Searching for the invisible: how dark forces shape our Universe

by

Katharine Schutz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Hitoshi Murayama, Chair
Professor Lawrence Hall
Professor Chung-Pei Ma

Spring 2019

Searching for the invisible: how dark forces shape our Universe

Abstract

Searching for the invisible: how dark forces shape our Universe

by

Katharine Schutz

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Hitoshi Murayama, Chair

Astrophysical observations on a wide range of scales indicate that the majority of matter in our Universe seems to be approximately inert and non-luminous. The existence of this dark matter implies the existence of an undiscovered particle, since there is no viable dark matter candidate within the Standard Model. Many terrestrial searches for dark matter particles are underway; however, there is no evidence to date that the dark matter interacts with particles in the Standard Model except through gravity. It highly conceivable that the dark matter exists as part of a rich hidden sector with diverse matter content and its own dark forces (in analogy to the Standard Model) which would imply that terrestrial searches may not pose an optimal path to discovering dark matter. Instead, observing astrophysical systems — where dark matter is known to be present through its gravitational influence — would be the best available way to test theories of dark matter where dark forces play a role in altering the properties of those systems. The complementarity of observing various astrophysical systems is a powerful asset for exploring the physics of dark sectors: one can explore broad classes of theories with dark forces in different environments, on different length scales, and at different epochs in the history of our Universe. This dissertation explores several scenarios where astrophysical observations inform our understanding of dark forces in a way that would not necessarily be possible on Earth. In particular, we consider dark sector energy dissipation, dark matter self-interaction, and the early Universe production of dark matter through dark channels. We propagate the implications of these effects for stars, supernovae, the Milky Way stellar disk, dwarf galaxies, galaxy clusters, large-scale structure, the epoch of reionization, and the cosmic microwave background.

For Adrian Liu, my dearest and most loyal friend.

Acknowledgments

This work would not have been possible without the generous support of the National Science Foundation and the Hertz Foundation.

I thank my advisor, Hitoshi Murayama, as well as Tracy Slatyer and Tongyan Lin who acted *in loco praeceptoris* on multiple occaisions. I also thank my thesis committee members, Lawrence Hall and Chung-Pei Ma, and qualifying exam committee member Matt Pyle. These individuals played a critical role in my pursuit of a doctoral degree.

I acknowledge the additional collaborators on the work appearing in this dissertaion: Cora Dvorkin, Yonit Hochberg, Eric Kuflik, Robert McGehee, Benjamin R. Safdi, Mark Vogelsberger, Chih-Liang Wu, and Jesús Zavala.

The work in this dissertation was improved through conversations with Masha Baryakhtar, Asher Berlin, Ana Bonaca, Jo Bovy, Babette Döbrich, Doug Finkbeiner, Eder Izaguirre, Simon Knapen, Eric David Kramer, Jung-Tsung Li, Mariangela Lisanti, Aneesh Manohar, Sam McDermott, Christopher McKee, Julian Muñoz, Lina Necib, Lisa Randall, Eddie Schlafly, Alexandra Shelest, Tomer Volansky, Jay Wacker, Neal Weiner, and Hai-Bo Yu.

Finally, I thank my family and friends for their support and for providing much-needed distraction during difficult times.

The work in Chapter 2 of this dissertation made use of data from the European Space Agency (ESA) mission *Gaia*,[1] processed by the *Gaia* Data Processing and Analysis Consortium.[2] Funding for this Consortium has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work also made use of the `Multinest` nested sampling package [1] through its `Python` interface [2], the Powell minimization algorithm [3] implemented in `SciPy` [4], and the `gaia_tools` package [5]. The analysis appearing in this chapter was performed using the Flux computing cluster at the University of Michigan.

The work in Chapter 3 of this dissertation made use of `Arepo` [6], access for which was provided by Volker Springel. The simulations appearing in this chapter were performed on the joint computing cluster supported by the Massachusetts Institute of Technology Kavli Institute and the Harvard Faculty of Arts and Sciences.

Part of the work in Chapter 4 of this dissertation was completed during visits to the Kavli Institute for the Physics and Mathematics of the Universe at the University of Tokyo and the Hebrew University of Jerusalem.

Part of the work in Chapter 5 of this dissertation was completed during visits to the University of California in San Diego, Princeton University, and New York University.

[1] https://www.cosmos.esa.int/gaia

[2] https://www.cosmos.esa.int/web/gaia/dpac/consortium

Chapter 1

Motivation

The preponderance of evidence for the existence of dark matter (DM) in a wide range of environments is one of the greatest triumphs of modern astrophysics. Cosmological observations indicate that DM was present in the first few minutes after the Big Bang on the scale of our entire observable Universe; astronomical observations also show that DM makes up the majority of matter in nearby dwarf galaxies at the present day; DM is also seen in many intermediate lengthscales and epochs on our cosmological timeline. The concordance between different observations in such different regimes is striking, and in the aggregate the data conclusively indicate that DM makes up approximately 85% of the matter in our Universe. In spite of this resounding agreement between datasets about the *existence* of DM, little is known about its composition and its physical properties.

Historically, DM was first discovered because the gravitational potential wells of galaxy clusters and galaxies [7–11] are not deep enough (based on the distribution of luminous matter) to support the fast rotation observed at large radii: without DM, clusters and galaxies would fly apart. More extreme examples with hidden sources of gravitational pull include low surface brightness galaxies [12], dwarf galaxies [13–18], and substructure inferred by strong gravitational lensing [19–23], all of which are environments that appear to be overwhelmingly dominated by DM. Other observations exclude the possibility that DM is comprised of ordinary baryonic matter that merely happens to be dimmed by dust or by the failed ignition of would-be stars (*i.e.* clouds of gas). For instance, in merging galaxy clusters the spatial distribution of mass (inferred from weak gravitational lensing) and the spatial distribution of supersonic gas (which can be seen in X-rays) are rather distinct, with the gas (which makes up most of the ordinary matter) lagging behind as the clusters pass through each other [24–26]. This implies that the gas, which is slowed down by viscous drag, is lagging behind some component of inert matter which does not feel the underlying forces responsible for the drag. The same kind of behavior can be seen in the primordial plasma that sources the cosmic microwave background (CMB), where baryons undergo acoustic oscillations due to the competing forces of gravity and radiation pressure. The addition of non-baryonic matter that does not experience radiation pressure deepens gravitational wells and enhances the compression of high-density regions relative to a scenario where all gravitating matter is influenced by radiation pressure. Thus, evidence for DM can be seen in the pattern of baryon acoustic oscillations (BAO) measured in CMB power spectra [27–31]. The BAO are then imprinted onto

large scale structure (LSS) [32, 33], the growth of which is seeded by DM that was able to cluster at early times without being washed out by radiation. Observations of LSS on different scales and at different epochs independently indicate the existence of non-baryonic DM through this BAO feature [34–41]. The evidence is bolstered by considering how the abundance of elements produced during big bang nucleosynthesis (BBN) would be affected by the baryon-to-photon ratio; measurements of primordial elemental abundances indicate that the total density of baryons is lower than the total matter density measured in the CMB and in LSS [42, 43].

The existence of DM poses one of the greatest outstanding questions of modern particle physics[1] due to the lack of a viable DM candidate within the Standard Model (SM).[2] Moreover, there are relatively few model-independent restrictions on possible particle DM candidates. DM must not interact appreciably with light (see Ref. [90] for a recent compilation of the limits), must have enough mass so that the characteristic length scale (which depends on particle spin statistics) is smaller than the smallest observed dwarf galaxies [89, 91, 92], must be stable on cosmological timescales (see for instance Ref. [93]), must not be moving relativistically during epochs that are relevant for the growth of LSS [94, 95], and must be inert enough to leave LSS [96] and merging clusters [97] unaffected by DM collisions. These requirements are easily satisfied from a model-building perspective, leaving little theoretical guidance in the absence of other questions and anomalies. However, there are other unresolved theoretical questions in particle physics, many of which have solutions that include DM candidates. For instance,

- the weakly interacting massive particle (WIMP) is a DM candidate that is naturally present in many solutions of the electroweak hierarchy problem and that has a thermal relic abundance set by weak-scale annihilations [98–101],
- the axion is a DM candidate that also solves the strong-CP problem in quantum chromodynamics [102–105], and
- the sterile neutrino is a DM candidate that addresses the origin of the SM neutrino mass [106–110].

From an empirical perspective, anomalous observations or experimental results can also guide DM model-building. Examples include

[1] It is not strictly necessary that DM be comprised of an undiscovered particle; for instance, black holes could be a component of (or all of) the DM because they are collisionless, dark, and massive. In order to be consistent with the baryon-to-photon ratio measured at BBN, the black holes must form before the onset of BBN if they are a large fraction of the DM today. Understanding the formation of primordial black holes and their observational signatures is an active area of research [44–81] that we do not explore further in this dissertation.

[2] Massive SM neutrinos would pose an appealing candidate to explain DM if they did not move at relativistic speeds in the early Universe after thermally decoupling from the rest of the SM plasma. Because relativistic particles are not easily captured by nascent gravitational potential wells, neutrinos as a component of a fixed matter budget actually suppress the growth of structure their free-streaming scale [82–84]. Null observations of such a suppression indicate that neutrinos are not a large component of DM and these observations can actually be translated to a strong limit on the neutrino mass from cosmology [85–87]. Furthermore, even without cosmological upper bounds, terrestrial limits on the neutrino mass [88] imply that neutrinos cannot be a substantial fraction of the DM inside dwarf galaxies because of Fermi packing [89].

- gamma rays from the Galactic center [111] and extragalactic sources of X-rays [112, 113],
- the rising positron fraction in high-energy cosmic rays [114] and the sharp break in the cosmic ray spectrum at ∼TeV energies [115],
- the low density and number of local subhalos and satellites (see Chapter 3 for further discussion),
- the cold spin temperature of neutral hydrogen observed by the EDGES 21 cm global signal experiment at a redshift $z \sim 20$ [116],
- the LSND [117] and MiniBooNE [118] excesses,
- and the anomalous magnetic moment of the muon [119].

As one of the most promising potential windows onto new fundamental physics, there are many efforts underway to probe the particle nature of DM and its interactions with the SM, but to date it has only been detected through its gravitational influence on visible matter. Most experimental and observational efforts thus far have focused on the most theoretically well-motivated DM candidates, with a strong emphasis on WIMPs and axions. Much of the available parameter space has been explored by direct detection (*e.g.* in Refs. [120–124]), indirect detection (*e.g.* in Refs. [125, 126]), and collider searches (*e.g.* in Refs. [127–130]) for these DM candidates. Taken together, the wide variety of null searches for particle DM strongly motivates taking a broader view of potential models. In particular, it is possible that DM does not dovetail with other theoretical questions like the hierarchy problem or the strong-CP problem, and solutions to those problems involving a DM candidate may have no bearing on the true nature of DM. If we abandon the theoretical prior on DM paradigms that solve other problems in fundamental physics, then DM sectors naturally emerge as a compelling alternative (for an overview of dark sectors and some of their signatures, see *e.g.* Refs. [131–133] and references contained therein). In analogy to the SM, dark sectors can contain multiple states in the spectrum and can have their own forces with nontrivial gauge structure. Dark sectors can optionally have portal interactions with the SM, which would make DM searches involving those portals compelling; alternatively, dark sectors can be completely secluded, in which case they can only ever be detected via gravitational interactions.

By its very nature, the study of dark sectors is an undertaking which requires a diverse set of theoretical and observational tools. As these sectors have thus far evaded detection, there are few restrictions on the kinds of dark sector structures that are possible. Thus dark sectors provide a sandbox for the realm of conceivable DM models. Moreover, dark sectors can be manifest in select situations but undetectable in others, indicating that complementarity between observables is necessary for making any conclusive determinations about dark sectors. This is particularly true of astrophysical observables, which provide such a diversity of environments, energy scales, and sheer size over which dark sectors can have an impact. This dissertation will make no attempt to give a comprehensive overview of the status of dark sector models and searches. Rather, we will try to illustrate the constraining power of astrophysics by considering a few different systems and a few types of hidden sector theories.

The theories we consider in this dissertation range from having strong, nearly non-perturbative interactions to having extremely feeble, sub-Hubble interactions. Some theories considered in this work are thermal DM candidates produced in the early Universe, much like SM particles. Other theories considered in this work are so highly secluded that the DM must have a non-thermal origin. We demonstrate that these theories can be tested by considering their impact at different cosmological epochs and in different astrophysical environments. In some cases, small portal couplings between the dark sector and SM are constrained when a large quantity of SM material is aggregated, for instance in stars. Alternatively, dark sectors may be best seen via their influence on the spatial distribution of DM in galaxies and halos (for instance, if there are collisions or dissipations in the dark sector) and the subsequent gravitational effect on visible matter. The mechanism for setting the DM relic abundance in a dark sector may provide yet another observational handle through gravitational and non-gravitational imprints on the CMB and formation of LSS. The work in this dissertation includes analytic calculations, numerical simulations, and the analysis of high-precision data, which illustrates the range of techniques that can be used in the pursuit of uncovering dark sectors.

The remaining chapters in this dissertation are as follows:

Chapter 2 — If a component of the DM has dissipative interactions, it could collapse to form a thin dark disk in our Galaxy that is coplanar with the baryonic disk. It has been suggested that dark disks could explain a variety of observed phenomena, including periodic comet impacts. Using the first data release from the *Gaia* space observatory, we search for a dark disk via its effect on stellar kinematics in the Milky Way. Our new limits disfavor the presence of a thin DM disk, and we present updated measurements on the total matter density in the solar neighborhood. This chapter is based on the work appearing in Ref. [134].

Chapter 3 — Self-interacting DM is a promising alternative to the cold DM paradigm that can modify small-scale galaxy formation, making the predicted properties of small DM halos more consistent with observations when implemented in N-body simulations. Nearly all self-interacting DM simulations so far have considered purely elastic collisions. However, the presence of internal structure for the DM — for example, a nearly-degenerate state in the spectrum that could be collisionally excited or de-excited — can potentially enhance the agreement between simulation and observation by converting mass to kinetic energy. We present simulations of a galactic halo within a realistic DM model with inelastic scattering using a novel numerical implementation in the `Arepo` code, which is based on the development of accurate approximations in the non-perturbative regime. When compared to purely elastic DM self-scattering, inelastic scattering can create larger cores inside DM halos, lower the abundance of subhalos, reduce the mass of halos, inject the equivalent energy of 100 million Type II supernovae, and avoid the gravothermal catastrophe. The cross sections needed to achieve these effects are much smaller than for elastic self-scattering, implying that well-established limits on DM self-interaction must be revised if inelastic collisions are the dominant mode of scattering. This chapter is based on the work appearing in Refs. [135] and [136].

Chapter 4 — DM could be a thermal relic comprised of strongly interacting massive particles (SIMPs), where $3 \to 2$ interactions set the relic abundance. Such interactions generically arise in theories of chiral symmetry breaking via the Wess-Zumino-Witten term. We show that an axion-like particle can successfully maintain kinetic equilibrium between the DM and the visible sector, allowing the requisite entropy transfer that is crucial for SIMPs to be a cold DM candidate. Constraints on this scenario arise from beam dump and collider experiments, from the cosmic microwave background, and from supernovae. We find a viable parameter space when the axion-like particle is close in mass to the SIMP DM, with strong-scale masses of order a few hundred MeV. Many planned experiments are set to probe the parameter space in the near future. This chapter is based on the work appearing in Ref. [137].

Chapter 5 — DM could couple to particles in the SM through a light vector mediator. In the limit of small coupling, this portal could be responsible for producing the observed DM abundance through a mechanism known as freeze-in. Furthermore, the requisite DM-SM couplings provide a concrete benchmark for direct and indirect searches for DM. We present updated calculations of the relic abundance for DM produced by freeze-in through a light vector mediator, identifying an additional production channel: the decay of photons that acquire an in-medium plasma mass. These plasmon decays are a dominant channel for DM production for sub-MeV DM masses, and including this channel leads to a significant reduction in the predicted signal strength for DM searches. Accounting for production from both plasmon decays and annihilations of SM fermions, the DM acquires a highly non-thermal phase space distribution which impacts the cosmology at later times. This chapter is based on the work appearing in Ref. [138].

Chapter 6 — We summarize our outlook based on the work presented here in the context of the broader search for DM.

Chapter 2

Dark sector dissipation in the Milky Way

2.1 Introduction

As discussed in Chapter 1, many recently-proposed models posit that DM is part of a dark sector, containing interactions or particles that lead to non-trivial dynamics on astrophysical scales [139–154]. Meanwhile, the *Gaia* satellite [155] has been observing one billion stars in the local Milky Way (MW) with high precision astrometry, which will allow for a vast improvement in our understanding of DM substructure in our Galaxy and its possible origins from dark sectors. In this Chapter, we apply the first *Gaia* data release [156] to constrain the possibility that DM can dissipate energy through interactions in a dark sector. Existing constraints imply that the entire dark sector cannot have strong self interactions, since this would lead to deviations from the predictions of cold DM that are inconsistent with cosmological observations [97, 157–159]. However, it is possible that only a subset of the dark sector interacts strongly (in analogy with the Standard Model) or that DM interactions are only strong in low-velocity environments [160–162]. In these cases, there is leeway in cosmological bounds and one must make use of smaller scale observables to constrain such interactions [163, 164]. If the DM component can dissipate energy through emission or upscattering (see *e.g.* [135, 154, 165–171] for examples of mechanisms), then it can cool and collapse to form DM substructure. These interactions could result in a striking feature in our Galaxy: a thin DM disk (DD) [152, 153] that is coplanar with the baryonic disk.

A thin DD may be accompanied by a range of observational signatures. For instance, DDs may be responsible for the $\sim$30 million year periodicity of comet impacts [172], the co-rotation of Andromeda's satellites [173, 174], the point-like nature of the inner Galaxy GeV excess [175, 176], the orbital evolution of binary pulsars [177], and the formation of massive black holes [178], in addition to having implications for DM direct detection [179, 180]. Typically, a DD surface density of $\Sigma_{DD} \sim 10\,\mathrm{M}_\odot/\mathrm{pc}^2$ and a scale height of $h_{DD} \sim 10$ pc are required to meaningfully impact the above phenomena.[1]

Here we present a comprehensive search for a local DD, using tracer stars as a probe of the

[1] A thicker DD with $h_{DD} \gtrsim 30$ pc can cause periodic cratering [181], however a larger surface density $\Sigma_{DD} \sim$ 15-20 $\mathrm{M}_\odot/\mathrm{pc}^2$ is required to be consistent with paleoclimactic constraints [182].

local gravitational potential. Specifically, we use the *Tycho-Gaia* Astrometric Solution (TGAS) [183, 184] catalog, which provides measured distances and proper motions for ∼2 million stars in common with the *Tycho-2* catalog [185]. Previous work searching for a DD with stellar kinematics used data from the *Hipparcos* astrometric catalog [186] and excluded local surface densities $\Sigma_{DD} \gtrsim 14\,\mathrm{M}_\odot/\mathrm{pc}^2$ for dark disks with thickness $h_{DD} \sim 10$ pc [187]. As compared with *Hipparcos*, TGAS contains ∼20 times more stars with three dimensional positions and proper motions within a larger observed volume, which allows for a significant increase in sensitivity. Our analysis also improves on previous work by including a comprehensive set of confounding factors that were previously not all accounted for, such as uncertainties on the local density of baryonic matter and the tracer star velocity distribution. We exclude $\Sigma_{DD} \gtrsim 6\,\mathrm{M}_\odot/\mathrm{pc}^2$ for $h_{DD} \sim 10$ pc, and our results put tension on the DD parameter space of interest for explaining astrophysical anomalies [172–178].

2.2 Vertical kinematic modeling

We use the framework developed in Ref. [188] (and extended in Ref. [187]) to describe the kinematics of TGAS tracer stars in the presence of a DD. This formalism improves upon previous constraints on a DD that did not self-consistently model the profiles of the baryonic components in the presence of a DD [189–193]. These earlier bounds typically compared the total surface density of the Galactic disk, measured from the dynamics of a tracer population above the disk, to a model of the surface density of the baryons based on extrapolating measurements from the Galactic midplane. However, the models did not include the pinching effect of the DD on the distribution of the baryons, which would lower the inferred baryon surface density for fixed midplane density. Instead, Refs. [187, 188] consider the dynamics close to the disk and self-consistently model the baryonic components for fixed DD surface density and scale heights. We summarize the key components below.

In a collisionless self-gravitating system, such as a collection of stars, particles obey Liouville's theorem. In particular, for a population labeled with the upper case roman index I, the phase space distribution function $f_I(\mathbf{x}, \mathbf{v})$ obeys the collisionless Boltzmann equation

$$\frac{d\,f_I}{d\,t} = \partial_t f_I + \partial_{\mathbf{x}} f_I \cdot \mathbf{v} - \partial_{\mathbf{v}} f_I \cdot \partial_{\mathbf{x}} \Phi = 0\,, \tag{2.1}$$

where Φ is the total gravitational potential summed over all populations and where we have dropped the explicit dependence of f_I on phase space coordinates. Instead of working directly with f_I, for understanding the bulk behavior of the system it is often sufficient to describe mo-

ments of the phase space distribution,

$$\nu_I = \int d^3v f_I \tag{2.2}$$

$$\bar{v}_{I,i} = \frac{1}{\nu_I}\int d^3v\, v_i\, f_I \tag{2.3}$$

$$\sigma^2_{I,ij} = \left(\frac{1}{\nu_I}\int d^3v\, v_i v_j\, f_I\right) - \bar{v}_i\bar{v}_j\,, \tag{2.4}$$

such that ν_I represents the number density, $\bar{v}_{i,I}$ represents the mean velocity, and $\sigma^2_{I,ij}$ represents the velocity dispersion tensor for stars in population I. Note that lower case roman letters here denote spatial indices. Equipped with these definitions and assuming the system is in equilibrium (*i.e.* time derivatives vanish) we can integrate moments of the Boltzmann equation in cylindrical coordinates. The first non-vanishing moment (the Jeans equation) takes the form

$$\frac{1}{r\nu_I}\partial_r(r\nu_I\sigma^2_{I,rz}) + \frac{1}{\nu_I}\partial_z(\nu_I\sigma^2_{I,zz}) + \partial_z\Phi = 0\,, \tag{2.5}$$

assuming axisymmetry. In general, in the Milky Way the assumption of axisymmetry (and of dropping the related time-dependence) does not hold due to the presence of the rotating spiral arms; however for the volume considered in this analysis such effects can be neglected [194–196]. The first term of Eq. (2.5) (commonly known as the tilt term) can be ignored when dealing with dynamics near the disk since radial derivatives are much smaller than vertical ones (see *e.g.* Refs. [197, 198]). However, this term can be important above the scale height of the disk [194]. We further assume that each population is gravitationally well-equilibrated near the Galactic plane with constant $\sigma^2_{I,zz}$ (*i.e.* that Maxwell-Boltzmann velocity statistics for an isothermal population are satisfied [199]). With these simplifying assumptions, the solution to the vertical Jeans equation is $\nu_I(z) = \nu_{I,0}\, e^{-\Phi(z)/\sigma^2_I}$, where we impose $\Phi(0) = 0$ and define $\nu_{I,0} \equiv \nu_I(0)$ and $\sigma_I \equiv \sigma_{I,zz}$. For populations composed of roughly equal mass constituents (including gaseous populations), we then make the assumption that the number density and mass density are proportional, *i.e.* $\rho_I(z) = \rho_{I,0}\, e^{-\Phi(z)/\sigma^2_I}$.

To compute the gravitational potential from the mass density of the system, we use the Poisson equation for standard Newtonian gravity,

$$\nabla^2\Phi = \partial_z^2\Phi + \frac{1}{r}\partial_r(r\partial_r\Phi) = 4\pi G\rho\,, \tag{2.6}$$

where ρ is the total mass density of the system, summed over all gravitating populations indexed by I. Under the assumption of axisymmetry, orbits are circular and we can relate the second term to measured quantities as

$$\frac{1}{r}\partial_r^2(r\partial_r\Phi) = \frac{1}{r}\partial_r^2 v_c^2 = 2(B^2 - A^2)\,, \tag{2.7}$$

where v_c is the circular orbital velocity, and A and B are Oort's constants, defined as

$$A \equiv \frac{1}{2}\left(-\frac{\partial v_c}{\partial r} + \frac{v_c}{r}\right) \qquad B \equiv -\frac{1}{2}\left(\frac{\partial v_c}{\partial r} + \frac{v_c}{r}\right). \tag{2.8}$$

One of the most recent measurements of these constants comes from *Gaia* DR1 [200], so for self-consistency we adopt these values of $A = 15.3 \pm 0.4$ km/s/kpc and $B = -11.9 \pm 0.4$ km/s/kpc. This means that the radial contribution to the effective vertical density is

$$\rho_{\rm eff}(z) = \rho(z) - \frac{B^2 - A^2}{2\pi G} = \rho(z) + (3.4 \pm 0.6) \times 10^{-3} \ \ {\rm M}_\odot/\ {\rm pc}^3, \tag{2.9}$$

assuming uncorrelated errors. This additional effective contribution to the density is roughly one third of previous measurements of the local DM density and can be subtracted off at the end of the analysis to obtain a measure of the physical DM density alone [201]. Note that in principle Oort's constants can vary as a function of z; however, in the region we consider which is close to the galactic plane, this effect is smaller than the uncertainties on their measured values [202].

Combining the Jeans and Poisson equations assuming reflection symmetry in the z direction yields the integral equation

$$\nu_I(z) = \nu_{I,0} \exp\left(- \sum_J \frac{4\pi G}{\sigma^2_{J,zz}} \int_0^z dz' \int_0^{z'} dz'' \rho_J(z'') \right) . \tag{2.10}$$

Here we are again assuming that the mass density of population J, ρ_J, is proportional to its number density ν_J, *i.e.* that while there may be some scatter in the mean mass of tracer stars, this does not have any dependence on z. In the limiting case with only one population, the solution to this integral equation is $\rho(z) = \rho_0 \, {\rm sech}^2(\sqrt{2\pi G \rho_0}\, z/\sigma)$. For multiple populations, the solution for the density profiles and Φ must be determined numerically. We use an iterative solver with two steps per iteration. On the $n^{\rm th}$ iteration, we compute the total potential

$$\Phi^{(n)}(z) = 4\pi G \sum_I \int_0^z dz' \int_0^{z'} dz'' \rho_I^{(n)}(z'') \tag{2.11}$$

and update the density profile for the $I^{\rm th}$ population,

$$\rho_I^{(n+1)}(z) = \rho_{0,I} \, e^{-\Phi^{(n)}(z)/\sigma_I^2} . \tag{2.12}$$

We find that this solver converges within a few iterations when implementing the mass model described in the next Section. Adding more gravitating populations compresses the density profile relative to the single-population case.

Once equipped with a converged gravitational potential for the combined components in the Galactic disk, we can predict the vertical density profile for a given tracer population. Since our focus is on vertical motion, we assume a form for the distribution function where the motion in the z direction is separable from the other components, which is equivalent to dropping the tilt, azimuthal, and rotation curve terms as we have done above. Then for this z component of the distribution function, the Boltzmann equation reads

$$v_z \partial_z f_K - \partial_z \Phi \partial_{v_z} f_K = 0 \, , \tag{2.13}$$

Baryonic Component	$\rho(0)$ [$M_\odot$/pc^3]	σ [km/s]
Molecular Gas (H_2)	0.0104 ± 0.00312	3.7 ± 0.2
Cold Atomic Gas ($H_I(1)$)	0.0277 ± 0.00554	7.1 ± 0.5
Warm Atomic Gas ($H_I(2)$)	0.0073 ± 0.0007	22.1 ± 2.4
Hot Ionized Gas (H_{II})	0.0005 ± 0.00003	39.0 ± 4.0
Giant Stars	0.0006 ± 0.00006	15.5 ± 1.6
$M_V < 3$	0.0018 ± 0.00018	7.5 ± 2.0
$3 < M_V < 4$	0.0018 ± 0.00018	12.0 ± 2.4
$4 < M_V < 5$	0.0029 ± 0.00029	18.0 ± 1.8
$5 < M_V < 8$	0.0072 ± 0.00072	18.5 ± 1.9
$M_V > 8$ (M Dwarfs)	0.0216 ± 0.0028	18.5 ± 4.0
White Dwarfs	0.0056 ± 0.001	20.0 ± 5.0
Brown Dwarfs	0.0015 ± 0.0005	20.0 ± 5.0
Total	0.0889 ± 0.0071	—

Table 2.1: The baryonic mass model that informs our priors.

where we have indexed tracer (rather than gravitating) populations by K and where again we are dropping time derivatives. Any function of the form $\mathcal{F}(v_z^2/2 + \Phi(z))$ will satisfy the above differential equation. We also note that separability implies

$$\int dv_z f_K(z, v_z) = \nu_K(z) \quad \Rightarrow \quad f_K(z, v_z) = \nu_K(z) f_{K,z}(v_z)\,, \tag{2.14}$$

where $f_{K,z}(v_z)$ is the velocity distribution function at some fixed height z, normalized to unity. One can show that in the limit of a single self-gravitating system in equilibrium, the velocity distribution near the midplane is Gaussian [203]. In general and with the above definitions in mind, we can write

$$\begin{aligned} \nu_K(z) &= \int dv_z f_K(z, v_z) \\ &= \int dv_z f_K\left(0, \sqrt{v_z^2 + 2\Phi(z)}\right) \\ &= \nu_K(0) \int dv_z f_{K,0}\left(\sqrt{v_z^2 + 2\Phi(z)}\right). \end{aligned} \tag{2.15}$$

Therefore, once we know the gravitational potential and the velocity distribution function for tracers at the midplane, we can solve for the tracer profile. In our analysis we determine $f_{K,0}$ empirically and do not assume that our tracer population is necessarily isothermal.

2.3 Mass Model

In order to solve for the gravitational potential, we must have an independent model for the baryons. In Table tab:massmodel, we compile some of the most up-to-date measurements of $\rho_{A,0}$ and σ_A

for the local stars and gas, primarily drawing from the composite model of Ref. [193] and supplementing with velocity dispersions from Refs. [204, 205] as described below.

The local gas densities are determined from measurements of the column density of hydrogen in its various forms, assuming a fixed scale height. A correction factor of 1.4 is applied to account for the presence of heavier elements whose abundance relative to hydrogen is known. For the case of molecular hydrogen (H_2), there is a 30% uncertainty on the density, while for the ionized hydrogen (H_{II}) the uncertainty is 5%. The atomic hydrogen (H_I) is split into warm and cold components: the uncertainty on the density of the warm component is 10% while for the cold component the uncertainty is 20% due to optical depth corrections.

For the giant stars and stars with $M_V < 8$, we conservatively estimate the uncertainty on the density as 10%, which is commensurate with the relative uncertainty on the *total* surface density of these populations for fixed scale height in the analysis of Ref. [193]. Note that the analysis of Ref. [5] found very similar values for the density of giant stars and main sequence stars with $M_V < 8$, within the margins of uncertainty reported by Ref. [193]. Since these stars are such a subdominant component of the mass model, we do not perform a separate analysis with measured densities of Ref. [5] but instead perform our analysis with those from Ref. [193], which have more conservative uncertainties. For the M dwarfs the uncertainty on the surface density (again, for fixed scale height) is 13%, which we translates to the same relative uncertainty for the density [193]. The uncertainty on the density of white dwarfs is given explicitly in Ref. [193], while for the brown dwarfs we assume a 35% uncertainty on the density coming from the uncertainty in the spectral index of the brown dwarf mass function. Note that for the densities reported here, the thick stellar disk is already taken into account as described further in Ref. [193].

For the velocity dispersions, we aggregate measurements and uncertainties from several sources. The velocity dispersions for the gas, giant stars, and stars with $M_V < 4$ came from the revised estimates provided in Refs. [187, 204], while for the other category we use velocity dispersions from Refs. [205, 206]. In particular the updated velocity dispersions and uncertainties we quote for the gas include additional non-thermal contributions to the pressure, including the effects of magnetic fields. Where available, the uncertainties on the velocity dispersion were drawn from Refs. [205, 206]. In some cases, the populations were categorized differently, in which case a best attempt was made to consolidate different references. For instance, Ref. [187] estimated the velocity dispersion for the category $M_V < 3$ as being the same for the category $M_V < 2.5$, under the assumption that the contribution from stars with $2.5 < M_V < 3$ is small compared to the error on the velocity dispersion. For the stars with $3 < M_V < 8$ and the M dwarfs, we estimate the error on the velocity dispersions as coming from the variance of different measurements of the scale height aggregated in Ref. [193], which yield similar error estimates as those reported in Refs. [205, 206]. Finally, in other cases where the updated stellar velocity dispersions from Ref. [187] were different from the values given in Refs. [205, 206], we assume that the relative size of the error on the revised values is the same as for the previous values. For instance, for the giant stars the error is assumed to be 10% for the new value of the velocity dispersion, as it was for the old value.

We also include the energy density from the smooth DM halo and a possible thin DD component. We model the bulk collisionless DM halo of the MW as a constant local density $\rho_{\rm DM}$. Current measurements favor $\rho_{\rm DM} \sim 0.01\text{-}0.02\,{\rm M}_\odot/{\rm pc}^3$ at ∼kpc heights above the Galactic plane

[206], though we will treat $\rho_{\rm DM}$ as a nuisance parameter in our analysis. For the thin DD, we parametrize the density as

$$\rho_{DD}(z) = \frac{\Sigma_{DD}}{4h_{DD}} \,\mathrm{sech}^2\left(\frac{z}{2h_{DD}}\right) , \tag{2.16}$$

with Σ_{DD} and h_{DD} the DD model parameters. In constraining these model parameters, we include the uncertainties tabulated in Table 2.1 and profile over these nuisance parameters in our analysis.

2.4 Tracer Star Sample

As outlined above, we will constrain a DD by considering the distribution of tracer stars in the underlying gravitational potential, as sourced by the baryonic disk and possibly a DD. In an ideal situation, we would consider a sample volume as large and inclusive as possible so as to improve statistics with a larger number of tracer stars. However, we wish to avoid biasing the sample with systematic errors that can influence our results. In this section, we demonstrate the effects of different cuts on meeting the above goals and discuss how we estimate the errors on the data.

For our analysis of the tracer density profile, we select TGAS stars within a cylinder about the Sun with radius $R_{\rm max}$ and which extends 200 pc above and below the Galactic plane. This range in z ensures that we are within the regime of validity for several key assumptions made above: that the tilt term is negligible, that the various mass components have constant velocity dispersions, and that the radial and vertical motions are separable. Indeed, Ref. [197] showed in simulations and with data that these assumptions are satisfied within the $z \sim 200$ pc half-mass height of the disk. Using the color cuts defined in Ref. [207], we consider main sequence stars of spectral types A0-G4. Older spectral types have density profiles that are closer to constant (*i.e.* less peaked) in our volume, and thus less are useful in constraining a DD. In the rest of this Section, we explore how different sample volumes, parallax uncertainty cuts, and selection functions impact the tracers.

When determining the midplane vertical velocity distributions of our tracer stars, we impose latitude cuts to ensure that the unknown line-of-sight velocities have a small projection in the vertical direction. In the rest of this Section, we show how different Galactic latitudes and typical line-of-sight velocities affect the tracer velocity distributions.

Parallax Uncertainties and Sample Geometry

Perhaps the largest source of error in determining the shape of the tracer density profiles is parallax uncertainty. In order to reduce this uncertainty, we can impose a parallax cut on the data. Since cutting on individual stars could bias our sample, we instead make cuts on the *mean* parallax uncertainty in a given region of the sky rather than cutting individual stars with poor parallax measurements. In order to retain sufficient statistics, we make a modest cut on regions of the sky with average parallax error greater than 0.4 mas. In addition to these statistical parallax uncertainties, there is a reported systematic parallax uncertainty on all TGAS data of 0.3 mas. In total, this translates to a $\sim 10\%$ uncertainty on z near the edge of our sample volume at $z = 200$ pc, as can be

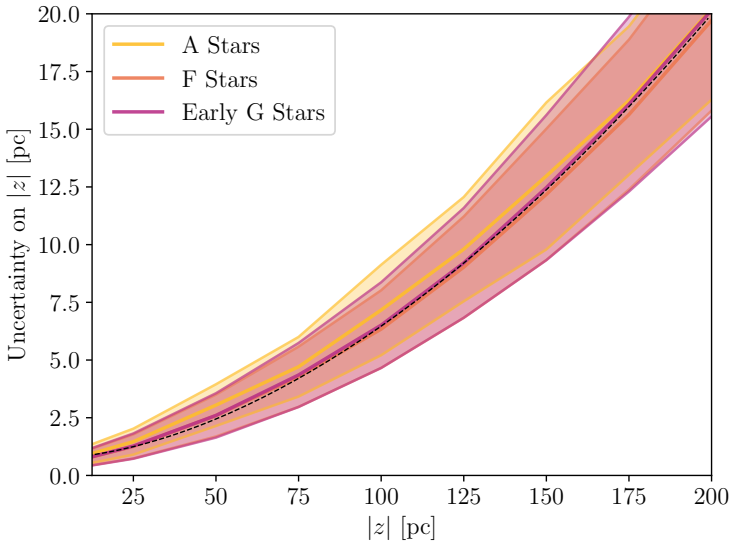

Figure 2.1: The uncertainty in the height above the Galactic plane z as a function of z for a cylinder centered at Earth with radius 150 pc. The mean errors are shown as thick lines and the bands represent the spread in the error. The error in z incurred from parallax uncertainties is similar between different stellar populations. The fitting function of Eq. (2.17) is shown as a black dotted line.

seen in Fig. 2.1. These uncertainties can be mitigated to an extent by combining TGAS with other catalogs [208, 209], but the issue remains that parallax uncertainties are the main contaminant in our density profiles. In our analysis, we use the fitting form

$$\sigma(z) = 0.00183\,|z|^{1.7478} + 0.74 \text{ pc} \tag{2.17}$$

for the uncertainty on z at different heights above the Galactic plane, as shown in Fig. 2.1.

Due to the uncertainties in parallax error, we would ideally choose bins in z that are larger than the uncertainties. However, one possible effect of choosing large z bins is to artificially broaden the profile through Eddington bias (see below). This could cause our constraints on a DD to be overly restrictive, since we are searching for the pinching effect from a DD in the profile. To check for the dependence on z binning, we run our analysis on the data with both fine 8 pc bins and coarse 25 pc bins, as described in Section 2.7. Parallax uncertainties also feed into the midplane vertical velocities through the projection of the proper motions onto the z direction. The RMS uncertainty on midplane vertical velocities from parallax uncertainties is $\sim$0.6 km/s. Again, to check for artifacts from binning we run our analysis pipeline with fine 2 km/s velocity bins and coarse 4 km/s velocity bins in Section 2.7.

In Fig. 2.2, we show the effects of the cut on the average parallax error and the choice of binning in z on the tracer star density profile, as functions of the radial cut R_{max}. Fitting the density of F and early G stars to the profile $\nu(z) = \nu_0\,\text{sech}^2((z + z_{\text{sun}})/2h)$, we find systematic deviations

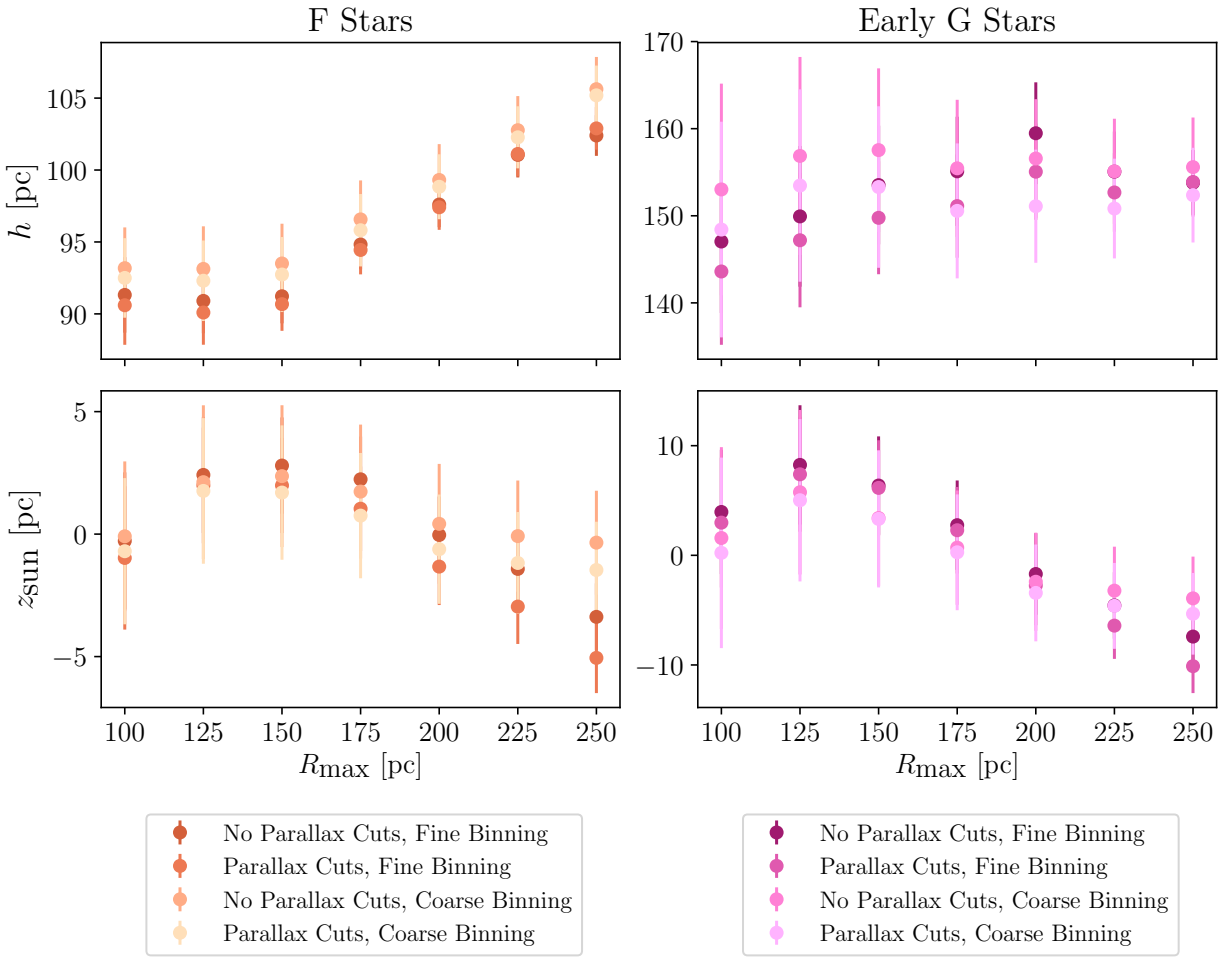

Figure 2.2: Best fit parameters for the profile $\nu(z) = \nu_0 \text{sech}^2((z + z_{\text{sun}})/2h)$ of F and Early G type stars for cylindrical sample volumes with different radial cuts R_{max}. We consider data with different vertical bin sizes (8 pc and 25 pc) as well as data with and without mean parallax uncertainty cuts. Generally, placing a parallax cut or using finer binning yields a narrower inferred profile. The different choices of cuts show increasing disagreement in the best fit values for $R_{\text{max}} \gtrsim 150$ pc, which is suggestive of systematic effects due to parallax uncertainty.

in the scale height h and height of the sun z_{sun} depending on these choices. For both fine (8 pc) and coarse (25 pc) binning, omitting the parallax cut described in the previous section biases the profile to be broader due to larger uncertainties on z. The profiles are also broader with coarser binning. These differences in the density profile are mitigated by considering a smaller sample volume. We demonstrate this by showing the dependence of the fit results on the maximum radius R_{max} of our sample volume. The most salient feature is the emergence of R_{max} =150 pc as the approximate radius beyond which the profiles start to systematically broaden more with increased radius. This

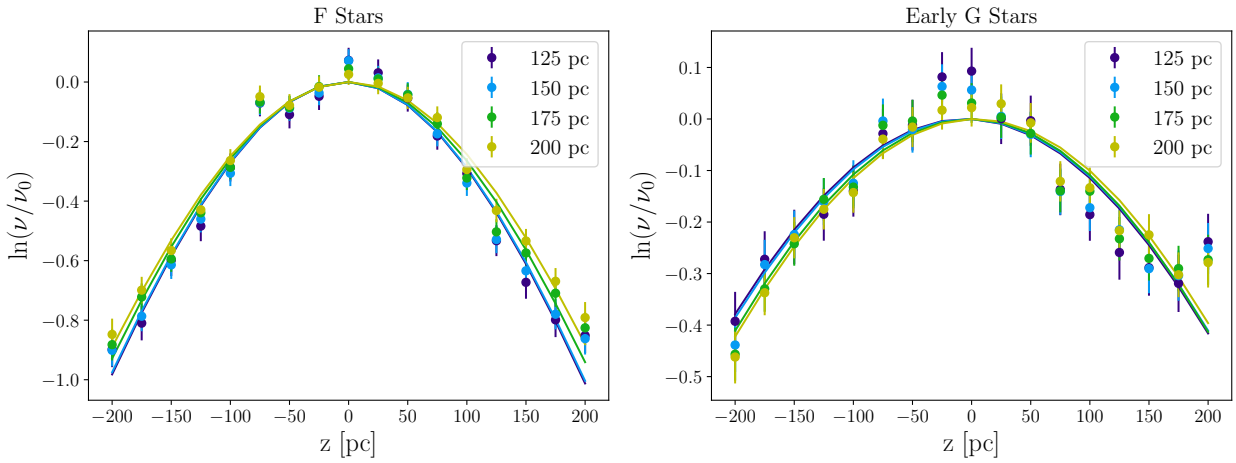

Figure 2.3: The density profiles $\nu(z)$ for F and early G stars in sample volumes with different values of $R_{\rm max}$ and including parallax cuts. Overlaid on top of the data are the best-fit sech^2 profiles. As the sample volumes grow past $R_{\rm max} = 150$ pc, the inferred profiles broaden and shift, which we attribute to larger parallax uncertainty at the edges of the sample volume and Eddington bias. Most notably, the profiles inferred with $R_{\rm max} = 125$ pc and $R_{\rm max} = 150$ pc are in good agreement, while the profiles inferred with $R_{\rm max} = 175$ pc and $R_{\rm max} = 200$ pc are different from the low-$R_{\rm max}$ profiles and from each other.

is illustrated further in Fig. 2.3 where a convergence of sech^2 profiles is seen below $R_{\rm max}$=150 pc for the coarsely-binned data with parallax cuts.

We interpret the broadening of the number density profile as being due to Eddington bias: due to the increasing parallax uncertainties with distance of the star, a larger sample volume leads to smearing of the density profiles at large z. Mechanistically this arises because for a peaked distribution, large uncertainties are more likely to cause excursions from high-count to low-count bins than the reverse, meaning that excursions that broaden the distribution are not compensated by ones that would cause it to narrow. In order to maximize the statistics of our analysis while reducing these systematic effects, we restrict all of the analyses in this Chapter to $R_{max} = 150$ pc. Since we are searching for pinching due to the presence of a DD, including data beyond $r \sim 150$ pc could artificially make our constraints overly restrictive. Even with this cut, there remains non-negligible errors in the z-values of the stars due to parallax error, as shown in Fig. 2.1. We account for this in our analysis by applying a Gaussian kernel smoothing to our model predictions when comparing to data, as discussed further below.

Completeness and the Selection Function

Within the spatial cuts outlined above, the TGAS sample is not complete. For instance, the *Gaia* scanning pattern causes the relative completeness to vary over the sky because some regions do not

yet have sufficient observations to yield a high-quality astrometric solution [183]. To account for this, we use the results of Ref. [5], which compared the TGAS catalog counts to those of the Two Micron All-Sky Survey (2MASS) catalog [210] in order to determine the effective completeness as a function of position, color, and magnitude. In particular, we used the implementation provided in the `gaia_tools` package.[2] In the default settings of this package, cuts are placed on regions of the sky in order to select for those regions with reduced measurement uncertainties. In particular, the "good" part of the sky is defined as those pixels with sufficient number of observations in the *Gaia* scan strategy, small enough variations in the scan strategy for stars in that pixel, and an ecliptic latitude cut. After these default cuts, approximately 48% of the sky is selected, with a typical mean parallax uncertainty < 0.45 mas. As discussed in the previous subsection, we also impose further parallax uncertainty cuts and geometric sample region cuts. When taking the effective volume completeness into account, the tracer counts yield an optimal estimator for the true density with Poisson-distributed uncertainties [5].

To determine the number density of stars for a given spectral type, we then obtain the completeness for each spectral subtype (*e.g.* for A stars, we consider the A0, A1, ..., A9 stars individually), correct the observed number counts of stars using the effective completeness, and finally sum the subtypes. In total, our sample contains 1599 A stars, 16302 F stars, and 14252 early G stars, as compared with ~ 2000 stars that were used in the analysis of Ref. [187]. When including a three dimensional model of dust in the selection function, Ref. [5] found that the difference in stellar density distributions is typically 1-2% for a similar sample volume. We conservatively include a 3% systematic uncertainty on the density in each z-bin, which also includes uncertainties in the selection function as estimated in Ref. [5]. This is in addition to the statistical uncertainty in the stellar counts. In Fig. 2.4, we show the completeness for each of our tracer populations in the fiducial volume with $R_{\rm max} = 150$ pc. The selection efficiencies are fairly flat in z within our sample volume, such that we are never sampling the steeply falling efficiencies at large distances. It can be seen in Figs. 3-4 of Ref. [5] that the effects of dust extinction (which, as discussed above, are included in the 3% systematic error budget) are also small for these distances. The dashed lines in Fig. 2.4 are the result for the default selection on "good" parts of the sky, while the solid lines show the effect of including an additional cut on regions of the sky with average parallax uncertainty greater than 0.4 mas.

Midplane Vertical Velocities

In determining the vertical velocity distribution function, the primary source of uncertainty is on the radial (line of sight) velocities, which are not included in the TGAS catalog. In our analysis, we assume that on average the stars are co-rotating with the rest of the disk and thus the mean apparent radial velocities are simply given by Earth's proper motion relative to the disk projected along the line of sight to any given star. The mean radial velocity is given by $\langle v_r \rangle = -v_{x,0} \cos l \cos b - v_{y,0} \sin l \cos b - v_{z,0} \sin b$, where l is the Galactic longitude, b is the Galactic latitude, and where $v_{x,0} = 11.1 \pm 0.7^{\,\text{stat.}} \pm 1.0^{\,\text{sys.}}$ km/s and $v_{y,0} = 12.24 \pm 0.47^{\,\text{stat.}} \pm 2.0^{\,\text{sys.}}$ km/s capture the proper

[2] https://github.com/jobovy/gaia_tools

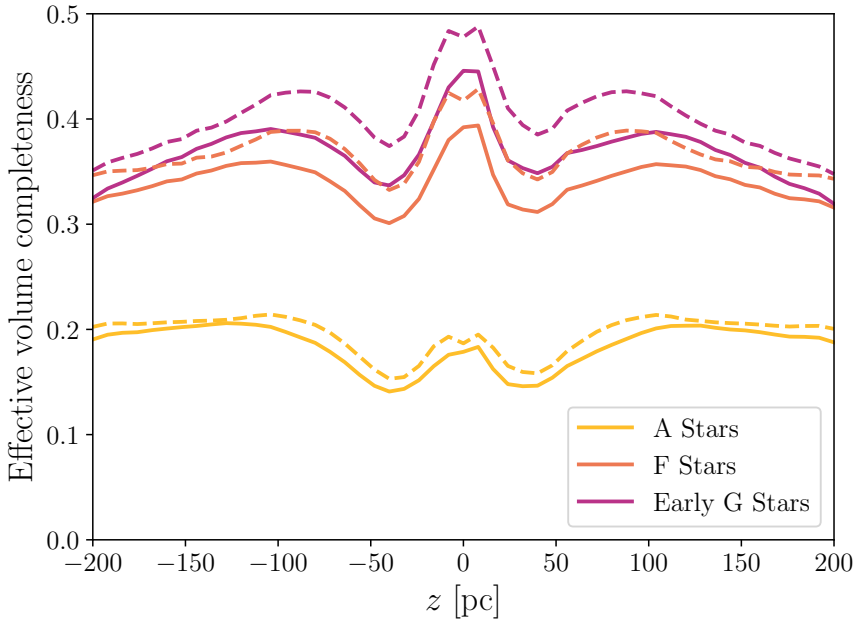

Figure 2.4: Effectiveness completeness for our fiducial selection volume, with radial cut of $R_{\rm max} = 150$ pc. The dashed lines show the result using the default cuts in `gaia_tools`, while the solid lines include an additional cut on regions of the sky with average statistical parallax uncertainty greater than 0.4 mas.

motion of the Sun inside the disk [211]. This mean radial velocity is as large as $\sim$ 10 km/s, depending on the angular position on the sky, and the projection onto motion in the z direction depends primarily on the Galactic latitude.

Thus, in determining the midplane velocity distributions of our tracers $f_{i,0}(|v|)$ we impose a latitude cut on stars so that the projection of the radial velocity onto the z direction is small. The projected vertical motions are given by $v_z = v_{z,0} + (\mu_b \kappa \cos b)/\tilde{\pi} + v_r \sin b$, where $v_{z,0}$ is the vertical velocity of the Sun, μ_b is the proper motion in Galactic latitude b in mas/yr, κ=4.74 is a prefactor which converts this term to units of km/s, $\tilde{\pi}$ is the parallax in mas, and v_r is the radial velocity in km/s. Note that this procedure is in contrast with other methods in the literature, such as the deconvolution approach used in Ref. [212]. For a latitude cut of $|b| < 5°$, the *largest* contribution to the vertical velocity from the mean radial velocity is $\Delta v_z \sim 1$ km/s. This large contribution is for the special case where the line of sight to the star is roughly in the same direction as Earth's motion relative to the Galactic plane. In other parts of the sky, the radial velocity contributes a smaller uncertainty to the vertical velocity with this latitude cut.

The effect of the radial velocity uncertainty decreases as we make tighter latitude cuts. However, tighter cuts also reduce our statistics. In Fig. 2.5 we compare the $f_0(v)$ for the F stars inferred from the data with different latitude cuts. We find good agreement between the different latitude

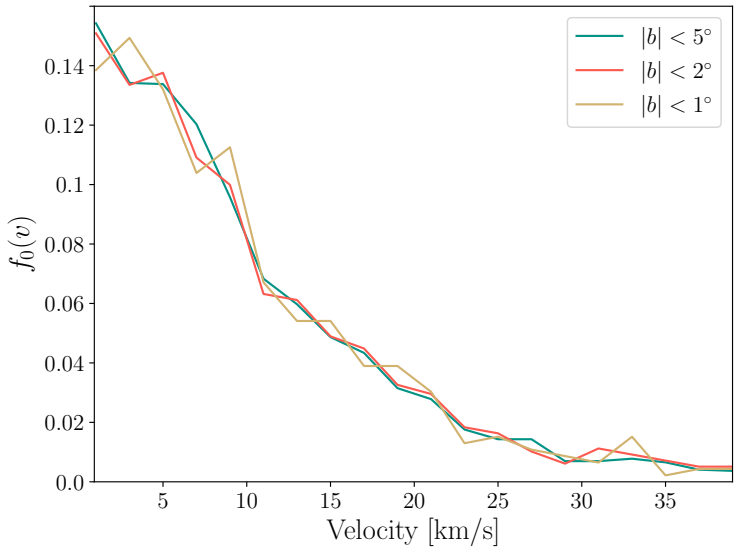

Figure 2.5: Using the F stars as a representative case, we illustrate the effect of different latitude cuts on the inferred midplane velocity distribution $f_0(|v|)$. Tighter b cuts help to mitigate the uncertainty in the radial velocity but add statistical noise. We find that within statistical uncertainties, the distribution function is not sensitive to the different latitude cuts shown here so for our analysis we use $|b| < 5°$.

cuts within statistical uncertainties, so we take $|b| < 5°$ by default. Still, the uncertainty in the radial velocities cannot be neglected; this can be seen in the left panel of Fig. 2.6, where we compare the case assuming average radial velocities to the case with radial velocities set to zero. We use the difference between these two to estimate the size of the systematic error due to radial velocities.

Also shown in Fig. 2.6 are the effects of the dependence of the selection function on spectral type. The selection function does not depend on velocity, although the effective completeness does vary for different spectral types (see for example Fig. 2.4). A good consistency check is to make sure that when subdividing stellar types into subtypes (*e.g.* dividing A stars into types A0-A9) and weighting $f_0(v)$ accordingly by the subtype selection functions that we do not find a discrepancy with the unweighted $f_0(v)$. We find that the data are consistent between these two procedures, with the largest difference being for A stars, which have the lowest counts. We include the difference between the $f_0(v)$ data with and without the selection function weighting as part of our estimate of the systematic uncertainty in $f_0(v)$.

Out-of-equilibrium effects

A key assumption of our analysis is that the Galactic disk is locally in equilibrium. However, there are observations that suggest the presence of out-of-equilibrium features, such as bulk velocities, asymmetric density profiles about the Galactic plane, breathing density modes, and vertical offsets

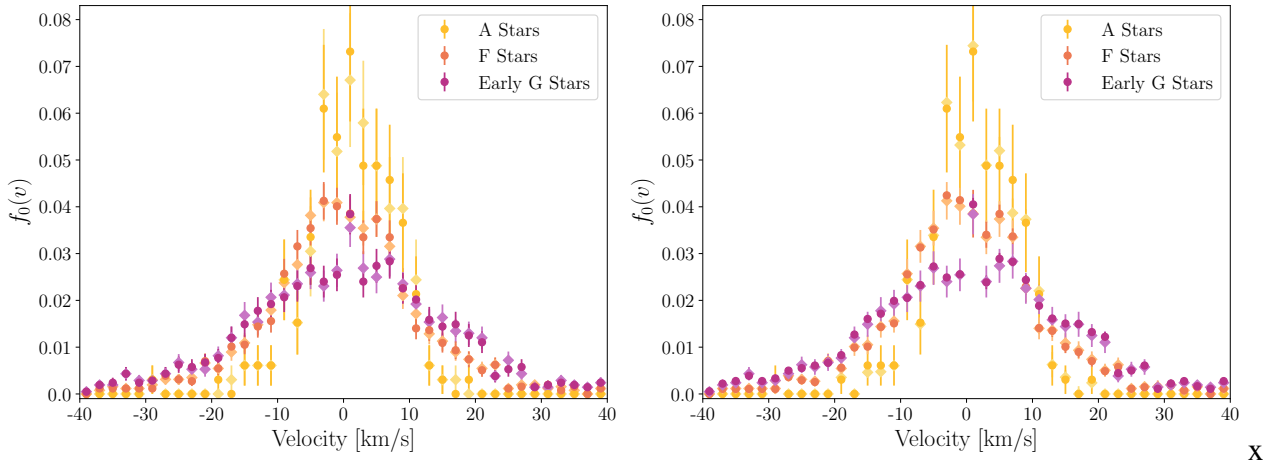

Figure 2.6: Left: The midplane velocity distribution function for different stellar types assuming mean $v_r = \langle v_r \rangle$ (dark circles) and $v_r = 0$ (light diamonds). The latitude cut $|b| < 5°$ accounts for the smallness of the difference between these situations, which is well within the purely statistical uncertainties shown here as error bars. Right: The midplane velocity distribution function for different stellar types when appropriately weighting the stellar subtypes by their selection functions (dark circles) and when ignoring the subtype selection functions (light diamonds). Since the selection functions do not correlate with velocity but rather with position and magnitude, the two procedures are expected to be very consistent within the purely statistical uncertainties shown here as error bars. Indeed, the two procedures give very similar determinations of $f_0(v)$, with the biggest difference coming from the low-statistics A stars.

between populations [213–216]. Such features could also be present in the DM components. While these effects are typically manifest higher above the Galactic plane than what we consider, it is still important to account for the possibility that the disk is not in equilibrium.

However, our tracer samples appear to obey the criteria for an equilibrium disk. When adjusting for the height of the Sun above the Galactic plane (which we find to be -1.3 ± 4.6 pc, consistent with Refs. [5] and [217]), we do not find any significant asymmetry in the density profile above and below the Galactic plane. We find no significant difference in the vertical velocity distribution function above and below the Galactic plane, unlike Ref. [218] which claimed evidence for a contracting mode. We also find that the midplane velocity distribution function is symmetric about $v = 0$. We find the vertical Solar velocity $vz_{,0} = 6.8 \pm 0.2$ km/s, consistent with the measurement of Ref. [211], and has the expected Gaussian profile of a static isothermal population [203]. To be conservative, we take the systematic uncertainty from non-equilibrium dynamics to be the difference between the $+|v|$ and $-|v|$ values of $f_0(|v|)$. These estimated non-equilibrium systematics, in combination with the systematic uncertainties from radial velocities and selection function weighting, are included along with statistical uncertainties in Fig. 2.7 for 2 km/s velocity bins.

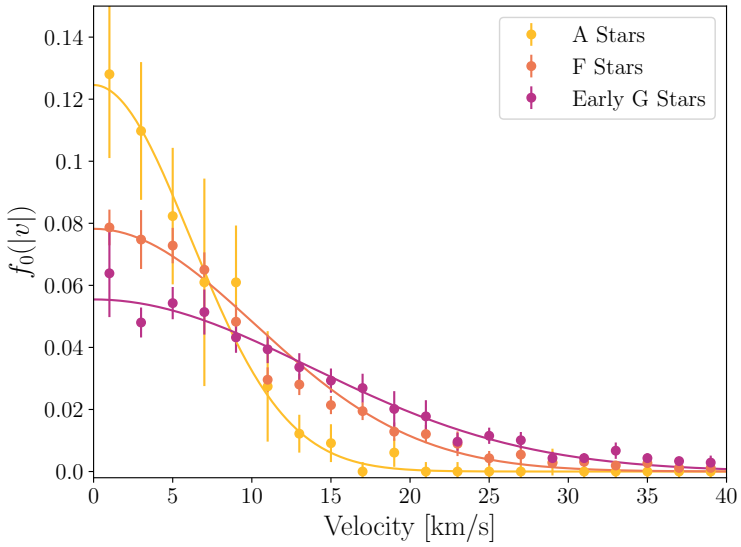

Figure 2.7: The midplane velocity distribution function $f_0(|v|)$ used in our analysis. The error bars are a combination of statistical and systematic uncertainties stemming from radial velocity uncertainties, selection function weighting, and non-equilibrium-induced asymmetry in $f_0(v)$. Also shown are the best-fit Gaussian functions, which have velocity dispersions of 6.1 ± 0.1, 10.2 ± 0.2 and 13.7 ± 0.1 km/s for the A, F and G stars, respectively.

Our treatment differs from the out-of-equilibrium analysis of Ref. [187], which evolves the *observed* tracer density profile as it oscillates up and down through the spatially-fixed potential of other mass components (including a DD), while determining the error on this evolution through bootstrapping. This results in a band of possible tracer profiles that could be caused by a DD in the presence of disequilibria. In contrast, our approach treats all the data on equal footing. Since changing $f_0(v)$ can potentially mimic the pinching effect from a DD, our analysis accounts for the possibility that pinching arises from fluctuations or systematics in $f_0(v)$. Thus, our analysis also scans over an analogous band of tracer profiles.

We perform a final consistency check by breaking down our tracer sample into subpopulations with different velocity dispersions, which are affected differently by disequilibria due to their different mixing timescales. In the presence of out-of-equilibrium features, separate analyses of these different subpopulations could yield discrepant parameters [219]. As detailed in the SM, however, we find broad agreement between the subpopulations.

2.5 Likelihood analysis

We search for evidence of a thin DD by combining the model and datasets described above with a likelihood function. Here we summarize our statistical analysis.

We use a likelihood function to fit a model $\mathcal{M}$ to the data d, where the model has parameters $\boldsymbol{\theta} = \{\boldsymbol{\psi}, \boldsymbol{\zeta}\}$; the $\boldsymbol{\psi}$ are the parameters of interest and the $\boldsymbol{\zeta}$ are the nuisance parameters. For a single tracer population, the dataset d consists of the log of the binned vertical star counts $\ln \nu_i^{\rm dat}$, where i labels one of the z bins, and the midplane binned velocity distribution $f_j^{\rm dat}$, where j labels one of the velocity bins. Additionally, the dataset contains uncertainties $\sigma_{\ln \nu_i}$ and σ_{f_j} on the number density and velocity distribution measurements, respectively, that account for both statistical and systematic uncertainties. Our nuisance parameters include a parameter $z_{\rm sun}$ for the height of the Sun above the disk, the local DM density from the bulk halo $\rho_{\rm DM}$, baryonic nuisance parameters ρ_k and σ_{v_k} for the densities of velocity dispersions of each of the N_b baryonic components, indexed by k, listed in Tab. 2.1, parameters f_ν that describe the overall normalization of the vertical star counts distribution for each tracer population, and nuisance parameters f_j that describe the normalization of the midplane velocity distribution of the N_t tracer populations in each of the N_v velocity bins. Thus, the total number of nuisance parameters is $N_t + N_t N_v + 2\,N_b + 2$. Our parameters of interest $\boldsymbol{\psi}$ are the thin DD surface density Σ_{DD} and scale height h_{DD}. Note that given a parameter space vector $\boldsymbol{\theta}$, we calculate the number density of stars $\nu_i(\boldsymbol{\theta})$ in each of the N_z z bins though the iterative procedure described in Section 2.2.

More concretely, the predicted z-distribution of stars is a function of the DD model parameters (namley the DD scale height and surface density) and nuisance parameters, which consist of: *(i)* the 12 baryonic densities in Tab. 2.1, along with their velocity dispersions; *(ii)* the local DM density in the halo $\rho_{\rm DM}$; *(iii)* the height of the Sun; *(iv)* the midplane stellar velocity distribution $f_0(v_j)$, where j indexes the velocity bins. The velocity distributions are given Gaussian priors in each velocity bin with central values and widths as described in the previous Section. The baryon densities and velocity dispersions are also given Gaussian priors with the parameters in Tab. 2.1. The height of the Sun above the disk and local DM density are given linear priors that encompass a broad range of previous measurements, $z_{\rm sun} \in [-30, 30]$ pc and $\rho_{DM} \in [0, 0.06]\,{\rm M}_\odot/{\rm pc}^3$ [5, 206, 217, 220, 221]. When combining stellar populations, we use a shared mass model but compute the densities of the A, F, and G stars independently and give their velocity distributions independent nuisance parameters. In analyzing all three stellar populations, we have 89 nuisance parameters.

As shown in Fig. 2.1, parallax uncertainties lead to a larger uncertainty on the height above the midplane at larger heights. Using this information, we can model what the true density profile would look like *on average* after a simulated TGAS measurement in the presence of parallax uncertainty. We apply a Gaussian kernel to smooth the $\nu_i(\boldsymbol{\theta})$ by the TGAS parallax uncertainties, with the dispersion of the kernel varying as a function of z as inferred from Eq. (2.17), before comparing the model predictions for $\nu_i(\boldsymbol{\theta})$ to data. After taking the effects of parallax uncertainty into account in our prediction, we compare to the data through the likelihood function

$$p_\nu(d|\mathcal{M}, \boldsymbol{\theta}) = \prod_{i=1}^{N_z} \frac{1}{\sqrt{2\pi\sigma^2_{\ln \nu_i}}} \exp\left[-\frac{\left[\ln(f_\nu\,\nu_i(\boldsymbol{\theta})) - \ln \nu_i^{\rm dat}\right]^2}{2\sigma^2_{\ln \nu_i}}\right] . \tag{2.18}$$

The total likelihood function for a single population in isolation is then given by the above multi-

plied by the appropriate prior distributions for the baryons and the stellar velocities:

$$
\begin{aligned}
p(d|\mathcal{M},\boldsymbol{\theta}) &= p_\nu(d|\mathcal{M},\boldsymbol{\theta}) \times p_f(d|\mathcal{M},\boldsymbol{\zeta}) \times p_b(d|\mathcal{M},\boldsymbol{\zeta})\,, \\
p_f(d|\mathcal{M},\boldsymbol{\zeta}) &= \prod_{j=1}^{N_v} \frac{1}{\sqrt{2\pi\sigma_{f_j}^2}} \exp\left[-\frac{(f_j - f_j^{\rm dat})^2}{2\sigma_{f_j}^2}\right]\,; \\
p_b(d|\mathcal{M},\boldsymbol{\zeta}) &= \prod_{k=1}^{N_b} \left(\frac{1}{\sqrt{2\pi\sigma_{\rho_k}^2}} \exp\left[-\frac{(\rho_k - \rho_k^{\rm dat})^2}{2\sigma_{\rho_k}^2}\right]\right)\left(\frac{1}{\sqrt{2\pi\sigma_{\sigma_{v_k}}^2}} \exp\left[-\frac{(\sigma_{v_k} - \sigma_{v_k}^{\rm dat})^2}{2\sigma_{\sigma_{v_k}}^2}\right]\right).
\end{aligned}
\tag{2.19}
$$

Note that the prior distributions on the stellar velocities p_f and baryons p_b are only functions of the nuisance parameters. When we combine N_t tracer populations, indexed by ℓ, the likelihood function instead becomes

$$
p(d|\mathcal{M},\boldsymbol{\theta}) = p_{\rm baryon}(d|\mathcal{M},\boldsymbol{\zeta}) \times \left[\prod_{\ell=1}^{N_t} p_\nu^\ell(d|\mathcal{M},\boldsymbol{\theta}) \times p_f^\ell(d|\mathcal{M},\boldsymbol{\zeta})\right]. \tag{2.20}
$$

Note that baryonic nuisance parameters and the nuisance parameter for the local DM density in the bulk halo are shared between all N_t tracer populations, as is the nuisance parameters for the position of the Sun, while each population is given separate nuisance parameters for the binned velocities and normalization for $\nu(z)$.

Given the likelihood function, we construct likelihood profiles $\lambda(\Sigma_{DD})$ at fixed DD scale heights h_{DD}:

$$
\lambda(\Sigma_{DD}) = 2\big[\max_{\boldsymbol{\zeta}} \log p(d|\mathcal{M},\boldsymbol{\theta}) - \max_{\boldsymbol{\zeta},\Sigma_{DD}} \log p(d|\mathcal{M},\boldsymbol{\theta})\big]\,. \tag{2.21}
$$

Above, the second term denotes the maximum log likelihood taken by maximizing over the nuisance parameters and Σ_{DD}, at fixed h_{DD}, while the first term is the maximum log likelihood at fixed h_{DD} and Σ_{DD}. The likelihood profile is only strictly defined for Σ_{DD} greater than the Σ_{DD} of maximum likelihood. The 95% upper limit on Σ_{DD} is given by the value of Σ_{DD} for which $\lambda(\Sigma_{DD}) = -2.71$ [222]. The TS, on the other hand, is used to quantify the significance of a detection, and is defined analogously:

$$
\mathrm{TS} = 2\big[\max_{\boldsymbol{\theta}} \log p(d|\mathcal{M},\boldsymbol{\theta}) - \max_{\boldsymbol{\zeta}} \log p(d|\mathcal{M},\boldsymbol{\theta})|_{\Sigma_{DD}=0\,M_\odot/\mathrm{pc}^2}\big]\,. \tag{2.22}
$$

That is, the TS is twice the log-likelihood difference between the best-fit DD model and the null model, which has $\Sigma_{DD} = 0\,M_\odot/\mathrm{pc}^2$.

We present our results in terms of both frequentist analyses, following the statistical treatment described above, and Bayesian analyses. Both types of statistical analyses use the likelihood function in Eq. (2.21). The Bayesian analyses proceed through Bayes' theorem:

$$
p(\theta|\mathcal{M},d) = \frac{p_\nu(d|\mathcal{M},\theta)\,p(\theta|\mathcal{M})}{p(d|\mathcal{M})}\,. \tag{2.23}
$$

Above, $p(\theta|\mathcal{M})$ denotes the prior distribution; the combination of our prior and likelihood functions, $p_\nu(d|\mathcal{M},\theta)\,p(\theta|\mathcal{M})$, is given in Eq. (2.20). The posterior distribution is given by $p(\theta|\mathcal{M},d)$, while the Bayesian evidence (sometimes also referred to as the marginal likelihood or model likelihood) is found through the integral

$$p(d|\mathcal{M}) = \int d\theta\, p(d|\mathcal{M},\theta)\, p(\theta|\mathcal{M})\,. \tag{2.24}$$

When performing parameter estimation in the Bayesian framework, we integrate the posterior distribution over all model and nuisance parameters except for our specific parameter of interest and then calculate the appropriate percentiles of the resulting one-dimensional posterior distribution. We also, in the main Letter, compare nested models using the Bayes factor. The Bayes factor in preference for a model $\mathcal{M}_A$ over a model $\mathcal{M}_B$ is given by the evidence ratio

$$\mathcal{B}_{AB} = \frac{p(d|\mathcal{M}_A)}{p(d|\mathcal{M}_B)}\,. \tag{2.25}$$

2.6 Validating our Analysis with Mock Data

In this section we test our analysis framework on simulated TGAS data. By analyzing simulated data with an injected DD signal, we show that our analysis framework is able to appropriately reconstruct the injected parameters and importantly we demonstrate that the resulting limits to do not exclude the true DD parameters.

Mock Data Generation

We create an ensemble of mock datasets in the following way. First, we assume a mass model. For the baryons, we simply take the centers of the priors for the midplane densities and velocity dispersions. For every mock dataset we set $z_{\rm sun} = 0$ pc and halo DM density $\rho_{\rm DM} = 0.01\,{\rm M}_\odot/{\rm pc}^3$. We optionally include a DD in the mass model for testing the recovery of injected DD parameters.

The velocity distributions of the three tracer populations are drawn assuming Poisson fluctuations in each velocity bin, where we assume the true $f_0(|v|)$ is given by the measured distribution of that population. Using the mass model and the randomly drawn velocity distribution function, we then calculate the gravitational potential and density profile $\nu(z)$ for each tracer population. The calculated density profile is normalized to the observed number of tracer stars in the TGAS data over the range $z \in [-200, 200]$ pc; then, using this distribution, we randomly generate both the z-positions of stars as well as the total number of stars assuming Poisson noise. Distance uncertainties due to parallax error are also included in this data set. For every star with distance z_i we introduce a random Gaussian smearing on z_i with width given by $\sigma(z_i)$ in Eq. 2.17. Finally, we assign both statistical and systematic uncertainties to the mock data: for the number densities, we assume a 3% systematic as for the density profiles, while for the velocity distributions we assume 15%, 5%, and 7% systematic errors for A, F, and Early G stars, respectively, which roughly reproduce the uncertainty levels we find in the real TGAS data.

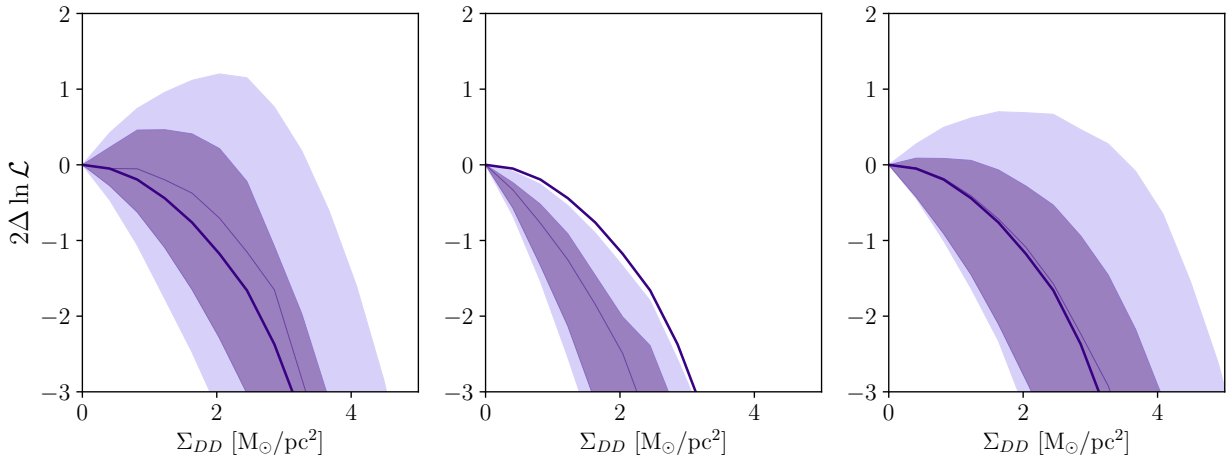

Figure 2.8: Left: The likelihood profile when no parallax error is introduced into the mock datasets and with no Gaussian kernel density smoothing. Center: Same as the left but with parallax error introduced into the mock datasets, causing artificially tight constraints. Right: The recovered likelihood profile when parallax error is present in the mock data and when the Gaussian kernel is used in the analysis to model the effects of parallax error. The likelihood profile using the Gaussian kernel has similar statistics and yields a similar limit on a DD under the null hypothesis as the case with no parallax error. In all three panels, we take $h_{DD} = 5$ pc and no DD is injected. The dark (light) shaded regions correspond to the 68% (95%) containment regions, and the thick line corresponds to the results of the Asimov analysis.

In addition to generating mock datasets with statistical fluctuations, we also generate the Asimov dataset [222]. The Asimov dataset consists of data that are identical to the prediction under the hypothesis being tested, with no statistical fluctuations, even if this means using fractional counts in cases where the data consists of binned counts. In this case, we generate data that lies exactly on top of the predicted density profile for a given mass model. As shown in Ref. [222], the mean likelihood profile over an ensemble of simulated datasets converges to the likelihood profile obtained with the Asimov dataset in the vicinity of the point of maximum likelihood. This provides a way of cross-checking both the likelihood function and the mock-data generation framework.

In Fig. 2.8, we show the likelihood profiles for a sample of mock datasets, as compared with the likelihood profiles computed on the Asimov dataset. When the datasets are generated without scatter due to parallax uncertainty and compared to models without Gaussian kernel smearing, the resulting likelihood profiles behave as expected from the Asimov dataset (left panel). However, when datasets are generated *with* scatter from parallax uncertainty and compared to models *without* Gaussian kernel smoothing, we obtain artificially steep likelihood profiles (center panel). Finally, we analyze mock datasets with parallax uncertainty and Gaussian kernel smoothing (right panel), finding agreement with the expectation from the Asimov dataset. These results indicate the importance of including the Gaussian kernel smoothing for obtaining the correct likelihood profiles, and

this also verifies that we have adequately accounted for these effects in our modeling.

Results of Mock Data Analysis

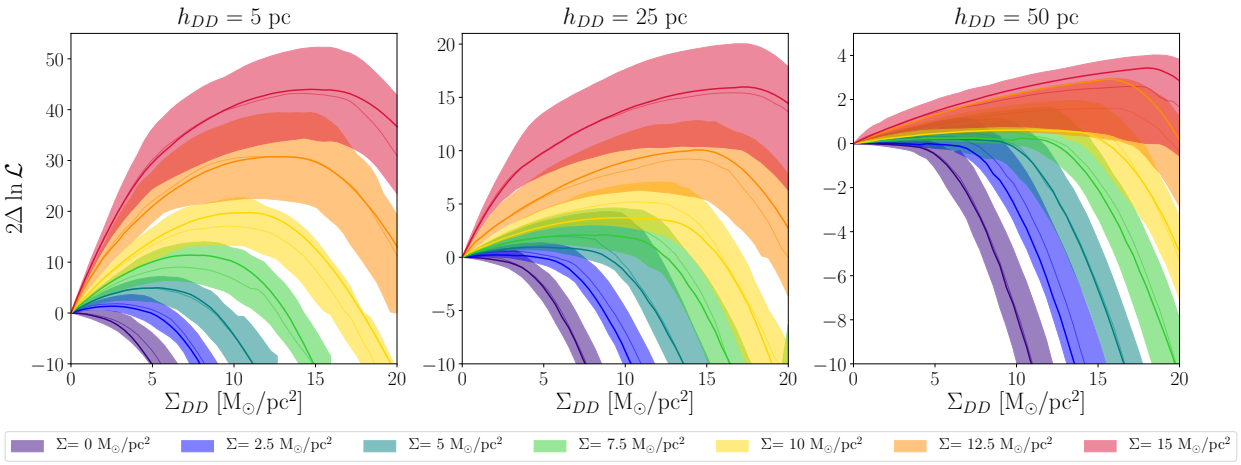

Figure 2.9: The likelihood profiles $2\Delta \ln \mathcal{L}$ for mock data with injected DD signals. Different colors correspond to different values of the injected DD surface density. The bands show the 68% containment regions for the 50 mock datasets for each set of parameters. The thick line is the likelihood profile of the Asimov dataset, while the thin line is the median likelihood profile of the 50 mock datasets.

