Signals of Particle Dark Matter

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:9527316">http://nrs.harvard.edu/urn-3:HUL.InstRepos:9527316</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Abstract

This thesis explores methods of detecting dark matter particles, with some emphasis on several dark matter models of current interest. Detection in this context means observation of an experimental signature correlated with dark matter interactions with Standard Model particles. This includes recoils of nuclei or electrons from dark matter scattering events, and direct or indirect observation of particles produced by dark matter annihilation.
Contents

Title Page ................................................................. i
Abstract ................................................................. iii
Table of Contents ......................................................... iv
Citations to Previously Published Work ................................ vii
Acknowledgments ......................................................... viii

1 Introduction ......................................................... 1
  1.1 Structure of this thesis ............................................. 3

2 Inelastic Dark Matter: An *Experimentum Crucis* .............. 6
  2.1 Introduction ..................................................... 6
    2.1.1 The DAMA/iDM Scenario ..................................... 8
    2.1.2 Advantages of Directional Sensitivity ..................... 9
  2.2 Experimental Setup ............................................ 11
    2.2.1 Experimental Design ...................................... 13
  2.3 Recoil Spectrum ................................................ 15
  2.4 Sensitivity ...................................................... 17
    2.4.1 Detectability ................................................ 17
    2.4.2 Parameter Estimation ...................................... 19
  2.5 Conclusions ..................................................... 25

3 Directional Signals of Magnetic Inelastic Dark Matter ....... 29
  3.1 Introduction ..................................................... 29
  3.2 Directional Detection ........................................... 33
    3.2.1 XENON100 ................................................... 35
    3.2.2 Detector Efficiency ........................................ 36
  3.3 Recoil Spectrum ................................................ 38
  3.4 Sensitivity ...................................................... 41
    3.4.1 Directional Detection ...................................... 42
    3.4.2 Parameter Estimation ...................................... 43
    3.4.3 Measurement of Both Recoils .............................. 46
## Contents

3.5 Conclusions .................................................. 46

4 Cosmic Ray, Gamma Ray, and Microwave Signals .......................... 48
   4.1 Introduction ............................................... 48
   4.2 Signals .................................................. 54
      4.2.1 Data ................................................ 59
      4.2.2 Solar Modulation .................................... 61
   4.3 Fitting procedure ......................................... 63
      4.3.1 Uncertainties ....................................... 66
      4.3.2 $\chi^2$ minimization and regularization ................. 69
   4.4 Results .................................................. 72
      4.4.1 Annihilating Dark Matter Results ...................... 75
      4.4.2 Decaying Dark Matter Results ........................ 78
      4.4.3 Pulsar Results ...................................... 82
      4.4.4 Combination Results .................................. 83
   4.5 Conclusions ................................................ 83

5 CMB Constraints on Dark Matter Annihilation ............................ 86
   5.1 Introduction ............................................... 86
   5.2 The Effect of Energy Injection ................................ 92
      5.2.1 Brief review of the Fisher matrix ...................... 97
      5.2.2 Experimental parameters ............................... 100
      5.2.3 Numerical stability of derivatives and linearity ...... 101
   5.3 Principal Component Analysis ................................ 103
      5.3.1 The principal components .............................. 104
      5.3.2 Mapping into $\delta C_\ell$ space ..................... 107
   5.4 Detectability ................................................ 110
      5.4.1 Estimating limits from the Fisher matrix ............... 111
      5.4.2 Sensitivity of future experiments ..................... 113
      5.4.3 Biases to the cosmological parameters ................. 117
   5.5 A universal $p_{\text{ann}}(z)$ for WIMP annihilation .......... 119
   5.6 CosmoMC Results ........................................... 119
   5.7 Conclusion ................................................ 120

6 Asymmetric and Symmetric Light Dark Matter .......................... 123
   6.1 Introduction ............................................... 123
   6.2 Relic Density for Symmetric and Asymmetric Dark Matter .......... 127
   6.3 CMB Constraints ............................................ 130
   6.4 Light Mediators ............................................ 136
      6.4.1 Collider and Direct Detection Constraints on Light DM with Heavy Mediators ......................... 137
      6.4.2 Light Dark Matter with Light Mediators ............... 139
6.5 Halo Shape Constraints on the Mediator Mass ........................................... 143
6.6 Direct Detection ......................................................................................... 148
   6.6.1 Nucleon Scattering ........................................................................... 149
   6.6.2 Electron Scattering ........................................................................... 153
6.7 Conclusions .............................................................................................. 158

7 Conclusions .................................................................................................. 160

Bibliography .................................................................................................... 161
Citations to Previously Published Work

Chapter 2 has been published as


Chapter 3 has been published as


Chapter 4 has been published as


Chapter 5 appeared in the following:


Finally, Chapter 6 has been published as


Electronic preprints (shown in typewriter font) are available on the Internet at the following URL:

http://arXiv.org
Acknowledgments

The research completed in this thesis has been partially funded by an NSF Graduate Fellowship and NASA Theory grant NNX10AD85G.

The primary acknowledgment goes to my advisor Doug Finkbeiner. He has been an inspiring and supportive mentor, both in guiding this work and also providing me ample opportunity to explore all kinds of physics. It has also been a pleasure to complete the work in this thesis with my collaborators: Greg Dobler, Silvia Galli, Tracy Slatyer, Neal Weiner, Hai-bo Yu, and Kathryn Zurek. I enjoyed working on a project with Aqil Sajjad and John Mason, although that paper is outside the scope of this thesis.

Many other people have provided interesting conversations, half-complete collaborations, helpful feedback, or support along the way, including: Dionysios Anninos, Yang-Ting Chien, Tim Cohen, Mboyo Esole, Melissa Franklin, Monica Guica, Bram Gaasbeek, David Krohn, Vijay Kumar, Ramalingam Loganayagam, Suvarat Raju, Lisa Randall, Eddie Schlafly, Matt Schwartz, Brian Shuve, David Simmons-Duffin, Meng Su, Chin Lin Wong, and too many others to list here. I would like to thank the CfA and high energy theory group for both accommodating me. Finally, I am grateful to Yaim Cooper, Vijay Kumar, and my parents for their friendship and support.
Chapter 1

Introduction

The nature of dark matter is one of the foremost mysteries of cosmology and particle physics. There is abundant evidence for the existence of dark matter [224, 54], from galaxy rotation curves to the measurement of cosmic abundances from the Cosmic Microwave Background (CMB). From structure formation it is inferred that dark matter must have been cold (non-relativistic) since the time when the photon bath had a temperature $T \sim \text{keV}$. However, we know very little about the particle physics of dark matter, in particular its relation to the Standard Model of particle physics.

Experimental searches for dark matter interactions with Standard Model particles have rapidly improved in sensitivity in recent years, yielding some unexpected results. Direct detection experiments in particular strongly constrain WIMP (weakly interacting massive particle) dark matter, historically the leading dark matter candidate [180]. In addition to the null results, there have been a number of anomalous excesses in direct detection and astrophysical data which have some characteristics expected for a dark matter signal, but are also apparently in contradiction with either the null results or with expectations for a
conventional WIMP.

The DAMA/LIBRA collaboration has detected an annual modulation of the recoil rate in NaI crystals with the phase expected for dark matter scattering events [48, 49, 53]. More recently CoGeNT [2, 1] has also claimed to observe recoil events and an annual modulation consistent with that of DAMA/LIBRA. This annual modulation signal is dramatically inconsistent with upper limits from other experiments for elastically scattering WIMPs, however.

Meanwhile, recent cosmic ray, gamma ray, and microwave signals observed by Fermi[196, 3, 109, 5], PAMELA [7], and WMAP [108] have suggested the presence of an unexpected primary source of $e^+e^-$ at 10-1000 GeV. In particular, PAMELA has observed a rise in the positron excess above 10 GeV (recently corroborated by Fermi [5]), while gamma rays and microwaves observed by Fermi and WMAP are consistent with inverse Compton scattering and synchrotron radiation of the energetic electrons and positrons observed in PAMELA.

While dark matter annihilation in the Milky Way could produce these hard $e^+e^-$, the size of the excess observed is about $\sim 100$ times larger than expected for WIMP dark matter. Furthermore, the absence of a proton excess suggests that dark matter preferentially couples to electrons or muons rather than quarks.

These signals have provoked much recent work and interest in particle dark matter models beyond WIMP dark matter. This thesis is concerned with ongoing efforts to observe a signature of such dark matter models. While the anomalous results discussed above may turn out to be unrelated to dark matter, they have reminded us how limited our knowledge really is, and how the data may yield unexpected results. Many of the models considered have

\footnote{However, recent work on the Fermi Bubbles [263] has shown that a significant fraction Fermi and WMAP signals are likely due to transient phenomena such as AGN activity, rather than from dark matter annihilation.}
novel interactions or are in a different mass range than weak-scale dark matter. In some cases, additional structure in dark sector leads to interesting new observational signals, or enhanced signals that previously were thought to be out of experimental reach. It is imperative that we explore as many avenues as possible in the search to understand dark matter.

1.1 Structure of this thesis

Chapters 2-3 study directional detection of inelastic dark matter (iDM), a model proposed to reconcile the DAMA/LIBRA annual modulation signal with null results from other direct detection experiments (as of 2009). The crucial test of the iDM explanation of DAMA – an experimentum crucis – is an experiment with directional sensitivity, which can measure the daily modulation in direction. Because the contrast can be 100%, it is a sharper test than the much smaller annual modulation in the rate.

In Chapter 2 we estimate the significance of such an experiment as a function of the WIMP mass, cross section, background rate, and other parameters. The proposed experiment severely constrains the DAMA/iDM scenario even with modest exposure ($\sim 1000 \text{ kg} \cdot \text{day}$) on gaseous xenon.

Chapter 3 focuses on the case of magnetic inelastic dark matter (MiDM), in which dark matter inelastically scatters off nuclei through a magnetic dipole interaction. We explore a unique signature of MiDM, which allows for the directional detection with an ordinary direct detection experiment. In MiDM, after the dark matter scatters into its excited state, it decays with a lifetime of order 1 $\mu$s and emits a photon with energy $\sim 100 \text{ keV}$. Both the nuclear recoil and the corresponding emitted photon can be detected by studying delayed coincidence events. The recoil track and velocity of the excited state can be reconstructed
from the nuclear interaction vertex and the photon event vertex. It is therefore possible to observe the directional modulation of WIMP-nucleon scattering without a large-volume gaseous directional detection experiment.

We turn to astrophysical signals of dark matter in the Milky Way in Chapter 4. We fit the Fermi, PAMELA, and WMAP data to “standard backgrounds” plus a new source, assumed to be a separable function of position and energy. For the spatial part, we consider three cases: annihilating dark matter, decaying dark matter, and pulsars. In addition, we consider arbitrary modifications to the energy spectrum of the “ordinary” primary source function, fixing its spatial part, finding this alone to be inadequate to explain the PAMELA or WMAP signals. Dark matter annihilation fits well, where our fit finds a mass of \( \sim 1 \text{ TeV} \) and a boost factor times energy fraction of \( \sim 70 \). While it is possible for dark matter decay and pulsars to fit the data, unconventionally high magnetic fields and radiation densities are required near the Galactic Center to counter the relative shallowness of the assumed spatial profiles.

In Chapter 5 we study the effect of dark matter annihilation and decay during the epoch of recombination through its effect on the CMB. Precision measurements of the temperature and polarization anisotropies of the CMB have been employed to set robust constraints on dark matter annihilation during recombination. We improve and generalize these constraints to apply to energy deposition with arbitrary redshift dependence. Our approach also provides more rigorous and model-independent bounds on dark matter annihilation and decay scenarios.

Finally, Chapter 6 examines cosmological, astrophysical and collider constraints on light dark matter. Models of light dark matter have received much interest recently as explana-
tions of the DAMA/LIBRA and CoGeNT signals. Here light dark matter means thermal
dark matter (DM) with mass $m_X$ in the range $\sim 1$ MeV – 10 GeV. CMB observations, which
severely constrain light symmetric DM, can be evaded if the DM relic density is sufficiently
asymmetric. We determine the minimum annihilation cross section for achieving these asym-
metries subject to the relic density constraint; these cross sections are larger than the usual
thermal annihilation cross section. On account of collider constraints, such annihilation cross
sections can only be obtained by invoking light mediators. These light mediators can give
rise to significant DM self-interactions, and we derive a lower bound on the mediator mass
from elliptical DM halo shape constraints. We map all of these constraints to the parameter
space of DM-electron and DM-nucleon scattering cross sections for direct detection. For
DM-electron scattering, a significant fraction of the parameter space is already ruled out by
beam-dump and supernova cooling constraints.
Chapter 2

Inelastic Dark Matter: An 

Experimentum Crucis

2.1 Introduction

Despite decades of direct detection efforts [137], the nature of dark matter interactions with regular matter remains elusive. The results from the DAMA/NaI and DAMA/LIBRA collaborations suggest that such interactions may be more intricate than originally expected. DAMA has observed an annual modulation in NaI crystals for the past decade [48, 49, 53], with the expected phase for WIMP-nuclei interactions. This claim has long appeared to be in conflict with non-detections in other experiments [137] for conventional spin-independent elastic scattering of WIMPs on nuclei. Though recent limits by XENON10 [22, 23] and CDMS II [11] appear to rule out the DAMA region of parameter space by a factor of 100 in cross section, DAMA/LIBRA [49] has recently confirmed their previous annual modulation result and increased the significance to 8.2σ. This conflict has motivated serious discussion
of models beyond the simplest elastic scattering of weak-scale WIMPs, with the hope of accommodating DAMA as well as the other limits.

At least four approaches have been considered: 1. electron scattering [50]; 2. spin dependent scattering [268, 45, 242]; 3. light dark matter [60, 147]; and 4. inelastic scattering [255]. The first hypothesizes that the signal in DAMA is scattering of WIMPs off of electrons. Significant momentum can be transferred to the electron during the small fraction of the time ($< 0.1\%$) that it finds itself near the nucleus and at moderately relativistic speeds. However, this small fraction must be balanced by an uncomfortably large cross section, which is almost certainly ruled out by early Universe (CMB) constraints.

The spin-dependent scattering argument attempts to circumvent limits from CDMS in Si for example by positing that the cross section is strongly dependent on nuclear spin. However, recent experiments [43] have significantly tightened constraints on this scenario, and the allowed regions require a significant drop in the background in the signal region [240]. While small regions of parameter space are still allowed, we do not consider this here.

Another suggestion is that the DAMA recoil events are not in the energy range first suspected. Assuming recoils off of iodine, the quenching factor of 0.09 implies that the $2 - 6$ keVee observed energy corresponds to a recoil energy of $22 - 66$ keVr. It has recently been suggested that “channeling”, i.e. alignment of the recoil with principal directions in the crystal lattice, creates an effective quenching factor of unity for some fraction of the events [51]. In this case, there is a small amount of parameter space available for lighter WIMPs ($\sim 5$ GeV) still compatible with other limits [61, 227, 240]. In general, light WIMPs have difficulty with constraints from the energy spectrum of the unmodulated DAMA signal [73, 113]. While further exploration of light WIMPs may be warranted, we do not consider
2.1.1 The DAMA/iDM Scenario

The inelastic scattering scenario of Tucker-Smith & Weiner [255, 267, 72] takes a different approach: inelastic dark matter (iDM) has an excited state some $\delta \sim 100$ keV above the ground state. The origin of this excited state is unimportant for the present arguments; see [33] for one realization of this idea. Elastic scatterings off of nuclei are suppressed by at least two orders of magnitude with respect to the inelastic scatterings, leading to a preferred energy threshold with few events at low energies. The high sensitivity of e.g. XENON10 to low-energy scatterings (which dominate in the standard elastic scattering models) means that even a small exposure time (316 kg day) can place record-beating limits on the elastic cross section. Because iDM does not produce such low-energy events, it is plausible that the much larger combined exposure time of DAMA/LIBRA and DAMA/NaI (300,000 kg day) could see the higher energy events invisible in the other experiments.

Models of iDM are simple to construct, for instance a fourth-generation (vector-like) neutrino, coupling through the Z-boson [267], a mixed sneutrino [255], KK states in RS theories [101], in composite models [19], or in theories with light mediators [32], see also [230, 87, 42, 79, 101, 185, 130, 41, 19, 219, 77, 182]. In fact, off-diagonal couplings are very natural in dark matter theories, with only the small splitting $\delta$ remaining to be explained.

In an annual modulation experiment, iDM enjoys an additional enhancement relative to elastic models because only WIMPs on the high velocity tail scatter. The modulation can be much larger than the 2-3% expected for elastic scattering, partially compensating for the fact that the majority of WIMPs are below threshold and do not scatter.
If the direct detection data from DAMA and others are taken at face value as nuclear WIMP scattering events, they argue strongly for further experiments designed to test iDM. The experiment must make predictions beyond the already observed annual modulation so that a positive result would add substantially to the believability of the result. Such a make-or-break experiment is known as a “critical experiment,” or *experimentum crucis* ¹. In the next section we describe such an experiment and discuss the limits obtained.

### 2.1.2 Advantages of Directional Sensitivity

The DAMA result is compelling enough to motivate further experiments involving iodine or other nuclei of similar mass. Direct detection experiments generally fall into 3 categories, based on their background rejection strategy. Some (CDMS II, XENON10, etc.) reject individual electron scattering events and look for the residual signal from WIMP scattering.

Another strategy for dealing with background is to search for the annual modulation of the signal (DAMA) brought about by the Earth’s velocity around the Sun, added to the velocity of the Sun around the Galaxy. The assumption is that the WIMP velocities are nearly isotropic, and the Sun moves through the WIMPs at roughly 200 km/s. The Earth moves around the Sun at $v_{\text{orb}} \approx 30$ km/s in an orbit inclined by $i \approx 60^\circ$ with respect to the Sun’s velocity, introducing a modulation of $v_{\text{orb}} \cos(i) \approx 15$ km/s. This method has the virtue of ignoring all steady state instrumental backgrounds, but is vulnerable to backgrounds that vary with the seasons. Though DAMA has placed stringent limits on variations in temperature, humidity, radon gas, line voltage, and anything else known to vary by season [49], this remains a persistent concern.

¹The term *experimentum crucis* was first used by Isaac Newton in a 1672 letter about his Theory of Light and Colors.
A third strategy is to use directional information \[258\]. Because the scattering events should originate, on average, from a specific direction on the sky \((\ell = 90^\circ, b = 0^\circ)\), a daily modulation in direction due to the rotation of the Earth is a sharp test of the WIMP scattering model. As with the annual modulation, many other backgrounds may be expected to vary on a daily timescale. However, as the Earth orbits around the Sun, the angle between the Sun direction and the WIMP signal varies from \(60^\circ \) (\(\sim 7\) March) to \(120^\circ \) (\(\sim 9\) September). Also, any Sun-related oscillation \((365.25\) yr\(^{-1}\)) is orthogonal to the WIMP signal \((366.25\) yr\(^{-1}\)) over one year. This separation allows a much sharper test than the annual modulation alone, even in the limit of low statistics. Furthermore, directional detectors have excellent background rejection and can distinguish between recoils of nuclei and other particles by correlating the length and energy of recoil tracks.

In the context of iDM, a directional experiment has another advantage. The minimum velocity \(v_{\text{min}}\) for a WIMP to scatter with a nuclear recoil of energy \(E_R\) is:

\[
v_{\text{min}} = \sqrt{\frac{1}{2m_NE_R} \left( \frac{m_NE_R}{\mu} + \delta \right)}
\]

(2.1)

where \(\mu\) is the nucleus-WIMP reduced mass \(m_\chi m_N/(m_\chi + m_N)\) and \(m_\chi\) is the WIMP mass. Because of the energy threshold, most events result from WIMPs in the high velocity tail of the WIMP velocity distribution, and therefore most events happen near threshold. This is advantageous because events at threshold have a sharply peaked angular distribution, making the directional discrimination even more pronounced. The energy-dependent maximum recoil angle is

\[
\cos \gamma_{\text{max}}(E_R) = \frac{v_{\text{esc}} - v_{\text{min}}(E_R, \delta)}{v_E}
\]

(2.2)

Here \(\gamma\) is the angle between the velocity of the Earth and the recoil velocity in the Earth frame, and \(v_{\text{esc}}\) is the Galactic escape velocity from the Solar neighborhood. For the bench-
mark models considered here, $\gamma$ is constrained to be within $\sim 100$ degrees of the Earth’s direction. Furthermore, as with annual modulation, the total number of events should vary through the year in a predictable way. These advantages allow a decisive test of the DAMA/iDM scenario with modest experimental effort.

In this article, we evaluate the sensitivities for the DAMA/iDM scenario as a function of WIMP mass $m_\chi$, $\delta$, and other parameters. We focus on a set of benchmark models, given in Table 3.1, that can simultaneously explain DAMA and satisfy constraints from other experiments [72]. Note that the $m_\chi = 70$ GeV benchmark cannot actually explain the DAMA data because of the predicted asymmetry in the modulation amplitude during summer and winter. However, we include the benchmark as a worst-case scenario, as there is flexibility in the WIMP parameters due to the uncertainty in the halo distribution and astrophysical parameters [208]. These benchmarks give the general features and sensitivities (within an order of magnitude) of a directional experiment to the available parameter space of iDM. We find that in most parts of parameter space, 1000 kg days of exposure is sufficient to confirm or refute DAMA/iDM at high confidence.

### 2.2 Experimental Setup

Before discussing the specifics of the experiment, we can address a few basic questions of exposure and energy range. DAMA/LIBRA reports a cumulative modulation in the $2 - 6$ keVee range of 0.052 counts per day per kg, (cpd/kg). The quoted energy range is related to the nuclear recoil energy by a quenching factor $q = E_{ee}/E_{NR} \approx 0.09$ for iodine. Thus, $2 - 6$ keVee $\approx 22 - 66$ keVr.

In the extreme case where the modulation is 100% (i.e., no scattering at all occurs in
Table 2.1: Benchmark models for $v_{esc} = 500$ km/s, $v_0 = 220$ km/s [72]. In the last row we have listed the benchmark model for $m_\chi = 150$ GeV at $v_{esc} = 600$ km/s.

<table>
<thead>
<tr>
<th>$m_\chi$ (GeV)</th>
<th>$\delta$ (keV)</th>
<th>$\sigma_n$ ($10^{-40}$ cm$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>119</td>
<td>11.85</td>
</tr>
<tr>
<td>150</td>
<td>126</td>
<td>2.92</td>
</tr>
<tr>
<td>700</td>
<td>128</td>
<td>4.5</td>
</tr>
<tr>
<td>150*</td>
<td>130</td>
<td>4</td>
</tr>
</tbody>
</table>

The winter), the signal is essentially directional. One would need approximately 400 kg · day in the summer to yield 20 events of signal, roughly the number of events needed for an unambiguous detection at zero background, as we will discuss in Section 3.4.1. Consistency with other experiments is also possible with $\sim 20\%$ modulation [72], with only 40 kg · day needed for a clear discovery.

