

Self-Interacting Dark Matter

Jeremie Choquette

Doctor of Philosophy

Physics

McGill University

Montreal, Quebec

April 2, 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Theoretical Physics

©Jeremie Choquette 2019

ACKNOWLEDGEMENTS

I thank my supervisor, James Cline, for his support and patience. He, along with my other co-author Jonathan Cornell, were always there to discuss my research when I was stuck in a rut, and occasionally to give a much needed shove. Finally I'd like to thank all the friends and family who had to put up with me talking (and in some cases complaining) about my thesis. I assure you it was worth it.

ABSTRACT

Self-interacting dark matter models have grown in influence over the last decade as an alternative to cold dark matter, resolving key problems in the distribution of dark matter structure on sub-galactic scales. In this manuscript-based thesis, we present four papers related to dark matter astrophysics, with an emphasis on self-interacting dark matter. In the first, we consider a dark sector in which a hidden $SU(2)$ gauge symmetry, which breaks to $U(1)$ via a Higgs-like doublet, results in a nonabelian model of atomic dark matter. The model has an interesting phenomenology and can result in the correct relic density. In the second and third papers, we explore a potential signal from annihilating dark matter in the Galactic Center. It has been claimed that the Galactic Center excess is at odds with observations of dwarf spheroidal galaxies that show no signal, however we present two scenarios in which the signal from dwarf galaxies would be diminished by a sufficient margin to alleviate the tension. In the first, the dark matter annihilates with a velocity dependent cross section, decreasing the annihilation rate in dwarf galaxies. In the second, we consider the fact that dwarf spheroidal galaxies likely have density profiles which have a central core of near-constant density (rather than a centrally peaked or ‘cuspy’ profile), and show that density profiles consistent with self-interacting dark matter as well as observed density profiles would weaken the signal to undetectable levels. In the fourth paper, we explore the formation of high-redshift supermassive black holes from two-component dark matter models, determining the self-interaction cross section and self-interacting dark matter fraction required to provide an alternative explanation to the observed early formation of supermassive black holes.

ABRÉGÉ

Les modèles de matière noire auto-interagissante ont pris une importance croissante au cours de la dernière décennie en tant qu'alternative à la matière noire froide, résolvant ainsi les principaux problèmes de répartition de la structure de la matière noire à des échelles sous-galactiques. Dans cette thèse par articles, nous présentons quatre articles sur l'astrophysique de la matière noire, en mettant l'accent sur la matière noire auto-interagissante. Dans la première, nous considérons un secteur noir dans lequel une symétrie de jauge $SU(2)$ cachée, qui se casse à $U(1)$ par un doublet de type Higgs, donne un modèle non abélien de matière noire atomique. Le modèle a une phénoménologie intéressante et peut aboutir à une densité relique correcte. Dans les deuxième et troisième articles, nous explorons un signal potentiel provenant de la destruction de la matière noire au centre galactique. On a prétendu que l'excès au centre galactique était en contradiction avec l'observation de galaxies naines sphéroïdales qui ne montrent aucun signal. Cependant, nous présentons deux scénarios dans lesquels le signal des galaxies naines serait diminué d'une marge suffisante pour atténuer la tension. Dans le premier cas, la matière noire s'anéantit avec une section efficace dépendante de la vitesse, ce qui diminue le taux d'annihilation dans les galaxies naines. Dans le deuxième, nous considérons le fait que les galaxies naines sphéroïdales ont probablement des profils de densité ayant une région centrale de densité presque constante (plutôt qu'un profil froncé), et nous montrons que ces profils de densité, qui sont cohérents avec la matière noire auto-interagissante et les observations, affaiblissent le signal à des niveaux indétectables. Dans le quatrième article, nous explorons la formation de trous noirs supermassifs fortement décalés vers le rouge à partir de modèles de matière noire à deux composantes, en

déterminant la section efficace auto-interagissante et la fraction de matière noire auto-interagissante requises pour fournir une explication alternative à la formation précoce observée de trous noirs supermassifs.

TABLE OF CONTENTS

	ACKNOWLEDGEMENTS	ii
	ABSTRACT	iii
	ABRÉGÉ	iv
1	Introduction	2
	1.1 Evidence for Dark Matter	3
	1.2 Cold Dark Matter Theories	5
	1.2.1 WIMP Dark Matter	6
	1.2.2 Axions	7
	1.2.3 Primordial Black Holes	8
	1.3 Dark Matter Detection	9
	1.3.1 Direct Detection of Dark Matter	9
	1.3.2 Indirect Detection of Dark Matter	11
	1.3.3 Collider Production of Dark Matter	13
	1.4 Problems for the CDM Paradigm	15
	1.4.1 The Core/Cusp Problem	17
	1.4.2 The Missing Satellites Problem	18
	1.4.3 The Too-Big-To-Fail Problem	18
	1.4.4 Warm Dark Matter	20
	1.5 Self-Interacting Dark Matter	20
	1.5.1 Constraints on Dark Matter Self-Interactions	23
2	A Model of Atomic Dark Matter	25
	2.1 Hidden Sector Dark Matter	25
	2.2 Dark Atoms	27
	2.3 Leptogenesis	29
	2.4 Our Work	29
	2.4.1 Addendum:	71
3	The Galactic Center Excess	72
	3.1 The Galactic Center Excess: A Potential Dark Matter Signal?	72
	3.2 p -Wave Annihilating Dark Matter	74
	3.2.1 Addendum:	114
	3.3 The Effect of Dark Matter Density Profiles on The GCE	114

4	Supermassive Black Holes from Self-Interacting Dark Matter . . .	144
4.1	High-Redshift SMBHs	144
4.2	Baryonic Explanations for High-Redshift SMBHs	145
4.2.1	Population III Stars	145
4.2.2	Direct Collapse Black Holes	146
4.2.3	Mergers	147
4.3	High-Reshift SMBHs From SIDM	148
4.4	Our Work	150
5	Discussion	187
6	Conclusion	196

Contribution To Original Knowledge

The four manuscripts presented in this thesis constitute original scholarship and are distinct contributions to knowledge.

CHAPTER 1

Introduction

The nature of dark matter has been one of the defining unanswered questions in the fields of astrophysics and particle physics in the past half century. Despite evidence showing that it comprises approximately 23% [1] of the energy content of the universe, very little is known about the nature of dark matter. There is no consensus on its properties or its relationship to the Standard Model of particle physics. Decades of searches have not led to any confirmed direct or indirect detection of dark matter, nor have collider experiments been able to produce it. Apart from its gravitational effects on baryonic matter and contribution to the overall matter content of the universe, very few details are known about dark matter.

In this manuscript-based thesis, we present four manuscripts related to dark matter, with a special focus on models of self-interacting dark matter. We begin in section 1.1 with an introduction to the evidence for dark matter, placing current dark matter research in its historical context. We then briefly review the most popular category of dark matter models, cold dark matter, in section 1.2. In section 1.3 we review the three main categories of dark matter detection methods: direct detection, indirect detection, and collider production. In section 1.4 we list some of the problems faced by cold dark matter relating to its structure on small scales. These motivate models of self-interacting dark matter, as discussed in section 1.5.

We then present our four manuscripts. In chapter 2 we introduce hidden sector dark matter models, models in which additional gauge symmetries form a ‘dark sector’ which interacts with the Standard model in either a very limited

fashion or not at all, as well as dark atom models, in which bound states form between dark matter particles oppositely charged under a dark $U(1)$ gauge symmetry. We present our own work, a model of nonabelian atomic dark matter.

In chapter 3 we introduce and discuss the Galactic Center excess, a γ -ray signal observed in the center of the Milky Way which is consistent with annihilating dark matter. We present two potential resolutions to the purported tension between the Galactic Center excess and observations of dwarf spheroidal galaxies: p -wave annihilating dark matter, and strongly cored density profiles (consistent with self-interacting dark matter) for dwarf spheroidal galaxies.

Self-interacting dark matter, in addition to resolving some problems with the small scale structure of dark matter, can also result in the formation of supermassive black holes in the early universe. In chapter 4 we review some of the current theories on supermassive black hole formation at high redshifts, including from self-interacting dark matter. We present our own paper in which we determine the parameter space that allows supermassive black holes to form sufficiently early in the universe for three distinct self-interacting dark matter models.

In chapter 5 we discuss the results of these four papers and potential future work in the field, including the impact of future dark matter searches on our results.

1.1 Evidence for Dark Matter

Dark matter was arguably first discovered in 1933 by Fritz Zwicky [2], who observed the orbital velocities of member galaxies of the Coma cluster. Zwicky determined that the velocities of these galaxies were far too large to be

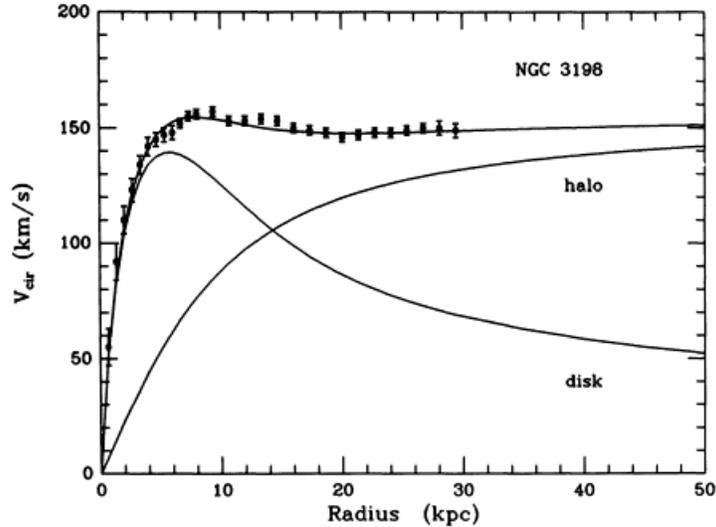


Figure 1–1: Plot of the rotation curve of NGC 3198 from [9]. The observed data is shown as dots, and the expected curve from visible (non-dark) matter is plotted, as well as the expected curve from a dark matter halo. The curve resulting from the sum of the two is shown to be an excellent fit to the data.

virialized given the visible matter content of the Coma cluster, and proposed that some unknown non-luminous matter was responsible for the discrepancy.

It was, however, Vera Rubin’s detailed measurements [3] of galactic rotation curves using the 21 cm line in the 1970’s which revealed the extent of this discrepancy. By measuring the line of sight velocities in nearby galaxies, Rubin was able to demonstrate that the rotation curve (the velocity as a function of radius) diverged significantly from that expected given the amount of luminous matter (stars, gas, and dust) in the host galaxies. This implied the existence of an extended ‘halo’ of non-luminous matter extending well beyond the disk of stars, gas, and dust, with approximately five times the mass of the luminous matter (Figure 1–1). This missing matter has since become known as dark matter. These results were borne out by subsequent measurements [4–10] which demonstrated missing matter is present at all scales from dwarf galaxies to galaxy clusters.

Further evidence of dark matter can be found in the baryon acoustic oscillations of our universe. Baryon acoustic oscillations are a feature of the matter power spectrum caused by sound waves in baryonic matter in the early universe, which cause a characteristic scale in the distribution and clustering of matter. This can be seen in the large-scale structure of matter nearer the present era [11–13]. Since their wavelength is related to both the matter density and baryon density, they can be used to place constraints on both the total gravitating matter content of the universe and the baryonic content. The characteristic scale may also be seen in the cosmic microwave background as seen by WMAP or Planck [14, 15] (Figure 1–2), where the variation in density causes recombination to occur at slightly later or earlier times. Like the galactic rotation curve, the power spectrum of the large scale structure and the CMB both show that the vast majority of matter is composed of non-luminous or ‘dark’ matter, and show that this has been true since before the era of recombination.

A third, critical line of evidence comes from the era of big bang nucleosynthesis, when the universe cooled enough for light elements to be generated and the abundances of these elements were set. The observed abundances of various isotopes (particularly helium-4 and deuterium) place constraints on the baryon to photon ratio in the universe [16]. These constraints preclude baryonic matter as a candidate for dark matter; if baryonic matter made up the missing mass during big bang nucleosynthesis the fraction of deuterium and helium-4 would be significantly altered from their observed values [17–19].

1.2 Cold Dark Matter Theories

The search for the missing matter has led to large number of theories. Since very little is known about the properties of dark matter, a wide variety of theories can be accommodated so long as they result in the correct

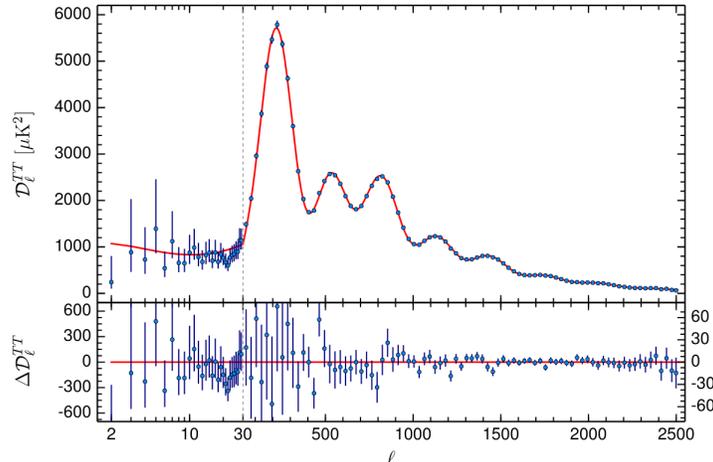


Figure 1–2: Temperature power spectrum of the cosmic microwave background as seen by the Planck space telescope, showing speaks in the power as a function of multipole (ℓ). The red line is the best-fit Λ CDM theoretical spectrum. The location of the peaks is highly sensitive to the content of the universe at early times. The lower panel shows the residuals when compared to this model. Image from the Planck 2015 results [15].

relic abundance and evade detection experiments. Given the constraints from big bang nucleosynthesis (and in many cases constraints from detector experiments), viable dark matter candidates require physics beyond the Standard Model. Most of these are examples of cold dark matter (CDM), matter which interacts very weakly electromagnetically, is non-relativistic when it decouples from the plasma in the early universe, and interacts with itself only weakly.

1.2.1 WIMP Dark Matter

Weakly-interacting massive particles (WIMPs) are a class of dark matter candidates which have weak-scale interactions and masses around 100 GeV. The interactions can either be through the weak nuclear force or simply have a weak-scale coupling to the Standard Model. This interaction scale and mass leads to the so-called ‘WIMP miracle’ by giving the particles an annihilation cross section of order $\langle\sigma v\rangle \approx 3 \times 10^{-26} \text{ cm}^3\text{s}^{-1}$ [20–22]. This annihilation cross section leads to the production of the correct relic abundance of dark

matter upon freeze-out [23], though there is a good deal of flexibility in both the interaction scale and the mass scale so long as the ratio $\sigma \propto \frac{g^4}{m_\chi^2}$ is kept roughly constant [24, 25], or if the annihilation proceeds through a mediator. Furthermore, particles with these properties tend to arise naturally in minimal supersymmetric theories, and dark matter could therefore be the lightest supersymmetric particle (LSP)¹.

Examples of WIMP dark matter are the neutralino [26, 27], extra Higgs particles [28], and a panoply of others. Besides predicting the correct relic abundance, these models have the benefit of being well-motivated given that supersymmetry is a popular method of solving other problems in high energy theory (such as incorporating gravity in the standard model and the gauge hierarchy problem). For a review of WIMP dark matter candidates (and the status of current and future WIMP searches), see Ref. [29].

1.2.2 Axions

Like WIMPs, axions are an appealing dark matter candidate because they are already well-motivated. The QCD axion is used in the Peccei-Quinn mechanism to solve the strong CP problem [30]. The strong CP problem is that the QCD Lagrangian allows a CP-violating² term, however the coefficient of this term is bounded by neutron electric dipole measurements to be at most $\sim 10^{-10}$. This can be solved by considering the coefficient to instead be a dynamic field with a global U(1) symmetry which becomes broken. The axion is the pseudo-Nambu-Goldstone boson resulting from the broken symmetry [31],

¹ Discrete symmetries, such as supersymmetry, ensure that a dark matter candidate charged under such a symmetry is stable so long as it is the lightest such particle.

² CP symmetry is the product of two symmetries, charge conjugation and parity.

however it acquires a mass due to non-perturbative effects and drives the field to zero, solving the strong CP problem.

Axions interact in an interesting way with the electromagnetic field, allowing the conversion of axions to photons in the presence of a large external magnetic field [32]. This leads to interesting possibilities for detection in the lab, such as the Axion Dark Matter eXperiment (ADMX), which uses a microwave cavity in a superconducting magnet with magnetic fields of up to 7 T to search for axionic dark matter [33].

Axions act as cold dark matter despite their very small masses [34], which are predicted to be in the $\mu\text{eV} - \text{eV}$ range. If axions are to be dark matter, the relic abundance must be produced non-thermally (not through a freeze-out mechanism) [35]. This can be done through vacuum misalignment (in which the field does not begin at its minimum, leading to oscillations of the field about its minimum as it is driven to zero) [34], or through the decay of topological defects such as cosmic strings and domain walls [36, 37].

1.2.3 Primordial Black Holes

Primordial black holes — black holes which formed in the very early universe ($t \sim 10^{-43} \text{ s} - 1 \text{ s}$) [38] — can arise due to overdense regions in a wide variety of scenarios [39–41], and can form with masses of anywhere from the Planck mass to $10^5 M_\odot$ depending on the time of formation. Primordial black holes are the last remaining viable candidate in what used to be a large class of promising theories, massive compact halo objects (MACHOs). These theories posited a large population of massive objects too dim to be observed by telescopes, such as neutron stars, black holes, brown dwarfs, or rogue³ Jupiter-like objects. Note that most of these are composed of baryonic matter

³ Rogue planets refer to those not closely associated with a star.

and therefore inconsistent with the constraints from big bang nucleosynthesis. Primordial black holes, on the other hand, would have formed prior to big bang nucleosynthesis and are therefore compatible with nucleosynthesis constraints.

Though they have largely been ruled out by microlensing surveys (for small black holes) [42, 43] and by the stability of compact stellar systems as well as the CMB [44] (for large black holes), a small window of weaker constraints exists in the $\sim 10 - 60 M_{\odot}$ range. Interestingly, this is the range in which the LIGO-Virgo collaboration has recently observed several black hole mergers [45–48], reigniting interest in the possibility of primordial black holes as dark matter [41, 49–51]. Recent measurements of type Ia supernovae, and specifically the lack of lensing signatures, seem to constrain primordial black holes to make up less than a third of dark matter [52]. This, along with previous constraints, disfavors primordial black holes as dark matter.

1.3 Dark Matter Detection

Searches for dark matter typically fall into three categories: direct detection (detection of dark matter through possible scattering events with baryonic matter such as atomic nuclei in detectors), indirect detection (in which dark matter annihilates or decays into Standard Model particles such as photons), or collider production (in which dark matter is produced from Standard Model particles in collider experiments). See figure 1–3 for a diagrammatic representation. So far, no definitive signals have been observed from any of the above methods, however each leads to constraints on dark matter interactions with the Standard Model.

1.3.1 Direct Detection of Dark Matter

In direct detection of dark matter, the goal is typically to detect nuclear recoils from collisions between a dark matter particle and a detector’s target nuclei (such as Xenon). In many models the scattering is mediated by the

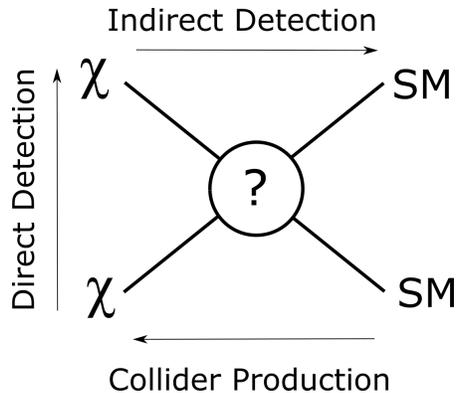


Figure 1–3: Representation of the three primary methods of dark matter detection. SM refers to any Standard Model particles, while χ represents the dark matter. The exact nature of the interaction is model-dependent.

weak force or a scalar or pseudoscalar, but could also occur through the Higgs portal or photons depending on the dark matter model. Steps must be taken to minimize backgrounds in direct detection experiments, and the detectors are therefore constructed deep underground.

Even with precautionary steps, determining whether or not a signal is from dark matter (rather than a background) is difficult. Dark matter searches through direct detection must search for characteristic signatures of dark matter. The most common method is to measure the energy dependence of dark matter interactions. Another signature characteristic of possible dark matter direct detection signals is an annual modulation of the signal [53]. As the Earth orbits the Sun, the velocity of the Earth relative to the Milky Way halo changes depending on whether the Earth and Sun’s velocities are in the same or opposite directions. The number of dark matter particles with energies high enough to produce detectable recoils would be maximized when these add (leading to the greatest relative velocity of the Earth with the Galactic halo) and smallest when they are opposite. Related to this is the directionality of the expected signal. A dark matter signal should have a greater event rate from particles coming from the direction opposite the sun’s motion through

the Galactic halo. As the Sun passes through the halo, it encounters a “dark matter wind”, which leads to not only a greater flux of dark matter particles from the forward direction but also higher energies [54].

Direct detection searches have found no definitive evidence of dark matter,⁴ but stringent bounds can be placed on the nucleon scattering cross sections. These are split into the spin-independent (SI) cross section and spin-dependent (SD) cross section. If the scattering is spin-independent, each nucleon in the target nucleus contributes to the scattering cross section, whereas if the interaction is spin-dependent only single unpaired nucleons will contribute, leading to a suppressed scattering cross section. Different targets are best suited for each. Some of the most stringent constraints on both the SI and SD cross sections are shown in figure 1–4.

For a more detailed review of dark matter direct detection, see Ref. [59].

1.3.2 Indirect Detection of Dark Matter

Telescope searches may be able to detect annihilating or decaying dark matter through the detection of photons or electrons/positrons by observing an excess in radiation (see Ref. [60] for a review).⁵

The expected signal changes depending on whether the dark matter annihilates directly to photons ($\chi\chi \rightarrow \gamma\gamma$) - which would produce a line [64–66],

⁴ Though see Ref. [55] for an overview of controversial potential signals, including that from DAMA which shows an annual modulation consistent with dark matter [56]. These potential signals, while consistent with dark matter, are in tension with null results from other experiments such as LUX and XENON.

⁵ Dark matter annihilation and decay can also affect the CMB through energy injection in the early universe [61–63], placing constraints on the annihilation and decay rate at this time. Here, however, we focus on searches for objects at $z \approx 0$, such as the Galactic Center, which will be discussed in chapter 3.

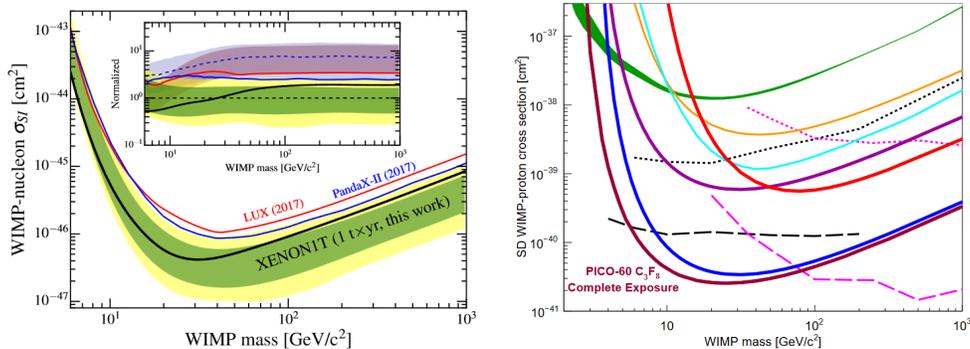


Figure 1–4: Left: Constraints on the WIMP-nucleon spin-independent scattering cross section from various experiments, including XENON1T (from which the figure is reproduced) [57]. The grey contours represent the favoured region of a specific supersymmetric model. Right: Constraints on the WIMP-nucleon spin-dependent scattering cross section from various experiments, including PICO-60 (from which the figure is reproduced) [58].

light leptons ($\chi\chi \rightarrow e^+e^-$) which would produce a line,⁶ or to quarks or heavy leptons (for example $\chi\chi \rightarrow b\bar{b}$), which can decay to a large number of final products at varying energies, and therefore produce a signal with an extended spectrum [71].

The primary difficulty in conducting such searches is that one must be able to identify a signal above a background produced by a wide variety of astrophysical processes. This requires not only detailed knowledge of the astrophysical processes involved, but also the selection of good observational

⁶ The line produced by annihilation to light leptons is broadened by interactions with the interstellar medium as well as through Bremsstrahlung radiation and inverse Compton scattering with starlight and the CMB. In this work we focus on the detection of γ -ray signals, however light leptons such as those produced in these processes could be directly measured by experiments such as AMS-02, a magnetic spectrometer on board the International Space Station designed to detect positron and antiproton fluxes [67, 68]. Possible signals from annihilation dark matter have been reported in the AMS-02 data [69, 70].

targets in which the dark matter density is expected to be high and background processes are minimized. The annihilation rate, in particular, is proportional to the square of the dark matter density, making it very sensitive to the distribution of the dark matter.

A second difficulty is that the signal is further affected by propagation through the galactic or intergalactic medium (depending on the source), and for chaotic environments such as galactic centers secondary processes such as inverse Compton scattering and Bremsstrahlung radiation play an important role in determining the final spectrum [72]. These environments are not always well known, and different choices of variables such as the magnetic field or radiation background can have a significant effect on the spectrum.

Ideal targets for indirect detection are ones in which the dark matter density is high and the background from astrophysical processes is small or well-known. Galaxy clusters, which are dominated by dark matter, make excellent targets [73]. Galactic haloes have high central densities, making the Galactic Center of the Milky Way another high priority target for experiments such as H.E.S.S. [74] and Fermi-LAT [75] despite the fact that the background is poorly known. Finally, dwarf spheroidal galaxies [76] and low surface brightness galaxies [77] make extremely good targets as they have relatively high densities of dark matter and are almost entirely dominated by it, leading to minimal backgrounds. Sample limits on the velocity-averaged annihilation cross section from Fermi-LAT data are shown in figure 1–5.

1.3.3 Collider Production of Dark Matter

Collider production relies on high energy collisions of Standard Model particles such as electrons or protons to produce dark matter in a manner similar to the production of heavier Standard Model particles such as the Higgs. In collider production, dark matter is not directly detected but rather

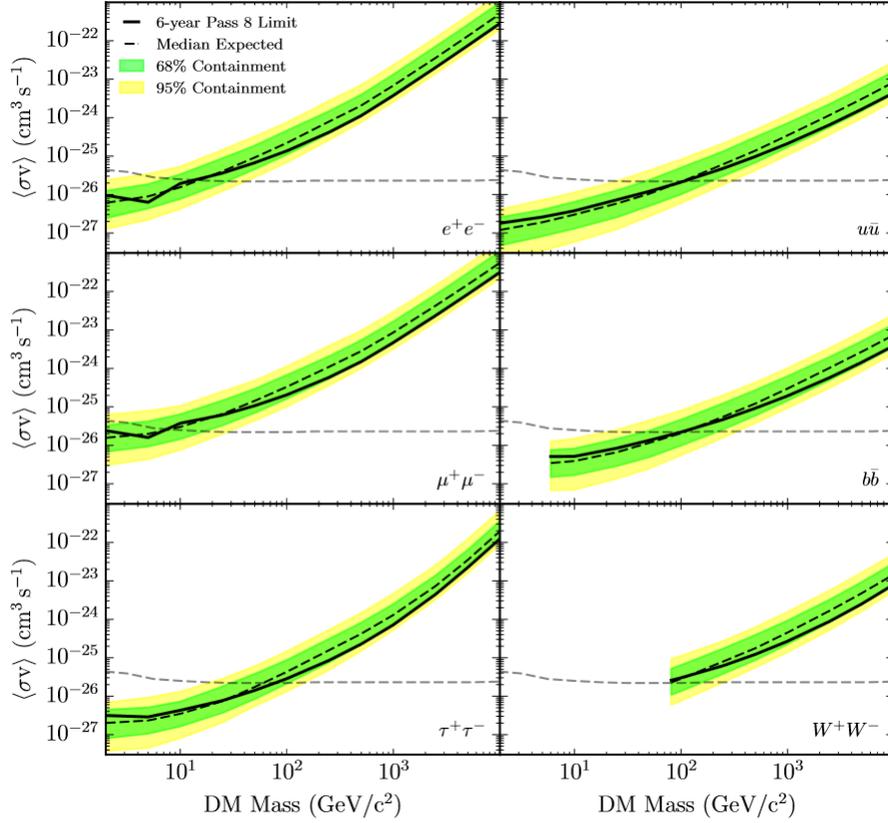


Figure 1–5: Combined limits on the velocity-averaged annihilation cross section to various channels from 6 years of observation of 15 dwarf spheroidal galaxies by Fermi-LAT. The dashed-line represents the annihilation cross section which reproduces the WIMP miracle. The figure is reproduced from Ref. [78].

inferred from missing energy and momentum accompanied by an energetic jet. The exact nature of the search depends on the specific model, but generally the production proceeds as $p + p \rightarrow \chi + \bar{\chi} + \text{SM}$, where SM is a photon, a hadronic jet, a Higgs, or a Z or W .

The most stringent bounds on collider production of dark matter are from CERN’s Large Hadron Collider (LHC), which searches for dark matter using the ATLAS and CMS experiments by colliding protons with energies of 13 TeV [79, 80]. Collider constraints can be translated to constraints on specific models or classes of models, as different models will result in the production of very different final states [59]. Rather than place universal bounds on dark matter parameters, the bounds will depend on both the nature of the dark matter candidate in question (for example scalar, fermion, or vector boson) and on the mediator between dark matter and the Standard Model. For many models, these constraints are competitive with those from direct detection experiments. See figure 1–6 for an example of constraints from the ATLAS experiment. For a review of collider searches for dark matter, see Ref. [81].

1.4 Problems for the CDM Paradigm

Detection of dark matter (see section 1.3) typically relies on its interactions with the Standard Model, which are known to be very weak and may even be nonexistent. Beyond this, one of the few means we have of studying dark matter is to examine the astrophysical properties of dark matter halos and subhalos (for example, dwarf galaxies). A characteristic property of CDM is that it has a tendency to form hierarchical structures; each halo contains a large number of smaller subhalos. Furthermore, these halos are self-similar, all sharing the same overall density profile [83]. CDM also makes key predictions about the distribution of subhalos within their host. The number and

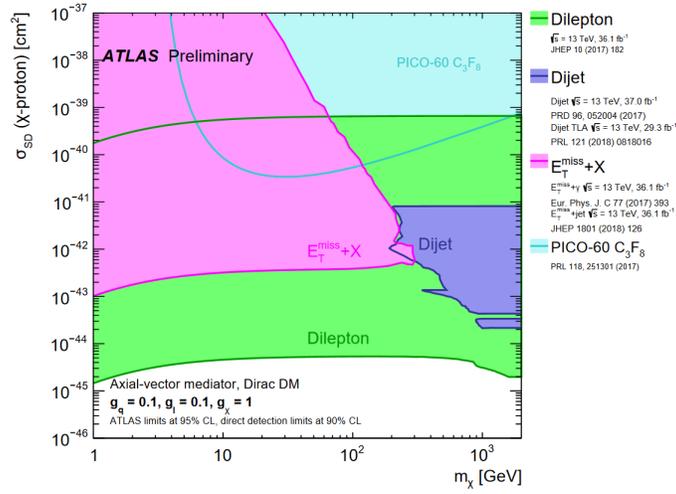


Figure 1–6: Limits from various searches with the ATLAS experiment on the spin-dependent WIMP-neutron scattering cross section for axial-vector mediated couplings to the Standard Model for Dirac dark matter. Specific couplings have been assumed for the model in order to derive the constraints. The pink region is ruled out by searches for missing energy in conjunction with observed particle(s) (X), the purple by dijet events, and the green by dilepton events. In dijet and dilepton searches the spectrum of events from at least two jets or leptons is analyzed in search of an excess that could be the result of a resonance from dark matter contributions to the diagrams. The blue region shows a direct detection constraint on spin-independent events from PICO. Figure is reproduced from Ref. [82], see original paper for more details.

statistical distribution of subhalos can be observed and compared to the results of N-body simulations, as can the density profile of the halos. The result of these comparisons seems to point to several difficulties for the cold dark matter (CDM) paradigm related to its structure on small scales. Here we will review three key ones: the core/cusp problem, the missing satellites problem, and the ‘too-big-to-fail’ problem.

1.4.1 The Core/Cusp Problem

The core/cusp problem refers to the inner part of a dark matter halo’s density profile. N-body simulations point to ‘cuspy’ profiles, well fit by functions which diverge at small radius [83–89]. For example the Navarro-Frenk-White profile [85],

$$\rho_{\text{NFW}} = \frac{\rho_s}{\frac{r}{R_s} \left(1 + \frac{r}{R_s}\right)^2}, \quad (1.1)$$

in which ρ_s and R_s are the scale density and radius respectively, diverges as r^{-1} at small radii (a cusp). This can be contrasted to a profile such as the Burkert profile [90],

$$\rho_{\text{Burkert}} = \frac{\rho_s}{\left(1 + \frac{r}{R_s}\right) \left(1 + \left(\frac{r}{R_s}\right)^2\right)}, \quad (1.2)$$

which does not diverge in the center, and instead flattens to a near-constant density when $r < R_s$ (a core).

Although N-body simulations of halo formation point to CDM forming cuspy halos, with $\gamma = d \ln \rho / d \ln r \gtrsim 1$ [83–89], observations of dwarf galaxies point to cored profiles $\gamma < 1$ [90–97]. Some solutions to this problem from baryonic effects have been proposed, most notably baryonic feedback from supernovae or active galactic nuclei which transfer energy to the dark matter halo, resulting in a more cored profile [98, 99]. In very dark matter-dominated

environments, however, such as low surface brightness galaxies [77], these effects should be minimal. In spite of this, LSBs are found to have significantly cored profiles [100, 101].

1.4.2 The Missing Satellites Problem

The missing satellites problem arises from the observation that despite correctly predicting the number and distribution of large (galaxy-sized) halos in the universe, CDM simulations predict a much larger number of galactic subhalos [102] than the observed number of dwarf galaxies in the Milky Way (by several orders of magnitude) [103] (see figure 1–7). This suggests that some process is preventing the formation of small subhalos specifically, disrupting the CDM prediction of self-similarity. It is possible tidal stripping or mergers may act to suppress the formation of satellites, however there is disagreement on whether this effect is large enough to account for the disparity, and the effect is also observed for galaxies in the field (not within a cluster or group) where these effects should be minimal [104, 105].

Although this is troubling for cold dark matter models, it is also possible that these halos exist but have simply not been observed because they do not contain large quantities of baryonic matter. This possibility is supported by the observation of ultra-faint dwarf galaxies by the Keck Space Telescope in 2007. These dwarf galaxies have extremely high mass to light ratios and are almost entirely dominated by dark matter [106]. More dwarf galaxies have been observed since then by the Sloan Digital Sky Survey and Dark Energy Survey [107–111]. This suggests that the ‘missing’ satellites could simply be very dim and therefore as yet unobserved.

1.4.3 The Too-Big-To-Fail Problem

A related problem is the ‘too-big-to-fail’ problem. Instead of focusing on the smaller subhalos as in the missing satellites problem, the too-big-to-fail

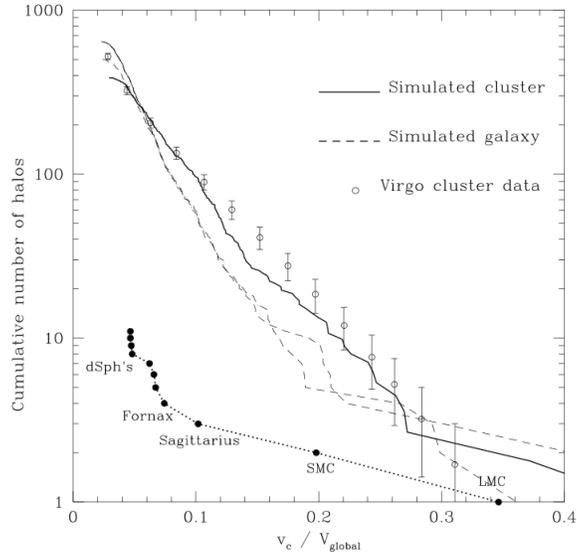


Figure 1–7: Figure from [103], showing the discrepancy between predicted CDM results and observed results for the cumulative number of sub-halos as a function of their circular velocity v_c (in terms of the circular velocity of their parent halo V_{global}). On the largest (cluster) scales, the observed distribution (open circles) is well-reproduced by CDM simulations (solid line), however on galactic scales the observed number of halos (solid circles) is much lower than that produced in CDM simulations (dotted lines, each showing the results at different epochs to demonstrate that these predictions do not change over time).

problem refers to the discrepancy between the observed densities of the largest Milky Way satellites, which are lower than those predicted by CDM simulations [103, 112]. In other words, the largest, densest galaxies expected to be produced by CDM models are missing. Though this problem could potentially be alleviated by tidal stripping effects in a similar fashion to the missing satellites problem, this problem has also been shown to apply to galaxies in the field (not strongly associated with clusters or groups) [113]. This suggests that tidal stripping is not a sufficient explanation as these galaxies have evolved in a more isolated environment. It has also been suggested that other baryonic feedback effects, such as gas outflows from supernovae, are insufficient to solve this problem for galaxies in the field [113].

1.4.4 Warm Dark Matter

One potential solution to the small-scale structure problems is warm dark matter — dark matter that was semi-relativistic during freeze-out [114, 115]. This could, for example, be a sterile neutrino with mass of a few keV (an appealing prospect given the observation of an anomalous 3.55 keV line observed in some clusters) [116]. This allows dark matter to free stream out of sufficiently small halos, suppressing their formation and effectively smoothing dark matter substructure [117, 118]. Warm dark matter models also form cores in small halos set by the free streaming length, erasing the cuspy inner region of the profile [119–124].

Warm dark matter, however, may be in tension with constraints from the Lyman- α forest [125–128] that restrict it to masses too large to form sizeable cores [122], and may even be too effective at erasing small-scale structure [129]. We therefore consider an alternative explanation: self-interacting dark matter.

1.5 Self-Interacting Dark Matter

Although there is disagreement about the degree to which baryonic effects can solve the aforementioned problems encountered in CDM-only simulations, a common solution to all three is to instead consider self-interacting dark matter (SIDM)⁷. For a thorough review of SIDM, see the review by Tulin & Yu [130]. SIDM was originally proposed by Spergel & Steinhardt to solve both the core/cusp problem and the missing satellites problem [131]. If dark matter is permitted to scatter with itself with a cross section on the order of

⁷ Although the dark matter in these models is usually ‘cold’ in the sense that it is non-relativistic during the era of kinetic decoupling, the term ‘cold dark matter’ is typically reserved for collisionless dark matter.

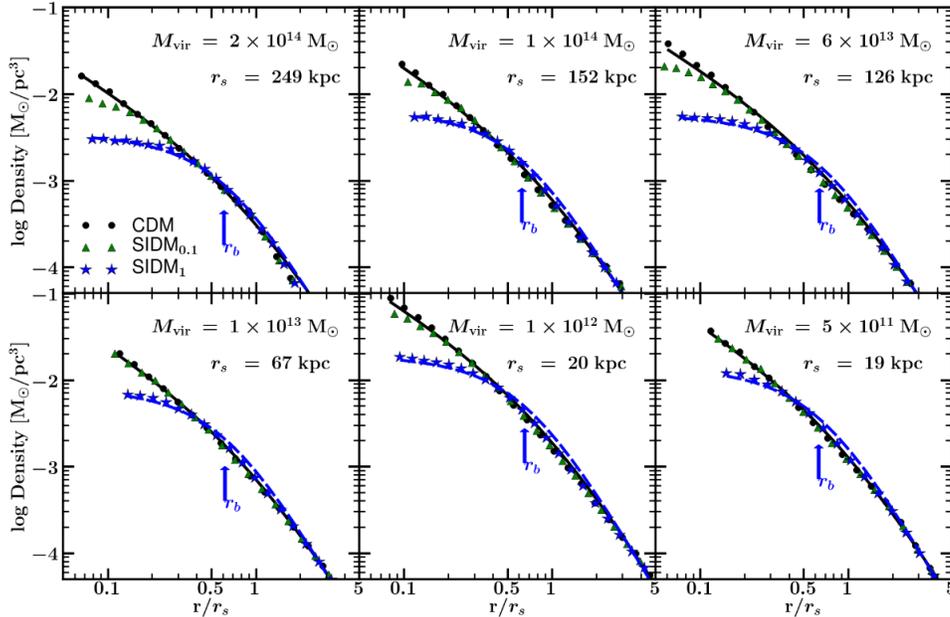


Figure 1–8: N-body simulation results from Rocha *et al.* comparing the density profiles of CDM (black circles), which are well fit by a cuspy NFW profile (black line) and those of SIDM. The green triangles represent $\sigma/m_\chi = 0.1 \text{ cm}^2/\text{g}$ and blue stars correspond to $\sigma/m_\chi = 1.0 \text{ cm}^2/\text{g}$. The latter are well fit by cored Burkert profiles. Here M_{vir} is the virial mass of the halo, r_s is the scale radius of the NFW fit and r_b is the Burkert profile’s core radius. The SIDM is seen form cored profiles for the larger scattering cross section.

$\sigma/m_\chi \approx 0.1 - 1 \text{ cm}^2/\text{g}$, cored profiles form during structure formation and the formation of subhalos is suppressed.

The resolution of the core/cusp problem relies on the fact that dark matter collisions are able to conduct heat throughout the halo. Heat can flow throughout the dark matter halo to thermalize the inner region, increasing its velocity dispersion and leading to the generation of a central core [131–133]. This effect can be seen in figure 1–8, which shows the results of N-body simulations [132]. The formation of cored profiles can be seen for models of SIDM with a self-interaction cross section of $\sigma/m_\chi = 1 \text{ cm}^2/\text{g}$ (and to a lesser extent for $\sigma/m_\chi = 0.1 \text{ cm}^2/\text{g}$).

Despite it being an initial motivator for SIDM, SIDM can address the missing satellites problem only with some difficulty. Smaller halos have an extended core and are much less concentrated, increasing the effects of tidal stripping relative to CDM halos and disrupting the subhalos through stripping as they pass through the host halo [131, 134]. Although cross sections of order $\sigma/m_\chi = 10 \text{ cm}^2/\text{g}$ are required in order to achieve this effect (which would be in strong tension with bounds from the Bullet Cluster — see below) [132, 135], a velocity-dependent cross section which is inversely proportional to the relative velocity would significantly suppress scattering in the Bullet Cluster merger event and in larger halos — preserving their agreement with CDM simulations — relative to dwarf galaxy halos [135, 136]. Such a model can be theoretically motivated by a sector in which the scattering is mediated by a new boson, leading to a scattering cross section proportional to v^{-2} [135]. More complex models of SIDM which do not scatter only elastically may also still be able to address this problem [137]. One such model is that of atomic dark matter [138], discussed further in chapter 2.

Self-interacting dark matter is, however, able to resolve the ‘too-big-to-fail’ problem. CDM simulations predict a large number of very dense subhalos in Milky-Way like halos. Self-interacting dark matter with a scattering cross section of $\sigma/m_\chi \approx 1.0 \text{ cm}^2/\text{g}$, by generating more cored profiles, decreases the central density of dwarf galaxies [131–133]. The galaxies which would have the highest densities under CDM (and are not observed) have their central densities reduced to levels similar to the most massive dwarf galaxies in the Milky Way. This resolves the ‘too-big-to-fail’ problem, creating a distribution of dwarf galaxies consistent with observations [135, 136].

1.5.1 Constraints on Dark Matter Self-Interactions

The most stringent constraint on the self-interaction cross section of dark matter comes from the Bullet Cluster, which constrains the scattering cross section per unit mass to $\sigma/m_\chi < 0.7 \text{ cm}^2/\text{g}$ [139]. The Bullet Cluster is a merging galaxy cluster in which a subcluster has passed through the main cluster. Since most of the baryonic matter is made up of collisional gas, it is slowed by the ram pressure as the subcluster passes through the main cluster resulting in bow shocks [140]. If dark matter were to have a very high scattering cross section, it would also be slowed; if it were completely non-collisional the dark matter halo of the subcluster would pass through that of the main cluster without any of its mass being stripped. Gravitational lensing can be used to map the mass of the cluster as a whole [141], while X-ray observations can be used to map the gas (luminous matter). There is a visible offset between the two (Figure 1–9), with the gas lagging behind the mass of the subcluster⁸. By measuring the mass-to-light ratio of the clusters, constraints can be placed on the amount of dark matter which has been stripped away from the cluster and therefore on the scattering cross section [139, 142]. It is worth noting that the constraint on the self-interaction cross section derived from the bullet cluster applies at speeds of $\sim 10^3 \text{ km/s}$; velocity-dependent cross sections could result in greater self-interactions at smaller speeds such as those found in dwarf galaxies ($\sim 10 \text{ km/s}$).

By resolving or mitigating these three problems (to varying degrees) while maintaining agreement with the bounds from the Bullet Cluster, models of

⁸ Additionally, this acts as direct evidence for dark matter; the offset between the luminous matter and the mass indicates that the majority of matter within the cluster is non-luminous and either non-collisional or weakly collisional.

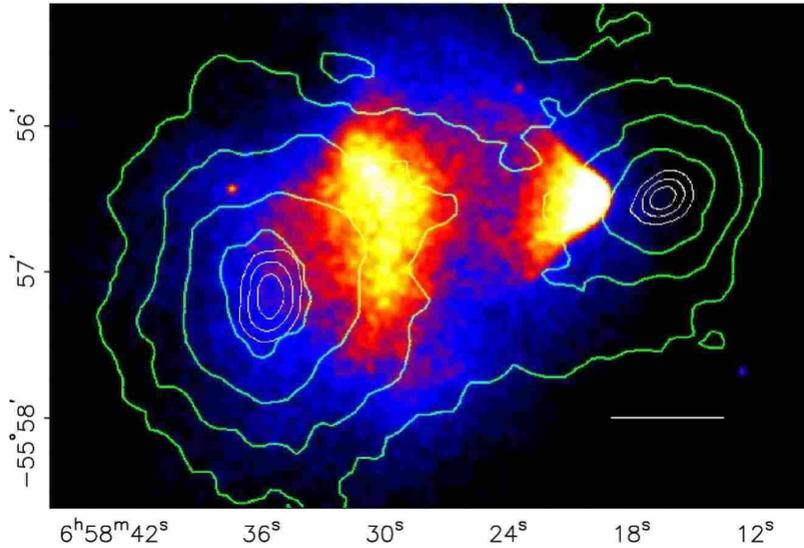


Figure 1–9: Image showing the mass offset of the Bullet Cluster. The color image shows the distribution of gas in the cluster, with the bow shocks visible in both the main cluster (left) and subcluster (right). The contours show the distribution of mass from weak lensing surveys [143].

self-interacting dark matter have become an area of intense research. Self-interacting dark matter allows for rich model building and creates opportunities for detection through portals to the Standard Model, which can be naturally created by the dark force mediating the self-interactions (for example Higgs portal DM [144], or kinetic mixing of a dark photon [145] or Z' [146]). In the following chapters, we explore in more detail several models of self-interacting dark matter, its effects on the structure of galactic halos and subhalos, and related topics.

CHAPTER 2

A Model of Atomic Dark Matter

Self-interacting dark matter presents an opportunity for rich model-building. One class of SIDM models is that of atomic dark matter, in which bound states between oppositely charged dark matter particles result in a large self-interacting cross section and interesting phenomenology.

Significant self-interactions, however, often result in large interactions with the Standard Model (through scattering with or annihilation to Standard Model particles). This is because large self-interaction cross sections (whether scattering or annihilation) under Standard Model gauge groups require large couplings which will inevitably lead to sizeable interactions with Standard Model particles. These can be in tension with detection experiments (as discussed in section 1.3). This can be resolved by placing dark matter within a ‘hidden’ sector: additional gauge symmetries under which dark matter can transform. If the model is to maintain interesting opportunities for detection, it should allow some portal to the Standard Model while still reproducing the successes of traditional CDM models.

We present one such model of atomic dark matter here. In section 2.1, we give a brief review of hidden sector dark matter, and in 2.2 we shift our view to a specific subset of hidden sector models: dark atoms. In section 2.3 we give a very brief overview of leptogenesis. In section 2.4 we present our work, a minimal nonabelian model of atomic dark matter.

2.1 Hidden Sector Dark Matter

A growing class of dark matter theories is that of hidden sector dark matter. In hidden sector models, dark matter is composed of particle species

which do not transform under any of the Standard Model gauge groups [147–152]. This leads to limited (or non-existent) interactions with Standard Model particles beyond gravity. If the dark matter in these models interacts with the Standard Model at all, it is generally through kinetic mixing [150] (for example the mixing of a dark photon with the Standard Model photon) or through a Higgs portal [149]. These models can range from very simple to extremely rich, with a wide variety of particle species interacting under their own gauge group(s). For example, early models of hidden sector dark matter, known as mirror dark matter, included complete or near-complete copies of the Standard Model, resulting from a gauge group $G \otimes G$ with $G = SU(3) \otimes SU(2) \otimes U(1)$ with a Z_2 symmetry interchanging between the two sectors [147, 148].

Hidden sector models of dark matter have two key advantages: their ability to more easily accommodate large self-interaction cross sections and their use of additional symmetries by which heavy particles can be made stable.

Due to their secluded nature, self-interaction cross sections between hidden sector dark matter particles are much less heavily constrained than those in models that make use of the Standard Model gauge groups. As mentioned above, couplings through Standard Model gauge groups large enough to result in the desired scattering cross section will result in significant interactions with Standard Model particles. These interactions could be detected through direct detection, indirect detection, or even production in colliders. Hidden sector dark matter avoids these limits by allowing small (or even no) couplings to Standard Model gauge groups without compromising self-interactions.

The second point — the stability of heavy particles — is a result of the fact that most theories of dark matter introduce new heavy particles with masses > 1 GeV. If a dark matter candidate is allowed to interact with the Standard Model, new symmetries must be present to bar it from decaying to

Standard Model particles (or, in the rare cases where it is lighter than Standard Model particles, to prevent their decay into dark matter). For example, several models of WIMP dark matter rely on supersymmetry (for example neutralino, sneutrino, or gravitino dark matter) [26, 27] to allow stable massive particles, as the lightest supersymmetric particle would necessarily be stable unless R -parity is violated. In hidden sector models, new gauge groups create new symmetries which must be observed, allowing massive particles to be stable.

Although working within the framework of a hidden sector allows a large variety of models, this does risk leading to runaway unconstrained model-building without any predictive power. It can also leave dark matter models without some of the more interesting features of traditional WIMP models, such as their connection to supersymmetric theories or the gauge hierarchy problem. Motivating specific models of hidden sector dark matter generally requires consideration of the ways in which the model would lead to detectable signals through direct or indirect detection, as well as how the observed dark matter relic density is produced.

2.2 Dark Atoms

Dark atoms are a well-studied example of hidden sector dark matter [138, 153–171]. In atomic dark matter, the dark matter consists of two oppositely charged particle species under a $U(1)$ gauge symmetry. These particles would attract through their interactions mediated by the hidden sector (or ‘dark’) photon, and behave similarly to the electron and proton of the Standard model, though they could have entirely different masses and need not have the other properties of electrons and protons (for example, both could be scalars or vector bosons). Both remain stable due to charge conservation — they are

the lightest positively and negatively charged particles under the dark U(1) symmetry.