Using the methods outlined above, we generate mock datasets with DDs of varying surface density between 0 and 15 M$_\odot$/pc^2 and scale heights $h_{DD} = 5, 25$, and 50 pc. For each set of DD parameters, we generate 50 such datasets including the effects of parallax uncertainties and run these datasets through our analysis with Gaussian kernel smoothing. In Fig. 2.9, for each value of h_{DD} we show the resulting likelihood profiles in Σ_{DD}. Thinner DDs with larger surface density (and corresponding higher midplane density) are detected with much higher significance using our analysis. For the thickest DD we inject into the data, even large surface density DDs are not detected with high significance, in part because the density $\sim \Sigma_{DD}/4h_{DD}$ is lower and also because there is a possible degeneracy of the DD with the baryonic disk and halo DM.

The top panel of Fig. 2.10 compares the recovered best fit Σ_{DD} (green band) with the injected value (dashed line), where we find good agreement. We also show the 95% limit obtained from the data sets with injected signals, indicated by the blue band. As desired, we never rule out an injected signal. The fact that even the 95% containment region of the limit (light blue) does not have any overlap with the injected signal (dashed line) means that our "95%" exclusion — which is determined by assuming that the likelihood profile is χ^2-distributed in the vicinity of the point of maximum likelihood — is likely conservative.

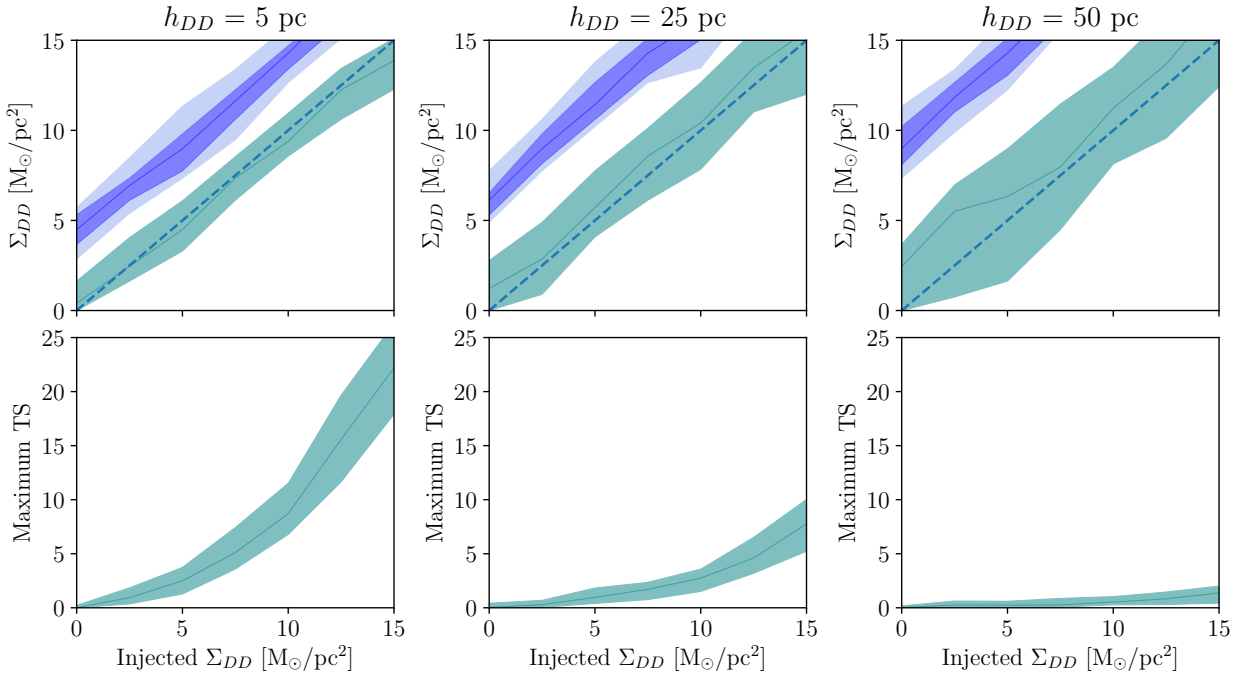

Figure 2.10: Top: For mock data with injected DD signals, the recovered value of Σ_{DD} (green band, at 68% containment) and the 95% upper limit on Σ_{DD} as determined by $2\Delta \ln \mathcal{L} = -2.71$ (blue bands). The dark (light) blue bands represent the 68% (95%) containment regions for the upper limits determined by 50 mock datasets. Crucially, we never exclude the injected signal, which is indicated by the dashed diagonal line. Bottom: The 68% containment region for the maximum TS obtained in our analysis, which gives a measure of the detection strength of the recovered signal.

2.7 Results

In this Section, we provide the results of our analysis procedure on the real TGAS data. We show our composite frequentist results as well as Bayesian and frequentist results for individual tracer populations. We also show how the analysis is affected by alternate binning and gas densities.

Main Results

As detailed in Section 2.5, we compute likelihood profiles as functions of Σ_{DD} by profiling over the nuisance parameters for fixed h_{DD}. From the likelihood profiles we compute the 95% one-sided limit on Σ_{DD}, which is shown in Fig. 2.11. We also compare our limit to the expectation under the null hypothesis, which is generated by analyzing multiple simulated TGAS datasets where we assume the fiducial baryonic mass model and include $\rho_{\rm DM} = 0.014$ M$_\odot$/pc^3. We present the

68% and 95% containment region for the expected limits at each h_{DD} value. The TGAS limit is consistent with the Monte Carlo expectations at high h_{DD} but becomes weaker at low h_{DD}. This deviation is also manifest in the test-statistic (TS), which is defined as twice the difference in log-likelihood between the maximum-likelihood DD model and the null hypothesis. We find TS ~ 5 at $h_{DD} \sim 5$ pc and $\Sigma_{DD} \sim 4$ $M_\odot/pc^2$; while this does indicate that the best-fit point has a nonzero DD density, the TS is not statistically significant. Moreover we cannot exclude the possibility that, the true evidence in favor of the DD is much lower due to the possible existence of systematic uncertainties that are not captured by our analysis.

For the limit given in Fig. 2.11, we also show the complete likelihood profiles over the range $h_{DD} = 5 - 100$ pc in Fig. 2.12 (left panel). The excess seen in Fig. 2.11 is also manifest in the test-statistic (TS), which is defined as twice the difference in log-likelihood between the maximum-likelihood DD model and the null hypothesis. In the right panel of Fig. 2.12 we find TS ~ 5 at $h_{DD} \sim 5$ pc and $\Sigma_{DD} \sim 4$ $M_\odot/pc^2$; while this does indicate that the best-fit point has a nonzero DD density, the TS is not statistically significant. In this panel we also show the expectation from 100 mock datasets generated with no DD, similar to the band shown in Fig. 2.11. Although our TS value is well outside of the 95% containment region of the mock data (light green band), it is not easy to interpret the statistical significance of this result due to the possibility of hidden systematics. If the fiducial model is adjusted to have larger baryonic densities or larger halo DM density, then the expected band of TS values will also increase compared to that shown. It is also possible that the assumptions going into our vertical Poisson-Jeans modeling do not fully capture the dynamics.

Component	A Stars	F Stars	Early G Stars
baryons	$0.088^{+0.006}_{-0.006}$	$0.088^{+0.007}_{-0.007}$	$0.085^{+0.007}_{-0.006}$
DM	$0.038^{+0.012}_{-0.015}$	$0.019^{+0.012}_{-0.011}$	$0.004^{+0.01}_{-0.004}$

Table 2.2: Posteriors on the total baryonic and DM halo density at the midplane, in units of $M_\odot/pc^3$.

The model without the thin DD provides insight into the baryonic mass model and the local properties of the bulk DM halo. While the DD limits described above were computed in a frequentist framework, we analyze the model without a DD within a Bayesian framework for the purposes of parameter estimation and model comparison. The marginalized Bayesian posterior values for the total baryonic density and local DM density are given in Tab. 2.2 for analyses considering the three stellar populations in isolation. Despite only analyzing data in a small sample volume, we find mild evidence in favor of halo DM: for the model with halo DM compared to that without, the Bayes factors are $\sim$8.4 and 1.9 using the A and F stars, respectively, while for early G stars the Bayes factor $\sim$0.4 is inconclusive. We show the best-fit profiles of our tracer stars in Fig. 2.13 with statistical and systematic uncertainties on the tracer density and velocity profiles; these profiles for different tracer populations are generated with the same underlying maximum likelihood baryonic parameters assuming no DD. As can be seen, the model with no DD reproduces the tracer profiles remarkably well.

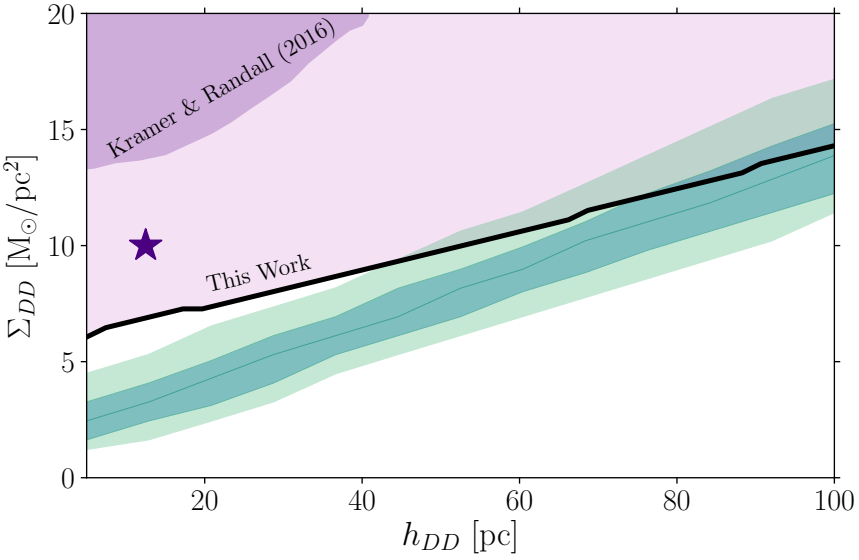

Figure 2.11: The 95% constraint on the DD surface density Σ_{DD} as a function of the scale height h_{DD}, as found in this work and in Ref. [187] (Kramer & Randall 2016). The star indicates fiducial DD parameters that can account for phenomena such as periodic comet impacts [172]. Also shown is a comparison of the limit to the 68% and 95% containment regions (in dark and light green, respectively) on the expected limit from simulated data generated under the null hypothesis of no DD.

Separate Analyses of Tracer Populations

Here we compare the joint frequentist analysis including A, F, and early G stars to the frequentist limits separately obtained on the individual tracer populations in the left panel of Fig. 2.14. Here we separately marginalize over the height of the Sun, $\rho_{\rm DM}$ and baryons for each tracer population, as described in Eq. 2.19. Due to the factor of $\sim$10 fewer stars in the A star sample, our limits are mainly driven by the F and early G stars. The F and early G stars have similar statistics and systematic errors, however the F star data favors a higher DD density and the limits are overall weaker. The right panel of Fig. 2.14 shows the TS values obtained from the separate analyses, which are lower than that from the joint analysis since the joint result requires the baryon parameters are shared for the three populations.

Using the nested sampling code `Multinest` [1], we have also separately analyzed the individual tracer populations to obtain posteriors on the DD and nuisance parameters. This allows us to check consistency of results between the tracers, while differences could be sensitive to nonequilibrium effects as discussed in Section 2.4. Including all nuisance parameters and a thin DD, we find the posterior distributions shown in Figs. 2.15-2.17 for the DD parameters, halo DM $\rho_{\rm DM}$, and the total baryon density ρ_b. There is a degeneracy between halo DM and thin disk DM, as well as a broad posterior distribution for h_{DD}, such that the z-dependence of the DM profile is poorly determined near the midplane. In Fig. 2.18, we overlay the results for the DD parameters and $\rho_{\rm DM}$ for

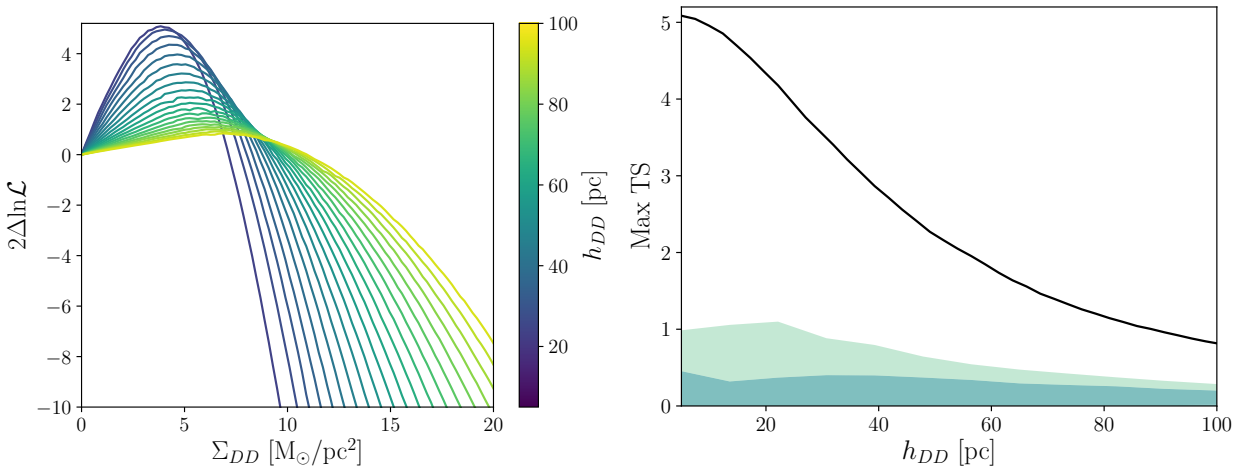

Figure 2.12: Left: For each h_{DD}, we show the likelihood profile used to set 95% one-sided bounds on Σ_{DD} in our fiducial analysis. Right: For the same analysis, we show the maximum TS as a function of h_{DD} (black line). The dark and light green regions are 68% and 95% containment regions for simulated data generated under the null hypothesis of no DD.

the three populations, finding that they agree within the $\sim$1-2 σ containment regions. The results in a Bayesian framework are consistent with the limits obtained with the profile likelihood method: for a thin dark disk $h_{DD} \approx 10$ pc, we see that the maximum DD surface density within the 2σ posterior distribution is $\Sigma_{DD} \approx 6 - 7$ M$_\odot$/pc^2 for all tracer populations. We also find consistency in the height of the sun above the Galactic plane for all three tracer populations: for the A, F, and early G stars we find $z_{\text{sun}} = -4.0^{+2.4}_{-2.6}$, $1.8^{+1.9}_{-2.0}$, and $-1.6^{+3.7}_{-3.7}$ pc, respectively.

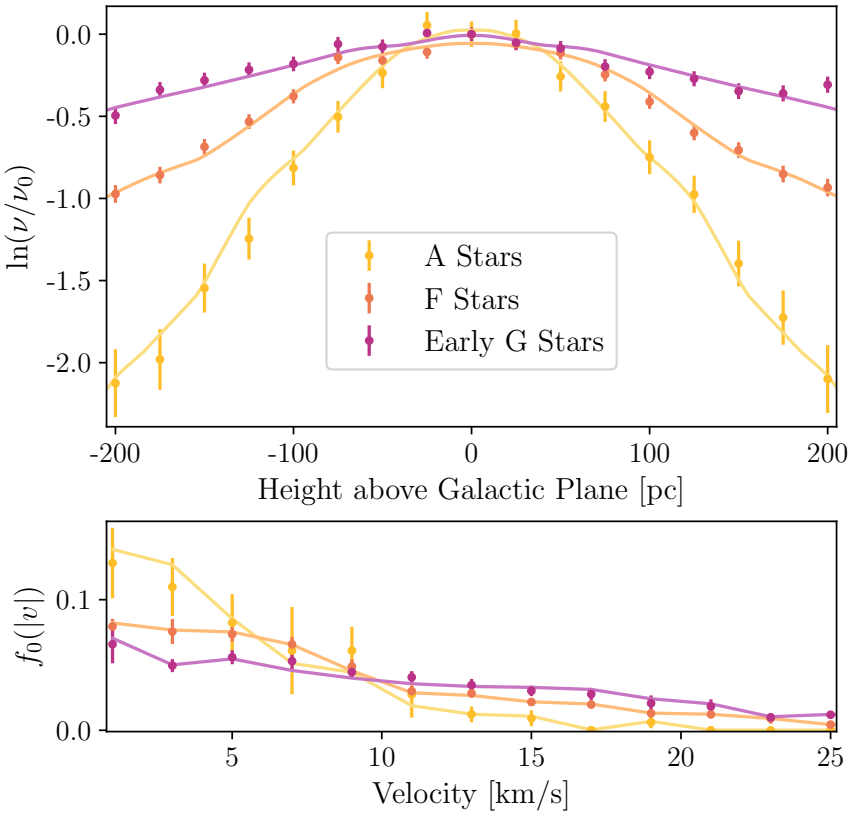

Figure 2.13: The measured number density profiles (top) and velocity distributions (bottom) for the tracer stars in our sample volume, subdivided by spectral type. The solid lines are the best-fit from our analysis assuming no DD.

In Fig. 2.19 we show posterior distributions in $\rho_{\rm DM}$ and ρ_b for an analysis where no DD is included, which gives a measure of the total midplane density under the null hypothesis. The corresponding posteriors on $\rho_{\rm DM}$ and ρ_b are given in Table. 2.2. Again, we find consistency between the three populations for the $\sim$1-2 σ containment regions.We have additionally used these distributions to compute the posterior on the baryon surface density in our sample volume as $\Sigma = 25.4^{+1.6}_{-1.6}$, $26.0^{+1.9}_{-1.9}$, and $26.1^{+1.6}_{-1.5}$ $M_\odot/pc^2$ for the A, F, and G stars, respectively. We can also extrapolate our baryon density profiles out to $z = 1.1$ kpc to compute $\Sigma_{1.1}$, a metric often quoted in the literature. In doing this, we extend the assumption of isothermal mass components to beyond the regime for our analysis, where it is expected that velocity dispersions are z-dependent. Nevertheless, we find surface densities that agree well with those in the literature. For the A, F, and early G stars we find $\Sigma_{1.1} = 40.7^{+4.0}_{-4.9}$, $44.6^{+4.9}_{-5.5}$, and $48.5^{+5.0}_{-4.8}$ $M_\odot/pc^2$, respectively. The larger errors at large z reflect the fact that our extrapolation is sensitive to small variations in the profile within $z < 200$ pc.

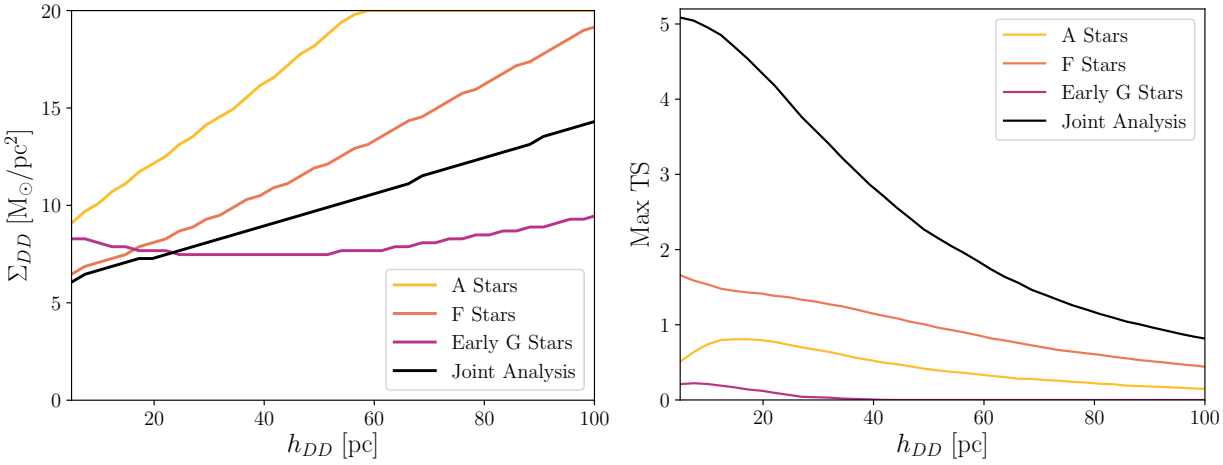

Figure 2.14: Left: The joint analysis on all three tracer populations is compared with the limit obtained from separate analyses on the individual populations. While the limits from only A stars are noticeably weaker, there are roughly ten times fewer stars than either F or Early G stars. Right: The maximum TS as a function of h_{DD} for each tracer population as well as the joint analysis.

Dependence of Results on Binning and Gas Densities

As discussed in Section 2.4, our choice of binning for the data was motivated by the size of typical measurement uncertainties in z and vertical velocities v. In Fig. 2.20 (left panel), we show how the limits change with alternate binning choices, finding our conclusions are robust to these differences. Although the limits are somewhat weaker at large h_{DD} with different z-bin or v-bin widths, the bounds for a thin DD at $h_{DD} \sim 10$ pc are unchanged.

The right panel of Fig. 2.20 gives limits under the assumption of a different mass model. Our main mass model in Table 2.1 drew on gas densities obtained in Ref. [193] (McKee et al.), while the revised estimates of Ref. [204] (Kramer & Randall) found lower gas densities. The comparison is given in Table 2.3. With the Kramer & Randall gas densities, a larger DD density is allowed and we find moderately weaker limits; however, the 95% bound still excludes surface densities of $\Sigma_{DD} \sim 10$ M$_\odot$/pc^2 for a thin DD with $h_{DD} \sim 10$ pc.

Gas Component	Kramer & Randall $\rho(0)$ [M$_\odot$/pc^3]	McKee et al. $\rho(0)$ [M$_\odot$/pc^3]
Molecular Gas (H_2)	0.014 ± 0.005	0.01 ± 0.003
Cold Atomic Gas ($H_I(1)$)	0.015 ± 0.003	0.028 ± 0.006
Warm Atomic Gas ($H_I(2)$)	0.005 ± 0.001	0.007 ± 0.001
Hot Ionized Gas (H_{II})	0.0011 ± 0.0003	0.0005 ± 0.00002

Table 2.3: Midplane gas densities from the Kramer & Randall [204] compared with those from McKee et al. [193].

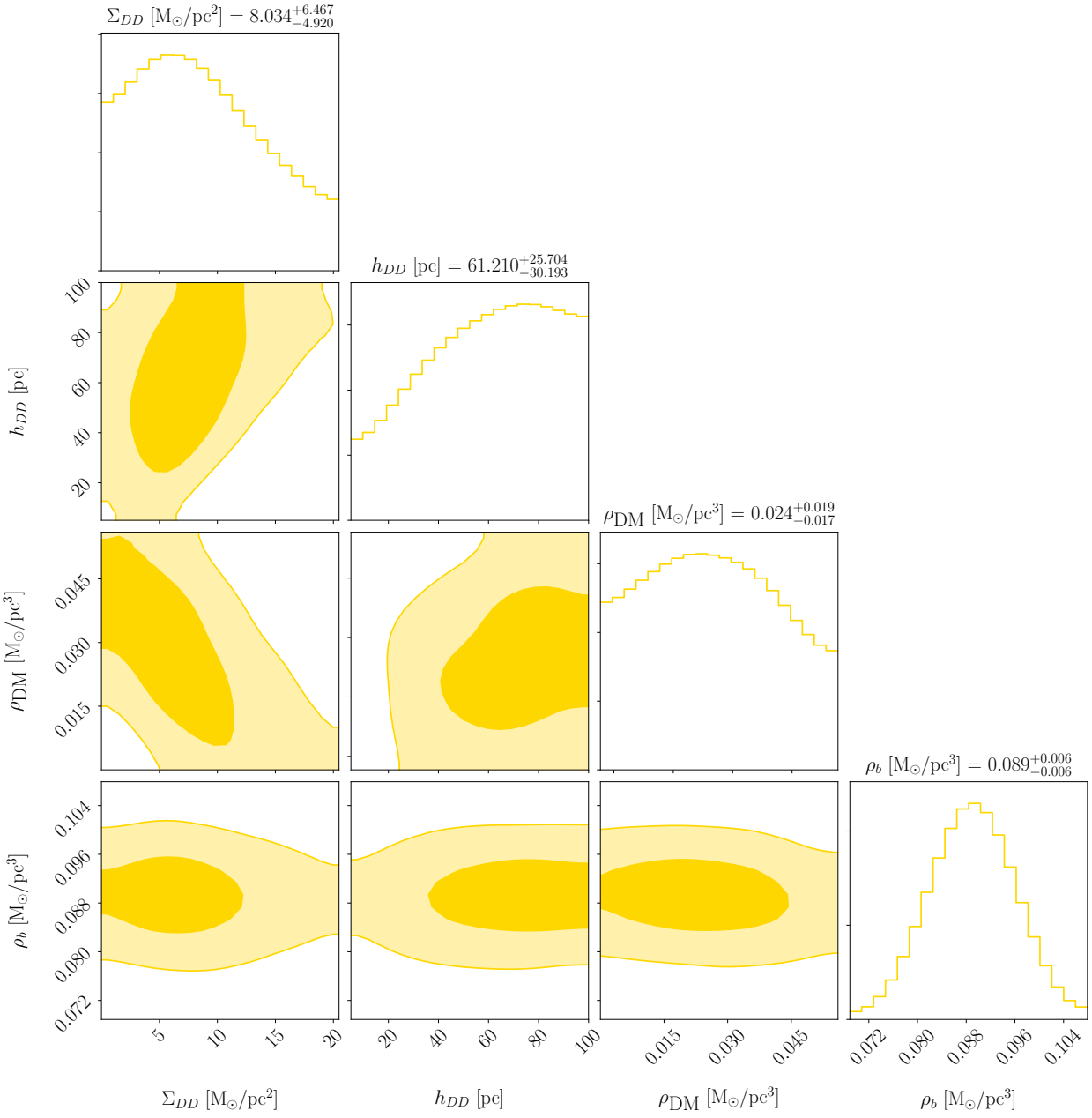

Figure 2.15: Posterior distributions in DD parameters, the local density of the DM halo $\rho_{\rm DM}$, and the total baryon density ρ_b for an analysis including only the A stars. The dark (light) regions correspond to 1 σ (2 σ) containment regions.

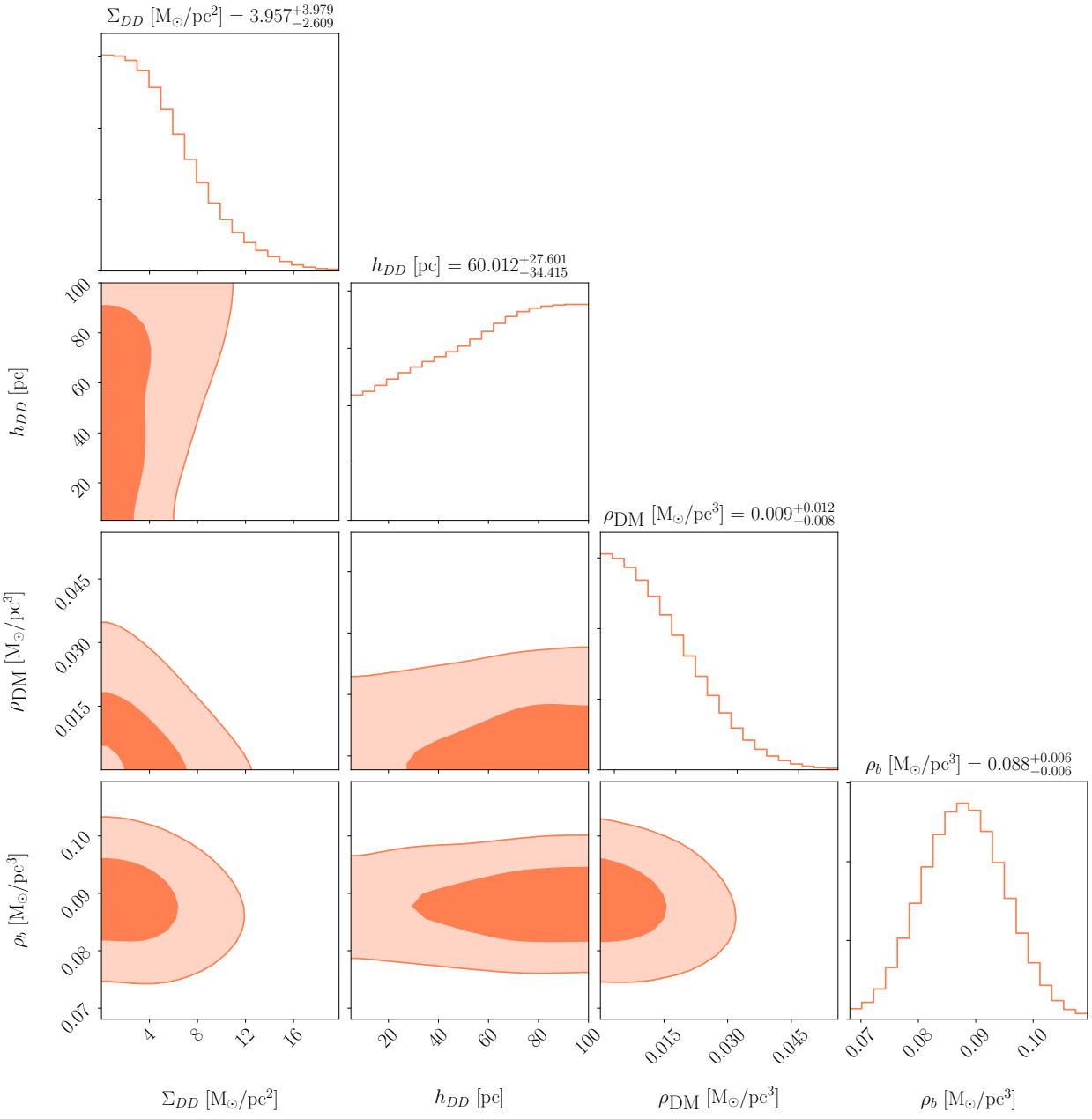

Figure 2.16: Same as Fig. 2.15 except for the F stars.

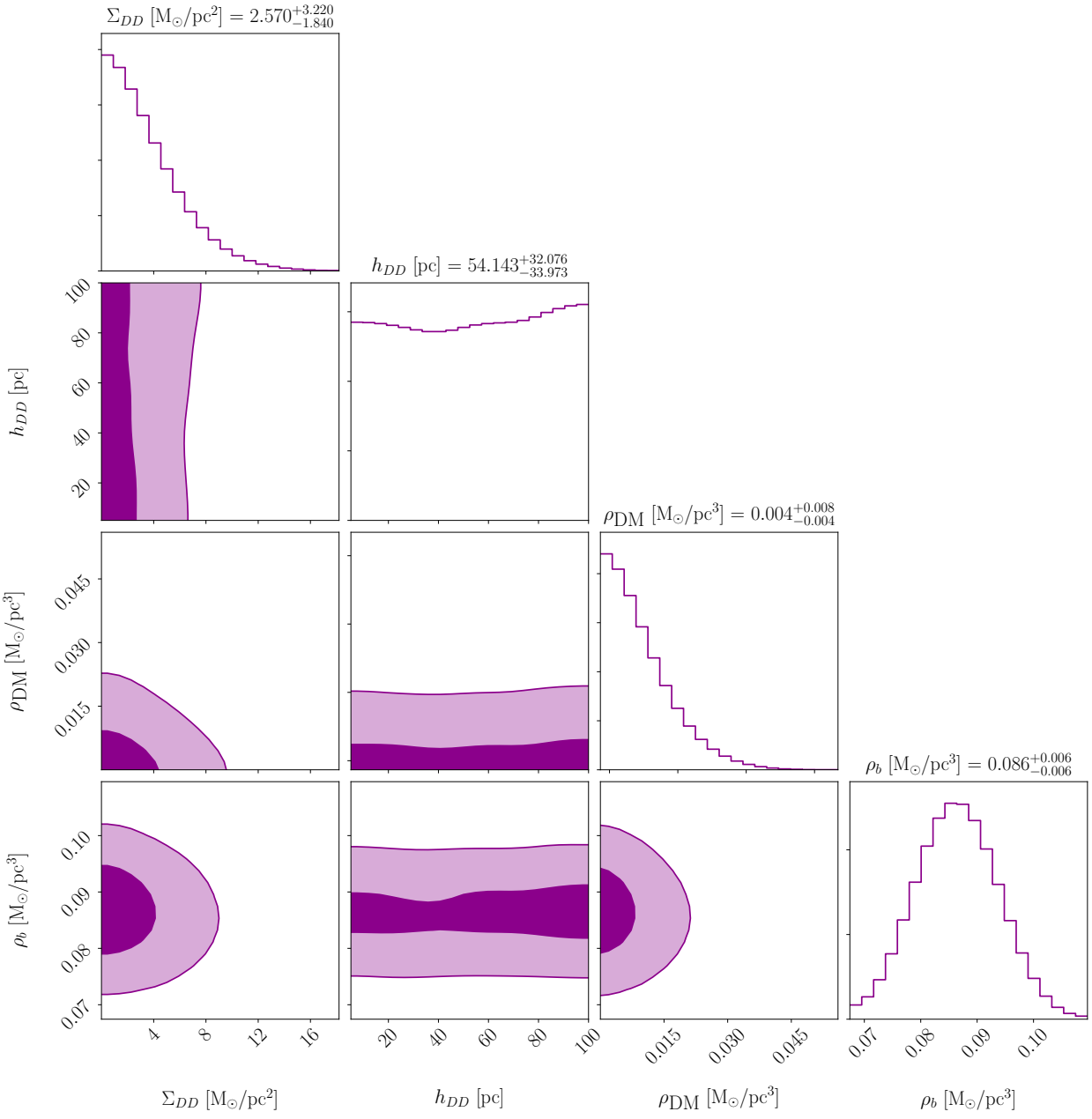

Figure 2.17: Same as Fig. 2.15 except for the early G stars.

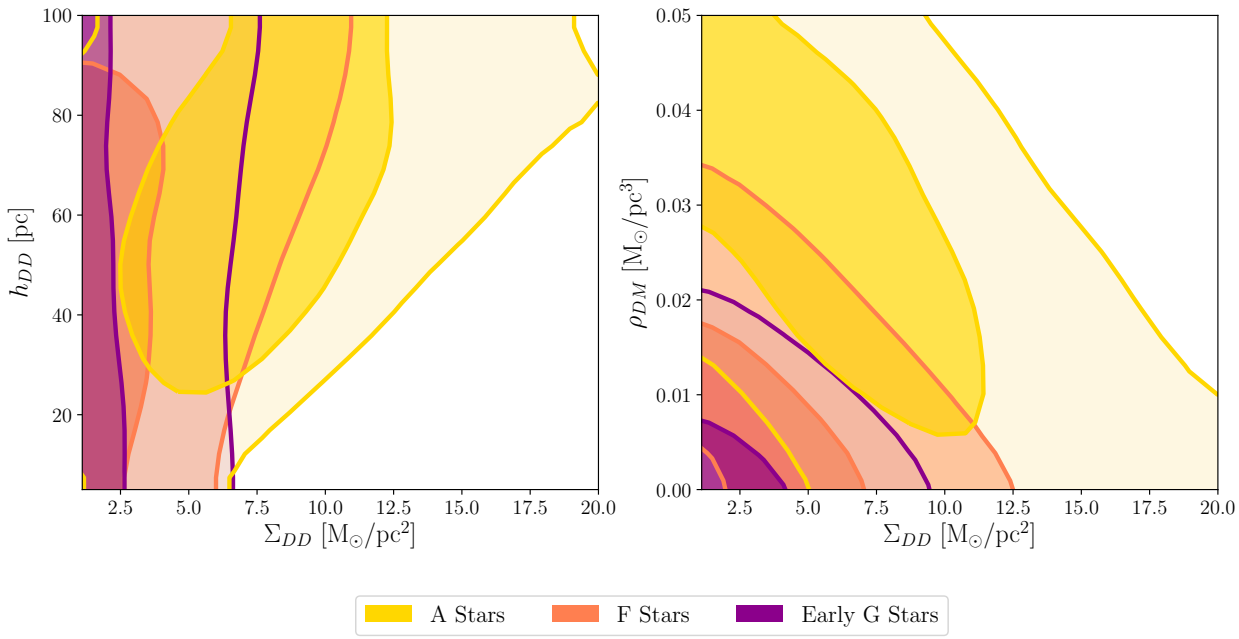

Figure 2.18: Here we overlay the posteriors for the DD parameters and halo DM $\rho_{\rm DM}$, shown in Figs. 2.15-2.17, for the three tracer populations. The shaded contours are the $1\,\sigma$ (darker) and $2\,\sigma$ (lighter) containment regions.

2.8 Stability Analysis of Allowed Dark Disks

Here we briefly comment on the stability of the thin DDs constrained by our analysis. As a basic check, we apply Toomre's criterion to the thin DD in two limiting cases: a collisionless gas, or a cold fluid with speed of sound c_s.

For the case of a self-gravitating collisionless gas, the Toomre stability parameter is

$$Q = \sigma_R \kappa / (3.36\, G\, \Sigma_{DD}), \tag{2.26}$$

where σ_R is the radial velocity dispersion and κ is the epicycle frequency. For a self-gravitating isothermal disk, the vertical velocity dispersion is related to the height and surface density as

$$\sigma_z = \sqrt{2\pi G\, \Sigma_{DD} h_{DD}}. \tag{2.27}$$

Assuming that $\sigma_R \simeq 2\sigma_z$ as for the stars in the MW, then the condition on stability can be written in terms of dark disk parameters as

$$Q \simeq \frac{2\sqrt{2\pi}\kappa}{3.36} \sqrt{\frac{h_{DD}}{G\, \Sigma_{DD}}} > 1. \tag{2.28}$$

Taking measured values of κ [203], we find that all the DDs we constrain are stable (except the thinnest DDs with $h_{DD} \lesssim 7$ pc) in this case.

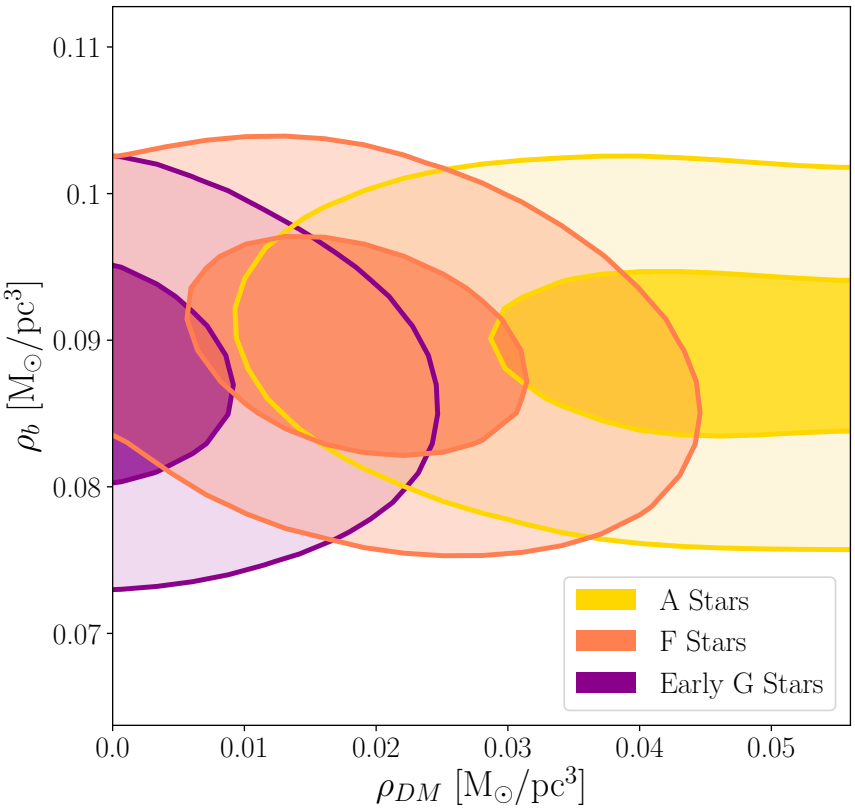

Figure 2.19: Assuming no DD is present, marginalized posteriors for the midplane baryon density ρ_b and halo DM $\rho_{\rm DM}$ for each of the tracer populations. The shaded contours are the $1\,\sigma$ (darker) and $2\,\sigma$ (lighter) containment regions. The posterior on ρ_b and $\rho_{\rm DM}$ is reported in Table 2.2.

In the latter case of a cold fluid, the Toomre stability parameter is

$$Q = c_s \kappa / (\pi G\, \Sigma_{DD}). \tag{2.29}$$

For a self-gravitating fluid in hydrostatic equilibrium, if we assume that the speed of sound is related to the scale height as $c_s = \sqrt{2\pi G \Sigma_{DD} h_{DD}}$ similar to the collisionless case, then we find that the DD is less stable; however, the DDs we constrain with $h_{DD} \gtrsim 13$ pc are within the regime of stability. For an isothermal fluid rotating about a central mass, the scale height is instead given by $h_{DD} = c_s/(\sqrt{2}\Omega)$ where Ω is the rotational angular velocity of circular orbits at the Sun's Galactocentric radius. The condition for stability can then be written as

$$Q = \sqrt{2} h_{DD} \Omega\, \kappa / (\pi G\, \Sigma_{DD}) \gtrsim 1. \tag{2.30}$$

For $h_{DD} = 10$ pc, this implies that only dark disks of $\Sigma_{DD} \lesssim 1 - 2 M_\odot/\text{pc}^2$ are stable, in contrast to the other scenarios. As demonstrated, the stability of the thin DD depends sensitively on the

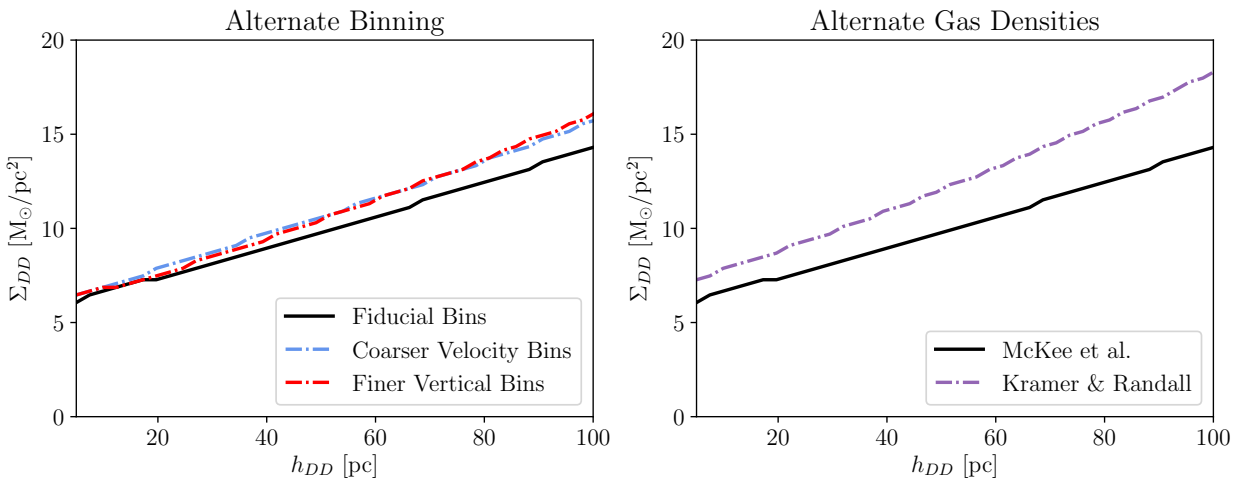

Figure 2.20: Left: Dependence of the limit on the choice of bin sizes. While for the fiducial analysis (black line), we use z-bins of width 25 pc and v-bins of width 2 km/s, there is a mild difference in the limits when using finer z-bins of width 8 pc (red dashed) or when using coarser v-bins of width 4 km/s (blue dashed). Right: Dependence of the limit on different models for the gas densities, given in Table 2.3. Our main analysis uses gas densities from Ref. [193] (McKee et al.), but we also consider the lower gas densities obtained in Ref. [204] (Kramer & Randall).

dynamics leading to DD formation and the properties of the dark sector. In addition, the presence of additional mass components could change these estimates substantially by adding a stabilizing potential [223].

2.9 Discussion

The main results of our analysis, shown in Fig. 2.11, strongly constrain the presence of a DD massive enough to account for phenomena such as periodic comet impacts. If we assume that the DD radial profile is identical to that of the baryonic disk (which need not be the case) then we can set a limit on the fraction of DM with strong dissipations. Taking the baryon scale radius $R_s = 2.15$ kpc [192], the Galactocentric radius of the Earth to be 8.3 kpc [224] and the MW halo mass to be 10^{12}M$_\odot$ [225], then dissipative disk DM can account for at most $\sim$1% of DM in the MW, for $h_{DD} \lesssim 20$ pc. Previous analyses which made this assumption found that up to $\sim$5% of the DM in the MW could be in the disk [152]. DDs that are marginally allowed by our analysis are not necessarily stable as per Toomre's criterion [182, 193, 226], although this depends on the collisional properties of the DM [227] and on the presence of other disk components [223, 228].

For the purpose of comparing our results with previous limits, the time-dependent analysis of Ref. [187] is the most similar to this work: although obtained in different ways, both analyses search over a band of possible density profiles that could arise from systematic effects. Our analysis

is more conservative, in that we search over multiple nuisance parameters, such as in the baryon mass model. However, we set a more stringent limit owing to the increased statistics of *Gaia* over *Hipparcos*.

Our analysis was limited by uncertainties that can be better understood with the second *Gaia* data release (DR2), which will have more proper motions, spectra for measuring line-of-sight velocities, and reduced measurement errors. The improved data will reduce the systematic uncertainty on $f_0(v)$ and the line-of-sight motions will (for the first time) allow for crucial checks on isothermality and any coupling between radial and vertical motions. Thus, the lack of evidence for out-of-equilibrium features in DR1 can be validated with DR2.

Chapter 3

Multi-state, Self-interacting DM in Halos

3.1 Introduction

As discussed in Chapters 1 and 2, the formation and dynamics of DM structures can be highly sensitive to the microphysics of DM. There are stringent constraints on DM interactions with SM particles, but the interactions of DM within its own dark sector (at most weakly coupled to the SM) are far less constrained. Increasingly detailed probes of the distribution of DM in halos may allow us to explore and constrain such interactions, independent of the coupling of the dark sector to known particles. It is therefore important to understand the potential observational signatures of dark sector physics. This is particularly true on small scales, where the CDM paradigm has yet to be verified.

Already, several apparent discrepancies have been identified between observations on sub-galactic scales and the predictions of collisionless cold DM (CCDM) simulations. Among the most relevant challenges are: the under-abundance of dwarf galaxies in the Milky Way (the missing satellites problem, [229, 230]) and in the field [231–233], the unexpected inner DM density profile in low surface brightness galaxies and dwarf galaxies (the cusp-core problem, e.g. [234, 235]), the deficit of DM in the inner regions of massive dwarf galaxies (the too-big-to-fail problem, e.g. [236, 237]), and the unexpected diversity in the shape of dwarf rotation curves [238, 239]. In general, observed dwarf galaxies appear to be less abundant, less massive and less centrally concentrated than predicted from CCDM simulations.

Most of these problems have been identified by contrasting CCDM-only simulations with observations, which is an oversimplified comparison that does not take into account the complex baryonic physics at play. It is therefore possible that some or even all of these challenges to CDM can be solved through the proper modeling of baryonic physics and by carefully considering observational biases. Plausible solutions have been presented separately for each problem. For instance, the existence of DM cores could be explained by the gravitational transfer of energy from supernovae into the orbits of DM particles (e.g. [240–244]). The resulting inner DM densities, in combination with strong environmental effects like tidal stripping and heating, have been invoked to alleviate the too-big-to-fail problem in the Milky Way (e.g. [245–248]). Furthermore, account-

ing for observational biases influences the interpretation of the severity of the dwarf abundance problem (in the Milky Way satellites e.g. [249, 250] and in the field e.g. [251]), as well as the too-big-to-fail problem in the field (e.g. [252, 253]), and more recently the diversity problem of dwarf rotation curves [254].

Nevertheless, a comprehensive solution to these CDM challenges remains elusive since there is presently no consensus on a definitive implementation of baryonic physics galaxy formation models. In particular, it is uncertain how to couple the sub-resolution physics of supernova explosions with the effective energy injection into the interstellar medium, and ultimately into the DM distribution. This coupling between supernovae and DM, which is seemingly crucial in solving the CDM challenges, depends on the stellar mass content of the galaxy relative to the depth of the potential well (e.g. [255, 256]), which sets the energy requirements for the cusp-core transformation: the smaller the galaxy, the less likely it is that this transformation is viable. Another key ingredient is how efficiently the available energy is deposited in the interstellar medium, and on which timescales it is injected relative to the dynamical time scales of the inner regions of the halo. Large modifications to the DM distribution are associated with "bursty" star formation histories, which have been shown to prevail in galaxies with stellar masses in the range $10^8 - 10^{10}\,\mathrm{M}_\odot$ [257]. However, at lower masses, where the CDM challenges are more severe, the time resolution needed to settle this issue remains inaccessible [258].

Given these outstanding CDM challenges, it is conceivable that these problems actually have a root in the assumptions about the underlying nature of DM. DM self-interaction (via some novel dark sector physics) can alleviate the discrepancies by transferring energy inward, thermalizing the inner DM distribution [259, 260]. This process has been shown to create $\mathcal{O}(\mathrm{kpc})$ size cores in the centre of haloes with allowed transfer cross sections per unit mass $\sigma_T/m_\chi \sim 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$. This is enough to alleviate the core-cusp problem and the too-big-to-fail problem [261–263], and it seemingly enhances the diversity of rotation curve shapes relative to CDM for a fixed baryonic physics implementation [264, 265]. Consistency with existing limits on DM self-interaction from merging clusters (which have much greater virial velocities than dwarf galaxies) is most easily achieved if the scattering cross-section has a velocity dependence (growing larger at low velocities); however, constant cross sections in the range $\sigma/m_\chi \sim 0.1 - 1$ cm^2/g remain viable [266]. The effects of DM interaction across broad classes of models can be studied generically within the framework of a generalized theory of structure formation (see the ETHOS framework, [267–269]). For a recent review of DM self-interaction see [270].

With few exceptions [271, 272], structure formation with DM self-interaction has been restricted to elastic scattering between degenerate DM states (i.e. either DM is a single Majorana fermion, or a Dirac fermion with particle-particle and particle-antiparticle scatterings). However, the inclusion of multiple DM states is another simple modification to DM physics with potentially striking observational signatures, especially in the presence of non-negligible self-interaction. Exothermic scatterings from an excited state would deliver velocity "kicks" to DM particles, which could be comparable to or larger than the escape velocity in bound DM structures, especially in the slow-moving environs of dwarf galaxies. In this way inelastic DM scattering can dilute dense cusps or dissipate dwarf halos entirely [273]. Velocity kicks from late-time *decays* of a metastable excited state have been considered in e.g. [274–279], but unlike the decay rate, the scattering rate

can be velocity-dependent and enhanced in regions of high density. Inelastic DM scattering from either DM or baryonic matter, whether exothermic or endothermic, can also have interesting signatures in direct and indirect searches for DM (e.g. [179, 280, 281] for direct detection, [282–284] for indirect detection).

The magnitude of the velocity kick in exothermic scattering depends on the exact energy splitting, δ, between the ground state (with mass m_χ) and the excited state, $v_{\rm kick} = \sqrt{2\delta/m_\chi}$. The velocity kick would have a substantial impact on the orbits of DM particles if the mass splitting is comparable to their orbital kinetic energy. Following [285], if we take the typical velocity in dwarf galaxies to be $\sim 10\,{\rm km\,s^{-1}} \sim 3\times10^{-5}\,c$, this then implies $\delta/m_\chi > 10^{-9}$ in order for down-scattering to have a relevant effect. Scenarios with such small mass splittings are easily constructed from a theoretical perspective. If DM is charged under some new dark gauge symmetry which is broken, then the states in the DM multiplet can naturally acquire a small mass splitting, regardless of whether the gauge group is Abelian or non-Abelian (e.g. [139, 286–288]). Additionally, in any case where the mediator of the DM self-interaction is light, the scattering cross section is automatically enhanced at low velocities, ensuring a larger impact on dwarf galaxies relative to Milky-Way-size galaxies and clusters. Even in situations where the scattering is purely elastic (i.e. there is no transition to a state of different mass), as we will show in this Chapter, the presence of a nearly-degenerate state in the spectrum can significantly modify any resonances present in the low-velocity scattering rate.

The goals of this Chapter are (1) to derive an analytic approximation for the DM-DM scattering cross section induced by an off-diagonal Yukawa interaction in the presence of a single excited state and (2) to study the impact of such a DM scenario on galactic halos using simulations. This DM scenario corresponds to the case where the dark gauge group is $U(1)$ and provides a simple and illustrative toy model for inelastic DM self-scattering more generally. As we will show, our expressions give good agreement with the numerical solution to the appropriate multi-state Schrödinger equation, but are very quick to compute and provide intuitive insight into the scattering behavior. The efficiency of evaluating these analytic cross sections makes it feasible to implement the behavior of this DM candidate in simulations. Our analytic results are applicable to the low-velocity regime where s-wave scattering dominates the total cross section; we defer study of the "classical" regime, where many partial waves contribute, to later work. We identify regions of parameter space with particular relevance to dwarf-galaxy-sized halos, and consequently to the discrepancies described above. We then show that in simulations the DM self-interactions do indeed modify the properties of halos. We also determine regions of parameter space which could potentially provide a DM explanation for the observed 3.5 keV spectral line from clusters [112, 113] via collisional excitation followed by decay, as in the "XrayDM" scenario [284].

We begin, in Section 3.2, by describing the general parameters and properties of our toy model. We derive our analytic scattering cross sections (following the method of Ref. [289]), and their behavior in several limiting regimes of interest, in Section 3.3. We simultaneously verify the validity of our approximate solutions by direct comparison to the numerical results. In Section 3.4 we provide first simple estimates of the parameter space where such scatterings could affect the internal structure and dynamics of dwarf galaxies, or yield the appropriate upscattering cross section to account for the 3.5 keV line in the XrayDM scenario. In Section 3.5 we show the

results of high-resolution simulations of a Milky Way-like halo within this inelastic DM model and quantitatively explore the impact on the DM distribution and substructure. Concluding remarks follow in Section 3.6.

3.2 A Simple Model of DM with Inelastic Scattering

We consider the case of a Yukawa-like interaction coupling two states with some small mass splitting, δ. We take the DM to be a pseudo-Dirac fermion charged under a dark $U(1)$. At high energies, where the $U(1)$ symmetry is unbroken, the DM is a charged Dirac fermion; at low energies, a small Majorana mass splits the Dirac fermion into two nearly-degenerate Majorana states χ^1 and χ^2. Since these states cannot carry conserved charge, their couplings with the $U(1)$ vector boson A_D are purely off-diagonal: that is, there is no vertex of the schematic form $\chi^1\chi^1 A_D$ or $\chi^2\chi^2 A_D$, only $\chi^1\chi^2 A_D$.

We will not specify the high-energy physics that gives rise to the Majorana masses, since we restrict ourselves to the extremely non-relativistic environment relevant for present-day DM scattering. The phenomenology of such models in the context of DM annihilation and inelastic scattering has been discussed in e.g. [139, 280, 282, 289, 290]. Inelastic scattering between states with small mass splittings can also naturally be generated in other contexts, for example composite DM [142] and sneutrino DM [291].

Since we work always in the non-relativistic limit, the interaction between the DM-like states can be characterized entirely in terms of a matrix potential that couples two-body states. The scattering cross sections can then be calculated using the methods of quantum mechanics (we refer the reader to e.g. [292] for a field-theoretic discussion using the Bethe-Salpeter formalism, and [293] for a derivation using effective field theory). The potential matrix coupling the $|11\rangle$ and $|22\rangle$ two-body states (corresponding to both particles being in the ground state or both particles being in the excited state) is [289]:

$$\mathcal{V}(r) = \begin{pmatrix} 0 & -\hbar c\alpha \frac{e^{-m_\phi c r/\hbar}}{r} \\ -\hbar c\alpha \frac{e^{-m_\phi c r/\hbar}}{r} & 2\delta c^2 \end{pmatrix} \tag{3.1}$$

where α is the coupling between the DM and the mediator (the dark fine structure constant), m_ϕ is the mass of the mediator, δ is the mass splitting between the ground and excited states, r is the separation between the particles, and the first row corresponds to the ground state $|11\rangle$. The two-body Schrödinger equation for the relative motion (factoring out the overall free motion of the system as usual) then takes the form:

$$\hbar^2 \frac{\nabla^2 \Psi(\mathbf{r})}{m_\chi} = \left(\mathcal{V}(r) - m_\chi v^2\right) \Psi(\mathbf{r}). \tag{3.2}$$

Here m_χ is the mass of the DM, v is the individual velocity of either of the DM particles in the center-of-mass frame (half the relative velocity), and $\Psi(\mathbf{r})$ is the wavefunction.

As mentioned above, the interaction between the ground state, $|1\rangle$ and the excited state, $|2\rangle$, is purely off-diagonal. This is an automatic consequence of taking the force carrier to be a vector,

as the mass eigenstates are 45° rotations of the high-energy gauge eigenstates, and do not carry a conserved charge. As a result, two particles initially in the same state (ground or excited) can only scatter into the two-body states where they are both in the ground state, or both in the excited state. Such scatterings are described by the off-diagonal terms in the $\mathcal{V}(r)$ matrix. If the initial state is $|12\rangle$ (i.e. one particle is in the ground state and the other in the excited state) then their scattering decouples from the other two-body states and is elastic, with the final state being $|12\rangle$ or $|21\rangle$. This case can be treated by the existing methods in the literature (e.g. [149]).

In this article we will treat only s-wave scattering (the methods developed in [289] do not generalize to higher partial waves). Accordingly we will make the ansatz of spherical symmetry in solving the Schrödinger equation (more formally, as discussed in [289], we expand in partial waves and only keep the $\ell = 0$ term), and write $r\Psi(\mathbf{r}) = \psi(r)$.

We define the dimensionless parameters:

$$\epsilon_v \equiv \frac{v}{c\alpha}, \quad \epsilon_\delta \equiv \sqrt{\frac{2\delta}{m_\chi \alpha^2}}, \quad \epsilon_\phi \equiv \frac{m_\phi}{m_\chi \alpha} \tag{3.3}$$

so that rescaling r by $\alpha m_\chi c/\hbar$ gives the s-wave Schrödinger equation:

$$\psi''(r) = \begin{pmatrix} -\epsilon_v^2 & -\frac{e^{-\epsilon_\phi r}}{r} \\ -\frac{e^{-\epsilon_\phi r}}{r} & \epsilon_\delta^2 - \epsilon_v^2 \end{pmatrix} \psi(r) \equiv \bar{V}(r)\psi(r) \equiv \begin{pmatrix} -\epsilon_v^2 & -V(r) \\ -V(r) & \epsilon_\delta^2 - \epsilon_v^2 \end{pmatrix} \psi(r). \tag{3.4}$$

In the last line we have defined the matrix potential $\bar{V}(r)$ and the scalar Yukawa potential $V(r) = e^{-\epsilon_\phi r}/r$. Solving this equation with scattering boundary conditions will be the focus of Section 3.3.

The model presented above is by no means the only model where DM could have multiple interacting states. Small mass splittings between states, similar to those we consider, could also be present in the context of atomic DM (e.g. [294–298]), non-Abelian dark sectors where the DM is part of a nearly-degenerate multiplet of states (e.g. [299, 300]), or in scenarios where the DM forms stable bound states (e.g. [301]). However, it should be noted that the scenario we consider requires that transitions from the heavier to the lighter state occur primarily through collisional de-excitation; for this type of transition it is thus necessary but not sufficient to have small mass splittings. In models that include a very light force carrier, such as classic dark atom models, the excited states are typically depleted by radiative decay rather than collisions, leading to different effects on structure formation (e.g. [302]). Our calculation has other qualitative features that are not universal to dark-sector models with small mass splittings: for example, that there are only two states that significantly participate in the phenomenology, and that de-excitation requires two excited-state particles to simultaneously de-excite. Thus while this model serves as a valuable illustrative example of the effects of inelastic self-scattering, detailed quantitative results for other dark-sector models would generally require dedicated analyses.

We note that in the generic framework of the model we are considering here, the presence of a potential mediated by light vector exchange automatically implies that both the ground and excited states can annihilate into the light vectors. For some of the parameter space considered in this work, the annihilation cross section can be much larger than the thermal relic cross section,

implying a depletion of the DM states in the early Universe. Additionally, to justify a large primordial abundance of the excited state (which would give the largest impact on structure formation), a non-thermal origin for these species could be invoked, which could occur through the decay of thermally produced heavier species, preferentially into the excited state (as discussed in [303]). Plausible models with late-time decays would need to avoid a series of constraints in the early Universe, for instance in their gravitational effects on the cosmic microwave background radiation (e.g. [304]) and large scale structure. Such constraints may point toward models where the metastable decaying species is only slightly heavier than the DM. Since our focus in this paper is on the implications for structure formation of a large metastable population of excited states in the early Universe, we leave a detailed study of possible models for later work. Models with large late-time populations of a metastable excited state have been previously discussed in the context of indirect-detection and direct-detection signals (e.g. [305, 306]).

3.3 Scattering Cross Sections

Approximate Wavefunctions

The eigenvalues $\lambda_\pm$ and eigenvectors $\psi_\pm$ of the matrix $\bar{V}(r)$ are

$$\lambda_\pm = -\epsilon_v^2 + \frac{\epsilon_\delta^2}{2} \pm \sqrt{\frac{\epsilon_\delta^2}{2} + \frac{e^{-2\epsilon_\phi r}}{r^2}}, \quad \psi_\pm = \frac{1}{\sqrt{2}} \begin{pmatrix} \mp\sqrt{1 \mp \frac{1}{\sqrt{1+(4e^{-2\epsilon_\phi r})/(r^2\epsilon_\delta^4)}}} \\ \sqrt{1 \pm \frac{1}{\sqrt{1+(4e^{-2\epsilon_\phi r})/(r^2\epsilon_\delta^4)}}} \end{pmatrix}. \tag{3.5}$$

There is a transition in the behavior of the eigenvalues and eigenvectors when $\frac{e^{-\epsilon_\phi r}}{r} \sim \frac{\epsilon_\delta^2}{2}$. In the regime where $\frac{e^{-\epsilon_\phi r}}{r} \gg \frac{\epsilon_\delta^2}{2}$, the eigenvalues and eigenvectors can be approximated as

$$\lambda_\pm \approx -\epsilon_v^2 + \frac{\epsilon_\delta^2}{2} \pm \frac{e^{-\epsilon_\phi r}}{r}, \qquad \psi_\pm \approx \frac{1}{\sqrt{2}} \begin{pmatrix} \mp 1 \\ 1 \end{pmatrix}. \tag{3.6}$$

When $\frac{e^{-\epsilon_\phi r}}{r} \ll \frac{\epsilon_\delta^2}{2}$, the eigenvalues and eigenvectors can be approximated as

$$\lambda_+ \approx -\epsilon_v^2 + \epsilon_\delta^2 + \frac{e^{-2\epsilon_\phi r}}{r^2\epsilon_\delta^2}, \quad \lambda_- \approx -\epsilon_v^2 - \frac{e^{-2\epsilon_\phi r}}{r^2\epsilon_\delta^2}, \quad \psi_+ \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \psi_- \approx \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{3.7}$$

In other words, at small radii the symmetry is restored and the energy eigenstates become the gauge eigenstates, as at high energies; at large radii the energy eigenstates are the mass eigenstates.

At small and large r, the diagonalization of the potential matrix is roughly independent of r. If $\phi_\pm''(r) = \lambda_\pm \phi_\pm(r)$, then $\phi_\pm(r)\psi_\pm$ is an approximate solution to the matrix Schrödinger equation in those regimes, and we can use standard techniques — in particular, the WKB approximation — to solve for the scalar wavefunctions $\phi_\pm(r)$. However, in the transition region where $\frac{e^{-\epsilon_\phi r}}{r} \sim \frac{\epsilon_\delta^2}{2}$, the eigenvectors will vary as a function of r.

Provided that $\epsilon_\delta^2/2 \lesssim \epsilon_\phi$, this transition occurs at $r \gtrsim 1/\epsilon_\phi$, where the exponential behavior of the potential dominates: in this case we can use an *exact* solution for the two-state exponential

Regimes for the wavefunction	
$V(r) \approx 1/r$ for small r (exact solution)	$V(r) \gg \epsilon_\delta^2/2, V(r) \gtrsim \epsilon_v^2,\ \ell = 0$
WKB approximation for intermediate r	$V(r) \gtrsim \epsilon_\phi^2, \epsilon_\delta^2/2$
$V(r) \approx V_0 e^{-\mu r}$ for large r (exact solution)	$r \gtrsim 1/\epsilon_\phi$
Match onto large-r asymptotic eigenstates	$V(r) \ll \epsilon_\delta^2/2$
Conditions on $\boldsymbol{\epsilon_v}$, $\boldsymbol{\epsilon_\phi}$, and $\boldsymbol{\epsilon_\delta}$	
2nd and 3rd regimes above overlap	$\epsilon_\delta^2/2 \lesssim \epsilon_\phi$
s-wave dominates, simple matching	$\epsilon_v \lesssim \epsilon_\phi$
Scattering cross section enhanced by dark force	$\epsilon_v, \epsilon_\phi, \epsilon_\delta \lesssim 1$

Table 3.1: A summary of the various approximations and assumptions used for deriving the wavefunction in different regimes. The top four rows characterize the different regimes in r where various techniques can be applied to approximate the wavefunction in the vacinity of a Yukawa potential; the lower three rows describe constraints on the parameters which allow the approximation to hold and cause the scattering cross section to be enhanced. Despite these restrictions, the various regimes of validity overlap substantially, making these approximations useful in large swatches of parameter space.

potential (taken from [307]) to cross the transition region and match the WKB solution at small r to the large-r asymptotic solution. Where $\epsilon_\delta^2/2 > \epsilon_\phi$, we can still use this approach, but the exponential potential is a poor approximation to the true $V(r)$ at the radius corresponding to the transition region, and so the results are less reliable.

The full procedure for determining the approximate wavefunction $\psi(r)$ can be summarized as:

- For $r \lesssim 1$, we approximate the Yukawa potential as $V(r) \sim 1/r$, assume the potential term dominates (since ϵ_v^2 and ϵ_δ^2 are assumed to be small), and solve the resulting Schrödinger equation exactly.

- For $r \gtrsim 1$, we propagate the small-r solution outward using a WKB approximation. The validity of the WKB approximation in the potential-dominated regime requires that

$$\left|\sqrt{V'(r)}/V(r)\right| = \frac{1}{2}\sqrt{r}\, e^{\epsilon_\phi r/2}(\epsilon_\phi + 1/r) \ll 1, \tag{3.8}$$

  which is equivalent to requiring that the spatial variation of the local de Broglie wavelength is sufficiently gradual. For $r > 1$, the condition for validity of the approximation is that $\epsilon_\phi/\sqrt{V(r)} \lesssim 1$, so the approximation breaks down when $V(r) \sim \epsilon_\phi^2$.

- Where the WKB approximation breaks down (at $V(r) \lesssim \epsilon_\phi^2$) or where the diagonalization approximation fails (at $V(r) \lesssim \epsilon_\delta^2/2$), we must match the WKB solution onto a large-r solution. Therefore, in this work, we will choose the matching radius r_M such that $\frac{e^{-\epsilon_\phi r_M}}{r_M} = \max\left(\frac{\epsilon_\delta^2}{2}, \epsilon_\phi^2\right)$. For large r ($r \gtrsim 1/\epsilon_\phi$, so the exponential behavior dominates) we can approximate the Yukawa potential as an exponential potential $V(r) \sim V_0 e^{-\mu r}$, which leads to a Schrödinger equation that can be solved exactly even in the two-state case [307].

We summarize the different regimes in Figure 3.1. The large-r solution involves hypergeometric functions which can only be analytically matched at r_M by using an asymptotic expansion. As alluded to in Table 3.1 (noting that $\mu \sim \epsilon_\phi$), this matching is simplest if $\epsilon_v \lesssim \mu$; if $\epsilon_v \gtrsim \mu$, there is a term in the expansion that appears exponentially suppressed but can acquire an exponentially large prefactor. The mathematical condition that $\epsilon_v \lesssim \mu$ is physically equivalent to ignoring the contribution from the small-r eigenstates of the $1/r$ potential that experience a repulsive interaction. Additionally, for $\epsilon_v \gtrsim \mu$, we expect the s-wave term we compute here to be subdominant [149][1]. The method developed here does not generalize straightforwardly to higher partial waves [289] — it is useful for the "resonant" regime described by [149], not the "classical" regime, which even in the elastic-scattering case demands a different approach. Consequently we focus on the part of parameter space where $\epsilon_v \lesssim \mu \sim \epsilon_\phi$.

We also require that ϵ_v, ϵ_δ, and ϵ_ϕ are all less than 1 in order to see substantial enhancement to the s-wave cross section, beyond the geometric cross section associated with the mass of the DM particle (not the force carrier). If $\epsilon_v \gtrsim 1$, the kinetic energy is large compared to the potential energy; if $\epsilon_\phi \gtrsim 1$, the range of the interaction is short; in both cases the presence of the potential does not significantly deform the wave-function. If $\epsilon_\delta \gtrsim 1$, the mass splitting is large compared to the Bohr potential energy ($\sim \alpha^2 m_\chi$, leading to a suppression of virtual excitations.) Since the potential is purely off-diagonal, this suppresses the elastic scattering cross section as well.

For completeness, we present a full description of the approximate wavefunctions below, including the WKB matching between the small-r and large-r wavefunctions. A more in-depth

[1] Numerical studies of inelastic DM scattering in this regime confirm that the higher partial waves provide large contributions [308].

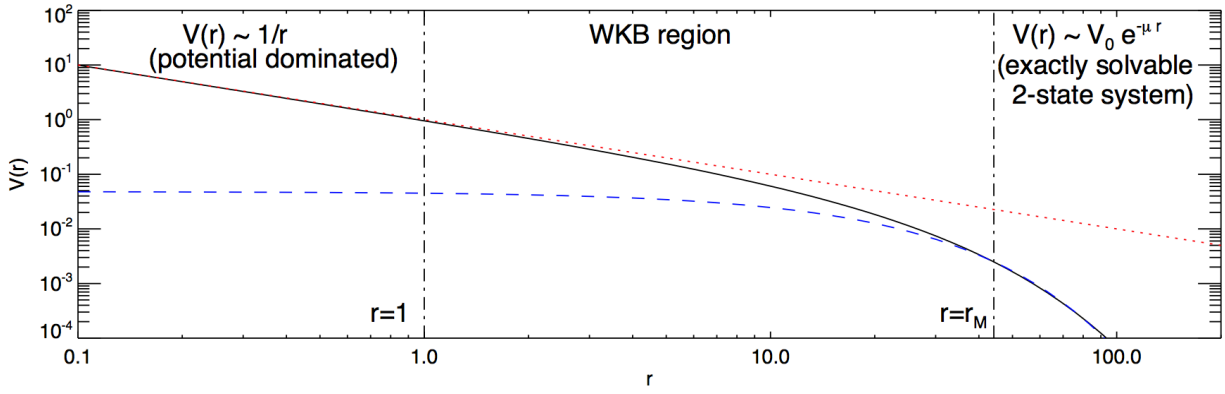

Figure 3.1: Example of the different r-regimes and matching points for a sample parameter set ($\epsilon_v = 0.1$, $\epsilon_\delta = 0.02$, $\epsilon_\phi = 0.05$), following [289]. The plot shows the exact Yukawa potential (*solid black line*) and the approximate potentials we employ, in their regimes of validity. In the $r \lesssim r_M$ region where the eigenstates are decoupled, for $r \lesssim 1$ the potential dominates the kinetic energy and mass splitting and is well approximated by $V(r) \approx 1/r$ (*red dotted line*), whereas for $1 \lesssim r \lesssim r_M$ the WKB approximation is employed to obtain an approximate wavefunction. At $r \gtrsim r_M$ the WKB approximation may break down, but there $V(r) \approx V_0 e^{-\mu r}$ (*dashed blue line*).

derivation of these wavefunctions (including more extensive discussion of the regimes of validity) can be found in [289], albeit with different boundary conditions.

For small r, we can approximate the Yukawa potential as $1/r$, which therefore dominates the eigenvalues at small r so $\lambda_\pm \approx \pm 1/r$. The s-wave solutions to the rescaled Schrödinger equation can be expressed in terms of Bessel functions:

$$\begin{aligned} \phi_-(r) &= A_- \sqrt{r} J_1(2\sqrt{r}) - \pi \phi_-(0) \sqrt{r} Y_1(2\sqrt{r}) \\ \phi_+(r) &= A_+ \sqrt{r} I_1(2\sqrt{r}) + 2\phi_+(0) \sqrt{r} K_1(2\sqrt{r}), \end{aligned} \tag{3.9}$$

where A_+ and A_- are the coefficients of the repulsed and attracted eigenstates, respectively. Moving radially outward (but still within the regime of validity for the $V \sim 1/r$ approximation), the large-r asymptotics of the Bessel functions give

$$\phi_+(r) = \frac{1}{\lambda_+^{1/4}} \left(\frac{A_+}{2\sqrt{\pi}} e^{\int_0^r \sqrt{\lambda_+(r')}\, dr'} + \left(\sqrt{\pi}\phi_+(0) - \frac{iA_+}{2\sqrt{\pi}} \right) e^{-\int_0^r \sqrt{\lambda_+(r')}\, dr'} \right) \tag{3.10}$$

$$\phi_-(r) = \frac{1}{\lambda_-^{1/4}} \left(-\frac{1}{2\sqrt{\pi}} \left(A_- + i\pi\phi_-(0) \right) e^{\int_0^r \sqrt{\lambda_-(r')}\, dr'} + \frac{i}{2\sqrt{\pi}} \left(A_- - i\pi\phi_-(0) \right) e^{-\int_0^r \sqrt{\lambda_-(r')}\, dr'} \right).$$

If the particle velocity is high enough (above threshold) such that there exists a radius r^* in this regime where $V(r^*) = \epsilon_v \sqrt{\epsilon_v^2 - \epsilon_\delta^2}$ then $\lambda_+(r^*) = 0$ and we must perform a WKB approximation about the turning point. Linearizing the potential and matching the wavefunctions on either side

using the connection formulae yields:

$$\phi_+ = \frac{1}{|\lambda_+|^{1/4}}\Bigg[\left(\frac{-iA_+}{2\sqrt{\pi}} + \frac{1}{2}\left(\sqrt{\pi}\phi_+(0) - \frac{iA_+}{2\sqrt{\pi}}\right)e^{-2\int_0^{r^*}\sqrt{\lambda_+(r')}\,dr'}\right)e^{\frac{i\pi}{4}+\int_0^r\sqrt{\lambda_+(r')}\,dr'} \\ + \left(\frac{iA_+}{2\sqrt{\pi}}e^{2\int_0^{r^*}\sqrt{\lambda_+(r')}\,dr'} + \frac{1}{2}\left(\sqrt{\pi}\phi_+(0) - \frac{iA_+}{2\sqrt{\pi}}\right)\right)e^{-\frac{i\pi}{4}-\int_0^r\sqrt{\lambda_+(r')}\,dr'}\Bigg]. \tag{3.11}$$

In either case (above or below threshold), the small-r wavefunctions are in a form that will match smoothly onto the WKB wavefunctions.

At large r, we can approximate the Yukawa potential as a purely exponential potential of the form $V_0\, e^{\mu r}$. We impose conditions on V_0 and μ by requiring that the exponential potential mimic the Yukawa for $r > r_M$, where r_M is the matching radius which we have chosen such that $V(r_M) = \max\left(\frac{\epsilon_\delta^2}{2}, \epsilon_\phi^2\right)$. The potentials should match at r_M so $\frac{e^{-\epsilon_\phi r_M}}{r_M} = V_0 e^{-\mu\, r_M}$. We also require that $\int_{r_M}^\infty e^{-\epsilon_\phi r}\,dr = \int_{r_M}^\infty r\, V_0\, e^{-\mu r}\,dr$, which comes from solving the Lippman-Schwinger form of the Schrödinger equation and requiring that the rescaled wavefunctions from both potentials match to first order in the coupling constant α (following [292]). The parameters μ and V_0 are therefore given by

$$\mu = \epsilon_\phi\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{4}{\epsilon_\phi r_M}}\right) \qquad V_0 = \frac{e^{\epsilon_\phi r_M\left(-\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{4}{\epsilon_\phi r_M}}\right)}}{r_M}. \tag{3.12}$$

The wavefunctions for an exponential potential can be solved for exactly in terms of ${}_0F_3$ hypergeometric functions [307] as follows:

$$\psi_1 = C_1\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{\frac{i\epsilon_v}{\mu}}\,{}_0F_3\left[\left\{1 + \frac{i\epsilon_v}{\mu}, \frac{1}{2} + \frac{i(\epsilon_v - \epsilon_\Delta)}{2\mu}, \frac{1}{2} + \frac{i(\epsilon_v + \epsilon_\Delta)}{2\mu}\right\}, \left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right] \\ + C_2\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{-\frac{i\epsilon_v}{\mu}}\,{}_0F_3\left[\left\{1 - \frac{i\epsilon_v}{\mu}, \frac{1}{2} - \frac{i(\epsilon_v + \epsilon_\Delta)}{2\mu}, \frac{1}{2} - \frac{i(\epsilon_v - \epsilon_\Delta)}{2\mu}\right\}, \left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right] \\ + C_3\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{\frac{i\epsilon_\Delta}{\mu}+1}\,{}_0F_3\left[\left\{\frac{3}{2} - \frac{i(\epsilon_v - \epsilon_\Delta)}{2\mu}, \frac{3}{2} + \frac{i(\epsilon_v + \epsilon_\Delta)}{2\mu}, 1 + \frac{i\epsilon_\Delta}{\mu}\right\}, \left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right] \\ + C_4\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{-\frac{i\epsilon_\Delta}{\mu}+1}\,{}_0F_3\left[\left\{\frac{3}{2} - \frac{i(\epsilon_v + \epsilon_\Delta)}{2\mu}, \frac{3}{2} + \frac{i(\epsilon_v - \epsilon_\Delta)}{2\mu}, 1 - \frac{i\epsilon_\Delta}{\mu}\right\}, \left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right] \tag{3.13}$$

$$\psi_2 = -\frac{C_1\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{\frac{i\epsilon_v}{\mu}+1}}{\left(\frac{1}{2}+\frac{i\epsilon_v}{2\mu}\right)^2+\frac{\epsilon_\Delta^2}{4\mu^2}}\,{}_0F_3\left[\left\{\frac{3}{2}+\frac{i(\epsilon_v-\epsilon_\Delta)}{2\mu},\,\frac{3}{2}+\frac{i(\epsilon_v+\epsilon_\Delta)}{2\mu},\,1+\frac{i\epsilon_v}{\mu}\right\},\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right]$$
$$-\frac{C_2\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{-\frac{i\epsilon_v}{\mu}+1}}{\left(\frac{1}{2}-\frac{i\epsilon_v}{2\mu}\right)^2+\frac{\epsilon_\Delta^2}{4\mu^2}}\,{}_0F_3\left[\left\{\frac{3}{2}-\frac{i(\epsilon_v+\epsilon_\Delta)}{2\mu},\,\frac{3}{2}-\frac{i(\epsilon_v-\epsilon_\Delta)}{2\mu},\,1-\frac{i\epsilon_v}{\mu}\right\},\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right]$$
$$-C_3\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{\frac{i\epsilon_\Delta}{\mu}}\left(\frac{\epsilon_v^2}{4\mu^2}+\left(\frac{1}{2}+\frac{i\epsilon_\Delta}{2\mu}\right)^2\right)$$
$$\times\ {}_0F_3\left[\left\{1+\frac{i\epsilon_\Delta}{\mu},\,\frac{1}{2}-\frac{i(\epsilon_v-\epsilon_\Delta)}{2\mu},\,\frac{1}{2}+\frac{i(\epsilon_v+\epsilon_\Delta)}{2\mu}\right\},\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right]$$
$$-C_4\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^{-\frac{i\epsilon_\Delta}{\mu}}\left(\frac{\epsilon_v^2}{4\mu^2}+\left(\frac{1}{2}-\frac{i\epsilon_\Delta}{2\mu}\right)^2\right)$$
$$\times\ {}_0F_3\left[\left\{1-\frac{i\epsilon_\Delta}{\mu},\,\frac{1}{2}-\frac{i(\epsilon_v+\epsilon_\Delta)}{2\mu},\,\frac{1}{2}+\frac{i(\epsilon_v-\epsilon_\Delta)}{2\mu}\right\},\left(\frac{V_0 e^{-\mu r}}{4\mu^2}\right)^2\right], \tag{3.14}$$

where we have defined $\epsilon_\Delta \equiv \sqrt{\epsilon_v^2-\epsilon_\delta^2}$. The wavefunctions are expressed in terms of four linearly-independent solutions, corresponding to ingoing or outgoing particles in the ground or excited states. In particular,

- the C_1 term represents an ingoing wave in the ground state,
- the C_2 term represents an outgoing wave in the ground state,
- the C_3 term represents an ingoing wave in the excited state,
- the C_4 term represents an outgoing wave in the excited state.

