/m_{\rm DM} \lesssim 1\ {\rm cm}^2{\rm g}^{-1} \sim (60\ {\rm MeV})^{-3}\,. \tag{4.4.1}$$

  It is straightforward to check that for the high confinement scales we are studying, this upper bound is easily satisfied.

- One can also search for signals coming from the inhomogeneities in the DM density that were produced during the phase transition when DM was compressed by contracting pockets, but this seems unlikely. By performing a Jeans stability analysis we find that the internal baryon pressure easily overcomes the self-gravity of these overdensities. Pockets therefore do not seed self-gravitating DM clumps. One might also look for modifications to the matter power spectrum due to these overdensities, but initial estimates indicate that the matter power spectrum would only be modified at unobservably small mass scales if at all. Specifically, the total DM mass within a horizon radius soon after the phase-transition epoch (after which the comoving abundance is fixed) can be estimated as the DM density multiplied by H^{-3}, with a DM density crudely approximated (ignoring changes in the number of relativistic degrees of freedom over time) as $\sim (\Lambda/T_{\rm CMB,0})^3\times$ the present cosmological density of DM, where $T_{\rm CMB,0} \sim 2\times10^{-4}$ eV is the present-day temperature of the radiation bath. This gives an enclosed mass:

$$M_{\rm enc} \sim (\Lambda^2/M_{\rm pl})^{-3}(\Lambda/T_{\rm CMB,0})^3 \times 10^{-6}{\rm GeV/cm}^3 \sim \left(\frac{1{\rm TeV}}{\Lambda}\right)^3 \times 10{\rm kg} \tag{4.4.2}$$

  Thus for phase transitions at the TeV scale and above, we would expect phase-transition-induced inhomogeneities to affect DM clumps at the kg scale and below. Even if these clumps survived, this mass scale is vastly lower than can be probed by any possible observational constraints on the matter power spectrum, which are currently exploring halo masses of order $10^{7-8}M_\odot$ (e.g. [250–252]).

Because of its low number density, the dark matter in our setup can have significant interactions with the SM particles and still have escaped detection so far. Creative new search strategies will be needed to explore this possibility. Novel ideas for direct detection of such a scenario have been put forward in Ref. [236], and interesting signals in heavy isotope searches [253] could arise if our dark baryons can bind to SM atoms and nuclei.

In addition to the above signals, which should exist for any specific realization of the DM-SM portal, there may exist additional portal-dependent signatures. We also find, using the results of Ref. [184], that depending on the type of the portal to the SM the glueballs lifetime could be larger than the Hubble time at $T = \Lambda$. In such a scenario, the delayed decay of the glueballs can further dilute the DM abundance [186, 189] in the parameter space that we have studied, thus pushing the relic abundance line in Figs. 4-8-4-9 to even higher DM masses. A proper study of this effect, as well as other signals from any specific portal, is left for future works.

4.5 Conclusion

In this work we studied the consequences of a first-order phase transition in a confining dark sector with a single heavy quark in the fundamental representation. We assumed a portal exists between our sector and the dark sector that keeps the two sectors at kinetic equilibrium at the time of the phase transition and respects dark baryon number conservation. The arguments we presented do not depend on further details of the portal.

We argued that the bubbles of the confined phase, after nucleation, expand very slowly. Soon after the bubbles come in contact and coalesce, pockets of the deconfined phase form, and are submerged in a sea of the confined phase. The quarks are trapped inside these isolated and ever-contracting deconfined phase pockets. There is always an accidental asymmetry in the net dark baryon number in a given pocket, due to local stochastic fluctuations in the number of quarks and anti-quarks at the onset of pocket formation. As the pockets contract, the enclosed quarks compress until formerly frozen-out interactions recouple, giving rise to a second stage of annihilation.

We wrote down the complete set of Boltzmann equations with all $2 \to 2$ interactions between lowest lying bound states of the heavy quark. By solving these equations, we were able to calculate the fraction of dark quarks that survive the sec-

ond annihilation event. These surviving quarks bind into stable, color-singlet states that comprise the DM abundance we see today. We find that these Boltzmann equations predict a dramatic suppression in the DM relic abundance. This suppression is sensitive to the initial size of the pocket, the density of the quarks trapped within, and the pocket wall velocity. While there is a large uncertainty in determining these parameters, we showed that for virtually any reasonable values of these parameters there is a significant suppression in the DM relic abundance.

We find the effect of quark pressure on the pocket wall velocity difficult to model. However, we do know that this effect will further slow down the pocket wall, which we showed will imply a smaller survival factor. We calculated the relic abundance of DM in this setup in two extreme scenarios: (i) the *zero quark pressure* scenario, and (ii) when we assume the quark pressure is so severe that the *asymmetry* bound on the survival factor is saturated. These two limiting scenarios bracket the range over which the relic abundance line can move when the quark pressure effects are properly taken into account. We found that for a fixed dark confinement scale Λ, the DM mass in this setup only changes by $\mathcal{O}(1)$ factors between these two scenarios.

After identifying the parts of the $m_q - \Lambda$ parameter space that predict the observed present-day DM abundance, we found that this large suppression opens up parts of the parameter space that were previously thought to be ruled out. In particular, we found a DM mass scale well above the often-quoted unitarity bound. Our calculation also suggests an upper bound on the dark confinement scale $\Lambda \sim \mathcal{O}(1) - \mathcal{O}(100)$ TeV. For any Λ above this bound DM is overproduced, despite the dramatic suppression of its abundance during the phase transition. Depending on the value of Λ, the dark baryon mass that can explain the observed DM abundance varies roughly between 10^3 to 10^5 TeV.

There are many possible signals that our setup can give rise to. With the exception of gravitational waves, all the other potentially detectable signals depend on the specific form of this model's portal to the SM. It will be interesting to investigate the signatures of specific portals and their constraints, which we leave to future work.

There are numerous ways in which our analysis can be improved. To decrease the uncertainties in our results it will be important to perform more detailed numerical simulations of the *macroscopic* bubble dynamics during the phase transition and the *microscopic* strong dynamics that determine the particle interactions. The most im-

portant quantities to be calculated would be the initial pocket size and its subsequent contraction rate, and the cross sections and binding energies included in our Boltzmann equations. Additionally, it will also be interesting to study the relic abundance calculation for other gauge groups and different quark representations. Furthermore, for any specific portal we should study the potential DM dilution due to a delayed glueball decay after the phase transition.

Confining sectors are natural dark sector candidates. In this chapter we focused on such sectors with only a single species of heavy dark quark. We have pointed out the dramatic effect that this model's first-order phase transition has on the relic DM abundance of such a sector. The dynamics lead us to a sharp prediction about the natural mass scale of such DM candidates, 10^{3-5} TeV. It is of paramount importance to study variations of these minimal dark sectors in greater detail and their potential signatures in various upcoming experiments.

Chapter 5

Conclusion

In this thesis, I have offered a glimpse into the power of early universe studies of exotic dark matter. A wide range of scenarios, like dark matter conversion to standard model matter or a first-order phase transition in the dark sector, produce significant effects in the early universe. These effects can then have observable consequences, as was the case in Chapter 2 and Chapter 3, or expand our conception of what can constitute dark matter, as was the case in Chapter 4.

There are many directions in which to extend the ideas presented in this thesis. Specifically, `DarkHistory` can be used to explore the effects of dark matter energy injection on other early universe observables. By improving the treatment of low-energy electrons and tracking the transitions between the different excited states of Hydrogen, `DarkHistory` will be able to calculate the spectral distortion to the CMB due to dark matter energy injection. By relaxing the assumption of homogeneity of energy injection and deposition and also interfacing with 21-cm codes, `DarkHistory` will also be able to calculate the effect that dark matter energy injection has on the 21-cm power spectrum.

Hopefully some of the general ideas presented in this thesis, like model-dependent and model-independent studies of dark matter or the use of early universe observables, will lead to a better understanding of dark matter. With so much high-quality cosmological and astrophysical data coming our way – coming from new telescopes, 21-cm experiments, state-of-the-art N-body simulations, gravitational wave detectors, etc. – I am optimistic that we will eventually be able to crack the dark matter mystery.

Appendix A

DarkHistory

A.1 Inverse Compton Scattering

In this appendix, we discuss in detail the methods used to compute the spectra of photons that are produced by the cooling of electrons through ICS. We restore $\hbar$, c and k_B in this appendix, since the exact numerical value of these spectra is important.

A.1.1 Scattered Spectra

We begin with some preliminaries that will be important in understanding our subsequent discussion of ICS. The goal is to determine the secondary photon spectrum produced on average by multiple scatterings of a single electron.

Consider an electron with energy E_e and corresponding Lorentz factor γ incident on some distribution of photons $n(\epsilon)$ with initial energy ϵ in the comoving frame. Since we are only interested in ICS off the CMB, we will only consider an isotropic photon bath in the co-moving frame, distributed as a blackbody with some temperature T. The electron has some probability per unit time of scattering the photons into some outgoing energy ϵ_1, with some probability distribution $dN_\gamma/(d\epsilon\, d\epsilon_1\, dt)$, which we call the "differential scattered photon spectrum". This quantity is proportional to the number density per unit energy of the photon bath $n(\epsilon)$, so that integrating over ϵ also integrates over the distribution of these photons. This can be interpreted as a normalized scattered photon spectrum for ICS by many electrons with the same energy. Integrating the differential scattered photon spectrum with respect to ϵ gives us the "scattered photon spectrum",

$$\frac{dN_\gamma}{d\epsilon_1\, dt}(E_e, T, \epsilon_1) = \int_{\epsilon_{\min}}^{\epsilon_{\max}} d\epsilon \frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(E_e, T, \epsilon, \epsilon_1)\,, \tag{A.1.1}$$

with $\epsilon_{\rm min}$ and $\epsilon_{\rm max}$ determined by the kinematics of ICS.

We further define the "scattered photon energy loss spectrum",

$$\frac{dN_\gamma}{d\Delta\, dt}(E_e, T, \Delta) = \int d\epsilon \frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(E_e, T, \epsilon, \epsilon_1 = \epsilon + \Delta), \tag{A.1.2}$$

where Δ is the change in energy of a photon scattering by a single electron. This is simply the distribution of scattered photons as a function of the energy gained or lost by the photon during the scattering.

Now, consider some arbitrary injection spectrum of electrons $d\tilde{N}_e/dE_1$. The tilde serves to remind the reader that this is a distribution of electrons, and not a normalized quantity. From the definition of Eq. (A.1.2), we define the "scattered electron spectrum" as

$$\frac{d\tilde{N}_e}{dE_1\, dt} = \int_0^\infty dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, T, \Delta = E - E_1)\,, \tag{A.1.3}$$

where E_1 is the energy of the scattered electron. However, this result allows some electrons to gain energy after scattering, significantly complicating our calculations. Intuitively, we expect electrons that upscatter from $E \to E_1$ to partially cancel with downscatters from $E_1 - E$, justifying an approximate treatment where we simply cancel out photons that downscatter (and upscatters an electron) with photons that upscatter (and downscatters an electron). We leave a full justification of this to the end of this section, but for now, we will accordingly define the "scattered electron net energy loss spectrum",

$$\frac{dN_e}{d\Delta\, dt}(\beta, T, \Delta) = \frac{dN_\gamma}{d\Delta\, dt}(\beta, T, \Delta) - \frac{dN_\gamma}{d\Delta\, dt}(\beta, T, -\Delta), \tag{A.1.4}$$

with $\Delta \geq 0$ in the expression above. For relativistic electrons, the average energy lost due to an upscattering a photon is much larger than the average energy gained due to downscattering a photon, and it is therefore a good approximation to consider only scattering events where electrons lose their energy [254]. The upscattered photons also have outgoing energy $\epsilon_1 \gg \epsilon$, and so a reasonable approximation to make in the relativistic limit is

$$\left.\frac{dN_e}{d\Delta\, dt}\right|_{\beta\to 1} \approx \left.\frac{dN_\gamma}{d\epsilon_1\, dt}\right|_{\beta\to 1}. \tag{A.1.5}$$

We now turn our attention to justifying the approximation laid out in Eq. (A.1.4).

First, we split the exact integral in Eq. (A.1.3) into an integral from 0 to E_1, and from E_1 to ∞. The first integral can be rewritten as (dropping the T dependence for clarity)

$$\int_0^{E_1} dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, \Delta = E - E_1) \\ = -\int_{E_1}^{2E_1} dx \frac{d\tilde{N}}{dx} \frac{dN_\gamma}{d\Delta\, dt}(E = 2E_1 - x, \Delta = E_1 - x)\,, \quad \text{(A.1.6)}$$

where we have simply made the substitution $x = 2E_1 - E$. In this part of the integral, we are dealing with upscattered electrons and downscattered photons, and so we know that $dN_\gamma/(d\Delta\, dt)$ only has support when $E - E_1 \sim T_{\rm CMB} \ll E, E_1$, since ICS is only included for electrons with $E > 3\,\text{keV}$ [82]. This implies that the integral only has support near $x = E_1$, and we can therefore make the following approximation:

$$\int_0^{E_1} dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, \Delta = E - E_1) \\ \approx -\int_{E_1}^{\infty} dx \frac{d\tilde{N}_e}{dx} \frac{dN_\gamma}{d\Delta\, dt}(E = x, \Delta = E_1 - x) \\ = -\int_{E_1}^{\infty} dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, \Delta = E_1 - E)\,, \quad \text{(A.1.7)}$$

where in the last step we have trivially relabeled $x \to E$. We have therefore shown that

$$\frac{d\tilde{N}'_e}{dE_1\, dt} \approx -\int_{E_1}^{\infty} dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, \Delta = E_1 - E) \\ + \int_{E_1}^{\infty} dE\, \frac{d\tilde{N}_e}{dE} \frac{dN_\gamma}{d\Delta\, dt}(E, \Delta = E - E_1)\,, \quad \text{(A.1.8)}$$

and that is a good approximation due to the relatively low temperature of the CMB.

With these definitions in mind, we are now ready to understand how to compute these scattered spectra when the electron is in two limits. For $\gamma > 20$, the spectra are computed in the relativistic limit, while below that, scattering with the CMB at all relevant redshifts lie well within the Thomson regime. Together, they cover all relevant kinematic regimes that we consider in our code.

Relativistic Electrons

The differential upscattered photon spectrum produced by ICS between an electron and the CMB blackbody spectrum in the relativistic regime ($\gamma \gg 1$) is given by [254]

$$\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt} = \frac{2\pi r_0^2 c}{\gamma^2}\frac{n(\epsilon,T)}{\epsilon}\left[2q\log q + (1+2q)(1-q) + \frac{1}{2}\frac{(\Gamma(\epsilon)q)^2}{1+\Gamma(\epsilon)q}(1-q)\right], \quad \text{(A.1.9)}$$

where r_0 is the classical electron radius, m_e is the electron mass, ϵ is the incident photon energy in the comoving frame, and ϵ_1 is the scattered photon energy in the same frame, and we have defined

$$\Gamma(\epsilon) = \frac{4\epsilon\gamma}{m_e c^2}, \quad q = \frac{\epsilon_1}{\gamma m_e c^2 - \epsilon_1}\frac{1}{\Gamma(\epsilon)}. \quad \text{(A.1.10)}$$

We stress that Eq. (A.1.9) is strictly only correct when photons are upscattered by the incoming electron, which corresponds to the kinematic regime $\epsilon \leq \epsilon_1 \leq 4\epsilon\gamma^2/(1+4\epsilon\gamma/m)$. In the opposite regime where $\epsilon/(4\gamma^2) \leq \epsilon_1 < \epsilon$ and photons get downscattered, we have [255]

$$\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt} = \frac{\pi r_0^2 c}{2\gamma^4\epsilon}\left(\frac{4\gamma^2\epsilon_1}{\epsilon} - 1\right) n(\epsilon,T). \quad \text{(A.1.11)}$$

For ICS off CMB photons, the $n(\epsilon)$ is the number density of photons per unit energy; for a blackbody, this is

$$n_{\rm BB}(\epsilon,T) = \frac{1}{\pi^2\hbar^3 c^3}\frac{\epsilon^2}{\exp(\epsilon/k_B T) - 1}, \quad \text{(A.1.12)}$$

where T is the temperature of the CMB.

The complete upscattered photon spectrum for ICS off the CMB is therefore obtained by performing the integral in Eq. (A.1.9) over ϵ, with the kinematic limits given by $1/4\gamma^2 \leq q \leq 1$ [254]. Since the CMB photons at $z \lesssim 3000$ have energies less than 1 eV, the amount of energy transferred by an electron is always completely dominated by Eq. (A.1.9). Furthermore, one can check that at $q = 1/4\gamma^2$, $\epsilon \gg T$. We can therefore make the approximation that Eq. (A.1.9) gives the full ICS spectrum while neglecting Eq. (A.1.11), and take the integral limits to be $0 \leq q \leq 1$ instead. This assumption is made in the ICS transfer functions provided as part of the downloaded data, but options are available in the `ics` module to turn these various assumptions off.

The quantity $\Gamma(\epsilon)$ separates the two kinematic regimes of Compton scattering:

$\Gamma \gg 1$ for the Klein-Nishina regime, where Compton scattering in the electron rest frame is highly inelastic, and $\Gamma \ll 1$ for the Thomson regime, where it is almost elastic instead.[1] Eq. (A.1.9) applies to both regimes, with the only assumption being $\gamma \gg 1$.

To avoid computing the scattered photon spectrum repeatedly in the code, we use the following relation between spectra at different temperatures:

$$\frac{dN_\gamma}{d\epsilon_1\, dt}(E_e, yT, \epsilon_1) = y^4 \frac{dN_\gamma}{d\epsilon_1\, dt}(yE_e, T, y\epsilon_1)\,, \tag{A.1.13}$$

for any real positive number y, even if yE_e is unphysical.[2] In `DarkHistory`, we evaluate the scattered photon spectrum at $1 + z = 400$, and use this relation to compute the subsequent spectra at lower redshifts by a straightforward interpolation.

Thomson Regime

In the Thomson regime, the rate at which photons are scattered is given by [254]

$$\frac{dN_\gamma}{dt} = \sigma_T c N_{\text{rad}}, \tag{A.1.14}$$

where N_{rad} is the total number density of incident photons, with $\sigma_T = 8\pi r_0^2/3$ being the Thomson cross section. Note that the scattering rate is independent on the incident photon energy. The energy loss rate of the electron is [254]

$$\frac{dE_e}{dt} = \frac{4}{3}\sigma_T c \gamma^2 \beta^2 U_{\text{rad}}, \tag{A.1.15}$$

where β is the velocity of the electron, with U_{rad} being the total energy density of the incident photons.

While Eqs. (A.1.14) and (A.1.15) are well-known, the actual spectrum of scattered photons in the Thomson regime is much less so. The complete expression for the differential scattered photon spectrum with no further assumptions is, as far as the authors know, first given in Ref. [256], and we reproduce their final result here for

[1] Although the scattering process is almost elastic in the initial electron rest frame, it is certainly not elastic in the co-moving frame. In the co-moving frame, the electron loses a small fraction of its energy per collision, but each collision can upscatter a CMB photon by a significant factor.

[2] This trick can only be performed by integrating over $0 \leq q \leq 1$, and is the key reason for making such an approximation.

completeness. For $(1-\beta)\epsilon_1/(1+\beta) < \epsilon < \epsilon_1$, we have

$$\begin{aligned}\left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(\beta,T,\epsilon,\epsilon_1)\right|_{\epsilon<\epsilon_1} =& \frac{\pi r_0^2 c n(\epsilon,T)}{4\beta^6\gamma^2\epsilon}\left\{\frac{1}{\gamma^4}\frac{\epsilon}{\epsilon_1} - \frac{1}{\gamma^4}\frac{\epsilon_1^2}{\epsilon^2}\right.\\ &+(1+\beta)\left[\beta(\beta^2+3)+\frac{1}{\gamma^2}(9-4\beta^2)\right]\\ &+(1-\beta)\left[\beta(\beta^2+3)-\frac{1}{\gamma^2}(9-4\beta^2)\right]\frac{\epsilon_1}{\epsilon}\\ &\left.-\frac{2}{\gamma^2}(3-\beta^2)\left(1+\frac{\epsilon_1}{\epsilon}\right)\log\left(\frac{1+\beta}{1-\beta}\frac{\epsilon}{\epsilon_1}\right)\right\}, \end{aligned} \tag{A.1.16}$$

and for $\epsilon_1 < \epsilon < (1+\beta)\epsilon_1/(1-\beta)$,

$$\left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(\beta,T,\epsilon,\epsilon_1)\right|_{\epsilon\geq\epsilon_1} = -\left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(-\beta,T,\epsilon,\epsilon_1)\right|_{\epsilon<\epsilon_1}. \tag{A.1.17}$$

All other values of ϵ outside of the ranges specified are kinematically forbidden, and so to find the spectrum, we need to integrate over ϵ with $n(\epsilon) = n_{\rm BB}(\epsilon)$ in the finite range specified above, i.e.

$$\begin{aligned}\frac{dN_\gamma}{d\epsilon_1\, dt}(\beta,T,\epsilon_1) =& \int_{\frac{1-\beta}{1+\beta}\epsilon_1}^{\epsilon_1} d\epsilon\, \left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(\beta,T,\epsilon,\epsilon_1)\right|_{\epsilon<\epsilon_1}\\ &-\int_{\epsilon_1}^{\frac{1+\beta}{1-\beta}\epsilon_1} d\epsilon\, \left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(-\beta,T,\epsilon,\epsilon_1)\right|_{\epsilon<\epsilon_1}. \end{aligned} \tag{A.1.18}$$

The relationship between spectra at different temperatures is given by

$$\frac{dN_\gamma}{d\epsilon_1\, dt}(\beta,yT,\epsilon_1) = y^2\frac{dN_\gamma}{d\epsilon_1\, dt}(\beta,T,\epsilon_1/y)\,. \tag{A.1.19}$$

The scattered photon energy loss spectrum $dN_\gamma/(d\Delta\, dt)$ is similarly given by

$$\frac{dN_\gamma}{d\Delta dt}(\beta,T,\Delta) = \begin{cases}\int_{\frac{1-\beta}{2\beta}\Delta}^{\infty} d\epsilon\, \left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(\beta,T,\epsilon,\epsilon+\Delta)\right|_{\epsilon<\epsilon_1}, & \Delta > 0,\\ \int_{-\frac{1+\beta}{2\beta}\Delta}^{\infty} d\epsilon\, \left.\frac{dN_\gamma}{d\epsilon\, d\epsilon_1\, dt}(\beta,T,\epsilon,\epsilon+\Delta)\right|_{\epsilon\geq\epsilon_1}, & \Delta \leq 0.\end{cases} \tag{A.1.20}$$

The relation shown in Eq. (A.1.19) between scattered photon spectra of different temperatures also holds for the energy loss spectrum, with $\epsilon_1 \to \Delta$.

A.1.2 Numerical Methods

Computationally, to evaluate all of the scattered spectra, we need to perform numerical quadrature over a large range of electron and scattered photon energies; using a

standard grid of 5000×5000 energy values, the grid would take the standard `numpy` integrator over a day to populate. While a substantial speed-up may be obtained by using packages like `Cython` [257], numerical quadrature for ICS in the Thomson regime is also subject to significant numerical errors when the electron is nonrelativistic due to the existence of catastrophic cancellations. A semi-analytic approach provides both a faster method and a way to avoid such errors in a robust manner.

Thomson and Relativistic Regime: Large β

For $\beta \gtrsim 0.1$, we can obtain the scattered photon spectrum in Eq. (A.1.16) in the Thomson regime or Eq. (A.1.9) in the relativistic regime, as well as the scattered electron energy loss spectrum in the Thomson regime in Eq. (A.1.4), by direct integration.

The problem of integrating these expressions reduces to obtaining an expression for indefinite integrals over the Bose-Einstein distribution of the form

$$P_f(y) \equiv \int \frac{f(y)\, dy}{e^y - 1}\,. \tag{A.1.21}$$

Throughout this appendix, we ignore the constant of integration for such indefinite integrals, since we will ultimately be taking differences of such expressions to find definite integrals. For $f(y) \equiv y^n$ with integer $n \geq 0$, the indefinite integral is known explicitly:

$$P_{y^n}(x) = -n! \sum_{s=0}^{n} \frac{x^s}{s!} \mathrm{Li}_{n-s+1}(e^{-x}) \quad (n = 0, 1, 2, \cdots), \tag{A.1.22}$$

where $\mathrm{Li}_m(z)$ is the polylogarithm function of order m with argument z (see Appendix A.1.4 for the definition). Note however that NumPy does not have a numerical function for the polylogarithm of order $m > 2$, and so the semi-analytic method that we describe below is still necessary for P_{y^n}, $n \geq 2$ due to this limitation.

For other functions $f(y)$, closed-form solutions do not exist. However, an expression for the indefinite integral as an infinite series can be obtained [258]. Importantly, more than one series expression exists for all of the integrals $P_f(x)$ of interest in both the relativistic and nonrelativistic regimes, so that it is always possible to find a series expression that converges quickly for any integration limit. We tabulate the series expressions already found in Ref. [258] for completeness, together with the many new

series expressions derived in this appendix required for the nonrelativistic limit in Appendix A.1.4.

Thomson Regime: Small β

In the Thomson regime for $\beta \lesssim 0.1$, catastrophic cancellations between terms in the integral make even the method described above insufficient. After integrating Eq. (A.1.18) over ϵ to get the scattered photon spectrum, for example, the final result must be $\mathcal{O}(\beta^0)$, even though the prefactor in Eq. (A.1.16) is $\mathcal{O}(\beta^{-6})$. The integrals of all of the terms in the curly braces of Eq. (A.1.16) and their analog from Eq. (A.1.17) must therefore cancel among themselves to 1 part in β^{-6}; such a computation is impossible to perform for $\beta \lesssim 0.003$ due to floating point inaccuracy, even with double precision.

We avoid this problem by expanding the scattered photon spectrum in Eq. (A.1.18) and the mean electron energy loss spectrum in Eq. (A.1.4). Eq. (A.1.18) can be expanded straightforwardly in β, but Eq. (A.1.4) must be expanded in both β and $\xi \equiv \Delta/T$, since catastrophic cancellations occur when either variable is small. In `DarkHistory`, we expand these expressions up to $\mathcal{O}(\beta^6)$ and $\mathcal{O}(\xi^6)$, but the precision of this calculation is systematically improvable by adding more terms to the code as desired. The exact expressions for the expansions, details of their derivations and several consistency checks for these expressions can be found in Appendix A.1.4.

A.1.3 Results

Figs. A-1 and A-2 show the scattered photon spectrum in the Thomson and relativistic regimes respectively as a function of electron energy, at a CMB temperature of 0.25 eV, corresponding to a redshift of $z \approx 1065$ that is near recombination. By default, `DarkHistory` transitions between these two limits at $\gamma = 20$. Fig. A-3 shows the mean electron energy loss spectrum in the Thomson regime. Above $\gamma = 20$, `DarkHistory` uses the approximation shown in Eq. (A.1.5). Finally, the computed secondary photon spectrum after completely cooling of all electrons and positrons through ICS is shown in Fig. A-4.

All results shown here are computed using a 500×500 grid of electron and photon energies/energy loss, and each can be completed under ten seconds on a typical personal computer.