Initially modelling dark matter as the stau (from a hidden sector copy of the minimal supersymmetric standard model), Feng et al. [138] showed that dark atoms have several desirable qualities, including making dark matter self-interacting through Rutherford scattering and suppressing small-scale structure formation due to Compton scattering (which results in delayed kinetic decoupling). This allows dark atoms, unlike many other models of self-interacting dark matter, to address the missing satellites problem (see section 1.4).

Interactions with the Standard Model could appear through the Higgs portal or if the dark photon mixes kinetically with the normal photon. Higgs portal dark matter allows dark matter to interact with the Higgs, allowing Higgs-mediated interactions with Standard Model particles. Kinetic mixing of the dark photon occurs when the dark photon is allowed to couple directly to the Standard Model photon, though the term $\mathcal{L}_{\text{k.m.}} \propto -B^{\mu\nu}Y_{\mu\nu}$ (where $Y_{\mu\nu}$ is the Standard Model photon gauge field and $B^{\mu\nu}$ is the dark photon gauge field). The dark photon can be massive or massless depending on the model. In the latter case, dark matter can be considered to be ‘millicharged’, having a small electromagnetic charge of $\epsilon \lesssim 10^{-3}e$ [145]. This opens the possibility of indirect detection (through the production of SM photons) [172, 173], production in collider searches [174–177], and even direct detection [178–180], though in the case of dark atoms the mutual screening of the millicharged constituents reduces their scattering cross section with nucleons significantly if they have roughly equal masses.

2.3 Leptogenesis

An unsolved problem in cosmology is the origin of the baryon and lepton asymmetries, in other words why the universe contains more matter than antimatter. The baryon asymmetry in particular is thought to be generated dynamically (through symmetry breaking) rather than as an initial condition (which would be erased by inflation), and must be generated prior to Big Bang Nucleosynthesis in order to yield the correct proportions of light elements [181]. The process through which this occurs is generically known as baryogenesis. In order to produce the asymmetry, new interactions must exist that violate both baryon number and CP symmetry; additionally the baryons must be produced out of equilibrium otherwise no asymmetry will be produced [182].

Leptogenesis is a mechanism for producing the baryon asymmetry without relying on Grand Unified Theories (for a review, see Ref. [181]). In leptogenesis, an asymmetry is first produced in leptons and is then translated to a baryon asymmetry. The mechanism relies on the addition of very heavy right-handed neutrinos. Such right-handed neutrinos are often introduced in the context of explaining the small masses of the three known neutrinos through the see-saw mechanism [181]. As the heavy right-handed neutrinos decay out of equilibrium in a CP-violating manner, they produce a lepton asymmetry. The lepton asymmetry is then converted to a baryon asymmetry through sphaleron processes [183], producing the observed baryon asymmetry sufficiently early in the universe.

2.4 Our Work

A goal of our work was to produce a consistent model of atomic dark matter which could also simultaneously explain the baryonic and dark matter asymmetries through leptogenesis. This minimal model introduces a dark

SU(2) gauge group which is broken to U(1) in an identical manner to the Standard Model SU(2), with a dark sector Higgs boson providing the symmetry breaking (rendering the dark photon massless). A pair of fermions, Ψ — which gain opposite charges under U(1) and can therefore bind into dark atoms — and a heavy scalar doublet η complete the model.

In one regime, the dark atoms are made up of the new Dirac fermion doublet, with dark matter composed of Ψ_1^+ , Ψ_2^- . However if these are massive enough one of the fermions can decay into stable vector bosons with double charges. In this case, novel bound states are formed which are analogous to a Hydrogen molecule (\mathbf{H}_2), but with both ‘electrons’ instead composed of a single doubly charged vector boson. We verify the existence of these three-body bound states and determine their binding energies.

The model has three means of interacting with the standard model: through the Higgs, photons, and the heavy neutrino. This makes the model phenomenologically interesting by creating many opportunities for direct detection, indirect detection, or production.

Our work neglected to take into account constraints from dark acoustic oscillations, which have been derived in the discussion of this thesis (chapter 5).

This manuscript was published in Physical Review D in 2015 [184]. The introduction was written collaboratively. Sections 2 and 5.3 were primarily written by my co-author, James Cline; the rest is primarily my own work.

Minimal Nonabelian Model of Atomic Dark Matter

Jeremie Choquette*¹ and James M. Cline^{†1,2}

¹*Department of Physics, McGill University, 3600 Rue University, Montréal, Québec, Canada H3A 2T8* ²*Niels Bohr International Academy and Discovery Center Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark*

Abstract

A dark sector resembling the standard model, where the abundance of matter is explained by baryon and lepton asymmetries, and stable constituents bind to form atoms, is a theoretically appealing possibility. We show that a minimal model with a hidden $SU(2)$ gauge symmetry broken to $U(1)$, with a Dirac fermion doublet, suffices to realize this scenario. Supplemented with a dark Higgs doublet that gets no VEV, we readily achieve the dark matter asymmetry through leptogenesis. The model can simultaneously have three portals to the standard model, through the Higgs, nonabelian kinetic mixing, and the heavy neutrino, with interesting phenomenology for direct and collider searches, as well as cosmologically relevant DM self-interactions. Exotic bound states consisting of two fermions and a doubly-charged vector boson can exist in one phase of the theory.

Dark matter (DM) from a hidden sector has been a popular alternative to supersymmetric weakly interacting massive particles in recent years [1, 2]. A widely studied example is dark atoms, where the DM consists of two species with opposite charges under an unbroken $U(1)_h$ hidden sector gauge symmetry [3–9]. This class of models presents rich possibilities for direct detection [10–13], as well as cosmological imprints [14–19]. If the hidden photon has kinetic mixing with the normal photon, the dark constituents acquire electric millicharges [20], leading to further constraints and prospects for detection [21–23].

Simplified models of atomic dark matter are easy to construct, consisting of just two fermions and the gauge boson in the hidden sector, but such examples are necessarily incomplete descriptions of the new physics required. First, it is desirable for the DM to be asymmetric, otherwise the long-range $U(1)_h$ interaction would leave too small a relic abundance unless the DM mass exceeds ~ 400 GeV [24]¹. Simplified models do not explain the origin of the asymmetry. Second, the $U(1)_h$ gauge interaction leads to a Landau pole at high energies, so it would be desirable to find a more UV-complete version of the theory. Third, dark constituent millicharges greater than $\sim 10^{-7}e$ (of interest for collider searches) require the atomic constituents to be nearly equal

¹ For lower masses the DM self-interactions violate bounds from structure formation. This argument assumes that the DM remains ionized, which turns out to be valid for the gauge coupling strength needed to get the right relic density from thermal freezeout.

in mass, which is a rather ad hoc requirement in the simplified models. In this work we present a model that is still relatively simple, but addresses both of these issues, and makes a number of interesting experimental predictions. It relies upon breaking a nonabelian (hence asymptotically free) gauge symmetry $SU(2)_h$ down to $U(1)_h$ to explain the origin of the massless dark photon. The approximate equality of the dark constituents, if desired, can be explained as a remnant of the gauge symmetry.

There have been many proposals for mechanisms that link the asymmetries of the hidden and visible sectors. In general, they tend to be complicated. A notable exception is to use the out-of-equilibrium decays of heavy neutrinos to generate both asymmetries via leptogenesis and its analog in the hidden sector [5, 25–30]. We adopt this approach here.

The model presents opportunities for direct detection, either through Higgs portal interactions or nonabelian gauge kinetic mixing. The latter can arise through a dimension-5 operator involving the triplet Higgs field that breaks the $SU(2)_h$ gauge symmetry [31]. This results in electric millicharges for the dark matter constituents, that normally must be very small to avoid direct detection, but can be sizable if the dark constituents have equal mass, which is a symmetry limit of the theory presented here. Moreover the self-interactions of the dark atoms can be of the right magnitude for addressing problems of small-scale structure formation in standard noninteracting Λ CDM cosmology.

In the following we introduce the model (section 1) and then estimate the dark matter and baryon asymmetries that can arise in a generic scenario for leptogenesis (section 2). Limits from direct searches are worked out in section 3. In section 4 we consider the region of parameter space in which the vector bosons are stable, leading to a markedly different dark sector. In

sect. 5 we discuss constraints pertaining to the ionization fraction, dark atom self-interactions, and searches for millicharged particles. Conclusions are given in sect. 6.

1 The model

The new-physics content of the model (summarized in table 1) is a hidden $SU(2)_h$ gauge boson B_μ with field strength $B_{\mu\nu}^a$, a real scalar triplet ϕ that spontaneously breaks $SU(2)_h$ by getting a VEV, a scalar doublet η that does not get a VEV, two Weyl fermion doublets ψ_i^α (with gauge index α and flavor index i) and the heavy right-handed neutrinos N_j that also interact in the usual way with the standard model neutrinos. An even number of fermion doublets is required to avoid Witten’s global $SU(2)$ anomaly [32]. They can be combined into a Dirac doublet fermion $\Psi = (\psi_{1L}, \psi_{2R}^c)$ where the conjugate is defined as $\psi_{2R}^c = \sigma_2 \tau_2 (\psi_{2L})^*$, *i.e.*, the epsilon tensor is applied both to the spin and to the $SU(2)_h$ gauge indices. Without loss of generality the VEV of ϕ can be rotated to the 3rd component, $\langle \phi^a \rangle = (0, 0, \sigma)$.

particle VEV	$B^{0,++,--}$	$\phi^a \rightarrow$ (0, 0, $\sigma +$ ϕ)	$\eta^{+,-}$	$\Psi_1^{+,-}$	$\Psi_2^{+,-}$	N_j
Spin	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$SU(2)_h$	3	3	2	2	2	1
$U(1)_h$	0, +2, -2	0	+1, -1	+1, -1	+1, -1	0

Table 1: New particle content in the model, showing the Lorentz, hidden $SU(2)$ and hidden $U(1)$ (after breaking of $SU(2)_h \rightarrow U(1)_h$) quantum numbers.

The relevant terms in the Lagrangian are

$$\mathcal{L} = -\frac{1}{4}B_{\mu\nu}^a B_a^{\mu\nu} + \frac{1}{2}(D_\mu\phi)^2 - \frac{1}{\Lambda}\phi^a B_{\mu\nu}^a Y^{\mu\nu} \quad (1)$$

$$+ \bar{\Psi}(i\not{D} - m_\psi)\Psi - \bar{\Psi}(y_1 + iy_2\gamma_5)(\vec{\phi} \cdot \vec{\tau})\Psi \quad (2)$$

$$- |D_\mu\eta|^2 - V(H, \phi, \eta)$$

$$- (\bar{\psi}_L^i \eta) y_\psi^{ij} P_R N_j + \text{h.c.}$$

where the covariant derivative is $D_\mu\phi^a = \partial_\mu\phi^a - g\epsilon_{abc}B_\mu^b\phi^c$ or $D_\mu\Psi = (\partial_\mu - i(g/2)\vec{B}_\mu \cdot \vec{\tau})\Psi$, g is the $SU(2)_h$ gauge coupling, and $Y_{\mu\nu}$ is the Standard Model hypercharge gauge field strength. In the Yukawa interactions with the sterile neutrino we use the Weyl fermion notation since the analogy to leptogenesis via neutrino physics is more clear in this way.

The triplet scalar VEV breaks $SU(2)_h$ to $U(1)_h$, mediated by the massless gauge boson B_3^μ , while $B^{\pm\pm} = (B^1 \pm iB^2)/\sqrt{2}$ obtain mass $m_B = g\sigma$. The upper and lower components $\Psi_{1,2}$ of the fermion doublet are also charged under the $U(1)_h$ (with half the charge of $B^{\pm\pm}$). Their masses are split by the Yukawa interaction, $m_{1,2} = ((m_\psi \pm y_1\sigma)^2 + (y_2\sigma)^2)^{1/2}$. We used the freedom to perform a chiral rotation on Ψ so that m_ψ is real (has no γ_5 component).

In section 2.4 we discuss the decay of the scalars through $\eta \rightarrow \psi\nu$ via the dimension-5 operator

$$\bar{\psi}_{i,L}\eta y_\psi^{ij} M_j^{-1} y_\nu^{jk} P_R (H^T \bar{L}_k^T) + \text{h.c.} \quad (3)$$

where M_j is the heavy neutrino mass (in a basis where its mass matrix is diagonal), y_ν is the neutrino Yukawa matrix, H is the SM Higgs doublet, and L_k are the lepton doublets.

We will initially consider the case where decays $\Psi_1 \rightarrow B^{++}\Psi_2$ are not kinematically allowed. They would lead to a dark sector consisting of stable B^{++} vector bosons and Ψ_2^- fermions. (The alternative case in which these

decays are allowed is considered in section 4.) This leaves two species of stable dark matter, the Dirac fermions $\Psi_1 = (\psi_{1L}^1, \psi_{2R}^{2c})^T$ and $\Psi_2 = (\psi_{1L}^2, \psi_{2R}^{1c})^T$ with charges ± 1 under the unbroken $U(1)_h$. The long-range force mediated by the dark photon $B_3 \equiv \gamma'$ causes the symmetric component of the DM densities to be at least partially depleted by annihilations, and the asymmetric components of $\Psi_{1,2}$ to bind into dark atoms. The efficiency of these processes depends upon the gauge coupling g and the dark atom mass $m_{\mathbf{H}}$, as we will discuss in section 5.

For simplicity we impose a softly broken $U(1)$ symmetry under which $\psi_i \rightarrow e^{i\theta} \psi_i$, $\eta \rightarrow e^{i\theta} \eta$, which forbids the interactions $(\bar{\psi}_i \tilde{\eta}) N_j$, with $\tilde{\eta} = \tau_2 \eta^*$. The symmetry is broken by the Dirac mass term, which takes the form

$$- m_\psi (\bar{\psi}_{2R}^{2c} \psi_{1L}^1 + \bar{\psi}_{2R}^{1c} \psi_{1L}^2) + \text{h.c.} \quad (4)$$

If the symmetry were exact, then the subsequent decays $\eta \rightarrow \psi$ mediated by N_i would completely erase any produced DM asymmetry. However the chirality flips induced by the mass term prevent this erasure, as we will explain in more detail in section 2. There is an unbroken discrete Z_2 remnant of this symmetry, where $\psi_i \rightarrow -\psi_i$ and $\eta \rightarrow -\eta$, that ensures the stability of the dark matter.

The potential V is assumed to give rise to the VEV of ϕ and it generically also includes the Higgs portal coupling $\frac{1}{2} \lambda_{h\phi} |H|^2 \phi^2$. Once ϕ gets its VEV, the nonabelian kinetic mixing operator can be written as

$$- \frac{1}{2} \sin \tilde{\epsilon} B_{\mu\nu}^3 Y^{\mu\nu} \quad (5)$$

where $\sin \tilde{\epsilon} = 2\sigma/\Lambda$. It could arise from integrating out a heavy vector-like fermion χ that carries hypercharge and transforms as a doublet under $SU(2)_h$. The interaction $y_\chi \bar{\chi} \phi_a \sigma_a \chi$ leads to the diagram in fig. 1, implying $\Lambda^{-1} \sim gg_1 y_\chi / m_\chi$ where g_1 is the hypercharge coupling. The kinetic mixing gives rise

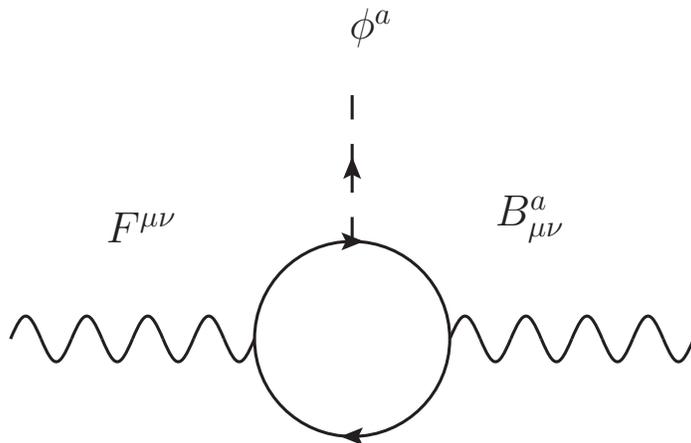


Figure 1: Loop contribution to the nonabelian kinetic mixing operator.

to electric millicharges $\pm\tilde{\epsilon}g \equiv \pm\epsilon e$ for the fermions $\Psi_{1,2}$. This or alternatively the Higgs portal interaction allows for direct detection of the dark atoms, as we discuss in section 3.

2 Origin of dark matter asymmetry

Our setup allows for heavy neutrinos to decay in a CP-violating manner into an excess of dark matter versus its antiparticles in close analogy to leptogenesis. The structure of Yukawa couplings is similar to that of neutrinos except that we have only two light fermionic DM species, $\Psi_{1,2}$ as compared to the three light neutrinos. The dark Higgs doublet η does not have a VEV, so it also gets an asymmetry, which will be determined by those in Ψ_i .

The asymmetry in the decay of the j th heavy neutrino into $\psi_i^* \eta$ versus $\psi_i \eta^*$ (recall that ψ_i denotes the Weyl doublet states) is given by

$$\begin{aligned} \epsilon_\psi^{ji} &= \frac{\Gamma(N_j \rightarrow \psi_i^* \eta) - \Gamma(N_j \rightarrow \psi_i \eta^*)}{\Gamma(N_i \rightarrow \text{any})} \\ &= \frac{1}{8\pi} \sum_{k \neq j} \left[\frac{\text{Im} \left[(y_\psi^\dagger y_\psi)_{kj} y_\psi^{ik} y_\psi^{ij*} \right]}{(y_\psi^\dagger y_\psi + y_\nu^\dagger y_\nu)_{jj}} g(M_k^2/M_j^2) \right. \\ &\quad \left. + \frac{\text{Im} \left[(y_\nu^\dagger y_\nu)_{kj} y_\psi^{ik} y_\psi^{ij*} \right]}{(y_\psi^\dagger y_\psi + y_\nu^\dagger y_\nu)_{jj}} g'(M_k^2/M_j^2) \right] \end{aligned} \quad (6)$$

where $g(x) = \sqrt{x} [1/(1-x) + 1 - (1+x) \ln(1+1/x)]$ and $g'(x) = \sqrt{x}/(1-x)$. This differs from the standard leptogenesis expression because the denominator must take into account decays of N_j both into neutrinos and dark matter, and there is a mixed term of order $y_\nu^2 y_\psi^2$ from the self-energy correction of N_j by the SM Yukawa interaction.

For definiteness, we will focus on decay of the lightest heavy neutrino N_1 . In the simplest scenario of leptogenesis, where $M_1 \ll M_{2,3}$ and the reheat temperature is in between, $M_1 < T_{rh} < M_{2,3}$, this is the only relevant decay since the heavier neutrinos are not present. In this case the functions in eq. (6) can be approximated as $g \cong -3/2\sqrt{x}$ and $g' \cong -1/\sqrt{x}$ with $x = (M_2/M_1)^2 \gg 1$.

Initially, we can expect independent asymmetries $Y_{1,2}$ for ψ_1 and ψ_2 , where $Y_i = (n_{\psi_i} - n_{\bar{\psi}_i})/s$ is the dark matter to entropy ratio, since $\epsilon_\psi^{11} \neq \epsilon_\psi^{12}$. However the Dirac mass term takes the form $\psi_1^T \sigma_2 \tau_2 \psi_2$, which implies that mass effects will cause the asymmetries of ψ_1 and ψ_2 to become equal and opposite. This projects the net asymmetry of the fermions onto the difference between the initial ones, $Y_\psi = Y_1 - Y_2$, at temperatures where the helicity-flipping interactions due to m_ψ come into equilibrium.

On the other hand, the η boson gets a different asymmetry, proportional to $\epsilon_\psi^{11} + \epsilon_\psi^{12}$. Eventually it will decay into ψ_i . For simplicity, we consider the

case $\epsilon_\psi^{11} \sim -\epsilon_\psi^{12}$. Then not only does the initial asymmetry in η tend to be small, but so also is its contribution to the final asymmetry in ψ_i , and we can estimate the net asymmetry in ψ from N_1 decays as

$$\epsilon_{\psi 1} \sim \epsilon_\psi^{11} - \epsilon_\psi^{12} \sim 2\epsilon_\psi^{11} \quad (7)$$

The sign difference is in contrast to the CP asymmetries for decays into neutrinos, $\epsilon_{\nu 1} = \sum_i \epsilon_\nu^{1i}$ familiar from leptogenesis.

2.1 Dark matter asymmetry estimate

The initial asymmetries depend upon an efficiency factor κ_ψ that quantifies the amount of washout (see for example [33] for a review). The contribution from N_1 decay is

$$Y_\psi = \frac{45}{\pi^4} \frac{\epsilon_{\psi 1} \kappa_\psi}{g_*} \quad (8)$$

where $\kappa_\psi \cong \min(0.25 (m_*/\tilde{m}_{\psi 1})^{1.1}, 1)$ with $\tilde{m}_{\psi 1} = 2(y_\psi^\dagger y_\psi)_{11} v^2 / M_1$, $m_* = 10^{-3}$ eV and $v = 174$ GeV. The Higgs VEV v has no direct physical relevance for the dark matter abundance, but $\tilde{m}_{\psi 1}/m_*$ gives $\Gamma(N_1 \rightarrow \psi \eta^{(*)})/H(M_1)$ (the ratio of the partial decay width to the Hubble rate), just like $\tilde{m}_{\nu 1}/m_*$ does for the decays into νh . The dark sphalerons associated to the $SU(2)_h$ gauge interactions have the same effect as (an increase in) the Dirac mass term for Ψ and therefore do not require additional consideration for the dark asymmetry.

If there is no hierarchical structure to the couplings y_ψ^{ij} and their phases are large, we can estimate $(y_\psi^\dagger y_\psi)_{kj} \sim (\tilde{y}_\psi^\dagger \tilde{y}_\psi)_{kj}$ or its imaginary part by some average value \bar{y}_ψ^2 . Further defining $\bar{y}_\nu^2 = (y_\nu^\dagger y_\nu)_{11}$ and assuming that the terms of order y_ν^2 in the numerator of (6) can be estimated as \bar{y}_ν^2 , we find that the CP asymmetry for ψ is of order

$$\epsilon_{\psi 1} \sim \frac{2 \bar{y}_\psi^2}{8\pi \sqrt{x}} \left(\frac{1 + \frac{3}{2}r}{1 + r} \right) \quad (9)$$

where we define $r = \bar{y}_\psi^2/\bar{y}_\nu^2$, and assume that $\epsilon_\psi^{12} \sim -\epsilon_\psi^{11}$ in (8). It is evident that eq. (9) has only mild dependence upon r . Combining with the efficiency factor κ_ψ (where we approximate the exponent 1.1 by 1) leads to the estimate²

$$Y_\psi \cong 1.4 \times 10^{-12} \left(\frac{M_1}{10^{10} \text{ GeV}} \right) \left(\frac{10}{x^{1/2}} \right) \quad (10)$$

ignoring r dependence.

2.2 Baryon asymmetry estimate

We wish to explain the baryon asymmetry simultaneously with that of dark matter. Analogously to (8), it is given by

$$Y_B = \frac{28}{79} \cdot \frac{45}{\pi^4} \frac{\epsilon_{\nu,1} \kappa_\nu}{g_*} \quad (11)$$

where the prefactor $28/79$ is due to redistribution of the initial lepton asymmetry into baryons via sphaleron interactions. The CP asymmetry $\epsilon_{\nu,1}$ is defined as $\epsilon_{\nu,1} = \sum_i \epsilon_{\nu,1i}$ in the usual way for leptogenesis. Similarly to our estimate in (9), we expect the well-known D-I bound [34] to be modified by a function of $r = \bar{y}_\psi^2/\bar{y}_\nu^2$,

$$\begin{aligned} |\epsilon_{\nu,1}| &\leq \frac{3}{16\pi} \frac{M_1}{v^2} \frac{\Delta m_{\text{atm}}^2}{m_{\nu_3}} \left(\frac{1 + \frac{2}{3}r}{1+r} \right) \\ &\cong 10^{-6} \left(\frac{M_1}{10^{10} \text{ GeV}} \right) \equiv 10^{-6} M_{10} \end{aligned} \quad (12)$$

where $\Delta m_{\text{atm}}^2 = m_{\nu_3}^2 - m_{\nu_2}^2$, which we assume to be $\cong m_{\nu_3}^2 \cong (0.05 \text{ eV})^2$. Again the dependence upon r is mild, and we will ignore the effect of the DM Yukawa coupling on leptogenesis in the visible sector. To estimate the

² This is valid for parameters such that $\kappa_\psi < 1$ hence $\tilde{m} \gtrsim 4m_*$. Using eq. (13) this implies $\bar{y}^2 \gtrsim 10^{-7}(M_1/10^{10} \text{ GeV})$. We will assume that this restriction holds in the following.

efficiency factor $\kappa_\nu \cong 0.25 (m_*/\tilde{m}_\nu)$, with $\tilde{m}_\nu = (y_\nu^\dagger y_\nu)_{11} v^2 / M_1$, we use the Casas-Ibarra parametrization of y_ν ,

$$\begin{aligned} (y_\nu^\dagger y_\nu)_{11} &\cong U_{1i} m_{\nu_i}^{1/2} R_{ik}^\dagger \frac{M_k}{v^2} R_{kj} m_{\nu_j}^{1/2} U_{j1}^\dagger \\ &\cong 10^{-6} M_{10} \end{aligned} \quad (13)$$

(the same result as eq. (12)) where U is the PMNS matrix and R is an arbitrary SU(3) transformation. We assumed that $R_{ik}^\dagger M_k R_{kj} \sim M_1$ since we take the heavy neutrino masses to be of the same order, and $|U_{12}|^2 m_{\nu_2} + |U_{13}|^2 m_{\nu_3} = 0.003$ eV (taking m_{ν_1} to be much less than the solar neutrino mass splitting). This gives $\kappa_\nu \cong 1/12$ and

$$Y_B \cong 1.4 \times 10^{-10} M_{10} \epsilon_{\text{DI}} \quad (14)$$

where we have introduced a parameter ϵ_{DI} to quantify how much $\epsilon_{\nu,1}$ falls below the D-I bound, *i.e.*, ϵ_{DI} is $|\epsilon_{\nu,1}|$ over its maximum value. Equating Y_B to its measured value, we find $\epsilon_{\text{DI}} M_{10} = 0.7$.

2.3 DM to baryon constraint

We can combine the above results to get a constraint on the model parameters from the known ratio of baryon and dark matter energy densities, $\Omega_B/\Omega_{DM} = m_p Y_B / (m_{\mathbf{H}} Y_\psi) = 0.18$. Here $m_{\mathbf{H}} = m_1 + m_2$, the mass of the dark atom (neglecting its binding energy). Then we find $m_{\mathbf{H}}/m_p = 166 \epsilon_{\text{DI}} (x/10)^{1/2}$. We can eliminate ϵ_{DI} using eq. (14) and M_1 using (13) to obtain

$$\frac{m_{\mathbf{H}}}{m_p} = 560 \epsilon_{\text{DI}} \left(\frac{x^{1/2}}{10} \right) = \frac{360}{M_{10}} \left(\frac{x^{1/2}}{10} \right) \quad (15)$$

Eq. (15) reveals part of our motivation for the choice $M_1 \sim 10^{10}$ GeV: it gives dark atom masses in a range that is interesting for direct detection and

consistent with our prejudice for the new physics scale to not be far below the weak scale. It is interesting that the same mass scale is also consistent with the observed baryon asymmetry for generic choices of the neutrino CP asymmetry, $\epsilon_{DI} \lesssim 1$.

Notably absent from our estimates is any explicit dependence upon the Yukawa couplings \bar{y}_ψ^2 and \bar{y}_ν^2 . This is because of the cancellation between the CP asymmetry ϵ and the efficiency factor κ , which only occurs for couplings such that $\kappa < 1$. We verified this condition for $\kappa_\nu = 0.08$. It is also satisfied by κ_ψ so long as $\bar{y}_\psi^2 \gtrsim \bar{y}_\nu^2/12$. We will make this technical assumption to keep the analysis simple. For smaller values of \bar{y}_ψ^2 , there would be a suppression of Y_ψ and the need for correspondingly larger values of $m_{\mathbf{H}}$.

As an example, we take $\epsilon_{DI} = 0.65$, $\bar{y}_\psi^2 = \bar{y}_\nu^2 = 10^{-6}$, $m_{\mathbf{H}} \cong 83 \text{ GeV}$, $M_1 = 10^{10} \text{ GeV}$, $M_2 = \sqrt{x}M_1 = 2 \times 10^{11} \text{ GeV}$. Larger or smaller values of $m_{\mathbf{H}}$ can be obtained by adjusting M_2/M_1^2 , using eq. (15).

2.4 Decay of dark scalars

An interesting feature of our model is that the seesaw mechanism produces the new dimension-5 operator (3) that allows the dark scalars η to decay [29]. When the SM Higgs takes its vacuum expectation value, this allows the η to decay directly into $\nu\psi$. The decay rate is of order

$$\Gamma \sim \frac{\bar{y}_\psi^2 \bar{y}_\nu^2 m_\eta v^2}{8\pi M_1^2} \quad (16)$$

$$\sim 3 \times 10^{-3} \text{ s}^{-1} \cdot \left(\frac{m_\eta}{150 \text{ GeV}} \right) \quad (17)$$

where for the numerical estimate of Γ we used the exemplary values specified at the end of the previous section (ignoring the mass of ψ in the phase space integral).

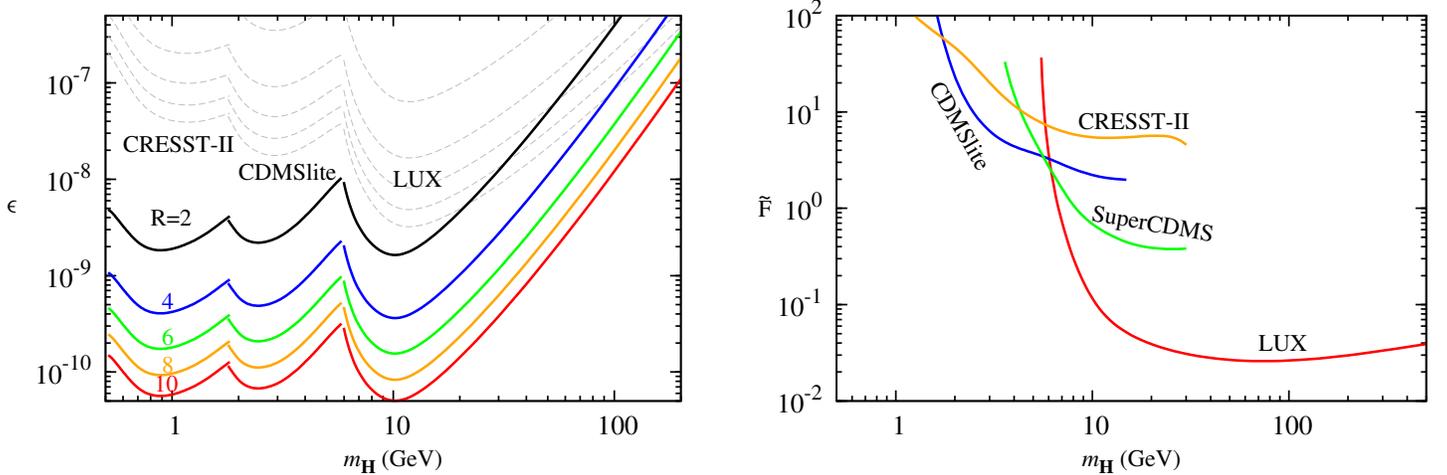


Figure 2: Left: CRESST-II, CDMSlite and LUX limits on millicharge ϵ of dark atom constituents, with constituent mass ratios $m_2/m_1 = R = 2, 4, \dots, 10$ as indicated, for photon-mediated scattering of dark atoms on protons. For clarity, only the most constraining limit is shown for any DM atom mass $m_{\mathbf{H}}$. Gauge coupling is set to $\alpha_g = \alpha_{\text{ion}}$, eq. (19) for solid curves, and fixed at $\alpha_g = 0.06$ for light dashed curves. Right: Corresponding limits on $\tilde{F} = y_1 \theta |1 - m_h^2/m_\phi^2| F_\psi(0)$ from Higgs portal scattering, where θ is the ϕ -Higgs mixing angle.

Such decays must occur sufficiently early so that the decay products are fully thermalized before they can distort the CMB. Ref. [35] shows that this occurs if the lifetime is below $\sim 10^{12}$ s. Eq. (16) implies that our model easily satisfies this bound.

3 Direct Detection

There are two portals through which our dark atoms to interact with nuclei. The kinetic mixing allows for photon exchange, which has been discussed in refs. [21, 22]. The ensuing constraints on the electric millicharge ϵ are weakened for atoms compared to ions because of the screening of electric charge. In the special case where $m_1 = m_2$ this screening is perfect, and the

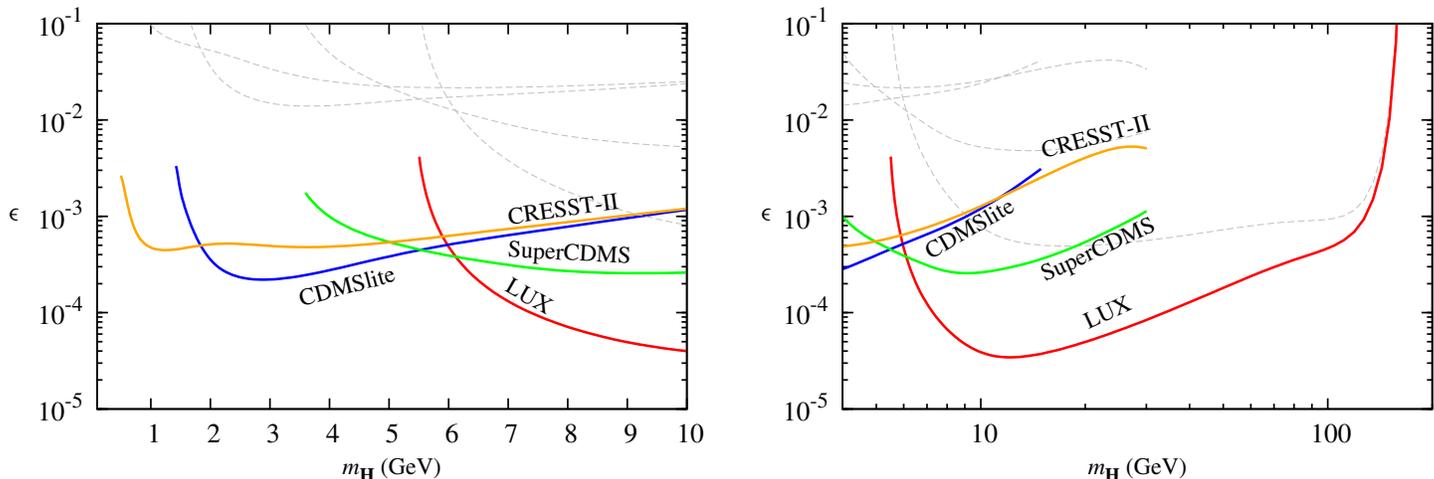


Figure 3: Direct detection constraints on kinetic mixing parameter ϵ versus dark atom mass $m_{\mathbf{H}}$ for case of equal-mass constituents $m_1 = m_2 = m_{\mathbf{H}}/2$, when interaction is magnetic inelastic. Hyperfine mass splitting is chosen as a function of $m_{\mathbf{H}}$ as described in text. Solid and dashed curves refer to choice of α_g as in fig. 2. Left: low mass region; right: larger mass region.

interaction becomes magnetic dipole, further weakening the limits [17]. Here we extend results of ref. [17] for the $m_1 = m_2$ case to higher DM masses.

In addition, there is the Higgs portal induced by mixing of h and ϕ_3 through the operator $\frac{1}{2}\lambda_{h\phi}|H|^2|\phi|^2$. ϕ_3 interacts with the dark atom constituents through the operator $\bar{\Psi}(y_1 + iy_2\gamma_5)(\vec{\phi}\cdot\vec{\tau})\Psi$ that splits the $\Psi_{1,2}$ masses.

3.1 Kinetic mixing portal

Unequal-mass constituents

As discussed in ref. [21], the mutual screening of the electric charges of the Ψ_1 and Ψ_2 constituents results in a scattering matrix element where the $1/q^2$ of the photon propagator is canceled by q^2 in the form factor for the charge

density. The cross section for scattering of dark atoms on a proton is

$$\begin{aligned}\sigma_p &= 4\pi \frac{\alpha^2 \epsilon^2 \mu_n^2}{\alpha_g^4} \left(\frac{1}{m_1^2} - \frac{1}{m_2^2} \right)^2 \\ &= 4\pi \frac{\alpha^2 \epsilon^2 \mu_n^2}{\alpha_g^4 m_{\mathbf{H}}^4} f_0(R)\end{aligned}\tag{18}$$

where μ_n is the reduced mass of the dark atom and nucleon system, and $f_0(R) = (1 + 1/R)^4 (R^2 - 1)^2$. Here we have generalized the result of ref. [21] where the approximation of large R was made. The expression (18) is valid if R is not too close to 1. The question of “how close?” is discussed below.

For $R \neq 1$, the resulting upper limits on ϵ is illustrated in fig. 2(a) showing the most constraining limit from the LUX [36], CRESST-II [37] or CDMSlite [38] experiments, at any given dark atom mass $m_{\mathbf{H}}$. In section 5 we will see that the requirement of sufficiently small ionization fraction in the dark sector leads to the constraint

$$\alpha_g \geq \alpha_{\text{ion}} \equiv 5 \times 10^{-3} \left(\frac{m_{\mathbf{H}}}{\text{GeV}} \right)^{1/2} f_2^{-1/4}(R)\tag{19}$$

where

$$f_2(R) = R + 2 + R^{-1}\tag{20}$$

The solid curves are derived for the parameter choice which saturates this bound, $\alpha_g = \alpha_{\text{ion}}$, and R ranging from 2 to 10, while the dashed ones assume a fixed value of $\alpha_g = 0.06$. This value satisfies the constraint $\alpha_g > \alpha_{\text{ion}}$ over the entire range of R and $m_{\mathbf{H}}$ shown on the plots. (The unusual sensitivity of the solid curves to light DM masses is due to the decrease of $\alpha_g = \alpha_{\text{ion}}$ with $m_{\mathbf{H}}$, and consequent increase in the dark Bohr radius, leading to larger cross sections.) The nominal constraints from the experiments are weakened by factors of $(A/Z)^2 = 5.9, 5.2$ and 4 respectively to account for the coupling to protons only. For CRESST this corresponds to collisions with the oxygen

atoms that dominate the sensitivity to low-mass dark matter. The strongest constraints occur for $m_{\mathbf{H}} \cong 1 - 10 \text{ GeV}$, in the range $\epsilon \lesssim 10^{-10} - 10^{-8}$. For conventional abelian kinetic mixing, such small values of ϵ could be difficult to achieve since the loop diagram that generates it is not suppressed by any large mass scales, since in this case the kinetic mixing operator is marginal. However for nonabelian kinetic mixing, ϵ is suppressed by the mass m_χ of the heavy particle in the loop, as well as its Yukawa coupling y_χ . For example if the couplings described below eq. (5) are $y_\chi = 0.1$, $g_1 = g$, $R = 10$, $\alpha_g = \alpha_{\text{ion}}$ and $\sigma = 30 \text{ GeV}$, we require $m_\chi \gtrsim 3 \times 10^{11} \text{ GeV}$ to satisfy the LUX bound on 10 GeV dark atoms.

Equal-mass constituents

For $R \cong 1$, there is perfect screening of charge because of the complete overlap of the wave functions of the two constituents, and the magnetic dipole interaction that we have neglected in (18) becomes important. This case was considered in detail in ref. [21]. The magnetic scattering is inelastic because of the hyperfine transition of the dark atom, requiring energy $\delta E = \frac{1}{6}\alpha_g^4 m_{\mathbf{H}}$, hence a minimum DM velocity of $v_{\text{min}} = q/(2\mu_N) + \delta E/q > \sqrt{2\delta E/\mu_N}$ for momentum transfer q and dark atom-nucleus reduced mass μ_N . There is a q - and v -dependent form factor $F = (q_0/q)^2(v^2 - v_{\text{min}}^2)/v_0^2$ that is of order unity for typical values $v \sim v_0$ and $q \sim q_0$, as long as $v_0 \gtrsim v_{\text{min}}$. The cross section on protons is of order

$$\sigma_{p,0} \equiv \frac{64\pi\epsilon^2\alpha^2\mu_n^2v_0^2}{m_{\mathbf{H}}^2q_0^2} \quad (21)$$

in that case, where μ_n is the proton-atom reduced mass.

More quantitatively, the actual cross section for a given scattering event is $\sigma_p = \sigma_{p,0} F(q, v)$ and the detection rate is proportional to

$$R \propto Z^2 \int_{E_{\min}}^{E_{\max}} dE_R \int_{v_{\min}}^{v_{\text{esc}}} \frac{d^3 \vec{v}}{v} f(\vec{v}) \sigma_p \quad (22)$$

$$\propto Z^2 \sigma_{p,0} I_F, \quad (23)$$

with

$$\begin{aligned} E_{\min} &= \frac{1}{2} m_{\mathbf{H}} v_{\min}^2 \\ E_{\max} &= p_{\max}^2 / (2m_N) \\ p_{\max} &= \sqrt{\mu_n^2 (v_{\text{esc}} + v_0)^2 - 2\delta E \mu_N + \mu_N (v_{\text{esc}} + v_0)} \\ f(\vec{v}) &\propto e^{-(\vec{v} + \vec{v}_e)^2 / v_0^2} - e^{-v_{\text{esc}}^2 / v_0^2}, \end{aligned} \quad (24)$$

$$I_F \equiv \int_{E_{\min}}^{E_{\max}} dE_R \int_{v_{\min}}^{v_{\text{esc}}} \frac{d^3 \vec{v}}{v} f(\vec{v}) F(q, v). \quad (25)$$

Here \vec{v}_e is the Earth's speed relative to the DM halo, $v_0 \approx 220 \text{ km s}^{-1}$ is the mean DM velocity, $v_{\text{esc}} \approx 450 \text{ km s}^{-1}$ is the approximate escape velocity of the DM halo (we see no significant variation in the results for values in the range $400 - 500 \text{ km s}^{-1}$), and E_{esc} is the maximum recoil energy from a DM particle with the escape velocity.

We compare the rate for our model to that of generic DM scattering with a constant cross section σ_n , for which the corresponding expressions are

$$R \propto A^2 \sigma_n I_0, \quad (26)$$

$$I_0 \equiv \int_{E_{\min}}^{E_{\max}^{(0)}} dE_R \int_{v_{\min}}^{v_{\text{esc}}} \frac{d^3 \vec{v}}{v} f(\vec{v}). \quad (27)$$

where $E_{\max}^{(0)} = E_{\max}$ evaluated at $\delta E = 0$. Therefore the magnetic inelastic cross section (21) is bounded from above as

$$\sigma_{p,0} < \frac{A^2 I_0}{Z^2 I_F} \sigma_{n,\text{lim}} \quad (28)$$

where $\sigma_{n,\text{lim}}$ is the experimental upper limit on the cross section for a generic DM model. Notice that the arbitrary quantity $(v_0/q_0)^2$ appears in the same way on both sides of (28) and hence can be divided out.

Although the gauge coupling α_g does not appear in (21), the mass splitting $\delta E = \alpha_g^4 m_{\mathbf{H}}/6$ depends upon it. For definiteness, we have chosen the value $\alpha_g = \alpha_{\text{ion}}$ in eq. (19) from the requirement of sufficiently small dark ionization fraction. This fixes δE as a function of $m_{\mathbf{H}}$.

We plot the ensuing limits on ϵ in figure 3, using results from the LUX [36], SuperCDMS [39], CRESST-II [37], and CDMSlite [38] experiments. The mass dependence of $\alpha_g = \alpha_{\text{ion}}$ changes the shape of the exclusion curves relative to those on the cross section itself, since the mass splitting δE rises rapidly with $m_{\mathbf{H}}$, nullifying the signal for $m_{\mathbf{H}} \gtrsim 100 \text{ GeV}$.

Transition from $R = 1$ to $R > 1$

We have noted that inelastic magnetic transitions dominate for equal constituent masses, $R = 1$, while elastic charge-charge interactions dominate when $R > 1$. One may wonder how sharp the transition is between the two regimes; how small must $R - 1$ be for inelastic transitions to dominate? We have calculated the ratio of the two cross sections as a function of R for a particular value of $m_{\mathbf{H}} = 10 \text{ GeV}$ as an example, taking the mass splitting δE as described above. The result is graphed in fig. 4, which shows that only for $R < 1.0001$ do the inelastic transitions dominate. Hence the most natural situation corresponding to this case is that where $R = 1$ exactly. There are two limits in our model that give $R = 1$: either $y_1 = 0$, or $m_{\psi} = 0$. The latter is a point of enhanced SU(2) flavor symmetry for the two chiral (doublet) fermions.

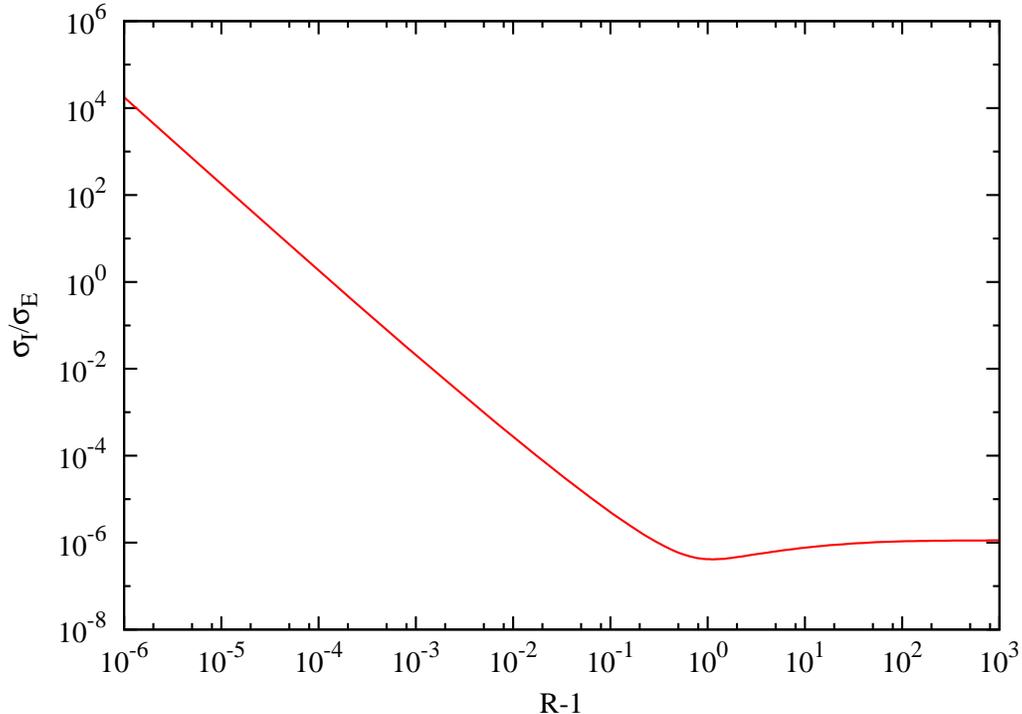


Figure 4: Ratio of the magnetic inelastic and elastic cross section for scattering of dark atoms on protons as a function of the constituent mass ratio R (its deviation from unity), for dark atom mass $m_{\mathbf{H}} = 10$ GeV and mass splitting described in section 3.1.

3.2 Higgs portal

The interaction of dark atoms with the Higgs through ϕ_3 - H mixing also undergoes screening because of the coupling τ_3 which has opposite sign for Ψ_1 and Ψ_2 . At low energies, the dark atoms can be described by a Dirac field \mathbf{H} whose coupling to the virtual ϕ_3 or h carrying momentum q is given by the amplitude

$$y_1 \bar{u}_{\mathbf{H}} u_{\mathbf{H}} F_{\psi}(q) \quad (29)$$

We have neglected the y_2 contribution that is suppressed by the dark matter velocity. By matching onto the scattering amplitudes in the high-energy

theory, we infer that

$$F_\psi(q) = \frac{1}{m_{\mathbf{H}}} \left[\frac{m_2}{\left(1 + \frac{1}{4}q^2 a_2^2\right)^2} - \frac{m_1}{\left(1 + \frac{1}{4}q^2 a_1^2\right)^2} \right] \quad (30)$$

with $a_i = (\alpha_g m_i)^{-1}$. Thus the coupling vanishes in the limit $R = 1$ ($m_1 = m_2$). If θ is the h - ϕ_3 mixing angle, then the amplitude for scattering of dark atoms on nucleons is

$$\begin{aligned} \mathcal{M} &= y_1 \bar{u}_{\mathbf{H}}(p_3) u_{\mathbf{H}}(p_1) \cdot \bar{u}_n(p_4) u_n(p_2) \left(\frac{y_n m_n}{v} \right) \\ &\times c_\theta s_\theta \left(\frac{1}{m_h^2} - \frac{1}{m_\phi^2} \right) F_\psi(q) \end{aligned} \quad (31)$$

where $(y_n m_n/v)$ with $y_n \cong 0.3$ [40] is the coupling of the Higgs to nucleons.

If α_g is not too small, we can take the $q = 0$ limit of the form factor. In this case the cross section for dark atom-nucleon scattering is

$$\sigma_n \cong \frac{1}{\pi v^2} [y_1 y_n m_n \mu_{n\mathbf{H}} \theta F_\psi(0)]^2 (m_\phi^{-2} - m_h^{-2})^2 \quad (32)$$

in terms of the \mathbf{H} -nucleon reduced mass, and taking $\theta \ll 1$. The LUX upper limit on the dimensionless combination $\tilde{F} = y_1 \theta |1 - m_h^2/m_\phi^2| F_\psi(0)$ is plotted in fig. 2. The Yukawa coupling y_1 is related to the mass splitting in the dark sector since $m_2^2 - m_1^2 = y_1^2 \sigma^2$. Moreover it is straightforward to show that $F_\psi(0) = (m_2^2 - m_1^2)/m_{\mathbf{H}}^2$. If $m_\phi < m_h$ then $\tilde{F} \cong y_1^3 \theta (m_h^2 \sigma^2)/(m_\phi^2 m_{\mathbf{H}}^2)$. We expect $m_\phi \sim \sigma$, similarly to $m_h \sim v$ in the visible sector, and $\theta \lesssim 0.01$ to satisfy LEP constraints [41] on mixing of a light scalar with the Higgs. The largest dark atom mass range for saturating the LUX bound shown in fig. 2 with $|y_1| \lesssim 1$ is $m_{\mathbf{H}} \lesssim 70$ GeV.

4 Stable Vector Bosons

Up to now we have assumed that the Ψ_1 - Ψ_2 mass splitting is sufficiently small to prohibit the decay $\Psi_2^- \rightarrow B^{--}\Psi_1^+$, corresponding to the condition

$$|y_1| < \frac{m_1 + m_2}{4m_\psi} g \quad (33)$$

However this need not be the case, and the model is also compatible with a universe where charge neutrality in the dark sector is achieved by having two Ψ_1^+ particles for every B^{--} . This leads to a very different kind of dark atom that is reminiscent of the H_2 molecule, except that the two “protons” are bound together by a single charge -2 “electron”. We will refer to these variant dark atoms as \mathbf{H}_2 . In the absence of fine-tuning, the stable vector boson is typically lighter than Ψ_1 , prompting us to define the ratio

$$R_2 = \frac{m_1}{m_B} \geq 1 \quad (34)$$

in analogy to $R = m_2/m_1$ for \mathbf{H} atoms.³

4.1 Bound States

To verify the existence of the 3-body \mathbf{H}_2 bound states, we make some ansätze for its wave function and use the variational method to prove that the energy is minimized at a negative value. We consider trial wave functions where the positions of the three particles are given by

$$\vec{x}_\psi = \pm \vec{\Delta}/2, \quad \vec{x}_B = \vec{r} \quad (35)$$

³ To get the opposite situation where $m_B > m_1$, we need $(g\sigma)^2 > (m_\psi - y_1\sigma)^2 + (y_2\sigma)^2$. This requires not only y_2 to be small, but also an accidental cancellation between m_ψ and $y_1\sigma$.