To match the large-r wavefunctions to the small-r wavefunctions, one can use the WKB approximation to propagate the known wavefunctions of the exponential potential into the transition region. We write the large-r WKB solutions as

$$\phi_\pm = \frac{1}{\left|\tilde{\lambda}_\pm\right|^{1/4}}\left(E_\pm e^{\int_0^r\sqrt{\tilde{\lambda}_\pm(r')}dr'}+F_\pm e^{-\int_0^r\sqrt{\tilde{\lambda}_\pm(r')}dr'}\right). \tag{3.15}$$

where $\tilde{\lambda}_\pm$ are the eigenvalues of the matrix Schrödinger equation with an exponential potential rather than a Yukawa. In order to match the WKB solution with the exact solution for the large-r

exponential potential, we define the following convenient quantities:

$$\begin{aligned}
\eta &\equiv \frac{1}{4}\left[\frac{\epsilon_v^2}{\mu^2} + \left(1 + \frac{i\epsilon_\Delta}{\mu}\right)^2\right] \\
\Gamma_v &\equiv \Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right)\Gamma\left(\frac{i\epsilon_v - i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right)\Gamma\left(\frac{i\epsilon_v + i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right) \\
\Gamma_\Delta &\equiv \Gamma\left(1 + \frac{i\epsilon_\Delta}{\mu}\right)\Gamma\left(\frac{i\epsilon_\Delta - i\epsilon_v}{2\mu} + \frac{1}{2}\right)\Gamma\left(\frac{i\epsilon_v + i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right).
\end{aligned} \tag{3.16}$$

Then, deriving expressions for the WKB coefficients $E_\pm$ and $F_\pm$ to match onto the exponential wavefunctions is a matter of using the asymptotic behavior of the ${}_0F_3$ hypergeometric functions in the $r \to -\infty$ limit (see [289] for details), and using the WKB approximation again to propagate these asymptotic solutions into the matching region. We find that

$$\begin{aligned}
E_+ &= 0 \\
F_+ &= -\frac{\sqrt{\mu}}{(2\pi)^{3/2}} e^{\int_0^{r_s}\sqrt{\tilde{\lambda}_+(r')}\,dr'} e^{\frac{2i\sqrt{V_0\,e^{-\mu r_s}}}{\mu}} \left(C_1\Gamma_v + C_2\Gamma_v^* + C_3\,\eta\,\Gamma_\Delta + C_4\,\eta^*\,\Gamma_\Delta^*\right) \\
E_- &= \frac{\sqrt{\mu}}{(2\pi)^{3/2}} e^{\int_{r_s}^{0}\sqrt{\tilde{\lambda}_-(r')}\,dr' + \frac{i\pi}{4}} e^{-\frac{2i\sqrt{V_0\,e^{-\mu r_s}}}{\mu}} \\
&\times \left(C_1\Gamma_v\, e^{-\frac{\pi\epsilon_v}{\mu}} + C_2\Gamma_v^*\, e^{\frac{\pi\epsilon_v}{\mu}} - C_3\,\eta\,\Gamma_\Delta e^{-\frac{\pi\epsilon_\Delta}{\mu}} - C_4\,\eta^*\,\Gamma_\Delta^* e^{\frac{\pi\epsilon_\Delta}{\mu}}\right) \\
F_- &= \frac{\sqrt{\mu}}{(2\pi)^{3/2}} e^{\int_0^{r_s}\sqrt{\tilde{\lambda}_-(r')}\,dr' - \frac{i\pi}{4}} e^{\frac{2i\sqrt{V_0\,e^{-\mu r_s}}}{\mu}} \\
&\times \left(C_1\Gamma_v e^{\frac{\pi\epsilon_v}{\mu}} + C_2\Gamma_v^* e^{-\frac{\pi\epsilon_v}{\mu}} - C_3\,\eta\,\Gamma_\Delta e^{\frac{\pi\epsilon_\Delta}{\mu}} - C_4\,\eta^*\,\Gamma_\Delta^* e^{-\frac{\pi\epsilon_\Delta}{\mu}}\right)
\end{aligned} \tag{3.17}$$

where r_s is some (possibly negative) radius chosen such that (a) $V_0 e^{-\mu r} \gg \mu^2$ (required for the asymptotic expansion to be valid) and (b) the potential dominates the kinetic and mass-splitting terms, i.e. $V_0\, e^{-\mu r_s} \gg \epsilon_v^2,\, \epsilon_\delta^2$. There is no reason not to choose r_s arbitrarily large and negative, however, and as discussed in the main text, we generally take $r_s \to -\infty$. (Of course, this does not correspond to a physical region of real space, but this matching is simply a mathematical trick to translate the hypergeometric functions into a form that facilitates matching to the previously derived WKB solutions.)

Note that this procedure sets $E_+ = 0$. This originates from the neglect of a particular (exponentially suppressed) term in the asymptotic expansion for the hypergeometric functions. However, this term can gain an exponentially large prefactor when $\epsilon_v \gtrsim \mu$, as discussed in [289].

In order to match the WKB solution with the small-r solution below threshold, we equate

(3.15) with (3.10), which gives

$$\begin{aligned} E_+ &= \frac{A_+}{2\sqrt{\pi}}\, e^{\int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr'} \\ F+ &= \left(\sqrt{\pi}\phi_+(0) - \frac{iA_+}{2\sqrt{\pi}}\right) e^{-\int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr'} \\ E_- &= -\frac{(-1)^{1/4}}{2\sqrt{\pi}}\left(A_- + i\pi\phi_-(0)\right) e^{\int_0^{r_M}\left(\sqrt{\lambda_-}-\sqrt{\tilde{\lambda}_-}\right)dr'} \\ F_- &= \frac{i(-1)^{1/4}}{2\sqrt{\pi}}\left(A_- - i\pi\phi_-(0)\right) e^{-\int_0^{r_M}\left(\sqrt{\lambda_-}-\sqrt{\tilde{\lambda}_-}\right)dr'} \end{aligned} \tag{3.18}$$

and similarly, above threshold, equating (3.15) with (3.11) gives

$$\begin{aligned} E_+ &= \frac{iA_+}{2\sqrt{\pi}}\left[e^{2\int_0^{r^*}\sqrt{\lambda_+}\,dr - \int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr} - e^{2\int_0^{r^\dagger}\sqrt{\tilde{\lambda}_+}\,dr + \int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr}\right] \\ &+ \frac{1}{2}\left(\sqrt{\pi}\phi_+(0) - \frac{iA_+}{2\sqrt{\pi}}\right)\left[e^{-\int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr} + e^{2\int_0^{r^\dagger}\sqrt{\tilde{\lambda}_+}\,dr - 2\int_0^{r^*}\sqrt{\lambda_+}\,dr + \int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr}\right] \\ F_+ &= \frac{A_+}{4\sqrt{\pi}}\left[e^{2\int_0^{r^*}\sqrt{\lambda_+}\,dr - 2\int_0^{r^\dagger}\sqrt{\tilde{\lambda}_+}\,dr - \int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr} + e^{\int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr}\right] \\ &+ \left(\frac{\sqrt{\pi}\phi_+(0)}{4i} - \frac{A_+}{8\sqrt{\pi}}\right)\left[e^{-\int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr - 2\int_0^{r^\dagger}\sqrt{\tilde{\lambda}_+}\,dr} - e^{-2\int_0^{r^*}\sqrt{\lambda_+}\,dr + \int_0^{r_M}\left(\sqrt{\lambda_+}-\sqrt{\tilde{\lambda}_+}\right)dr}\right], \end{aligned} \tag{3.19}$$

where $r^\dagger$ is the radius above threshold at which the eigenvalue of the repulsed eigenstate of the exponential potential passes through zero, defined by $V_0\, e^{-\mu r^\dagger} = \epsilon_v\sqrt{\epsilon_v^2 - \epsilon_\delta^2}$.

We can then equate the coefficients from matching the WKB solution onto the small-r solution with the coefficients from matching the WKB solution with the large-r solutions. After imposing appropriate boundary conditions, we can then determine the coefficients for the various physical solutions to the Schrödinger equation. For instance, we can calculate the coefficients for the repulsed and attracted eigenstates for the small-r solution. We can also determine the coefficients for the ingoing and outgoing spherical waves in the ground or excited states for the large-r solution.

Boundary Conditions

We will impose one of two sets of boundary conditions: the radially ingoing particles will either be purely in the ground state or purely in the excited state since the case of a ground state particle interacting directly with an excited state particle can be treated using existing methods in the literature [149]. Additionally, in the case of scattering, we impose regular boundary conditions at the origin, $\phi_+(0) = \phi_-(0) = 0$. This corresponds to setting $\psi(0) = 0$, which is equivalent to requiring that the physical wavefunction $\Psi(0)$ is finite at the origin. By contrast, in Ref. [289], the Sommerfeld enhancement to annihilation was extracted from irregular solutions with $\phi_\pm(0) \neq 0$.

We then define the following useful phases:

$$\begin{aligned} i\theta &\equiv \int_0^{r_M} \sqrt{\lambda_+}dr + \int_{r_M}^{r_s} \sqrt{\tilde{\lambda}_+}dr + \frac{2i\sqrt{V_0}}{\mu}\, e^{-\mu r_s/2} \\ i\varphi &\equiv \int_0^{r_M} \sqrt{\lambda_-}dr + \int_{r_M}^{r_s} \sqrt{\tilde{\lambda}_-}dr + \frac{2i\sqrt{V_0}}{\mu}\, e^{-\mu r_s/2}, \end{aligned} \tag{3.20}$$

where r_s is some radius chosen such that $V_0\, e^{-\mu r_s} \gg \epsilon_v^2,\ \epsilon_\delta^2$, as in (3.17). Note that the phase φ can be rewritten in the form

$$\begin{aligned} \varphi = -i\int_0^{r_M} \sqrt{\lambda_-(r')}dr' + \left(\frac{\max(16\epsilon_\phi^4, 4\epsilon_\delta^4)}{\mu^4}\right)^{1/4} \\ - \int_{\max(\epsilon_\phi^4,\epsilon_\delta^4/4)/(16\mu^4)}^{z_s} \frac{1}{2\mu z}\left(\sqrt{\epsilon_v^2 - \frac{\epsilon_\delta^2}{2} + \sqrt{\left(\frac{\epsilon_\delta^2}{2}\right)^2 + 16\mu^4 z}} - 2\mu z^{1/4}\right) dz \end{aligned} \tag{3.21}$$

where we have defined $z_s = V_0^2 e^{-2\mu r_s} 16\mu^4$. We can now simply take $z_s \to \infty$. Furthermore, the second integral is analytically tractable, yielding (noting that λ_- is always negative):

$$\begin{aligned} \varphi = \int_0^{r_M} \sqrt{|\lambda_-(r')|}dr' - \frac{1}{\mu}\Bigg(-2\sqrt{\frac{\epsilon_\Delta^2+\epsilon_v^2}{2} + \sqrt{\frac{\epsilon_\delta^4}{4} + \max(\epsilon_\phi^4, \epsilon_\delta^4/4)}} + \frac{i\pi}{2}(\epsilon_v + \epsilon_\Delta) \\ + \epsilon_\Delta \operatorname{arctanh}\left[\frac{\sqrt{\frac{\epsilon_v^2+\epsilon_\Delta^2}{2} + \sqrt{\frac{\epsilon_\delta^4}{4} + \max(\epsilon_\phi^4, \epsilon_\delta^4/4)}}}{\epsilon_\Delta}\right] \\ + \epsilon_v \operatorname{arctanh}\left[\frac{\sqrt{\frac{\epsilon_v^2+\epsilon_\Delta^2}{2} + \sqrt{\frac{\epsilon_\delta^4}{4} + \max(\epsilon_\phi^4, \epsilon_\delta^4/4)}}}{\epsilon_v}\right]\Bigg), \end{aligned} \tag{3.22}$$

where we have chosen the convention for the branch cut such that $\operatorname{arctanh}(x) \to -i\pi/2$ as $x \to \infty$, for x on the positive real line. There is still one numerical integral to perform, but it is fast and stable. Then, for the below-threshold case, equating (3.17) with (3.18) at the matching radius, r_M

gives

$$
\begin{aligned}
E_+ &= \frac{A_+}{2\sqrt{\pi}} e^{i\theta} = 0 \\
F_+ &= \frac{iA_+}{2\sqrt{\pi}} e^{-i\theta} = \frac{\sqrt{\mu}}{(2\pi)^{3/2}} \left[C_1 \Gamma_v + C_2 \Gamma_v^* + C_3\, \eta\, \Gamma_\Delta + C_4\, \eta^*\, \Gamma_\Delta^* \right] \\
E_- &= -\frac{A_-\, e^{i\varphi}}{2\sqrt{\pi}} = \frac{\sqrt{\mu}}{(2\pi)^{3/2}} \left[C_1 \Gamma_v e^{-\frac{\pi\epsilon_v}{\mu}} + C_2 \Gamma_v^* e^{\frac{\pi\epsilon_v}{\mu}} - C_3\, \eta\, \Gamma_\Delta e^{-\frac{\pi\epsilon_\Delta}{\mu}} - C_4\, \eta^*\, \Gamma_\Delta^* e^{\frac{\pi\epsilon_\Delta}{\mu}} \right] \\
F_- &= -\frac{A_-\, e^{-i\varphi}}{2\sqrt{\pi}} = \frac{\sqrt{\mu}}{(2\pi)^{3/2}} \left[C_1 \Gamma_v e^{\frac{\pi\epsilon_v}{\mu}} + C_2 \Gamma_v^* e^{-\frac{\pi\epsilon_v}{\mu}} - C_3\, \eta\, \Gamma_\Delta e^{\frac{\pi\epsilon_\Delta}{\mu}} - C_4\, \eta^*\, \Gamma_\Delta^* e^{-\frac{\pi\epsilon_\Delta}{\mu}} \right],
\end{aligned}
\tag{3.23}
$$

where we have defined $\epsilon_\Delta \equiv \sqrt{\epsilon_v^2 - \epsilon_\delta^2}$ and Γ_v and Γ_Δ are defined in (3.16). Similarly, above threshold, equating (3.17) with (3.19) gives

$$
\begin{aligned}
E_+ &= \frac{iA_+}{2\sqrt{\pi}} \Bigg[\left(e^{2\int_0^{r^*} \sqrt{\lambda_+}\, dr - \int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} - e^{2\int_0^{r^\dagger} \sqrt{\tilde{\lambda}_+}\, dr + \int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} \right) \\
&\quad - \frac{1}{2} \left(e^{-\int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} + e^{2\int_0^{r^\dagger} \sqrt{\tilde{\lambda}_+}\, dr - 2\int_0^{r^*} \sqrt{\lambda_+}\, dr + \int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} \right) \Bigg] = 0 \\
F_+ &= \frac{iA_+}{4\sqrt{\pi}} \Bigg\{ \left[e^{2\int_0^{r^*} \sqrt{\lambda_+}\, dr - 2\int_0^{r^\dagger} \sqrt{\tilde{\lambda}_+}\, dr - \int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} + e^{\int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} \right] \\
&\quad - \frac{1}{2} \left[e^{-\int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr - 2\int_0^{r^\dagger} \sqrt{\tilde{\lambda}_+}\, dr} - e^{-2\int_0^{r^*} \sqrt{\lambda_+}\, dr + \int_0^{r_M} \left(\sqrt{\lambda_+} - \sqrt{\tilde{\lambda}_+}\right) dr} \right] \Bigg\} \\
&= \frac{\sqrt{\mu}}{(2\pi)^{3/2}} \left[C_1 \Gamma_v + C_2 \Gamma_v^* + C_3\, \eta\, \Gamma_\Delta + C_4\, \eta^*\, \Gamma_\Delta^* \right]
\end{aligned}
\tag{3.24}
$$

Both above and below threshold, the E_+ equation gives us $A_+ = 0$, which is akin to neglecting the contribution from the repulsed eigenstate at small radii. This indicates that repulsive scattering occurs most significantly at large r, where the exponential part of the Yukawa dominates the behavior of the wavefunction. This makes intuitive sense since we are interested in the low-velocity limit, which means that incoming particles must climb up to the classically disallowed region of a repulsive potential in order to even reach the small-r region. Recall that the result $E_+ = 0$ was a consequence of assuming $\epsilon_v \lesssim \mu$: when this assumption is not satisfied, we cannot expect $A_+ = 0$ on physical grounds. Since $A_+ = 0$, the F_+ equation gives

$$
C_1 \Gamma_v + C_2 \Gamma_v^* + C_3\, \eta\, \Gamma_\Delta + C_4\, \eta\, \Gamma_\Delta^* = 0, \tag{3.25}
$$

both above and below threshold.

Cross Sections

Once we have imposed the relevant boundary conditions on the large-r wavefunctions, we can extract the scattering amplitudes and cross sections by reading off the coefficients of the outgoing solutions. Since the hypergeometric functions asymptote to 1 as $r \to \infty$, the large-r ground and excited wavefunctions approach ingoing and outgoing spherical waves:

$$\begin{aligned}\psi_1 &= C_1 \left(\frac{V_0}{4\mu^2}\right)^{i\epsilon_v/\mu} e^{-i\epsilon_v r} + C_2 \left(\frac{V_0}{4\mu^2}\right)^{-i\epsilon_v/\mu} e^{i\epsilon_v r} \\ \psi_2 &= -C_3\,\eta \left(\frac{V_0}{4\mu^2}\right)^{i\epsilon_\Delta/\mu} e^{-i\epsilon_\Delta r} - C_4\,\eta^* \left(\frac{V_0}{4\mu^2}\right)^{-i\epsilon_\Delta/\mu} e^{i\epsilon_\Delta r}.\end{aligned} \tag{3.26}$$

More generally, consider wavefunctions for nondegenerate states X and Y given by

$$\begin{aligned}\psi_X &= (A+B)e^{ikr} - Ae^{-ikr} \\ \psi_Y &= Ce^{ik'r}\end{aligned} \tag{3.27}$$

where the A terms represent the unscattered wavefunction in state X, the B term represents the elastically scattered wavefunction in state X, and the C term represents the inelastically scattered wavefunction in state Y (hence the wavenumber k' as distinct from k.) In this case, conservation of probability current dictates that $k\,|A|^2 = k\,|A+B|^2 + k'\,|C|^2$. We can reformulate these wavefunctions (recall that all wavefunctions used in this Chapter are rescaled by r) in the context of a 3-dimensional scattering problem as an ingoing cylindrical wave and a scattered outgoing spherical wave:

$$\psi = N\left(r\begin{pmatrix}e^{ikz}\\0\end{pmatrix} + \begin{pmatrix}f_X(\theta)e^{ikr}\\f_Y(\theta)e^{ik'r}\end{pmatrix}\right) = N\left(\frac{1}{2ik}\begin{pmatrix}e^{ikr}-e^{-ikr}\\0\end{pmatrix} + \begin{pmatrix}f_X(\theta)e^{ikr}\\f_Y(\theta)e^{ik'r}\end{pmatrix}\right) \tag{3.28}$$

where the second equality follows because for the s-wave,

$$e^{ikz} \to j_0(kr)P_0(\cos\theta) = \frac{\sin kr}{kr}. \tag{3.29}$$

While the interaction considered in this Chapter is purely off-diagonal, it is still possible for particles to scatter elastically, i.e. two particles incoming in the ground (excited) state may scatter off each other while remaining in the ground (excited) state. This process does not occur at tree level, in the perturbative regime (see the Subsection on the Born regime), but beyond the perturbative regime it is not suppressed. Figure 3.2 shows schematic perturbative Feynman diagrams for elastic and inelastic scattering; the non-perturbative regime requires resumming such ladder diagrams with arbitrary numbers of vector boson exchanges. Thus, below we present results for ground-to-ground and excited-to-excited elastic scattering, as well as the upscattering and downscattering rates.

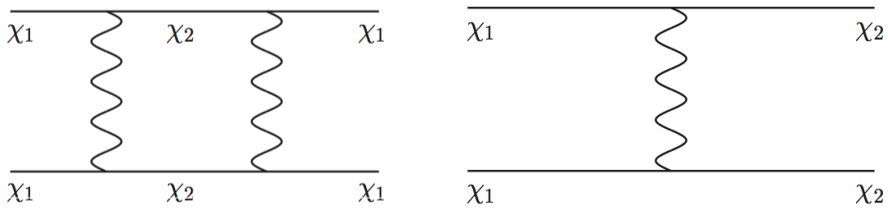

Figure 3.2: Schematic Feynman diagrams at lowest order for (left panel) "ground-to-ground" scattering (i.e. $\chi^1\chi^1 \to \chi^1\chi^1$) and (right panel) "ground-to-excited" scattering (i.e. $\chi^1\chi^1 \to \chi^2\chi^2$): swapping the fermion labels $1 \leftrightarrow 2$ gives the "excited-to-excited" and "excited-to-ground" diagrams. The u-channel diagrams also contribute.

In general to get a differential cross section, we relate the incident probability flux through an area to the scattered outgoing probability flux through a solid angle:

$$\begin{aligned} \frac{dP_{\text{in}}}{d\sigma\, dt} &= |\psi_{\text{in}}|^2 v_{\text{in}} = |N|^2 \frac{\hbar k_{\text{in}}}{m}, \\ \frac{dP_{\text{out}}}{r^2 d\Omega\, dt} &= |\psi_{\text{out}}|^2 v_{\text{out}} = \frac{|N|^2 |f(\theta)|^2}{r^2} \frac{\hbar k_{\text{in}}}{m} \\ \Rightarrow \frac{d\sigma}{d\Omega} &= \frac{k_{\text{out}}}{k_{\text{in}}} |f(\theta)|^2 \end{aligned} \tag{3.30}$$

Equating the top row of (3.28) with ψ_X from (3.27) gives

$$A = \frac{N}{2ik},\ B = N f_X(\theta) \quad \Rightarrow \quad \frac{d\sigma_{\text{elastic}}}{d\Omega} = |f_X(\theta)|^2 = \frac{1}{4k^2}\frac{|A|^2}{|B|^2} \quad \Rightarrow \quad \sigma_{\text{elastic}} = \frac{\pi}{k^2}\frac{|B|^2}{|A|^2}. \tag{3.31}$$

Similarly, equating the bottom row of (3.28) with ψ_Y from (3.27) gives

$$C = N f_Y(\theta) \quad \Rightarrow \quad \frac{d\sigma_{\text{inelastic}}}{d\Omega} = \frac{k'}{k}|f_Y(\theta)|^2 = \frac{k'}{4k^3}\frac{|C|^2}{|A|^2} \quad \Rightarrow \quad \sigma_{\text{inelastic}} = \frac{\pi k'}{k^3}\frac{|C|^2}{|A|^2} \tag{3.32}$$

If we apply the analogy to our wavefunctions for the case where we begin purely in the ground state (which corresponds to setting C_3 to zero), then the elastic scattering cross section is

$$\sigma_{\text{elastic}} = \frac{\pi}{\epsilon_v^2}\left| C_2 \left(\frac{V_0}{4\mu^2}\right)^{-i\epsilon_v/\mu} + \left(\frac{V_0}{4\mu^2}\right)^{i\epsilon_v/\mu} \right|^2 \tag{3.33}$$

and the inelastic scattering cross section is

$$\sigma_{\text{inelastic}} = \frac{\pi\epsilon_\Delta}{\epsilon_v^3}\left|C_4\, \eta^*\right|^2. \tag{3.34}$$

Similarly, for the case where we begin purely in the excited state (which corresponds to setting C_1 to zero), then the elastic scattering cross section is

$$\sigma_{\text{elastic}} = \frac{\pi \left| C_4\, \eta^* \left(\frac{V_0}{4\mu^2}\right)^{-i\epsilon_\Delta/\mu} + \eta \left(\frac{V_0}{4\mu^2}\right)^{i\epsilon_\Delta/\mu} \right|^2}{\epsilon_\Delta^2 \left|\eta\right|^2} \tag{3.35}$$

and the inelastic scattering cross section is

$$\sigma_{\text{inelastic}} = \frac{\pi \epsilon_v \left|C_2\right|^2}{\epsilon_\Delta^3 \left|\eta\right|^2}\,. \tag{3.36}$$

These cross sections must be multiplied by $\hbar^2/(c^2\alpha^2 m_\chi^2)$ to obtain the physical cross sections, since we initially rescaled r by $\alpha m_\chi c/\hbar$.

Incoming in the Ground State

We will impose boundary conditions such that the ingoing wave is purely in the ground state, which implies that $C_3 = 0$. We are free to set $C_1 = 1$ up to some overall normalization. Dividing the E_- equation by the F_- equation of (3.23) yields

$$e^{2i\varphi} = \frac{\Gamma_v e^{-\frac{\pi\epsilon_v}{\mu}} + C_2 \Gamma_v^* e^{\frac{\pi\epsilon_v}{\mu}} - C_4\, \eta^*\, \Gamma_\Delta^* e^{\frac{\pi\epsilon_\Delta}{\mu}}}{\Gamma_v e^{\frac{\pi\epsilon_v}{\mu}} + C_2 \Gamma_v^* e^{-\frac{\pi\epsilon_v}{\mu}} - C_4\, \eta^*\, \Gamma_\Delta^* e^{-\frac{\pi\epsilon_\Delta}{\mu}}} \tag{3.37}$$

and combining this with (3.25) gives

$$\begin{aligned} C_4 &= \frac{-2\,\Gamma_v \sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\left[1 + e^{2i\varphi}\right]}{\eta^*\,\Gamma_\Delta^* \left[\left(e^{\frac{\pi\epsilon_\Delta}{\mu}} + e^{\frac{\pi\epsilon_v}{\mu}}\right) - e^{2i\varphi}\left(e^{-\frac{\pi\epsilon_\Delta}{\mu}} + e^{-\frac{\pi\epsilon_v}{\mu}}\right)\right]}\,. \\ C_2 &= \frac{\Gamma_v}{\Gamma_v^*}\left[\frac{2\sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\left[1 + e^{2i\varphi}\right]}{\left(e^{\frac{\pi\epsilon_\Delta}{\mu}} + e^{\frac{\pi\epsilon_v}{\mu}}\right) - e^{2i\varphi}\left(e^{-\frac{\pi\epsilon_\Delta}{\mu}} + e^{-\frac{\pi\epsilon_v}{\mu}}\right)} - 1\right] \end{aligned} \tag{3.38}$$

So by (3.33) and denoting the initial and final states as "gr" for "ground" and "ex" for "excited", the elastic scattering cross section is

$$\sigma_{\text{gr}\to\text{gr}} = \frac{\pi}{\epsilon_v^2}\left|1 + \left(\frac{V_0}{4\mu^2}\right)^{-\frac{2i\epsilon_v}{\mu}}\left(\frac{\Gamma_v}{\Gamma_v^*}\right)\left[\frac{\cosh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_v-\epsilon_\Delta)}{2\mu} + i\varphi\right)}{\cosh\left(\frac{\pi(\epsilon_v-\epsilon_\Delta)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu} - i\varphi\right)}\right]\right|^2 \tag{3.39}$$

and by (3.34), the inelastic scattering cross section is

$$\sigma_{\text{gr}\to\text{ex}} = \frac{2\pi\cos^2\varphi \sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\sinh\left(\frac{\pi\epsilon_\Delta}{\mu}\right)}{\epsilon_v^2 \cosh^2\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\left(\cosh\left(\frac{\pi(\epsilon_v+\epsilon_\Delta)}{\mu}\right) - \cos(2\varphi)\right)}\,. \tag{3.40}$$

Incoming in the Excited State

We will impose boundary conditions such that the ingoing wave is purely in the excited state, which implies that $C_1 = 0$. We are free to set $C_3 = 1$ up to some overall normalization. Dividing the E_- equation by the F_- equation of (3.23) yields

$$e^{2i\varphi} = \frac{C_2\Gamma_v^* e^{\frac{\pi\epsilon_v}{\mu}} - \eta\,\Gamma_\Delta e^{-\frac{\pi\epsilon_\Delta}{\mu}} - C_4\,\eta^*\,\Gamma_\Delta^* e^{\frac{\pi\epsilon_\Delta}{\mu}}}{C_2\Gamma_v^* e^{-\frac{\pi\epsilon_v}{\mu}} - \eta\,\Gamma_\Delta e^{\frac{\pi\epsilon_\Delta}{\mu}} - C_4\,\eta^*\,\Gamma_\Delta^* e^{-\frac{\pi\epsilon_\Delta}{\mu}}} \tag{3.41}$$

and combining this with (3.25) gives

$$\begin{aligned} C_4 &= \frac{\eta\,\Gamma_\Delta\left[e^{2i\varphi}\left(e^{\frac{\pi\epsilon_\Delta}{\mu}} + e^{-\frac{\pi\epsilon_v}{\mu}}\right) - \left(e^{\frac{\pi\epsilon_v}{\mu}} + e^{-\frac{\pi\epsilon_\Delta}{\mu}}\right)\right]}{\eta^*\,\Gamma_\Delta^*\left[\left(e^{\frac{\pi\epsilon_\Delta}{\mu}} + e^{\frac{\pi\epsilon_v}{\mu}}\right) - e^{2i\varphi}\left(e^{-\frac{\pi\epsilon_\Delta}{\mu}} + e^{-\frac{\pi\epsilon_v}{\mu}}\right)\right]} \\ C_2 &= \frac{-2\,\eta\,\Gamma_\Delta \sinh\left(\frac{\pi\epsilon_\Delta}{\mu}\right)\left[1 + e^{2i\varphi}\right]}{\Gamma_v^*\left[\left(e^{\frac{\pi\epsilon_\Delta}{\mu}} + e^{\frac{\pi\epsilon_v}{\mu}}\right) - e^{2i\varphi}\left(e^{-\frac{\pi\epsilon_\Delta}{\mu}} + e^{-\frac{\pi\epsilon_v}{\mu}}\right)\right]} \end{aligned} \tag{3.42}$$

So by (3.35), the elastic scattering cross section is

$$\sigma_{\text{ex}\to\text{ex}} = \frac{\pi}{\epsilon_\Delta^2}\left|1 + \left(\frac{V_0}{4\mu^2}\right)^{-\frac{2i\epsilon_\Delta}{\mu}}\left(\frac{\Gamma_\Delta}{\Gamma_\Delta^*}\right)\left[\frac{\cosh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu} + i\varphi\right)}{\cosh\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu} - i\varphi\right)}\right]\right|^2 \tag{3.43}$$

and by (3.36), the inelastic scattering cross section is

$$\sigma_{\text{ex}\to\text{gr}} = \frac{2\pi\cos^2\varphi\,\sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\sinh\left(\frac{\pi\epsilon_\Delta}{\mu}\right)}{\epsilon_\Delta^2\cosh^2\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\left(\cosh\left(\frac{\pi(\epsilon_v+\epsilon_\Delta)}{\mu}\right) - \cos(2\varphi)\right)}. \tag{3.44}$$

Relation to the Transfer Cross Section

When considering the effects of DM scattering on structure formation in general, the total cross section may acquire an unphysical forward divergence. Generally, the literature has instead employed the transfer cross section σ_T, which determines the longitudinal momentum transfer:

$$\sigma_T = \int d\Omega\,(1 - \cos\theta)\,\frac{d\sigma}{d\Omega}. \tag{3.45}$$

Since our differential cross sections are angle-independent, we can pull those out of the integral. Since the $\cos\theta$ term is orthogonal to the $\sin\theta$ term in the $d\Omega$ Jacobian, the remaining integral just gives

$$\sigma_T = 4\pi\,\frac{d\sigma}{d\Omega}, \tag{3.46}$$

which is the same as the cross section that we computed. (Note that [149] argues for the use of the viscosity cross section instead, but computes σ_T in order to make contact with the literature – we follow their approach, albeit we reiterate that for the s-wave component of the amplitude, the distinction is a trivial one.)

Numerical Cross-Checks

Here we compare the results of our analytic approximation to the exact s-wave cross sections, obtained by numerically solving the matrix Schrödinger equation using `Mathematica`. Numerically solving for the scattering amplitudes proved computationally expensive and unstable in certain regions of parameter space, and was in all cases several orders of magnitude slower than computing the semi-analytic results, further motivating the use of these approximations. The results are shown in Figures 3.3-3.6, for two sample choices of the mass splitting parameter $\epsilon_\delta = 0.01, 0.05$, and for elastic scattering in the ground and excited states (i.e. "gr→gr" and "ex→ex" respectively), upscattering ("gr→ex") and downscattering ("ex→gr"). We show only results for $\epsilon_v > \epsilon_\delta$ in this figure, since if this is not the case only ground-to-ground scattering is possible, and also restrict $\epsilon_v < \epsilon_\phi$ since otherwise we expect both that our approximation may break down and that higher partial waves will become important. We note that in all cases, notable resonances and antiresonances develop at particular values of ϵ_ϕ.

We find that our approximations agree with the numerics to within 10% away from resonances (at resonances, minor shifts can cause huge fractional disagreement in spite of the fact that the approximation actually does capture the behavior quite well), and correctly describe the resonance positions. (Note that the minor numerical artifacts in these plots reflect the instability of the numerical calculation, and should be ignored for purposes of comparison.)

Figure 3.7 shows the analytic calculation of the effect of a non-zero mass splitting on elastic scattering in the ground state, including for $\epsilon_v < \epsilon_\delta$, for the same two choices of ϵ_δ. For ground-state elastic scattering, the main effect of the mass splitting is to cause the positions of the resonances to shift. The effect is much more pronounced for small ϵ_ϕ. We will understand this behavior in the following subsection, by studying analytically tractable limits of eqs. 3.86-3.89. The numerical results again agree with the analytic results in this regime (we will show an explicit comparison in the case of small ϵ_v in Figure 3.9).

Features and Limits of the Scattering Cross Sections

We will now study various limits of the scattering cross sections using our analytic approximation. First, however, let us emphasize a point regarding our definition of ϵ_v. This quantity relates to the energy of the particle *relative to the ground state*. A particle freely propagating in the excited state will *always* have $\epsilon_v > \epsilon_\delta$. While, as noted above, the "uphill" and "downhill" scattering *amplitudes* are identical above the kinematic threshold given by the mass splitting, the cross section for downscattering diverges as $1/\epsilon_\Delta$ close to threshold (except at anti-resonances) while the cross section for upscattering goes identically to zero. These different behaviors originate solely from the very different phase-space factors near threshold, and are demonstrated in Figures 3.3-3.7.

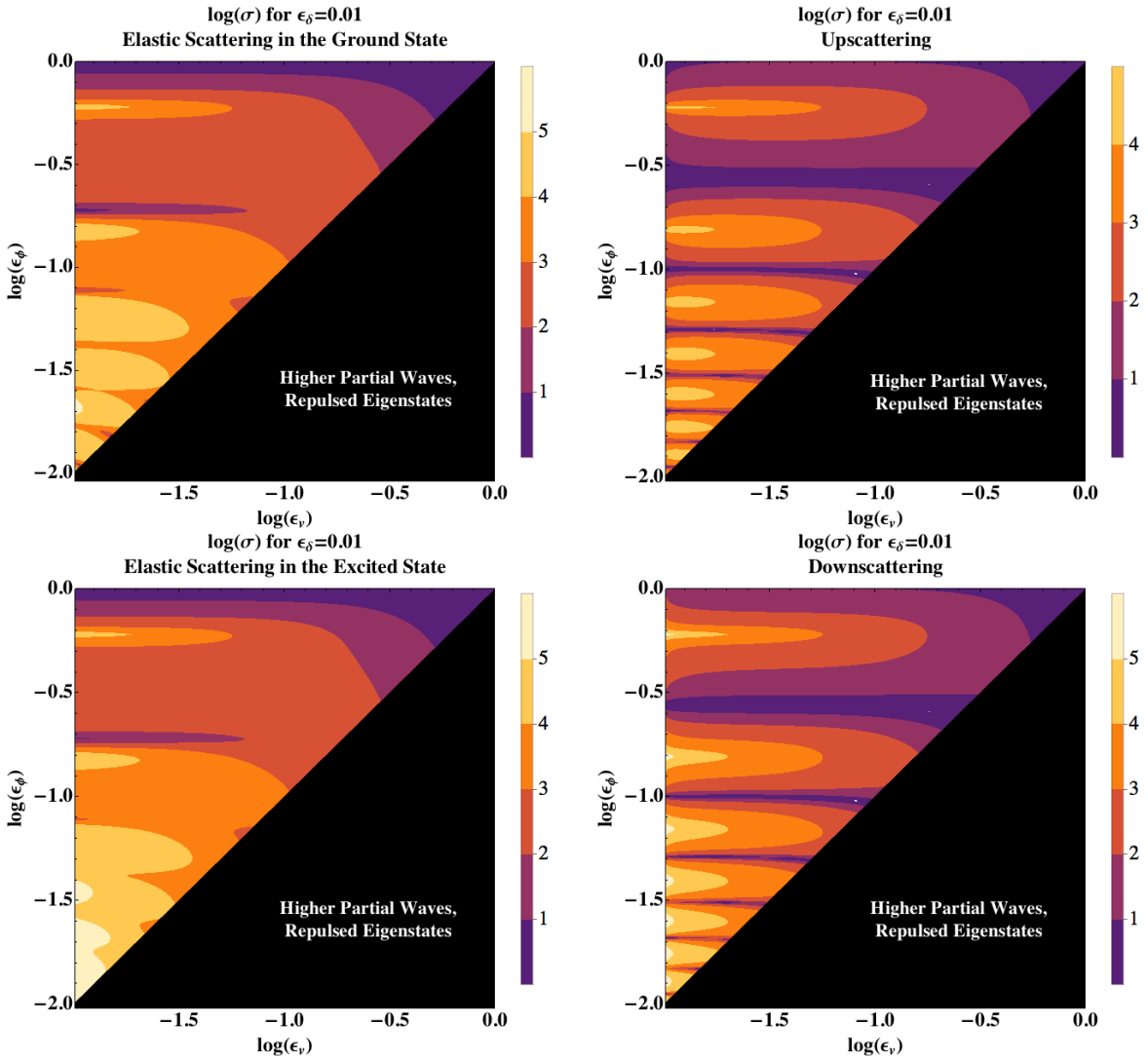

Figure 3.3: The dimensionless cross sections computed using our approximations, for $\epsilon_\delta = 0.01$. Note that "log" indicates base 10.

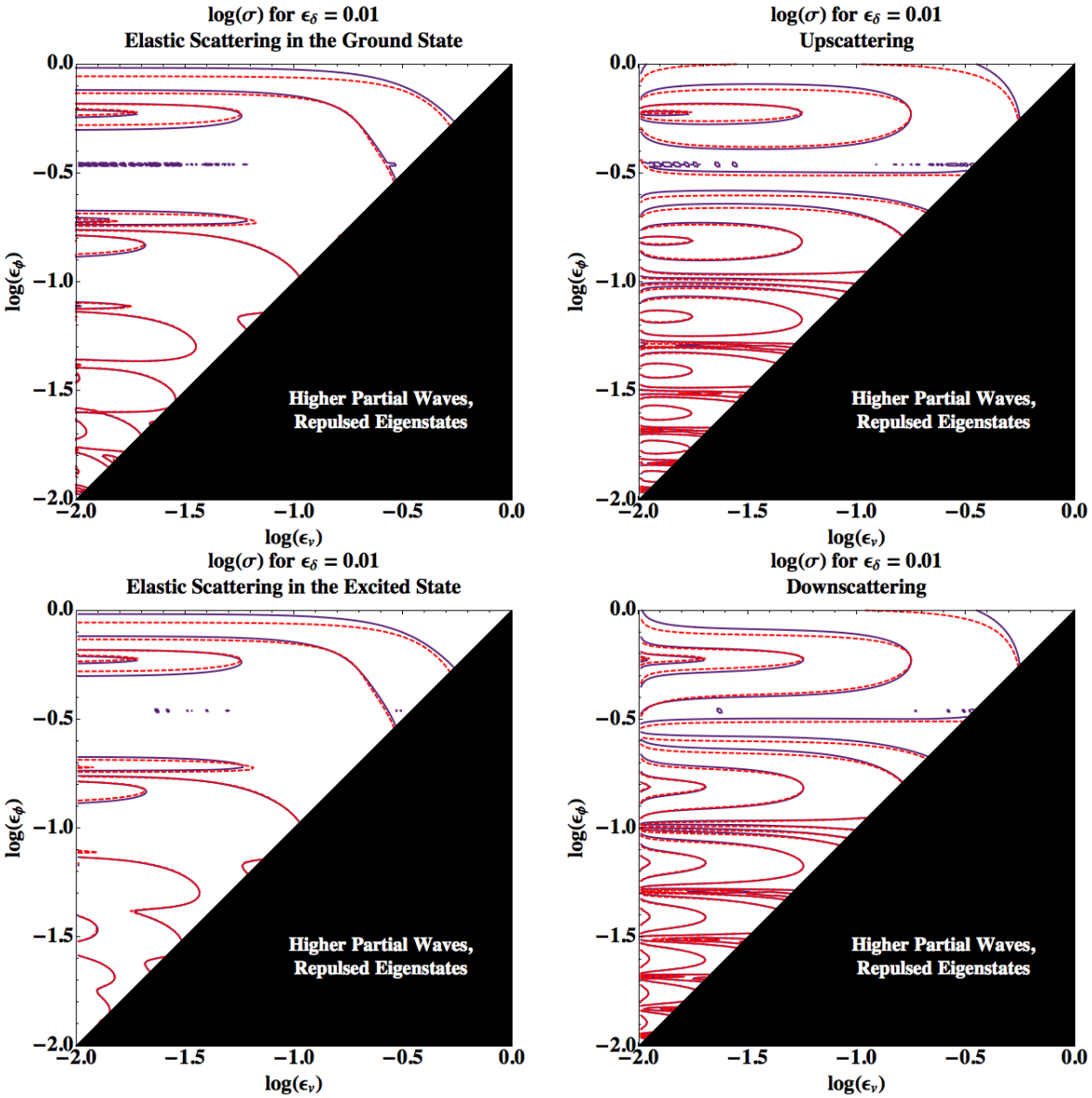

Figure 3.4: The same as Fig. 3.3 except with a comparison of these results (red dashed lines) to the exact numerical results (solid lines).

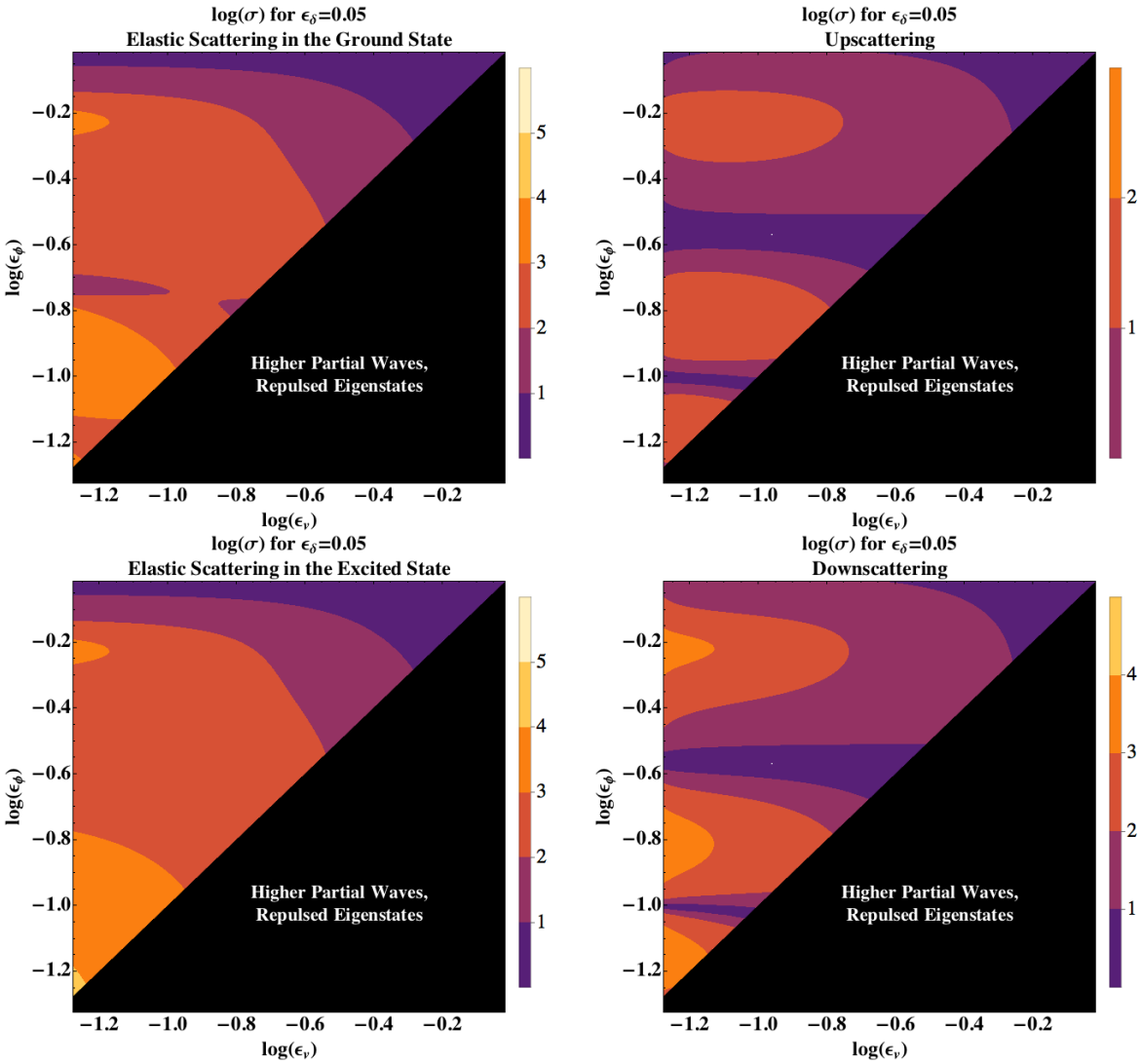

Figure 3.5: As Figure 3.3, but with $\epsilon_\delta = 0.05$.

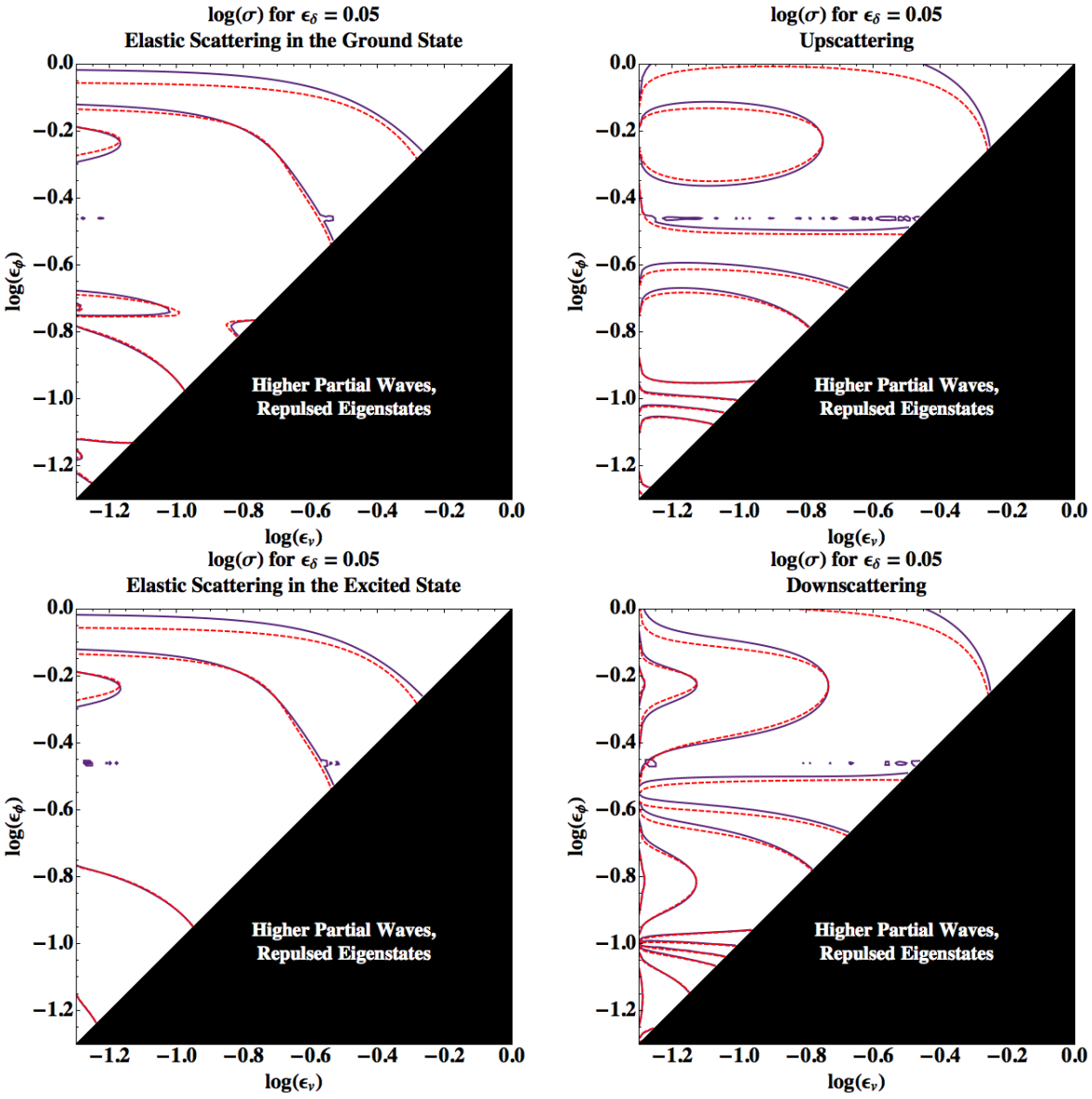

Figure 3.6: As Figure 3.4, but with $\epsilon_\delta = 0.05$.

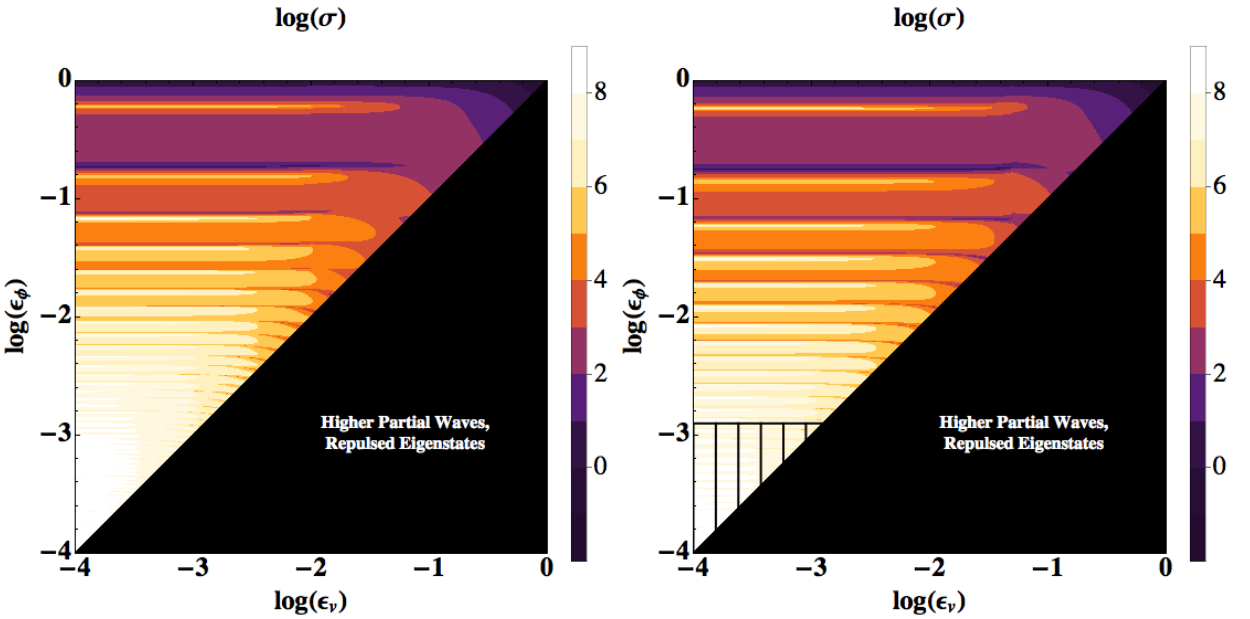

Figure 3.7: An analytic calculation of the elastic ground→ground dimensionless cross section in the ϵ_ϕ vs. ϵ_v plane, with $\epsilon_\delta = 0.01$ (*left*) and $\epsilon_\delta = 0.05$ (*right*). The resonance structures in the upper left are very similar to those depicted in Figure 1 of [149]. The black areas to the lower right of the diagonal on the plots indicates where our approximation is no longer valid because of the repulsed eigenstates and higher partial waves (i.e. when $\epsilon_v > \epsilon_\phi$.) The black cross-hatching on the bottom of the $\epsilon_\delta = 0.05$ plot indicates where $\epsilon_\phi < \epsilon_\delta^2/2$ which is a region where our approximation is less accurate (see Table 3.1.) Note that the use of "log" indicates the base 10 logarithm.

For larger ϵ_v, upscattering and downscattering are equally likely, since there is relatively little energetic overhead to upscattering. As well as Figures 3.3-3.7, the threshold behavior is shown for a constant-ϵ_ϕ slice in Figure 3.10.

These cross sections, with the same ϵ_v, correspond to different physical scenarios in the context of a virialized DM halo — for instance, an *excited* state particle in a virialized halo will give a larger ground state velocity and thus correspond to a larger resulting ϵ_v than for virialized *ground* state particles in the halo. We further discuss the astrophysical relevance of these scenarios in Section 3.4.

The Degenerate Limit

In the limit where $\delta \to 0$, analytic expressions for the scattering cross section in a repulsive or attractive potential have been previously derived [149]. In this limit, our "elastic" and "inelastic" cross sections refer to scatterings between particular linear combinations of the attracted and repulsed two-body eigenstates, and it is natural to switch to the basis of (dark) charge eigenstates accordingly (which in this limit are also mass eigenstates).

In the $\delta \to 0$ limit, the potential matrix can be diagonalized, yielding the exact eigenstate basis

$\psi_+ = \frac{1}{\sqrt{2}}(-1,1)$, $\psi_- = \frac{1}{\sqrt{2}}(1,1)$ (as in Eq. 3.6). Since the potential is now diagonal, scatterings from ψ_+ to the ψ_- (and vice versa) do not occur: the two eigenstates are decoupled. The ψ_+ and ψ_- eigenstates experience, respectively, a repulsive and attractive potential.

For the $-$ state, let the scattering solution $\phi_-(r)$ (for the coefficient of the ψ_- eigenvector) have the asymptotic form $\phi_-(r) = e^{i(kr+2\delta_-)} + e^{-ikr}$. The phase shift δ_- characterizes the scattering amplitude, which is given by $\left|1 - e^{2i\delta_-}\right|^2$. Likewise, for the $+$ state, let the phase shift be δ_+.

Since the differential equation is linear, any linear combination of these solutions ($A\phi_-(r)\psi_- + B\phi_+(r)\psi_+$) is also a solution. In particular, if we set $A = B = 1/\sqrt{2}$ and $A = -B = 1/\sqrt{2}$, we obtain the two solutions:

$$\psi(r) = \begin{pmatrix} e^{ikr}\left(\frac{e^{2i\delta_-} - e^{2i\delta_+}}{2}\right) \\ e^{ikr}\left(\frac{e^{2i\delta_-} + e^{2i\delta_+}}{2}\right) + e^{-ikr} \end{pmatrix}, \quad \psi(r) = \begin{pmatrix} e^{ikr}\left(\frac{e^{2i\delta_-} + e^{2i\delta_+}}{2}\right) + e^{-ikr} \\ e^{ikr}\left(\frac{e^{2i\delta_-} - e^{2i\delta_+}}{2}\right) \end{pmatrix}. \tag{3.47}$$

These correspond to the cases we studied above, where the particles are initially purely in the ground or excited states. So we see that by calculating the phase shifts for these initial conditions (given by the A, B and C coefficients in Eq. 3.31 and Eq. 3.32), we can recover the values for δ_- and δ_+, and vice versa. Our cross sections are given by:

$$\sigma(\text{ground} \to \text{ground}) = \frac{\pi}{k^2}\left|1 - \frac{e^{2i\delta_-} + e^{2i\delta_+}}{2}\right|^2, \quad \sigma(\text{ground} \to \text{excited}) = \frac{\pi}{k^2}\left|\frac{e^{2i\delta_-} - e^{2i\delta_+}}{2}\right|^2, \tag{3.48}$$

and in this case, since $\delta = 0$, swapping the identifications of "ground" and "excited" states has no effect. Note that the sum of these cross sections gives $\sigma_{\text{tot}} = \frac{\pi}{k^2}\left(\left|1 - e^{2i\delta_-}\right|^2 + \left|1 - e^{2i\delta_+}\right|^2\right) = \sigma_- + \sigma_+$, as it must — the total scattering rate cannot depend on the choice of basis.

In the limit where the phase shifts are small (which we will see is the case at low velocities and away from resonances), we can expand:

$$\sigma(\text{ground} \to \text{ground}) \approx \frac{\pi}{k^2}\left|\delta_- + \delta_+\right|^2, \quad \sigma(\text{ground} \to \text{excited}) = \frac{\pi}{k^2}\left|\delta_- - \delta_+\right|^2. \tag{3.49}$$

The authors of [149] define a quantity a which corresponds to our ϵ_v, and a quantity $c = 1/(\kappa\epsilon_\phi)$, where κ is set to 1.6. The phase shifts derived for the repulsive and attractive case by [149] in the low-velocity limit are then given by:

$$\delta_- \approx -[2\gamma + \psi(1+\sqrt{c}) + \psi(1-\sqrt{c})]ac, \quad \delta_+ \approx -[2\gamma + \psi(1+i\sqrt{c}) + \psi(1-i\sqrt{c})]ac, \tag{3.50}$$

where γ is the Euler-Mascheroni constant and $\psi(z)$ is the digamma function. Note these phase shifts become small in the low-velocity limit due to the scaling with a, as claimed above.

The asymptotic expansions of the digamma function, as $|z| \to \infty$, are $\psi(z) \approx \ln(z) + i\pi(i\cot(\pi z) - 1)\lfloor |\arg(z)|/\pi \rfloor$ for z not a negative integer. Thus in the large-c limit (corresponding to $\epsilon_\phi \ll 1$, which is necessary for our approximations to hold), and neglecting terms of $\mathcal{O}(1/c)$ and higher, these phase shifts approach:

$$\delta_- = -\left[2\gamma + \ln(c) + \pi\cot(\pi\sqrt{c})\right]ac, \quad \delta_+ = -\left[2\gamma + \ln(c)\right]ac. \tag{3.51}$$

So the cross sections for ground-ground and ground-excited scattering should be set by:

$$\begin{aligned}\sigma(\text{ground} \to \text{ground}) &= \pi \left[4\gamma + 2\ln(c) + \pi \cot(\pi\sqrt{c})\right]^2 c^2, \\ \sigma(\text{ground} \to \text{excited}) &= \pi^3 \cot^2(\pi\sqrt{c}) c^2.\end{aligned} \tag{3.52}$$

Note the prefactor $1/k^2$ has canceled out the factors of a in the phase shifts.

We now take the same limits (first $\delta \to 0$, and then $v \to 0$) in our semi-analytic approximation. Setting $\epsilon_v = \epsilon_\Delta$ we obtain:

$$\Gamma_v \to \sqrt{\pi}\, \Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right) \Gamma\left(\frac{1}{2} + \frac{i\epsilon_v}{2}\right) = \pi\, 2^{-\frac{2i\epsilon_v}{\mu}}\, \Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right), \tag{3.53}$$

where the second equality comes from the gamma function identities $\Gamma(1+z) = z\,\Gamma(z)$, $\Gamma(1-z)\,\Gamma(z) = \pi/\sin(\pi z)$, and $\Gamma(z)\Gamma(z+1/2) = 2^{1-2z}\sqrt{\pi}\,\Gamma(2z)$. So the elastic scattering cross section to first order in ϵ_δ (for both the ground and excited state, since they are now degenerate) becomes:

$$\sigma_{\text{elastic}} = \frac{\pi}{\epsilon_v^2} \left| 1 + \frac{\Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right)}{\Gamma\left(1 - \frac{i\epsilon_v}{\mu}\right)} \left(\frac{i \sin\varphi \cosh\left(\frac{\pi\epsilon_v}{\mu}\right)}{\sinh\left(\frac{\pi\epsilon_v}{\mu} - i\varphi\right)} \right) \left(\frac{V_0}{\mu^2}\right)^{-2i\epsilon_v/\mu} \right|^2. \tag{3.54}$$

Off-resonance, in the limit as $\epsilon_v \to 0$, $\sinh\left(\frac{\pi\epsilon_v}{\mu} - i\varphi\right) \to -i\sin\varphi$, and the scattering amplitude approaches zero. More precisely, a Taylor expansion yields:

$$\sigma_{\text{elastic}} \to \frac{\pi}{\mu^2} \left[2\gamma + 2\ln\left(\frac{V_0}{\mu^2}\right) + \pi \cot\varphi \right]^2, \tag{3.55}$$

where γ is the Euler-Mascheroni constant. On-resonance, where $\sin\varphi = 0$, the elastic scattering amplitude is simply 1.

Meanwhile, for the inelastic case, setting $\epsilon_v = \epsilon_\Delta$ yields

$$\sigma_{\text{inelastic}} = \frac{2\pi \cos^2\varphi \sinh^2\left(\frac{\pi\epsilon_v}{\mu}\right)}{\epsilon_v^2 \left(\cosh\left(\frac{2\pi\epsilon_v}{\mu}\right) - \cos(2\varphi)\right)}. \tag{3.56}$$

Off-resonance, as $\epsilon_v \to 0$, this probability approaches

$$\sigma_{\text{inelastic}} \to \pi \left(\frac{\pi \cot\varphi}{\mu}\right)^2. \tag{3.57}$$

On-resonance, where $\cos(2\varphi) = 1$, the inelastic scattering amplitude simply approaches 1.

We see that these would agree precisely with the approximate forms of the cross sections derived from the results of [149] if we made the replacements:

$$\mu \to 1/c, \quad \varphi \to \pi\sqrt{c}, \quad \ln(V_0/\mu) \to \gamma. \tag{3.58}$$

These replacements are parametrically correct — $\mu \sim \epsilon_\phi \sim 1/c$ up to $\mathcal{O}(1)$ factors, likewise $V_0 \sim \mu$ up to $\mathcal{O}(1)$ corrections. The results are most sensitive to the identification $\varphi \to \pi\sqrt{c}$, since this sets the resonance positions: taking $c = 1/(\kappa\epsilon_\phi)$, and our approximate expression $\varphi \sim \sqrt{2\pi/\epsilon_\phi}$, we see that they agree exactly if $\kappa = \pi/2 \approx 1.57$. The value of $\kappa = 1.6$ chosen by [149] therefore leads to percent-level agreement in the resonance positions.

Perfect agreement between the two analyses should not be expected, since they use different potentials (albeit with similar properties), but our approach agrees both qualitatively and quantitatively with the results of [149] in the region of parameter space where they can both be used.

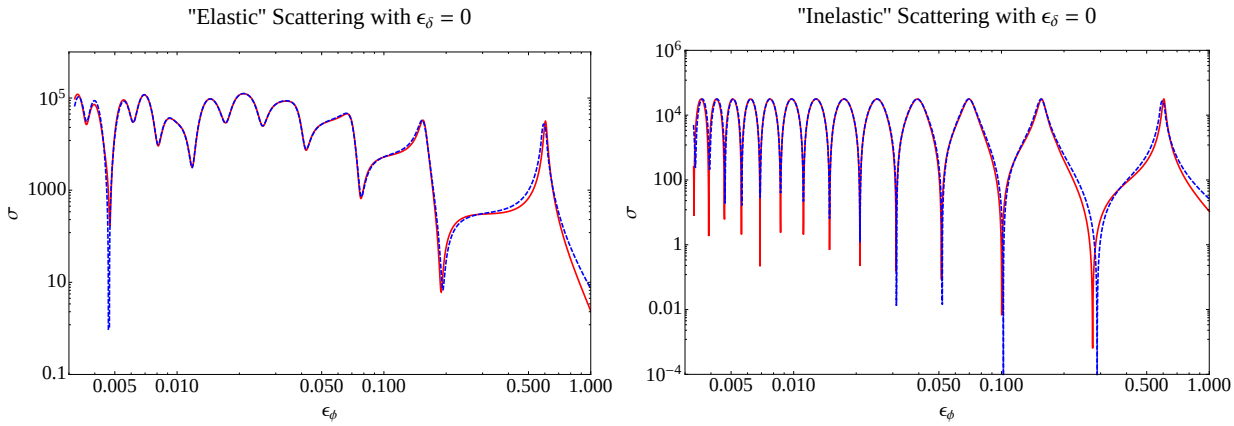

Figure 3.8: Scattering with $\epsilon_\delta = 0$ and $\epsilon_v = 0.01$, where the red solid curve comes from our analytic approximation and the blue dashed curve comes from numerically solving the Schrödinger equation. The quotation marks in the plot titles serve as a reminder that with $\epsilon_\delta = 0$ there is no inelastic scattering because the states are degenerate. Resonances occur as predicted by our Taylor expansions.

We plot our results for the cross section in this limit in Figure 3.8, again finding good agreement between our full semi-analytic approximation and the numerical results for $\epsilon_\phi < 1$. These results can be used more generally for elastic and inelastic scattering where ϵ_δ is small relative to all the other parameters in the problem.

The Low-Velocity Limit

Now let us consider the case where $\epsilon_v \to 0$, *without* first setting $\epsilon_\delta \to 0$. In this limit, scattering into the excited state is forbidden, so we will only examine elastic scattering in the ground state. Expanding $\sigma_{\text{gr}\to\text{gr}}$ to first order in ϵ_v yields

$$\sigma_{\text{gr}\to\text{gr}} \approx \frac{\pi}{\mu^2}\left(\pi\cot\left(\varphi - \frac{\pi\epsilon_\delta}{2\mu}\right) + 2\ln\left(\frac{V_0}{\mu^2}\right) + 2\gamma - 4\ln 2 - 2\psi_0\left(\frac{\epsilon_\delta}{2\mu} + \frac{1}{2}\right)\right)^2 \tag{3.59}$$

where ψ_0 is the digamma function and γ is the Euler-Mascheroni constant.

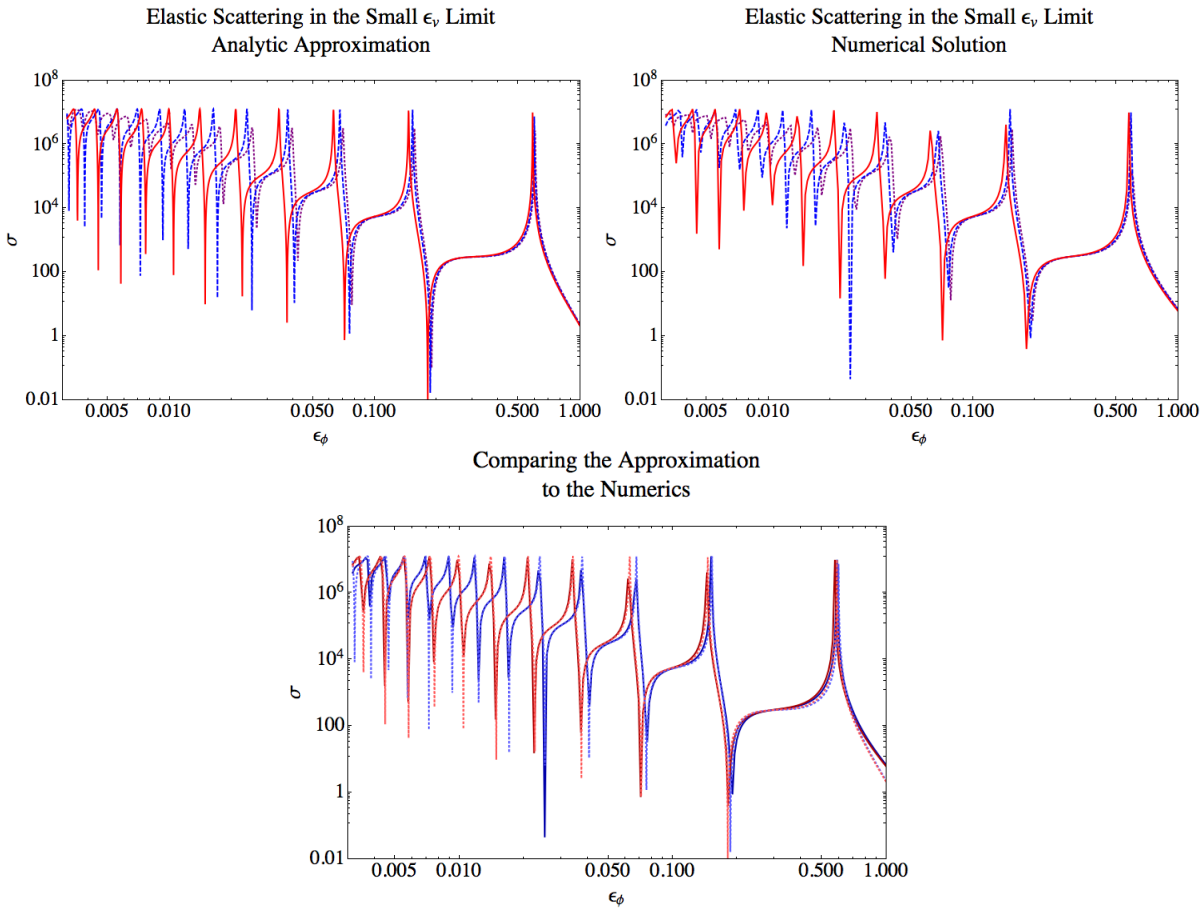

Figure 3.9: Scattering from the ground state to the ground state with $\epsilon_v = 0.001$. Shown here are different scenarios with $\epsilon_\delta = 0$ (purple, dotted), $\epsilon_\delta = 0.01$ (dashed, blue), and $\epsilon_\delta = 0.03$ (solid, red). The upper left panel shows our analytic approximation and the upper right panel shows the numerical solution, which are in close agreement. To show the agreement between the two calculations explicitly, in the lower panel we have overlaid the analytic and numerical results for the $\epsilon_\delta = 0.01$ and $\epsilon_\delta = 0.03$ cases, with the solid dark lines representing the numerical computation and the lighter dashed lines representing our approximation. The analytic result accurately reproduces the numerics, and we can see that the presence of a mass splitting — even a small one— shifts the resonance positions, most noticeably for small ϵ_ϕ.

We see that the cross section does not vary with ϵ_v in this low-velocity limit, with the cross section approaching the expected geometric size of $\sim \pi/m_\phi^2$, up to constant prefactors, once we convert to dimensionful parameters (making the approximation $\mu \sim \epsilon_\phi = m_\phi/(\alpha m_\chi)$, and then multiplying the dimensionless cross section by $1/(\alpha m_\chi)^2$ as usual). Resonances occur when $\varphi = (n + \epsilon_\delta/2\mu)\pi$, and in the case where $\delta = 0$ (as mentioned previously) the resonance positions are $\varphi = n\pi$. Thus the presence of a mass splitting induces a shift to the resonance positions at velocities below the threshold, to smaller ϵ_ϕ: Figure 3.9 shows the impact of the mass splitting on the resonances at low velocity for several choices of δ. The effect on the resonances is the same as found for the case of annihilation [289]. We also demonstrate in Figure 3.9 that our analytic approximation accurately reproduces the numerical results for low-ϵ_v scattering below threshold.

Except for the shift in resonance positions, this cross section is very similar in form to (3.55); in the limit as $\epsilon_\delta \to 0$ (but $\epsilon_v < \epsilon_\delta$) they are identical, except that (3.59) has a 4γ term rather than 2γ (two extra γ's come from $\psi_0(\frac{1}{2})$). This is a subdominant correction; generally larger contributions will arise from the cot and log terms. So we see that for elastic scattering in the ground state, the effect of the mass splitting is primarily just to shift the resonance positions; this contrasts with the case of annihilation where switching on the mass splitting can lead to a generic enhancement of the cross section by a factor of 2-4 at low velocities [289].

The Threshold ($\epsilon_v = \epsilon_\delta$) Limit

Scattering amplitudes involving the excited state will be suppressed by ϵ_Δ as ϵ_Δ approaches zero from above, but the corresponding cross sections need not vanish. The case where $\epsilon_v \approx \epsilon_\delta$ corresponds, for particles initially in the excited state, to very low physical velocities. We perform a Taylor expansion in small (but real and positive) ϵ_Δ, finding for the cross sections:

$$\sigma_{\text{gr}\to\text{ex}} \approx \frac{4\pi^2\epsilon_\Delta \cos^2\varphi \tanh\left(\frac{\pi\epsilon_\delta}{2\mu}\right)}{\epsilon_\delta^2\,\mu\left(\cosh\left(\frac{\pi\epsilon_\delta}{\mu}\right) - \cos(2\varphi)\right)} \tag{3.60}$$

$$\sigma_{\text{ex}\to\text{gr}} \approx \frac{4\pi^2 \cos^2\varphi \tanh\left(\frac{\pi\epsilon_\delta}{2\mu}\right)}{\epsilon_\Delta\,\mu\left(\cosh\left(\frac{\pi\epsilon_\delta}{\mu}\right) - \cos(2\varphi)\right)} \tag{3.61}$$

$$\sigma_{\text{ex}\to\text{ex}} \approx \frac{1}{\mu^2}\left(\zeta^2 + \frac{2\pi\cos\varphi}{\cosh\left(\frac{\pi\epsilon_\delta}{\mu}\right) - \cos(2\varphi)}\left(2\,\zeta\sin\varphi + \cos\varphi\,\text{sech}^2\left(\frac{\pi\epsilon_\delta}{2\mu}\right)\right)\right)^2 \tag{3.62}$$

where for convenience, we have defined the subdominant $\mathcal{O}(1)$ term

$$\zeta \equiv 2\gamma - 2\psi_0\left(\frac{i\epsilon_\delta}{2\mu} + \frac{1}{2}\right) + i\pi\tanh\left(\frac{\pi\epsilon_\delta}{2\mu}\right) + \ln\left(\frac{V_0^2}{16\mu^4}\right), \tag{3.63}$$

which is a real quantity because the term with the hyperbolic tangent cancels out the imaginary part of the digamma function.

We see that in all cases there is a potentially large enhancement corresponding to the zero-δ resonances, $\varphi = n\pi$ so $\cos(2\phi) = 1$. The cross section does not actually diverge at these pseudo-resonances, but scales as $1/(\cosh(\pi\epsilon_\delta/\mu) - 1)$, and so is large when $\epsilon_\delta \ll \mu$.

The upscattering cross section vanishes as $\epsilon_\Delta \to 0$, as expected, as the phase space for newly-excited particles shrinks to zero. The elastic scattering cross section for the particles in the excited state scales parametrically as $1/\mu^2$, except close to the resonances, where it instead scales as $1/\epsilon_\delta^2$ if $\epsilon_\delta \ll \mu$. Both these behaviors correspond to geometric cross sections, one governed by the range of the force and one by the momentum transfer associated with virtual de-excitation to the ground state. The downscattering cross section — which is perhaps most interesting for scenarios where an abundant relic population of DM in the halo exists in the excited state — diverges as $1/\epsilon_\Delta$, meaning that if $v_{\rm ex} = \alpha\epsilon_\Delta$ is the physical velocity of the incoming particles in the excited state, $\sigma v_{\rm ex}$ will approach a constant value at low velocities. For $\epsilon_\delta \ll \mu$, the cross section scales as $\epsilon_\delta/(\epsilon_\Delta\mu^2)$ away from the resonances, and $1/(\epsilon_\Delta\epsilon_\delta)$ close to the resonances. Inserting the dimensionful prefactors, the physical cross sections for downscattering and elastic scattering in the excited state have the following scaling behavior:

$$\sigma_{\rm ex\to gr}\, v_{\rm ex} \propto \frac{1}{m_\phi^2}\sqrt{\frac{\delta}{m_\chi}}\ \text{off-resonance}, \quad \frac{1}{m_\chi\delta}\sqrt{\frac{\delta}{m_\chi}}\ \text{near-resonance} \tag{3.64}$$

$$\sigma_{\rm ex\to ex} \propto \frac{1}{m_\phi^2}\ \text{off-resonance}, \quad \frac{1}{m_\chi\delta}\ \text{near-resonance}. \tag{3.65}$$

We note that for slow-moving particles initially in the excited state, inelastic downscattering will generally dominate over elastic scattering (due to the $1/v$ scaling). The constant σv for downscattering implies that the argument given in [273] (which predicts a constant density core in dwarf galaxies as a direct result of a constant σv for exothermic interactions) holds even at low velocities where the perturbative approach used in that work is not valid. However, the scaling of the constant σv with the parameters of the model is quite different to the perturbative case. Note in particular that in regions of parameter space close to a resonance, large scattering cross sections can be achieved even for large m_ϕ (provided $m_\phi \lesssim \alpha m_\chi$ so our approximation holds), depending only on the mass splitting and the DM mass rather than the mediator mass.

The cross section for elastic scattering in the ground state does not have a simple behavior close to threshold, since there is nothing special about $\epsilon_v \approx \epsilon_\delta$ from the perspective of the ground state. Setting $\epsilon_\Delta = 0$ we obtain:

$$\Gamma_v \to \Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right)\Gamma^2\left(\frac{i\epsilon_v}{2\mu} + \frac{1}{2}\right), \tag{3.66}$$

$$\sigma_{\rm gr\to gr} = \frac{\pi}{\epsilon_v^2}\left|1 + \left(\frac{V_0}{4\mu^2}\right)^{-\frac{2i\epsilon_v}{\mu}}\left(\frac{\Gamma_v}{\Gamma_v^*}\right)\left[\frac{\sinh\left(\frac{\pi\epsilon_v}{2\mu} + i\varphi\right)}{\sinh\left(\frac{\pi\epsilon_v}{2\mu} - i\varphi\right)}\right]\right|^2. \tag{3.67}$$

The term involving $\sinh$'s approaches 1 when $\varphi \to n\pi$ and -1 when $\varphi \to (n+1)\pi/2$, which gives rise to the characteristic resonances and anti-resonances in the low-v limit. More explicitly, we can

perform a Taylor expansion in the low-velocity limit $\epsilon_v \ll \mu$ (here having already set $\epsilon_v = \epsilon_\delta$), obtaining:

$$\sigma_{\text{gr}\to\text{gr}} \approx \frac{\pi}{\mu^2}\left(\pi\cot(\varphi) + 2\ln\left(\frac{V_0}{\mu^2}\right) + 4\gamma\right)^2 . \tag{3.68}$$

We see that this cross section has the same form as the other low-velocity and low-mass-splitting limits we have studied; it is identical to the expression obtained by first taking $\epsilon_v \to 0$ and then $\epsilon_\delta \to 0$.

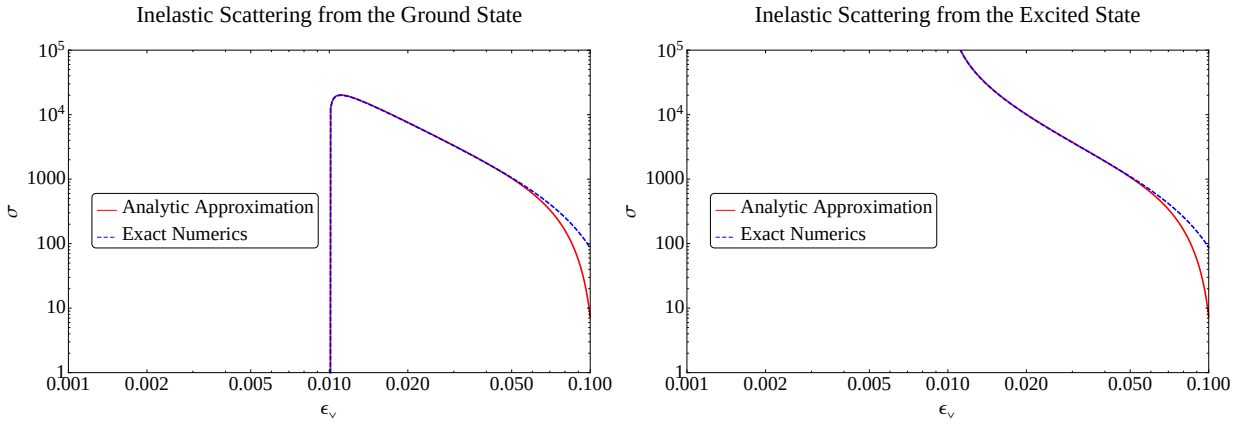

Figure 3.10: A comparison of the numerical vs analytic results for s-wave inelastic scattering, for upscattering (left) and downscattering (right), with ϵ_δ = 0.01 and ϵ_ϕ = 0.04. The two diverge for $\epsilon_v \gtrsim \epsilon_\phi$, as expected.