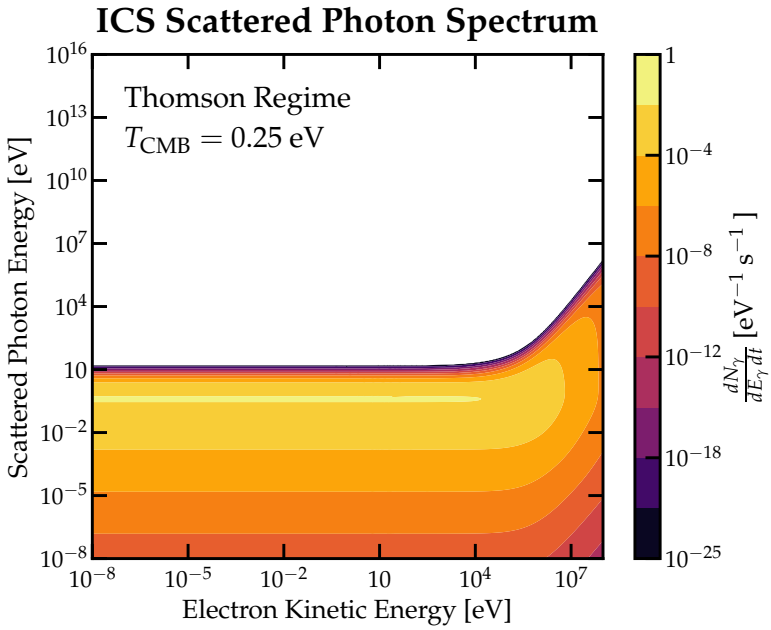


Figure A-1: The ICS scattered photon spectrum in the Thomson regime, with $T_{\rm CMB}$ = 0.25 eV.

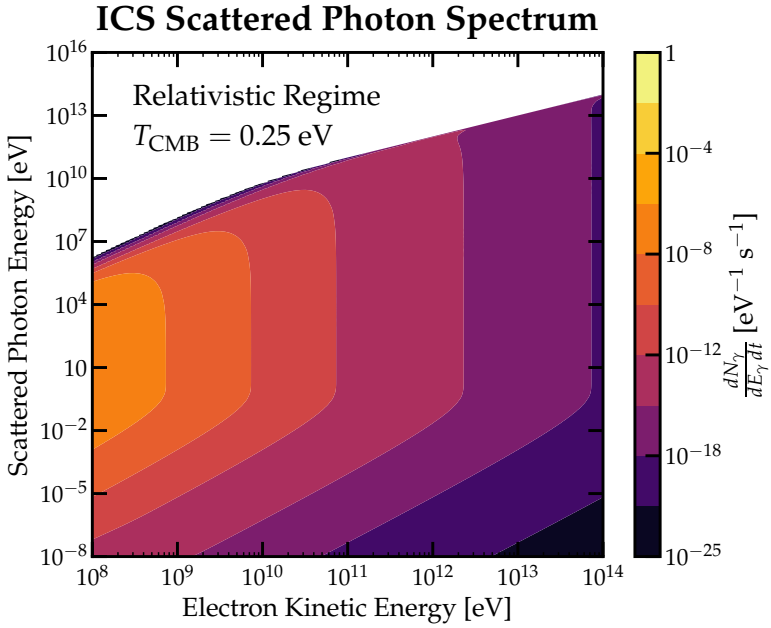


Figure A-2: The ICS scattered photon spectrum in the relativistic regime, with $T_{\rm CMB}$ = 0.25 eV.

A.1.4 Integrals and Series Expansions

We are now ready to detail the integrals and series expansions used in the numerical methods described above.

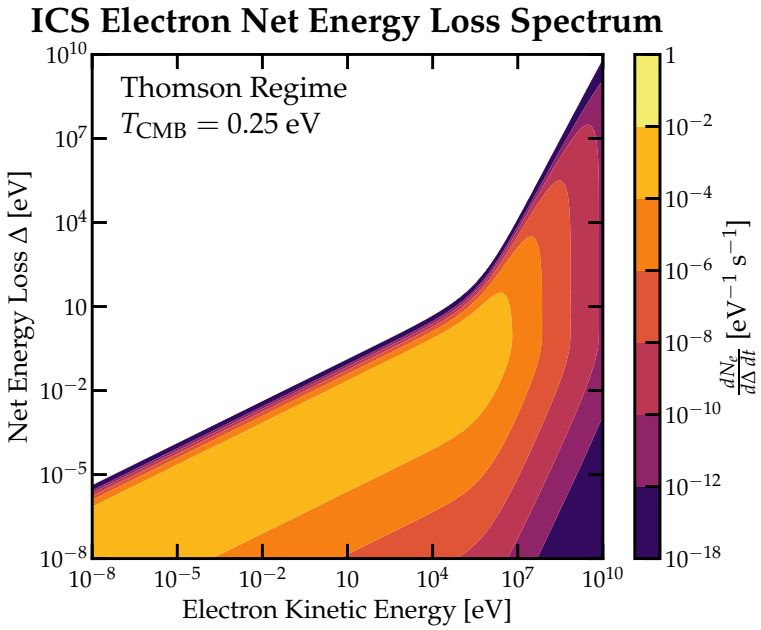


Figure A-3: The ICS mean electron energy loss spectrum in the Thomson regime, with $T_{\rm CMB} = 0.25\,\text{eV}$.

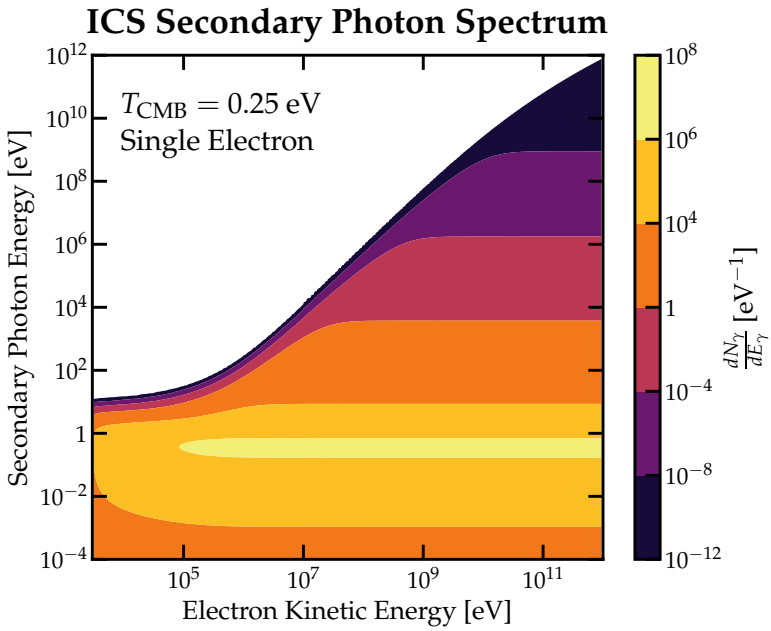


Figure A-4: The ICS secondary photon spectrum after complete cooling of a single electron, with $T_{\rm CMB} = 0.25\,\text{eV}$.

Bose Einstein Integrals

Each $P_f(x)$ that is of interest has a series that converges quickly for small values of x, and another that converges quickly for large values of x. `DarkHistory` by default chooses $x = 2$ as the value to switch between the two expressions.

Suppose we approximate the indefinite integral in Eq. (A.1.21) $P_f(x)$ by the first

N terms of its series expression, which we denote $S_N(x)$. Let $S_N^s(x)$ and $S_N^l(x)$ be the series expressions we obtain for $x < 2$ and $x \geq 2$ respectively. In all of the cases we are interested in, $S^l_{N\to\infty}(x \to \infty) = 0$ (with the constant of integration taken to be zero) due to the exponential function in the denominator of the original integral, and so

$$S^l_{N\to\infty}(b > 2) = -\int_b^\infty \frac{f(y)\,dy}{e^y - 1}\,. \tag{A.1.23}$$

Then defining $\Delta S_N^{s,l}(a,b) = S_N^{s,l}(b) - S_N^{s,l}(a)$, the definite integral is evaluated as

$$\int_a^b \frac{f(y)\,dy}{e^y-1} = \begin{cases} \Delta S^s_{N\to\infty}(a,b), & a < 2, b < 2; \\ \Delta S^s_{N\to\infty}(a,2) + \Delta S^l_{N\to\infty}(2,b), & a < 2, b \geq 2; \\ \Delta S^l_{N\to\infty}(a,b), & a \geq 2, b \geq 2. \end{cases} \tag{A.1.24}$$

Terms are added sequentially until the next contribution to the full integral falls below a given relative tolerance; the default value for this tolerance used by `DarkHistory` is 10^{-10}.

Before listing the series expressions, we must first introduce some notation that will be relevant. The numbers and analytic functions defined below are all well-known, but are often defined with different normalizations or given different names. We explicitly define all relevant functions used here for clarity.

B_n are the Bernoulli numbers, defined through the following exponential generating function:

$$\frac{x}{e^x - 1} \equiv \sum_{n=0}^\infty \frac{B_n x^n}{n!}, \tag{A.1.25}$$

with $B_{0,1,2,\cdots} = 1, -1/2, 1/6, \cdots$. Note that $B_{2j+1} = 0$ for all integers $j > 0$.

Next, we define the generalized exponential integrals

$$E_n(x) \equiv \int_1^\infty \frac{e^{-xt}}{t^n}\,dt \tag{A.1.26}$$

and the closely related incomplete gamma function

$$\Gamma(n,x) \equiv \int_x^\infty t^{n-1}e^{-t}\,dt. \tag{A.1.27}$$

The polylogarithm of order m, denoted $\mathrm{Li}_m(z)$, is defined as

$$\mathrm{Li}_m(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^m}\,. \tag{A.1.28}$$

Finally, we define ${}_2F_1(a,b;c;z)$, the Gaussian hypergeometric function, as

$${}_2F_1(a,b;c;z) \equiv 1 + \frac{ab}{1!c}z + \frac{a(a+1)b(b+1)}{2!c(c+1)}z^2 + \cdots = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, \tag{A.1.29}$$

where $(x)_n \equiv x(x+1)\cdots(x+n-1)$ is the Pochhammer symbol. This function only appears in the form $R(n,x) \equiv \mathrm{Re}[{}_2F_1(1,n+1;n+2;x)]$, where Re denotes the real part; to avoid the slow evaluation of the `hyp2f1` function in NumPy, we use instead the following relation:

$$\begin{aligned} R(n,x) \equiv\ & \mathrm{Re}[{}_2F_1(1,n+1;n+2;x)] \\ =& -(n+1)x^{-(n+1)}\log(|1-x|) - \sum_{j=1}^{n} \frac{n+1}{j} x^{j-n-1}\,. \end{aligned} \tag{A.1.30}$$

The list of all of the series expressions that we use, including those already derived in [258], are shown in Tables A.1 and A.2 for $x < 2$ and $x \geq 2$ respectively.

Nonrelativistic Thomson Limit: Small Parameter Expansion

The expression for the scattered photon spectrum in the Thomson limit, shown in Eq. (A.1.18), can be expanded in the small β limit, to obtain

$$\frac{dN_\gamma}{d\epsilon_1 dt_1} = \frac{3\sigma_T k_B^2 T^2}{32\pi^2\hbar^3 c^2} \sum_{n=0}^{\infty} \sum_{j=1}^{2n} \frac{A_n \beta^{2n} x_1^3 P_{n,j}(x_1) e^{-jx_1}}{(1-e^{-x_1})^{2n+1}}\,, \tag{A.1.31}$$

where $x_1 = \epsilon_1/T$, A_n is a constant, and $P_{j,n}(x_1)$ is some rational or polynomial function in x_1. These quantities are as follows. For $n = 0$ i.e. $\mathcal{O}(\beta^0)$,

$$A_0 = \frac{32}{3}, \quad P_{0,1}(x) = \frac{1}{x}\,. \tag{A.1.32}$$

For $n = 1$, i.e. $\mathcal{O}(\beta^2)$,

$$A_1 = \frac{32}{9}, \quad P_{1,1}(x) = x - 4, \quad P_{1,2}(x) = x + 4\,. \tag{A.1.33}$$

$f(y)$	$P_f,\ x < 2$
$y^n,\ n \geq 1$	$\displaystyle\sum_{k=0}^{\infty} \frac{B_k x^{k+n}}{k!(k+n)}$
$y \log y$	$\displaystyle x \log x - x + \sum_{k=1}^{\infty} \frac{B_k x^{k+1}}{k!(k+1)} \left[\log x - \frac{1}{k+1}\right]$
$y \log(y+a),\ a > -x$	$\displaystyle\sum_{k=0}^{\infty} \frac{B_k x^{k+1}}{k!(k+1)} \left[\log(x+a) + \frac{R(k,-x/a)}{k+1} - \frac{1}{k+1}\right]$
1	$\log(1-e^{-x})$
$\log y$	$\displaystyle\frac{1}{2}\log^2 x + \sum_{k=1}^{\infty} \frac{B_k x^k}{k!k} \left[\log x - \frac{1}{k}\right]$
$\log(y+a),\ a > 0$	$\displaystyle\log x \log a - \mathrm{Li}_2(-x/a) + \sum_{k=1}^{\infty} \frac{B_k x^k}{k!k} \left[\log(x+a) - \frac{x}{a(k+1)} R(k,-x/a)\right]$
$\log(y+a),\ -x < a < 0$	$\displaystyle\log(-x/a)\log(x+a) - \mathrm{Li}_2(1+x/a) + \sum_{k=1}^{\infty} \frac{B_k x^k}{k!k} \left[\log(x+a) - \frac{x}{a(k+1)} R(k,-x/a)\right]$
$\dfrac{1}{y+a},\ a > -x$	$\displaystyle\frac{1}{a}\log\left(\frac{x}{x+a}\right) + \sum_{k=1}^{\infty} \frac{B_k x^k}{k!k} \left[\frac{1}{a} - \frac{kx}{(n+1)a^2} R(k,-x/a)\right]$
$y^{-n},\ n \geq 1$	$\displaystyle\sum_{k=0}^{n-1} \frac{B_k}{k!} \frac{x^{k-n}}{k-n} + \frac{B_n}{n!}\log x + \sum_{k=1}^{\infty} \frac{B_{k+n}}{(k+n)!} \frac{x^k}{k}$

Table A.1: Series expressions for the relevant indefinite integrals of the form shown in Eq. (A.1.21). Here, y is the integration variable, and x denotes the integration limit of interest. These expressions are used for $x < 2$.

$f(y)$	$P_f,\ x \geq 2$
$(y+a)^n,\ \forall n \in \mathbb{Z},\ a > -x$	$\displaystyle\sum_{k=1}^{\infty} \frac{e^{ka}\Gamma\big(n+1, k(x+a)\big)}{k^{n+1}} = \sum_{k=1}^{\infty} \frac{e^{ka} E_{-n}\big(k(x+a)\big)}{(x+a)^{-k-1}}$
$y\log(y+a),\ a > -x$	$\displaystyle\sum_{k=1}^{\infty} \frac{e^{ka}}{k^2}\left[(1+kx)e^{-k(x+a)}\log(x+a) + (1+kx)E_1\big(k(x+a)\big) + E_2\big(k(x+a)\big)\right]$
$\log(y+a),\ a > -x$	$\displaystyle\sum_{k=1}^{\infty} \frac{e^{ka}}{k}\left[e^{-k(x+a)}\log(x+a) + E_1\big(k(x+a)\big)\right]$

Table A.2: Series expressions for the relevant indefinite integrals of the form shown in Eq. (A.1.21). Here, y is the integration variable, and x denotes the integration limit of interest. These expressions are used for $x \geq 2$.

For $n = 2$, i.e. $\mathcal{O}(\beta^4)$,

$$\begin{aligned} A_2 &= \frac{16}{225}, \\ P_{2,1}(x) &= 7x^3 - 84x^2 + 260x - 200, \\ P_{2,2}(x) &= 77x^3 - 252x^2 - 260x + 600, \\ P_{2,3}(x) &= 77x^3 + 252x^2 - 260x - 600, \\ P_{2,4}(x) &= 7x^3 + 84x^2 + 260x + 200, \end{aligned} \tag{A.1.34}$$

and finally for $n = 3$, i.e. $\mathcal{O}(\beta^6)$,

$$\begin{aligned} A_3 &= \frac{16}{4725}, \\ P_{3,1}(x) &= 11x^5 - 264x^4 + 2142x^3 - 7224x^2 + 9870x - 4200, \\ P_{3,2}(x) &= 3(209x^5 - 2200x^4 + 6426x^3 - 2408x^2 - 9870x + 7000), \\ P_{3,3}(x) &= 2(1661x^5 - 5280x^4 - 10710x^3 + 28896x^2 + 9870x - 21000), \\ P_{3,4}(x) &= 2(1661x^5 + 5280x^4 - 10710x^3 - 28896x^2 + 9870x + 21000), \\ P_{3,5}(x) &= 3(209x^5 + 2200x^4 + 6426x^3 + 2408x^2 - 9870x - 7000), \\ P_{3,6}(x) &= 11x^5 + 264x^4 + 2142x^3 + 7224x^2 + 9870x + 4200. \end{aligned} \tag{A.1.35}$$

Furthermore, when x_1 is small, it becomes numerically advantageous to expand Eq. (A.1.31) in x_1 as well, leaving a simple polynomial in x_1 and β, i.e.

$$\frac{dN_1}{d\epsilon_1 dt_1} = \frac{3\sigma_T k_B^2 T^2}{32\pi^2 \hbar^3 c^2} \sum_{n=0}^{\infty} \sum_{j=1}^{\infty} C_{n,j} \beta^{2n} x_1^j \,. \tag{A.1.36}$$

The values of $C_{n,j}$ are shown in Table A.3.

Three checks can be performed to verify that this is indeed the correct expansion in β. First, taking $\beta \to 0$, the scattered photon spectrum simply becomes $dN_\gamma/(d\epsilon_1\, dt_1) = n_{\rm BB}(\epsilon_1, T)\sigma_T c$, which is exactly the expected result for Thomson scattering in the rest frame of the electron: all photons simply scatter elastically at a rate governed by the Thomson scattering cross section, thus remaining in a blackbody distribution. Second, a more non-trivial check is to integrate Eq. (A.1.31) with respect to ϵ_1, giving the total Thomson scattering rate given in Eq. (A.1.14). Since the scattering rate is independent of β, the $\mathcal{O}(\beta^0)$ term in the series should integrate to exactly $\sigma_T c N_{\rm rad}$ where $N_{\rm rad}$ is the number density of the blackbody photons, while

$C_{n,j}$	x_1	x_1^2	x_1^3	x_1^5	x_1^7	x_1^9	x_1^{11}
β^0	32/3	-16/3	8/9	-2/135	1/2835	-1/113400	1/4490640
β^2	-64/9	0	32/27	-4/45	8/1701	-1/4860	1/124740
β^4	-256/225	0	32/27	-296/1125	1208/42525	-64/30375	389/3118500
β^6	-832/1575	0	32/27	-1828/3375	31352/297675	-10669/850500	10267/9355500

Table A.3: List of coefficients $C_{n,j}$ for use in Eq. (A.1.36).

the other higher order terms should integrate to exactly zero. This is indeed the case for the series expansion shown here. Lastly, one can check that Eq. (A.1.31) agrees with the energy loss expression Eq. (A.1.15), by noting that

$$\int \frac{dN_\gamma}{d\epsilon_1\, dt_1}\epsilon_1\, d\epsilon_1 = \sigma_T c u_{\rm BB}(T) + \frac{4}{3}\sigma_T c\beta^2\gamma^2 u_{\rm BB}(T)\,, \tag{A.1.37}$$

where $u_{\rm BB}(T)$ is the blackbody energy density with temperature T, i.e. the produced secondary photon spectrum must have the same energy as the upscattered CMB photons plus the energy lost from the scattering electron. This check has also been performed for the series expansions shown here.

For the scattered electron energy loss spectrum shown in Eq. (A.1.4), the small β and ξ expansion can be written as

$$\frac{dN_e}{d\Delta\, dt} = \frac{3\sigma_T k_B^2 T^2}{32\pi^2\hbar^3 c^2}\sum_{n=0}^{\infty}\left[\sum_{j=1}^{2n}\frac{A^{j+1}Q_{n,j}(e^{-A})}{(1-e^{-A})^j\beta^{-2n}} + R_n(A)\right], \tag{A.1.38}$$

where $Q_{n,j}(x)$ is a polynomial, $A \equiv \Delta/(2\beta T) = \xi/(2\beta)$, and $R_n(A)$ is a sum of integrals of the form

$$P_k(A) = A^{k+1}\int_A^\infty \frac{x^{-k}\,dx}{e^x - 1}\,. \tag{A.1.39}$$

These integrals can be evaluated using the same methods detailed in Appendix A.1.4. The list of polynomials $Q_{n,j}$ and of $R_n(A)$ is given below. All expressions not listed should be taken to be zero. For $n = 0$,

$$R_0(A) = \frac{176}{15}P_0 - \frac{64}{3}P_3 + \frac{128}{5}P_5\,. \tag{A.1.40}$$

For $n = 1$,

$$\begin{aligned} Q_{1,1}(x) &= -\frac{32}{3}x\,, \qquad Q_{1,2} = \frac{8}{3}x\,, \\ R_1(A) &= -\frac{1168}{105}P_0 + \frac{128}{3}P_3 - \frac{2176}{15}P_5 + \frac{1280}{7}P_7\,. \end{aligned} \tag{A.1.41}$$

For $n = 2$,

$$\begin{aligned} Q_{2,1}(x) &= -\frac{512}{15}x\,, \qquad Q_{2,2}(x) = \frac{8}{5}x\,, \\ Q_{2,3}(x) &= -\frac{8}{15}x(1+x)\,, \\ Q_{2,4}(x) &= \frac{2}{15}(x + 4x^2 + x^3)\,, \end{aligned}$$

$$R_2(A) = -\frac{64}{3}P_3 + \frac{640}{3}P_5 - 768P_7 + \frac{14336}{15}P_9\,. \tag{A.1.42}$$

And finally for $n = 3$,

$$\begin{aligned}
Q_{3,1}(x) &= -\frac{416}{3}x\,, \qquad Q_{3,2}(x) = \frac{1184}{105}x\,,\\
Q_{3,3}(x) &= -\frac{256}{315}(x + x^2)\,,\\
Q_{3,4}(x) &= -\frac{2}{63}(x + 4x^2 + x^3)\,,\\
Q_{3,5}(x) &= -\frac{4}{315}(x + 11x^2 + 11x^3 + x^4)\,,\\
Q_{3,6}(x) &= \frac{1}{315}(x + 26x^2 + 66x^3 + 26x^4 + x^5)\,,\\
R_3(A) &= -\frac{512}{3465}P_0 - \frac{1408}{15}P_5 + \frac{6912}{7}P_7 - \frac{161792}{45}P_9 + \frac{49152}{11}P_{11}\,.
\end{aligned} \tag{A.1.43}$$

These are all the terms necessary to work at order $\mathcal{O}(\beta^6)$ and $\mathcal{O}(\xi^6)$. As before, if A becomes small, we should expand Eq. (A.1.38) as

$$\frac{dN_e}{d\Delta\, dt} = \frac{3\sigma_T k_B^2 T^2}{32\pi^2\hbar^3 c^2}\sum_{n=0}^{\infty}\left[\sum_{j=0}^{\infty} D_{n,j}\beta^{2n}A^j + R_n(A)\right], \tag{A.1.44}$$

with the values of $D_{n,j}$ shown in Table A.4. These expressions are complicated, but can be checked in a similar fashion as the scattered photon spectrum by integrating over $\Delta\, d\Delta$ to obtain the mean energy loss rate of electrons scattering of a blackbody spectrum, given exactly in Eq. (A.1.15). Using the fact that

$$\int_0^\infty d\Delta\, \Delta\, P_k(A) = \frac{4\pi^4\beta^2T^2}{15(n+2)}\,, \tag{A.1.45}$$

one can verify that integrating the $\mathcal{O}(\beta^6)$ expansion gives

$$\frac{dE_e}{dt} = \frac{4}{3}\sigma_T c U_{\rm rad}\beta^2(1 + \beta^2 + \beta^4 + \beta^6)\,, \tag{A.1.46}$$

which is precisely the Taylor expansion of Eq. (A.1.15) in powers of β.

$D_{n,j}$	A	A^2	A^3	A^5	A^7	A^9	A^{11}
β^2	-8	16/3	-10/9	7/270	-1/1260	11/453600	-13/17962560
β^4	-164/5	256/15	-134/45	161/2700	-19/9450	359/4536000	-289/89812800
β^6	-40676/315	208/3	-1312/105	4651/18900	-416/59535	989/4536000	-173/22453200

Table A.4: List of coefficients $D_{n,j}$ for use in Eq. (A.1.44).

A.2 Positronium Annihilation Spectra

The spin-triplet $^3S_1$ state of positronium annihilates to three photons, producing a photon spectrum per annihilation given by [259]

$$\left.\frac{dN_\gamma}{dE_\gamma}\right|_{^3S_1} = \frac{6}{(\pi^2-9)m_e}\left\{\frac{2-x}{x} + \frac{x(1-x)}{(2-x)^2} + 2\log(1-x)\left[\frac{1-x}{x^2} - \frac{(1-x)^2}{(2-x)^3}\right]\right\}, \tag{A.2.1}$$

where $x \equiv E_\gamma/m_e$. The kinematically allowed range is $0 \leq x \leq 1$. Assuming that the formation of positronium by low energy positrons populates all of the degenerate ground states equally, the averaged photon spectrum per annihilation is

$$\left.\frac{dN_\gamma}{dE_\gamma}\right|_{\text{Ps}} = \frac{1}{4}\delta(E_\gamma - m_e) + \frac{3}{4}\left.\frac{dN_\gamma}{dE_\gamma}\right|_{^3S_1}. \tag{A.2.2}$$

A.3 Cross Checks

A.3.1 Helium Deposition

In this section, we compare the various helium energy deposition methods discussed in Sec. 2.3.5. We pick a dark matter candidate which decays to two photons with a lifetime of 3×10^{24} s as an example, but the results are similar across different dark matter masses and energy injection rates.

Switching between methods can be done with the parameter `compute_fs_method` passed to `evolve()`, with the following strings for each method: (1) `'no_He'`, (2) `'He_recomb'` and (3) `'He'`, e.g.

```
helium_method_alt = main.evolve(
    DM_process='decay', mDM=1e8, lifetime=3e24,
    primary='phot_delta', start_rs=3000., backreaction=True,
    helium_TLA=True, compute_fs_method='He_recomb'
)
```

Fig. A-5 shows the helium ionization fraction x_{HeII} as a function of redshift for each of the different methods. In method (1), x_{HeII} is simply the baseline helium ionization fraction, which is almost entirely neutral once helium recombination is complete. No energy is assigned to helium iondization at all. Method (2) has no contribution to helium ionization from photons, since every ionized helium atom is assumed to recombine, producing a photon that photoionizes hydrogen instead (i.e.