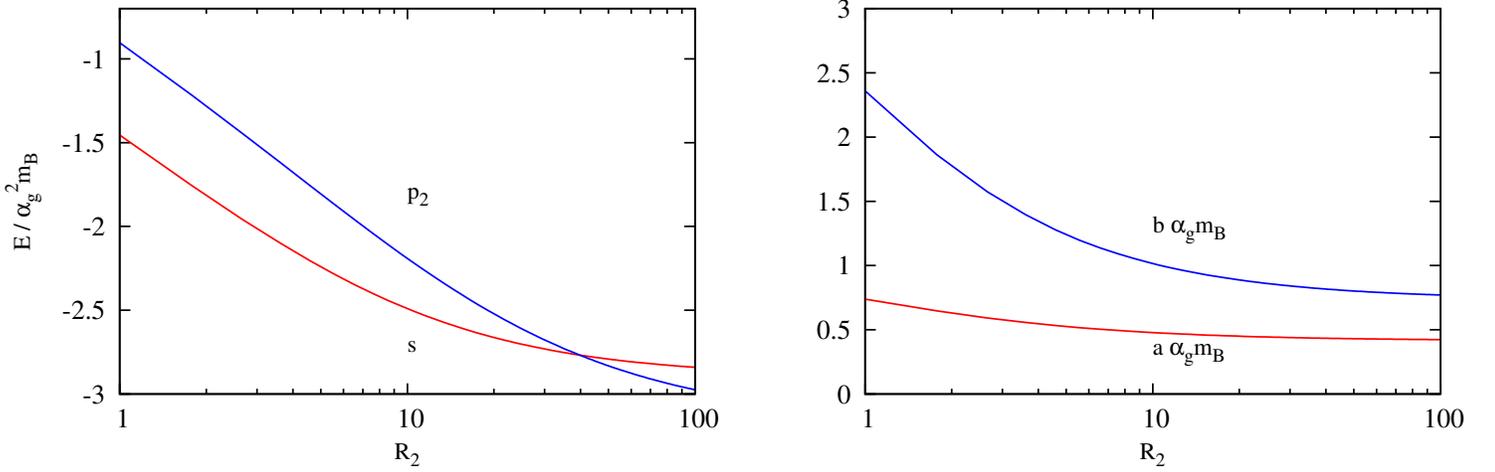


Figure 5: Left: energy obtained from variational method as a function of R_2 for the trial wavefunctions for \mathbf{H}_2 bound states $\psi_{\mathbf{H}_2,s}$ and $\psi_{\mathbf{H}_2,p2}$. Right: Corresponding values of the parameters a, b (shown in dimensionless combinations with $\alpha_g m_B$) that determine the spatial distributions of the wave functions, for the s -wave.

i.e., we work in the center-of-mass frame of the two Ψ_1 particles, with $\vec{\Delta}$ being their relative separation. In analogy to the H_2 molecule, it could be expected that the wavefunction for Δ is approximately that of a 3D harmonic oscillator, $e^{-\Delta^2/b^2}$ for some scale b . For simplicity we take the wave function for r to be hydrogen-like, $e^{-r/a}$ for some other scale a . We consider three possible states, an s -wave and two p -waves,

$$\begin{aligned}
 \psi_{\mathbf{H}_2,s}(\vec{r}, \vec{\Delta}) &= N_s e^{-\Delta^2/b^2 - r/a} \\
 \psi_{\mathbf{H}_2,p1}(\vec{r}, \vec{\Delta}) &= N_{p1} r_z e^{-\Delta^2/b^2 - r/a} \\
 \psi_{\mathbf{H}_2,p2}(\vec{r}, \vec{\Delta}) &= N_{p2} \Delta_z e^{-\Delta^2/b^2 - r/a}
 \end{aligned} \tag{36}$$

where r_z (Δ_z) is the z -component of \vec{r} ($\vec{\Delta}$).

It is convenient to work in the analog of atomic units by rescaling to dimensionless coordinates $r = r'/(\alpha_g m_B)$, $\Delta = \Delta'/(\alpha_g m_B)$. Then the Hamiltonian can be written as $H = (\alpha_g^2 m_B) H'$, where the dimensionless H' is

$$H' = -\frac{1}{R_2} \nabla_{\Delta'}^2 - \frac{1}{2} \nabla_{r'}^2 + \frac{1}{\Delta'} - \sum_{\pm} \frac{2}{|\vec{r}' \pm \vec{\Delta}'/2|}. \quad (37)$$

By minimizing the expectation values $E = \langle \psi_{\mathbf{H}_2} | H | \psi_{\mathbf{H}_2} \rangle$ with respect to a, b and varying over a range of R_2 values, we find that bound states (having $E < 0$) exist for all three trial wave functions, but $\Psi_{\mathbf{H}_2, p1}$ is always more weakly bound than the other two. Moreover the s -wave has lower energy than $p2$ only for $R_2 \lesssim 40$; for $R_2 > 40$ the $p2$ state is lower, as shown in fig. 5. Taking as an example the values $m_1 = 60 \text{ GeV}$, $R_2 = 10$, $\alpha_g = 3 \times 10^{-2}$, the three-constituent atoms have binding energies of approximately $E \approx -15 \text{ MeV}$.

4.2 Direct Detection

Dark \mathbf{H}_2 atoms interact similarly with nucleons relative to our treatment for \mathbf{H} atoms in section 3, but there are some qualitative differences, due to the more complicated wave function. In particular, there is no longer any special case like $R = 1$ for \mathbf{H} atoms in which the electric millicharge clouds of the constituents give exactly canceling contributions to the total charge density. This can be seen by computing the form factor, which is the Fourier transform of the charge density

$$\begin{aligned} \rho(x) &= \int d^3 \Delta d^3 r |\Psi(\vec{r}, \vec{\Delta})|^2 \\ &\times \left(\sum_{\pm} \delta(\vec{x} \pm \vec{\Delta}/2) - 2\delta(\vec{x} - \vec{r}) \right) \end{aligned} \quad (38)$$

Using $\psi_{\mathbf{H}_2, s}$ from eq. (36), the form factor is

$$\begin{aligned} F(q) &= 2 \left(-e^{-b^2 q^2/32} + \frac{1}{(1 + a^2 q^2/4)^2} \right) \\ &\cong q^2 \left(\frac{b^2}{16} - a^2 \right), \end{aligned} \quad (39)$$

where the approximation is for low momentum transfer q .

In computing the cross section for scattering on protons, the factor of q^2 in the form factor cancels the $1/q^2$ of the propagator like before, giving

$$\sigma_p = 16\pi\alpha^2\epsilon^2\mu_n^2 \left(a^2 - \frac{b^2}{16} \right)^2. \quad (40)$$

at low momentum transfer. (The normalization can be deduced by considering the limits $a = 0, b \rightarrow \infty$ or *vice versa* where the usual Feynman rules for the amplitude with no form factor apply.) The direct detection limits from LUX [36], CRESST-II [37] and CDMSlite [38] through the kinetic mixing portal are shown for various values of R_2 in figure 6, assuming α_g saturates the constraint (51) from ionization of \mathbf{H}_2 atoms that we will derive in the next section. (We also show the constraints for the fixed value of $\alpha_g = 0.01$ as dashed curves.) Unlike with the dark atoms, the form factor never vanishes for any value of R_2 (since b is always $< 4a$).

For the Higgs portal, we follow the procedure in section 3. The amplitude and cross section are

$$\begin{aligned} \mathcal{M} &= \bar{u}_{\mathbf{H}_2}(p_3)u_{\mathbf{H}_2}(p_1) \cdot \bar{u}_n(p_4)u_n(p_2) \left(\frac{y_n m_n}{v} \right) \\ &\quad \times \left(\frac{c_\theta s_\theta}{m_h^2} - \frac{c_\theta s_\theta}{m_\phi^2} \right) (y_1 F_\psi(q) + g F_B(q)) \end{aligned} \quad (41)$$

$$\begin{aligned} \sigma_n &\cong \frac{1}{\pi v^2} [(y_1 F_\psi(0) + g F_B(0)) y_n m_n \mu_n \theta]^2 \\ &\quad \times (m_\phi^{-2} - m_h^{-2})^2. \end{aligned} \quad (42)$$

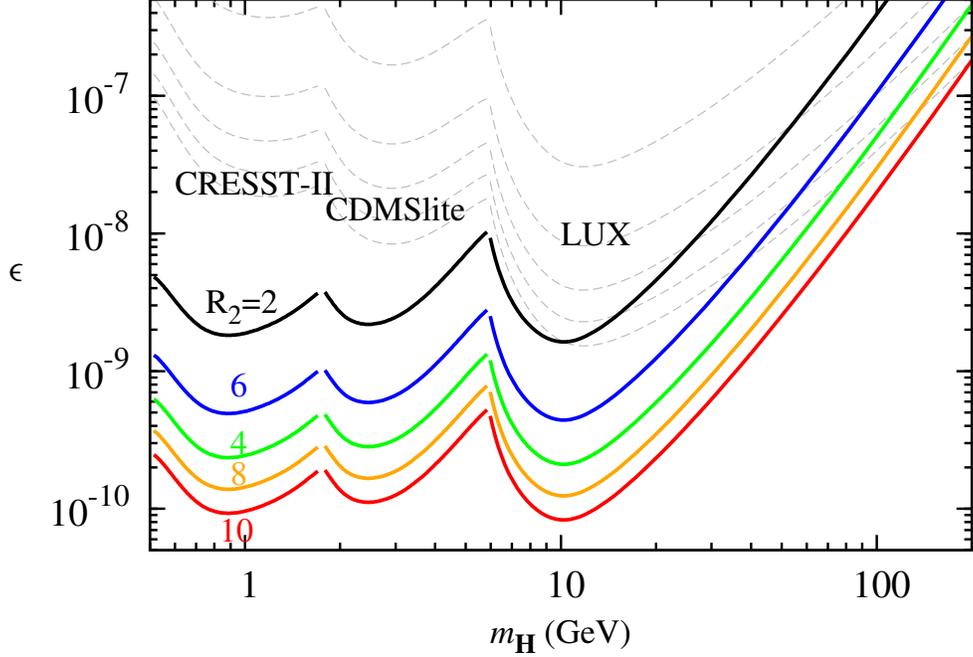


Figure 6: Direct detection constraints on kinetic mixing parameter as in fig. 2, but for \mathbf{H}_2 atoms with $m_1/m_B \equiv R_2 = 2, 4, \dots, 10$.

We have again made the approximation $\theta \ll 1$ and assumed a small momentum transfer. μ_n is the \mathbf{H}_2 -nucleon reduced mass, and $y_n \cong 0.3$ is the Higgs coupling to nucleons (modulo m_n/v). The form factors are given by

$$\begin{aligned}
 F_\psi(q) &= \frac{2m_\Psi}{m_{\mathbf{H}_2}} e^{-b^2 q^2/32} \\
 F_B(q) &= \frac{m_B}{2m_{\mathbf{H}_2}} \frac{1}{(1 + \frac{1}{4}a^2 q^2)^2}.
 \end{aligned} \tag{43}$$

Redefining $\tilde{F} = (y_1 F_\psi(0) + g F_B(0))\theta|1 - m_h^2/m_\phi^2|$, the constraint on \tilde{F} from the LUX, CRESST-II and CDMSlite experiments takes the same form as was previously shown shown in figure 2 (right), where $m_{\mathbf{H}}$ is reinterpreted as $m_{\mathbf{H}_2}$.

4.3 Neutron Star Constraints

Tight constraints exist on the cross section for asymmetric bosonic dark matter scattering on nucleons from the existence of long-lived neutron stars [42, 43]. If the rate of dark matter accretion is large enough, it can collapse to form a black hole that would consume the progenitor, on time scales shorter than the ages of neutron stars observed in globular clusters. In our model it is important that we have only one kind of stable bosonic dark matter constituent carrying dark $U(1)_h$ charge. In the case of \mathbf{H} atoms with only fermionic constituents, the would-be scalar constituents decayed early in the cosmological history, leaving no asymmetric scalars. For \mathbf{H}_2 atoms, on the other hand, the vector bosons are mostly bound inside of atoms that resist collapse because of the degeneracy pressure of their fermionic constituents. The ionized fraction also resists collapse because of dark Coulomb repulsion. In contrast, in a model containing two species of bosons carrying different $U(1)_h$ charges, nothing would prevent the collapse of the combined bosonic fluid.

In more detail, we first note that the dark atoms remain bound once they start to accumulate in the neutron star. From figure 5, the binding energy is given by

$$E_b \approx 2 \alpha_g^2 m_B = \frac{2 \alpha_g^2 m_{\mathbf{H}_2}}{1 + 2R_2} \quad (44)$$

Using the dark ionization constraint (51), we find that $E_b > 130 \text{ eV}$ even for the extreme parameter choices $m_{\mathbf{H}_2} = 1 \text{ GeV}$, $R_2 = 100$, which is higher than the temperature of the star, of order 100 eV [44]. Moreover fermions within a neutron star are supported by their degeneracy pressure, given by

$$p = \frac{(3\pi^2)^{2/3}}{5 m_\psi} n_\psi^{5/3}, \quad (45)$$

where n is the number density. A larger fermion mass decreases the pressure, and therefore the dark atoms will tend to remain bound.

As for any ionized bosons that accumulate within the neutron star, their repulsive self-interaction greatly weakens the bounds on scattering with nucleons by preventing their collapse into a black hole. Ref. [45] finds that a repulsive scattering cross section exceeding 10^{-50} cm^2 is sufficient to avoid neutron star constraints for $m_B < 1 \text{ TeV}$. In our case the cross section corresponding to dark Rutherford scattering is infrared divergent, but if we make it finite by multiplying $d\sigma/d\Omega$ by $(1 - \cos\theta)^2$ (thus taking into account only scatterings with significant momentum transfer), it is of order $\alpha_g^2/m_B^2 \gtrsim 10^{-34} \text{ cm}^2$, where we used (51) and $m_B \lesssim 100 \text{ GeV}$. This satisfies the requirements of [45] by many orders of magnitude.

5 Other constraints

Dark atoms, dark matter with millicharges, and models with asymmetric dark bosons are subject to further constraints from cosmological, astrophysical and laboratory probes. Here we discuss those coming from dark recombination, self-interactions of the dark matter and accumulation in neutron stars, and searches for millicharged particles.

5.1 Dark ions

If the constituents of the hidden sector fail to combine into atoms, they can scatter very strongly with each other through the dark Coulomb interaction, contradicting the normally assumed properties of collisionless cold dark matter. From fitting to results of ref. [4], one finds that the ionization fraction

can be estimated as [14, 21, 23]

$$X_e \cong \left(1 + 10^{10} f_2(R) \xi^{-1} \alpha_g^4 \frac{\text{GeV}^2}{m_{\mathbf{H}}^2} \right)^{-1} \quad (46)$$

where $f_2(R) = R + 2 + 1/R$ (introduced in eq. (20)), and ξ is the ratio of dark sector to SM sector temperatures.

In [4] it was argued that observations of the Bullet Cluster rule out $X_e \gtrsim 0.1$, leading to the conservative lower limit $\alpha_g > \alpha_{\text{ion}}$ (19) that we already incorporated in our analysis of direct detection constraints. Ref. [14] estimates that there is a factor of 10 uncertainty in (46). We note that this leads to only a factor of 1.8 uncertainty in the expression for α_{ion} .

The ratio between temperatures can be found using the relation [14]

$$\xi = \left(\frac{g_{*S,SM}^0 g_{*S,D}^{\text{dec}}}{g_{*S,SM}^{\text{dec}} g_{*S,D}^0} \right)^{1/3}, \quad (47)$$

with $g_{*S,SM}$ and $g_{*S,D}$ denoting the number of degrees of freedom in the visible and dark sectors, and the superscripts 0, dec indicating the respective values today and at the time the two sectors decouple kinetically. The temperature at which this decoupling occurs is therefore relevant. We find that mixed Compton scattering with one dark and one SM photons is the most important process for maintaining kinetic equilibrium. It goes out of equilibrium when $H = n_\gamma \langle \sigma v \rangle$, leading to the estimate

$$1.66 g_* \frac{T^2}{m_{\text{Pl}}} \sim g_* T^3 \frac{8\pi}{3} \frac{\epsilon^2 \alpha^2}{m_{\mathbf{H}}^2}, \quad (48)$$

Thus mixed Compton scattering keeps the two sectors at the same temperature until

$$T_{\text{dec}} = \frac{3 \times 10^{-6} \text{ eV}}{\epsilon^2} \left(\frac{m_{\mathbf{H}}}{\text{GeV}} \right) \quad (49)$$

The lowest value of T_{dec} is obtained by saturating the direct detection limits on ϵ as a function of $m_{\mathbf{H}}$, as shown in figs. 2-3. In the case of $R = 1$ (equal mass dark atom constituents), this can be much lower than the dark recombination temperature T_{rec} , so that in fact $T_{\text{dec}} = T_{\text{rec}}$, since Compton scattering is no longer efficient on neutral atoms. For $R > 1$ on the other hand, the constraints on ϵ are sufficiently strong that the decoupling temperature is limited to $T_{\text{dec}} > 300 \text{ TeV}$.

As long as $T_{\text{dec}} \gg 1 \text{ TeV}$, all particles are relativistic except for the heavy neutrinos. We therefore use the values $g_{*S,SM}^0 = 3.94$ [15], $g_{*S,SM}^{\text{Dec}} = 106.75$, $g_{*S,D}^0 = 2$, and $g_{*S,D}^{\text{Dec}} = 18$. The resulting temperature ratio is $\xi \approx 0.71$. At the other extreme, decoupling occurs after electrons have frozen out. This corresponds to $g_{*S,SM}^{\text{Dec}} = 7.25$, $g_{*S,DM}^{\text{Dec}} = 2$, and $\xi \approx 0.81$. Even at the two extremes, therefore, the difference is minimal, and is further mitigated by the fact that ξ is raised to the $1/4$ power in calculating α_{ion} . We therefore adopt the value $\xi = 0.71$ in eq. (19) so that α_{ion} remains a reasonable lower limit for α_g .

There are certain cases that can lead to a lower temperature ratio, with the smallest being that in which all dark content apart from the dark photon has frozen out prior to the freeze-out of the top quark, with decoupling occurring some time between these; in this case the dark temperature could be as low as 0.3. These cases, however, are unrepresentative and only apply to a narrow range of values of ϵ . Even in the extreme case of $\xi \approx 0.3$, the estimate on α_{ion} would only differ by a factor of ≈ 0.8 , which is smaller than the error due to the uncertainty in the ionization fraction.

H₂ ionization

For the case where Ψ_2 can decay to Ψ_1 and the vector boson B , to make a rough estimate of the ionization fraction, we assume that recombination will typically happen in two steps: in the first, unbound Ψ_1 's combine with the free B 's to make a Ψ - B ion, while in the second these ions bind with a second Ψ_1 . The first step is similar to hydrogen atom recombination with the substitution $\alpha_g \rightarrow 2\alpha_g$ due to B having charge 2. In the second step, the potential at long range is like that for hydrogen atom recombination. Equation (46) then becomes

$$\begin{aligned} X_{e1} &\cong \left(1 + \xi^{-1} 16 \times 10^{10} \alpha_g^4 \frac{\text{GeV}}{m_1 m_B} \right)^{-1} \\ X_{e2} &\cong \left(1 + \xi^{-1} 10^{10} \alpha_g^4 \frac{\text{GeV}}{m_1 (m_B + m_1)} \right)^{-1} \\ X_{e,\text{tot}} &= X_{e1} + X_{e2} \cong X_{e2} \end{aligned} \quad (50)$$

The constraint on the ionization fraction ($X_{e,\text{tot}} \lesssim 0.1$) from [4] is therefore

$$\alpha_g \gtrsim \xi^{1/4} 4 \times 10^{-3} \left(\frac{m_{\mathbf{H}_2}}{\text{GeV}} \right)^{1/2} f_3^{-1/4}(R_2) \quad (51)$$

where $f_3(R_2) = (R_2 + 1/2)^2 / (R_2 + R_2^2)$.

5.2 Self interactions

Although standard cold dark matter is considered to be noninteracting with itself, there has been interest in variant theories where dark matter has an elastic self-scattering cross section of order 1b per GeV of DM mass. This has been motivated by persistent discrepancies between predictions of N -body simulations and observed properties of dark matter halos. While simulations tend to predict cuspy density profiles for galaxies, there is some observational evidence for cored profiles, especially in dwarf spheroidals. Simulations also

tend to predict too many high-mass satellite galaxies accompanying Milky-Way like progenitors compared to observations. For a review of these problems and their possible resolutions, see ref. [46]. A number of studies have been done indicating that the small-scale structure problems can be alleviated by invoking dark matter elastic scattering with $\sigma/m \sim 1\text{b}/\text{GeV}$. Dark atoms can naturally accommodate such large cross sections since they can have a significant geometric size.

The elastic scattering of dark atoms on each other has been studied very quantitatively, thanks to the fact that the problem can be mapped onto that of normal atom scattering with appropriate rescalings of parameters [16]. A useful rough estimate is that the scattering cross section goes as $\sigma \cong 100 a_0'^2 \cong 100 \alpha_g^{-2} f_2^2(R) m_{\mathbf{H}}^{-2}$. A cosmologically interesting level of self-scattering requires $\sigma/m_{\mathbf{H}} \sim 1.1 \text{ b}/\text{GeV} \cong 2800 \text{ GeV}^{-3}$ [47] in order to address the structure formation problems of cold dark matter. This corresponds to a gauge coupling of

$$\alpha_g = 0.2 f_2(R) (m_{\mathbf{H}}/\text{GeV})^{-3/2} \quad (52)$$

The criterion (52) can be compatible with the ionization constraint (19) if $m_{\mathbf{H}}$ is sufficiently small,

$$m_{\mathbf{H}} \lesssim 14 \text{ GeV} \left(\frac{f(R)}{4} \right)^{5/8} \quad (53)$$

obtained from eliminating α_g from the two relations. Very large values of R would be unnatural in our model, since it would require a fine-tuned cancellation between two contributions to $m_1^2 = (m_\psi - y_1\sigma)^2 + (y_2\sigma)^2$, as well as a small value of y_2 . An accidental cancellation at the level of $R = 10$ would allow for $m_{\mathbf{H}}$ as large as 28 GeV.

\mathbf{H}_2 self-interactions

In the \mathbf{H}_2 phase of the theory, the size of the atom is determined by the length scale a that describes the vector boson part of the wave function, rather than the characteristic distance b between the fermions, even though $b \sim 2a$. This is because the expectation values are $\langle r \rangle = 1.5a$, $\langle \Delta/s \rangle = 0.4b$. Therefore in parallel to the \mathbf{H} atom case, we can estimate the elastic cross section for atom-atom scattering as $\sigma \cong 100 a^2 \cong 100 \alpha_g^{-2} m_{\mathbf{H}_2}^{-2} R_2^2 f_4^2(R_2)$, where $f_4(R) = 1 + (2R_2)^{-1}$.

The gauge coupling corresponding to the desired scattering cross section of $\sigma/m_{\mathbf{H}_2} = 1.1 \text{ b/GeV}$ is therefore

$$\alpha_g = 0.19 R_2 f_4(R_2) \left(\frac{m_{\mathbf{H}_2}}{\text{GeV}} \right)^{-3/2}. \quad (54)$$

When combined with the constraint (51) on the ionization fraction, the result is

$$m_{\mathbf{H}_2} \lesssim 6.9 \text{ GeV} R_2 f_4 f_3^{-1/4}, \quad (55)$$

which is similar to the expression found for the \mathbf{H} case. The primary difference here is that large values of R_2 can be obtained without fine-tuning of model parameters, allowing for a larger natural range of masses consistent with both the ionization fraction and self-interaction constraints. (Notice that $f_{3,4} \rightarrow 1$ as R_2 becomes large.) Even with a moderate hierarchy $R_2 = 10$, we can reach masses as large as $m_{\mathbf{H}_2} \sim 70 \text{ GeV}$.

5.3 Laboratory millicharge searches

Pair production of $\bar{\Psi}_i \Psi_i$ is possible in accelerator experiments from the coupling of the photon to the dark matter millicharge. The resulting constraints on ϵ are quite weak in the mass range relevant for our model, $m_{\mathbf{H}} \sim$

1 – 100 GeV, as we show in fig. 7. The existing constraints are taken from tables in ref. [48] for the ASP and trident production limits, the E613 beam dump limit [49], ALEPH limits on the Z decay width [50] and a recent CMS search for particles of charge 1/3 or 2/3 [51]. We also show the reach of a new proposed experiment for LHC (dashed curve) [52]. These constraints are considerably weaker than that coming from direct detection, fig. (3), which is replotted as the dashed curve on fig. 7. Only at low ($m_{\mathbf{H}} \lesssim 4 \text{ GeV}$) or high ($m_{\mathbf{H}} < 100 \text{ GeV}$) masses, outside the sensitivity of direct detection, do they become dominant.

Possibly more significant constraints on millicharged particles arise from searches for exotic isotopes, bound states of normal nuclei with the charged DM constituents. Very stringent limits on the concentration of heavy isotopes of hydrogen or oxygen from sea water have been derived; for example ref. [53]) obtains an upper bound of 10^{-28} for the concentration of anomalously heavy H. These experiments assume integer-charged ions, but a recent experiment geared toward millicharged particles with $\epsilon > 10^{-5}$ set a limit of 10^{-14} on the abundance per nucleon. Naively such results would seem to rule out almost any values of $\epsilon \gtrsim 10^{-3}$ such that the binding energy $E_b \cong \frac{1}{2}(\alpha\epsilon)^2 m_p$ (for anomalous H) exceeds kT at room temperature, since we expect some fraction of ψ particles to remain ionized and thus be able to contaminate normal matter.

However to translate these limits on abundances into bounds on ϵ requires many considerations, including the expected flux of ψ particles, their capture cross section on the elements in question, the shielding of the earth and the galaxy from charged particles by magnetic fields, expulsion of charged particles by supernova winds, the process of purification of the samples studied, and the question of whether they apply to noninteger charged isotopes [21]. A

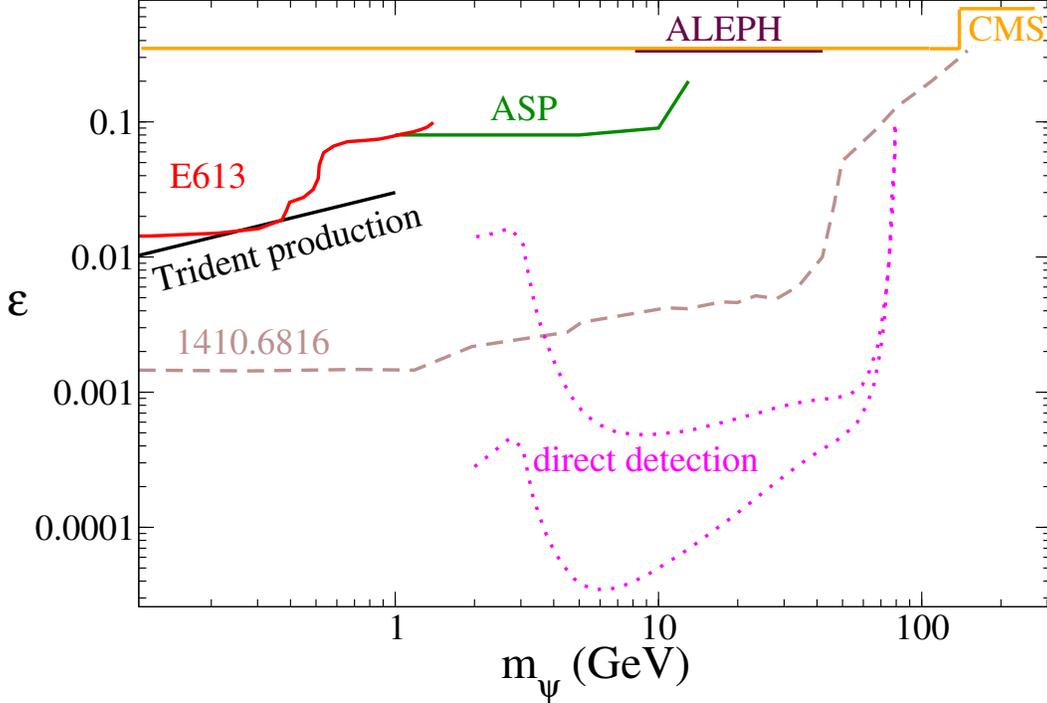


Figure 7: Solid curves: existing collider constraints on millicharge versus mass; dashed curve: expected reach of experiment proposed in ref. [52]. Dotted curves: our direct detection limits from fig. 3, depending on choice of $\alpha_g = 0.06$ (upper curve) or $\alpha_g = \alpha_{\text{ion}}$ (lower).

recent study of these issues was presented in ref. [54]. Here we take the view that there may be room for evading the anomalous isotope searches, but this question should be revisited if positive evidence for millicharges is found.

6 Conclusion

In this work we have tried to strike a balance between simplicity and realism in the construction of an atomic dark matter model. Our nonabelian construction is sufficiently rich to explain a unified origin of the massless dark photon and charged (under the hidden $U(1)_h$ interaction) DM constituents Ψ_i as a consequence of symmetry breaking $SU(2)_h \rightarrow U(1)_h$ by a scalar triplet VEV in the dark sector. With the addition of a dark Higgs doublet, we have

the necessary ingredients to explain the Ψ_i asymmetry through leptogenesis, simultaneously with the baryon asymmetry. Electric millicharges of Ψ_i , while not a necessary ingredient, can arise naturally through heavy states carrying both electric and $U(1)_h$ charge. Higgs portal interactions are also optional, but are allowed by a dimension-4 interaction of Ψ_i with the dark Higgs triplet and its mixing with the SM Higgs.

The model is mainly testable by direct detection. For sufficiently light or heavy constituents, the DM could also be discovered in an experiment proposed for LHC to probe millicharged particles. It can accommodate strong DM self-interactions as suggested by problems of Λ CDM simulations to correctly predict the small-scale structure of galaxies, if the dark atoms are not too heavy. Because of the requirement $\alpha_g \gtrsim \alpha_{\text{ion}}$, needed to make the ionization fraction in the dark sector sufficiently small, the symmetric component of the dark matter is highly suppressed due to annihilations into dark photons, making any indirect signals too weak to be detected.

Our model has a number of features that distinguish it from simplified atomic dark matter models. For example in the latter, the ratio R of the masses of the atomic constituents (which plays an important role) can be arbitrarily large, whereas here it is naturally of order 1, and requires fine-tuning to be much greater.

If the new Yukawa coupling y_1 exceeds the gauge coupling g , the stable dark matter particles can be the lighter fermion Ψ_1 and the doubly charged (under $U(1)_h$) vector boson B^{--} , leading to novel three-body $B\Psi\Psi$ bound states, where the mass ratio of the constituents m_1/m_B could be large without tuning of parameters (other than the usual hierarchy problem of light bosons). The properties of these unusual atoms for direct detection, as well as

for DM self-interactions, are qualitatively similar to those of the more conventional two-constituent atoms. This demonstrates a loophole for strong neutron star constraints on asymmetric bosonic dark matter, since the dark Coulomb repulsion prevents Bose condensation in this model.

For future work, these models suggest a potential novel signal for direct detection, due to the possible simultaneous presence of both dark atoms and a subdominant component of ionized or symmetric constituents. This would allow for the detection of both types of dark matter, typically having similar but distinct masses and interaction cross sections. Our analysis of leptogenesis as a common origin of the visible and hidden asymmetries is approximate, and it might also be interesting to undertake a more refined treatment for future studies.

Acknowledgments. We thank Sacha Davidson, Kimmo Kainulainen, Tim Linden, Zuowei Liu, Wei Xue and Wells Wulsin for helpful correspondence or discussion. We acknowledge support of the Natural Sciences and Engineering Research Council of Canada. We are grateful to NBIA for its generous hospitality while we were completing this work.

REFERENCES

- [1] M. Pospelov, A. Ritz, and M. B. Voloshin, *Phys.Lett.* **B662**, 53 (2008), 0711.4866.
- [2] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, *Phys.Rev.* **D79**, 015014 (2009), 0810.0713.
- [3] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, *JCAP* **0907**, 004 (2009), 0905.3039.
- [4] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, *JCAP* **1005**, 021 (2010), 0909.0753.
- [5] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, *JCAP* **1110**, 011 (2011), 1105.2073.
- [6] J. Fan, A. Katz, L. Randall, and M. Reece, *Phys.Rev.Lett.* **110**, 211302 (2013), 1303.3271.
- [7] J. Fan, A. Katz, L. Randall, and M. Reece, *Phys.Dark Univ.* **2**, 139 (2013), 1303.1521.
- [8] R. Foot, *Int. J. Mod. Phys.* **A29**, 1430013 (2014), 1401.3965.
- [9] I. García García, R. Lasenby, and J. March-Russell, (2015), 1505.07410.
- [10] J. Fan, A. Katz, and J. Shelton, *JCAP* **1406**, 059 (2014), 1312.1336.
- [11] R. Foot and S. Vagnozzi, *Phys. Lett.* **B748**, 61 (2015), 1412.0762.
- [12] M. T. Frandsen, F. Sannino, I. M. Shoemaker, and O. Svendsen, *JCAP* **1405**, 033 (2014), 1403.1570.
- [13] L. Pearce, K. Petraki, and A. Kusenko, *Phys. Rev.* **D91**, 083532 (2015), 1502.01755.

- [14] F.-Y. Cyr-Racine and K. Sigurdson, Phys. Rev. **D87**, 103515 (2013), 1209.5752.
- [15] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, Phys. Rev. **D89**, 063517 (2014), 1310.3278.
- [16] J. M. Cline, Z. Liu, G. Moore, and W. Xue, Phys. Rev. **D89**, 043514 (2014), 1311.6468.
- [17] J. M. Cline, Z. Liu, G. Moore, and W. Xue, Phys. Rev. **D90**, 015023 (2014), 1312.3325.
- [18] R. Foot and S. Vagnozzi, Phys. Rev. **D91**, 023512 (2015), 1409.7174.
- [19] K. Petraki, L. Pearce, and A. Kusenko, JCAP **1407**, 039 (2014), 1403.1077.
- [20] B. Holdom, Phys.Lett. **B166**, 196 (1986).
- [21] J. M. Cline, Z. Liu, and W. Xue, Phys.Rev. **D85**, 101302 (2012), 1201.4858.
- [22] J. M. Cline, Z. Liu, and W. Xue, Phys. Rev. **D87**, 015001 (2013), 1207.3039.
- [23] J. M. Cline, Y. Farzan, Z. Liu, G. D. Moore, and W. Xue, Phys. Rev. **D89**, 121302 (2014), 1404.3729.
- [24] L. Ackerman, M. R. Buckley, S. M. Carroll, and M. Kamionkowski, Phys. Rev. **D79**, 023519 (2009), 0810.5126, [,277(2008)].
- [25] H. An, S.-L. Chen, R. N. Mohapatra, and Y. Zhang, JHEP **03**, 124 (2010), 0911.4463.
- [26] E. J. Chun, Phys. Rev. **D83**, 053004 (2011), 1009.0983.
- [27] E. J. Chun, JHEP **03**, 098 (2011), 1102.3455.
- [28] C. Arina and N. Sahu, Nucl. Phys. **B854**, 666 (2012), 1108.3967.
- [29] A. Falkowski, J. T. Ruderman, and T. Volansky, JHEP **1105**, 106 (2011), 1101.4936.

- [30] W.-Z. Feng and P. Nath, Phys. Lett. **B731**, 43 (2014), 1312.1334.
- [31] F. Chen, J. M. Cline, and A. R. Frey, Phys.Rev. **D80**, 083516 (2009), 0907.4746.
- [32] E. Witten, Phys.Lett. **B117**, 324 (1982).
- [33] W. Buchmuller, P. Di Bari, and M. Plumacher, Annals Phys. **315**, 305 (2005), hep-ph/0401240.
- [34] S. Davidson and A. Ibarra, Phys. Lett. **B535**, 25 (2002), hep-ph/0202239.
- [35] T. R. Slatyer, Phys. Rev. **D87**, 123513 (2013), 1211.0283.
- [36] LUX Collaboration, D. Akerib *et al.*, Phys.Rev.Lett. **112**, 091303 (2014), 1310.8214.
- [37] CRESST, G. Angloher *et al.*, (2015), 1509.01515.
- [38] SuperCDMS, R. Agnese *et al.*, Submitted to: Phys. Rev. Lett. (2015), 1509.02448.
- [39] SuperCDMS, R. Agnese *et al.*, Phys. Rev. Lett. **112**, 241302 (2014), 1402.7137.
- [40] J. M. Cline, K. Kainulainen, P. Scott, and C. Weniger, Phys.Rev. **D88**, 055025 (2013), 1306.4710.
- [41] DELPHI, OPAL, ALEPH, LEP Working Group for Higgs Boson Searches, L3, S. Schael *et al.*, Eur. Phys. J. **C47**, 547 (2006), hep-ex/0602042.
- [42] I. Goldman and S. Nussinov, Phys. Rev. **D40**, 3221 (1989).
- [43] S. D. McDermott, H.-B. Yu, and K. M. Zurek, Phys. Rev. **D85**, 023519 (2012), 1103.5472.
- [44] K. M. Zurek, Phys.Rept. **537**, 91 (2014), 1308.0338.
- [45] C. Kouvaris and P. Tinyakov, Phys. Rev. Lett. **107**, 091301 (2011), 1104.0382.
- [46] D. H. Weinberg, J. S. Bullock, F. Governato, R. K. de Naray, and A. H. G. Peter, (2013), 1306.0913.

- [47] J. Zavala, M. Vogelsberger, and M. G. Walker, Monthly Notices of the Royal Astronomical Society: Letters **431**, L20 (2013), 1211.6426.
- [48] S. Davidson, B. Campbell, and D. Bailey, Phys. Rev. D **43**, 2314 (1991).
- [49] E. Golowich and R. W. Robinett, Phys. Rev. D **35**, 391 (1987).
- [50] ALEPH, D. Buskulic *et al.*, Phys. Lett. **B303**, 198 (1993).
- [51] CMS, S. Chatrchyan *et al.*, Phys. Rev. **D87**, 092008 (2013), 1210.2311.
- [52] A. Haas, C. S. Hill, E. Izaguirre, and I. Yavin, Phys. Lett. **B746**, 117 (2015), 1410.6816.
- [53] P. F. Smith *et al.*, Nucl. Phys. **B206**, 333 (1982).
- [54] C. Kouvaris, Phys. Rev. **D88**, 015001 (2013), 1304.7476.

2.4.1 Addendum:

The following clarifications to the paper should be added:

- The operator P_R , used in eq. (3), was not defined. This is the right chiral projection operator, whose definition is $P_R = \frac{1+\gamma^5}{2}$.
- In eq. (8) of the above manuscript, g_* is undefined. g_* refers to the effective relativistic degrees of freedom at the time of heavy neutrino decay.
- In eqs. (22) and (28), the variable R is used to represent scattering rates. This is distinct from the mass ratio of the dark atom constituents, which is denoted as R throughout the rest of the manuscript.

CHAPTER 3

The Galactic Center Excess

The Galactic Center excess, first observed by the Fermi-LAT space telescope in 2009 [75], is a γ -ray signal potentially originating from annihilating dark matter in the center of the Milky Way. Despite the signal's consistency with annihilating dark matter [75, 185–192], both in its spectrum and morphology, there is tension between the annihilation cross section required to explain the signal and observations of dwarf spheroidal galaxies at the frequencies in question [173, 193], which show no such signal [78].

In this chapter, we begin with a brief overview of the Galactic Center excess in section 3.1. In sections 3.2 and 3.3 we present two manuscripts which suggest possible methods of alleviating the tension between the Galactic Center Excess and dwarf spheroidal galaxy observations. In the former (section 3.2), we suggest p -wave (velocity-dependent) annihilating dark matter as a way of decreasing the annihilating rate in dwarf galaxies relative to the Milky Way. In the latter (section 3.3) we investigate the effect of cored dwarf spheroidal galaxy profiles on the bounds derived from their observation, showing that profiles consisted with self-interacting dark matter alleviate the tension.

3.1 The Galactic Center Excess: A Potential Dark Matter Signal?

The Galactic Center excess is a potential indirect detection (see section 1.3.2) signal observed by the Fermi-LAT space telescope, originating from the center of our galaxy [75, 185–192]. The signal appears as an excess in gamma rays above the expected astrophysical background in the range of 1–100 GeV. The signal is consistent with that expected from annihilating dark matter with a mass of approximately 50–80 GeV annihilating to $b\bar{b}$ or $\tau\bar{\tau}$ with

an annihilation cross section of $\langle\sigma v\rangle \approx 1.7 \times 10^{-26} \text{ cm}^3/\text{s}$ [189–191]. This mass range is very similar to that expected for WIMP dark matter, and the annihilation cross section is very near that which results in the correct relic density for dark matter: $\langle\sigma v\rangle \sim 3 \times 10^{-26} \text{ cm}^3/\text{s}$.

Although the signal has several promising characteristics, there are many uncertainties involved in modelling the gamma ray background of the Galactic Center, and several astrophysical alternatives have been proposed. One possibility is that the signal is produced by a previously unknown population of millisecond pulsars [194–196]. Millisecond pulsars produce a gamma ray flux with a similar shape to that produced by dark matter annihilating to heavy quarks or leptons, and the morphology (spatial distribution) of the signal is also well explained by millisecond pulsars [197, 198]. This would, however, require many more millisecond pulsars to be present in the Galactic Center than have been observed. Although this is not itself a problem for the theory given the difficulty of observing the Galactic Center, it has been argued that the number of millisecond pulsars can be inferred from the number of low-mass X-ray binaries, which would rule out millisecond pulsars as the source of more than $\sim 4 - 23\%$ of the Galactic Center excess [199]. These results, however, assume that the relative numbers of millisecond pulsars and low-mass X-ray binaries are the same in the Galactic Center as in globular clusters, leaving millisecond pulsars a viable candidate for the signal if the assumption does not hold. Another proposed source of such a signal is cosmic rays injected within the last million years, for example from supernova remnants [200–202]. These cosmic rays would diffuse through the Milky Way center and could reproduce the Galactic Center excess.

These astrophysical explanations are bolstered by the apparent tension between the dark matter explanation for the Galactic Center excess and observations of dwarf spheroidal galaxies [173, 193]. Dwarf spheroidal galaxies are very small galaxies (of order $10^9 M_\odot$) which are strongly dominated by dark matter. Because they lack the complex astrophysical processes of the Galactic Center and have very low luminosities — combined with their high dark matter density — these are ideal targets to observe annihilating dark matter. Observations of 25 dwarf spheroidal galaxies by Fermi-LAT have not observed a similar gamma-ray signal [78], and the constraints on dark matter annihilation within the dwarf spheroidals seemingly exclude dark matter as an explanation for the Galactic Center Excess¹.

Both papers included in the sections below propose ways to resolve this tension. We take two approaches: since the dark matter annihilation rate depends on both the annihilation cross section and the dark matter density, the tension is resolved if either the annihilation rate or density in the dwarf spheroidals is decreased relative to that in the Galactic Center.

3.2 *p*-Wave Annihilating Dark Matter

Our first approach to resolving the tension between the Galactic Center excess and observations of dwarf spheroidal galaxies is to consider a model in which the annihilation cross section would be smaller for the dwarf spheroidal galaxies. This can be achieved when dark matter annihilates through a velocity-dependent process; specifically we consider *p*-wave annihilation, in which the cross section scales as $\langle\sigma v\rangle \propto v^2$. Since the velocity dispersions in dwarf

¹ These observations were conducted on the dwarf spheroidal galaxies known at the time. The more recently discovered ultra-faint dwarf spheroidal galaxy Reticulum II, on the other hand, also exhibits a γ -ray excess [203, 204], which is consistent with the spectral shape of the Galactic Center Excess [205].

spheroidal galaxies are of the order $\sigma_v \sim 10 \text{ km/s}$,², a factor of 10 smaller than those of galaxies such as the Milky Way (which have $\sigma_v \sim 100 \text{ km/s}$), the annihilation rate is decreased by a factor of roughly 100. This is more than enough to alleviate the tension between the Galactic Center excess and null observations of dwarf spheroidal galaxies. It also presents an interesting possibility for detection given that it increases the expected signal from galaxy clusters (which have $\sigma_v \sim 1000 \text{ km/s}$) by a factor of 100.

The difficulty is formulating a model of p -wave annihilating dark matter which is consistent with observations. Making the model consistent with direct detection constraints necessitates limiting the nucleon scattering rate, which can be done by having dark matter annihilate first to a light mediator which subsequently decays to standard model particles. This allows the dark matter's Standard Model couplings to be made very weak while still having a large annihilation cross section, as the annihilation cross section depends only on the coupling to the mediator rather than to the Standard Model. p -wave annihilating dark matter also results in the wrong relic density unless it is produced non-thermally; we therefore consider a model in which a thermally-produced s -wave annihilating (velocity independent) predecessor is produced in the usual way before decaying to the dark matter which currently dominates.

We simulate the expected flux from p -wave annihilating dark matter in the Galactic Center and show that the dwarf spheroidal constraints are sufficiently relaxed to allow this model. We then determine the bounds on the

² The actual mean velocity is roughly half this[206] however the signal would be dominated by the high velocity end of the distribution.

parameter space of the model from a variety of sources, including direct detection experiments, the cosmic microwave background, big bang nucleosynthesis, production of the correct relic density, and collider constraints.

This manuscript was published in Physical Review D in 2016 [207]. Sections 3.2-3.3 and 4 were primarily written by my co-authors; the rest is primarily my own work.

p -wave Annihilating Dark Matter from a Decaying Predecessor and the Galactic Center Excess

Jeremie Choquette^{*1}, James M. Cline^{†1}, and Jonathan M. Cornell^{‡1}

¹*Department of Physics, McGill University, 3600 Rue University, Montréal,
Québec, Canada H3A 2T8*

Abstract

Dark matter (DM) annihilations have been widely studied as a possible explanation of excess gamma rays from the galactic center seen by Fermi/LAT. However most such models are in conflict with constraints from dwarf spheroidals. Motivated by this tension, we show that p -wave annihilating dark matter can easily accommodate both sets of observations due to the lower DM velocity dispersion in dwarf galaxies. Explaining the DM relic abundance is then challenging. We outline a scenario in which the usual thermal abundance is obtained through s -wave annihilations of a metastable particle, that eventually decays into the p -wave annihilating DM of the present epoch. The couplings and lifetime of the decaying particle are constrained by big bang nucleosynthesis, the cosmic microwave background and direct detection, but significant regions of parameter space are viable. A sufficiently large p -wave cross section can be found by annihilation into light mediators, that also give rise to Sommerfeld enhancement. A prediction of the scenario is enhanced annihilations in galaxy clusters.

Fermi-LAT Observations of the galactic center (GC) provide evidence of a gamma-ray excess in the multi-GeV energy range [1–9]. Millisecond pulsars (MSPs) are a favored astrophysical source to explain the signal [10–12], but there is debate in the literature as to whether the required numbers of MSPs in the GC for this purpose is consistent with the number resolved by Fermi [13] or expected on theoretical or empirical grounds [14]. Other possible astrophysical explanations have been presented [15–20], but dark matter annihilation into charged particles that lead to gamma rays remains a possibility that has attracted great interest. Further data should eventually be able to distinguish between the different possibilities [21, 22].

There is tension between most dark matter (DM) explanations of the galactic center excess (GCE) and constraints on dark matter annihilation coming from observations of dwarf spheroidal galaxies [23, 24].¹ (Further complementary constraints come from searches for GeV emission in the large Magellanic cloud [31] or subhalos of the Milky Way [32].) The best fits for DM mass and annihilation cross section for the GCE lie in regions that tend to be excluded by factors of a few by the dwarf spheroidal limits. A possible way of alleviating this tension is to assume that the annihilation is into electrons, a scenario in which the GCE is primarily produced through inverse Compton scattering which is suppressed in dwarfs because of their dilute radiation fields [33–35]. An additional idea to explain the discrepancy is a model of asymmetric DM where anti-DM is produced at late times via decays, leading to particles with enough kinetic energy to escape a dwarf galaxy but not the galactic center, where they annihilate with DM particles [36].

¹ Analyses of known dwarf spheroidal galaxies have revealed no significant excess gamma-ray emission. However, there have been claims of possible signals from the recently discovered [25, 26] dwarf spheroidal candidates Reticulum II [27, 28] and Turcana III [29]. These results are somewhat in dispute, with a Fermi-LAT analysis of Reticulum II using more data [30] claiming no significant excess.

In this work we explore a different possibility, noting that the tension can be avoided if the dark matter annihilation rates are velocity-dependent. Since the velocity dispersion in the galactic center is significantly higher than that in dwarf galaxies, the GCE can be consistent with the lack of signals from dwarf spheroidal galaxies provided that the annihilation cross section increases with velocity. This is the case in models where p -wave annihilations dominate, which is the subject of the present work. This scenario has recently been explored [37] to alleviate tension between the dwarf spheroidal constraints and DM explanations of the AMS-02 positron excess. We take a similar approach for the GCE. An immediate challenge is how to obtain the right relic density since the cross section needed for the GCE is of order $\langle\sigma v\rangle \sim 3 \times 10^{26} \text{cm}^3/\text{s}$, the usual value associated with a thermal origin for the relic density. But if $\langle\sigma v\rangle$ has such a value in the GC today, it would have been orders of magnitude larger in the early universe, leading to a negligible thermal abundance. We address this by showing how the current generation of p -wave annihilating dark matter could have arisen through the decays of a metastable predecessor DM particle that has a thermal origin. The decays can take place at temperatures ranging from $\sim 1 \text{eV}$ to several GeV. By this time the p -wave annihilations would be out of equilibrium despite their relatively large cross section.

The annihilation cross section needed to explain the GCE requires large couplings to compensate for the p -wave suppression. Such large couplings would generically tend to also give strong interactions of dark matter with nuclei. However constraints from direct detection can be satisfied if the dark matter annihilates into light mediators [38, 39] that subsequently decay into standard model particles. The light mediators also lead to Sommerfeld enhanced annihilation, allowing us to avoid nonperturbatively large couplings. In this way we are able to find viable models that have reasonably small couplings.

In section 1 we parametrize the p -wave annihilation cross section in the Milky Way (MW) and in dwarf spheroidal galaxies, in terms of assumed velocity dispersion profiles, leading to modified J -factors that are relevant for comparison to observations. In section 2 we give the results of the galactic propagation simulations used to compute the expected signal from the galactic center, including the effects of inverse Compton scattering and Bremsstrahlung radiation. This yields fits to the data in the plane of DM mass versus annihilation cross section σv . We then derive upper limits on σv in the same plane from dwarf spheroidals and galaxy clusters. In section 3 we show that p -wave annihilations of the desired strength would lead to strong suppression of the DM abundance at freeze-out, unless some nonthermal origin prevails. Here we present the scenario of decaying DM whose density is determined by the usual s -wave process, and the conditions under which this provides a consistent description. Three examples of decay channels leading to different phenomenology are presented, to illustrate the range of possibilities. In section 4 we systematically explore observational constraints on these models coming from cosmology, astrophysical line searches, direct searches, and colliders. In section 5 we provide a concrete model of χ annihilation into light scalar mediators to show that the desired large cross section can be achieved with reasonable values of the couplings in a renormalizable model. Conclusions are given in section 6.