However, this estimate assumes that the signal occurs in an energy range which is detectable at a directional experiment, and this, we shall see, is very unlikely to be the case. A directional experiment will likely have a higher energy threshold.

The DAMA/LIBRA signal peaks near $E_R \approx 3$ keVee, after which it falls significantly. Above 5 keVee, the total modulation is $0.0034 \pm 0.0024$ cpd/kg, which is consistent with zero. The signal above 4 keVee yields a signal at DAMA of $0.014 \pm 0.004$ cpd/kg, which requires approximately 1400 kg · day of exposure for 20 events. Moreover, it is possible that the actual signal is at 3.5 keVee and below, and the signal at apparently higher energy is due to the resolution of the DAMA detector [52].

There is a significant uncertainty in the quenching factor as well. While $q = 0.09$ is a
commonly used value, the measurements are uncertain, and values $q = 0.10$ and slightly higher are possible. Since the range of WIMP parameters allowed arises from fitting the DAMA peak, the uncertainty in this factor is hidden from our analyses here. Nonetheless, the presence of a larger quenching factor would result in a lower range of energies for the signal. Thus, it is clear that a robust test of the DAMA result involves pushing the energy threshold as low as possible. While the models that we consider generally do have signal above 50 keVr, this cannot be guaranteed, especially in situations where form factors might suppress the higher energy events [19]. In the event the experiments as we describe are performed and no signal is seen, we would advocate lowering the threshold, even at the cost of exposure from reduced pressure, to whatever extent possible.

With these important caveats in mind, we can proceed to discuss the details of what such an experiment would look like.

### 2.2.1 Experimental Design

Gaseous detectors can resolve the nuclear recoil tracks, which have lengths of several millimeters at sufficiently low pressures. Several gaseous directional detection experiments are already underway, including DMTPC [245], NEWAGE [215], DRIFT [67], and MIMAC [239], which employ time-projection chambers to reconstruct tracks. However, these experiments are typically focused on spin-dependent WIMP-nucleus interactions and use the gas CF$_4$ as a detector, with the exception of DRIFT, which uses CS$_2$. For a review of the various detector technologies, see [9, 246, 247].

We suggest using a gas containing xenon or another heavy element. This increases sensitivity to spin-independent interactions because scattering rates are kinematically highly
suppressed for lighter nuclei in the iDM scenario, in addition to the overall factor of $A^2$ that appears in the cross section. However, heavier elements have shorter recoil tracks which are more difficult to resolve. Furthermore, the gas should allow for good electron (or ion) drift and also have good scintillation properties (at least for DMTPC). Choosing a gas will involve some compromise between these properties. We note that for a splitting of $\delta \sim 120$ keV, $A$ must be greater than 75 to see any signal for the mass range $m_\chi \sim 100 - 1000$ GeV for an earth velocity of 225 km/s and an escape velocity of 500 km/s.

According to preliminary work of the directional detection experiments mentioned above, in order to resolve the angles of the tracks, the gas chamber must be at a pressure of around 50 torr. Furthermore if the recoil energies are too low (below $\sim 50$ keVr), it is difficult to detect the sense (head-tail discrimination) of the track, which reduces sensitivity significantly [218, 100, 154]. The directional resolution of DMTPC is currently estimated to be around 15 degrees at 100 keVr and improves by several degrees at higher energies [110].

The dominant irreducible background is neutron recoils arising from radioactive materials near or in the detector. Simulations suggest background rejection is excellent for gamma-rays, electrons, and $\alpha$’s [257] (see also Fig. 7 of [245]). The DRIFT collaboration has reported on neutron backgrounds; however, they found a radioactive source ($^{222}$Rn) inside the detector [66]. The NEWAGE experiment at Kamioka estimated their primary background to come from the fast neutron flux which, when shielded by 50 cm of water, would contribute only a few events per year [264].
2.3 Recoil Spectrum

We derive the differential nuclear recoil spectrum in recoil energy $E_R$ and $\cos \gamma$, which is defined as $\cos \gamma = \hat{v}_E \cdot \hat{v}_R$. The Earth’s motion in the halo rest frame is $\vec{v}_E$ and the vector $\vec{v}_R$ is the nuclear recoil velocity in the Earth’s frame. Let $\vec{v}$ be the incoming WIMP velocity in the Earth’s frame.
The single nucleon scattering cross section is:
\[ d\sigma = \frac{\sigma_n m_n}{2\mu_n^2} \frac{1}{v^2} dE_R \, d\cos \gamma \, \delta^{(1)} \left( \hat{v} \cdot \hat{v}_R - \frac{v_{\text{min}}}{v} \right) \] (2.3)

where \( \mu_n \) is the WIMP-nucleon reduced mass and \( \sigma_n \) is a reference cross section that is assumed to be the same for all nucleons. \( m_n \) is nucleon mass. The minimum velocity \( v_{\text{min}} \) for a WIMP to scatter with a nuclear recoil of energy \( E_R \) was given in Eq. 2.1.

The differential recoil rate for WIMP-nucleus scattering is
\[ \frac{dR}{dE_R d\cos \gamma} = N_T \frac{\rho_\chi}{m_\chi} \int d^3 v \, v \, f(\vec{v} + \vec{v}_E) \frac{d\sigma}{dE_R d\cos \gamma} \] (2.4)

where \( f(\vec{v}) \), the WIMP distribution in the galaxy frame, is boosted to the Earth frame. \( N_T \) is the number of target nuclei per kg and \( \rho_\chi \) is the local WIMP energy density. We are now using the differential scattering cross section \( d\sigma \) for the whole nucleus. Define the constant \( \kappa \):
\[ \kappa = N_T \frac{\rho_\chi}{m_\chi} \frac{\sigma_n m_N}{2\mu_n^2} \frac{(f_p Z + (A - Z)f_n)^2}{f_n^2}. \] (2.5)

Changing variables to \( \vec{v}' = \vec{v} + \vec{v}_E \) gives:
\[ \frac{dR}{dE_R d\cos \gamma} = \kappa F^2(E_R) \int d^3 v \, f(\vec{v}) \, \delta^{(1)} \left( \vec{v} \cdot \hat{v}_R - \vec{v}_E \cdot \hat{v}_R - v_{\text{min}}(E_R, \delta) \right) \] (2.6)

and \( F^2(E_R) \) is the Helm form factor given in [198]. This formula is discussed in detail (in the context of Radon transforms) in [145]. Thus we can see that at fixed \( E_R \), the signal peaks where the delta function is nonzero over the largest portion of the phase space, or \( \cos \gamma = \hat{v}_E \cdot \hat{v}_R = -1 \). The peak in \( E_R \) and fixed \( \gamma \) is determined by the competition between the form factor (which pushes the signal to lower energies) and the inelasticity (whereby the minimum velocity produces a minimum value of \( E_R \)).
Chapter 2: Inelastic Dark Matter: An Experimentum Crucis

Following [72], we use the truncated Maxwell-Boltzmann distribution in the rest of this work:

$$f(\vec{v}) = \frac{1}{n(v_0, v_{esc})} \exp\left(-\frac{\vec{v}^2}{v_0^2}\right) \Theta(v_{esc} - |\vec{v}|)$$

where \(n(v_0, v_{esc})\) normalizes \(\int d^3v f\) to 1. The resulting spectrum is:

$$\frac{dR}{dE_R d\cos \gamma} = \frac{\kappa F^2(E_R)}{n(v_0, v_{esc}) \pi v_0^2} \left(\exp \left(-\frac{(\vec{v}_E \cdot \hat{v}_R + v_{min}(E_R, \delta))^2}{v_0^2}\right) - \exp \left(-\frac{v_{esc}^2}{v_0^2}\right)\right) \Theta(v_{esc} - |\vec{v}_E \cdot \hat{v}_R + v_{min}(E_R, \delta)|)$$

The values we use for the astrophysical parameters are: \(v_0 = 220 \text{ km/s}, v_E = 225 \text{ km/s}, v_{esc} = 500 - 600 \text{ km/s}\) [256], and \(\rho_\chi = 0.3 \text{ GeV/cm}^3\). The normalized rate spectrum of several benchmark models is shown in Fig. 3.2.

### 2.4 Sensitivity

A robust detection of a directional modulation is possible with surprisingly few events, and does not require use of the rate formulas in the previous section. In fact, a full likelihood analysis based on the correct model is only a factor of \(\sim 2\) better than a simple technique, and for a convincing detection, simpler is better. In this section we assume the detection gas has \(A = 127\) (for iodine; Xe with \(A = 131\) would be similar) and focus on the energy range \(E_R \in [50, 80] \text{ keVr}\).

#### 2.4.1 Detectability

For a model-independent statistic we follow [218, 153] and use the dipole of the recoil direction, \(\langle \cos \gamma \rangle\). This is motivated by the fact that the rate should depend only on \(\cos \gamma\)
Figure 2.2: Exposure to obtain a $5\sigma$ measurement of $\langle \cos \gamma \rangle$ 90% of the time the experiment is conducted on Earth. The energy range of the experiment is $E_R \in [50, 80]$ keVr. $dR_{BG}$ is the background rate; the DAMA unmodulated background rate is indicated by the solid vertical line at 0.085. The bands shown give the exposures necessary as the rates modulate throughout a year. Since the annual modulation is asymmetric in summer and winter for low mass dark matter, the average exposure for $m_\chi = 70$ GeV is indicated by the dashed line. In (a) we show three mass benchmarks from Table 3.1 and in (b) we show the effect of decreasing the angular resolution of the detector to 60 degrees and of lowering the energy threshold to 40 keVr. (Darker regions indicate where the bands overlap.)

and $E_R$, so the directional part can be expanded in spherical harmonics.

Our detection criterion is a measurement of $\langle \cos \gamma \rangle$ that is $5\sigma$ relative to the distribution of $\langle \cos \gamma \rangle$ for the same number of randomly distributed events. For a fixed exposure, we generate many random sets of model data (constrained by the DAMA benchmarks in Table 3.1), and then demand that 90% of the time the result is $5\sigma$ from the null hypothesis. The background is modeled as uniform in recoil energy and angle. We assume the detector has an angular resolution of 15 degrees.

In Fig. 3.3(a) we show the exposures necessary for such conditions, as a function of the background rate, for a few benchmark models. At zero background, roughly 18 events are needed for all benchmark models, on average. Fig. 3.3(b) shows the effect of decreasing the angular resolution to 60 degrees and lowering the energy threshold of the experiment to $E_R = 40$ keVr. Because of the sharp angular profile of the recoil spectrum, a poor angular
resolution does not significantly reduce the possibility of a detection. However, achieving
an energy threshold of 30-40 keVr dramatically lowers the necessary exposures because the
peak of the recoil spectrum occurs at 30-40 keVr and falls off exponentially.

2.4.2 Parameter Estimation

We also perform a likelihood analysis as a measure of sensitivity of the experiment to
the parameters of the model, assuming perfect energy and angular resolution. From our
analysis in the previous section, we expect this assumption does not affect the results sig-
nificantly. (See also [100], which shows the sensitivity dependence on angular resolution.)
The parameters we consider here are \( m_\chi, \delta, \) and \( \sigma_n, \) which we denote together simply by \( p. \)
Define

\[
\mu(x; p) \equiv \frac{dR}{dE_R d\cos\gamma}(x; p) + \frac{dR_{BG}}{2},
\]

which is the rate (cpd/kg/keVr per \( \cos\gamma \)) at a given recoil energy and angle (denoted together
by \( x \)) for parameters \( p. \) We assume the background rate, \( dR_{BG}, \) in units of cpd/keVr/kg, is
known.

The likelihood is the probability of parameters \( p \) given the events \( \{x_i\}. \) Given events
\( \{x_i\}, \) bin the events such that in each bin there is only 0 or 1 event and label the bins with
one count by \( \{X_\alpha\} \) and the empty bins by \( \{X_\beta\}. \) The expected number of counts in a bin is

\[
E(X; p) = \mathcal{E} \mu(x; p) \Delta x
\]

where \( \mathcal{E} \) is the exposure. Then the (log) likelihood is

\[
\ln \mathcal{L}_{tot}(p) = \sum_\alpha \ln \left( e^{-E(X_\alpha; p)} E(X_\alpha; p) \right) + \sum_\beta \ln e^{-E(X_\beta; p)}
\]

(2.11)
which is the log of the Poisson probability of obtaining 0 or 1 event in each bin. To find the expected average $\ln L_{tot}$ for a given exposure $E$ and true parameters $p_0$, we compute

$$
\ln L_{tot}(p) = E \int dx \left( \mu(x; p_0) \ln \mu(x; p) - \mu(x; p) \right)
$$

(2.12)

which is the continuum, noiseless limit of Eq. 2.11. Since we can only compare differences in log likelihood, in this equation we have subtracted an arbitrary constant in $p$ which takes care of the units in $\ln \mu(x, p)$.

In Figs. 2.3-2.9 we show confidence levels of (68, 90, 95, 99, and 99.9%) on the WIMP parameters for an exposure of 1000 kg · day. To obtain the probability, or likelihood, at a point in the $m_\chi - \delta$ plane, we either: 1) find the likelihood as a function of $\sigma_n$ and maximize with respect to $\sigma_n$ or 2) assume $\sigma_n$ is exactly known from some other experiment. We can do the same also for points in $m_\chi - \sigma_n$ plane and $\sigma_n - \delta$ plane. The full log likelihood function lives in the full 3 dimensional parameter space. Here we show possible slices through that space.

For each possible slice, we have shown several variations on the real WIMP parameters or experimental parameters. In the default scenario, we consider the $m_\chi = 150$ GeV benchmark with $E_{th} = 50$ keVr, a background rate of $dR_{BG} = 10^{-3}$ cpd/kg/keVr, and $v_{esc} = 500$ km/s. We consider the following independent variations:

- Lower energy threshold ($E_{th} \rightarrow 40$ keVr)
- Higher background ($dR_{BG} \rightarrow 10^{-2}$ cpd/kg/keVr)
- Higher escape velocity ($v_{esc} = 600$ km/s)
- Lower WIMP mass ($m_\chi \rightarrow 70$ GeV benchmark)
Higher WIMP mass ($m_\chi \to 700$ GeV benchmark)

In each case, as $m_\chi$ and $v_{esc}$ vary, $\sigma_n$ and $\delta$ are adjusted to agree with benchmark fits to DAMA, using the parameters in Table 3.1. At masses above 250 GeV, there is increasing tension between the DAMA result and other experiments, notably CDMS. This tension is highly dependent on the high velocity tail of the WIMP velocity distribution, and can be alleviated by considering non-Maxwellian velocity distributions, for instance from the Via Lactea simulation [208, 192]. Thus, we consider these points, but it should be emphasized that the non-Maxwellian halos generally tend to lead to a larger signal at DAMA (relative to the other experiments), and thus on a xenon target (because of the similar kinematics), and thus we expect that our use of a Maxwellian distribution is conservative for these points.
Chapter 2: Inelastic Dark Matter: An Experimentum Crucis

Figure 2.4: Confidence levels for determining $m_\chi$ and $\sigma_n$, where $\delta$ is known with an exposure of 1000 kg · day. $\sigma_0 = 10^{-40}$ cm$^2$.

At masses much larger than the nucleus mass, the threshold velocity $v_{min}$ is independent of mass and the spectrum depends on $m_\chi$ only through the local WIMP density $\rho_\chi/m_\chi$. In these regions $m_\chi$ and $\sigma_n$ are completely degenerate since only the combination $\rho_\chi \sigma_n/m_\chi$ ever appears, as a prefactor determining the overall rate. This can be clearly seen in Fig. 2.3, which shows confidence intervals in the $m_\chi - \sigma_n$ plane. Note that because the contours never close, we have imposed the (rather conservative) constraint that $m_\chi < 100$ TeV based on the unitarity bound [155] for a thermal relic.

The effects of the $m_\chi - \sigma_n$ degeneracy can also be seen in the $m_\chi - \delta$ plane, shown in Fig. 2.5. Here high masses are all equally likely (given a fixed $\delta$) because $\sigma_n$ can be adjusted accordingly.
Chapter 2: Inelastic Dark Matter: An Experimentum Crucis

Figure 2.5: Confidence levels for determining $m_\chi$ and $\delta$, where $\sigma_n$ is unknown, with an exposure of 1000 kg \cdot day.

In the $\delta - \sigma_n$ plane, Fig. 2.7, there is a sharp discontinuity since low masses are favored at smaller $\sigma_n$ and very high masses are favored at high $\sigma_n$. This is because at low scattering cross section, in order to boost the rates such that it matches the observed number of events, one can lower $\delta$ or adjust the mass to optimize the number of rates. (The scattering rate is maximized when the mass of the WIMP $\sim$ the mass of the nuclei.) However, at high scattering cross section, one can increase $\delta$ but only increase the mass to very high masses to reduce the rates. Though lowering the mass drastically also decreases the rate, the angular shape at very low masses is very distinct (see Fig. 3.2) and thus unfavored. The cutoff in Fig. 2.7 at high $\sigma_n$ is a result of the unitarity bound on the mass.

These effects can make it difficult to constrain the WIMP mass at low exposures; however,
Figure 2.6: Confidence levels for determining $m_\chi$ and $\delta$, where $\sigma_n$ is known, with an exposure of 1000 kg · day.

It is easier to constrain the ratio $m_\chi/\sigma_n$, which we have shown in Fig. 2.9.

Finally, we note that in these figures we have assumed the earth velocity is unmodulated. For the benchmark where $m_\chi = 70$ GeV, our worst-case scenario, the effects of the annual modulation in velocity can improve the confidence levels significantly if the experiment is done during the summer.

The disadvantage of the likelihood analysis is its model dependence. We used the truncated Maxwell-Boltzmann profile, whereas in reality it is likely there is more structure in the dark matter profile. However we expect the results to roughly be the same for many more complicated velocity distributions, and in fact can improve for inelastic dark matter, as mentioned above. Furthermore, because of the velocity threshold due to $\delta$, the inelastic
Figure 2.7: Confidence levels for determining $\delta$ and $\sigma$, where $m_\chi$ is unknown, with an exposure of $1000\;\text{kg} \cdot \text{day}$. $\sigma_0 = 10^{-40}\;\text{cm}^2$.

scenario is not very sensitive to streams because most streams are below the threshold velocity. Anisotropies in the halo profile do not significantly affect the results here. To see the effect of using less simplistic halo models on the elastic scattering spectrum and sensitivity, see [16] and [99].

### 2.5 Conclusions

Motivated by the finding [72] that inelastic dark matter (iDM) is compatible with both the DAMA annual modulation signal at $22 - 66\;\text{keVr}$ and limits from other experiments at lower energies, we have investigated prospects for directional detection in the context of the
Figure 2.8: Confidence levels for determining $\delta$ and $\sigma_n$, where $m_\chi$ is known, with an exposure of 1000 kg · day. $\sigma_0 = 10^{-40} cm^2$.

iDM model. We are encouraged by the fact that ZEPLIN-III has also detected a number of events in the 40 – 80 keVr range [197]. This has not been claimed as evidence of WIMP scattering, but makes it impossible to rule out iDM with such data. In the near future, LUX$^2$ and XENON100 [31] will have greatly improved sensitivity and lower backgrounds, and will provide a sharp test of the iDM/DAMA scenario. If these experiments also detect an excess of events above background in the appropriate energy range, a major effort in directional detection will be justified.

Directional detection with a gaseous detector containing a heavy gas (e.g. Xe) may not require the huge exposure times implied by the elastic scattering limits. For a threshold

$^2$http://lux.brown.edu
Figure 2.9: Confidence levels for determining $\delta$ and $m_\chi/\sigma_n$, where $m_\chi$ is unknown, with an exposure of 1000 kg · day, taking $\sigma_0 = 10^{-40}$ cm$^2$. Over most of the parameter space, some value of $m_\chi$ (and therefore $\sigma_n$) can be found to produce enough events for the given $\delta$. However, in the case of large $\delta$ and large $m_\chi/\sigma_n$, no solution is possible in some cases.

energy of $E_{th} = 50$ keVr, we find that exposures of order $\sim 1000$ kg · day in a directional experiment can convincingly refute or support the claims of DAMA in the context of the inelastic dark matter model. At zero background, roughly 18 events are needed for a clear detection of WIMP scattering. Even with larger backgrounds, the required exposure is a few hundred kg · day, over most of the iDM parameter space that can explain both DAMA and other direct detection experiments. With roughly 1000 kg · day, it is possible to obtain a measurement of $\delta > 0$ at high significance and also the parameter $m_\chi/\sigma_n$ to within an order of magnitude.

Furthermore, if it is possible to roughly determine one of the WIMP parameters, for
example $\delta \sim 120$ keV, via another experiment, the mass and nucleon scattering cross section are highly constrained with an exposure of a few hundred kg \cdot day because of the distinctive shape of the energy-angle recoil spectrum.

Significantly lower exposures are needed if the threshold energy is decreased. As discussed in Section 2.2, because of the uncertainties in the nuclear recoil energies of the DAMA signal, it is crucial to reduce the threshold energy as much as possible. For low masses, the recoil spectrum is sharply distributed in energy and angle. However, typical recoil energies are smaller. Thus with an energy threshold of $E_{th} = 50$ keVr most of the events for $m_\chi = 70$ GeV are not seen. With an energy threshold of 100 keVr and $m_\chi = 70$ GeV, none of the WIMP recoils can be seen. Though the required volume increases and angular resolution decreases when $E_{th}$ is lowered, we found that a poor angular resolution ($\sim 60^\circ$) does not significantly affect the results, assuming that 3D reconstruction of the track and determining the sense is still possible.
Chapter 3

Directional Signals of Magnetic Inelastic Dark Matter

3.1 Introduction

The basic inelastic dark matter (iDM), described in Section 2.1.1, is now tightly con- strained [244] by the latest results from CRESST [271], ZEPLIN-III [14], XENON [24], and CDMS. By introducing more ingredients in this model, one can increase the dark matter scattering rate off the NaI used in DAMA, relative the nuclei used in other direct detection experiments. In particular, we focus on the fact that iodine is special in having both a relatively large mass and a relatively large magnetic moment [74]. Inelastic scattering takes advantage of the large iodine mass.