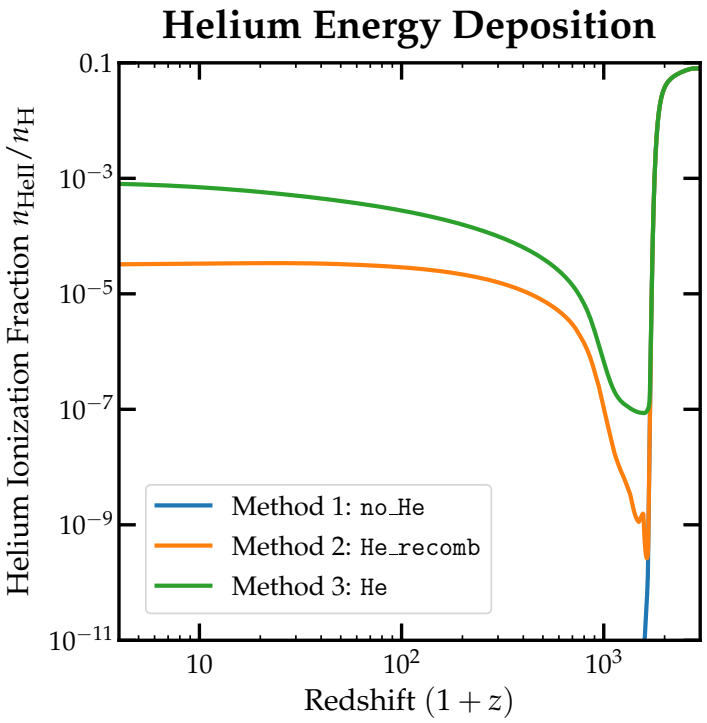


Figure A-5: Helium ionization fraction with different helium energy deposition methods: (1) no tracking of the helium evolution (i.e. $x_{\rm HeII}$ takes its baseline value) (blue) (2) all photoionized HeI atoms recombine, producing a photon that photoionizes hydrogen (orange), and (3) photoionized HeI atoms remain photoionized (green). The energy injection corresponds to 100 MeV DM decaying through $\chi \to \gamma\gamma$ with a lifetime of 3×10^{24} s.

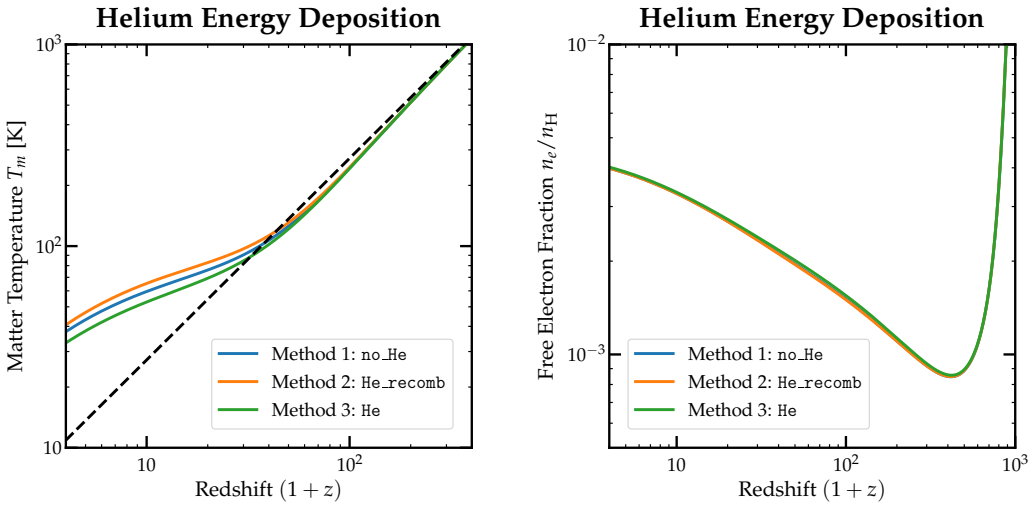


Figure A-6: Matter temperature (left) and free electron fraction (right) evolution with different helium energy deposition methods: (1) no tracking of the helium evolution (i.e. $x_{\rm HeII}$ takes its baseline value) (blue) (2) all photoionized HeI atoms recombine, producing a photon that photoionizes hydrogen (orange), and (3) photoionized HeI atoms remain photoionized (green). The CMB temperature is shown for reference (black, dashed). The energy injection corresponds to 100 MeV DM decaying through $\chi \to \gamma\gamma$ with a lifetime of 3×10^{24} s.

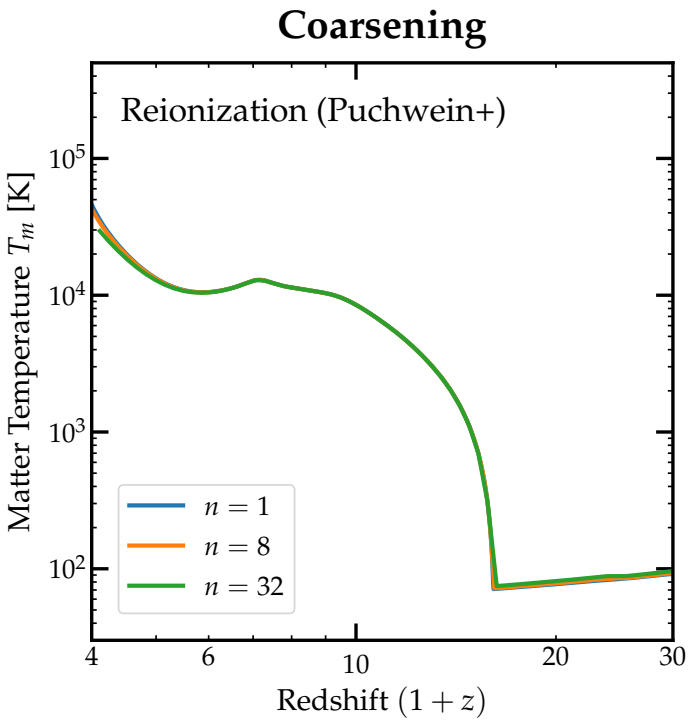


Figure A-7: Matter temperature evolution with the default reionization model with no coarsening (blue), a coarsening factor of 8 (orange) and 32 (green).

setting $q^{\gamma}_{\mathrm{He}} = 0$ in Eq. (2.42)). The helium ionization level therefore deviates from the baseline only from energy injection in the $\mathrm{He}_{\mathrm{ion}}$ channel from low-energy electrons. On the other hand, method (3) assumes that all helium atoms that get photoionized stay ionized, maximizing the amount of energy into $\mathrm{He}_{\mathrm{ion}}$ from photons (i.e. setting $q^{\gamma}_{\mathrm{He}} = 1 - q$ in Eq. (2.42)). This explains the higher x_{HeII} obtained.

Despite these differences in x_{HeII}, the evolution of x_e remains almost identical, due to the fact that the total number of ionization events between both hydrogen and helium remains the same regardless of method used. This in turn ensures only a small difference in T_m between the methods. The ionization and temperature histories for all three methods for the particular channel we have chosen are shown in Fig. A-6. Users may bracket the uncertainty in the treatment of helium with methods (2) and (3).

A.3.2 Coarsening

In the absence of reionization, a coarsening factor of up to 32 has been found to yield a small relative difference of between 5–10% in the values of $f_c(z)$ across the full range of redshifts used in `DarkHistory`. With reionization, however, T_m evolves more rapidly and attains larger values, and too much coarsening can lead to large absolute

differences and somewhat larger relative differences in T_m, since we are averaging over the T_m evolution over many redshift points. Fig. A-7 shows the resultant temperature evolution as a function of redshift for the same model used in the previous section but with the default reionization model turned on, with coarsening factors of 1, 8 and 32. Once reionization starts, the difference in T_m is $\sim 15\%$ for $n = 32$ compared to the uncoarsened result at $z \sim 4$, corresponding to an absolute error of $\sim 5000\,\mathrm{K}$. Prior to reionization, the relative errors are slightly smaller at $\lesssim 10\%$.

We therefore recommend using a coarsening factor of up to 32 if no reionization models are used, depending on the level of precision desired, and to use coarsening with care once reionization is included. We also emphasize that when using coarsening, it is best to check for convergence by comparing the result with less coarsening.

A.3.3 $f_c(z)$ Contours

Fig. A-8 show the computed $f_c(z)$ contours within `DarkHistory` for all channels of interest without any backreaction. The new $f_c(z)$ calculation by `DarkHistory` makes several small physics and numerical improvements over the previous calculation of these results [30, 82], but still agree to within less than 10% when methodologies (cosmological parameters, methods of interpolation etc.) are standardized between the code used in Ref. [82] and `DarkHistory`. The new calculation also corrects a bug in earlier work in the treatment of prompt energy deposition from nonrelativistic and mildly relativistic injected electrons. This accounts for the bulk of the visible differences in Fig. A-8 between the current contours and those of Refs. [30, 82], which are most pronounced for DM annihilation/decay to electrons and low injected particle energies.

A.4 Table of Definitions

Table A.5 shows a list of variables and their definitions for reference.

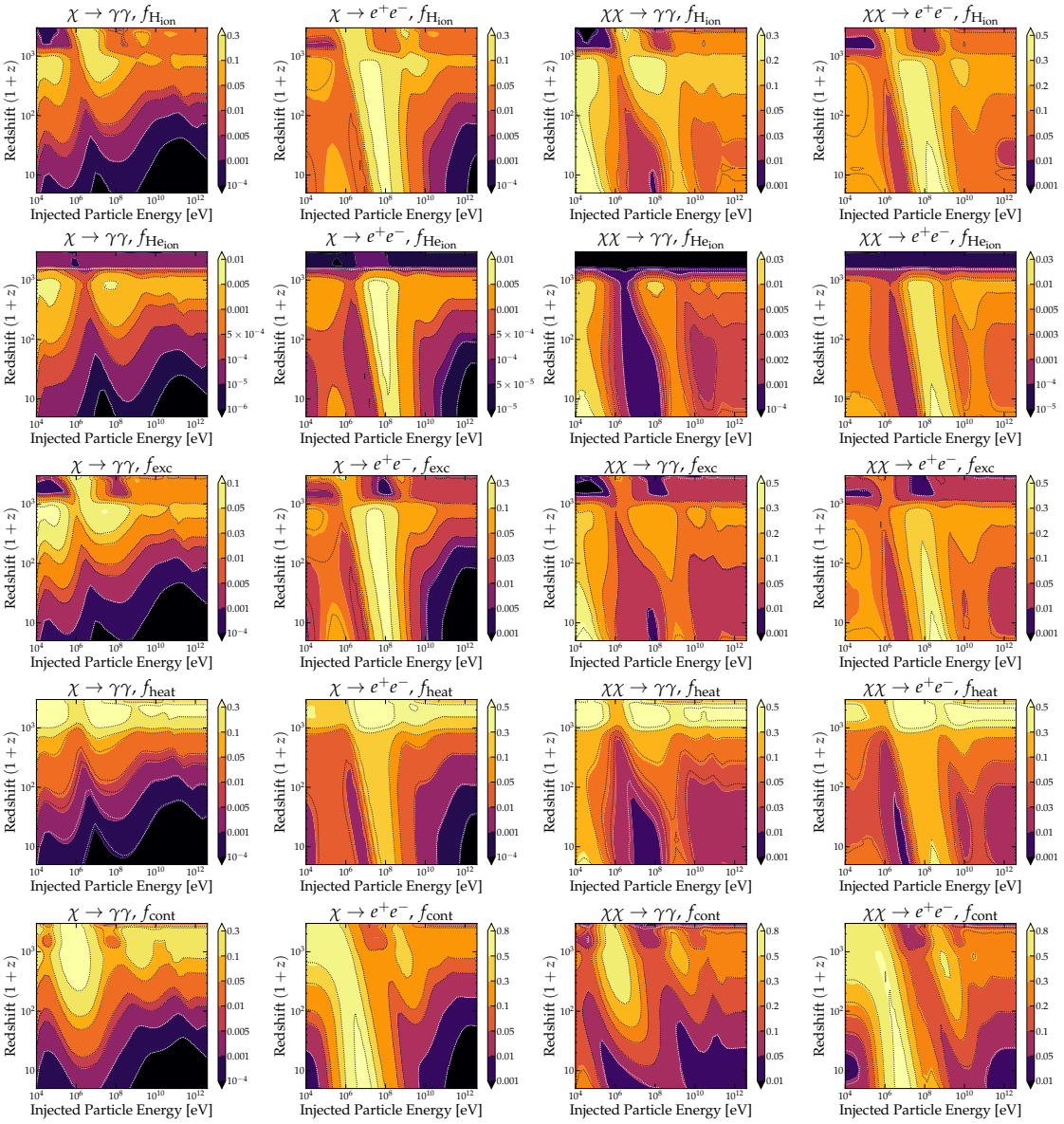


Figure A-8: Computed $f_c(z)$ values without backreaction with `DarkHistory` for (from left to right) $\chi \to \gamma\gamma$ decays, $\chi \to e^+e^-$ decays, $\chi\chi \to \gamma\gamma$ annihilations and $\chi\chi \to e^+e^-$ annihilations (with no boost factor). The results from Refs. [30, 82] are shown for comparison (dashed lines). These contour plots agree with the previous results to within 10% if all calculation methods are standardized between `DarkHistory` and Refs. [30, 82], and represent an improved calculation of $f_c(z)$ neglecting backreaction.

Category	Symbol	Definition
General	y	log-redshift, $y \equiv \log(1+z)$.
	Δy, Δt	log-redshift step size and associated time step size.
	$\mathbf{x}$	Ionization levels: $\mathbf{x} \equiv (x_{\rm HII}, x_{\rm HeII}, x_{\rm HeIII}) \equiv (n_{\rm HII}/n_{\rm H}, n_{\rm HeII}/n_{\rm H}, n_{\rm HeIII}/n_{\rm H})$ i.e. the fractional abundance of ionized hydrogen atoms, singly-ionized helium atoms and doubly-ionized helium atoms with respect to the number of hydrogen atoms (both neutral and ionized).
	ζ_i	$\operatorname{arctanh}\left[(2/\chi_i)\,(n_i/n_{\rm H} - \chi_i/2)\right]$ where $i \in \{\rm HII, HeII, HeIII\}$, convenient reparametrization of $\mathbf{x}$ introduced for numerical purposes.
	T_m	Temperature of the IGM.
	$T_m^{(0)}$, $x_{\rm HII}^{(0)}(z)$	Baseline temperature and ionization histories, obtained from Eq. (2.1).
	m_χ, τ, $\langle\sigma v\rangle$	Dark matter mass, lifetime, and velocity-averaged annihilation cross section.
	$\left(\frac{dE}{dV\,dt}\right)_{\rm inj}$	Energy injection rate per volume for exotic forms of energy injection, given for dark matter annihilation/decay in Eq. (2.3).
Spectra	$G(z)$	Conversion factor between the rate of injected events per volume to the number of injected events per baryon within a log-redshift step, as defined in Eq. (2.10).

	$\overline{\mathbf{N}}^{\alpha}_{\text{inj}}[E'_{\alpha,i}]$	Spectrum containing number of particles of type $\alpha \in \{\gamma, e\}$ injected into energy bin $E'_{\alpha,i}$ per annihilation event.
	$\mathbf{N}^{\alpha}_{\text{inj}}[E'_{\alpha,i}, y']$	Spectrum containing the number of particles per baryon in a log-redshift step of type α injected into energy bin $E'_{\alpha,i}$ at log-redshift y', as defined in Eq. (2.9).
	$\overline{\mathbf{N}}^{\gamma}_{\text{pos}}[E'_{\gamma,i}]$	Spectrum of photons produced from a single positronium annihilation event.
	$\mathbf{N}^{\gamma}_{\text{new}}[E'_{\gamma,i}, y']$	Sum of the spectra of primary injected photons, and secondary photons produced by the cooling of electrons, as defined in Eq. (2.23).
	$\mathbf{N}^{\gamma}_{\text{prop}}[E'_{\gamma,i}, y']$	Spectrum of propagating photons with energies greater than 13.6 eV that do not photoionize or get otherwise deposited into low-energy photons.
	$\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$	$\mathbf{N}^{\gamma}_{\text{prop}}[E'_{\gamma,i}, y'] + \mathbf{N}^{\gamma}_{\text{new}}[E'_{\gamma,i}, y']$, as defined in Eq. (2.28).
	$\mathbf{N}^{\alpha}_{\text{low}}[E_{\alpha,i}, y]$	Low-energy photons ($\alpha = \gamma$) or electrons ($\alpha = e$) at log-redshift y.
Photon Cooling	$\overline{\mathsf{P}}^{\gamma}[E'_{\gamma,i}, E_{\gamma,j}, y', \Delta y, \mathbf{x}]$	Transfer function for propagating photons, which multiplies $\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$ and produces $\mathbf{N}^{\gamma}_{\text{prop}}[E_{\gamma,j}, y' - \Delta y]$, as defined in Eq. (2.27).
	$\overline{\mathsf{D}}^{e}[E'_{\gamma,i}, E_{e,j}, y', \Delta y, \mathbf{x}]$	Low-energy electron deposition transfer function, which multiplies $\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$ and produces $\mathbf{N}^{e}_{\text{low}}[E_{e,j}, y' - \Delta y]$, as defined in Eq. (2.29).

	$\overline{\mathsf{D}}^{\gamma}[E'_{\gamma,i}, E_{\gamma,j}, y', \Delta y, \mathbf{x}]$	Low-energy photon deposition transfer function, which multiplies $\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$ and produces $\mathbf{N}^{\gamma}_{\text{low}}[E_{\gamma,j}, y' - \Delta y]$, as defined in Eq. (2.31).
	$\overline{\mathsf{D}}^{\text{high}}_{c}[E'_{\gamma,i}, y', \Delta y, \mathbf{x}]$	High-energy deposition transfer matrix, which multiplies $\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$ and returns the total energy that greater than 3 keV electrons produce during the cooling process deposit into channel $c \in$ {'ion', 'exc', 'heat'}, as defined in Eq. (2.32) in the next log-redshift step at $y' - \Delta y$.
	$\left(\overline{\mathsf{P}}^{\gamma}_{1/2}\right)^{n}$	Coarsened propagating photon transfer function with a coarsening factor of n, as defined in Eq. (2.33), which multiplies $\mathbf{N}^{\gamma}[E'_{\gamma,i}, y']$ and produces $\mathbf{N}^{\gamma}_{\text{prop}}[E_{\gamma,j}, y' - n\Delta y]$.
Electron Cooling	$\overline{\mathsf{N}}[E'_{e,i}, E_{e,j}]$	Spectrum of secondary electrons produced due to the cooling of a single injected electron with initial energy $E'_{e,i}$.
	$\overline{\mathbf{R}}_{c}[E'_{e,i}]$	High-energy deposition vector containing the total energy deposited into channel $c \in$ {'ion', 'exc', 'heat'} by a single injected electron with kinetic energy $E'_{e,i}$, as defined in Eq. (2.13).
	$\overline{\mathbf{R}}_{\text{CMB}}[E'_{e,i}]$	Total initial energy of CMB photons that are upscattered via ICS due to the cooling of a single electron of energy $E'_{e,i}$.
	$\overline{\mathsf{T}}_{\text{ICS},0}[E'_{e,i}, E_{\gamma,j}]$	Spectrum of photons produced with energy $E_{\gamma,j}$ due to the cooling of a single electron of energy $E'_{e,i}$, as defined in Eq. (2.14).

	$\overline{\mathsf{T}}_{\rm ICS}[E'_{e,i}, E_{\gamma,j}]$	The same as $\overline{T}_{\rm ICS,0}$, but with the pre-scattering spectrum of upscattered CMB photons subtracted out, as defined in Eq. (2.16).
	$\overline{\mathsf{T}}_{e}[E'_{e,i}, E_{e,j}]$	Low-energy electron spectrum produced due to the cooling of a single electron of energy $E'_{e,i}$, as defined in Eq. (2.17).
Low-Energy Deposition	$f_c(z, \mathbf{x})$	Ratio of deposited to injected energy, as a function of redshift z and the ionization level $\mathbf{x}$, into channels $c \in$ {'H ion', 'He ion', 'exc', 'heat', 'cont}, as defined in Eq. (2.4).
	$\left(\frac{dE^\alpha}{dV\,dt}\right)_c$	Energy deposited per volume and time by low-energy photons ($\alpha = \gamma$) or electrons ($\alpha = e$) into channel c.
	$E_c^{\rm high}[y]$	Total amount of high-energy deposition into channels $c \in$ {'ion', 'exc', 'heat'} at log-redshift y.

Table A.5: A list of the important definitions used in `DarkHistory`. In this table, all spectra are discretized spectra as described in Sec. 2.3.2. Spectra without overlines are normalized so that their entries contain number (per baryon) of particles produced in a redshift step. A primed energy denotes the energy of an injected particle, and by energy we mean kinetic energy. In this table, when we refer to electrons we will always mean electrons plus positrons.

Appendix B

Lyman-α Constraints

B.1 Terms in the Evolution Equations

In this appendix we provide explicit expressions for the terms appearing in Eq. (3.1) and (3.3) and explicitly write down the helium ionization evolution equations. Starting with the non-DM temperature sources,

$$\begin{aligned} \dot{T}_{\rm adia} &= -2HT_{\rm m}\,, \\ \dot{T}_{\rm C} &= -\Gamma_C(T_{\rm CMB} - T_{\rm m})\,, \end{aligned} \tag{B.1.1}$$

where H is the Hubble parameter, $T_{\rm CMB}$ is the temperature of the CMB, and Γ_C is the Compton cooling rate

$$\Gamma_C = \frac{x_{\rm e}}{1+\chi+x_{\rm e}} \frac{8\sigma_T a_r T_{\rm CMB}^4}{3m_{\rm e}}\,. \tag{B.1.2}$$

Here, σ_T is the Thomson cross section, a_r is the radiation constant, and $m_{\rm e}$ is the electron mass. The DM temperature source is given by

$$\dot{T}_{\rm DM} = \frac{2f_{\rm heat}(z,\mathbf{x})}{3(1+\chi+x_{\rm e})n_{\rm H}} \left(\frac{dE}{dV\,dt}\right)^{\rm inj} \tag{B.1.3}$$

where $f_{\rm heat}(z,\mathbf{x})$ is the deposition efficiency fraction into heating of the IGM as a function of redshift z and a vector, $\mathbf{x}$, storing the ionization levels of HI and HeII, which is computed by `DarkHistory`. $\left(\frac{dE}{dV\,dt}\right)^{\rm inj}$ is the total amount of energy injected per volume per time through DM decays or annihilations. Finally, $\dot{T}_{\rm atom}$ is given by the sum of the recombination, collisional ionization, collisional excitation, and bremsstrahlung cooling rate fitting functions given in Appendix B4 of Ref. [92]. In Fig. B-1, we plot these rates for a model of DM decaying to photons with a lifetime

of $2 \times 10^{22}\,\mathrm{s}$ and $m_\chi = 800\,\mathrm{MeV}$. We set x_e^Pl to Planck's latest FlexKnot ionization history and use the 'conservative' treatment for the photoheating term. Fig. B-1 demonstrates that in a hot and reionized universe, cooling processes that were once negligible become important and possibly dominant.

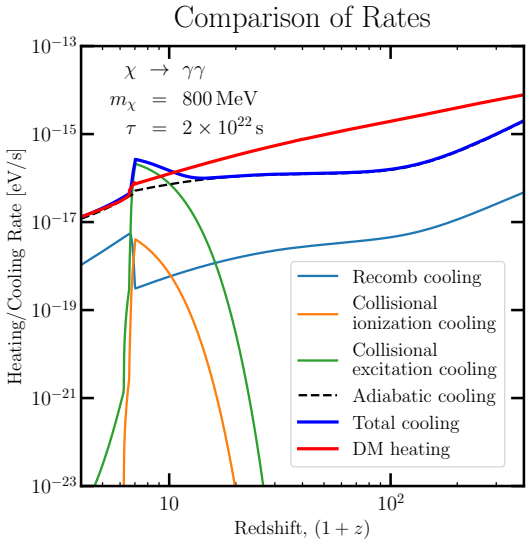


Figure B-1: The absolute value of the atomic cooling rates included in $\dot{T}_\mathrm{atom}$, the adiabatic cooling rate, and the DM heating rate. We assume a model of DM decaying to photons with a lifetime of $2 \times 10^{22}\,\mathrm{s}$ and $m_\chi = 800\,\mathrm{MeV}$. The blue line corresponds to the sum of all cooling rates while the red corresponds to the DM heating rate, the only source of heating in the 'conservative' treatment.

Moving on to the ionization equations, we write down the helium version of Eq. (3.1),

$$\begin{aligned} \dot{x}_\mathrm{HeII} &= \dot{x}_\mathrm{HeII}^\mathrm{atom} + \dot{x}_\mathrm{HeII}^\mathrm{DM} + \dot{x}_\mathrm{HeII}^\star\,, \\ x_\mathrm{HeIII} &= 0\,, \end{aligned} \tag{B.1.4}$$

where $x_\mathrm{HeII} \equiv n_\mathrm{HeII}/n_\mathrm{H}$ is the density of singly-ionized helium atoms in the IGM normalized to the density of hydrogen atoms, and x_HeIII is defined similarly. As explained above, the second of these two equations reflects the fact that there are negligibly few fully ionized helium atoms in the IGM over the redshifts under consideration in our analysis. Therefore we only need to keep track of the relative levels of HeI and HeII using the first equation. Similarly to the $\dot{x}_\mathrm{HII}^\star$ term, we have engineered the astrophysical reionization source term to turn off for $z > z^\star$ and produce a helium ionization

curve that is equal to $\frac{\chi}{1+\chi} x_{\rm e}^{\rm Pl}(z)$ for $z < z^\star$. In other words,

$$\begin{cases} \dot{x}_{\rm HeII} = \dot{x}_{\rm HeII}^{\rm atom} + \dot{x}_{\rm HeII}^{\rm DM}\,, & z > z^\star\,, \\ x_{\rm HeII} = \dfrac{\chi}{1+\chi} x_{\rm e}^{\rm Pl}(z)\,, & z < z^\star\,. \end{cases} \tag{B.1.5}$$

Notice that we do not need to know the explicit form of $\dot{x}_{\rm HeII}^\star$ in contrast to $\dot{x}_{\rm HII}^\star$, which we need to compute to evaluate $\dot{T}^\star$ in Eq. (3.4). Due to this simplified treatment, $x_{\rm HeII}$ can be discontinuous at $z^\star$; we have tested alternative prescriptions and found negligible effects on our constraints.

The atomic sources contain a contribution from photoionization and a contribution from recombination. For $z > z^\star$, we assume a case-B scenario [38, 71, 73],

$$\begin{aligned} \dot{x}_{\rm HII}^{\rm atom} &= 4\,\mathcal{C}_{\rm H}\,\left[(1-x_{\rm HII})\,\beta_{\rm H}^B e^{-E_{\rm H}/T_{\rm CMB}} - n_{\rm H}\,x_{\rm e}\,x_{\rm HII}\,\alpha_{\rm H}^B\right] \\ \dot{x}_{\rm HeII}^{\rm atom} &= 4\sum_s \mathcal{C}_{{\rm HeII},s}\,\Big[g_s(\chi - x_{\rm HeI})\,\beta_{{\rm HeI},s}^B e^{-E_{{\rm HeI},s}/T_{\rm CMB}} \\ &\qquad - n_{\rm H}\,x_{\rm e}\,x_{\rm HeII}\,\alpha_{{\rm HeI},s}^B\Big]\,, \end{aligned} \tag{B.1.6}$$

where E_i, β_i^B, α_i^B, and $\mathcal{C}_i$ are, respectively, the binding energy, case-B photoionization coefficient (including the gaussian fudge factor used in RECFAST v1.5.2 [38, 71]), case-B recombination coefficient, and Peebles $\mathcal{C}_i$ factor for species $i \in \{\text{H; HeI, singlet; HeI, triplet}\}$ [37]. Notice, there is a sum over both spin states of the two electrons in the excited HeI atom. For the spin singlet, $g_1 = 1$ and $E_{\rm HeI,1} = 20.616\,\text{eV}$, while for the spin triplet state $g_3 = 3$ and $E_{\rm HeI,3} = 19.820\,\text{eV}$.