1 Annihilation Cross Section

The expected signal from either the GC or dwarf spheroidals is proportional to the phase-space averaged cross section,

$$\langle \sigma v \rangle = \frac{1}{2} \int_0^{v_{\text{esc}}} dv_1 \int_0^{v_{\text{esc}}} dv_2 \int_{-1}^1 d \cos \theta f(v_1) f(v_2) \sigma v_{\text{rel}} \quad (1)$$

for a velocity distribution $f(v)$, where $v_{\text{rel}} = \sqrt{v_1^2 + v_2^2 - 2v_1v_2 \cos \theta}$ is the relative velocity between the two annihilating particles and the escape velocity

v_{esc} depends upon radial position r in the galaxy. In this work we consider Dirac fermion dark matter. Self-annihilating Majorana dark matter would introduce an additional factor of $1/2$ into equation (1). Following [40] and others, we adopt a Maxwell-Boltzmann distribution,

$$f(v) = \frac{3\sqrt{6}}{\sqrt{\pi}\sigma_v^3} v^2 e^{-3v^2/2\sigma_v^2}, \quad (2)$$

where σ_v is the velocity dispersion at the given r . The normalization factor in (2) is appropriate in the limit of large escape velocity, $v_{\text{esc}} \gg \sigma_v$. Numerically we find that this approximation is well-suited to the present applications.

We will be interested in p -wave annihilation for which at low velocities $\sigma v \cong \frac{1}{2}C_\sigma(v/c)^2$ with C_σ a constant. The phase-space averaged value is then

$$\langle \sigma v \rangle = C_\sigma(\sigma_v/c)^2 \quad (3)$$

In general σ_v is a function of r . This dependence is potentially significant in the MW, unlike in dwarf spheroidals, whose radial dependence has been observed to be roughly constant. Regardless of these details, it is however clear that $\langle \sigma v \rangle$ is several orders of magnitude lower in dwarf spheroidals (dSph) than in the MW if the cross section is p -wave suppressed. Measured values of σ_v are less than 15 km/s in MW dSph satellites [41], whereas most estimates of σ_v near the GC are $\gtrsim 130$ km/s (see for example refs. [42, 43]). On the other hand, Fermi upper limits on $\langle \sigma v \rangle$ from dSph observations are at most a factor of a few more stringent than the values of $\langle \sigma v \rangle$ needed to fit the GCE.

1.1 The Milky Way

The Milky Way, though composed predominantly of dark matter, has inner regions such as the bulge and bar (as well as Sagittarius A*) which are dominated by baryonic matter or otherwise do not follow an Navarro-Frenk-White (NFW) [44] profile. The velocity dispersion of dark matter in the Milky Way is difficult to measure directly in the inner region, hence we rely on

simulations and theoretical estimates. In order to quantify the uncertainties associated with choosing a velocity dispersion profile, we base our profiles on the results of simulations [45] that include baryonic matter to study the evolution of the Milky Way’s profile.

If the Milky Way contained no baryonic matter, it could be suitably modeled by an NFW profile. The resulting velocity dispersion, from fits to the aforementioned simulation, is

$$\sigma_v^3(r) = v_0^3 \left(\frac{r}{R_s} \right)^\chi \left(\frac{\rho(r)}{\rho_0} \right) \quad (4)$$

with $\chi = 1.87$ [46]. When baryons are included, however, a slope of $\chi = 1.64$ provides a better fit to the simulations [47]. We use the value $v_0 = 130 \text{ km s}^{-1}$, consistent with the results of [42, 43].

A second possibility that we consider is that the velocity dispersion of the Milky Way scales as a simple power law,

$$\sigma_v = v_0 \left(\frac{r}{R_s} \right)^\alpha, \quad (5)$$

as suggested by the results of ref. [45]. A numerical fit to those results gives $\alpha \cong -1/4$ [46] and, using our convention, a value of $v_0 = 104 \text{ km s}^{-1}$, which results in the same velocity dispersion at $r = R_\odot$ as eq. (4). Ref. [45] resolves only down to radii $r > 1 \text{ kpc}$, so (5) need not hold at smaller radii. Nevertheless we extrapolate it to $r < 1 \text{ kpc}$ to estimate the theoretical upper bound on the predicted GCE signal, which is greater for the ansatz (5) than for eq. (4). Since the observed signal is averaged over volume with r^2 weighting, the difference for the predicted GCE excess between the two assumptions is relatively small despite the fact that σ_v has very different behavior between the two as $r \rightarrow 0$.

1.2 Dwarf Spheroidal Galaxies

Dwarf spheroidal galaxies tend to have relatively flat observed velocity dispersion profiles out to large radii [41]. We therefore approximate them as

Galaxy	σ_v (km/s)	$\log_{10} J$	$\log_{10} J_p$	Ref.
Carina	7.5	18.1 ± 0.23	8.9	[49]
Draco	13	18.8 ± 0.16	10.1	[50]
Fornax	11.1	18.2 ± 18.2	9.3	[51]
Leo I	9.9	17.7 ± 17.7	8.7	[52]
Leo II	6.8	17.6 ± 0.18	8.3	[53]
Sculptor	9	18.6 ± 0.18	9.6	[54]
Sextans	8	18.4 ± 0.27	9.3	[49]
Ursa Minor	12	18.8 ± 0.19	10.0	[50]
Bootes I	6.6	18.8 ± 0.22	9.5	[55]
Canes Venatici I	7.6	17.7 ± 0.26	8.5	[56]
Canes Venatici II	4.6	17.9 ± 0.25	8.3	[56]
Coma Berenices	4.6	19.0 ± 0.25	9.4	[56]
Hercules	5.1	18.1 ± 0.25	8.6	[56]
Leo IV	3.3	17.9 ± 0.28	8.0	[56]
Segue 1	4.3	19.5 ± 0.29	9.8	[57]
Ursa Major I	7.6	18.3 ± 0.24	9.1	[56]
Ursa Major II	6.7	19.3 ± 0.28	10.0	[56]
Willman 1	4.0	19.1 ± 0.31	9.3	[58]

Table 1: J -factors for dwarf spheroidal galaxies with kinematic data [48] and velocity dispersion (with associated reference). J and J_p are given in $\text{GeV}^2 \text{cm}^{-5} \text{sr}$.

being constant, independent of radius. In this case, the J -factor for p -wave annihilation is simply proportional to that for s -wave. We define the former to be

$$J_p \equiv \int_{\Delta\Omega} \int_{\text{l.o.s.}} \rho(x)^2 \left(\frac{\sigma_v(r)}{c} \right)^2 dl d\Omega' \quad (6)$$

In ref. [48], the s -wave J -factors of the 18 dwarf spheroidal galaxies for which kinematic data was available were computed. We use these to determine J_p through the relation $J_p = J(\sigma_v/c)^2$. Table 1 shows the velocity dispersions and J -factors of the dwarf galaxies used.

2 Simulations and Indirect Limits

The observed gamma ray excess, if it originates from dark matter, can be the result of annihilations to SM particles. It has been shown that the

observed flux can be fit by annihilations with a large branching ratio to $b\bar{b}$, as would be expected for Higgs portal dark matter [6, 7, 59]. Although most of these gamma rays are prompt (decay products of the b quarks), a significant fraction comes from inverse Compton scattering and, to a lesser extent, from bremsstrahlung. While the prompt signal can be relatively easily computed, the ICS and bremsstrahlung contributions are more involved. To this end, we use the DRAGON [60] code to simulate cosmic ray production and propagation from dark matter annihilations, and the GammaSky program which implements GALPROP [61] to simulate the ICS and bremsstrahlung contributions along the line of sight. GammaSky is as yet unreleased, but some results have been given in [62].

We have modified DRAGON to account for p -wave annihilating dark matter, replacing the constant cross section appearing with the DM density by $\sigma v \rho(r) \rightarrow \rho(r) C_\sigma (\sigma_v(r)/c)^2$. We also incorporate a generalized NFW profile

$$\rho(r) = \frac{\rho_0}{\left(\frac{r}{R_s}\right)^\gamma \left(1 + \frac{r}{R_s}\right)^{3-\gamma}} \quad (7)$$

and the galactic diffusion parameters and magnetic field model used in ref. [6], corresponding to their best-fit model (therein called Model F). The NFW parameters are taken to be $\rho_0 = 0.3 \text{ GeV cm}^{-3}$ (giving a local DM density of 0.4 GeV cm^{-3}), $R_s = 20 \text{ kpc}$, and $\gamma = 1.2$ (the best-fit value for the GCE found in [6, 7]). The electron injection spectrum is taken from PPC 4 [63, 64], as is the photon spectrum used in calculating the prompt contribution.

We will focus on models in which DM annihilates into on-shell scalar mediators ϕ that subsequently decay into SM particles, primarily $b\bar{b}$. The prompt photon and electron spectra must be boosted with respect to those from DM annihilating at rest, to account for the velocity of ϕ when it decays. The decay spectrum into particles of type $i = \gamma, e$ in the rest frame of the ϕ is denoted by $dN_i^{(\phi)}/dE$. It is related to the spectrum in the center of mass

frame of the $\chi\bar{\chi}$ system by [65, 66]

$$\frac{dN_i^{(\chi)}}{dE} = \frac{2}{(x_+ - x_-)} \int_{E x_-}^{E x_+} \frac{dE'}{E'} \frac{dN_i^{(\phi)}}{dE'}, \quad (8)$$

where $x_{\pm} = m_{\chi}/m_{\phi} \pm \sqrt{(m_{\chi}/m_{\phi})^2 - 1}$. This expression assumes that the final state particles are massless, which is approximately true for the electrons as well as the photons injected from b decays.

The prompt photon spectrum can be calculated independently of the DRAGON simulation. Its integrated spectral flux (in units of photons \cdot $\text{cm}^{-2} \text{s}^{-1}$) is given by

$$\frac{d\Phi_{\text{prompt}}}{dE} = \frac{C_{\sigma}}{8\pi m_{\chi}^2} \frac{dN_{\gamma}}{dE} \times J_p, \quad (9)$$

with J_p defined in eq. (6). The total observed spectrum is equal to the sum of $d\Phi_{\text{prompt}}/dE$ and the ICS+Bremsstrahlung spectrum determined from the simulations.

2.1 Simulation Results

We simulated the gamma ray flux for a range of dark matter masses ($20 \text{ GeV} \leq m_{\chi} \leq 200 \text{ GeV}$) and compared the results to the GCE signals estimated in refs. [6–8]. The best fit regions are presented in fig. 1, which show the confidence intervals in the C_{σ} - m_{χ} plane for four different values for the mediator mass, $m_{\phi} = 12, 20, 30, 50 \text{ GeV}$. The contours are generated by minimizing the χ^2 of our simulated spectrum with respect to each dataset in the C_{σ} - m_{χ} plane, and contours are then drawn at $\chi_{\text{min}}^2 + 2.30$, $\chi_{\text{min}}^2 + 6.18$, and $\chi_{\text{min}}^2 + 11.83$, corresponding to 1σ , 2σ , and 3σ . The minimum values of χ^2 and the corresponding model parameters are given in table 2, which shows that the fit results are relatively insensitive to the mediator mass (the fits to the Fermi data display a mild preference for heavier mediators). Reasonably

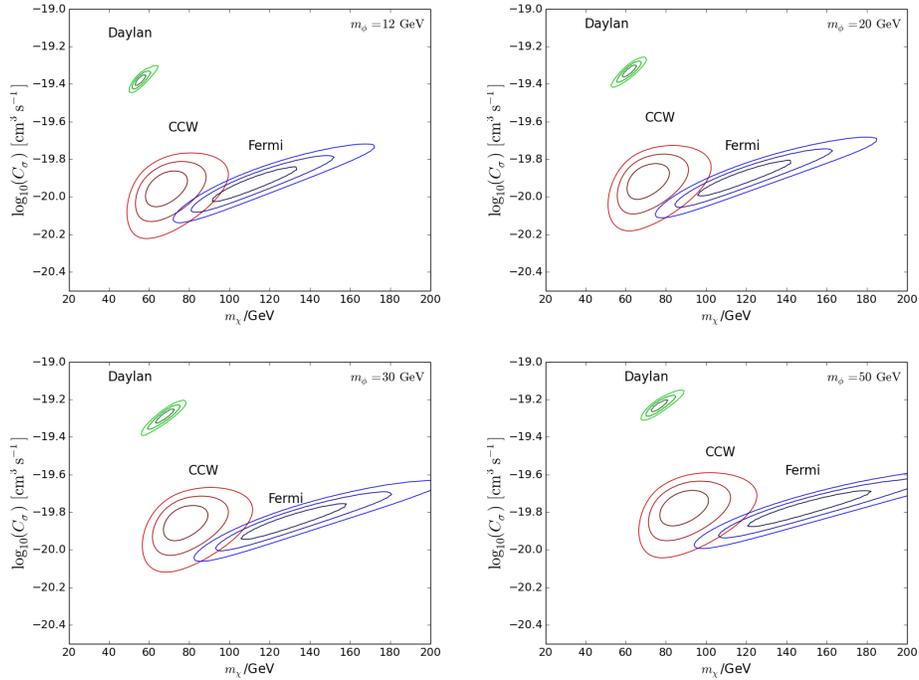


Figure 1: 1σ , 2σ , and 3σ contours for the CCW [6], Daylan *et al.* [7], and Fermi Collaboration [9] data. The results are shown for annihilation into on-shell scalar mediators, followed by decay into $b\bar{b}$, with a mediator mass of $m_\phi = 12, 20, 30,$ and 50 GeV. Shaded regions in upper left corner indicate the constraint from the Virgo cluster.

good fits to the Fermi and CCW data sets are obtained, with

$$m_\chi \sim 90 \text{ GeV}, \quad C_\sigma \sim 10^{-20} \text{ cm}^3 \text{ s}^{-1} \quad (10)$$

whereas the fit to the Daylan *et al.* data is poor. The data are compared to the simulated observed spectrum from the GC in fig. 2 for representative values of m_χ and C_σ , taking a mediator mass of $m_\phi = 12$ GeV.

The previous results are based upon the assumption of eq. (4) for the DM velocity dispersion in the MW. The effect of using higher σ_v , using eq. 5, is shown in fig. 3, which results in somewhat lower central values of $C_\sigma \sim 0.2 \times 10^{-20} \text{ cm}^3 \text{ s}^{-1}$ for the cross section and $m_\chi \sim 80$ GeV for the mass.

m_ϕ	CCW ($N = 21$)			Fermi ($N = 20$)			Daylan ($N = 25$)		
	χ^2_{\min}	m_χ	$\log_{10} C_\sigma$	χ^2_{\min}	m_χ	$\log_{10} C_\sigma$	χ^2_{\min}	m_χ	$\log_{10} C_\sigma$
12	29.7	68	-20.0	24.9	109	-19.9	54.1	56	-19.4
20	29.9	70	-19.9	23.7	116	-19.9	65.3	62	-19.3
30	29.9	76	-19.9	22.7	128	-19.9	71.3	67	-19.3
50	30.6	88	-19.8	22.0	146	-19.8	76.8	76	-19.2

Table 2: Minimum χ^2 values for fits to the three datasets, (number of data points N indicated). Masses are in GeV and C_σ is in cm^3s^{-1} . The confidence regions are shown in fig. 1.

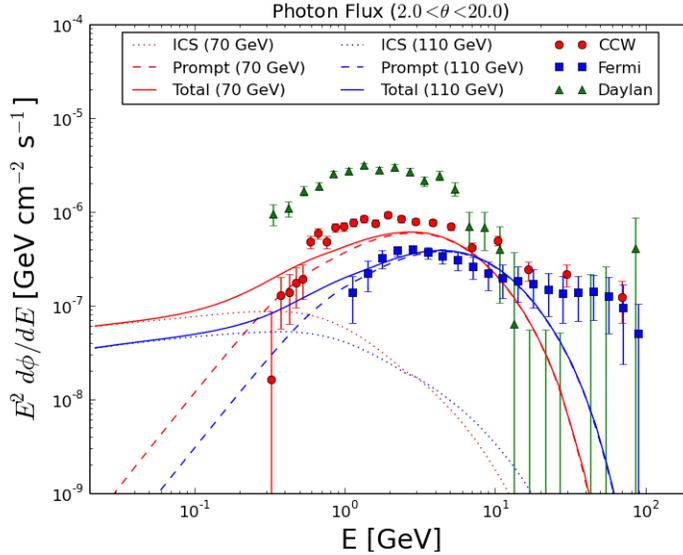


Figure 2: Simulated observed photon energy flux for p -wave annihilating dark matter with $m_\chi = 70$ GeV (red, upper curved) or $m_\chi = 110$ GeV (blue, lower curves), mediator mass $m_\phi = 12$ GeV and cross section coefficient $C_\sigma = 10^{-20} \text{cm}^3\text{s}^{-1}$. The observed region is the disk-like region $2.0 < \theta < 20.0$, where θ is the viewing angle as measured from the galactic center. The ICS+bremsstrahlung (dotted) and prompt (dashed) components are shown individually. Also shown are the three datasets of observed fluxes; the values of m_χ are chosen to demonstrate the best fits to two of the individual datasets.

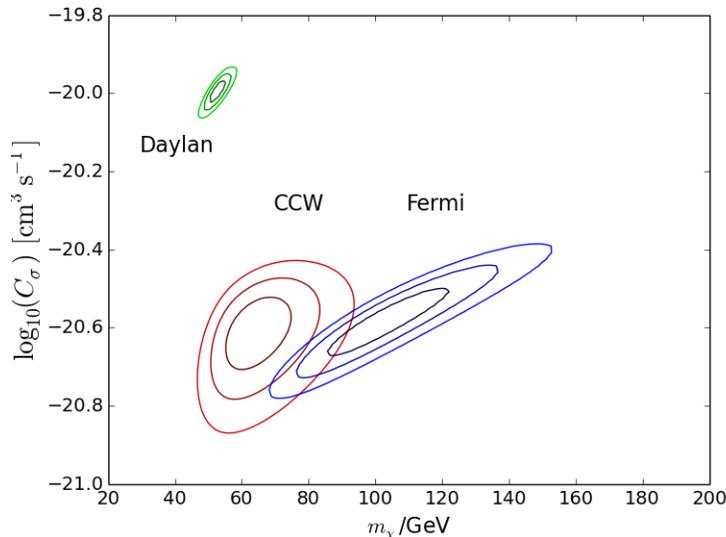


Figure 3: Like fig. 1, but using the velocity dispersion profile in equation 5, with a mediator mass of $m_\phi = 12$.

2.2 Limits from Dwarf Spheroidal Galaxies

An upper limit on the gamma-ray flux from DM annihilation in 18 dwarf spheroidal galaxies with kinematic data has been determined using Fermi-LAT data [48]. This can be used in conjunction with the J -factors presented in table 1 to obtain an upper limit on C_σ . The strongest such constraint comes from the dwarf galaxy Draco. At a distance of 80 kpc and with a relatively large J -factor and high velocity dispersion, it would be the most likely to exhibit signs of p -wave annihilating dark matter.

Ackermann *et al.* give the combined limit on $\langle\sigma v\rangle_{b\bar{b}}$ (annihilation into $b\bar{b}$) at 95% C.L. for 15 dwarf spheroidal galaxies. In our model, DM annihilates to $b\bar{b}b\bar{b}$, leading to a different gamma-ray spectrum, but in this section and the next we assume the resulting limit on the annihilation rate in both cases is approximately the same. (Note that the total energy deposition in the two cases is the same.)

The previously derived limit assumes s -wave annihilation and therefore cannot be directly converted into a limit from p -wave annihilation, as the

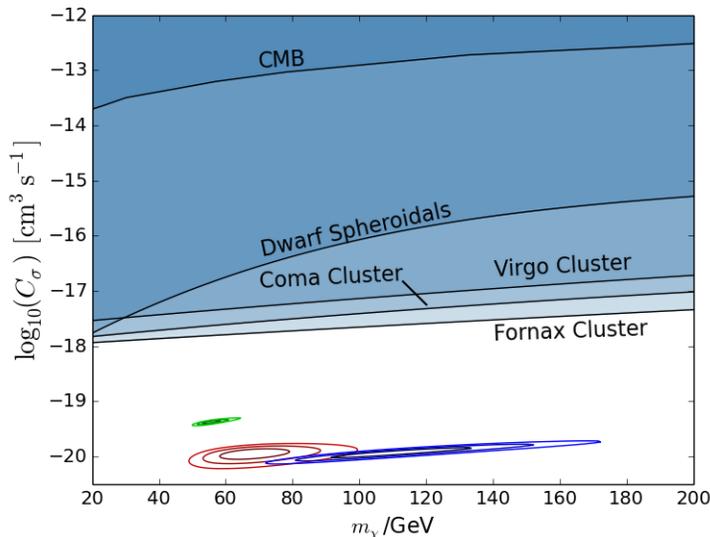


Figure 4: Like figure 1, including 95% C.L. upper limits on C_σ from the five most constraining dwarf spheroidals, the Virgo, Fornax, and Coma clusters, and the CMB. The fits to the GCE for p -wave annihilating dark matter are well below the limits. The CMB constraint is taken from ref. [67], for the case of annihilations to e^+e^- .

different velocity dispersions of the dwarf spheroidals would have to be taken into account individually. If, however, we make the simplifying assumption that all dwarf spheroidal galaxies have velocity dispersions equal to the greatest value (that of Draco, with $\sigma_v = 13$ km/s), we can then use equation 3 to directly convert the limit to one on C_σ . This will lead to a constraint that is slightly more stringent than the true value, but sufficient for our purpose of showing that there is no tension with the GCE. The resulting upper limit on C_σ as a function of m_χ is shown in figure 4, along with the GCE best-fit regions. The weaker CMB constraint from energy injection at recombination [67] (also discussed in section 4.2) is also indicated there.

We see that the assumption of p -wave annihilation rather than s -wave completely eliminates the tension between the dwarf spheroidal constraints and the GCE. The former are softened by a factor of $\sim \sigma_{v,\text{dwarf}}^2 / \sigma_{v,\text{MW}}^2 \sim (13/130)^2 \sim 10^{-2}$ relative to the GCE signal. The constraints depend on the velocity dispersion profile assumed for the dwarfs, but even taking into account

the uncertainties, the limiting cross section from dwarf spheroidal galaxies is far above the values required to explain the GCE.

2.3 Galaxy Cluster Limits

Searches for gamma rays from galaxy clusters can place more stringent constraints on our scenario. Although dwarf spheroidal constraints were weakened due to their smaller velocity dispersion, the converse is true for clusters: their larger velocity dispersions amplify the signal from p -wave annihilating dark matter, relative to smaller systems.

Observations of the Coma [68] and Virgo [69] clusters have recently been analyzed by the Fermi-LAT Collaboration. The first of these references gave no limits on annihilating dark matter, while the second did so for s -wave annihilations. We therefore derive the bound on p -wave annihilating DM arising from the latter. For this purpose we adopt a value for the velocity dispersion of 643 km/s for Virgo [70].

Limits on $\langle\sigma v\rangle_{b\bar{b}}$ are derived at 95% C.L. for the Virgo cluster in [69], using a background model taking into account all Fermi 2-year catalog point sources as well as diffuse galactic and extragalactic spectra. We have converted them directly into limits on C_σ using equation 3, with one caveat: dark matter substructure—subhalos residing within the larger host halo—is expected to significantly boost the signal strength from s -wave dark matter annihilation over what would be expected from the host halo alone. The constraints in ref. [69] for the more conservative limit given assume a boost factor of $b = 33$ from the substructure of the cluster. The substructure is not expected to have the same velocity dispersion as the host halo however, making the simple rescaling described in the previous section inapplicable for p -wave annihilation. Subhalos generically have a significantly smaller velocity dispersion than the host halo, due to the fact that the velocity dispersion depends on the total mass of the subhalo where the dark matter is virially bound, not on that of the host halo. This can be seen in simulations such as RHAPSODY [71, 72],

in which the number of galaxy cluster subhalos is found to drop off sharply with increasing maximum circular velocity (a power law index of -2.95) with no subhalos exceeding a third of the host halo's maximum circular velocity. The contribution to the signal from subhalos is therefore weakened due to the velocity dependence of the annihilation cross section, offsetting the gains that come from the increased dark matter density. Ultimately, we choose a conservative approach and rescale the limits from [69] by a factor of $b + 1$ to remove the boost from the substructure for a self-consistent limit. The upper limits on C_σ are shown in fig. 4.

Similar limits have been found for several other clusters, including Coma and Fornax, using earlier Fermi data [73]. The Fornax cluster was subsequently reanalyzed with specific attention to the effects of subhalos and contraction due to baryonic infall [74], leading to a more stringent upper bound on $\langle\sigma v\rangle_{b\bar{b}}$. As with the Virgo cluster, from this work we use the conservative limits neglecting the effect of substructure, which in ref. [74] are given alongside the more optimistic limits. We convert the constraints on Coma [73] (which does not account for substructure) and Fornax [74] directly into limits on C_σ , using velocity dispersions of 913 km/s [70] and 370 km/s [75] respectively; these are also included in fig. 4.

Although our best-fit parameters are consistent with older bounds from the Virgo and Fornax clusters, more recent observations of the Coma cluster are expected to give more stringent constraints due to its high dark matter density and larger velocity dispersion. Currently there are no limits on dark matter annihilation rates from the more recent observations, and such a study is beyond the scope of the present work.

2.4 Isotropic Gamma-Ray Background

The isotropic gamma-ray background (IGRB) could further constrain our scenario. As part of the DM annihilation contribution to this signal could be

from even larger halos than the ones surrounding the galaxy clusters we considered in the previous section, it is possible that it could be enhanced if the annihilation cross section is velocity dependent. The most recent measurements of the IGRB can constrain the s -wave annihilation cross-section to $\langle\sigma v\rangle \lesssim 10^{-24} \text{ cm}^3 \text{ s}^{-1}$ for conservative limits and $\langle\sigma v\rangle \lesssim \langle\sigma v\rangle_{\text{thermal}}$ for more optimistic limits corresponding to our adopted best-fit value of $m_\chi = 80 \text{ GeV}$ [76]. Converting these limits into constraints on C_σ is not a simple matter, as arriving at an expected IGRB signal requires taking into account how the velocity dispersion varies for halos of different sizes and at different redshifts. Such a detailed analysis is beyond the scope of this work but would be interesting for future investigation.

3 Relic abundance from decaying dark matter

An immediate problem with p -wave annihilating DM in the galactic center is that the corresponding cross section in the early universe would have been orders of magnitude greater, due to the larger relative velocities, leading to a highly suppressed relic density. The form of the Boltzmann equation which describes the time evolution of the number density for Dirac dark matter χ is

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\langle\sigma v\rangle \left(n_\chi n_{\bar{\chi}} - n_\chi^{\text{EQ}} n_{\bar{\chi}}^{\text{EQ}} \right), \quad (11)$$

where n^{EQ} is the number density of a particle in thermal equilibrium with the photon bath. The equation for the evolution of the number density of the antiparticle $\bar{\chi}$ is of the same form. We assume that there is no asymmetry between n_χ and $n_{\bar{\chi}}$, and therefore the total number density $n = n_\chi + n_{\bar{\chi}}$ is given by

$$\frac{dn}{dt} + 3Hn = -\frac{\langle\sigma v\rangle}{2} (n^2 - n_{\text{EQ}}^2). \quad (12)$$

Following the procedure of ref. [77], an approximate solution of the Boltzmann equation for the relic density is given by

$$\Omega_\chi h^2 = \frac{\rho_\chi}{\rho_c} h^2 = 2.14 \times 10^9 \frac{(n+1)x_f^{(n+1)}}{\left(g_{*S}/g_*^{1/2}\right) M_{\text{Pl}} \sigma_0} \text{GeV}^{-1}. \quad (13)$$

where $x_f = m_\chi/T_f$, T_f is the freeze-out temperature, and the effective degrees of freedom g_* and g_{*S} are evaluated at T_f . The thermally averaged cross section takes the form $\langle\sigma v\rangle = \sigma_0 x_f^{-n}$; hence $n = 1$ and $\sigma_0 = 3C_\sigma$ for our p -wave annihilation scenario where

$$\langle\sigma v\rangle = 3C_\sigma \frac{T}{m_\chi} \quad (14)$$

An approximate solution for x_f is given by

$$\begin{aligned} x_f &= \ln y_f - (n+1/2) \ln \ln y_f, \\ y_f &= 0.038 (n+1) g_*^{-1/2} M_{\text{Pl}} m_\chi \sigma_0 \end{aligned} \quad (15)$$

Our fiducial fit, eq. (10), implies

$$x_f = 32.3 \quad \Omega_\chi = \frac{\rho_\chi}{\rho_c} = 3.6 \times 10^{-5} \quad (16)$$

to be compared to the observed value $\Omega_{DM} = 0.26$ [78]. Hence the thermally produced abundance is approximately 7000 times too small; we need a nonthermal production mechanism.

3.1 Decaying dark matter

A conceptually simple solution, similar to the superWIMP model proposed in [79], is to suppose that today's dark matter χ is the product of a heavier metastable state ψ , that decayed into χ at temperatures below freeze-out of $\bar{\chi}\chi$ annihilations. For $m_\chi \sim 90 \text{ GeV}$, this occurs at $T_f \sim m_\chi/32 \sim 3 \text{ GeV}$ according to (16). Hence we need for ψ to have a lifetime exceeding 10^{-6} s . Such long lifetimes are suggestive of an analog of weak interactions in the dark

sector. We consider representative effective interactions giving rise to decays $\psi \rightarrow \chi\gamma$, $\psi \rightarrow \chi e^+ e^-$ or $\psi \rightarrow \chi b\bar{b}$, of the form

$$\frac{1}{\Lambda_\gamma} \bar{\chi} \sigma_{\mu\nu} \psi F^{\mu\nu}, \quad \frac{(\bar{\chi} \gamma^\mu \psi)(\bar{e} \gamma_\mu e)}{\Lambda_e^2}, \quad \frac{(\bar{\chi} \gamma^\mu \psi)(\bar{b} \gamma_\mu b)}{\Lambda_b^2}, \quad (17)$$

where $\Lambda_{e,\gamma,b}$ are heavy scales. Each operator is also accompanied by its Hermitian conjugate, which leads to decays of $\bar{\psi}$. These decay channels are chosen to illustrate constraints that can arise from big bang nucleosynthesis (BBN) and the cosmic microwave background (CMB). An alternate channel $\psi \rightarrow \chi \nu \bar{\nu}$ would be safe from these constraints. The decay rates corresponding to the first two operators are given by

$$\Gamma_\gamma = \frac{4 \delta m^3}{\pi \Lambda_\gamma^2}, \quad \Gamma_{ee} \cong \frac{\delta m^5}{60 \pi^3 \Lambda_e^4} \quad (18)$$

where the mass splitting $\delta m = m_\psi - m_\chi$ is considered to be much less than $m_\psi \cong m_\chi$, but greater than $2m_e$ for decays into electrons. (We ignore phase space effects in the small region of parameter space where $\delta m \gtrsim 2m_e$.) For the third operator, we are interested in larger mass splittings since δm must be at least $2m_b$. We use numerical results for its decay rate where needed. A fairly good fit is given by $\Gamma_b \cong A_0(m_\psi^{A_1} - m_\chi^{A_1})^{A_2} / \Lambda_b^4$ where for $\Gamma_b, m_{\psi,\chi}$ in GeV units, $A_{0,1,2} = (3.60, 1.33, 2.30)$.

To obtain the relic density of the parent particle ψ , we assume for definiteness an effective interaction

$$\frac{(\bar{\psi} \gamma^\mu \psi)(\bar{f} \gamma_\mu f)}{\Lambda_f^2}, \quad (19)$$

giving rise to $\psi\bar{\psi} \rightarrow f\bar{f}$, where f can be a light fermion of the standard model or in a hidden sector. The annihilation cross section for $\psi\bar{\psi} \rightarrow f\bar{f}$ is

$$\langle \sigma v \rangle \cong \frac{m_\psi^2}{\pi \Lambda_f^4} \quad (20)$$

To determine the relic density in this scenario, we again use eqs. 13 and 16 but now with $n = 0$, since the vector current operators of eq. (17) lead

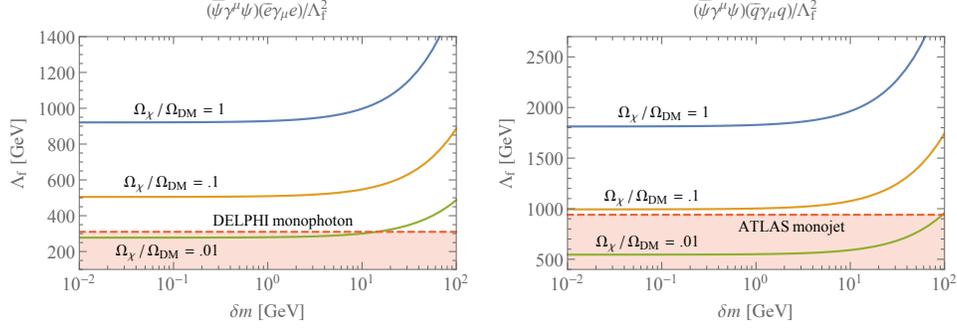


Figure 5: Contours of constant relic density for a dark matter mass of $m_\chi = 90$ GeV assuming that the $\psi\chi$ coannihilation rate is negligible relative to the $\psi\bar{\psi}$ annihilation rate. In the left plot, ψ and $\bar{\psi}$ couple to e^+e^- and in the right they couple to quarks. The shaded region in the left plot is excluded at 90% C.L. from a DELPHI search for monophotons, while an ATLAS search for monojets excludes the shaded region in the right plot at 95% C.L.

to s -wave annihilation. The ultimate relic density of χ particles is related to the prior abundance of ψ by $\Omega_\chi = (m_\chi/m_\psi)\Omega_\psi$. Curves of constant Ω_χ in the δm - Λ_f plane for $m_\chi = 90$ GeV are shown in figure 5. Here we consider two different scenarios: $\psi\bar{\psi}$ annihilations to electrons and positrons and to quark-antiquark pairs. Large mass splittings $\delta m \gtrsim 1$ GeV lead to a reduction in Ω_χ that must be compensated by reducing the cross section by increasing Λ_f . These estimates assume that coannihilations $\psi\chi \rightarrow \bar{f}f$ as well as inelastic scatters $\psi f \rightarrow \chi f$ are unimportant for determining the DM relic density. This will be true (as we explore in detail in the following subsections) as long as $\Lambda_e \gg \Lambda_f$, which is also consistent with the need for ψ to be relatively long-lived. For small $\delta m \lesssim 1$ GeV, the desired relic density for ψ and χ is independent of δm and requires $\Lambda_f \cong 920$ GeV when ψ and $\bar{\psi}$ couple to e^+e^- and $\Lambda_f \cong 1810$ GeV when they couple to $q\bar{q}$.

3.2 Coannihilations

Coannihilation processes can reduce the relic density of ψ , which was assumed to be a small effect in the previous treatment. When the splitting between m_ψ and m_χ is small, leading to $n_\psi \approx n_\chi$, the effect can be estimated

by replacing $\langle\sigma v\rangle$ in eq. 12 with [80]

$$\langle\sigma_{\text{eff}} v\rangle = \langle\sigma_{\psi\bar{\psi}\rightarrow X\bar{X}} v\rangle + \langle\sigma_{\psi\chi\rightarrow X\bar{X}} v\rangle. \quad (21)$$

Here X represents any standard model particle, so the first term in the above equation is the total $\psi\bar{\psi}$ annihilation cross section and the second is the total $\psi\chi$ coannihilation cross section. Eq. (12) with this effective cross section only describes the number density of ψ until the freeze-out of this species, since after that point decays and inelastic scatterings can have a significant impact on the ψ number density. In this low mass splitting limit, the relevant $\psi\chi \rightarrow f\bar{f}$ coannihilation cross section for the 4-fermion operator in eq. (17) has the same form as eq. (20), while the dipole operator gives [81]

$$\langle\sigma_{\psi\chi\rightarrow f\bar{f}} v\rangle = \frac{4\alpha Q_f^2}{\Lambda_\gamma^2}, \quad (22)$$

where Q_f is the electric charge of the fermions in the final state and α is the fine-structure constant.

In the scenario where the relic density is determined entirely by coannihilation processes, *i.e.*, when the operator in eq. (19) is not present, the correct relic density requires $\Lambda_e \geq 920$ GeV or $\Lambda_\gamma \geq 8000$ GeV. These are lower bounds, since increasing the strength of coannihilation processes would lead to underproduction of DM, while the larger relic density induced by decreased coannihilation can be offset by increased $\psi\bar{\psi}$ annihilation.

The resulting limits are shown in fig. 6. Decays of ψ to b -quarks require a relatively large mass splitting and consequently a more sophisticated calculation than the one described here, but the limits on Λ_b from suppressing coannihilations are greatly subdominant to those from demanding that ψ decays after χ freeze-out. We also note that the operators in eq. (17) lead to additional annihilation processes from the ones we have considered above, including $\psi\bar{\psi} \rightarrow \bar{f}f\bar{f}f$ for the four-fermion operator as well as $\psi\bar{\psi} \rightarrow \gamma\gamma$ and $\psi\bar{\psi} \rightarrow \gamma\phi$ for the magnetic dipole operator. We have checked that these are negligible when the other constraints considered are satisfied.

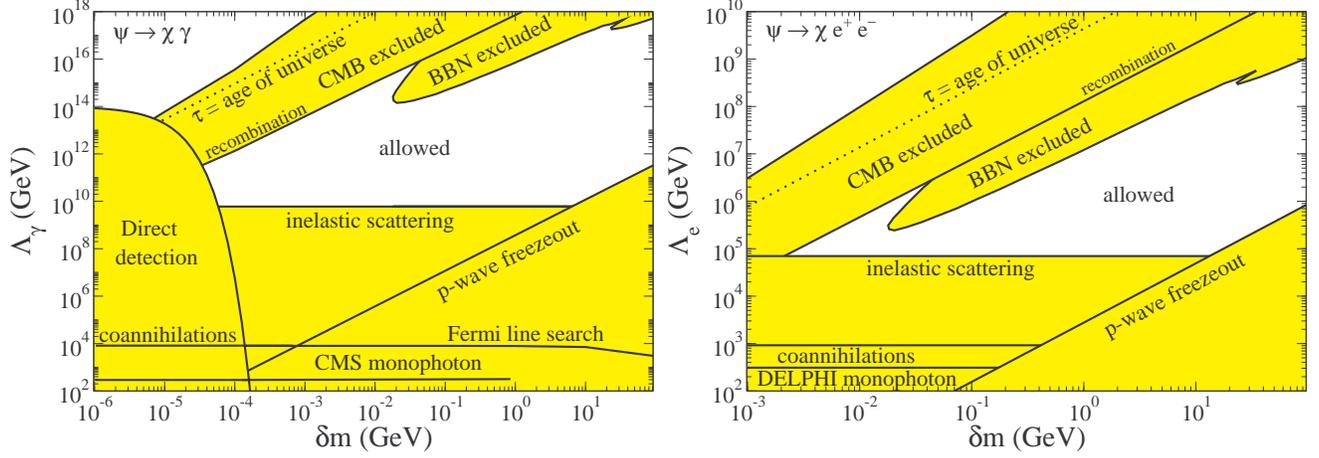


Figure 6: Left: Excluded (shaded) and allowed (unshaded) regions of parameter space for $\psi \rightarrow \chi\gamma$ decays in the δm - Λ_γ plane. In the upper-left regions, the lifetime of the ψ is too great, causing its decays to interfere with BBN, CMB, or exceed the age of the universe; in the lower right regions ψ decays before the $\chi\bar{\chi}$ annihilations freeze out, erasing any excess above the standard χ relic abundance produced via pair annihilation. Right: corresponding result for $\psi \rightarrow \chi e^+e^-$. Dark matter mass $m_\chi = 90$ GeV was assumed for determining the number density of decaying ψ particles.

3.3 Inelastic scattering

A further requirement for consistency of our relic density determination is that inelastic scatterings $\psi f \rightarrow \chi f$ induced by the decay operators (17) are not important during the epoch between ψ freeze-out and the significantly later χ freeze-out. Otherwise further depletion of the final abundance would occur due to scattering-induced $\psi \rightarrow \chi$ transitions followed by $\chi\bar{\chi}$ annihilations. This leads to the criterion

$$(n_e + n_{\bar{e}})\langle\sigma v\rangle_{\psi e\rightarrow\chi e} < n_\psi\langle\sigma v\rangle_{\psi\bar{\psi}\rightarrow f\bar{f}} \quad (23)$$

for the $(\bar{\chi}\gamma^\mu\psi)(\bar{e}\gamma_\mu e)$ operator. We ignore the effect of $\chi \rightarrow \psi$ transitions because the number density of χ relative to ψ is extremely suppressed in our scenario at χ freeze-out. For $m_\chi = 90$ GeV, freeze-out occurs at $T_f \cong 3$ GeV, for which it is sufficient to compute the inelastic cross section in the elastic

limit $\delta m = 0$, and also approximating $m_e \cong 0$. We find that

$$(\sigma v)_{\psi e \rightarrow \chi e} \cong \frac{E_e^2}{2\pi\Lambda_e^4} \quad (24)$$

at the relevant energies. Performing the thermal average over electron energies gives $\langle E_e^2 \rangle \cong 12.9 T_f^2$, and we find from (23) the limit

$$\Lambda_e \gtrsim 1.9 x_f^{-1/2} Y_{\psi/e}^{-1/4} \Lambda_f \cong 70 \text{ TeV} \quad (25)$$

where $Y_{\psi/e} \cong 3.4 \times 10^{-11}$ is the abundance of 90 GeV DM relative to electrons at T_f .

From the magnetic dipole operator, one has photon-mediated scattering from all charged particles that are in equilibrium at $T_f \sim 3 \text{ GeV}$, which we take to be $f = e, \mu, \tau, u, d, s, c$ plus their antiparticles. The cross section has a logarithmic infrared divergence in the limit $m_f \delta m \rightarrow 0$ from low-angle scattering. For $m_f = m_e$, it is regulated more effectively by Debye screening than by the small value of $m_e \delta m$, giving

$$\begin{aligned} (\sigma v)_{\psi f \rightarrow \chi f} &\cong \frac{Q_f^2 e^2}{\pi\Lambda_\gamma^2} \left(2 \left(\frac{m_\chi^2}{s} - 1 \right) \right. \\ &\quad \left. + \ln \left(1 + \frac{(s - 2m_\chi^2)^2}{s m_D^2} \right) \right) \end{aligned} \quad (26)$$

For simplicity we cut off the divergence for all species using the Debye mass $m_D = (\sum_f Q_f^2 n_f/T)^{1/2} \cong 1.5 e T \cong 1.4 \text{ GeV}$. The thermal average of (26) is $0.13/\Lambda_\gamma^2$ for the parameters of interest. The resulting bound analogous to (25) is

$$\Lambda_\gamma \gtrsim 4 \times 10^9 \text{ GeV} \quad (27)$$

No similar constraint arises for Λ_b since b quarks are not present in the plasma at temperature T_f .

The bounds on Λ_e and Λ_γ are shown in fig. 6. In both cases the limits derived from suppressing inelastic scattering are much stronger than those from suppressing coannihilation processes. This is because the number density

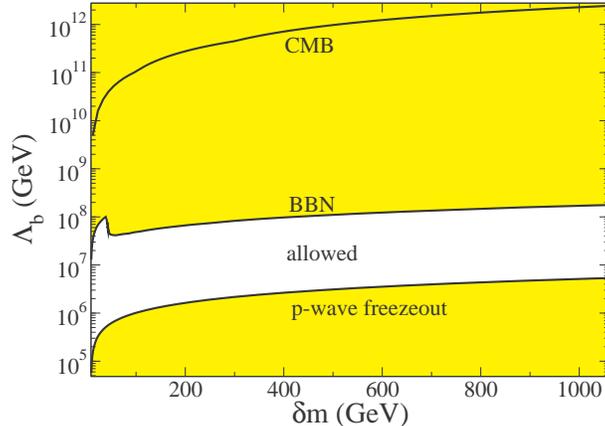


Figure 7: Similar to fig. 6, but for decays $\psi \rightarrow \chi b\bar{b}$.

of relativistic standard model scattering partners is much greater than the Boltzmann-suppressed number density of χ at ψ freeze-out.

4 Constraints on decaying DM

To ensure that ψ decays occur after freeze-out of p -wave annihilations estimated in (16), we assume that $\Gamma < H(T_f)$ for the relevant decay rate, with Hubble parameter $H(T_f) = 1.66\sqrt{g_*}(m_\chi/x_f)^2/M_p$ and $g_* \cong 76$ for $T_f \cong 3$ GeV. Comparing H to the decay rates (18), we obtain constraints on the parameter space in figs. 6,7, shown in the lower regions of the plots. In the unshaded central regions, decays occur after freeze-out and before big bang nucleosynthesis (BBN) or recombination. In the upper shaded regions, decays will disrupt BBN or the cosmic microwave background (CMB), due to the deposition of electromagnetic energy, as well as hadrons in the case of decays to $b\bar{b}$, as we consider in the following subsections.

4.1 BBN constraints

For the first two operators of (17), leading to decays into photons or electrons, only the total energy deposited in the plasma is relevant for photo-production or dissociation of light elements produced by BBN. We take the

combined constraints from ref. [82] (see fig. 8 of that reference). An upper limit on $\zeta \equiv (n_\chi/n_\gamma) \delta m$ as a function of lifetime is derived there, which we convert into a limit on $\Lambda_{\gamma,e}$ as a function of δm , shown in fig. 6 for $m_\chi = 90$ GeV. (The choice of m_χ determines n_χ/n_γ .) Since the limit on ζ is not monotonic in lifetime, BBN excludes a range of $\Lambda_{\gamma,e}$ for a given value of δm .

The third operator of (17) leading to $b\bar{b}$ pairs entails somewhat more stringent constraints because of hadronic interactions that can more efficiently disturb light element abundances [83]. The limits depend not only upon the total amount of energy deposited, but also the energy per decay. By interpolating between the constraints of [83] calculated for different masses of decaying DM, we find the BBN lower limit on Λ_b versus δm shown in fig. 7.² The role of DM mass in that reference (where the DM particle was assumed to decay completely to standard model particles) is played by δm in the present context.

4.2 CMB constraints

For lifetimes $\tau > 10^{12}$ s, electromagnetic energy deposition starts to distort the cosmic microwave background, superseding BBN constraints. We have computed the Planck-projected upper limits on the injected energy fraction $\delta\Omega_\chi/\Omega_\chi = \delta m/m_\chi$ as a function of lifetime using the tools of ref. [84] (see also ref. [85]), where transfer functions $T_{\gamma,e}(z', z, E)$ are provided for computing the efficiency of energy deposition as a function of redshift z for injections of photons or electrons at z' . For $\chi \rightarrow \psi \gamma$, T_γ can be used directly since the spectrum is monochromatic. For $\chi \rightarrow \psi e^+ e^-$, T_e must be convolved with the normalized energy spectrum of electrons from the 3-body decay, which in the limit of $\delta m \ll m_\chi$ takes the form $dN/d \ln x = 60 x^2(1-x)^3$, where $x = E/\delta m$.

² The relevant constraints are inferred from figs. 9-10 of [83], in the region $\tau < 100$ s, which is insensitive to uncertainties in the observed ${}^6\text{Li}/{}^7\text{Li}$ abundance.

Converting the limits on $\delta m/m_\chi$ versus τ into the δm - Λ_e plane results in the excluded regions shown in fig. 6. These extend to lifetimes greater than the age of the universe, not of interest in the present context, since ψ would still be the principal component of the dark matter.

Projected Planck limits on the lifetimes for decays into $b\bar{b}$ have been given in ref. [85] for several DM masses. Interpolating those results we translate them into 95% C.L. limits on Λ_b as a function of δm , shown in fig. 7.

4.3 Direct detection

For $\delta m \lesssim 100$ keV, it is possible to have direct detection through inelastic scattering on nuclei, $\chi N \rightarrow \psi N$. This is relevant for the magnetic dipole operator for which such small mass splittings are in the allowed region of fig. 6. We have roughly indicated the region excluded by direct searches there by taking the scattering rate to scale as $\Lambda_\gamma^{-2} f(v_{\min}) \sim \Lambda_\gamma^{-2} e^{-3v_{\min}^2/2\sigma_v^2}$ where v_{\min} is the minimum velocity for an inelastic transition. It is given in terms of the DM-nucleus reduced mass $\mu_{\chi N}$ as $v_{\min} = \sqrt{2\delta m/\mu_{\chi N}}$. Therefore the experimental limit on Λ_γ takes the form $\Lambda_\gamma > \Lambda_0 e^{-\delta m/\delta m_0}$ for some reference mass splitting, which we estimate to be $\delta m_0 \cong 6.3$ keV by comparison to recent constraints on magnetic inelastic dark matter found in ref. [86]. The coefficient Λ_0 corresponds to the limit from elastic scattering ($\delta m = 0$), which we take to be $\Lambda_0 \cong 10^{14}$ GeV by rescaling the constraints on dipolar dark matter from ref. [87] according to the latest limits from the LUX experiment [88].

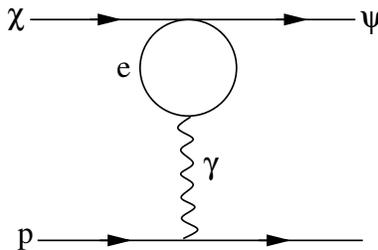


Figure 8: Loop-induced operator leading to inelastic scattering of DM on protons.

In principle, the four-fermion operators in (17) could give rise to inelastic scattering on nucleons, by forming a loop from the electrons or quarks and considering virtual photon exchange between the loop and protons in the nucleus (see fig. 8). However the scattering rate is negligible since the required mass splitting $\delta m > 2m_e$ or $2m_b$ is too large to be excited in direct detection experiments. For smaller δm , there is an electron-loop mediated decay $\psi \rightarrow \chi + 3\gamma$ (decays into 1 and 2 photons are forbidden by gauge invariance or Furry's theorem), but this is too slow to be of interest for $\delta m < 2m_e$, since the lifetime exceeds 10^{12} s and puts the model into the CMB-excluded region.

4.4 Fermi gamma ray line search

The magnetic dipole operator in eq. (17) induces $\chi\bar{\chi}$ annihilation to monochromatic gamma rays through t -channel exchange of a ψ particle. The cross section for this process has been calculated in [81]:

$$\langle\sigma v\rangle = \frac{16m_\chi^4}{\pi\Lambda_\gamma^4} \left(\frac{m_\chi + \delta m}{(m_\chi + \delta m)^2 + m_\chi^2} \right)^2. \quad (28)$$

The Fermi-LAT collaboration has searched for such signals of DM annihilation in the Milky Way halo [89]. In the left plot in figure 6 we show the limits at 95% C.L. on the magnetic dipole operator from their search, assuming that the DM density follows a generalized NFW profile with $\gamma = 1.2$ and corresponding to a region of interest of 3° around the galactic center to maximize the expected signal [89]. The line search limit is $\Lambda_\gamma \gtrsim 8000$ GeV, roughly equivalent to the bound we obtained from coannihilations.