If dark matter has (weak) electromagnetic moments [35, 232], it can interact through the charge and magnetic dipole moment of the nuclei. For a summary of the interaction strengths for various nuclei used in direct detection experiments, see [37]. This type of interaction has
Table 3.1: In the first three (starred) rows, we give the best fit benchmark models of MiDM, with $v_{\text{esc}} = 550 \text{ km/s}$ and $v_0 = 220 \text{ km/s}$ [74]. We also list parameters within the 90\% CL region of each best fit value, for which the lifetime, $\tau$, can be a factor of a few larger or smaller. $\lambda$ is the average recoil track length. $\eta_{15}$ is an estimate of the efficiency of XENON100 to detect delayed coincidence events, as described in Section 3.2.2. The 'angular' rate is the rate for delayed coincidence events with a nuclear recoil in the energy range $10^{-80} \text{ keVr}$, followed by a photon with $\delta$ keVee. This is obtained from multiplying the total rate by $\eta_{15}$. We also show the expected number of nuclear recoil events for the published XENON100 non-blind analysis.

been used to explain some recent direct detection results [212, 21, 82, 39, 131, 37], including the positive claim of DAMA. However, there are strong constraints from CDMS [12] and XENON [24, 28] on this explanation of DAMA.

We focus on magnetic inelastic dark matter (MiDM), because it has a unique and interesting directional signature. Chang et al. [74] showed MiDM could explain both DAMA and other null results. The model takes advantage of both the magnetic moment and large mass
of iodine. In MiDM, the dark matter couples off-diagonally to the photon:

$$\mathcal{L} \supset \left(\frac{\mu_\chi}{2}\right) \chi^* \sigma_{\mu\nu} F^{\mu\nu} \chi + \text{c.c.}$$

where the mass of $\chi$ and $\chi^*$ are split by $\delta \sim 100$ keV. The off-diagonal coupling is natural if the dark matter is a Majorana fermion. The excited state has a lifetime $\tau = \pi/(\delta^3 \mu_\chi^2) \sim 1 - 10\mu$s, and emits a photon when it decays. This short lifetime makes it possible to observe both the nuclear recoil and the emitted photon with a meter-scale detector. The two interaction vertices allow reconstruction of the excited state track. Both the velocity and angle can be measured, enabling directional detection even without a directional detector.

A dark matter particle with a permanent electromagnetic dipole moment generally can be constrained by, e.g., gamma-ray measurements, the CMB, or precision Standard Model tests [148, 251, 141]. However, the strongest bounds tend to come from direct detection experiments themselves, at least in the 100 GeV mass range. Furthermore, in MiDM, the inelastic nature of the interaction suppresses interactions with photons and baryons at low energies. If the dark matter is a composite particle, a low compositeness scale can also suppress annihilation to photons.

There are some variants of the MiDM idea. In [118], the parameter values were taken to be $m_\chi \sim 1$ GeV and $\delta \sim 3$keV. The DAMA signal is produced by the emitted photon. This explanation evades constraints from other direct detection experiments because such low-energy electromagnetic events are typically rejected or not seen by other detectors.

It is also possible that the dark matter couples to a new ‘dark’ $U(1)$, with gauge boson mass $m_A \neq 0$ [161, 231, 21]. Here the dark matter has a large dark dipole. If the dark gauge boson couples to regular electromagnetic currents, a sizable interaction with nuclei can be generated. However, the decay rate of the excited state is suppressed because there
is no direct interaction with the photon. While these ideas are interesting explanations of the DAMA signals, we do not consider them further because the excited state has a long lifetime.

We study MiDM benchmarks, given in Table 3.1, which are good fits to the DAMA annual modulation signal [74]. MiDM models with $m_\chi$ greater than $\sim 300$ GeV are severely constrained by ZEPLIN-III [14], KIMS [190], and XENON100 [28].

The benchmarks are subject to form factor and velocity distribution uncertainties [208, 192, 201, 270, 18], especially for larger masses. The directional signal prediction can change wildly depending on the lifetime and rate.

In order to explore the parameter space, we also considered two extreme points within the DAMA 90% confidence level region found by [74], for each of the three masses. For the point with highest $\delta$ and $\mu_\chi$, the expectation for directional detection is better. The point with lowest $\delta$ and $\mu_\chi$, which would not result in many delayed coincidence events, is in any case already tightly constrained by the XENON100 non-blind analysis.

In this work, we show that the current generation of direct detection experiments can observe a directional signal from MiDM. For concreteness we focus on a XENON100-like detector, for two reasons. First, XENON100 will soon place strong constraints on the MiDM parameter space, making it the most relevant experiment to consider. Second, we wish to emphasize the feasibility of detecting a directional signal with experiments that are currently running.

We compute the distribution of recoil track angles and velocities from MiDM benchmarks. The sensitivity of XENON100 to the MiDM parameter space depends strongly on the lifetime of the excited state. For the benchmark lifetimes of $\sim 1-10\mu s$, XENON100 can measure the
directional modulation at high significance and obtain sharp constraints on the parameter space with just tens of events. This is achievable with around 5000 kg \cdot day in the energy range 10 – 80 keVr.

### 3.2 Directional Detection

Directional detection can clearly test whether any signal comes from WIMP interactions [258]. Due to the Earth’s motion in the Galaxy, there is a “WIMP wind” which is opposite the motion of the Earth. There is a daily modulation in the angle of recoil tracks in the lab frame. This modulation depends only on the rotation of the Earth relative to the WIMP wind, and can be disentangled from the daily rotation of the Earth with respect to the Sun. The experimental directional detection effort focuses on measuring the nuclear recoil track with large-volume, gaseous detectors [9, 246, 247].

Angular information is a particularly powerful discriminant of WIMP scattering for iDM [128, 202]. Because inelastic interactions have a high velocity threshold, the angular distribution of the nuclear recoil tracks is sharply peaked in the direction of the WIMP wind. There is a kinematic constraint on the recoil angle of the nucleus:

$$\left( \cos \gamma \right)_{\text{max}}(E_R) = \frac{v_{\text{esc}} - v_{\text{min}}(E_R; \delta)}{v_E}. \quad (3.2)$$

Here $\gamma$ is the angle between the velocity of the Earth and the recoil velocity in the Earth frame, $v_E$ is the Earth’s velocity in the Galactic frame, and $v_{\text{esc}}$ is the Galactic escape velocity from the Solar neighborhood. For typical iDM models considered in the literature, $\gamma$ is constrained to be within $\sim 100$ degrees of the WIMP wind [128]. However, because the signal goes to zero at the bound in Eq. 3.2, the precise location of this kinematic constraint
can be difficult to pinpoint.

MiDM has better directional detection prospects at XENON100, compared to directional detection of iDM. Current directional detectors focus on spin-dependent scattering and use light targets such as CF$_4$ [10, 103, 156, 215]. Thus, they would not see inelastic scattering events. In the MiDM case, there is also much more event information and thus more sensitivity to the parameter space. One can measure both the velocity ($v_*$) and the angle ($\cos \gamma_*$) of the WIMP recoil track. Once again, this recoil angle is with respect to the Earth’s motion. The tracks are sharply peaked in angle opposite the motion of the Earth.

For the WIMP recoil angle, there is also an energy-dependent maximum recoil angle, which we give in Sec. 3.3. The most important bound is on the WIMP recoil velocity,

$$v_*^{\min}(E_R) = \left| \frac{(E_R(m_N/m_\chi - 1) - \delta)}{\sqrt{2m_NE_R}} \right|. \quad (3.3)$$

Here the signal peaks near the kinematic bound because most events occur near the threshold velocity in Eq. 2.1. Thus having information on both $v_*$ and $E_R$ is an extremely sensitive probe of the model parameters. There is a remaining degeneracy: if $\delta$ and $m_\chi$ are shifted in opposite directions, the bound can remain roughly the same. However, one can fit $\delta$ separately from the spectrum of the nuclear recoils, and from the energy of the emitted photons.

There is also a maximum velocity for the excited state,

$$v_*^{\max}(E_R) = \sqrt{(v_E + v_{esc})^2 - 2(E_R + \delta)/m_\chi}, \quad (3.4)$$

but the rate is exponentially suppressed at this bound.
3.2.1 XENON100

We model directional detection in XENON100 with a simplified XENON100-like experiment. XENON10 [22, 24] had 316.4 kg·day of data in the energy range 4.5-75 keVr. XENON100 has a 40 kg fiducial mass, at even lower backgrounds. The initial 170 kg·day non-blind run already constrains the MiDM parameter space (at low $\delta$).

The XENON100 detector is a cylinder, with a radius of 15.3 cm and a height of 30.6 cm. The fiducial volume has a radius of 13.5 cm and height of 24.3 cm. The primary scintillation (S1) and ionization (S2) signals of an event are measured. For more details, see [30]. The S2 signal is observed 15-140 $\mu$s after the S1 signal, for events in the fiducial volume.

The signature of MiDM is two S1 signals separated by roughly .5 $\mu$s in time, followed at least 15 $\mu$s later by two S2 signals. The photon event is identified from the second S1 signal and an S2 peak with energy of $\sim$100 keVee. At 100 keVee, a photon is clearly distinguishable from a nuclear recoil by S2/S1. The other event should be consistent with a nuclear recoil. The time separation of the two S2 signals depends on how the WIMP recoils along the cylinder axis, $z$. In XENON10, events with multiple S2 events at different $z$ positions were rejected.

We refer to the track connecting the two events as the decay track. Events can be localized to a 3D spatial resolution of 3 mm (though the absorption length for the 100 keV photon may blur this) and timing resolution of 10 ns. Meanwhile, the track should be at least 10 cm long. This makes it possible to measure direction and velocity of the decay track to an extremely high accuracy. The head-tail discrimination of the track can be determined using timing information and the S1/S2 ratio.

We wish to obtain the $\chi^*$ recoil track from the decay track. However, because the photon
can travel up to \( \sim 1 \) cm after emission, this introduces systematic uncertainties. The observed
decay track can be blurred by a few degrees, relative to the \( \chi^* \) recoil track direction. This
also introduces an uncertainty in the velocity of the \( \chi^* \) of roughly 10%.

There are some specific event geometries that can result in more ambiguous events. For
example, it could be difficult to resolve the two S2 signals if the decay track is perpendicular
to the \( z \) axis. Then the two S2 signals arrive at nearly the same time. S2 signals generally
have a time width of \( \sim 1 \) \( \mu \)s and the PMT spatial resolution is only \( \sim 2.5 \) cm. However,
because the drift velocity is 2mm/\( \mu \)s, this is a small fraction of the total solid angle.

Thus directional events are in principle detectable at XENON100. The background for
such delayed coincidence events with both a nuclear recoil and a photon of energy \( \sim 100 \)
keV should be extremely low. There are other ‘mixed’ delayed coincidence events from Bi
and Kr contamination, and excitation of metastable states of Xe [30]. However, these have
very different energies and decay times. It may be possible to extend the fiducial volume
when searching for directional events.

### 3.2.2 Detector Efficiency

The typical decay length is \( 1 - 10 \) m in these models, relatively large compared to
XENON100. Thus the WIMP can recoil inside the detector volume, but decay outside
the detector \(^1\). The effective exposure for delayed coincidence events is, in general, lower
than the exposure for nuclear recoils because of this geometric effect. Here we compute the
detector efficiency, as a function of typical detector size, for the MiDM benchmarks.

\(^1\)The reverse can also happen, similar to the idea in [118]. The rate depends on whether the material
within \( \sim 10 \) m of the Xe detector mostly consists of light or heavy nuclei. Aside from a 20cm layer of lead,
the shielding for XENON100 consists of polyethylene, water, and copper.
Figure 3.1: The efficiency $\eta$ for the best-fit benchmarks from Table 3.1. $R_0$ is the size of a spherical detector. We approximate the XENON100 fiducial volume as a sphere with radius $R_0 = 0.15$ m, marked by the vertical black line. The thinner lines show the corresponding results with highest $\delta$, within the 90% CL region of the best fit.

The efficiency is

$$
\eta(t) = \int d^3 \vec{v}_s f(\vec{v}_s, t) \int \frac{dt'}{\tau} e^{-t'/\tau} \left( \int_V \frac{d^3 \vec{x}}{V} H(t, \vec{x}, \vec{d}, t') \right)
$$

The term in brackets comprises detector effects. The spatial integral is over the detector volume. $H(t, \vec{x}, \vec{d}, t')$ is the efficiency for observing a WIMP decay, given that a nuclear recoil was observed. This depends on the time of the year $t$, the location of the WIMP-nucleus interaction inside the detector, $\vec{x}$, the decay vector, $\vec{d}$, and the WIMP decay time (coincidence time), $t'$. Whether a given WIMP decay track is located inside the detector depends on the orientation of the detector with respect to the Earth’s velocity, the decay vector, and the efficiency for the particular event geometry.

The astrophysics and particle physics is captured by the integral over $t'$ and $\vec{v}_s$. $\tau$ is the lifetime of the excited state. The distribution of recoils depends on the WIMP recoil velocity distribution, $f(\vec{v}_s)$, and the decay time distribution. We assume that $\vec{v}_s$ is defined with respect to the Earth’s velocity vector so that $f(\vec{v}_s)$ does not depend on detector orientation.

For the calculation below, we model the detector as a single sphere of size $R_0$. We
assume that $H$ depends only on the interaction position $\vec{x}$ and the decay length $L = v_* t'$. Here we neglect the smearing arising from the mean free path of the emitted photon, since the emission is isotropic. There is also no dependence on $t$ or recoil angle in this approximation. Then the expression for efficiency above can be simplified to

$$\eta = \int dL \; g(L) \int_0^{R_0} \frac{3 R^2 dR}{R_0^3} H(R, L)$$

where $L$ is the recoil length. The recoil length distribution $g(L)$ is

$$g(L) = \int dv_* \frac{f(v_*)}{v_* \tau} \exp \left( -\frac{L}{v_* \tau} \right)$$

where now $f(v_*)$ is the distribution for $v_*$, not $\vec{v}_*$. A good approximation is $g(L) = \exp(-L/\lambda)/\lambda$, where $\lambda = \langle v_* \rangle \tau$ is the average recoil length. Typical $\lambda$ values are given in Table 3.1.

We approximate the XENON100 detector as a sphere. The fiducial volume has radius $R_0 = 15$ cm, with efficiency $\eta_{15}$. Results are shown in Fig. 3.1. The precise efficiency depends on specifics of the detector, and must take into account the effects mentioned in Sec. 3.2.1.

### 3.3 Recoil Spectrum

There are two electromagnetic scattering channels for magnetic dark matter: dipole-dipole and dipole-charge. In the dipole-dipole scattering case, the dark matter interacts with the magnetic moment of the nucleus. The matrix element is

$$\frac{|\mathcal{M}|^2}{32\pi m_N m_X^2} = 16\pi \alpha^2 m_N \left( \frac{\mu_{nuc}}{e} \right)^2 \left( \frac{\mu_X}{e} \right)^2 \frac{S_N + 1}{3 S_N}$$

(3.7)
for each isotope. We sum over all isotopes, weighted by their abundances \[37\]. There is, in general, also a factor of \((S_\chi + 1)/(3S_\chi)\) for the spin of the dark matter. We take \(S_\chi = 1/2\).

In the dipole-charge scattering case, the dark matter interacts with the electric charge of the nucleus. The matrix element is

\[
\frac{|M|^2}{32\pi m_N m_\chi^2} = \frac{4\pi Z^2 \alpha^2}{E_R} \left(\frac{\mu_\chi}{e}\right)^2 \left[ v^2 - E_R \left(\frac{1}{2m_N} + \frac{1}{m_\chi}\right) \right. \\
- \delta \left(\frac{1}{\mu_{N\chi}} + \frac{\delta}{2m_NE_R}\right)],
\]

(3.8)

where \(v\) is the initial velocity of the WIMP in the lab frame. We have again assumed \(S_\chi = 1/2\).

The differential scattering rate for measuring both nuclear recoil energy and WIMP recoil track is

\[
\frac{dR}{dE_R dv_* dx_*} = \eta N_T \rho_\chi \int d^3\vec{v} f(\vec{v} + \vec{v}_E) \frac{d\sigma}{dE_R dv_* dx_*} \]

(3.9)

where we have abbreviated \(x_* = \cos\gamma_*\). The three-dimensional WIMP velocity distribution is given by \(f(\vec{v})\). \(N_T\) is the number of target nuclei per kg and \(\rho_\chi\) is the local WIMP energy density, which we fix to be 0.4 GeV/cm\(^3\) \[70\].

As in \[128\], we expand \(d\sigma\) and change variables to \(\vec{v}' = \vec{v} + \vec{v}_E\). The trivial integral over \(\vec{v}'\) imposes the condition

\[
\vec{v}' = \vec{q}/m_\chi + \vec{v}_* + \vec{v}_E.
\]

(3.10)

\(\vec{q}\) is the recoil momentum of the nucleus. The resulting differential rate is

\[
dR = \frac{\eta N_T \rho_\chi}{m_\chi} d^3v_* d^3\vec{q} f(\vec{v}') \left(\frac{|M|^2}{64\pi^2 m_\chi^2 m_N^2}\right) F^2[E_R] \\
\times \delta^{(i)} \left(\frac{q^2}{2m_\chi} + \vec{q} \cdot \vec{v}_* - E_R - \delta\right).
\]

(3.11)

\(F^2[E_R]\) is a nuclear form factor which depends on the type of interaction.
Chapter 3: Directional Signals of Magnetic Inelastic Dark Matter

Figure 3.2: Differential rates $dR/(dE_R d\gamma d\cos \gamma d\nu)$ for (a) $m_\chi = 140$ GeV MiDM benchmark and (b) $m_\chi = 70$ GeV MiDM benchmark. The $m_\chi = 300$ GeV benchmark looks similar to the $m_\chi = 140$ GeV benchmark. In each case we show three two-dimensional distributions, where we have integrated over the third variable. All rates are computed assuming scattering on Xe, and benchmarks are given in Table 3.1. The differential rate is normalized so that the total rate is unity.

For a xenon target, dipole-charge scattering, Eq. 3.8, dominates. For this we use the standard nuclear Helm form factor. Dipole-dipole scattering, Eq. 3.7, is roughly 20% of the total rate. To calculate dipole-dipole scattering a magnetic moment form factor is necessary. The nuclear magnetic moment receives contributions from both spin and angular momentum. We use the spin form factor from [237]. The angular momentum component is $\sim 20 - 30\%$ at zero momentum for Xe. Since dipole-dipole scattering is already subdominant for Xe, and since we do not have accurate angular momentum form factors, we approximate the entire magnetic moment form factor with the spin component.

We now specialize to the case where $f(\bar{\nu})$ is a normalized, truncated Maxwell-Boltzmann distribution, with $v_{esc} = 550$ km/s [256] and $v_0 = 220$ km/s. We assume $v_E = 240$ km/s on
average and label the normalization factor of the distribution as \( n(v_0, v_{\text{esc}}) \). The result is

\[
\frac{dR}{dE_R dv_* dx_*} = \eta N_T \rho v_* \frac{|M|^2}{m_\chi} \frac{32 \pi m_N m_\chi^2}{v_0} F^2[E_R] \Theta(1 - |x_q|) \\
\times \int d\phi \frac{e^{-(v')^2/v_0^2}}{n(v_0, v_{\text{esc}})} \Theta(v_{\text{esc}} - |\vec{v}'|) \tag{3.12}
\]

with the following definitions:

\[
x_q = -\frac{(E_R(m_N/m_\chi - 1) - \delta)}{qv_*}, \quad \text{and} \tag{3.13}
\]

\[
(v')^2 = v_E^2 + q^2/m_\chi^2 + v_*^2 + 2v_E v_* x_* + 2x_q v_* q/m_\chi \\
+ 2v_E q/m_\chi \left(x_q x_* + \sqrt{1 - x_q^2} \sqrt{1 - x_*^2} \cos \phi \right).
\]

An upper bound on \( x_* \) can be extracted from setting \( v' = v_{\text{esc}} \), with \( \cos \phi = -1 \). The bound depends on both \( v_* \) and \( E_R \).

Finally, the matrix elements are given in Eq. 3.7 or in Eq. 3.8. Note that in the dipole-charge scattering case we need to replace \( v \) in Eq. 3.8 using the energy conservation relation,

\[ m_\chi v^2 = 2E_R + m_\chi v_*^2 + 2\delta. \]

The normalized total rate spectrum of several benchmark models is shown in Fig. 3.2.

### 3.4 Sensitivity

XENON100 is collecting several thousand kg·day of exposure. We assume a total exposure of 5000 kg·day on a 40.6 kg fiducial target, in a nuclear recoil energy range of 10-80 keVr. This is consistent with scaling up the results from XENON10 and with preliminary results reported by XENON100.

For the best-fit parameters listed in Table 3.1, this would imply a minimum of \( \sim 100 \) nuclear recoils observable by XENON100. Only \( \sim 10 \) delayed coincidence events are expected,
due to the small size of the detector relative to the average recoil track length. Despite these low efficiencies, a study of the delayed coincidence events is still vastly more informative in two ways: (a) it establishes a directional signal correlated with the WIMP wind, and (b) it is much more sensitive to the parameter space.

### 3.4.1 Directional Detection

We first determine the exposures required to establish a correlation with the WIMP wind. The average nuclear recoil angle with respect to the Earth’s motion, \( \langle \cos \gamma \rangle \), is a robust model-independent statistic for directional detection [218, 153, 128]. Here we use \( \langle \cos \gamma_\ast \rangle \), the WIMP recoil angle with respect to the Earth’s motion. Because of the rotation of the Earth, on average \( \langle \cos \gamma \rangle \) or \( \langle \cos \gamma_\ast \rangle \) should be consistent with 0 for standard backgrounds.

Because XENON100 has excellent spatial resolution, we assume that the recoil track angle can be determined to 10 degrees. We compute the exposures required to obtain a 5\( \sigma \) result for \( \langle \cos \gamma_\ast \rangle \) 90% of the time. We allow for a uniform (isotropic) background, though
the XENON100 background should be negligibly low. The results are shown in Fig. 3.3. The required exposures roughly correspond to a minimum of 16 events at zero background.

### 3.4.2 Parameter Estimation

The predicted rate for delayed coincidence events at XENON100 is only a few counts per 1000 kg · day. However, the additional recoil track information makes it possible to obtain an excellent measurement of the model parameters.