When $z < z^\star$, we assume a case-A scenario, which is applicable during reionization [92]:

$$\dot{x}_{\rm HII}^{\rm atom} = n_{\rm H}\,(1-x_{\rm HII})\,x_{\rm e}\,\Gamma_{\rm eHI} - n_{\rm H}\,x_{\rm e}\,x_{\rm HII}\,\alpha_{\rm HII}^A\,. \tag{B.1.7}$$

The collisional ionization rate, $\Gamma_{\rm eHI}$, and case-A recombination coefficient, $\alpha_{\rm HII}^A$, can be found in Ref. [92]. Notice that the case-A photoionization term from CMB photons is not included because it is exponentially suppressed at these low redshifts, and that the photoionization term from astrophysical reionization sources is already accounted for in $\dot{x}_{\rm HII}^\star$. Additionally, we do not need the analogous HeII version of Eq. (B.1.7) since at these redshifts we have assumed $x_{\rm HeII} = \chi\,x_{\rm HII}$.

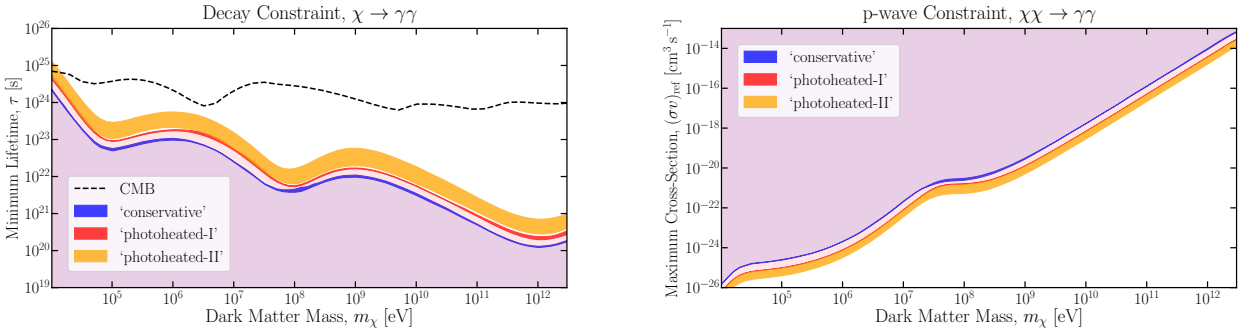


Figure B-2: Constraints for decay (left) or p-wave annihilation (right) to $\gamma\gamma$ pairs with $v_{\rm ref} = 100\,\mathrm{km\,s^{-1}}$. We show our constraints using the 'conservative' (blue band), 'photoheated-I' (red band), and 'photoheated-II' (orange band) treatments. We also include the CMB constraint for decay [59] (dashed-black). Telescope constraints [260–264] are many orders of magnitude stronger than ours, and are not shown for clarity.

The DM ionization source terms are given by

$$\begin{aligned}
\dot{x}^{\rm DM}_{\rm HII} &= \left[\frac{f_{\rm H\ ion}(z,\mathbf{x})}{E_{\rm H} n_{\rm H}} + \frac{(1-\mathcal{C}_{\rm H}) f_{\rm exc}(z,\mathbf{x})}{0.75 E_{\rm H} n_{\rm H}}\right]\left(\frac{dE}{dV\,dt}\right)^{\rm inj}, \\
\dot{x}^{\rm DM}_{\rm HeII} &= \frac{f_{\rm He\ ion}(z,\mathbf{x})}{E_{\rm HeI} n_{\rm He}}\left(\frac{dE}{dV\,dt}\right)^{\rm inj}, \\
\dot{x}^{\rm DM}_{\rm HeIII} &= 0\,,
\end{aligned} \tag{B.1.8}$$

where $f_{\rm H\ ion}(z,\mathbf{x})$, $f_{\rm He\ ion}(z,\mathbf{x})$, $f_{\rm exc}(z,\mathbf{x})$ are the deposition efficiency fractions into hydrogen ionization, single neutral helium ionization, and hydrogen excitation calculated by `DarkHistory`.

B.2 Other Final States

In this appendix we provide constraints for DM decay and p-wave annihilation to $\gamma\gamma$, $\mu^+\mu^-$, $\pi^+\pi^-$, and $\pi^0\pi^0$.

B.2.1 Photons

Fig. B-2 shows decay and annihilation constraints for $\gamma\gamma$ final states using the 'conservative' (blue), 'photoheated-I' (red), or 'photoheated-II' treatments (orange). As in the main text, the p-wave annihilation cross-section is defined by $\sigma v = (\sigma v)_{\rm ref} \times (v/v_{\rm ref})^2$ with $v_{\rm ref} = 100\,\mathrm{km\,s^{-1}}$ and we use the NFW boost factor for p-wave annihilation calculated in Ref. [130], which accounts for enhanced annihilation due to increased DM density and dispersion velocity in halos. Just as in the main text, the darkly shaded blue, red, and orange bands show the variation of our constraints as

we vary $x_e^{\rm Pl}$ in Eq. (3.2) over the 95% confidence region of Planck's FlexKnot and Tanh late-time ionization curves. As before, the 'conservative' and 'photoheated-I' bands are narrow, demonstrating an insensitivity to the precise form of the reionization curve, while the 'photoheated-II' curve is broader for the reasons discussed in the main text.

The photon final state constraints are less competitive with existing constraints than are the e^+e^- constraints. For example, CMB constraints [59] are stronger for all masses in the decay channel. Additionally, telescope constraints (see e.g. Refs. [260–264]) are many orders of magnitude stronger than ours since telescopes can search directly for the produced photons, in contrast to our temperature constraints that indirectly look for the effects that these photons have on the IGM.

Our $\gamma\gamma$ constraints are weaker than our e^+e^- constraints because the photoionization probability is small (equivalently, the path length is long) for the redshifts and photon energies of interest. In contrast, electrons can efficiently heat the gas either through direct Coulomb interactions (for non-relativistic and mildly relativistic electrons) or through inverse Compton scattering that produces efficiently-ionizing photons (for higher-energy electrons).

B.2.2 Muons and Pions

While we could also consider any other Standard Model particle final state, our results from the previous section and the main text indicate that our constraints are most competitive at masses below 10 GeV. Therefore, we consider some of the most important final states that are available to sub-GeV DM besides those already considered: muons, charged pions, and neutral pions. To compute the final spectra of e^+e^- and γ produced by the decay of pions or muons, we use the PPPC4DMID for DM masses above 10 GeV.

For DM masses below 10 GeV, we follow the method described in Ref. [160]. We start with the spectrum of electrons in the muon rest frame, which is given by

$$\frac{dN_e^{\mu\to e\nu\bar{\nu}}}{dE_e} = \frac{4\sqrt{\xi^2-4\varrho^2}}{m_\mu}[\xi(3-2\xi)+\varrho^2(3\xi-4)] \tag{B.2.1}$$

between energies of $m_e < E_e < (m_\mu^2+m_e^2)/(2m_\mu)$ and is otherwise zero. In this equation, $\xi = 2E_e/m_\mu$ and $\varrho = m_e/m_\mu$. For a particle A with mass m_A decaying with some spectrum dN/dE' in its rest frame, the spectrum dN/dE in an arbitrary frame

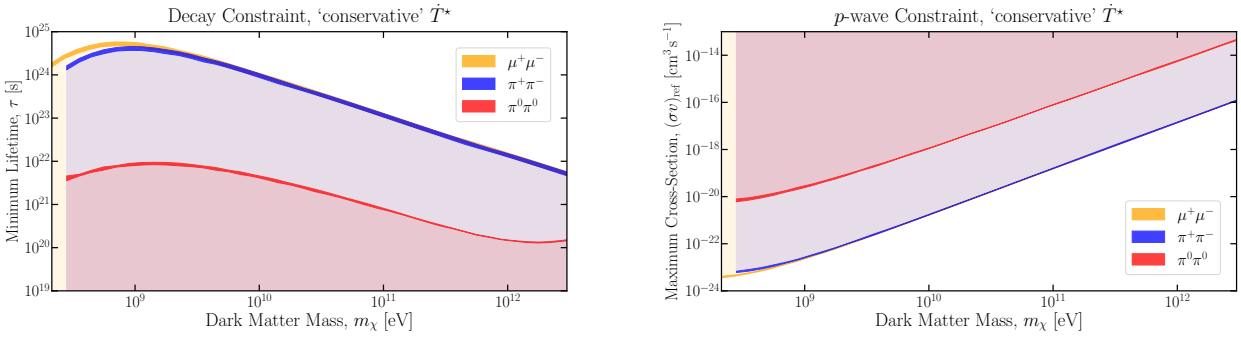


Figure B-3: Constraints for decay (left) or p-wave annihilation (right) to $\mu^+\mu^-$ (yellow), $\pi^+\pi^-$ (blue), and $\pi^0\pi^0$ (red) pairs with $v_{\rm ref} = 100\,{\rm km\,s^{-1}}$. We show our constraints only using the 'conservative' treatments.

where A has energy E_A is given by

$$\frac{dN}{dE} = \frac{1}{2\beta\gamma}\int_{E'_{\rm min}}^{E'_{\rm max}} \frac{dE'}{p'}\frac{dN}{dE'}, \tag{B.2.2}$$

where $\gamma = E_A/m_A$ is the Lorentz factor, $\beta = \sqrt{1-\gamma^{-2}}$, and $E'_{\rm max/min} = \gamma(E \pm \beta p)$. In the case of decay to muons, we can use this equation to boost from the muon frame to the dark matter frame, where the muon has energy m_χ for annihilations or $m_\chi/2$ for decays. For decay to pions, we first boost to the pion rest frame where the muon has energy $(m_\pi^2 + m_\mu^2)/(2m_\pi)$, and then the dark matter frame where the pion similarly has energy m_χ for annihilations or $m_\chi/2$ for decays.

We plot our constraints in Fig. B-3. $\mu^+\mu^-$ and $\pi^+\pi^-$ ultimately decay to e^+e^- and neutrinos, meaning that these constraints are comparable to the e^+e^- constraints, though somewhat weaker because the produced electrons share at most an $\mathcal{O}(1)$ fraction of the total DM injected energy with the other neutrinos. $\pi^0\pi^0$ decays almost exclusively to 4γ, so the photons carry half the energy as compared to photons that result from $\chi \to \gamma\gamma$ decays. Thus, the pion constraints look exactly like the $\gamma\gamma$ constraints shifted by a factor of 2 to the left.

B.3 Cross Checks

Here, we provide cross checks to validate the assumptions we made in our analysis. First, we will validate maintaining $x_{\rm HeIII} = 0$ after H and HeI reionization despite DM injecting HeII ionizing photons. Second, we will check that our p-wave constraints are insensitive to the uncertainty in the halo boost factor coming from the halo profile. Finally, we will validate our use of ionization histories that feature significant

ionization levels prior to reionization, by checking that they do not violate constraints on the total $z < 50$ optical depth.

B.3.1 Treatment of HeIII

In calculating the constraints shown in Fig. 3-2, we assume that there is no ionization of HeII to HeIII – i.e. $x_{\text{HeIII}} = 0$ – consistent with the assumptions that went into the making of `DarkHistory`'s transfer functions. We still account for energy deposition through ionization of HeII by allowing photons with energies $E_\gamma > 54.4\,\text{eV}$ to be absorbed by HeII atoms, thus producing electrons of energy $E_\gamma - 54.4\,\text{eV}$ that thermalize with the IGM. This is not entirely self-consistent because these photoionization events would gradually increase the fraction of HeII atoms as they convert into HeIII atoms. Having fewer HeII atoms could then affect our constraints by decreasing the heating deposition fraction, since fewer photoionized electrons could be produced and thermalize with the IGM.

We test our sensitivity of our constraints to this approximation by adding a new $\dot{x}_{\text{HeIII}}$ source term and accounting for recombination photons once HI/HeI reionization is complete. We restrict this correction to after HI/HeI reionization because it is expected to make the biggest difference in the heating rate then, since HeII atoms are the only possible source of photoionized electrons at this point, and because the temperature data we use are primarily in this redshift range.

To apply our correction, we first modify Eq. (B.1.4) to track the fully ionized helium fraction,

$$\begin{aligned}\dot{x}_{\text{HeIII}} =& \frac{f_{\text{He ion}}(z, \mathbf{x})}{4 E_{\text{H}} n_{\text{He}}} \left(\frac{dE}{dV\, dt}\right)^{\text{inj}} \\ &+ n_{\text{H}} \left(\chi - x_{\text{HeIII}}\right) x_{\text{e}}\, \Gamma_{\text{eHeII}} \\ &- n_{\text{H}}\, x_{\text{e}}\, x_{\text{HeIII}}\, \alpha^{A}_{\text{HeIII}}\,, \end{aligned} \tag{B.3.1}$$

where the deposition fraction $f_{\text{He ion}}(z, \mathbf{x})$ computed by `DarkHistory` accounts for the total energy deposited into HeII ionization and Γ_{eHeII} is the collisional ionization rate of HeII [92]. We then compute the fraction of HeIII atoms that recombines within a timestep of the code, Δt,

$$f^{\text{HeIII}}_{\text{recomb}} = 1 - e^{-\alpha^{\text{A}}_{\text{HeIII}} x_{\text{HeIII}} n_e \Delta t}\,. \tag{B.3.2}$$

We convert this fraction to the number of 54.4 eV photons per baryon emitted by HeIII atoms in this time step

$$N_{\rm recomb}^{\rm HeIII} = f_{\rm recomb}^{\rm HeIII} n_{\rm HeIII}/n_{\rm B}\,, \tag{B.3.3}$$

then add these photons to `DarkHistory`'s low energy photon spectrum within that time step.

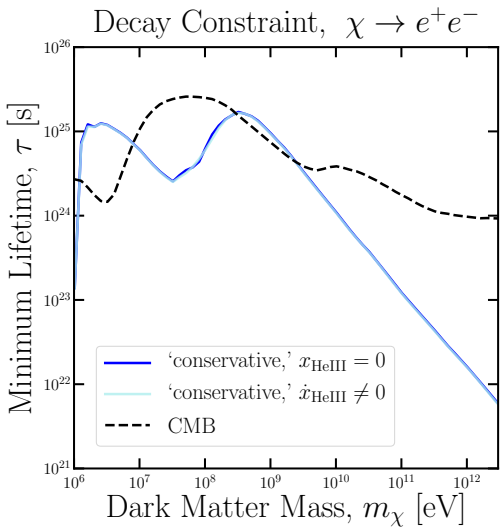


Figure B-4: Comparison of 'conservative' constraints for decay to electrons, including and not including the effects of HeII ionization after HI/HeI reionization. Both constraints were generated assuming Planck's earliest Tanh reionization history.

Fig. B-4 shows a comparison of constraints for dark matter decaying to electrons, where the two curves either allow for a non-zero HeIII fraction (light blue) or do not (blue). The difference in constraints is always less than 1%, and so is not an important source of error in our analysis.

B.3.2 Boost factor for p-wave annihilation

The boost factor due to enhanced density and velocity dispersion in halos depends on the halo profile chosen. However, in Ref. [130], the boost factor was found to be highly robust to this choice, since the main contribution to the boost factor comes from the largest halos, which are fully resolved in N-body simulations. We find that the difference in our constraints made by using the Einasto p-wave boost factor rather than the NFW p-wave boost factor from Ref. [130] is negligibly small, resulting in a modification of no more than 0.5% to our constraints. Notice that the two boost factors only vary over the halo mass function and halo profile, and do not include

uncertainties due to mergers, asphericity, etc.

B.3.3 Optical depth

In this appendix, we discuss the relation between temperature and ionization constraints, focusing in particular on the complementarity of these constraints. One might worry that scenarios excluded by excess heating of the IGM are strictly a subset of those excluded by the ionization history. In some cases, the DM contribution to the optical depth τ before reionization, combined with one of the Planck reionization models, can exceed the Planck limit on τ. DM energy injection starts to increase the ionization fraction and temperature immediately after recombination, and so our computed ionization histories will always be in excess of Planck's reionization curves at early enough redshifts.

To some extent, these worries have already been addressed by the fact that the temperature constraints can sometimes be stronger than the CMB power spectrum constraints for DM decays as derived in Refs. [59, 67], which account for the effect of excess ionization on the full multipole structure of the CMB power spectrum. For simplicity, however, we would like to compare the IGM temperature constraints derived in the main body with limits on the ionization history coming simply from the Planck upper limit on τ.

Given an ionization history $x_{\rm e}(z)$, the optical depth is

$$\tau = n_{\rm H,0}\sigma_T \int_0^{z_{\rm max}} dz\, x_{\rm e}(z) \frac{(1+z)^2}{H(z)}, \tag{B.3.4}$$

where σ_T is the Thomson cross-section and $z_{\rm max}$ is set to 50, as is done in Ref. [121]. The 68% upper bound on the optical depth from Planck assuming a tanh function reionization history is $\tau = 0.0549$ [121]. To derive a constraint, we compute an ionization history in the presence of DM energy injection and exclude it if the history's optical depth is greater than 0.0549.

Clearly, these optical depth constraints will be highly sensitive to the reionization curve we choose. For example, if we were to use the earliest Tanh reionization curve that already saturates the optical depth bound we would rule out all DM models since they all increase τ. On the other hand, we saw that our temperature constraints were very weakly dependent on the choice of reionization curve. For a fair comparison, we choose a reionization history with the smallest optical depth. While we could choose

the latest Tanh reionization curve, we instead follow the instantaneous reionization method described in Ref. [130] so that we can compare to older optical depth constraints. We will assume an instantaneous HI/HeI reionization at $z = 6$, then an instantaneous HeII reionization at $z = 3$, but no other sources of reionization other than DM for $z > 6$. The optical depth contributed by the range $0 < z < 6$ is 0.384,[1] meaning that DM models that contribute more than $\delta\tau = 0.0165$ to the optical depth within the range $6 < z < 50$ will be ruled out.

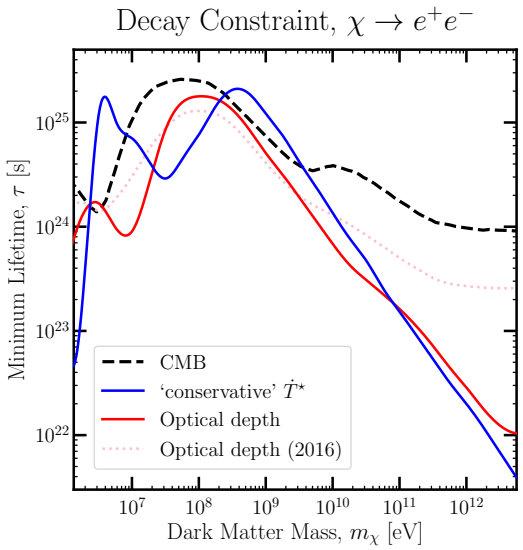


Figure B-5: Constraints obtained from the IGM temperature (red) and optical depth (blue, solid), as well as previous bounds derived in Ref. [130] from the optical depth (pink, dashed). The black line shows the constraints derived using a principal component analysis of CMB data [59].

Fig. B-5 shows a comparison of the optical depth constraint in blue to the IGM temperature constraint in red for a model of DM decay to e^+e^-. For reference, we show the CMB power spectrum constraint in dashed-black [59], and also show an older optical depth constraint made with Planck intermediate results [102] in dotted-pink [130]. This latter constraint calculated $\tau = 0.058 \pm 0.012$ by following the method in Ref. [129] and integrating over the excess ionization fraction over the standard three-level atom result up to recombination. We see that across most of the mass range, the two methods of constraining dark matter lifetimes are comparable, but there is a large range of DM masses over which the temperature constraints do better than the optical depth limits. Had we considered a model of p-wave annihilation to e^+e^- instead, we would find that the heating constraints are generically stronger than

[1] This is nearly equivalent to using the earliest Tanh reionization history, which has an optical depth contribution of 0.383 to the same redshift range.

those from the optical depth since, as compared to the decay case, the power from p-wave annihilation is enhanced at low redshifts where temperature limits are most sensitive. To summarize, since the IGM temperature constraints are insensitive to the exact ionization history during reionization, they probe a different aspect of energy injection from DM that is distinct from ionization-based constraints like optical depth and the CMB power spectrum.

B.4 Test Statistics

In this appendix, we derive the distribution of the modified χ^2-like test statistic (TS) that we use in conjunction with the 'conservative' treatment of the $\dot{T}^\star$ photoheating term (i.e. $\dot{T}^\star = 0$). We are working in a frequentist framework, so we wish to evaluate the probability distribution for the TS defined in Eq. (3.5), when assuming a certain pattern of heating due to DM energy injection. We can then say that this scenario is excluded if the TS observed in the real data is sufficiently unlikely. We make the assumption that the data points in different redshift bins are independent and Gaussian distributed.

Suppose that there are N redshift bins, and in the ith bin the temperature value $T_{i,\mathrm{data}}$ is drawn from a Gaussian distribution with mean $T_{i,\mathrm{pred}}$ and standard deviation $\sigma_{i,\mathrm{data}}$. There is then a 50% chance that $T_{i,\mathrm{data}} > T_{i,\mathrm{pred}}$, so the probability distribution for TS_i as defined in Eq. (3.5) is:

$$\begin{aligned} f(\mathrm{TS}_i|T_{i,\mathrm{pred}}) &= \frac{1}{2}\delta(\mathrm{TS}_i) + P(T_{i,\mathrm{data}})\frac{d(T_{i,\mathrm{data}})}{d(\mathrm{TS}_i)} \\ &= \frac{1}{2}\delta(\mathrm{TS}_i) + \frac{1}{2\sqrt{2\pi}}\mathrm{TS}_i^{-1/2}\exp(-\mathrm{TS}_i/2)\,, \end{aligned} \tag{B.4.1}$$

where δ is the Dirac delta function. Let the χ^2 probability distribution function with j degrees of freedom be denoted by $f_{\chi^2}(\mathrm{TS}; j)$. Then one can rewrite this distribution in terms of the χ^2 distribution with one degree of freedom:

$$f(\mathrm{TS}_i|T_{i,\mathrm{pred}}) = \frac{1}{2}\delta(\mathrm{TS}_i) + \frac{1}{2}f_{\chi^2}(\mathrm{TS}_i; 1)\,. \tag{B.4.2}$$

Now we want to know the distribution for the total TS value from combining the bins (assuming uncorrelated data), $\mathrm{TS} \equiv \sum_i \mathrm{TS}_i$. We can write:

$$
\begin{aligned}
f(\mathrm{TS}|\{T_{i,\mathrm{pred}}\}) &= \left[\prod_{i=1}^{N}\int_0^\infty d\mathrm{TS}_i\, f(\mathrm{TS}_i|T_{i,\mathrm{pred}})\right] \\
&\quad\times \delta(\mathrm{TS}-\sum_{j=1}^{N}\mathrm{TS}_j) \\
&= \left[\prod_{i=1}^{N}\int_0^\infty d\mathrm{TS}_i\, \left[\delta(\mathrm{TS}_i)+f_{\chi^2}(\mathrm{TS}_i;1)\right]\right] \\
&\quad\times \frac{1}{2^N}\delta(\mathrm{TS}-\sum_{j=1}^{N}\mathrm{TS}_j)\,. \qquad \text{(B.4.3)}
\end{aligned}
$$

Expanding the product inside the integrals gives a sum of terms, which each consist of a product of delta-functions and f_{χ^2} functions. For a term with n delta-functions, the delta-functions can be used to do n of the integrals, resulting in a term of the form:

$$
\left[\prod_{j=n+1}^{N}\int_0^\infty d\mathrm{TS}_{i_j} f_{\chi^2}(\mathrm{TS}_{i_j};1)\right]\delta(\mathrm{TS}-\sum_{k=n+1}^{N}\mathrm{TS}_{i_k})\,, \qquad \text{(B.4.4)}
$$

where n can take on values from 0 to N, and $i_{n+1}, i_{n+2}, \cdots, i_N$ are a collection of indices between 1 and N for this particular term. However, this is exactly the standard probability distribution function for the sum of the χ^2 test statistic over $N-n$ bins, so we can write it as $f_{\chi^2}(\mathrm{TS};N-n)$.

The coefficient of each such term will be the number of ways of choosing which n indices correspond to δ-function terms as opposed to the $N-n$ indices labeling $f_{\chi^2}(\mathrm{TS}_i;1)$ contributions – which is the binomial coefficient $\binom{N}{n}$. Since $\binom{N}{n}=\binom{N}{N-n}$, we can write:

$$
f(\mathrm{TS}|\{T_{i,\mathrm{pred}}\}) = \frac{1}{2^N}\sum_{n=0}^{N}\binom{N}{n} f_{\chi^2}(\mathrm{TS};n)\,. \qquad \text{(B.4.5)}
$$

This completes the proof of Eq. (3.6).

Note that this expression integrates correctly to 1, as $\int d\mathrm{TS} f_{\chi^2}(\mathrm{TS};n)=1$ and $\sum_{n=0}^{N}\binom{N}{n}=2^N$. The largest binomial coefficients $\binom{N}{n}$ will occur for $n\approx N/2$, and so we may approximate the distribution as a χ^2 distribution with $N/2$ degrees of freedom. However, for the actual constraints in the main text we use the full distribution, rather than this approximation.

We can also understand this distribution by thinking of TS as the standard χ^2 test statistic, in the presence of a model for the data where each redshift bin contains an irreducible (dark matter) contribution plus a non-negative but otherwise arbitrary

increase to the temperature from photoheating. If we profile over the nuisance parameters describing the unknown astrophysics, we see that the minimum χ^2 will be attained when:

- in bins where the irreducible contribution from dark matter already exceeds the measured temperature, extra contributions from photoheating are set to zero; the contribution to the TS is the usual χ^2 computed using the irreducible model and the data,
- in bins where the irreducible contribution from dark matter does not exceed the measured temperature, the additional photoheating contribution is chosen to precisely match the data, and consequently the contribution to the TS is zero.

This is exactly the prescription for our modified TS, Eq. (3.5).

Because this is a standard χ^2 test, just with a flexible background model, the probability distribution for the TS should follow that of a χ^2 distribution with $N-m$ degrees of freedom, where m is the number of floated parameters in the fit. The number of floated parameters for this signal model is the number of bins where the data is greater than the irreducible model, which can vary from 0 to N; thus the full probability distribution is obtained as a linear combination of χ^2 distributions with degrees of freedom varying from 0 to N.

Appendix C

Thermal Squeezeout

C.1 Thermodynamics of a First-order Phase Transition

In this Appendix we collect results that are helpful for understanding the dynamics of a first-order phase transition. We also describe a numerical simulation of the confining phase transition from the main text, which we perform to fix the initial pocket radius and its contraction rate. The code we used to perform this simulation can be found at https://github.com/gridgway/ConfiningPT_Bubbles.

C.1.1 Standard thermodynamics

The defining characteristic of a first-order phase transition is that a first derivative of the free energy, or the free energy density f, is discontinuous at a critical point. In our case, the entropy density, $s = -\frac{\partial f}{\partial T}$, is discontinuous, corresponding to a non-zero latent heat release due to phase conversion. In contrast, f itself is continuous, meaning that the free energy of the two phases are the same at the critical point, $f_{\text{deconf}} = f_{\text{conf}}$.