4.5 Collider constraints

The operators we consider are also constrained by collider searches. For the $(\bar{\chi}\gamma^\mu\psi)(\bar{e}\gamma_\mu e)$ operator in eq. (17), the relevant limits come from LEP, where the characteristic signature is missing energy and a photon which is radiated off the initial e^+ or e^- . For our fiducial case of $m_\chi = 90$ GeV,

DELPHI monophoton searches constrain $\Lambda_e \gtrsim 310$ GeV at 90% C.L. [90]. For the magnetic dipole operator, the current most stringent constraint is from LHC monojet searches [91] requiring $\Lambda_\gamma \gtrsim 280$ GeV at 95% C.L., a limit which is only slightly more constraining than searches for monophotons at LEP [92] or the LHC [91]. The collider-disfavored region for the magnetic moment operator is more strongly excluded in our scenario by direct detection and the ψ lifetime constraints. The $(\bar{\chi}\gamma^\mu\psi)(\bar{b}\gamma_\mu b)$ operator is in principle limited by LHC monojet searches, but the small b -quark content of the proton makes such limits very weak.

All of the exclusions discussed here were derived under the assumption that e^+e^- or pp collisions lead to stable final-state dark sector particles. Although it is possible for the ψ produced in these collisions to decay inside the detector, in the regions of parameter space for which the collider limits are relevant, the mass splitting is so small that the softness of the decay products would render them undetectable.

For the operator of eq. (19), the relevant limits are again from LEP monophoton searches when $f = e$ and LHC monojet searches when $f = q$. Limits from an ATLAS monojet search [93] as well as that from the previously mentioned DELPHI monophoton search are shown in fig. 5. In either case the correct relic density is compatible with current collider limits.

5 Mediator couplings

A large p -wave cross section would generically run afoul of direct and indirect detection constraints if the DM χ coupled directly to SM particles. On the other hand, annihilation to light mediators ϕ , that subsequently decay into SM particles, can avoid this problem. If χ couples to ϕ as $g\phi\bar{\chi}\chi$, the resulting p -wave cross section at low velocities is given by

$$\sigma v \cong \frac{3v^2 g^4}{32\pi m_\chi^2} \quad (29)$$

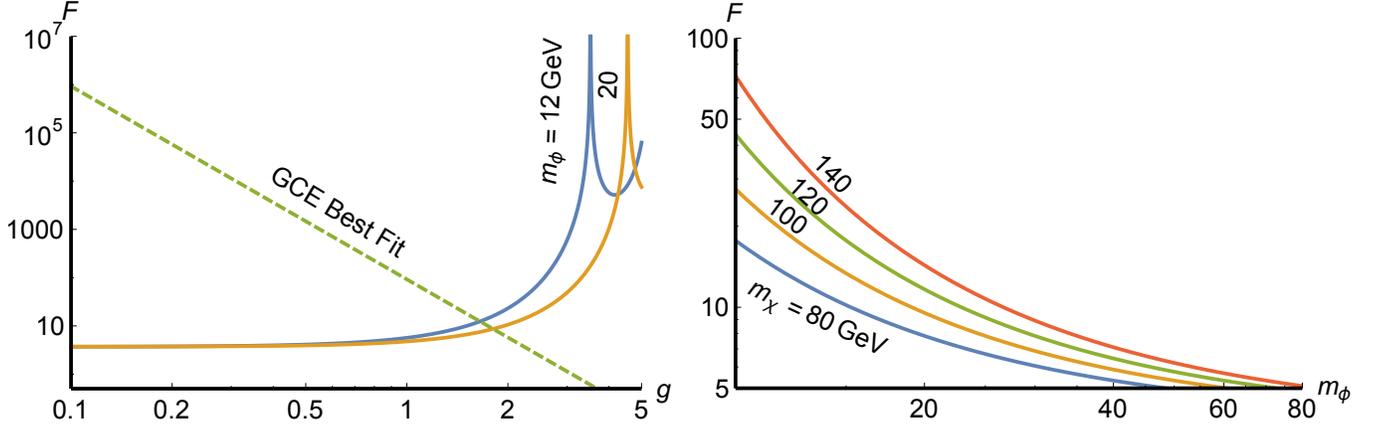


Figure 9: Left: Phase-space averaged enhancement factor F versus g for $m_\chi = 80 \text{ GeV}$ and $\sigma_v = 250 \text{ km s}^{-1}$, representative of the Milky Way. The corresponding result for $\sigma_v = 10 \text{ km s}^{-1}$, appropriate for dwarf spheroidals, looks very similar. The solid lines show $F(g)$ with for the cases $m_\phi = 12 \text{ GeV}$ and $m_\phi = 20 \text{ GeV}$. The dashed line is the value of $F(g)$ required to give a sufficiently large annihilation cross-section for the GCE. Right: Dependence of F on m_ϕ for a fixed value of $g = 0.8$ (the intersection point in the top left panel) with masses $m_\chi = 80, 100, 120,$ and 140 GeV .

An uncomfortably large coupling $g \sim 3.7$ would be needed to match the fit to the GC excess.

However smaller values of g can be sufficient if the cross section is Sommerfeld enhanced, which can naturally occur if the mediator ϕ is light. Defining $\alpha_g = g^2/4\pi$, an analytic approximation to the enhancement factor is given by [94]

$$S_l \cong \left| \frac{\Gamma(a^+) \Gamma(a^-)}{\Gamma(1+l+2iw)} \right|^2. \quad (30)$$

for partial wave l scattering, where

$$\begin{aligned} a^\pm &= 1 + l + iw \left(1 \pm \sqrt{1 - x/w} \right) \\ x &= \frac{\alpha_g}{\beta}, \quad w = \frac{6\beta m_\chi}{\pi^2 m_\phi} \end{aligned} \quad (31)$$

with velocity $v = \beta c$ in the center of mass frame. For a p -wave process we take $l = 1$. S_1 is nonvanishing in the limit $v \rightarrow 0$, so that velocity suppression of

the p -wave cross section is still present despite the enhancement, and S_1 has quasiperiodic resonant behavior as a function of α_g .

The enhancement factor depends on the relative velocity of the particles, which in principle must be averaged over phase space. Ignoring the radial dependence of the annihilation cross-section, we can find an estimate of the average enhancement, which is given by

$$F(g) = \frac{\langle S_1 \sigma v_{\text{rel}} \rangle}{\langle \sigma v_{\text{rel}} \rangle} \quad (32)$$

using eq. (1). However for the parameter values of interest, we find that the dependence on v is very weak and one can simply use ~ 10 - 1000 km/s with negligible error from omitting the average. To match the desired value of the cross section in (10), we need

$$g^4 F(g) = 144 \left(\frac{m_\chi}{100 \text{ GeV}} \right)^2 \quad (33)$$

This relation as a function of g is shown as the dashed line in fig. 9, while the analytic approximation for F is indicated by the curves for two different values of the mediator mass.

By comparing $F(g)$ to the required cross section, eq. (10), we find that the coupling constant can be reduced to a more comfortable value of $g \cong 2$. It can be somewhat further reduced by taking larger values of m_χ/m_ϕ , as can be seen in figure 9; the left panel shows $F(g)$ decreasing with m_ϕ , while the right shows $F(g)$ increasing with m_χ . It was recently pointed out [95] that approximations to the enhancement factor such as (30) may fail to satisfy partial wave unitarity in the resonant regions. We have checked that we are very far from any such violation however, for the parameter values of interest.

Finally, it has been noted in [96] that it is possible for two DM particles to capture into a bound state and then annihilate to mediators. The bound state formation process dominantly occurs in the s -wave, so, if possible, it dominates over the direct p -wave annihilation to mediators. In forming a bound state a mediator is emitted, so the mass of the mediator must be less than the binding

energy of the ground state for this to occur, i.e.

$$m_\phi \leq \frac{g^4 m_\chi}{64\pi^2}. \quad (34)$$

For the values of g and m_χ that we consider to explain the GC excess, $m_\phi \lesssim 2.2$ GeV for a bound state to form. In this work we have only considered mediator masses above this limit.

6 Conclusion

We have presented a scenario in which p -wave annihilating dark matter could have significant indirect signals from the galactic center despite having a velocity-suppressed cross section. Although our immediate motivation was to reconcile a dark matter interpretation of the observed GC gamma ray excess with conflicting constraints from dwarf spheroidal galaxies, the framework presented here could be of more general interest.

Our key idea is to assume that the current generation of p -wave annihilating DM χ is the decay product of a metastable predecessor particle ψ , which has a thermal origin and decays after p -wave annihilations of the stable DM have frozen out. This allows a large range of lifetimes that depend upon the ψ - χ mass splitting δm and the mass scale Λ of the effective interaction which leads to the decay. A number of constraints must be satisfied, including those coming from BBN, the CMB, direct detection, photon line searches, mono-X searches at colliders, and $\psi f \rightarrow \chi f$ scattering in the early universe (which could deplete the DM abundance). They depend strongly upon the nature of the decays, which we have illustrated using the examples of $\psi \rightarrow \chi e^+ e^-$, $\psi \rightarrow \chi \gamma$, and $\psi \rightarrow \chi b \bar{b}$, but in all cases there is a significant region of the δm - Λ parameter space in which all of the requirements can be satisfied.

The annihilation cross sections of interest for explaining the galactic center excess are larger than would generically occur in the presence of p -wave

suppression because of the low DM velocity in the GC. We nevertheless demonstrated a working example using reasonable coupling strengths, where the DM annihilates into light scalar mediators that mix with the Higgs boson and subsequently decay into b quarks. We have shown that such models give a reasonably good fit to the observed GCE, while satisfying constraints from dwarf spheroidals by a comfortable margin.

Collider tests of our scenario are currently weaker than the consistency requirement that inelastic $\psi f \rightarrow \chi f$ scatterings on standard model particles do not deplete the DM relic density in the early universe (due to strong p -wave annihilations of χ). For a narrow window of mass splittings $\delta m \sim 0.1$ MeV, direct SM searches provide a possible means of detection in the case of magnetic inelastic transitions.

Fermi observations of gamma rays from galaxy clusters may provide a more sensitive test of our scenario, due to the large velocity dispersion in clusters. We have shown that limits on DM annihilation from the Virgo cluster, while significantly stronger than limits from dwarf spheroidals, still are far from being in tension with this interpretation of the GCE. We hope that our work will motivate further studies of limits on DM annihilation in the Coma cluster, which has the potential to be more constraining because of its relatively high velocity dispersion.

Acknowledgments. We thank Kimmo Kainulainen, Maxim Pospelov, Subir Sarkar, Tracy Slatyer, Alex Drlica-Wagner and Wei Xue for helpful discussions or correspondence. We acknowledge support of the McGill Space Institute and the Natural Sciences and Engineering Research Council of Canada. We thank Niels Bohr International Academy for its generous hospitality during the initial stages of this work.

REFERENCES

- [1] L. Goodenough and D. Hooper, (2009), 0910.2998.
- [2] D. Hooper and L. Goodenough, Phys. Lett. **B697**, 412 (2011), 1010.2752.
- [3] D. Hooper and T. Linden, Phys. Rev. **D84**, 123005 (2011), 1110.0006.
- [4] K. N. Abazajian and M. Kaplinghat, Phys. Rev. **D86**, 083511 (2012), 1207.6047, [Erratum: Phys. Rev.D87,129902(2013)].
- [5] B. Zhou *et al.*, Phys. Rev. **D91**, 123010 (2015), 1406.6948.
- [6] F. Calore, I. Cholis, and C. Weniger, JCAP **1503**, 038 (2015), 1409.0042.
- [7] T. Daylan *et al.*, Phys. Dark Univ. **12**, 1 (2016), 1402.6703.
- [8] S. Murgia, presented at Fifth Fermi Symposium, 20- 24 Oct. 2014, <http://fermi.gsfc.nasa.gov/science/mtgs/symposia/2014/program/08Murgia.pdf>.
- [9] Fermi-LAT, M. Ajello *et al.*, (2015), 1511.02938.
- [10] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, (2015), 1506.05124.
- [11] R. Bartels, S. Krishnamurthy, and C. Weniger, (2015), 1506.05104.
- [12] R. M. O’Leary, M. D. Kistler, M. Kerr, and J. Dexter, (2015), 1504.02477.
- [13] T. Linden, Phys. Rev. **D93**, 063003 (2016), 1509.02928.
- [14] I. Cholis, D. Hooper, and T. Linden, JCAP **1506**, 043 (2015), 1407.5625.
- [15] E. Carlson and S. Profumo, Phys. Rev. **D90**, 023015 (2014), 1405.7685.
- [16] J. Petrović, P. D. Serpico, and G. Zaharijaš, JCAP **1410**, 052 (2014), 1405.7928.
- [17] I. Cholis *et al.*, JCAP **1512**, 005 (2015), 1506.05119.

- [18] D. Gaggero, M. Taoso, A. Urbano, M. Valli, and P. Ullio, (2015), 1507.06129.
- [19] W. de Boer, I. Gebauer, S. Kunz, and A. Neumann, (2015), 1509.05310.
- [20] E. Carlson, T. Linden, and S. Profumo, (2016), 1603.06584.
- [21] Q. Yuan and K. Ioka, *Astrophys. J.* **802**, 124 (2015), 1411.4363.
- [22] S. K. Lee, M. Lisanti, and B. R. Safdi, *JCAP* **1505**, 056 (2015), 1412.6099.
- [23] Fermi-LAT, M. Ackermann *et al.*, *Phys. Rev. Lett.* **115**, 231301 (2015), 1503.02641.
- [24] K. N. Abazajian and R. E. Keeley, (2015), 1510.06424.
- [25] DES, K. Bechtol *et al.*, *Astrophys. J.* **807**, 50 (2015), 1503.02584.
- [26] DES, A. Drlica-Wagner *et al.*, *Astrophys. J.* **813**, 109 (2015), 1508.03622.
- [27] A. Geringer-Sameth *et al.*, *Phys. Rev. Lett.* **115**, 081101 (2015), 1503.02320.
- [28] D. Hooper and T. Linden, *JCAP* **1509**, 016 (2015), 1503.06209.
- [29] S. Li *et al.*, *Phys. Rev.* **D93**, 043518 (2016), 1511.09252.
- [30] DES, Fermi-LAT, A. Drlica-Wagner *et al.*, *Astrophys. J.* **809**, L4 (2015), 1503.02632.
- [31] M. R. Buckley *et al.*, *Phys. Rev.* **D91**, 102001 (2015), 1502.01020.
- [32] B. Bertoni, D. Hooper, and T. Linden, *JCAP* **1512**, 035 (2015), 1504.02087.
- [33] J. Liu, N. Weiner, and W. Xue, *JHEP* **08**, 050 (2015), 1412.1485.
- [34] F. Calore, I. Cholis, C. McCabe, and C. Weniger, *Phys. Rev.* **D91**, 063003 (2015), 1411.4647.
- [35] M. Kaplinghat, T. Linden, and H.-B. Yu, *Phys. Rev. Lett.* **114**, 211303 (2015), 1501.03507.
- [36] E. Hardy, R. Lasenby, and J. Unwin, *JHEP* **07**, 049 (2014), 1402.4500.
- [37] Y. Zhao, X.-J. Bi, H.-Y. Jia, P.-F. Yin, and F.-R. Zhu, (2016), 1601.02181.

- [38] M. Abdullah *et al.*, Phys. Rev. **D90**, 035004 (2014), 1404.6528.
- [39] G. Elor, N. L. Rodd, and T. R. Slatyer, Phys. Rev. **D91**, 103531 (2015), 1503.01773.
- [40] B. Robertson and A. Zentner, Phys. Rev. **D79**, 083525 (2009), 0902.0362.
- [41] M. G. Walker *et al.*, Astrophys. J. **667**, L53 (2007), 0708.0010.
- [42] G. Battaglia *et al.*, Mon. Not. Roy. Astron. Soc. **364**, 433 (2005), astro-ph/0506102, [Erratum: Mon. Not. Roy. Astron. Soc.370,1055(2006)].
- [43] W. Dehnen, D. McLaughlin, and J. Sachania, Mon. Not. Roy. Astron. Soc. **369**, 1688 (2006), astro-ph/0603825.
- [44] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996), astro-ph/9508025.
- [45] E. Romano-Diaz, I. Shlosman, Y. Hoffman, and C. Heller, Astrophys. J. **685**, L105 (2008), 0808.0195.
- [46] F. Chen, J. M. Cline, A. Fradette, A. R. Frey, and C. Rabideau, Phys. Rev. **D81**, 043523 (2010), 0911.2222.
- [47] M. Cirelli and J. M. Cline, Phys. Rev. **D82**, 023503 (2010), 1005.1779.
- [48] Fermi-LAT, M. Ackermann *et al.*, Phys. Rev. **D89**, 042001 (2014), 1310.0828.
- [49] M. I. Wilkinson *et al.*, (2006), astro-ph/0602186, [EAS Publ. Ser.20,105(2006)].
- [50] M. I. Wilkinson *et al.*, Astrophys. J. **611**, L21 (2004), astro-ph/0406520.
- [51] M. G. Walker *et al.*, Astron. J. **131**, 2114 (2006), astro-ph/0511465, [Erratum: Astron. J.132,968(2006)].
- [52] A. Koch *et al.*, Astrophys. J. **657**, 241 (2007), astro-ph/0611372.
- [53] M. G. Coleman, K. Jordi, H.-W. Rix, E. K. Grebel, and A. Koch, Astron. J. **134**, 1938 (2007), 0708.1853.
- [54] K. B. Westfall *et al.*, Astron. J. **131**, 375 (2006), astro-ph/0508091.

- [55] V. Belokurov *et al.*, *Astrophys. J.* **647**, L111 (2006), astro-ph/0604355.
- [56] J. D. Simon and M. Geha, *Astrophys. J.* **670**, 313 (2007), 0706.0516.
- [57] M. Geha *et al.*, *Astrophys. J.* **692**, 1464 (2009), 0809.2781.
- [58] B. Willman *et al.*, *Astron. J.* **142**, 128 (2011), 1007.3499.
- [59] A. Cuoco, B. Eiteneuer, J. Heisig, and M. Krämer, (2016), 1603.08228.
- [60] C. Evoli, D. Gaggero, D. Grasso, and L. Maccione, *JCAP* **0810**, 018 (2008), 0807.4730.
- [61] A. W. Strong, I. V. Moskalenko, and O. Reimer, *Astrophys. J.* **537**, 763 (2000), astro-ph/9811296, [Erratum: *Astrophys. J.*541,1109(2000)].
- [62] G. Di Bernardo, C. Evoli, D. Gaggero, D. Grasso, and L. Maccione, *JCAP* **1303**, 036 (2013), 1210.4546.
- [63] M. Cirelli *et al.*, *JCAP* **1103**, 051 (2011), 1012.4515, [Erratum: *JCAP*1210,E01(2012)].
- [64] P. Ciafaloni *et al.*, *JCAP* **1103**, 019 (2011), 1009.0224.
- [65] P. Agrawal, B. Batell, P. J. Fox, and R. Harnik, *JCAP* **1505**, 011 (2015), 1411.2592.
- [66] J. M. Cline, G. Dupuis, Z. Liu, and W. Xue, *Phys. Rev.* **D91**, 115010 (2015), 1503.08213.
- [67] R. Diamanti, L. Lopez-Honorez, O. Mena, S. Palomares-Ruiz, and A. C. Vincent, *JCAP* **1402**, 017 (2014), 1308.2578.
- [68] Fermi-LAT, M. Ackermann *et al.*, *Astrophys. J.* **819**, 149 (2016), 1507.08995.
- [69] Fermi-LAT, M. Ackermann *et al.*, *Astrophys. J.* **812**, 159 (2015), 1510.00004.
- [70] M. Girardi *et al.*, *Astrophys. J.* **457**, 61 (1996), astro-ph/9507031.
- [71] H.-Y. Wu, O. Hahn, R. H. Wechsler, Y.-Y. Mao, and P. S. Behroozi, *Astrophys. J.* **763**, 70 (2013), 1209.3309.

- [72] H.-Y. Wu, O. Hahn, R. H. Wechsler, P. S. Behroozi, and Y.-Y. Mao, *Astrophys. J.* **767**, 23 (2013), 1210.6358, [*Astrophys. J.*767,23(2013)].
- [73] M. Ackermann *et al.*, *JCAP* **1005**, 025 (2010), 1002.2239.
- [74] S. Ando and D. Nagai, *JCAP* **1207**, 017 (2012), 1201.0753.
- [75] M. J. Drinkwater, M. D. Gregg, and M. Colless, *Astrophys. J.* **548**, L139 (2001), astro-ph/0012415.
- [76] W. Liu, X.-J. Bi, S.-J. Lin, and P.-F. Yin, (2016), 1602.01012.
- [77] E. W. Kolb and M. S. Turner, *Front. Phys.* **69**, 1 (1990).
- [78] Planck, P. A. R. Ade *et al.*, (2015), 1502.01589.
- [79] J. L. Feng, A. Rajaraman, and F. Takayama, *Phys. Rev. Lett.* **91**, 011302 (2003), hep-ph/0302215.
- [80] K. Griest and D. Seckel, *Phys. Rev.* **D43**, 3191 (1991).
- [81] N. Weiner and I. Yavin, *Phys. Rev.* **D86**, 075021 (2012), 1206.2910.
- [82] R. H. Cyburt, J. R. Ellis, B. D. Fields, and K. A. Olive, *Phys. Rev.* **D67**, 103521 (2003), astro-ph/0211258.
- [83] K. Jedamzik, *Phys. Rev.* **D74**, 103509 (2006), hep-ph/0604251.
- [84] T. R. Slatyer, *Phys. Rev.* **D87**, 123513 (2013), 1211.0283.
- [85] J. M. Cline and P. Scott, *JCAP* **1303**, 044 (2013), 1301.5908, [Erratum: *JCAP*1305,E01(2013)].
- [86] G. Barello, S. Chang, and C. A. Newby, *Phys. Rev.* **D90**, 094027 (2014), 1409.0536.
- [87] K. Sigurdson, M. Doran, A. Kurylov, R. R. Caldwell, and M. Kamionkowski, *Phys. Rev.* **D70**, 083501 (2004), astro-ph/0406355, [Erratum: *Phys. Rev.*D73,089903(2006)].
- [88] LUX, D. S. Akerib *et al.*, (2015), 1512.03506.
- [89] Fermi-LAT, M. Ackermann *et al.*, *Phys. Rev.* **D91**, 122002 (2015), 1506.00013.

- [90] P. J. Fox, R. Harnik, J. Kopp, and Y. Tsai, Phys. Rev. **D84**, 014028 (2011), 1103.0240.
- [91] V. Barger, W.-Y. Keung, D. Marfatia, and P.-Y. Tseng, Phys. Lett. **B717**, 219 (2012), 1206.0640.
- [92] J.-F. Fortin and T. M. P. Tait, Phys. Rev. **D85**, 063506 (2012), 1103.3289.
- [93] ATLAS, G. Aad *et al.*, Eur. Phys. J. **C75**, 299 (2015), 1502.01518, [Erratum: Eur. Phys. J.C75,no.9,408(2015)].
- [94] S. Cassel, J. Phys. **G37**, 105009 (2010), 0903.5307.
- [95] K. Blum, R. Sato, and T. R. Slatyer, (2016), 1603.01383.
- [96] H. An, M. B. Wise, and Y. Zhang, Phys. Lett. **B773**, 121 (2017), 1606.02305.

3.2.1 Addendum:

The following clarifications to the paper should be added:

- In Table 1, there is a typo in the uncertainties for the J -factors of Fornax and Leo I. The entry for Fornax should read 18.2 ± 0.21 , and that for Leo I should read 17.7 ± 0.18 . The correct values were used in the analysis, and this typo has no bearing on the results.

3.3 The Effect of Dark Matter Density Profiles on The GCE

Unlike the previous paper, in which we considered a model of dark matter in which the cross section would vary depending on the size of the halo, in the following paper I considered the effect of having the density profile vary depending on the size of the halo. Given that the annihilation rate depends on the square of the dark matter density, the strength of a signal depends not only on the average density but also on the specific density profile. A halo in which the dark matter is strongly concentrated in the center (a ‘cuspy’ profile) will have an enhanced signal relative to a ‘cored’ profile in which the central density is less sharply peaked.

In the paper I consider two possible dark matter halo profiles for the both the Milky Way and Dwarf Spheroidals — the generalized Navarro-Frenk-White profile (NFW) [85] and the Einasto profile [208]:

$$\rho_{\text{NFW}}(r) = \frac{\rho_s}{\left(\frac{r}{R_s}\right)^\gamma \left(1 + \frac{r}{R_s}\right)^{3-\gamma}} \quad (3.1)$$

$$\rho_{\text{Einasto}}(r) = \rho_s e^{-\frac{2}{\alpha} \left(\left(\frac{r}{R_s}\right)^\alpha - 1\right)}. \quad (3.2)$$

These profiles have the advantage of each having a single parameter which determines how cuspy or cored the profile is (γ for the NFW and α for the Einasto profile). The logic of the paper is thus: If we assume the Galactic Center excess is in fact a signal from annihilating dark matter, then for a given

value of γ_{MW} (or α_{MW}) for the Milky Way we can determine how cored the dwarf spheroidal galaxies must be in order to produce no detectable signal, placing bounds on γ_{dsph} (or α_{dsph}). Alternatively, we determine how cored dwarf spheroidal profiles must be in order to be consistent with the Galactic Center excess.

This strength of the signal is characterized by the J -factor of the dark matter halo, which encapsulates the astrophysical component of the strength of a signal from dark matter annihilation:

$$J = \int_{\Delta\Omega} \int_{\text{l.o.s.}} \rho^2 dl d\Omega. \quad (3.3)$$

Since the J -factor is dependent on the values ρ_s and R_s , the characteristic density and radius, these need to be determined for each dwarf spheroidal galaxy. These values are determined using stellar kinematic data, and are dependent upon the choice of profile and on γ or α . As all previous work assumes a specific profile when determining these values (usually $\gamma = 1$), I determined these values from stellar kinematic data for a range of values of γ and α .

These results are used to place bounds on γ_{dsph} and α_{dsph} , showing that if the Galactic Center excess is a signal from annihilating dark matter, dwarf galaxies are generally constrained to be more cored than the halo of the Milky Way.

This has interesting implications for the nature of dark matter should the Galactic Center excess prove to be a real dark matter signal. First, it would favour strongly cored profiles in dwarf galaxies which, as discussed in the paper, would lend further credence to models of self-interacting dark matter. It also suggests that larger halos such as that of the Milky Way are less heavily cored than those of dwarf spheroidals. This is predicted by models of warm dark

matter, such as sterile neutrinos, in which density fluctuations during galaxy formation are smoothed out on scales below the free streaming length [114, 115]. It is also predicted in models of SIDM in which the scattering cross section is inversely proportional to the velocity, such as those in which dark matter interacts via a light mediator [137], as this causes the self-interaction cross section to be enhanced in dwarf galaxies relative to that of the Milky Way.

This manuscript was published in Physical Review D in 2018 [209].

Constraining Dwarf Spheroidal Dark Matter Halos With the Galactic Center Excess

Jeremie Choquette^{*1}

¹*Department of Physics, McGill University, 3600 Rue University, Montréal,
Québec, Canada H3A 2T8*

Abstract

If the gamma-ray excess from the Galactic Center reported by Fermi-LAT is a signal from annihilating dark matter, one must question why a similar excess has not been observed in dwarf spheroidal galaxies. We use this observation to place constraints on the density profile of dwarf spheroidal galaxies under the assumption that the Galactic Center excess is in fact a signal from annihilating dark matter. We place constraints on the generalized Navarro-Frenk-White (NFW) parameter γ and the Einasto profile parameter α which control the logarithmic slope of the inner regions of the halo's density profile. The best-fit halo parameters R_s and ρ_s are determined using stellar kinematic data for a range of γ and α . We determine that under these assumptions the Galactic Center excess is inconsistent with the standard NFW profile (and other “cuspy” profiles) for dwarf spheroidal galaxies, but is consistent with observations of cored dwarf galaxy profiles. Specifically, we find that dwarf spheroidal profiles must be less cuspy than that of the Milky Way. Models of dark matter which self-interacts through a light mediator can achieve this.

1 Introduction

Observations by Fermi-LAT have indicated an excess of gamma-rays in the center of the Milky Way galaxy in the range of a few GeV [1–9]. Interpretations of the galactic center excess (GCE) differ, with likely candidates including dark matter annihilations and known astrophysical phenomena. On the astrophysical side, the spectrum and morphology of the signal from millisecond pulsars provides a good fit to the observed excess [10–12], but this would require a much greater number of millisecond pulsars than are observed [13, 14]. The Fermi-LAT collaboration has more recently completed an analysis of the purported signal and has concluded that the morphology of the signal is more consistent with millisecond pulsars than with the dark matter interpretation [15, 16]. It is concluded that the dark matter interpretation is strongly disfavoured relative to other interpretations of the excess. In a recent paper, however, Haggard et al. argue that a sufficiently large population of millisecond pulsars would also imply a large population of observable low-mass X-ray binaries, limiting the contribution of millisecond pulsars to the galactic center excess to $\sim 4\% - 23\%$ [17], though it should be noted that these results are predicated on the assumption that the relative populations of low-mass X-ray binaries and millisecond pulsars are the same in the inner galaxy as in globular clusters. This nonetheless makes annihilating dark matter an interesting possibility.

It is also well known that there is tension between dark matter explanations of the galactic center excess and observations of dwarf spheroidal galaxies. Dwarf spheroidal galaxies show no corresponding signal, with the constraints seeming to exclude dark matter annihilation as a viable explanation for the galactic center excess[18, 19]. The analysis of [18] (upon which [19] is based), however, assumes a Navarro-Frenk-White (NFW) profile for the dwarf spheroidals. The NFW profile has a sharp cusp at the center, leading to an

enhanced signal relative to more ‘cored’ dark matter distributions. We consider two profiles here: the generalized NFW profile and the Einasto profile, defined in equations (3) and (4) respectively.

The exact distribution of dark matter in dwarf galaxies is not well known, but there is a large body of evidence pointing to cored profiles (see section 4), or profiles with inner radii with slope smaller than the $\rho \propto r^{-1}$ predicted by cold dark matter simulations and exemplified by the NFW profile.

The logarithmic slope of the inner dark matter halo can have a significant impact on its J -factor, a measure of the rate of dark matter annihilations within the halo. We will show that the tension between the dwarf galaxy observations and the GCE are reduced when considering more cored profiles. As the tension is moderate to begin with, it can be erased entirely for sufficiently cored profiles.

It also follows, therefore, that if the GCE signal were assumed to indeed originate from dark matter annihilations, constraints could be placed on the central slope of the dark matter profiles of the dwarf spheroidals. In section 2 we simulate the GCE signal from dark matter to find best fit values for the dark matter mass and annihilation cross section. In section 3 we use these adopted values to place limits on the parameters γ and α which control how cuspy the dwarf spheroidals are. In section 4 we compare these values to those found through observation of dwarf spheroidals and simulations of cold dark matter (CDM) halos. In section 5 we discuss the implications for the CDM paradigm, should the GCE prove to indeed be a signal from annihilating dark matter.

2 Simulation of Signal

It has been shown that the observed gamma ray excess is well fit by models of annihilating dark matter in which the dark matter predominantly annihilates to $b\bar{b}$. The signal, however, consists of multiple components: the

prompt gamma rays (from the b decay products), inverse Compton scattering (ICS, caused by the upscattering of starlight and CMB photons by the e^+/e^- produced as b decay products) and a small amount of bremsstrahlung radiation (also from the decay products). These three sources combine to produce the total signal.

The prompt signal is easiest to compute numerically, as it depends only on the J -factor and average spectrum from a single annihilation, taken from PPPC 4 [20, 21]:

$$\frac{d\Phi_{\text{prompt}}}{dE} = \frac{\langle\sigma v\rangle}{8\pi m_\chi^2} \frac{dN_\gamma}{dE} \times J, \quad (1)$$

$$J = \int_{\Delta\Omega} \int_{\text{l.o.s.}} \rho^2 dl d\Omega, \quad (2)$$

with the integral along the line of sight and angular extent of the observed system. The J -factor can then be computed numerically by assuming a density profile for the dark matter halo.

One way to parametrize the cusiness of a galaxy is through the inner slope of the profile. If we assume a generalized NFW profile:

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{R_s}\right)^\gamma \left(1 + \frac{r}{R_s}\right)^{3-\gamma}}, \quad (3)$$

then the parameter γ corresponds to the negative slope at $r = 0$. Larger values of γ correspond to a more cuspy profile, whereas smaller values correspond to a more cored profile. Following [6] we choose a generalized NFW profile with $R_s = 20$ kpc and $\rho_\odot = 0.40 \text{ GeVcm}^{-31}$ (the local dark matter density, which for $\gamma = 1$ corresponds to a scale density of $\rho_s = 0.26 \text{ GeVcm}^{-3}$). γ is typically

¹ Measurements of the local dark matter density vary greatly, but tend to range from $0.2 - 0.5 \text{ GeVcm}^{-3}$ [22]. Some analyses, however, indicate even larger values of up to $0.5 - 0.7 \text{ GeVcm}^{-3}$ [23]. We adopt the value $\rho_\odot = 0.40 \text{ GeVcm}^{-3}$ in part to ease comparison with the results of Calore *et al.*[6]. Adopting a larger value for the local dark matter density would have the effect of easing the limits on the dwarf spheroidal galaxies by decreasing

taken to be somewhere on the order of $1.0 - 1.5$, with $\gamma = 1.0$ corresponding to the classic NFW profile, but in our analysis we allow it to vary from $0.1 - 1.4$.

Another popular profile that is easily parametrized in terms of the inner slope is the Einasto profile:

$$\rho(r) = \rho_s e^{-\frac{2}{\alpha} \left(\frac{r}{R_s} \right)^\alpha - 1}. \quad (4)$$

The proportionality constant is chosen maintain the same slope and density at R_s as the NFW profile. Although the parameter α does not exactly correspond to the inner log slope, it does control the extent to which the profile is concentrated toward the center, with greater concentrations at smaller α . We therefore consider both Einasto and NFW profiles in our analysis, using γ and α to control how cuspy the profile is.

The ICS and bremsstrahlung components, particularly the ICS, have previously been found to contribute significantly to the signal, dominating it at lower energies ($E \lesssim 1 \text{ GeV}$)[6]. This is especially true for gamma rays originating near or in the galactic disk[7, 8]. This is due to the b -quark products decaying to high energy electrons and positrons, which in turn propagate through the interstellar medium and upscatter photons into the GeV range.

For the ICS and bremsstrahlung predictions, we use simulations to account for the propagation of decay products through the Milky Way and the distribution of gas and photons. We use the DRAGON code [24] to simulate the injection and propagation of high energy electrons from DM annihilation, and the GammaSky program to compute the ICS and bremsstrahlung contributions resulting from these cosmic rays. GammaSky is as yet unreleased, though some results have been given [25]. GammaSky implements GALPROP in the calculation of photon production and upscattering along the line of sight.

the best-fit annihilation cross section required to produce the purported GCE signal.

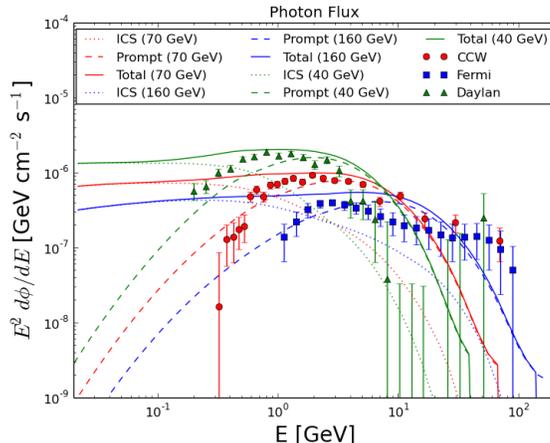


Figure 1: Example of simulated GCE signal (NFW profile, $\gamma_{\text{MW}} = 1$) compared to that observed by [6] (red), [7] (green), and [8] (blue). The simulated signal is shown for the individual best fit values in table 1. The spectra are masked to include only the region of interest considered in each dataset ($2^\circ < |b| < 20^\circ$ and $|l| < 20^\circ$, $1^\circ < |b| < 20^\circ$ and $|l| < 20^\circ$, and $15^\circ \times 15^\circ$ respectively).

The magnitude of the ICS component is highly model-dependent. In the interest of consistency with previous work, we use the model parameters — describing the galactic magnetic field strength and shape and the galactic diffusion model used to compute the resulting inverse Compton scattering rates — adopted in [6], labelled Model F, which is found to perform particularly well in explaining the GCE signal. This results in an ICS component of the same magnitude as that found by the authors. We compare the results for a range of dark matter masses ($20 \text{ GeV} \leq m_\chi \leq 200 \text{ GeV}$) to the GCE signals estimated in [6–8], as shown in 1.

We compare our predicted spectra to those observed in Refs. [6–8] by minimizing the χ^2 in the $\langle\sigma v\rangle$ - m_χ plane to determine the best fit values for both. For the first reference, [6], we use the full covariance matrix; for the other two datasets the published fluxes and errorbars are used. Our results agree with those found in the original references [6, 7], and in Refs. [9, 19] for each dataset, though we find the best fit mass to be slightly higher than Calore *et al.* (70 GeV in our analysis, as opposed to 50 GeV). All find a best-fit cross section of approximately $\langle\sigma v\rangle = 1.7 \times 10^{-26} \text{ cm}^3\text{s}^{-1}$.

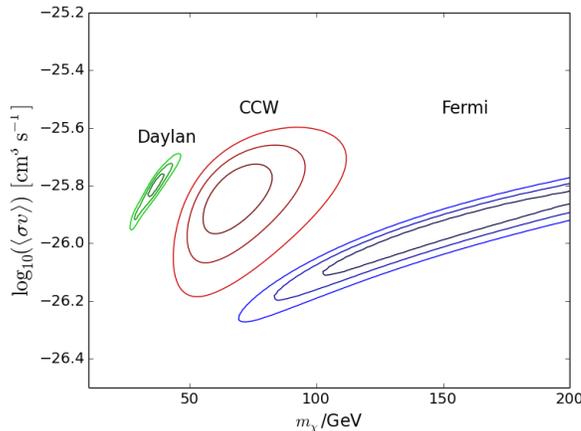


Figure 2: Example of best fit χ^2 1σ , 2σ and 3σ contours for [6] (red), [7] (green), and [8] (blue). This example is for an NFW profile, $\gamma_{\text{MW}} = 1$.

Figure 2 shows the best fit regions for $\gamma_{\text{MW}} = 1$, showing the $1 - \sigma$, $2 - \sigma$, and $3 - \sigma$ confidence intervals generated by minimizing the χ^2 and creating contours at $\chi_{\text{min}}^2 + 2.30$, $+ 6.18$, and $+11.93$. This gives us our best fit values which we will adopt when placing limits on the dwarf galaxy profiles. An example, for $\gamma_{\text{MW}} = 1$, is shown in table 1.

Dataset	$\langle\sigma v\rangle$ [cm^3s^{-1}]	m_χ [GeV]
CCW	1.5×10^{-26}	70
Fermi	1.3×10^{-26}	160
Daylan	1.7×10^{-26}	40

Table 1: Best fit values found for $\gamma_{\text{MW}} = 1$.

3 The Dwarf Spheroidal J Factors

Given the assumption that the GCE signal is indeed the result of annihilating dark matter, our adopted values can be used to place constraints on the density profiles of dwarf spheroidal galaxies. We once again assume an NFW or Einasto profile, allowing the parameters γ_{dpsph} and α_{dpsph} to range from $0.1 - 1.2$ and $0.01 - 1.0$ respectively. In all cases the region of interest $\Delta\Omega$

is taken to be the entire sky in the integration. Observations of the gamma ray flux from the dwarf spheroidals by Fermi-LAT use a $10^\circ \times 10^\circ$ region of interest[26]. Due to the small angular size of dwarf spheroidals (with an angular size less than 1 degree) this is large enough to be indistinguishable from the entire sky.

The exact halo parameters R_s and ρ_s of the dwarf spheroidals are not well known for either profile. Given the difficulty of measuring a large enough population of stars in the galaxies combined with the fact that they are very dark-matter dominated, stellar kinematic surveys tend to give us a view of the profiles of only the innermost regions of many dwarf spheroidals. Furthermore, these parameters themselves depend on the shape of the profile assumed; a given dwarf spheroidal will have different values for its characteristic radius and density depending on what value of γ_{dsph} or α_{dsph} is chosen. We therefore derive best fit parameters for individual values of γ_{dsph} and α_{dsph} using the maximum likelihood method described in appendix 1, using stellar kinematic data from the 18 dwarf spheroidals for which data was available [27–35].

We see from appendix 1 that varying the inner slope from $\gamma = 1.0$ to $\gamma = 0.2$ results in approximately an average reduction of 30% in the J-factor. This is twice as large as the 15% reduction found in the original Fermi-LAT dwarf spheroidal analysis [36], but in line with that found in the later analysis based on six years of Fermi-LAT data where a 20 – 40% difference was found [26]. Notably, however, large reductions are found for several dwarf spheroidals with particularly large J-factors; for dwarf spheroidals with $J_{\text{NFW}} > 19.0$ the average reduction is 40%. For the Einasto profile the difference is even more marked, with an average reduction of 70% from $\alpha = 0.2$ to $\alpha = 1.0$.

With our adopted value for the annihilation cross section from the fit to the GCE data, we can find the expected signal from any individual dwarf galaxy as a function of the dark matter mass m_χ using equation (2). Note that we only consider the prompt signal for dwarf spheroidal galaxies as they are

much cleaner environments and therefore have negligible contributions from inverse Compton scattering or bremsstrahlung radiation.

The Fermi-LAT collaboration has released the upper limits on the observed flux from a large number Milky Way dwarf spheroidal galaxies based on 6 years of observation [26]. We compare our simulated observed flux to these reported limits, assuming an observed flux of 0 and taking their 95% C.L. limit as twice the $1 - \sigma$ deviation. Computing the χ^2 of our simulations versus their observations, we obtain a 95% C.L. constraint on the halo parameters as a function of mass by finding the contour along which $\chi^2 = \chi_{\min}^2 + 6.18$. The resulting constraints are shown in figure 3. An upper limit can also be placed on the dark matter annihilation cross section $\langle\sigma v\rangle$ in the same manner. This upper limit is claimed to be in tension with the observed GCE flux [19]. In figure 5 we demonstrate the reduction of this limit as γ and α are varied.

In the analysis described so far, we have assumed $\gamma_{\text{MW}} = 1.0$. If a smaller inner slope were chosen, we would expect an increase in the best-fit annihilation cross-section for the signal. This would lead to correspondingly more stringent constraints on the dwarf spheroidals. We therefore repeat the calculation for several values of γ_{MW} , as well as for Einasto profiles with parameter α_{MW} to produce constraints in the $\gamma_{\text{d sph}} - \gamma_{\text{MW}}$ plane and $\alpha_{\text{d sph}} - \alpha_{\text{MW}}$ plane, shown in figure 4.

4 Comparison To Simulations and Observation

It has long been suspected that there is a discrepancy between the observed profiles of dwarf galaxies and those produced in CDM-only simulations. For a review of observational evidence and evidence from numerical simulations, see [37]. Early attempts to fit the observational data to an analytic profile [38, 39] showed that dwarf galaxies are well characterized as having a

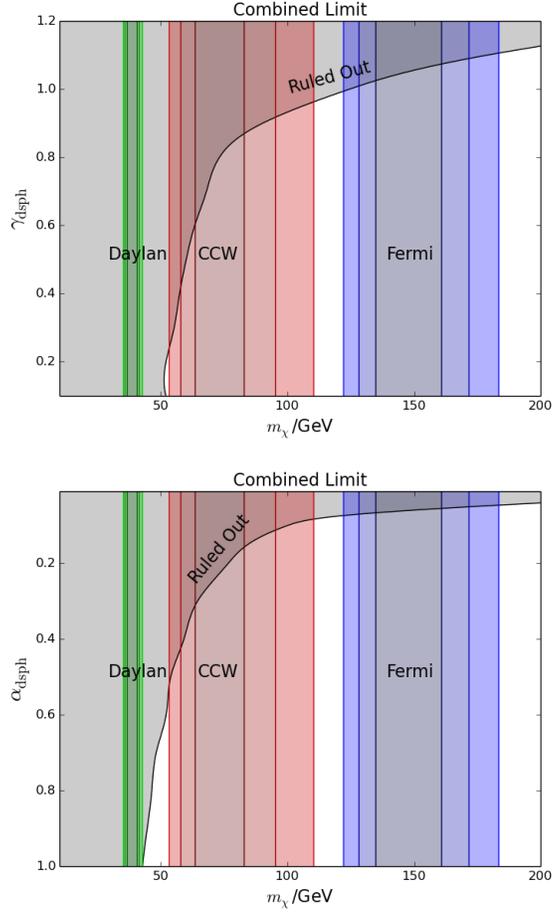


Figure 3: 95% C.L. constraints on γ_{dsph} and α_{dsph} for the generalized NFW (top) and Einasto (bottom) profiles respectively. The best-fit contours for the fit to the GCE are shown in red [6], green [7], and blue [8]. We assume $\gamma_{\text{MW}} = 1.0$ and $\langle \sigma v \rangle = 1.7 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$.

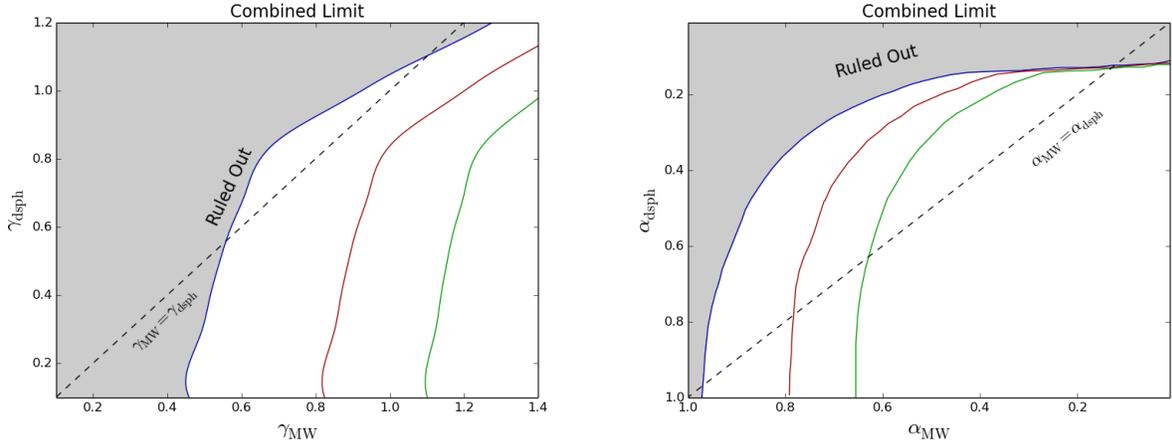


Figure 4: 95% C.L. constraints on γ for both the Milky way and the Dwarf Spheroidals. The signals are calculated for the individual best-fit masses and annihilation cross sections for each of the three datasets, as shown in Figure 2.

constant density core ($\gamma = 0$) following an isothermal profile:

$$\rho_I = \frac{\rho_0}{1 + (r/R_C)^2}, \quad (5)$$

where ρ_0 is the central density and R_C is the core radius. A variation on the isothermal profile, the Burkert profile [40] was later introduced to account for observations indicating that the density falls off as r^{-3} at large radii:

$$\rho_B = \frac{\rho_0}{(1 + r/R_C)(1 + (r/R_C)^2)}. \quad (6)$$

Numerous other groups have found evidence for cored, rather than cuspy, halos in dwarf galaxies [41–45]

Few studies present a numerical best fit value for the inner slope, instead typically comparing the NFW ($\gamma = 1$) model to an isothermal or Burkert profile ($\gamma = 0$). Those that do (several examples of which are listed below) tend to find values of $\gamma \sim 0.2$. Spekkens et al. [46] have derived density profiles for 165 low-mass galaxies including dwarf galaxies based on their rotation curves to find median inner slopes of $\gamma = 0.22 \pm 0.08$ to 0.28 ± 0.06 depending on the subsample considered.

Numerical simulations of cold dark matter (CDM) halos, on the other hand, have typically found values of the inner slope greater than $\gamma = 1$. Early numerical simulations of CDM halos were well characterized by the NFW profile of equation (3) with $\gamma \sim 1$ [47–49] for halos of all sizes. Others pointed towards an even steeper slope of $\gamma \sim 1.5$ [50, 51] or an intermediate value of $\gamma \sim 1.2$ [52]. Despite this variation, there is general agreement that pure CDM simulations result in inner slopes of $\gamma \geq 1$.

Some simulations instead found that the slope continues to become more shallow at smaller radii but does not converge[53, 54]. The Einasto profile, eq. (4)[55, 56], parameterizes this kind of behaviour. It describes a cored profile at large values of α and becomes cuspier for small values of order 0.1. Ref. [57] found CDM simulations are well described by $\alpha \approx 0.17$, which even at $r/r_s = 10^{-3}$ provide a slope of $\gamma \sim 1$, and therefore for our purposes represents a cuspy profile.

It is clear that our results for the inner slopes of dwarf spheroidal halos, while compatible with observation, are not compatible with traditional CDM simulations. Our results favour values of $\gamma_{\text{dsph}} < 1.0$. They also favour $\gamma_{\text{dsph}} < \gamma_{\text{MW}}$, which would suggest that the inner slope of the Milky Way’s profile is steeper than that of dwarf spheroidals.

5 Discussion

The core/cusp controversy is by no means new, and [58] reviews it in great detail. Many mechanisms have been proposed through which baryonic matter can have a feedback effect on the dark matter halo in the hopes of giving a more cored halo, but the results have been mixed. These mechanisms include rotating bars[59] (however later studies argue that this might actually have the opposite effect[60]) and the heating of cusps by dynamical friction[61–63] (however again, others find that this process is insufficient to explain cored profiles[64]). Another possibility is feedback from supernovae[65, 66]; in these

simulations repeated feedback from supernovae can turn a cusp into a core. Although viable baryonic mechanisms have been proposed to explain the discrepancy, its ultimate source remains an open question.

Although the standard CDM paradigm is difficult to render consistent with cored profiles, some dark matter models address this issue. Models of warm dark matter (WDM) such as sterile neutrinos rely on the particles having large velocities during structure formation, giving them a free-streaming length with a similar scale to galaxies. This smooths out density fluctuations on scales less than the free streaming length, and is borne out in simulations of WDM halos, giving dwarf sized halos a more cored profile[67–72]. Ref. [73] compare CDM and WDM simulations and find $\gamma = 1.18-1.46$ for CDM and $\gamma = 0.25-0.66$ for WDM. It should be noted, however, that WDM faces many challenges, including conflict with the small scale power spectrum [74], tension with strong-lens system observations which show evidence for a larger subhalo population than would be produced by WDM [75], and recent conservative estimates of the number of dwarf galaxies in the Milky Way restrict the WDM mass to $m_\chi \gtrsim 4$ keV, and may even restrict it to $m_\chi \gtrsim 8$ keV in the near future[76]. There are also challenges from observations of the Lyman- α forest which set a lower limit on the dark matter mass of a few keV [77, 78]. These requirements may be inconsistent with the formation of sizable cores, which requires WDM masses of $m_\chi \lesssim 1 - 2$ keV[70].

Another solution to the cusp-core problem is self-interacting dark matter (SIDM), in which cold dark matter has weak-scale interactions or no interactions at all with baryonic matter but a large self-interaction cross section. When the scattering cross section is of the order $\sigma/m_\chi \sim 0.1-1$ cm²g⁻¹, dark matter halos naturally form cores[79–81].