We perform a likelihood analysis, as described in [128], over the parameter space of $m_\chi, \delta$, and $\mu_\chi$. We compute the (relative) log likelihoods for $E$ kg·day on Xe, with nuclear recoil energy range $10^{-8}$ 80 keVr. We neglect the effects of imperfect angular and energy resolution. (The XENON100 energy resolution is $\sim 10\%$ in this energy range, and we estimate an angular resolution of 10 degrees.) The log likelihood is

$$\ln L_{\text{tot}}(p) = E \int dx \left( \mu(x; p_0) \ln \mu(x; p) - \mu(x; p) \right)$$

where $p$ refers to $(m_\chi, \delta, \mu_\chi)$ and $x$ refers generically to the event space of either $E_R$ or $(E_R, v_\ast, \cos \gamma_\ast)$. $p_0$ are the true model parameters. $\mu$ is the rate for parameters $p$. If there is only nuclear recoil energy information,

$$\mu(E_R; p) \equiv \frac{dR}{dE_R}(E_R; p) + dR_{BG},$$

in cpd/kg/keVr for parameters $p$. We assume the background rate, $dR_{BG}$ is known and negligibly small.

If there is directional information,

$$\mu(E_R, v_\ast, x_\ast; p) \equiv \eta_{15}(p) \frac{dR}{dE_R dv_\ast dx_\ast}(E_R, v_\ast, x_\ast; p) + \frac{dR_{BG}}{dv_\ast dx_\ast},$$

where $\eta_{15}(p)$ is a function that depends on the parameters $p$. This expression takes into account the additional directional information provided by the experiment.
Chapter 3: Directional Signals of Magnetic Inelastic Dark Matter

Figure 3.4: Confidence levels for determining $m_\chi$, $\delta$, and $\mu_\chi$, marginalized over the third parameter for each two-dimensional slice. We assume an exposure of 5000 kg·day on Xe in the energy range 10 – 80 keVr. The plots show sensitivity to the MiDM parameter space, using only delayed coincidence data, for the (a) $m_\chi = 70$ GeV benchmark, and (b) $m_\chi = 140$ GeV benchmark. (c) shows the sensitivity using only nuclear recoil events, for the $m_\chi = 140$ GeV benchmark. The directional information is a better test of $m_\chi$ and $\mu_\chi$. The case with $m_\chi = 300$ GeV looks similar to $m_\chi = 140$ GeV.
where \( x_\ast = \cos \gamma_\ast \). \( \eta_{15}(p) \) is the efficiency, for parameters \( p \), at XENON100.

In Fig. 3.4 we show the sensitivity to MiDM parameters if (1) only nuclear recoil information is used and (2) if only delayed coincidence events are considered for 5000 kg \cdot day. We show confidence levels of (68, 90, 95, 99, and 99.9\%). We neglect the Earth’s velocity about the Sun since a livetime of order a year is needed for 5000 kg \cdot day.

Despite the reduction by a factor of 10-50 in events, the directional data is a much stronger constraint on \( m_\chi \) and \( \mu_\chi \). \( \delta \) can also be determined from the \( E_R \) data or the photon energies. In Fig. 3.5 we show the sensitivity to \( m_\chi \) and \( \mu_\chi \) for the \( m_\chi = 140 \) GeV benchmark, assuming that \( \delta \) is already known. The directional information breaks the degeneracy in \( m_\chi \) and \( \mu_\chi \) when only nuclear recoil information is used.

![Figure 3.5: Confidence levels for determining \( m_\chi \) and \( \mu_\chi \) using WIMP recoil tracks, assuming \( \delta \) is already measured from the photon energies or nuclear recoil spectrum. Here we take the \( m_\chi = 140 \) GeV benchmark and assume an exposure of 5000 kg \cdot day on Xe in the energy range 10 – 80 keVr.](image-url)
3.4.3 Measurement of Both Recoils

So far, we considered measurement of the WIMP recoil velocity vector from delayed coincidence events. With a gaseous directional detector, it is also possible to obtain the recoil angle of the nucleus. Then $m_\chi$ and $\delta$ are highly constrained. For such events there are 4 equations and 5 unknowns: $m_\chi$, $\delta$, and $\vec{v}$. However, one can obtain $\delta$ from the energy peak of the coincident photons. Then it is possible to measure the WIMP mass and velocity with just 1 WIMP scattering event. The mass is determined by the following equation:

$$m_\chi = \frac{2m_NE_R}{2(\delta + E_R)} - \sqrt{2m_NE_R \vec{q} \cdot \vec{v}_s}.$$  (3.16)

Since $\vec{q}$ and $\vec{v}_s$ are measured, the initial WIMP velocity $\vec{v}$ is then fixed by momentum conservation. A direct measurement of the WIMP velocity distribution is then also possible.

3.5 Conclusions

The magnetic inelastic dark matter model has an interesting and previously unstudied signature at direct detection experiments: a delayed coincident photon with energy $\delta$. Observation of such photons would also allow current direct detection experiments to become excellent directional detectors.

Motivated by the MiDM setup, we studied several benchmark model parameters that can fit the combined DAMA/NaI and DAMA/LIBRA data. Given the rapidly improving constraints from other experiments, we feel that MiDM is currently the best hope for a dark matter interpretation of DAMA – and it predicts a low-background signature detectable with current experiments.

With 5000 kg·day of exposure, XENON100 can detect the angular modulation of the
recoils and determine the MiDM model parameters. While we focused on benchmarks in MiDM, we emphasize that such a delayed coincidence signal is worth searching for in general. Such events, if found, carry much more information than simple nuclear recoils, and would provide more direct access to the WIMP velocity distribution in our halo.
Chapter 4
Cosmic Ray, Gamma Ray, and Microwave Signals

4.1 Introduction

Several apparent anomalies in recent astrophysical data hint at a new source of high energy electrons, positrons, and possibly gamma rays, at the 10 GeV to 1 TeV scale. The cosmic ray signals observed by Fermi [196, 194, 205, 142] and PAMELA [7] are direct evidence for these energetic electrons and positrons ($e^+e^-$), which would lose their energy primarily through synchrotron radiation and inverse Compton scattering (IC). If the number density of these $e^+e^-$ rises towards the Galactic Center (GC), then this synchrotron and IC could explain the WMAP microwave “haze” [108] and the Fermi diffuse gamma ray “haze” [109], respectively.

It is difficult to explain these signals within the conventional diffusive propagation model and with standard assumptions about the interstellar medium (ISM). In this framework, the
positron signal arises from secondary production from spallation of proton cosmic rays on the ISM. Assuming that 1. positrons and electrons have the same energy losses, 2. primary electrons and protons have the same production spectrum, and 3. the proton escape time decreases with energy, then the predicted positron fraction generically falls with energy, in contrast to the rising fraction observed by PAMELA. Katz et al. [186] point out these assumptions can be wrong, and explore alternative scenarios. Indeed, secondary production at shock fronts could explain the $e^+$ excess [56, 57], but this would also imply an excess of anti-protons, which is not observed. We will not consider these alternatives further.

We examine here whether a new primary source of $e^+e^-$ is a viable explanation of the signals. First, the rise in the positron fraction measured by PAMELA suggests the presence of a new hard source of positrons [248]. Second, the WMAP “haze” is consistent with a hard synchrotron signal in the inner galaxy, in addition to a soft-spectrum synchrotron component traced by Haslam. Though this decomposition is not unique, it is a good fit to the WMAP data. Third, the Fermi gamma ray “haze” similarly extends to $|b| > 30^\circ$ above and below the plane in the inner galaxy. Neither haze correlates with the morphology of any known astrophysical objects or the ISM. (See Fig. 4.1.)

Many attempts to explain the data operate by including a new component of high energy particles and gamma rays originating from one of the following sources:

1. Annihilation of TeV-scale dark matter,

2. Decay of TeV-scale dark matter, or

3. An astrophysical source such as pulsars.

These sources can produce energetic electrons, positrons, and gamma rays. In addition,
Figure 4.1: Maps of the (a) WMAP haze at 23 GHz and (b) Fermi gamma-ray haze at 5-10 GeV for the region $-90 < \ell < 90$ and $-45 < b < 45$. The gamma-ray haze is obtained by subtracting the Fermi diffuse model from the data. All maps are centered on the GC. The data are compared to normalized maps of (c) pulsar synchrotron at 23 GHz and (d) synchrotron at 23 GHz from dark matter annihilation with an Einasto profile. The magnetic field has the form of Eq. 4.5 with $r_B = 6.5$ kpc. The morphology of the haze signals more closely resembles the signals from dark matter than from pulsars. We show the corresponding results for $r_B = 4.5$ kpc in (e) and (f). Choosing a steeper magnetic field can change the morphology, but this is not preferred by the Haslam data (see Fig. 4.2). The region $-5 \leq b \leq 5$ is masked out because we only fit the region $b \leq -5$.

the dark matter distribution in the Galaxy is expected to be roughly spherical, providing at least qualitative agreement with the morphology of the gamma-ray and microwave haze. Nevertheless, each explanation above has drawbacks.
While annihilating dark matter may seem natural given a weak-scale WIMP which has a thermal freeze-out annihilation cross section, this vanilla scenario cannot explain the observed signals. Boost factors in the annihilation rate, arising from substructure or particle physics enhancement, of order 100-1000 are typically needed, depending on the annihilation channels and dark matter mass. Significant model-building effort is also required to explain the lack of excess in the observed $\bar{p}/p$ flux [6]. For examples, see [88, 89, 32, 83, 209].

In the decaying dark matter scenario, dark matter has the freeze-out annihilation cross section but also decays with lifetime $\tau_\chi \sim 10^{26}$ s. These models also must explain why there is no excess in $\bar{p}/p$, though no boost factors are required. Examples include [22, 210, 34, 172, 238, 76, 272].

The pulsar explanation is the least exotic, but there are significant astrophysical uncertainties in pulsar distributions and $e^+e^-$ emission spectra. The Fermi cosmic ray signals can be explained by the presence of one or more nearby pulsars with hard $e^+e^-$ emission spectra [162, 233, 274, 206, 143]. However, pulsars are generally expected to be concentrated in the disk and it can be difficult to explain the shape of the WMAP and Fermi “haze” signals, which are much more spherical. See also [184, 159].

In this work we quantify how well each of these three scenarios can explain the data described above without resorting to model-dependent details of the particle physics or pulsars. Rather we use the data to determine the best-fit injection spectrum of electrons and positrons produced by each new source. We also show that a simple modification to the background electron injection can explain the Fermi $e^+e^-$ spectrum and the Fermi gamma ray spectrum but not the rest of the data.

The standard procedure to analyse whether a model can explain the astrophysical signals
Table 4.1: Typical propagation parameters consistent with low-energy cosmic ray data [106]. We use the “Default” parameters and show the effect of using M1 and MED in Fig 4.5.

<table>
<thead>
<tr>
<th></th>
<th>$K_0$ [kpc$^2$/Myr]</th>
<th>$\delta$</th>
<th>L [kpc]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>0.097</td>
<td>0.43</td>
<td>4</td>
</tr>
<tr>
<td>M1</td>
<td>0.0765</td>
<td>0.46</td>
<td>15</td>
</tr>
<tr>
<td>MED</td>
<td>0.0112</td>
<td>0.70</td>
<td>4</td>
</tr>
<tr>
<td>M2</td>
<td>0.00595</td>
<td>0.55</td>
<td>1</td>
</tr>
</tbody>
</table>

Often, one fits a specific dark matter or pulsar model to a subset of the “anomalous” signals described above. For dark matter, the particle physics model is usually processed through Pythia [253] to generate a spectrum of $e^+e^-$. The injection spectrum is the spectrum of $e^+e^-$ produced per unit source times the rate of production of $e^+e^-$ per source and the spatial distribution of the source. These injected $e^+e^-$ are propagated through the Galaxy to obtain a steady-state solution. The signals are then compared with data.

While some analyses have studied the cases above in a less model-dependent way, the
injection spectrum is assumed to have one of a few common forms [89, 38, 276].

In this work we effectively reverse the arrows in the procedure above. We fit the data from Fermi, PAMELA, and WMAP to expected backgrounds plus a new source which produces positrons and electrons. We assume the injection of the new source is separable in position and space. Rather than specifying the spectrum of $e^+e^-$ injected by the new source, we fit for this injection spectrum as a function of energy. The shape of the spectrum is the same everywhere, and the spatial distribution is varied over several conventional models. Therefore, for our purposes, the three scenarios listed above differ only in their spatial distributions.

In the pulsar case, the assumption that the pulsar injection is a separable function of position and energy is a crude approximation that allows us to fit the data without specifying the details of pulsar physics, since the position dependence of pulsar populations and their $e^+e^-$ injection spectra is very uncertain.

In our fits of the injection spectra, we simultaneously account for possible variations in the conventionally assumed spectrum and spatial distribution of the background injected electrons, as well as propagation parameters, magnetic field, and starlight densities. This takes into account the uncertainties in current models of the Galactic backgrounds.

We describe the signals and their expected backgrounds in more detail in the next section. We then present the overall framework of the analysis. Predictions are computed using GALPROP, and we allow for variations in the background model. We then present the best-fit injection spectrum for each of the three scenarios above, as well as the best-fit of the data to an arbitrary modification of the background electron injection spectrum. Finally we present injection spectra for linear combinations of these scenarios.
4.2 Signals

In this section we review the method of computing the signals and standard assumptions made in modeling the astrophysical backgrounds. However, in our fits we allow for variations in many of these assumptions. This is discussed in more detail in Sec. 4.3.1.

In the conventional diffusive propagation model, the $e^{-}$ cosmic ray density, $dn(\vec{x}, E)/dE$, is the steady-state solution to the diffusion and energy loss equation:

$$\frac{\partial}{\partial t} \left( \frac{dn(\vec{x}, E)}{dE} \right) = 0 \quad (4.1)$$

$$= \vec{\nabla} \cdot \left[ K(E) \vec{\nabla} \frac{dn}{dE} \right] + \frac{\partial}{\partial E} \left[ b(E, \vec{x}) \frac{dn}{dE} \right] + Q(E, \vec{x})$$

where the first term represents diffusion, the second term energy loss, and the third term the source term. $K(E)$ is the diffusion coefficient and $b(E, \vec{x})$ is the energy loss rate. This equation holds separately for positrons. For both electrons and positrons, diffusive re-acceleration and galactic convection are negligible above a few GeV. Those effects are often relevant for other cosmic rays, which are governed by similar equations. We use GALPROP v50p.1 to solve for steady state cosmic ray densities. For a review, see [262].

For electrons, the source term includes primary electrons produced by supernovae and secondary electrons produced by collisions of proton cosmic rays on the ISM. We denote these sources by $Q_0(E, \vec{x})$. The source term can also include any new source of electrons, $Q_1(E, \vec{x})$. For positrons, the source term includes only secondary positrons and any new source of positrons. The spectrum of injected secondary $e^{+}e^{-}$ is determined by the astrophysics of proton cosmic rays and their interactions.

The injected primary electron spectrum is usually assumed to have the following energy
dependence:

\[
\frac{dN}{dE} \propto \begin{cases} 
E^{1.6}, & E < 4 \text{ GeV} \\
E^\gamma_e, & 4 \text{ GeV} < E < 2.2 \text{ TeV} \\
E^{3.3}, & E > 2.2 \text{ TeV}
\end{cases}
\] (4.2)

where \( \gamma_e \) can vary. \( dN/dE \) is the spectrum of \( e^- \) per unit source and is continuous. Eq. 4.2 has often been adopted in the past because the resulting cosmic ray fluxes approximately agreed with the available data. Though we use this form as a default, we will also fit for an arbitrary modification to \( dN/dE \).

The number density for the supernovae that inject these electrons is commonly parametrized as

\[
n_s(x) \propto r^\alpha \exp \left( -\beta \frac{r}{r_\odot} - \frac{|z|}{2 \text{ kpc}} \right) \Theta(r_{max} - r)
\] (4.3)

where \( r \) is distance to the center of Galaxy, projected on the galactic plane, and \( z \) is distance perpendicular to the galactic plane. The default GALPROP parameters are \( \alpha = 2.35, \beta = 5.56283, \) and \( r_{max} = 15 \text{ kpc} \) [261, 204].

The default normalization of the product \( n_s \times dN/dE \) is fixed such that the observed local flux from the primary electrons satisfies

\[
\frac{c}{4\pi} \frac{dn}{dE}(34.5 \text{ GeV}, z = 0, r = r_\odot) = 3.15922 \times 10^{-7} (\text{cm}^2 \cdot \text{sr} \cdot \text{s GeV})^{-1}
\] (4.4)

which is consistent with the flux observed by Fermi.

The diffusion of the injected \( e^+e^- \) is governed by the diffusion coefficient, \( K(E) \), and \( L \), the escape distance out of the galactic plane. \( K(E) \) represents the random walk of a charged particle in tangled magnetic fields, and is approximated as constant in space. It is
generally assumed that \( K(E) = K_0(E/\text{GeV})^\delta \), where \( K_0 \) and \( \delta \) are propagation parameters. In Table 4.1 we give some commonly used values of \( K_0, \delta, \) and \( L \) \([106, 102]\). Our default model assumes \( K_0 = 0.097 \text{kpc}^2/\text{Myr}, \delta = 0.43, \) and \( L = 4 \text{kpc} \), though we will vary these parameters later. This choice matches cosmic ray data for protons, the B/C cosmic ray ratio, and was used in \([85]\).

As injected electrons and positrons propagate, they lose energy. The energy loss rate \( b(E, \vec{x}) \) includes energy loss mechanisms. The path length for an electron or positron losing \( 1/e \) of its energy is typically given by \( \sim \sqrt{KE/b} \), which is \( \sim 1 \text{kpc} \) around 1 GeV and becomes shorter for higher energies, at least until the Klein-Nishina limit \([259]\). The dominant mechanisms for energy loss are IC scattering and synchrotron, where \( b(E) \propto E^2 \). Bremsstrahlung (brem), for which \( b(E) \propto E \), is sub-dominant above \( \sim 1 \text{GeV} \) and is far more localized to the disk. For a new high energy component of \( e^+e^- \), we neglect bremsstrahlung.

The IC rate depends strongly on the model for the interstellar radiation field. We use the default GALPROP model \([220]\) as a baseline. Meanwhile, the synchrotron depends on the model for the magnetic field. We assume a standard parametrization of the field,

\[
|B| = B_0 \exp \left( -\frac{r - r_\odot}{r_B} \right) \exp \left( -\frac{|z|}{z_B} \right),
\]

\( r \) is the distance to the center of the Galaxy, projected on the galactic plane. Typical parameters are \( B_0 = 5\mu\text{G}, r_B \sim 5 - 10 \text{kpc}, \) and \( z_B = 2 \text{kpc} \). \(^1\) For our default propagation parameters, the GALPROP synchrotron prediction at 408 MHz best matches the Haslam 408 MHz map if \( r_B \approx 8.5 \text{kpc} \); see Fig. 4.2.

This parameterization is consistent with observations of the large-scale (ordered) mag-

\(^1\)The documentation for GALPROP v50p incorrectly states that their parameter \( B_0 \) is the magnetic field in the center of the galaxy.
Figure 4.2: (a) Haslam 408 MHz map. In the other panels we subtract the default GALPROP model for (b) \( r_B = 4.5 \) kpc, (c) \( r_B = 6.5 \) kpc, and (d) \( r_B = 8.5 \) kpc from the Haslam map. The GALPROP model is normalized such that the total emission in the region \( \ell \in [-10, 10], b \in [-90, -5] \) matches the Haslam 408 MHz intensity in the same region. The constant offset is subtracted from the Haslam 408 MHz data. Note that local features like the North polar spur are not modeled in GALPROP and hence are not fit.

The solution to Eq. 4.1 is the steady-state cosmic ray density, which then determines the photon signals. The gamma ray flux includes decay of \( \pi^0 \)s produced in proton cosmic ray collisions with the gas in the ISM, IC scattering of \( e^\pm \) on interstellar photons, and bremsstrahlung of \( e^\pm \) colliding with the ISM. The gamma-ray power in a given direction
scales as:

\[ P_{\text{x}^0} \propto \int n_{\text{gas}}(s) n_p(s) \, ds , \quad (4.6) \]
\[ P_{\text{IC}} \propto \int n_s(s) n_{e^\pm}(s) \, ds , \quad (4.7) \]
\[ P_{\text{brem}} \propto \int n_{\text{gas}}(s) n_{e^\pm}(s) \, ds. \quad (4.8) \]

where \( s \) is the coordinate along the line of sight. The \( \sim 23 \) GHz microwave flux off the Galactic plane is primarily synchrotron radiation of electrons and positrons

\[ P_{\text{synch}} \propto \int |B(s)|^2 n_{e^\pm}(s) \, ds \quad (4.9) \]

where \( B \) is the magnetic field.

A new source such as dark matter or pulsars can inject high energy electrons and positrons at 10 GeV to 1 TeV. These new sources are included in \( Q_1(E, \vec{x}) \). Here we solve for the \( e^+e^- \) injection spectrum which, after propagation, yields the observed cosmic ray spectrum and gives rise to gamma rays and synchrotron radiation. Our fit will essentially determine \( Q_1(E, \vec{x}_0) \), where \( \vec{x}_0 \) is the Earth’s location. The spatial dependence of \( Q \) is fixed to be one of a few conventional models.

These sources can also directly inject photons. There are primary photons from pulsars which are important at lower gamma-ray energies. Given our energy range of interest, we do not consider these further.

In the case of dark matter annihilation or decay, generally there are many channels through which dark matter produces Standard Model particles, which can then decay on short time scales. The end products are \( e^\pm \), neutrinos, and photons. However, we do not consider these direct gamma rays any further. These can be produced from \( \pi^0 \)’s, final state
radiation\textsuperscript{2} from $\tau^\pm$s or $\mu^\pm$s, or a direct photon channel. For TeV-scale dark matter, these gamma rays can have higher energies than those observed by Fermi. Furthermore, in the fits below it is not difficult to produce enough gamma ray signal above 10-100 GeV. In fact, direct gamma ray production can be rather constrained by observations [90, 226].

4.2.1 Data

We fit to the following data:

- PAMELA $J(e^+)/\left(J(e^-) + J(e^+)\right)$ positron fraction, which displays a steep rise from 10-100 GeV [7]

- Fermi $(e^+ + e^-)$ cosmic ray spectrum, which shows a slight hardening of the spectrum at a few hundred GeV [196, 194, 205]

- Fermi gamma ray spectrum, which shows a hardening of the spectrum at around 10-100 GeV, averaged over the haze region $-15 < \ell < 15$ and $10 < |b| < 30$. Note the pion signal has been subtracted from the data [109]. Our background models match the pion component, shown in Fig. 11 of [109]. This is not affected by the inclusion of new sources of electrons.