Higher Partial Waves

Figure 3.10 shows a particular slice through the parameter space of Figure 3.3, for inelastic scattering, but extending to $\epsilon_v > \epsilon_\phi$. As expected, our analytic approximation breaks down in this regime (note that μ and ϵ_ϕ are generally equal up to a $\mathcal{O}(1)$ factor). Conveniently, $\epsilon_v \lesssim \epsilon_\phi$ is also precisely the condition for s-wave scattering to dominate over the higher partial waves. Consequently, while a more careful treatment of the matching between the WKB and large-r regimes (see Figure 3.1) might allow extension of our approximation for the s-wave to the region with $\epsilon_v \gtrsim \epsilon_\phi$, at that point it would be necessary to include the higher partial waves as well.

This can be easily seen by comparing the relevant length scales: for the ℓth partial wave, the vacuum solution is proportional to the Bessel function $j_\ell(\epsilon_v r)$, which peaks when $\ell \sim \epsilon_v r$, i.e. $r \sim \ell/\epsilon_v$. In order for scattering of the ℓth partial wave to be significant, this peak must lie within the range of the potential, i.e. $r \lesssim 1/\epsilon_\phi$, and so we must have $\ell \lesssim \epsilon_v/\epsilon_\phi$. If $\epsilon_v/\epsilon_\phi \lesssim 1$, then only the s-wave term can penetrate the potential far enough to experience significant scattering.

In the case where δ is non-zero, this argument still holds — for particles in the excited state, the asymptotic wave function is now $j_\ell(\sqrt{\epsilon_v^2 - \epsilon_\delta^2})$, but since $\sqrt{\epsilon_v^2 - \epsilon_\delta^2} < \epsilon_v$, requiring $\epsilon_v \ll \epsilon_\phi$

is certainly *sufficient* to ensure that the potential cuts off at smaller r than the peak of the higher-ℓ wavefunctions. Since whenever particles in the excited state are present and the scattering rate is significant, their downscatterings will populate the ground state, we will generally consider $\epsilon_v \lesssim \epsilon_\phi$ to be both a necessary and sufficient condition for our approximate solution to be useful.

The Adiabatic Regime

There is a regime where $\epsilon_v \gtrsim \epsilon_\phi$ but our approximate solution remains valid, although the s-wave does not generally dominate scattering in this part of parameter space, and so we caution that our s-wave result should not be used as a proxy for the total scattering cross section. However, it can be used as a lower bound.

As discussed above, our method essentially neglects repulsive scattering at small distances, which is valid when the range of the potential is relatively short and so the scattering wavefunction for the repulsed eigenstate is peaked outside its range (note this is the same reason we can ignore the higher partial waves in this regime).

There is another regime where this approximation is valid, for a different reason. Suppose the system starts with both DM particles in the ground state (i.e. the state of lowest energy). At short distances, the lowest-energy eigenstate is the one that experiences an attractive potential (corresponding to the $+-$ two-body state at high energies). If the transition from long distances to short distances is adiabatic — i.e. this transition occurs slowly relative to the scale associated with the splitting between the eigenstates — then particles in the lowest-energy eigenstate at long distances will find themselves entirely in the attracted eigenstate at short distances, in analogy to the adiabatic theorem, and ignoring the repulsed eigenstate will be valid because it will simply never be populated.

The splitting between the eigenstates corresponds to an energy scale of ϵ_δ^2 in our dimensionless coordinates, and hence to a time scale of $1/\epsilon_\delta^2$; the corresponding distance scale, for an inward-moving wavepacket, would be $\sim \epsilon_v/\epsilon_\delta^2$. The rotation of the eigenstates (from mass eigenstates to gauge eigenstates of the high-energy potential), as described in Section 3.2, occurs when $V(r) = e^{-\epsilon_\phi r}/r$ becomes comparable to $\epsilon_\delta^2/2$. If the cause of this transition is the exponential cutoff, i.e. $r \sim 1/\epsilon_\phi$, then the transition occurs over a radius $\Delta r \sim 1/\epsilon_\phi$; if $\epsilon_\phi \lesssim \epsilon_\delta^2$ then it occurs when $V(r) \sim 1/r$ and over a range $\Delta r \sim 1/\epsilon_\delta^2$.

So the criterion for adiabaticity is $\epsilon_v/\epsilon_\delta^2 \lesssim 1/\epsilon_\phi$ if $\epsilon_\phi \gtrsim \epsilon_\delta^2$, or $\epsilon_v/\epsilon_\delta^2 \lesssim 1/\epsilon_\delta^2$ otherwise. In the first case, the transition is adiabatic for $\epsilon_v\epsilon_\phi \lesssim \epsilon_\delta^2$; in the second case adiabaticity always holds for $\epsilon_v \lesssim 1$ (however, note that $\epsilon_\phi \lesssim \epsilon_\delta^2$ is a regime where the approximations we use are known to be less accurate [289]), which by the condition $\epsilon_\phi \lesssim \epsilon_\delta^2$ implies $\epsilon_\phi\epsilon_v \lesssim \epsilon_\delta^2$. We can summarize this by saying adiabaticity holds if and only if $\epsilon_\phi\epsilon_v \lesssim \epsilon_\delta^2$, provided our other assumptions hold (that is, $\epsilon_v, \epsilon_\delta, \epsilon_\phi \lesssim 1$).

This mechanism is also responsible for the enhancement in the annihilation rate noted in [289] for the case with a mass splitting, compared to the case where the mass splitting is negligible relative to the kinetic energy and can be ignored; under adiabatic conditions, the presence of the mass splitting causes particles initially in the ground state to transition into a purely attracted state, rather than a equal linear combination of attracted and repulsed states.

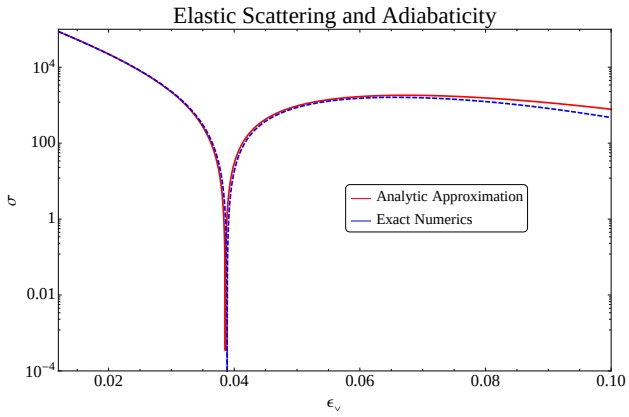

Figure 3.11: A scan through ϵ_v with $\epsilon_\phi = 0.03$ and $\epsilon_\delta = 0.04$. This demonstrates the shift from the transition to small-r being adiabatic vs. nonadiabatic. We can see the breakdown near $\epsilon_\delta^2 \sim \epsilon_v \epsilon_\phi$, which happens near $\epsilon_v \sim 0.06$.

This argument cannot be applied to scattering from the excited state or into the excited state, as if the excited state is populated then this implies the repulsed eigenstate will also be populated and cannot be ignored. But provided we are only interested in elastic scattering from the ground state (i.e. for the below-threshold case $\epsilon_v \lesssim \epsilon_\delta$), we expect our results to be accurate (for the s-wave) even when $\epsilon_v \gtrsim \epsilon_\phi$, in the event that $\epsilon_\phi \epsilon_v \lesssim \epsilon_\delta^2$.

The Perturbative Regime

The Born approximation can be straightforwardly applied to this multi-state system, by writing the three-dimensional Schrödinger equation in the form:

$$\Psi(\mathbf{r}) = \Psi_0(\mathbf{r}) - \frac{m_\chi}{4\pi} \int \frac{d^3\mathbf{r_0}}{|\mathbf{r}-\mathbf{r_0}|} \begin{pmatrix} e^{ik|\mathbf{r}-\mathbf{r_0}|} & 0 \\ 0 & e^{ik'|\mathbf{r}-\mathbf{r_0}|} \end{pmatrix} \cdot \bar{V}(\mathbf{r_0}) \cdot \Psi(\mathbf{r_0}). \tag{3.69}$$

Here $\Psi(\mathbf{r}) = \begin{pmatrix} \Psi_1(\mathbf{r_0}) \\ \Psi_2(\mathbf{r_0}) \end{pmatrix}$ is the full, three-dimensional physical wavefunction for the two-body state; for the spherically symmetric potential and s-wave scattering we consider, it is related to the wavefunction $\psi(r)$ by $\psi(r) = r\Psi(\mathbf{r})$. $\Psi_0(\mathbf{r})$ is the equivalent wavefunction for the unperturbed system with no potential, satisfying $\nabla^2\Psi_0(\mathbf{r}) = -\begin{pmatrix} k^2 & 0 \\ 0 & k'^2 \end{pmatrix} \cdot \Psi_0(\mathbf{r})$. Note, since $\Psi(\mathbf{r})$ describes a two-body state, we have replaced the mass in the Schrödinger equation with the reduced mass $m_\chi/2$.) $\bar{V}(\mathbf{r})$ is the matrix potential given in Eq. 3.4, but now with the diagonal $2\delta c^2$ term omitted (we instead incorporate the effect of this term in the difference between k' and k). Here we use units where $c = \hbar = 1$.

The first-order Born contribution to the scattering amplitude is solely inelastic, since the matrix $\bar{V}$ is purely off-diagonal. For an initial state consisting of a plane wave in a single state (purely for

notational purposes, we here choose the upper row of the matrix to correspond to the initial state), we obtain:

$$\begin{aligned}\Psi(\mathbf{r}) &= \begin{pmatrix} Ae^{ikz} \\ 0 \end{pmatrix} - \frac{m_\chi}{4\pi}\int \frac{d^3\mathbf{r_0}}{|\mathbf{r}-\mathbf{r_0}|}\begin{pmatrix} e^{ik|\mathbf{r}-\mathbf{r_0}|} & 0 \\ 0 & e^{ik'|\mathbf{r}-\mathbf{r_0}|}\end{pmatrix}\cdot\bar{V}(\mathbf{r_0})\cdot\begin{pmatrix} Ae^{ikz_0} \\ 0\end{pmatrix} \\ &= \begin{pmatrix} Ae^{ikz} \\ 0 \end{pmatrix} - \frac{m_\chi}{4\pi}\int \frac{d^3\mathbf{r_0}}{|\mathbf{r}-\mathbf{r_0}|}e^{ik'|\mathbf{r}-\mathbf{r_0}|}\left(-\frac{\alpha}{r}e^{-m_\phi r}\right)\begin{pmatrix} 0 \\ Ae^{ikz_0}\end{pmatrix}. \end{aligned} \tag{3.70}$$

Writing $\mathbf{k} = k\hat{r}$, $\mathbf{k}' = k'\hat{r}$, $\mathbf{p} = k\hat{z}$, $\mathbf{p}' = k'\hat{z}$, and making the long-distance approximation $e^{ik|\mathbf{r}-\mathbf{r_0}|}/|\mathbf{r}-\mathbf{r_0}| \approx (e^{ikr}/r)e^{-i\mathbf{k}\cdot\mathbf{r_0}}$, we obtain:

$$\begin{aligned}\Psi(\mathbf{r}) &\approx A\left[\begin{pmatrix} e^{ikz} \\ 0\end{pmatrix} + \frac{\alpha m_\chi}{4\pi}\frac{e^{ik'r}}{r}\begin{pmatrix} 0 \\ 1\end{pmatrix}\int d^3\mathbf{r_0}\left(\frac{1}{r}e^{-m_\phi r}\right)e^{i(\mathbf{p}-\mathbf{k}')\cdot\mathbf{r_0}}\right] \\ &= A\left[\begin{pmatrix} e^{ikz} \\ 0\end{pmatrix} + \frac{\alpha m_\chi}{4\pi}\frac{e^{ik'r}}{r}\begin{pmatrix} 0 \\ 1\end{pmatrix}\frac{4\pi}{|\mathbf{p}-\mathbf{k}'|^2+m_\phi^2}\right]. \end{aligned} \tag{3.71}$$

We can read off the scattering amplitudes as

$$f_X = 0, \quad f_Y = \frac{\alpha m_\chi}{|\mathbf{p}-\mathbf{k}'|^2+m_\phi^2}. \tag{3.72}$$

Now $|\mathbf{p}-\mathbf{k}'|^2 = k^2 + (k')^2 - 2kk'\cos\theta$, where θ describes the angle between $\hat{r}$ and $\hat{z}$, and we have:

$$\frac{d\sigma^{\text{Born}}_{\text{inelastic}}}{d\Omega} = \frac{k'}{k}f_Y^2 = \frac{k'}{k}\frac{\alpha^2 m_\chi^2}{\left(k^2+(k')^2-2kk'\cos\theta+m_\phi^2\right)^2}. \tag{3.73}$$

Performing the angular integral yields:

$$\sigma^{\text{Born}}_{\text{inelastic}} = \frac{k'}{k}\frac{4\pi\alpha^2 m_\chi^2}{\left(m_\phi^2+(k')^2-k^2\right)^2+4k^2m_\phi^2}. \tag{3.74}$$

This result applies to both particles initially in the ground or excited states, with the proper choices of k and k'. For upscattering, $k = m_\chi v$ and $k' = \sqrt{m_\chi^2 v^2 - 2\delta m_\chi}$, where v is the speed of a single particle in the center-of-mass frame in the initial state. If we use the convention in the body of the text where v always refers to the speed of the particle in the ground state, then for downscattering these choices of k and k' are simply reversed. If instead we take v to be the speed of the particle in the initial state, for downscattering we have $k = m_\chi v$ and $k' = \sqrt{m_\chi^2 v^2 + 2\delta m_\chi}$. Thus we obtain the upscattering cross section,

$$\sigma^{\text{Born}}_{\text{gr}\to\text{ex}} = \frac{4\pi\alpha^2 m_\chi^2\sqrt{1-\frac{2\delta}{m_\chi v^2}}}{m_\phi^4\left[\left(1-2\delta m_\chi/m_\phi^2\right)^2+4m_\chi^2v^2/m_\phi^2\right]} \tag{3.75}$$

and the downscattering cross section (taking v to be the speed of the ground-state particle):

$$\sigma^{\text{Born}}_{\text{ex}\to\text{gr}} = \frac{4\pi\alpha^2 m_\chi^2}{\sqrt{1-\frac{2\delta}{m_\chi v^2}}\, m_\phi^4 \left[\left(1-2\delta m_\chi/m_\phi^2\right)^2 + 4m_\chi^2 v^2/m_\phi^2\right]}. \tag{3.76}$$

The corresponding dimensionless cross sections are respectively:

$$\sigma^{\text{Born}}_{\text{gr}\to\text{ex}} = \frac{\epsilon_\Delta}{\epsilon_v}\frac{4\pi}{\epsilon_\phi^4(1-\epsilon_\delta^2/\epsilon_\phi^2)^2 + 4\epsilon_v^2/\epsilon_\phi^2}, \quad \sigma^{\text{Born}}_{\text{ex}\to\text{gr}} = \frac{\epsilon_v}{\epsilon_\Delta}\frac{4\pi}{\epsilon_\phi^4(1-\epsilon_\delta^2/\epsilon_\phi^2)^2 + 4\epsilon_v^2/\epsilon_\phi^2}. \tag{3.77}$$

The lowest-order contribution to elastic scattering comes via the second term in the Born series. For the initial condition above, the contribution to the wavefunction is:

$$\begin{aligned}
\Psi_{\text{2nd-order}}(\mathbf{r}) &= \left(\frac{\alpha m_\chi}{4\pi}\right)^2 \int \frac{d^3\mathbf{r_0}}{|\mathbf{r}-\mathbf{r_0}|}\begin{pmatrix} e^{ik|\mathbf{r}-\mathbf{r_0}|} & 0 \\ 0 & e^{ik'|\mathbf{r}-\mathbf{r_0}|}\end{pmatrix}\cdot \bar{V}(\mathbf{r_0}) \\
&\times \int \frac{d^3\mathbf{r}_1}{|\mathbf{r_0}-\mathbf{r_1}|}\begin{pmatrix} e^{ik|\mathbf{r_0}-\mathbf{r_1}|} & 0 \\ 0 & e^{ik'|\mathbf{r_0}-\mathbf{r_1}|}\end{pmatrix}\cdot \bar{V}(\mathbf{r_1})\cdot \Psi_0(\mathbf{r_1}) \\
&= A\left(\frac{\alpha m_\chi}{4\pi}\right)^2 \begin{pmatrix}1\\0\end{pmatrix} \int \frac{d^3\mathbf{r_0}}{|\mathbf{r}-\mathbf{r_0}|} e^{ik|\mathbf{r}-\mathbf{r_0}|}\frac{e^{-m_\phi r_0}}{r_0}\int \frac{d^3\mathbf{r}_1}{|\mathbf{r_0}-\mathbf{r_1}|} e^{ik'|\mathbf{r_0}-\mathbf{r_1}|}\frac{e^{-m_\phi r_1}}{r_1}e^{ikz_1}.
\end{aligned} \tag{3.78}$$

Again making the long-distance approximation we can read off the scattering amplitudes (at this order) as:

$$f_X = \left(\frac{\alpha m_\chi}{4\pi}\right)^2 \int d^3\mathbf{r_0} e^{-i\mathbf{k}\cdot\mathbf{r_0}}\frac{e^{-m_\phi r_0}}{r_0}\int \frac{d^3\mathbf{r}_1}{|\mathbf{r_0}-\mathbf{r_1}|} e^{ik'|\mathbf{r_0}-\mathbf{r_1}|}\frac{e^{-m_\phi r_1}}{r_1}e^{ikz_1}, \quad f_Y = 0. \tag{3.79}$$

In the limit where the initial particle is very slow-moving, so we can take $k = 0$ (but not $k' = 0$), this amplitude has a simple analytic form: for the integral over $\mathbf{r_1}$, we can choose coordinates where $\mathbf{r_0}$ lies along the z-axis to simplify the angular integral, obtaining:

$$\int \frac{d^3\mathbf{r}_1}{|\mathbf{r_0}-\mathbf{r_1}|} e^{ik'|\mathbf{r_0}-\mathbf{r_1}|}\frac{e^{-m_\phi r_1}}{r_1} = \frac{4\pi}{m_\phi^2+(k')^2}\frac{e^{ik'r_0}-e^{-m_\phi r_0}}{r_0}. \tag{3.80}$$

The amplitude in the $k = 0$ limit then becomes,

$$f_X = \left(\frac{\alpha m_\chi}{4\pi}\right)^2 \frac{(4\pi)^2}{m_\phi^2+(k')^2}\left(\frac{1}{m_\phi - ik'} - \frac{1}{2m_\phi}\right). \tag{3.81}$$

For scattering where the initial particles are in the excited state, $k' = \sqrt{2\delta m_\chi}$ for $k = 0$, and the amplitude is given by:

$$\frac{d\sigma^{\text{Born}}_{\text{ex}\to\text{ex}}}{d\Omega} = |f_X|^2 = \frac{\alpha^4}{4m_\phi^2}\left(\frac{m_\chi^2}{m_\phi^2+2\delta m_\chi}\right)^2. \tag{3.82}$$

For scattering where the initial particles are in the ground state and $k = 0$, $k' = i\sqrt{2\delta m_\chi}$, and we obtain:

$$\frac{d\sigma^{\text{Born}}_{\text{gr}\to\text{gr}}}{d\Omega} = |f_X|^2 = \frac{\alpha^4}{4m_\phi^2}\left(\frac{m_\chi^2}{m_\phi^2 - 2\delta m_\chi}\right)^2\left(\frac{m_\phi - \sqrt{2\delta m_\chi}}{m_\phi + \sqrt{2\delta m_\chi}}\right)^2 = \frac{\alpha^4}{4m_\phi^2}\frac{m_\chi^4}{\left(m_\phi + \sqrt{2\delta m_\chi}\right)^4}. \tag{3.83}$$

Thus the cross sections for ground-state elastic scattering and excited-state elastic scattering, in the limit of slow-moving initial particles, are respectively,

$$\sigma^{\text{Born}}_{\text{gr}\to\text{gr}} = \frac{\pi\alpha^4 m_\chi^4}{m_\phi^2(m_\phi + \sqrt{2\delta m_\chi})^4}, \quad \sigma^{\text{Born}}_{\text{ex}\to\text{ex}} = \frac{\pi\alpha^4 m_\chi^4}{m_\phi^2(m_\phi^2 + 2\delta m_\chi)^2}. \tag{3.84}$$

The corresponding dimensionless cross sections are:

$$\sigma^{\text{Born}}_{\text{gr}\to\text{gr}} = \frac{\pi}{\epsilon_\phi^6}\frac{1}{\left(1 + \epsilon_\delta/\epsilon_\phi\right)^4}, \quad \sigma^{\text{Born}}_{\text{ex}\to\text{ex}} = \frac{\pi}{\epsilon_\phi^6}\frac{1}{\left(1 + \epsilon_\delta^2/\epsilon_\phi^2\right)^2}. \tag{3.85}$$

Note that these second-order cross sections become comparable to (or larger than) the first-order inelastic scattering cross sections for $\epsilon_\phi \lesssim 1$, signaling the breakdown of the Born approximation and the need to transition to the resonant regime described in the previous Subsections.

Summary of the Calculation

In the calculations of the previous subsections, we have obtained the following dimensionless scattering cross sections:

$$\sigma_{\text{gr}\to\text{gr}} = \frac{\pi}{\epsilon_v^2}\left|1 + \left(\frac{V_0}{4\mu^2}\right)^{-\frac{2i\epsilon_v}{\mu}}\left(\frac{\Gamma_v}{\Gamma_v^*}\right)\left[\frac{\cosh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_v-\epsilon_\Delta)}{2\mu} + i\varphi\right)}{\cosh\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu} - i\varphi\right)}\right]\right|^2 \tag{3.86}$$

$$\sigma_{\text{ex}\to\text{ex}} = \frac{\pi}{\epsilon_\Delta^2}\left|1 + \left(\frac{V_0}{4\mu^2}\right)^{-\frac{2i\epsilon_\Delta}{\mu}}\left(\frac{\Gamma_\Delta}{\Gamma_\Delta^*}\right)\left[\frac{\cosh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu} + i\varphi\right)}{\cosh\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\sinh\left(\frac{\pi(\epsilon_\Delta+\epsilon_v)}{2\mu} - i\varphi\right)}\right]\right|^2 \tag{3.87}$$

$$\sigma_{\text{gr}\to\text{ex}} = \frac{2\pi\cos^2\varphi\sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\sinh\left(\frac{\pi\epsilon_\Delta}{\mu}\right)}{\epsilon_v^2\cosh^2\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\left(\cosh\left(\frac{\pi(\epsilon_v+\epsilon_\Delta)}{\mu}\right) - \cos(2\varphi)\right)} \tag{3.88}$$

$$\sigma_{\text{ex}\to\text{gr}} = \frac{2\pi\cos^2\varphi\sinh\left(\frac{\pi\epsilon_v}{\mu}\right)\sinh\left(\frac{\pi\epsilon_\Delta}{\mu}\right)}{\epsilon_\Delta^2\cosh^2\left(\frac{\pi(\epsilon_\Delta-\epsilon_v)}{2\mu}\right)\left(\cosh\left(\frac{\pi(\epsilon_v+\epsilon_\Delta)}{\mu}\right) - \cos(2\varphi)\right)} \tag{3.89}$$

where we have defined $\epsilon_\Delta \equiv \sqrt{\epsilon_v^2 - \epsilon_\delta^2}$, and μ and V_0 are the defining parameters for the exponential potential $V_0 e^{-\mu r}$, given by:

$$\mu = \epsilon_\phi \left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{4}{\epsilon_\phi r_M}} \right), \quad V_0 = \frac{e^{\epsilon_\phi r_M \left(-\frac{1}{2} + \frac{1}{2}\sqrt{1+\frac{4}{\epsilon_\phi r_M}} \right)}}{r_M}, \tag{3.90}$$

with r_M chosen so that $\frac{e^{-\epsilon_\phi r_M}}{r_M} = \max\left(\frac{\epsilon_\delta^2}{2}, \epsilon_\phi^2\right)$. The terms Γ_v and Γ_Δ come from matching the WKB wavefunction onto the wavefunction for the exponential potential, and are defined by,

$$\begin{aligned} \Gamma_v &\equiv \Gamma\left(1 + \frac{i\epsilon_v}{\mu}\right) \Gamma\left(\frac{i\epsilon_v - i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right) \Gamma\left(\frac{i\epsilon_v + i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right) \\ \Gamma_\Delta &\equiv \Gamma\left(1 + \frac{i\epsilon_\Delta}{\mu}\right) \Gamma\left(\frac{i\epsilon_\Delta - i\epsilon_v}{2\mu} + \frac{1}{2}\right) \Gamma\left(\frac{i\epsilon_v + i\epsilon_\Delta}{2\mu} + \frac{1}{2}\right), \end{aligned} \tag{3.91}$$

where Γ denotes the gamma function. Finally, φ is a phase that comes from extending the WKB solution to the matching region. Its full form is rather complicated, but where $\epsilon_\delta \lesssim \epsilon_\phi$ and our other approximations hold (i.e. $\epsilon_\phi \lesssim 1$, $\epsilon_v \lesssim \mu$), the phase φ can be accurately approximated by $\varphi = \sqrt{2\pi/\epsilon_\phi}$ (as noted in [289]). More generally, it is given by:

$$i\varphi \equiv \int_0^{r_M} \sqrt{\lambda_-} dr' + \int_{r_M}^{r_s} \sqrt{\tilde{\lambda}_-} dr' + \frac{2i\sqrt{V_0}}{\mu} e^{-\mu r_s/2}. \tag{3.92}$$

Here r_s must be chosen such that $V_0\, e^{-\mu r_s} \gg \mu^2, \epsilon_v^2, \epsilon_\delta^2$, but formally should be taken to $-\infty$.

We emphasize again that these cross sections are only good approximations in the regimes described in the lower half of Table I. Since we are only computing the s-wave piece of the scattering amplitude, which is angle-independent, the viscosity and transfer cross sections are trivially related to σ (see [149] for a discussion of the different cross sections and their relevance for the problem at hand). We have verified numerically that our approximations are accurate and we have explored various limits and regimes, finding that they make physical sense and match on to other scenarios, including the degenerate limit and the Born limit.

The astute reader will also notice the following salient feature of the scattering cross sections: the elastic and inelastic cross sections are the same whether the system starts in the ground state or the excited state, modulo a swap of ϵ_v with ϵ_Δ (assuming that ϵ_Δ is the same in both cases, which requires the system to be above threshold.) The scattering *amplitudes* are identical. This reflects the identical interactions of the ground and excited states: swapping $\epsilon_v \leftrightarrow \epsilon_\Delta$ simply corresponds to relabeling the states. The result also agrees with our intuition from quantum mechanical scattering off a 1D step potential: the transmission and reflection probabilities are the same for "downhill" and "uphill" scattering when the particle's energy is greater than the potential barrier, and the same is true in this system above the mass-splitting threshold.

3.4 Expected Effect on DM Halos

Parameters of Interest for Modification of DM Halos

On dwarf galaxy scales, an elastic scattering cross section of roughly $\sigma/m_\chi \gtrsim 0.1$ cm^2/g is required in order for DM self-scattering to have a significant impact on the internal structure [266]. This corresponds to particles in the core interacting once on average over the age of the universe [273], and so is likely also a necessary condition for exothermic downscattering to be relevant. We will thus use this cross section as a benchmark.

As discussed in [273], requiring a significant relic population of particles in the excited state at late times (that was not depleted by scatterings in the early universe) requires m_χ at the MeV scale or lighter. However, the excited state might be populated non-thermally, in which case much heavier DM masses might also be viable.

For the non-degeneracy of the excited state to have a significant impact on scattering in dwarf galaxies, the mass splitting should be significant compared to the typical kinetic energy of the DM particles. Taking the typical velocity in dwarf galaxies to be 10 km/s$\sim 3\times10^{-5}c$ [309], this implies $\delta \gtrsim 10^{-9}m_\chi$ in order to see differences from purely elastic scattering. Our requirements that $\epsilon_\delta \lesssim 1$, $\epsilon_\phi \lesssim 1$ impose that $\delta \lesssim \alpha^2 m_\chi$ and $\alpha m_\phi \lesssim \alpha^2 m_\chi$[2]. So in order for our approximation to be valid and the mass splitting to be interestingly large, we will focus on the range $\alpha \gtrsim 10^{-4}$ (or higher for larger δ: $\alpha \gtrsim \sqrt{\delta/m_\chi}$), which will also guarantee $\epsilon_v = v/\alpha \lesssim 1$ as required. For a vector mediator, this is in agreement with broad expectations from the Standard Model if the coupling is not fine-tuned to be small.

In general, we will treat α as a free parameter within the range $10^{-4} \lesssim \alpha \lesssim 1$, since the constraints on it are rather model-dependent. In particular, we do *not* impose the constraint that annihilations of the DM to the force carrier should generate the correct relic density. The self-interacting DM could be non-thermal or a sub-dominant component of the total DM if α is higher than the value expected for a thermal relic, or annihilation channels not involved in the self-scattering could prevent the overclosure of the universe if α is too small. However, for calibration, [290] found typical values for α (yielding the correct relic density) of a few times 10^{-2} for TeV-scale DM in a model with the same potential as the one employed in this work, and the annihilation of the DM to the force carriers has a cross section scaling as α^2/m_χ^2 at freezeout, so lighter DM will generally imply a smaller value of α if the DM is indeed a thermal relic.

There are few model-independent constraints on the mediator mass m_ϕ; the coupling of the force carrier to Standard Model particles is independent of its role here of mediating DM scattering. For significant scattering we require that $m_\phi \lesssim \alpha m_\chi$, and for s-wave scattering to dominate and our approximation to be valid we will generally require that $\epsilon_v \lesssim \epsilon_\phi$ in dwarf galaxies, i.e. $m_\phi \gtrsim m_\chi v \sim 3\times10^{-5}m_\chi$.

When we consider the exothermic scenario with a significant population of particles initially in the excited state, their scatterings have $\epsilon_v \sim \epsilon_\delta$ in our notation, assuming the kinetic energy of

[2]Natural scales for the splitting can include $\alpha^2 m_\chi$ (analogous to the splitting between atomic energy levels), αm_ϕ or $\alpha^2 m_\phi$ (if the splitting is generated by loops of the mediator) [139]; if the mass splitting is generated by some higher-dimension operator as in [290], then its size depends on the heavy mass scale.

the excited-state particles (limited by the escape velocity of the dwarf) is small compared to the mass splitting. Thus for this scenario we will also require $\epsilon_\delta \lesssim \epsilon_\phi$, i.e. $2\delta/\alpha^2 m_\chi \lesssim m_\phi^2/\alpha^2 m_\chi^2 \Rightarrow m_\phi \gtrsim \sqrt{\delta m_\chi}$. The requirement that $m_\phi \lesssim \alpha m_\chi$ means that the higher α is above its lower bound of $\sqrt{\delta/m_\chi}$, the more valid parameter space there will be for m_ϕ (although raising m_ϕ above the upper bound simply means we should use the Born approximation.)

With the above reasoning as our guide, we performed several cuts through parameter space to estimate phenomenologically interesting regions in DM mass, mediator mass, mass-splitting, and coupling. We selected a few fiducial velocity scales which correspond to the virial velocities of dwarf galaxies, galaxies the size of the Milky Way (MW), and clusters. Practically speaking, the DM scattering rate within a halo is not determined by one fixed velocity, but rather some distribution of velocities as DM particles pass through different parts of the halo. In order to understand these dynamics in detail, one must perform numerical simulations which lie outside the scope of this work. However, the simple estimates presented here can serve as benchmarks for simulations of the dynamical behavior of inelastically scattering DM in an actual halo.

The results of our parameter sweeps for four different scattering scenarios (ground to ground, ground to excited, excited to excited, and excited to ground) can be found in Figures 3.12-3.15, respectively. All plots depict σ_T/m_χ in nine slices of the m_χ vs. m_ϕ plane with α and δ held fixed (note we label the cross section as the transfer cross section σ_T for comparison to the literature, but in our s-wave approximation it is identical to the cross section we have discussed so far). Each slice has different values of α and δ, with α decreasing from left to right and δ increasing from top to bottom (with each subplot labelled accordingly.) We also include three velocity scales: 10 km/s for dwarf galaxies, 200 km/s for MW-sized galaxies, and 1000 km/s for clusters. In terms of our model, these velocities correspond to ϵ_v for situations where the particles are initially in the ground state and ϵ_Δ for situations where the particles are initially in the excited state. Additional features of the plots are listed as follows:

- Black regions show where our approximation breaks down. Since we know that our approximations will give wrong or misleading cross sections in these regions, we cover them entirely. Black, triangular regions in the lower right corner show when $\epsilon_\phi > 1$. Black, horizontal strips across the bottom of the plots (when present) show when $\epsilon_\delta > 1$. Black, vertical strips across the left side of the plots (when present) show when $\epsilon_\delta^2/2 > \epsilon_\phi$. Black regions in the upper left show when higher partial waves and repulsed eigenstates can no longer be ignored, which happens when $\epsilon_v > \epsilon_\phi$. Note that we chose to black out the region with ϵ_v corresponding to dwarf scales, e.g. the blacked out regions have $\epsilon_\phi < 3 \times 10^{-5}/\alpha$ for scenarios where the particles are incoming in the ground state. Since this criterion is velocity-dependent, we only plot cross sections for DM in the MW and clusters when our approximations are valid at those velocity scales (rather than blacking out regions where the approximation is not valid for those scales.)

- Yellow regions show 0.1 cm^2/g $< \sigma_T/m_\chi <$ 1 cm^2/g (light), 1 cm^2/g $< \sigma_T/m_\chi <$ 10 cm^2/g (medium), and 10 cm^2/g $< \sigma_T/m_\chi <$ 100 cm^2/g (dark) on dwarf velocity scales with $v \sim 10$ km/s.

- Blue contours represent σ_T/m_χ = 0.1 cm^2/g (light) and σ_T/m_χ = 1 cm^2/g (dark) on MW velocity scales with $v \sim 200$ km/s. We emphasize that the contours are only plotted when $\epsilon_v < \epsilon_\phi$ on these scales, which ensures that the repulsed eigenstates and higher partial waves are safely neglected.

- Red contours represent σ_T/m_χ = 0.1 cm^2/g (light) and σ_T/m_χ = 1 cm^2/g (dark) on cluster velocity scales with $v \sim 1000$ km/s. We emphasize that the contours are only plotted when $\epsilon_v < \epsilon_\phi$ on these scales, which ensures that the repulsed eigenstates and higher partial waves are safely neglected. The strongest robust constraints on DM self-interaction come from observations of merging clusters, with the upper bound being around $\sigma_T/m_\chi \sim 1$ cm^2/g (e.g. [310]). Thus, parameter space to the lower left of the dark red contours is effectively ruled out for our model at the selected masses and couplings.

- Teal contours represent the Born approximation with σ_T/m_χ = 0.1 cm^2/g (light) and σ_T/m_χ = 1 cm^2/g (dark). The Born approximation for the elastic cases is shown in dot-dashed lines (we wish to emphasize that we have taken a low-velocity limit in computing the Born approximation, so the result does not actually depend on the velocity of the incoming particles; it is most likely to be valid for dwarfs where the virial velocities are very low).

  Meanwhile, for the inelastic cases, the Born approximation is shown in dotted lines for MW velocities and dashed lines for cluster velocities.

The results in Figures 3.12-3.15 show that in the parameter region where our approximation holds, parameters below the dark red line are definitively ruled out by cluster mergers while yellow regions are favored because they potentially solve the small-scale structure anomalies. Interestingly, tentative evidence from cluster mergers suggests that the DM self-interaction cross section may be nonzero, and the favored value of σ/m_χ is 0.8 cm^2/g at around 1σ [311]. Thus, the discovery of future mergers may further constrain self-interacting DM or even possibly pick out a strongly-preferred interaction cross section at that velocity scale.

The resonances occur in thin bands of roughly constant ϵ_ϕ from the lower left corner of the m_χ vs. m_ϕ plane up to the upper right corner. Decreasing α spaces the resonances further apart and broadens them due to the weak dependence of the ratio ϵ_δ/μ (which appears in our expressions for predicted resonance positions) on α. Increasing δ shifts the resonance positions slightly. For elastic scattering in the ground state, increasing δ elongates the resonance bands and makes them slightly more curved; the most pronounced curvature develops at small m_ϕ, as expected from the discussion in Section 3.3. Increasing δ has the opposite effect on elastic scattering in the excited state, in that it dampens the resonances (as discussed in Section 3.3) and so reduces their effect on the favored regions. Our approximations match smoothly onto the perturbative expressions as $\epsilon_\phi \to 1$, as expected.

We see that for these mass splittings, in the region of greatest astrophysical interest the ground-ground elastic scattering cross section is generally similar to the case where the DM states are degenerate [149]. Additionally, the favored regions are quite comparable for elastic scattering and downscattering, so in the context of this model it is possible for both effects to simultaneously

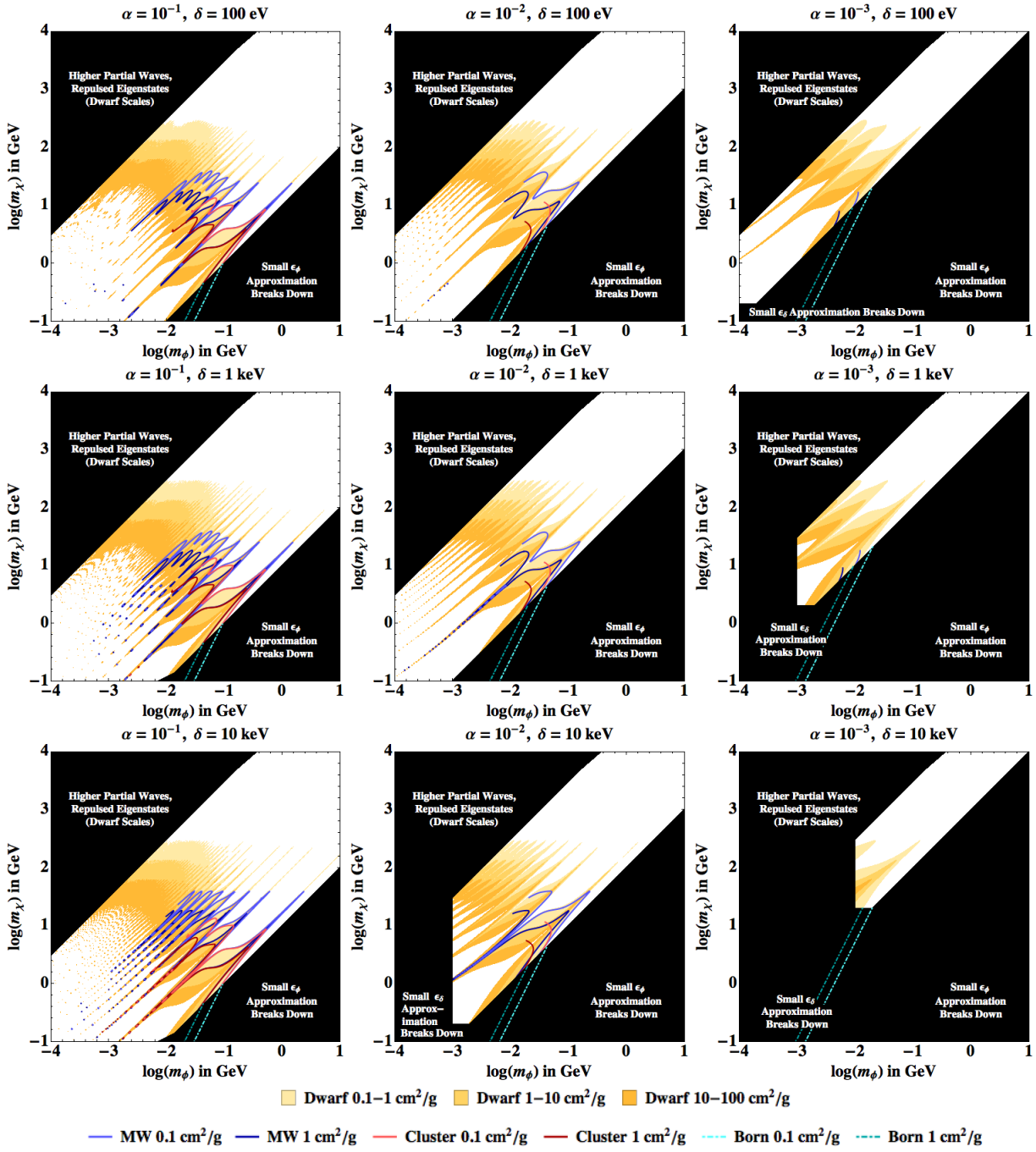

Figure 3.12: DM and mediator masses with ground→ground elastic scattering cross sections σ/m_χ in the 0.1-10 cm^2/g range, for couplings and mass splittings as labeled. Note that the use of "log" indicates the base 10 logarithm.

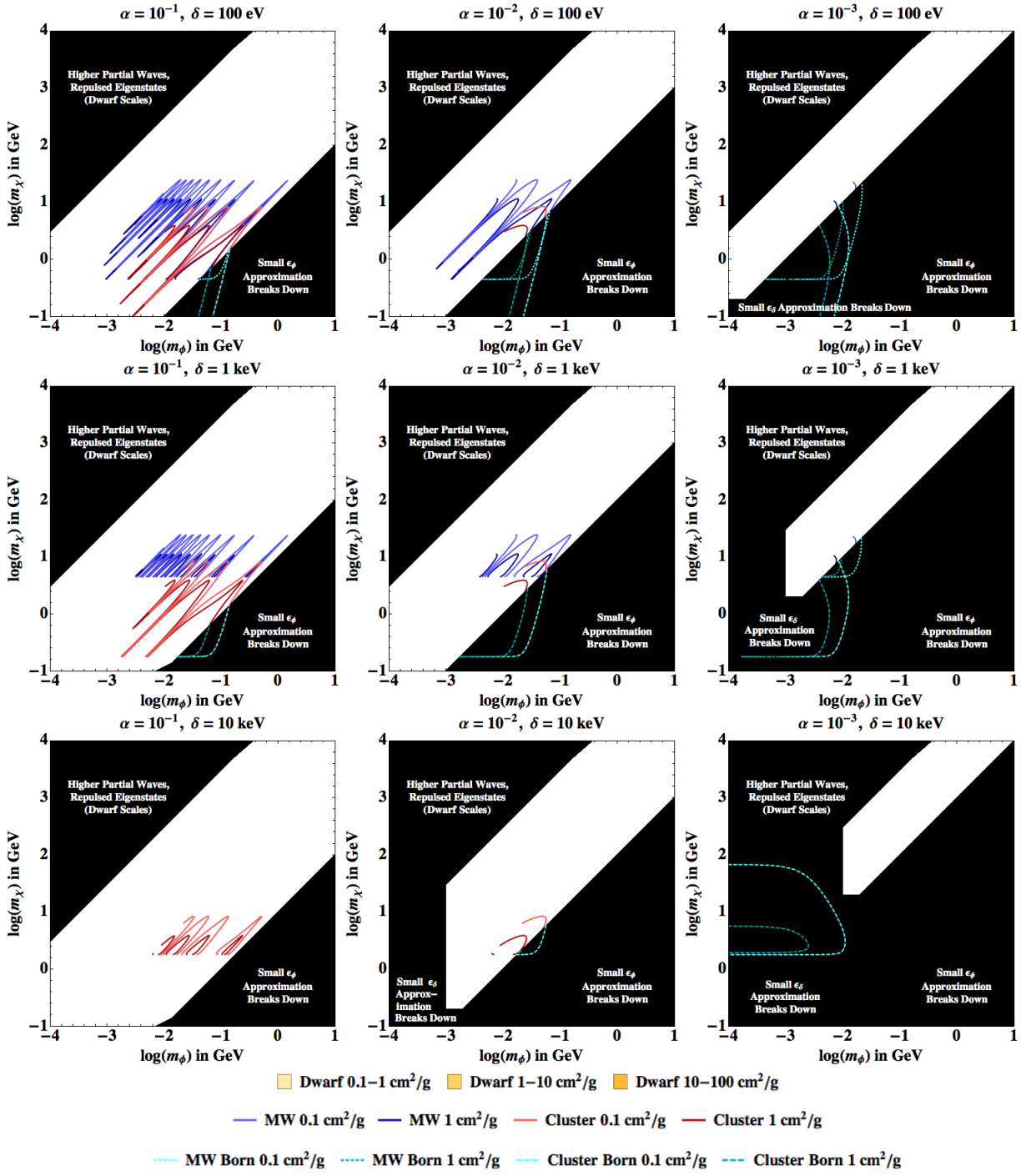

Figure 3.13: As Figure 3.12, except for inelastic scattering from the ground state to the excited state (upscattering). Upscattering in dwarf-sized halos, even where kinematically allowed, was never significant for the parameters we sampled (i.e. $\sigma/m_\chi \lesssim 0.01$ cm^2/g for all of the parameter space.) Note that the horizontal cutoffs of the contours come from the mass splitting threshold, as upscattering does not occur below threshold.

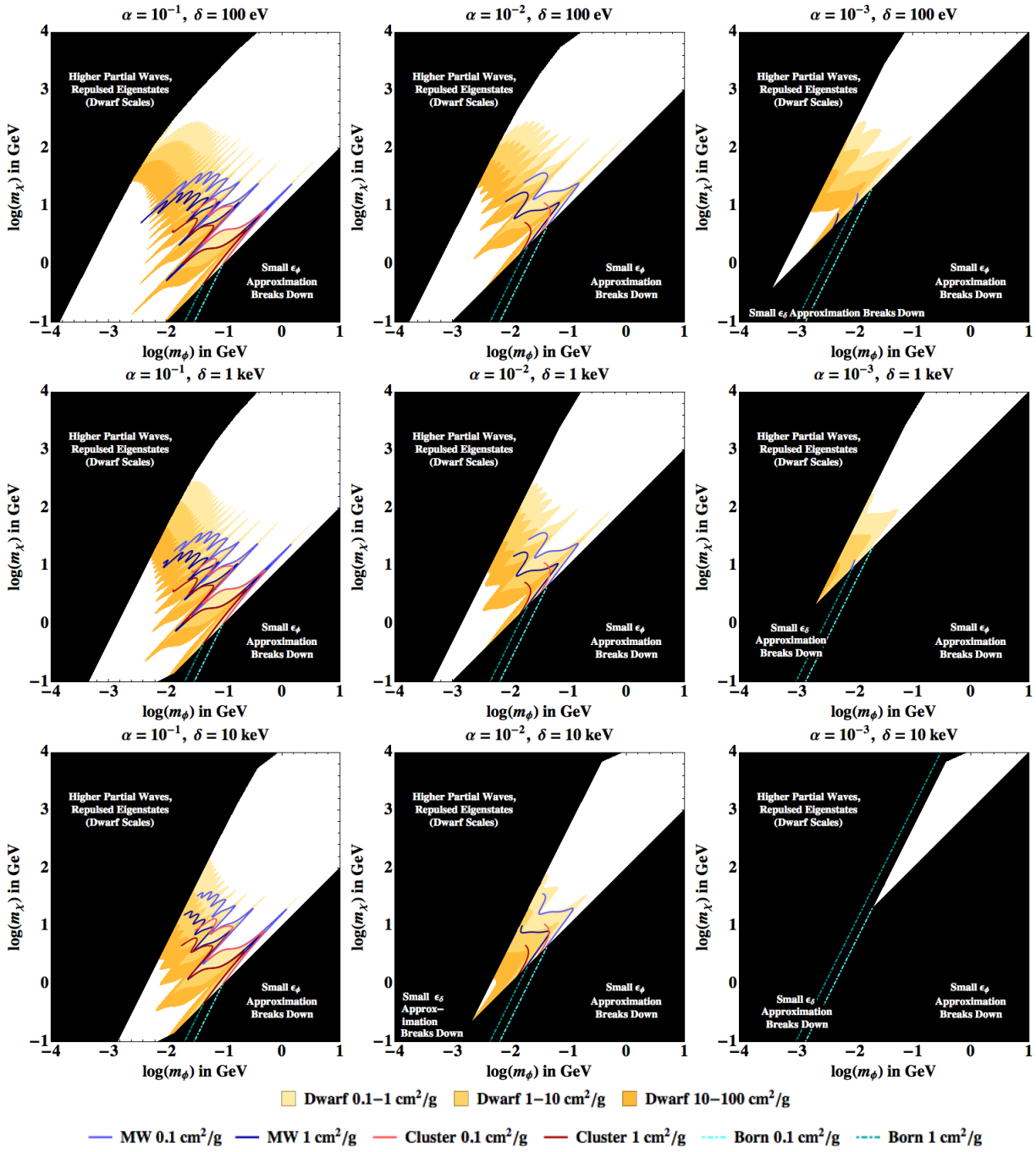

Figure 3.14: As Figure 3.12, except for elastic scattering from the excited state to the excited state.

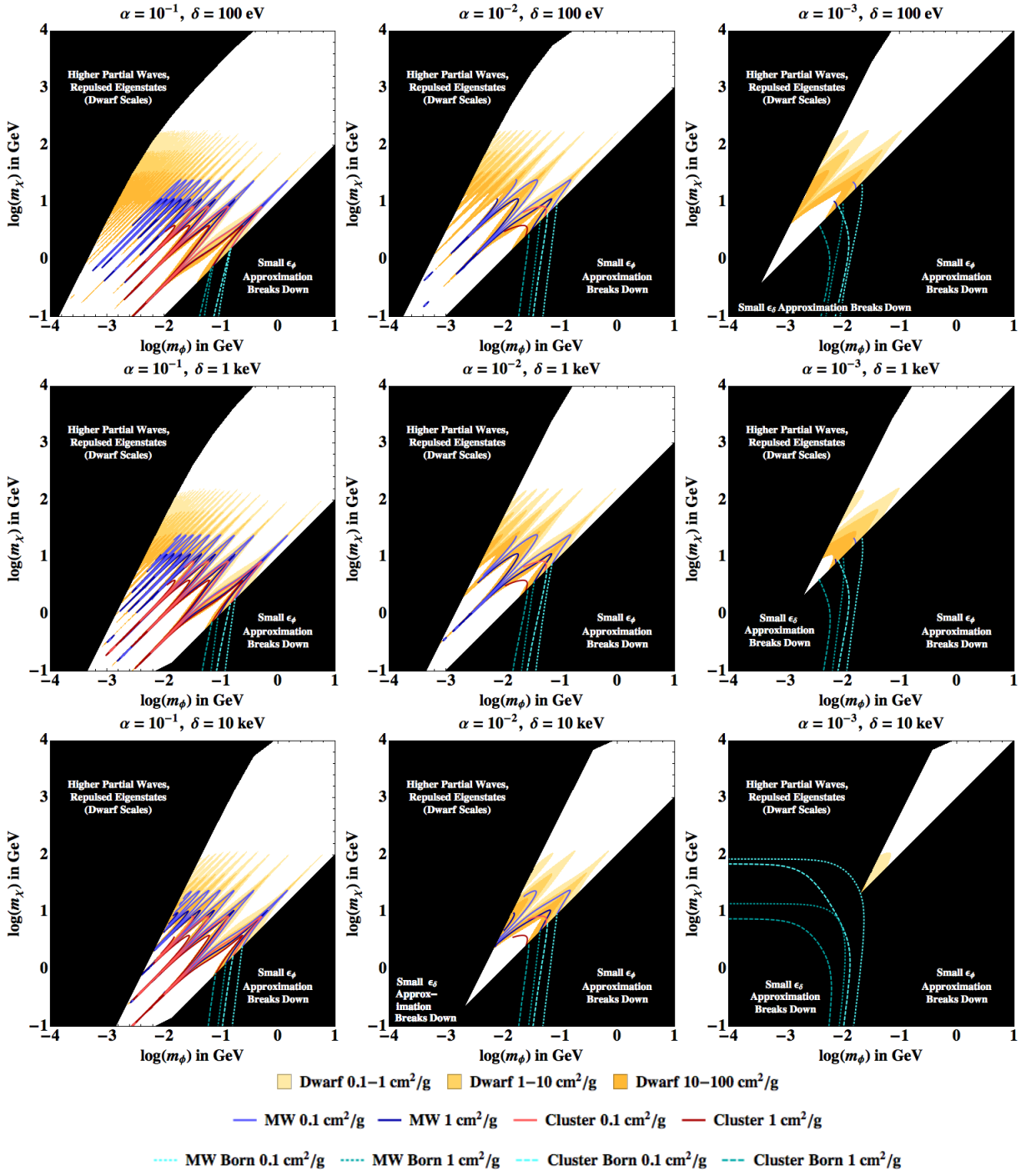

Figure 3.15: As Figure 3.12, except for inelastic scattering from the excited state to the ground state (downscattering). There are significant portions of parameter space where downscattering could be rapid and where the velocity "kick" to DM particles exceeds the halo escape velocity. Such downscattering could potentially dissipate halos and help solve the missing satellites problem.

contribute to the dynamics of DM within halos and help alleviate small-scale structure issues. This is a consequence of our choice of mass splittings fairly similar to the kinetic energy of virialized DM particles, since the (non-resonant) ratio of the scattering cross sections at low velocity scales as $\sqrt{\delta/(m_\chi v^2)}$ (see Eq. 3.64). However, for much higher mass splittings the scattering will most likely be in the classical high-velocity regime for most of the parameter space of interest, unless the mediator mass is also raised. However, as expected from our earlier discussion, at small m_ϕ the resonances occur for different parameters for downscattering, compared to elastic scattering in the ground state. Consequently, there are regions of parameter space where the ground-state elastic scattering cross section is large and the downscattering rate small, and vice versa.

Interestingly, for the case of upscattering there are no regions in the sampled parameter space where upscattering was significant in dwarf halos (i.e. there were no regions where upscattering exceeded 0.1 cm^2/g.) From the perspective of avoiding DM "cooling" in dwarf halos (which could potentially worsen the core-cusp problem, etc.) the ease of suppressing upscattering is an appealing feature. However, there are substantial regions of parameter space where upscattering is significant for the MW and for clusters. It is possible that upscattering could contribute to *steepening* of the density profile in the central parts of such large halos.

Regimes of Interest for XrayDM

There has been a great deal of recent interest in the detection of an apparent ~ 3.5 spectral keV line in radio observations of galaxies and galaxy clusters, as a potential signal from DM [112, 113]. However, there appears to be some tension between the interpretation of this line as originating from the decays of keV-scale DM, and its non-detection in the Virgo cluster. It has been proposed that the signal could instead originate from the upscattering of (weak-scale) DM to an excited state ~ 3.5 keV heavier than the ground state, followed by decay back to the ground state with emission of a photon [284]. This scenario was termed "XrayDM". Since the upscattering process involves two particles, the rate of excitations (and hence decays) scales with the density *squared* rather than the density (as would also be the case for annihilation, e.g. [312]), and also depends on the typical velocity of the DM particles: this can modify the relative strength of the signal in different regions[3].

The example model employed in the XrayDM scenario of [284] is *exactly* the simple model studied here: accordingly, we can now use our approximation to calculate which regions of parameter space can give rise to a sufficiently large cross section to explain the 3.5 keV line. Due to the large virial velocities of clusters, for small m_ϕ and/or large m_χ our approximation becomes invalid (as higher partial waves become important), but there is a significant region of interesting parameter space where the s-wave contribution dominates, as shown in Figure 3.16. Here we have

[3]Recent studies have also claimed tension with or exclusion of the interpretation of the signal as DM decay, based on samples of dwarfs and galaxies, e.g. [313, 314]; it is not clear how those exclusions generalize to cases with a different density/velocity dependence than decay, although one would expect a suppression of upscattering in dwarfs due to their lower virial velocities.

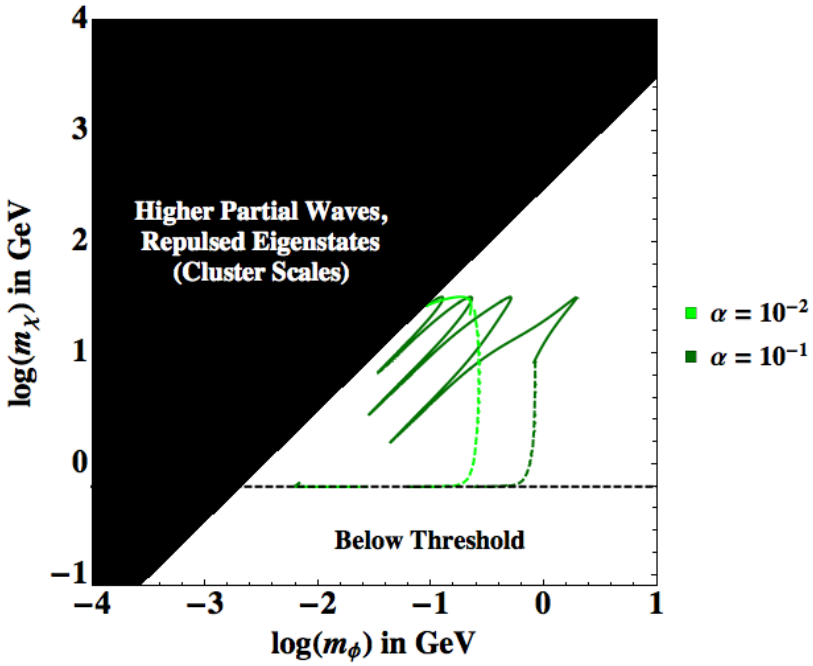

Figure 3.16: Contours satisfying (3.93) for the characteristic velocity scale of clusters, $v = 1000$ km/s. These lines serve as illustrative benchmarks – in reality, there will of course be some distribution of velocities in any cluster. The region where our approximation breaks down because $\epsilon_v > \epsilon_\phi$ is masked out, but it is likely that this part of parameter space can also furnish appropriate cross sections. The horizontal dashed line indicates the threshold for upscattering to be kinematically allowed, $\epsilon_v > \epsilon_\delta$.

imposed the criterion that:

$$\sigma_{\rm gr\to ex} = \left(\frac{10^{-19}\ {\rm cm}^3/{\rm s}}{1000\ {\rm km/s}}\right) \times \left(\frac{m_\chi}{10{\rm GeV}}\right)^2, \tag{3.93}$$

following [284], with $v = 1000$ km/s and $\delta = 3.5$ keV. As previously, we only show the result for a single velocity, rather than integrating over a distribution, since our purpose here is to provide estimates rather than a detailed analysis of the allowed parameter space.

Within the regime of validity of our approximation, we find that DM masses in the range of a few GeV to a few tens of GeV, moderately large values of α, and mediator masses in the range of 10 MeV – 1 GeV can naturally produce the required flux. We expect that similar DM masses and lower mediator masses will also provide viable explanations, but higher partial waves become important in these cases. We do not show results for smaller α as in that case, for interesting cross sections, the perturbative regime transitions directly to the classical regime (where high partial waves are important) without an intermediate resonant regime – this does not mean no parameter space is open for smaller α, just that it does not require the use of our results.

A similar analysis could be performed for the original "XDM" scenario [282]; we do not carry it out here because in that case the mass splitting is quite large (~ 1 MeV), meaning that only particles in the high-velocity tail of the velocity distribution (for DM in the Milky Way) are typically able to upscatter. The viability of this scenario thus depends critically on the details of the velocity distribution, and also typically requires a more complex model where both particles in a collision do not need to excite simultaneously (to reduce the energy requirement for upscattering).

3.5 Simulated Effect on DM Halos

Choice of Model Parameters

In the previous Section, we found that at the velocity scale of $\sim 10\,\mathrm{km\,s^{-1}}$, relevant for the dynamics of dwarf galaxies, the range of particle physics parameters that result in interesting cross sections are approximately: $m_\chi \in [0.1, 300]\,\mathrm{GeV}c^{-2}$, $m_\phi \in [10^{-4}, 1]\,\mathrm{GeV}c^{-2}$, $\alpha \in [10^{-3}, 10^{-1}]$, $\delta \in [0.1, 10]\,\mathrm{keV}$. For the purposes of simulation, we select one particular model from this range with $\delta = 10\,\mathrm{keV}$, $m_\chi = 10\,\mathrm{GeV}c^{-2}$, $m_\phi = 30\,\mathrm{MeV}$ and $\alpha = 0.1$ (see top panel of Fig. 3.17 to see the velocity dependence of the various scattering channels with this choice of parameters). This choice results in an elastic cross section per unit mass[4] of a few $\mathrm{cm^2 g^{-1}}$ at the scale of dwarf galaxies, while the velocity kick is of the order of $v_{\rm kick} \sim 424\,\mathrm{km\,s^{-1}}$. We remark that this model has not been fine-tuned, and only represent a benchmark point in the relevant parameter space. The model has five different reactions and corresponding cross sections, which are presented in the bottom panel of Fig. 3.17: elastic scattering in the ground state ($\chi^1 + \chi^1 \to \chi^1 + \chi^1$), elastic scattering in the excited state ($\chi^2 + \chi^2 \to \chi^2 + \chi^2$), endothermic up-scattering ($\chi^1 + \chi^1 \to \chi^2 + \chi^2$), exothermic down-scattering ($\chi^2 + \chi^2 \to \chi^1 + \chi^1$), and elastic Yukawa scattering ($\chi^1 + \chi^2 \to \chi^1 + \chi^2$). We note that our benchmark point practically forbids up-scattering for the typical velocities of DM particles in the Milky Way environment, since the required energy splitting is too large for this reaction to occur frequently. This can be seen in the bottom panel of Fig. 3.17, where the cross section for up-scattering is zero for relative velocities $v < 2\sqrt{2\delta/m_\chi} \cong 848\,\mathrm{km\,s^{-1}}$.

Aside from specifying the model parameters, we also have to specify the initial conditions for the abundance of DM species; i.e. what fraction of DM is in which state initially – ground state (χ^1) or excited state (χ^2). Putting all particles initially into the excited state will maximize the effect of energy release during structure formation. On the other hand, putting all particles initially in the ground state will do the opposite, and behave like a purely elastic self-interacting DM model with multiple different cross sections. To explore the relevant range for inelastic self-interacting DM, we consider in the following two initial configurations: all particles initially in the excited state ($\chi^2_{\rm init} = 100\%$), or half the particles initially in the excited state ($\chi^2_{\rm init} = 50\%$). For the large dark-sector coupling ($\alpha = 0.1$) and relatively light DM mass ($10\,\mathrm{GeV}c^{-2}$) considered in this simulation, the natural cross section for depletion of DM through annihilation to vectors

[4] We again emphasize that for the model considered, the relevant cross section, which is the transfer cross section is the same as the total cross section given the lack of angular dependence of the differential cross section. Hereafter we use the terms transfer cross section and cross section interchangeably.

is $\langle\sigma v\rangle \sim \alpha^2/m_\chi^2 \sim 10^{-21}\,\mathrm{cm^3 s^{-1}}$, which is much larger than the thermal relic cross section. As discussed in Section 3.2, the DM could be produced via a non-thermal mechanism. In order to have $\sim 100\%$ of all particles in the excited state by the starting redshift of the simulations $z = 127$, it is necessary to suppress down-scattering (or delay the production of the excited state) until around the epoch of recombination ($z = 1100$). If instead, down-scattering is allowed already by the time of matter-radiation equality ($z_{\rm eq} \sim 3400$), then we can make a simple estimate of the ratio of excited to ground states by the time the simulation starts at $z_{in} = 127$.

$$\chi = 1 - \int_{z_{\rm in}}^{z_{\rm eq}} \frac{\Gamma(z)}{H(z)(1+z)} dz \tag{3.94}$$

where $H(z)$ is the Hubble expanstion rate and $\Gamma(z)$ is the scattering rate for de-excitation:

$$\Gamma(z) = \rho_{\rm dm}(z) v_{\rm typ}(z) \sigma_T^{\rm de}(v_{\rm typ}(z))/m_\chi \tag{3.95}$$

where $v_{\rm typ}$ is the characteristic velocity of DM particles, which at redshifts prior to $z = 127$ is roughly equal to the velocity dispersion of unclustered DM particles $\sim T(z)/(m_\chi T_{\rm kd})^{1/2}$ where $T(z)$ is the radiation temperature and $T_{\rm kd}$ is the DM kinetic decoupling temperature. After decoupling, the DM temperature and radiation temperature scale as $T_\chi = T^2/T_{\rm kd}$. We fix $T_{\rm kd} = 10\,\mathrm{MeV}$, but note that the dependence of χ on $T_{\rm kd}$ is only mild. The cross section for de-excitation $\sigma_T^{\rm de}/m_\chi$ is extrapolated down to the very small typical velocities of DM particles at early redshifts, from the behaviour in the bottom panel of Fig. 3.17. This calculation results in $\chi \sim 0.76$.

Numerical Implementation

Although we study in this paper a specific two-state inelastic self-interacting DM model, we have implemented a more general multi-state DM framework into the `Arepo` code [6]. This framework is able to handle an arbitrary number of states with an arbitrary number of reactions and corresponding cross sections, and with arbitrary, also non-degenerate, energy level splittings. This code represents a generalisation and complete rewrite of the algorithm presented in [261] and [268], which has been employed in multiple previous self-interacting DM studies [315–320]. Here we briefly describe this new numerical implementation.

In the following we assume that each DM simulation particle i is in a specific state α, i.e. every simulation particle represents a single state and not a mixture of different states. The simulation volume is then filled with DM simulation particles in different states $(\alpha, \beta, \gamma, \delta)$ with a variety of possible two-body scatterings:

$$\chi_i^\alpha + \chi_j^\beta \rightarrow \chi_i^\gamma + \chi_j^\delta, \tag{3.96}$$

where two particles, i and j, with initial states α (particle i) and β (particle j) scatter into two new states γ (particle i) and δ (particle j). The particle states have masses m_i^α, m_j^β before the scattering and m_i^γ, m_j^δ after the scattering. We note that these are simulation particle masses, whereas the actual DM particle masses are m_{χ^ϵ}, with $\epsilon = (\alpha, \beta, \gamma, \delta)$, depending on the state. The velocity-dependent transfer cross section for reaction $\alpha\beta \rightarrow \gamma\delta$ is given by $\sigma_{\rm T}^{\alpha\beta\rightarrow\gamma\delta}(v^{\alpha\beta})$, where $v^{\alpha\beta}$ is the

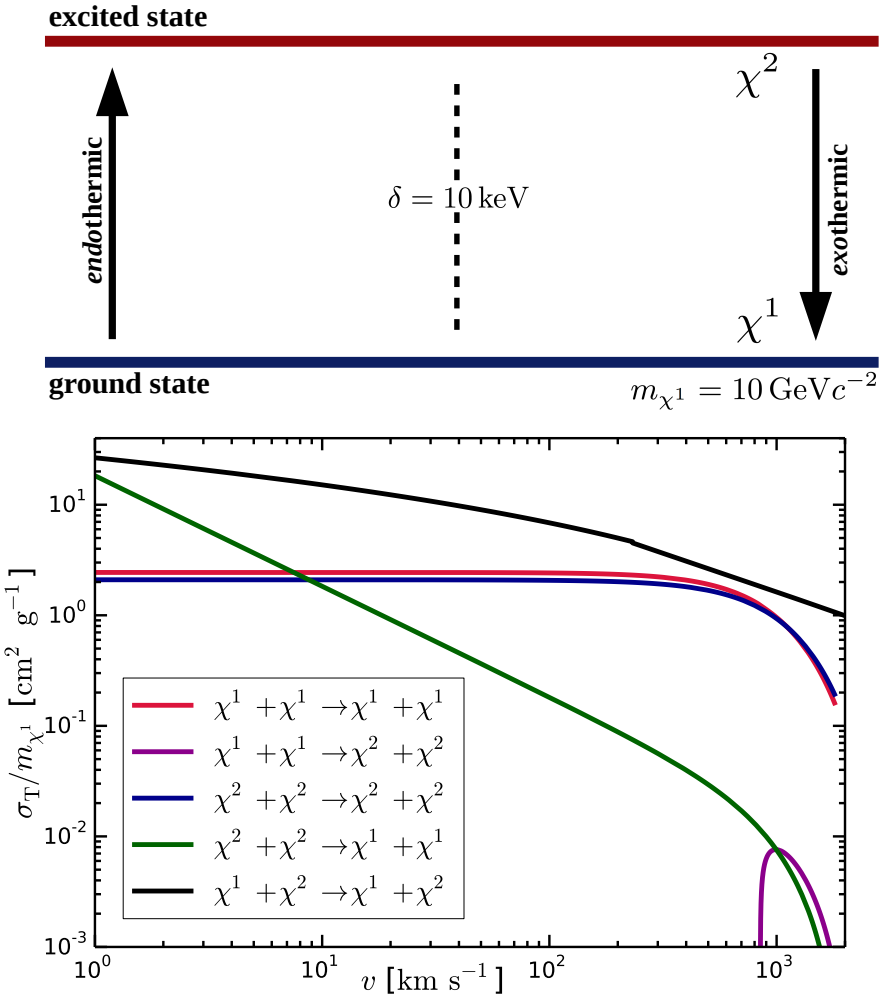

Figure 3.17: Inelastic self-interacting DM model. *Top panel:* Schematic overview of the two-state inelastic self-interacting DM model. The ground state (χ^1) and excited state (χ^2) are split by $\delta = 10\,\mathrm{keV}$. This two-state model allows for exo- and endothermic reactions. *Bottom panel:* Self-scattering cross section per unit mass for the different reactions of the two-state inelastic DM model. Up-scattering ($\chi^1 + \chi^1 \rightarrow \chi^2 + \chi^2$, purple) is suppressed in the Milky Way environment since it can only occur for large relative velocities, $v > 2\sqrt{2\delta/m_\chi^1} \cong 848\,\mathrm{km\,s^{-1}}$. The elastic cross sections ($\chi^1 + \chi^1 \rightarrow \chi^1 + \chi^1$, red) and ($\chi^2 + \chi^2 \rightarrow \chi^2 + \chi^2$, blue) are nearly flat over the whole velocity range relevant for the Milky Way environment. Down-scattering ($\chi^2 + \chi^2 \rightarrow \chi^1 + \chi^1$, green) mainly occurs for rather low velocities ($v \lesssim 10\,\mathrm{km\,s^{-1}}$). This reaction is exothermic corresponding to a velocity kick of $\sqrt{2\delta/m_\chi^1} \cong 424\,\mathrm{km\,s^{-1}}$. The largest cross sections occur for mixed-state elastic scattering ($\chi^1 + \chi^2 \rightarrow \chi^1 + \chi^2$, black), with a normalization that exceeds $\sim 10\,\mathrm{cm^2\,g^{-1}}$ for low velocities.

modulus of the relative velocity between particles in states α and β. The scattering rates for the different reaction channels are then given by:

$$R^{\alpha\beta\to\gamma\delta} = \frac{\rho^\beta}{m_{\chi^\beta}} \langle \sigma_{\mathrm{T}}^{\alpha\beta\to\gamma\delta}(v^{\alpha\beta})\, v^{\alpha\beta} \rangle, \tag{3.97}$$

where we take the thermal average of the product of the cross section times the relative velocities between particles. Here ρ^β measures the local density of DM particles in state β. We convert this mass density to a number density by dividing it by m_{χ^β}, the mass of the DM particle in state β. During a scattering reaction an energy $\Delta E^{\alpha\beta\to\gamma\delta}$ is released (exothermic) or absorbed (endothermic):

$$\Delta E^{\alpha\beta\to\gamma\delta} \begin{cases} = 0, & \text{elastic,} \\ < 0, & \text{inelastic: endothermic} \\ > 0, & \text{inelastic: exothermic.} \end{cases} \tag{3.98}$$

Once particle i in state α and j in state β have been selected to scatter into states γ and δ, we perform the scattering in the centre of mass frame and assign new velocities for the particles after the scattering:

$$\begin{aligned} \mathbf{v}_i &= \frac{m_i^\alpha + m_j^\beta}{m_i^\gamma + m_j^\delta} \mathbf{v}_{\mathrm{cm}} + \frac{m_j^\delta}{m_i^\gamma + m_j^\delta} \widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} v_{ij}\, \hat{\mathbf{e}}, \\ \mathbf{v}_j &= \frac{m_i^\alpha + m_j^\beta}{m_i^\gamma + m_j^\delta} \mathbf{v}_{\mathrm{cm}} - \frac{m_i^\gamma}{m_i^\gamma + m_j^\delta} \widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} v_{ij}\, \hat{\mathbf{e}}, \end{aligned} \tag{3.99}$$

where $\mathbf{v}_{\mathrm{cm}}$ is the centre of mass velocity of the two particles, v_{ij} their relative velocity, $\hat{\mathbf{e}}$ is a random vector on the unit sphere, and $\widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta}$ is a dimensionless velocity scale factor that depends on the energy splitting related to the reaction the two particles are undergoing:

$$\widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} = \sqrt{\frac{\mu_{ij}^{\alpha\beta}}{\mu_{ij}^{\gamma\delta}} \left(1 + \frac{2\Delta E^{\alpha\beta\to\gamma\delta}}{\mu_{ij}^{\alpha\beta} v_{ij}^2} \right)}, \tag{3.100}$$

where $\mu_{ij}^{\alpha\beta} = m_i^\alpha m_j^\beta / (m_i^\alpha + m_j^\beta)$ and $\mu_{ij}^{\gamma\delta} = m_i^\gamma m_j^\delta / (m_i^\gamma + m_j^\delta)$ are the reduced masses of the two particles before and after the scattering event; i.e., we take into account the change in particle mass during the scattering. We note however, that this change is usually tiny in the cases of interest where the mass splitting is small. For example, for our specific nearly-degenerate model, it is negligible (10^{-6}) such that the two reduced masses are essentially given by $1/2\, m_0$, assuming a constant DM simulation particle mass m_0. Based on the three cases for $\Delta E^{\alpha\beta\to\gamma\delta}$ we find accordingly for the dimensionless velocity scale factor:

$$0 \leq \widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} \begin{cases} = 1, & \text{elastic,} \\ < 1, & \text{inelastic: endothermic} \\ > 1, & \text{inelastic: exothermic.} \end{cases} \tag{3.101}$$

In the inelastic case this energy is either given or taken from the two scattering particles in equal parts. The endothermic regime is limited by the fully inelastic collision ($\widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} = 0$) after which both particles move with the centre of mass velocity. The exothermic case is not limited and can in principle inject an arbitrary amount of energy into the system depending on the energy level splitting.

To decide whether a certain scattering reaction occurs between two simulation particles, we have to evaluate the corresponding scattering probabilities. The pairwise scattering probability for a given reaction $\alpha\beta \to \gamma\delta$, and the total probability for scattering of a given particle with any of its neighbours are given respectively by:

$$P_{ij}^{\alpha\beta\to\gamma\delta} = \delta_{ij}^{\alpha\beta}\, m_j^{\beta}\, W_{ij}\, \frac{\sigma_{\mathrm{T}}^{\alpha\beta\to\gamma\delta}(v_{ij})}{m_{\chi^\beta}}\, \frac{v_{ij}}{2}\, \Delta t_i, \tag{3.102}$$

$$P_i^{\alpha\beta\to\gamma\delta} = \sum_{j=0}^{N_{\mathrm{ngb}}} P_{ij}^{\alpha\beta\to\gamma\delta}, \tag{3.103}$$

where N_{ngb} is the number of neighbor particles, Δt_i the individual time-step of particle i, and $W_{ij} = w(r_{ij}/h_i, h_i)$ is the cubic spline Kernel function with a 3D normalization:

$$w(q,h) = \frac{8}{\pi\, h^3} \begin{cases} 1 - 6q^2 + 6q^3, & 0 \le q \le \frac{1}{2}, \\ 2\,(1-q)^3\,, & \frac{1}{2} < q \le 1, \\ 0, & q > 1. \end{cases} \tag{3.104}$$

Here h_i is the smoothing length including the N_{ngb} neighbor particles j around particle i with spatial distance r_{ij}. We note that N_{ngb} does not distinguish the particle states; i.e., it is possible that particles of a certain state are not enclosed in the smoothing length. In this case scattering between the particle in question with this state is not possible. The factor $1/2$ in the calculation of $P_{ij}^{\alpha\beta\to\gamma\delta}$ accounts for the fact that a scatter event always involves two particles, and we therefore need to divide by two to reproduce the correct scattering rate. A given neighbour only contributes to the sum if the initial states of the reaction match the actual particle states. This is guaranteed by the Kronecker delta function $\delta_{ij}^{\alpha\beta}$, which is equal to 1 if particle i is in initial state α and particle j is in initial state β. Otherwise the function evaluates to zero.

To decide whether a particle is going to scatter and with which reaction, we first arbitrarily order the reactions according to $0 \le \zeta = (\alpha\beta \to \gamma\delta) \le \zeta_{\max}$, where $\zeta_{\max}$ denotes the total number of reactions. For each particle we then draw a random number $x_i^P \in U(0,1)$. A scattering of particle i occurs if there is a reaction $\zeta_i \le \zeta_{\max}$ with

$$\sum_{\zeta=0}^{\zeta_i - 1} P_i^{\zeta} < x_i^P < \sum_{\zeta=0}^{\zeta_i} P_i^{\zeta}. \tag{3.105}$$

If such a ζ_i exists, particle i will scatter with reaction $\zeta_i = (\alpha_i\beta_i \to \gamma_i\delta_i)$. This then determines the energetics of the scattering process (inelastic, exothermic or endothermic), the state of the scattering partner (β_i), and the final states of the reaction (γ_i, δ_i). Once the scattering reaction is

decided, a scattering partner j_i for particle i needs to be found. The exact partner depends on the reaction since it must be a particle which is currently in state β_i and will then scatter into state δ_i. The selection is done by finding the partner j_i that satisfies:

$$P_i^{\zeta<\zeta_i} + \sum_{j=0}^{j_i-1} P_{ij}^{\zeta_i} < x_i^P < P_i^{\zeta<\zeta_i} + \sum_{j=0}^{j_i} P_{ij}^{\zeta_i}, \tag{3.106}$$

where $P^{\zeta<\zeta_i}$ is the sum of all probabilities for reactions smaller than ζ_i, i.e. $P^{\zeta<\zeta_i} = \sum_{\zeta=0}^{\zeta_i-1} P_i^{\zeta}$.

Once the scattering partner j_i is found, the scattering can be performed by assigning new velocities to particle i and j_i based on the dimensionless velocity scale factor $\widetilde{v}_{ij_i}^{\zeta_i} = \widetilde{v}_{ij_i}^{\alpha_i\beta_i\to\gamma_i\delta_i}$. In a last step we assign the new states to the scattering particles, i.e., the state of particle i is changed to γ_i and the state of particle j_i is changed to δ_i. We also change the masses of the particles to reflect their state change, although we note that in our simulation this only results in a minimal mass change due to the small mass splitting for the nearly degenerate system presented above.

To avoid multiple scattering per particle in a single time-step, we impose a limit to the time-step of each particle i:

$$\Delta t_i < \kappa \min_{\alpha,\beta,\gamma,\delta} (\rho_i^\alpha \sigma_{\mathrm{T}}^{\alpha\beta\to\gamma\delta}(\sigma_i^\alpha)/m_{\chi^\alpha}\sigma_i^\alpha)^{-1}, \tag{3.107}$$

where ρ_i^α is the density of particles in state α at the location of particle i, and σ_i^α is the corresponding velocity dispersion. κ is a dimensionless parameter that can be adjusted to control the size of the minimal time-step. For our simulations we find that $\kappa = 0.0025$ is sufficient to avoid multiple scattering and usually sets the time-step to be smaller than the local dynamical time scale in the simulations.