To better understand the latent heat, first notice that in the absence of chemical potentials the critical point is specified by a single parameter, the critical temperature T_c, and the free energy density is minus the pressure, $f = -p$. Since the free energy of both phases are the same at the critical point, their pressures are the same. Using the Euler relation, $\rho = Ts - p$, where ρ is the energy density, one can take the difference in energy densities of the two phases at the critical temperature to find

$$\Delta\rho = T_c \Delta s \equiv l\,, \tag{C.1.1}$$

which defines the latent heat density. Whereas the *free* energy density is continuous, the energy density is not. Another important quantity that characterizes first-order phase transitions is the surface tension, σ. The surface tension is the energy cost per unit area of an interface separating the two phases.

The latent heat density and surface tension are calculable via lattice simulations. We assume that at temperatures below the quark mass, the thermodynamics of the system are insensitive to the heavy quark field and we need only consider lattice simulations of pure $SU(3)$ Yang-Mills theory. In particular, the authors of [48] calculate a latent heat density and surface tension in the infinite volume limit (see their tables 7 and 15)

$$\begin{aligned} l &= 1.413\, T_c^4\,, \\ \sigma &= .02\, T_c^3\,. \end{aligned} \tag{C.1.2}$$

These quantities have been computed elsewhere [49–52], and we find that the uncertainties on l and σ among these different lattice calculations are not large enough to qualitatively change our results.

As the universe expands, our thermal system, still in the deconfined phase, supercools.[1] Once the system is supercooled, the free energy of the confined phase is lower than the free energy of the deconfined phase, meaning that it is energetically favorable for a phase transition to take place. For any non-zero amount of supercooling, there now exists a critical radius, R_c, at which the energy cost of increasing a spherical bubble's surface area is exactly compensated by the free-energy decrease due to phase conversion,

$$\begin{aligned} \left.\frac{\partial F}{\partial R}\right|_{R_c} &= \left.\frac{\partial}{\partial R}\left(4\pi R^2\sigma - \frac{4\pi}{3}R^3\Delta f\right)\right|_{R_c} = 0 \\ \implies R_c &= \frac{2\sigma}{\Delta f}\,, \end{aligned} \tag{C.1.3}$$

where Δf is the confined phase free energy density minus that of the deconfined phase, and so is positive.

To relate Δf to the latent heat density, first recall that at T_c the entropy difference is given by $\Delta s = l/T_c$. If we then use the thermodynamic relation $\frac{\partial f}{\partial T} = -s$ we find $\left.\frac{\partial \Delta f}{\partial T}\right|_{T_c} = -l/T_c$. If we assume small supercooling we can use a Taylor expansion,

[1] Much of the following discussion can be found in Refs. [265] §162 and [266] §99-100.

which at leading order gives

$$\begin{aligned}\Delta f &= \frac{\partial \Delta f}{\partial T}\,(T - T_c) \\ &= l\,\frac{(T_c - T)}{T_c}\,.\end{aligned} \tag{C.1.4}$$

We then find

$$R_c = \frac{2\sigma\, T_c}{l\,(T_c - T)}\,. \tag{C.1.5}$$

The total free energy of a bubble at the critical radius is

$$\begin{aligned}F_c &= 4\pi R_c^2 \sigma - \frac{4\pi}{3} R_c^3 \Delta f \\ &= \frac{16\pi}{3}\frac{\sigma^3}{\Delta f^2} \\ &= \frac{16\pi}{3}\left(\frac{\sigma}{T_c^3}\right)^3 \left(\frac{l}{T_c^4}\right)^{-2} \frac{T_c^3}{(T_c - T)^2}.\end{aligned} \tag{C.1.6}$$

Thermal fluctuations will randomly convert regions of varying shapes and sizes from one phase to the other. When the system is supercooled, there is a chance that a converted region will be large enough that it expands rather than contracts. The probability per unit time and volume of converting a region with free energy F is determined primarily by the Boltzmann factor $e^{-F/T}$. Using dimensional analysis we have [267],

$$\Gamma = A T^4 e^{-\frac{F}{T_c}}\,, \tag{C.1.7}$$

where A is assumed to be some $\mathcal{O}(1)$ number that is roughly constant with respect to temperature and determined by the microscopic theory. Provided A is indeed an $\mathcal{O}(1)$ number, we find that the exponential is by far the more important factor for determining the behavior of the phase transition, so we set A to 1 without qualitatively changing our results. For an alternative though similar expression for Γ, See [266] §99.

Let us make the simplifying assumption that all bubbles can be approximated as spherical. Then bubbles that nucleate with radii below R_c are ephemeral, quickly shrinking due to surface tension, while bubbles with radii well above R_c are exponentially less likely to nucleate than critical bubbles by Eq. (C.1.7). Then to a good approximation, we can assume that only bubbles at the critical radius nucleate. Com-

bining Eqs. (C.1.6) and (C.1.7), we find the nucleation rate of these critical bubbles is

$$\Gamma = AT_c^4 e^{-\frac{\kappa T_c^2}{(T_c-T)^2}}\,, \tag{C.1.8}$$

$$\kappa = \frac{16\pi}{3}\left(\frac{\sigma}{T_c^3}\right)^3\left(\frac{l}{T_c^4}\right)^{-2}\,,$$

$$\sim 7\times 10^{-5}\,, \tag{C.1.9}$$

where in the last line we have used the lattice results from Eq. (C.1.2). Because the latent heat density is an order one number in units of T_c while the surface tension is a small number in units of T_c, κ turns out to be a very small number (which Ref. [48] indicates is generically true for $SU(N)$ theories). As a result, very little supercooling is required before bubble nucleation becomes efficient.

C.1.2 First half of the phase transition: bubble growth

After bubble nucleation begins, the phase transition proceeds via the nucleation of new bubbles and the expansion of old bubbles. To keep track of the progress of the phase transition, we define the fraction of the universe that is in the confined phase [267],

$$x(t) = \int_{t_c}^{t} dt'\Gamma(t')\frac{4\pi}{3}R^3(t,t')(1-x(t'))\,, \tag{C.1.10}$$

where t_c is the time at which the universe first reaches the critical temperature and $R(t,t')$ is the radius at time t of a bubble nucleated at time t'. Applying a time derivative yields

$$\dot{x}(t) = \Gamma(t)(1-x(t))\frac{4\pi}{3}R_c^3(t) + \int_{t_c}^{t} dt'\Gamma(t')4\pi R^2(t,t')\dot{R}(t,t')(1-x(t'))\,, \tag{C.1.11}$$

where the first term corresponds to the nucleation of new bubbles while the second corresponds to the expansion of old bubbles.

The temperature evolution of the early universe plasma includes the usual adiabatic cooling term due to Hubble expansion, but now it also includes a new heating term due to the steady release of latent heat as the deconfined phase converts to the confined phase. By Eq. (C.1.1), converting a fraction dx from the deconfined to the confined phase releases $d\rho = l dx$ energy per unit volume. The released energy

is absorbed in each phase with a temperature increase determined by each phase's respective specific heat, $d\rho/dT$. For now, we will focus on the large scale average temperature so that we do not have to deal with small scale temperature gradients between points near and far from the sites of latent heat release. If we assume that the portal interaction between the SM bath and dark sector leads to frequent enough interactions between the two sectors per Hubble time, then in each phase the specific heat is dominated by the many degrees of freedom of the high temperature SM bath. For example, the deconfined phase of the dark sector is found to contribute about 5% to the specific heat at the critical temperature [52]. We can then use

$$\rho(T) \approx \frac{g_*(T)\pi^2}{30} T^4 \,, \tag{C.1.12}$$

for both phases, where $g_*(T)$ is the effective number of relativistic degrees of freedom and is $g_*(T) \approx 106.75$ for all temperatures of interest in our analysis. We therefore approximate both phases as having the same specific heats and find

$$\begin{aligned} dT &\approx l \left(\frac{d\rho(x)}{dT}(T_c) \right)^{-1} dx \\ &\approx 10^{-2} T_c dx. \end{aligned} \tag{C.1.13}$$

The total temperature evolution of the universe during the phase transition is then given by

$$\dot{T} = -HT + 10^{-2}\, T_c \dot{x} \,, \tag{C.1.14}$$

where the first term comes from the adiabatic cooling of relativistic species due to Hubble expansion. Had we considered a model in which few interactions take place between the standard model and dark sector baths per Hubble time, then the two sectors would be thermally decoupled. T would refer to the dark sector's temperature, which would heat relative to the SM temperature, and we would have divided by the dark sector's specific heat, eliminating the factor of 10^{-2} in Eq. (C.1.14).

There is an important distinction between the temperature evolution of a weakly first-order phase transition and a strongly first-order phase transition. In a weakly first-order phase transition the latent heat in units of T_c is typically small (see [267]) so that the heating term in Eq. (C.1.14) is negligible and the amount of supercooling does not change much due to the added latent heat. In a strongly first-order phase

transition the latent heat in units of T_c can be order one or larger (see Eq. (C.1.2)) so that T can be driven back up to T_c before the phase transition completes. Furthermore, since κ in Eq. (C.1.9) scales inversely with l^2, the large value of the latent heat decreases the amount of supercooling needed to achieve efficient bubble nucleation as is seen in Eq. (C.1.8), making it easier for the universe to reheat to a point where nucleation is negligible. We will show in a simulation below that this reheating scenario is achieved in our phase transition.

Notice also that T cannot ever reheat all the way up to T_c. If it did, the critical radius would diverge and all bubbles would shrink. The second term in Eq. (C.1.14) would change sign and become a cooling term since latent heat would be absorbed, and hence T would be driven back below T_c. Instead, $|T-T_c|$ decreases from its maximum to an equilibrium value very close to zero at which the heating and cooling terms in Eq. (C.1.14) nearly balance one another, as we will show below. This equilibrium phase coexistence is exactly the regime described by the Maxwell construction for first-order phase transitions (see [265], §84).

As explained in the previous section, we assume that bubbles nucleated at time t_0 are of size $R_c(T(t_0))$, giving the initial condition $R(t = t_0, t_0) = R_c(T(t_0))$. To determine the radius at a future time, we need an expression for the bubble wall velocity, $\dot{R}(t, t')$. An accurate treatment of the bubble wall velocity, requires full 3+1 dimensional numerical simulations of bubble dynamics during the phase transition. However, even in simplified settings, various numerical simulations have not converged on a single, definitive answer [216, 220–222]. Instead, we will use a convenient, basic model of $\dot{R}(t, t')$. We require that critical bubbles not change their radius, and that larger bubbles expand while smaller bubbles contract. To capture this behavior, we use the simple functional form

$$\dot{R}(t, t_0) = v_w(t)\,\mathrm{sign}\left[R(t, t_0) - R_c(t)\right] , \tag{C.1.15}$$

where we define $\mathrm{sign}(0) = 0$.

To determine $v_w(t)$ would require a better understanding of the underlying strong dynamics. Instead, we will estimate an upper bound on $v_w(t)$ based on thermodynamic arguments. As argued in the main text, the larger v_w is, the less the DM relic abundance will be suppressed. So we will always set v_w to its upper bound, assuming it to be a conservative choice.

As a bubble expands, it releases latent heat near its wall, *locally* heating the plasma at the wall to a temperature $T_{\rm wall} > T$ and reducing the free energy difference at the interface. Since the free energy density is minus the pressure, this local heating reduces the net pressure acting on the wall (See Eq. (C.1.4)). Since the degree of supercooling is so small, the wall could potentially heat up to a temperature at which the net pressure balances against the surface tension. If the wall reached this temperature, bubble growth would no longer be thermodynamically favorable, and so the wall motion would come to a halt. By an argument completely analogous to the one that lead to Eq. (C.1.3), except we evaluate Eq. (C.1.4) at the wall temperature, we find this equilibrium wall temperature to be

$$T_{\rm wall}^{\rm eq} = T_c \left(1 - \frac{2\sigma}{l\,R}\right) . \tag{C.1.16}$$

We assume that as the wall temperature approaches $T_{\rm wall}^{\rm eq}$ its growth slows down gradually. Before the wall temperature reaches $T_{\rm wall}^{\rm eq}$ it will have slowed down to a steady state at which the rates of wall heating and cooling cancel one another. By estimating the rates of wall heating and cooling then setting them equal, we will determine an approximate expression for $v_w(t)$.

We start with the cooling rate. We will assume that $T_{\rm wall}$ is very close to $T_{\rm wall}^{\rm eq}$, which is in turn very close to T_c since the second term in Eq. (C.1.16) is very small compared to 1 for bubbles larger than Λ^{-1}. This assumption will lead to a faster v_w. The fractional temperature difference between points near and far from the wall is then $(T_c - T)/T_c$. We assume that the heat loss rate is given by a diffusion equation, $\dot{T}_{\rm cool} \sim -K\nabla^2 T$, and that the transport coefficient at T_c is of order $K \sim \Lambda^{-1}$. If we further assume that the length scale of the density gradient is Λ^{-1}, then we find $\dot{T}_{\rm cool} \sim -\Lambda^2\,(T_c - T)/T_c$.

Now we move on to the heating rate. We start with the energy injected per unit wall area and time, $l\,v_w$. If we assume that this energy is injected within a typical length Λ^{-1} of the wall, then the energy injected per unit volume is $l v_w \Lambda$. As before, dividing by the specific heat, $d\rho/dT$, converts the energy increase into a temperature increase. Rather than assume that the specific heat is dominated by the SM degrees of freedom as we did when deriving Eq. (C.1.13), we assume that for this process the SM degrees of freedom are irrelevant and the specific heat is dominated by the dark sector degrees of freedom, though this choice will not affect our final result. We do

so because we anticipate that for most models the portal interaction between the SM and dark sector will generically take place on timescales much slower than Λ^{-1}, and so will be inefficient compared to the interactions within the dark sector. Hence, we assume that the dark degrees of freedom disperse latent heat on a fast timescale of order Λ^{-1}, and only on much longer timescales does this heat find its way into the SM degrees of freedom.[2] Then we can use $d\rho_{\rm DS}/dT \sim T^3$ [52], which gives a heating rate of $\dot{T}_{\rm heat} \sim \Lambda^2 v_w$.[3] The wall velocity at which both rates are in balance is then

$$v_w = \left(\frac{T_c - T}{T_c}\right) . \tag{C.1.17}$$

As argued in Ref. [209], we see that v_w is suppressed by the small degree of supercooling during the phase transition.

We apply a similar argument to the wall velocity of contracting bubbles, which are smaller than the critical radius. As bubbles contract, latent heat is absorbed near the wall, decreasing the wall temperature, increasing the net pressure on the wall, and thus opposing further contraction. The contraction rate is therefore limited by the rate at which the cold wall can heat due to heat flow from the hotter surrounding plasma. Balancing the heating and cooling rates as before leads to a wall velocity given by Eq. (C.1.17) up to a relative sign, justifying the symmetric functional form of Eq. (C.1.15).

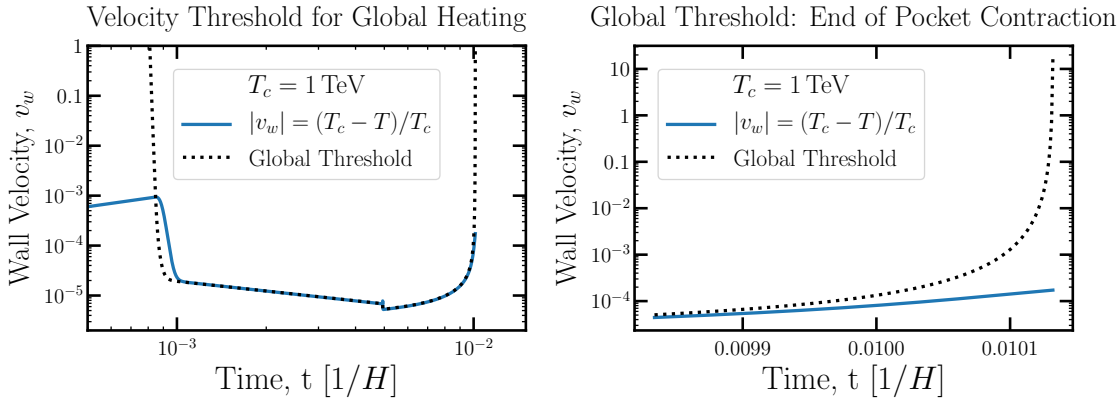
Again, the above expression for v_w is in reality an upper bound. An expanding bubble wall cannot move any faster than Eq. (C.1.17), at least for an extended period, because then it would *locally* overheat the wall. There is a similar consideration that the *global* rate of increase in temperature from bubble expansion should not significantly outpace the Hubble cooling, lest the whole universe be heated above $T = T_c$. Consequently, we will see that the value of v_w derived in Eq. (C.1.17) evolves so that it always lies below or close to this "global threshold".

The wall velocity (Eq. (C.1.17)) and global threshold are plotted in Fig. C-1. At

[2] We assume that the timescale over which interactions between the SM and dark sector baths exchange energy is much faster than the Hubble rate, justifying why the SM degrees of freedom are included in the heat capacity in Eq. (C.1.13).

[3] If the portal interaction is actually efficient enough to keep the two sectors in equilibrium on this short timescale, then the heating rate would be suppressed by a factor of $g_\star$. However, the cooling rate would also be suppressed by a factor of $g_\star$ since K is inversely proportional to the number density of interacting degrees of freedom, [268] §3.9. These two factors would then cancel in the expression for v_w. We therefore expect Eq. (C.1.17) to be independent of the specific portal interaction.

Figure C-1: The wall velocity (blue) and "global threshold" (black dashed) above which phase conversion is so fast that the universe experiences net heating. The right panel displays the same data as the left panel, but zooms in to the very end of the phase transition when pockets have contracted significantly. The discontinuity in the middle is an artifact of our modeling. It occurs when the spectrum of bubbles, which peaks at R_0, discontinuously jumps to a delta function spectrum of pockets centered at R_1.



the start of the phase transition, when bubbles are rare and their radii are small, the global heating from bubble expansion is very small even for $v_w \to c$, and so the global threshold velocity goes to infinity. The same is true at the end of the phase transition, when there is very little volume available for phase conversion. However, during the phase transition when the rate of phase conversion is rapid, this threshold becomes relevant.

Let us consider what happens when bubble coalescence and growth causes $v_w = (T_c - T)/T_c$ to exceed this global threshold (which depends on the bubble density and typical bubble radius) for the first time. If v_w overshoots the threshold, then the universe will begin to heat up on average, thus reducing v_w. The net effect is for the degree of supercooling to evolve such that v_w tracks the global threshold velocity. In practice, we observe that there is at first an abrupt drop in both v_w and the global threshold velocity, associated with a sharp increase in the bubble number density due to nucleation; in this epoch, v_w slightly exceeds the global threshold velocity and this supports a fairly rapid increase in T. Once nucleation becomes inefficient due to the rising temperature, the global threshold velocity evolves more slowly, driven by the expansion of the largest existing bubbles. In this epoch v_w tracks the global threshold closely and $\dot{T} \approx 0$, with a slow adiabatic increase in temperature toward T_c driven by the slow decrease in the global threshold velocity (which requires a corresponding

decrease in v_w and hence in the degree of supercooling). We can even derive the scaling of the global threshold (and v_w) with t during this period. By Eq. (C.1.14), $\dot{T} \approx 0$ implies $\dot{x} \approx 100H$. Using the fact that the spectrum of bubbles is strongly peaked at a single radius R, and the nucleation of bubbles is so suppressed that the number density of bubbles is constant, we have $x = 4\pi R^3 n_{\rm bub}/3$. Combining the two equations gives $dR/dt \propto R^{-2}$ so that $R = 3A(t-t_0)^{1/3}$ for some constants A and t_0. Therefore, $v_w = dR/dt = A(t-t_0)^{-2/3}$. Indeed, we find that the slope of the line in Fig. C-1 between times $.001/H$ and $.005/H$ is precisely $-2/3$.

The Hubble cooling can be relevant here, even though the phase transition takes place on timescales much smaller than a Hubble time, because the bubble expansion is so sensitive to the degree of supercooling; in contrast, we can freely drop e.g. density dilution terms corresponding to the Hubble expansion, as there is no comparably small density difference relevant to our calculation (see e.g. Eq. (4.3.2)). After coalescence, when the heating comes from pockets of shrinking radius rather than bubbles of expanding radius, the reverse process occurs, with a slow adiabatic decrease in the equilibrium temperature due to the decreasing size of the pockets. The temperature evolution eventually switches over to the standard Hubble cooling once $v_w = (T_c - T)/T_c$ can no longer reach the global velocity threshold (and hence $\dot{T} \approx 0$ cannot be maintained).

With Eqs. (C.1.11) and (C.1.14) through (C.1.17) in hand we are able to simulate the first half of the phase transition. This system of equations models the initial bubble nucleation and accompanying latent heat release, and the proceeding equilibrium regime up until percolation when bubbles begin to overlap and new dynamics must be included. Our initial conditions are that the universe has supercooled a little bit, $\frac{T-T_c}{T_c} = -10^{-4}$, the universe is fully in the deconfined phase, $x = 0$, and that no bubbles have nucleated yet. We evolve forward in small time steps, $\Delta t = 10^{-6}/H(T_c)$. In each time step we nucleate $\Gamma(T(t))(1-x(t))\Delta t$ bubbles per unit volume at radius $R_c(T(t))$ and add them to a list. We allow all other bubbles from previous steps to expand or contract by an amount $\dot{R}(t,t')\Delta t$, which only depends on the bubble size and temperature within that time step. This procedure produces bubbles with radii less than or equal to 0, so we set such bubbles' radii to zero and remove them from our list. Additionally, many time steps result in an additional number density of bubbles that is so exponentially small that the computer sets the number density

to zero. We remove these bubbles from our list, too. Each bubble nucleation and all bubble expansions increase x by $\dot{x}\Delta t$ and the temperature by $\dot{T}\Delta t$ according to Eqs. (C.1.11) and (C.1.14). We finish our evolution once $x = \frac{1}{2}$.

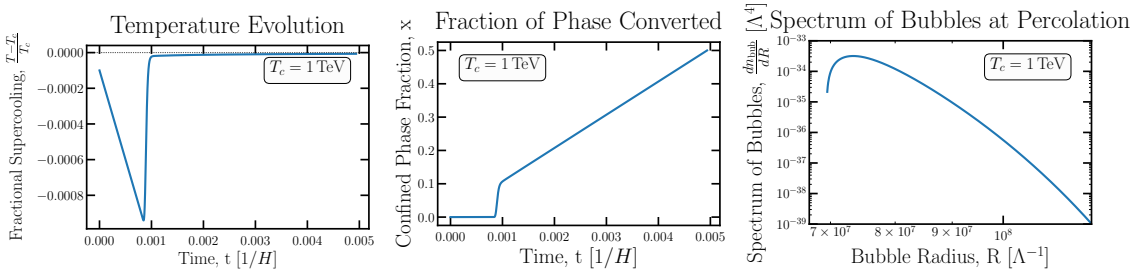
Outputs of our simulation are shown in Fig. C-2. The left plot shows the degree of supercooling before percolation and the middle plot shows the fraction of phase converted, x. These plots make clear that the first half of the phase transition can be divided into three distinct stages. In the first stage, the degree of supercooling is so small that the bubble nucleation rate is too suppressed to have a significant effect on the simulation. During this period $x = 0$ and the universe cools through Hubble expansion. In the second stage, the supercooling reaches a point at which nucleation becomes efficient. These nucleated bubbles quickly grow and inject heat, corresponding to the sudden jump in the temperature and x a little before $t = 0.001/H$. The temperature reaches a point very close to T_c at which nucleation of new bubbles becomes inefficient again, leading to the third stage. This stage is exactly the equilibrium phase coexistence regime described above. There are a fixed number density of large bubbles that grow and inject latent heat at such a rate as to cancel Hubble cooling. The net effect is that $\dot{T} \approx 0$. By Eq. (C.1.14), we have

$$\dot{x} \approx 10^2 H\,. \tag{C.1.18}$$

Since the temperature is constant during this stage, the Hubble rate is as well, meaning that x grows linearly in time, which can be seen in our middle plot. This equation explains why the phase transition occurs over a time scale of $10^{-2}/H$. Had we assumed instead that the portal interaction between the SM bath and dark sector was very weak and led to few scatters per Hubble time, then the dark sector would not have access to the SM heat capacity and the above factor of 10^2 would be replaced by an $\mathcal{O}(1)$ factor instead.

The rightmost plot in Fig. C-2 shows the spectrum of bubble radii at the time of percolation, $x = 1/2$. The shape of this spectrum is a product of the preceding stages. The earliest bubbles were nucleated during the first stage. They had the longest time to grow, but were nucleated at a time of relatively small supercooling, meaning that their number density was exponentially suppressed. So as R increases away to the right of the peak of the spectrum, $dn_{\rm bub}/dR$ decreases. Just as the second stage begins the universe is maximally supercooled. The bubbles produced at this

Figure C-2: The degree of supercooling as a function of time (left), the confined phase fraction as a function of time (middle), and the spectrum of bubble radii at percolation, when $x = 1/2$ (right).



point are the most numerous and constitute the peak of the distribution. In the rest of the second stage, the supercooling quickly diminishes, producing an exponentially suppressed population of bubbles that are smaller than the peak radius since they have less time to grow, leading to the sharp decrease of the spectrum to the left of the peak.

We define the peak in dn_{bub}/dR to occur at R_0 and empirically find that it is very well fit by the function

$$R_0\,\Lambda = 10^{-6} \times \left(\frac{\Lambda}{M_{\mathrm{pl}}}\right)^{-0.9} = 6.7 \times 10^7 \times \left(\frac{\Lambda}{\mathrm{TeV}}\right)^{-0.9} . \tag{C.1.19}$$

R_0, however, is not the only relevant length scale for the bubbles. An additional length scale, R_1, emerges from the dynamics of bubble coalescence.

At percolation, bubbles frequently come into contact with one another and begin to coalesce. To model the coalescence dynamics, we borrow another argument from [209]. When two spherical bubbles of radius R coalesce into one larger-radius bubble, they decrease their surface area and therefore reach a more energetically favorable configuration due to surface tension. The energy difference between the two configurations is $\Delta E \sim 4\pi R^2(2 - 2^{\frac{2}{3}})\sigma = 4\pi R^2(2 - 2^{\frac{2}{3}}) \times .02T_c^3$. This change in energy is achieved by applying a force to the mass M in the bubbles over a characteristic distance R. So $F \sim \Delta E/R \sim Ma$, where a is the acceleration of the material in the bubbles. This acceleration can also be estimated as $a \sim R/t^2_{\mathrm{coalesce}}$, where t_{coalesce} is the coalescence timescale. If we then use that the total mass in the two bubbles is $M = 2 \times \frac{4}{3}\pi R^3 \rho_{\mathrm{deconf}} \approx \frac{8}{3}\pi R^3 T_c^4$, where we used that $\rho_{\mathrm{deconf}}(T_c) \approx T_c^4$ [52], then we

find

$$t_{\text{coalesce}} \sim \left(\frac{MR^2}{\Delta E}\right)^{\frac{1}{2}} \tag{C.1.20}$$

$$\sim 10 T_c^{\frac{1}{2}} R^{\frac{3}{2}} \,. \tag{C.1.21}$$

The above equation shows that small bubbles coalesce quicker than large ones. Therefore, at percolation small bubbles will quickly coalesce until they reach a size R_1 past which t_{coalesce} takes longer than the timescale of percolation. From Fig. C-2 we can estimate the timescale of percolation as the time it takes x to change by a couple percent at around $x = \frac{1}{2}$. We find $t_{\text{perc}} \sim 10^{-3} H^{-1} \sim 10^{-3} \frac{M_{\text{pl}}}{T_c^2}$ where $M_{\text{pl}} = (8\pi G)^{-1/2} \sim 2.4 \times 10^{18}\text{GeV}$ is the reduced Planck mass and G is Newton's constant. Setting the percolation timescale equal to t_{coalesce} then yields the critical bubble size of,

$$R_1 \, \Lambda \approx 10^{-8/3} \left(\frac{M_{\text{pl}}}{\Lambda}\right)^{2/3} \approx 4 \times 10^7 \left(\frac{\text{TeV}}{\Lambda}\right)^{2/3} \,, \tag{C.1.22}$$

in units of $T_c = \Lambda$.