- WMAP synchrotron at 23, 33, and 41 GHz averaged over $-10 < \ell < 10$ for $-90 < b < -5$, in 2 degree bins. We also fit to the same data averaged over $10 < |\ell| < 30$, which we call the “high $\ell$” region of the WMAP data and is incorporated to include morphological information from the microwave haze. [108]

\textsuperscript{2}In some papers, final state radiation is referred to as internal bremsstrahlung. We use “bremsstrahlung” exclusively to mean \textit{e}\textsuperscript{+}e\textsuperscript{-} cosmic rays colliding with the ISM.
Figure 4.3: The fits in this work are consistent with Fermi observations of the gamma ray spectrum in the region $|\ell| \leq 3$, $|b| \leq 3$. Point sources have been subtracted. Solid colored lines show the predicted total signal for the best fits of the new sources considered here, including backgrounds. The dashed lines show the contribution from only the new source.

These data describe the “anomalous signals”, which suggest the presence of a new source of electrons and positrons, and possibly gamma rays, at roughly 10-1000 GeV.

The Fermi LAT collaboration has provided a reference model for the diffuse emission [126], a detailed fit that includes a reference GALPROP model for IC and models for a number of residual local features giving rise to bremsstrahlung at lower energies. Since we are not studying the detailed structure of the diffuse gamma rays and because IC and pions dominate at high energies, it is sufficient for us to use GALPROP to model the diffuse gamma ray emission in the haze region.

We also do not attempt to fit the Fermi gamma-ray spectrum near the Galactic Center region nor the Fermi isotropic gamma rays. The signal near the GC suffers from large uncertainties in both the dark matter profile and the astrophysical backgrounds. The isotropic
signal is extremely sensitive to the halo mass function. Some recent analyses have used these sets of data to constrain dark matter explanations of cosmic-ray signals, for a variety of dark matter models and spatial distributions [90, 91, 226, 75, 171]. In Fig. 4.3 we show that for the best fit spectra and spatial distributions in this work, there is little or no tension between the predicted total signal and the data in the GC region.

Finally, we also include the following data in order to help regularize the fit at lower energies and higher energies:

- AMS $e^+ + e^-$ cosmic ray spectrum below 10 GeV [15]

- HESS $e^+ + e^-$ cosmic ray spectrum above 900 GeV [8]

Fitting to the data from AMS ensures that the background models are consistent with the low energy cosmic ray data.

We include systematic errors in our analysis and treat them as statistical errors because we do not have the full covariance matrix. The energy calibration error of the Fermi data points is $^{+5\%}_{-10\%}$, but rather than effectively increasing the error bars, we allow for freedom in the normalization of the background, discussed in the next section. The 15% energy calibration error has been included in the error bars used for HESS.

### 4.2.2 Solar Modulation

Though our focus is on high energy data, we ensure that our results are consistent with the low energy (below $\sim$10 GeV) cosmic ray data from PAMELA and AMS. However, this data is extremely sensitive to the very local propagation in the heliosphere. To relate the data to GALPROP predictions for the local interstellar (LIS) spectrum outside the heliosphere, it
is necessary to apply solar modulation corrections to $e^+ + e^-$ spectra. The solar modulation correction in the force-field approximation is

$$\frac{J_\odot(E)}{E^2 - m_e^2} = \frac{J_{LIS}(E + \Phi)}{(E + \Phi)^2 - m_e^2}$$  \hfill (4.10)$$

where $\Phi$ is the solar modulation parameter and $J$ is the differential intensity $dn/dE$ [144]. Because of the uncertainty in the force field approximation, we reduce the weight of the PAMELA and AMS data points below 10 GeV, effectively multiplying error bars by a factor of 3. This is adequate to stabilize the fits at low energy.

The solar modulation correction is applied to the GALPROP outputs. We also use the correction when converting the positron fraction data of PAMELA into a positron flux data, using the AMS data on the intensity of $e^+ + e^-$. This will allow the fit to be linear below.

However, these two data sets correspond to different parts of the solar cycle. We thus apply an inverse solar modulation correction to the AMS data to obtain the unmodulated positron intensity. Denoting the solar modulation correction by $\hat{S}_\Phi$, then the positron signal is obtained from

$$J_{PAM}(e^+) = \left( \frac{J(e^+)}{J(e^-) + J(e^+)} \right)_{PAM} \times \hat{S}_{\Phi PAM} \left( \hat{S}_{\Phi AMS}^{-1} \left[ J_{AMS}(e^+ + e^-) \right] \right)$$  \hfill (4.11)$$

where $\Phi$'s are solar modulation parameters. $\Phi_{PAM}^-$ is the solar modulation parameter for the PAMELA electrons, which we allow to be different from $\Phi_{PAM}^+$ for the positrons. This approximately captures the charge dependence of the solar physics, visible in the time-dependent positron fraction at lower energies. In the above equation we applied $\hat{S}_{\Phi PAM}^-$ to the total $e^+ + e^-$ signal. Because the positrons are at most $\sim 10\%$ of the total flux, this approximation is justified.
4.3 Fitting procedure

We fit for the $e^+e^-$ injection spectrum that, when combined with a background model, best matches the cosmic ray, gamma ray, and microwave observations. The steady-state $e^+e^-$ density is linear in the source function $Q(E, \vec{x})$, so we take a Green’s function approach in energy space. The spatial dependence is fixed to be one of a few conventional models in each of the cases below.

We inject delta functions of $e^+e^-$ at various energies and compute the signal from each delta function with GALPROP. Since GALPROP is discretized, in practice this amounts to propagating an appropriately normalized bin of energy. For each of these delta functions, GALPROP computes the steady-state $e^+e^-$ spectrum as well as maps of synchrotron and IC radiation at various energies. We solve for the linear combination of these outputs that best matches the data. The best-fit injection spectrum solution is simply the same linear combination of delta function injections (or in our case, energy bins).

We inject $e^+e^-$ via the source term $Q(E, x)$ in the propagation equation, Eq. 4.1. For dark matter annihilation, dark matter decay, and pulsars, the new source function $Q_1(E, \vec{x})$ of both positrons and electrons can be written as

$$Q_1(E, \vec{x}) = \begin{cases} 
\frac{dN}{dE} \langle \sigma v \rangle_0 BF \frac{\langle \rho^2 \rangle}{m^2_\chi} \frac{f_E}{2}, \text{ ann} \\
\frac{dN}{dE} \tau^{-1}_\chi \frac{\rho_\chi}{m_\chi} \frac{f_E}{2}, \text{ decay} \\
\frac{dN}{dE} \tau^{-1}_p n_p, \text{ pulsar.}
\end{cases}$$

(4.12)

Here $dN/dE$ is the spectrum of electrons or positrons produced per unit source, normalized such that all the power per unit source goes into electrons. The specific condition can be found in Sec. 4.4.1 for dark matter annihilation and Sec. 4.4.2 for dark matter decay.
density and mass of the dark matter. $(\sigma v)_0$ is the thermal freeze-out cross section for annihilation, $3 \cdot 10^{-26}\text{cm}^3/\text{s}$. $BF$ is a boost factor (from either particle physics or astrophysics such as substructure enhancement). $\tau_\chi$ is the lifetime in the case of dark matter decay. $\tau_p$ and $n_p(\vec{x})$ are rate and density parameters associated with pulsar emission rate and number density. Finally $f_E = f_E(e^+ + e^-)$ is the fraction of energy going to electron-positron pairs. If $f_E = 1$, then the total energy of the electrons will be equal to $m_\chi$ for dark matter annihilation and $m_\chi/2$ for dark matter decay.

We also consider arbitrary modifications to the energy dependence of the background primary electron injection, Eq. 4.2. To accomplish this, we include an extra source of only electrons which has the same spatial distribution as the supernovae, Eq. 4.3:

$$Q_1(E, \vec{x}) = \frac{dN}{dE} \tau_s^{-1} n_s, \text{ supernova} \quad (4.13)$$

where $\tau_s$ is an arbitrary rate parameter that is fixed by matching to the data.

Because we do not wish to a priori specify model parameters, we instead implement the scenarios above with the following electron injection:

$$Q_1(E, \vec{x}) = \begin{cases} 
Q_1(E, \vec{x}_0) \left( \frac{n_\chi(\vec{x})}{n_\chi(\vec{x}_0)} \right)^2, & \text{ann} \\
Q_1(E, \vec{x}_0) \left( \frac{\rho_\chi(\vec{x})}{\rho_\chi(\vec{x}_0)} \right), & \text{decay} \\
Q_1(E, \vec{x}_0) \left( \frac{n_p(\vec{x})}{n_p(\vec{x}_0)} \right), & \text{pulsar} \\
Q_1(E, \vec{x}_0) \left( \frac{n_s(\vec{x})}{n_s(\vec{x}_0)} \right), & \text{supernova}. 
\end{cases} \quad (4.14)$$

where the local injection, $Q_1(E, \vec{x}_0)$ will be determined by the fit ($\vec{x}_0$ is our location in the galaxy). The positron injection is the same, except in the case of the source injection where there are no positrons injected. Only the spatial profiles distinguish dark matter annihilation, dark matter decay, or pulsars, in our fits.
We bin the energies of the new source, $Q_1(E, \vec{x})$, and treat the particles in each energy bin independently. For example, we generally consider the energy range $\sim 5$-5000 GeV with 17 log spaced bins. The propagation of a given injection spectrum is just a linear combination of the propagation of each of the energy bins.

The problem can be treated linearly because high energy $e^+e^-$ are a tiny perturbation to the matter and radiation of the Galaxy. High energy $e^+e^-$ also almost never interact with each other or other cosmic rays; they dominantly interact with the ISM, radiation, and magnetic fields. In GALPROP, the magnetic field is fixed and the usual feedback between cosmic rays and $B$ field is absent. In this limit, the propagation of the individual energy bins is independent.

We use this linearity to invert the propagation problem and determine the injected spectrum $Q_1(E, \vec{x}_0)$, given some assumptions about the spatial density of dark matter or pulsars. Define the vector $\mathbf{x}$ by $x_i = Q_1(E_i, \vec{x}_0)$ for energy bin $E_i$. The injection everywhere else is determined by the assumed spatial distribution. Also, let $b_j$ be the $j$th data point minus the galactic background, computed by GALPROP, for that data point.

For each $x_i$, we propagate the injection and obtain a signal $A_{ji}$. Thus $\mathbf{A}$ is a matrix which maps $\mathbf{x}$ to the predicted signal, and the columns of $\mathbf{A}$ give the predicted signal from each energy bin. We wish to compare the signal from the new source, $\mathbf{A} \cdot \mathbf{x}$, with the background-subtracted data, $\mathbf{b}$.

The best fit $\mathbf{x}$ is determined by a goodness-of-fit test, which for a linear problem is a quadratic in the fit parameters:

$$\chi^2 = (\mathbf{A} \cdot \mathbf{x} - \mathbf{b})^T \mathbf{C}^{-1} (\mathbf{A} \cdot \mathbf{x} - \mathbf{b})$$

(4.15)

where $\mathbf{C}$ is a covariance matrix containing the errors on the data. This is just a quadratic
minimization problem. Note that we also include several other parameters in \( x \) that allows us to slightly modify the background predictions and improve the fit. This is described further in the following subsection.

Finally, it is possible to obtain \( dN/dE \) and the other parameters in Eq. 4.12. This will be possible for the dark matter scenarios with additional constraints on \( \int E \, (dN/dE) \, dE \) and where \( dN/dE \) cuts off. This is discussed more below when we describe the results for the scenarios above.

### 4.3.1 Uncertainties

The predictions for the signals discussed in this chapter can have significant theoretical and astrophysical uncertainty. To capture the effects of these uncertainties, we include several parameters in the fits that essentially allow for (small) variations in the background model.

The main uncertainties are in the background primary and secondary \( e^+e^- \) injection, since we are fitting all the data that could possibly constrain this. (For a more detailed discussion of uncertainties in the cosmic rays signals, see [105, 252].)

In the primary injection, we allow \( \gamma_e \) in Eq. 4.2 to vary in discrete steps. We also fit for an arbitrary normalization factor \( N_p \) relative to the condition in Eq. 4.4. Usually we find \( N_p \approx 1.0 \) because the condition was chosen to approximately match the Fermi cosmic ray data. The source spatial distribution for primary electrons, Eq. 4.3, is also uncertain. Rather than considering the full range of possible source distributions, we simply allow for a different normalization factor of the primary electron spectrum near the Galactic Center, relative to Eq. 4.4. Because the diffusion length is \( \sim 1 \) kpc, this will not affect the local cosmic ray
Figure 4.4: All fit results for the three scenarios, over a $3 \times 3$ grid in background electron injection index ($\gamma_e = 2.45, 2.55, 2.65$) and scale for the magnetic field $r_B = 4.5, 6.5, \text{ and } 8.5 \text{ kpc}$. These spectra were obtained from non-negative fits; the interpolated local injection density is plotted. Despite a wide range of assumptions about the background model, the results remain the same, qualitatively, for each scenario.

signal. This normalization factor, $N_h$, is fixed by matching the synchrotron background prediction onto the Haslam 408 MHz synchrotron map [160] for $-90 < b < -5$, averaged over $-10 < \ell < 10$. The contribution from the new high-energy source is negligible at this frequency.

As for the secondary $e^+e^-$, it was shown in [106] that modifying the propagation parameters effectively changes the overall normalization of the local steady-state secondary positron flux by up to an order of magnitude. Thus rather than scanning over a large set of propagation parameters consistent with all the low energy cosmic ray data, we allow the
normalization of the secondaries to be a fit parameter, $N_s$.\(^4\)

There are several other adjustable parameters that can improve the fit and allow for variations in the background model. We give the complete list below. None of these will change the predictions for other cosmic ray data.

The following parameters characterize the uncertainties of our theoretical models. We fit for these parameters simultaneously with the injection spectrum, as their effects can also be treated linearly:

- $N_{IC}$: The normalization of the background IC signal near the center of the galaxy, relative to the GALPROP prediction. There are many uncertainties in the starlight density and spatial variations in the primary electron density near the galactic center.

- $\Delta S$: Zero-points of the WMAP signal. We allow a different one for each data set.

- $N_s$: Normalization of secondary local electron spectrum, relative to the GALPROP output for our choice of propagation parameters. This can vary by up to an order of magnitude given theoretical uncertainties [104, 252].

- $N_p$: Normalization of local primary electron spectrum, relative to Eq. 4.2. As mentioned above, this factor does not have to be the same as $N_h$.

We include these parameters in $x$, and $A$ is enlarged to include extra columns corresponding to each of these background signals.

The signals are not linear in the following parameters, so we scan over a discrete set of these:

\(^4\)We show the effect of changing propagation parameters on some fits in Fig. 4.5. The result does not differ significantly relative to the error bars.
• $r_B$: The $r$-scale of the galactic $B$ field, where the local $B$ field is fixed to 5 $\mu$G. See Eq. 4.5. We include $r_B = 4.5, 6.5$, and 8.5 kpc, corresponding to $B = 33, 18$, and 14 $\mu$G in the center of the Galaxy. We used $z_B = 2$ kpc.

• $\gamma_e$: The index of the primary electron injection spectrum above 4 GeV. We include $\gamma_e = 2.45, 2.50, 2.55, 2.60, 2.65$, and 2.70.

• $\Phi_{AMS}, \Phi^+_PAM, \Phi^-_{PAM}$: Solar modulation parameters for AMS and PAMELA, in the force-field approximation, as described in Section 4.2.2.

Though we allow these to be fit parameters, clearly in reality they have some definite form independent of our model. In Fig. 4.4 we show the best fit for a grid in $r_B$ and $\gamma_e$. Though the spectra do change, the qualitative features remain roughly the same.

4.3.2 $\chi^2$ minimization and regularization

We are minimizing

$$\chi^2 = (A \cdot x - b)^T C^{-1} (A \cdot x - b)$$

(4.16)

where $x$ is a vector of parameters we fit for, containing the injection spectrum as well as the normalization parameters and offsets described above. $C$ is a covariance matrix, so it is symmetric and positive. It can then be shown that the matrix $A^T C^{-1} A$ is positive semi-definite.

Ideally the spectrum we derive is smooth and non-negative. However, the existence of null (or nearly null) eigenvalues of $A^T C^{-1} A$ means that there are directions in the parameter space where we can modify the spectrum by large values with little change to the observed signals. This corresponds to, for example, changing the spectrum for two neighboring energy
Figure 4.5: In the left column we show the space of possible solutions within $\Delta \chi^2 = \pm 1$ of the best-fit solution, which was obtained from a non-negative fit. The red lines are the best fits for (a) dark matter annihilation, (c) dark matter decay, and (e) pulsars. The spectra shown are interpolated between the bins (marked by red stars). In the right column we show other best fits obtained from using different propagation parameters, given in Table 4.1, or a different fitting regulator that enforces smoothness of the solution (from Eq. 4.17). Our results are robust to very different propagation parameters. The fit for M2 is not shown because $L$ is only 1 kpc. Since the haze signals extend out to $\sim 4$ kpc or more, it is impossible for this set of propagation parameters to produce the haze.
bins by a large positive and negative amount respectively, such that the observed signal remains nearly the same.

We regularize the spectrum by using only 17 log-spaced bins between 5 GeV and 5000 GeV. We also perform a non-negative quadratic fit following the algorithm in [249]. All of the fit parameters should be positive except for $\Delta I$ which we find is always of the same sign for our data, so we can choose a convention where it is positive.

To obtain errors on the spectrum, we find the eigenvectors and eigenvalues of $A^T C^{-1} A$. This allows us to change basis from the parameter space in $x$ to a new parameter space $y$ where $\chi^2$ is separately parabolic in each parameter. The variance of these new parameters $y$ is determined by computing the allowed shift of each parameter, relative to the best fit, such that $\Delta \chi^2 = \pm 1$. Even though a non-negative constraint was imposed for the best fit, we consider the entire space of solutions within $\Delta \chi^2 = \pm 1$.

Each $x_i$ is a linear combination of the $y_i$, so we sum the squares of the contribution from each $y_i$ to find the variance in $x_i$. The quoted error on each $x_i$ is the square root of the variance. Because we are performing a non-negative fit, however, the positive and negative errors can be different.

In Fig. 4.5 we show the entire range of possible variations of the best fit injection spectrum with $\Delta \chi^2 = \pm 1$. We add to the best-fit spectrum all possible variations along the eigendirections, or all independent variations of $y_i$. (We do not show the background normalization coefficients and WMAP offsets, though they are simultaneously varying with the injection spectrum.)

We also considered several alternative methods of regularization, rather than non-negativity.
As an example, we can impose smoothness by adding terms to $\chi^2$:

$$
\chi_{eff}^2 = \chi^2 + \eta_1 (D_E \cdot x)^T \cdot (D_E \cdot x)
+ \eta_2 (D_E^2 \cdot x)^T \cdot (D_E^2 \cdot x)
$$

(4.17)

where $D_E$ and $D_E^2$ are finite difference and second-difference matrices, respectively. These matrices only act on the injection spectrum and not the other fit parameters such as normalization factors and WMAP offsets. $\eta_1$ and $\eta_2$ are tunable parameters that control the smoothness of the fit. In Fig. 4.5 we show the best fit using this regulator instead of the non-negative regulator above. For an appropriate range of $\eta_1$ and $\eta_2$ the solution is qualitatively similar to the non-negative result. Similarly, we tested several other regulation techniques, such as suppressing variations in nearly null eigendirections. Again, for “reasonable” regulators, the result is qualitatively similar.

4.4 Results

We determined the best-fit injection spectrum for 350 data points from Fermi, PAMELA, WMAP, AMS, and HESS. There are 29 fit parameters coming from 17 energy bins, 3 normalization factors, 6 WMAP offsets, and 3 solar modulation parameters. Including $r_B$ and $\gamma_e$, then there are 31 fit parameters. Our results are summarized in Table 4.2 and in Figs. 4.7-??.

The details of the fit results for each scenario can be found in the following sections.

In each of the following figures, we show the fits to the

- $e^+e^-$ flux data from Fermi, AMS, and HESS
- positron flux obtained from combining the AMS and PAMELA data in Eq. 4.11
Table 4.2: Best fit parameters for annihilating dark matter, decaying dark matter, and pulsar cases to 350 data points. Ann1 and Ann2 had nearly the same $\chi^2$ but had different $r_B$ so both results are displayed. In the supernova (SN) injection case there were 91 data points. We obtained mass, boost factor, and lifetime parameters from the best fit. In the last two columns we show fit results for linear combinations of these three scenarios. The fit errors on the normalization parameters $N$ are less than 5-10% and thus are not shown. †For the combination cases, we fixed the dark matter profiles to have Einasto $\alpha = 0.17$.

- positron flux fraction $J(e^+)/ (J(e^-) + J(e^+))$ from PAMELA for comparison, though we did not directly fit to this data
• WMAP synchrotron data at 23 GHz and 33 GHz, and 23 GHz at high $\ell$; the data for 41 GHz and the high $\ell$ data for 33 GHz and 41 GHz are included in the fit, but not shown because the fit looks extremely similar to the plots already shown

• Haslam 408 MHz data, used to fix $N_h$, as discussed in Sec. 4.3.1

• Fermi gamma ray data, where the $\pi^0$ background has been subtracted [109] along with the best-fit local injection, $E^2 Q_1(E, \bar{x}_0)$.

Before discussing the fits in detail, we emphasize that the results in Table 4.2 and the spectra plotted here are not meant to be taken as precise answers but as qualitative guidelines for the necessary spectra, for each scenario, in order to explain the data. As shown in Fig. 4.4, the spectra vary with the background model, but the general features remain the same. Errors and variations in the solution were discussed in Section 4.3.2. In addition, the effect of changing propagation parameters is shown in Fig. 4.5.

Specific bumps and features in the spectra we find are more likely signs that the smooth background models we have assumed are not adequate. If there is any large systematic or unmodeled effect in the Fermi cosmic ray data, for example, it can change the features in best fit spectrum significantly. In particular, note that the shape of the high energy region of each spectrum above $\sim$100 GeV is only constrained by the high energy Fermi cosmic ray data since the Fermi gamma ray data is primarily only sensitive to the total power in this energy range. The other data are almost completely insensitive to such high energy particles. Thus the error bars on these bins are typically the largest. Furthermore, the high energy spectrum is more sensitive to changes in $\gamma_e$ (see Fig. 4.4).

The low energy part of the spectrum is more severely constrained by all of the data. However, the low energy spectrum is also extremely sensitive to the bumps and features in
the Fermi cosmic ray spectrum at low energies. This is very likely a sign that some features in the Fermi cosmic ray spectrum have not been properly included in the background model. For example, in Fig. ?? we show a fit which allows both dark matter annihilation and an arbitrary modification to the energy dependence of the supernovae-injected electrons. The low-energy features can be fit by a modification of the supernova electron spectrum, while dark matter annihilation is still necessary to explain the signals above 10-20 GeV.