An interesting possibility is that of dark matter self-interacting through a light mediator. This results in a scattering cross section inversely proportional to velocity, causing greater self-interactions in dwarf galaxies than in galaxies or clusters[82]. For some choices of parameters, the cross section can be up to

100 times greater at velocities typically found in dwarf galaxies than for larger galaxies, which allows cored profiles to form for dwarfs but not for larger halos. These results correspond well to those presented here: the dwarf spheroidal halos are constrained to be more cored than that of the Milky Way. This ‘dark force’ scattering can be further enhanced at dwarf-scale velocities by resonances, and the coupling can even be chosen such that the correct relic density is reproduced [83–85], though the simplest s -wave models are ruled out by CMB constraints [86].

As WDM and SIDM are able to create cored halos, our results are consistent with these models which depart from the traditional CDM model. This implies that the GCE, if it does prove to originate from annihilating dark matter, would provide evidence in favour of these non-CDM cosmologies.

6 Conclusions

We have presented constraints on the density profiles of dwarf spheroidal galaxy dark matter halos under the assumption that the reported galactic center excess is due to annihilating dark matter. As there is currently tension between dark matter explanations of the GCE and observations of dwarf spheroidal galaxies (which do not exhibit any discernible excess), we can make this assumption and work backwards to determine the characteristics a dwarf spheroidal galaxy’s profile would need to satisfy in order to be consistent with the signal.

In the process of computing these constraints, we have determined the best fit characteristic radii and densities of 18 dwarf spheroidal galaxies for various possible NFW and Einasto profile slopes. This was accomplished by applying the log-likelihood method to stellar kinematic data. We find that choosing a cored profile over a cuspy one leads to a reduction in J-factor of approximately 30-50% in most cases, though the reduction varies for each individual dwarf.

We find that more cored profiles are favoured, and that the GCE is consistent with most observations of dwarf galaxies which show dwarf galaxies to be consistent with cored profiles. This could suggest that the GCE is more consistent with SIDM or WDM models than with the traditional CDM model.

Acknowledgments

I would like to thank James Cline and Jonathan Cornell for their ideas and assistance in preparing and reviewing this work, and Wei Xue for the use of and assistance with the GAMMASKY program. I also thank Matthew Walker for helpful correspondence and Marla Geha for providing us with the stellar kinematic data for several dwarf galaxies through private correspondence. This work was done with the support of the McGill Space Institute and the Natural Sciences and Engineering Research Council of Canada.

1 Maximum Likelihood Method

We adopt the method of Geringer-Sameth et. al [87] to calculate the halo parameters using the maximum likelihood method. They argue that the velocity data sample a Gaussian distribution, and therefore adopt the likelihood [87]:

$$L = \prod_{i=1}^N \frac{\exp \left[-\frac{1}{2} \frac{(u_i - \langle u \rangle)^2}{\delta_{u,i}^2 + \sigma^2(R_i)} \right]}{(2\pi)^{1/2} (\delta_{u,i}^2 + \sigma^2(R_i))^{1/2}}, \quad (7)$$

where u_i and $\delta_{u,i}$ are the observed line of sight velocity and uncertainty, $\langle u \rangle$ is the mean velocity of the dwarf, and $\sigma^2(R_i)$ is the velocity dispersion at the projected position of the observed star.

The velocity dispersion is the model dependent quantity, and has the form [88]

$$\sigma^2(R) = \frac{2G}{\Sigma(R)} \int_R^\infty \frac{v(s)M(s)}{s^2} \sqrt{s^2 - R^2} ds \quad (8)$$

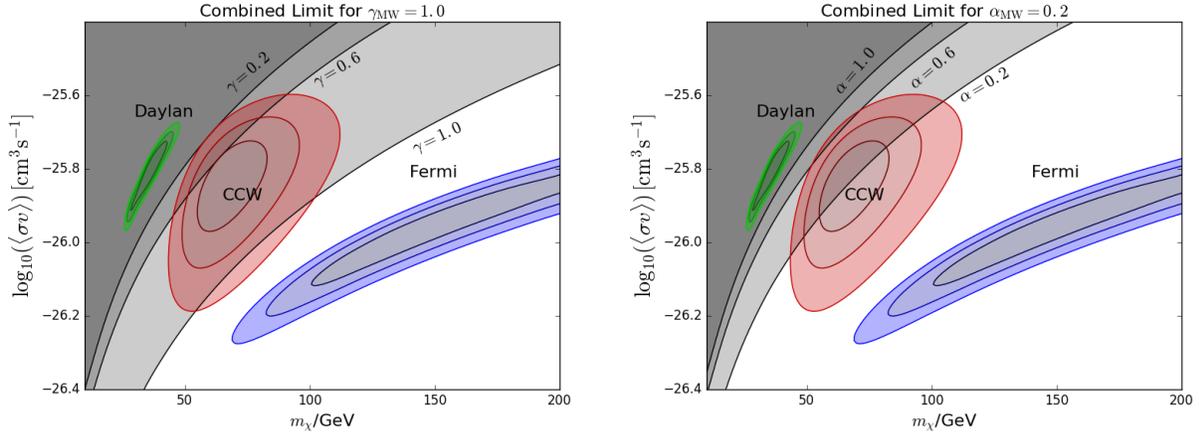


Figure 5: As Figure 2, however also shown is the combined 95% C.L. upper limit on $\langle\sigma v\rangle$ from the dwarf spheroidals. This is shown for several values of γ and α , demonstrating how the limit is weakened when considering more cored profiles (smaller γ or larger α).

for an isotropic halo (where the parameter describing the velocity anisotropy, β_a , has been set to 0). $M(r)$ is the mass contained within the given radius, and $v(r)$ and $\Sigma(R)$ are the stellar density and luminosity profiles respectively.

For a halo in which stars are distributed according to a Plummer profile [87], the ratio of these profiles is given by

$$\frac{v(r)}{\Sigma(R)} = \frac{3}{4r_{1/2}} \frac{1}{\sqrt{1 + r^2/r_{1/2}^2}}, \quad (9)$$

where $r_{1/2}$ is the half-light radius.

Note that we make two assumptions about the stellar profiles: the first is that the Plummer profile accurately describes the stellar distribution and the second that the velocity anisotropy is 0. For any given dwarf spheroidal both of these assumptions will be violated to some degree. The Plummer profile, however, is generally recognized as a good fit to the stellar distributions of dwarf spheroidals [89–92]. In the analysis of [87] most dwarf spheroidals are found to have nearly isotropic orbital velocities. We therefore adopt the simplifying assumption of $\beta_a=0$ for all.

Dwarf Galaxy	$\gamma = 0.2$		$\gamma = 0.4$		$\gamma = 0.6$		$\gamma = 0.8$		$\gamma = 1.0^b$		Ref.
	R_s [kpc]	ρ_s [GeV/cm ³]	R_s	ρ_s	R_s	ρ_s	R_s	ρ_s	R_s	ρ_s	
Carina	0.68	2.4	0.79	1.6	0.93	1.0	1.1	0.6	1.4	0.32	[27]
Draco	1.4	4.7	1.8	2.7	2.6	1.4	4.6	0.49	—	—	[28]
Fornax	0.66	6.6	0.74	4.8	0.84	3.3	0.98	2.1	1.2	1.3	[27]
Leo I	1.1	3.4	1.4	2.1	1.9	1.2	2.8	0.53	5.5	0.15	[29]
Leo II	1.1	3.1	1.4	1.9	1.9	1.1	2.8	0.48	6.1	0.13	[30]
Sculptor	0.57	6.6	0.65	4.5	0.76	2.9	0.92	1.8	1.2	0.98	[27]
Sextans	0.59	3.5	0.68	2.4	0.8	1.6	0.97	0.94	1.2	0.52	[27]
Bootes I	1.7	6.4	2.4	3.6	3.9	1.6	36	0.13	—	—	[31]
Hercules	1.0	0.62	1.2	0.39	1.6	0.22	2.2	0.11	3.7	0.039	[32] ^a
Leo V	2.0	1.3	2.8	0.73	5.5	0.3	40	0.012	—	—	[33]
Segue 1	1.1	4.4	1.4	2.7	1.9	1.5	2.8	0.67	6.4	0.17	[34]
Segue 2	1.4	4.9	1.8	2.9	2.6	1.5	4.9	0.54	—	—	[35]
Canes Venatici I	1.9	1.1	2.6	0.59	4.3	0.26	16	0.047	—	—	[32] ^a
Canes Venatici II	1.5	5.1	2	4.1	2.9	2	6.3	0.66	—	—	[32] ^a
Coma Berenices	1.4	6	1.9	3.5	2.7	1.8	5.4	0.62	—	—	[32] ^a
Leo T ^c	0.076	210	0.088	140	0.1	86	0.13	50	0.16	27	[32] ^a
UrsaMajor I	0.16	30	0.18	21	0.21	14	0.25	8.3	0.31	4.8	[32] ^a
UrsaMajor II	1.6	3.6	2.1	2.1	3.2	0.99	8	0.27	—	—	[32] ^a

^a Unpublished, provided by private correspondence.

^b For missing data, see explanation in text.

^c Due to lack of FERMI-LAT data, this dwarf is excluded from constraints on γ .

Table 2: Best-fit NFW parameters for various values of γ_{dsph} . Typical values of the relative uncertainties are $\sim 10\%$ for R_s and $\sim 15\%$ for ρ_s .

Dwarf Galaxy	$\alpha = 0.2$		$\alpha = 0.4$		$\alpha = 0.6$		$\alpha = 0.8$		$\alpha = 1.0$		Ref.
	R_s [kpc]	ρ_s [GeV/cm ³]	R_s	ρ_s	R_s	ρ_s	R_s	ρ_s	R_s	ρ_s	
Carina	1.6	0.061	1.3	0.11	1.2	0.16	1.2	0.20	1.1	0.24	[27]
Draco	15	0.016	2.9	0.19	1.8	0.44	1.4	0.67	1.2	0.85	[28]
Fornax	1.1	0.34	1.4	0.24	1.5	0.21	1.6	0.2	1.7	0.19	[27]
Leo I	8.0	0.02	2.2	0.16	1.5	0.3	1.2	0.41	1.1	0.48	[29]
Leo II	8.2	0.018	2.2	0.14	1.5	0.27	1.2	0.36	1.1	0.42	[30]
Sculptor	1.3	0.19	1.2	0.24	1.2	0.26	1.2	0.28	1.2	0.29	[27]
Sextans	1.3	0.11	1.2	0.14	1.2	0.15	1.2	0.16	1.2	0.17	[27]
Bootes I	38	0.013	4.1	0.24	2.1	0.59	1.6	0.88	1.3	1.1	[31]
Hercules	5.6	0.0048	1.8	0.03	1.3	0.052	1.1	0.068	1.5	0.62	[32] ^a
Leo V	63	0.0021	5.1	0.047	2.4	0.12	1.7	0.18	1.4	0.23	[33]
Segue 1	8.7	0.025	2.2	0.2	1.4	0.38	1.2	0.51	1.1	0.58	[34]
Segue 2	17	0.017	3	0.2	1.7	0.52	1.4	0.69	1.2	0.8	[35]
Canes Venatici I	40	0.002	4.5	0.039	2.3	0.1	1.7	0.16	1.4	0.2	[32] ^a
Canes Venatici II	22	0.021	3.3	0.28	1.8	0.48	1.4	0.65	1.3	0.76	[32] ^a
Coma Berenices	19	0.019	3.1	0.25	1.8	0.54	1.4	0.77	1.2	0.92	[32] ^a
Leo T ^c	0.16	6.2	0.16	7.3	0.17	7.6	0.18	8.2	0.18	9.4	[32] ^a
UrsaMajor I	0.32	1.1	0.35	0.96	0.41	0.72	0.47	0.56	0.52	0.46	[32] ^a
UrsaMajor II	26	0.0092	3.5	0.14	1.9	0.33	1.5	0.48	1.3	0.59	[32] ^a

^a Unpublished, provided by private correspondence.

^c Due to lack of FERMI-LAT data, this dwarf is excluded from constraints on γ .

Table 3: Best-fit Einasto parameters for various values of $\alpha_{\text{d sph}}$. Typical values of the relative uncertainties are $\sim 10\%$ for R_s and $\sim 15\%$ for ρ_s .

Dwarf Galaxy	$\gamma = 0.2 = \alpha$		$\gamma = 0.6 = \alpha$		$\gamma = 1.0 = \alpha$		$J_{\gamma=0.2}/J_{\gamma=1.0}$ NFW	$J_{\alpha=1.0}/J_{\alpha=0.2}$ Einasto	Ref.
	J_{NFW}	J_{Einasto}	J_{NFW}	J_{Einasto}	J_{NFW}	J_{Einasto}			
Carina	17.4	17.6	17.4	17.6	17.4	17.7	0.844	1.3	[27]
Draco	18.8	19.6	18.9	19.2	18.9	19.2	0.924	0.403	[28]
Fornax	18	18.3	18	17.9	18.1	17.8	0.711	0.32	[27]
Leo I	17.6	17.9	17.7	17.6	17.7	17.6	0.843	0.443	[29]
Leo II	17.6	17.9	17.6	17.6	17.7	17.5	0.795	0.387	[30]
Sculptor	18.2	18.4	18.2	18.2	18.3	18.1	0.777	0.539	[27]
Sextans	17.7	17.9	17.7	17.7	17.8	17.7	0.771	0.524	[27]
Bootes I	19.4	20.7	19.4	19.8	19.4	19.7	0.77	0.0859	[31]
Hercules	16.5	16.8	16.5	16.5	16.5	18.7	0.858	91.4 ^b	[32] ^a
Leo V	17.7	19	17.8	17.8	16.8	18.5	7.94 ^b	0.331	[33]
Segue 1	19	20.3	19.2	19.9	19.3	19.8	0.507	0.308	[34]
Segue 2	19.1	20.5	19.3	20	19.4	19.8	0.555	0.215	[35]
Canes Venatici I	17.3	18.2	17.4	17.4	17.5	17.2	0.694	0.12	[32] ^a
Canes Venatici II	18.6	19.7	19	18.7	18.9	18.5	0.567	0.0739	[32] ^a
Coma Berenices	19.3	20.5	19.4	19.9	19.5	19.8	0.6	0.165	[32] ^a
Leo T ^c	17	16.9	17	14.7	17	13.8	1.15	0.000671 ^b	[32] ^a
UrsaMajor I	17.8	18	17.9	17.5	17.9	17.4	0.842	0.238	[32] ^a
UrsaMajor II	19	20.6	19.1	19.8	19.1	19.7	0.747	0.132	[32] ^a

^a Unpublished, provided by private correspondence.

^b See explanation in text.

^c Due to lack of FERMI-LAT data, this dwarf is excluded from constraints on γ .

Table 4: J-factors derived from best-fit parameters for various values of γ_{dsph} .

The mass contained within a given radius is attained by integrating the chosen density profile:

$$M(s) = \int_0^s 4\pi r^2 \rho_s(r, R_s, \rho_s, \gamma) dr. \quad (10)$$

For each dwarf spheroidal we minimize the negative log likelihood for $0.1 \leq \gamma \leq 1.2$ and again for $0.1 \leq \alpha \leq 1.0$ (for the NFW and Einasto profiles respectively) over the parameters R_s and ρ_s . This is accomplished using the downhill simplex method over the two parameters. The best fit values of R_s and ρ_s are shown in table 2 for several values of γ and in table 3 for α , the Einasto profile parameter.

Best fit values are not available for some dwarf galaxies for $\gamma = 1.0$ (or greater). The likelihood in these cases approaches its maximum value only as $r_s \rightarrow \infty$ and $\rho_s \rightarrow \infty$. This is due to the nature of the NFW profile, which has its shallowest log slope at $r = 0$, with the slope becoming steeper at greater distances. In these cases, therefore, the slope $\gamma = 1.0$ is inconsistent with the stellar kinematic data. In these cases the fit can always be made better by increasing r_s to grant a smaller log slope (approaching a uniform log slope of 1.0) and reducing the density to compensate.

Typical relative errors on R_s are approximately $\sim \pm 10\%$, while those on ρ_s are somewhat larger ($\sim \pm 15\%$). These are found by varying each parameter from the best fit until the criteria $\Delta \ln \mathcal{L} = 2.6 / \ln 2$ is satisfied. The fits are not good enough to discriminate between values of γ or α , with the log likelihood varying by only by $\Delta \ln \mathcal{L} \sim 0.5$ for all dwarf spheroidals between $0.1 \leq \gamma \leq 1.2$ and $0.1 \leq \alpha \leq 1.0$.

In table 4 we present the J-factors corresponding to the best fit halo parameters for $\gamma, \alpha = 0.2, 0.6, \text{ and } 1.0$, and Figure 5 shows the corresponding effect on the upper limit on the annihilation cross section. Note that even in the cases of $\gamma = 1.0$ where the halo parameters cannot be found, the J-factor is nonetheless convergent. We also demonstrate the reduction of the J-factors for both profiles as the inner slope is changed. For the NFW profiles,

$J_{\gamma=0.2}/J_{\gamma=1.0} \sim 0.7$ on average, and in Einasto profiles the difference is even more marked with $J_{\alpha=1.0}/J_{\gamma=0.2} \sim 0.4$. A few of the Dwarf spheroidals show an opposite trend, with the J-factor increasing as the profile is made more cored, notably Hercules (for the Einasto profile) with $J_{\alpha=1.0}/J_{\gamma=0.2} = 91.4$ and Leo V (for the NFW profile) with $J_{\gamma=0.2}/J_{\gamma=1.0} = 7.94$. Little stellar kinematical data exists for Hercules and Leo V, and our analyses are therefore based on very few data points, which could explain this discrepancy. The same is true of Leo T, which displays the opposite behaviour with an unrealistically large decrease in J-factor.

REFERENCES

- [1] L. Goodenough and D. Hooper, (2009), 0910.2998.
- [2] D. Hooper and L. Goodenough, Phys. Lett. **B697**, 412 (2011), 1010.2752.
- [3] D. Hooper and T. Linden, Phys. Rev. **D84**, 123005 (2011), 1110.0006.
- [4] K. N. Abazajian and M. Kaplinghat, Phys. Rev. **D86**, 083511 (2012), 1207.6047, [Erratum: Phys. Rev.D87,129902(2013)].
- [5] B. Zhou *et al.*, Phys. Rev. **D91**, 123010 (2015), 1406.6948.
- [6] F. Calore, I. Cholis, and C. Weniger, JCAP **1503**, 038 (2015), 1409.0042.
- [7] T. Daylan *et al.*, Phys. Dark Univ. **12**, 1 (2016), 1402.6703.
- [8] Fermi-LAT, M. Ajello *et al.*, (2015), 1511.02938.
- [9] C. Karwin, S. Murgia, T. M. P. Tait, T. A. Porter, and P. Tanedo, Phys. Rev. **D95**, 103005 (2017), 1612.05687.
- [10] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, (2015), 1506.05124.
- [11] R. Bartels, S. Krishnamurthy, and C. Weniger, (2015), 1506.05104.
- [12] R. M. O’Leary, M. D. Kistler, M. Kerr, and J. Dexter, (2015), 1504.02477.
- [13] T. Linden, Phys. Rev. **D93**, 063003 (2016), 1509.02928.
- [14] I. Cholis, D. Hooper, and T. Linden, JCAP **1506**, 043 (2015), 1407.5625.
- [15] Fermi-LAT, M. Ackermann *et al.*, Astrophys. J. **840**, 43 (2017), 1704.03910.
- [16] Fermi-LAT, M. Ajello *et al.*, Submitted to: Astrophys. J. (2017), 1705.00009.
- [17] D. Haggard, C. Heinke, D. Hooper, and T. Linden, (2017), 1701.02726.

- [18] Fermi-LAT, M. Ackermann *et al.*, Phys. Rev. **D91**, 122002 (2015), 1506.00013.
- [19] K. N. Abazajian and R. E. Keeley, (2015), 1510.06424.
- [20] M. Cirelli *et al.*, JCAP **1103**, 051 (2011), 1012.4515, [Erratum: JCAP1210,E01(2012)].
- [21] P. Ciafaloni *et al.*, JCAP **1103**, 019 (2011), 1009.0224.
- [22] J. I. Read, J. Phys. **G41**, 063101 (2014), 1404.1938.
- [23] B. Burch and R. Cowsik, Astrophys. J. **779**, 35 (2013), 1306.1920.
- [24] C. Evoli, D. Gaggero, D. Grasso, and L. Maccione, JCAP **0810**, 018 (2008), 0807.4730.
- [25] G. Di Bernardo, C. Evoli, D. Gaggero, D. Grasso, and L. Maccione, JCAP **1303**, 036 (2013), 1210.4546.
- [26] Fermi-LAT, M. Ackermann *et al.*, Phys. Rev. Lett. **115**, 231301 (2015), 1503.02641.
- [27] M. G. Walker, M. Mateo, and E. Olszewski, Astron. J. **137**, 3100 (2009), 0811.0118.
- [28] M. G. Walker, E. W. Olszewski, and M. Mateo, Mon. Not. Roy. Astron. Soc. **448**, 2717 (2015), 1503.02589.
- [29] M. Mateo, E. W. Olszewski, and M. G. Walker, Astrophys. J. **675**, 201 (2008), 0708.1327.
- [30] A. Koch *et al.*, Astron. J. **134**, 566 (2007), 0704.3437.
- [31] S. E. Koposov *et al.*, Astrophys. J. **736**, 146 (2011), 1105.4102.
- [32] J. D. Simon and M. Geha, Astrophys. J. **670**, 313 (2007), 0706.0516.
- [33] M. G. Walker *et al.*, Astrophys. J. **694**, L144 (2009), 0902.3003.
- [34] J. D. Simon *et al.*, Astrophys. J. **733**, 46 (2011), 1007.4198.
- [35] E. N. Kirby *et al.*, Astrophys. J. **770**, 16 (2013), 1304.6080.

- [36] Fermi-LAT, M. Ackermann *et al.*, Phys. Rev. **D89**, 042001 (2014), 1310.0828.
- [37] W. J. G. de Blok, Adv. Astron. **2010**, 789293 (2010), 0910.3538.
- [38] B. Moore, Nature **370**, 629 (1994).
- [39] R. A. Flores and J. R. Primack, Astrophys. J. **427**, L1 (1994), astro-ph/9402004.
- [40] A. Burkert, IAU Symp. **171**, 175 (1996), astro-ph/9504041, [Astrophys. J.447,L25(1995)].
- [41] A. Burkert and J. Silk, Astrophys. J. **488**, L55 (1997), astro-ph/9707343.
- [42] F. C. van den Bosch and R. A. Swaters, Mon. Not. Roy. Astron. Soc. **325**, 1017 (2001), astro-ph/0006048.
- [43] P. Salucci, F. Walter, and A. Borriello, Astron. Astrophys. **409**, 53 (2003), astro-ph/0206304.
- [44] G. Gentile, P. Salucci, U. Klein, and G. L. Granato, Mon. Not. Roy. Astron. Soc. **375**, 199 (2007), astro-ph/0611355.
- [45] R. Cowsik, K. Wagoner, E. Berti, and A. Sircar, Astrophys. J. **699**, 1389 (2009), 0904.0451.
- [46] K. Spekkens and R. Giovanelli, Astron. J. **129**, 2119 (2005), astro-ph/0502166.
- [47] J. Dubinski and R. G. Carlberg, Astrophys. J. **378**, 496 (1991).
- [48] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490**, 493 (1997), astro-ph/9611107.
- [49] P. Colin, A. Klypin, O. Valenzuela, and S. Gottlober, Astrophys. J. **612**, 50 (2004), astro-ph/0308348.
- [50] B. Moore, T. R. Quinn, F. Governato, J. Stadel, and G. Lake, Mon. Not. Roy. Astron. Soc. **310**, 1147 (1999), astro-ph/9903164.

- [51] A. Klypin, A. V. Kravtsov, J. Bullock, and J. Primack, *Astrophys. J.* **554**, 903 (2001), astro-ph/0006343.
- [52] J. Diemand, M. Zemp, B. Moore, J. Stadel, and M. Carollo, *Mon. Not. Roy. Astron. Soc.* **364**, 665 (2005), astro-ph/0504215.
- [53] J. F. Navarro *et al.*, *Mon. Not. Roy. Astron. Soc.* **349**, 1039 (2004), astro-ph/0311231.
- [54] E. Hayashi *et al.*, *Mon. Not. Roy. Astron. Soc.* **355**, 794 (2004), astro-ph/0310576.
- [55] D. Merritt, J. F. Navarro, A. Ludlow, and A. Jenkins, *Astrophys. J.* **624**, L85 (2005), astro-ph/0502515.
- [56] A. W. Graham, D. Merritt, B. Moore, J. Diemand, and B. Terzic, *Astron. J.* **132**, 2701 (2006), astro-ph/0608613.
- [57] J. F. Navarro *et al.*, *Mon. Not. Roy. Astron. Soc.* **402**, 21 (2010), 0810.1522.
- [58] D. H. Weinberg, J. S. Bullock, F. Governato, R. Kuzio de Naray, and A. H. G. Peter, *Proc. Nat. Acad. Sci.* **112**, 12249 (2014), 1306.0913.
- [59] M. D. Weinberg and N. Katz, *Astrophys. J.* **580**, 627 (2002), astro-ph/0110632.
- [60] J. Dubinski, I. Berentzen, and I. Shlosman, *IAU Symp.* **254**, 165 (2009), 0807.4197.
- [61] A. El-Zant, I. Shlosman, and Y. Hoffman, *Astrophys. J.* **560**, 636 (2001), astro-ph/0103386.
- [62] C. Tonini and A. Lapi, *Astrophys. J.* **649**, 591 (2006), astro-ph/0603051.
- [63] E. Romano-Diaz, I. Shlosman, Y. Hoffman, and C. Heller, *Astrophys. J.* **685**, L105 (2008), 0808.0195.
- [64] J. R. Jardel and J. A. Sellwood, *Astrophys. J.* **691**, 1300 (2009), 0808.3449.

- [65] A. Pontzen and F. Governato, *Mon. Not. Roy. Astron. Soc.* **421**, 3464 (2012), 1106.0499.
- [66] F. Governato *et al.*, *Mon. Not. Roy. Astron. Soc.* **422**, 1231 (2012), 1202.0554.
- [67] V. Avila-Reese, P. Colin, O. Valenzuela, E. D’Onghia, and C. Firmani, *Astrophys. J.* **559**, 516 (2001), astro-ph/0010525.
- [68] P. Colin, O. Valenzuela, and V. Avila-Reese, *Astrophys. J.* **673**, 203 (2008), 0709.4027.
- [69] R. M. Dunstan, K. N. Abazajian, E. Polisensky, and M. Ricotti, (2011), 1109.6291.
- [70] A. V. Maccio, S. Paduroiu, D. Anderhalden, A. Schneider, and B. Moore, *Mon. Not. Roy. Astron. Soc.* **424**, 1105 (2012), 1202.1282.
- [71] A. Schneider, D. Anderhalden, A. Maccio, and J. Diemand, *Mon. Not. Roy. Astron. Soc.* **441**, 6 (2014), 1309.5960.
- [72] R. E. Angulo, O. Hahn, and T. Abel, *Mon. Not. Roy. Astron. Soc.* **434**, 3337 (2013), 1304.2406.
- [73] A. Gonzalez-Samaniego, V. Avila-Reese, and P. Colin, *Astrophys. J.* **819**, 101 (2016), 1512.03538.
- [74] E. Polisensky and M. Ricotti, *Phys. Rev.* **D83**, 043506 (2011), 1004.1459.
- [75] N. Dalal and C. S. Kochanek, *Astrophys. J.* **572**, 25 (2002), astro-ph/0111456.
- [76] S. Y. Kim, A. H. G. Peter, and J. R. Hargis, (2017), 1711.06267.
- [77] U. Seljak, A. Makarov, P. McDonald, and H. Trac, *Phys. Rev. Lett.* **97**, 191303 (2006), astro-ph/0602430.
- [78] M. Viel *et al.*, *Phys. Rev. Lett.* **100**, 041304 (2008), 0709.0131.
- [79] D. N. Spergel and P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000), astro-ph/9909386.

- [80] M. Rocha *et al.*, Mon. Not. Roy. Astron. Soc. **430**, 81 (2013), 1208.3025.
- [81] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, Mon. Not. Roy. Astron. Soc. **430**, 105 (2013), 1208.3026.
- [82] A. Loeb and N. Weiner, Phys. Rev. Lett. **106**, 171302 (2011), 1011.6374.
- [83] S. Tulin, H.-B. Yu, and K. M. Zurek, Phys. Rev. Lett. **110**, 111301 (2013), 1210.0900.
- [84] S. Tulin, H.-B. Yu, and K. M. Zurek, Phys. Rev. **D87**, 115007 (2013), 1302.3898.
- [85] M. Kaplinghat, S. Tulin, and H.-B. Yu, Phys. Rev. Lett. **116**, 041302 (2016), 1508.03339.
- [86] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, Phys. Rev. Lett. **118**, 141802 (2017), 1612.00845.
- [87] A. Geringer-Sameth, S. M. Koushiappas, and M. Walker, Astrophys. J. **801**, 74 (2015), 1408.0002.
- [88] G. A. Mamon and E. L. Lokas, Mon. Not. Roy. Astron. Soc. **363**, 705 (2005), astro-ph/0405491, [Addendum: Mon. Not. Roy. Astron. Soc.370,1582(2006)].
- [89] M. Irwin and D. Hatzidimitriou, Mon. Not. Roy. Astron. Soc. **277**, 1354 (1995).
- [90] S. Mashchenko, A. Sills, and H. M. P. Couchman, Astrophys. J. **640**, 252 (2006), astro-ph/0511567.
- [91] M. G. Walker *et al.*, Astrophys. J. **704**, 1274 (2009), 0906.0341, [Erratum: Astrophys. J.710,886(2010)].
- [92] A. McConnachie and M. Irwin, Mon. Not. Roy. Astron. Soc. **365**, 1263 (2006), astro-ph/0511004.

CHAPTER 4

Supermassive Black Holes from Self-Interacting Dark Matter

Direct and indirect detection of dark matter can potentially be used to observe dark matter, but despite decades of searches no conclusive evidence of dark matter signals has been found. Signals such as the Galactic Center excess tend to rise to prominence only to be eventually explained by baryonic matter. The only solid evidence for dark matter is through its gravitational effects. These gravitational effects can be seen both at low and high redshifts, and dark matter is thought to have a significant effect on structure formation in the early universe [210]. It is possible dark matter is the key to several mysteries in the early universe, including the formation of supermassive black holes; the observation of these black holes at high redshifts is difficult to explain.

In section 4.1, we give an overview of the problem of high-redshift supermassive black holes, and in section 4.2 we summarize the current baryonic (non-dark matter) explanations for their existence. In section 4.3 we discuss an alternative explanation: the formation of supermassive black holes from self-interacting dark matter. In section 4.4 we present our own work, investigating the allowed parameter space of three models in which supermassive black holes form from self-interacting dark matter sufficiently early to explain observations.

4.1 High-Redshift SMBHs

Supermassive black holes, found at the centers of galaxies, can be observed as AGNs (Active Galactic Nuclei) as matter accretes onto the SMBH. The mass of the SMBH can be estimated based on the $H\alpha$ and $H\beta$ line widths, which are related to the velocity dispersion of the accretion disk and therefore

the virial mass [211]. Most of these are found to be near the Eddington luminosity — the luminosity at which the gravitational potential is balanced by the radiation pressure of the AGN (see, for example, Ref. [212]).

Supermassive black holes have been observed at redshifts of up to $z \sim 7.5$ with masses of up to $M_{\text{SMBH}} \sim 10^9 M_{\odot}$ [213]. This is at odds with the fact that they are typically slow to grow, as their accretion rate is restricted by the Eddington luminosity (sometimes referred to as the Eddington limit on accretion) [214]. If the SMBHs are made up primarily of baryonic matter, this leaves three possibilities: either small seed black holes form much earlier than previously thought (as early as $z \sim 30$), the black holes have some method of super-Eddington accretion, or there is some way in which large ($M_{\text{SMBH}} \sim 10^4 - 10^5 M_{\odot}$) black hole seeds form without accretion.

4.2 Baryonic Explanations for High-Redshift SMBHs

Most current explanations for these high-redshift SMBHs fall under three categories: light seeds from Population III stars, intermediate seeds formed through collisions of lighter seeds, and heavy seeds formed through direct collapse of gas clouds.

4.2.1 Population III Stars

Population III stars are the hypothetical first stars from which the first heavy elements are formed. They form at $z \sim 20 - 50$, and can become very massive so long as the metallicity of the cloud is poor enough to slow fragmentation of the gas cloud [215]. Metals provide a very efficient cooling mechanism; since cold clouds are more prone to fragmentation the presence of metals will result in a larger number of smaller stars. The evolution of such stars with masses of over $260 M_{\odot}$ can result in the formation of massive black holes [216]. Above this mass, the collapse cannot be reversed by the radiation pressure generated by nuclear burning, causing approximately half of the star

to collapse into a black hole. This results in SMBHs of $M \sim 10^2 M_\odot$. If these form sufficiently early, they could accrete enough matter to become SMBHs by a redshift of $z \sim 7$. It is unknown, however, if Population III stars actually formed above this mass threshold; there are a large number of factors which could prevent their formation such as fragmentation of the gas cloud (leading to the formation of more than one Population III star per gas cloud) [217] or feedback from the collapsing gas cloud [218].

4.2.2 Direct Collapse Black Holes

An alternate route to creating black holes is through Direct Collapse. If halos are exposed to large quantities of dissociating UV radiation, the formation of molecular hydrogen (H_2) is suppressed in the early universe. This, along with the lack of metals, prevents gas clouds from efficiently cooling [219]. Large dense gas clouds in galactic centers with suppressed H_2 and metallicity are unable to cool efficiently beyond $T \sim 4000$ K, and this inability to cool prevents fragmentation of the cloud. The result is the isothermal collapse of the gas cloud.

In most cases the cloud has sufficient angular momentum to halt the collapse, but if the gas cloud happens to have a very small angular momentum [220], or if the angular momentum can be transported to its surroundings (for example through the formation of bars [221]), the collapse continues until the Eddington limit is reached and a massive central object forms — a supermassive star with mass $10^4 - 10^6 M_\odot$. The supermassive star forms a black hole at its center which can accrete mass, and does so much more rapidly than the Eddington limit since the accretion rate of the gas is limited by the size of the full star rather than the black hole itself. Up to 90% of the mass eventually forms the black hole [222], resulting in very large black hole seeds

which can then accrete more gas until they reach the masses of those seen at high redshifts.

Much like the scenario in which Population III stars form the first SMBH seeds, this process requires specific conditions to be met in the earliest halos, and there is a good deal of uncertainty about the environments in which direct collapse black holes (DCBHs) could form. For example, this process requires the presence of much more UV radiation than is thought to be present in most halos, and the fraction of galaxies in which these conditions would occur has been estimated to be $\sim 10^{-6}$ [223]. It is also heavily reliant on mechanisms to strip sufficient angular momentum from the collapsing halo while maintaining its large mass and without causing it to fragment.

4.2.3 Mergers

Neither of the above scenarios can occur once stars have formed, as they require large quantities of metal-free gas to prevent cooling and therefore fragmentation. If the first halos are able to cool quickly (for example through H_2 cooling), then by the time large enough gas clouds have formed they will already have been enriched by the first stars and will fragment to form lower mass stars [224]. Although this prevents the formation of large black hole seeds through the first two methods, it may produce an environment conducive to SMBH formation through mergers.

If the gas has cooled significantly, stars can form in very dense clusters of up to $\sim 10^5 M_\odot$ with a half-mass radius of ~ 1 pc [225]. Runaway collisions between the stars can lead to the formation of a $\sim 10^3 M_\odot$ black hole, intermediate to the two other scenarios.

There are a number of possible baryonic explanations for these high-redshift SMBHs, but each relies on rather specific conditions in the early universe, which may or may not be achieved. We have therefore explored another

proposed solution: the formation of seed black holes through the collapse of a self-interacting dark matter halo.

4.3 High-Redshift SMBHs From SIDM

Dark matter undoubtedly plays a role in the formation of supermassive black holes as the dominant factor affecting the formation of large-scale structures [210]. It is possible, however, that it is directly responsible for the formation of SMBHs. This possibility was explored by Pollack *et al.* [226], who proposed that strong self-interactions between dark matter particles could result in the formation of a SMBH through the gravothermal catastrophe. This is an attractive possibility. The three scenarios discussed above require very particular conditions (such as ultra-dense star clusters or large quantities of dissociating UV radiation), but detailed knowledge of the newly formed structures at these times is lacking. The formation of a black hole through the gravothermal catastrophe, on the other hand, requires simply that the dark matter halo be relatively undisturbed by mergers or collisions throughout its early history.

It is well known that a dark matter scattering cross section of $\sigma \sim 1 \text{ cm}^2/\text{g}$ results in cored halos in dwarf and low surface brightness galaxies [131–133], but on sufficiently large timescales even elastic scatterings can result in the transfer of energy to the outer halo [227], causing the collapse of the inner halo and the eventual formation of a black hole. A halo subject to elastic scatterings between particles undergoes gravothermal collapse as heat flows from the inner region (with high temperature) to the outer (with low temperature) [227]. As the temperature of the inner halo would decrease, it instead contracts, increasing its temperature as particles move to a lower potential energy. Somewhat counter-intuitively, the inner halo’s temperature increases as heat flows out of it, leading to a runaway process through which the core

of the halo will eventually collapse to a sufficiently small and hot region that the radial instability causes the formation of a black hole — the gravothermal catastrophe. This process is a known mechanism responsible for the creation of globular clusters, as stars scatter elastically through the gravitational force and collapse into a small dense region. Unlike SIDM, the formation of stellar binaries eventually halts the collapse by providing an alternative energy sink [228], avoiding the gravothermal catastrophe.

This process bypasses the Eddington limit by using dark matter rather than baryonic matter to form the initial black hole seed. So long as there is no dark radiation (as in elastic scattering) or the dark radiation exerts a much smaller pressure than that of baryonic matter, accretion proceeds much more quickly. Once the black hole seed is formed, baryonic matter can accrete onto it as normal to further increase its mass.

Black hole seeds may also form through inelastic collisions between dark matter particles. Dissipative dark matter, such as a mirror sector that includes dark atoms, can result in the formation of intermediate-mass black hole seeds ($\sim 10^4 - 10^5 M_\odot$) [229]. Dark atoms, like ordinary atoms, can scatter inelastically by exciting energy transitions such as the hyperfine transition, or can form bound H_2 states which will then cool efficiently through their own mutual scatterings. These dissipative interactions speed the transfer of heat to the outer halo in the form of dark radiation. In the paper by D’Amico *et al.* [229], the dark atoms have a lower temperature than the ordinary matter, allowing them to form larger gas clouds which direct collapse to form massive black holes.

Both these models rely on dark matter self-interactions, however, on which there are significant constraints. Although a cross section of $\sigma \sim 1 \text{ cm}^2/\text{g}$ forms the observed cores in small halos, larger self-interactions are ruled out by

bounds from the Bullet Cluster. With this scattering cross section, we would expect a typical halo to form a central black hole through the gravothermal catastrophe only on the scale of $\sim 10^3 - 10^4$ Gyr [227, 230], far exceeding the Hubble time let alone the ~ 1 Gyr timescale required to explain high redshift SMBHs. To produce the black hole seeds more quickly, a much higher scattering cross section is required. For example, if the entire dark sector were made of mirror dark atoms with properties similar to ordinary atoms, the scattering cross section would be at least $\sigma \sim a_0^2/m_H \approx 10^7 \text{ cm}^2/\text{g}$, where a_0 is the Bohr radius and m_H is the Hydrogen atom mass.

In both cases, therefore, two dark matter components are required. The Bullet Cluster bounds are weakened if the dark matter has a CDM component alongside a sub-dominant SIDM component. If the total SIDM abundance is restricted to $M_{\text{SIDM}}/M_{\text{CDM}} \lesssim 0.1$, the Bullet Cluster bounds can be effectively evaded for arbitrarily large scattering cross sections [226], allowing us to consider models in which such a sub-dominant component is responsible for the formation of the SMBH even at very high redshifts.

4.4 Our Work

The previous work in high-redshift SMBHs has so far relied on analytic or semi-analytic models solving the gravothermal fluid equations. The primary advantage of these over N-body simulations is that they are relatively computationally inexpensive: a large number of simulations can be completed in a short time, allowing a wide range of parameter space to be investigated. These simulations, however, struggle to accommodate multiple components of dark matter, a problem which Pollack *et al.* solve by considering the self-interacting component's evolution only, and doing so with a static background. These models also must be calibrated individually for each possible scattering type, making them difficult to extend beyond elastic scattering.

N-body simulations, while much more computationally expensive, can easily accommodate both multiple dark matter components and different scattering types without ignoring the effect of the SIDM on the dominant component. In our work, we consider the evolution of a dark matter halo in three models: elastic scattering, inelastically scattering dark matter, and dark matter which is able to form bound states.

Our simulations are done with the help of the GADGET code [231], a parallel N-body simulation code which has previously been modified to accommodate dark matter self-interactions [230] and which we have further modified to accommodate our own models. We consider not only the evolution of the halo but also its formation, which for large self-interaction cross sections can be affected by the collisions. Our primary goal in the work is to determine the exact parameter space in which dark matter self-interactions can lead to the production of high-redshift SMBHs without reliance on baryonic effects beyond the subsequent accretion of gas onto the black hole seed.

This manuscript is currently in preparation for publication. It is primarily my own work, with the co-authors writing section 6.2 and providing assistance in writing, planning, confirming results, and editing the existing GADGET code.

Early formation of supermassive black holes via dark matter self-interactions

Jeremie Choquette^{*1}, James M. Cline^{†1}, and Jonathan M. Cornell^{‡2}

¹*Department of Physics, McGill University, 3600 Rue University, Montréal,
Québec, Canada H3A 2T8*

²*Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221,
USA*

Abstract

The existence of supermassive black holes at high redshifts ($z \sim 7$) is difficult to accommodate in standard astrophysical scenarios. It has been shown that dark matter models with a subdominant self-interacting component are able to produce early seeds for supermassive black holes through the gravothermal catastrophe. Previous studies used a fluid equation approach, requiring some limiting assumptions. Here we reconsider the problem using N -body gravitational simulations starting from the formation of the initial dark matter halo. We consider both elastic and dissipative scattering, and elucidate the interplay between the dark matter microphysics and subsequent accretion of the black hole needed to match the properties of observed high redshift supermassive black holes.

1 Introduction

Supermassive black holes (SMBHs) are now known to be ubiquitous in the centers of Milky way-like and larger galaxies. Although our own galaxy's SMBH is quiescent, those in active galactic nuclei (quasars) are highly luminous due to radiation from accretion, outshining their entire host galaxy. In recent years, quasars containing SMBHs with masses of order $10^9 M_\odot$ have been discovered at redshifts of up to 7.5 [4, 13, 32]. In standard scenarios for structure formation, it is difficult to account for these large masses at such early times, since the progenitors must start out significantly lighter and only acquire their observed masses through accretion. The rate of accretion is bounded by the Eddington limit, which is the maximum allowed by the balance of gravitational force versus radiative pressure. This restricts the rate of growth to an e-folding time of order 50 Myr [40].

Alternative astrophysical mechanisms have been proposed for producing early SMBHs, that typically rely upon boosting the mass of the progenitor to order $\gtrsim 100 M_\odot$, so that less accretion time is needed. These include early Population III stars, collisions of stellar-mass black holes and stars in stellar clusters to form black holes with mass $\sim 10^3$ - $10^4 M_\odot$, or the direct collapse of low metallicity gas clouds into black holes. For a review of these mechanisms, see ref. [50].

The mass of an accreting black hole as a function of time is given by [40, 50]

$$M(t) = M_0 \exp\left(\frac{1 - \epsilon_r}{\epsilon_r} \frac{t}{0.45 \text{ Gyr}}\right). \quad (1)$$

The radiative efficiency is typically taken to be $\epsilon_r \approx 0.1$ [50]. Then a black hole seed of $M_0 = 10^2 M_\odot$ would take at least 0.81 Gyr to develop into a $10^9 M_\odot$ SMBH, whereas a seed with $M_0 = 10^5 M_\odot$ would take 0.46 Gyr. The age of the universe at $z = 7$ is approximately 0.76 Gyr. This means the seeds must either form very early, *e.g.*, $z = 13.5$ in the case of $M_0 = 10^5 M_\odot$, or be very large, presenting a challenge even for the above mechanisms.

An alternative mechanism is the gravothermal collapse of a self-interacting dark matter (SIDM) halo, as shown in ref. [37], hereafter called PSS14. Gravothermal collapse is the process believed to be the origin of globular clusters, through gravitational interactions that eject more energetic stars, allowing the gravitationally bound system to contract [26]. Such systems have negative specific heat, and the process can run away unless halted by some interaction that prevents further outflow of energy. In the case of globular clusters, formation of binary systems may halt runaway collapse.

Self-interactions of dark matter (DM) can cause the analogous process in DM halos. In this case, there need not be anything that halts the collapse, which results in a black hole. Several early studies of halo formation with SIDM considered this process [1, 2, 22], in the context of using SIDM to solve the core-cusp problem of halo density profiles, rather than trying to explain SMBH formation. Refs. [1, 22] showed that, with proper cosmological boundary conditions applied to the halo, gravothermal collapse would not occur within a Hubble time unless the cross section per DM mass is much larger than that required to match observations of halo profiles, or allowed by constraints from the Bullet Cluster [27, 38], $\sigma/m \sim 1 \text{ cm}^2/\text{g}$, where σ is the elastic scattering cross section and m is the DM mass.

Nevertheless, a subdominant component of strongly interacting DM could still initiate collapse of SMBH seeds while remaining consistent with such bounds, as was first claimed by PSS14, in a study limited to the effects of elastic scattering. More recently refs. [12, 24] investigated this general idea within the framework of mirror dark matter, assuming a large fraction $f \sim 0.2$ of dissipative SIDM. However the mechanism of collapse for mirror SIDM is not the gravothermal catastrophe, but rather a modified version of ordinary SMBH formation, accelerated by lowering the temperature of the dark sector.

In PSS14, the gravothermal collapse was modeled using a set of fluid equations for spherically symmetric distributions of mass, temperature, velocity dispersion and radiated heat. To implement the fluid approach with two

DM components, it was necessary for ref. PSS14 to make some simplifying assumptions: first that the initial density for the dominant component followed the usual Navarro-Frenk-White (NFW) profile [33], despite the possible influence of the SIDM component, and second that during the subsequent evolution the two densities should maintain the same profile shape, apart from the different normalizations. One might question whether these assumptions are really innocuous as regards the main features of gravothermal collapse, and to what extent they are borne out in a more exact treatment.

To overcome the limitations of the fluid approach, in this work we reconsider the problem by simulating the gravothermal collapse of a partially SIDM halo using an N -body code, initially developed in ref. [22]. We aim for a generic, model-independent treatment, exploring the effects of both elastic and dissipative scattering for the production of SMBHs. Our simplified models of dissipative interactions are designed to mimic energy loss through excitation followed by emission of dark radiation, or the formation of DM bound states.

In section 2, we review the process of gravothermal collapse, introduce the framework of two-component dark matter and summarize the previous results of ref. [37] (hereafter referred to as PSS14) on SMBH formation from elastically scattering DM. In section 3 we describe our N -body simulation methodology and present the results of simulations for an elastically scattering subdominant DM component. We show that it is not consistent to assume an initial NFW profile, and that one must instead simulate the full halo formation process. Moreover we show that elastic scattering cannot produce early SMBHs unless the cross section is large, $\sigma/m \gtrsim 10^3 \text{ cm}^2/\text{g}$. In section 4 we turn our attention to two simplified models of dissipative DM, which greatly speeds up the process of collapse, allowing smaller σ/m to explain high-redshift SMBHs. In section 5 we combine these results with a model of subsequent accretion to illustrate a range of possible working parameters in the three classes of interactions considered, comparing to the properties of three observed high-redshift SMBHs. We briefly consider the possible formation of black holes in smaller

systems, namely dwarf galaxies. Discussion of these results is given in sect. 6 and conclusions in sect. 7.

2 Gravothermal collapse and the gravothermal catastrophe

Gravothermal collapse can occur when heat and matter are transferred out of a virialized, gravitationally bound system of point masses. The virial theorem states that $U = -2T$, where U is potential and T kinetic energy, so that the total energy is $E = U + T = -T$. Such systems therefore have a negative specific heat: when energy is added they become less strongly gravitationally bound (and therefore the kinetic energy, or temperature, decreases), and when energy is removed they become more strongly bound, increasing the temperature.

In a halo with a negative radial temperature gradient, heat and mass will flow radially outward as it evolves towards equilibrium. This causes the inner part of the halo to shrink and further increase in temperature. If the specific heat of the outer halo is smaller than that of the inner, eventually the two regions reach equilibrium and the inner halo stops contracting. If it is larger, the process instead continues in a runaway fashion known as the gravothermal catastrophe [26]. Collapse occurs on a timescale related to the relaxation time t_r , the average time between collisions for a particle in the halo.

During the contraction, particles may eventually reach relativistic speeds and form a black hole through the radial instability. This occurs on a dynamical timescale,

$$t_d = r_c/v_{\text{rms}} \ll t_r, \tag{2}$$

where r_c is the core radius and v_{rms} the core r.m.s. speed. Once the core reaches relativistic speeds it very quickly collapses into a black hole [43].

This process requires the conduction of heat, which can happen through elastic scattering. A classic example is globular clusters, where heat is transferred by the gravitational interactions of stars, in particular when a higher-energy star is scattered outward to a larger radius while the lower-energy star falls inward toward the center of the halo (increasing its kinetic energy in the process). In contrast, the DM particles in a cold dark matter (CDM) halo are typically not massive enough for gravitational self-scattering to lead to gravothermal collapse. But the nongravitational self-interactions of SIDM can be much stronger, as we discuss next.

2.1 Self-interacting dark matter

While standard CDM is defined to be collisionless, self-interacting models have garnered much interest in recent years. DM scattering with cross sections per DM mass of order $\sigma/m \sim 1.0 \text{ cm}^2/\text{g}$ have been shown to ameliorate several problems in CDM small scale structure predictions, including the cusp/core and missing satellite problems [36, 39, 44, 48]. The former refers to the tendency of CDM simulations to produce ‘cuspy’ halos whose densities diverge at small radii [9, 14, 15, 21, 31, 33], in contrast to observations of dwarf and low surface brightness galaxies that indicate a flattening density profile at small radii (cored) [7, 10, 18, 19, 29, 41, 49]. The latter refers to the observation that CDM, while correctly predicting large scale structure and the number and distribution of large halos, predicts far more small satellite halos than are observed [30].

The required cross section for SIDM to solve the small-scale structure problems is of the same order as the upper bound coming from observations of the Bullet Cluster [27, 38],

$$\frac{\sigma}{m} \lesssim 0.7 \text{ cm}^2/\text{g}. \quad (3)$$

This is a pair of merging clusters that passed through each other. The collisional baryonic gas component, mapped through X-ray observations, is seen to have been stripped from the dominant DM component, mapped by gravitational lensing, creating an offset between the two. The observed offset would be diminished if DM self-interactions were too strong, since in that case the DM would tend to behave more similarly to the baryons. Measurements of the mass-to-light ratio of the clusters can be used to place an upper limit on the amount of DM that could have undergone such self-scatterings, which combined with theoretical estimates leads to the limit (3). Since larger values are needed for gravothermal collapse at early times [3, 22], we are motivated to consider models with two components of DM, that make it possible to evade (3).

2.2 Two-component dark matter

The Bullet Cluster bound (3) assumes that all the DM has the same self-interaction cross section, but if DM consists of two (or more) species, the smaller component could have a practically unconstrained value of σ/m . Stemming from observational uncertainties, it is estimated that the colliding DM subcluster could have lost as much as 23% of its mass in the collision [38]. One could then imagine that a fraction of arbitrarily strongly self-interacting DM as large as $f \sim 0.23$ is allowed. A stronger, complementary bound of $f < 0.05$ arises if the DM is significantly coupled to dark radiation, which could lead to dark acoustic oscillations in the matter power spectrum for large scale structure [11]. This however is more model-dependent and can be evaded if dark radiation is absent or suppressed.