Impact of inelastic DM interactions on an isolated halo

We test our inelastic self-interacting DM implementation by setting up an isolated halo in equilibrium [321] with $M_{200} = 10^{10}\,\mathrm{M}_\odot$ and a concentration[5] of $c = 8$ for the benchmark DM model presented above. The halo is sampled with 10^7 particles and gravitational forces are softened with a Plummer-equivalent softening length of $\epsilon = 10\,\mathrm{pc}$. We simulate this halo in three different DM models: CDM, elastic self-interacting DM and inelastic self-interacting DM. The CDM case just follows the evolution of collisionless CDM, whereas the other two cases employ the two-state model with the cross sections discussed above (see bottom panel of Fig. 3.17) using the numerical implementation described in the previous section. The elastic simulation artificially suppresses the energy release during (de-)excitations, but is otherwise identical to the inelastic self-interacting DM model. This simulation can then be compared to the inelastic case, where velocity kicks play a relevant role. To be more specific, the elastic case is realised by simply setting $\widetilde{v}_{ij}^{\alpha\beta\to\gamma\delta} = 1$ during any scattering process, i.e., $\Delta E^{\alpha\beta\to\gamma\delta} = 0$. As mentioned above, we explore two different initial configurations for the non-CDM models. In the first configuration all particles are initially in the

[5]Defined as the ratio of r_{200}, the radius where the mean density of the halo is equal to 200 times the critical density, and r_{-2}, the radius where the logarithmic slope of the density profile is equal to -2.

		Model				
			elastic scattering		inelastic scattering	
		CDM	$\chi^2_{\rm init} = 100\%$	$\chi^2_{\rm init} = 50\%$	$\chi^2_{\rm init} = 100\%$	$\chi^2_{\rm init} = 50\%$
M_{200}	$[10^{12}\,M_\odot]$	1.609	1.600	1.600	1.478	1.569
r_{200}	[kpc]	243.85	243.39	243.38	237.05	241.81
$V_{\rm max}$	$[{\rm km\,s^{-1}}]$	174.87	178.12	177.51	164.14	172.44
$R_{\rm max}$	[kpc]	66.14	62.22	63.39	77.21	64.76
$N_{\rm sub}$	-	16,264	14,486	14,675	13,171	14,409
χ^1	[%] ($r < 300\,$kpc)	-	7.51	52.23	0.65	51.12
Energy	[SNIIs ($10^{51}\,$erg)]	-	-	-	755×10^6	268×10^6

Table 3.2: Basic properties of the Milky Way-sized halo simulated in the different DM models. We list the virial mass (M_{200}), virial radius (r_{200}), maximum circular velocity ($V_{\rm max}$), radius where the maximum circular velocity is reached ($R_{\rm max}$), the number of resolved subhaloes within $300\,$kpc ($N_{\rm sub}$), the ground state population of all DM particles within $300\,$kpc, and the injected energy (in units of $10^{51}\,$erg, i.e., the canonical energy of a SNII) measured based on the population split between the two DM states of all high resolution particles at $z = 0$.

excited state ($\chi^2_{\rm init} = 100\%$), while in the second only half of the particles are initially in the excited state ($\chi^2_{\rm init} = 50\%$).

In Fig. 3.18 we present the time evolution of the population split in the halo for the two initial conditions for the inelastic self-interacting DM case, in the left and right panels, respectively. The right axis in each panel indicates the total injected energy into the system (solid white lines) in units of the canonical SNII energy ($10^{51}\,$erg). This measures the total energy released due to de-excitation that is transformed into kinetic energy by the cumulative effect of velocity kicks in each scattering event. As expected, in both cases the ground state population increases over time while the excited state gets de-populated. As mentioned above, this is a consequence of the employed cross sections, where up-scattering is strongly suppressed (see bottom panel of Fig. 3.17), and does not occur given the typical relative velocities of particles in a $10^{10}\,M_\odot$ halo. After a few Gyr, the exothermic reaction has already injected the equivalent of more than one million SNII for the $\chi^2_{\rm init} = 100\%$ configuration. After $10\,$Gyr, the cumulative energy injection of down-scatterings has reached $\sim 2 \times 10^{57}\,$erg of energy into the system for the $\chi^2_{\rm init} = 100\%$ configuration, and $\sim 7 \times 10^{56}\,$erg for the $\chi^2_{\rm init} = 50\%$ configuration. The ground state population increases from 0% to 10% for the $\chi^2_{\rm init} = 100\%$ initial configuration over the simulation time span of $10\,$Gyr. Similarly also the ground state population for the $\chi^2_{\rm init} = 50\%$ initial configuration increases as a function of time, although the total increase is lower in this case. As a reference, we note that the minimum energy required to transform a cusp into a 1 kpc core for a Navarro-Frenk-White (NFW) halo with a mass of $10^{10}\,M_\odot$ calculated from the virial theorem assuming initial and final equilibrium configuration is $\sim 10^{55}\,$erg [255]. Depending on the stellar mass content, efficiency

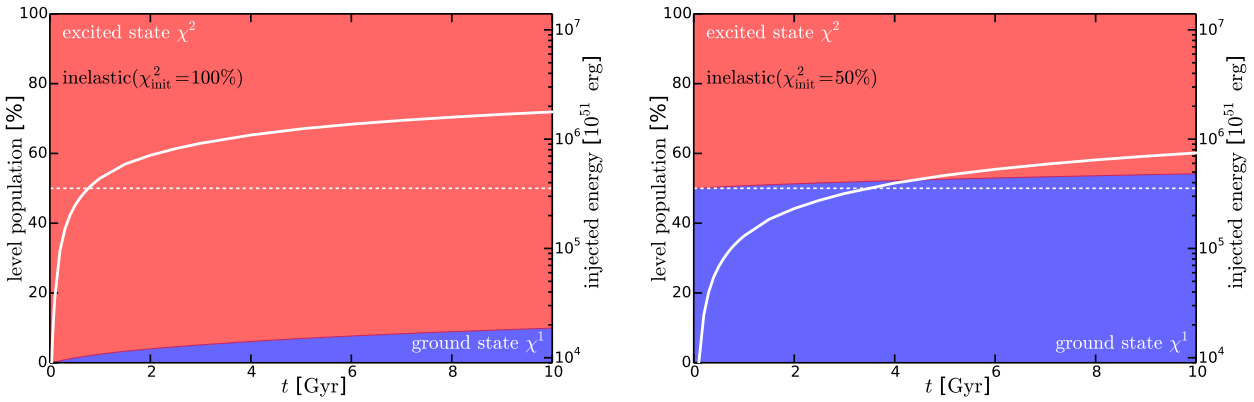

Figure 3.18: Time evolution of the level population for the inelastic self-interacting DM model with different initial population split. For the $\chi^2_{\rm init} = 100\%$ configuration (*left panel*) all particles are initially in the excited state (χ^2), whereas for the $\chi^2_{\rm init} = 50\%$ configuration (*right panel*) only 50% of the particles are initially excited and the rest is in the ground state (χ^1). The split between the two level populations is shown by the filled areas. The solid white lines present the total energy injected into the halo through level decay. This energy is plotted in units of the canonical energy of a single SNII (10^{51} erg) with the scale shown on the right vertical axis. After $10\,$Gyr, de-excitation has injected $\sim 2 \times 10^{57}$ erg ($\sim 7 \times 10^{56}$ erg) for the $\chi^2_{\rm init} = 100\%$ ($\chi^2_{\rm init} = 50\%$) configuration.

of energy injection, and star formation history of a given galaxy living in a halo of this size, this energetic requirement might or might not be satisfied by the SNe-driven core formation scenario (e.g. [255, 322, 323]). In the inelastic self-interacting DM model explored here the cumulative energy available easily exceeds this minimum energy requirement.

The mass splitting, which corresponds to a velocity kick of $424\,{\rm km\,s^{-1}}$ is sufficient to unbind particles, i.e., to efficiently remove DM particles from the $10^{10}\,{\rm M}_\odot$ halo. This is demonstrated in Fig. 3.19, where we show the dimensionless energy distribution of all gravitationally unbound particles after $10\,$Gyr. The CDM case (black lines) has a very small fraction of gravitationally unbound particles $\mathcal{O}(0.01\%)$, caused by numerical noise over the equilibrium configuration. This population serves as a comparison baseline with the other models. On the other hand, the inelastic model (red lines) has a much larger fraction of gravitationally unbound particles ($\sim 5\% - 10\%$), with a significant number of them populating a nearly log-normally distributed peak at high energies. These are the particles that were predominantly expelled from the halo centre during the velocity kicks imparted in down-scatterings. Notice that this population is absent in the elastic case (blue lines), which shows a very similar distribution as the CDM case, but shifted towards higher energies and with a higher fraction of gravitationally unbound particles, of $\mathcal{O}(1\%)$. This distinct population of gravitationally unbound particles in the inelastic case is key to understand the further reduction of the core density that happens in this case compared to pure elastic self-interacting DM.

We can roughly estimate the energy shifts in the elastic and inelastic self-interacting DM cases.

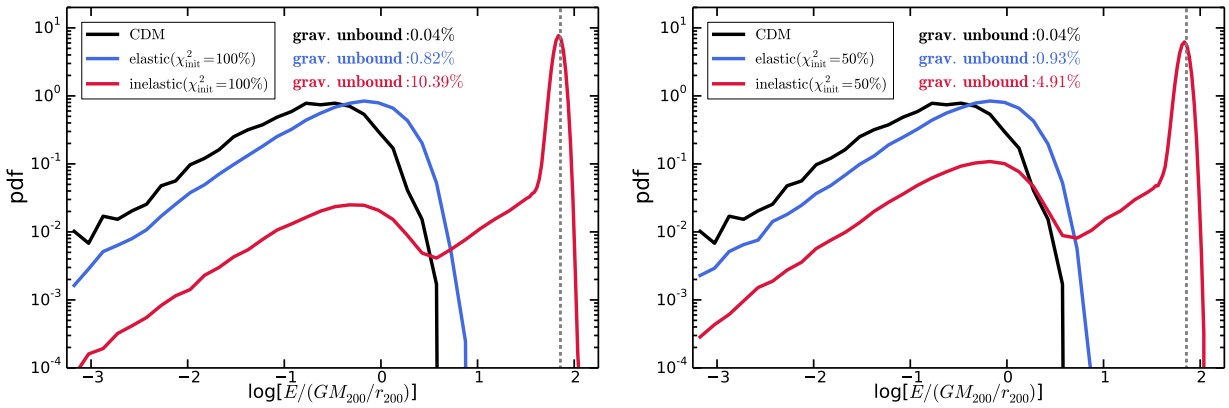

Figure 3.19: Dimensionless energy distribution of gravitationally unbound particles after 10 Gyr for the inelastic models in Fig. 3.18. In the *left panel*, 100% of the particles are initially in the excited state, while on the *right panel*, 50% of the particles are initially excited and the rest is in the ground state. The legends on the panels also show the total fraction of gravitationally unbound particles in the different models corresponding to the black (CDM), blue (elastic) and red (inelastic) lines. The inelastic self-interacting DM model leads to a significant removal of particles, predominantly from the halo centre, and has a peak at an energy around $\sim 74\,GM_{200}/r_{200}$ due to the exothermic reactions (gray, dashed vertical line; see text for an analytical estimate). The elastic self-interacting DM model has a distribution similar to the CDM distribution, but slightly shifted towards larger energies. The fraction of gravitationally unbound particles is about twice as large for the $\chi^2_{\rm init} = 100\%$ initial configuration compared to the $\chi^2_{\rm init} = 50\%$ case.

For the elastic case we assume that the shift seen between the blue and black lines is related to the energy gained by the particles that were barely bound in the inner halo, and that become gravitationally unbound after elastic scattering with particles with more kinetic energy. Since scattering is more frequent deep within the potential, this energy gain is approximately bounded by the the r.m.s. velocity at the maximum of the velocity dispersion profile, which occurs at $r_{\sigma,\,\rm max} \sim a/5$, where a is the scale radius of the Hernquist profile. In [261] we found that the r.m.s. velocity at the core is $v^2_{\rm rms}(r_{\sigma,\,\rm max}) \sim 3\sigma^2(r_{\sigma,\,\rm max}) \sim 0.96\,GM_{200}/a$ (see Fig. 2 of [261]). This means that those particles that were barely bound become gravitationally unbound gaining an additional kinetic energy $0.48\,GM_{200}/a \sim 3.42\,GM_{200}/r_{200}$, which is roughly 0.5 dex to the right, relative to the CDM distribution where this effect is absent. We note that our analytic estimate is an oversimplification given the radially dependent re-distribution of energy. For the inelastic case the elastic collisions still cause a similar shift as in the elastic case since the reaction ($\chi^1 + \chi^2 \rightarrow \chi^1 + \chi^2$) is very frequent (it has the highest cross section). In addition, during down-scatterings, the resulting ground state particles receive a velocity kick $\sqrt{2\delta/m_\chi}$, which results in a gain in kinetic energy of $\sim 74\,GM_{200}/r_{200}$, which is shown in Fig. 3.19 as a vertical dashed line. This estimate describes the numerical result well, demonstrating that the numerical implementation

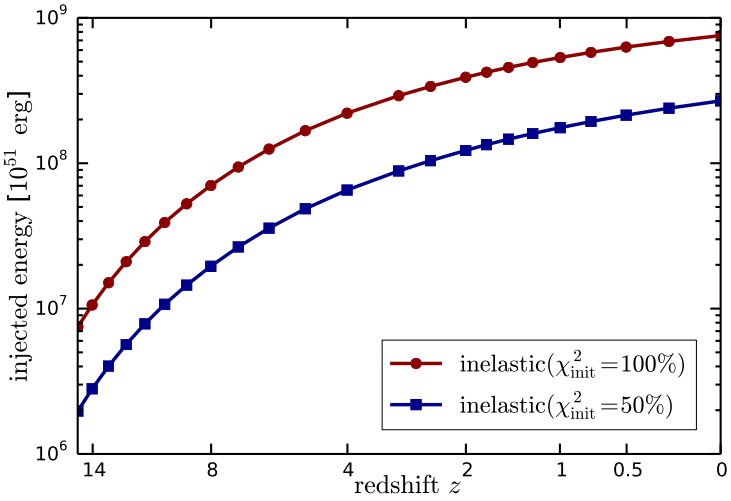

Figure 3.20: Injected energy into the Milky Way-like halo due to level de-excitation. For both initial configurations ($\chi^2_{\rm init} = 100\%$ and $\chi^2_{\rm init} = 50\%$) the total injected energy into the halo due to down-scattering of the excited states is equivalent to hundreds of millions of SNII (10^{51} erg). This energy leads to an increased core formation compared to elastic self-interacting DM models.

behaves correctly.

Impact of inelastic self-interacting DM on a Milky Way-like halo

Next we explore the impact of inelastic self-interacting DM on a galactic halo in a cosmological context. We resimulated the galactic halo ($M_{200} = 1.6 \times 10^{12}\,{\rm M}_\odot$) presented in [268] within inelastic self-interacting DM for the benchmark model presented above with two different initialisations, $\chi^2_{\rm init} = 100\%$ and $\chi^2_{\rm init} = 50\%$. The simulations employ the following cosmological parameters: $\Omega_m = 0.302$, $\Omega_\Lambda = 0.698$, $\Omega_b = 0.046$, $h = 0.69$, $\sigma_8 = 0.839$ and $n_s = 0.967$, which are consistent with recent Planck data [324, 325]. The initial conditions are generated at $z = 127$. The gravitational softening length is fixed in comoving coordinates until $z = 9$, and is then fixed in physical units until $z = 0$. The DM particle mass resolution is $2.756 \times 10^4\,{\rm M}_\odot$ with a Plummer-equivalent softening length of $72.4\,{\rm pc}$ at $z = 0$.

The main properties of this halo are presented in Table 3.2 for the four different self-interacting DM models, elastic and inelastic with the two different initial state populations explored in this paper. We can see that the virial mass, M_{200}, is reduced by nearly $\sim 10\%$ for the inelastic model with initially all particles in the excited state ($\chi^2_{\rm init} = 100\%$). This mass loss is a consequence of unbinding ground state particles after down-scattering. We also find a rather significant reduction in $V_{\rm max}$ of about $\sim 8.5\%$. The total number of resolved subhaloes within $300\,{\rm kpc}$ is also reduced by $\sim 23\%$ for the inelastic model with $\chi^2_{\rm init} = 100\%$. Here we count all subhaloes that are found by our structure finder [326] with more than 20 particles corresponding to a lower mass limit of $5.512 \times 10^5\,{\rm M}_\odot$. We note that there is a significant reduction in the abundance of subhaloes in

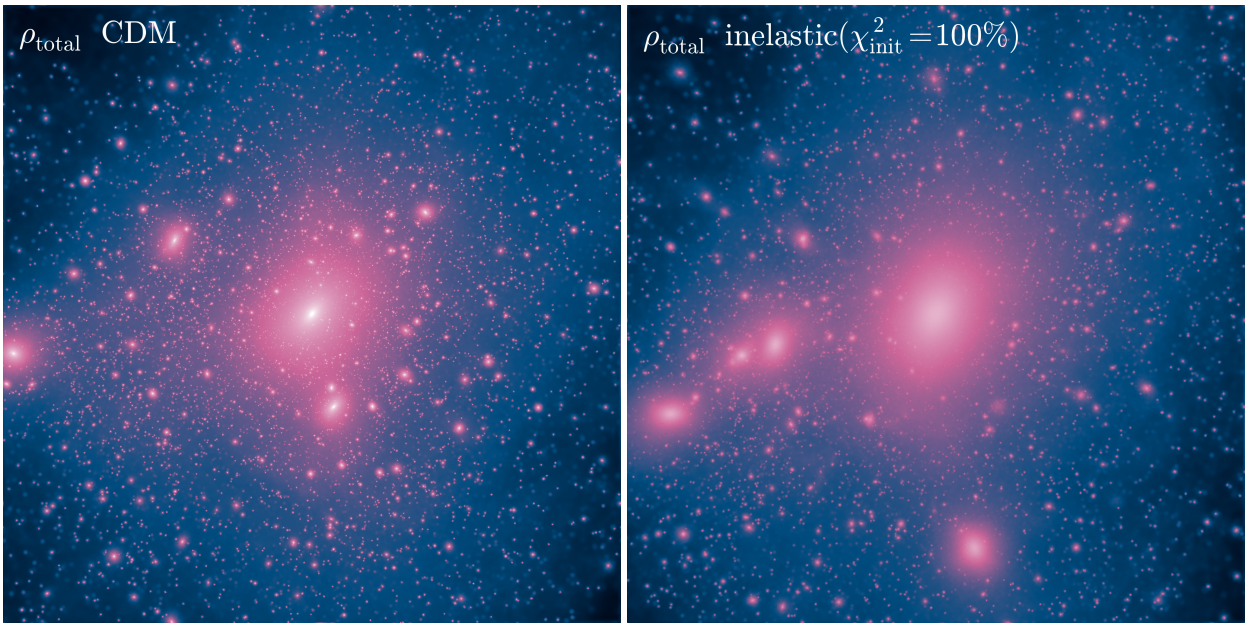

Figure 3.21: Projected DM density for the Milky Way-sized halo in CDM and for inelastic self-interacting DM. The *left panel* shows the CDM case, whereas in the *right panel* we present the inelastic self-interacting DM case where 100% of the particles are initially in the excited state ($\chi^2_{\rm init} = 100\%$). The spatial extent of the maps is $500\,$kpc across with a depth of $300\,$kpc, with a colour scale related to the projected value of ρ^2. It is apparent that the inelastic self-interacting DM model leads to reduced core densities and an overall reduction in the abundance of subhaloes due to down-scattering. The virial mass for the halo in the inelastic case is reduced by about 10% compared to the CDM case due to the removal of DM particles in the ground state following de-excitation. The efficient removal of ground state particles leads to a very small fraction, $\lesssim 1\%$, of such particles within $300\,$kpc.

both the elastic and inelastic self-interacting DM model. The fact that this also happens for the elastic model is caused by the rather large elastic cross sections that we employ in our benchmark model (see bottom panel of Fig. 3.17). In particular, the cross section for the ($\chi^1 + \chi^2 \to \chi^1 + \chi^2$) reaction reaches values larger than $10\,{\rm cm}^2\,{\rm g}^{-1}$ for low relative velocities, and it is around $5\,{\rm cm}^2\,{\rm g}^{-1}$ at relative velocities $\sim 200\,{\rm km\,s}^{-1}$, which are the relevant ones for subhalo evaporation. For instance, in [261] it was demonstrated that an elastic cross section of $\sim 10\,{\rm cm}^2\,{\rm g}^{-1}$ on galactic scales leads to a significant suppression of substructure. As anticipated, this reduction in the abundance of subhaloes is clearly enhanced if inelastic scattering is included. As is clear from Table 3.2, qualitatively similar trends are observed for the $\chi^2_{\rm init} = 50\%$ configuration. The second-to-last column of Table 3.2 shows the present-day population of ground state particles within a radius of $300\,$kpc from the galactic centre. We find that for the inelastic model the ground state population increases to $\sim 0.7\%$ for the $\chi^2_{\rm init} = 100\%$ initial configuration and to $\sim 51.1\%$ for the $\chi^2_{\rm init} = 50\%$ initial

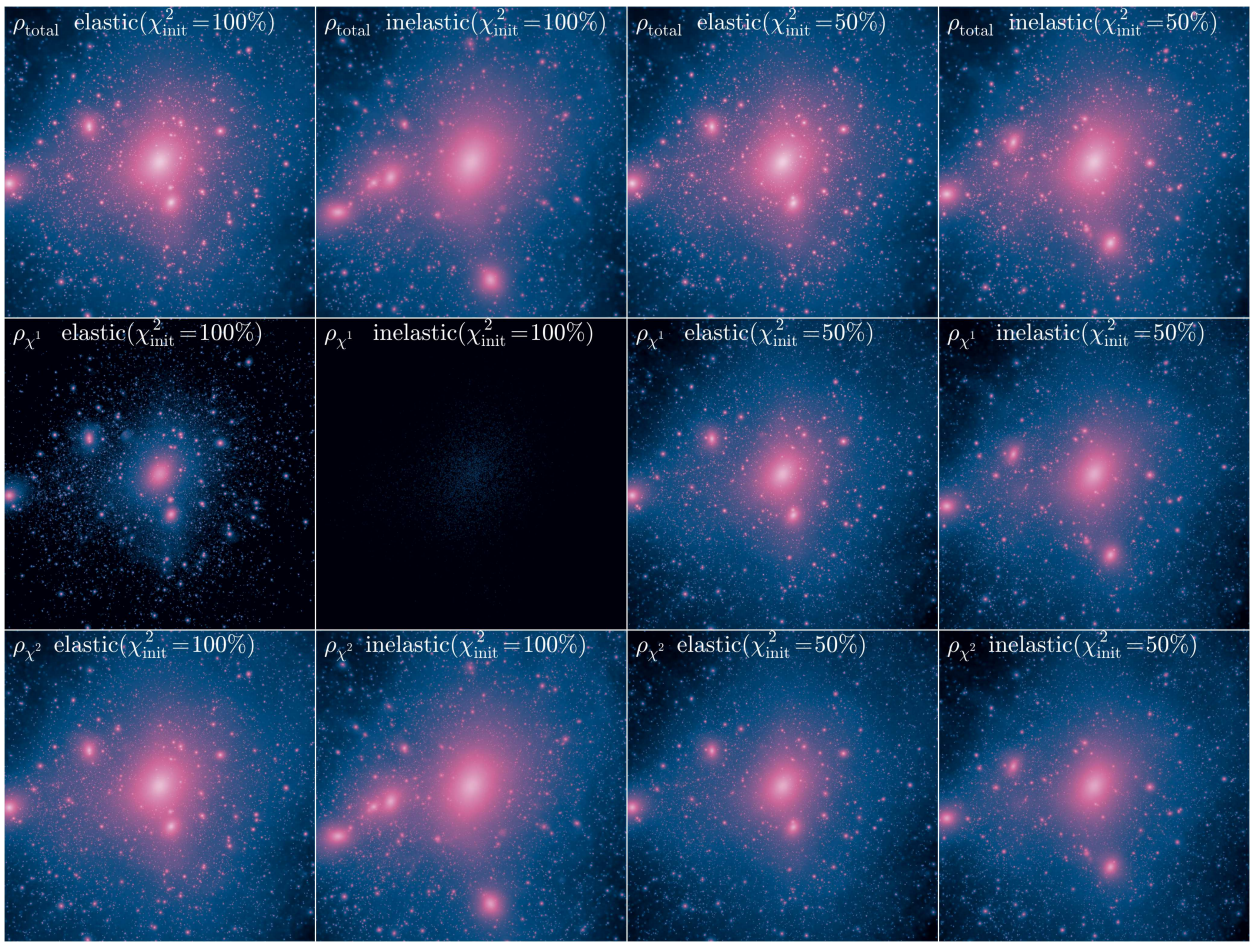

Figure 3.22: Projected DM density for the total DM density, the lower level, and the upper level densities for the self-interacting DM models. The total extent of the maps is $500\,$kpc with a depth of $300\,$kpc, and we project ρ^2. *Top row panels:* Total DM density fields for the halo in different self-interacting DM models as indicated. The largest difference compared to the CDM case occurs for the inelastic model with the $\chi^2_{\rm init} = 100\%$ initial configuration. The reduction of core densities and subhalo abundance is smaller for the $\chi^2_{\rm init} = 50\%$ initial configuration compared to the $\chi^2_{\rm init} = 100\%$ case. *Middle row panels:* DM densities of the ground state (χ^1) for the different models. For the inelastic case nearly all excited particles have been removed from the halo centre due to the energy injection during de-excitation of the state. This is not the case for the elastic model, where the ground state particles stay close to the halo centre and are only redistributed due to scattering events. *Bottom row panels:* DM densities of the excited state (χ^2) for the different models.

configuration. On the other hand, for the elastic case we find larger fractions of particles in the ground state at $z = 0$ than in the inelastic case, $\sim 7.5\%$ and $\sim 52.2\%$ for the $\chi^2_{\rm init} = 100\%$ and $\chi^2_{\rm init} = 50\%$ configurations, respectively. This is because the elastic configuration has been set on purpose to have the same reactions as the inelastic benchmark case, but without the energy release (see Section 3.5), i.e., there is no velocity kick associated to down-scattering, which means that the associated ground state particles remain bound to the halo. For the inelastic case, we provide in the last column of the table the cumulative injected energy due to de-excitations. For the fully excited initial configuration we find that a total of $\sim 8 \times 10^8 \times 10^{51}$ erg are injected. This value is a factor of a few lower ($\sim 3 \times 10^8 \times 10^{51}$ erg) for the configuration where only 50% of the particles are initially in the upper state. The redshift dependence of the injected energy is presented in Fig. 3.20. The magnitude of these energies suggest that inelastic self-interacting DM can have a substantial impact on the galactic halo, both in terms of abundance of substructure, and their density structure.

The projected DM density distribution of the simulated halo for a CDM universe is shown in the left panel of Fig. 3.21, while the inelastic self-interacting DM model with $\chi^2_{\rm init} = 100\%$ is shown in the right panel. These plots clearly show that in this specific inelastic self-interacting DM model, both the central (sub)halo densities and subhalo abundance are significantly reduced. This latter property is quite distinct from typical elastic self-interacting DM models where a relevant difference in the subhalo abundance relative to CDM is only possible for rather large elastic cross sections on galactic scales ($\sim 10\,{\rm cm}^2\,{\rm g}^{-1}$).

The inelastic self-interacting DM model with $\chi^2_{\rm init} = 100\%$ represents the most extreme scenario of all our simulations. To show the larger diversity of the non-CDM cases we explored, we present the corresponding maps of all of them in the four top panels of Fig. 3.22. These four panels show the total DM density, i.e., taking into account the ground and excited states. Overall, we see a consistent trend on the abundance of substructure being more suppressed in the inelastic cases compared to the elastic ones. The effect is however, considerably stronger in the case where all particles are initially in the excited state (the two leftmost panels are visually distinct, while the rightmost panels are more alike). In the middle and bottom panels of Fig. 3.22 we show the density maps but only considering particles in the ground state, and excited state, respectively. The inelastic model with $\chi^2_{\rm init} = 100\%$ has a very low ground state density. De-excited particles escape from the halo centres in that model due to the strong velocity kicks. For the associated elastic case, these particles are not removed and therefore stay near the halo centres and are visible in the maps. A visual comparison between the two leftmost panels of Fig. 3.22 shows very clearly the striking difference between elastic and inelastic self-interacting DM models. For the $\chi^2_{\rm init} = 50\%$ configuration, the differences between the ground and excited state populations, and between the elastic and inelastic cases are barely visible in the maps since the increase of ground state particles due to de-excitations is only at the per cent level.

To quantify in detail the distribution of the two-state population in the self-interacting DM models, we show in the top panels of Fig. 3.23 the radial profiles of the ground state population centered in the Milky Way-sized halo. The differences between the central distributions of the ground and excited state populations are striking, particularly for the case where all particles are initially in the excited state ($\chi^2_{\rm init} = 100\%$, left panel). As mentioned above this is due to the large energy imparted to the ground state particles during down-scattering, which is large enough to

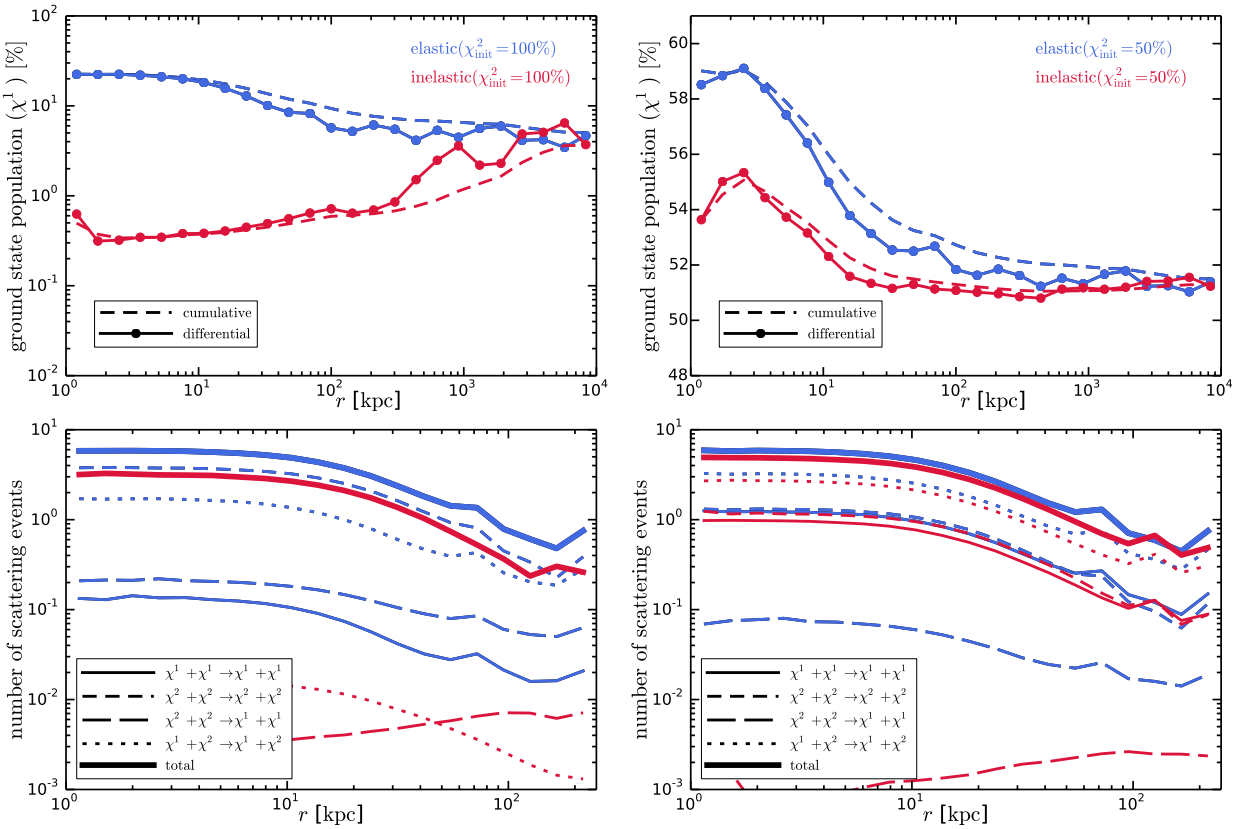

Figure 3.23: Level population and scatter statistics of the Milky Way-sized halo. *Top panels:* Radial profiles of the ground state population for the $\chi^2_{\rm init} = 100\%$ configuration in the left, and $\chi^2_{\rm init} = 50\%$ in the right (solid: differential, dashed: cumulative). We show the elastic and inelastic cases with blue and red lines, respectively. For the inelastic case, down-scattered particles escape the centre of the halo, which causes a strong suppression of ground state particles. *Bottom panels:* Radial profiles of the mass-weighted average number of scattering events in each reaction channel according to the legend. The number of scattering events in each channel depends on the initial configuration and the cross section of each channel. For the $\chi^2_{\rm init} = 50\%$ case (right panel), the distribution of scattering events reflects the ranking of the cross section (see bottom panel of Fig. 3.17). Comparatively, this is different for the $\chi^2_{\rm init} = 100\%$ case (left panel), where the reactions ($\chi^1+\chi^2 \to \chi^1+\chi^2$) and ($\chi^1+\chi^1 \to \chi^1+\chi^1$) are suppressed since no ground state scattering partners are available initially. The thick solid lines show the total number of scatters summed over all channels. We note that we do not show reactions with less then 10^{-3} average scattering events. Within the relevant radial range, i.e., within $r = R_{\rm max} \sim 60$ kpc, all the cases have $\mathcal{O}(1)$ scattering events per particle (including all channels) by $z = 0$.

remove these particles from the halo. These removed particles are actually visible in Fig. 3.23 far away from the halo centre in the differential ground state population distribution. They appear as peaks in this distribution around and beyond 1 Mpc from the halo centre. Since the velocity kick in our benchmark model is $424\,\mathrm{km\,s^{-1}}$, the particle would travel ~ 1 Mpc already within 3 Gyr. For the elastic case on the the other hand, the ground state population rises towards the centre since the down-scattering rate is higher towards the denser and colder halo centre (see bottom panel of Fig. 3.17), unimpeded due to the lack of energy release. Here we find that the ground state population reaches more than 20% towards the centre of the halo. Even at a distance of $\sim 100\,\mathrm{kpc}$ we still find a cumulative fraction of $\sim 10\%$ of all particles in the ground state for the elastic mode, whereas for the inelastic case, this fraction is much smaller and less than a per cent even out to ~ 1 Mpc. The right panel of Fig. 3.23 shows the profiles for the case where 50% of all particles are initially in the excited state ($\chi^2_{\mathrm{init}} = 50\%$). We find here a similar trend when comparing the elastic and inelastic cases, i.e., the ground state is more densely populated for the elastic case in the inner halo. For the elastic case about 55% of all particles in the centre are in the ground state. Lastly, we note that the cumulative fraction of ground state particles within ~ 10 Mpc agrees between the elastic and inelastic cases for both initial state configurations. This population is at the per cent level above the initial ground state population. This radial distance is large enough to include even those particles that were ejected due to de-excitation in the inelastic self-interacting DM model, and thus, the ground state level populations converge to similar values at large radii. Any remaining deviations are due to differences in the detailed scatter reactions that occurred during the assembly history of the halo.

The lower panels of Fig. 3.23 show the mass-weighted average number of scattering events per particle in each reaction channel. For the $\chi^2_{\mathrm{init}} = 50\%$ configuration (right panel) we find that the average number of scattering events follows closely the overall ranking of the cross sections presented in the bottom panel of Fig. 3.17. For example, the mixed elastic reaction ($\chi^1 + \chi^2 \rightarrow \chi^1 + \chi^2$) has the largest cross section of all channels and, correspondingly, it also contributes to the largest number of average scattering events at all radii. The exothermic reaction ($\chi^2 + \chi^2 \rightarrow \chi^1 + \chi^1$) on the other hand, has the smallest cross section, except for very small relative velocities, and therefore leads to the smallest number of scattering events. The inelastic model shows a lower number of scattering events along that channel (long-dashed red lines) compared to the elastic case (long-dashed blue lines), because of the removal of particles once they de-excite into the ground state. The situation is quite different for the $\chi^2_{\mathrm{init}} = 100\%$ configuration (left panel) since some channels are initially suppressed because no ground state scattering partners can be found (e.g. $\chi^1 + \chi^2 \rightarrow \chi^1 + \chi^2$). Also, given that all particles are initially in the excited state, the number of scattering events for the $\chi^2 + \chi^2 \rightarrow \chi^2 + \chi^2$ channel is the highest despite having a smaller cross section than the mixed channel. The thick solid lines show the total number of scatters summed over all channels. These are nearly the same for the elastic and inelastic cases in the $\chi^2_{\mathrm{init}} = 50\%$ configuration, whereas there is a clear difference in the $\chi^2_{\mathrm{init}} = 100\%$ case. Again, this is caused by the removal of ground state particles following de-excitation. We note that for all cases, on average, only a few scatters per particle occur in a Hubble time within the inner halo. This number is nearly constant within the central $\sim 20\,\mathrm{kpc}$ of the halo. Beyond $\sim 100\,\mathrm{kpc}$, the number of scattering events drops below one rapidly.

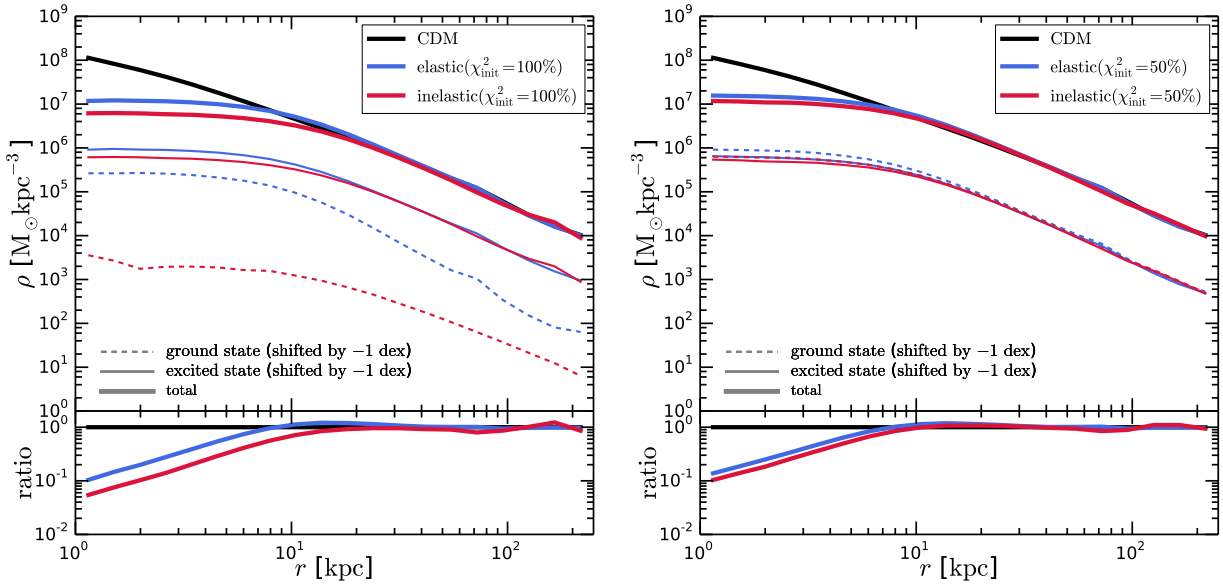

Figure 3.24: Radial density profiles for the Milky Way-sized halo split in the two-state populations. The inelastic self-interacting DM models lead to larger cores with lower central densities compared to the elastic self-interacting DM models. This is caused by the removal of ground state particles after exothermic reactions, which is largest for the $\chi^2_{\rm init} = 100\%$ case, where more energy is available for release. For the inelastic case the total density almost never exceeds the CDM density at intermediate radii. This is distinct from the typical behaviour of elastic self-interacting DM models, where a small density enhancement over the CDM density is always observed due to the redistribution of particles from the inner to the intermediate regions. The ground state density profile for the $\chi^2_{\rm init} = 100\%$ configuration is very different for the elastic and inelastic cases due to the removal of de-excited particles from the halo centre. For the $\chi^2_{\rm init} = 50\%$ configuration, on the other hand, the profiles are very similar since the de-excited particles represent only a small fraction of the total ground state population. The bottom panels show the ratio of the total density profiles relative to CDM.

The net impact of energy release due to de-excitations is seen in Fig. 3.24 where we show the spherically averaged total radial density profiles (thick lines) for the CDM (black), elastic (blue) and inelastic (red) self-interacting DM models. As in previous figures, on the left (right) we show the $\chi^2_{\rm init} = 100\%$ ($\chi^2_{\rm init} = 50\%$) case. One can immediately see that the inelastic case leads to a stronger depletion of the central density (a larger density core) than the elastic case. We stress that the elastic and inelastic cases have exactly the same reaction channels and cross sections, with the only difference being the energy release during down-scattering in the inelastic case. An interesting implication of this result is that it is possible in inelastic models to create a core of the same size and density as in the elastic case but with a smaller scattering cross section. This is an important point since it indicates that in the inelastic case, a wider range of cross section normalizations could significantly modify the inner DM density while remaining consistent with other constraints, compared with the range preferred by elastic self-interacting DM simulations.

The thin lines in Fig. 3.24 show the density profiles of the individual states, ground state (χ^1) and excited state (χ^2) with solid and dashed lines, respectively. For clarity, we have shifted those profiles down by one dex, relative to the total profiles. As anticipated, the density profile for the ground state in the inelastic $\chi^2_{\rm init} = 100\%$ case (left panel, dashed red line) is strongly suppressed. The average density of the ground state is more than two orders of magnitude lower than in the elastic case, where particles are not kicked out from the halo. This is not the case for the $\chi^2_{\rm init} = 50\%$ configuration, where the profiles for the different state populations are rather similar and nearly the same between the elastic and inelastic cases.

The lower panels in Fig. 3.24 present the actual density reduction compared to the CDM case for the four different self-interacting DM scenarios. Elastic collisions alone reduce the central density at 1 kpc already by an order of magnitude compared to the CDM case. As stated, the strongest reduction is for the inelastic $\chi^2_{\rm init} = 100\%$ model where by $z = 0$, the density is lower by nearly a factor of ~ 25 at 1 kpc. Comparing the elastic-CDM and inelastic-CDM ratio profiles in more detail, it becomes also clear that, contrary to the elastic case, the inelastic case does not show an enhancement over the CDM density at intermediate radii (for the $\chi^2_{\rm init} = 100\%$ case). This is because in inelastic models the particles are not only redistributed within the halo due to the effective inside-out heat transport caused by elastic scattering, but they can, at least for our benchmark model, also be removed, and not contribute to the halo density profiles anymore. This is a distinctive signature between elastic and inelastic self-interacting DM models since all elastic models that produce a core also lead to such a density enhancement at intermediate radii. We finally note that the impact of inelastic down-scattering in the halo density profiles strongly depends on the initial level population of the excited state. For the 50% case (right panel of Fig. 3.24), we see only a rather small effect compared to the elastic case.

The total velocity dispersion profile for the $\chi^2_{\rm init} = 100\%$ initial configuration is presented in Figure 3.25. An isothermal core of very similar characteristics is formed in both the elastic and inelastic self-interacting DM models, the key differences are that the inelastic case has a colder core (to compensate for the lower central density) and a population of ground state particles that is essentially unbound, moving with large velocities due to the energy injection during down-scattering.

We expect that inelastic self-interacting DM models have a stronger impact on the abundance

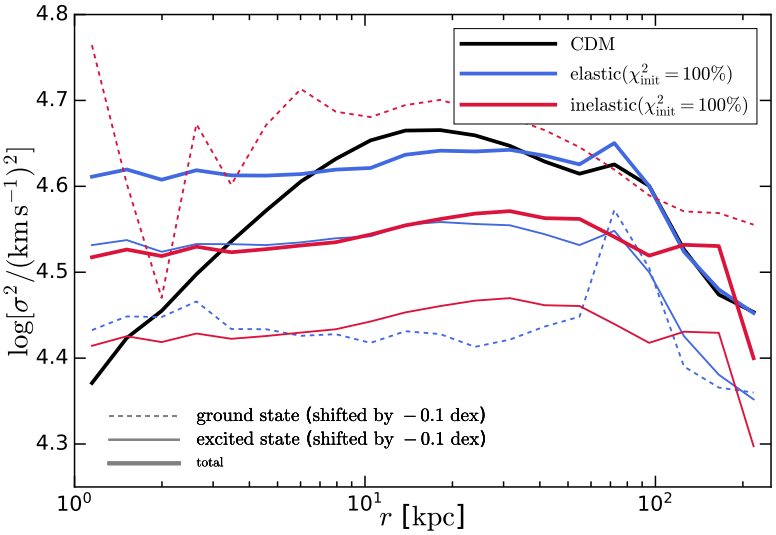

Figure 3.25: Total velocity dispersion profiles for the Milky Way-sized halo split in the two-state populations. Both elastic and inelastic self-interacting DM lead to isothermal profiles in the centre. The ground state velocity dispersion is signficantly higher for the inelastic model due to the injected kinetic energy gained during the level decay.

of subhaloes compared to purely elastic models. In fact, for elastic models a quite large cross section on galactic scales ($\sim 10\,\mathrm{cm^2\,g^{-1}}$) is needed to create a relevant difference relative to CDM. We present the subhalo velocity ($V_{\rm max}$) function for our MW-size simulations in the top panels of Fig. 3.26. Given the relatively large cross sections of our benchmark model, we already see a mild impact on the abundance of subhaloes for the elastic case. However, the effect is much larger for the inelastic case. Specifically, we find that for $V_{\rm max} \gtrsim 30\,\mathrm{km\,s^{-1}}$ ($V_{\rm max} \gtrsim 20$ km/s) the number of subhaloes is reduced by ~ 3 (~ 4) for the inelastic model with $\chi^2_{\rm init} = 100\%$. For the $\chi^2_{\rm init} = 50\%$ configuration we find a smaller effect, roughly reduced by a factor of 2.

In the bottom panel of Fig. 3.26 we show the median (solid lines) and 1σ region (shaded areas) of the distribution of density profiles of the ten most massive subhaloes at $z = 0$. Similarly to the main halo density profile, we find that inelastic self-interacting DM models lead to subhaloes with larger cores and smaller densities compared to the elastic case. We note however, that the effect of inelastic collisions is larger for these smaller systems with lower velocity dispersion than the main MW halo since the cross section for down scattering strongly increases towards lower relative velocities as shown in the bottom panel of Fig. 3.17, where the cross section exceeds the equal state elastic cross sections for relative velocities below $\sim 10\,\mathrm{km\,s^{-1}}$.

For the $\chi^2_{\rm init} = 100\%$ ($\chi^2_{\rm init} = 50\%$) configuration we find that the core density is reduced by ~ 4 (2) compared to the elastic case. Therefore, the inelastic self-interacting DM cross sections can be smaller by a factor of a few compared to the elastic case while creating a core of similar size and density. We quantify this reduction in Fig. 3.27, where we compare the elastic and inelastic model for the $\chi^2_{\rm init} = 100\%$ initial configuration with an elastic model, where we have increased the

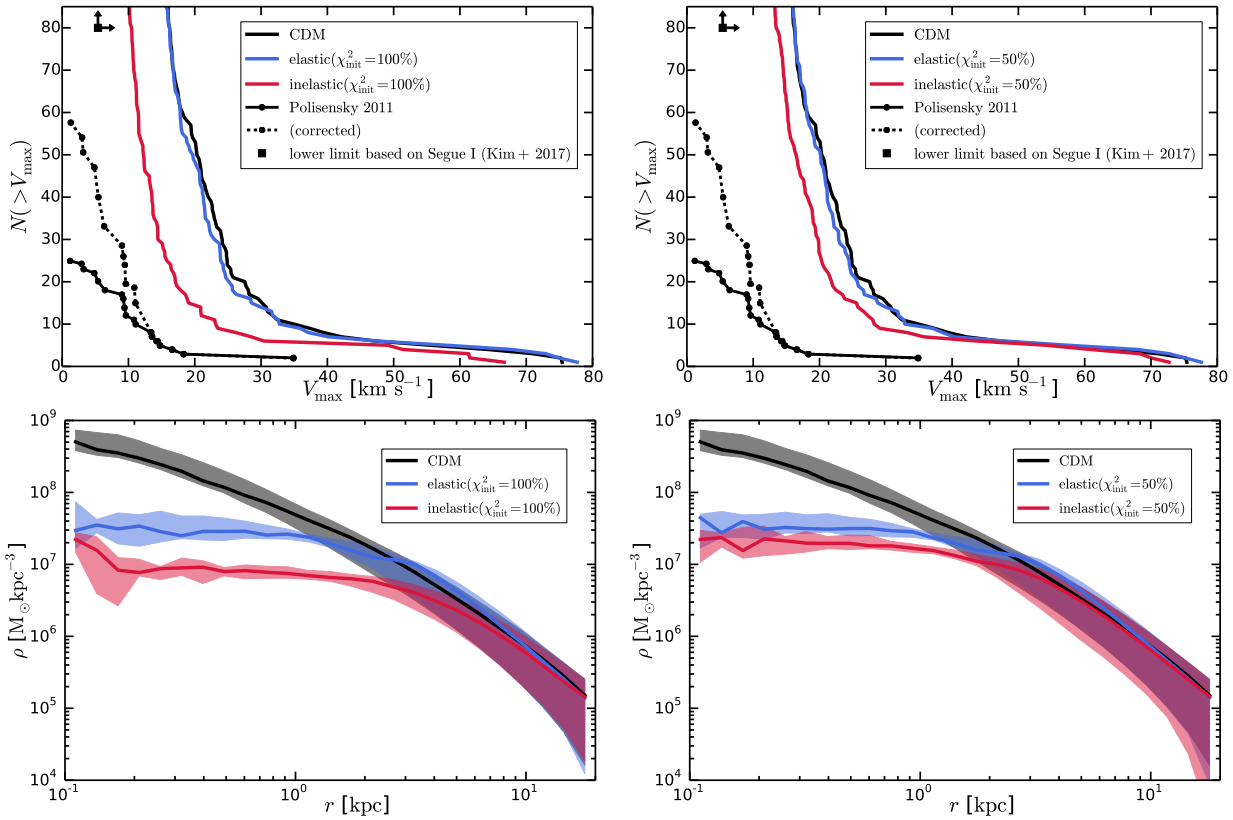

Figure 3.26: Abundance and inner structure of subhaloes in CDM and (in)elastic self-interacting DM models. *Top panels:* Cumulative subhalo velocity ($V_{\rm max}$) function for the different models and observed satellites of the Milky Way including a sky coverage correction [327]. Despite the relatively large cross sections (see bottom panel of Fig. 3.17), the elastic model leads only to a minor reduction in subhalo abundance relative to CDM. A substantially larger effect is visible for the inelastic case. For $V_{\rm max} \gtrsim 30\,{\rm km\,s^{-1}}$ ($V_{\rm max} \gtrsim 20\,{\rm km\,s^{-1}}$) the number of subhaloes is reduced by ~ 3 (~ 4) for the inelastic model with $\chi^2_{\rm init} = 100\%$. *Bottom panels:* Subhalo density profiles. The solid lines show the median profile of the ten most massive subhaloes at $z = 0$, while the shaded region indicates the 1σ scatter of the distribution. Inelastic self-interacting DM leads to significantly larger and lower density cores compared to the elastic case. For the $\chi^2_{\rm init} = 100\%$ configuration the core density is reduced by ~ 4 compared to the elastic case. This implies that previous estimates on cross sections within elastic self-interacting DM simulations based on certain requirements for core densities, are altered in the presence of inelastic self-interacting DM reactions. We also note that such an inelastic self-interacting DM model will not suffer from the gravothermal catastrophe if the elastic scattering cross sections are sufficiently small.

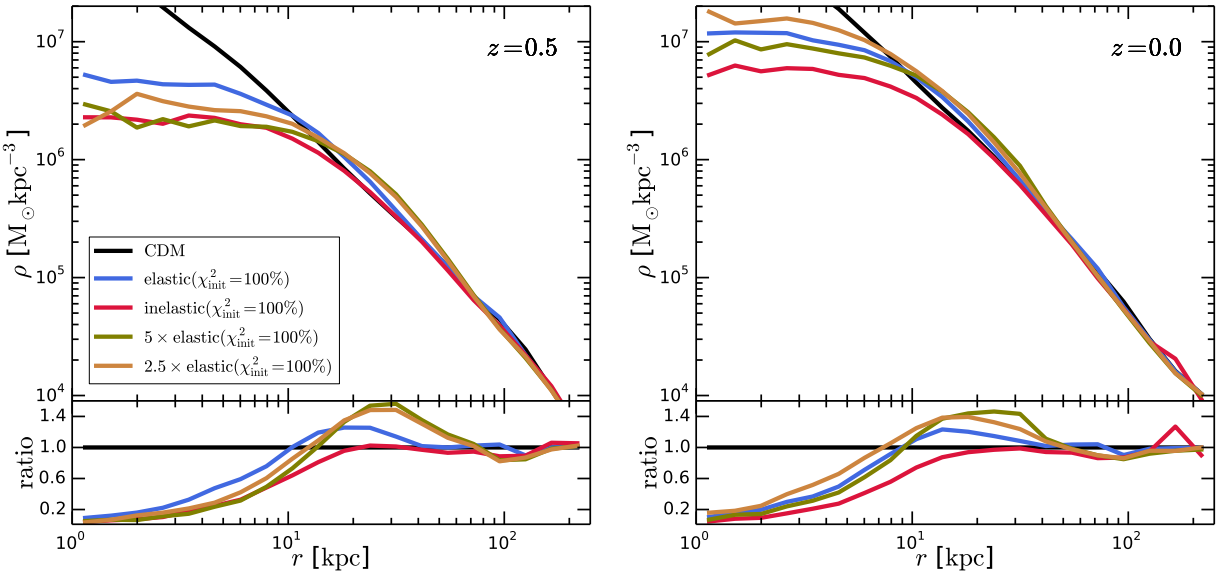

Figure 3.27: Radial density profiles for the Milky Way-sized halo for scaled elastic cross section models. An elastic model with a 5 times larger cross section leads to a central density reduction similar to the inelastic model at $z = 0.5$ (*left panel*). The scaled elastic models core collapse at later times, but not the inelastic model (*right panel*). Inelastic models are therefore more efficient in core formation, and furthermore avoid the core collapse phase of elastic models. This implies that current constraints on the normalization of elastic cross sections need to be revised if inelastic scatterings are the dominant mode of self-interactions. These simulations were performed at level-3 resolution, which is sufficiently converged.

cross section normalization by factors of 2.5 and 5. These simulations were performed at level-3 resolution, which is sufficiently converged for this test. The figure shows that the central density reduction for the main halo density profile is comparable for the inelastic model and the elastic model with a five times larger cross section (left panel). This implies that previous conclusions obtained from elastic self-interacting DM simulations on the requirements for core sizes and densities given a certain cross section normalization, are strongly altered in the presence of inelastic scattering. We also note that inelastic self-interacting DM models will not suffer from the gravothermal catastrophe, as long as down-scattering and the resultant energy release is not suppressed. This is demonstrated in the right panel, where at $z = 0$ the scaled elastic models already started to core-collapse, which does not occur for the inelastic model. This absence of core collapse is distinctly different from elastic self-interacting DM models, where the runaway collapse of the core is unavoidable on a time scale that depends on the normalization of the cross section [328]. As for the main halo density profile, we find that the inelastic model does not lead to a density enhancement relative to the CDM case at intermediate radii. As we have argued, this is due to the expulsion of de-excited ground state particles from the halo.

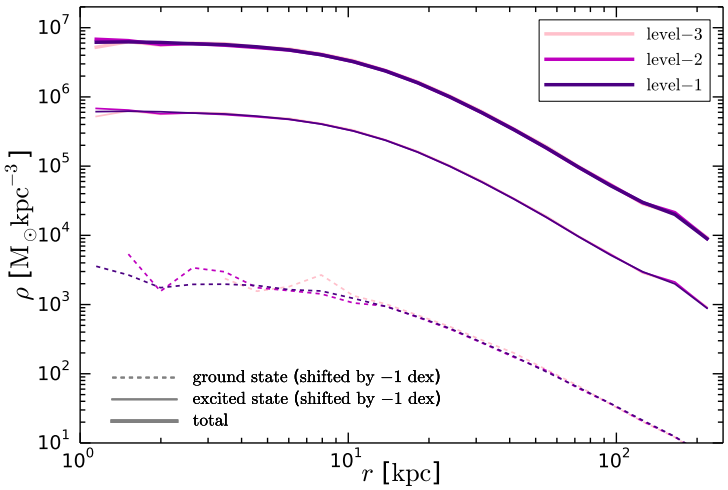

Figure 3.28: Numerical convergence of inelastic self-interacting DM simulations. Density profiles for total density and state density split by ground and excited state for three different resolution levels of the $\chi^2_{\rm init} = 100\%$ initial configuration.

Finally, we remark that the formation of a core in the inelastic case is a combination of the redistribution of energy following elastic scatterings and loss of DM particles during inelastic down-scattering. The degree to which effect is more important depends on the primordial fraction of particles in the excited state, and the hierarchy of the velocity-dependent cross sections of the different scattering channels. A substantial mass loss in the halo centre is a distinct feature of inelastic models relative to elastic self-interacting DM, and it might be of particular relevance for constraining the (inelastic) energy deposition. In elastic self-interacting DM for instance, the mass redistribution in the halo due to elastic scattering can be reversed if the galaxy within is compact enough (e.g. [329]) or if the gravothermal collapse phase has been triggered. This reversal has been invoked to argue that in elastic self-interacting DM, not all haloes are expected to be cored, and in fact the natural diversity of the inner DM densities reflected in the rotation curves of dwarf galaxies [238] is to be expected for elastic self-interacting DM with $\sigma_{\rm T}/m \sim 1\,{\rm cm}^2\,{\rm g}^{-1}$ [264]. This reversal of cored profiles into dense profiles would have only a limited extent in the case of inelastic self-interacting DM since the central mass loss is irreversible and the gravothermal collapse cannot be triggered. Thus, it will be quite relevant to explore the interplay of inelastic self-interacting DM and baryonic physics to check under which conditions they can develop cuspy profiles, which is a promising avenue to constrain inelastic self-interacting DM.

Numerical Convergence

We test the numerical convergence of the inelastic self-interacting DM implementation in Figure 3.28. We present the density profiles for total density and state density split in ground and excited state for three different resolution levels of the $\chi^2_{\rm init} = 100\%$ initial configurations. The

three resolution levels differ each by factors of 8 in mass resolution, and a factor of 2 in softening length. Level-1 corresponds to the highest resolution that has been employed for the paper. We find good convergence of the inelastic self-interacting DM implementation.

3.6 Discussion

In this chapter we have derived a semi-analytic approximation for the scattering cross sections for DM interacting via an off-diagonal dark Yukawa potential. This physical model demonstrates several new features relative to a system with purely elastic scattering, and we have presented simple analytic forms for the upscattering and downscattering rates in the nonperturbative resonant regime. Within the regime of validity of our approximations, we readily obtain cross sections that appear large enough to modify the inner regions of dwarf halos or give rise to interesting signatures in indirect DM searches.

Our approximations are valid when $v \lesssim m_\phi/m_\chi \lesssim \alpha$ and (for particles initially in the excited state) $\delta \lesssim m_\phi^2/m_\chi \lesssim \alpha^2 m_\chi$, where m_χ is the DM mass and m_ϕ the mass of the mediator. Within this regime, the scattering cross section exhibits resonant enhancement at particular values of the parameter $m_\phi/(\alpha m_\chi)$. For particles initially in the excited state, or scatterings of ground-state particles with energies above the excitation threshold, the resonance positions are very similar to those in the case with no mass splitting: however, for scatterings in the ground state below the excitation threshold, the resonances are shifted by the presence of the mass splitting when δ becomes comparable to m_ϕ^2/m_χ. Consequently, the relationship between the elastic and inelastic scattering cross sections can be quite complex. Away from the resonances, we recover the expected geometric cross sections at low velocities: $\sigma \propto 1/m_\phi^2$ for elastic scattering, $\sigma v \propto \sqrt{\delta/m_\chi}(1/m_\phi^2)$ for downscattering. Our approximations match smoothly onto the corresponding perturbative expressions when $m_\phi \approx \alpha m_\chi$.

The incorporation of these semi-analytic scattering cross sections into numerical simulations has allowed the first detailed studies of halos containing inelastically scattering DM. To this end, we have performed high resolution simulations of a galactic halo within an inelastic self-interacting DM model using a newly developed self-interaction DM implementation in the `Arepo` code to study multi-state DM models. We simulated a generic inelastic model of a nearly degenerate two-state DM particle with an energy level splitting of 10^{-6} with respect to its ground state, and an interaction cross section of a few $\mathrm{cm}^2\mathrm{g}^{-1}$ on the scale of dwarf galaxies.

We demonstrated that the physics of inelastic self-interacting DM leads to interesting new effects, which are distinctly different or absent in elastic self-interacting DM models. Most importantly, DM particles can be removed from potential wells, which causes large core formation with lower densities compared to elastic models for the same interaction cross section. Furthermore, inelastic collisions also lead to subhalo evaporation reducing efficiently the abundance of subhaloes. Our main findings are:

- Inelastic self-interacting DM creates larger and lower density cores compared to elastic models for the same cross section. Specifically, a ~ 5 times larger elastic cross section would

be required to achieve the same density reduction and core formation as the inelastic model. The exact factor depends on the underlying inelastic DM model. This implies that currently existing constraints on self-interaction cross sections have to be revised if inelastic reactions are taken into account.

- Similar to models with a power spectrum cutoff, we find that inelastic self-interacting DM can reduce the amount of substructure for rather small interaction cross sections. The reduction of the subhalo abundance is due to gravitational particle unbinding such that for $V_{\rm max} \gtrsim 30\,{\rm km\,s^{-1}}$ ($V_{\rm max} \gtrsim 20\,{\rm km\,s^{-1}}$) the number of subhaloes is reduced by ~ 3 (~ 4) compared to the cold DM case.

- The de-excitation of the upper level to the ground state can inject the energy equivalent of $\mathcal{O}(100)$ million Type II supernovae in Milky Way-like haloes.

- The gravothermal catastrophe that can occur in elastic self-interacting DM models, can be avoided in inelastic self-interacting DM models. This is possible because of the unbinding of particles that de-excite during the scattering process. If this process dominates over elastic scattering, a central density increase is not possible anymore since the DM mass is efficiently removed.

- The virial mass of a Milky Way-like halo is reduced by about 10% compared to the cold DM case for inelastic self-interacting DM.

- Density profiles for elastic and inelastic self-interacting DM models are distinctly different. Elastic models lead to an increased DM density at intermediate radii caused by particles being transferred from the core region to outer parts. This effect is not present for inelastic models that can unbind particles. For such models the density profile follows the CDM prediction outside of the core region.

We conclude that inelastic self-interacting DM models have some properties, like the removal of substructure and enhanced core formation, that are absent in typically studied elastic self-interacting DM models. Those features can have a significant impact on properties of Milky Way-like DM haloes and can address many of the small-scale challenges of the cold DM paradigm. Furthermore, inelastic self-interacting DM models are a rather generic feature of many DM particle physics models beyond the canonical WIMPs, which have still not been detected. It is therefore high time to consider also alternative DM particles, like inelastic self-interacting DM, and understand how structure formation progresses in those models. In the future it will be interesting to study these inelastic models with baryonic galaxy formation models (e.g. [330, 331]) to understand their impact on galaxy formation in more detail. Furthermore, we also expect changes in the detailed DM phase-space structure (e.g. [315, 332]) due to the energy injection and particle removal in inelastic self-interacting models.

Chapter 4

Strongly Interacting DM and an Axion Portal

4.1 Introduction

As discussed in Chapter 1, DM might be a thermal relic from the early Universe. The most well-studied thermal scenario is that DM is comprised of WIMPs. The number density of WIMPs is set by $2 \to 2$ annihilations of the DM into Standard Model (SM) particles, and the observed DM relic abundance is achieved when both the DM mass and coupling to SM particles are near the scales relevant for electroweak processes.

An alternative thermal setup was proposed in Ref. [147] where $3 \to 2$ DM self-interactions set its abundance. In this scenario, the observed relic density indicates that the DM mass and self-coupling should be near the strong scale. This mechanism of strongly interacting massive particles (SIMPs) was shown to be generic in strongly coupled theories of chiral symmetry breaking, where the pions play the role of DM [333]. The $3 \to 2$ interactions are then sourced by the well known Wess-Zumino-Witten (WZW) action [334–336]. This provides a simple and calculable realization of the SIMP mechanism, although by no means the only one [337–347].

In addition to providing a novel thermal mechanism for explaining the DM abundance, SIMPs also offer a possible explanation for issues related to small-scale structure formation (see discussion in Chapter 3). To summarize, observed DM subhalos tend to be less dense than in simulations (see Ref. [348] for a recent review). While many of these issues may be resolved with better understanding of astrophysical processes (for instance in Ref. [349]), it is also possible to mitigate these issues if the DM can self-scatter (see Ref. [150] for a recent review). The strong self-annihilations of SIMP DM imply that their self-scatterings are also large, such that they naturally address these small-scale puzzles [147, 333].

The $3 \to 2$ DM annihilations would raise the temperature of the residual DM due to conservation of comoving entropy. Therefore, the DM must be in thermal equilibrium with a heat sink, such as the SM bath, until after freeze-out [147]. Otherwise, the $3 \to 2$ DM annihilations would cause the steady depletion of DM particles and heating of the remaining DM, a scenario referred to as

cannibalization. While cannibalization was originally proposed to provide a class of DM models intermediate between hot and cold DM, such models are not observationally-viable [350, 351]. Obtaining the observed DM abundance inevitably leads to an unacceptable washout of small-scale structure.

To allow for adequate thermalization between the DM and the SM, Refs. [339, 340] explored the kinetically-mixed hidden photon portal. Here, we explore the possibility of a pseudoscalar portal using axion-like particles to accomplish the entropy transfer to photons. For brevity, we refer to axion-like particles simply as "axions" throughout the Chapter. We note that Ref. [352] also considered an axion portal, but focused on the regime where semi-annihilations set the relic abundance. In contrast, we focus on the SIMP regime where $3 \to 2$ annihilations determine the relic density. For concreteness, we will use the SIMPlest pion realization of the DM based on an Sp($2N_c$) gauge theory with four doublet Weyl fermions following Ref. [333]. Sp($2N_c$) gauge groups with a larger number of flavors or SU(N_c) and SO(N_c) gauge groups allow for semi-annihilations which can control the relic abundance, although there may still be parameter space where $3 \to 2$ annihilations determine the DM density.

The rest of this Chapter is organized as follows. In Section 4.2, we describe the framework and identify the interactions responsible for setting the correct DM relic abundance and cooling the DM via the axion portal. In order to cool the DM effectively, the axion must be in thermal equilibrium with both the DM and the SM; the general equations governing these conditions are introduced in Section 4.3. In Section 4.4, we illustrate the theoretical and empirical requirements on the coupling strength between pions and axions, while in Section 4.5 we do the same for the coupling strength between axions and the SM bath of photons. Concluding remarks and discussions follow in Section 4.6.

4.2 The framework

Our starting point is an Sp($2N_c$) gauge theory with $2N_f$ Weyl fermions that couple to an axion-like field a as

$$\mathcal{L}_{aq} = -\frac{1}{2}m_a^2 a^2 - \left(\frac{1}{2}m_Q e^{ia/f_{a\pi}} J^{ij} q_i q_j + \text{ h.c.}\right) \tag{4.1}$$

where m_a is the axion mass, m_Q is the quark mass matrix, q_i are the confining quarks and J is the Sp($2N_f$) group invariant.[1] Upon dynamical chiral symmetry breaking, the ground state is expected to be given by

$$\langle q_i q_j \rangle = \mu^3 J_{ij}. \tag{4.2}$$

Any transformation by the flavor symmetry $V \in \text{SU}(N_f)$ would also be a ground state, and in general

$$\langle q_i q_j \rangle = \mu^3 (V J V^T)_{ij}. \tag{4.3}$$

[1] Note that in terms of a 4-component spinor $\psi^T = (q \quad q^\dagger)$, the identities $i\bar{\psi}\gamma_5\psi = -iqq + iq^\dagger q^\dagger$ and $\bar{\psi}\psi = qq + q^\dagger q^\dagger$ hold.

Switching the description to the chiral Lagrangian, a spacetime-dependent flavor rotation gives the low-energy excitations,

$$\langle qq\rangle \to \mu^3\Sigma\,, \quad \Sigma \equiv VJV^T\,, \quad V = \exp(i\pi/f_\pi)\,, \tag{4.4}$$

where $\pi \equiv \pi^b T^b$, T^b are the Sp($2N_f$) generators and f_π is the pion decay constant. We use the normalization Tr $T^bT^c = 2\delta^{bc}$ for the generators. In terms of the pion fields,

$$\begin{aligned} -\tfrac{i}{2}m_Q J^{ij}q_iq_j + h.c. \quad &\Rightarrow \quad \tfrac{m_\pi^2}{6f_\pi}\,\text{Tr}\,\pi^3 + \mathcal{O}(\pi^5)\,, \\ -\tfrac{1}{2}m_Q J^{ij}q_iq_j + h.c. \quad &\Rightarrow \quad -\tfrac{m_\pi^2}{4}\,\text{Tr}\,\pi^2 + \mathcal{O}(\pi^4)\,. \end{aligned} \tag{4.5}$$

The theory has an SU($2N_f$)/Sp($2N_f$) flavor structure, where the residual Sp($2N_f$) is exact due to the quark masses' proportionality to J. For $N_f \geq 2$, the fifth homotopy group of the coset space is non-vanishing and the WZW term exists [334–336],

$$\mathcal{S}_{\text{WZW}} = \frac{-iN_c}{240\pi^2}\int \text{Tr}\,(\Sigma^\dagger d\Sigma)^5\,. \tag{4.6}$$

Generally, both a Tr (π^3) and a^2 Tr (π^2) terms can appear in the interaction Lagrangian. However, the former introduces semi-annihilations of pions into a pion and an axion which might contribute to determining the relic abundance of the DM. Here, we are interested in exploring the role of an axion mediator in the SIMP mechanism of $3 \to 2$ self-annihilations of pions. Consequently, we focus on an Sp($2N_c$) gauge theory with $N_f = 2$ fermions, where the flavor symmetry is SU(4)/Sp(4) and $N_\pi = (N_f - 1)(2N_f + 1) = 5$ pions emerge. In this theory, the semi-annihilation process is absent since Tr $(\pi^3) = 0$, and pure $3 \to 2$ annihilations of pions via the WZW term are guaranteed to control the relic abundance of DM. For more flavors, or for other gauge groups, $3 \to 2$ annihilation may still control the relic abundance, though in a smaller region of parameter space. To leading order in pion fields, the WZW term for our choice of gauge group takes the form

$$\mathcal{L}_{\text{WZW}} = \frac{2N_c}{15\pi^2 f_\pi^5}\epsilon^{\mu\nu\rho\sigma}\text{Tr}[\pi\partial_\mu\pi\partial_\nu\pi\partial_\rho\pi\partial_\sigma\pi] = \frac{8N_c}{15\pi^2 f_\pi^5}\epsilon^{\mu\nu\rho\sigma}\epsilon_{abcde}\pi^a\partial_\mu\pi^b\partial_\nu\pi^c\partial_\rho\pi^d\partial_\sigma\pi^e. \tag{4.7}$$

The excess kinetic energy generated in the dark sector from $3 \to 2$ annihilations needs to be transferred out, which can be obtained through kinetic coupling of the pions to the axions and the axions to the SM bath. Since the semi-annihilation term is absent for our flavor group of choice, the interaction Lagrangian between pions and axions is

$$\mathcal{L}_{a\pi} \supset \frac{\kappa}{4}a^2\pi^b\pi^c\delta^{bc}\,. \tag{4.8}$$

If the axion coupling to the pions arises in a similar manner to what occurs in QCD, as in Eq. (4.1), the mass term for the hidden quarks q in the Sp($2N_c$) gauge theory gives rise to an axion potential:

$$\begin{aligned} \mathcal{L}_{a\pi} &= -\frac{1}{2}m_a^2a^2 - \frac{1}{2}m_Q\mu^3 e^{ia/f_{a\pi}}\,\text{Tr}J\Sigma + \text{ h.c.} \\ &= -\frac{1}{2}\left(m_a^2a^2 + \frac{2m_\pi^2f_\pi^2}{f_{a\pi}^2}\right)a^2 + \frac{m_\pi^2}{8f_{a\pi}^2}a^2\,\text{Tr}\pi^2 + \cdots \end{aligned} \tag{4.9}$$

where m_π is the pion mass. Using the normalization of $\mathrm{Tr}\pi^2 = 2\pi^b\pi^c\delta^{bc}$, we identify the Feynman rule for the $aa\pi^b\pi^c$ vertex in Eq. (4.8) as

$$i\kappa\delta^{bc} = i\frac{m_\pi^2}{f_{a\pi}^2}\delta^{bc}\,. \tag{4.10}$$

Meanwhile, the interaction Lagrangian between the axions and SM photons is

$$\mathcal{L}_{a\gamma} = \frac{1}{4f_{a\gamma}} aF^{\mu\nu}\tilde{F}_{\mu\nu}\,. \tag{4.11}$$

As long as the kinetic equilibrium between the pions and the SM is maintained through the interactions of Eqs. (4.9) and (4.11), the preferred mass for the DM is $m_\pi \approx 300$ MeV [333] with $m_\pi \sim 2\pi f_\pi$ to set the observed relic abundance. We find below that viable axion masses are around the same scale, $10\text{ MeV} \lesssim m_a \lesssim 1$ GeV. Couplings that satisfy $m_\pi \sim 2\pi f_\pi$ correspond to the strongly-interacting regime of the theory, where self-interactions are important on astrophysical scales. In this regime, $\mathcal{O}(1)$ corrections to perturbative results are expected, and therefore should be thought of as a proxy for the scales involved. Phenomenologically interesting pion masses lie at the edge of perturbativity, where higher order corrections and vector meson effects can impact the range of observationally-viable pion masses [347, 353].

4.3 Early Universe Thermodynamics

As mentioned in Section 4.1, if the DM is in a secluded sector the $3 \to 2$ interactions that set the relic abundance of DM will also raise the DM temperature due to conservation of entropy. To ensure that the DM stays relatively cold, there must be a heat sink, which can be provided by the SM thermal bath. The transfer of entropy while the comoving DM number density is changing can be described by Boltzmann equations that govern the evolution of the phase space $f_X(p,t)$ for particle X,

$$\frac{\partial f_X}{\partial t} - H\frac{p^2}{E}\frac{\partial f_X}{\partial E} = C[f_X] \tag{4.12}$$

where the left hand side is the relativistic Liouville operator in a Friedmann-Robertson-Walker spacetime and $C[f_X]$ is the collision term. In the regime where $2 \to 2$ annihilations are negligible, the relevant terms which appear in the collision term are the $3 \to 2$ interactions which set the relic abundance, the DM self-interactions which give it a thermal distribution, the elastic interactions which transfer energy between the pions and axions, and decays (and inverse decays) of axions to SM particles. We neglect axions converting to photons (and vice versa) via t-channel scattering off electrons, which is less efficient than decays (and inverse decays) at thermalizing the axions with the SM. In the parameter space of interest (with the exception of axions in the out-of-equilibrium decay scenario), all particles will be interacting sufficiently frequently so that they have thermal distributions,

$$f_X = \frac{1}{e^{(E-\mu)/T} \pm 1}, \tag{4.13}$$

where the $+$ sign ($-$ sign) is for Fermi-Dirac (Bose-Einstein) statistics and μ is the chemical potential. At temperatures below the mass of a given particle, the effects of quantum statistics become negligible and the Maxwell-Boltzmann (MB) distribution is recovered.

To solve the Boltzmann equations, it is most useful to look at two moments of the phase space distribution, which correspond to the number density and energy density

$$n_X = g_X \int đ^3p\, f_X \qquad\qquad \rho_X = g_X \int đ^3p\, E\, f_X, \tag{4.14}$$

where g_X is the number of degrees of freedom and$đ^3p \equiv d^3p/(2\pi)^3$. The Boltzmann equations for axion-mediated SIMPs are

$$\frac{\partial n_\pi}{\partial t} + 3Hn_\pi = -\langle\sigma_{3\to2}v^2\rangle\left(n_\pi^3 - n_\pi^2 n_\pi^{\text{eq},T_\pi}\right), \tag{4.15}$$

$$\frac{\partial \rho_\pi}{\partial t} + 3H\left(\rho_\pi + p_\pi\right) = -\langle\sigma_{\text{el.}}v\delta E\rangle n_\pi n_a, \tag{4.16}$$

$$\frac{\partial n_a}{\partial t} + 3Hn_a = -\left(\langle\Gamma\rangle_{T_a} n_a - \langle\Gamma\rangle_{T_{\text{SM}}} n_a^{\text{eq},T_{SM}}\right), \tag{4.17}$$

$$\frac{\partial \rho_a}{\partial t} + 3H\left(\rho_a + p_a\right) = \langle\sigma_{\text{el.}}v\,\Delta E\rangle n_\pi n_a - m_a\Gamma\left(n_a - n_a^{\text{eq},T_{SM}}\right). \tag{4.18}$$

where p_X is the pressure densities of species X (which is related to the energy density through the equation of state w_X), and $n_X^{\text{eq},T}$ is the thermal equilibrium density for species X at temperature T. Additionally, $\langle\Gamma\rangle_T = \Gamma\,\langle m_a/E_a\rangle_T$ is the thermally averaged decay rate of the axion at temperature T, $\langle\sigma_{3\to2}v^2\rangle$ is the thermally averaged $3\to2$ cross-section of the pions for this choice of gauge group, labeled $i = 1\ldots5$,

$$\langle\sigma_{3\to2}v^2\rangle = \frac{1}{3!2!n_\pi^3}\int\left(\prod_{i=1}^{5}\frac{đ^3p_i}{2E_i}\right) f_1f_2f_3\,|\mathcal{M}_{123\to45}|^2\,(2\pi)^4\delta^{(4)}\left(\sum p\right) = \frac{6N_c^2}{\sqrt{5}\pi^5}\frac{m_\pi^3T_F^2}{f_\pi^{10}} \tag{4.19}$$

and $n_\pi n_a\langle\sigma_{\text{el.}}v\delta E\rangle$ is the energy transfer rate between the pions and axions (with the initial and final states labeled as 1 and 2),

$$n_\pi n_a\langle\sigma_{\text{el.}}v\,\delta E\rangle = \int\frac{đ^3p_{\pi_1}}{2E_{\pi_1}}\frac{đ^3p_{a_1}}{2E_{a_1}}\frac{đ^3p_{\pi_2}}{2E_{\pi_2}}\frac{đ^3p_{a_2}}{2E_{a_2}}\,\delta E_\pi\, f_{\pi_1}f_{a_1}\,|\mathcal{M}_{\pi a\to\pi a}|^2\,(2\pi)^4\delta^{(4)}\left(\sum p\right). \tag{4.20}$$

For MB statistics, the equilibrium values for the number, energy, and pressure densities for a particle of mass m and temperature T with g degrees of freedom are

$$n^{\text{eq}} = g\int đ^3p\, f^{\text{eq}} = \frac{gm^2T}{2\pi^2}K_2\left(\frac{m}{T}\right) \tag{4.21}$$

$$\rho^{\text{eq}} = g\int đ^3p\, Ef^{\text{eq}} = \frac{gm^2T}{2\pi^2}\left(mK_1\left(\frac{m}{T}\right) + 3TK_2\left(\frac{m}{T}\right)\right) \tag{4.22}$$

$$p^{\text{eq}} = g\int đ^3p\frac{p^2}{3E}f^{\text{eq}} = \frac{gm^2T^2}{2\pi^2}K_2\left(\frac{m}{T}\right) \tag{4.23}$$

and the boost factor appearing in the thermally averaged decay rate is

$$\left\langle \frac{m}{E} \right\rangle_T = \frac{1}{n^{\text{eq}}} \int g \frac{d^3p}{(2\pi)^3} \frac{m}{E} f^{\text{eq}} = \frac{K_1(m/T)}{K_2(m/T)}. \tag{4.24}$$

4.4 Coupling between Axions and pions

The interaction Lagrangian in Eq. (4.9) leads to annihilations of pions into axions and to elastic scattering of axions off of pions. The SIMP mechanism requires the former process to be suppressed at the time of $3 \to 2$ freeze-out (so as not to be dominated by a WIMP process) while the latter is active (in order to transfer entropy via elastic scattering).