Fig. 4.6 shows the “supernova” fit of the low energy PAMELA data, all of the $e^+e^-$ data, and the gamma ray data to a disk-like source with only electrons. This corresponds to a modification of the background primary electron spectrum and is implemented using the injection in Eq. 4.14. The best fit spectrum we found is a hardening of the injection up to 1 TeV. Though this source modification can easily match the cosmic ray or IC data, the disk-like spatial profile and lack of positrons produced are starkly inconsistent with the synchrotron signal and the PAMELA data. A new source is required.

4.4.1 Annihilating Dark Matter Results

The form of the injection for annihilating dark matter was given in Eq. 4.14 and Eq. 4.12. We assume the local dark matter density is $\rho_0 = 0.4$ GeV/cm$^3$ [70].

Conventionally used dark matter halo density profiles are obtained by simulations and can be approximated by an Einasto profile, with $0.12 \lesssim \alpha \lesssim 0.22$ and $\alpha \approx 0.17$ on average [222]. This does not include substructure effects which can modify the effective spatial profile used in Eq. 4.12, as in [193].

We allow values of $\alpha = 0.12, 0.17,$ and $0.22$, with a core radius of $r_{-2} = 25$ kpc. In practice the shallower profile with $\alpha = 0.22$ is always the best fit to avoid overproducing
The gamma ray signal. These profiles only differ by a factor of $\sim 2$ at .1 kpc from the center of the galaxy. Though NFW profiles are also commonly used, their signatures can be approximated by one of these Einasto profile. We also considered spatial profiles which were Einasto squared times an $r^{1/4}$ or $r^{1/2}$ scaling, corresponding to an $r$ dependent cross section [86]. Using these profiles can improve the $\chi^2$ by 5-10, but the injection spectrum does not
Figure 4.7: Best fit for the annihilating dark matter scenario, with $\chi^2_{\text{red}} \approx 0.44$. The spatial profile of the dark matter is Einasto with $\alpha = 0.22$. Black dashed lines are the background prediction for a model with $\gamma_e = 2.5$ and $r_B = 8.5$ kpc. Red dashed lines give the contribution of the new source injection, and solid red lines are the total. The gray shaded region is the error estimate on the best-fit injection spectrum.

change significantly.

In the annihilating case we found best fits with magnetic fields of $r_B = 4.5$, 6.5, and 8.5 kpc, all with $\chi^2$ around 140 and $\chi^2_{\text{red}} \approx 0.44$. Conventional magnetic field models have $r_B$ closer to 8.5 kpc. Furthermore, in this case, the normalization factors $N$ are $\sim 1$, so that the model is self-consistent. Thus we show the fit with $r_B = 8.5$ kpc in Fig. 4.7. In Table 4.2 we
give the fit parameters for $r_B = 6.5$ kpc under the column “Ann2”.

For all three magnetic fields above, we found that an injection index of $\gamma_e = 2.5$ for the primary electron signal optimized the ratio between the PAMELA and the Fermi $e^+ + e^-$ apparent dark matter components. However, for $r_B = 8.5$ kpc the fit does not match the PAMELA data as well, as an excess of cosmic rays above 100 GeV can produce too many gamma rays through IC scattering.

We can estimate several model parameters from the best-fit spectrum by relating Eq. 4.14 and Eq. 4.12. To find the dark matter mass, we assume $dN/dE$ cuts off at around $m_\chi$. Though this estimate of $m_\chi$ depends on the rather uncertain high-energy part of the injection spectrum, values of roughly 1 TeV are expected given the turnover in the $e^+ + e^-$ data around 600-1000 GeV and the turnover in the gamma-ray spectrum at 100-200 GeV.

Next, $dN/dE$ was defined such that that the total energy of the emitted electrons sums to the mass of dark matter:

$$\int E \frac{dN}{dE} dE = m_\chi. \quad (4.18)$$

Therefore, integrating the local injection multiplied by energy gives

$$\int E Q_1(E, \vec{x}_0) dE = \langle \sigma v \rangle_0 BF \frac{(\rho_0)^2}{m_\chi} f_E(e^+ + e^-). \quad (4.19)$$

Given an estimate of $m_\chi$, we can therefore estimate $BF \times f_E(e^+ + e^-)$ in terms of the best fit local injection and known parameters.

### 4.4.2 Decaying Dark Matter Results

For the decaying dark matter case, we assume the same range of dark matter density profiles as in the annihilating case. Again, in practice we will be limited to the case where
Figure 4.8: Best fit for the decaying dark matter scenario, with $\chi^2_{red} \approx 0.41$. The spatial profile of the dark matter is Einasto with $\alpha = 0.12$. Black dashed lines are the background prediction for a model with $\gamma_\chi = 2.6$ and $r_B = 4.5$ kpc. Red dashed lines give the contribution of the new source injection, and solid red lines are the total. The gray shaded region is the error estimate on the best-fit injection spectrum.

$\alpha = 0.12$. This time a steeper profile is required to produce sufficient synchrotron signal to fit the WMAP data.

The model parameters can be determined from Eq. 4.12 and Eq. 4.14. We assume $dN/dE$ cuts off at around $m_\chi/2$ this time. Again, this cutoff is rather sensitive to the high-energy part of the spectrum, which has large error bars, but values of $\gtrsim 2$ TeV are expected given
Figure 4.9: Best fit for the pulsar scenario, with $\chi^2_{\text{red}} \approx 0.46$. The pulsar profile is given by Eq. 4.22. Black dashed lines are the background prediction for a model with $\gamma_e = 2.6$ and $r_B = 4.5$ kpc. Red dashed lines give the contribution of the new source injection, and solid red lines are the total. The gray shaded region is the error estimate on the best-fit injection spectrum.

By definition, $dN/dE$ satisfies

$$
\int E \frac{dN}{dE} \, dE = m_\chi.
$$

(4.20)
Again, we integrate the local injection multiplied by energy, giving
\[
\int E Q_1(E, \vec{x}_0) \, dE = \tau^{-1}_\chi \rho_0 \frac{f_E(e^+ + e^-)}{2}. \tag{4.21}
\]
This allows us to determine the dark matter lifetime over the energy fraction. However, note that in many cases, \(dN/dE\) does not cut off in the energy ranges we consider and the spectrum is essentially unconstrained at higher energies. Then we only obtain bounds on the mass and lifetime.

The best fit is shown in Fig. 4.8. There is no clear mass cutoff in the best-fit spectrum, so the mass of the particle can be from \(\sim 4\) TeV to greater than 16 TeV.

Because in the decaying scenario the injected power is proportional to \(\rho_\chi\) and not \(\rho_\chi^2\), generally it is harder to generate enough synchrotron and IC signal. Both of these signals are in regions at least 5 degrees off of the galactic plane. The steeper dark matter profile with \(\alpha = 0.12\) is not enough to produce the signals.

We found \(r_B = 4.5\) kpc can increase synchrotron near the center of the galaxy, but this gives a somewhat unconventionally high value of the magnetic field in the GC, 33\(\mu\)G. Fig. 4.2 shows that \(r_B = 4.5\) kpc also gives the poorest fit to the Haslam data, especially compared to \(r_B = 8.5\) kpc. In addition, a somewhat large injection of low energy electrons and positrons is required. However, for this large magnetic field, the IC signal drops. Thus the normalization \(N_{IC}\) is rather large, \(N_{IC} \sim 2.5\). Even for fits with \(r_B = 6.5\) kpc, it was necessary for \(N_{IC} \sim 2\) to obtain sufficient IC signal. This corresponds to rather high starlight density.

While it is possible that the decaying dark matter can also produce gamma rays directly or through FSR, these signals are typically at higher energies, above 10-100 GeV. In this case, the large \(N_{IC}\) factor for the background IC signal indicates that there is a depletion of
gamma rays at low energies, below 10 GeV.

Though the decaying scenario nominally gives the best $\chi^2 \approx 130$ and $\chi^2_{\text{red}} \approx .41$, the large normalization factors demand a more self-consistent modeling of backgrounds and uncertainties in order to fully justify the goodness of fit.

### 4.4.3 Pulsar Results

The range of types of pulsars, their spatial distributions, and their $e^+e^-$ spectra is not well determined. As a crude model, we posit some spatial profile for the number density of pulsars and assume the spectrum of electrons and positrons has the same energy dependence everywhere. Generally pulsars are concentrated in the galactic disk, making it difficult to produce the haze. In Fig. 4.1 we compare the morphology of the synchrotron haze produced by pulsars to that produced by dark matter. It is possible, however, that certain types of pulsars have a more spherical distribution [207].

We consider the following range of pulsar profiles, which span those typically used in the literature. (See [159, 116] for examples and references.) Assume a density profile of the form

$$n_p \propto \exp \left(-\frac{|z|}{z_p}\right) n_\rho(\rho)$$  \hspace{1cm} (4.22)

where $\rho = \sqrt{x^2 + y^2}$ and the origin is at the center of the galaxy. We include profiles with $z_p = 0.08, 0.5, \text{ and } 2 \text{ kpc}$. For the radial profile,

$$n_\rho \propto \exp \left(-\frac{\rho}{4.5 \text{ kpc}}\right).$$  \hspace{1cm} (4.23)

In practice, the three cases above look nearly identical because of diffusion. Another commonly used profile has $n_\rho \sim \rho \exp (-\rho/4.0 \text{ kpc})$. However we do not consider this option
further because the suppressed density near the center of the galaxy makes it even more
difficult to produce the haze.

Qualitatively, the pulsar results, Fig. 4.9, are rather similar to the decaying dark matter
results, though the fits are even worse because of the disk-like rather than spherical profile.
The best fit has $r_B = 4.5$ kpc with significant low-energy injection and large normalization
factor $N_{IC}$ of 2.6. Though it is possible that pulsars can produce many low-energy gamma
rays, it is unlikely these gamma rays can compensate for the background gamma-ray signal
being $2 - 3$ times too low. For $r_B = 6.5$ kpc or 8.5 kpc, the pulsar scenario cannot produce
sufficient synchrotron signal.

### 4.4.4 Combination Results

We fit for linear combinations of annihilation, decay, pulsars, and supernova injections,
which not surprisingly can provide better fits and alleviate the problems of each individual
scenario. However, this extra freedom means that fits are much less constrained and errors in
the spectra. Results of these fits are presented in [199]; the best fit parameters in summarized
in Table 4.2.

### 4.5 Conclusions

We have thoroughly examined the annihilating dark matter, decaying dark matter, and
pulsar explanations of the recent anomalous cosmic ray, gamma ray, and synchrotron signals.
We investigated whether each scenario can fit all of these data simultaneously. Our analysis
is independent of the particle physics or pulsar model details of each scenario and only de-
pends on the spatial profiles and background models. We determined the necessary injection
spectrum of electrons and positrons in each case in order to reproduce the data, including the effects of different background models, propagation models, and solar modulation.

Though decaying dark matter is the best fit, the large normalization factors suggest that it will be difficult to find a fully self-consistent model with decaying dark matter that can explain the data, without changing some aspect of our model by a large amount. In particular, it may be necessary to find either a radiation field model with roughly twice as much starlight to produce enough IC, a rather large magnetic field of $33\mu G$ in the Galactic center, enormous amounts of low energy electrons or gamma rays injected, a much steeper dark matter profile, or a combination of these.

Pulsars give the worst fit; the disk-like profile makes it nearly impossible to produce both the gamma ray and synchrotron signals. Much like the decaying dark matter case, this suggests that dramatic re-assessments of backgrounds and models are necessary to find a self-consistent interpretation of the data.

Annihilating dark matter, however, has self-consistent fits with conventional astrophysical background models. Though we had to choose a somewhat shallower dark matter halo profile with Einasto $\alpha = 0.22$, it is still within the current range of profiles found by simulations. Furthermore, we can satisfy the gamma ray constraints from the GC. The boost factors are $\sim 70/f_E$, which at first seems much lower than the boost factors of $\sim 1000$ often used in the literature. Several factors enter in this difference: our use of the updated $\rho_0 = .4 \text{ GeV/cm}^3$ rather than $\rho_0 = .3 \text{ GeV/cm}^3$ [70], the relatively hard spectrum allowed by the fit, and our assumption that the cutoff of the spectrum is $m_{\chi}$. Given these factors, our result of $\sim 70/f_E$ is typical of the models discussed in the introduction of this chapter. However, the shape of the spectrum, combined with a lack of $\pi^0$ or $\bar{p}$ production, may still be difficult to achieve
in current particle physics models of dark matter.

Our results should be regarded as qualitative guidelines to injection spectra. The specifics will necessarily change as both Galactic models and data are refined. The WMAP “haze” data will be superseded by data from Planck [228], while data from Fermi and PAMELA will improve. In addition, cosmic ray data from AMS-02 [20] may also soon be available. If the data does not change substantially, and if current models indeed describe Galactic propagation and interactions, then the qualitative results of this work will remain valid.
Chapter 5

CMB Constraints on Dark Matter Annihilation

5.1 Introduction

Measurements of the cosmic microwave background (CMB) in the past decade by experiments including WMAP, ACBAR and BOOMERANG [191, 236, 217], and more recently SPT, QUaD and ACT [69, 62, 96], have provided an unprecedented window onto the universe around redshift 1000. With the advent of the Planck Surveyor [228], the successor experiment to WMAP, percent-level modifications to recombination will be observable. Planck has already completed three sky surveys and begun a fourth, and cosmological data are expected to be released publicly in 2012-13.

Accurate measurements of the CMB have the potential to probe the physics of dark matter (DM) beyond its gravitational interactions. In the large class of models where the DM is a thermal relic, its cosmological abundance is determined by its annihilation rate in
the early universe: the correct relic density ($\sim 22\%$ of the energy density of the universe) is obtained for an $s$-wave annihilation cross section of $\langle \sigma v \rangle \approx 3 \times 10^{-26} \text{ cm}^3/\text{s}$ during freezeout.

DM annihilation at this rate modifies the ionization history of the universe and has a potentially measurable effect on the CMB. During the epoch of recombination, DM annihilation produces high-energy photons and electrons, which heat and ionize the hydrogen and helium gas as they cool. The result is an increased residual ionization fraction after recombination, giving rise to a low-redshift tail in the last scattering surface. The broader last scattering surface damps correlations between temperature fluctuations, while enhancing low-$\ell$ correlations between polarization fluctuations.

The resulting constraints on the dark matter annihilation rate have been studied by several authors [225, 139, 254, 80, 138, 170, 140]. These bounds have a notable advantage over other indirect constraints on dark matter annihilation, in that they are independent of the DM distribution in the present day, and do not suffer from uncertainties associated with Galactic astrophysics. They depend only on the cosmological DM density, which is well measured; the DM mass; the annihilation rates to the final states; and the standard physics of recombination. Recombination modeling, while not simple, involves only well-understood conventional physics, and the latest models are thought to be accurate at the sub-percent level required for Planck data [81, 17].

Current limits from WMAP already significantly constrain models of light dark matter with masses of around a few GeV and below, if the annihilation rate at recombination is the thermal relic cross section. Heavier DM is also constrained if the annihilation rate is enhanced at low velocities or for other reasons is much larger than the thermal relic cross section at recombination. Models lying in these general categories are also of significant
interest for their possible connection with experimental anomalies.

The DAMA/LIBRA [47] and CoGeNT [2, 1] direct detection experiments have reported excess events and annual modulation that may have a consistent explanation as originating from scattering of $O(5 - 10)$ GeV WIMPs (e.g. [163]). Results from the XENON10, XENON100 and CDMS experiments are in tension with this interpretation [241, 13], but there is ongoing debate on the sensitivity of these experiments to the very low-energy nuclear recoils in question (see e.g. [97, 98]).

The PAMELA [7], Fermi [3], PPB-BETS [266], ATIC [71] and H.E.S.S [8] have measured electron and positron cosmic rays in the neighborhood of the Earth, and found results consistent with a new primary source of $e^\pm$ in the $10 - 1000$ GeV energy range. If the signal is attributed to dark matter annihilation then the annihilation rate in the Galactic halo today must be 1-3 orders of magnitude above the canonical thermal relic value [84, 83]. This has motivated models of dark matter with enhanced annihilation at low velocities [32, 230]. While this enhancement would not be significant during freezeout, it would be effective at recombination when the typical velocity of dark matter is $v \sim 10^{-8}c$ [139].

With the release of data from Planck expected in the next two years, models falling into these categories should either be robustly ruled out, or give rise to a measurable signal [139, 254, 170]. If no signal is observed, the sensitivity of Planck will allow us to probe regions of parameter space relevant for supersymmetric models, where the DM is a thermal relic with mass of several tens of GeV. It is timely to explore improvements to these constraints.

The approach of previous studies has been to specify the energy deposition history (redshift dependence) and then calculate the effect on the ionization history and anisotropy spectrum using public codes such as RECFAST and CAMB. A single parameter describing the
normalization of the signal is then added to the standard likelihood analysis using \texttt{CosmoMC}, and bounded by WMAP observations. The redshift dependence has been studied in two cases: in the “on-the-spot” case, assuming that the amount of energy deposited to the gas precisely tracks the rate of dark matter annihilation (e.g. \cite{225, 139, 140}), or employing detailed energy deposition histories for specific models (e.g. \cite{254, 170, 140}). In the first case, model-independent constraints are obtained, but without a precise way to connect the bounds to any particular model. The second case only precisely constrains specific models.

While these analyses have been adequate for simple estimates of whether a model is strongly ruled out, easily allowed, or on the borderline, the upcoming high-precision data from \textit{Planck} demand a more careful model-independent analysis. Such an analysis can also be applied to more general classes of energy deposition histories during and after recombination: for example, the energy deposited by a late-decaying particle species, decay from an excited state of the dark matter, or dark matter annihilation in models where the redshift dependence of the annihilation rate has an unusual form (as in some models of asymmetric dark matter).

In this work we exploit the fact that the effects of energy deposition at different redshifts are not uncorrelated. Any arbitrary energy deposition history can be decomposed into a linear combination of orthogonal basis vectors, with orthogonal effects on the observed CMB power spectra ($C_{\ell}$’s). For a broad range of smooth energy deposition histories, the vast majority of the effect on the $C_{\ell}$’s can be described by a small number of independent parameters, corresponding to the coefficients of the first few vectors in a well-chosen basis. These parameters in turn can be expressed as (orthogonal) weighted averages of the energy deposition history over redshift.

We employ principal component analysis (PCA) to make this statement quantitative and
derive the relevant weight functions, and the corresponding perturbations to the $C_\ell$ spectra. Our approach in principle generalizes to all possible energy deposition histories. To investigate the number of observable parameters, we consider generic perturbations about two physically interesting fiducial cases. We focus primarily on the example of dark matter annihilation, or any other scenario where the power deposited per volume scales approximately as $(1 + z)^6$ (i.e. as density squared), as an energy deposition mechanism, but also show results for the case of dark matter decay, or similar scenarios where the power deposited scales as $(1 + z)^3$.

Our computation of the effects of energy deposition on the CMB anisotropies, and the approximations we use for estimating the significance of these effects in experimental data, are described in §5.2. In §5.3 we present our principal component analysis for both “annihilation-like” and “decay-like” general energy deposition histories\(^1\). There are significant degeneracies between energy deposition and perturbations to the cosmological parameters, and so we marginalize over the standard cosmological parameters when deriving the principal components\(^2\).

We then address the constraints on and detectability of the principal components in current and future experiments. Given a $C_\ell$ spectrum observed by an experiment (e.g. Planck), we can measure the residual with respect to the best-fit standard ΛCDM model, and then project this residual onto the $C_\ell$-space directions corresponding to the principal components. Given any model for the energy deposition history, we can then ask if the reconstructed coef-

\(^1\)Files containing the results of these analyses are available online at http://nebel.rc.fas.harvard.edu/epsilon/.

\(^2\)We test the effect of including additional cosmological parameters (running of the scalar spectral index, the number of massless neutrino species, and the primordial He fraction) and find no large degeneracy with energy injection, justifying our neglect of these additional parameters in our main analysis.
ficients for the various principal components are consistent with the model. Of course, for the later principal components the effect on $C_\ell$'s is so small that very little information on their coefficients can be recovered. In §5.4, we make this statement quantitative, and estimate the number of principal components whose coefficients could be detectable in Planck and an ideal cosmic variance limited (CVL) experiment, subject to constraints from WMAP 7. The CVL case presents a hard upper limit on the number of independent parameters describing the energy deposition history that can profitably be retained in the analysis. We also discuss the bias to the standard cosmological parameters, in the case where there is non-zero energy deposition that is neglected in the analysis; in our framework it is straightforward to characterize the biases to the cosmological parameters for an arbitrary energy deposition history.

In §5.5 we present a separate principal component analysis for the more limited case of conventional GeV-TeV WIMPs annihilating to Standard Model final states. We demonstrate that in this case, all the effect of dark matter annihilation can be captured by one parameter only, i.e. the amplitude of the first principal component.

Finally, in §5.6, we estimate the constraints on the principal components obtainable with current (WMAP 7) and future (Planck, CVL) experiments with a full likelihood analysis using the CosmoMC code. We employ here the principal components obtained with the Fisher matrix analysis – which assumes that the effect on the CMB scales linearly with the energy deposition. We illustrate the range of validity of this assumption for the different experiments considered. We check that the constraints previously obtained with our Fisher matrix analysis – which assumes Gaussian likelihood functions – are compatible with the ones obtained with the CosmoMC analysis. We check that the constraints on a given energy deposition
history can be reconstructed from the constraints on the principal components.

In [127], we also consider the effects on the analysis of changing various assumptions and conventions, including the effect of additional cosmological parameters and using different codes to calculate the ionization histories. We find that the only such choice that non-negligibly modifies the early (detectable) principal components is the treatment of Lyman-α photons, although the inclusion of additional cosmological parameters can change the constraints at the $\sim 10\%$ level.