In Fig. 4.2.2 we plot R_0 and R_1 as a function of the confinement scale. We find that for $\Lambda \gtrsim 1\,\text{TeV}$, R_0 is less than or equal to R_1. Therefore, for this range of Λ, once our simulation finishes at $x = \frac{1}{2}$ we assume that all bubbles begin coalescing and quickly grow to radius R_1. Since x does not change during this process, we assume that T remains fixed, too. For $\Lambda < 1\,\text{TeV}$, we assume that coalescence is inefficient and bubbles remain at radius R_0.

C.1.3 Second half of the phase transition: pocket contraction

After percolation, most bubbles are in contact with one another. The confined regions form a web and the deconfined regions form isolated pockets. We assume these pockets quickly attain spherical symmetry due to surface tension, and also that the typical size of a bubble before percolation is equal to the typical size of a pocket after percolation, R_1.

Since all pockets are at the same initial radius, we can solve for the initial density of pockets. The number density of pockets that are all of radius R satisfies

$$1 - x = \frac{4\pi}{3} R^3 n_{\text{pocket}} \,. \tag{C.1.23}$$

Since $x = \frac{1}{2}$ at percolation, we have

$$n_{\text{pocket}} = \frac{3}{8\pi R_1^3}\,. \tag{C.1.24}$$

We will find that the degree of supercooling continues to be so small during pocket contraction that the nucleation of more bubbles is completely suppressed. Therefore n_{pocket} remains constant and we find

$$x = 1 - \frac{R^3}{2R_1^3}\,. \tag{C.1.25}$$

As before the contraction rate of the pockets is limited due to the latent heat release near the pocket wall. However, the wall velocity estimate for pockets is slightly different than it was for bubbles. Whereas supercooling results in a net pressure outward for bubbles, supercooling results in a net pressure inward for pockets, since the two phases are on opposite sides of the wall in either case. With this caveat in mind we can repeat our argument leading to Eq. (C.1.17).

As before, we argue that as it expands, the wall quickly heats up, approaching a threshold temperature at which pressure and surface tension are in equilibrium. This threshold now corresponds to slight superheating at the temperature $T_c\left(1 + \frac{2\sigma}{lR}\right)$. Before T_{wall} reaches this threshold it achieves a steady state at which heating from latent heat injection cancels against cooling from heat diffusion from the wall. Using the same arguments as before, we find that this steady state corresponds to a velocity of

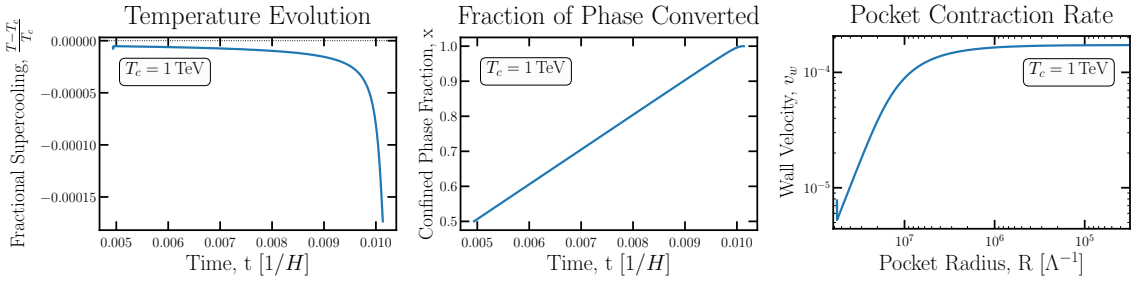
$$\frac{dR}{dt} = -(T_c - T)/T_c\,. \tag{C.1.26}$$

Again, we used that $\frac{2\sigma}{lR} = \frac{.03}{R\Lambda} \ll 1$, so we approximate the temperature difference between the wall and its surroundings appearing in the above equation as $T_c - T$, which leads to an overestimate of the wall velocity.

There is another much more important new effect modifying the contraction rate of a pocket: quark pressure. The density of quarks trapped within pockets increases as they are forced within ever-shrinking volumes. This pressure opposes contraction, slowing down the wall velocity relative to Eq. (C.1.26). For now, to build intuition, we will ignore the effect of quark pressure, and we will consider it in the next section.

With Eqs. (C.1.14), (C.1.25), and (C.1.26) we have all we need to simulate the

Figure C-3: The degree of supercooling as a function of time (**left**), the confined phase fraction as a function of time (**middle**), and the pocket wall velocity as a function of bubble radius during the second half of the phase transition (**right**).



second half of the phase transition. We use the same method and parameters as we did for the first half of the phase transition. In Fig. C-3 we show the results of our simulation. The first instant of the simulation features a discontinuity in the temperature (and thus v_w). This discontinuity results from our discontinuous change of the spectrum of bubbles from the smooth form of dn/dR in the right panel of Fig. C-2 to a delta function at R_1. As the pockets get smaller, their heating rate diminishes. Hubble cooling becomes relatively more important over time leading to the steady decrease in T over time. By the very end of the phase transition, the temperature evolution is purely determined by Hubble cooling.

The right panel shows the pocket wall velocity as a function of the pocket radius. As a function of R, v_w asymptotes to a well-defined value near 2×10^{-4} for the choice of Λ shown in the figure. Naively, this plot seems to be at odds with the left panel, since the two plots are equivalent up to a minus sign. Whereas the velocity seems to asymptote to a constant value at late times, the degree of supercooling seems to vary greatly at late times. The apparent discrepancy is a result of the different x-axis scales. Whereas the x-axis in the left panel is linearly scaled in t, the x-axis in the right panel is log scaled in R. The majority of the simulation takes place when the pocket radii are very large. The pocket radii are much smaller than their initial value only for a very short time at the end of the simulation, at time scales much shorter than the $1/H$. In fact, the small timestep $10^{-6}/H$ was chosen to resolve this smaller timescale. At such small timescales and pocket radii, very little Hubble cooling or latent heat injection takes place, leading to the plateau in v_w. Since most of the quark interactions recouple at pocket radii much smaller than R_1 (see e.g. Fig. 4-6), we are

justified in treating v_w as constant within the Boltzmann equations of Sec. 4.3.[4] The asymptotic v_w value as a function of Λ is plotted in Fig. 4.2.2 and is well fit by

$$v_w(R \ll R_1) = 0.2 \left(\frac{\Lambda}{M_{\rm pl}}\right)^{0.2} = 1.7 \times 10^{-4} \times \left(\frac{\Lambda}{\rm TeV}\right)^{0.2} . \tag{C.1.27}$$

C.1.4 The effect of quark pressure

Up to this point we have neglected the effect of quark pressure on the phase transition,

$$p_q = n_q T \, . \tag{C.1.28}$$

For the first half of the phase transition, this approximate treatment was justified. At the start of the phase transition, p_q is suppressed compared to the gluon pressure. During the phase transition n_q grows, and so the quark pressure could potentially become large enough to affect the bubble dynamics. During the first half of the phase transition, however, the quark density, and hence p_q, grows by only a factor of two. Including this factor of two enhancement, we find that the quark pressure is sub-dominant compared to the net gluon pressure during the bubble growth stage of the phase transition, and hence can be neglected.

On the other hand, during the second half of the phase transition, the quarks are compressed much more. We find that for every point in the DM parameter space we consider, the quark pressure eventually becomes comparable to the other forces acting on the wall. Most likely, the increased quark pressure would oppose further contraction and slow v_w down. Unfortunately, the process by which p_q gradually grows and, in response, v_w gradually shrinks, is a non-equilibrium, strong physics problem for which we have no solution. Nevertheless, we can still use thermodynamic arguments as before to understand the possible limiting behavior of v_w.

Consider the scenario in which the enhanced quark pressure forces the pocket into a state of mechanical equilibrium. Mechanical equilibrium is achieved when the four forces on the wall – gluon pressure inside the pocket, glueball pressure outside the pocket, surface tension, and quark pressure – are in balance. At this equilibrium

[4] Recall that by "recoupling" we mean the point when the rates of quark interactions become large compared to the contraction rate of the pocket, so the quark density evolution is dominated by interactions.

point, we must have

$$\begin{aligned} 0 &= dA\,\sigma + dV\,\left(\sum p\right)\Big|_{\text{wall}} \\ &= 8\pi\sigma R + 4\pi\Delta f(T_{\text{wall}})R^2 - 4\pi p_q R^2 \\ &= 8\pi\sigma R + 4\pi\frac{(T_c - T_{\text{wall}})}{T_c} l\, R^2 - 3N_q T_{\text{wall}} R^{-1}\,, \end{aligned} \tag{C.1.29}$$

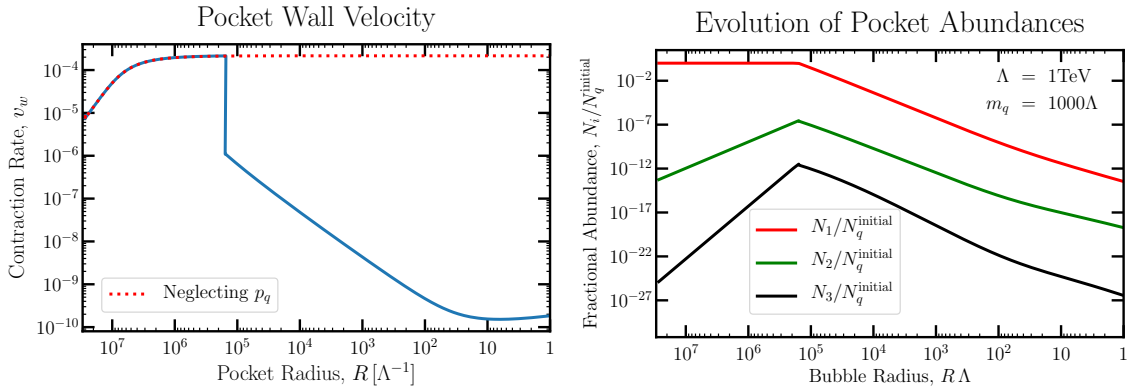
where Δf still refers to the confined phase minus the deconfined phase gluonic pressures. The temperatures are all evaluated at the local wall temperature and we have used $N_q = \frac{4\pi}{3}R^3\, n_q$. If the system ever reached this equilibrium point, it would be a stable equilibrium: if R were to contract the increased quark density would oppose it; if R were to expand the surface tension would increase, the quark pressure would decrease, and the wall would absorb latent heat and increase the net gluon pressure difference, all of which oppose further expansion. After achieving mechanical equilibrium, the wall would proceed to contract adiabatically as quark annihilation decreases the quark pressure and wall cooling increases the net gluon pressure compressing the wall. By differentiating Eq. (C.1.29) with respect to time and defining $v_w = -\dot{R}$, we find

$$v_w = \frac{-3\dot{N}_q T_{\text{wall}} - \left(N_q + \frac{4\pi l R^3}{T_c}\right)\dot{T}_{\text{wall}}}{\left(8\pi\sigma R + 8\pi\frac{(T_c - T_{\text{wall}})}{T_c} l R^2 + 3N_q T_{\text{wall}} R^{-1}\right)}\,. \tag{C.1.30}$$

Since we do not keep track of T_{wall}, nor can we calculate v_w as a function of T_{wall} when the wall is out of mechanical equilibrium, we have no way of knowing when or if mechanical equilibrium is achieved.

However, to develop some intuition for the possible effects of quark pressure, we can perform a crude approximate calculation. For this calculation, we perform our pocket contraction simulation while simultaneously keeping track of the quark abundance within each step using the Boltzmann equations of Sec. 4.3. Since the quark pressure is sub-dominant initially, we start with v_w given by Eq. (C.1.26). Eventually there comes a radius when the quark pressure is so large that it is able to oppose the combined forces of surface tension and the net gluonic pressure, even when the latter pressure is at its maximum (which is attained when the wall temperature is at its minimum, $T_{\text{wall}} = T$). At this radius, we say that the wall has suddenly attained mechanical equilibrium. We then switch over to a wall velocity of Eq. (C.1.30) for the

Figure C-4: The pocket wall velocity (left) and particle abundances (right) within the pocket using a model that crudely incorporates the effects of quark pressure on v_w. We choose a confinement scale of 1 TeV and dark matter mass of 10^3 TeV. We begin the simulation neglecting the quark pressure, allowing us to use Eq. (C.1.26) to determine the contraction rate (red-dotted line). Eventually the quark pressure is so large that it is able to come into mechanical equilibrium with the other forces acting on the wall, which happens at the discontinuity near $R\,\Lambda = 10^5$. At this point, we switch over to a contraction rate given by Eq. (C.1.30), leading to the discontinuous drop in v_w in the left panel and the sudden depletion of all particle abundances in the right panel.



rest of the simulation. We assume that the system maintains mechanical equilibrium and $T_{\rm wall} = T$ until the end of the simulation.

In Fig. C-4 we plot the velocity and particle abundances within the pocket as functions of R for this new simulation. One can see that the velocity discontinuously drops at a radius of $R \sim 10^5\Lambda^{-1}$ when the pocket abruptly achieves mechanical equilibrium. When this happens, further contraction is allowed only by subsequent quark annihilations; the quark depletion processes immediately recouple and dominate the density evolution due to the sharp drop in the contraction rate v_w/R. At this same radius, the baryon abundance is at a maximum and we find that the DM relic abundance is set. Afterwards, the pocket slowly contracts and all particle numbers deplete until the pocket vanishes. Importantly, we find that the pocket asymmetry is saturated in this simulation (see Eq. (4.2.16)), and is also saturated for every other point in the DM parameter space that we consider. Of course, this is a crude approximation – a realistic scenario could have a smoother evolution of v_w, or if the wall velocity falls sharply, this could induce plasma shock waves which modify the pocket evolution. However, this explicit example supports our intuition that turning on quark pressure will drive the system rapidly into the regime where the asymmetry is saturated, and once in this regime the details of the evolution do not affect the final relic abundance.

C.2 Cross Sections

For the computation of the baryon survival factor, multiple processes are relevant. We distinguish between three classes of interactions:

- Annihilation process, i.e. a direct annihilation of free quarks into dark gluons: $1 + (-1) \to 0 + 0$.
- Capture processes, where a dark gluon is emitted for momentum conservation, for example: $1 + 1 \to 2 + 0$.
- Rearrangement processes, where no gluon is emitted and only quark constituents exchanged, for example: $2 + 2 \to 3 + 1$.

Similar to Tab. 4.1, in writing these equations we use each relic's quark number.

For the values of the annihilation and capture cross sections explicit calculations taking into account group theory factors have been performed in [47, 184, 185]. The cross sections scale as

$$\langle \sigma_{\text{ann./cap.}} v \rangle = \zeta \, \frac{\pi \alpha^2}{m_q^2} \equiv \zeta \, \sigma_0 \,, \tag{C.2.1}$$

where ζ is a numerical factor that depends on the number of colors and flavors in the theory, and the coupling strength α is evaluated at the scale of the momentum transferred in the annihilation process, which is m_q. In addition at low interaction energies the bound state formation and the annihilation process experience enhancement due to non-perturbative Sommerfeld corrections. The cumulative effect of those non-perturbative effects at finite temperature can be taken into account by the effective cross section $\langle \sigma_{\text{eff}} v_{\text{rel}} \rangle$ defined in [47, 184, 185]. In Fig. C-5 the factor ζ is shown for different annihilation and capture processes in our full set of Boltzmann equations.

Thermal masses of the dark gluons prevent bound state formation at large temperatures, an effect that has been confirmed by additional investigations of the process at hand in a non-equilibrium field theory treatment [47, 269, 270].

The rearrangement process is more complex and requires taking into account non-perturbative effects. Here simulations and comparisons to hydrogen−anti-hydrogen annihilation have been performed in [184]. The resulting cross section scales with the area set by the Bohr radius of the colliding bound states and contains a suppression

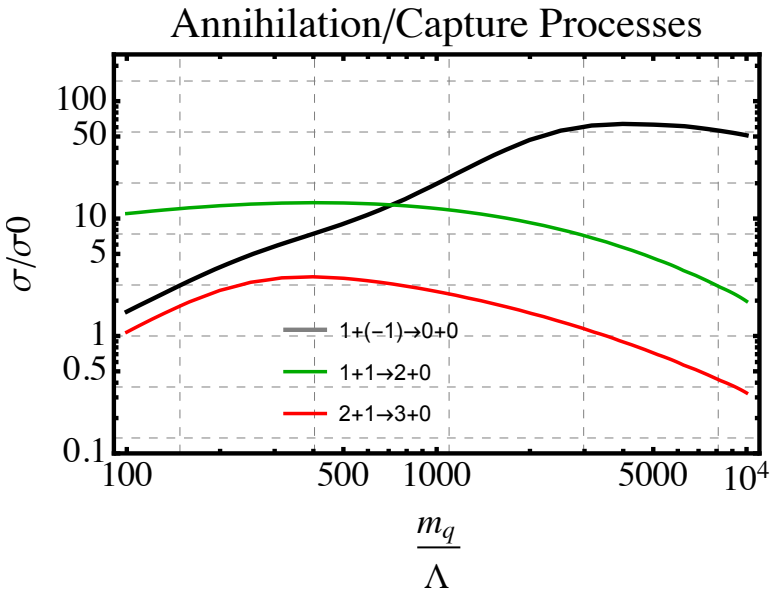


Figure C-5: Reproducing the results of [47, 184, 185] for ζ factor in Eq. (C.2.1) for various annihilation or capture processes entering our full set of Boltzmann equations.

factor which becomes effective once the kinetic energy exceeds the available binding energy:

$$\sigma_{\rm RA} = \frac{\pi R_{\rm Bohr}^2}{\sqrt{E_{\rm kin}/E_{\rm E_B}}}. \tag{C.2.2}$$

This results in a constant σv cross section, which is expected for an exothermic reaction

$$\langle \sigma_{\rm RA} v \rangle = \frac{1}{C_N \alpha} \frac{\pi}{m_q^2} = \frac{\sigma_0}{C_N \alpha^3}\,, \tag{C.2.3}$$

with C_N being the quadratic Casimir of the dark quark representation ($C_N = 4/3$ for quarks in the fundamental representation of $SU(N)$), that controls the interquark attraction in a non-abelian theory. The overall scaling is thus $\sigma_{\rm RA} v \sim \sigma_{\rm ann./cap.} v/\alpha^3$, in agreement with the numerical results of [271]. This non-perturbative enhancement results from taking into account the finite size of the colliding bound states.

We summarize all the cross sections entering Eq. (4.3.10) in Tab. C.1. Notice that in this table some processes involving gluons (denoted by 0) are listed as rearrangement. These processes, in fact, have an intermediate step in which the quarks rearrange and make pions ($\bar{q}q$), which can promptly decay into gluons, see Eq. (4.3.1).

Class	Process	Cross section
Annihilation	$1 + (-1) \to 0 + 0$	Fig. C-5 & Eq. (C.2.1)
Capture	$1 + 1 \to 2 + 0$	
	$2 + 1 \to 3 + 0$	
Rearrangement	$(-3) + 1 \to (-1) + (-1)$	Eq. (C.2.3)
	$(-3) + 1 \to (-2) + 0$	
	$(-2) + 1 \to (-1) + 0$	
	$3 + (-2) \to 1 + 0$	
	$2 + 2 \to 3 + 1$	
	$3 + (-2) \to 2 + (-1)$	
	$3 + (-3) \to 2 + (-2)$	
	$2 + (-2) \to 1 + (-1)$	
	$3 + (-3) \to 1 + (-1)$	
	$3 + (-3) \to 0 + 0$	
	$2 + (-2) \to 0 + 0$	

Table C.1: The processes and cross section classes involved in the annihilation and baryon formation. Bound states are donated by their baryon number. Direct annihilation takes place if multiple gluons need to be emitted in the final state. If one gluon is radiated, then perturbative capture cross section calculations apply. In the case that the final states have no gluons that would be needed for momentum conservation, we use the geometric rearrangement cross sections discussed below. The 0 in the rearrangement processes denotes a pion $\bar{q}q$ that promptly decays to gluons. Generally, all processes above have equivalent reactions, where all particles are replaced by anti-particles. For those we assume the same cross sections, i.e. assuming CP is conserved.

C.3 Binding Energies

The binding energies of several types of dark states enter the full set of Boltzmann equations. For the two quark states exact results are available. Since we work in the limit $m_q \gg \Lambda_{\rm DC}$ the Coulomb potential approximation is valid. For the baryon binding energy variational methods are needed, and numerical evaluations were performed in [184]. We focus on the case that $N = 3$. The resulting binding energies are:

- Binding energy of a singlet diquark, or meson $\bar{q}q$: $E_B^{\bar{q}q} = \frac{1}{4}\alpha^2 C_N^2 m_q$.
- Binding energy of a bound non-singlet diquark state qq in a binding configuration is: $E_B^{qq} = \frac{1}{4}E_B^{\bar{q}q}$.
- Binding energy of a baron qqq singlet sate is: $E_B^{qqq} = 0.26\alpha^2 C_N^2 m_q$.

Here α is the gauge coupling of the confining group given by Eq. (4.2.4). The relevant scale at which the coupling should be evaluated is the Bohr momentum αm_q, which can be determined iteratively, starting from the value of $\alpha(m_q)$. It has been shown that the corrections due to the linear (confining) part of the potential between quarks has negligible effect on these binding energies (see Eq. (4) of [185]).

Note that the binding energies inside and outside the pockets differ only by the subdominant contribution of the Cornell potential $\Lambda^2 r$, evaluated at the interquark spacing $R_b \sim 1/(\alpha m_q)$. That contribution is suppressed by $\Lambda^2/m_q^2 \ll 1$ with respect to the ground state binding energy, which is of the order $\alpha^2 m_q$.

Bibliography

[1] S. Morisi and S. Boucenna, "Theories relating baryon asymmetry and dark matter," Frontiers in Physics **1** (2014), 10.3389/fphy.2013.00033.

[2] R. Catena and L. Covi, "SUSY dark matter(s)," The European Physical Journal C **74** (2014), 10.1140/epjc/s10052-013-2703-4.

[3] L. J. Rosenberg, "Dark-matter qcd-axion searches," Proceedings of the National Academy of Sciences **112**, 12278–12281 (2015).

[4] B. Carr and F. Kühnel, "Primordial black holes as dark matter: Recent developments," Annual Review of Nuclear and Particle Science **70**, 355–394 (2020), https://doi.org/10.1146/annurev-nucl-050520-125911 .

[5] B. Carr and F. Kuhnel, "Primordial black holes as dark matter candidates," (2021).

[6] A. D. POPOLO, "NONBARYONIC DARK MATTER IN COSMOLOGY," International Journal of Modern Physics D **23**, 1430005 (2014).

[7] Particle Data Group, "Review of Particle Physics," Progress of Theoretical and Experimental Physics **2020** (2020), 10.1093/ptep/ptaa104.

[8] H. W. Babcock, "The rotation of the Andromeda Nebula," Lick Observatory Bulletin **498**, 41–51 (1939).

[9] V. C. Rubin and J. Ford, W. Kent, "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," Astrophys. J. **159**, 379 (1970).

[10] V. C. Rubin, J. Ford, W. K., and N. Thonnard, "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." Astrophys. J. **238**, 471–487 (1980).

[11] A. Bosma, "21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types." The Astrophysical Journal **86**, 1825–1846 (1981).

[12] . A. A. Abdo, "OBSERVATIONS OF MILKY WAY DWARF SPHEROIDAL GALAXIES WITH THE<i>FERMI<i>-LARGE AREA TELESCOPE DETECTOR AND CONSTRAINTS ON DARK MATTER MODELS," The Astrophysical Journal **712**, 147–158 (2010).

[13] C. L. Sarazin, "X-ray emission from clusters of galaxies," Rev. Mod. Phys. **58**, 1–115 (1986).

[14] D. Clowe, M. Bradač, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter," The Astrophysical Journal **648**, L109–L113 (2006).

[15] R. Massey, T. Kitching, and J. Richard, "The dark matter of gravitational lensing," Reports on Progress in Physics **73**, 086901 (2010).

[16] . Volker Springel, "Simulations of the formation, evolution and clustering of galaxies and quasars," Nature **435**, 629–636 (2005).

[17] G. Blumenthal, S. Faber, J. Primack, and M. REES, "Formation of galaxies and large-scale structure with cold dark matter," Nature **313** (1985), 10.1038/313072a0.

[18] N. Aghanim *et al.* (Planck), "Planck 2018 results. VI. Cosmological parameters," (2018), arXiv:1807.06209 [astro-ph.CO] .

[19] M. Schumann, "Direct detection of WIMP dark matter: concepts and status," Journal of Physics G: Nuclear and Particle Physics **46**, 103003 (2019).

[20] G. Steigman, B. Dasgupta, and J. F. Beacom, "Precise relic wimp abundance and its impact on searches for dark matter annihilation," Physical Review D **86** (2012), 10.1103/physrevd.86.023506.

[21] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, "Mechanism for Thermal Relic Dark Matter of Strongly Interacting Massive Particles," Phys. Rev. Lett. **113**, 171301 (2014), arXiv:1402.5143 [hep-ph] .

[22] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, "Model for Thermal Relic Dark Matter of Strongly Interacting Massive Particles," Phys. Rev. Lett. **115**, 021301 (2015), arXiv:1411.3727 [hep-ph] .

[23] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, "Freeze-In Production of FIMP Dark Matter," JHEP **03**, 080 (2010), arXiv:0911.1120 [hep-ph] .

[24] Y. Mambrini and K. A. Olive, "Gravitational production of dark matter during reheating," Physical Review D **103** (2021), 10.1103/physrevd.103.115009.

[25] M. Kuster, *Axions - Theory, Cosmology and Experimental Searches*, Vol. 741 (2008).

[26] M. H. Namjoo, T. R. Slatyer, and C.-L. Wu, "Enhanced n-body annihilation of dark matter and its indirect signatures," JHEP **03**, 077 (2019), arXiv:1810.09455 [astro-ph.CO] .

[27] M. Lisanti, "Lectures on dark matter physics," in *New Frontiers in Fields and Strings* (WORLD SCIENTIFIC, 2016).

[28] A. Berlin, D. Hooper, G. Krnjaic, and S. D. McDermott, "Severely Constraining Dark Matter Interpretations of the 21-cm Anomaly," (2018), arXiv:1803.02804 [hep-ph] .

[29] E. W. Kolb and M. S. Turner, "The Early Universe," Front. Phys. **69**, 1–547 (1990).

[30] H. Liu, T. R. Slatyer, and J. Zavala, "Contributions to cosmic reionization from dark matter annihilation and decay," Phys. Rev. **D94**, 063507 (2016), arXiv:1604.02457 [astro-ph.CO] .

[31] S. Schön, K. J. Mack, C. A. Avram, J. B. Wyithe, and E. Barberio, "Dark matter annihilation in the first galaxy haloes," Mon. Not. Roy. Astron. Soc. **451**, 2840–2850 (2015), arXiv:1411.3783 [astro-ph.CO] .