In PSS14, a two-component scenario is investigated using a fluid approach, starting from an initial NFW profile and evolving it according to the gravothermal fluid equations. A generalized NFW profile can be defined as:

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{R_s}\right)^\gamma \left(1 + \frac{r}{R_s}\right)^{3-\gamma}}, \quad (4)$$

with R_s the scale radius and ρ_s the scale density. The parameter γ controls the extent to which the profile is cuspy or cored, with $\gamma = 1$ corresponding to the original NFW profile.

The results from the fluid formalism are given in terms of the relaxation time,

$$t_r = \frac{m}{af\sigma\rho_s v_s}, \quad (5)$$

where $a = 4/\sqrt{\pi}$ for hard-sphere interactions and v_s is the velocity dispersion at the characteristic radius,

$$v_s = \sqrt{4\pi G\rho_s R_s}, \quad (6)$$

For reference, we will ultimately be interested in halos with parameters mass $\sim 10^{12} M_\odot$ and NFW parameters $\rho_s \sim 10^{11} M_\odot/\text{kpc}^3$, $R_s \sim 1 \text{ kpc}$, leading to $v_s = 2300 \text{ km/s}$ and a relaxation time of

$$t_r = 0.28 \text{ Myr} \left(\frac{1 \text{ cm}^2/\text{g}}{f\sigma/m}\right) \left(\frac{10^{11} M_\odot/\text{kpc}^3}{\rho_s}\right)^{3/2} \left(\frac{1 \text{ kpc}}{R_s}\right). \quad (7)$$

The initial choice of an NFW profile is justified so long as the halo is optically thin at its scale radius,

$$\frac{\sigma f}{m} \lesssim \frac{1}{\rho_s R_s}. \quad (8)$$

This follows from demanding that the relaxation time (5) is greater than the dynamical timescale for the halo, R_s/v_s (analogous to that for the core, eq. (2)), ensuring that the initial halo structure is not strongly perturbed by the SIDM component. Eq. 8 implies that the optical depth of the halo to DM self-interactions is larger than the halo size. PSS14 finds that

- The gravothermal catastrophe occurs (and therefore the SMBH forms) after approximately $450 t_r$ regardless of cross section or SIDM fraction f . Therefore, the time taken depends only on the combination σf .
- The SMBH contains 2.5% of the SIDM component. Therefore for a halo with mass M_0 , $M_{\text{SMBH}} = 0.025 f M_0$.
- There is a region of parameter space in which SMBHs of the correct size may form early enough to accommodate observations ($z = 7$).

In the following we will obtain different results: the gravothermal catastrophe occurs after approximately $480 f^{-2} t_r$, greatly increasing the time until collapse for halos with a small SIDM fraction, and the SMBH contains a smaller fraction of the total SIDM component, $M_{\text{SMBH}}/M_{\text{SIDM}} \approx 0.6\%$. Due to the additional dependence on f of the time of collapse, we will find that although there is still a region of parameter space in which SMBHs of the correct size form by $z = 7$, the scattering cross sections required are much larger, unless dissipative interactions are introduced. For these larger elastic cross sections, the consistency requirement (8) is no longer satisfied, invalidating the assumption of an initial NFW halo.

3 N -body simulations of elastically scattering two-component dark matter

Both N -body and hydrodynamical simulations are frequently used to study the collapse of a DM halo. The former have the disadvantage of being quite computationally expensive, as the gravitational potential must be calculated for a large number ($N \sim 500000$ in our case) of particles, which must then be individually evolved forward in time. Scattering probabilities between neighbouring particles must be calculated, along with the resulting velocities if a scattering does occur [22].

Hydrodynamical simulations instead discretize space into a series of radial shells, keeping track of the amount of DM in each shell. This formulates the problem as a set of coupled partial differential equations. When key constants have been correctly calibrated, it can reproduce the results of N -body simulations [3, 22]. In this formalism it is difficult to accommodate two-component DM, which is crucial to the formation of high redshift SMBHs. Each DM component requires its own set of shells since the self-interactions differ between the two, but when computing the gravitational potential one would have to interpolate between the shells. Errors in interpolation grow quickly between successive timesteps, making this approach impractical. To circumvent these difficulties, ref. PSS14 applied the hydrodynamical simulation to the SIDM component only, while assuming a gravitational potential consistent with an NFW profile, *i.e.*, the SIDM component does not significantly affect the overall gravitational potential or distribution of CDM. The validity of this assumption is not obvious, motivating our use of N -body simulations that are not limited in this way.

3.1 Simulation of gravothermal collapse from an initial NFW halo

As a first step we employed the GADGET N -body simulation code [46, 47] to simulate the gravothermal collapse of an initial NFW halo; below we will consider formation of the halo from a primordial overdensity. GADGET is capable of simulating both noninteracting DM and baryonic gas. Baryonic simulations are much more computationally intensive; because of limited computer time we consider only DM. For ref. [22], GADGET was modified to include DM self-interactions between nearest neighbor particles, and the modified code is available online [20]. We further developed it to allow for dissipative (in addition to elastic) scattering of a subdominant DM component.

To test the code we first considered a single DM component with hard-sphere scattering, having a velocity independent cross section of $\sigma/m = 38 \text{ cm}^2/\text{g}$, to facilitate comparison with previous work [22, 37] that used this value. The initial conditions are that of an isolated NFW halo, as used in ref. [22], which has a total mass $M_0 = 10^{11} M_\odot$, and NFW parameters

$$R_s = 11.1 \text{ kpc}, \quad \rho_s = 1.49 \times 10^6 M_\odot \text{ kpc}^{-3} \quad (9)$$

and a maximum halo radius $R_{\text{max}} = 100 R_s$, at which we place a reflective boundary, reversing the radial velocity of particles which exceed this value. This is chosen to be a sufficiently large cut-off that it has no effect on the dynamics and evolution of the inner halo. From eq. (7), the relaxation time is $t_r = 0.37/f \text{ Gyr}$, which is too long to allow for SMBH formation by $z \sim 7$, for realistic values of f . We will consider more promising examples later, in section 5.

As the halo evolves, mass flows inward as expected for gravothermal collapse, until the central density begins to very rapidly increase and causes the timestep Δt to approach zero. This occurs because Δt goes inversely to the density in the modified code, $\Delta t \sim 1/\rho(r)$, and $\rho(0)$ diverges as the core collapses. For practical purposes, we identify the time at which Δt falls to 10^{-5} of its initial value as marking the onset of the gravothermal catastrophe, and formation of the black hole seed. At this moment, the inner part of the density profile increases quite suddenly, following a long period of slow evolution. The mass in the central region quickly contracts, leading to a flattening of $M(r)$, the mass enclosed within radius r , shown in figure 1). These qualitative observations are consistent with the results of hydrodynamical treatments, where the halo shows very little change over most of its history, followed by a sudden contraction [16, 37].

Our results roughly agree with those of PSS14 for the limiting case of single-component SIDM, $f = 1$, with the gravothermal catastrophe occurring after approximately 480 relaxation times (close to their result of $\sim 450 t_r$), as

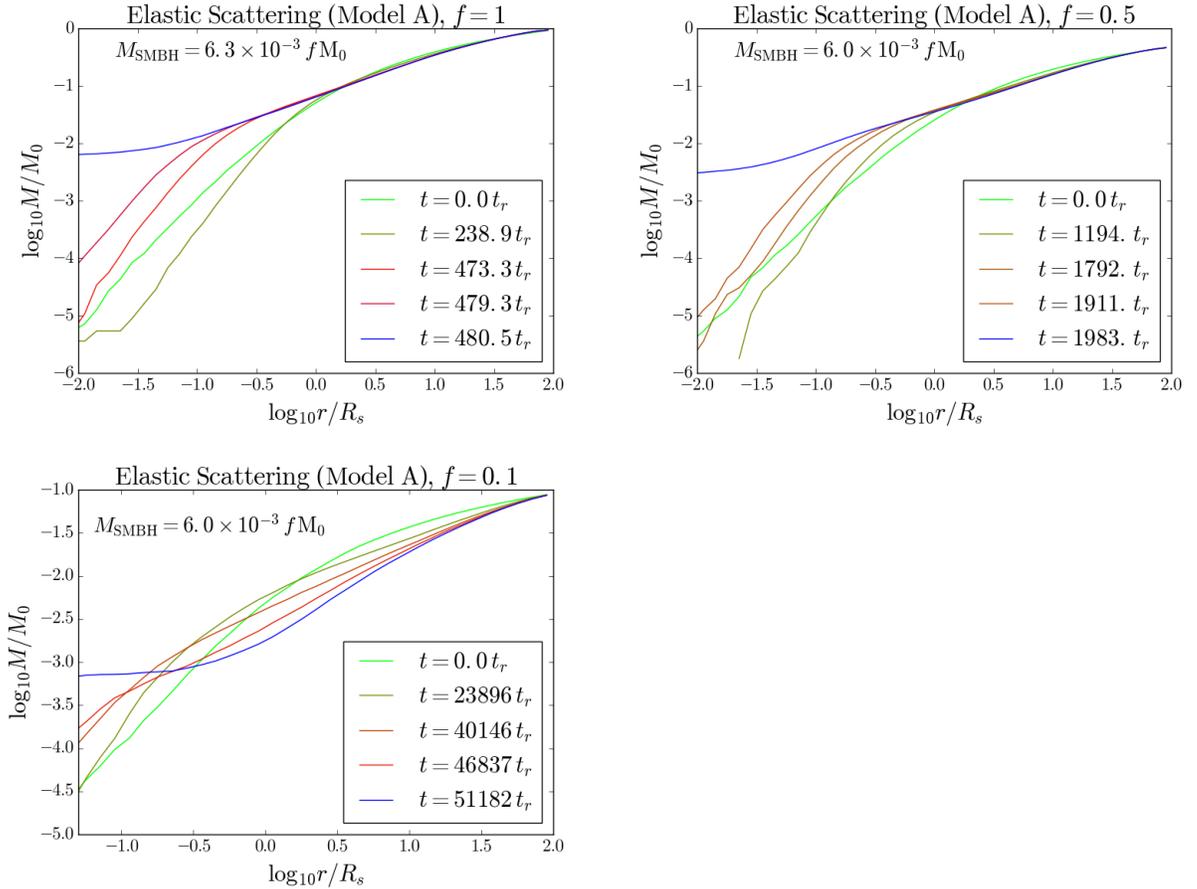


Figure 1: *Top left:* Halo evolution versus time for elastically scattering dark matter from an initial NFW halo with $f = 1$ and $\sigma/m = 38 \text{ cm}^2/\text{g}$. The plotted value is the mass enclosed at the given radius. The gravothermal catastrophe begins at $t_{\text{grav}} \cong 470 t_r$ and the black hole forms around $t_{\text{col}} \cong 482 t_r$. *Top right:* As above with $f = 0.5$ and $\sigma/m = 38 \text{ cm}^2/\text{g}$. Here we show only the SIDM component. $t_{\text{grav}} \cong 1790 t_r$ and $t_{\text{col}} \cong 1980 t_r$. *Bottom left:* As above with $f = 0.1$ and $\sigma/m = 38 \text{ cm}^2/\text{g}$. $t_{\text{grav}} \cong 4.0 \times 10^4 t_r$ and $t_{\text{col}} \cong 5.1 \times 10^4 t_r$.

can be seen in figure 1. But for smaller values $f = 0.5$ and $f = 0.1$, with the combination $f\sigma$ held constant, we find that the gravothermal catastrophe occurs after 1980 or 49000 relaxation times respectively. The results of these two simulations are shown in figure 1. The dependence upon f has a simple form, expressed by the empirical observation that if $f^3\sigma$ is held fixed, the time of SMBH formation remains nearly constant.¹

We thus find that the time of collapse does not simply scale with the relaxation time (5), but rather as $1/(f^3\sigma)$. This is at first surprising, since one would naively expect that the scattering rate of the SIDM component, proportional to $f^2\sigma$ (also at variance with the findings of PSS14), should control heat conduction through the halo. But this heat takes the form of kinetic energy of the SIDM particles, which also scales with their total mass, bringing an additional factor of f .²

The final fraction of the halo mass that becomes part of the supermassive black hole is $\sim 0.6\%$. This can be seen in figure 1, where the interior mass eventually levels off at small radii, showing that a fixed amount of the SIDM has collapsed to a central region smaller than our minimum resolvable radius.³

¹ To achieve greater numerical accuracy for small values of f , which would have large statistical fluctuations if the number of SIDM particles was simply reduced, we simulate the normal and SIDM components using equal numbers of particles, but with the SIDM mass adjusted so that the total SIDM mass is only a fraction f of the total DM mass, and σ/m is also rescaled accordingly. The N -body code is designed to treat these configurations as being physically equivalent.

² Another way of understanding the additional factor of f could be that gravothermal collapse proceeds through the formation of a shrinking core. If the SIDM is only a small fraction of the overall halo, the core cannot become as massive; its mass scales as f . This weakens the gravitational potential of the core proportionally to f , slowing its growth and resulting in a total proportionality of f^3 .

³ The gravitational smoothing, which roughly corresponds to the minimum resolvable radius, is taken to be $0.01R_s$ for $f = 1$ or $f = 0.5$, and $0.06R_s$ for

The SMBH mass, M_{SMBH} , is defined as the mass inside this radius at the time of its formation. The fraction of the SIDM that forms the SMBH is independent of f . This f -independence agrees with the results of PSS14, except that the final value of M_{SMBH} is smaller than their estimate of $2.5\% \times M_0$.

Combining these results, we can compare to the limit of $f\sigma/m \geq 0.336 \text{ cm}^2/\text{g}$ advocated by PSS14 to explain observations of high redshift SMBHs. This has some overlap with the constraint from eq. (8), that implies $f\sigma/m < 0.425 \text{ cm}^2/\text{g}$. Our numerical values scale with the relative number of relaxation times before collapse, $480/450 = 1.07$, but more importantly, our required value for SMBH formation scales as $f^3\sigma$, in contrast to the optical depth bound which goes as $f\sigma$. Since $f \lesssim 0.2$ from the Bullet Cluster constraint, there is no longer any overlap between the two inequalities. Hence the assumption of an initial NFW halo with common shape for both the CDM and SIDM components cannot be justified, since the SIDM scatterings could alter both distributions. This motivates our subsequent investigation, where we model the collapse of the halo to determine the impact of violating (8) on the initial halo profile.

3.2 Halo formation in a two-component universe

Since the assumption of an initial NFW profile may not be justified, we use GADGET to simulate the formation of a two-component halo using a simple spherical collapse model [23]. An initial spherically symmetric overdensity in

$f = 0.1$ to compensate for the much greater computational time required at small f .

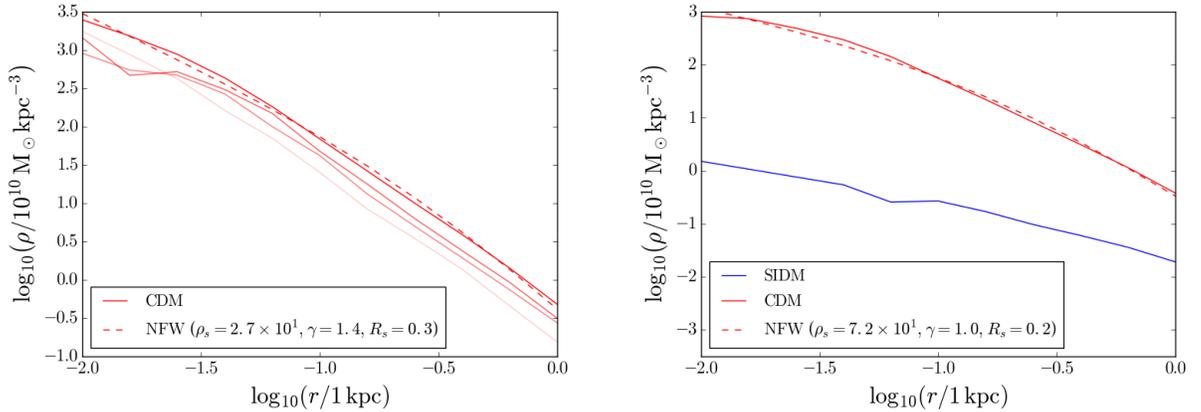


Figure 2: *Left:* Density profile of the dark matter halo at $z = 15$ for a single-component CDM halo (solid line) compared to the best-fit NFW halo (dotted line). The top solid line corresponds to a starting redshift of $z = 63$, while the others are at $z = 40, 30$, and 20 . The results are largely insensitive to the choice of starting redshift within this range. *Right:* Density profile of the dark matter halo at $z = 15$ for a two-component CDM halo with $f = 0.1, \sigma/m = 380 \text{ cm}^2/\text{g}$ (solid lines). The best-fit NFW profile for the CDM component is also shown (dashed line).

the early universe is given by

$$\rho(r) = \begin{cases} \rho_i > \rho_{\text{crit}}, & r < r_i \\ \rho_o < \rho_{\text{crit}}, & r_o > r > r_i \\ \rho_{\text{crit}}, & r > r_o \end{cases} \quad (10)$$

Well outside the overdense region, the universe behaves as a flat expanding universe, whereas inside it acts like a closed universe that undergoes expansion to a maximum local scale factor. The density contrast at the time of maximum expansion is $\rho/\rho_{\text{crit}} = 5.55$, after which the overdensity begins to collapse.

We simulate these conditions by implementing periodic boundary conditions within a cube of length $L = (2000 \text{ kpc})/(1+z)$ on each side. Within the cube is a spherical region of uniform density with $r_i = (372 \text{ kpc})/(1+z)$ and $\rho = 5.55 \rho_{\text{crit}}$. Outside the sphere, the density is chosen such that the total average density within the cube is ρ_{crit} . Due to the periodic boundary conditions, far from the overdense region the universe is effectively flat. The size

of the cube and overdensity are chosen such that the latter contains $10^{11} M_{\odot}$ of DM, facilitating comparison with our prior simulations, that used the same halo mass. We begin the simulation at $z = 63$,⁴ and the initial condition file is constructed using the GADGET initial condition generator [46, 47].

The simulation is allowed to continue until $z = 15$, by which point the halo will have virialized into an NFW profile. This expectation is borne out by the Milli-Millennium database [25], derived from Millennium Simulation [45] structure formation results for Λ CDM universes. The largest halo in the dataset at $z \sim 7$ has total mass $M_0 \gtrsim 10^{12} M_{\odot}$, and formed at $z \sim 15$. We therefore expect that smaller halos will also have virialized by $z = 15$. The results for both CDM and the two-component model are shown in figure 2. In the CDM-only simulation, the DM halo collapses into a NFW profile with $\gamma = 1.4$ (see equation 4) by $z = 15$.

We then performed a two-component simulation with $\sigma/m = 380 \text{ cm}^2/\text{g}$ and $f = 0.1$. The scaling law for the time of SMBH formation found above, $t \sim m/(f^3\sigma)$, shows that this is nearly the minimum value expected to produce a SMBH by $z = 7$, given our choice of halo parameters. Fig. 2(b) shows that the CDM component again collapses into a NFW profile by $z = 15$, but the influence of the SIDM leads to a less cuspy profile for the CDM with $\gamma = 1.0$. The SIDM component itself is far more cored, and is poorly fit by an NFW profile. Hence for the interesting region of parameter space where $f^3\sigma/m \gtrsim 1 \text{ cm}^2/\text{g}$, the full collapse of the halo must be simulated, rather than assuming an NFW profile. Given that the two components evolve very differently from each other, the hydrodynamical approach may not be well suited to modelling the gravothermal collapse of a two-component DM halo.

⁴ This value is sufficiently early that the halo virializes by $z \sim 15$. Other simulations were done beginning at redshifts of $z = 40, 30$, and 20 . The results are shown in figure 2 to be largely insensitive to the choice of starting redshift.

A proper treatment would require separate sets of mass shells for the two components, not implemented in PSS14.

4 Dissipative dark matter

We have found that large elastic cross sections $\sigma/m \gg 1 \text{ cm}^2/\text{g}$ are required for early SMBH formation, but one expects that gravothermal collapse could be accelerated by instead using dissipative (inelastic) scattering. Such processes can greatly increase the heat flow from the inner halo to the outer, hastening the collapse of the DM halo, for example through the emission of dark radiation. Ref. [12] showed that a subdominant mirror sector could effectively seed SMBHs during structure formation.⁵ There is one important caveat: if the dark radiation exerts a significant pressure on the collapsing halo, it can slow or even halt the collapse. In the present work we circumvent this potential issue, by assuming that any radiation or light particles produced during inelastic collisions are free to exit the halo: the optical thickness is larger than the halo size.

4.1 Dissipative dark matter models

In the interests of making a model-independent analysis, we consider two simplified models of inelastic scattering, that could plausibly capture the essential features of more realistic models. We will refer to them as models *B* and *C*, with *A* denoting simple elastic scattering.

In Model *B*, the SIDM loses a fixed quantity of kinetic energy in each scattering event, if sufficient energy is available. This can approximate the effect of creating an excited DM state, that subsequently decays by of radiation

⁵ Upper limits on dissipative scattering were obtained by ref. [16], in the context of a single component of DM.

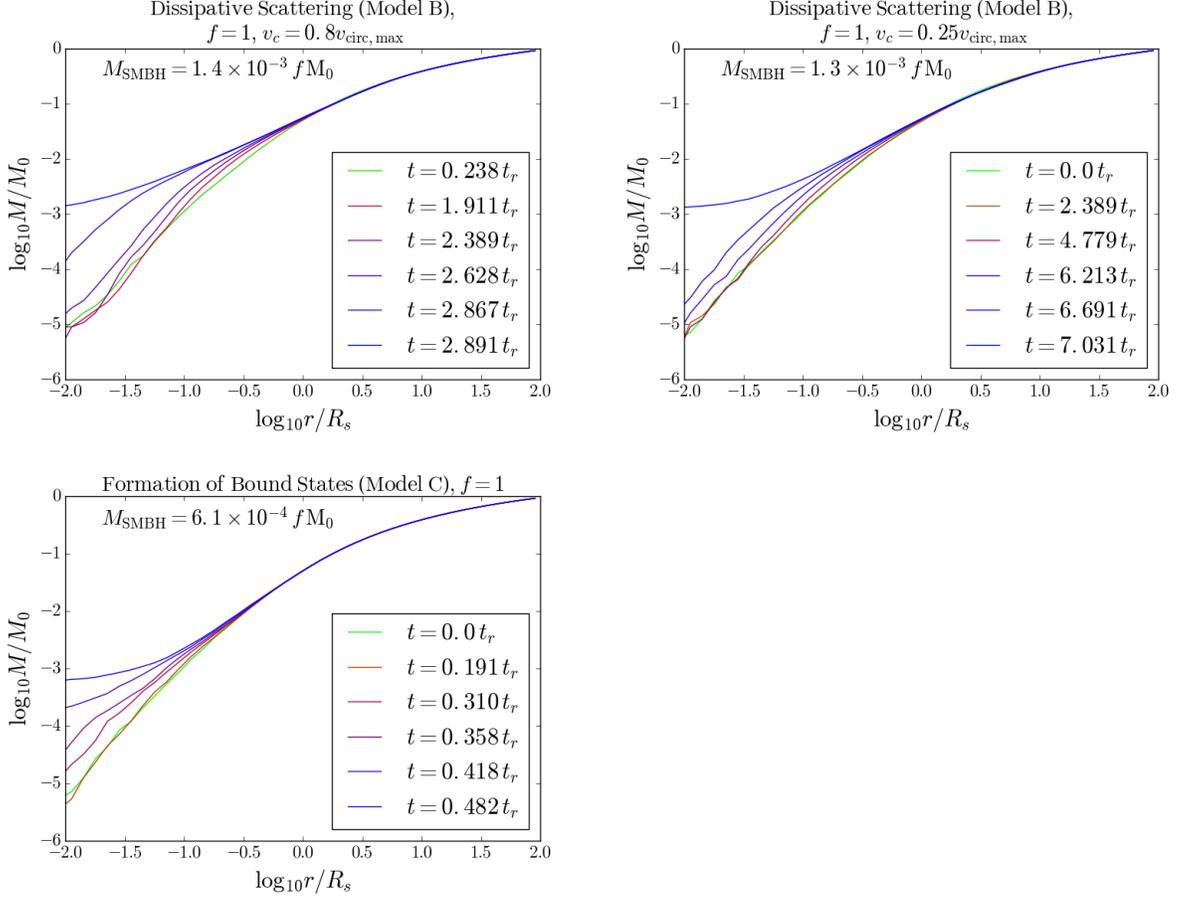


Figure 3: *Top left:* Halo mass interior to radius r as a function of time for Model B, assuming an initial NFW halo with $f = 1$, $\sigma/m = 38 \text{ cm}^2/\text{g}$ and $v_c = 0.8v_{\text{circ,max}}$. *Top right:* As above but with $v_c = 0.25v_{\text{circ,max}}$. *Bottom left:* As above for Model C (note the parameter v_c does not apply here).

or a light particle. For example, a dark atom collision could result in hyperfine excited states with fast radiative decays. Ref. [5] notes that selection rules require both atoms to become excited. Accordingly, we assume that the SIDM scatters elastically if its center of mass (c.m.) kinetic energy per particle is $< \Delta E$, and inelastically otherwise, in which case each particle loses energy equal to ΔE in the c.m. frame. The final c.m. speed of the SIDM particles after scattering inelastically is given by:

$$v_f = \begin{cases} v_i & v_i < v_c \\ \sqrt{v_i^2 - 2\Delta E/m} & v_i > v_c. \end{cases} \quad (11)$$

Model	Description	$\lambda = M_{\text{SMBH}}/M_{\text{SIDM}}$
<i>A</i>	Elastic scattering	6×10^{-3}
<i>B</i>	Inelastic above cut-off $v_c = \sqrt{2\Delta E/m}$, elastic below v_c	1×10^{-3}
<i>C</i>	Perfectly inelastic scattering, elastic scattering once bound state is formed	6×10^{-4}

Table 1: A summary of the three SIDM models considered in this work. The last column shows the results of the simulations for the approximate value of the fraction λ of total SIDM mass that forms the SMBH within each model (see figures 1 and 3).

The cutoff velocity $v_c = \sqrt{2\Delta E/m}$ plays an important role: to have any inelastic collisions, it must be less than the velocity dispersion v_s of the halos of interest. On the other hand, if v_c is too low, very little energy is lost in the collisions, making the inelasticity less effective. It could also lead to gravothermal collapse in dwarf galaxies or LSBs which could lead to cuspy DM profiles [16] contrary to observations. By choosing v_c to be greater than typical velocities of LSBs (~ 100 km/s, yet lower than that of our galaxies of interest (~ 500 km/s, see section 3.2), we can avoid formation of SMBHs in dwarf spheroidals. We discuss this in more detail below.

To make the simulations scale-independent, it is useful to express the v_c in units of the maximum circular velocity of the halo, which for an NFW halo extending to $\sim 100 R_s$ (as in our initial conditions in section 3) is

$$v_{\text{circ,max}} \cong 0.244\sqrt{GM_0/R_s} \quad (12)$$

In the second simplified model, denoted *C*, the DM interacts perfectly inelastically, as through forming a bound state, whose subsequent scatterings are assumed to be purely elastic, taking the same cross section for simplicity. This could mimic mirror dark matter models in which the formation of dark H₂ molecules is the primary mechanism for dissipating energy [12]. A summary of the models is given in table 1.

Having established the scaling of gravothermal collapse time with the SIDM fraction f in the previous section, we can reduce the noise associated

with large relative fluctuations in the scattering rate by taking $f = 1$, since this choice maximizes the probability for scattering. The results of three such dissipative simulations, starting from the same initial halo as in section 3, are shown in figure 3. We find that the SMBH forms within $\sim 3\text{-}7 t_r$ for Model *B* and $\sim 0.4 t_r$ for Model *C*, in contrast to the elastic scattering result $\sim 450 t_r$. The inelastic scenarios however result in smaller SMBHs, with mass approximately 0.1% of the SIDM total mass for Model *B* and 0.06% for Model *C*.

The time required for collapse is thus greatly reduced relative to that found for elastically scattering DM, consistent with the results found by refs. [12, 16]. However direct comparison with previous studies is hampered by key differences between the approaches. In ref. [12] the SIDM component was taken to be a perfect mirror sector of the Standard Model (SM) with fraction $f \cong 0.2$. Only because the mirror sector is taken to have a lower temperature than the SM, the mirror baryons can behave differently than their SM particle counterparts. The main dissipative process is formation of mirror H_2 molecules by $\text{H} + e^- \rightarrow \text{H}^- + \gamma'$ and $\text{H}^- + \text{H} \rightarrow \text{H}_2 + e^-$, which is sensitive to the dark photon temperature and cannot be adequately modeled by our simplified treatment.

Ref. [16] also considered the gravothermal collapse of a halo of dissipative DM, but for a single-component model with $f = 1$. Constraints on the cross section are derived by demanding that gravothermal collapse does not occur in dwarf galaxies and low surface brightness galaxies (LSBs), which would create cuspy density profiles unlike those that are observed. There are two means by which SIDM can escape these bounds while remaining relevant for SMBHs. The first is by taking the SIDM fraction to be sufficiently small, so that even if the SIDM component undergoes gravothermal collapse it will have little impact on the combined profile. Exactly how small it should be remains a problem for further investigation. The second is by adjusting the cutoff velocity v_c appropriately in a model with a threshold for inelasticity, like our

model B , as mentioned in section 3.2. For observed SMBHs, we are interested in halos of mass $M_0 \sim 10^{12}$ kpc that form by $z = 15$, giving a scale radius of $R_s \sim 1$ kpc, and a maximum circular velocity of $v_{\text{circ,max}} = 506$ km/s (see eq. (12)). For sufficiently large values of v_c , we can evade the bounds placed by ref. [16], as the constraints disappear for $v_c > 200$ km/s (their parameter v_{loss} coincides with v_c). We therefore require that $v_c \gtrsim 0.40 v_{\text{circ,max}}$.

We can also compare our predicted timescale for collapse t_{col} with that of ref. [16], which like us finds accelerated collapse from dissipative relative to elastic interactions, modeling dissipation similarly to our Model B . For $v_c = 13$ km/s, $M_0 = 8 \times 10^{10} M_\odot$, $R_s = 6.5$ kpc, they find t_{col} is reduced by a factor of 20 relative to elastic scattering. In our parametrization, this corresponds to $v_c = 0.23 v_{\text{circ,max}}$ for such a halo, while for the nearby value $v_c = 0.25 v_{\text{circ,max}}$ we find t_{col} is reduced by a greater factor of ~ 70 . The difference may be due to the fact that we allow the DM to collide elastically for $v < v_c$. Moreover in ref. [16] the DM scatters only if its velocity in the halo rest frame is $v > v_c$, whereas we impose the weaker requirement $v_{\text{rel}} > 2v_c$.

5 Comparison to observations

We now discuss simulations similar to those described in section 3 to constrain the parameters f and σ/m with respect to seeding SMBHs like those observed at high redshifts [4, 13, 32]. Because of limited computational resources, we restrict this preliminary study to a unique initial halo mass, subject to the varying scattering scenarios of our models A , B , C . The most favorable initial condition for explaining the observed SMBHs is a very massive halo that virializes sufficiently early.

The Milli-Millennium database includes a publicly available subset (1/512 fraction of the total volume) of the data [25] from the Millennium Simulation [45], a large-scale structure formation simulation using Λ CDM cosmology. The largest halo in the dataset at $z \sim 7$ has total mass $M_0 \gtrsim 10^{12} M_\odot$. Its

history suggests that it virializes by $z \sim 15$. We take this to be the most favorable candidate for early SMBH formation. It is atypical, having a much higher ρ_s and smaller R_s ,

$$\begin{aligned}\rho_s &\cong 2 \times 10^{10} M_\odot/\text{kpc}^3 \\ R_s &\cong 1 \text{ kpc}\end{aligned}\tag{13}$$

than halos of similar mass that form later. Eq. (12) gives a maximum circular velocity of

$$v_{\text{circ,max}} = 506 \text{ km/s}\tag{14}$$

We therefore simulate halo formation starting at $z = 63$ (see footnote 4) from an overdensity with mass $10^{12} M_\odot$, that will produce a halo of this mass before $z \sim 7$. Model *B* requires a choice of v_c , that we take to be $v_c = 0.25 v_{\text{circ,max}}$ and $v_c = 0.6 v_{\text{circ,max}}$, using eq. (14). The simulations are carried out on a grid in the plane of σ/m versus f , at $f = 0.01, 0.02, 0.05, 0.1, 0.5$ and integer values of $\log_{10} \sigma$, for models *A*, *B* and *C*. For each simulation, the redshift of SMBH formation is calculated, leading to contours labeled by z as shown in fig. 4.

Table 2 lists the properties of the three high- z SMBHs that we would like to explain by the simulations. To do so requires taking account of a degeneracy: the observed SMBH mass can be partly due to accretion after its initial formation. Taking the commonly assumed value $\epsilon_r = 0.1$ for the radiative efficiency in eq. 1, this growth is described by

$$M_{\text{SMBH}} = M_{\text{seed}} \exp\left(\frac{t_{\text{obs}} - t_{\text{col}}}{50 \text{ Myr}}\right).\tag{15}$$

where t_{col} is the time of collapse.⁶ In sect. 3 we saw that the black hole seed mass is a fixed fraction λ of the total SIDM mass, depending on the model;

⁶ This accretion rate could be affected by the dissipative interactions, an effect which we do not consider here but which has been explored in [35].

Galaxy	Redshift	M_{SMBH}
J1342+0928[4]	7.5413 ± 0.0007	$7.8_{-1.9}^{+3.3} \times 10^8$
J1120+0641[32]	7.085 ± 0.003	$2.0_{-0.7}^{+1.5} \times 10^9$
J2348-3054[13]	$6.889_{-0.006}^{+0.007}$	$2.1_{-0.5}^{+0.5} \times 10^9$

Table 2: The redshifts and masses of the three highest- z SMBHs, which we use to compare our results to observations.

see table 1):

$$M_{\text{seed}} \approx \lambda f M_0 \quad (16)$$

where M_0 is the total mass of the host halo.

Because of possible accretion, an observed SMBH can be explained by values of f and σ/m lying on curves, parametrized by the number of e -foldings of growth following the collapse. These are shown in fig. 4, with heavy dots marking successive e -foldings for the three observed SMBHs. Since the timescale for growth is faster than the Hubble rate, these curves cross the constant- z contours at a shallow angle. Points where the trajectories are terminated by stars indicate the limiting cases where no accretion has occurred and the observed mass is entirely due to the initial collapse. These curves should be interpreted as lower limits on the cross section needed to explain a given SMBH, since they assume that the rate of accretion saturates the Eddington limit, and we ignore disturbances such as mergers or tidal stripping by dwarf galaxies or sub-halos that could slow SMBH formation by revirializing the halo.

It is encouraging that the trajectories for the three different SMBHs are nearly coincident, which need not have been the case. It implies that all three systems could be explained by a single DM model, albeit with different amounts of accretion. In particular, J1342+0928 requires significantly less growth for given values of f and σ than the others because of its smaller mass. This is to be expected, since it was observed at a significantly higher redshift and thus had less time to accrete.

It is also apparent from the two plots for Model B that changing v_c has little difference other than shifting the required value of σ/m . Small changes

in v_c , therefore, are equivalent to a change in σ/m . For very large or small values, however, Model B becomes equivalent to Model A.

6 Discussion

We have demonstrated that large regions of SIDM parameter space can consistently explain early SMBH formation. As can be seen from figure 4, there is a large section of the parameter space in which SIDM can produce the observed high redshift supermassive black holes. In realistic settings, one could expect larger values of $f\sigma/m$ than our idealized simulations will be required, since accretion may be less efficient than assumed in eq. (15). For example gas may become depleted within the vicinity of the SMBH, interrupting accretion. This may explain why not all halos with the minimal properties develop early SMBHs, making them rare events. From a particle physics perspective, very large values of σ/m (compared for example to the Bullet Cluster limit) need not strain credulity. Atomic dark matter generically has σ of order $\gtrsim 10 a_0^2$, where a_0 is the mirror Bohr radius [8]. For an exact mirror of the standard model, this gives

$$\frac{\sigma}{m} \sim 10^8 \text{ cm}^2/\text{g}. \quad (17)$$

Although SIDM-induced gravothermal collapse is capable of forming very massive high- z SMBHs, it will not necessarily do so in all galaxies. Our simulations assumed an isolated halo corresponding to a galaxy in the field, but most galaxies form in more chaotic environments. Mergers and the stripping of the SIDM could slow or even halt the gravothermal collapse of the halo by injecting energy and revirializing the halo, leading to the delayed formation of a SMBH. Moreover we took a special case in which the halo forms unusually early.

The SIDM mechanism of SMBH formation is not mutually exclusive with others. For example, Population III stars are able to form large black holes

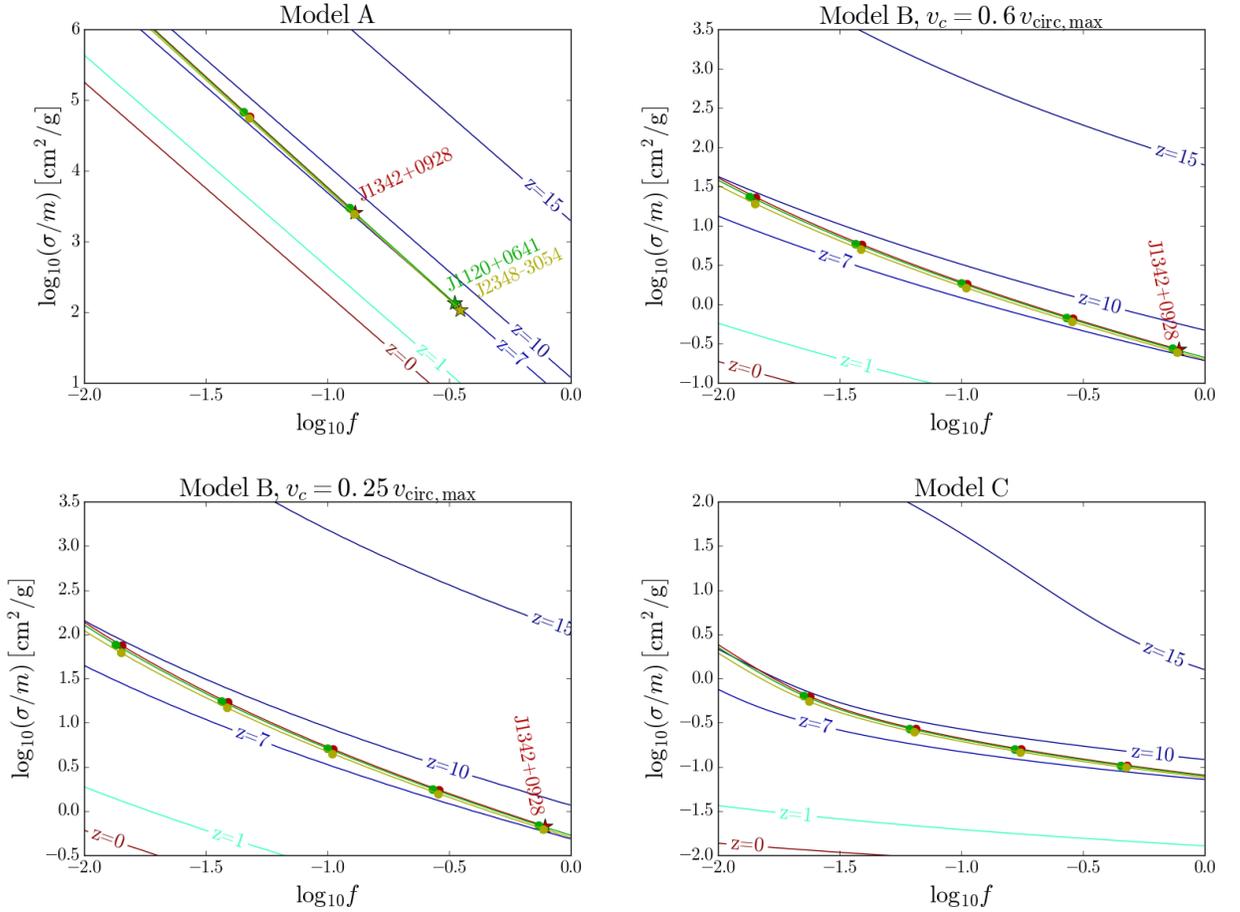


Figure 4: Contours in the f - σ plane showing the redshift of SMBH formation for a halo with $M_0 = 10^{15} M_\odot$ for Models A, B, and C. The red, green and yellow lines indicate the parameters compatible with the three observed high-redshift SMBHs. The (labelled) stars that terminate these lines show the parameters for which the SMBH is formed at the time of observation. The dots indicate the parameters for which a smaller SMBH seed initially forms, and then accretes mass by an integral number of e -foldings (see text for explanation). The SMBHs are massive enough relative to the total halo mass that for models B and C, even with $f = 1$ the SMBH must have undergone a modest amount of accretion, hence the absence of SMBH labels (and stars) for these plots. For Model B, the rightmost unlabelled dots correspond to 1 e -fold, whereas for Model C the rightmost dots correspond to 1 e -fold (J1342+0928) or 2 e -folds (J1120+064 and J2348-3054).

($\sim 100 M_\odot$) at high redshifts, but unless they form extraordinarily early, super-Eddington accretion is required to grow them to $\sim 10^9 M_\odot$ by redshift ~ 7 [6]. The gravothermal collapse of a SIDM cloud provides a natural mechanism for super-Eddington accretion, as the radiation pressure can be far smaller or absent in the dark sector. Simulation of the accretion of an SIDM halo onto a pre-existing SMBH could be interesting for a future study,

6.1 Connection to CDM small-scale structure

Another interesting question is whether two-component SIDM is capable of addressing the small-scale structure problems of CDM that provided one of the original motivations for (single-component) SIDM [36, 39, 44, 48]. Although one may suspect that with small enough fraction f there should be little effect on the central part of the DM density profile, this could depend upon σ/m for the SIDM component, and thus far no N -body studies have been carried out to address this question for typical halos. It is therefore possible that the scenario we present could also have an impact on the cusp-core problem.

In fact, our inelastic models B and C can produce SMBHs even for $\sigma < 1.0 \text{ cm}^2/\text{g}$ with $f = 1$, which obeys the Bullet Cluster constraint. (Although Model B ostensibly requires $f \lesssim 0.8$ to form J1342+0928 in figure 4, considering a slightly smaller initial halo would likely resolve this discrepancy.) Of course another simple way to combine the two mechanisms is to allow the principal DM component to have elastic $\sigma/m \sim 1 \text{ cm}^2/\text{g}$, which would match the usual requirements of one-component SIDM without invalidating our results, since the dominant component would experience gravothermal collapse only on a timescale of $500 f^{-2} t_r$, much greater than the Hubble time.

6.2 Dark disk formation

An aspect of dissipative matter that has been vigorously studied is its propensity to form a DM disk, that would overlap with the baryonic disk in Milky-Way-like galaxies [17, 28]. The existence of a dark disk in the Milky Way (MW) is strongly constrained by an analysis of recent *Gaia* data [42], though this constraint can be evaded if the local MW halo is out of equilibrium, for example through a recent tidal disruption. So far no N -body simulations of dissipative DM have been done to investigate formation of a dark disk.

Ref. [17] studied dark disk formation assuming the SIDM component consisted of ionized dark atoms, leading to dissipation via bremsstrahlung interactions amongst the massive dark particles and inverse Compton scattering off a dark photon background. Here we make an order of magnitude estimate for the timescale t_{dd} for dark disk formation, within our models B and C . Defining \mathcal{E} to be the kinetic energy density of the SIDM component and dP/dV to be the kinetic energy lost per unit time and volume.

$$t_{dd} = \frac{\mathcal{E}}{dP/dV}, \quad (18)$$

We take $dP/dV = 2n_{\chi'}^2 \sigma v \Delta E$, where $n_{\chi'}$ is the average SIDM number density inside the virial radius, v is its average velocity which we estimate as $v = \sqrt{3T_{\text{vir}}/m}$, $E = (3/2)T_{\text{vir}}n_{\chi'}$, and $2\Delta E$ is the average energy lost in each collision. The MW has a virial mass of approximately $1.5 \times 10^{12} M_{\odot}$ (taking the overdensity constant $\Delta_c = 200$) [34] corresponding to a virial radius of 240 kpc and hence

$$T_{\text{vir}} = \frac{1}{5} \frac{G_N M_{\text{vir}} m}{R_{\text{vir}}} = 5.9 \times 10^{-8} m. \quad (19)$$

Combining these relations we determine that for the Milky Way

$$t \approx 6 \times 10^3 \text{ Gyr} \left(\frac{0.1}{f} \right) \left(\frac{1 \text{ cm}^2/\text{g}}{\sigma/m} \right) \left(\frac{10^{-7}}{\Delta E/m} \right) \quad (20)$$

which can be longer than the age of the universe, 13.8 Gyr, for values of $f\Delta E/m$ that are compatible with early SMBH formation as discussed in

sec. 4.1. For example with $v_c = 200$ km/s to evade constraints of ref. [16], $\Delta E/m \cong 2 \times 10^{-7}$.

More realistic SIDM scenarios than our simplified models could have interactions between dark atoms and a dark radiation bath that might change this conclusion, but such effects are model-dependent and beyond the scope of this work. Such models must have a dark sector temperature substantially lower than that of the visible sector, to avoid dark acoustic oscillations and modifications of the matter power spectrum [11].

7 Conclusion

We have conducted the first N -body study of gravothermal collapse of a subdominant fraction f of self-interacting dark matter, coexisting with a dominant component of cold dark matter, as a means of seeding the early formation of supermassive black holes. This was motivated by technical limitations of an earlier hydrodynamical study, PSS14, that artificially required the normal CDM and SIDM components to maintain proportional density profiles, and which also confined its investigation to elastic scattering. Although we validate their results for the limiting case $f = 1$, we find an important difference in the timescale for collapse, going as f^{-3} instead of f^{-1} . Moreover we extended our study to include simplified models of dissipative interactions, showing that they are more effective than elastic scattering, at a fixed cross section σ .

We find that three observed SMBH's with masses $\sim 10^9 M_\odot$ and redshifts $z \sim 7$ can be simultaneously explained with reasonable values of $f\sigma/m$, allowing for different amounts of accretion subsequent to collapse. Moreover, if the scattering is dissipative, a possible choice is $f = 1$, $\sigma/m \cong 1 \text{ cm}^2/\text{g}$, which can be marginally consistent with constraints from the Bullet Cluster, while addressing puzzles about small scale structure in CDM, like the core-cusp problem.

There are several simplifying assumptions that could be improved upon in a future study. We incorporated simplified models of dissipation that are meant to capture the main features of more realistic models, where DM might form bound states, or excited states that decay by radiative emission. Our results are based upon a rare initial halo that is exceptionally large and early-forming, although still realistic in that it was taken from a large-scale cosmological simulation. We took an idealized model of subsequent accretion to describe the SMBH after initial collapse, assuming saturation of the Eddington limit, and ignoring complications such as mergers or collisions that could interrupt the SMBH growth.

More generally, the effects of dissipative interactions on structure formation is a subject that has not yet been explored in a very quantitative way, in the context of N -body simulations. Issues like the formation of a dark disk or distinctive effects of inelastic collisions on the small-scale structure problems represent interesting targets for future study.

Acknowledgments. We thank Guido D’Amico, JiJi Fan, Jun Koda, Roya Mohayaee, Paolo Panci, Jason Pollack, Matt Reece and Takashi Toma for very helpful discussions or correspondence. We acknowledge Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca) for supercomputing resources. JC thanks the Niels Bohr International Academy for its hospitality during the inception of this work, which was also supported by the McGill Space Institute and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- [1] K. Ahn and P. R. Shapiro. Formation and evolution of self-interacting dark matter haloes. *Mon. Not. Roy. Astron. Soc.*, 363:1092–1110, November 2005.
- [2] S. Balberg, S. L. Shapiro, and S. Inagaki. Self-Interacting Dark Matter Halos and the Gravothermal Catastrophe. *Astrophys. J.*, 568:475–487, April 2002.
- [3] Shmuel Balberg, Stuart L. Shapiro, and Shogo Inagaki. Selfinteracting dark matter halos and the gravothermal catastrophe. *Astrophys. J.*, 568:475–487, 2002.
- [4] Eduardo Banados et al. An 800-million-solar-mass black hole in a significantly neutral Universe at redshift 7.5. *Nature*, 553(7689):473–476, 2018.
- [5] Kimberly K. Boddy, Manoj Kaplinghat, Anna Kwa, and Annika H. G. Peter. Hidden Sector Hydrogen as Dark Matter: Small-scale Structure Formation Predictions and the Importance of Hyperfine Interactions. *Phys. Rev.*, D94(12):123017, 2016.
- [6] J. R. Bond, W. D. Arnett, and Bernard J. Carr. The Evolution and fate of Very Massive Objects. *Astrophys. J.*, 280:825–847, 1984.
- [7] Andreas Burkert and Joseph Silk. Dark baryons and rotation curves. *Astrophys. J.*, 488:L55, 1997.
- [8] James M. Cline, Zuowei Liu, Guy Moore, and Wei Xue. Scattering properties of dark atoms and molecules. *Phys. Rev.*, D89(4):043514, 2014.

- [9] Pedro Colin, A. Klypin, O. Valenzuela, and Stefan Gottlober. Dwarf dark matter halos. *Astrophys. J.*, 612:50–57, 2004.
- [10] Ramanath Cowsik, Kasey Wagoner, Emanuele Berti, and Amit Sircar. Internal dynamics and dynamical friction effects in the dwarf spheroidal galaxy in Fornax. *Astrophys. J.*, 699:1389–1394, 2009.
- [11] Francis-Yan Cyr-Racine, Roland de Putter, Alvise Raccanelli, and Kris Sigurdson. Constraints on Large-Scale Dark Acoustic Oscillations from Cosmology. *Phys. Rev.*, D89(6):063517, 2014.
- [12] Guido D’Amico, Paolo Panci, Alessandro Lupi, Stefano Bovino, and Joseph Silk. Massive Black Holes from Dissipative Dark Matter. *Mon. Not. Roy. Astron. Soc.*, 473(1):328–335, 2018.
- [13] G. De Rosa et al. Black hole mass estimates and emission-line properties of a sample of redshift $z < 6.5$ quasars. *Astrophys. J.*, 790:145, 2014.
- [14] Jürg Diemand, Marcel Zemp, Ben Moore, Joachim Stadel, and Marcella Carollo. Cusps in cold dark matter haloes. *Mon. Not. Roy. Astron. Soc.*, 364:665, 2005.
- [15] John Dubinski and R. G. Carlberg. The Structure of cold dark matter halos. *Astrophys. J.*, 378:496, 1991.
- [16] Rouven Essig, Hai-Bo Yu, Yi-Ming Zhong, and Samuel D. McDermott. Constraining Dissipative Dark Matter Self-Interactions. 2018.
- [17] JiJi Fan, Andrey Katz, Lisa Randall, and Matthew Reece. Double-Disk Dark Matter. *Phys. Dark Univ.*, 2:139–156, 2013.
- [18] Ricardo A. Flores and Joel R. Primack. Observational and theoretical constraints on singular dark matter halos. *Astrophys. J.*, 427:L1–4, 1994.
- [19] Gianfranco Gentile, Paolo Salucci, Uli Klein, and Gian Luigi Granato. NGC 3741: Dark halo profile from the most extended rotation curve. *Mon. Not. Roy. Astron. Soc.*, 375:199–212, 2007.