### 5.2 The Effect of Energy Injection

We begin by considering DM annihilation-like or decay-like energy deposition histories. The energy injection from these sources scales respectively as density squared and density,
so these cases cover the generic scenarios where energy is injected by two-body or one-body processes. It is convenient to express the energy injection as a slowly varying function of $z$ that depends on the source of the energy injection (e.g. the WIMP model) and a term containing cosmological parameters. We parameterize the energy deposition histories, respectively, as,

$$\left(\frac{dE}{dt \, dV}\right)_{\text{ann}} = p_{\text{ann}}(z)c^2\Omega_{DM}^2\rho_c^2(1 + z)^6,$$

$$\left(\frac{dE}{dt \, dV}\right)_{\text{dec}} = p_{\text{dec}}(z)c^2\Omega_{DM}\rho_c(1 + z)^3,$$

(5.1)

where $p_{\text{ann}}(z)$ (or $p_{\text{dec}}(z)$) contains all of the information about the source of energy injection and the efficiency with which that energy ionizes the gas. We generically refer to $p_{\text{ann}}$ and $p_{\text{dec}}$ as the “energy deposition yield.” For consistency with [140], we express $p_{\text{ann}}(z)$ in units of cm$^3$/s/GeV, while the units of $p_{\text{dec}}(z)$ are s$^{-1}$. If the energy injection is due to DM annihilation, $p_{\text{ann}} = f(z)\langle\sigma v\rangle/m_{DM}$ [140], where $f(z)$ is an $O(1)$ dimensionless efficiency factor [254]; if the energy injection is due to DM decay, $p_{\text{dec}}(z) = f(z)/\tau$, where $\tau$ is the decay lifetime. Other authors have written $p_{\text{ann}}$ in units of m$^3$/s/kg [139], or parameterized the energy deposition in eV/s/baryon [225, 254, 80]. For calibration, the energy deposition from a 100 GeV thermal relic WIMP with $f(z) = 1$ corresponds to $p_{\text{ann}} \approx 3 \times 10^{-28}$cm$^3$/s/GeV $\approx 1.7 \times 10^{-7}$m$^3$/s/kg, or an energy deposition of $2.1 \times 10^{-24}$ eV/s/H, assuming the WMAP 7 best-fit cosmology. Throughout this work, we employ the cosmological parameters from [195] as a baseline: explicitly, $\omega_b = 2.258 \times 10^{-2}$, $\omega_c = 0.1109$, $A_s(k=0.002 \text{ Mpc}^{-1}) = 2.43 \times 10^{-9}$, $n_s = 0.963$, $\tau = 0.088$, $H_0 = 71.0$ km/s/Mpc.

Energy deposition during recombination primarily affects the CMB through additional ionizations, as studied in [225, 78]; the modified ionization history then leads to an increased width for the surface of last scattering, which in turn modifies the temperature and polariz-
tion anisotropies of the CMB. Electrons and photons injected at high energies – where the cross section for direct ionization is small – scatter on the CMB and on the gas, partitioning their energy into (1) many low-energy electrons and photons that efficiently ionize, excite and heat the H and He, and (2) X-ray and gamma-ray photons that free-stream to the present day (a detailed study of the relevant processes is given in [254]). It is the first component (ionization + excitation + heating) that we refer to as “deposited energy”; as shown in [254], the bulk of the energy injected in photons and electrons/positrons is deposited. Restricting our attention to this “deposited” component, the scattered electrons from excitations and ionizations in turn re-scatter, rapidly partitioning their energy between excitation, ionization and heating in a ratio that depends on the ambient ionization fraction, but has little dependence on the initial spectrum of electrons and photons [250]. Thus the effect on the CMB is completely determined by the redshift dependence of the energy deposition, which we refer to as the “energy deposition history”; further details of the energy injection are largely irrelevant. The excitations created by energy deposition can modify recombination via additional Lyman-alpha photons, but the ionizations have the greatest direct effect on the ionization history, the surface of last scattering and the CMB anisotropies.

Energy deposition also results in $\mu$-type spectral distortions of the CMB, if energy is deposited at redshifts $z \gtrsim 5 \times 10^4$, as well as $y$-type distortions from energy injected at lower redshifts [174, 158, 275]; an order of magnitude estimate of the effect is $\delta \rho_\gamma / \rho_\gamma \sim (dE/dt)/H/\rho_\gamma$. However, the bounds on $|\mu|$ and $|y|$ from COBE/FIRAS [132] give a weaker constraint on DM annihilation by a factor of $\sim 10^5$, compared to the limits from WMAP measurements of CMB anisotropies.

An alternate approach to studying generic energy deposition histories might be to study
generic ionization histories [167], since the former can be directly mapped to the latter. We frame the problem in terms of energy deposition histories because they can be more directly mapped to physical energy injection models.

Suppose we are interested primarily in a class of energy deposition histories for which the energy deposition yield \( p(z) \) (that is, \( p_{\text{ann}}(z) \) or \( p_{\text{dec}}(z) \), as appropriate) is not very rapidly varying. Then we can discretize \( p(z) \) as a sum over a basis of \( N \) \( \delta \)-function-like energy deposition histories, \( p(z) = \sum_{i=1}^{N} \alpha_i G_i(z) \). The basis functions \( G_i(z) \) are (by default) Gaussians with \( \sigma = \Delta z/4 \), centered on \( z_i \) (\( i = 1..N \)), where \( \Delta z \) is the spacing between the \( z_i \). They are normalized such that \( \int dz G_i(z) = \Delta z \). For example, in the limit of large \( N \) an energy deposition history with constant \( p(z) = p_0 \) corresponds to \( \alpha_i = p_0 \) for all \( i \).

If the energy deposition is small enough, the effect on the CMB anisotropy power spectrum is linear in the energy depositions at different redshifts,

\[
\delta C_l(p(z)) = \delta C_l(\sum_{i=1}^{N} \alpha_i G_i(z)) = \sum_{i=1}^{N} \delta C_l(\alpha_i G_i(z)),
\]

and in the amount of energy deposition at any redshift,

\[
\delta C_l(\sum_{i=1}^{N} \alpha_i G_i(z)) = \sum_{i=1}^{N} \alpha_i \delta C_l(G_i(z)).
\]

Then the effect of an arbitrary energy deposition history can be determined simply from studying the basis functions \( G_i(z) \). We will assume linearity throughout this work; we justify that assumption in §5.2.3.

Of course, given any annihilation-like energy deposition history, it can be rewritten in decay-like form with a strongly redshift-dependent \( p(z) \), and vice versa. The basis of \( G_i(z) \) functions can describe any energy deposition history, at least in the large-\( N \) limit. However, the very different “underlying” redshift dependence in the two cases, and the uncertainties
associated with the annihilation rate at low redshift (due to the onset of structure formation), motivate us to study different redshift ranges in the two cases.

For each $G_i$, we can compute the effect on the ionization history and the anisotropy spectrum in the limit of small energy deposition. We determine $\partial C_{TT}^{\ell}/\partial \alpha_i$, $\partial C_{EE}^{\ell}/\partial \alpha_i$, $\partial C_{TE}^{\ell}/\partial \alpha_i$ $\forall i, \ell$. In our default analysis we employ the CosmoRec and CAMB codes, with the prescription for including the extra energy deposition laid out in [78, 225]. If there are $N$ basis functions and we take $n_{\ell}$ spherical harmonics into account, this yields an $n_{\ell} \times N$ transfer matrix $T$ whose $(\ell, i)$th element is,

$$\frac{\partial C_{\ell}}{\partial \alpha_i} = \left\{ \frac{\partial C_{TT}^{\ell}}{\partial \alpha_i}, \frac{\partial C_{EE}^{\ell}}{\partial \alpha_i}, \frac{\partial C_{TE}^{\ell}}{\partial \alpha_i} \right\}. \quad (5.2)$$

In this work we focus primarily on annihilation-like energy deposition histories, for which we restrict ourselves to the $80 < z < 1300$ range; as a default, we will take 50 redshift bins covering this range. At higher redshifts the universe is ionized and so the effect of energy deposition on the ionization history is negligible, while at lower redshifts the DM number density becomes so small that the energy injected from annihilation is insignificant, as shown in Figure 5.1. This in turn justifies our neglect of DM structure formation: while for $z \lesssim 100$, DM clumps start to form and the annihilation rate no longer tracks the square of the average relic density, the energy injection is already sufficiently suppressed that the signal remains negligible.

For DM decay, the signal is not nearly so suppressed at low redshifts, and so we consider the redshift range $10 < z < 1300$. With this expanded redshift range, we switch from linear to log binning, with 90 bins covering this redshift range\textsuperscript{3}; we take the basis functions

\textsuperscript{3}Log binning can of course also be employed for the annihilation-like case; there is no clear best choice there, so we will use linear binning as the default but show results for both options. See Appendix A of [127] for a discussion.
\(G_i(\ln(1 + z))\) to be Gaussians in \(\ln(1 + z)\), normalized so that their integral with respect to \(d \ln(1 + z)\) is given by the spacing between the log bins \(\Delta \ln(1 + z)\). With these choices, again an energy deposition history with constant \(p(z) = p_0\) corresponds (in the large-\(N\) limit) to \(p_0 \sum_{i=1} G_i\).

We again ignore structure formation in the decay-like case, where the total power injected depends only on the average density. The universe is rather transparent to the products of DM decay and annihilation at these redshifts, so even a very spatially non-uniform distribution of energy injection would not be expected to cause ionization or temperature hot-spots (at least for particles injected at weak-scale energies; de-excitation of nearly-degenerate states or annihilation of very light DM might change this conclusion to some degree). Modeling of reionization may pose a more significant challenge for analyses relying on low redshifts \((z \sim 10)\); note, however, that the transparency of the universe at these redshifts means that in realistic scenarios (even decay-like scenarios) the bulk of the effect on the CMB comes from earlier times.

### 5.2.1 Brief review of the Fisher matrix

The degree to which energy deposition is observable in the CMB can be captured by the Fisher matrix for energy deposition, denoted \(F_e\), which is obtained by contracting the transfer matrix \(T\) (Equation 5.2) with the appropriate covariance matrix for the \(C_\ell\)'s (e.g.
\[ \Sigma_\ell = \frac{2}{2\ell + 1} \times \left( \begin{array}{ccc} (C_{TT}^\ell)^2 & (C_{TE}^\ell)^2 & C_{TT}^\ell C_{TE}^\ell \\ (C_{TE}^\ell)^2 & (C_{EE}^\ell)^2 & C_{EE}^\ell C_{TE}^\ell \\ C_{TT}^\ell C_{TE}^\ell & C_{EE}^\ell C_{TE}^\ell & \left( (C_{TE}^\ell)^2 + C_{TT}^\ell C_{EE}^\ell \right) \end{array} \right), \]

\[(F_e)_{ij} = \sum_\ell \left( \frac{\partial C_\ell}{\partial \alpha_i} \right)^T \cdot \Sigma_\ell^{-1} \cdot \frac{\partial C_\ell}{\partial \alpha_j}. \tag{5.3} \]

For experiments other than the perfect cosmic variance limited (CVL) case, noise is included by replacing \( C_{TT,EE}^\ell \rightarrow C_{TT,EE}^\ell + N_{TT,EE}^\ell \), where \( N_\ell \) is the effective noise power spectrum and is given by:

\[ N_\ell = (\omega_p)^{-1} e^{\ell(\ell+1)\theta^2} \tag{5.4} \]

Here \( \theta \) describes the beam width (FWHM = \( \theta \sqrt{8 \ln 2} \)), and the raw sensitivity is \( (\omega_p)^{-1} = (\Delta T \times FWHM)^2 \), with all angles in radians. The standard deviation of the parameter \( \alpha_i \), marginalized over uncertainties in the other parameters, is given by \( \sigma_{\alpha_i} \geq (F_{-1})_{ii}^{1/2} \). The parameter \( \alpha_i \) is then detectable at 1\( \sigma \) if its signal-to-noise \( \alpha_i/\sigma_{\alpha_i} \) is larger than 1.

So far, we have not taken into account covariance between the standard cosmological parameters and the energy deposition parameters, but in fact there are significant degeneracies between them. In particular, shifting the primordial scalar spectral index \( n_s \) can absorb much of the effect of energy deposition [225, 139]. Therefore we must marginalize over the cosmological parameters, since the naively most measurable energy deposition history may be strongly degenerate with them and thus difficult to constrain. We parameterize the usual six-dimensional cosmological parameter space by the following set of parameters: the physical baryon density, \( \omega_b \equiv \Omega_b h^2 \), the physical CDM density, \( \omega_c \equiv \Omega_c h^2 \), the primordial scalar
spectral index, $n_s$, the normalization, $A_s(k = 0.002/\text{Mpc})$, the optical depth to reionization, $\tau$, and the Hubble parameter $H_0$.

Using exactly the same machinery as described above for the energy deposition histories, we determine the derivatives of the $C_\ell$’s with respect to changes in the cosmological parameters, again assuming that these changes are in the linear regime. Then these $C_\ell$ derivatives are vectors spanning an $n_c$-dimensional subspace of the space of all $C_\ell$ derivatives (where for the standard parameter set $n_c = 6$); only directions orthogonal to this subspace can be constrained. We can regard marginalization over the cosmological parameters as simply projecting out the components of the energy deposition derivatives orthogonal to this subspace\(^4\).

In analogy with Equation 5.3, we now use the derivatives with respect to both energy deposition and the cosmological parameters to construct the full Fisher matrix,

$$ F_0 = \begin{pmatrix} F_e & F_v \\ F_v^T & F_c \end{pmatrix}, $$

(5.5)

where $F_e$ is the Fisher matrix for solely the energy deposition parameters, $F_c$ is the Fisher matrix of the cosmological parameters, and $F_v$ contains the cross terms. The usual prescription for marginalization is to invert the Fisher matrix, remove the rows and columns corresponding to the cosmological parameters, and invert the resulting submatrix to obtain the marginalized Fisher matrix $F$ (e.g. [269]). When the number of energy deposition parameters is much greater than the number of cosmological parameters, it is convenient to take advantage of the block-matrix inversion,

\(^4\)See Appendix B of [127] for a detailed explanation of this projection and how it relates to the standard marginalization prescription.
\[ F_0^{-1} = \begin{pmatrix} (F_e - F_v F_c^{-1} F_v^T)^{-1} & - (F_e - F_v F_c^{-1} F_v^T)^{-1} F_v F_c^{-1} \\ -F_c^{-1} F_v^T (F_e - F_v F_c^{-1} F_v^T)^{-1} F_c^{-1} \left( 1 + F_v^T (F_e - F_v F_c^{-1} F_v^T)^{-1} F_v F_c^{-1} \right) \end{pmatrix} \]

We can now read off the marginalized Fisher matrix as \( F = F_e - F_v F_c^{-1} F_v^T \) (note that \( F \) has the same units as \( F_e \)).

The Fisher matrix approach to estimate detectability is optimistic in the sense that it assumes the likelihood function is Gaussian about its maximum; for non-Gaussian likelihoods, the significance of a given energy deposition history will generally be smaller, and any constraints on the amount of energy deposition will be weakened [269]. We verify that the Fisher matrix method gives results consistent with previous studies of WMAP limits on constant \( p_{\text{ann}} \) [139, 140].

### 5.2.2 Experimental parameters

For comparison to the existing literature and constraint forecasting, we consider the WMAP 5, WMAP 7 and Planck experiments, as well as a theoretical experiment that is CVL up to \( \ell = 2500 \). The beam width and sensitivity parameters for WMAP and Planck are given in Table 5.1. We use only the \( W \) band for WMAP and the 143 GHz band for Planck, under the conservative assumption that the other bands will be used to remove systematics. The effect of partial sky coverage is included by dividing \( \Sigma_\ell \) by \( f_{\text{sky}} = 0.65 \).
### Table 5.1: Detector sensitivities and beams for different CMB temperature and polarization experiments.

Results for WMAP temperature sensitivity are taken from [177], with the noise reduced by $\sqrt{5/4}$ ($\sqrt{7/4}$ for WMAP 7) to account for the longer integration time. The polarization noise for WMAP is taken to be $\sqrt{2}$ times the temperature noise. WMAP beam widths are taken from [46]. The sensitivity and beam width for Planck are taken from the Planck Blue Book, available at [http://www.rssd.esa.int/SA/PLANCK/docs](http://www.rssd.esa.int/SA/PLANCK/docs), and assume 14 months of Planck data.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Beam FWHM (arcmin)</th>
<th>$10^6 \Delta T/T$ (I)</th>
<th>$10^6 \Delta T/T$ (Q,U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMAP (5 yr, Q band)</td>
<td>29</td>
<td>6.7</td>
<td>9.5</td>
</tr>
<tr>
<td>WMAP (5 yr, V band)</td>
<td>20</td>
<td>7.9</td>
<td>11.1</td>
</tr>
<tr>
<td>WMAP (5 yr, W band)</td>
<td>13</td>
<td>7.6</td>
<td>10.7</td>
</tr>
<tr>
<td>Planck (100 GHz)</td>
<td>10</td>
<td>2.5</td>
<td>4.0</td>
</tr>
<tr>
<td>Planck (143 GHz)</td>
<td>7.1</td>
<td>2.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Planck (217 GHz)</td>
<td>5.0</td>
<td>4.8</td>
<td>9.8</td>
</tr>
</tbody>
</table>

5.2.3 Numerical stability of derivatives and linearity

When dealing with general energy deposition histories, we hope to work in a regime where the effect of deposition on the CMB is linear, so that the effect of a general energy deposition history can be described in terms of a linear combination of basis energy deposition histories. This is the idea behind characterizing the effect of new parameters entirely in terms of the transfer matrix of derivatives, $T$, and the Fisher matrix $F$ derived from it. Equivalently, linearity means it is sensible to speak of a single transfer matrix $T$ largely independent of the “fiducial” energy deposition history about which the derivatives $\partial C_\ell / \partial \alpha_i$ are taken (our default assumption is that this “fiducial” energy deposition is zero). If the energy deposition
Figure 5.2: The first three principal components for WMAP 7, Planck and a CVL experiment, both before and after marginalization over the cosmological parameters.

history being studied is too great a perturbation away from the fiducial, the first derivatives will no longer accurately describe its effect on the $C_\ell$’s, and the Fisher matrix estimate of its significance will break down. In this subsection we discuss the numerical stability of the derivatives, and the degree to which they describe the effect of arbitrary energy deposition histories on the $C_\ell$’s.

The derivatives used in the Fisher matrix are evaluated at the fiducial cosmology (with no energy deposition). The assumption of linearity is that these derivatives are still correct away from the fiducial. For the standard set of six cosmological parameters, the biases to the cosmological parameters induced by the maximum permitted energy deposition from WMAP 5 generally lie well within the linear regime.

For large energy deposition, the effect on the $C_\ell$’s is nonlinear, i.e. not directly proportional to the deposited power as parameterized by the $\alpha_i$; equivalently, the derivatives about a fiducial large energy deposition are not the same as for zero energy deposition. Our polynomial fits for the derivatives, described above, also allow us to check the extent to which nonlinearity may become important: that is, the extent to which $O(\alpha_i^2)$ corrections
to the effect on the $C_\ell$’s are non-negligible.

The amount of energy deposition such that nonlinearities become important depends on redshift $z$. This can be estimated by the fractional rate of ionization per Hubble time, $(dn_{ion}/dt)/(n_{ion}^0 H(z))$, arising from the energy deposition (where $dn_{ion}/dt$ is related to $dE/dtdV$ according to the prescription of [250]). For two fiducial cases this quantity is shown in Figure 5.1. Conversely, the energy deposition at redshift $z$ such that $(dn_{ion}/dt)/(n_{ion}^0 H(z)) = 1$ gives a measure of what energy deposition is required before nonlinearities may become significant. For each redshift bin, we use the polynomial fits of $\delta C_\ell(\alpha_i)$ to numerically calculate the derivatives at this level of energy deposition. We then find 1% corrections (averaged over $\ell$) to the fiducial derivative $(\partial C_\ell/\partial \alpha_i)|_{\alpha_i=0}$.

### 5.3 Principal Component Analysis

The effects of energy deposition at different redshifts on the $C_\ell$’s are highly correlated, and so the effects of a large class of energy deposition histories can be characterized by a small number of parameters. Principal component analysis provides a convenient basis into which energy deposition histories can be decomposed, with the later terms in the decomposition contributing almost nothing to the effect on the $C_\ell$’s. It thus allows generalization of constraints on energy deposition to a wide range of models (subject to the linearity assumption discussed above).
Figure 5.3: The first six principal components for Planck after marginalization, in the case of (left) annihilation-like redshift dependence with linear binning, (center) annihilation-like redshift dependence with log binning, and (right) decay-like redshift dependence with log binning. Note that for decay-like energy deposition histories, the redshift range is extended down to \( z = 10 \) in order to fully capture the effect on the CMB - see §5.2. This larger redshift range makes linear binning impractical.

5.3.1 The principal components

Having obtained the marginalized Fisher matrix \( F \), diagonalizing \( F \): 

\[
F = W^T \Lambda W, \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)
\]  

yields a convenient basis of eigenvectors or “principal components”. \( W \) is an orthogonal matrix in which the \( i \)-th row contains the eigenvector corresponding to the eigenvalue \( \lambda_i \). If we compute derivatives for \( N \) redshift bins, then the \( N \times N \) Fisher matrix has \( N \) principal components. The eigenvectors are orthonormal in the space of vectors \( \{\alpha_i\}, i = 1..N \). Let us label these vectors \( e_i \), with corresponding eigenvalues \( \lambda_i, i = 1..N \). Our convention is to rank the principal components by decreasing eigenvalue, such that \( e_1 \) has the largest eigenvalue.

Note that the principal components may be significantly different from the un marginalized principal components, or the eigenvectors of \( F_e \). Figure 5.2 shows the first three principal components for WMAP 7, Planck and a CVL experiment, both before and after marginaliza-
tion, for the annihilation-like case \( \frac{dE}{dt} \propto p_{\text{ann}}(z)(1+z)^6 \) with 50 linearly-spaced redshift bins. We see that while the shapes of the PCs are qualitatively similar, marginalization produces noticeable changes to the PCs, as does changing from one set of experiment parameters to another. The differences become more pronounced for higher PCs.

Note that the shapes of the principal components can be affected by a number of other different factors: choice of binning, choice of ionization history calculator, energy deposition model, fiducial cosmological model considered, etc. We discuss these effects in Appendix A of [127].

In Figure 5.3 we show the first six marginalized PCs for \textit{Planck}, for annihilation-like \( \frac{dE}{dt} \propto p_{\text{ann}}(z)(1+z)^6 \) and decay-like \( \frac{dE}{dt} \propto p_{\text{dec}}(z)(1+z)^3 \) energy deposition histories. We show the annihilation-like case with both log and linear binning. We note that the first principal component is always largely or completely non-negative, and (in the annihilating case) peaked around redshift 600. The first PC can be thought of as a weighting function, describing the effect of energy deposition on the CMB (orthogonal to the effect of shifting the cosmological parameters), as a function of redshift\(^5\).

In Figure 5.4 we show the effect on the ionization history for the first three \textit{Planck} PCs in the annihilation case, with each PC multiplied by an energy deposition coefficient of \( \varepsilon = 2 \times 10^{-27} \text{ cm}^3/\text{s/GeV} \) to obtain \( p_{\text{ann}}(z) \). Note that this energy deposition is too large to be strictly in the linear regime; this figure illustrates the shape and size of the effect in the linear regime, the true effect for this value of \( \varepsilon \) will be somewhat smaller.