[32] S. Schön, K. J. Mack, and S. B. Wyithe, "Dark Matter Annihilation in the Circumgalactic Medium at High Redshifts," Mon. Not. Roy. Astron. Soc. **474**, 3067–3079 (2018), arXiv:1706.04327 [astro-ph.CO] .

[33] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, P. Panci, M. Raidal, F. Sala, and A. Strumia, "PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection," JCAP **1103**, 051 (2011), [Erratum: JCAP1210,E01(2012)], arXiv:1012.4515 [hep-ph] .

[34] C. Weniger, P. D. Serpico, F. Iocco, and G. Bertone, "CMB bounds on dark matter annihilation: Nucleon energy-losses after recombination," Phys. Rev. **D87**, 123008 (2013), arXiv:1303.0942 [astro-ph.CO] .

[35] D. Grin and C. M. Hirata, "Cosmological hydrogen recombination: The effect of extremely high-n states," Physical Review D **81** (2010), 10.1103/physrevd.81.083005.

[36] Y. Ali-Haïmoud and C. M. Hirata, "Ultrafast effective multilevel atom method for primordial hydrogen recombination," Physical Review D **82** (2010), 10.1103/physrevd.82.063521.

[37] P. J. E. Peebles, "Recombination of the Primeval Plasma," Astrophys. J. **153**, 1 (1968).

[38] S. Seager, D. D. Sasselov, and D. Scott, "How exactly did the universe become neutral?" Astrophys. J. Suppl. **128**, 407–430 (2000), arXiv:astro-ph/9912182 [astro-ph] .

[39] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, "The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Set," Astrophys. J. **473**, 576 (1996), arXiv:astro-ph/9605054 [astro-ph] .

[40] R. Essig *et al.*, "Working Group Report: New Light Weakly Coupled Particles," in *Community Summer Study 2013: Snowmass on the Mississippi* (2013) arXiv:1311.0029 [hep-ph] .

[41] F.-Y. Cyr-Racine and K. Sigurdson, "Cosmology of atomic dark matter," Physical Review D **87** (2013), 10.1103/physrevd.87.103515.

[42] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, "Model for thermal relic dark matter of strongly interacting massive particles," Physical Review Letters **115** (2015), 10.1103/physrevlett.115.021301.

[43] R. Foot, "Mirror dark matter: Cosmology, galaxy structure and direct detection," International Journal of Modern Physics A **29**, 1430013 (2014).

[44] A. Soni and Y. Zhang, "Hidden su(n) glueball dark matter," **93** (2016), 10.1103/physrevd.93.115025.

[45] C. Bonati, P. de Forcrand, M. D'Elia, O. Philipsen, and F. Sanfilippo, "Constraints on the two-flavor QCD phase diagram from imaginary chemical potential," PoS **LATTICE2011**, 189 (2011), arXiv:1201.2769 [hep-lat] .

[46] O. Kaczmarek, F. Karsch, E. Laermann, and M. Lutgemeier, "Heavy quark potentials in quenched QCD at high temperature," Phys. Rev. D **62**, 034021 (2000), arXiv:hep-lat/9908010 .

[47] A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, "Cosmological Implications of Dark Matter Bound States," JCAP **05**, 006 (2017), arXiv:1702.01141 [hep-ph] .

[48] B. Lucini, M. Teper, and U. Wenger, "Properties of the deconfining phase transition in SU(N) gauge theories," JHEP **02**, 033 (2005), arXiv:hep-lat/0502003 .

[49] T. Celik, J. Engels, and H. Satz, "The Order of the Deconfinement Transition in SU(3) Yang-Mills Theory," Phys. Lett. **125B**, 411–414 (1983).

[50] N. H. Christ and H.-Q. Ding, "Latent heat and critical temperature of the color-deconfining phase transition," Phys. Rev. Lett. **60**, 1367–1370 (1988).

[51] B. Lucini, P. de Forcrand, and M. Vettorazzo, "Measuring interface tensions in 4d su(n) lattice gauge theories," (2004), arXiv:hep-lat/0409148 [hep-lat] .

[52] L. Giusti and M. Pepe, "Equation of state of the su(3) yang–mills theory: A precise determination from a moving frame," Physics Letters B **769**, 385–390 (2017), arXiv:1612.00265 [hep-lat] .

[53] H. Liu, G. W. Ridgway, and T. R. Slatyer, "Code package for calculating modified cosmic ionization and thermal histories with dark matter and other exotic energy injections," Phys. Rev. **D101**, 023530 (2020), arXiv:1904.09296 [astro-ph.CO] .

[54] H. Liu, W. Qin, G. W. Ridgway, and T. R. Slatyer, "Lyman-α constraints on cosmic heating from dark matter annihilation and decay," Phys. Rev. D **104**, 043514 (2021), arXiv:2008.01084 [astro-ph.CO] .

[55] P. Asadi, E. D. Kramer, E. Kuflik, G. W. Ridgway, T. R. Slatyer, and J. Smirnov, "Thermal squeezeout of dark matter," Phys. Rev. D **104**, 095013 (2021), arXiv:2103.09827 [hep-ph] .

[56] P. Asadi, E. D. Kramer, E. Kuflik, G. W. Ridgway, T. R. Slatyer, and J. Smirnov, "Accidentally Asymmetric Dark Matter," Phys. Rev. Lett. **127**, 211101 (2021), arXiv:2103.09822 [hep-ph] .

[57] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, "CMB Constraints on WIMP Annihilation: Energy Absorption During the Recombination Epoch," Phys. Rev. **D80**, 043526 (2009), arXiv:0906.1197 [astro-ph.CO] .

[58] T. R. Slatyer, "Energy Injection And Absorption In The Cosmic Dark Ages," Phys. Rev. **D87**, 123513 (2013), arXiv:1211.0283 [astro-ph.CO] .

[59] T. R. Slatyer and C.-L. Wu, "General Constraints on Dark Matter Decay from the Cosmic Microwave Background," Phys. Rev. **D95**, 023010 (2017), arXiv:1610.06933 [astro-ph.CO] .

[60] T. Kanzaki, M. Kawasaki, and K. Nakayama, "Effects of Dark Matter Annihilation on the Cosmic Microwave Background," Prog. Theor. Phys. **123**, 853–865 (2010), arXiv:0907.3985 [astro-ph.CO] .

[61] L. Lopez-Honorez, O. Mena, A. Moliné, S. Palomares-Ruiz, and A. C. Vincent, "The 21 cm signal and the interplay between dark matter annihilations and astrophysical processes," JCAP **1608**, 004 (2016), arXiv:1603.06795 [astro-ph.CO] .

[62] H. Liu and T. R. Slatyer, "Implications of a 21-cm signal for dark matter annihilation and decay," Phys. Rev. **D98**, 023501 (2018), arXiv:1803.09739 [astro-ph.CO] .

[63] R. Diamanti, L. Lopez-Honorez, O. Mena, S. Palomares-Ruiz, and A. C. Vincent, "Constraining Dark Matter Late-Time Energy Injection: Decays and P-Wave Annihilations," JCAP **1402**, 017 (2014), arXiv:1308.2578 [astro-ph.CO] .

[64] P. Stöcker, M. Krämer, J. Lesgourgues, and V. Poulin, "Exotic energy injection with ExoCLASS: Application to the Higgs portal model and evaporating black holes," JCAP **1803**, 018 (2018), arXiv:1801.01871 [astro-ph.CO] .

[65] L. Lopez-Honorez, O. Mena, S. Palomares-Ruiz, and A. C. Vincent, "Constraints on dark matter annihilation from CMB observations before Planck," JCAP **1307**, 046 (2013), arXiv:1303.5094 [astro-ph.CO] .

[66] V. Poulin, P. D. Serpico, and J. Lesgourgues, "Dark Matter annihilations in halos and high-redshift sources of reionization of the universe," JCAP **1512**, 041 (2015), arXiv:1508.01370 [astro-ph.CO] .

[67] V. Poulin, J. Lesgourgues, and P. D. Serpico, "Cosmological constraints on exotic injection of electromagnetic energy," JCAP **1703**, 043 (2017), arXiv:1610.10051 [astro-ph.CO] .

[68] S. J. Clark, B. Dutta, Y. Gao, Y.-Z. Ma, and L. E. Strigari, "21 cm limits on decaying dark matter and primordial black holes," Phys. Rev. **D98**, 043006 (2018), arXiv:1803.09390 [astro-ph.HE] .

[69] A. Hektor, G. Hütsi, L. Marzola, M. Raidal, V. Vaskonen, and H. Veermäe, "Constraining Primordial Black Holes with the EDGES 21-cm Absorption Signal," Phys. Rev. **D98**, 023503 (2018), arXiv:1803.09697 [astro-ph.CO] .

[70] A. Hektor, G. Hütsi, L. Marzola, and V. Vaskonen, "Constraints on ALPs and excited dark matter from the EDGES 21-cm absorption signal," Phys. Lett. **B785**, 429–433 (2018), arXiv:1805.09319 [hep-ph] .

[71] S. Seager, D. D. Sasselov, and D. Scott, "A new calculation of the recombination epoch," Astrophys. J. **523**, L1–L5 (1999), arXiv:astro-ph/9909275 [astro-ph] .

[72] Y. Ali-Haimoud and C. M. Hirata, "HyRec: A fast and highly accurate primordial hydrogen and helium recombination code," Phys. Rev. **D83**, 043513 (2011), arXiv:1011.3758 [astro-ph.CO] .

[73] W. Y. Wong, A. Moss, and D. Scott, "How well do we understand cosmological recombination?" Mon. Not. Roy. Astron. Soc. **386**, 1023–1028 (2008), arXiv:0711.1357 [astro-ph] .

[74] Ya. B. Zeldovich, V. G. Kurt, and R. A. Sunyaev, "Recombination of hydrogen in the hot model of the universe," Sov. Phys. JETP **28**, 146 (1969), [Zh. Eksp. Teor. Fiz.55,278(1968)].

[75] J. Chluba, D. Paoletti, F. Finelli, and J.-A. Rubiño-Martín, "Effect of primordial magnetic fields on the ionization history," Mon. Not. Roy. Astron. Soc. **451**, 2244–2250 (2015), arXiv:1503.04827 [astro-ph.CO] .

[76] J. Chluba and R. M. Thomas, "Towards a complete treatment of the cosmological recombination problem," Mon. Not. Roy. Astron. Soc. **412**, 748 (2011), arXiv:1010.3631 [astro-ph.CO] .

[77] S. Furlanetto and S. J. Stoever, "Secondary ionization and heating by fast electrons," Mon. Not. Roy. Astron. Soc. **404**, 1869 (2010), arXiv:0910.4410 [astro-ph.CO] .

[78] M. Valdes, C. Evoli, and A. Ferrara, "Particle energy cascade in the Intergalactic Medium," Mon. Not. Roy. Astron. Soc. **404**, 1569–1582 (2010), arXiv:0911.1125 [astro-ph.CO] .

[79] C. Evoli, M. Valdes, A. Ferrara, and N. Yoshida, "Energy deposition by weakly interacting massive particles: a comprehensiv e study," Mon. Not. Roy. Astron. Soc. **422**, 420–433 (2012).

[80] S. Galli, T. R. Slatyer, M. Valdes, and F. Iocco, "Systematic Uncertainties In Constraining Dark Matter Annihilation From The Cosmic Microwave Background," Phys. Rev. **D88**, 063502 (2013), arXiv:1306.0563 [astro-ph.CO] .

[81] C. Evoli, A. Mesinger, and A. Ferrara, "Unveiling the nature of dark matter with high redshift 21 cm line experiments," JCAP **1411**, 024 (2014), arXiv:1408.1109 [astro-ph.HE] .

[82] T. R. Slatyer, "Indirect Dark Matter Signatures in the Cosmic Dark Ages II. Ionization, Heating and Photon Production from Arbitrary Energy Injections," Phys. Rev. **D93**, 023521 (2016), arXiv:1506.03812 [astro-ph.CO] .

[83] T. Kanzaki and M. Kawasaki, "Electron and Photon Energy Deposition in Universe," Phys. Rev. **D78**, 103004 (2008), arXiv:0805.3969 [astro-ph] .

[84] M. Kawasaki, K. Nakayama, and T. Sekiguchi, "CMB Constraint on Dark Matter Annihilation after Planck 2015," Phys. Lett. **B756**, 212–215 (2016), arXiv:1512.08015 [astro-ph.CO] .

[85] A. Ferrara, M. Valdés, N. Yoshida, and C. Evoli, "Energy deposition by weakly interacting massive particles: a comprehensive study," Monthly Notices of the Royal Astronomical Society **422**, 420–433 (2012),

http://oup.prod.sis.lan/mnras/article-pdf/422/1/420/18597009/mnras0422-0420.pdf .

[86] S. Furlanetto, S. P. Oh, and F. Briggs, "Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe," Phys. Rept. **433**, 181–301 (2006), arXiv:astro-ph/0608032 [astro-ph] .

[87] S. H. Hansen and Z. Haiman, "Do we need stars to reionize the universe at high redshifts? Early reionization by decaying heavy sterile neutrinos," Astrophys. J. **600**, 26–31 (2004), arXiv:astro-ph/0305126 [astro-ph] .

[88] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, "HeII->HeI Recombination of Primordial Helium Plasma Including the Effect of Neutral Hydrogen," Astron. Lett. **34**, 725–738 (2008), arXiv:0812.3067 [astro-ph] .

[89] E. E. Kholupenko, A. V. Ivanchik, and D. A. Varshalovich, "Rapid HeII->HeI recombination and radiation concerned with this process," Mon. Not. Roy. Astron. Soc. **378**, 39 (2007), arXiv:astro-ph/0703438 [astro-ph] .

[90] E. Puchwein, F. Haardt, M. G. Haehnelt, and P. Madau, "Consistent modelling of the meta-galactic UV background and the thermal/ionization history of the intergalactic medium," (2018), arXiv:1801.04931 [astro-ph.GA] .

[91] F. Haardt and P. Madau, "Radiative transfer in a clumpy universe: IV. New synthesis models of the cosmic UV/X-ray background," Astrophys. J. **746**, 125 (2012), arXiv:1105.2039 [astro-ph.CO] .

[92] J. S. Bolton and M. G. Haehnelt, "The nature and evolution of the highly ionized near-zones in the absorption spectra of z =6 quasars," Mon. Not. Roy. Astron. Soc. **374**, 493–514 (2007), arXiv:astro-ph/0607331 [astro-ph] .

[93] P. Ciafaloni, D. Comelli, A. Riotto, F. Sala, A. Strumia, and A. Urbano, "Weak Corrections are Relevant for Dark Matter Indirect Detection," JCAP **1103**, 019 (2011), arXiv:1009.0224 [hep-ph] .

[94] M. Tanabashi *et al.* (Particle Data Group), "Review of Particle Physics," Phys. Rev. **D98**, 030001 (2018).

[95] F. Calore, I. Cholis, C. McCabe, and C. Weniger, "A Tale of Tails: Dark Matter Interpretations of the Fermi GeV Excess in Light of Background Model Systematics," Phys. Rev. **D91**, 063003 (2015), arXiv:1411.4647 [hep-ph] .

[96] M.-Y. Cui, Q. Yuan, Y.-L. S. Tsai, and Y.-Z. Fan, "Possible dark matter annihilation signal in the AMS-02 antiproton data," Phys. Rev. Lett. **118**, 191101 (2017), arXiv:1610.03840 [astro-ph.HE] .

[97] A. Cuoco, M. Krämer, and M. Korsmeier, "Novel Dark Matter Constraints from Antiprotons in Light of AMS-02," Phys. Rev. Lett. **118**, 191102 (2017), arXiv:1610.03071 [astro-ph.HE] .

[98] T. R. Slatyer, "Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s-wave dark matter annihilation from Planck results," Phys. Rev. **D93**, 023527 (2016), arXiv:1506.03811 [hep-ph] .

[99] J. R. Pritchard and A. Loeb, "21-cm cosmology," Rept. Prog. Phys. **75**, 086901 (2012), arXiv:1109.6012 [astro-ph.CO] .

[100] J. D. Bowman, A. E. E. Rogers, R. A. Monsalve, T. J. Mozdzen, and N. Mahesh, "An absorption profile centred at 78 megahertz in the sky-averaged spectrum," Nature **555**, 67–70 (2018).

[101] A. Lewis, "Cosmological parameters from WMAP 5-year temperature maps," Phys. Rev. **D78**, 023002 (2008), arXiv:0804.3865 [astro-ph] .

[102] R. Adam *et al.* (Planck), "Planck intermediate results. XLVII. Planck constraints on reionization history," Astron. Astrophys. **596**, A108 (2016), arXiv:1605.03507 [astro-ph.CO] .

[103] Y. Ali-Haïmoud, J. Chluba, and M. Kamionkowski, "Constraints on Dark Matter Interactions with Standard Model Particles from Cosmic Microwave Background Spectral Distortions," Phys. Rev. Lett. **115**, 071304 (2015), arXiv:1506.04745 [astro-ph.CO] .

[104] S. K. Acharya and R. Khatri, "Rich structure of non-thermal relativistic CMB spectral distortions from high energy particle cascades at redshifts $z \lesssim 2 \times 10^5$," Phys. Rev. **D99**, 043520 (2019), arXiv:1808.02897 [astro-ph.CO] .

[105] S. R. Furlanetto, S. P. Oh, and E. Pierpaoli, "The Effects of Dark Matter Decay and Annihilation on the High-Redshift 21 cm Background," Phys. Rev. **D74**, 103502 (2006), arXiv:astro-ph/0608385 [astro-ph] .

[106] E. Ripamonti, M. Mapelli, and A. Ferrara, "The impact of dark matter decays and annihilations on the formation of the first structures," Monthly Notices of the Royal Astronomical Society **375**, 1399–1408 (2007), http://oup.prod.sis.lan/mnras/article-pdf/375/4/1399/18675257/mnras0375-1399.pdf .

[107] R. J. Smith, F. Iocco, S. C. O. Glover, D. R. G. Schleicher, R. S. Klessen, S. Hirano, and N. Yoshida, "Weakly Interacting Massive Particle Dark Matter and First Stars: Suppression of Fragmentation in Primordial Star Formation," Astrophys. J. **761**, 154 (2012), arXiv:1210.1582 .

[108] A. Stacy, A. H. Pawlik, V. Bromm, and A. Loeb, "The Mutual Interaction Between Population III Stars and Self-Annihilating Dark Matter," Mon. Not. Roy. Astron. Soc. **441**, 822–836 (2014), arXiv:1312.3117 [astro-ph.CO] .

[109] J. A. Adams, S. Sarkar, and D. Sciama, "CMB anisotropy in the decaying neutrino cosmology," Mon. Not. Roy. Astron. Soc. **301**, 210–214 (1998), arXiv:astro-ph/9805108 .

[110] X.-L. Chen and M. Kamionkowski, "Particle decays during the cosmic dark ages," Phys. Rev. D **70**, 043502 (2004), arXiv:astro-ph/0310473 .

[111] N. Padmanabhan and D. P. Finkbeiner, "Detecting dark matter annihilation with CMB polarization: Signatures and experimental prospects," Phys. Rev. D **72**, 023508 (2005), arXiv:astro-ph/0503486 .

[112] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, "CMB constraints on Dark Matter models with large annihilation cross-section," Phys. Rev. D **80**, 023505 (2009), arXiv:0905.0003 [astro-ph.CO] .

[113] J. Hisano, M. Kawasaki, K. Kohri, T. Moroi, K. Nakayama, and T. Sekiguchi, "Cosmological constraints on dark matter models with velocity-dependent annihilation cross section," Phys. Rev. D **83**, 123511 (2011), arXiv:1102.4658 [hep-ph] .

[114] G. Hutsi, J. Chluba, A. Hektor, and M. Raidal, "WMAP7 and future CMB constraints on annihilating dark matter: implications on GeV-scale WIMPs," Astron. Astrophys. **535**, A26 (2011), arXiv:1103.2766 [astro-ph.CO] .

[115] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, "Updated CMB constraints on Dark Matter annihilation cross-sections," Phys. Rev. D **84**, 027302 (2011), arXiv:1106.1528 [astro-ph.CO] .

[116] D. P. Finkbeiner, S. Galli, T. Lin, and T. R. Slatyer, "Searching for Dark Matter in the CMB: A Compact Parameterization of Energy Injection from New Physics," Phys. Rev. D **85**, 043522 (2012), arXiv:1109.6322 [astro-ph.CO] .

[117] M. S. Madhavacheril, N. Sehgal, and T. R. Slatyer, "Current Dark Matter Annihilation Constraints from CMB and Low-Redshift Data," Phys. Rev. D **89**, 103508 (2014), arXiv:1310.3815 [astro-ph.CO] .

[118] T. R. Slatyer, "Indirect Dark Matter Signatures in the Cosmic Dark Ages II. Ionization, Heating and Photon Production from Arbitrary Energy Injections," Phys. Rev. **D93**, 023521 (2016), arXiv:1506.03812 [astro-ph.CO] .

[119] L. Zhang, X. Chen, M. Kamionkowski, Z.-g. Si, and Z. Zheng, "Constraints on radiative dark-matter decay from the cosmic microwave background," Phys. Rev. D **76**, 061301 (2007), arXiv:0704.2444 [astro-ph] .

[120] S. K. Acharya and R. Khatri, "CMB anisotropy and BBN constraints on pre-recombination decay of dark matter to visible particles," JCAP **12**, 046 (2019), arXiv:1910.06272 [astro-ph.CO] .

[121] N. Aghanim *et al.* (Planck), "Planck 2018 results. VI. Cosmological parameters," (2018), arXiv:1807.06209 [astro-ph.CO] .

[122] G. D'Amico, P. Panci, and A. Strumia, "Bounds on Dark Matter annihilations from 21 cm data," (2018), arXiv:1803.03629 [astro-ph.CO] .

[123] K. Cheung, J.-L. Kuo, K.-W. Ng, and Y.-L. S. Tsai, "The impact of EDGES 21-cm data on dark matter interactions," (2018), arXiv:1803.09398 [astro-ph.CO] .

[124] A. Mitridate and A. Podo, "Bounds on Dark Matter decay from 21 cm line," (2018), arXiv:1803.11169 [hep-ph] .

[125] J. Schaye, T. Theuns, M. Rauch, G. Efstathiou, and W. L. Sargent, "The Thermal history of the intergalactic medium," Mon. Not. Roy. Astron. Soc. **318**, 817 (2000), arXiv:astro-ph/9912432 .

[126] G. D. Becker, J. S. Bolton, M. G. Haehnelt, and W. L. Sargent, "Detection of Extended He II Reionization in the Temperature Evolution of the Intergalactic Medium," Mon. Not. Roy. Astron. Soc. **410**, 1096 (2011), arXiv:1008.2622 [astro-ph.CO] .

[127] J. S. Bolton, G. D. Becker, S. Raskutti, J. B. Wyithe, M. G. Haehnelt, and W. L. Sargent, "Improved measurements of the intergalactic medium temperature around quasars: possible evidence for the initial stages of He-II reionisation at z˜6," Mon. Not. Roy. Astron. Soc. **419**, 2880–2892 (2012), arXiv:1110.0539 [astro-ph.CO] .

[128] J. S. Bolton, G. D. Becker, J. B. Wyithe, M. G. Haehnelt, and W. L. Sargent, "A first direct measurement of the intergalactic medium temperature around a quasar at z=6," Mon. Not. Roy. Astron. Soc. **406**, 612 (2010), arXiv:1001.3415 [astro-ph.CO] .

[129] M. Cirelli, F. Iocco, and P. Panci, "Constraints on Dark Matter annihilations from reionization and heating of the intergalactic gas," JCAP **0910**, 009 (2009), arXiv:0907.0719 [astro-ph.CO] .

[130] H. Liu, T. R. Slatyer, and J. Zavala, "Contributions to cosmic reionization from dark matter annihilation and decay," Phys. Rev. **D94**, 063507 (2016), arXiv:1604.02457 [astro-ph.CO] .

[131] S. J. Witte, S. Rosauro-Alcaraz, S. D. McDermott, and V. Poulin, "Dark Photon Dark Matter in the Presence of Inhomogeneous Structure," (2020), arXiv:2003.13698 [astro-ph.CO] .

[132] A. Caputo, H. Liu, S. Mishra-Sharma, and J. T. Ruderman, "Dark Photon Oscillations in Our Inhomogeneous Universe," (2020), arXiv:2002.05165 [astro-ph.CO] .

[133] A. Caputo, H. Liu, S. Mishra-Sharma, and J. T. Ruderman, "Modeling Dark Photon Oscillations in Our Inhomogeneous Universe," (2020), arXiv:2004.06733 [astro-ph.CO] .

[134] J. B. Muñoz and A. Loeb, "Constraints on Dark Matter-Baryon Scattering from the Temperature Evolution of the Intergalactic Medium," JCAP **11**, 043 (2017), arXiv:1708.08923 [astro-ph.CO] .

[135] H. Kim, "A constraint on light primordial black holes from the interstellar medium temperature," (2020), arXiv:2007.07739 [hep-ph] .

[136] W. Hu and G. P. Holder, "Model - independent reionization observables in the CMB," Phys. Rev. **D68**, 023001 (2003), arXiv:astro-ph/0303400 [astro-ph] .

[137] M. Millea and F. Bouchet, "Cosmic Microwave Background Constraints in Light of Priors Over Reionization Histories," Astron. Astrophys. **617**, A96 (2018), arXiv:1804.08476 [astro-ph.CO] .

[138] M. Walther, J. F. Hennawi, H. Hiss, J. Oñorbe, K.-G. Lee, A. Rorai, and J. O'Meara, "A New Precision Measurement of the Small-scale Line-of-sight Power Spectrum of the Lyα Forest," Astrophys. J. **852**, 22 (2018), arXiv:1709.07354 [astro-ph.CO] .

[139] N. Palanque-Delabrouille *et al.*, "The one-dimensional Ly-alpha forest power spectrum from BOSS," Astron. Astrophys. **559**, A85 (2013), arXiv:1306.5896 [astro-ph.CO] .

[140] J. O'Meara, N. Lehner, J. Howk, J. Prochaska, A. Fox, M. Swain, C. Gelino, G. Berriman, and H. Tran, "The First Data Release of the KODIAQ Survey," Astron. J. **150**, 111 (2015), arXiv:1505.03529 [astro-ph.CO] .

[141] J. M. O'Meara, N. Lehner, J. C. Howk, J. X. Prochaska, A. J. Fox, M. S. Peeples, J. Tumlinson, and B. W. O'Shea, "The Second Data Release of the KODIAQ Survey," Astron. J. **154**, 114 (2017), arXiv:1707.07905 [astro-ph.GA] .

[142] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, "Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data," Phys. Rev. **D88**, 043502 (2013), arXiv:1306.2314 [astro-ph.CO] .

[143] V. Iršič *et al.*, "The Lyman α forest power spectrum from the XQ-100 Legacy Survey," Mon. Not. Roy. Astron. Soc. **466**, 4332–4345 (2017), arXiv:1702.01761 [astro-ph.CO] .

[144] M. Walther, J. Oñorbe, J. F. Hennawi, and Z. Lukić, "New Constraints on IGM Thermal Evolution from the Lyα Forest Power Spectrum," Astrophys. J. **872**, 13 (2019), arXiv:1808.04367 [astro-ph.CO] .

[145] E. Boera, M. T. Murphy, G. D. Becker, and J. S. Bolton, "The thermal history of the intergalactic medium down to redshift z = 1.5: a new curvature measurement," Mon. Not. Roy. Astron. Soc. **441**, 1916–1933 (2014), arXiv:1404.1083 [astro-ph.CO] .