SIMP Mechanism Maintained

The requirement in order for the $3 \to 2$ pion self-annihilations to dominate the $2 \to 2$ annihilations of pions into axions at the time of freeze-out is

$$n_\pi \langle \sigma v \rangle_{\text{ann}} \lesssim H|_{T_F} \,, \tag{4.25}$$

where $\langle \sigma v \rangle_{\text{ann}}$ is the thermally averaged cross section for the annihilation process $\pi\pi \to aa$. The Hubble parameter at freeze-out is given by

$$H|_{T_F} = \sqrt{\frac{g_{*,F}\,\pi^2}{90}} \frac{T_F^2}{M_{\text{Pl}}}, \tag{4.26}$$

where T_F is the freeze-out temperature of $3 \to 2$ interactions (typically $T_F \sim m_\pi/20$ in the SIMP setup), M_{Pl} is the Planck mass, and $g_{*,F}$ is the effective number of relativistic degrees of freedom at the time of freeze-out. We have verified numerically that this requirement on the annihilation rate does maintain the correct relic abundance as set by the $3 \to 2$ SIMP mechanism.

The thermally averaged annihilation cross section that appears in Eq. (4.25) can be readily calculated. For a trivial matrix element, $\mathcal{M}$, of a process $12 \to 34$ (as is relevant for the Lagrangian of Eq. (4.9)) in which all states obey Maxwell-Boltzmann statistics, the thermally averaged cross section entering the Boltzmann equations [354] is expressed in terms of

$$\gamma_{12\to34} = \frac{g_1 g_2 g_3 g_4 T |\mathcal{M}|^2}{2^9 \pi^5} \int_{s_{\min}}^{\infty} ds\, \sqrt{s}\, \lambda^{1/2}(\sqrt{s}, m_1, m_2) \lambda^{1/2}(\sqrt{s}, m_3, m_4) K_1(\sqrt{s}/T)\,, \tag{4.27}$$

where g_i counts degrees of freedom for particle i,

$$s_{\min} = \max\{(m_1 + m_2)^2, (m_3 + m_4)^2\}, \tag{4.28}$$

$$\lambda(x, y, z) \equiv (1 - (y + z)^2/x^2)(1 - (y - z)^2/x^2), \tag{4.29}$$

and K_1 is the first order modified Bessel function of the second kind. The amplitude $|\mathcal{M}|^2$ which appears is averaged over *all* degrees of freedom. For pion-axion scattering and pion annihilation to axions, the relevant amplitude is therefore

$$|\mathcal{M}|^2 \equiv \frac{1}{g_\pi}\frac{1}{g_\pi}\sum_{b,c}\kappa^2\delta^{bc} = \frac{m_\pi^4}{g_\pi f_{a\pi}^4} \tag{4.30}$$

since the trace requires that the pions be the same. The thermally averaged cross section for the annihilation process $\pi\pi \to aa$ is

$$\langle\sigma v\rangle_{\text{ann}} = \frac{1}{2}\frac{\gamma_{\text{ann}}}{(n_\pi^{\text{eq}})^2}\,, \tag{4.31}$$

where n_i^{eq} denotes the number density of particle i in equilibrium,

$$n_i^{\text{eq}} = \frac{g_i}{2\pi^2}T_i m_i^2 K_2(m_i/T_i)\,, \tag{4.32}$$

where K_2 is the second order modified Bessel function of the second kind. In Eq. (4.31), the phase-space factor of $1/2$ for identical initial particles is cancelled because the number density changes by two particles per annihilation. For $m_a \ll m_\pi$, Eq. (4.31) simplifies to

$$\langle\sigma v\rangle_{\text{ann}} \approx \frac{g_a^2 m_\pi^2}{64 g_\pi \pi f_{a\pi}^4}\,. \tag{4.33}$$

Pion-Axion Kinetic Equilibrium

In addition to the suppression of the $2 \to 2$ annihilations, the SIMP mechanism requires that the rate of energy transfer in the scattering process $a\pi \to a\pi$ is fast enough to successfully cool the DM. We therefore impose the requirement that the pions and axions are in kinetic (thermal) equilibrium during the entire time over which the $3 \to 2$ process is active. This can be recast as a requirement that the thermal decoupling temperature between the two sectors is lower than the temperature at which the $3 \to 2$ process freezes out $T_D < T_F$. In this range of temperatures, the pions are guaranteed to be nonrelativistic since $T_F \sim m_\pi/20$. Therefore, the energy transfer rate can be re-written as

$$n_\pi n_a \langle\sigma_{\text{el.}} v\,\Delta E\rangle \simeq -\int \frac{d^3p_{\pi_1}}{2E_{\pi_1}}\frac{p_{\pi_1}^2}{2m_\pi}C[f_{\pi_1}]. \tag{4.34}$$

The form of the collision term in the integrand will depend on whether the axions are relativistic or not at around the time of freeze-out, as detailed below.

In the limit of small m_a (where axions are still relativistic and Bose-Einstein distributed at the time of freeze-out) there are well known methods for computing the collision term [355] which have been implemented for this type of model in Ref. [356]. For pion-axion scattering in this regime, the collision term takes the form

$$C[f_{\pi_1}] = \frac{\pi g_a^2 m_\pi^6}{360 f_{a\pi}^4}\left(\frac{T_a}{m_\pi}\right)^4\left(m_\pi T_a\,\nabla_{p_{\pi_1}}^2 + \mathbf{p}_{\pi_1}\cdot\nabla_{p_{\pi_1}} + 3\right)f_{\pi_1}(T_\pi). \tag{4.35}$$

Integrating over the pion phase space in Eq. (4.34) yields

$$n_\pi n_a \langle \sigma_{\text{el.}} v \, \Delta E \rangle = \frac{\pi g_a^2 m_\pi^5}{120 f_{a\pi}^4} \left(\frac{T_a}{m_\pi} \right)^4 (T_\pi - T_a). \tag{4.36}$$

While the $3 \to 2$ is actively depleting the number density, the pions are nonrelativistic and follow MB statistics, which means that their energy density Boltzmann equation can then be expressed as

$$\frac{\partial T_\pi}{\partial T_a} = 3 \frac{T_\pi^2}{m_\pi T_a} + \frac{\pi g_a^2 m_\pi^5}{120 f_{a\pi}^4 H|_{T=m_\pi}} \frac{T_a T_\pi^2 (T_\pi - T_a)}{m_\pi^4}. \tag{4.37}$$

The first term on the right-hand side comes from the $3 \to 2$ and causes the pion temperature to increase, while the second term comes from pion-axion elastic scattering and pushes $T_\pi \to T_a$. The second term cannot keep up with the first as the temperature drops and the pions and axions decouple. The temperature of decoupling is

$$T_D \simeq m_\pi \left(\frac{\pi g_a^2 m_\pi^5}{120 f_{a\pi}^4 H|_{T=m_\pi}} \right)^{-1/4}. \tag{4.38}$$

Further details of this derivation, including an order unity normalization factor of $4^{-1/4}\Gamma(3/4)$, can be found in Ref. [356]. We have checked numerically that this requirement on thermalization between the axions and pions does keep the DM cool.

In the regime where $T_F < m_a \lesssim m_\pi$, we generalize the approach laid out in Ref. [355], which only strictly applies when the axion is relativistic. However, similar methods are valid for particles scattering in the limit of low momentum transfer. This criterion should be met in the regime where the axion mass is still smaller than the pion mass (which kinematically enforces that the momentum transferred in a single collision is relatively small) but where the axion is sufficiently heavier than the freeze-out temperature $T_F \sim m_\pi/20$ such that it becomes Boltzmann suppressed. Most generally, the collision term for $2 \to 2$ scattering of pions and axions can be written as

$$C = \frac{1}{2} \int \frac{đ^3 p_{a_1}}{2E_{a_1}} \frac{đ^3 p_{a_2}}{2E_{a_2}} \frac{đ^3 p_{\pi_2}}{2E_{\pi_2}} (2\pi)^4 \delta^{(4)}(p_{a_1} + p_{\pi_1} - p_{a_2} - p_{\pi_2}) \, |\mathcal{M}|^2 \, J \tag{4.39}$$

where J is the relevant combination of phase space factors. In the regime of interest, everything is MB distributed at thermal decoupling so

$$J = e^{-E_{\pi_1}/T_\pi} e^{-E_{a_1}/T_a} - e^{-E_{\pi_2}/T_\pi} e^{-E_{a_2}/T_a}. \tag{4.40}$$

The collision term can be written as an expansion in the momentum transfer, $C = \sum C^j$ where

$$C^j = \frac{(2\pi)^4}{2j!} \int \frac{đ^3 p_{a_1}}{2E_{a_1}} \frac{đ^3 p_{a_2}}{2E_{a_2}} \frac{đ^3 p_{\pi_2}}{2E_{\pi_2}} \, \delta(E_{a_1} + E_{\pi_1} - E_{a_2} - E_{\pi_2}) \, |\mathcal{M}|^2 \\ J \big((\mathbf{p}_{a_2} - \mathbf{p}_{a_1}) \cdot \nabla_{p_{\pi_2}} \big)^j \delta^{(3)}(p_{\pi_1} - p_{\pi_2}). \tag{4.41}$$

In this expansion, C^0 vanishes simply because if the momentum transfer is zero and the number of a species does not change, the collision term is identically zero. For a contact interaction which has no angular dependence (as in the scenario we consider here, $|\mathcal{M}|^2 \sim$ const.), C^1 also vanishes because the angular integral contains the integrand $(\mathbf{p_{a_2}} - \mathbf{p_{a_1}}) \cdot \mathbf{p}_\pi \equiv \mathbf{q} \cdot \mathbf{p}_{\pi_1}$, which is odd over the angular domain. Therefore, the leading-order term is C^2. The momentum transfer scales like $\Delta p_a \sim (m_\pi p_a - m_a p_\pi)/(m_a + m_\pi)$, and plugging in thermal values for the typical momentum indicates that when truncated at $\mathcal{O}((\Delta p_a/p_\pi)^2)$, the expansion in momentum is accurate at the $\sim$10% level when the axion mass is $m_a \lesssim m_\pi/3$. As we will show in the following Subsections, considerations of the parameter space will restrict us to this mass range, indicating that this truncation is sufficiently accurate for the purposes of this scenario. Plugging in the matrix element of Eq. (4.30), the leading order piece of the collision term is then

$$\begin{aligned} C^2 = \frac{\pi\, m_\pi^4}{8 g_\pi (2\pi)^3 f_{a\pi}^4} \int \frac{d^3 p_{a_1}}{2E_{a_1}} \int d\Omega_2\, dE_{a_2}\, p_{a_2} \Bigg[& \frac{2(\mathbf{q}\cdot\mathbf{p}_{\pi_1})^2}{E_{\pi_1}^3} J\, \partial_{E_{a_2}}^2 \delta(E_{a_1} - E_{a_2}) \\ & + \left(\frac{q^2}{E_{\pi_1}^2} J - \frac{3(\mathbf{q}\cdot\mathbf{p}_{\pi_1})^2}{E_{\pi_1}^4} J + \frac{2(\mathbf{q}\cdot\mathbf{p}_{\pi_1})^2}{E_{\pi_1}^3 T_\pi} J' \right) \partial_{E_{a_2}} \delta(E_{a_1} - E_{a_2}) \\ & + \left(\frac{q^2}{E_{\pi_1}^2 T_\pi} - \frac{(\mathbf{q}\cdot\mathbf{p}_{\pi_1})^2}{E_{\pi_1}^3 T_\pi^2} - \frac{3(\mathbf{q}\cdot\mathbf{p}_{\pi_1})^2}{E_{\pi_1}^4 T_\pi} \right) J' \delta(E_{a_1} - E_{a_2}) \Bigg] \end{aligned} \tag{4.42}$$

$$= \frac{2 m_\pi^4 e^{-E_{\pi_1}/T_\pi} e^{-m_a/T_a} (T_a - T_\pi) T_a^3 (m_a^2 + 3 m_a T_a + 3T_a^2)}{g_\pi (2\pi)^3 f_{a\pi}^4 E_{\pi_1}^2 T_a T_\pi} \tag{4.43}$$

where $J' \equiv e^{-E_{\pi_2}/T_\pi} e^{-E_{a_2}/T_a}$. This collision term into the calculation of the energy transfer rate of Eq. (4.34) under the assumption that the pions are non-relativistic,

$$\begin{aligned} n_\pi n_a \langle \sigma v \Delta E \rangle &= \frac{6 e^{-m_\pi/T_\pi} e^{-m_a/T_a} (T_\pi - T_a) T_a^3 (m_a^2 + 3 m_a T_a + 3T_a^2) \sqrt{\pi (m_\pi T_\pi)^5}}{g_\pi \sqrt{2} (2\pi)^5 f_{a\pi}^4 T_a T_\pi} \\ &= \frac{3 m_\pi e^{-m_a/T_a} (T_\pi - T_a) T_a^2 (m_a^2 + 3 m_a T_a + 3T_a^2)}{g_\pi (2\pi)^3 f_{a\pi}^4}\, n_\pi \end{aligned} \tag{4.44}$$

By analogy to Eq. (4.37), the term in the Boltzmann equations that equalizes temperatures between the two sectors is

$$\frac{3 g_a^2 m_\pi^3 m_a^2}{(2\pi)^3 f_{a\pi}^4 H|_{T=m_\pi}} \left(\frac{T_a}{m_\pi}\right)^{-1} \left(\frac{T_\pi}{m_\pi}\right)^2 \left(\frac{T_\pi - T_a}{m_\pi}\right) e^{-m_a/T_a} \tag{4.45}$$

and decoupling happens when this is order unity. The requirement that thermal decoupling happens after pion freeze-out can be recast as a requirement on $f_{a\pi}$

$$f_{a\pi} \lesssim \left(\frac{3 g_a^2 m_\pi^3 m_a^2}{(2\pi)^3 H|_{T=m_\pi}} \left(\frac{T_F}{m_\pi}\right)^2 e^{-m_a/T_F} \right)^{1/4}. \tag{4.46}$$

Eqs. (4.46) and (4.38) do not match exactly due to a difference in numerical prefactors for BE vs. MB statistics. When that relative factor $\pi^4/90$ is taken into account, then the two match exactly.

Theoretical Considerations

In addition to the above requirements, the decay constant $f_{a\pi}$ must be greater than the cutoff scale of chiral symmetry breaking. Otherwise, the description in Eq. (4.1) breaks down. We require that $f_{a\pi} \gtrsim 2\pi f_\pi$, where f_π is determined for a given m_π from the solution to the Boltzmann equation. Since the Sp($2N_c$) gauge theory with $N_f = 2$ we discuss here points to the strongly interacting regime where $m_\pi \sim 2\pi f_\pi$, we require that $f_{a\pi} \gtrsim m_\pi$. In practice, however, suppressing $2 \to 2$ annihilations at freeze-out is always a stronger requirement.

An additional preference, though not a requirement, comes from considering how chiral symmetry breaking contributes to the axion mass in Eq. (4.9),

$$\Delta m_a^2 = \frac{2m_\pi^2 f_\pi^2}{f_{a\pi}^2} \,. \tag{4.47}$$

The natural range for the axion mass-squared is therefore where $\Delta m_a^2 \lesssim m_a^2$, such that no fine tuning is present against an unspecified negative contribution, possibly from another confining gauge theory with $\theta \approx \pi$.

Empirical requirements

Having established the theoretical requirements on the axion-SIMP parameter space, we now move to the observational constraints coming from the cosmic microwave background (CMB). In standard cold DM cosmology, the intergalactic medium (IGM) is almost entirely neutral after recombination and CMB photons free stream. If some fraction of the DM annihilates to SM particles and partially ionizes the IGM, this will cause some CMB photons to re-scatter which modifies CMB anisotropies in a characteristic way. For the scenario we consider in this work, the process of interest is $\pi\pi \to aa \to 4\gamma$. In the parameter space where there is sufficient thermalization between axions and the SM (see Section 4.5), the decay of the intermediate-state axions happens immediately. Thus, the use of the narrow-width approximation is appropriate and the cross section for this process is set by the cross section for the annihilation process $\pi\pi \to aa$. We use limits derived in Ref. [357], which are not very sensitive to whether there are two final-state photons or four.[2]

Pion-Axion Parameter Space

Satisfying the above requirements on $f_{a\pi}$ as a function of m_a for a variety of DM masses m_π yields the SIMP regions depicted in Fig. 4.1. We take $g_{*,F} = 10.75$ at freeze-out since for the DM masses we consider, freeze-out happens below the temperature of muon-antimuon annihilation. We learn that SIMP-axion thermalization is achieved over a broad range of axion masses and couplings $f_{a\pi}$. The observational upper limits from the CMB are also shown in Fig. 4.1 as a set of thick, solid lines corresponding to the different depicted pion masses. We thus find a viable SIMP-axion parameter space below the CMB curve and above the thermalization curve.

[2]This statement is based on private communication with T. R. Slatyer.

We note that elastically decoupling relic (ELDER) DM [356, 358] is obtained along the thermalization curve in Fig. 4.1. For ELDER DM, the kinetic decoupling between the DM and SM baths occurs before $3 \to 2$ pion self-annihilations freeze out. This causes the relic abundance to be exponentially sensitive to this elastic scattering while being relatively insensitive to the strength of the $3 \to 2$ pion process. On the thermalization curve in Fig. 4.1, the elastic scattering of pions off of axions dominates over the $3 \to 2$ pion self-annihilation process in setting the relic abundance.

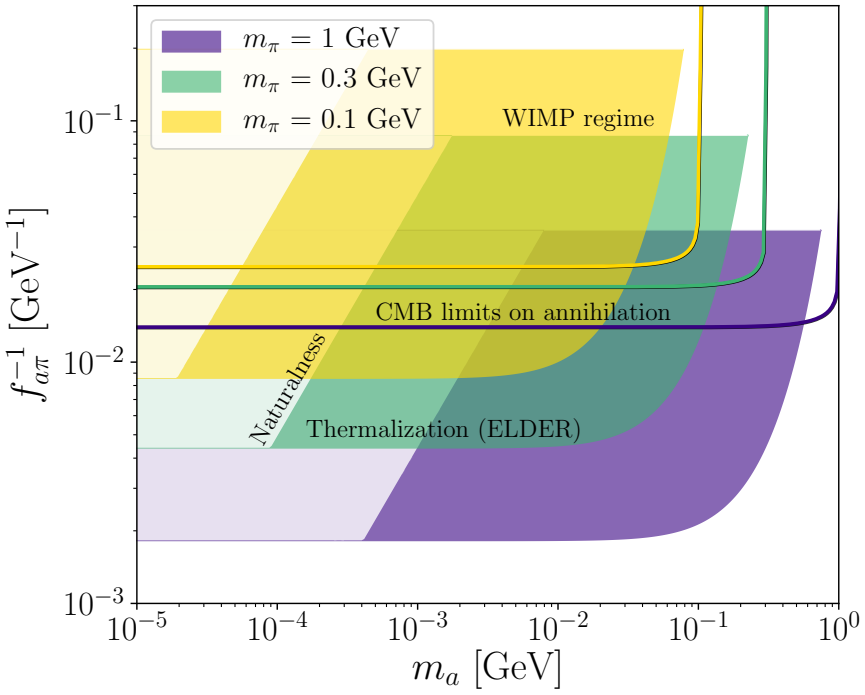

Figure 4.1: The parameter space for axions coupling to pions. The shaded regions correspond to the regions where the SIMP mechanism is theoretically viable for a given DM mass m_π. The boundaries of this region are set by requiring that the $3 \to 2$ rather than the $2 \to 2$ process sets the relic abundance (labeled as "WIMP regime") and that the pions can transfer their excess heat to the axions and hence the SM sector (labeled as "Thermalization"). Note that the parameter space along the "Thermalization" curve corresponds to the scenario where DM is an elastically decoupled relic (ELDER). We also indicate the natural mass range where the axion mass is at least as large as its contribution from chiral symmetry breaking (labeled as "Naturalness"). Also shown are the empirical upper limits on pion annihilation from energy injection into the CMB (thick solid lines labeled as "CMB limits on annihilation").

4.5 Coupling between Axions and Photons

Fig. 4.1 presents the viable region where the SIMP and axion maintain thermal equilibrium. If the axion and SM also maintain thermal equilibrium via the axion-photon coupling in Eq. (4.11), then the pions will share a temperature with the SM. For most of the axion masses we consider, decays and inverse decays into SM photons are the most efficient processes for kinetic equilibrium with the SM at freeze-out. The rate for these decays at rest is

$$\Gamma_a = \frac{m_a^3}{64\pi f_{a\gamma}^2}\,. \tag{4.48}$$

For the axions to thermalize with the SM during freeze-out, two conditions must be satisfied: the decays and inverse decays of the axions need to thermalize the axions with the SM, and the axions need to lose kinetic energy via decays faster than they gain energy from the pion $3 \to 2$ heating. We have verified numerically that as long as these conditions are satisfied at the freeze-out temperature of the pion, the relic abundance of DM is unaffected and the pions constitute cold DM.

For axions with masses below the freeze-out temperature, the scattering process $ae \to \gamma e$ is more efficient than the decays considered below. This arises because the rate of scattering is enhanced relative to decays like $\sim (T/m_a)^3$ for axions that are relativistic at the time of freeze-out, while the rate of decays is suppressed due to the boost factor of $\sim (m_a/T)$. We find that the parameter space with $m_a < T_F$ is tightly constrained for the pion masses we consider and therefore do not include the scattering process $ae \to \gamma e$ in our analytics or numerics, since it is expected to be subdominant for axion masses with $m_a \gtrsim T_F$. Note that by including decays as the only channel to transfer entropy, we are being conservative since adding the $ae \to \gamma e$ channel would only lower the required coupling strength between axions and the SM.

Axions in thermal equilibrium with photons

To understand the requirement that axions maintain thermal contact with the SM, we can ignore the pions and consider only the relevant Boltzmann equations for the axions:

$$\frac{\partial n_a}{\partial t} + 3Hn_a = -\Gamma_a m_a \left(\left\langle E_a^{-1} \right\rangle_{T_a} n_a - \left\langle E_a^{-1} \right\rangle_{T_{\rm SM}} n_a^{{\rm eq},T_{\rm SM}} \right) \tag{4.49}$$

$$\frac{\partial \rho_a}{\partial t} + 3H\left(\rho_a + P_a\right) = -m_a\Gamma_a \left(n_a - n_a^{{\rm eq},T_{\rm SM}} \right) \tag{4.50}$$

where the average axion energy is $\langle E_a \rangle = \rho_a/n_a$. To make the notation less cumbersome for the remainder of this Chapter, the label for equilibrium distributions will denote chemical equilibrium *and* kinetic equilibrium between the axions and SM, *i.e.* $T_a = T_{\rm SM} \equiv T$. With this notation, these equations can be re-expressed as

$$-T\frac{\partial n_a}{\partial T} + 3n_a = -\frac{m_a\Gamma_a}{H} n_a \left(\left\langle E_a^{-1} \right\rangle - \left\langle E_a^{-1} \right\rangle^{\rm eq} \frac{n_a^{\rm eq}}{n_a} \right) \equiv -\frac{m_a\Gamma_a}{H} n_a c_n \tag{4.51}$$

$$-T\frac{\partial \left\langle E_a \right\rangle n_a}{\partial T} + 3\left\langle E_a \right\rangle n_a \left(1 + w_a\right) = -\frac{m_a\Gamma_a}{H} n_a \left(1 - \frac{n_a^{\rm eq}}{n_a} \right) \equiv -\frac{m_a\Gamma_a}{H} n_a c_\rho, \tag{4.52}$$

where w_a is the equation of state of the axion, which is a function of time and axion temperature $w_a = w_a(T_a)$. In order to eliminate $\partial n_a/\partial T$, Eqs. (4.51) and (4.52) can be combined to give a differential equation for $\langle E_a \rangle$:

$$\frac{\partial \langle E_a \rangle}{\partial T} = 3w_a \frac{\langle E_a \rangle}{T} - \frac{m_a \Gamma_a}{TH} \left(\langle E_a \rangle c_n - c_\rho \right). \tag{4.53}$$

In the first term, the expansion is driving the change in the average energy, while in the second term, the decay is driving the average energy. One can check that the first term matches the expectations for a decoupled particle. Meanwhile, for further examination of the second term, we define the variable $\alpha = \alpha(T_a)$ such that

$$\left\langle E_a^{-1} \right\rangle \equiv \alpha / \langle E_a \rangle . \tag{4.54}$$

The value of α changes monotonically $\alpha \in [1, \pi^6/(360\,\zeta(3)^2)]$ as T_a goes from 0 to ∞. Using this definition, we find

$$\frac{\partial \langle E_a \rangle}{\partial T} = 3w_a \frac{\langle E_a \rangle}{T} + \frac{m_a \Gamma_a}{TH} \left[(1-\alpha) - \frac{n_a^{\rm eq}}{n_a} \left(1 - \alpha \frac{\langle E_a \rangle}{\langle E_a \rangle^{\rm eq}} \right) \right]. \tag{4.55}$$

The second term vanishes when the particle is in equilibrium, i.e., $n_a = n_a^{\rm eq}$ with $T_a = T$. If the particle is driven out of equilibrium (for instance by the expansion), this term will push it back into equilibrium. In order to overcome the expansion, Γ_a needs to be large enough so that the second term is larger than the first.

First we consider the case that $m_a \ll T$ such that $w_a = 1/3$ and $\alpha = \pi^6/(360\,\zeta(3)^2)$. Assuming the axion is near equilibrium and expanding Eq. (4.55) around $T_a = T$ to leading order gives

$$\frac{\partial \langle E_a \rangle}{\partial T} = \frac{\langle E_a \rangle}{T} - \frac{m_a \Gamma_a}{TH} (2\alpha - 3) \frac{(T_a - T)}{T}. \tag{4.56}$$

If the two temperatures differ, then the second term will drive the system back into equilibrium if it is comparable to the first,

$$\frac{m_a \Gamma_a}{TH} \gtrsim \frac{\langle E_a \rangle}{(2\alpha - 3)T} \simeq 4 \qquad m_a \ll T. \tag{4.57}$$

In the next Subsection, we find the strongest requirement on kinetic equilibrium when $m_a \gtrsim m_\pi$. For intermediate masses $T \lesssim m_a \lesssim m_\pi$, we analytically continue the condition in Eq. (4.57) and conservatively require

$$\frac{m_a \Gamma_a}{TH} e^{-m_a/2T} \gtrsim 4 \qquad m_a \lesssim m_\pi \tag{4.58}$$

when evaluated at the freeze-out temperature. We have checked numerically that this requirement ensures thermal equilibrium between the axions and SM for the entire region $m_a \lesssim m_\pi$.

Decays and inverse decays come into equilibrium at late times relative to elastic processes. *A priori*, this could suggest the need for Eq. (4.58) to hold prior to freeze-out in order to sufficiently transfer entropy from the annihilating pions into the SM particles. However, we verified numerically that this is not the case: the SIMP relic abundance is unaffected if decays and inverse decays into SM particles only come into equilibrium close to the time of freeze-out.

Energy transfer through decays faster than from cannibalization

We also require that the kinetic energy transferred to the axions from pion $3 \to 2$ can be compensated by axion decays. In other words, the axions must decay quickly enough to transfer entropy that has accumulated from the pion sector to the SM. This condition will only be stronger than the thermalization condition when the axion is heavier than the pion and has a smaller number density, such that each axion decay must transfer several pions' entropy. The rate of kinetic energy density loss through decays for non-relativistic axions is $\Gamma_a n_a T^2/m_a$. Meanwhile, for the axions to sufficiently cool the pions, we require that $\langle \sigma v \Delta E \rangle \, n_a n_\pi \gtrsim m_\pi \dot{n}_\pi \sim H m_\pi^2 n_\pi / T$ at freeze-out when the pions are still barely in chemical equilibrium. Therefore, the requirement is

$$\frac{\Gamma_a T^2 n_a}{m_a} \gtrsim \frac{H m_\pi^2 n_\pi}{T} \tag{4.59}$$

at freeze-out. We have numerically checked that this condition keeps the pion, axion, and SM at the same temperature in the regime $m_a \gtrsim m_\pi$. We have also numerically verified that enforcing the condition only at relatively late times (corresponding to T_F) is sufficient for entropy transfer.

Out-of-equilibrium Decays

The conditions outlined above amount to requiring that the axions and SM have the same temperature at freeze-out; however, there is another possibility which we outline here. For the DM to cool, it only needs to transfer entropy to the axions throughout the freeze-out process. Then, instead of keeping the axions and SM at the same temperature, axions can decay out of equilibrium into the SM at some later time. Relative to the scenario where the axions are thermalized with the SM, the pion–axion sector will be slightly hotter than the SM and the relic abundance will be slighter larger for the same value of $3 \to 2$ rate. Therefore if the axions and SM are not thermalized at freeze-out, the value of f_π must be increased slightly to give the right relic abundance. For sufficiently large m_a, the universe undergoes a brief matter-dominated phase where the axions dominate the energy density of the universe. When the axions decay, they reheat the SM components, the universe becomes radiation-dominated again, and the pion abundance is diluted. This happens when

$$H(T_{\rm RH}) = \sqrt{\frac{g_{*,\rm RH}\pi^2}{90}} \frac{T_{\rm RH}^2}{M_{\rm Pl}} \sim \Gamma_a. \tag{4.60}$$

We require that the reheat temperature be larger than the temperature of neutrino decoupling: if the reheat temperature is lower, then the photons get preferentially heated and the effective number of relativistic neutrinos ($N_{\rm eff}$) becomes smaller than allowed by observations of the CMB [359]. We take the neutrino decoupling temperature to be $\sim$3 MeV [360], at which point $g_* = 10.75$. Having such a high reheat temperature also enforces that the decay products do not affect Big Bang Nucleosynthesis (BBN). Therefore, with the considerations outlined here in mind, couplings that are smaller than what was discussed in the previous Subsections may be viable with a slightly different $3 \to 2$ cross-section than in the standard SIMP scenario. More dedicated numerical studies are necessary in this case and will be explored in future work.

Empirical requirements

Having established a theoretically-viable parameter space, we must check whether it is allowed by current experiments and observations. Constraints arise from early universe cosmology, astrophysical bodies, and terrestrial experiments as described in the remainder of this Subsection.

If light axions are in thermal equilibrium with the SM bath, a bound on their mass arises from their effect on the temperature ratio T_ν/T_γ after neutrinos have decoupled. This difference alters the effective number of neutrino species contributing to the radiation density, N_{eff}, which can be measured in the CMB by comparing the photon diffusion scale to the sound horizon scale [359, 361]. Such constraints are relevant for light particles in equilibrium with the photon or electron plasma beneath the temperature of neutrino decoupling unless the particle couples to neutrinos as well. When applicable, this bound is stronger than the BBN bound of $\sim$ MeV, which comes from the fact that changes to N_{eff} change the expansion history and hence modify the abundance of the light elements. Because of this bound, only values of $m_a > 2.6$ MeV are considered.

The direct coupling of the axion to photons can also lead to excess emission from supernovae (SN) via the Primakoff scattering process [362]. When the coupling between axions and the SM is sufficiently strong, the scattering of $e\gamma \to ea$ produces axions in the stellar medium which leads to excess cooling if the axions escape the SN. However, if the coupling is too strong, then trapping occurs via the inverse $ea \to e\gamma$ process along with axion decays, in which case the axion does not carry any energy out of the star. SN cooling primarily proceeds through neutrino emission; due to the observed neutrino signal from SN 1987A, any new SN cooling process must carry away less energy than the neutrinos, $\sim 3 \times 10^{53}$ ergs. For photon couplings that are too weak to produce significant energy loss in the supernova, there are still constraints from escaping axions decaying into an observable burst of photons [363].

The couplings between axions and SM particles are further constrained by terrestrial experiments. However, these constraints often come with assumptions about how the axions interact with the SM. We classify constraints on the axion-photon coupling based on different assumptions about its fundamental origin, namely that the photon coupling arises from:

(a) solely coupling to $U(1)_Y$;

(b) solely coupling to $SU(2)_W$;

(c) equal couplings to $U(1)_Y$ and $SU(2)_W$, in which case, the $aZ\gamma$ coupling vanishes.

Measurements from the LEP collider and CDF constrain the decay $Z \to \gamma\gamma(\gamma)$ [364–366]. BaBar also constrains the decay $Z \to \gamma +$ invisible [366]. In the third case above, in which the $aZ\gamma$ coupling vanishes due to equal couplings to $U(1)_Y$ and $SU(2)_W$, both of these Z decay constraints are alleviated. In their place, there is a LEP bound on $e^+e^- \to \gamma\gamma$ [365] and a BaBar bound on $e^+e^- \to \gamma +$ inv [366]. Constraints from electron beam dump experiments SLAC 137, SLAC 141 [366–369], and proton beam dump experiments CHARM and NuCal [370, 371] apply for axions coupled to photons regardless of how the coupling arises. Constraints from $K_L \to \pi^0 a$ and $K^\pm \to \pi^\pm a$ with $a \to \gamma\gamma$ assume the axion couples to $SU(2)_W$. These kaon results were obtained

in Ref. [372] from analyses of fixed-target kaon rare decay experiments by the E949 [373], NA62 and NA48/2 [374], and KTeV [375] collaborations.

In addition to existing constraints, we consider the projected reach of several future experiments and analyses on the photon coupling to an axion. Experiments such as SHiP [371, 376], NA62 [371], BaBar [372], Belle II [366, 372], SeaQuest [377] and FASER [378] will affect the parameter space. In principle there could be a constraint from a process involving the aZZ coupling for all three scenarios, though we expect it would be weaker than the constraints we present and at this time we are not aware of any existing or projected limits from such a process.

Axion-Photon Parameter Space

In Fig. 4.2, we depict the requirement on $f_{a\gamma}$ such that decays and inverse decays sufficiently transfer entropy between the sectors. Each solid curve corresponds to the lower bound on $f_{a\gamma}^{-1}$ to maintain thermal contact for a fixed pion mass. We use the full Boltzmann equations and full energy transfer rates. A crossover between two regimes occurs at $m_a \sim m_\pi$, where the lower axion number density starts to matter and Eq. (4.59) becomes a stronger condition than Eq. (4.58). In the regime where $m_a \gtrsim m_\pi$, kinetic equilibrium is maintained by axions in the exponential tail of the distribution, which causes the precipitous increase in $f_{a\gamma}^{-1}$. As is evident, kinetic equilibrium between the axion and the SM through decays and inverse decays is possible over a range of axion masses. Transfer of entropy through out-of-equilibrium decays is also shown as a solid line, where we require that reheating occurs before the SM neutrinos decouple so as to keep N_{eff} unaffected. We also show the existing and projected constraints on the parameter space as outlined in the previous Subsection.

4.6 Discussion

In this Chapter, we have considered the pion realization of SIMP DM in strongly coupled gauge theories, and have shown that it can be realized with axions as the thermalization portal between the DM and SM. Throughout this work, we have required that all three sectors — the SIMPs, axions, and SM — share the same temperature as the $3 \to 2$ annihilations freeze out. This requirement sets a target range of masses and couplings for this mechanism to be theoretically viable.

In examining the couplings between the SIMPs and axions, we have required that the coupling is strong enough to thermalize the two sectors via $2 \to 2$ scattering. At the same time, we require that the coupling not be strong enough for $2 \to 2$ annihilations to overwhelm the $3 \to 2$ process that is the hallmark of the SIMP mechanism. Combined, these requirements lead to a well-defined range of couplings between the pion DM and the axions such that the SIMP mechanism can work. Constraints on annihilation coming from the CMB narrow the allowed range of couplings, though a broad parameter space remains. It is possible that a future CMB spectral distortions experiment can probe this parameter space further, though exploring this possibility is beyond the scope of this work.

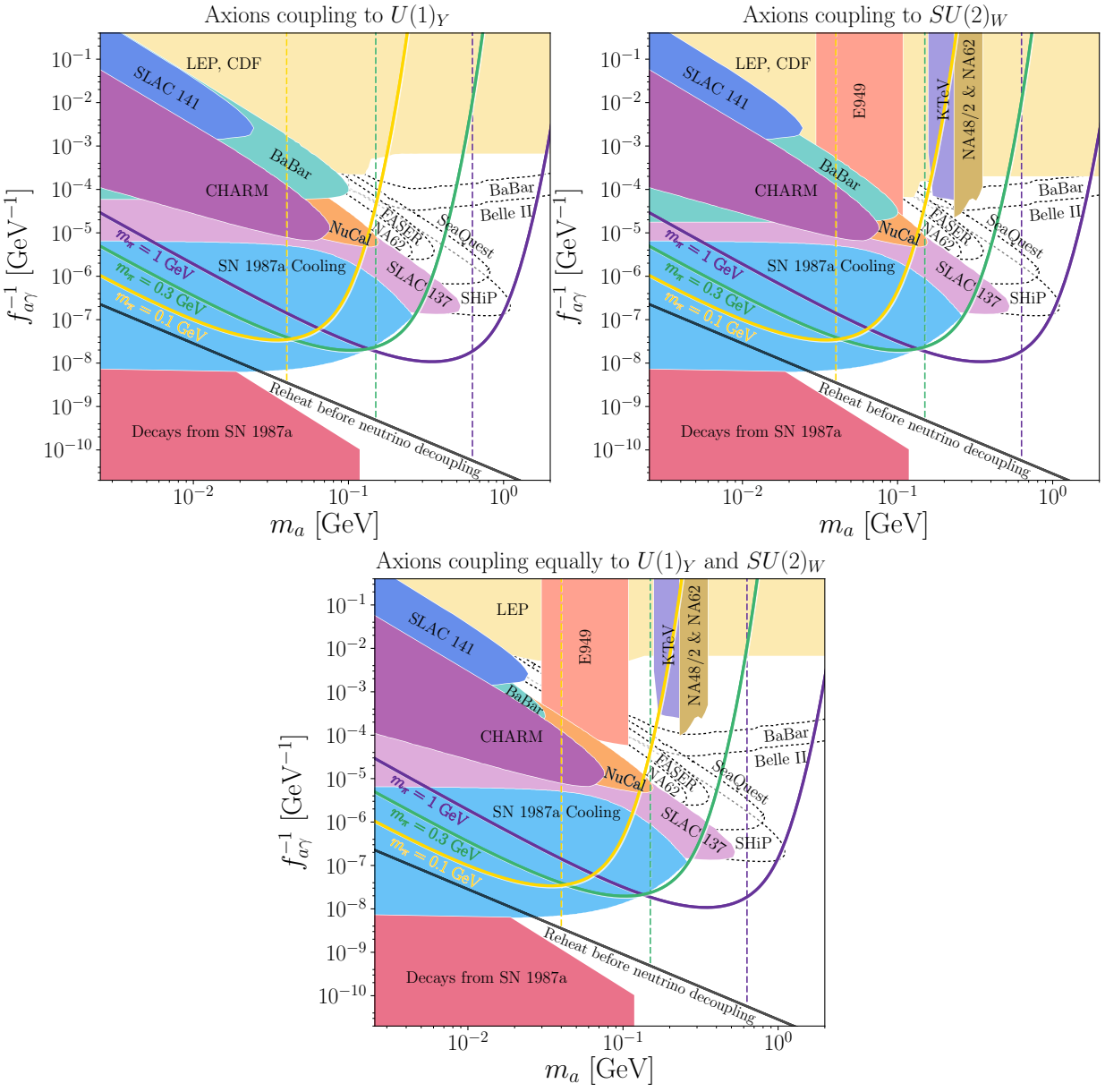

Figure 4.2: The parameter space for axions coupling to photons through $\mathrm{U}(1)_Y$ (top left), $\mathrm{SU}(2)_W$ (top right), and both in equal amounts (bottom). Solid curves correspond to the lower bound on the decay rate for thermalization between the axions and SM for various pion masses. Vertical, dashed lines correspond to the largest axion mass allowed by CMB constraints for a given pion mass (see Fig. 4.1). Below the thermalization curves, the SIMP mechanism may still be viable down to the black solid line if the axion decays out of equilibrium and reheats above the neutrino decoupling temperature. Solid, filled regions correspond to existing constraints. Regions enclosed by dotted, black lines correspond to projected reach by various proposed beam dump experiments. See the main text for more details.

In considering the couplings between the axions and the SM, we focused on the coupling to photons. For a given pion mass, there is a range of axion masses allowed for maintaining SIMP-axion equilibrium. Within this axion mass range, the main requirement for axion-SM thermalization is that the axions decay quickly enough to successfully transfer the entropy from the pions to the SM, which can easily be achieved. The relevant couplings to photons can be probed in a multitude of ways. The range of axion masses considered here are at an energy scale that is relevant for supernovae, which constrains part of the parameter space. Additionally, terrestrial beam dump and collider experiments have probed complementary parameter space. We find that the SIMP mechanism can be realized in a broad region of parameter space that is not excluded by current constraints. Several upcoming experiments are forecast to probe much of the viable parameter space that is currently allowed, providing an excellent handle for testing the framework.

There are several possible ways to extend the parameter space for axion-mediated SIMPs. Some of these possibilities are already excluded by existing limits. For instance, heavy axions which mediate the entropy transfer through off-shell interactions (both through the interactions we consider here and through the CP-violating interaction $\mathcal{L} \supset \frac{f_{\text{CPV}}}{2} a\pi^b\pi^b$) are excluded by the LEP constraint shown in Fig. 4.2. Another heavily-constrained scenario is that the axions couple to all fermions through a universal Yukawa coupling, which is almost entirely ruled out by SLAC 137, CHARM, kaon decays, B decays, supernova 1987a, BaBar and the muon anomalous magnetic moment [379–381]. A more promising possibility is that axions couple only to electrons or to charged leptons; in this case, there are known limits from SLAC 137 and the muon anomalous magnetic moment [379, 381]. In addition, the axion-electron parameter space should be constrained by limits from supernova 1987a and from loop-induced couplings to photons. However, such constraints have not been explored in the literature and will be the subject of future work. Another possibility is that the axions have a long enough lifetime that they decay out of equilibrium, dumping entropy to the SM and diluting the SIMPs— this too will be explored in future work.

Chapter 5

Freeze-in from Plasma Effects

5.1 Introduction

As discussed in Chapters 1 and 4, one of the most well-studied mechanisms for setting the observed dark matter (DM) abundance is thermal freeze-out, where DM is in equilibrium with the Standard Model (SM) thermal bath at very early times. The DM abundance is then depleted through annihilations at later times until the DM drops out of chemical equilibrium. The appeal of this mechanism is that the final relic abundance is generally independent of the high-temperature initial conditions at reheating. Furthermore, producing the observed relic abundance requires a particular thermally averaged annihilation cross section in most thermal freeze-out scenarios, $\langle \sigma v \rangle \sim 10^{-26}$ cm^3/s. This weak-scale cross section provides a target that can be probed by direct and indirect detection experiments. Assuming the relic abundance is set by annihilations to SM particles, then consistency with Big Bang Nucleosynthesis (BBN) generally requires that thermal freeze-out candidates have masses $m_\chi \gtrsim 1$ MeV [361, 382, 383]. The appealing simplicity of this scenario has led to an enormous number of DM searches targeting the thermal freeze-out mechanism, with a particular emphasis on WIMP candidates in the $m_\chi \sim$ GeV$-$TeV mass range. More recently, there has been a growing interest in $m_\chi \sim$ keV$-$GeV thermal candidates where interactions with the SM or within a hidden sector deplete the DM density to the observed value [137, 147, 333, 345, 384–393].

The freeze-in mechanism for DM production is a compelling alternative to thermal freeze-out, where DM is instead produced by feeble, sub-Hubble interactions of SM particles [394–399]. If the dominant freeze-in process is annihilation of SM particles into DM via a light mediator, then many of the appealing features of thermal freeze-out are maintained. For annihilation through a mediator lighter than the DM, the thermal cross section typically scales as $\langle \sigma v \rangle \sim g_\chi^2 g_{\rm SM}^2/(4\pi T)^2$ where g_χ is the mediator-DM coupling, $g_{\rm SM}$ is the mediator-SM coupling, and T is the SM temperature. With this scaling, DM freeze-in dominantly occurs at the lowest temperature where the process is kinematically accessible, and thus the mechanism is not sensitive to the reheat scale.[1]

Freeze-in through a light vector mediator has emerged as a key benchmark for sub-GeV direct

[1] We assume the minimal scenario where the dark sector is not populated in abundance at reheating.

detection experiments. Producing the observed DM relic abundance implies a tiny value for the coupling constants, which is difficult to target with accelerator searches. However, sufficiently light mediators give rise to scattering cross sections that scale as $\sigma \propto 1/v^4$ for relative velocity v, implying that the kinematics of the Milky Way (where $v \sim 10^{-3}$) can enhance the detectability of DM coupling to a light mediator. If the mediator also couples to charged SM fermions, then the DM can scatter off of electrons or nuclei and may be detectable with the next generations of direct detection experiments [400–411] (see also Ref. [131] for a recent review). Indeed, recent experimental results by SENSEI [412, 413], SuperCDMS [414], and DarkSide [415] are demonstrating significant progress towards achieving the sensitivity needed in the MeV-GeV mass range. It was also shown recently that Xenon1T [121] is for the first time constraining freeze-in in the GeV-TeV mass range [416].

In the keV−MeV DM mass range, freeze-in through a light mediator is the leading scenario that could be tested by proposed low-threshold direct detection experiments. Refs. [417, 418] studied the possible direct detection cross sections in models of sub-MeV DM, finding that it would be difficult to observe thermal freeze-out scenarios (even purely within a dark sector) due to a combination of BBN, CMB, fifth force, and stellar emission constraints. Obtaining accurate predictions of freeze-in is thus an important step in the program to search for low-mass DM. While freeze-in from electron-positron annihilations via a light vector mediator has been studied in the past [400, 419], in this work we thoroughly explore a previously overlooked production mechanism: freeze-in through plasma effects. The contribution of plasma effects to dark sector thermalization was estimated earlier in Refs. [420, 421] and the effect on freeze-in via a heavier mediator was recently considered in Ref. [422] as we were in the late stages of completing this work, but it was not included in previous studies of freeze-in through a light vector mediator. We find that the plasma production of DM is a dominant channel for sub-MeV DM masses, and will therefore restrict our discussion to this mass range. The additional contribution to the relic abundance implies that the target cross section for direct detection is lower by roughly an order of magnitude for the lowest experimentally accessible DM masses.

The rest of this Chapter is organized as follows. We begin in Section 5.2 by reviewing the arguments for the simplest viable freeze-in models in the keV-MeV mass range: either pure millicharged DM arising from a DM hypercharge or *effectively* millicharged DM that is coupled to an ultralight dark photon mediator. These two scenarios are almost phenomenologically identical, with the key difference being that DM-DM scattering can be parametrically larger when dark photon interactions are present. These DM candidates have recently received considerable attention in the context of the anomalous 21 cm global signal [90, 116, 423–425]. In Section 5.3 we compute the DM relic abundance from freeze-in via a light mediator. We include the effects of plasmon decays for the first time, and show the impact for direct detection. We then present the calculation of the phase space distribution for freeze-in DM in Section 5.4. A summary of our results can be found in Section 5.5. In a forthcoming paper [426], we will apply the calculations of the phase-space distribution to cosmological observables, showing that the cosmic microwave background (CMB) and probes of large-scale structure (LSS) provide a strong complementary test of DM freeze-in for $m_\chi \sim$ keV−MeV. In particular, we find that existing cosmological constraints restrict $m_\chi \gtrsim$ tens of keV for freeze-in via a light mediator, and it will be possible to probe even

higher masses with planned experiments.

5.2 Models for sub-MeV freeze-in

The case for light vector mediators

The simplest observationally viable models for sub-MeV freeze-in through a light mediator can be divided into two classes, where (1) the DM only has interactions mediated by the SM photon or (2) the DM has interactions with an ultralight kinetically mixed dark photon. We note that models of millicharged DM [420, 427] can fall under either category: they can arise as a limit of the dark-photon model where the dark photon is nearly massless, or they could be present as Dirac fermions with a tiny hypercharge.[2]

For sub-MeV freeze-in to be relevant for direct detection, vector mediators are the only observationally viable option due to stringent constraints on other light mediators with the requisite couplings to the SM, as outlined below. For direct detection of freeze-in, the mediator masses must be sufficiently small compared to the typical momentum transfer for scattering processes. If the mediators are heavier, then they do not give rise to the v^{-4} enhancement that would render extremely feeble DM-SM interactions detectable on Earth. For nuclear recoils the relevant momentum scale is set by galactic kinematics $q \sim m_\chi v \sim 10^{-3} m_\chi$, while for electron recoils the typical electron momentum in the target material is most relevant $q \sim \alpha m_e \approx 4$ keV, where m_e is the electron mass and α is the electromagnetic fine structure constant. Thus for sub-MeV DM, the experimentally relevant mediators have masses below $\mathcal{O}(1)$ keV.

Assuming an annihilation cross section of SM fermions into DM with the form $\langle \sigma v \rangle \sim g_\chi^2 g_{\rm SM}^2 / (4\pi T)^2$, the relic abundance can be estimated as

$$Y_\chi = \frac{n_\chi}{s} \sim \frac{n_{\rm SM}^2 \langle \sigma v \rangle}{sH} \sim 2 \times 10^{-4} \, \frac{g_\chi^2 g_{\rm SM}^2 M_{\rm Pl}}{T}, \tag{5.1}$$

where $M_{\rm Pl} = 1/\sqrt{8\pi G}$ is the reduced Planck mass and we assumed $T \sim$ MeV. Then for $m_\chi \sim$ MeV, we find that $g_\chi g_{\rm SM} \simeq 10^{-12}$ to saturate the relic abundance. This order-of-magnitude estimate is in agreement with more detailed calculations below. Since obtaining the relic abundance from freeze-in requires $g_\chi g_{\rm SM} \sim 10^{-12}$, $g_{\rm SM}$ must be greater than 10^{-12} if we require the dark sector to be perturbative (i.e. $g_\chi \lesssim 1$). Weakly coupled, sub-keV mediators can be emitted in stars, affecting their luminosity and lifetime. The observed properties of stars lead to strong bounds on such mediators, which we summarize here (see also Refs. [417, 418] where these bounds are collected and discussed in the context of sub-MeV DM models):

- *Scalars and pseudoscalars coupled to electrons* – The strongest bound on a light scalar with interaction $g_{\phi ee} \phi \bar{e} e$ comes from helium ignition in red giants, with $g_{\phi ee} \lesssim 7 \times 10^{-16}$ for sub-keV masses [430]. For a sub-keV pseudoscalar, observations of white dwarfs give typical

[2] Other models that have been considered in the past require giving neutrinos small charges as well [428], which we do not consider further due to strong experimental bounds on neutrino charge [429].

constraints of $g_{aee} \lesssim 2 \times 10^{-13}$ [126, 431, 432]. A caveat for most stellar emission bounds is that when the coupling is increased, the new particle may be trapped within the star and would not lead to anomalous energy loss. However, this would still affect energy transport in the star, which can be constrained for the range of couplings relevant for freeze-in through this mediator [433, 434].

- *Scalars and pseudoscalars coupled to nucleons* – Similar to the case of mediators coupling to electrons, red giants constrain $g_{\phi nn} \lesssim 10^{-12}$ for a scalar [430] and $g_{ann} \lesssim$ few $\times 10^{-10}$ for a pseudoscalar [126, 435]. While the latter coupling appears at face value to be sufficiently large, freeze-in through baryons is largely suppressed after the QCD phase transition due to the low baryon number density. Therefore, in this case our estimate for the minimum $g_{\rm SM}$ with $T \sim 1$ MeV is much too low and freeze-in would have to occur with a larger value of $g_{\rm SM}$ that is in tension with stellar bounds.

- *Scalar mixing with the Higgs* – The bounds here are similar to those in the two previous cases, and it has been shown in Ref. [436] that freeze-in through this portal is only a viable mechanism for producing all of the DM for DM masses above a few hundred MeV.

- *Kinetically mixed dark photon* – In this case, the stellar constraints on $g_{\rm SM}$ decrease linearly with the mediator mass for masses below ~ 100 eV [437, 438] because of the in-medium plasma mass suppression of producing dark photons from SM interactions, as detailed in Eq. (5.5) and the surrounding discussion in Section 5.2. From the collected bounds on dark photons from Refs. [439], a dark photon can have $g_{\rm SM} > 10^{-12}$ when its mass is well below 1 eV. At even lower masses, the coupling could be $\sim 10^{-3}$ for masses below $\lesssim 10^{-14}$ eV.

- *$B - L$ vector* – Stellar constraints on a $B - L$ vector are similar to that for the dark photon. However, for eV-scale and lighter mediator masses, a $B - L$ vector is also strongly constrained by fifth force searches (e.g. [440, 441]), which limits the mediator-SM coupling to below 10^{-12}.

Since we are focusing on the simplest benchmarks for direct detection, we do not consider more exotic possibilities with additional particles and interactions. From the bounds on new particles with the couplings described above, we conclude that freeze-in through a light mediator is viable either when the mediator is (1) the SM photon, and the DM has a tiny electric charge, or (2) when the mediator is an ultralight kinetically mixed dark photon.

We discuss these two closely related scenarios in the rest of the section. In both cases, DM has an effective charge Qe (or millicharge Q) with respect to the SM photon. This parameter determines the relic abundance, irrespective of which of the two models is under consideration. Both models allow for heat and momentum transfer between SM particles and DM during epochs when the typical relative velocities are low (as discussed in Section 5.4), which is relevant to observations of the CMB [442–448] and the cosmological 21 cm global signal [90, 116, 423–425]. The main phenomenological difference between these two possibilities is that DM-DM scattering via a dark photon can be parametrically larger than DM-DM scattering mediated by the SM photon,

as discussed below. If present at a sufficient level, the DM self-scattering can play an important role in determining the DM phase space distribution at late times, well after freeze-in.

DM with photon-mediated interactions

If the DM is a Dirac fermion χ with a tiny hypercharge Q_Y (the only gauge-invariant, renormalizable operator leading to a bare millicharge), then it can interact via the SM photon. After electroweak symmetry breaking, the DM obtains an electric charge given by $eQ_Y \equiv eQ$ (taking the convention where the Gell-Mann Nishijima formula reads $Q = I_3 + Y$). Although there are also Z-mediated DM interactions, they are negligible for the relevant epochs where $T \ll m_Z$. This gives the simplest model of millicharged DM. It is difficult to incorporate such matter content into a Grand Unified Theory (GUT) [449]; however, this scenario is economical in that it requires that no additional particles be introduced to the SM aside from the DM itself.

The possibility that this DM candidate obtains its relic abundance by thermal freeze-out has been considered before in Ref. [450], where it was shown to be excluded by structure formation when all of the DM is produced this way. Thus, freeze-in is the simplest remaining possibility for producing this DM candidate, with $g_\chi = eQ$ and $g_{\rm SM} = e$ in the language of the previous Subsections.

There are stellar emission bounds on this DM candidate because the DM can be pair produced by the decay of plasmons in stars, leading to additional energy loss [420, 421]. These bounds are shown as the shaded region in our summary plot, Fig. 5.9. Constraints on DM pair produced in SN1987a were derived in Refs. [420, 451] and require $Q \lesssim 10^{-9}$ for m_χ up to a few MeV, which does not impact freeze-in. However, there are constraints on freeze-in with m_χ below $\mathcal{O}(10)$ keV from emission in white dwarfs, horizontal branch stars, and red giants (see Appendix B of Ref. [421] for more details). Note that the range of m_χ where stellar emission can constrain freeze-in is exponentially sensitive to assumptions about temperatures within the stars. In addition, the bounds derived are applicable in the weak coupling limit where the DM escapes cleanly from the star. For sufficiently large Q, DM emission could contribute to energy transport within the star and the effects have not been carefully studied in this regime. The couplings for freeze-in are large enough that they could be in this regime and stellar bounds on freeze-in should be regarded with care.

The relevant interactions for the relic abundance and phase space distribution in this model are SM annihilations and plasma decay into the DM. DM-SM scattering can become important at late times but, as we discuss in Section 5.4, the effect must be small to be consistent with limits from the CMB. The DM self-scattering cross section is proportional to Q^4, and we find it to be irrelevant for the phase space. Finally DM-photon scattering is also proportional to Q^4 and is not enhanced in the low-velocity limit, so it is also irrelevant.

DM with dark photon interactions

We next consider Dirac fermion DM coupled to a kinetically mixed dark photon A', with the vacuum Lagrangian given by

$$\begin{aligned}\mathcal{L} \supset & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{\kappa}{2}F_{\mu\nu}F'^{\mu\nu} - \frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{1}{2}m_{A'}^2 A'_\mu A'^\mu \\ & + eJ^\mu_{\rm EM}A_\mu + g_\chi \bar{\chi}\gamma^\mu \chi A'_\mu + \bar{\chi}(i\partial - m_\chi)\chi,\end{aligned} \tag{5.2}$$

where A is the SM photon, κ is the kinetic mixing parameter and χ is Dirac fermion DM. For the purposes of this discussion, we consider Abelian kinetic mixing, noting that non-Abelian kinetic mixing is also possible [452, 453]. The mixing parameter κ could have any number of origins; for instance, it could be generated as a result of loop diagrams with heavy matter fields charged under both A and A' [454] or from certain compactifications of type IIB strings [455, 456]. Since the kinetic mixing term is a marginal operator, we take the point of view of a bottom-up effective field theory and we will treat it here as a small free parameter without specifying its origin. In this model, the combination of couplings relevant for the relic abundance is $g_\chi g_{\rm SM} = g_\chi \kappa e$.

As discussed in Section 5.2, the dark photon mass must satisfy $m_{A'} \lesssim 1$ eV in order to give a sufficient coupling for freeze-in while also evading existing bounds on stellar energy loss [439]. However, the requirements are even more stringent because unlike the model presented in Section 5.2 there could be large A'-mediated DM self interaction. For $m_{A'} <$ eV, the mediator would be light enough to give rise to v^{-4} enhanced DM self-scattering in astrophysical environments, with a rate proportional to g_χ^4. Furthermore, as mentioned above, the freeze-in relic abundance is determined by $(g_\chi \kappa e)^2$ entering into Eq. (5.1) (with $\kappa e = g_{\rm SM}$), meaning that large g_χ can be compensated by reducing κ to give the same observed relic abundance. Thus a sizable DM self-interaction is possible, and could be relevant to astrophysical probes of self-interacting DM (SIDM). The effects of SIDM are typically parameterized by the momentum-transfer self-scattering cross section, which in the limit of a very light vector mediator is given by [457]

$$\sigma_{T,\chi\chi} = \int d\cos\theta_{\rm CM} \frac{d\sigma_{\chi\chi}}{d\cos\theta_{\rm CM}}(1-\cos\theta_{\rm CM}) \approx \frac{8\pi\alpha_\chi^2}{m_\chi^2 v^4}\ln\frac{(m_\chi v)^2}{m_{A'}^2}, \tag{5.3}$$

where $\theta_{\rm CM}$ is the scattering angle in the center-of-mass (CM) frame, $\sigma_{\chi\chi}$ is the self-interaction cross section, and α_χ is the dark equivalent of the electromagnetic fine structure constant, $\alpha_\chi \equiv g_\chi^2/4\pi$. Typical bounds on SIDM require $\sigma_{\chi\chi}/m_\chi < 1-10$ cm^2/g for systems ranging from dwarf galaxies where $v \sim 10^{-4}$ to merging clusters where $v \sim 10^{-2}$ (for a recent review, see Ref. [150]). While few simulation-based studies of self-interactions have been done in the ultralight mediator limit (see for instance Ref. [458]), we can estimate the expected bound. Taking the more restrictive limit of $\sigma_{\chi\chi}/m_\chi \sim$1 cm^2/g, the bound is

$$g_\chi \lesssim 4\times10^{-5}\times\left(\frac{v}{10^{-3}}\right)\times\left(\frac{m_\chi}{1\,\text{MeV}}\right)^{3/4}\times\left(\frac{10}{\ln\left(m_\chi^2 v^2/m_{A'}^2\right)}\right)^{1/4}. \tag{5.4}$$

Since $\kappa e g_\chi \gtrsim 10^{-12}$ is needed for sub-MeV freeze-in, the SIDM bounds imply that the kinetic mixing is $\kappa \gtrsim 10^{-7}$ for MeV-scale DM. For sub-eV dark photons, such large kinetic mixing is only possible when $m_{A'} \lesssim 10^{-10}$ eV [439]. For even lighter DM, g_χ is even more restricted so $\kappa \gtrsim 10^{-5}$ is required for freeze-in, which is possible when $m_{A'} \lesssim 10^{-14}$ eV. Therefore, we are required to consider an "ultralight" dark photon [417]. Note that black hole superradiance constrains dark photons being present in the mass spectrum (in the small-coupling limit where dark photon interactions do not affect the superradiance) between $\sim 10^{-14} - 10^{-11}$ eV and preliminarily between $\sim 10^{-19} - 10^{-17}$ eV [459].

Such a light dark photon is phenomenologically equivalent to the massless dark photon limit for all processes considered in this Chapter because the $m_{A'}$ is much lower than the effective in-medium photon mass m_A in the early universe. Then, following Appendix D of Ref. [417], the vacuum Lagrangian in Eq. (5.2) is modified with an additional term $m_A^2 A^\mu A_\mu/2$.[3] Rotating away the mixing term in the presence of m_A and $m_{A'}$ and rewriting in terms of the mass eigenstates $\tilde{A}$ and $\tilde{A}'$, the in-medium Lagrangian is given by

$$\begin{aligned}\mathcal{L}_{\rm IM} \supset & -\frac{1}{4}\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{1}{4}\tilde{F}'_{\mu\nu}\tilde{F}'^{\mu\nu} + \frac{m_A^2}{2}\tilde{A}^\mu\tilde{A}_\mu + \frac{m_{A'}^2}{2}\tilde{A}'^\mu\tilde{A}'_\mu \\ & + J^\mu_{\rm EM}\left(e\tilde{A}_\mu + \frac{e\kappa\, m_{A'}^2}{m_{A'}^2 - m_A^2}\tilde{A}'_\mu\right) + g_\chi\bar{\chi}\gamma^\mu\chi\left(\tilde{A}'_\mu - \frac{\kappa m_A^2}{m_{A'}^2 - m_A^2}\tilde{A}_\mu\right). \end{aligned} \tag{5.5}$$

From this, we see that when $m_A \gg m_{A'}$, the interaction terms above reduce to

$$\mathcal{L}_{\rm IM} \supset J^\mu_{\rm EM}\left(e\tilde{A}_\mu\right) + g_\chi\bar{\chi}\gamma^\mu\chi\left(\tilde{A}'_\mu + \kappa\tilde{A}_\mu\right), \tag{5.6}$$

meaning that DM has an effective millicharge parameter $Q = \kappa g_\chi/e$, and the interactions are identical to those for a massless dark photon. Note that this suppression of the A'-SM coupling in the $m_{A'} \ll m_A$ limit is the source of the in-medium (plasma mass) suppression of the stellar constraints on dark photons [437, 438] discussed in Section 5.2. Also note that this suppression means that the dark photon is not abundantly produced by SM interactions in the early universe and does not contribute to the effective number of relativistic species, $N_{\rm eff}$.

In the exactly massless A' limit, we are free to perform a field redefinition on $A' \to A' + \kappa A$ in the vacuum Lagrangian, Eq. (5.2), which eliminates the kinetic mixing term and generates a DM interaction term $g_\chi\bar{\chi}\gamma^\mu\chi(A'_\mu + \kappa A_\mu)$, which is again identical to having a millicharge $Q = \kappa g_\chi/e$ under $U(1)_{EM}$.

The model considered here thus provides another realization of millicharged DM, and all of the stellar constraints discussed in the previous section apply. The only difference is the additional DM self-interaction via the A', which potentially leads to sizeable self-interactions.

[3] For simplicity we consider a constant m_A^2 for the schematic purposes of this discussion, although the photon polarization tensor $\Pi^{\mu\nu}(\mathbf{q},\omega)$ (which gives rise to the in-medium effective mass) depends on the photon momentum $\mathbf{q}$, energy ω, polarization, and thermal properties of the medium. For an on-shell mode with $\omega \sim |\mathbf{q}|$, m_A^2 would correspond to the plasma mass, as discussed in Section 5.3. For scattering processes with a highly off-shell mode, $|\mathbf{q}| \gg \omega$, m_A^2 is given by the Debye mass [460].

5.3 Relic abundance from freeze-in

Here we compute the relic abundance of DM from freeze-in. We begin by reproducing the contribution from annihilation of SM fermions $f\bar{f} \to \chi\bar{\chi}$ that was previously calculated in Refs. [400, 419]. Because freeze-in is peaked at low temperatures and this Chapter concerns sub-MeV DM, electrons are the primary source of DM for this channel; in the rest of this section we explicitly refer to freeze-in off electrons, noting we have numerically checked that adding heavier fermions (for instance muons) to the calculation changes the results by less than 1%. In addition to freeze-in off electrons, there is a contribution from plasmon decays, $\gamma^* \to \chi\bar{\chi}$, which we calculate for the first time. Photon annihilation into DM $\gamma\gamma \to \chi\bar{\chi}$ is suppressed by an additional factor of Q^2 and can be safely neglected.

In what follows, we take the observed present-day relic DM abundance to be $\omega_c \equiv \Omega_c h^2 = 0.12$ [31]. After freeze-in, the DM density should scale like a^{-3} and it is common practice to compare this to another quantity that has the same scaling irrespective of changes to the SM bath temperature. In this work we choose to compare the number density to the entropy density. Taking the present-day CMB temperature to be 2.73 K, the observed yield is then

$$Y \equiv n_\chi/s = 4.35 \times 10^{-7} \times \left(\frac{1\,\mathrm{MeV}}{m_\chi}\right). \tag{5.7}$$

For $m_\chi \gtrsim 1$ keV, the DM yield is much lower than the order unity yield for relativistic species, such that DM contributes negligibly to $N_{\rm eff}$. This is in contrast to other DM models, such as thermal freeze-out, where sub-MeV DM would generically inject a considerable amount of entropy to the photon or neutrino sectors and would violate observational bounds on $N_{\rm eff}$.[4]

The low DM occupation number also implies that it is possible to self-consistently ignore back-reactions that would reduce the DM number density, namely DM annihilation to electrons and inverse decays to plasmons. For instance, if we ignore the back-reaction, the solution for the number density of DM is significantly lower than the electron number density during the entirety of freeze-in in spite of the fact that the latter is becoming Boltzmann suppressed. Depletion of the DM number density through annihilation to dark photons $\chi\bar{\chi} \to \gamma'\gamma'$ is negligible for the same reason. In what follows, we solve the $0^{\rm th}$ moment of the Boltzmann equation ignoring back-reactions, noting that we have numerically checked that they are negligible. The relevant equation is then

$$\frac{dn_{\rm DM}}{da} + \frac{3n_{\rm DM}}{a} = \frac{2}{aH}\left(\langle\sigma v\rangle_{e^+e^-\to\chi\bar{\chi}}\, n_e^2 + \langle\Gamma\rangle_{\gamma^*\to\chi\bar{\chi}}\, n_{\gamma^*}\right). \tag{5.8}$$

Here we are using a as our time variable. The relationship between a and the SM temperature T (which determines the DM production rate) is not adiabatic during freeze-in because the electrons are leaving the thermal bath at this time, as discussed below. Note that we are solving for the

[4] An exception for thermal, sub-MeV DM was pointed out in Ref. [393], where the DM thermalizes with the SM thermal bath *after* neutrino-photon decoupling, reducing the contribution to $N_{\rm eff}$. Furthermore, in this model changes to $N_{\rm eff}$ that occur after DM thermalization are compensated by decoupling at a later time.

total DM density which includes both χ and $\bar{\chi}$ in the matter budget; assuming zero DM chemical potential, $n_{\rm DM} = 2n_\chi = 2n_{\bar{\chi}}$, which accounts for the factor of two in Eq. (5.8).[5]

To evaluate the relationship between a and T, we consider the properties of the SM thermal bath to be given by their equilibrium values at zero chemical potential. The photons and neutrinos are relativistic gases with energy and entropy densities

$$\rho_\gamma = \frac{\pi^2 T^4}{15}, \qquad s_\gamma = \frac{4\rho_\gamma}{3T}, \qquad \rho_\nu = \frac{7\pi^2 T_\nu^4}{40}, \qquad s_\nu = \frac{4\rho_\nu}{3T_\nu}. \tag{5.9}$$

Here we distinguish between the neutrino and SM bath temperatures T and T_ν; in this work we assume that the neutrinos kinetically decouple at a temperature that is higher than relevant for sub-MeV freeze-in and that their temperature evolves adiabatically $T_\nu \sim 1/a$ during this epoch, which is a good approximation at the percent level. We also ignore the negligible neutrino masses. Meanwhile, the electrons are transitioning from being relativistic to being non-relativistic during freeze-in, so we use the unapproximated expressions for the energy and entropy density,

$$\rho_e = \frac{2}{\pi^2}\int_{m_e}^\infty dE\, \frac{E^2(E^2-m_e^2)^{1/2}}{e^{E/T}+1}, \quad p_e = \frac{2}{3\pi^2}\int_{m_e}^\infty dE\, \frac{(E^2-m_e^2)^{3/2}}{e^{E/T}+1}, \quad s_e = \frac{p_e+\rho_e}{T}. \tag{5.10}$$

Throughout the evolution of the SM bath, we require conservation of entropy. Since we are assuming adiabatic evolution of the neutrino temperature, its entropy $s_\nu(T_\nu)a(T)^3$ is constant by definition. The DM has negligible entropy because of its low occupation number. The remaining constraint equation on the temperature evolution is then

$$\left(s_\gamma(T) + s_e(T)\right) a^3 = \text{const.}, \tag{5.11}$$

which yields a smooth temperature evolution $T(a)$, as shown in Fig. 5.1. After the electrons have fully left the bath, we recover the usual result $T_\nu = (4/11)^{1/3}T$. We can then use this temperature evolution to evolve the Hubble parameter smoothly through the transition as the electrons leave the thermal bath,

$$H^2(a) = \frac{\rho_e(T(a)) + \rho_\gamma(T(a)) + \rho_\nu(T_\nu(a))}{3M_{\rm Pl}^2} \tag{5.12}$$

with $M_{\rm Pl}$ the reduced Planck mass. The non-adiabatic temperature and Hubble evolution affect the DM relic abundance and phase space, as discussed in Section 5.4.

Annihilations

In computing the DM relic abundance from annhilations of electron-positron pairs, we use the well-known results of Ref. [354]. We provide an alternate derivation of these results in this Subsection,

[5] This factor is related to the usual factor of $1/2$ that appears in the Boltzmann equation for Dirac fermions [354, 461]; however, unlike the ordinary case of thermal DM, the change in the comoving DM density for freeze-in is independent of the DM number density (i.e. there is no factor of $n_{\rm DM}^2$ appearing in Eq. (5.8)) which accounts for the factor of four difference.

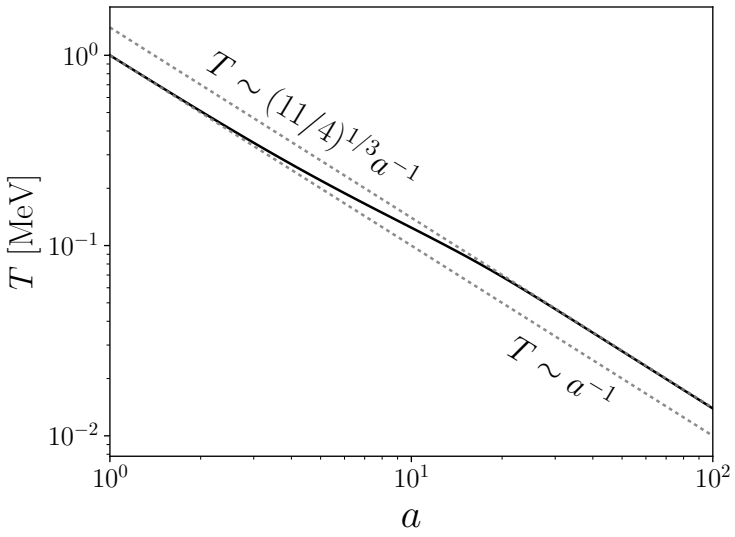

Figure 5.1: The non-adiabatic temperature evolution of the SM thermal bath during freeze-in.

which we validate here here because we use similar techniques to derive the full collision term for annihilation in Section 5.4.

In our calculation, we treat the two scenarios discussed in Section 5.2 as indistinguishable in the limit that $m_{A'} \to 0$. We also ignore the in-medium photon mass for this process, which we find to be a percent level effect for s-channel annihilations happening at the relevant range of temperatures. In this limit, the matrix element squared is

$$\sum_{\text{d.o.f.}} |\mathcal{M}|^2_{e^+e^- \leftrightarrow \chi\bar{\chi}} = \frac{32Q^2e^4}{(p_{e^+} + p_{e^-})^4}\Big((p_{e^+} \cdot p_\chi)(p_{e^-} \cdot p_{\bar{\chi}}) + (p_{e^+} \cdot p_{\bar{\chi}})(p_{e^-} \cdot p_\chi) + m_e^2(p_\chi \cdot p_{\bar{\chi}}) + m_\chi^2(p_{e^+} \cdot p_{e^-}) + 2m_e^2 m_\chi^2 \Big), \tag{5.13}$$

where we sum over both initial *and* final spin degrees of freedom (d.o.f.) without averaging and where Q is the effective millicharge in the dark photon case, $Q = \kappa g_\chi / e$. The thermally averaged cross section appearing in Eq. (5.8) for this process is given by

$$\langle \sigma v \rangle_{e^+e^- \to \chi\bar{\chi}} \, n_e^2 = \int \frac{\bar{d}^3 p_{e^+}}{2E_{e^+}} \frac{\bar{d}^3 p_{e^-}}{2E_{e^-}} \frac{\bar{d}^3 p_\chi}{2E_\chi} \frac{\bar{d}^3 p_{\bar{\chi}}}{2E_{\bar{\chi}}} \, e^{-(E_{e^+} + E_{e^-})/T} \times \sum_{\text{d.o.f.}} |\mathcal{M}|^2_{e^+e^- \to \chi\bar{\chi}} \, (2\pi)^4 \delta^{(4)}(p_{e^+} + p_{e^-} - p_\chi - p_{\bar{\chi}}) \tag{5.14}$$

where $\bar{d}^3 p \equiv d^3 p / (2\pi)^3$. We assume that from the onset of freeze-in, the electrons have entered the non-relativistic regime where their phase space is given by a Maxwell-Boltzmann distribution with temperature T and zero chemical potential. As we will show, sub-MeV DM freeze-in through the

annihilation channel is most effective at temperatures $T \lesssim m_e$ where the effects of Fermi-Dirac statistics can be neglected. We also ignore Pauli blocking of the DM due to its low occupation number.

To evaluate the thermal cross section, we note that the primordial plasma has a preferred rest frame (where bulk motions average to zero), which breaks Lorentz invariance. The phase space factors of Eq. (5.14) are evaluated in a frame that is comoving with the plasma. Practically, we can perform the integration by inserting factors of unity,

$$\int \frac{d^3q_{12}ds_{12}}{2E_{12}}\delta^{(4)}(q_{12}-p_1-p_2)=1, \tag{5.15}$$

where q_{12} is the effective bulk 4-momentum of the particles labelled 1 and 2 and s_{12} can be thought of as the effective (Lorentz invariant) mass-squared of a single particle with that bulk 3-momentum and energy (*i.e.* here $E_{12}=\sqrt{s_{12}+\mathbf{q}_{12}^2}$). Inserting such a factor into Eq. (5.14) gives

$$\begin{aligned}\langle\sigma v\rangle_{e^+e^-\to\chi\bar\chi}\, n_e^2 = &\int \frac{d^3q_{\chi\bar\chi}ds_{\chi\bar\chi}}{2E_{\chi\bar\chi}}\int \frac{\bar{d}^3p_{e^+}}{2E_{e^+}}\frac{\bar{d}^3p_{e^-}}{2E_{e^-}}\frac{\bar{d}^3p_\chi}{2E_\chi}\frac{\bar{d}^3p_{\bar\chi}}{2E_{\bar\chi}}\, e^{-(E_{e^+}+E_{e^-})/T} \\ &\times\sum_{\text{d.o.f.}}|\mathcal{M}|^2_{e^+e^-\to\chi\bar\chi}\,(2\pi)^4\delta^{(4)}(p_{e^+}+p_{e^-}-p_\chi-p_{\bar\chi})\delta^{(4)}(q_{\chi\bar\chi}-p_\chi-p_{\bar\chi}).\end{aligned} \tag{5.16}$$

The integral over p_χ and $p_{\bar\chi}$ does not depend on the frame of $q_{\chi\bar\chi}$, so the two-body phase space of p_χ and $p_{\bar\chi}$ can be evaluated in the CM frame of $q_{\chi\bar\chi}$. We define

$$\begin{aligned}\Phi_{\chi\bar\chi}(s_{\chi\bar\chi})\,|\mathcal{M}|^2_{\text{CM}}\,(s_{\chi\bar\chi}) &\equiv \int\frac{\bar{d}^3p_\chi}{2E_\chi}\int\frac{\bar{d}^3p_{\bar\chi}}{2E_{\bar\chi}}(2\pi)^4\delta^{(4)}(q_{\chi\bar\chi}-p_\chi-p_{\bar\chi})\sum_{\text{d.o.f.}}|\mathcal{M}|^2_{e^+e^-\to\chi\bar\chi} \\ &= \frac{Q^2e^4}{2\pi s_{\chi\bar\chi}^2}\sqrt{1-\frac{4m_\chi^2}{s_{\chi\bar\chi}}}\left(s_{\chi\bar\chi}^2+\frac{1}{3}(s_{\chi\bar\chi}-4m_e^2)(s_{\chi\bar\chi}-4m_\chi^2)+4s_{\chi\bar\chi}(m_\chi^2+m_e^2)\right),\end{aligned} \tag{5.17}$$

and insert this into the expression for the thermally averaged cross section

$$\begin{aligned}\langle\sigma v\rangle_{e^+e^-\to\chi\bar\chi}\, n_e^2 = &\int\frac{d^3q_{\chi\bar\chi}ds_{\chi\bar\chi}}{2E_{\chi\bar\chi}}\, e^{-E_{\chi\bar\chi}/T}\Phi_{\chi\bar\chi}(s_{\chi\bar\chi})\,|\mathcal{M}|^2_{\text{CM}}\,(s_{\chi\bar\chi}) \\ &\times\int\frac{\bar{d}^3p_{e^+}}{2E_{e^+}}\frac{\bar{d}^3p_{e^-}}{2E_{e^-}}\delta^{(4)}(p_{e^+}+p_{e^-}-q_{\chi\bar\chi}).\end{aligned} \tag{5.18}$$

Again, we can evaluate the integral over p_{e^+} and p_{e^-} in the center-of-mass frame. Defining

$$\Phi_{e^+e^-}(s_{\chi\bar\chi})\equiv\frac{1}{8\pi}\sqrt{1-\frac{4m_e^2}{s_{\chi\bar\chi}}}, \tag{5.19}$$

the thermally averaged cross section becomes

$$\langle\sigma v\rangle_{e^+e^-\to\chi\bar\chi}\, n_e^2 = \frac{1}{(2\pi)^4}\int\frac{d^3q_{\chi\bar\chi}ds_{\chi\bar\chi}}{2E_{\chi\bar\chi}}e^{-E_{\chi\bar\chi}/T}\Phi_{e^+e^-}(s_{\chi\bar\chi})\Phi_{\chi\bar\chi}(s_{\chi\bar\chi})\,|\mathcal{M}|^2_{\text{CM}}\,(s_{\chi\bar\chi}). \tag{5.20}$$

We can write this result in terms of the first order modified Bessel function of the second kind $K_1(z) = z\int_1^\infty du\, e^{-zu}\sqrt{u^2-1}$ with $u = \sqrt{1+q^2_{\chi\bar{\chi}}/s_{\chi\bar{\chi}}}$:

$$\langle\sigma v\rangle_{e^+e^-\to\chi\bar{\chi}}\, n_e^2 = \frac{T}{(2\pi)^3}\int ds\,\sqrt{s}\,\Phi_{e^+e^-}(s)\,\Phi_{\chi\bar{\chi}}(s)\,|\mathcal{M}|^2_{\rm CM}(s)\,K_1(\sqrt{s}/T) \tag{5.21}$$

where we have dropped the subscript on the integration variable s. Note that s is restricted to $s > 4\max\left(m_e^2, m_\chi^2\right)$. Our derivation indeed reproduces the result of Ref. [354] as claimed.

Plasmon decay

The early Universe is an optically thick plasma where photons acquire an in-medium mass; this can be understood classically as arising from the electrons' oscillatory response to a propagating electric field and the dynamical shielding of that electric field. This effective mass is also manifest in the photon propagator and the polarization vectors of external photon legs in the medium; in other words, the photon mass and wavefunction are renormalized in the plasma. The effective masses and dressed polarization functions for the transverse and longitudinal "plasmon" modes are shown in Fig. 5.2. The effective mass for plasmons is closely related to the plasma frequency. For a relativistic plasma at zero chemical potential, the plasma frequency is $\omega_p = eT/3 \approx 0.1T$ where e is electric charge.