- [20] <https://github.com/junkoda/sidm-nbody>.
- [21] Anatoly Klypin, Andrey V. Kravtsov, James Bullock, and Joel Primack. Resolving the structure of cold dark matter halos. *Astrophys. J.*, 554:903–915, 2001.
- [22] Jun Koda and Paul R. Shapiro. Gravo-thermal collapse of isolated self-interacting dark matter haloes: N-body simulation versus the fluid model. *Mon. Not. Roy. Astron. Soc.*, 415:1125, 2011.
- [23] Edward W. Kolb and Michael S. Turner. The Early Universe. *Front. Phys.*, 69:1–547, 1990.
- [24] M. A. Latif, A. Lupi, D. R. G. Schleicher, G. D’Amico, P. Panci, and S. Bovino. Black hole formation in the context of dissipative dark matter. 2018.
- [25] Gerard Lemson. Halo and Galaxy Formation Histories from the Millennium Simulation: Public release of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the Lambda-CDM cosmogony. 2006.
- [26] D. Lynden-Bell and Roger Wood. The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems. *Mon. Not. Roy. Astron. Soc.*, 138:495, 1968.
- [27] Maxim Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56. *Astrophys. J.*, 606:819–824, 2004.
- [28] Matthew McCullough and Lisa Randall. Exothermic Double-Disk Dark Matter. *JCAP*, 1310:058, 2013.
- [29] B. Moore. Evidence against dissipationless dark matter from observations of galaxy haloes. *Nature*, 370:629, 1994.

- [30] B. Moore, S. Ghigna, F. Governato, G. Lake, Thomas R. Quinn, J. Stadel, and P. Tozzi. Dark matter substructure within galactic halos. *Astrophys. J.*, 524:L19–L22, 1999.
- [31] Ben Moore, Thomas R. Quinn, Fabio Governato, Joachim Stadel, and George Lake. Cold collapse and the core catastrophe. *Mon. Not. Roy. Astron. Soc.*, 310:1147–1152, 1999.
- [32] Daniel J. Mortlock et al. A luminous quasar at a redshift of $z = 7.085$. *Nature*, 474:616, 2011.
- [33] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal density profile from hierarchical clustering. *Astrophys. J.*, 490:493–508, 1997.
- [34] Fabrizio Nesti and Paolo Salucci. The Dark Matter halo of the Milky Way, AD 2013. *JCAP*, 1307:016, 2013.
- [35] Nadav Joseph Outmezguine, Oren Slone, Walter Tangarife, Lorenzo Ubaldi, and Tomer Volansky. Accretion of Dissipative Dark Matter onto Active Galactic Nuclei. *JHEP*, 11:005, 2018.
- [36] Annika H. G. Peter, Miguel Rocha, James S. Bullock, and Manoj Kaplinghat. Cosmological Simulations with Self-Interacting Dark Matter II: Halo Shapes vs. Observations. *Mon. Not. Roy. Astron. Soc.*, 430:105, 2013.
- [37] Jason Pollack, David N. Spergel, and Paul J. Steinhardt. Supermassive Black Holes from Ultra-Strongly Self-Interacting Dark Matter. *Astrophys. J.*, 804(2):131, 2015.
- [38] Scott W. Randall, Maxim Markevitch, Douglas Clowe, Anthony H. Gonzalez, and Marusa 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.

- [39] Miguel Rocha, Annika H. G. Peter, James S. Bullock, Manoj Kaplinghat, Shea Garrison-Kimmel, Jose Onorbe, and Leonidas A. Moustakas. Cosmological Simulations with Self-Interacting Dark Matter I: Constant Density Cores and Substructure. *Mon. Not. Roy. Astron. Soc.*, 430:81–104, 2013.
- [40] E. E. Salpeter. Accretion of Interstellar Matter by Massive Objects. *Astrophys. J.*, 140:796–800, 1964.
- [41] P. Salucci, F. Walter, and A. Borriello. The distribution of dark matter in galaxies: The Constant density halo around DDO 47. *Astron. Astrophys.*, 409:53–56, 2003.
- [42] Katelin Schutz, Tongyan Lin, Benjamin R. Safdi, and Chih-Liang Wu. Constraining a Thin Dark Matter Disk with Gaia. *Phys. Rev. Lett.*, 121(8):081101, 2018.
- [43] S. L. Shapiro and S. A. Teukolsky. The collapse of dense star clusters to supermassive black holes: the origin of quasars and AGNs. *Astrophys. J.*, 292:L41–L44, 1985.
- [44] David N. Spergel and Paul J. Steinhardt. Observational evidence for selfinteracting cold dark matter. *Phys. Rev. Lett.*, 84:3760–3763, 2000.
- [45] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005.
- [46] Volker Springel. The Cosmological simulation code GADGET-2. *Mon. Not. Roy. Astron. Soc.*, 364:1105–1134, 2005.
- [47] Volker Springel, Naoki Yoshida, and Simon D. M. White. GADGET: A Code for collisionless and gasdynamical cosmological simulations. *New*

- Astron.*, 6:79, 2001.
- [48] Sean Tulin and Hai-Bo Yu. Dark Matter Self-interactions and Small Scale Structure. *Phys. Rept.*, 730:1–57, 2018.
- [49] Frank C. van den Bosch and Rob A. Swaters. Dwarf galaxy rotation curves and the core problem of dark matter halos. *Mon. Not. Roy. Astron. Soc.*, 325:1017, 2001.
- [50] Marta Volonteri. Formation of Supermassive Black Holes. *Astron. Astrophys. Rev.*, 18:279–315, 2010.

CHAPTER 5

Discussion

In this thesis a wide range of models of dark matter have been discussed, demonstrating some of the alternatives to the standard WIMP paradigm of dark matter. Self-interacting dark matter provides a good explanation for the observed morphologies of small dark matter halos (such as those of dwarf galaxies and low surface brightness galaxies) by effectively smoothing the density distribution at the center of the halos. It also helps alleviate the too-big-to-fail problem in the same manner by decreasing the central densities of the most massive sub-halos predicted by CDM simulations, and for some models can also alleviate the missing satellites problem. But beyond the requirement of self-interaction cross sections of order $\sigma \sim 1 \text{ cm}^2/\text{g}$ and consistency with the lack of conclusive observed signals, there is a great deal of freedom in constructing self-interacting dark matter models, which can range from mirror dark matter models that are a complete copy of the Standard Model to entirely different gauge groups resulting in novel behaviour. Our research suggests that self-interacting dark matter could be the missing piece in solving other fundamental problems in physics and astrophysics, such as the origin of the baryon asymmetry, the source of the Galactic Center Excess or the existence of high-redshift supermassive black holes.

A model of dark matter has less interest if it is only consistent with the various bounds; it should also be motivated by unsolved problems in high energy theory or astrophysics. One such problem is that of baryogenesis, and our model presented in chapter 2 naturally solves both the issue of baryogenesis (in the Standard Model sector) and those associated with self-interacting dark

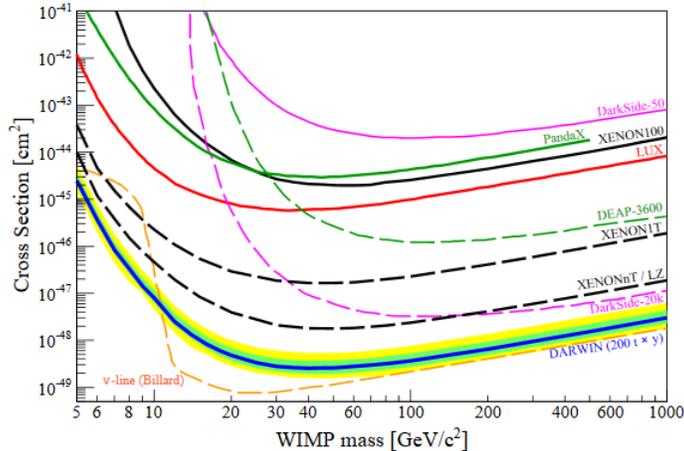


Figure 5–1: Expected sensitivity of the upcoming DARWIN direct detection experiment, with comparison to other past and current (solid lines) and upcoming (dashed lines) experiments. The dashed orange line represents the neutrino floor. Figure is reproduced from ref. [232]; for more information see the original paper and references therein.

matter, demonstrating that minimal models of atomic dark matter can lead to the production of the observed asymmetry in both sectors. Models such as this can be constrained through their interactions with the Standard Model, and these constraints are expected to strengthen, probing even weaker interactions, as experiments improve. One particularly powerful upcoming direct detection experiment is DARWIN, a 40 ton liquid xenon detector expected to begin operation in 2023 [232]. DARWIN could detect dark matter interacting through a kinetic mixing or Higgs portal like those described in our work, and in fact reaches sensitivities great enough to reach the so-called ‘neutrino floor’, at which point coherent neutrino scattering with nuclei produces a background that will make detection of dark matter significantly more challenging. The expected limits from DARWIN are shown in figure 5–1.

Another key constraint is that from dark acoustic oscillations, which we had not taken into account in our paper presented in section 2.4. Like the baryon acoustic oscillations, dark acoustic oscillations imprint a characteristic

scale on the matter power spectrum if the dark matter couples strongly to dark radiation. The lack thereof can be used to place constraints on the coupling of dark radiation with dark matter and on the temperature ratio between the dark and standard model sectors, ξ [158]. This bounds $\xi \leq 0.6$ if $\Sigma_{\text{DAO}} \leq 10^{-4.5}$ and $\xi \leq 0.2$ otherwise. Our model accommodates $\Sigma_{\text{DAO}} \leq 10^{-4.5}$ for $\alpha_d \lesssim 1$, but the constraint $\xi \leq 0.6$ is still challenging. In our model, we found that $\xi \geq 0.7$ for most of the parameter space, seeming to rule our model out. In fact, this does rule out the case of $R \neq 1$, as this result in $\xi \approx 0.71$ for the small values of the kinetic mixing parameter ϵ demanded by direct detection constraints (as per our discussion in section 5.1 of the paper). We therefore may only consider $R = 1$. We can derive new bounds on ϵ by demanding that $\xi \leq 0.6$, which according to equation (47) of the paper requires the effective degrees of freedom at the time of kinetic decoupling of the two sectors for the standard model to be large. Specifically, the requirement is:

$$\xi = \left(\frac{g_{*S,SM}^0 g_{*S,D}^{\text{dec}}}{g_{*S,SM}^{\text{dec}} g_{*S,D}^0} \right)^{1/3} \leq 0.6. \quad (5.1)$$

If we assume $g_{*S,SM}^0 = 3.94$, $g_{*S,D}^0 = 2$, $g_{*S,D}^{\text{dec}} = 2$, in other words assuming the dark matter (apart from the dark photon) has all frozen out prior to the decoupling of the two sectors, then $g_{*S,SM}^{\text{dec}} \geq 18$. If the sectors decouple after the strange quark but before the down and up quarks¹, giving $g_{*S,SM} \geq 31.25$ so long as $T_{\text{dec}} \gtrsim 10 \text{ MeV}$. Using the equation

$$T_{\text{dec}} = \frac{3 \times 10^{-6} \text{ eV}}{\epsilon^2} \left(\frac{m_{\mathbf{H}}}{\text{GeV}} \right), \quad (5.2)$$

¹ We in theory could get closer to $g_{*S,SM} = 18$ by having the the decoupling occur between the freeze-out of the up and down quarks, however as they freeze out at nearly the same temperature it is safer to choose a temperature scale prior to both

the resulting constraint is $\epsilon \leq 5.5 \times 10^{-7} \left(\frac{m_{\mathbf{H}}}{\text{GeV}}\right)^{1/2}$, which is significantly stronger than those we derive for direct detection. In order for the dark matter to have frozen out, we require that $T_{\text{dec}} \lesssim m_{\mathbf{H}}$, which results in $\epsilon \gtrsim 5.5 \times 10^{-8}$. These combined constraints only allow $m_{\mathbf{H}} \lesssim 100 \text{ GeV}$, creating only a small region of allowable parameter space.

Alternatively, we can consider the case where the dark atoms have not yet frozen out, but the vector bosons have. This sets $g_{*S,D}^{\text{dec}} = 10$. Equation 5.1 then requires $g_{*S,SM}^{\text{dec}} \geq 91$, which is accomplished only if the sectors decouple prior to the τ . Our constraints on ϵ become $\epsilon \leq 4.2 \times 10^{-8} \left(\frac{m_{\mathbf{H}}}{\text{GeV}}\right)^{1/2}$ (from the requirement that decoupling occurs before the τ freezes out), $\epsilon \leq 5.5 \times 10^{-8}$ (from the requirement that the dark atoms have not yet frozen out), and $\epsilon \gtrsim 5.5 \times 10^{-8} \left(\frac{m_{\mathbf{H}}}{m_B}\right)^{1/2}$ from the requirement that the lightest vector boson has (m_B here is the mass of the lightest vector boson). Though this does create an allowable region of parameter space, we have $\epsilon_{\text{max}}/\epsilon_{\text{min}} = \sqrt{m_B/m_{\mathbf{H}}}$, meaning that even if the lightest vector boson is 100 times more massive than the dark atoms, ϵ may only vary over a factor of 10. This demonstrates the power of dark acoustic oscillations in constraining models with dark radiation; had we taken these constraints into account we would have had a much smaller region of allowed parameter space.

Our work with the Galactic Center excess has shown that the tension with observations of dwarf spheroidal galaxies can be relieved for some models, such as self-interacting or p -wave annihilating dark matter. Since the writing of these papers, the debate over the GCE has continued, and the dark matter explanation has lost some prominence. Like many potential dark matter signals, the potential for explaining the signal well within the Standard Model (in this case with a previously undiscovered population millisecond pulsars)

weakens the case for a dark matter signal, though the matter is far from settled. In spite of this, the work is still relevant as in both cases we re-evaluate the measured J -factors for cored profiles and for p -wave annihilating dark matter respectively for the dwarf spheroidal galaxies for which sufficient data exists. In the former case in particular this was done directly from the stellar kinematic data, allowing us to place updated profile-dependent bounds on 18 dwarf spheroidal galaxies which could be used in future dark matter searches.

Future searches for dark matter through indirect detection include the Cherenkov Telescope Array, which by 2023 is expected to reach unprecedented sensitivity in the 100 GeV – 1 TeV range [233] (see figure 5–2) by observing the Galactic halo. As Fermi-LAT continues to run, the bounds on dark matter annihilations in dwarf spheroidals will also strengthen [78, 234], and systematic uncertainties in observations of the Galactic Center excess will decrease as our understanding of these environments evolves. For example, Fermi-LAT data should resolve millisecond pulsars in the future if they exist in quantities great enough to account for the Galactic Center excess [234]. In the case of p -wave annihilating dark matter, future searches of galaxy clusters are expected to yield results, as the annihilation cross section is enhanced by a factor of ~ 100 , as their velocity dispersions are typically a factor of ~ 10 greater, as mentioned in our paper in section 3.2.

The discoveries of new ultra-faint dwarf galaxies, such as Reticulum II [110], grant new targets in the search for dark matter annihilations [78]. Ultra-faint dwarfs in particular can be almost entirely composed of dark matter, making them ideal places to search for a signal. A potential Gamma-ray signal has been observed in Reticulum II using the Fermi-LAT telescope which is compatible with the GCE [204, 205]. These new dwarf galaxies, along with continued observation and the deployment of new telescopes with increased

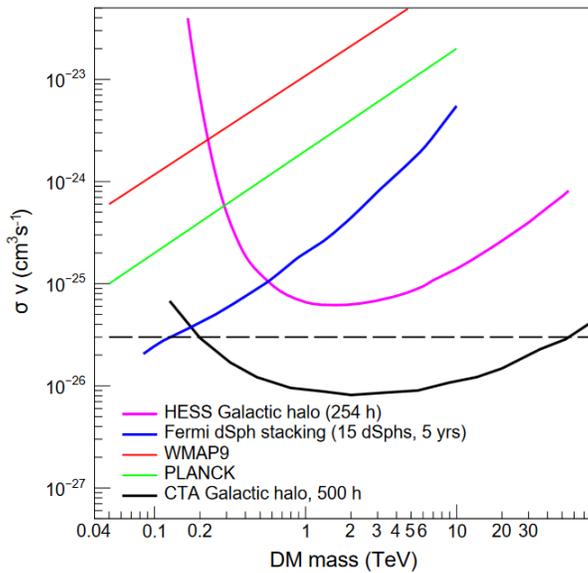


Figure 5–2: Expected sensitivity of the Cherenkov Telescope Array from observations of the Galactic halo, compared to current limits from HESS and Fermi-LAT. These values assume annihilation to W^+W^- , but are comparable in magnitude to those for annihilation to $b\bar{b}$. The dashed line at $\sigma v = 3.0 \times 10^{26} \text{ cm}^3/\text{s}$ represents the value required for the so-called ‘WIMP miracle’. Figure is reproduced from ref. [233]; for more information see the original paper and references therein.

sensitivity will likely allow the GCE to either be fully ruled out or confirmed. If confirmed, it would be the first conclusive dark matter signal discovered, and could constitute further evidence in favour of self-interacting dark matter if it is found to be attenuated in dwarf spheroidal galaxies due to their more cored profiles.

Depending on the model, these searches could observe signals (or confirm existing ones such as the Galactic Center excess), or we may eventually find that dark matter exists in a hidden sector that can only be probed gravitationally or at energies unreachable by collider experiments. If so, the best method to determine the nature of dark matter is to examine its effects on astrophysical events and structure formation through gravitational probes. One

possible probe is supermassive black holes. Though non-dark matter explanations for high-redshift SMBHs exist, each relies on very specific conditions in order to form the seed black holes early enough that they may accrete enough matter to reach billions of solar masses by redshifts of 7 or more. Since the early history of structure formation is still relatively unknown, this makes such explanations somewhat tenuous. We have demonstrated that self-interacting dark matter, especially dissipative models of self-interacting dark matter, provide an alternative explanation. Future work along these lines could involve determining the frequency with which the SMBHs form and comparing their mass distribution and number to those observed, or integrating more realistic models of dissipative SIDM.

Even if dark matter is not fully responsible for the formation of SMBHs, or has no significant role in the formation of the SMBH seeds, it may have a role in their accretion. SMBHs may prove to be a significant probe in the study of SIDM models as future research investigates the possible relationship between SIDM and SMBHs. For example, a massive black hole seed could be formed from a Population III star and accrete self-interacting dark matter to accelerate its growth beyond the Eddington limit. For dissipative interactions, this might not even require a scattering cross section beyond the bullet cluster bounds. Gravitational probes of dark matter have, to date, been the only to succeed, and the addition of a new one would greatly propel the field of dark matter astrophysics and particle physics.

A related line of potential future work is related to double disk dark matter [160]. A subdominant component of self-interacting dark matter which scatters dissipatively through a light mediator is able to form a second ‘dark’ disk parallel to that of the Milky Way. This formation of the double disk cannot be seen in our simulations, likely due to the somewhat simplistic spherical

collapse model chosen, which results in a halo that has no overall angular momentum. Additionally, we do not consider baryonic matter at all in our simulation, further reducing the tendency of a dark disk to form. More realistic N-body simulations of dissipative dark matter (specifically what we term Model B) which include baryonic matter and a more accurate history of structure formation could show the formation of a dark disk. These models enhance prospects for both direct and indirect detection by localizing dark matter within the Galactic Plane [160].

A more general probe of self-interacting dark matter, however, continues to be its effects on structure of halos, for example through the creation of a core. More detailed knowledge of the way in which dark matter is distributed in halos will provide further insight into its nature, and one of the most promising current projects related to this field is the *Gaia* satellite.

Gaia is an ambitious project which is sampling more than a billion stars in the Milky Way to reconstruct the distribution of structure in our galaxy [235]. This will provide a unique probe into the exact distribution of dark matter in our galaxy, allowing the dark matter density profile and local dark matter density to be determined with much greater accuracy than ever before, and potentially giving further insight into dark matter self-interactions².

With powerful new experiments such as DARWIN, CTA, and *Gaia* placing new constraints on dark matter detection and the distribution of dark matter in the Milky Way, the prospects for dark matter research warrant

² For example, if the SIDM fraction is large enough in the double disk scenario, this would have a measurable effect on stellar kinematic data and could be seen by *Gaia* [161].

optimism, and we will be better poised than ever to confirm or rule out self-interacting models of dark matter.

CHAPTER 6

Conclusion

We have studied several topics in dark matter, for the most part centered around self-interacting dark matter models. In chapter 2 we took a very theoretical approach to construct a minimal nonabelian hidden sector in which dark atoms could form. In this model the dark sector is composed of a dark SU(2) gauge field which is broken to U(1) through a dark Higgs boson. The dark matter is composed of a Dirac fermion doublet which is charged under the U(1) symmetry, allowing it to form bound states. The dark photon mixes kinematically with the standard model one, allowing the dark matter to acquire an electric millicharge (transforming under the Standard Model U(1) symmetry).

In chapter 3 we presented two papers related to the Galactic Center excess in which we attempted to reconcile the purported dark matter annihilation signal with the lack thereof in dwarf spheroidal galaxies. We showed that the tension between these two observations can be alleviated if a velocity-dependent annihilation cross section is considered, reducing the cross section by a factor of ~ 10 in dwarf spheroidals relative to that in the Milky Way. Since it is more challenging to explain the relic abundance of dark matter if it is p -wave annihilating, we demonstrate that the correct relic abundance can be produced if the dark matter is produced non-thermally from a decaying predecessor. The GCE-dwarf spheroidal tension can also be alleviated if the dwarf spheroidal galaxies are assumed to have sufficiently cored profiles, and we quantify the degree to which dwarf spheroidals must be cored in order to reduce the expected signal sufficiently. Our results are consistent with the

cored profiles produced by self-interacting dark matter and with observations of dwarf galaxies.

In section 4, we showed that self-interacting dark matter may present an alternative explanation for supermassive black holes which are observed at very high redshifts. We presented a two-component model in which the bulk of the dark matter is made up of either cold dark matter or self-interacting dark matter with a small scattering cross section, while a small fraction of it ($f \lesssim 0.1$) is composed of self-interacting dark matter with a very high self-interaction cross section. We showed using N-body simulations that for sufficiently high elastic self-interaction cross section the gravothermal catastrophe is able to form supermassive black holes within a short enough time to account for observed high-redshift AGN, and this process is further accelerated if the dark matter scatters inelastically or forms bound states.

References

- [1] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **571**, A1 (2014), 1303.5062.
- [2] F. Zwicky, *Helv. Phys. Acta* **6**, 110 (1933), [Gen. Rel. Grav.41,207(2009)].
- [3] V. C. Rubin and W. K. Ford, Jr., *Astrophys. J.* **159**, 379 (1970).
- [4] K. C. Freeman, *Astrophys. J.* **160**, 811 (1970).
- [5] M. Roberts and A. H. Rots, *Astron. Astrophys.* **26**, 483 (1973).
- [6] M. Roberts and R. Whitehurst, *Astrophys. J.* **201**, 327 (1975).
- [7] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr., *Astrophys. J.* **238**, 471 (1980).
- [8] A. Bosma, *Astron. J.* **86**, 1825 (1981).
- [9] T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi, *Astrophys. J.* **295**, 305 (1985).
- [10] Y. Sofue and V. Rubin, *Ann. Rev. Astron. Astrophys.* **39**, 137 (2001), astro-ph/0010594.
- [11] SDSS, M. Tegmark *et al.*, *Astrophys. J.* **606**, 702 (2004), astro-ph/0310725.
- [12] SDSS, D. J. Eisenstein *et al.*, *Astrophys. J.* **633**, 560 (2005), astro-ph/0501171.
- [13] SDSS, W. J. Percival *et al.*, *Mon. Not. Roy. Astron. Soc.* **401**, 2148 (2010), 0907.1660.
- [14] WMAP, E. Komatsu *et al.*, *Astrophys. J. Suppl.* **192**, 18 (2011), 1001.4538.

- [15] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **594**, A13 (2016), 1502.01589.
- [16] R. H. Cyburt, *Phys. Rev.* **D70**, 023505 (2004), astro-ph/0401091.
- [17] M. Kawasaki, K. Kohri, and T. Moroi, *Phys. Lett.* **B625**, 7 (2005), astro-ph/0402490.
- [18] M. Kawasaki, K. Kohri, and T. Moroi, *Phys. Rev.* **D71**, 083502 (2005), astro-ph/0408426.
- [19] K. Jedamzik, *Phys. Rev.* **D74**, 103509 (2006), hep-ph/0604251.
- [20] P. Gondolo and G. Gelmini, *Nucl. Phys.* **B360**, 145 (1991).
- [21] G. Jungman, M. Kamionkowski, and K. Griest, *Phys. Rept.* **267**, 195 (1996), hep-ph/9506380.
- [22] T. Nihei, L. Roszkowski, and R. Ruiz de Austri, *JHEP* **05**, 063 (2001), hep-ph/0102308.
- [23] G. Steigman, B. Dasgupta, and J. F. Beacom, *Phys. Rev.* **D86**, 023506 (2012), 1204.3622.
- [24] J. L. Feng and J. Kumar, *Phys. Rev. Lett.* **101**, 231301 (2008), 0803.4196.
- [25] H. Baer, K.-Y. Choi, J. E. Kim, and L. Roszkowski, *Phys. Rept.* **555**, 1 (2015), 1407.0017.
- [26] H. Goldberg, *Phys. Rev. Lett.* **50**, 1419 (1983), [,219(1983)].
- [27] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, *Nucl. Phys.* **B238**, 453 (1984), [,223(1983)].
- [28] N. G. Deshpande and E. Ma, *Phys. Rev.* **D18**, 2574 (1978).
- [29] L. Roszkowski, E. M. Sessolo, and S. Trojanowski, *Rept. Prog. Phys.* **81**, 066201 (2018), 1707.06277.
- [30] R. D. Peccei and H. R. Quinn, *Phys. Rev. Lett.* **38**, 1440 (1977), [,328(1977)].

- [31] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
- [32] H. Primakoff, Phys. Rev. **81**, 899 (1951).
- [33] ADMX, N. Du *et al.*, Phys. Rev. Lett. **120**, 151301 (2018), 1804.05750.
- [34] L. F. Abbott and P. Sikivie, Phys. Lett. **B120**, 133 (1983), [URL(1982)].
- [35] E. W. Kolb and M. S. Turner, Front. Phys. **69**, 1 (1990).
- [36] R. L. Davis, Phys. Rev. **D32**, 3172 (1985).
- [37] S. Chang, C. Hagmann, and P. Sikivie, Phys. Rev. **D59**, 023505 (1999), hep-ph/9807374.
- [38] Y. B. Zel'dovich and I. D. Novikov, Astronomicheskii Zhurnal **43**, 758 (1966).
- [39] S. Hawking, Mon. Not. Roy. Astron. Soc. **152**, 75 (1971).
- [40] T. Harada, C.-M. Yoo, and K. Kohri, Phys. Rev. **D88**, 084051 (2013), 1309.4201, [Erratum: Phys. Rev.D89,no.2,029903(2014)].
- [41] S. Clesse and J. Garca-Bellido, Phys. Rev. **D92**, 023524 (2015), 1501.07565.
- [42] L. Wyrzykowski *et al.*, Mon. Not. Roy. Astron. Soc. **413**, 493 (2011), 1012.1154.
- [43] E. Mediavilla, J. Jimnez-Vicente, J. A. Muoz, H. Vives-Arias, and J. Caldern-Infante, Astrophys. J. **836**, L18 (2017), 1702.00947.
- [44] J. Luis Bernal, N. Bellomo, A. Raccanelli, and L. Verde, JCAP **1710**, 052 (2017), 1709.07465.
- [45] Virgo, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016), 1602.03837.
- [46] Virgo, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 241103 (2016), 1606.04855.
- [47] VIRGO, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **118**, 221101 (2017), 1706.01812.

- [48] Virgo, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **119**, 141101 (2017), 1709.09660.
- [49] S. Bird *et al.*, Phys. Rev. Lett. **116**, 201301 (2016), 1603.00464.
- [50] S. Clesse and J. Garca-Bellido, Phys. Dark Univ. **15**, 142 (2017), 1603.05234.
- [51] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Phys. Rev. Lett. **117**, 061101 (2016), 1603.08338.
- [52] M. Zumalacarregui and U. Seljak, (2017), 1712.02240.
- [53] A. K. Drukier, K. Freese, and D. N. Spergel, Phys. Rev. **D33**, 3495 (1986).
- [54] D. N. Spergel, Phys. Rev. **D37**, 1353 (1988).
- [55] J. H. Davis, Int. J. Mod. Phys. **A30**, 1530038 (2015), 1506.03924.
- [56] R. Bernabei *et al.*, Eur. Phys. J. **C73**, 2648 (2013), 1308.5109.
- [57] XENON, E. Aprile *et al.*, Phys. Rev. Lett. **121**, 111302 (2018), 1805.12562.
- [58] PICO, C. Amole *et al.*, (2019), 1902.04031.
- [59] T. Marrodn Undagoitia and L. Rauch, J. Phys. **G43**, 013001 (2016), 1509.08767.
- [60] A. Berlin, P. Gratia, D. Hooper, and S. D. McDermott, Phys. Rev. **D90**, 015032 (2014), 1405.5204.
- [61] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, Phys. Rev. D **80**, 023505 (2009), 0905.0003.
- [62] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, Phys. Rev. D **80**, 043526 (2009), 0906.1197.
- [63] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, Phys. Rev. D **84**, 027302 (2011), 1106.1528.
- [64] L. Bergstrom, Phys. Lett. **B225**, 372 (1989).

- [65] L. Bergstrom and J. Kaplan, *Astropart. Phys.* **2**, 261 (1994), hep-ph/9403239.
- [66] L. Bergstrom and P. Ullio, *Nucl. Phys.* **B504**, 27 (1997), hep-ph/9706232.
- [67] AMS, M. Aguilar *et al.*, *Phys. Rev. Lett.* **110**, 141102 (2013).
- [68] AMS, M. Aguilar *et al.*, *Phys. Rev. Lett.* **117**, 091103 (2016).
- [69] M.-Y. Cui, Q. Yuan, Y.-L. S. Tsai, and Y.-Z. Fan, *Phys. Rev. Lett.* **118**, 191101 (2017), 1610.03840.
- [70] AMS, N. Zimmermann, *PoS EPS-HEP2017*, 090 (2017).
- [71] M. Cirelli *et al.*, *JCAP* **1103**, 051 (2011), 1012.4515, [Erratum: *JCAP*1210,E01(2012)].
- [72] M. Regis and P. Ullio, *Phys. Rev.* **D78**, 043505 (2008), 0802.0234.
- [73] L. Dugger, T. E. Jeltema, and S. Profumo, *JCAP* **12**, 015 (2010), 1009.5988.
- [74] H.E.S.S., F. Aharonian *et al.*, *Phys. Rev. Lett.* **97**, 221102 (2006), astro-ph/0610509, [Erratum: *Phys. Rev. Lett.*97,249901(2006)].
- [75] L. Goodenough and D. Hooper, (2009), 0910.2998.
- [76] J. Diemand, M. Kuhlen, and P. Madau, *Astrophys. J.* **657**, 262 (2007), astro-ph/0611370.
- [77] W. J. G. de Blok and S. S. McGaugh, *Mon. Not. Roy. Astron. Soc.* **290**, 533 (1997), astro-ph/9704274.
- [78] Fermi-LAT, M. Ackermann *et al.*, (2015), 1503.02641.
- [79] ATLAS, M. Aaboud *et al.*, *Phys. Rev.* **D94**, 032005 (2016), 1604.07773.
- [80] CMS, A. M. Sirunyan *et al.*, *JHEP* **07**, 014 (2017), 1703.01651.
- [81] A. Boveia and C. Doglioni, *Ann. Rev. Nucl. Part. Sci.* **68**, 429 (2018), 1810.12238.
- [82] ATLAS, T. A. collaboration, (2018).

- [83] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996), astro-ph/9508025.
- [84] J. Dubinski and R. G. Carlberg, *Astrophys. J.* **378**, 496 (1991).
- [85] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **490**, 493 (1997), astro-ph/9611107.
- [86] B. Moore, T. R. Quinn, F. Governato, J. Stadel, and G. Lake, *Mon. Not. Roy. Astron. Soc.* **310**, 1147 (1999), astro-ph/9903164.
- [87] A. Klypin, A. V. Kravtsov, J. Bullock, and J. Primack, *Astrophys. J.* **554**, 903 (2001), astro-ph/0006343.
- [88] P. Colin, A. Klypin, O. Valenzuela, and S. Gottlober, *Astrophys. J.* **612**, 50 (2004), astro-ph/0308348.
- [89] J. Diemand, M. Zemp, B. Moore, J. Stadel, and M. Carollo, *Mon. Not. Roy. Astron. Soc.* **364**, 665 (2005), astro-ph/0504215.
- [90] A. Burkert, *IAU Symp.* **171**, 175 (1996), astro-ph/9504041, [*Astrophys. J.*447,L25(1995)].
- [91] R. A. Flores and J. R. Primack, *Astrophys. J.* **427**, L1 (1994), astro-ph/9402004.
- [92] B. Moore, *Nature* **370**, 629 (1994).
- [93] A. Burkert and J. Silk, *Astrophys. J.* **488**, L55 (1997), astro-ph/9707343.
- [94] F. C. van den Bosch and R. A. Swaters, *Mon. Not. Roy. Astron. Soc.* **325**, 1017 (2001), astro-ph/0006048.
- [95] P. Salucci, F. Walter, and A. Borriello, *Astron. Astrophys.* **409**, 53 (2003), astro-ph/0206304.
- [96] G. Gentile, P. Salucci, U. Klein, and G. L. Granato, *Mon. Not. Roy. Astron. Soc.* **375**, 199 (2007), astro-ph/0611355.
- [97] R. Cowsik, K. Wagoner, E. Berti, and A. Sircar, *Astrophys. J.* **699**, 1389 (2009), 0904.0451.

- [98] J. F. Navarro, V. R. Eke, and C. S. Frenk, *Mon. Not. Roy. Astron. Soc.* **283**, L72 (1996), astro-ph/9610187.
- [99] A. Pontzen and F. Governato, *Mon. Not. Roy. Astron. Soc.* **421**, 3464 (2012), 1106.0499.
- [100] S. S. McGaugh and W. J. G. de Blok, *Astrophys. J.* **499**, 41 (1998), astro-ph/9801123.
- [101] W. J. G. de Blok, S. S. McGaugh, A. Bosma, and V. C. Rubin, *Astrophys. J.* **552**, L23 (2001), astro-ph/0103102.
- [102] G. Kauffmann, S. D. M. White, and B. Guiderdoni, *Mon. Not. Roy. Astron. Soc.* **264**, 201 (1993).
- [103] B. Moore *et al.*, *Astrophys. J.* **524**, L19 (1999), astro-ph/9907411.
- [104] J. Zavala *et al.*, *Astrophys. J.* **700**, 1779 (2009), 0906.0585.
- [105] S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky, *Astrophys. J.* **742**, 16 (2011), 1005.1289.
- [106] J. D. Simon and M. Geha, *Astrophys. J.* **670**, 313 (2007), 0706.0516.
- [107] S. Walsh, B. Willman, and H. Jerjen, *Astron. J.* **137**, 450 (2009), 0807.3345.
- [108] J. S. Bullock, K. R. Stewart, M. Kaplinghat, and E. J. Tollerud, *Astrophys. J.* **717**, 1043 (2010), 0912.1873.
- [109] DES, K. Bechtol *et al.*, *Astrophys. J.* **807**, 50 (2015), 1503.02584.
- [110] S. E. Koposov, V. Belokurov, G. Torrealba, and N. W. Evans, *Astrophys. J.* **805**, 130 (2015), 1503.02079.
- [111] DES, A. Drlica-Wagner *et al.*, *Astrophys. J.* **813**, 109 (2015), 1508.03622.
- [112] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, *Mon. Not. Roy. Astron. Soc.* **415**, L40 (2011), 1103.0007.

- [113] E. Papastergis, R. Giovanelli, M. P. Haynes, and F. Shankar, *Astron. Astrophys.* **574**, A113 (2015), 1407.4665.
- [114] S. Colombi, S. Dodelson, and L. M. Widrow, *Astrophys. J.* **458**, 1 (1996), astro-ph/9505029.
- [115] P. Bode, J. P. Ostriker, and N. Turok, *Astrophys. J.* **556**, 93 (2001), astro-ph/0010389.
- [116] S. Horiuchi *et al.*, *Mon. Not. Roy. Astron. Soc.* **456**, 4346 (2016), 1512.04548.
- [117] M. R. Lovell *et al.*, *Mon. Not. Roy. Astron. Soc.* **420**, 2318 (2012), 1104.2929.
- [118] M. R. Lovell *et al.*, *Mon. Not. Roy. Astron. Soc.* **439**, 300 (2014), 1308.1399.
- [119] V. Avila-Reese, P. Colin, O. Valenzuela, E. D’Onghia, and C. Firmani, *Astrophys. J.* **559**, 516 (2001), astro-ph/0010525.
- [120] P. Colin, O. Valenzuela, and V. Avila-Reese, *Astrophys. J.* **673**, 203 (2008), 0709.4027.
- [121] R. M. Dunstan, K. N. Abazajian, E. Polisensky, and M. Ricotti, (2011), 1109.6291.
- [122] A. V. Maccio, S. Paduroiu, D. Anderhalden, A. Schneider, and B. Moore, *Mon. Not. Roy. Astron. Soc.* **424**, 1105 (2012), 1202.1282.
- [123] A. Schneider, D. Anderhalden, A. Maccio, and J. Diemand, *Mon. Not. Roy. Astron. Soc.* **441**, 6 (2014), 1309.5960.
- [124] R. E. Angulo, O. Hahn, and T. Abel, *Mon. Not. Roy. Astron. Soc.* **434**, 3337 (2013), 1304.2406.
- [125] U. Seljak, A. Makarov, P. McDonald, and H. Trac, *Phys. Rev. Lett.* **97**, 191303 (2006), astro-ph/0602430.
- [126] M. Viel *et al.*, *Phys. Rev. Lett.* **100**, 041304 (2008), 0709.0131.

- [127] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, *Phys. Rev.* **D88**, 043502 (2013), 1306.2314.
- [128] V. Iri *et al.*, *Phys. Rev.* **D96**, 023522 (2017), 1702.01764.
- [129] N. Dalal and C. S. Kochanek, *Astrophys. J.* **572**, 25 (2002), astro-ph/0111456.
- [130] S. Tulin and H.-B. Yu, *Phys. Rept.* **730**, 1 (2018), 1705.02358.
- [131] D. N. Spergel and P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000), astro-ph/9909386.
- [132] M. Rocha *et al.*, *Mon. Not. Roy. Astron. Soc.* **430**, 81 (2013), 1208.3025.
- [133] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, *Mon. Not. Roy. Astron. Soc.* **430**, 105 (2013), 1208.3026.
- [134] R. Dave, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, *Astrophys. J.* **547**, 574 (2001), astro-ph/0006218.
- [135] M. Vogelsberger, J. Zavala, and A. Loeb, *Mon. Not. Roy. Astron. Soc.* **423**, 3740 (2012), 1201.5892.
- [136] J. Zavala, M. Vogelsberger, and M. G. Walker, *Monthly Notices of the Royal Astronomical Society: Letters* **431**, L20 (2013), 1211.6426.
- [137] A. Loeb and N. Weiner, *Phys. Rev. Lett.* **106**, 171302 (2011), 1011.6374.
- [138] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, *JCAP* **0907**, 004 (2009), 0905.3039.
- [139] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, *Astrophys. J.* **679**, 1173 (2008), 0704.0261.
- [140] M. Markevitch *et al.*, *Astrophys. J.* **567**, L27 (2002), astro-ph/0110468.
- [141] D. Clowe, A. Gonzalez, and M. Markevitch, *Astrophys. J.* **604**, 596 (2004), astro-ph/0312273.
- [142] M. Markevitch *et al.*, *Astrophys. J.* **606**, 819 (2004), astro-ph/0309303.
- [143] D. Clowe *et al.*, *Astrophys. J.* **648**, L109 (2006), astro-ph/0608407.

- [144] B. Patt and F. Wilczek, (2006), hep-ph/0605188.
- [145] B. Holdom, Phys.Lett. **B166**, 196 (1986).
- [146] H. An, X. Ji, and L.-T. Wang, JHEP **07**, 182 (2012), 1202.2894.
- [147] Z. Berezhiani, D. Comelli, and F. L. Villante, Phys. Lett. **B503**, 362 (2001), hep-ph/0008105.
- [148] A. Y. Ignatiev and R. Volkas, Phys.Rev. **D68**, 023518 (2003), hep-ph/0304260.
- [149] C. Burgess, M. Pospelov, and T. ter Veldhuis, Nucl.Phys. **B619**, 709 (2001), hep-ph/0011335.
- [150] D. Feldman, Z. Liu, and P. Nath, Phys.Rev. **D75**, 115001 (2007), hep-ph/0702123.
- [151] M. Pospelov, A. Ritz, and M. B. Voloshin, Phys.Lett. **B662**, 53 (2008), 0711.4866.
- [152] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys.Rev. **D79**, 015014 (2009), 0810.0713.
- [153] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, JCAP **1005**, 021 (2010), 0909.0753.
- [154] D. E. Kaplan, G. Z. Krnjaic, K. R. Rehermann, and C. M. Wells, JCAP **1110**, 011 (2011), 1105.2073.
- [155] J. M. Cline, Z. Liu, and W. Xue, Phys.Rev. **D85**, 101302 (2012), 1201.4858.
- [156] J. M. Cline, Z. Liu, and W. Xue, Phys. Rev. **D87**, 015001 (2013), 1207.3039.
- [157] F.-Y. Cyr-Racine and K. Sigurdson, Phys. Rev. **D87**, 103515 (2013), 1209.5752.
- [158] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli, and K. Sigurdson, Phys. Rev. **D89**, 063517 (2014), 1310.3278.

- [159] J. M. Cline, Z. Liu, G. Moore, and W. Xue, Phys. Rev. **D89**, 043514 (2014), 1311.6468.
- [160] J. Fan, A. Katz, L. Randall, and M. Reece, Phys.Rev.Lett. **110**, 211302 (2013), 1303.3271.
- [161] J. Fan, A. Katz, L. Randall, and M. Reece, Phys.Dark Univ. **2**, 139 (2013), 1303.1521.
- [162] J. Fan, A. Katz, and J. Shelton, JCAP **1406**, 059 (2014), 1312.1336.
- [163] J. M. Cline, Z. Liu, G. Moore, and W. Xue, Phys. Rev. **D90**, 015023 (2014), 1312.3325.
- [164] R. Foot and S. Vagnozzi, Phys. Rev. **D91**, 023512 (2015), 1409.7174.
- [165] K. Petraki, L. Pearce, and A. Kusenko, JCAP **1407**, 039 (2014), 1403.1077.
- [166] J. M. Cline, Y. Farzan, Z. Liu, G. D. Moore, and W. Xue, Phys. Rev. **D89**, 121302 (2014), 1404.3729.
- [167] R. Foot, Int. J. Mod. Phys. **A29**, 1430013 (2014), 1401.3965.
- [168] R. Foot and S. Vagnozzi, Phys. Lett. **B748**, 61 (2015), 1412.0762.
- [169] M. T. Frandsen, F. Sannino, I. M. Shoemaker, and O. Svendsen, JCAP **1405**, 033 (2014), 1403.1570.
- [170] I. Garca Garca, R. Lasenby, and J. March-Russell, (2015), 1505.07410.
- [171] L. Pearce, K. Petraki, and A. Kusenko, Phys. Rev. **D91**, 083532 (2015), 1502.01755.
- [172] J. M. Cline and P. Scott, JCAP **1303**, 044 (2013), 1301.5908.
- [173] M. Ackermann *et al.*, Phys. Rev. **D91**, 122002 (2015).
- [174] S. Davidson, B. Campbell, and D. Bailey, Phys. Rev. D **43**, 2314 (1991).
- [175] E. Golowich and R. W. Robinett, Phys. Rev. D **35**, 391 (1987).
- [176] CMS, S. Chatrchyan *et al.*, Phys. Rev. **D87**, 092008 (2013), 1210.2311.

- [177] A. Haas, C. S. Hill, E. Izaguirre, and I. Yavin, Phys. Lett. **B746**, 117 (2015), 1410.6816.
- [178] LUX Collaboration, D. Akerib *et al.*, Phys.Rev.Lett. **112**, 091303 (2014), 1310.8214.
- [179] SuperCDMS, R. Agnese *et al.*, Phys. Rev. Lett. **112**, 241302 (2014), 1402.7137.
- [180] SuperCDMS, R. Agnese *et al.*, Phys. Rev. Lett. **112**, 041302 (2014), 1309.3259.
- [181] S. Davidson, E. Nardi, and Y. Nir, Phys. Rept. **466**, 105 (2008), 0802.2962.
- [182] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967), [Usp. Fiz. Nauk161,no.5,61(1991)].
- [183] M. Fukugita and T. Yanagida, Phys. Lett. **B174**, 45 (1986).
- [184] J. Choquette and J. M. Cline, Phys. Rev. **D92**, 115011 (2015), 1509.05764.
- [185] D. Hooper and L. Goodenough, Phys. Lett. **B697**, 412 (2011), 1010.2752.
- [186] D. Hooper and T. Linden, Phys. Rev. **D84**, 123005 (2011), 1110.0006.
- [187] K. N. Abazajian and M. Kaplinghat, Phys. Rev. **D86**, 083511 (2012), 1207.6047, [Erratum: Phys. Rev.D87,129902(2013)].
- [188] B. Zhou *et al.*, Phys. Rev. **D91**, 123010 (2015), 1406.6948.
- [189] F. Calore, I. Cholis, and C. Weniger, JCAP **1503**, 038 (2015), 1409.0042.
- [190] T. Daylan *et al.*, Phys. Dark Univ. **12**, 1 (2016), 1402.6703.
- [191] Fermi-LAT, M. Ajello *et al.*, (2015), 1511.02938.
- [192] C. Karwin, S. Murgia, T. M. P. Tait, T. A. Porter, and P. Tanedo, Phys. Rev. **D95**, 103005 (2017), 1612.05687.
- [193] K. N. Abazajian and R. E. Keeley, (2015), 1510.06424.

- [194] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, (2015), 1506.05124.
- [195] R. Bartels, S. Krishnamurthy, and C. Weniger, (2015), 1506.05104.
- [196] R. M. O’Leary, M. D. Kistler, M. Kerr, and J. Dexter, (2015), 1504.02477.
- [197] Fermi-LAT, M. Ackermann *et al.*, *Astrophys. J.* **840**, 43 (2017), 1704.03910.
- [198] Fermi-LAT, M. Ajello *et al.*, Submitted to: *Astrophys. J.* (2017), 1705.00009.
- [199] D. Haggard, C. Heinke, D. Hooper, and T. Linden, (2017), 1701.02726.
- [200] E. Carlson and S. Profumo, *Phys. Rev.* **D90**, 023015 (2014), 1405.7685.
- [201] J. Petrovi, P. D. Serpico, and G. Zaharija, *JCAP* **1410**, 052 (2014), 1405.7928.
- [202] D. Gaggero, M. Taoso, A. Urbano, M. Valli, and P. Ullio, *JCAP* **1512**, 056 (2015), 1507.06129.
- [203] DES, Fermi-LAT, A. Drlica-Wagner *et al.*, Submitted to: *Astrophys. J.* (2015), 1503.02632.
- [204] A. Geringer-Sameth *et al.*, (2015), 1503.02320.
- [205] D. Hooper and T. Linden, (2015), 1503.06209.
- [206] J. D. Simon, arXiv e-prints , arXiv:1901.05465 (2019), 1901.05465.
- [207] J. Choquette, J. M. Cline, and J. M. Cornell, *Phys. Rev.* **D94**, 015018 (2016), 1604.01039.
- [208] A. W. Graham, D. Merritt, B. Moore, J. Diemand, and B. Terzic, *Astron. J.* **132**, 2685 (2006), astro-ph/0509417.
- [209] J. Choquette, *Phys. Rev.* **D97**, 043017 (2018), 1705.09676.
- [210] COBE, G. F. Smoot *et al.*, *Astrophys. J.* **396**, L1 (1992).

- [211] J. E. Greene and L. C. Ho, *Astrophys. J.* **630**, 122 (2005), astro-ph/0508335.
- [212] W. Zuo *et al.*, *Astrophys. J.* **799**, 189 (2015), 1412.2438.
- [213] E. Banados *et al.*, *Nature* **553**, 473 (2018), 1712.01860.
- [214] E. E. Salpeter, *Astrophys. J.* **140**, 796 (1964).
- [215] M. Tegmark *et al.*, *Astrophys. J.* **474**, 1 (1997), astro-ph/9603007.
- [216] J. R. Bond, W. D. Arnett, and B. J. Carr, *Astrophys. J.* **280**, 825 (1984).
- [217] P. C. Clark, S. C. O. Glover, and R. S. Klessen, *AIP Conf. Proc.* **990**, 79 (2008), 0706.0613.
- [218] C. F. McKee and J. C. Tan, *Astrophys. J.* **681**, 771 (2008), 0711.1377.
- [219] V. Bromm and A. Loeb, *Astrophys. J.* **596**, 34 (2003), astro-ph/0212400.
- [220] D. J. Eisenstein and A. Loeb, *Astrophys. J.* **443**, 11 (1995), astro-ph/9401016.
- [221] M. C. Begelman, M. Volonteri, and M. J. Rees, *Mon. Not. Roy. Astron. Soc.* **370**, 289 (2006), astro-ph/0602363.
- [222] J. H. Wise, M. J. Turk, and T. Abel, *Astrophys. J.* **682**, 745 (2008), 0710.1678.
- [223] M. Dijkstra, Z. Haiman, A. Mesinger, and S. Wyithe, *Mon. Not. Roy. Astron. Soc.* **391**, 1961 (2008), 0810.0014.
- [224] K. Omukai, R. Schneider, and Z. Haiman, *Astrophys. J.* **686**, 801 (2008), 0804.3141.
- [225] D. Bernadetta and V. Marta, *Astrophys. J.* **694**, 302 (2009), 0810.1057.
- [226] J. Pollack, D. N. Spergel, and P. J. Steinhardt, *Astrophys. J.* **804**, 131 (2015), 1501.00017.
- [227] S. Balberg, S. L. Shapiro, and S. Inagaki, *Astrophys. J.* **568**, 475 (2002), astro-ph/0110561.
- [228] A. P. Lightman and S. L. Shapiro, *Rev. Mod. Phys.* **50**, 437 (1978).

- [229] G. D’Amico, P. Panci, A. Lupi, S. Bovino, and J. Silk, *Mon. Not. Roy. Astron. Soc.* **473**, 328 (2018), 1707.03419.
- [230] J. Koda and P. R. Shapiro, *Mon. Not. Roy. Astron. Soc.* **415**, 1125 (2011), 1101.3097.
- [231] V. Springel, N. Yoshida, and S. D. M. White, *New Astron.* **6**, 79 (2001), astro-ph/0003162.
- [232] DARWIN, J. Aalbers *et al.*, *JCAP* **1611**, 017 (2016), 1606.07001.
- [233] CTA Consortium, A. Morselli, *PoS ICRC2017*, 921 (2018), 1709.01483.
- [234] Fermi-LAT, E. Charles *et al.*, *Phys. Rept.* **636**, 1 (2016), 1605.02016.
- [235] Gaia, T. Prusti *et al.*, *Astron. Astrophys.* **595**, A1 (2016), 1609.04153.