[146] P. Gaikwad *et al.*, "Probing the thermal state of the intergalactic medium at $z > 5$ with the transmission spikes in high-resolution Lyα forest spectra," Mon. Not. Roy. Astron. Soc. **494**, 5091–5109 (2020), arXiv:2001.10018 [astro-ph.CO] .

[147] R. Essig, E. Kuflik, S. D. McDermott, T. Volansky, and K. M. Zurek, "Constraining Light Dark Matter with Diffuse X-Ray and Gamma-Ray Observations," JHEP **11**, 193 (2013), arXiv:1309.4091 [hep-ph] .

[148] T. Cohen, K. Murase, N. L. Rodd, B. R. Safdi, and Y. Soreq, " -ray Constraints on Decaying Dark Matter and Implications for IceCube," Phys. Rev. Lett. **119**, 021102 (2017), arXiv:1612.05638 [hep-ph] .

[149] M. Boudaud, J. Lavalle, and P. Salati, "Novel cosmic-ray electron and positron constraints on MeV dark matter particles," Phys. Rev. Lett. **119**, 021103 (2017), arXiv:1612.07698 [astro-ph.HE] .

[150] M. Boudaud, T. Lacroix, M. Stref, and J. Lavalle, "Robust cosmic-ray constraints on p-wave annihilating MeV dark matter," Phys. Rev. D **99**, 061302 (2019), arXiv:1810.01680 [astro-ph.HE] .

[151] M. McQuinn, "The Evolution of the Intergalactic Medium," Ann. Rev. Astron. Astrophys. **54**, 313–362 (2016), arXiv:1512.00086 [astro-ph.CO] .

[152] J. Oñorbe, J. F. Hennawi, and Z. Lukić, "Self-Consistent Modeling of Reionization in Cosmological Hydrodynamical Simulations," Astrophys. J. **837**, 106 (2017), arXiv:1607.04218 [astro-ph.CO] .

[153] M. McQuinn and P. U. Sanderbeck, "On the intergalactic temperature–density relation," Mon. Not. Roy. Astron. Soc. **456**, 47–54 (2016), arXiv:1505.07875 [astro-ph.CO] .

[154] J. Miralda-Escudé and M. J. Rees, "Reionization and thermal evolution of a photoionized intergalactic medium." Mon. Not. Roy. Astron. Soc. **266**, 343–352 (1994).

[155] M. McQuinn, "Constraints on X-ray Emissions from the Reionization Era," Mon. Not. Roy. Astron. Soc. **426**, 1349–1360 (2012), arXiv:1206.1335 [astro-ph.CO] .

[156] P. R. U. Sanderbeck, A. D'Aloisio, and M. J. McQuinn, "Models of the Thermal Evolution of the Intergalactic Medium After Reionization," Mon. Not. Roy. Astron. Soc. **460**, 1885–1897 (2016), arXiv:1511.05992 [astro-ph.CO] .

[157] E. Puchwein, J. S. Bolton, M. G. Haehnelt, P. Madau, G. D. Becker, and F. Haardt, "The photoheating of the intergalactic medium in synthesis models of the UV background," Mon. Not. Roy. Astron. Soc. **450**, 4081–4097 (2015), arXiv:1410.1531 [astro-ph.CO] .

[158] R. Bouwens, G. Illingworth, P. Oesch, J. Caruana, B. Holwerda, R. Smit, and S. Wilkins, "Reionization after Planck: The Derived Growth of the Cosmic Ionizing Emissivity now matches the Growth of the Galaxy UV Luminosity Density," Astrophys. J. **811**, 140 (2015), arXiv:1503.08228 [astro-ph.CO] .

[159] A. Massari, E. Izaguirre, R. Essig, A. Albert, E. Bloom, and G. A. Gómez-Vargas, "Strong Optimized Conservative *Fermi*-LAT Constraints on Dark Matter Models from the Inclusive Photon Spectrum," Phys. Rev. D **91**, 083539 (2015), arXiv:1503.07169 [hep-ph] .

[160] M. Cirelli, N. Fornengo, B. J. Kavanagh, and E. Pinetti, "Integral X-ray constraints on sub-GeV Dark Matter," (2020), arXiv:2007.11493 [hep-ph] .

[161] Y. Zhao, X.-J. Bi, H.-Y. Jia, P.-F. Yin, and F.-R. Zhu, "Constraint on the velocity dependent dark matter annihilation cross section from Fermi-LAT observations of dwarf galaxies," Phys. Rev. D **93**, 083513 (2016), arXiv:1601.02181 [astro-ph.HE] .

[162] S. B. Gudnason, C. Kouvaris, and F. Sannino, "Dark Matter from new Technicolor Theories," Phys. Rev. D **74**, 095008 (2006), arXiv:hep-ph/0608055 .

[163] D. S. M. Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, "Composite Inelastic Dark Matter," Phys. Lett. B **692**, 323–326 (2010), arXiv:0903.3945 [hep-ph] .

[164] C. Kilic, T. Okui, and R. Sundrum, "Vectorlike Confinement at the LHC," JHEP **02**, 018 (2010), arXiv:0906.0577 [hep-ph] .

[165] T. Hambye and M. H. G. Tytgat, "Confined hidden vector dark matter," Phys. Lett. B **683**, 39–41 (2010), arXiv:0907.1007 [hep-ph] .

[166] G. D. Kribs, T. S. Roy, J. Terning, and K. M. Zurek, "Quirky Composite Dark Matter," Phys. Rev. D **81**, 095001 (2010), arXiv:0909.2034 [hep-ph] .

[167] D. Spier Moreira Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, "The Cosmology of Composite Inelastic Dark Matter," JHEP **06**, 113 (2010), arXiv:1003.4729 [hep-ph] .

[168] Y. Bai and R. J. Hill, "Weakly Interacting Stable Pions," Phys. Rev. D **82**, 111701 (2010), arXiv:1005.0008 [hep-ph] .

[169] J. L. Feng and Y. Shadmi, "WIMPless Dark Matter from Non-Abelian Hidden Sectors with Anomaly-Mediated Supersymmetry Breaking," Phys. Rev. D **83**, 095011 (2011), arXiv:1102.0282 [hep-ph] .

[170] R. Fok and G. D. Kribs, "Chiral Quirkonium Decays," Phys. Rev. D **84**, 035001 (2011), arXiv:1106.3101 [hep-ph] .

[171] R. Lewis, C. Pica, and F. Sannino, "Light Asymmetric Dark Matter on the Lattice: SU(2) Technicolor with Two Fundamental Flavors," Phys. Rev. D **85**, 014504 (2012), arXiv:1109.3513 [hep-ph] .

[172] M. Frigerio, A. Pomarol, F. Riva, and A. Urbano, "Composite Scalar Dark Matter," JHEP **07**, 015 (2012), arXiv:1204.2808 [hep-ph] .

[173] M. R. Buckley and E. T. Neil, "Thermal dark matter from a confining sector," Phys. Rev. D **87**, 043510 (2013), arXiv:1209.6054 [hep-ph] .

[174] T. Appelquist *et al.* (Lattice Strong Dynamics (LSD)), "Lattice Calculation of Composite Dark Matter Form Factors," Phys. Rev. D **88**, 014502 (2013), arXiv:1301.1693 [hep-ph] .

[175] S. Bhattacharya, B. Melić, and J. Wudka, "Pionic Dark Matter," JHEP **02**, 115 (2014), arXiv:1307.2647 [hep-ph] .

[176] J. M. Cline, Z. Liu, G. Moore, and W. Xue, "Composite strongly interacting dark matter," Phys. Rev. D **90**, 015023 (2014), arXiv:1312.3325 [hep-ph] .

[177] K. K. Boddy, J. L. Feng, M. Kaplinghat, and T. M. P. Tait, "Self-Interacting Dark Matter from a Non-Abelian Hidden Sector," Phys. Rev. D **89**, 115017 (2014), arXiv:1402.3629 [hep-ph] .

[178] K. K. Boddy, J. L. Feng, M. Kaplinghat, Y. Shadmi, and T. M. P. Tait, "Strongly interacting dark matter: Self-interactions and keV lines," Phys. Rev. D **90**, 095016 (2014), arXiv:1408.6532 [hep-ph] .

[179] T. Appelquist *et al.*, "Stealth Dark Matter: Dark scalar baryons through the Higgs portal," Phys. Rev. D **92**, 075030 (2015), arXiv:1503.04203 [hep-ph] .

[180] O. Antipin, M. Redi, A. Strumia, and E. Vigiani, "Accidental Composite Dark Matter," JHEP **07**, 039 (2015), arXiv:1503.08749 [hep-ph] .

[181] A. Soni and Y. Zhang, "Hidden SU(N) Glueball Dark Matter," Phys. Rev. D **93**, 115025 (2016), arXiv:1602.00714 [hep-ph] .

[182] G. D. Kribs and E. T. Neil, "Review of strongly-coupled composite dark matter models and lattice simulations," Int. J. Mod. Phys. A **31**, 1643004 (2016), arXiv:1604.04627 [hep-ph] .

[183] K. Harigaya, M. Ibe, K. Kaneta, W. Nakano, and M. Suzuki, "Thermal Relic Dark Matter Beyond the Unitarity Limit," JHEP **08**, 151 (2016), arXiv:1606.00159 [hep-ph] .

[184] A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, "Dark Matter as a weakly coupled Dark Baryon," JHEP **10**, 210 (2017), arXiv:1707.05380 [hep-ph] .

[185] V. De Luca, A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, "Colored Dark Matter," Phys. Rev. D **97**, 115024 (2018), arXiv:1801.01135 [hep-ph] .

[186] R. Contino, A. Mitridate, A. Podo, and M. Redi, "Gluequark Dark Matter," JHEP **02**, 187 (2019), arXiv:1811.06975 [hep-ph] .

[187] C. Gross, A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, "Cosmological Abundance of Colored Relics," Phys. Rev. D **99**, 016024 (2019), arXiv:1811.08418 [hep-ph] .

[188] V. Beylin, M. Y. Khlopov, V. Kuksa, and N. Volchanskiy, "Hadronic and Hadron-Like Physics of Dark Matter," Symmetry **11**, 587 (2019), arXiv:1904.12013 [hep-ph] .

[189] N. A. Dondi, F. Sannino, and J. Smirnov, "Thermal history of composite dark matter," Phys. Rev. D **101**, 103010 (2020), arXiv:1905.08810 [hep-ph] .

[190] D. Buttazzo, L. Di Luzio, P. Ghorbani, C. Gross, G. Landini, A. Strumia, D. Teresi, and J.-W. Wang, "Scalar gauge dynamics and Dark Matter," JHEP **01**, 130 (2020), arXiv:1911.04502 [hep-ph] .

[191] G. Landini and J.-W. Wang, "Dark Matter in scalar Sp($\mathcal{N}$) gauge dynamics," JHEP **06**, 167 (2020), arXiv:2004.03299 [hep-ph] .

[192] R. C. Brower *et al.* (Lattice Strong Dynamics), "Stealth dark matter confinement transition and gravitational waves," Phys. Rev. D **103**, 014505 (2021), arXiv:2006.16429 [hep-lat] .

[193] R. Contino, A. Podo, and F. Revello, "Composite Dark Matter from Strongly-Interacting Chiral Dynamics," (2020), arXiv:2008.10607 [hep-ph] .

[194] V. Beylin, M. Khlopov, V. Kuksa, and N. Volchanskiy, "New physics of strong interaction and Dark Universe," Universe **6**, 196 (2020), arXiv:2010.13678 [hep-ph] .

[195] S. M. Barr, "Baryogenesis, sphalerons, and the cogeneration of dark matter," Phys. Rev. D **44**, 3062–3066 (1991).

[196] B. Svetitsky and L. G. Yaffe, "Critical behavior at finite-temperature confinement transitions," Nuclear Physics B **210**, 423 – 447 (1982).

[197] C. Alexandrou, A. Boriçi, A. Feo, P. de Forcrand, A. Galli, F. Jegerlehner, and T. Takaishi, "Deconfinement phase transition in one-flavor qcd," Phys. Rev. D **60**, 034504 (1999).

[198] Y. Aoki, G. Endrodi, Z. Fodor, S. Katz, and K. Szabo, "The Order of the quantum chromodynamics transition predicted by the standard model of particle physics," Nature **443**, 675–678 (2006), arXiv:hep-lat/0611014 .

[199] H. Saito, S. Ejiri, S. Aoki, T. Hatsuda, K. Kanaya, Y. Maezawa, H. Ohno, and T. Umeda (WHOT-QCD), "Phase structure of finite temperature QCD in the heavy quark region," Phys. Rev. D **84**, 054502 (2011), [Erratum: Phys.Rev.D 85, 079902 (2012)], arXiv:1106.0974 [hep-lat] .

[200] B. Lucini, M. Teper, and U. Wenger, "The High temperature phase transition in SU(N) gauge theories," JHEP **01**, 061 (2004), arXiv:hep-lat/0307017 .

[201] M. J. Baker, J. Kopp, and A. J. Long, "Filtered Dark Matter at a First Order Phase Transition," Phys. Rev. Lett. **125**, 151102 (2020), arXiv:1912.02830 [hep-ph] .

[202] D. Chway, T. H. Jung, and C. S. Shin, "Dark matter filtering-out effect during a first-order phase transition," Phys. Rev. D **101**, 095019 (2020), arXiv:1912.04238 [hep-ph] .

[203] H. Davoudiasl and G. Mohlabeng, "Getting a THUMP from a WIMP," JHEP **04**, 177 (2020), arXiv:1912.05572 [hep-ph] .

[204] W. Chao, X.-F. Li, and L. Wang, "Filtered pseudo-scalar dark matter and gravitational waves from first order phase transition," (2020), arXiv:2012.15113 [hep-ph] .

[205] T. Konstandin and G. Servant, "Cosmological Consequences of Nearly Conformal Dynamics at the TeV scale," JCAP **12**, 009 (2011), arXiv:1104.4791 [hep-ph] .

[206] T. Hambye, A. Strumia, and D. Teresi, "Super-cool Dark Matter," JHEP **08**, 188 (2018), arXiv:1805.01473 [hep-ph] .

[207] P. Baratella, A. Pomarol, and F. Rompineve, "The Supercooled Universe," JHEP **03**, 100 (2019), arXiv:1812.06996 [hep-ph] .

[208] I. Baldes, Y. Gouttenoire, and F. Sala, "String Fragmentation in Supercooled Confinement and implications for Dark Matter," (2020), arXiv:2007.08440 [hep-ph] .

[209] E. Witten, "Cosmic Separation of Phases," Phys. Rev. D **30**, 272–285 (1984).

[210] K. Griest and M. Kamionkowski, "Unitarity Limits on the Mass and Radius of Dark Matter Particles," Phys. Rev. Lett. **64**, 615 (1990).

[211] B. von Harling and K. Petraki, "Bound-state formation for thermal relic dark matter and unitarity," JCAP **12**, 033 (2014), arXiv:1407.7874 [hep-ph] .

[212] J. Smirnov and J. F. Beacom, "TeV-Scale Thermal WIMPs: Unitarity and its Consequences," Phys. Rev. D **100**, 043029 (2019), arXiv:1904.11503 [hep-ph] .

[213] C. J. Moore, R. H. Cole, and C. P. L. Berry, "Gravitational-wave sensitivity curves," Class. Quant. Grav. **32**, 015014 (2015), arXiv:1408.0740 [gr-qc] .

[214] G. M. Pelaggi, A. D. Plascencia, A. Salvio, F. Sannino, J. Smirnov, and A. Strumia, "Asymptotically Safe Standard Model Extensions?" Phys. Rev. D **97**, 095013 (2018), arXiv:1708.00437 [hep-ph] .

[215] A. Patel, "A Flux Tube Model of the Finite Temperature Deconfining Transition in QCD," Nucl. Phys. B **243**, 411–422 (1984).

[216] S. Balaji, M. Spannowsky, and C. Tamarit, "Cosmological bubble friction in local equilibrium," (2020), arXiv:2010.08013 [hep-ph] .

[217] J. Kang and M. A. Luty, "Macroscopic Strings and 'Quirks' at Colliders," JHEP **11**, 065 (2009), arXiv:0805.4642 [hep-ph] .

[218] J. Schwinger, "On gauge invariance and vacuum polarization," Phys. Rev. **82**, 664–679 (1951).

[219] F. Karsch, "Lattice results on QCD thermodynamics," Nucl. Phys. A **698**, 199–208 (2002), arXiv:hep-ph/0103314 .

[220] J. Ignatius, K. Kajantie, H. Kurki-Suonio, and M. Laine, "The growth of bubbles in cosmological phase transitions," Phys. Rev. D **49**, 3854–3868 (1994), arXiv:astro-ph/9309059 .

[221] S. J. Huber and M. Sopena, "The bubble wall velocity in the minimal supersymmetric light stop scenario," Phys. Rev. D **85**, 103507 (2012), arXiv:1112.1888 [hep-ph] .

[222] J. Ellis, M. Lewicki, and J. M. No, "On the Maximal Strength of a First-Order Electroweak Phase Transition and its Gravitational Wave Signal," JCAP **04**, 003 (2019), arXiv:1809.08242 [hep-ph] .

[223] M. Redi and A. Tesi, "Cosmological Production of Dark Nuclei," JHEP **04**, 108 (2019), arXiv:1812.08784 [hep-ph] .

[224] Y. Bai, A. J. Long, and S. Lu, "Dark Quark Nuggets," Phys. Rev. D **99**, 055047 (2019), arXiv:1810.04360 [hep-ph] .

[225] C. Gross, G. Landini, A. Strumia, and D. Teresi, "Dark Matter as dark dwarfs and other macroscopic objects: multiverse relics?" (2021), arXiv:2105.02840 [hep-ph] .

[226] W.-C. Huang, M. Reichert, F. Sannino, and Z.-W. Wang, "Testing the Dark Confined Landscape: From Lattice to Gravitational Waves," (2020), arXiv:2012.11614 [hep-ph] .

[227] D. J. Weir, "Gravitational waves from a first order electroweak phase transition: a brief review," *Proceedings, Higgs cosmology: Theo Murphy meeting: Buckinghamshire, UK, March 27-28, 2017*, Phil. Trans. Roy. Soc. Lond. **A376**, 20170126 (2018), arXiv:1705.01783 [hep-ph] .

[228] C. Caprini *et al.*, "Science with the space-based interferometer eLISA. II: Gravitational waves from cosmological phase transitions," JCAP **1604**, 001 (2016), arXiv:1512.06239 [astro-ph.CO] .

[229] N. Seto, S. Kawamura, and T. Nakamura, "Possibility of direct measurement of the acceleration of the universe using 0.1-Hz band laser interferometer gravitational wave antenna in space," Phys. Rev. Lett. **87**, 221103 (2001), arXiv:astro-ph/0108011 .

[230] S. K. et al., "The japanese space gravitational wave antenna—DECIGO," Classical and Quantum Gravity **23**, S125–S131 (2006).

[231] G. Harry, P. Fritschel, D. Shaddock, W. Folkner, and E. Phinney, "Laser interferometry for the big bang observer," Class. Quant. Grav. **23**, 4887–4894 (2006), [Erratum: Class.Quant.Grav. 23, 7361 (2006)].

[232] M. Geller, A. Hook, R. Sundrum, and Y. Tsai, "Primordial Anisotropies in the Gravitational Wave Background from Cosmological Phase Transitions," Phys. Rev. Lett. **121**, 201303 (2018), arXiv:1803.10780 [hep-ph] .

[233] R. H. Cyburt, J. Ellis, B. D. Fields, F. Luo, K. A. Olive, and V. C. Spanos, "Nucleosynthesis Constraints on a Massive Gravitino in Neutralino Dark Matter Scenarios," JCAP **10**, 021 (2009), arXiv:0907.5003 [astro-ph.CO] .

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

[235] L. Forestell, D. E. Morrissey, and G. White, "Limits from BBN on Light Electromagnetic Decays," JHEP **01**, 074 (2019), arXiv:1809.01179 [hep-ph] .

[236] C. V. Cappiello, J. I. Collar, and J. F. Beacom, "New experimental constraints in a new landscape for composite dark matter," Phys. Rev. D **103**, 023019 (2021), arXiv:2008.10646 [hep-ex] .

[237] I. Goldman and S. Nussinov, "Weakly interacting massive particles and neutron stars," Phys. Rev. D **40**, 3221–3230 (1989).

[238] M. Baryakhtar, J. Bramante, S. W. Li, T. Linden, and N. Raj, "Dark Kinetic Heating of Neutron Stars and An Infrared Window On WIMPs, SIMPs, and Pure Higgsinos," Phys. Rev. Lett. **119**, 131801 (2017), arXiv:1704.01577 [hep-ph] .

[239] K. Van Tilburg, "Stellar Basins of Gravitationally Bound Particles," (2020), arXiv:2006.12431 [hep-ph] .

[240] R. K. Leane and J. Smirnov, "Exoplanets as New Sub-GeV Dark Matter Detectors," (2020), arXiv:2010.00015 [hep-ph] .

[241] R. K. Leane, K. C. Y. Ng, and J. F. Beacom, "Powerful Solar Signatures of Long-Lived Dark Mediators," Phys. Rev. D **95**, 123016 (2017), arXiv:1703.04629 [astro-ph.HE] .

[242] R. K. Leane, T. Linden, P. Mukhopadhyay, and N. Toro, "Celestial-Body Focused Dark Matter Annihilation Throughout the Galaxy," (2021), arXiv:2101.12213 [astro-ph.HE] .

[243] M. Markevitch, A. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker, "Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56," Astrophys. J. **606**, 819–824 (2004), arXiv:astro-ph/0309303 .

[244] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, "Hidden Charged Dark Matter," JCAP **07**, 004 (2009), arXiv:0905.3039 [hep-ph] .

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

[246] M. Vogelsberger, J. Zavala, and A. Loeb, "Subhaloes in self-interacting galactic dark matter haloes," Monthly Notices of the Royal Astronomical Society **423**, 3740–3752 (2012).

[247] J. Zavala, M. Vogelsberger, and M. G. Walker, "Constraining self-interacting dark matter with the Milky Way's dwarf spheroidals," Monthly Notices of the Royal Astronomical Society: Letters **431**, L20–L24 (2013).

[248] S. Tulin and H.-B. Yu, "Dark Matter Self-interactions and Small Scale Structure," Phys. Rept. **730**, 1–57 (2018), arXiv:1705.02358 [hep-ph] .

[249] K. Bondarenko, A. Sokolenko, A. Boyarsky, A. Robertson, D. Harvey, and Y. Revaz, "From dwarf galaxies to galaxy clusters: Self-Interacting Dark Matter over 7 orders of magnitude in halo mass," (2020), arXiv:2006.06623 [astro-ph.CO] .

[250] E. O. Nadler, V. Gluscevic, K. K. Boddy, and R. H. Wechsler, "Constraints on Dark Matter Microphysics from the Milky Way Satellite Population," Astrophys. J. Lett. **878**, 32 (2019), [Erratum: Astrophys.J.Lett. 897, L46 (2020), Erratum: Astrophys.J. 897, L46 (2020)], arXiv:1904.10000 [astro-ph.CO] .

[251] D. Gilman, S. Birrer, A. Nierenberg, T. Treu, X. Du, and A. Benson, "Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses," Mon. Not. Roy. Astron. Soc. **491**, 6077–6101 (2020), arXiv:1908.06983 [astro-ph.CO] .

[252] K. Schutz, "Subhalo mass function and ultralight bosonic dark matter," Phys. Rev. D **101**, 123026 (2020), arXiv:2001.05503 [astro-ph.CO] .

[253] S. Polikanov, C. S. Sastri, G. Herrmann, K. Lutzenkirchen, M. Overbeck, N. Trautmann, A. Breskin, R. Chechik, and Z. Frankel, "Search for supermassive nuclei in nature," Z. Phys. A **338**, 357–361 (1990).

[254] G. R. Blumenthal and R. J. Gould, "Bremsstrahlung, synchrotron radiation, and compton scattering of high-energy electrons traversing dilute gases," Rev. Mod. Phys. **42**, 237–270 (1970).

[255] F. C. Jones, "Calculated Spectrum of Inverse-Compton-Scattered Photons," Phys. Rev. **167**, 1159–1169 (1968).

[256] D. Fargion, R. V. Konoplich, and A. Salis, "Inverse Compton scattering on laser beam and monochromatic isotropic radiation," Z. Phys. **C74**, 571–576 (1997), arXiv:astro-ph/9606126 [astro-ph] .

[257] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith, "Cython: The best of both worlds," Computing in Science Engineering **13**, 31 –39 (2011).

[258] A. A. Zdziarski and P. Pjanka, "Compton scattering of blackbody photons by relativistic electrons," Mon. Not. Roy. Astron. Soc. **436**, 2950 (2013), arXiv:1307.6732 [astro-ph.HE] .

[259] A. Ore and J. L. Powell, "Three photon annihilation of an electron - positron pair," Phys. Rev. **75**, 1696–1699 (1949).

[260] M. Ackermann *et al.* (Fermi-LAT), "Fermi LAT Search for Dark Matter in Gamma-ray Lines and the Inclusive Photon Spectrum," Phys. Rev. D **86**, 022002 (2012), arXiv:1205.2739 [astro-ph.HE] .

[261] K. K. Boddy and J. Kumar, "Indirect Detection of Dark Matter Using MeV-Range Gamma-Ray Telescopes," Phys. Rev. D **92**, 023533 (2015), arXiv:1504.04024 [astro-ph.CO] .

[262] S. Archambault *et al.* (VERITAS), "Dark Matter Constraints from a Joint Analysis of Dwarf Spheroidal Galaxy Observations with VERITAS," Phys. Rev. D **95**, 082001 (2017), arXiv:1703.04937 [astro-ph.HE] .

[263] H. Abdallah *et al.* (HESS), "Search for γ-Ray Line Signals from Dark Matter Annihilations in the Inner Galactic Halo from 10 Years of Observations with H.E.S.S." Phys. Rev. Lett. **120**, 201101 (2018), arXiv:1805.05741 [astro-ph.HE] .

[264] V. Acciari *et al.* (MAGIC), "Constraining Dark Matter lifetime with a deep gamma-ray survey of the Perseus Galaxy Cluster with MAGIC," Phys. Dark Univ. **22**, 38–47 (2018), arXiv:1806.11063 [astro-ph.HE] .

[265] L. Landau and E. Lifshitz, *Statistical Physics, Part 1*, Course of Theoretical Physics, Vol. 5 (Pergamon Press, Oxford, 1980).

[266] L. Landau and E. Lifshitz, *Physical Kinetics*, Course of Theoretical Physics, Vol. 10 (Pergamon Press, Oxford, 1981).

[267] V. A. Rubakov and D. S. Gorbunov, *Introduction to the Theory of the Early Universe: Hot big bang theory* (World Scientific, Singapore, 2017).

[268] M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, 2007).

[269] T. Binder, K. Mukaida, and K. Petraki, "Rapid bound-state formation of Dark Matter in the Early Universe," Phys. Rev. Lett. **124**, 161102 (2020), arXiv:1910.11288 [hep-ph] .

[270] T. Binder, B. Blobel, J. Harz, and K. Mukaida, "Dark matter bound-state formation at higher order: a non-equilibrium quantum field theory approach," JHEP **09**, 086 (2020), arXiv:2002.07145 [hep-ph] .

[271] M. Geller, S. Iwamoto, G. Lee, Y. Shadmi, and O. Telem, "Dark quarkonium formation in the early universe," JHEP **06**, 135 (2018), arXiv:1802.07720 [hep-ph] .