More generally, the plasma dispersion relations can be determined at all temperatures and densities using the methods in Ref. [462]. The key approximation developed in that work was to evaluate thermal quantities at typical velocities, where thermal integrals have the most support. Specifically, the typical electron velocity is given by $v_* = \omega_1/\omega_p$, defined in terms of the first mode frequency and plasma frequency,

$$\omega_1^2 = \frac{4\alpha}{\pi}\int dp\frac{p^2}{E}\left(\frac{5}{3}v^2 - v^4\right) f_e(E) \tag{5.22}$$

$$\omega_p^2 = \frac{4\alpha}{\pi}\int dp\frac{p^2}{E}\left(1 - \frac{1}{3}v^2\right) f_e(E), \tag{5.23}$$

where f_e is the phase space density of electron-positron pairs. Protons can also be included but their contribution is negligible because protons are heavy and thus slow to respond to electric fields, and also because their number density is much lower than that of the electrons at the relevant epochs.

The propagation of the photon in the plasma can be determined from the electromagnetic polarization tensor, which can be written as a thermal integral and expressed in terms of the longitudinal and transverse polarization functions, Π_ℓ and Π_t, as

$$\begin{aligned}\Pi^{\mu\nu}(\omega,\mathbf{k}) = &\left(1,\frac{\omega}{k}\hat{k}\right)^\mu\left(1,\frac{\omega}{k}\hat{k}\right)^\nu\Pi_\ell(\omega,k)\\ &+\left((0,\epsilon_+)^\mu(0,\epsilon_+)^\nu + (0,\epsilon_-)^\mu(0,\epsilon_-)^\nu\right)\Pi_t(\omega,k),\end{aligned} \tag{5.24}$$

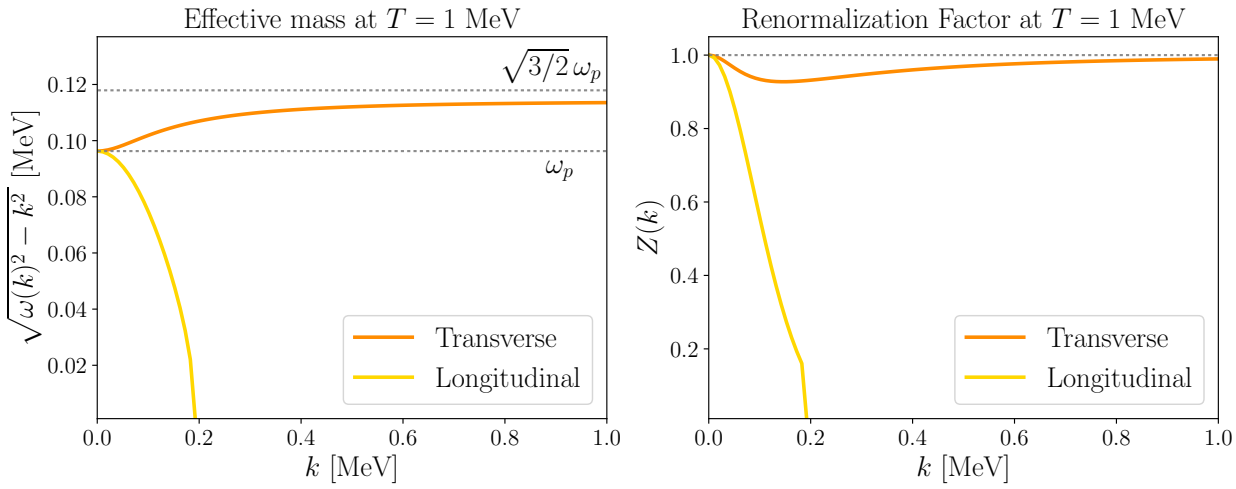

Figure 5.2: The effective in-medium mass (left) and wavefunction renormalization (right) for photons, as computed in Coulomb gauge for a plasma with $T = 1$ MeV and zero chemical potential. The transverse mode is relevant at all wavelengths while the longitudinal mode crosses the light-cone at high k and can thus only propagate at low k. Also shown are the low-k, low-T and high-k, high-T limits for the effective transverse mass, $m_t = \omega_p$ and $m_t = \sqrt{3/2}\omega_p$, respectively.

where ω and $\mathbf{k}$ are the plasmon energy and wavevector, and where the vacuum transverse polarization vectors $\epsilon_\pm$ are chosen to be orthogonal to the direction of propagation and normalized to unity. In terms of the quantities above, the polarization functions can be approximated as

$$\Pi_\ell(\omega, k) = \frac{3\omega_p^2}{v_*^2}\left(\frac{\omega}{2v_* k}\ln\left(\frac{\omega + v_* k}{\omega - v_* k}\right) - 1\right) \tag{5.25}$$

$$\Pi_t(\omega, k) = \frac{3\omega_p^2}{2v_*^2}\left(\frac{\omega^2}{k^2} - \frac{\omega(\omega^2 - v_*^2 k^2)}{2v_* k^3}\ln\left(\frac{\omega + v_* k}{\omega - v_* k}\right)\right). \tag{5.26}$$

These approximations are accurate up to $\mathcal{O}(\alpha)$ and up to $\mathcal{O}(k^2)$ at small k for all electron temperatures and densities [462]. The effective propagator can then be constructed; in Coulomb gauge, its nonzero components are

$$D^{00}(\omega, \mathbf{k}) = \frac{1}{k^2 - \Pi_\ell(\omega, k)} \tag{5.27}$$

$$D^{ij}(\omega, \mathbf{k}) = \frac{1}{\omega^2 - k^2 - \Pi_t(\omega, k)}\left(\delta^{ij} - \hat{k}^i \hat{k}^j\right). \tag{5.28}$$

The poles in the propagator yield the renormalized longitudinal and transverse dispersion relations for on-shell plasmons,

$$\omega_\ell(k)^2 = \frac{\omega_\ell(k)^2}{k^2}\Pi_\ell(\omega_\ell(k), k) \qquad \omega_t(k)^2 = k^2 + \Pi_t(\omega_t(k), k), \tag{5.29}$$

while the residues of the poles are identified as a combination of dressed polarization four-vectors, $\tilde{\epsilon}^\mu(k)\tilde{\epsilon}^\nu(k)^*$, for the appropriate polarization. The dressed polarization vectors are given by

$$\tilde{\epsilon}_L^\mu(k) = \frac{\omega_\ell(k)}{k}\sqrt{Z_\ell(k)}\,(1,\mathbf{0})^\mu \qquad \tilde{\epsilon}_\pm^\mu(k) = \sqrt{Z_t(k)}\,(0,\epsilon_\pm)^\mu\,. \tag{5.30}$$

Given the approximations for Π_ℓ and Π_t and the dispersion relations, the residue functions can be written as

$$Z_\ell(k) = \frac{2(\omega_\ell(k)^2 - v_*^2k^2)}{3\omega_p^2 - (\omega_\ell(k)^2 - v_*^2k^2)} \tag{5.31}$$

$$Z_t(k) = \frac{2\omega_t(k)^2(\omega_t(k)^2 - v_*^2k^2)}{3\omega_p^2\omega_t(k)^2 + (\omega_t(k)^2 + k^2)(\omega_t(k)^2 - v_*^2k^2) - 2\omega_t(k)^2(\omega_t(k)^2 - k^2)}. \tag{5.32}$$

Plasmons with the nontrivial dispersion relations expressed above can undergo decay provided that it is kinematically allowed. For instance, plasmons can decay to neutrino pairs through mixing with the Z boson [462]. Plasmons cannot decay to charged particles in the SM because their effective mass is also renormalized in the medium and it is always kinematically forbidden. However, this is not the case for millicharged DM where corrections to the mass are suppressed by powers of Q.

The effective matrix element that captures plasmons decaying to DM is

$$i\mathcal{M}_{\gamma^*\to\chi\bar{\chi}} = iQe\,\tilde{\epsilon}_\mu(k)\bar{u}(p_\chi)\gamma^\mu v(p_{\bar{\chi}}). \tag{5.33}$$

We express this process in terms of the DM effective millicharge Q, and the matrix element is the same for decays through a dark photon in the $m_{A'} \to 0$ limit. This is most easily seen in the basis where the dark photon has a coupling $e\kappa J^\mu_{\rm EM}A'_\mu$. In a thermal plasma, this coupling generates an in-medium mixing term in the Lagrangian given by $\kappa A_\mu \Pi^{\mu\nu}A'_\nu$ where $\Pi^{\mu\nu}$ is the electromagnetic polarization tensor. The matrix element in the dark photon basis is then given by

$$i\mathcal{M} = i\kappa g_\chi \tilde{\epsilon}^\mu(k)\Pi_{\mu\nu}\,(\omega,\mathbf{k})\,D^{\nu\alpha}_{A'}\,(\omega,\mathbf{k})\,\bar{u}(p_\chi)\gamma_\alpha v(p_{\bar{\chi}}) \equiv i\kappa g_\chi\tilde{\epsilon}^\mu(k)\bar{u}(p_\chi)\gamma_\alpha v(p_{\bar{\chi}})\,\Gamma^\alpha_\mu, \tag{5.34}$$

where $D^{\nu\alpha}_{A'}$ is the dark photon propagator. Taking the $m_{A'} = 0$ limit and again working in Coulomb gauge, the propagator is given by

$$D^{\nu\alpha}_{A'}\,(\omega,\mathbf{k}) = \frac{(1,\mathbf{0})^\nu(1,\mathbf{0})^\alpha}{k^2} + \frac{(0,\epsilon_+)^\nu\,(0,\epsilon_+)^\alpha + (0,\epsilon_-)^\nu\,(0,\epsilon_-)^\alpha}{\omega^2 - k^2}. \tag{5.35}$$

Here we are ignoring in-medium corrections on the dark photon propagator, which are suppressed by factors of κ^2. Contracting $D^{\nu\alpha}_{A'}$ with $\Pi_{\mu\nu}$ yields a vertex

$$\begin{aligned}\Gamma^{\alpha\mu}\,(\omega,\mathbf{k}) &= -\frac{\Pi_t(\omega,k)\,((0,\epsilon_+)^\mu\,(0,\epsilon_+)^\alpha + (0,\epsilon_-)^\mu\,(0,\epsilon_-)^\alpha)}{\omega^2-k^2} + \frac{\Pi_\ell(\omega,k)\left(1,\frac{\omega}{k}\hat{k}\right)^\mu(1,\mathbf{0})^\alpha}{k^2} \\ &= -\,(0,\epsilon_+)^\mu\,(0,\epsilon_+)^\alpha - (0,\epsilon_-)^\mu\,(0,\epsilon_-)^\alpha + \left(1,\frac{\omega}{k}\hat{k}\right)^\mu(1,\mathbf{0})^\alpha.\end{aligned} \tag{5.36}$$

In the second line, we have assumed on-shell transverse and longitudinal modes for the respective pieces of the vertex function and used the dispersion relations of Eq. (5.29). Contracting this with a dressed polarization vector for the external photon yields

$$\tilde{\epsilon}_L^\mu(k)\Gamma_\mu^\alpha\left(\omega_\ell,\mathbf{k}\right) = \frac{\omega_\ell(k)}{k}\sqrt{Z_\ell(k)}\left(1,\mathbf{0}\right)^\alpha \tag{5.37}$$

$$\tilde{\epsilon}_\pm^\mu(\mathbf{k})\Gamma_\mu^\alpha\left(\omega_t,\mathbf{k}\right) = \sqrt{Z_t(k)}\left(0,\epsilon_\pm\right)^\alpha, \tag{5.38}$$

which gives the same result as the vertex obtained in the millicharge basis. In squaring and summing over polarizations, only the diagonal terms (LL, $++$, and $--$) contribute,

$$\sum_{\text{d.o.f.}}|\mathcal{M}|^2_{\gamma^*\to\chi\bar{\chi}} = 4Q^2e^2\times\begin{cases}2Z_t(k)(p_\chi^2\sin^2\theta+\omega_t(k)E_\chi-kp_\chi\cos\theta) & ++\;\&\;--\\ Z_\ell(k)\frac{\omega_\ell(k)^2}{k^2}(\omega_\ell(k)E_\chi-2E_\chi^2+kp_\chi\cos\theta) & \text{LL},\end{cases} \tag{5.39}$$

where the DM four-momentum is given by $(E_\chi,\mathbf{p}_\chi)^\mu$ and θ is the angle between $\mathbf{k}$ and $\mathbf{p}_\chi$.

The thermally averaged decay rate is

$$\langle\Gamma\rangle_{\gamma^*\to\chi\bar{\chi}}\,n_{\gamma^*} = \int\frac{d^3k}{2\omega(k)}\frac{d^3p_\chi}{2E_\chi}\frac{d^3p_{\bar{\chi}}}{2E_{\bar{\chi}}}\,f\left(\omega(k)\right)(2\pi)^4\delta^{(4)}\left(K-p_\chi-p_{\bar{\chi}}\right)\sum_{\text{d.o.f.}}|\mathcal{M}|^2_{\gamma^*\to\chi\bar{\chi}}\,, \tag{5.40}$$

and can be evaluated directly. Taking the plasmons to be Bose-Einstein distributed, the longitudinal and transverse contributions to this rate are

$$\langle\Gamma\rangle_{\gamma_\ell^*\to\chi\bar{\chi}}\,n_{\gamma_\ell^*} = \frac{Q^2e^2}{(2\pi)^3}\int k^2\,dk\,\frac{Z_\ell(k)\omega_\ell(k)(m_\ell(k)^2+2m_\chi^2)\sqrt{m_\ell(k)^2(m_\ell(k)^2-4m_\chi^2)}}{3m_\ell(k)^4\left(e^{\omega_\ell(k)/T}-1\right)} \tag{5.41}$$

$$\langle\Gamma\rangle_{\gamma_t^*\to\chi\bar{\chi}}\,n_{\gamma_t^*} = \frac{4Q^2e^2}{(2\pi)^3}\int k^2\,dk\,\frac{Z_t(k)(m_t(k)^2-m_\chi^2)\sqrt{m_t(k)^2(m_t(k)^2-4m_\chi^2)}}{3\omega_t(k)\,m_t(k)^2\left(e^{\omega_t(k)/T}-1\right)}, \tag{5.42}$$

where the effective plasmon masses are $m_\ell(k)^2 = \omega_\ell(k)^2 - k^2$ for the longitudinal modes and $m_t(k)^2 = \omega_t(k)^2 - k^2$ for the tranverse ones (see Fig. 5.2). The final integrals over k can be computed numerically and the total plasmon contribution to decay is dominated by the transverse modes (note that we are working in Coulomb gauge). This is because the longitudinal mode has a finite range of k over which it can propagate, meaning that it has less available phase space than the transverse mode which has no restriction in k. Furthermore, the longitudinal mass and renormalization factors fall steeply within the range of k where this mode can propagate.

Couplings for freeze-in

In solving the zeroth moment of the Boltzmann equation for the DM relic abundance, we find that the relative contributions from e^+e^- annihilation and plasmon decays are starkly different in different mass ranges, as illustrated in Fig. 5.3. This can be understood by considering the

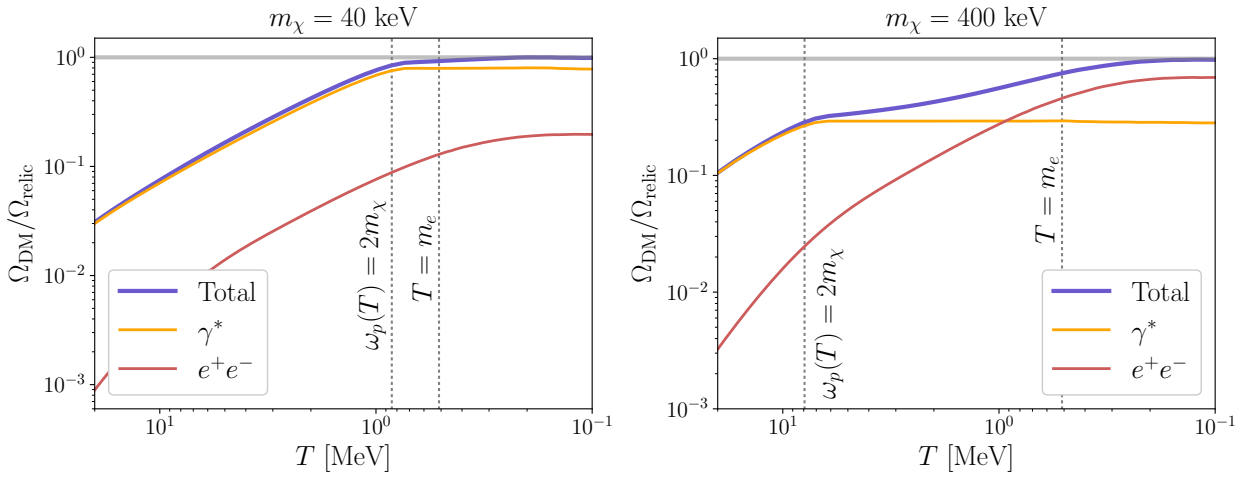

Figure 5.3: Evolution of the comoving DM number density for $m_\chi = 40$ keV (left) and $m_\chi = 400$ keV (right) as compared to the relic abundance of DM with that mass. Also shown are the relative contributions from electron-positron annihilations and plasmon decays, as discussed in the text.

fact that freeze-in is dominant at low temperatures, provided that it is kinematically allowed and that the population the DM is freezing in from has a sufficient abundance. For sub-MeV DM, freeze-in from e^+e^- annihilation is always kinematically allowed and this process only ends when the electron number density becomes Boltzmann suppressed, namely $T \lesssim m_e$. Meanwhile, the plasmon abundance is not Boltzmann suppressed but the mass runs with temperature, so freeze-in through plasmon decay ends when it is no longer kinematically allowed, namely when $m_{\gamma^*} \sim \omega_p = 2m_\chi$. Since $\omega_p \approx 0.1T$ in the relativistic limit, plasmon decay to millicharged DM shuts off at an earlier time compared to annihiliation. These two criteria are shown in Fig. 5.3 and indeed we see that plasmon decays are more dominant in determining the relic abundance for lower mass DM because the decays are active for a longer period of time.

In terms of the effective millicharge needed to produce the observed DM relic abundance, we find that including plasmon decays leads to a significant reduction in coupling for keV-mass DM while the effect is small once m_χ = MeV. The change to the freeze-in benchmark for direct detection is shown in Fig. 5.4, where the cross section for electron recoils is

$$\sigma_e = \frac{16\pi Q^2 \alpha^2 \mu_{\chi e}^2}{(\alpha m_e)^4}. \tag{5.43}$$

Here $\mu_{\chi e}$ is the electron-DM reduced mass, $\mu_{\chi e} = m_e m_\chi/(m_e + m_\chi)$. At the lowest mass where proposed low-threshold direct detection experiments are sensitive, the plasmon decay channel for DM production lowers the expected signal strength by roughly an order of magnitude.

It has been noted in the literature [463–465] that millicharged DM could be efficiently accelerated in supernova remnants, which would lead to an accelerated component of dark cosmic rays

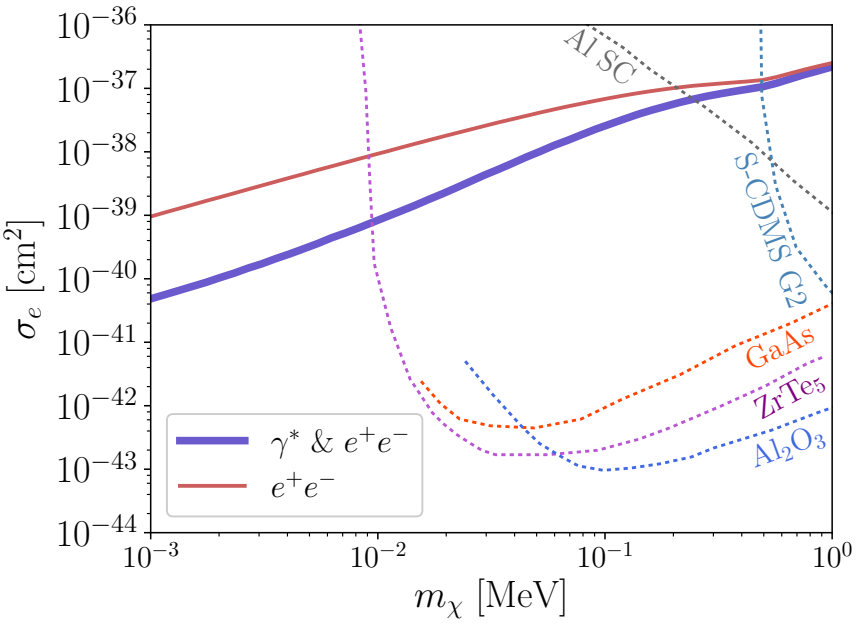

Figure 5.4: The effect of plasmon decays on the freeze-in benchmark for direct detection via electron recoils. Also shown are the projected sensitivities of low-threshold experiments with kg-day exposure, including a SuperCDMS G2 experiment [131] and proposals using polar materials (GaAs and Al_2O_3) [406, 407], Dirac materials ($ZrTe_5$) [405], or superconductors (Al SC) [411].

and eject DM from the disk. Both of these effects can lead to substantial changes to the predicted direct detection rates and sensitivities of proposed experiments shown above. However, the conclusions are highly sensitive to aspects of cosmic ray physics which are not fully understood, such as the injection of particles into the diffusive shock acceleration process. The predictions would also be sensitive to whether the DM obtains its effective millicharge through a kinetic mixing portal; in this case, the dark photon mass and couplings can affect the acceleration, and an exploration of these effects is beyond the scope of this work.

5.4 Dark matter phase space distribution

Since freeze-in DM is so weakly coupled to the SM, it does not thermalize with the SM during freeze-in and the phase space distribution can deviate substantially from a thermal distribution. While this has no clear impact on direct detection, since galaxy assembly is expected to significantly alter the DM velocity distribution, it does affect DM free-streaming and DM-SM scattering in the early universe. Here we compute the full phase space distributions needed to determine the cosmological observables; the signatures, constraints, and detection prospects will be presented in a forthcoming paper [426].

We must solve the full Boltzmann equation in an expanding background, given by

$$\frac{\partial f_\chi}{\partial t} - H\frac{p_\chi^2}{E_\chi}\frac{\partial f_\chi}{\partial E_\chi} = \frac{C(p_\chi, t)}{E_\chi}, \tag{5.44}$$

where $C(p_\chi, t)$ is the collision term, which encapsulates all interactions that affect the phase space. At early times, the interactions that determine the phase space evolution are e^+e^- annihilation and plasmon decay. We have checked numerically that heavier fermion annihilation processes (for instance the annihilation of muon-antimuon pairs) affect the phase space by a negligible amount because they occur only at early times when freeze-in is less efficient. Scattering has a negligible impact on the phase space during freeze-in since the DM occupation number is much smaller than that of electrons or plasmons. Neglecting the small effect of scattering during freeze-in, the collision term is independent of f_χ to leading order and the Boltzmann equation can be solved by direct integration [466],

$$f_\chi(p_\chi, t) = \int_{t_i}^{t} dt' \frac{C\left(\frac{a(t)}{a(t')}\, p_\chi, t'\right)}{\sqrt{\frac{a(t)^2}{a(t')^2}\, p_\chi^2 + m_\chi^2}} = \int_{a_i}^{a(t)} \frac{da'}{a' H(a')} \frac{C\left(\frac{a(t)}{a'}\, p_\chi, a'\right)}{\sqrt{\frac{a(t)^2}{a'^2}\, p_\chi^2 + m_\chi^2}}. \tag{5.45}$$

Here the factors of a in the integrand keep track of redshifting of momentum due to expansion. We use the scale factor a as our time variable rather than the common choice of using the SM temperature because it is not evolving adiabatically as the electron-positron pairs leave the bath during freeze-in, as discussed in Section 5.3.

After freeze-in ends, the DM momenta redshift and the phase space distribution is constant in comoving momentum. However, at late times DM-SM and DM-DM scattering eventually can become important since the scattering cross sections are peaked at low relative velocities. The effects of DM-SM scattering on the phase space are generally negligible for the allowed parameter space, but DM self-scattering can lead to thermalization of the DM phase space distribution. Whether this occurs is model-dependent, and we discuss the conditions for this to occur in Section 5.4.

Phase space from annihilation

The computation of the full collision term from annihilation proceeds similarly to the computation of its zeroth moment. Once again, inserting a factor of unity as defined in Eq. (5.15), we find

$$C(p_\chi, t)_{e^+e^-\to\chi\bar{\chi}} = \frac{1}{2(2\pi)^3}\int \frac{d^3q_{e^+e^-} ds_{e^+e^-}}{2E_{\bar{\chi}} 2E_{e^+e^-}} \delta(E_{e^+e^-} - E_\chi - E_{\bar{\chi}})\, e^{-E_{e^+e^-}/T} \times \Phi_{e^+e^-}(s_{e^+e^-})\, |\mathcal{M}|^2_{\rm CM}(s_{e^+e^-}), \tag{5.46}$$

where $E_{\bar{\chi}} = \sqrt{m_\chi^2 + p_\chi^2 + q_{e^+e^-}^2 - 2p_\chi q_{e^+e^-}\cos\theta}$, $E_{e^+e^-} = \sqrt{s_{e^+e^-} + q_{e^+e^-}^2}$ and θ is the angle that $\mathbf{q}_{e^+e^-}$ makes with the unconstrained, unintegrated $\mathbf{p}_\chi$. Defining $x \equiv \cos\theta$ and dropping the

subscript on the bulk electron momentum, we find

$$C(p_\chi,t)_{e^+e^-\to\chi\bar{\chi}} = \frac{1}{2(2\pi)^2 p_\chi}\int \frac{dx\,q dq\,ds}{4E}\delta\left(x-\frac{2E_\chi E - s}{2p_\chi q}\right) e^{-E/T}\Phi_{e^+e^-}(s)\,|\mathcal{M}|^2_{\rm CM}(s). \tag{5.47}$$

Requiring that $x\in[-1,1]$ and switching integration variables,

$$\begin{aligned} C(p_\chi,t)_{e^+e^-\to\chi\bar{\chi}} &= \frac{1}{8p_\chi(2\pi)^2}\int ds \int_{\frac{E_\chi s - p_\chi\sqrt{s(s-4m_\chi^2)}}{2m_\chi^2}}^{\frac{E_\chi s + p_\chi\sqrt{s(s-4m_\chi^2)}}{2m_\chi^2}} dE\, e^{-E/T}\Phi_{e^+e^-}(s)\,|\mathcal{M}|^2_{\rm CM}(s) \\ &= \frac{T}{4p_\chi(2\pi)^2}\int ds\, e^{-\frac{E_\chi s}{2m_\chi^2 T}}\sinh\left(\frac{p_\chi\sqrt{s(s-4m_\chi^2)}}{2m_\chi^2 T}\right)\Phi_{e^+e^-}(s)\,|\mathcal{M}|^2_{\rm CM}(s). \end{aligned} \tag{5.48}$$

Then, to solve for the final phase space from annihilation, we can combine Eqs. (5.45) and (5.48). Note that because p_χ is fixed (rather than an integration variable), s in the above integral is restricted to $s > \max\left(4m_e^2, 2m_\chi(E_\chi+m_\chi)\right)$ unlike in the integral for determining the thermally averaged cross section. The resulting evolution of the phase space distribution is shown in the left panel of Fig. 5.5.

Phase space from plasmon decay

The collision term from plasmon decay,

$$C(p_\chi,t)_{\gamma^*\to\chi\bar{\chi}} = \frac{1}{2}\int \frac{d^3k}{2\omega(k)}\frac{d^3p_{\bar{\chi}}}{2E_{\bar{\chi}}}\frac{1}{e^{\omega(k)/T}-1}(2\pi)^4\delta^{(4)}(K-p_\chi-p_{\bar{\chi}})\sum_{\rm d.o.f.}|\mathcal{M}|^2_{\gamma^*\to\chi\bar{\chi}} \tag{5.49}$$

proceeds through direct computation. We find

$$C(p_\chi,t)_{\gamma_\ell^*\to\chi\bar{\chi}} = \frac{Q^2e^2}{4\pi p_\chi}\int \frac{dk\,\omega_\ell(k)Z_\ell(k)}{k\left(e^{\omega_\ell(k)/T}-1\right)}\left(2E_\chi(\omega_\ell(k)-E_\chi) - m_\ell(k)^2/2\right) \tag{5.50}$$

$$C(p_\chi,t)_{\gamma_t^*\to\chi\bar{\chi}} = \frac{Q^2e^2}{4\pi p_\chi}\int \frac{dk\,kZ_t(k)}{\omega_t(k)(e^{\omega_t(k)/T}-1)}\left(2p_\chi^2 - \frac{(2E_\chi\omega_t(k)-m_t(k)^2)^2}{2k^2} + m_t(k)^2\right) \tag{5.51}$$

where the limits of the k integral are determined by the requirement that $x_0 = (2E_\chi\omega_{\ell,t}(k) - m_{\ell,t}(k)^2)/2kp_\chi$ lies in the range $[-1,1]$. The limits of integration cannot be solved for in closed form because of the nontrivial dispersion relations, so the phase space must be determined numerically.

The evolution of the phase space from plasmon decays is shown in the right panel of Fig. 5.5, and our results for the combined phase space can be found in Fig. 5.6. The distributions are noticeably nonthermal due to plasmon decays. Fig. 5.7 compares the average momentum and momentum-squared of the DM to the SM photons, which serves as a useful metric to determine the DM free-streaming and suppression of the growth of structure.

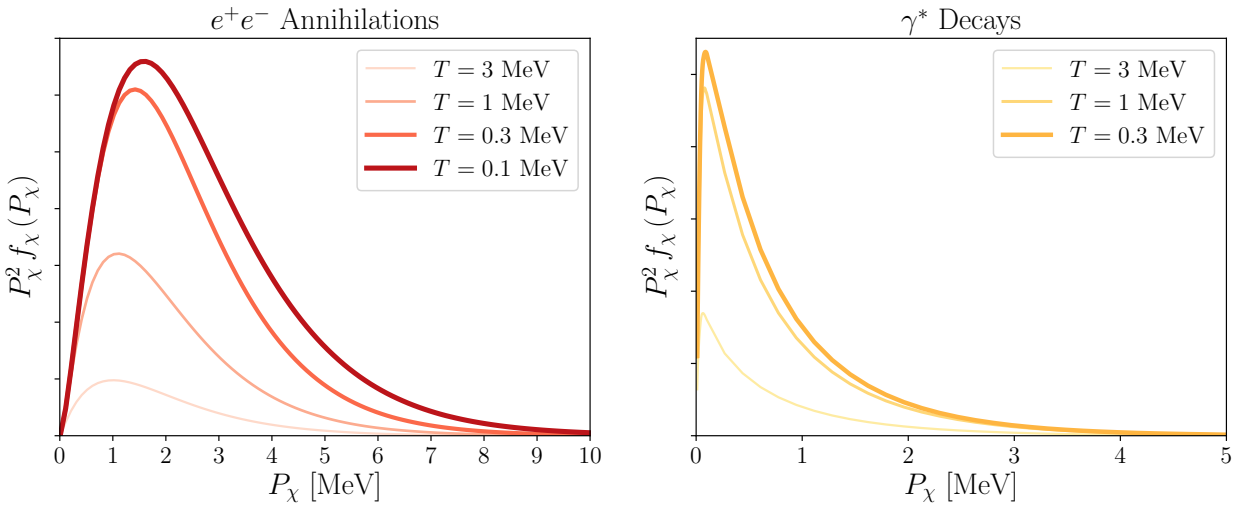

Figure 5.5: A comparison of the phase space evolution of DM being produced by e^+e^- annihilation (left) and γ^* decay (right) at $m_\chi = 40$ keV. The momenta shown here are comoving, $P_\chi \equiv a p_\chi$ where $a = 1$ corresponds to $T = 1$ MeV. The phase space is normalized arbitrarily for the purposes of comparing the P_χ-dependence side by side. Over time, the comoving phase space converges to its final frozen-in shape. The phase space from annihilation is similar to that of the thermal electrons from which they inherit their kinematics. Meanwhile, the phase space from plasmon decay is highly peaked at low P_χ because freeze-in through this channel occurs predominantly at threshold when $\omega_p \sim 2m_\chi$ and the decay is peaked when the plasmon is "at rest," $k \to 0$.

Effect of DM-SM scattering

We argue here that the effect of DM-SM scattering on the DM phase-space distribution is small from freeze-in until the onset of recombination. The relevant quantity is the momentum-transfer rate, which we estimate in the limits where the DM is relativistic and non-relativistic. We do not consider scattering by relativistic, charged SM particles because this is only relevant for electrons during freeze-in; during freeze-in, the number density of DM is many orders of magnitude smaller than the number density of electrons and the effect of electron-DM scattering is suppressed by n_χ/n_e relative to the dominant effect of electron-positron annihilations on the phase space. As outlined below, DM-SM scattering becomes more important at low velocities, corresponding to later cosmological times. This can affect CMB anisotropies and the cosmological 21 cm signal, and we provide more detailed calculations in that context in our forthcoming paper [426].

In the limit of relativistic DM scattering with non-relativistic SM particles (the case after freeze-in until $T_\gamma \sim m_\chi$), the differential cross section with respect to the center-of-mass scattering angle

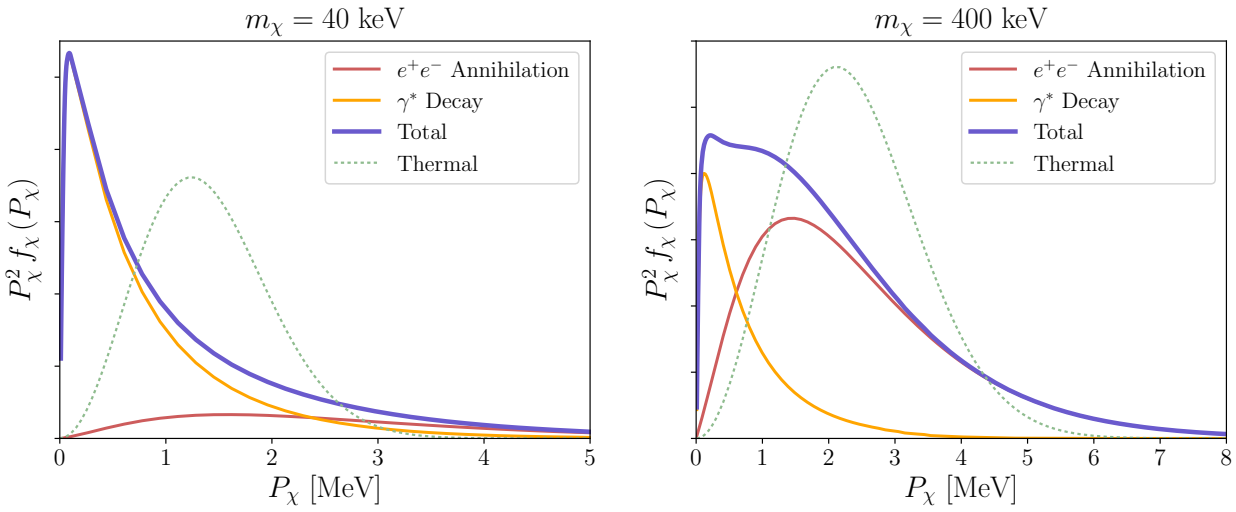

Figure 5.6: A comparison of the contributions to the phase space for $m_\chi = 40$ keV (left) and $m_\chi = 400$ keV (right). The momenta shown here are comoving, $P_\chi \equiv a p_\chi$ where $a = 1$ corresponds to $T = 1$ MeV. The phase space is normalized to the comoving DM relic abundance for each mass depicted. The plasmon contribution dominates more at low masses than at high masses because freeze-in through this channel persists for longer at lower masses, ending when the plasmon mass is at threshold, $\omega_p \sim 2m_\chi$. Also shown (dashed lines) are the phase space distributions that would arise if the DM could thermalize within its own sector, conserving $\langle P_\chi^2 \rangle$ for non-relativistic DM.

$\theta_{\rm CM}$ is given by

$$\frac{d\sigma_{\chi b}}{d\cos\theta_{\rm CM}} = \frac{\pi Q^2 \alpha^2}{p_{\rm CM}^2} \frac{(1+\cos\theta_{\rm CM})}{(1-\cos\theta_{\rm CM} + m_D^2/2p_{\rm CM}^2)^2}, \tag{5.52}$$

where $p_{\rm CM} \equiv |\mathbf{p}_{\rm CM}|$ is the momentum in the CM frame and m_D is the Debye mass. Here we have taken $p_\chi \ll m_e$, which is a good approximation after freeze-in has ended. In this approximation, the dependence on the SM particle mass drops out, making scattering with electrons and protons equally important (we refer to them collectively as "baryons," in the remainder of this discussion, hence the subscript b in the cross section).

The dependence on the Debye mass comes from considering the longitudinal polarization tensor of the plasma Π^{00} with the appropriate photon kinematics ($\omega \ll |\mathbf{q}|$) [460], which corresponds to screened Coulomb scattering. It can also be understood as the effective mass appearing in the screened electric potential, which takes the form of a Yukawa potential [460, 467, 468] or as a scale appearing in the electric form factor for a thermal Gibbs ensemble of charged particles in the plasma [468]. Note that the transverse polarization tensor Π^{ij}, which corresponds to the magnetic scattering mode, vanishes in the static $\omega \ll |\mathbf{q}|$ limit [460]; however, this mode of scattering is negligible for a non-relativistic plasma where its contribution is suppressed by factors of v [467,

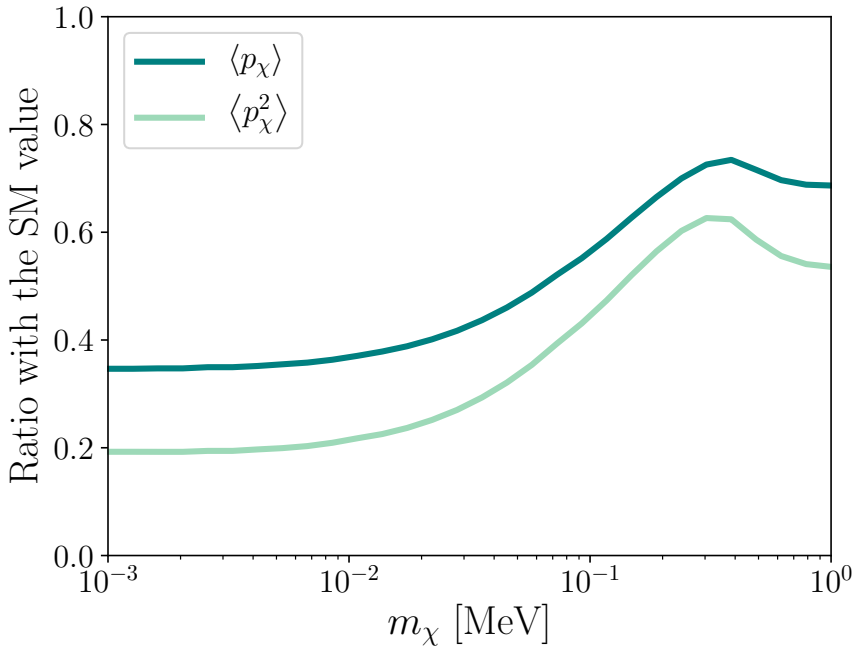

Figure 5.7: A comparison between moments of the DM phase space and the SM photon phase space as a function of DM mass. For reference, the moments for the SM photon are $\langle p_\gamma \rangle = 2.7\,T_\gamma$ and $\langle p_\gamma^2 \rangle = 10.35\,T_\gamma^2$. While the DM phase space is not thermal, these moments can be thought of as relating to the DM effective temperature, which will have ramifications for the subsequent cosmology. As the DM mass rises, the effective temperature increases because e^+e^- annihilations become more important than plasmon decays and have a comparatively fatter high-p_χ tail. At even larger masses where m_χ is comparable to m_e, that high-p_χ tail is suppressed because the DM mass becomes relevant to the kinematics of annihilation, causing the effective temperature to drop.

469]. The Debye mass automatically regulates the usual t-channel divergence in the forward-scattering direction, which is equivalent to cutting the angular integral at the Debye angle, defined as $\theta_D \equiv m_D/p_{\rm CM}$. Note that our treatment of the Debye angle differs from the one in the often-quoted Ref. [450]: that treatment did not include the Debye mass in the photon propagator. Once the plasma has become non-relativistic with $T_\gamma \lesssim m_e$, the Debye mass is given by

$$m_D = \sqrt{4\pi\alpha n_e/T_\gamma} = 3.7 \times 10^{-6}\,T_\gamma \tag{5.53}$$

in natural units, assuming $\Omega_b h^2 = 0.022$ [31] and that the ionization fraction is unity. The momentum transfer cross section is defined for DM self-scattering in Eq. (5.3) and the analogous definition applies for scattering between DM and SM particles. For relativistic DM, we find that

in the limit of the Debye angle $\theta_D \ll 1$

$$\sigma_{T,\chi b} = \frac{4\pi Q^2 \alpha^2}{p_\chi^2} \log \frac{2}{\theta_D}. \tag{5.54}$$

We emphasize that this Coulomb logarithm also agrees with other DM-baryon scattering rates found in the literature, for instance in Refs. [421, 427, 468, 470] but not with the treatment in Ref. [450]. Since $m_b \gg m_\chi$ and the baryons are non-relativistic, the DM momentum in the CM frame can be approximated by the DM momentum in the comoving frame, p_χ. As illustrated in Fig. 5.7, the typical DM momentum is comparable to the SM photon temperature, with both quantities redshifting after freeze-in. Therefore, we can estimate the momentum transfer rate per DM particle and per Hubble time as

$$\frac{n_p \sigma_{T,\chi b}}{H} \approx 5.3 \times 10^{-11} \left(\frac{Q}{10^{-10}}\right)^2 \left(\frac{\text{MeV}}{T_\gamma}\right), \tag{5.55}$$

where $n_p \approx 1.5 \times 10^{-10}\, T_\gamma^3$ and $p_\chi \approx 0.4\, p_\gamma \approx T_\gamma$. For T_γ in the keV-MeV range and $Q < 10^{-10}$ for freeze-in, this rate is tiny and thus scattering in this regime has a negligible effect on the DM phase space.

For scattering of non-relativistic DM and charged SM particles (corresponding to lower temperatures and later times), the differential cross section is instead given by

$$\frac{d\sigma_{\chi b}}{d\cos\theta_{\text{CM}}} = \frac{2\pi Q^2 \alpha^2}{\mu_{\chi b}^2 v^4} \frac{1}{(1 - \cos\theta_{\text{CM}} + m_D^2/2p_{\text{CM}}^2)^2}, \tag{5.56}$$

where $\mu_{\chi b}$ is the reduced mass of the DM and baryon, $\mu_{\chi b} = m_\chi m_b/(m_\chi + m_b)$, v is the relative velocity between DM and SM particles, and $p_{\text{CM}} = \mu_{\chi b} v$. The momentum transfer cross section is

$$\sigma_{T,\chi b} = \frac{4\pi\, Q^2 \alpha^2}{\mu_{\chi b}^2 v^4} \log \frac{2}{\theta_D}\,, \tag{5.57}$$

where again we take the $\theta_D \ll 1$ limit. We again emphasize that our approach differs from the treatment of the Coulomb logarithm in Ref. [450], which has been used for recent CMB constraints on millicharged DM [442–446]. In that work, the angular integral was cut by using the relation between impact parameter and scattering angle for (electric) Coulomb scattering, and requiring that the impact parameter for pairwise DM-baryon scattering not exceed the Debye length $\lambda_D = 1/m_D$. This translated to a minimum scattering angle that depended on the DM millicharge, with $\theta_{\text{min}} = 2Q\alpha/(3T\lambda_D)$. The corresponding minimum momentum transfer in that case would be $|\mathbf{q}|^2 = 4Q^2\alpha^2 p_{\text{CM}}^2 m_D^2/(9T^2)$. For freeze-in where $p_{\text{CM}} \approx T$ and $Q < 10^{-10}$, we see that $|\mathbf{q}|^2 \ll m_D^2$ and so we expect that the Yukawa-like form of the effective potential leads to a strong screening effect for modes of such large spatial size. In other words, the requirement of Ref. [450] may not be restrictive enough because DM-baryon scattering is suppressed by factors of Q relative to the strong collective effects in the plasma that give rise to the Debye mass. Because forward scattering

is so peaked, the resulting transfer cross section is highly sensitive to the limits of integration and their procedure yields a transfer cross section that is a factor of $\sim 2-3$ larger than the one obtained in Eq. (5.57) during times that are relevant for recombination. As a result, CMB limits on millicharged DM that use this result may be too strong.

Given the velocity scaling in Eq. (5.57), momentum transfer is most important at late times. For freeze-in couplings, there may be a substantial effect at the recombination epoch. In particular, momentum transfer during this epoch leads to a drag force between the DM and baryon fluids, which can affect CMB anisotropies [442, 446–448]. The CMB bounds *require* that the momentum transfer rate is slow compared to the rate of Hubble expansion at $z \approx 1100$, thus limiting the possible effect on the DM phase space. We calculate the bounds in detail in a forthcoming work [426], properly accounting for the velocity distribution for freeze-in DM with the updated Coulomb logarithm.

In addition to DM-baryon scattering as discussed above, DM-photon scattering is possible. However, these processes do not have the low-velocity v^{-4} enhancement in the rate and the cross section scales as Q^4, so the effects are negligible. In the model with a dark photon A', scattering processes such as $e^- + \gamma \to e^- + A'$ are also possible and scale only as kinetic mixing squared κ^2. However, these processes are still negligible compared to DM-baryon scattering since they lack the low-v enhancement and have an additional large suppression due to the in-medium kinetic mixing effects, as discussed in Section 5.2. Processes like $\chi + \gamma \to \chi + A'$ scale as $Q^2 g_\chi^2$; these also lack the v^{-4} enhancement and any enhancement (relative to DM-baryon scattering) from the large photon-to-baryon ratio is more than compensated by the factor of g_χ^2, even at the largest values of g_χ that saturate SIDM bounds.

Effect of DM-DM scattering

In the absence of a dark photon, DM self scattering is proportional to Q^4, rendering it entirely negligible. However, self-interactions of the DM can effectively thermalize the phase space distribution in the model with a dark photon. The rate for dark photon mediated DM scattering is proportional to g_χ^4, and thus may be important if g_χ is sufficiently large compared to κ. Similar to DM-baryon scattering, the cross section scales as $1/v^4$ and so these effects are most important at later times when the DM is cooler. Sufficient levels of self-scattering will convert a free-streaming phase space distribution into a Maxwell-Boltzmann or Gaussian velocity distribution. In the non-relativistic limit, the quantity $\langle a(t)^2 p_\chi^2 \rangle$ will remain the same after this process (by conservation of comoving energy), although other moments of the phase space differ.

To determine when self-scattering becomes important, we estimate the redshift z_{therm} when the momentum transfer rate per DM particle and per Hubble time is order unity:

$$\frac{n_\chi \sigma_{T,\chi\chi} v}{H(z_{\text{therm}})} = 1 \tag{5.58}$$

where v is the relative velocity between DM particles and $\sigma_{T,\chi\chi}$ is the self-scattering momentum transfer cross section given in Eq. (5.3), with the dark photon mass regulating the forward scattering instead of the Debye mass that is present for DM-baryon scattering. Using the ratio of the

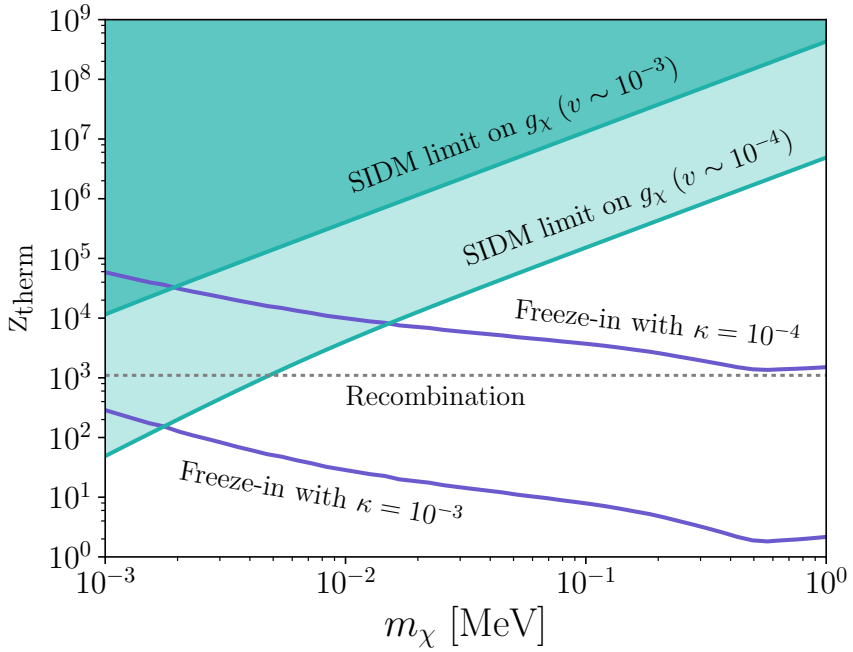

Figure 5.8: The approximate redshift when DM self-scattering becomes important, $z_{\rm therm}$, as a function of DM mass in the model with dark photon mediated interactions. The freeze-in relic abundance is determined by $Q = g_\chi \kappa / e$ and we show $z_{\rm therm}$ assuming two values of κ (where g_χ is fixed to obtain the DM relic abundance). The epoch when DM self-thermalization becomes relevant is highly sensitive to the choice of couplings, which can yield different results for CMB observables depending on whether thermalization occurs before recombination. Note that DM halo formation is neglected in this estimate. Also shown are bounds on DM self-thermalization which come from the SIDM limits on g_χ in Eq. (5.4). For illustration, we assume $\sigma_{T,\chi\chi} \lesssim 1$ cm^2/g for scattering via an ultralight mediator and show both $v \sim 10^{-3}$ and $v \sim 10^{-4}$, speeds relevant to a halo the size of the Milky Way and to a dwarf galaxy. In this figure we have taken $m_{A'} = 10^{-14}$ eV, which is sufficiently light that the constraints on the kinetic mixing parameter κ are rather weak.

average DM momentum to the photon momentum in Fig. 5.7, we approximate the relative velocity as $v \approx p_\chi / m_\chi \approx T_\gamma(z)/m_\chi$. In this estimate, we have assumed that DM is non-relativistic at the time self-interactions become important.

The self scattering randomizes the DM velocities while preserving the average kinetic energy $\frac{3}{2} T_\chi^{\rm eff}(z) \equiv \langle p_\chi^2 \rangle / (2 m_\chi)$, where p_χ is physical momentum and the average momentum-squared is given in Fig. 5.7. After self-scattering becomes significant, the DM phase space is described by a thermal Maxwell-Boltzmann distribution,

$$f_{\rm DM}(p_\chi, z) = n_{\rm DM}(z) \left(\frac{2\pi}{m_\chi T_\chi^{\rm eff}(z)} \right)^{3/2} 4\pi p_\chi^2 \exp\left(- \frac{p_\chi^2}{2 m_\chi T_\chi^{\rm eff}(z)} \right), \tag{5.59}$$

where $n_{\rm DM}(z)$ is the DM number density.

Fig. 5.8 shows the redshift of thermalization for two representative choices of κ (thus fixing g_χ to yield the observed relic abundance), where we see the assumption of non-relativistic DM is a reasonably good approximation in our estimates. Since the phase space calculations here will be an input to determining CMB constraints on freeze-in DM, we compare $z_{\rm therm}$ with the redshift of recombination $z \approx 1100$. For constraints from structure formation, a range of redshifts will be relevant. We also show some fiducial limits from SIDM, which give upper bounds on g_χ. Fig. 5.8 illustrates that the DM phase space at the time of recombination depends sensitively on the model parameters and on the robustness of SIDM limits in different astrophysical systems. For the largest values of g_χ consistent with the weaker assumed SIDM bounds, the DM phase space is described by a Maxwell-Boltzmann distribution at the time of recombination for all the DM masses we consider. However, for $\kappa = 10^{-3}$ (which is consistent with bounds on ultralight dark photons), g_χ is small enough that DM self-interactions are not important at recombination and the phase space is described by the results of Sections 5.4-5.4. The comparison of the free-streaming and thermalized phase space can be seen in Fig. 5.6.

5.5 Results and Discussion

In this Chapter, we have shown that DM freeze-in through a light vector mediator is substantially affected by plasmon decay, which constitutes a new production channel. This is an efficient way of producing sub-MeV DM and is dominant over SM fermion annihilation for masses below a few hundred keV. To account for this extra production channel, the couplings between the DM and the SM must be reduced in order to obtain the observed relic abundance of DM. For the lightest DM masses that are accessible to low-threshold direct detection experiments, the predicted cross section is lowered by roughly an order of magnitude. Updated predictions for freeze-in through a light vector mediator are shown in Fig. 5.9.

The presence of this channel also affects the DM phase space. In the absence of plasmon decays, the DM is never technically thermal but it acquires a distribution that appears thermal by inheriting the electron phase space distribution at the time of production. At early times $f_{\chi,e^+e^-}(p_\chi) \sim e^{-p_\chi/T_{\chi,e^+e^-}}$, where T_{χ,e^+e^-} is an effective DM temperature inherited from the electrons; at late times, this exponential distribution persists because the DM does *not* thermalize to give the Maxwell-Boltzmann distribution that would be expected for non-relativistic matter in equilibrium. On the other hand, the plasmon decay channel yields a DM phase space distribution that never appears thermal, which can be attributed to the running of the plasmon mass with temperature and the fact that plasmon decays occur dominantly as the plasmon wavenumber $k \to 0$. For DM masses where plasmon decays are the dominant production mode, the phase space is peaked at low momentum and has a long tail; for DM masses where contributions from both channels are important, the phase space distribution is bimodal.

Though the DM is born with a highly non-thermal distribution, it may be possible for the DM to thermalize with itself under the right circumstances. For DM that is only charged under the SM $U(1)_{EM}$ with millicharge Q, the thermalization rate is suppressed by a factor of Q^4 where the requisite Q to produce the DM relic abundance is $Q \sim \mathcal{O}(10^{-11})$. If the DM is also charged under

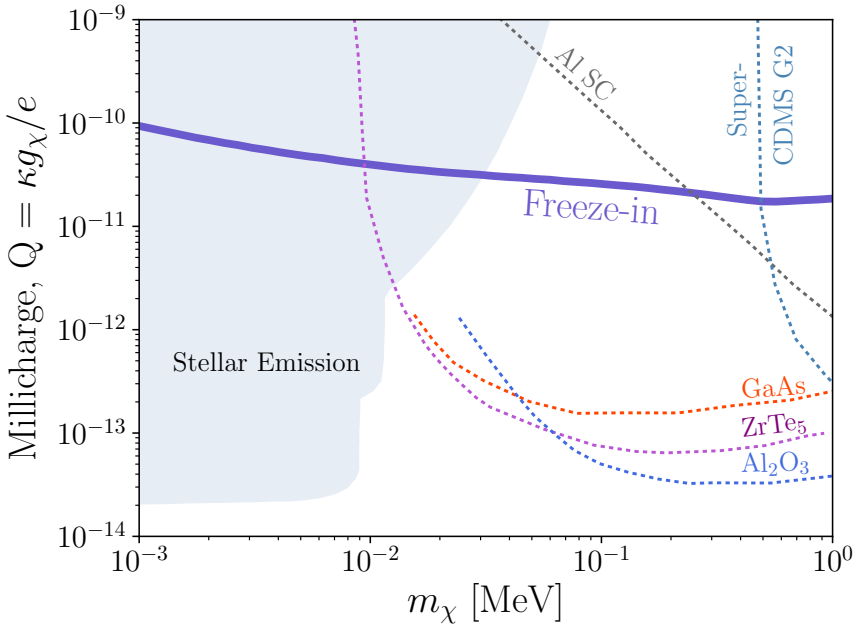

Figure 5.9: Summary plot including early-universe plasma effects for the parameter space of sub-MeV freeze-in DM. The correct DM relic abundance is obtained for couplings on the freeze-in line. We show constraints coming from emission of DM pairs in white dwarf, horizontal branch and red giant stars [421], while bounds from emission of DM pairs in supernovae apply for $Q \gtrsim 10^{-9}$ [451]. Dotted lines are projected sensitivities of proposed direct detection experiments as in Fig. 5.4.

a dark $U(1)$ gauge group that kinetically mixes with the SM $U(1)_{EM}$ (with mixing parameter κ), it may be possible for DM self-scattering to thermalize the DM phase space distribution. In this case, $Q = \kappa g_\chi / e$ (where κ can take on a wide range of values) and DM self-scattering via the dark photon scales as g_χ^4, meaning that with the appropriate choice of κ and g_χ it is possible to efficiently self-scatter while still producing the observed relic abundance. The coupling g_χ cannot be arbitrarily large due to observational limits on SIDM in astrophysical systems; however, there is a range of g_χ where self-scattering thermalizes the DM before recombination and where the SIDM bounds are simultaneously satisfied. Energy is conserved within the DM fluid, so for non-relativistic DM $\langle p_\chi^2 \rangle$ will be conserved and the resulting distribution has a well-defined notion of temperature.

Although the freeze-in DM phase space distribution may not be thermal, it is still informative to take moments of the distribution. When comparing the first and second moments of $f_\chi(p_\chi)$ to the equivalent quantities for the SM photon bath, we find that the typical DM momentum is similar to the typical photon momentum, $\langle p_\chi \rangle \approx (0.4 - 0.7) \times \langle p_\gamma \rangle$ depending on the DM mass. In other words, the DM is born considerably warmer than what is typically assumed for cold DM initial conditions. This will have ramifications for cosmology in two key ways:

- Freeze-in DM will behave like warm DM, leading to suppression of the matter power spectrum below some physical scale roughly corresponding to the free-streaming length. This effect is not already captured by existing limits on warm DM, where different DM phase space distributions are assumed. To understand this suppression quantitatively, a Boltzmann code is necessary that accounts for the potentially nonthermal phase space from freeze-in. Having understood this, it will be possible to constrain DM freeze-in via a light vector mediator using probes of the matter power spectrum and the halo mass function.

- Existing CMB limits on DM with an effective millicharge do not straightforwardly apply to the case of freeze-in. These limits stem from a DM-baryon drag; because the drag is highly sensitive to the relative DM-baryon velocity (the cross section scales like $\sim v^{-4}$), modifications to the DM phase space can substantially alter the size of the effect. Existing limits have made the assumption of cold dark matter, and the larger DM velocities for freeze-in will lead to reduced drag force. Taking into account the updated Debye logarithm (which may weaken existing limits by a factor of $\sim 2-3$), the limit on freeze-in will be further reduced compared to previously reported results.

Both of these effects will be thoroughly explored in our forthcoming paper [426], which will place restrictions on the range of masses where DM freeze-in via a light mediator is observationally viable.

Chapter 6

Summary

In this dissertation, we have demonstrated that insights from the study of astrophysical systems can shed light on hidden sector particle physics. In particular, we have shown that dark sector dissipations in the Milky Way can be constrained using kinematic analyses of astrometric data, that multi-state DM self-scattering can dramatically alter the structure and abundance of DM halos in N-body simulations, that strongly interacting DM can shed entropy through the axion portal and naturally produce the observed relic abundance of DM that self-scatters in DM halos, and that DM can be produced from the decay of collective plasma modes in the early Universe and observably alter the CMB and the formation of LSS. This thesis is far from comprehensive in the sense that DM remains undiscovered and other astrophysical probes of hidden sectors will certainly be explored in future works. The contents of our Universe are diverse, and there is yet room for creativity in harnessing the constraining power of astrophysics for dark sector physics motivations. Perhaps some new astrophysical anomaly will yield a positive discovery of new forces and particles, but at present we must be content to learn all we can from what observations of our Universe have already provided.

Bibliography

[1] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, "X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue", Astron. Astrophys. **564**, A125 (2014), arXiv:1402.0004 [astro-ph.HE].

[2] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, "X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue", Astron. Astrophys. **564**, A125, A125 (2014), arXiv:1402.0004 [astro-ph.HE].

[3] M. J. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives", The computer journal **7**, 155–162 (1964).

[4] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: open source scientific tools for Python*, 2001.

[5] J. Bovy, "Stellar inventory of the solar neighbourhood using Gaia DR1", Mon. Not. R. Astron. Soc. **470**, 1360–1387 (2017), arXiv:1704.05063.

[6] V. Springel, "E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh", Mon. Not. R. Astron. Soc. **401**, 791–851 (2010), arXiv:0901.4107 [astro-ph.CO].

[7] F. Zwicky, "Die rotverschiebung von extragalaktischen nebeln", Helvetica Physica Acta **6**, 110–127 (1933).

[8] V. C. Rubin and W. K. Ford Jr, "Rotation of the andromeda nebula from a spectroscopic survey of emission regions", The Astrophysical Journal **159**, 379 (1970).

[9] V. C. Rubin, N. Thonnard, and J. Ford W. K., "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4\,\mathrm{kpc}$) to UGC 2885 ($R = 122\,\mathrm{kpc}$)", Astrophys. J. **238**, 471 (1980).

[10] A. Bosma, "The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types", PhD thesis (Jan. 1978).

[11] V. Trimble, "Existence and nature of dark matter in the universe", Annual review of astronomy and astrophysics **25**, 425–472 (1987).

[12] A. Borriello and P. Salucci, "The Dark matter distribution in disk galaxies", Mon. Not. Roy. Astron. Soc. **323**, 285 (2001), arXiv:astro-ph/0001082 [astro-ph].

[13]M. Mateo, "Dwarf galaxies of the Local Group", Ann. Rev. Astron. Astrophys. **36**, 435–506 (1998), arXiv:astro-ph/9810070 [astro-ph].

[14]J. T. Kleyna, M. I. Wilkinson, N. W. Evans, and G. Gilmore, "Dark matter in dwarf spheroidals. 2. Observations and modeling of draco", Mon. Not. Roy. Astron. Soc. **330**, 792 (2002), arXiv:–astro-ph/0109450.

[15]M. G. Walker, M. Mateo, E. W. Olszewski, R. A. Bernstein, X. Wang, and M. Woodroofe, "Internal kinematics of the fornax dwarf spheroidal galaxy", Astron. J. **131**, [Erratum: Astron. J.132,968(2006)], 2114–2139 (2006), arXiv:astro-ph/0511465 [astro-ph].

[16]M. G. Walker, M. Mateo, E. W. Olszewski, O. Y. Gnedin, X. Wang, B. Sen, and M. Woodroofe, "Velocity dispersion profiles of seven dwarf spheroidal galaxies", The Astrophysical Journal Letters **667**, L53 (2007).

[17]J. D. Simon et al., "A Complete Spectroscopic Survey of the Milky Way Satellite Segue 1: The Darkest Galaxy", Astrophys. J. **733**, 46 (2011), arXiv:1007.4198 [astro-ph.GA].

[18]T. E. Armandroff, E. W. Olszewski, and C. Pryor, "The mass-to-light ratios of the draco and ursa minor dwarf spheroidal galaxies. i. radial velocities from multifiber spectroscopy", The Astronomical Journal **110**, 2131 (1995).

[19]N. Dalal and C. S. Kochanek, "Direct detection of CDM substructure", Astrophys. J. **572**, 25–33 (2002), arXiv:astro-ph/0111456 [astro-ph].

[20]R. B. Metcalf, L. A. Moustakas, A. J. Bunker, and I. R. Parry, "Spectroscopic gravitational lensing and limits on the dark matter substructure in Q2237+0305", Astrophys. J. **607**, 43–59 (2004), arXiv:astro-ph/0309738 [astro-ph].

[21]S. Vegetti, L. V. E. Koopmans, A. Bolton, T. Treu, and R. Gavazzi, "Detection of a Dark Substructure through Gravitational Imaging", Mon. Not. Roy. Astron. Soc. **408**, 1969 (2010), arXiv:0910.0760 [astro-ph.CO].

[22]S. Vegetti, D. J. Lagattuta, J. P. McKean, M. W. Auger, C. D. Fassnacht, and L. V. E. Koopmans, "Gravitational detection of a low-mass dark satellite at cosmological distance", Nature **481**, 341 (2012), arXiv:1201.3643 [astro-ph.CO].

[23]A. M. Nierenberg, T. Treu, S. A. Wright, C. D. Fassnacht, and M. W. Auger, "Detection of substructure with adaptive optics integral field spectroscopy of the gravitational lens B1422+231", Mon. Not. Roy. Astron. Soc. **442**, 2434–2445 (2014), arXiv:1402.1496 [astro-ph.GA].

[24]D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter", Astrophys. J. **648**, L109–L113 (2006), arXiv:astro-ph/0608407 [astro-ph].

[25]M. Bradac, S. W. Allen, T. Treu, H. Ebeling, R. Massey, R. G. Morris, A. von der Linden, and D. Applegate, "Revealing the properties of dark matter in the merging cluster MACSJ0025.4-1222", Astrophys. J. **687**, 959 (2008), arXiv:0806.2320 [astro-ph].

[26]D. Clowe, M. Markevitch, M. Bradac, A. H. Gonzalez, S. M. Chung, R. Massey, and D. Zaritsky, "On Dark Peaks and Missing Mass: A Weak Lensing Mass Reconstruction of the Merging Cluster System Abell 520", Astrophys. J. **758**, 128 (2012), arXiv:1209.2143 [astro-ph.CO].

[27]C.-P. Ma and E. Bertschinger, "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges", Astrophys. J. **455**, 7–25 (1995), arXiv:astro-ph/9506072 [astro-ph].

[28]W. Hu and N. Sugiyama, "Anisotropies in the cosmic microwave background: An Analytic approach", Astrophys. J. **444**, 489–506 (1995), arXiv:astro-ph/9407093 [astro-ph].

[29]W. Hu and S. Dodelson, "Cosmic microwave background anisotropies", Ann. Rev. Astron. Astrophys. **40**, 171–216 (2002), arXiv:astro-ph/0110414 [astro-ph].

[30]G. Hinshaw et al., "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results", Astrophys. J. Suppl. **208**, 19 (2013), arXiv:1212.5226 [astro-ph.CO].

[31]N. Aghanim et al., "Planck 2018 results. VI. Cosmological parameters", (2018), arXiv:1807.06209 [astro-ph.CO].

[32]D. J. Eisenstein and W. Hu, "Baryonic features in the matter transfer function", Astrophys. J. **496**, 605 (1998), arXiv:astro-ph/9709112 [astro-ph].

[33]D. J. Eisenstein and W. Hu, "Power spectra for cold dark matter and its variants", Astrophys. J. **511**, 5 (1997), arXiv:astro-ph/9710252 [astro-ph].

[34]M. Tegmark, M. Zaldarriaga, and A. J. S. Hamilton, "Towards a refined cosmic concordance model: Joint 11 parameter constraints from CMB and large scale structure", Phys. Rev. **D63**, 043007 (2001), arXiv:astro-ph/0008167 [astro-ph].

[35]M. Tegmark et al., "Cosmological parameters from SDSS and WMAP", Phys. Rev. **D69**, 103501 (2004), arXiv:astro-ph/0310723 [astro-ph].

[36]D. J. Eisenstein et al., "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies", Astrophys. J. **633**, 560–574 (2005), arXiv:astro-ph/0501171 [astro-ph].

[37]U. Seljak, A. Slosar, and P. McDonald, "Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints", JCAP **0610**, 014 (2006), arXiv:-astro-ph/0604335.

[38]A. Vikhlinin et al., "Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints", Astrophys. J. **692**, 1060–1074 (2009), arXiv:0812.2720 [astro-ph].

[39]F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant", Mon. Not. Roy. Astron. Soc. **416**, 3017–3032 (2011), arXiv:1106.3366 [astro-ph.CO].

[40]M. Betoule et al., "Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples", Astron. Astrophys. **568**, A22 (2014), arXiv:1401.4064.

[41]T. Abbott et al., "Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing", Phys. Rev. **D98**, 043526 (2018), arXiv:1708.01530.

[42]T. P. Walker, G. Steigman, D. N. Schramm, K. A. Olive, and H.-S. Kang, "Primordial nucleosynthesis redux", Astrophys. J. **376**, [,43(1991)], 51–69 (1991).

[43]C. J. Copi, D. N. Schramm, and M. S. Turner, "Big bang nucleosynthesis and the baryon density of the universe", Science **267**, 192–199 (1995), arXiv:astro-ph/9407006 [astro-ph].

[44]B. J. Carr and S. W. Hawking, "Black holes in the early Universe", Mon. Not. Roy. Astron. Soc. **168**, 399–415 (1974).

[45]S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, "Did LIGO detect dark matter?", Phys. Rev. Lett. **116**, 201301 (2016), arXiv:1603.00464 [astro-ph.CO].

[46]S. Clesse and J. García-Bellido, "The clustering of massive Primordial Black Holes as Dark Matter: measuring their mass distribution with Advanced LIGO", Phys. Dark Univ. **15**, 142–147 (2017), arXiv:1603.05234 [astro-ph.CO].

[47]A. Kashlinsky, "LIGO gravitational wave detection, primordial black holes and the near-IR cosmic infrared background anisotropies", Astrophys. J. **823**, L25 (2016), arXiv:1605.04023 [astro-ph.CO].

[48]M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, "Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914", Phys. Rev. Lett. **117**, 061101 (2016), arXiv:1603.08338 [astro-ph.CO].

[49]J. Garcia-Bellido, A. D. Linde, and D. Wands, "Density perturbations and black hole formation in hybrid inflation", Phys. Rev. **D54**, 6040–6058 (1996), arXiv:astro-ph/9605094 [astro-ph].

[50]D. H. Lyth, "Contribution of the hybrid inflation waterfall to the primordial curvature perturbation", JCAP **1107**, 035 (2011), arXiv:1012.4617 [astro-ph.CO].

[51]M. Kawasaki, N. Kitajima, and T. T. Yanagida, "Primordial black hole formation from an axionlike curvaton model", Phys. Rev. **D87**, 063519 (2013), arXiv:1207.2550 [hep-ph].

[52]K. Schutz and A. Liu, "Pulsar timing can constrain primordial black holes in the LIGO mass window", Phys. Rev. **D95**, 023002 (2017), arXiv:1610.04234 [astro-ph.CO].

[53]A. Polnarev and R. Zembowicz, "Formation of Primordial Black Holes by Cosmic Strings", Phys. Rev. **D43**, 1106–1109 (1991).

[54]R. A. Allsman et al., "MACHO project limits on black hole dark matter in the 1-30 solar mass range", Astrophys. J. **550**, L169 (2001), arXiv:astro-ph/0011506 [astro-ph].

[55]L. Wyrzykowski et al., "The OGLE View of Microlensing towards the Magellanic Clouds. IV. OGLE-III SMC Data and Final Conclusions on MACHOs", Mon. Not. Roy. Astron. Soc. **416**, 2949 (2011), arXiv:1106.2925 [astro-ph.GA].

[56]P. Tisserand et al., "Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds", Astron. Astrophys. **469**, 387–404 (2007), arXiv:astro-ph/0607207 [astro-ph].

[57]M. A. Monroy-Rodríguez and C. Allen, "The end of the MACHO era- revisited: new limits on MACHO masses from halo wide binaries", Astrophys. J. **790**, 159 (2014), arXiv:1406.5169 [astro-ph.GA].

[58]M. Ricotti, J. P. Ostriker, and K. J. Mack, "Effect of Primordial Black Holes on the Cosmic Microwave Background and Cosmological Parameter Estimates", Astrophys. J. **680**, 829 (2008), arXiv:0709.0524 [astro-ph].

[59]L. Chen, Q.-G. Huang, and K. Wang, "Constraint on the abundance of primordial black holes in dark matter from Planck data", JCAP **1612**, 044 (2016), arXiv:1608.02174.

[60]T. D. Brandt, "Constraints on MACHO Dark Matter from Compact Stellar Systems in Ultra-Faint Dwarf Galaxies", Astrophys. J. **824**, L31 (2016), arXiv:1605.03665 [astro-ph.GA].

[61]B. Carr, F. Kuhnel, and M. Sandstad, "Primordial Black Holes as Dark Matter", Phys. Rev. **D94**, 083504 (2016), arXiv:1607.06077 [astro-ph.CO].

[62]A. M. Green, "Microlensing and dynamical constraints on primordial black hole dark matter with an extended mass function", Phys. Rev. **D94**, 063530 (2016), arXiv:1609.01143 [astro-ph.CO].

[63]J. B. Muñoz, E. D. Kovetz, L. Dai, and M. Kamionkowski, "Lensing of Fast Radio Bursts as a Probe of Compact Dark Matter", Phys. Rev. Lett. **117**, 091301 (2016), arXiv:1605.00008 [astro-ph.CO].

[64]R. Saito and J. Yokoyama, "Gravitational wave background as a probe of the primordial black hole abundance", Phys. Rev. Lett. **102**, [Erratum: Phys. Rev. Lett.107,069901(2011)], 161101 (2009), arXiv:0812.4339 [astro-ph].

[65]E. Bugaev and P. Klimai, "Constraints on the induced gravitational wave background from primordial black holes", Phys. Rev. **D83**, 083521 (2011), arXiv:1012.4697 [astro-ph.CO].

[66]N. Seto and A. Cooray, "Searching for primordial black hole dark matter with pulsar timing arrays", Astrophys. J. **659**, L33–L36 (2007), arXiv:astro-ph/0702586 [astro-ph].

[67]K. Kashiyama and N. Seto, "Enhanced exploration for primordial black holes using pulsar timing arrays", Mon. Not. Roy. Astron. Soc. **426**, 1369 (2012), arXiv:1208.4101.

[68]M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, "Primordial black holes—perspectives in gravitational wave astronomy", Class. Quant. Grav. **35**, 063001 (2018), arXiv:1801.05235 [astro-ph.CO].

[69]M. P. Hertzberg and M. Yamada, "Primordial Black Holes from Polynomial Potentials in Single Field Inflation", Phys. Rev. **D97**, 083509 (2018), arXiv:1712.09750 [astro-ph.CO].

[70]M. Zumalacarregui and U. Seljak, "Limits on stellar-mass compact objects as dark matter from gravitational lensing of type Ia supernovae", Phys. Rev. Lett. **121**, 141101 (2018), arXiv:1712.02240 [astro-ph.CO].

[71]J. Luis Bernal, N. Bellomo, A. Raccanelli, and L. Verde, "Cosmological implications of Primordial Black Holes", JCAP **1710**, 052 (2017), arXiv:1709.07465 [astro-ph.CO].

[72]Y. Ali-Haïmoud, E. D. Kovetz, and M. Kamionkowski, "Merger rate of primordial black-hole binaries", Phys. Rev. **D96**, 123523 (2017), arXiv:1709.06576 [astro-ph.CO].

[73]V. Poulin, P. D. Serpico, F. Calore, S. Clesse, and K. Kohri, "CMB bounds on disk-accreting massive primordial black holes", Phys. Rev. **D96**, 083524 (2017), arXiv:1707.04206.

[74]B. Carr, M. Raidal, T. Tenkanen, V. Vaskonen, and H. Veermäe, "Primordial black hole constraints for extended mass functions", Phys. Rev. **D96**, 023514 (2017), arXiv:1705.05567 [astro-ph.CO].

[75]F. Kühnel and K. Freese, "Constraints on Primordial Black Holes with Extended Mass Functions", Phys. Rev. **D95**, 083508 (2017), arXiv:1701.07223 [astro-ph.CO].

[76]B. Horowitz, "Revisiting Primordial Black Holes Constraints from Ionization History", (2016), arXiv:1612.07264 [astro-ph.CO].

[77]T. Nakama, J. Silk, and M. Kamionkowski, "Stochastic gravitational waves associated with the formation of primordial black holes", Phys. Rev. **D95**, 043511 (2017), arXiv:1612.06264 [astro-ph.CO].

[78]Y. Ali-Haïmoud and M. Kamionkowski, "Cosmic microwave background limits on accreting primordial black holes", Phys. Rev. **D95**, 043534 (2017), arXiv:1612.05644.

[79]N. Orlofsky, A. Pierce, and J. D. Wells, "Inflationary theory and pulsar timing investigations of primordial black holes and gravitational waves", Phys. Rev. **D95**, 063518 (2017), arXiv:1612.05279 [astro-ph.CO].

[80]K. Inomata, M. Kawasaki, K. Mukaida, Y. Tada, and T. T. Yanagida, "Inflationary primordial black holes for the LIGO gravitational wave events and pulsar timing array experiments", Phys. Rev. **D95**, 123510 (2017), arXiv:1611.06130 [astro-ph.CO].

[81]S. Clesse and J. García-Bellido, "Detecting the gravitational wave background from primordial black hole dark matter", Phys. Dark Univ. **18**, 105–114 (2017), arXiv:1610.08479 [astro-ph.CO].

[82]J. R. Bond, G. Efstathiou, and J. Silk, "Massive Neutrinos and the Large Scale Structure of the Universe", Phys. Rev. Lett. **45**, [,61(1980)], 1980–1984 (1980).

[83]J. Lesgourgues and S. Pastor, "Massive neutrinos and cosmology", Phys. Rept. **429**, 307–379 (2006), arXiv:astro-ph/0603494 [astro-ph].

[84]S. D. M. White, C. S. Frenk, and M. Davis, "Clustering in a Neutrino Dominated Universe", Astrophys. J. **274**, [,80(1984)], L1–L5 (1983).

[85]A. J. Cuesta, V. Niro, and L. Verde, "Neutrino mass limits: robust information from the power spectrum of galaxy surveys", Phys. Dark Univ. **13**, 77–86 (2016), arXiv:1511.05983.

[86]N. Palanque-Delabrouille et al., "Neutrino masses and cosmology with Lyman-alpha forest power spectrum", JCAP **1511**, 011 (2015), arXiv:1506.05976 [astro-ph.CO].

[87]E. Di Valentino, E. Giusarma, O. Mena, A. Melchiorri, and J. Silk, "Cosmological limits on neutrino unknowns versus low redshift priors", Phys. Rev. **D93**, 083527 (2016), arXiv:1511.00975 [astro-ph.CO].

[88]G. Drexlin, V. Hannen, S. Mertens, and C. Weinheimer, "Current direct neutrino mass experiments", Adv. High Energy Phys. **2013**, 293986 (2013), arXiv:1307.0101.

[89]S. Tremaine and J. E. Gunn, "Dynamical Role of Light Neutral Leptons in Cosmology", Phys. Rev. Lett. **42**, [,66(1979)], 407–410 (1979).

[90]R. Barkana, N. J. Outmezguine, D. Redigolo, and T. Volansky, "Strong constraints on light dark matter interpretation of the EDGES signal", Phys. Rev. **D98**, 103005 (2018), arXiv:1803.03091 [hep-ph].

[91]W. Hu, R. Barkana, and A. Gruzinov, "Cold and fuzzy dark matter", Phys. Rev. Lett. **85**, 1158–1161 (2000), arXiv:astro-ph/0003365 [astro-ph].

[92]L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, "Ultralight scalars as cosmological dark matter", Phys. Rev. **D95**, 043541 (2017), arXiv:1610.08297 [astro-ph.CO].

[93]A. Ibarra, D. Tran, and C. Weniger, "Indirect Searches for Decaying Dark Matter", Int. J. Mod. Phys. **A28**, 1330040 (2013), arXiv:1307.6434 [hep-ph].

[94]V. K. Narayanan, D. N. Spergel, R. Dave, and C.-P. Ma, "Constraints on the mass of warm dark matter particles and the shape of the linear power spectrum from the Lyα forest", Astrophys. J. **543**, L103–L106 (2000), arXiv:astro-ph/0005095 [astro-ph].

[95]M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, "Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest", Phys. Rev. **D71**, 063534 (2005), arXiv:astro-ph/0501562 [astro-ph].

[96]F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, "ETHOS—an effective theory of structure formation: From dark particle physics to the matter distribution of the Universe", Phys. Rev. **D93**, 123527 (2016), arXiv:1512.05344 [astro-ph.CO].

[97]D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tittley, "The non-gravitational interactions of dark matter in colliding galaxy clusters", Science **347**, 1462–1465 (2015), arXiv:1503.07675 [astro-ph.CO].

[98]R. J. Scherrer and M. S. Turner, "On the relic, cosmic abundance of stable, weakly interacting massive particles", Physical Review D **33**, 1585 (1986).

[99]H. Goldberg, "Constraint on the photino mass from cosmology", Physical Review Letters **50**, 1419 (1983).

[100]J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. Olive, and M. Srednicki, "Supersymmetric relics from the big bang", Nuclear Physics B **238**, 453–476 (1984).

[101]G. Jungman, M. Kamionkowski, and K. Griest, "Supersymmetric dark matter", Phys. Rept. **267**, 195–373 (1996), arXiv:hep-ph/9506380 [hep-ph].

[102]R. D. Peccei and H. R. Quinn, "CP Conservation in the Presence of Instantons", Phys. Rev. Lett. **38**, [,328(1977)], 1440–1443 (1977).

[103]F. Wilczek, "Problem of Strong P and T Invariance in the Presence of Instantons", Phys. Rev. Lett. **40**, 279–282 (1978).

[104]S. Weinberg, "A New Light Boson?", Phys. Rev. Lett. **40**, 223–226 (1978).

[105]J. Preskill, M. B. Wise, and F. Wilczek, "Cosmology of the Invisible Axion", Phys. Lett. **B120**, [,URL(1982)], 127–132 (1983).

[106]S. Dodelson and L. M. Widrow, "Sterile-neutrinos as dark matter", Phys. Rev. Lett. **72**, 17–20 (1994), arXiv:hep-ph/9303287 [hep-ph].

[107]X.-D. Shi and G. M. Fuller, "A New dark matter candidate: Nonthermal sterile neutrinos", Phys. Rev. Lett. **82**, 2832–2835 (1999), arXiv:astro-ph/9810076 [astro-ph].

[108]K. Abazajian, G. M. Fuller, and M. Patel, "Sterile neutrino hot, warm, and cold dark matter", Phys. Rev. **D64**, 023501 (2001), arXiv:astro-ph/0101524 [astro-ph].

[109]T. Asaka, S. Blanchet, and M. Shaposhnikov, "The nuMSM, dark matter and neutrino masses", Phys. Lett. **B631**, 151–156 (2005), arXiv:hep-ph/0503065 [hep-ph].

[110]A. Boyarsky, O. Ruchayskiy, and M. Shaposhnikov, "The Role of sterile neutrinos in cosmology and astrophysics", Ann. Rev. Nucl. Part. Sci. **59**, 191–214 (2009), arXiv:0901.0011 [hep-ph].

[111]T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, "The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter", Phys. Dark Univ. **12**, 1–23 (2016), arXiv:1402.6703.

[112]E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, and S. W. Randall, "Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters", Astrophys. J. **789**, 13 (2014), arXiv:1402.2301 [astro-ph.CO].

[113]A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, and J. Franse, "Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster", Phys. Rev. Lett. **113**, 251301 (2014), arXiv:1402.4119 [astro-ph.CO].

[114]M. Aguilar et al., "First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV", Phys. Rev. Lett. **110**, 141102 (2013).

[115]G. Ambrosi et al., "Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons", Nature **552**, 63–66 (2017), arXiv:1711.10981 [astro-ph.HE].

[116]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).

[117]A. Aguilar-Arevalo et al., "Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam", Phys. Rev. **D64**, 112007 (2001), arXiv:hep-ex/0104049 [hep-ex].

[118]A. Aguilar-Arevalo et al., "Significant Excess of ElectronLike Events in the MiniBooNE Short-Baseline Neutrino Experiment", Phys. Rev. Lett. **121**, 221801 (2018), arXiv:1805.12028 [hep-ex].

[119]G. W. Bennett et al., "Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL", Phys. Rev. **D73**, 072003 (2006), arXiv:hep-ex/0602035 [hep-ex].

[120]D. S. Akerib et al., "Results from a search for dark matter in the complete LUX exposure", Phys. Rev. Lett. **118**, 021303 (2017), arXiv:1608.07648 [astro-ph.CO].

[121]E. Aprile et al., "Dark Matter Search Results from a One Ton-Year Exposure of XENON1T", Phys. Rev. Lett. **121**, 111302 (2018), arXiv:1805.12562 [astro-ph.CO].

[122]R. Agnese et al., "Silicon Detector Dark Matter Results from the Final Exposure of CDMS II", Phys. Rev. Lett. **111**, 251301 (2013), arXiv:1304.4279 [hep-ex].

[123]P. W. Graham and S. Rajendran, "New Observables for Direct Detection of Axion Dark Matter", Phys. Rev. **D88**, 035023 (2013), arXiv:1306.6088 [hep-ph].

[124]P. Sikivie, "Experimental Tests of the Invisible Axion", Phys. Rev. Lett. **51**, [,321(1983)], 1415–1417 (1983).

[125]R. K. Leane, T. R. Slatyer, J. F. Beacom, and K. C. Y. Ng, "GeV-scale thermal WIMPs: Not even slightly ruled out", Phys. Rev. **D98**, 023016 (2018), arXiv:1805.10305 [hep-ph].

[126]G. G. Raffelt, "Astrophysical axion bounds", Lect. Notes Phys. **741**, [,51(2006)], 51–71 (2008), arXiv:hep-ph/0611350 [hep-ph].

[127]V. Khachatryan et al., "Search for dark matter, extra dimensions, and unparticles in monojet events in proton–proton collisions at $\sqrt{s} = 8$ TeV", Eur. Phys. J. **C75**, 235 (2015), arXiv:1408.3583 [hep-ex].

[128]M. Aaboud et al., "Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector", Phys. Rev. **D94**, 032005 (2016), arXiv:1604.07773 [hep-ex].

[129]G. Aad et al., "Search for dark matter in events with a hadronically decaying W or Z boson and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", Phys. Rev. Lett. **112**, 041802 (2014), arXiv:1309.4017 [hep-ex].

[130]M. Bauer, M. Neubert, and A. Thamm, "Collider probes of axion-like particles", Journal of High Energy Physics **2017**, 44 (2017).

[131]M. Battaglieri et al., "US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report", (2017), arXiv:1707.04591 [hep-ph].

[132]J. Alexander et al., "Dark Sectors 2016 Workshop: Community Report", in (2016), arXiv:1608.08632 [hep-ph].

[133]R. Essig et al., "Working Group Report: New Light Weakly Coupled Particles", in Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013 (2013), arXiv:1311.0029 [hep-ph].

[134]K. Schutz, T. Lin, B. R. Safdi, and C.-L. Wu, "Constraining a Thin Dark Matter Disk with Gaia", Phys. Rev. Lett. **121**, 081101 (2018), arXiv:1711.03103 [astro-ph.GA].

[135]K. Schutz and T. R. Slatyer, "Self-Scattering for Dark Matter with an Excited State", JCAP **1501**, 021 (2015), arXiv:1409.2867 [hep-ph].

[136]M. Vogelsberger, J. Zavala, K. Schutz, and T. R. Slatyer, "Evaporating the Milky Way halo and its satellites with inelastic self-interacting dark matter", (2018), arXiv:1805.03203.

[137]Y. Hochberg, E. Kuflik, R. Mcgehee, H. Murayama, and K. Schutz, "Strongly interacting massive particles through the axion portal", Phys. Rev. **D98**, 115031 (2018), arXiv:1806.10139 [hep-ph].

[138]C. Dvorkin, T. Lin, and K. Schutz, "Making dark matter out of light: freeze-in from plasma effects", (2019), arXiv:1902.08623 [hep-ph].

[139]N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, "A Theory of Dark Matter", Phys. Rev. **D79**, 015014 (2009), arXiv:0810.0713 [hep-ph].

[140]D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, "Atomic Dark Matter", JCAP **1005**, 021 (2010), arXiv:0909.0753 [hep-ph].

[141]D. E. Kaplan, M. A. Luty, and K. M. Zurek, "Asymmetric Dark Matter", Phys. Rev. **D79**, 115016 (2009), arXiv:0901.4117 [hep-ph].

[142]D. S. M. Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, "Composite Inelastic Dark Matter", Phys. Lett. **B692**, 323–326 (2010), arXiv:0903.3945 [hep-ph].

[143]J. M. Cline, Z. Liu, and W. Xue, "Millicharged Atomic Dark Matter", Phys. Rev. **D85**, 101302 (2012), arXiv:1201.4858 [hep-ph].

[144]J. M. Cline, Z. Liu, G. Moore, and W. Xue, "Composite strongly interacting dark matter", Phys. Rev. **D90**, 015023 (2014), arXiv:1312.3325 [hep-ph].

[145]R. Foot, "Mirror dark matter: Cosmology, galaxy structure and direct detection", Int. J. Mod. Phys. **A29**, 1430013 (2014), arXiv:1401.3965 [astro-ph.CO].

[146]J. F. Cherry, A. Friedland, and I. M. Shoemaker, "Neutrino Portal Dark Matter: From Dwarf Galaxies to IceCube", (2014), arXiv:1411.1071 [hep-ph].

[147]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].

[148]M. A. Buen-Abad, G. Marques-Tavares, and M. Schmaltz, "Non-Abelian dark matter and dark radiation", Phys. Rev. **D92**, 023531 (2015), arXiv:1505.03542 [hep-ph].

[149]S. Tulin, H.-B. Yu, and K. M. Zurek, "Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure", Phys. Rev. **D87**, 115007 (2013), arXiv:1302.3898 [hep-ph].

[150]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].

[151]L. Randall, J. Scholtz, and J. Unwin, "Cores in Dwarf Galaxies from Fermi Repulsion", Mon. Not. Roy. Astron. Soc. **467**, 1515–1525 (2017), arXiv:1611.04590 [astro-ph.GA].

[152]J. Fan, A. Katz, L. Randall, and M. Reece, "Dark-Disk Universe", Phys. Rev. Lett. **110**, 211302 (2013), arXiv:1303.3271 [hep-ph].

[153]J. Fan, A. Katz, L. Randall, and M. Reece, "Double-Disk Dark Matter", Phys. Dark Univ. **2**, 139–156 (2013), arXiv:1303.1521 [astro-ph.CO].

[154]P. Agrawal, F.-Y. Cyr-Racine, L. Randall, and J. Scholtz, "Dark Catalysis", JCAP **1708**, 021 (2017), arXiv:1702.05482 [astro-ph.CO].

[155]Gaia Collaboration, Prusti, et al., "The Gaia mission", Astron. Astrophys. **595**, A1, A1 (2016), arXiv:1609.04153 [astro-ph.IM].

[156]Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, F. Mignard, R. Drimmel, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, and et al., "Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties", Astron. Astrophys. **595**, A2, A2 (2016), arXiv:1609.04172 [astro-ph.IM].

[157]F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, "Constraints on Large-Scale Dark Acoustic Oscillations from Cosmology", Phys. Rev. **D89**, 063517 (2014), arXiv:1310.3278 [astro-ph.CO].

[158]R. Massey et al., "The behaviour of dark matter associated with four bright cluster galaxies in the 10 kpc core of Abell 3827", Mon. Not. Roy. Astron. Soc. **449**, 3393–3406 (2015), arXiv:1504.03388 [astro-ph.CO].

[159]F. Kahlhoefer, K. Schmidt-Hoberg, J. Kummer, and S. Sarkar, "On the interpretation of dark matter self-interactions in Abell 3827", Mon. Not. Roy. Astron. Soc. **452**, L54–L58 (2015), arXiv:1504.06576 [astro-ph.CO].

[160]Z. Chacko, Y. Cui, S. Hong, T. Okui, and Y. Tsai, "Partially Acoustic Dark Matter, Interacting Dark Radiation, and Large Scale Structure", JHEP **12**, 108 (2016), arXiv:1609.03569 [astro-ph.CO].

[161]W. Fischler, D. Lorshbough, and W. Tangarife, "Supersymmetric Partially Interacting Dark Matter", Phys. Rev. **D91**, 025010 (2015), arXiv:1405.7708 [hep-ph].

[162]M. Raveri, W. Hu, T. Hoffman, and L.-T. Wang, "Partially Acoustic Dark Matter Cosmology and Cosmological Constraints", (2017), arXiv:1709.04877 [astro-ph.CO].

[163]R. Foot, "Galactic structure explained with dissipative mirror dark matter", Phys. Rev. **D88**, 023520 (2013), arXiv:1304.4717 [astro-ph.CO].

[164]K. Petraki, L. Pearce, and A. Kusenko, "Self-interacting asymmetric dark matter coupled to a light massive dark photon", JCAP **1407**, 039 (2014), arXiv:1403.1077 [hep-ph].

[165]E. Rosenberg and J. Fan, "Cooling in a Dissipative Dark Sector", (2017), arXiv:1705.10341 [astro-ph.GA].

[166]R. Foot and S. Vagnozzi, "Solving the small-scale structure puzzles with dissipative dark matter", JCAP **1607**, 013 (2016), arXiv:1602.02467 [astro-ph.CO].

[167]F.-Y. Cyr-Racine and K. Sigurdson, "Cosmology of atomic dark matter", Phys. Rev. **D87**, 103515 (2013), arXiv:1209.5752 [astro-ph.CO].

[168]K. K. Boddy, M. Kaplinghat, A. Kwa, and A. H. G. Peter, "Hidden Sector Hydrogen as Dark Matter: Small-scale Structure Formation Predictions and the Importance of Hyperfine Interactions", Phys. Rev. **D94**, 123017 (2016), arXiv:1609.03592 [hep-ph].

[169]J. M. Cline, Z. Liu, G. Moore, and W. Xue, "Scattering properties of dark atoms and molecules", Phys. Rev. **D89**, 043514 (2014), arXiv:1311.6468 [hep-ph].

[170]A. Das and B. Dasgupta, "New Dissipation Mechanisms from Multi-level Dark Matter Scattering", (2017), arXiv:1709.06577 [hep-ph].

[171]M. R. Buckley and A. DiFranzo, "Collapsed Dark Matter Structures", (2017), arXiv:1707.03829 [hep-ph].

[172]L. Randall and M. Reece, "Dark Matter as a Trigger for Periodic Comet Impacts", Phys. Rev. Lett. **112**, 161301 (2014), arXiv:1403.0576 [astro-ph.GA].

[173]L. Randall and J. Scholtz, "Dissipative Dark Matter and the Andromeda Plane of Satellites", JCAP **1509**, 057 (2015), arXiv:1412.1839 [astro-ph.GA].

[174]R. Foot and Z. K. Silagadze, "Thin disk of co-rotating dwarfs: A fingerprint of dissipative (mirror) dark matter?", Phys. Dark Univ. **2**, 163–165 (2013), arXiv:1306.1305 [astro-ph.GA].

[175]P. Agrawal and L. Randall, "Point Sources from Dissipative Dark Matter", (2017), arXiv:1706.04195 [hep-ph].

[176]S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, "Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy", Phys. Rev. Lett. **116**, 051103 (2016), arXiv:1506.05124 [astro-ph.HE].

[177]A. Caputo, J. Zavala, and D. Blas, "Binary pulsars as probes of a Galactic dark matter disk", Phys. Dark Univ. **19**, 1–11 (2018), arXiv:1709.03991 [astro-ph.HE].

[178]G. D'Amico, P. Panci, A. Lupi, S. Bovino, and J. Silk, "Massive Black Holes from Dissipative Dark Matter", (2017), arXiv:1707.03419 [astro-ph.CO].

[179]M. McCullough and L. Randall, "Exothermic Double-Disk Dark Matter", JCAP **1310**, 058 (2013), arXiv:1307.4095 [hep-ph].

[180]J. Fan, A. Katz, and J. Shelton, "Direct and indirect detection of dissipative dark matter", JCAP **1406**, 059 (2014), arXiv:1312.1336 [hep-ph].

[181]E. D. Kramer and M. Rowan, "Revisiting the Dark Matter - Comet Shower Connection", (2016), arXiv:1610.04239 [astro-ph.EP].

[182]N. J. Shaviv, "The Paleoclimatic evidence for Strongly Interacting Dark Matter Present in the Galactic Disk", (2016), arXiv:1606.02851 [astro-ph.GA].

[183]L. Lindegren et al., "Gaia Data Release 1. Astrometry: one billion positions, two million proper motions and parallaxes", Astron. Astrophys. **595**, A4, A4 (2016), arXiv:1609.04303.

[184]D. Michalik, L. Lindegren, and D. Hobbs, "The Tycho-Gaia astrometric solution . How to get 2.5 million parallaxes with less than one year of Gaia data", Astron. Astrophys. **574**, A115, A115 (2015), arXiv:1412.8770 [astro-ph.IM].

[185]E. Høg, C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, G. Wycoff, U. Bastian, P. Schwekendiek, and A. Wicenec, "The Tycho-2 catalogue of the 2.5 million brightest stars", Astron. Astrophys. **355**, L27–L30 (2000).

[186]M. A. C. Perryman et al., "The Hipparcos catalogue", Astron. Astrophys. **323**, L49–L52 (1997).

[187]E. D. Kramer and L. Randall, "Updated Kinematic Constraints on a Dark Disk", Astrophys. J. **824**, 116 (2016), arXiv:1604.01407 [astro-ph.GA].

[188]J. Holmberg and C. Flynn, "The local density of matter mapped by hipparcos", Mon. Not. Roy. Astron. Soc. **313**, 209–216 (2000), arXiv:astro-ph/9812404 [astro-ph].

[189]K. Kuijken and G. Gilmore, "The Mass Distribution in the Galactic Disc - II - Determination of the Surface Mass Density of the Galactic Disc Near the Sun", Mon. Not. R. Astron. Soc. **239**, 605–649 (1989).

[190]K. Kuijken and G. Gilmore, "The galactic disk surface mass density and the Galactic force K(z) at Z = 1.1 kiloparsecs", ApJL **367**, L9–L13 (1991).

[191]K. Kuijken and G. Gilmore, "The mass distribution in the galactic disc. I - A technique to determine the integral surface mass density of the disc near the sun.", MNRAS **239**, 571–603 (1989).

[192]J. Bovy and H.-W. Rix, "A Direct Dynamical Measurement of the Milky Way's Disk Surface Density Profile, Disk Scale Length, and Dark Matter Profile at 4 kpc $\lesssim$ R $\lesssim$ 9 kpc", Astrophys. J. **779**, 115 (2013), arXiv:1309.0809 [astro-ph.GA].

[193]C. F. McKee, A. Parravano, and D. J. Hollenbach, "Stars, Gas, and Dark Matter in the Solar Neighborhood", Astrophys. J. **814**, 13 (2015), arXiv:1509.05334 [astro-ph.GA].

[194]H. Silverwood, S. Sivertsson, P. Steger, J. I. Read, and G. Bertone, "A non-parametric method for measuring the local dark matter density", Mon. Not. Roy. Astron. Soc. **459**, 4191–4208 (2016), arXiv:1507.08581 [astro-ph.GA].

[195]Pasetto, S., Grebel, E. K., Zwitter, T., Chiosi, C., Bertelli, G., Bienayme, O., Seabroke, G., Bland-Hawthorn, J., Boeche, C., Gibson, B. K., Gilmore, G., Munari, U., Navarro, J. F., Parker, Q., Reid, W., Silviero, A., and Steinmetz, M., "Thin disk kinematics from rave and the solar motion", A&A **547**, A71 (2012).

[196]G. Monari, B. Famaey, and A. Siebert, "The vertical effects of disc non-axisymmetries from perturbation theory: the case of the Galactic bar", Mon. Not. Roy. Astron. Soc. **452**, 747–754 (2015), arXiv:1505.07456 [astro-ph.GA].

[197]S. Garbari, J. I. Read, and G. Lake, "Limits on the local dark matter density", Mon. Not. Roy. Astron. Soc. **416**, 2318–2340 (2011), arXiv:1105.6339 [astro-ph.GA].

[198]L. Zhang, H.-W. Rix, G. van de Ven, J. Bovy, C. Liu, and G. Zhao, "The Gravitational Potential Near the Sun From SEGUE K-dwarf Kinematics", Astrophys. J. **772**, 108 (2013), arXiv:1209.0256 [astro-ph.GA].

[199]J. N. Bahcall, "Self-consistent determinations of the total amount of matter near the sun", The Astrophysical Journal **276**, 169–181 (1984).

[200]J. Bovy, "Galactic rotation in Gaia DR1", Mon. Not. R. Astron. Soc. **468**, L63–L67 (2017), arXiv:1610.07610.

[201]S. Garbari, C. Liu, J. I. Read, and G. Lake, "A new determination of the local dark matter density from the kinematics of k dwarfs", Monthly Notices of the Royal Astronomical Society **425**, 1445–1458 (2012).

[202]J. Bovy and S. Tremaine, "On the local dark matter density", Astrophys. J. **756**, 89 (2012), arXiv:1205.4033 [astro-ph.GA].

[203]J. Binney and S. Tremaine, *Galactic dynamics* (Princeton university press, 2011).

[204]E. D. Kramer and L. Randall, "Interstellar Gas and a Dark Disk", Astrophys. J. **829**, 126 (2016), arXiv:1603.03058 [astro-ph.GA].

[205]C. Flynn, J. Holmberg, L. Portinari, B. Fuchs, and H. Jahreiss, "On the mass-to-light ratio of the local Galactic disc and the optical luminosity of the Galaxy", Mon. Not. Roy. Astron. Soc. **372**, 1149–1160 (2006), arXiv:astro-ph/0608193 [astro-ph].

[206]J. I. Read, "The Local Dark Matter Density", J. Phys. **G41**, 063101 (2014), arXiv:1404.1938 [astro-ph.GA].

[207]M. J. Pecaut and E. E. Mamajek, "Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars", ApJS **208**, 9, 9 (2013), arXiv:1307.2657 [astro-ph.SR].

[208]L. Anderson, D. W. Hogg, B. Leistedt, A. M. Price-Whelan, and J. Bovy, "Improving *Gaia* parallax precision with a data-driven model of stars", ArXiv e-prints (2017), arXiv:1706.05055.

[209]P. J. McMillan, G. Kordopatis, A. Kunder, J. Binney, J. Wojno, T. Zwitter, M. Steinmetz, J. Bland-Hawthorn, B. K. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, U. Munari, J. F. Navarro, Q. A. Parker, G. Seabroke, and R. F. G. Wyse, "Improved distances to stars common to TGAS and RAVE", ArXiv e-prints (2017), arXiv:1707.04554.

[210]M. F. Skrutskie et al., "The Two Micron All Sky Survey (2MASS)", Astron. J. **131**, 1163–1183 (2006).

[211]"Local kinematics and the local standard of rest", Monthly Notices of the Royal Astronomical Society **403**, 1829–1833 (2010).

[212]J. Bovy, D. Kawata, and J. A. S. Hunt, "Made-to-measure modeling of observed galaxy dynamics", ArXiv e-prints (2017), arXiv:1704.03884.

[213]L. M. Widrow, S. Gardner, B. Yanny, S. Dodelson, and H.-Y. Chen, "Galactoseismology: Discovery of Vertical Waves in the Galactic Disk", Astrophys. J. **750**, L41 (2012), arXiv:1203.6861 [astro-ph.GA].

[214]M. E. K. Williams et al., "The wobbly Galaxy: kinematics north and south with RAVE red clump giants", Mon. Not. Roy. Astron. Soc. **436**, 101 (2013), arXiv:1302.2468 [astro-ph.GA].

[215]J. L. Carlin et al., "Substructure in bulk velocities of Milky Way disk stars", Astrophys. J. **777**, L5 (2013), arXiv:1309.6314 [astro-ph.GA].

[216]B. Yanny and S. Gardner, "The Stellar Number Density Distribution in the Local Solar Neighborhood is North-South Asymmetric", Astrophys. J. **777**, 91 (2013), arXiv:1309.2300.

[217]J. Holmberg, C. Flynn, and L. Lindegren, "Towards an Improved Model of the Galaxy", in Hipparcos - venice '97, Vol. 402, edited by R. M. Bonnet, E. Høg, P. L. Bernacca, L. Emiliani, A. Blaauw, C. Turon, J. Kovalevsky, L. Lindegren, H. Hassan, M. Bouffard, B. Strim, D. Heger, M. A. C. Perryman, and L. Woltjer, ESA Special Publication (Aug. 1997), pp. 721–726.

[218]N. J. Shaviv, "A kinematic spiral arm shock signature: "Ringing" in the vertical motion of stars", (2016), arXiv:1606.02595 [astro-ph.GA].

[219]N. Banik, L. M. Widrow, and S. Dodelson, "Galactoseismology and the local density of dark matter", Mon. Not. R. Astron. Soc. **464**, 3775–3783 (2017), arXiv:1608.03338.

[220]Y. C. Joshi, "Displacement of the Sun from the Galactic plane", Mon. Not. R. Astron. Soc. **378**, 768–776 (2007), arXiv:0704.0950.

[221]D. Majaess, D. Turner, and D. Lane, "Characteristics of the Galaxy according to Cepheids", Mon. Not. Roy. Astron. Soc. **398**, 263–270 (2009), arXiv:0903.4206 [astro-ph.GA].

[222]G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. **C71**, [Erratum: Eur. Phys. J.C73,2501(2013)], 1554 (2011), arXiv:1007.1727 [physics.data-an].

[223]C. J. Jog, "Q criterion for disc stability modified by external tidal field", (2013), arXiv:1308.1754 [astro-ph.GA].

[224]S. Gillessen, P. M. Plewa, F. Eisenhauer, R. Sari, I. Waisberg, M. Habibi, O. Pfuhl, E. George, J. Dexter, S. von Fellenberg, T. Ott, and R. Genzel, "An Update on Monitoring Stellar Orbits in the Galactic Center", Astrophys. J. **837**, 30, 30 (2017), arXiv:1611.09144.

[225]P. J. McMillan, "The mass distribution and gravitational potential of the Milky Way", Mon. Not. R. Astron. Soc. **465**, 76–94 (2017), arXiv:1608.00971.

[226]A. Toomre, "On the gravitational stability of a disk of stars", The Astrophysical Journal **139**, 1217–1238 (1964).

[227]W. J. Quirk, "On the gas content of galaxies", The Astrophysical Journal **176**, L9 (1972).

[228]R. R. Rafikov, "Local axisymmetric instability criterion in the thin, rotating, multicomponent disk", Mon. Not. Roy. Astron. Soc. **323**, 445–452 (2001), arXiv:astro-ph/0007058.

[229]A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, "Where Are the Missing Galactic Satellites?", Astrophys. J. **522**, 82–92 (1999), arXiv:astro-ph/9901240 [astro-ph].

[230]B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi, "Dark Matter Substructure within Galactic Halos", Astrophys. J. **524**, L19–L22 (1999), arXiv:astro-ph/9907411 [astro-ph].

[231]J. Zavala, Y. P. Jing, A. Faltenbacher, G. Yepes, Y. Hoffman, S. Gottlöber, and B. Catinella, "The Velocity Function in the Local Environment from ΛCDM and ΛWDM Constrained Simulations", Astrophys. J. **700**, 1779–1793 (2009), arXiv:0906.0585 [astro-ph.CO].

[232]E. Papastergis, A. M. Martin, R. Giovanelli, and M. P. Haynes, "The Velocity Width Function of Galaxies from the 40% ALFALFA Survey: Shedding Light on the Cold Dark Matter Overabundance Problem", Astrophys. J. **739**, 38, 38 (2011), arXiv:1106.0710 [astro-ph.CO].

[233]A. Klypin, I. Karachentsev, D. Makarov, and O. Nasonova, "Abundance of field galaxies", Mon. Not. R. Astron. Soc. **454**, 1798–1810 (2015), arXiv:1405.4523 [astro-ph.CO].

[234]W. J. G. de Blok and S. S. McGaugh, "The dark and visible matter content of low surface brightness disc galaxies", Mon. Not. R. Astron. Soc. **290**, 533–552 (1997), arXiv:astro-ph/9704274 [astro-ph].

[235]M. G. Walker and J. Peñarrubia, "A Method for Measuring (Slopes of) the Mass Profiles of Dwarf Spheroidal Galaxies", Astrophys. J. **742**, 20, 20 (2011), arXiv:1108.2404.

[236]M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, "Too big to fail? The puzzling darkness of massive Milky Way subhaloes", Mon. Not. Roy. Astron. Soc. **415**, L40 (2011), arXiv:1103.0007 [astro-ph.CO].

[237]E. Papastergis, R. Giovanelli, M. P. Haynes, and F. Shankar, "Is there a "too big to fail" problem in the field?", Astron. Astrophys. **574**, A113, A113 (2015), arXiv:1407.4665.

[238]K. A. Oman, J. F. Navarro, A. Fattahi, C. S. Frenk, T. Sawala, S. D. M. White, R. Bower, R. A. Crain, M. Furlong, M. Schaller, J. Schaye, and T. Theuns, "The unexpected diversity of dwarf galaxy rotation curves", Mon. Not. R. Astron. Soc. **452**, 3650–3665 (2015), arXiv:1504.01437 [astro-ph.GA].

[239]K. A. Oman, J. F. Navarro, L. V. Sales, A. Fattahi, C. S. Frenk, T. Sawala, M. Schaller, and S. D. M. White, "Missing dark matter in dwarf galaxies?", Mon. Not. R. Astron. Soc. **460**, 3610–3623 (2016), arXiv:1601.01026 [astro-ph.GA].

[240]J. F. Navarro, V. R. Eke, and C. S. Frenk, "The cores of dwarf galaxy haloes", Mon. Not. R. Astron. Soc. **283**, L72–L78 (1996), arXiv:astro-ph/9610187 [astro-ph].

[241]F. Governato, A. Zolotov, A. Pontzen, C. Christensen, S. H. Oh, A. M. Brooks, T. Quinn, S. Shen, and J. Wadsley, "Cuspy no more: how outflows affect the central dark matter and baryon distribution in Λ cold dark matter galaxies", Mon. Not. R. Astron. Soc. **422**, 1231–1240 (2012), arXiv:1202.0554 [astro-ph.CO].

[242]J. Oñorbe, M. Boylan-Kolchin, J. S. Bullock, P. F. Hopkins, D. Keres, C.-A. Faucher-Giguère, E. Quataert, and N. Murray, "Forged in FIRE: cusps, cores and baryons in low-mass dwarf galaxies", Mon. Not. R. Astron. Soc. **454**, 2092–2106 (2015), arXiv:1502.02036.

[243]T. K. Chan, D. Keres, J. Oñorbe, P. F. Hopkins, A. L. Muratov, C. .-.-A. Faucher-Giguère, and E. Quataert, "The impact of baryonic physics on the structure of dark matter haloes: the view from the FIRE cosmological simulations", Mon. Not. R. Astron. Soc. **454**, 2981–3001 (2015), arXiv:1507.02282 [astro-ph.GA].

[244]J. I. Read, O. Agertz, and M. L. M. Collins, "Dark matter cores all the way down", Mon. Not. R. Astron. Soc. **459**, 2573–2590 (2016), arXiv:1508.04143 [astro-ph.GA].

[245]A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, and J. Wadsley, "Baryons Matter: Why Luminous Satellite Galaxies have Reduced Central Masses", Astrophys. J. **761**, 71, 71 (2012), arXiv:1207.0007.

[246]C. B. Brook and A. Di Cintio, "Expanded haloes, abundance matching and too-big-to-fail in the Local Group", Mon. Not. R. Astron. Soc. **450**, 3920–3934 (2015), arXiv:1410.3825 [astro-ph.GA].

[247]T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, R. G. Bower, R. A. Crain, C. Dalla Vecchia, M. Furlong, J. Helly, A. Jenkins, K. Oman, M. Schaller, J. Schaye, T. Theuns, J. Trayford, and S. D. M. White, "The APOSTLE simulations: solutions to the Local Group's cosmic puzzles", Mon. Not. R. Astron. Soc. **457**, 1931–1943 (2016), arXiv:1511.01098 [astro-ph.GA].

[248]A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguère, D. Keres, and E. Quataert, "Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy", Astrophys. J. **827**, L23, L23 (2016), arXiv:1602.05957 [astro-ph.GA].

[249]S. Koposov, V. Belokurov, N. W. Evans, P. C. Hewett, M. J. Irwin, G. Gilmore, D. B. Zucker, H. .-.-W. Rix, M. Fellhauer, E. F. Bell, and E. V. Glushkova, "The Luminosity Function of the Milky Way Satellites", Astrophys. J. **686**, 279–291 (2008), arXiv:0706.2687 [astro-ph].

[250]S. Y. Kim, A. H. G. Peter, and J. R. Hargis, "There is No Missing Satellites Problem", arXiv e-prints, arXiv:1711.06267, arXiv:1711.06267 (2017), arXiv:1711.06267 [astro-ph.CO].

[251]A. M. Brooks, E. Papastergis, C. R. Christensen, F. Governato, A. Stilp, T. R. Quinn, and J. Wadsley, "How to Reconcile the Observed Velocity Function of Galaxies with Theory", Astrophys. J. **850**, 97, 97 (2017), arXiv:1701.07835 [astro-ph.GA].

[252]A. A. Dutton, A. V. Macciò, J. Frings, L. Wang, G. S. Stinson, C. Penzo, and X. Kang, "NIHAO V: too big does not fail - reconciling the conflict between ΛCDM predictions and the circular velocities of nearby field galaxies", Mon. Not. R. Astron. Soc. **457**, L74–L78 (2016), arXiv:1512.00453 [astro-ph.GA].

[253]R. Verbeke, E. Papastergis, A. A. Ponomareva, S. Rathi, and S. De Rijcke, "A new astrophysical solution to the Too Big To Fail problem. Insights from the moria simulations", Astron. Astrophys. **607**, A13, A13 (2017), arXiv:1703.03810 [astro-ph.GA].

[254]I. M. Santos-Santos, A. Di Cintio, C. B. Brook, A. Macciò, A. Dutton, and R. Domínguez-Tenreiro, "NIHAO - XIV. Reproducing the observed diversity of dwarf galaxy rotation curve shapes in ΛCDM", Mon. Not. R. Astron. Soc. **473**, 4392–4403 (2018), arXiv:1706.04202 [astro-ph.GA].

[255]J. Peñarrubia, A. Pontzen, M. G. Walker, and S. E. Koposov, "The Coupling between the Core/Cusp and Missing Satellite Problems", Astrophys. J. **759**, L42, L42 (2012), arXiv:1207.2772 [astro-ph.GA].

[256]A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. Knebe, A. A. Dutton, and J. Wadsley, "The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores", Mon. Not. R. Astron. Soc. **437**, 415–423 (2014), arXiv:1306.0898 [astro-ph.CO].

[257]G. Kauffmann, "Quantitative constraints on starburst cycles in galaxies with stellar masses in the range 10^8-10^{10} $M_\odot$", Mon. Not. R. Astron. Soc. **441**, 2717–2724 (2014).

[258]D. R. Weisz, A. E. Dolphin, E. D. Skillman, J. Holtzman, K. M. Gilbert, J. J. Dalcanton, and B. F. Williams, "The Star Formation Histories of Local Group Dwarf Galaxies. I. Hubble Space Telescope/Wide Field Planetary Camera 2 Observations", Astrophys. J. **789**, 147, 147 (2014), arXiv:1404.7144 [astro-ph.GA].

[259]D. N. Spergel and P. J. Steinhardt, "Observational Evidence for Self-Interacting Cold Dark Matter", Physical Review Letters **84**, 3760–3763 (2000), eprint: astro-ph/9909386.

[260]P. Colín, V. Avila-Reese, O. Valenzuela, and C. Firmani, "Structure and Subhalo Population of Halos in a Self-interacting Dark Matter Cosmology", Astrophys. J. **581**, 777–793 (2002), arXiv:astro-ph/0205322 [astro-ph].

[261]M. Vogelsberger, J. Zavala, and A. Loeb, "Subhaloes in self-interacting galactic dark matter haloes", Mon. Not. R. Astron. Soc. **423**, 3740–3752 (2012), arXiv:1201.5892.

[262]M. Rocha, A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Oñorbe, and L. A. Moustakas, "Cosmological simulations with self-interacting dark matter - I. Constant-density cores and substructure", Mon. Not. R. Astron. Soc. **430**, 81–104 (2013), arXiv:1208.3025 [astro-ph.CO].

[263]J. Zavala, M. Vogelsberger, and M. G. Walker, "Constraining self-interacting dark matter with the Milky way's dwarf spheroidals.", Mon. Not. R. Astron. Soc. **431**, L20–L24 (2013), arXiv:-1211.6426 [astro-ph.CO].

[264]A. Kamada, M. Kaplinghat, A. B. Pace, and H.-B. Yu, "Self-Interacting Dark Matter Can Explain Diverse Galactic Rotation Curves", Phys. Rev. Lett. **119**, 111102, 111102 (2017), arXiv:-1611.02716 [astro-ph.GA].

[265]P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala, "Spreading out and staying sharp - creating diverse rotation curves via baryonic and self-interaction effects", Mon. Not. R. Astron. Soc. **468**, 2283–2295 (2017), arXiv:1612.03903 [astro-ph.GA].

[266]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), arXiv:1211.6426 [astro-ph.CO].

[267]F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, "ETHOS—an effective theory of structure formation: From dark particle physics to the matter distribution of the Universe", Phys. Rev. D **93**, 123527, 123527 (2016), arXiv:1512.05344 [astro-ph.CO].

[268]M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson, "ETHOS - an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems", Mon. Not. R. Astron. Soc. **460**, 1399–1416 (2016), arXiv:1512.05349 [astro-ph.CO].

[269]M. R. Lovell, J. Zavala, M. Vogelsberger, X. Shen, F.-Y. Cyr-Racine, C. Pfrommer, K. Sigurdson, M. Boylan-Kolchin, and A. Pillepich, "ETHOS - an effective theory of structure formation: predictions for the high-redshift Universe - abundance of galaxies and reionization", Mon. Not. R. Astron. Soc. **477**, 2886–2899 (2018), arXiv:1711.10497 [astro-ph.CO].

[270]S. Tulin and H.-B. Yu, "Dark matter self-interactions and small scale structure", Physics Reports **730**, 1–57 (2018), arXiv:1705.02358 [hep-ph].

[271]K. Todoroki and M. V. Medvedev, "Dark matter haloes in the multicomponent model - I. Substructure", Mon. Not. R. Astron. Soc. **483**, 3983–4003 (2019), arXiv:1711.11078.

[272]K. Todoroki and M. V. Medvedev, "Dark matter haloes in the multicomponent model - II. Density profiles of galactic haloes", Mon. Not. R. Astron. Soc. **483**, 4004–4019 (2019), arXiv:1711.11085 [astro-ph.CO].

[273]A. Loeb and N. Weiner, "Cores in Dwarf Galaxies from Dark Matter with a Yukawa Potential", Phys.Rev.Lett. **106**, 171302 (2011), arXiv:1011.6374 [astro-ph.CO].

[274]F. Sanchez-Salcedo, "Unstable cold dark matter and the cuspy halo problem in dwarf galaxies", Astrophys.J. **591**, L107–L110 (2003), arXiv:astro-ph/0305496 [astro-ph].

[275]M. Abdelqader and F. Melia, "Decaying Dark Matter and the Deficit of Dwarf Haloes", Mon. Not. Roy. Astron. Soc. **388**, 1869–1878 (2008), arXiv:0806.0602 [astro-ph].

[276]A. H. G. Peter, "Mapping the allowed parameter space for decaying dark matter models", Phys.Rev. **D81**, 083511 (2010), arXiv:1001.3870 [astro-ph.CO].

[277]A. H. G. Peter and A. J. Benson, "Dark-matter decays and Milky Way satellite galaxies", Phys.Rev. **D82**, 123521 (2010), arXiv:1009.1912 [astro-ph.GA].

[278]A. H. G. Peter, C. E. Moody, and M. Kamionkowski, "Dark-Matter Decays and Self-Gravitating Halos", Phys.Rev. **D81**, 103501 (2010), arXiv:1003.0419 [astro-ph.CO].

[279]N. F. Bell, A. J. Galea, and R. R. Volkas, "A Model For Late Dark Matter Decay", Phys.Rev. **D83**, 063504 (2011), arXiv:1012.0067 [hep-ph].

[280]D. R. Smith and N. Weiner, "Inelastic dark matter", Phys. Rev. **D64**, 043502 (2001), eprint: hep-ph/0101138.

[281]P. W. Graham, R. Harnik, S. Rajendran, and P. Saraswat, "Exothermic Dark Matter", Phys.Rev. **D82**, 063512 (2010), arXiv:1004.0937 [hep-ph].

[282]D. Finkbeiner and N. Weiner, "Exciting Dark Matter and the INTEGRAL/SPI 511 keV signal", Phys. Rev. **D76**, 083519 (2007), eprint: astro-ph/0702587.

[283]J. M. Cline, A. R. Frey, and F. Chen, "Metastable dark matter mechanisms for INTEGRAL 511 keV γ rays and DAMA/CoGeNT events", Phys. Rev. **D83**, 083511 (2011), arXiv:1008.1784 [hep-ph].

[284]D. Finkbeiner and N. Weiner, "X-ray line from exciting dark matter", Phys. Rev. **D94**, 083002 (2016), arXiv:1402.6671 [hep-ph].

[285]K. Schutz and T. R. Slatyer, "Self-scattering for Dark Matter with an excited state", Journal of Cosmology and Astro-Particle Physics **2015**, 021, 021 (2015), arXiv:1409.2867 [hep-ph].

[286]M. Baumgart, C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, "Non-Abelian Dark Sectors and Their Collider Signatures", JHEP **04**, 014 (2009), arXiv:0901.0283 [hep-ph].

[287]C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, "Kinetic Mixing as the Origin of Light Dark Scales", Phys. Rev. **D80**, 035008 (2009), arXiv:0902.3246 [hep-ph].

[288]F. Chen, J. M. Cline, and A. R. Frey, "Nonabelian dark matter: models and constraints", Phys. Rev. **D80**, 083516 (2009), arXiv:0907.4746 [hep-ph].

[289]T. R. Slatyer, "The Sommerfeld enhancement for dark matter with an excited state", JCAP **1002**, 028 (2010), arXiv:0910.5713 [hep-ph].

[290]D. P. Finkbeiner, L. Goodenough, T. R. Slatyer, M. Vogelsberger, and N. Weiner, "Consistent Scenarios for Cosmic-Ray Excesses from Sommerfeld Enhanced Dark Matter Annihilation", JCAP **1105**, 002 (2011), arXiv:1011.3082 [hep-ph].

[291]H. An, P. B. Dev, Y. Cai, and R. Mohapatra, "Sneutrino Dark Matter in Gauged Inverse Seesaw Models for Neutrinos", Phys.Rev.Lett. **108**, 081806 (2012), arXiv:1110.1366 [hep-ph].

[292]S. Cassel, "Sommerfeld factor for arbitrary partial wave processes", J.Phys. **G37**, 105009 (2010), arXiv:0903.5307 [hep-ph].

[293]J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, "Non-perturbative effect on dark matter annihilation and gamma ray signature from galactic center", Phys. Rev. **D71**, 063528 (2005), arXiv:hep-ph/0412403.

[294]D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, "Atomic dark matter", Journal of Cosmology and Astro-Particle Physics **2010**, 021, 021 (2010), arXiv:0909.0753.

[295]J. Fan, A. Katz, L. Randall, and M. Reece, "Double-Disk Dark Matter", Physics of the Dark Universe **2**, 139–156 (2013), arXiv:1303.1521 [astro-ph.CO].

[296]R. Foot and S. Vagnozzi, "Dissipative hidden sector dark matter", Phys. Rev. D **91**, 023512, 023512 (2015), arXiv:1409.7174 [hep-ph].

[297]J. Choquette and J. M. Cline, "Minimal non-Abelian model of atomic dark matter", Phys. Rev. D **92**, 115011, 115011 (2015), arXiv:1509.05764 [hep-ph].

[298]A. Ghalsasi and M. McQuinn, "Exploring the astrophysics of dark atoms", Phys. Rev. D **97**, 123018, 123018 (2018), arXiv:1712.04779 [astro-ph.GA].

[299]F. Chen, J. M. Cline, and A. R. Frey, "Non-Abelian dark matter: Models and constraints", Phys. Rev. D **80**, 083516, 083516 (2009), arXiv:0907.4746 [hep-ph].

[300]M. Cirelli and J. M. Cline, "Can multistate dark matter annihilation explain the high-energy cosmic ray lepton anomalies?", Phys. Rev. D **82**, 023503, 023503 (2010), arXiv:1005.1779 [hep-ph].

[301]M. B. Wise and Y. Zhang, "Stable bound states of asymmetric dark matter", Phys. Rev. D **90**, 055030, 055030 (2014), arXiv:1407.4121 [hep-ph].

[302]M.-Y. Wang, A. H. G. Peter, L. E. Strigari, A. R. Zentner, B. Arant, S. Garrison-Kimmel, and M. Rocha, "Cosmological simulations of decaying dark matter: implications for small-scale structure of dark matter haloes", Mon. Not. R. Astron. Soc. **445**, 614–629 (2014), arXiv:1406.0527 [astro-ph.CO].

[303]A. Loeb and N. Weiner, "Cores in Dwarf Galaxies from Dark Matter with a Yukawa Potential", Phys. Rev. Lett. **106**, 171302, 171302 (2011), arXiv:1011.6374 [astro-ph.CO].

[304]V. Poulin, P. D. Serpico, and J. Lesgourgues, "A fresh look at linear cosmological constraints on a decaying Dark Matter component", Journal of Cosmology and Astro-Particle Physics **2016**, 036, 036 (2016), arXiv:1606.02073 [astro-ph.CO].

[305]D. P. Finkbeiner, T. R. Slatyer, N. Weiner, and I. Yavin, "PAMELA, DAMA, INTEGRAL and signatures of metastable excited WIMPs", Journal of Cosmology and Astro-Particle Physics **2009**, 037, 037 (2009), arXiv:0903.1037 [hep-ph].

[306]J. M. Cline, A. R. Frey, and F. Chen, "Metastable dark matter mechanisms for INTEGRAL 511 keV γ rays and DAMA/CoGeNT events", Phys. Rev. D **83**, 083511, 083511 (2011), arXiv:1008.1784 [hep-ph].

[307]V. I. Osherov and A. I. Voronin, "Exact analytical solution of the quantum rosen-zener-demkov model", Phys. Rev. A **49**, 265–271 (1994).

[308]R. Morris and N. Weiner, "Low Energy INTEGRAL Positrons from eXciting Dark Matter", (2011), arXiv:1109.3747 [hep-ph].

[309]M. G. Walker, M. Mateo, E. W. Olszewski, J. Penarrubia, N. W. Evans, et al., "A Universal Mass Profile for Dwarf Spheroidal Galaxies", Astrophys.J. **704**, 1274–1287 (2009), arXiv:0906.0341 [astro-ph.CO].

[310]S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, "Constraints on the Self-Interaction Cross-Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657-56", Astrophys.J. **679**, 1173–1180 (2008), arXiv:0704.0261 [astro-ph].

[311]W. Dawson, "Cluster tests of dark matter self-interaction", 8th Harvard-Smithsonian Conference on Theoretical Astrophysics (2014).

[312]E. Dudas, L. Heurtier, and Y. Mambrini, "Generating X-ray lines from annihilating dark matter", Phys.Rev. **D90**, 035002 (2014), arXiv:1404.1927 [hep-ph].

[313]D. Malyshev, A. Neronov, and D. Eckert, "Constraints on 3.55 keV line emission from stacked observations of dwarf spheroidal galaxies", (2014), arXiv:1408.3531 [astro-ph.HE].

[314]M. E. Anderson, E. Churazov, and J. N. Bregman, "Non-Detection of X-Ray Emission From Sterile Neutrinos in Stacked Galaxy Spectra", (2014), arXiv:1408.4115 [astro-ph.HE].

[315]M. Vogelsberger and J. Zavala, "Direct detection of self-interacting dark matter", Mon. Not. R. Astron. Soc. **430**, 1722–1735 (2013), arXiv:1211.1377 [astro-ph.CO].

[316]M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist, "Properties of galaxies reproduced by a hydrodynamic simulation", Nature **509**, 177–182 (2014), arXiv:1405.1418 [astro-ph.CO].

[317]S. Vegetti and M. Vogelsberger, "On the density profile of dark matter substructure in gravitational lens galaxies", Mon. Not. R. Astron. Soc. **442**, 3598–3603 (2014), arXiv:1406.1170 [astro-ph.CO].

[318]G. A. Dooley, A. H. G. Peter, M. Vogelsberger, J. Zavala, and A. Frebel, "Enhanced tidal stripping of satellites in the galactic halo from dark matter self-interactions", Mon. Not. R. Astron. Soc. **461**, 710–727 (2016), arXiv:1603.08919 [astro-ph.GA].

[319]T. Brinckmann, J. Zavala, D. Rapetti, S. H. Hansen, and M. Vogelsberger, "The structure and assembly history of cluster-sized haloes in self-interacting dark matter", Mon. Not. R. Astron. Soc. **474**, 746–759 (2018), arXiv:1705.00623 [astro-ph.CO].

[320]O. Sameie, P. Creasey, H.-B. Yu, L. V. Sales, M. Vogelsberger, and J. Zavala, "The impact of baryonic discs on the shapes and profiles of self-interacting dark matter haloes", Mon. Not. R. Astron. Soc. **479**, 359–367 (2018), arXiv:1801.09682 [astro-ph.GA].

[321]L. Hernquist, "An Analytical Model for Spherical Galaxies and Bulges", Astrophys. J. **356**, 359 (1990).

[322]N. C. Amorisco, J. Zavala, and T. J. L. de Boer, "Dark Matter Cores in the Fornax and Sculptor Dwarf Galaxies: Joining Halo Assembly and Detailed Star Formation Histories", Astrophys. J. **782**, L39, L39 (2014), arXiv:1309.5958 [astro-ph.CO].

[323]A. J. Maxwell, J. Wadsley, and H. M. P. Couchman, "The Energetics of Cusp Destruction", Astrophys. J. **806**, 229, 229 (2015), arXiv:1505.00825 [astro-ph.GA].

[324]Planck Collaboration, "Planck 2013 results. XVI. Cosmological parameters", Astron. Astrophys. **571**, A16, A16 (2014), arXiv:1303.5076 [astro-ph.CO].

[325]D. N. Spergel, R. Flauger, and R. Hlozek, "Planck data reconsidered", Phys. Rev. D **91**, 023518, 023518 (2015), arXiv:1312.3313 [astro-ph.CO].

[326]V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann, "Populating a cluster of galaxies - I. Results at $z = 0$", Mon. Not. R. Astron. Soc. **328**, 726–750 (2001), arXiv:astro-ph/0012055 [astro-ph].

[327]E. Polisensky and M. Ricotti, "Constraints on the dark matter particle mass from the number of Milky Way satellites", Phys. Rev. D **83**, 043506, 043506 (2011), arXiv:1004.1459 [astro-ph.CO].

[328]J. Koda and P. R. Shapiro, "Gravothermal collapse of isolated self-interacting dark matter haloes: N-body simulation versus the fluid model", Mon. Not. R. Astron. Soc. **415**, 1125–1137 (2011), arXiv:1101.3097 [astro-ph.CO].

[329]O. D. Elbert, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, A. S. Graus, and M. Rocha, "A Testable Conspiracy: Simulating Baryonic Effects on Self-interacting Dark Matter Halos", Astrophys. J. **853**, 109, 109 (2018), arXiv:1609.08626 [astro-ph.GA].

[330]M. Vogelsberger, S. Genel, D. Sijacki, P. Torrey, V. Springel, and L. Hernquist, "A model for cosmological simulations of galaxy formation physics", Mon. Not. R. Astron. Soc. **436**, 3031–3067 (2013), arXiv:1305.2913 [astro-ph.CO].

[331]M. Vogelsberger, J. Zavala, C. Simpson, and A. Jenkins, "Dwarf galaxies in CDM and SIDM with baryons: observational probes of the nature of dark matter", Mon. Not. R. Astron. Soc. **444**, 3684–3698 (2014), arXiv:1405.5216 [astro-ph.CO].

[332]M. Vogelsberger, A. Helmi, V. Springel, S. D. M. White, J. Wang, C. S. Frenk, A. Jenkins, A. Ludlow, and J. F. Navarro, "Phase-space structure in the local dark matter distribution and its signature in direct detection experiments", Mon. Not. R. Astron. Soc. **395**, 797–811 (2009), arXiv:0812.0362 [astro-ph].

[333]Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. Wacker, "Model for Thermal Relic Dark Matter of Strongly Interacting Massive Particles", Phys. Rev. Lett. **115**, 021301 (2015), arXiv:1411.3727 [hep-ph].

[334]J. Wess and B. Zumino, "Consequences of Anomalous Ward Identities", Phys. Lett. **B37**, 95 (1971).

[335]E. Witten, "Current Algebra, Baryons, and Quark Confinement", Nucl. Phys. **B223**, 433–444 (1983).

[336]E. Witten, "Global Aspects of Current Algebra", Nucl. Phys. **B223**, 422–432 (1983).

[337]H. M. Lee and M.-S. Seo, "Models for Simp Dark Matter and Dark Photon", AIP Conf. Proc. **1743**, 060003 (2016), arXiv:1510.05116 [hep-ph].

[338]N. Bernal and X. Chu, "Z_2 Simp Dark Matter", JCAP **1601**, 006 (2016), arXiv:1510.08527 [hep-ph].

[339]Y. Hochberg, E. Kuflik, and H. Murayama, "SIMP Spectroscopy", JHEP **05**, 090 (2016), arXiv:-1512.07917 [hep-ph].

[340]H. M. Lee and M.-S. Seo, "Communication with Simp Dark Mesons via Z^2 -Portal", Phys. Lett. **B748**, 316–322 (2015), arXiv:1504.00745 [hep-ph].

[341]S.-M. Choi and H. M. Lee, "Resonant Simp Dark Matter", Phys. Lett. **B758**, 47–53 (2016), arXiv:1601.03566 [hep-ph].

[342]N. Bernal, X. Chu, and J. Pradler, "Simply Split Strongly Interacting Massive Particles", Phys. Rev. **D95**, 115023 (2017), arXiv:1702.04906 [hep-ph].

[343]S.-M. Choi, H. M. Lee, and M.-S. Seo, "Cosmic Abundances of Simp Dark Matter", JHEP **04**, 154 (2017), arXiv:1702.07860 [hep-ph].

[344]Y. Hochberg, E. Kuflik, and H. Murayama, "Dark Spectroscopy at Lepton Colliders", Phys. Rev. **D97**, 055030 (2018), arXiv:1706.05008 [hep-ph].

[345]S.-M. Choi, Y. Hochberg, E. Kuflik, H. M. Lee, Y. Mambrini, H. Murayama, and M. Pierre, "Vector SIMP dark matter", JHEP **10**, 162 (2017), arXiv:1707.01434 [hep-ph].

[346]A. Berlin, N. Blinov, S. Gori, P. Schuster, and N. Toro, "Cosmology and Accelerator Tests of Strongly Interacting Dark Matter", (2018), arXiv:1801.05805 [hep-ph].

[347]S.-M. Choi, H. M. Lee, P. Ko, and A. Natale, "Unitarizing Simp Scenario with Dark Vector Resonances", (2018), arXiv:1801.07726 [hep-ph].

[348]J. S. Bullock and M. Boylan-Kolchin, "Small-Scale Challenges to the ΛCDM Paradigm", Ann. Rev. Astron. Astrophys. **55**, 343–387 (2017), arXiv:1707.04256 [astro-ph.CO].

[349]A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguere, D. Keres, and E. Quataert, "Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites Around a Milky Way-Mass Galaxy", Astrophys. J. **827**, L23 (2016), arXiv:1602.05957 [astro-ph.GA].

[350]E. D. Carlson, M. E. Machacek, and L. J. Hall, "Self-Interacting Dark Matter", Astrophys. J. **398**, 43–52 (1992).

[351]A. A. de Laix, R. J. Scherrer, and R. K. Schaefer, "Constraints of Selfinteracting Dark Matter", Astrophys. J. **452**, 495 (1995), arXiv:astro-ph/9502087 [astro-ph].

[352]A. Kamada, H. Kim, and T. Sekiguchi, "Axionlike Particle Assisted Strongly Interacting Massive Particle", Phys. Rev. **D96**, 016007 (2017), arXiv:1704.04505 [hep-ph].

[353]M. Hansen, K. Langaeble, and F. Sannino, "Simp Model at NNLO in Chiral Perturbation Theory", Phys. Rev. **D92**, 075036 (2015), arXiv:1507.01590 [hep-ph].

[354]P. Gondolo and G. Gelmini, "Cosmic abundances of stable particles: Improved analysis", Nucl. Phys. **B360**, 145–179 (1991).

[355]T. Bringmann and S. Hofmann, "Thermal Decoupling of Wimps from First Principles", JCAP **0704**, [Erratum: JCAP1603,no.03,E02(2016)], 016 (2007), arXiv:hep-ph/0612238.

[356]E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, "Phenomenology of Elder Dark Matter", JHEP **08**, 078 (2017), arXiv:1706.05381 [hep-ph].

[357]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].

[358]E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, "Elastically Decoupling Dark Matter", Phys. Rev. Lett. **116**, 221302 (2016), arXiv:1512.04545 [hep-ph].

[359]P. A. R. Ade et al., "Planck 2015 results. XIII. Cosmological parameters", Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].

[360]K. Enqvist, K. Kainulainen, and V. Semikoz, "Neutrino annihilation in hot plasma", Nucl. Phys. **B374**, 392–404 (1992).

[361]C. Boehm, M. J. Dolan, and C. McCabe, "A Lower Bound on the Mass of Cold Thermal Dark Matter from Planck", JCAP **1308**, 041 (2013), arXiv:1303.6270 [hep-ph].

[362]E. Masso and R. Toldra, "On a Light Spinless Particle Coupled to Photons", Phys. Rev. **D52**, 1755–1763 (1995), arXiv:hep-ph/9503293 [hep-ph].

[363]J. Jaeckel, P. C. Malta, and J. Redondo, "Decay Photons from the Alp Burst of Type-Ii Supernovae", (2017), arXiv:1702.02964 [hep-ph].

[364]M. Acciarri et al., "Search for Anomalous Z –≫ Gamma Gamma Gamma Events at Lep", Phys. Lett. **B345**, 609–616 (1995).

[365]J. Jaeckel and M. Spannowsky, "Probing MeV to 90 GeV Axion-Like Particles with Lep and Lhc", Phys. Lett. **B753**, 482–487 (2016), arXiv:1509.00476 [hep-ph].

[366]M. J. Dolan, T. Ferber, C. Hearty, F. Kahlhoefer, and K. Schmidt-Hoberg, "Revised Constraints and Belle II Sensitivity for Visible and Invisible Axion-Like Particles", JHEP **12**, 094 (2017), arXiv:1709.00009 [hep-ph].

[367]J. D. Bjorken, S. Ecklund, W. R. Nelson, A. Abashian, C. Church, B. Lu, L. W. Mo, T. A. Nunamaker, and P. Rassmann, "Search for Neutral Metastable Penetrating Particles Produced in the Slac Beam Dump", Phys. Rev. **D38**, 3375 (1988).

[368]E. M. Riordan et al., "A Search for Short Lived Axions in an Electron Beam Dump Experiment", Phys. Rev. Lett. **59**, 755 (1987).

[369]B. Dobrich, "Axion-like Particles from Primakov production in beam-dumps", (2017), arXiv:-1708.05776 [hep-ph].

[370]J. L. Hewett et al., "Fundamental Physics at the Intensity Frontier", (2012), arXiv:1205.2671 [hep-ex].

[371]B. Döbrich, J. Jaeckel, F. Kahlhoefer, A. Ringwald, and K. Schmidt-Hoberg, "Alptraum: Alp Production in Proton Beam Dump Experiments", JHEP **02**, [JHEP02,018(2016)], 018 (2016), arXiv:1512.03069 [hep-ph].

[372]E. Izaguirre, T. Lin, and B. Shuve, "Searching for Axionlike Particles in Flavor-Changing Neutral Current Processes", Phys. Rev. Lett. **118**, 111802 (2017), arXiv:1611.09355 [hep-ph].

[373]A. V. Artamonov et al., "Search for the Decay K+ to Pi+ Gamma Gamma in the Pi+ Momentum Region P ≫ 213 Mev/C", Phys. Lett. **B623**, 192–199 (2005), arXiv:hep-ex/0505069 [hep-ex].

[374]C. Lazzeroni et al., "Study of the $K^{\pm} \to \pi^{\pm}\gamma\gamma$ Decay by the Na62 Experiment", Phys. Lett. **B732**, 65–74 (2014), arXiv:1402.4334 [hep-ex].

[375]E. Abouzaid et al., "Final Results from the Ktev Experiment on the Decay $K_L \to \pi^0\gamma\gamma$", Phys. Rev. **D77**, 112004 (2008), arXiv:0805.0031 [hep-ex].

[376]S. Alekhin et al., "A Facility to Search for Hidden Particles at the Cern Sps: the Ship Physics Case", Rept. Prog. Phys. **79**, 124201 (2016), arXiv:1504.04855 [hep-ph].

[377]A. Berlin, S. Gori, P. Schuster, and N. Toro, "Dark Sectors at the Fermilab Seaquest Experiment", (2018), arXiv:1804.00661 [hep-ph].

[378]J. L. Feng, I. Galon, F. Kling, and S. Trojanowski, "ALPs at FASER: The LHC as a Photon Beam Dump", (2018), arXiv:1806.02348 [hep-ph].

[379]R. Essig, R. Harnik, J. Kaplan, and N. Toro, "Discovering New Light States at Neutrino Experiments", Phys. Rev. **D82**, 113008 (2010), arXiv:1008.0636 [hep-ph].

[380]M. J. Dolan, F. Kahlhoefer, C. McCabe, and K. Schmidt-Hoberg, "A Taste of Dark Matter: Flavour Constraints on Pseudoscalar Mediators", JHEP **03**, [Erratum: JHEP07,103(2015)], 171 (2015), arXiv:1412.5174 [hep-ph].

[381]M. Bauer, M. Neubert, and A. Thamm, "Collider Probes of Axion-Like Particles", JHEP **12**, 044 (2017), arXiv:1708.00443 [hep-ph].

[382]K. M. Nollett and G. Steigman, "BBN And The CMB Constrain Light, Electromagnetically Coupled WIMPs", Phys. Rev. **D89**, 083508 (2014), arXiv:1312.5725 [astro-ph.CO].

[383]K. Nollett and G. Steigman, "BBN And The CMB Constrain Neutrino Coupled Light WIMPs", Phys. Rev. **D91**, 083505 (2015), arXiv:1411.6005 [astro-ph.CO].

[384]C. Boehm and P. Fayet, "Scalar dark matter candidates", Nucl. Phys. **B683**, 219–263 (2004), arXiv:hep-ph/0305261 [hep-ph].

[385]M. Pospelov, A. Ritz, and M. B. Voloshin, "Secluded WIMP Dark Matter", Phys. Lett. **B662**, 53–61 (2008), arXiv:0711.4866 [hep-ph].

[386]J. L. Feng and J. Kumar, "The WIMPless Miracle: Dark-Matter Particles without Weak- Scale Masses or Weak Interactions", Phys. Rev. Lett. **101**, 231301 (2008), arXiv:0803.4196.

[387]R. T. D'Agnolo, C. Mondino, J. T. Ruderman, and P.-J. Wang, "Exponentially Light Dark Matter from Coannihilation", JHEP **08**, 079 (2018), arXiv:1803.02901 [hep-ph].

[388]R. T. D'Agnolo, D. Pappadopulo, and J. T. Ruderman, "Fourth Exception in the Calculation of Relic Abundances", Phys. Rev. Lett. **119**, 061102 (2017), arXiv:1705.08450 [hep-ph].

[389]R. T. D'Agnolo and J. T. Ruderman, "Light Dark Matter from Forbidden Channels", Phys. Rev. Lett. **115**, 061301 (2015), arXiv:1505.07107 [hep-ph].

[390]D. Pappadopulo, J. T. Ruderman, and G. Trevisan, "Dark matter freeze-out in a nonrelativistic sector", Phys. Rev. **D94**, 035005 (2016), arXiv:1602.04219 [hep-ph].

[391]J. M. Cline, H. Liu, T. Slatyer, and W. Xue, "Enabling Forbidden Dark Matter", Phys. Rev. **D96**, 083521 (2017), arXiv:1702.07716 [hep-ph].

[392]J. Kopp, J. Liu, T. R. Slatyer, X.-P. Wang, and W. Xue, "Impeded Dark Matter", JHEP **12**, 033 (2016), arXiv:1609.02147 [hep-ph].

[393]A. Berlin and N. Blinov, "Thermal Dark Matter Below an MeV", Phys. Rev. Lett. **120**, 021801 (2018), arXiv:1706.07046 [hep-ph].

[394]T. Asaka, K. Ishiwata, and T. Moroi, "Right-handed sneutrino as cold dark matter", Phys. Rev. **D73**, 051301 (2006), arXiv:hep-ph/0512118 [hep-ph].

[395]T. Asaka, K. Ishiwata, and T. Moroi, "Right-handed sneutrino as cold dark matter of the universe", Phys. Rev. **D75**, 065001 (2007), arXiv:hep-ph/0612211 [hep-ph].

[396]S. Gopalakrishna, A. de Gouvea, and W. Porod, "Right-handed sneutrinos as nonthermal dark matter", JCAP **0605**, 005 (2006), arXiv:hep-ph/0602027 [hep-ph].

[397]V. Page, "Non-thermal right-handed sneutrino dark matter and the Omega(DM)/Omega(b) problem", JHEP **04**, 021 (2007), arXiv:hep-ph/0701266 [hep-ph].

[398]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].

[399]N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen, and V. Vaskonen, "The Dawn of FIMP Dark Matter: A Review of Models and Constraints", Int. J. Mod. Phys. **A32**, 1730023 (2017), arXiv:1706.07442 [hep-ph].

[400]R. Essig, J. Mardon, and T. Volansky, "Direct Detection of Sub-GeV Dark Matter", Phys. Rev. **D85**, 076007 (2012), arXiv:1108.5383 [hep-ph].

[401]R. Essig, A. Manalaysay, J. Mardon, P. Sorensen, and T. Volansky, "First Direct Detection Limits on sub-GeV Dark Matter from XENON10", Phys. Rev. Lett. **109**, 021301 (2012), arXiv:1206.2644 [astro-ph.CO].

[402]R. Essig, M. Fernandez-Serra, J. Mardon, A. Soto, T. Volansky, and T.-T. Yu, "Direct Detection of sub-GeV Dark Matter with Semiconductor Targets", JHEP **05**, 046 (2016), arXiv:1509.01598 [hep-ph].

[403]Y. Hochberg, Y. Kahn, M. Lisanti, C. G. Tully, and K. M. Zurek, "Directional detection of dark matter with two-dimensional targets", Phys. Lett. **B772**, 239–246 (2017), arXiv:1606.08849 [hep-ph].

[404]S. Derenzo, R. Essig, A. Massari, and T.-T. Soto Aand Yu, "Direct Detection of sub-GeV Dark Matter with Scintillating Targets", Phys. Rev. **D96**, 016026 (2017), arXiv:1607.01009.

[405]Y. Hochberg, Y. Kahn, M. Lisanti, K. M. Zurek, A. G. Grushin, R. Ilan, S. M. Griffin, Z.-F. Liu, S. F. Weber, and J. B. Neaton, "Detection of sub-MeV Dark Matter with Three-Dimensional Dirac Materials", Phys. Rev. **D97**, 015004 (2018), arXiv:1708.08929 [hep-ph].

[406]S. Knapen, T. Lin, M. Pyle, and K. M. Zurek, "Detection of Light Dark Matter With Optical Phonons in Polar Materials", (2017), arXiv:1712.06598 [hep-ph].

[407]S. Griffin, S. Knapen, T. Lin, and K. M. Zurek, "Directional Detection of Light Dark Matter with Polar Materials", (2018), arXiv:1807.10291 [hep-ph].

[408]K. Schutz and K. M. Zurek, "Detectability of Light Dark Matter with Superfluid Helium", Phys. Rev. Lett. **117**, 121302 (2016), arXiv:1604.08206 [hep-ph].

[409]S. Knapen, T. Lin, and K. M. Zurek, "Light Dark Matter in Superfluid Helium: Detection with Multi-excitation Production", Phys. Rev. **D95**, 056019 (2017), arXiv:1611.06228.

[410]Y. Hochberg, Y. Zhao, and K. M. Zurek, "Superconducting Detectors for Superlight Dark Matter", Phys. Rev. Lett. **116**, 011301 (2016), arXiv:1504.07237 [hep-ph].

[411]Y. Hochberg, M. Pyle, Y. Zhao, and K. M. Zurek, "Detecting Superlight Dark Matter with Fermi-Degenerate Materials", JHEP **08**, 057 (2016), arXiv:1512.04533 [hep-ph].

[412]M. Crisler, R. Essig, J. Estrada, G. Fernandez, J. Tiffenberg, M. Sofo haro, T. Volansky, and T.-T. Yu, "SENSEI: First Direct-Detection Constraints on sub-GeV Dark Matter from a Surface Run", Phys. Rev. Lett. **121**, 061803 (2018), arXiv:1804.00088 [hep-ex].

[413]O. Abramoff et al., "SENSEI: Direct-Detection Constraints on Sub-GeV Dark Matter from a Shallow Underground Run Using a Prototype Skipper-CCD", (2019), arXiv:1901.10478 [hep-ex].

[414]R. Agnese et al., "First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector", Phys. Rev. Lett. **121**, 051301 (2018), arXiv:1804.10697 [hep-ex].

[415]P. Agnes et al., "Constraints on Sub-GeV Dark-Matter-Electron Scattering from the DarkSide-50 Experiment", Phys. Rev. Lett. **121**, 111303 (2018), arXiv:1802.06998 [astro-ph.CO].

[416]T. Hambye, M. H. G. Tytgat, J. Vandecasteele, and L. Vanderheyden, "Direct Detection is testing Freeze-in", (2018), arXiv:1807.05022 [hep-ph].

[417]S. Knapen, T. Lin, and K. M. Zurek, "Light Dark Matter: Models and Constraints", (2017), arXiv:1709.07882 [hep-ph].

[418]D. Green and S. Rajendran, "The Cosmology of Sub-MeV Dark Matter", JHEP **10**, 013 (2017), arXiv:1701.08750 [hep-ph].

[419]X. Chu, T. Hambye, and M. H. G. Tytgat, "The Four Basic Ways of Creating Dark Matter Through a Portal", JCAP **1205**, 034 (2012), arXiv:1112.0493 [hep-ph].

[420]S. Davidson, S. Hannestad, and G. Raffelt, "Updated bounds on millicharged particles", JHEP **05**, 003 (2000), arXiv:hep-ph/0001179 [hep-ph].

[421]H. Vogel and J. Redondo, "Dark Radiation constraints on minicharged particles in models with a hidden photon", JCAP **1402**, 029 (2014), arXiv:1311.2600 [hep-ph].

[422]H. An, R. Huo, and W. Liu, "KeV Scale Frozen-in Self-Interacting Fermionic Dark Matter", (2018), arXiv:1812.05699 [hep-ph].

[423]R. Barkana, "Possible interaction between baryons and dark-matter particles revealed by the first stars", Nature **555**, 71 (2018).

[424]A. Berlin, D. Hooper, G. Krnjaic, and S. D. McDermott, "Severely Constraining Dark Matter Interpretations of the 21-cm Anomaly", Phys. Rev. Lett. **121**, 011102 (2018), arXiv:1803.02804 [hep-ph].

[425]J. B. Muñoz and A. Loeb, "A small amount of mini-charged dark matter could cool the baryons in the early Universe", Nature **557**, 684 (2018), arXiv:1802.10094 [astro-ph.CO].

[426]C. Dvorkin, T. Lin, and K. Schutz, "Cosmological Constraints on sub-MeV Dark Matter Freeze-in", in prep.

[427]S. L. Dubovsky, D. S. Gorbunov, and G. I. Rubtsov, "Narrowing the window for millicharged particles by CMB anisotropy", JETP Lett. **79**, [Pisma Zh. Eksp. Teor. Fiz.79,3(2004)], 1–5 (2004), arXiv:hep-ph/0311189 [hep-ph].

[428]R. Foot, H. Lew, and R. R. Volkas, "Electric charge quantization", J. Phys. **G19**, [Erratum: J. Phys.G19,1067(1993)], 361–372 (1993), arXiv:hep-ph/9209259 [hep-ph].

[429]J.-W. Chen, H.-C. Chi, H.-B. Li, C. .-.-P. Liu, L. Singh, H. T. Wong, C.-L. Wu, and C.-P. Wu, "Constraints on millicharged neutrinos via analysis of data from atomic ionizations with germanium detectors at sub-keV sensitivities", Phys. Rev. **D90**, 011301 (2014), arXiv:1405.7168 [hep-ph].

[430]E. Hardy and R. Lasenby, "Stellar cooling bounds on new light particles: plasma mixing effects", JHEP **02**, 033 (2017), arXiv:1611.05852 [hep-ph].

[431]N. Viaux, M. Catelan, P. B. Stetson, G. Raffelt, J. Redondo, A. A. R. Valcarce, and A. Weiss, "Neutrino and axion bounds from the globular cluster M5 (NGC 5904)", Phys. Rev. Lett. **111**, 231301 (2013), arXiv:1311.1669 [astro-ph.SR].

[432]M. M. Miller Bertolami, B. E. Melendez, L. G. Althaus, and J. Isern, "Revisiting the axion bounds from the Galactic white dwarf luminosity function", JCAP **1410**, 069 (2014), arXiv:-1406.7712 [hep-ph].

[433]E. D. Carlson and P. Salati, "Stellar Limits on keV Pseudoscalars", Phys. Lett. **B218**, 79–83 (1989).

[434]G. G. Raffelt and G. D. Starkman, "Stellar Energy Transfer by keV Mass Scalars", Phys. Rev. **D40**, 942 (1989).

[435]M. Giannotti, I. G. Irastorza, J. Redondo, A. Ringwald, and K. Saikawa, "Stellar Recipes for Axion Hunters", JCAP **1710**, 010 (2017), arXiv:1708.02111 [hep-ph].

[436]G. Krnjaic, "Freezing In, Heating Up, and Freezing Out: Predictive Nonthermal Dark Matter and Low-Mass Direct Detection", (2017), arXiv:1711.11038 [hep-ph].

[437]H. An, M. Pospelov, and J. Pradler, "New stellar constraints on dark photons", Phys. Lett. **B725**, 190–195 (2013), arXiv:1302.3884 [hep-ph].

[438]H. An, M. Pospelov, and J. Pradler, "Dark Matter Detectors as Dark Photon Helioscopes", Phys. Rev. Lett. **111**, 041302 (2013), arXiv:1304.3461 [hep-ph].

[439]J. Jaeckel, "A force beyond the Standard Model - Status of the quest for hidden photons", Frascati Phys. Ser. **56**, 172–192 (2012), arXiv:1303.1821 [hep-ph].

[440]J. Murata and S. Tanaka, "A review of short-range gravity experiments in the LHC era", Class. Quant. Grav. **32**, 033001 (2015), arXiv:1408.3588 [hep-ex].

[441]E. G. Adelberger, B. R. Heckel, and A. E. Nelson, "Tests of the gravitational inverse square law", Ann. Rev. Nucl. Part. Sci. **53**, 77–121 (2003), arXiv:hep-ph/0307284 [hep-ph].

[442]C. Dvorkin, K. Blum, and M. Kamionkowski, "Constraining Dark Matter-Baryon Scattering with Linear Cosmology", Phys. Rev. **D89**, 023519 (2014), arXiv:1311.2937.

[443]W. L. Xu, C. Dvorkin, and A. Chael, "Probing sub-GeV Dark Matter-Baryon Scattering with Cosmological Observables", Phys. Rev. **D97**, 103530 (2018), arXiv:1802.06788.

[444]T. R. Slatyer and C.-L. Wu, "Early-Universe constraints on dark matter-baryon scattering and their implications for a global 21 cm signal", Phys. Rev. **D98**, 023013 (2018), arXiv:1803.09734.

[445]E. D. Kovetz, V. Poulin, V. Gluscevic, K. K. Boddy, R. Barkana, and M. Kamionkowski, "Tighter Limits on Dark Matter Explanations of the Anomalous EDGES 21cm Signal", (2018), arXiv:1807.11482 [astro-ph.CO].

[446]K. K. Boddy, V. Gluscevic, V. Poulin, E. D. Kovetz, M. Kamionkowski, and R. Barkana, "A Critical Assessment of CMB Limits on Dark Matter-Baryon Scattering: New Treatment of the Relative Bulk Velocity", (2018), arXiv:1808.00001 [astro-ph.CO].

[447]S. L. Dubovsky and D. S. Gorbunov, "Small second acoustic peak from interacting cold dark matter?", Phys. Rev. **D64**, 123503 (2001), arXiv:astro-ph/0103122 [astro-ph].

[448]K. K. Boddy and V. Gluscevic, "First Cosmological Constraint on the Effective Theory of Dark Matter-Proton Interactions", Phys. Rev. **D98**, 083510 (2018), arXiv:1801.08609.

[449]L. B. Okun, M. B. Voloshin, and V. I. Zakharov, "Electrical Neutrality of Atoms and Grand Unification Models", Phys. Lett. **138B**, 115–120 (1984).

[450]S. D. McDermott, H.-B. Yu, and K. M. Zurek, "Turning off the Lights: How Dark is Dark Matter?", Phys. Rev. **D83**, 063509 (2011), arXiv:1011.2907 [hep-ph].

[451]J. H. Chang, R. Essig, and S. D. McDermott, "Supernova 1987A Constraints on Sub-GeV Dark Sectors, Millicharged Particles, the QCD Axion, and an Axion-like Particle", (2018), arXiv:1803.00993 [hep-ph].

[452]G. Barello, S. Chang, and C. A. Newby, "Correlated signals at the energy and intensity frontiers from non-Abelian kinetic mixing", Phys. Rev. **D94**, 055018 (2016), arXiv:1511.02865 [hep-ph].

[453]C. A. Argüelles, X.-G. He, G. Ovanesyan, T. Peng, and M. J. Ramsey-Musolf, "Dark Gauge Bosons: LHC Signatures of Non-Abelian Kinetic Mixing", Phys. Lett. **B770**, 101–107 (2017), arXiv:1604.00044 [hep-ph].

[454]K. R. Dienes, C. F. Kolda, and J. March-Russell, "Kinetic mixing and the supersymmetric gauge hierarchy", Nucl. Phys. **B492**, 104–118 (1997), arXiv:hep-ph/9610479 [hep-ph].

[455]S. A. Abel, M. D. Goodsell, J. Jaeckel, V. V. Khoze, and A. Ringwald, "Kinetic Mixing of the Photon with Hidden U(1)s in String Phenomenology", JHEP **07**, 124 (2008), arXiv:0803.1449 [hep-ph].

[456]M. Goodsell, J. Jaeckel, J. Redondo, and A. Ringwald, "Naturally Light Hidden Photons in LARGE Volume String Compactifications", JHEP **11**, 027 (2009), arXiv:0909.0515.

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

[458]J. Kummer, M. Brüggen, K. Dolag, F. Kahlhoefer, and K. Schmidt-Hoberg, "Simulations of core formation for frequent dark matter self-interactions", (2019), arXiv:1902.02330.

[459]M. Baryakhtar, R. Lasenby, and M. Teo, "Black Hole Superradiance Signatures of Ultralight Vectors", Phys. Rev. **D96**, 035019 (2017), arXiv:1704.05081 [hep-ph].

[460]J.-P. Blaizot, E. Iancu, and R. R. Parwani, "On the screening of static electromagnetic fields in hot QED plasmas", Phys. Rev. **D52**, 2543–2562 (1995), arXiv:hep-ph/9504408 [hep-ph].

[461]M. Srednicki, R. Watkins, and K. A. Olive, "Calculations of Relic Densities in the Early Universe", Nucl. Phys. **B310**, [,247(1988)], 693 (1988).

[462]E. Braaten and D. Segel, "Neutrino energy loss from the plasma process at all temperatures and densities", Phys. Rev. **D48**, 1478–1491 (1993), arXiv:hep-ph/9302213 [hep-ph].

[463]L. Chuzhoy and E. W. Kolb, "Reopening the window on charged dark matter", JCAP **0907**, 014 (2009), arXiv:0809.0436 [astro-ph].

[464]P.-K. Hu, A. Kusenko, and V. Takhistov, "Dark Cosmic Rays", Phys. Lett. **B768**, 18–22 (2017), arXiv:1611.04599 [hep-ph].

[465]D. Dunsky, L. J. Hall, and K. Harigaya, "CHAMP Cosmic Rays", (2018), arXiv:1812.11116 [astro-ph.HE].

[466]K. J. Bae, A. Kamada, S. P. Liew, and K. Yanagi, "Light axinos from freeze-in: production processes, phase space distributions, and Ly-α forest constraints", JCAP **1801**, 054 (2018), arXiv:1707.06418 [hep-ph].

[467]E. Braaten and T. C. Yuan, "Calculation of screening in a hot plasma", Physical review letters **66**, 2183 (1991).

[468]G. G. Raffelt, "Astrophysical axion bounds diminished by screening effects", Physical Review D **33**, 897 (1986).

[469]G. Baym, H. Monien, C. Pethick, and D. Ravenhall, "Transverse interactions and transport in relativistic quark-gluon and electromagnetic plasmas", Physical review letters **64**, 1867 (1990).

[470]S. Davidson, S. Hannestad, and G. Raffelt, "Updated bounds on milli-charged particles", Journal of High Energy Physics **2000**, 003 (2000).