For energy injections that do not greatly change the optical depth, the fractional change

\(^5\)Note that the shift in the peak position between log and linear binning is to be expected, as one “weighting function” would be integrated over \( dz \) and the other over \( d\ln(1+z) \); see Appendix A of [127] for further discussion.
Figure 5.4: Fractional change to the ionization fraction $x_e$ in the presence of energy deposition, for the first three (marginalized) principal components in Planck. The curve shown is extrapolated from the linear (small energy deposition) regime, with normalization factor $\varepsilon_{1,2,3} = 2 \times 10^{-27}$ cm$^3$/s/GeV.

to the visibility function can be read off directly from the fractional change to the ionization history shown in Figure 5.4. Defining the visibility function as $g(z) = \tau' e^{-\tau}$, where $\tau'(z) \equiv d\tau/dz = n_e \sigma_T c/((1+z)H)$ is the probability of scattering per unit redshift, $\tau$ is the optical depth, $\sigma_T$ is the Thomson scattering cross-section and $n_e$ is the free electron density, the perturbation to the visibility function is given by

$$\frac{\Delta g(z)}{g(z)} = (e^{-\Delta \tau} - 1) + \frac{\Delta \tau'}{\tau'} e^{-\Delta \tau} \sim \frac{\Delta \tau'}{\tau'} = \frac{\Delta x_e}{x_e},$$

provided $\Delta \tau \ll 1$ and so $e^{-\Delta \tau} \sim 1$. We have explicitly checked that the effect of the PCs on the visibility function is almost identical to their effect on the ionization history.

As previously, we have considered “annihilation-like” and “decay-like” energy deposition histories separately. If both analyses were performed over the same redshift range, then while the principal components might appear different, they would span the same space of energy deposition histories. If all principal components were retained, the difference between
the two would simply be equivalent to a change of basis, and provided sufficient principal components are retained, this will still be approximately true. However, a particular energy deposition history may be described by the early principal components much better in one case than in the other; in particular, energy deposition histories for which the effect on the CMB is dominated by low redshifts will not be well described by the (first few of the) default annihilation-like PCs. Thus we present results for both cases.

5.3.2 Mapping into $\delta C_\ell$ space

Let us consider the mapping into $\delta C_\ell$ space of these marginalized principal components. Applying the transfer matrix $T$ (Equation 5.2) to the eigenvectors yields a set of $N$ vectors in the space of $C_\ell$ perturbations, $\delta C_\ell = Te_i = h_i$. The $h_i$'s should be understood as $\delta C_\ell$'s per energy deposition, and have units of $C_\ell/p(z)$.

We can define a dot product on the space of $\delta C_\ell$'s by

$$h_i \cdot h_j = \sum_\ell h_i^T \Sigma^{-1}_\ell h_{j\ell} = e_i^T F e_j$$  (5.8)
Figure 5.6: The $\perp$ components of the first three principal components for Planck, after marginalization, mapped into $\delta C_\ell$ space. The normalization is the same as for Figure 5.5.

Figure 5.7: Decomposition of $\delta C_\ell$ from energy deposition with constant $p_{\text{ann}}(z)$ into parallel (||) components which can be absorbed by changes in the cosmological parameters, and perpendicular ($\perp$) components that cannot be absorbed by such changes. The overall effect of the energy deposition is suppression of high-$\ell$ modes, due to the increased optical depth, and enhancement of low-$\ell$ polarization modes, as discussed in [225]. The suppression at high $\ell$ is clearly seen in the TT and EE spectra; the effect is also present in the TE spectra, with the peaks of $\delta C^{TE}_\ell$ occurring at the troughs of $C^{TE}_\ell$, and vice versa. The normalization here is $p_{\text{ann}} = 2 \times 10^{-27}$ cm$^3$/s/GeV, comparable to the latest limits from WMAP 7+ACT [140]. This decomposition depends on the sensitivity of the experiment; the case shown is WMAP 7 single band.
We then see that while the PCs are orthogonal, the $h_i$ are in general not orthogonal to each other, nor to the $\delta C_\ell$’s from the cosmological parameters. They correspond to actual energy deposition histories, and in general, there is no such history that is precisely orthogonal to all the cosmological parameters.

However, we may decompose the $h_i$ into components parallel and perpendicular to the space spanned by varying the cosmological parameters, and denote the perpendicular components $h_i^\perp$. The projection operator that implements this decomposition is given in Appendix B of [127]. The $h_i^\perp$ vectors are now orthogonal amongst themselves, as well as to the cosmological parameters, and their norms are given by the square root of the marginalized eigenvalues $\lambda_i$. It is these $h_i^\perp$’s which determine the detectability of the marginalized principal components, and which form an orthogonal basis for residuals which cannot be absorbed by varying the cosmological parameters. The addition of the parallel components, to recover the $h_i$’s from the $h_i^\perp$’s, ensures that the $h_i$’s correspond to energy deposition histories, and so provide an orthogonal basis in redshift space.

In Figure 5.5, we show the mapping of the first three (marginalized) PCs for Planck into the space of $\delta C_\ell$’s; in Figure 5.6, we show the components of these $\delta C_\ell$’s which are orthogonal to the space spanned by varying the cosmological parameters. Figure 5.7 demonstrates this projection for a sample DM annihilation model, summing over principal components, and decomposing the effect on the $C_\ell$’s into components perpendicular and parallel to the cosmological parameters.

The eigenvectors of the Fisher matrix $\{e_i\}$ thus provide an orthogonal basis in both relevant spaces, and their eigenvalues precisely describe the measurability of a “unit norm” energy deposition history with $z$-dependence given by the eigenvector. For an arbitrary
energy deposition history which we now write as

\[ p(z) = \sum_{i=1}^{N} \varepsilon_i e_i(z), \tag{5.9} \]

the expected \( \Delta \chi^2 \) relative to the null hypothesis of no energy deposition is \( \sum_i \varepsilon_i^2 \lambda_i \). If the \( \varepsilon_i \) coefficients are comparable, the relative sizes of the eigenvalues describe the fractional variance attributable to each principal component (eigenvector).

A brief comment on unit conventions: we take the \( \{e_i\} \) and \( \{G_i\} \) to be dimensionless, with the units of \( p(z) \) (cm\(^3\)/s/GeV) carried by the coefficients \( \alpha_i, \varepsilon_i \). The derivatives (and transfer matrix) then have units of \( C_\ell/p(z) \), and the Fisher matrix and its eigenvalues have units of \( 1/p(z)^2 \) (since the covariance matrix \( \Sigma \) has units of \( C_\ell^2 \)). Note also that due to the units of the covariance matrix, the dot product defined above takes two vectors in \( C_\ell \)-space to a dimensionless number (if the vectors have units of \( C_\ell \)).

### 5.4 Detectability

For a general energy deposition history, the PCs provide a basis in which, by construction, the basis vectors are ranked by the significance of their effect on the \( C_\ell \)'s. The measurability of a generic (smooth, non-negative) energy deposition history can thus be accurately described by the first few PCs\(^6\). Equivalently, the coefficients of later principal components have extremely large error bars, and will be challenging to measure or constrain.

We now outline the method for reconstructing and constraining the PC coefficients, or

\(^6\)It is in principle possible for the coefficients \( \varepsilon_i \) to be zero for \( i < n \) for some \( n \), but if \( n \) is large this implies a very unphysical energy deposition history that oscillates rapidly between positive and negative values. While “negative energy deposition” might perhaps have a physical interpretation in terms of increased absorption of free electrons, such an interpretation is not at all obvious, and so we focus on smooth, non-negative energy deposition histories.
any specific energy deposition history, using the PCA formalism. We investigate the number of PCs that can generically be measured at $\geq 1\sigma$ by Planck and a CVL experiment, for arbitrary energy deposition, and show results for broad classes of example models. We also consider the biases to the cosmological parameters that are induced if energy deposition is present but ignored; we present results for each principal component, so the biases due to an arbitrary energy deposition history can be immediately calculated. Our estimates of detectability and the biases will be verified using CosmoMC in the §5.6.

5.4.1 Estimating limits from the Fisher matrix

As mentioned previously, the perpendicular components of the $\delta C_\ell$'s, $h_i^\perp$, are orthogonal with norms $\sqrt{\lambda_i}$. They are also orthogonal to the space spanned by varying the cosmological parameters. Given these results and a measurement of the temperature and polarization anisotropies, it is straightforward to estimate general constraints on the energy deposition history from the Fisher matrix formalism. Note that in a careful study, one would instead use CosmoMC to perform a full likelihood analysis, using the Fisher matrix results only to determine the optimal principal components, as we demonstrate in §5.6. We outline this simple method only to help build intuition and to clarify later comparisons between the Fisher matrix method and the CosmoMC results.

The first step is to extract any residual between the data and the best-fit model using the standard cosmological parameters; let us denote this residual by $R_{i}^{TT,EE,TE}$. Then we take the dot product (as defined in Equation 5.8) of this residual with the $h_i^\perp$ vectors, normalizing by the corresponding eigenvalues (this normalization is required because the
$h_i^\perp$’s are orthogonal, but not orthonormal; see Appendix B of [127]):

$$\bar{\varepsilon}_i = \frac{R \cdot h_i^\perp}{\lambda_i}. \quad (5.10)$$

The resulting $\bar{\varepsilon}_i$ are the model-independent reconstructed coefficients for the marginalized principal components. In the absence of energy deposition, we expect them to be zero (within uncertainties).

The individual 1σ uncertainties on each of these coefficients are $1/\sqrt{\lambda_i}$, in the sense that if a single coefficient is perturbed away from its best-fit value by $1/\sqrt{\lambda_i}$, the corresponding energy deposition history will be disfavored at 1σ. Thus it is possible to set a very general model-independent constraint on each of the coefficients, $\varepsilon_i = \bar{\varepsilon}_i \pm \frac{1}{\sqrt{\lambda_i}}$ (at 1σ).

Given an arbitrary energy deposition history, we can decompose it into the principal components, each with its own coefficient, and compare these coefficients $\varepsilon_i$ to the bounds. For any particular model, a stronger constraint can be set by noting that,

$$\Delta \chi^2 = \sum_i \lambda_i (\varepsilon_i - \bar{\varepsilon}_i)^2. \quad (5.11)$$

This $\Delta \chi^2$ is relative to the best-fit model including both energy deposition and the standard cosmological parameters; the $\Delta \chi^2$ relative to the best-fit standard cosmological model$^7$ is simply $\sum_i \lambda_i \varepsilon_i (\varepsilon_i - 2\bar{\varepsilon}_i)$.

This method has the usual deficiencies of the Fisher matrix approach: it assumes a Gaussian likelihood and also linearity of the derivatives, and so can only be used for an estimate. In §5.6 we will go beyond the Fisher matrix approach and present constraints derived from a likelihood analysis using CosmoMC: in the same way as this estimate, those

$^7$Of course, if the best-fit energy deposition history is everywhere zero, i.e $\varepsilon_i \approx 0$ for all $i$, these two quantities are identical.
limits can be expressed as bounds on (a simple combination of) the PC coefficients, and will therefore be immediately applicable to a wide range of models for energy deposition.

5.4.2 Sensitivity of future experiments

For an energy deposition history where the sizes of the coefficients, $|\varepsilon_i|$, are all similar, the respective detectability of the PCs are given simply by their eigenvalues. Literally taking all the coefficients to be the same does not give a physical energy deposition history (since the later eigenvectors are highly oscillatory), but it is in some sense a "generic" scenario: none of the PCs have coefficients that are fine-tuned to be small, so slight changes to $p(z)$ or the basis of PCs are unlikely to drastically change the detectability of the different components.

We define detectability of the PCs with respect to this "generic" case; of course, detectability of any particular model depends on the relative sizes of coefficients. We consider a number of physical examples below to illustrate that, in some sense, the generic case is a reasonable average over a wide class of models of interest.

As discussed previously, [254] derived a set of energy deposition profiles corresponding to a range of DM annihilation models. These models provide a convenient set of example energy deposition histories, although they all have very similar effects on the CMB (see §5.5). We adapt the code developed in [254] and discussed in detail there to obtain similar physical $f(z)$ curves for the case of decaying dark matter with a long lifetime.

While the DM itself must have a lifetime considerably longer than the age of the universe, there could be other metastable species which decay during the redshift range we study ($z \sim 10 - 1300$), or excited states of the dark matter which decay to the ground state + Standard Model particles (e.g. [129, 41, 130, 93, 44] and references therein). In this case
Figure 5.8: (a) The sensitivity for Planck (single-band), after marginalization, for various models subject to constraints from WMAP 7 single-band at 2σ. The models considered are shown in Figure 5.8. The left figure assumes annihilation-like energy deposition and the right figure assumes decay-like energy deposition. The top panels show: (1) assuming $p(z) \propto e_i(z)$ for each PC, (2) the generic case where all PC coefficients have equal magnitudes $|\varepsilon_i| = \varepsilon = 2/\sqrt{\sum_i \lambda_i^{WMAP}}$, (3) constant $p(z)$, and (4) taking $p(z) \propto f(z)$, with $f(z)$ from the models in [254]. For the left figure, the hatched region indicates the range of results from modifying the analysis slightly, as described in Appendix A of [127]. The bottom panels show some sample $z_\tau$ models for asymmetric annihilating dark matter (left) and decaying species (right), as discussed in §5.4.2 (the labels describe the initial particle mass, and the SM final state for annihilation or decay), and an extreme case where $p(z) = 0$ for $200 < z < 900$ and constant outside that range.
the decay rate would cut off exponentially for $z < z(\tau)$, although heating and ionization of the gas could continue for some time after that: we can again obtain detailed $p(z)$ curves for different decay lifetimes using the methods of [254]. Models of this type provide a simple class of examples suitable for use with the PCs derived for the case of decaying DM, since the underlying $dE/dt \propto (1 + z)^3$ redshift dependence is the same (although for models with lifetimes short enough that the energy deposition has ceased shortly after recombination, the PCs derived for the annihilation-like case may work better).

For the annihilating case, asymmetric dark matter scenarios can furnish a similar set of examples [165, 95, 183, 111, 94, 114]. In such scenarios the DM sector possesses an asymmetry analogous to that in the baryon sector, and it is this asymmetry which sets the DM relic density rather than the annihilation cross section. In the minimal case there is thus no requirement for an annihilation signal in the present day or during the epoch of recombination, but it is nonetheless possible to have a large late-time annihilation signal, by repopulation of the depleted component at late times, or by oscillations from the more-abundant to the less-abundant component [95, 68, 114]. As a simple example, we consider models where another species decays to repopulate the less-abundant DM state [114], thus...
causing the annihilation to “switch on” as \(1 - e^{-t/\tau}\) at a characteristic timescale \(\tau\) (with \(z_\tau\) being the corresponding redshift). We compute the \(p(z)\) curves for a range of \(\tau\). Finally, for both annihilation and decay we consider the constant \(p(z)\) case, studied in \([78]\) (for decay) and \([225, 139]\) (for annihilation), to facilitate comparison with the literature.

Figure 5.8 shows the detectability of the principal components in Planck and the ideal CVL experiment for these annihilating and decaying models, with the energy deposition normalized to lie at the 95% limit from WMAP 7. The energy deposition for these models is shown in Figure 5.9. In the “generic” case, we set the sizes of the coefficients of the Planck (or CVL) PCs to be \(|\varepsilon_i| = \varepsilon = 2/\sqrt{\sum_i \lambda_i^{WMAP}}\). The actual WMAP 7 signal-to-noise for the model is

\[
\frac{S}{N} = \left( \sum_i \lambda_i^{WMAP} \left[ \sum_j \varepsilon_j \epsilon_i^{Planck} \cdot \epsilon_j^{WMAP} \right]^2 \right)^{1/2}
\]

and thus depends on the signs of \(\varepsilon_i\), but the generic case is meant to indicate the typical detectability for a class of models, so we instead use the WMAP 7 constraints to set an overall scale for \(|\varepsilon_i|\).

We also show the detectability for each PC if \(p(z) \propto \varepsilon_i(z)\), or assuming the energy deposition history has zero overlap with all other PCs\(^8\). As mentioned previously, this is not a physical assumption (requiring an “energy deposition” oscillating rapidly between positive and negative values): in such a case the effect on the \(C_i\)’s is so small that the normalization of the “energy deposition” could be very large and still consistent with WMAP.

\(^8\)If the PCs were the same for the different experiments, this would give an upper bound on the detectability of the \(i\)th PC, given WMAP 7 2\(\sigma\) constraints. However, the PCs for different experiments are not orthogonal, \(\epsilon_i^{Planck} \cdot \epsilon_j^{WMAP} \neq \delta_{ij}\). A strict upper bound for the \(S/N\) of the \(i\)th Planck PC is given by \((S/N)_i^{Planck} \leq 2\sqrt{\lambda_i^{Planck}} \sum_j |\epsilon_i^{Planck} \cdot \epsilon_j^{WMAP} / \sqrt{\lambda_j^{WMAP}}|\), with the analogous result for a CVL experiment. However, this quantity is not very useful as an upper bound; for example, if \(p(z)\) is proportional to a high WMAP PC, the normalization of \(p(z)\) is essentially unconstrained, but the detectability for Planck may be very significant if there is even a small overlap with the first Planck PC.
Consequently, arbitrarily high PCs can be measured if they are the sole contributors to the energy deposition history.

We see that models with decay-like redshift dependence and those with annihilation-like redshift dependence tend to have roughly the same number of measurable parameters. In both cases, generally 2-3 components are potentially measurable in Planck and up to 5-7 for a CVL experiment.

As a side note, the improvement of these constraints between WMAP 7 and future experiments is in large part due to (anticipated) better measurements of the polarization. In the absence of polarization data (i.e. using the TT spectrum only), we would expect the constraints to weaken by a factor of \( \sim 3 \) for WMAP 7, \( \sim 7 \) for Planck, and \( \sim 14 \) for a CVL experiment. Here we have taken the square root of the eigenvalue of the first principal component as a proxy for sensitivity, which will be approximately true for models with a non-negligible overlap with the first PC.

### 5.4.3 Biases to the cosmological parameters

If energy deposition is present but neglected, it can bias the measurement of the cosmological parameters by a significant amount. For WMAP, the partial degeneracy between varying \( n_s \) and the effects of energy deposition means that the dominant bias is a 1\( \sigma \) negative shift to \( n_s \). The improved polarization sensitivity of Planck largely lifts the degeneracy with \( n_s \), but due to the smaller error bars of Planck other parameters develop non-negligible biases: at the maximum energy deposition allowed by WMAP 7 at 2\( \sigma \), Planck parameter estimates are generically biased at \( > 1\sigma \) for \( \omega_c, H_0, \) and \( A_s \).

Calculation of the biases is exactly complementary to calculating the marginalized Fisher
Figure 5.10: For the $i$th PC, the contribution to the bias to cosmological parameters in WMAP 7 (left panel) and Planck (right panel), relative to the error bars forecast from the Fisher matrix. The normalization is that of the “generic” case (see discussion in §5.4.2 or Figure 5.8), where each PC coefficient has the same absolute value and the overall normalization is the maximum allowed by WMAP 7 at 2σ. The total bias for the parameter $\theta$ is $\sum_i \delta \theta_i$.

Thus we can partition the biases into the bias per PC, which is shown in Figure 5.10 for WMAP 7 and Planck. For a generic energy deposition history, the total bias is dominated by the bias from the first few PCs, consistent with the fact that later PCs are undetectable and can essentially be neglected in any fit to the data. As expected from [138], the largest bias for WMAP 7 is to $n_s$. 
5.5 A universal $p_{\text{ann}}(z)$ for WIMP annihilation

Solutions for the redshift dependence of the efficiency function $f(z)$ (and hence the energy deposition history $p_{\text{ann}}(z)$), for 41 different combinations of dark matter mass and annihilation channel, were presented in [254]. We can use these 41 energy deposition histories, rather than $\delta$-functions in $z$, as the input states for a principal component analysis. Here we summarize the results presented in §V of [127].

After marginalization over the other cosmological parameters, we find that in this case the first eigenvalue completely dominates the later ones, accounting for 99.97% of the total variance in WMAP 7, Planck and the CVL forecast. Thus, to a very good approximation, for any of the DM models studied in [254] (or any linear combination of the final states studied there), the effect on the $C_\ell$'s is determined entirely by the dot product of $p_{\text{ann}}(z)$ with the first PC, with the $\ell$-dependence given by mapping the first PC to $C_\ell$-space. This conclusion agrees with the statements in [254, 170] that the effect of DM annihilation can be captured by a single parameter. The effective $f$-value of various WIMP annihilation models is just given by the dot product of their $f(z)$ curves with this first principal component. On our website, we provide the “universal” $f(z)$ curve for WIMP annihilation, as well as effective $f$-values for all models considered in [254].

5.6 CosmoMC Results

Everything we have done so far assumes both linearity and that the Fisher matrix is an adequate description of the likelihood function. §VI of [127] presents our results of a full likelihood analysis using the CosmoMC Markov chain Monte Carlo code, in particular
examining the biases to the cosmological parameters and the detectability of the PCs. To summarize, there we determine the constraints on parameters using the seven-year WMAP data [191] (temperature and polarization), for both general energy injection histories and for the universal $p_{\text{ann}}(z)$ for WIMP annihilation. We then generate simulated data for Planck and a CVL experiment using a fiducial cosmological model given by the best fit WMAP7 model, and show the constraints on the various parameters. In the Planck case, for example, we have confirmed that the constraints on the first three principal components can be used to recover the correct limit on a particular energy deposition history.

5.7 Conclusion

Principal component analysis provides a simple and effective parameterization for the effect of arbitrary energy deposition histories on anisotropies in the cosmic microwave background. We find that for DM annihilation-like energy deposition histories the first principal component, describing the bulk of the effect, is peaked around $z \sim 500 - 600$, at somewhat lower redshift than previously expected; the later principal components provide corrections to this basic weighting function.

The principal components, derived from a Fisher matrix approach, are stable against a wide variety of perturbations to the analysis, including choice of code calculating the ionization history, additions to the usual set of cosmological parameters, the inclusion or exclusion of ionization on helium, the range of included multipoles, and the choice of binning. The one significant potential change to the PCs arises from how deposited energy is attributed to additional Lyman-α photons: we have showed the effect of on one hand neglecting this channel, and on the other of assuming that all the energy attributed to “excitations” is con-
verted into Lyman-$\alpha$, which should bracket the true result. We eagerly await a more careful analysis of this problem.

Within the Fisher matrix formalism, it is straightforward to take into account degeneracies with the standard cosmological parameters. We have presented predictions for the (significant!) biases that would arise in Planck as a result of falsely assuming energy deposition to be zero, for each of the principal components. We have confirmed the previously noted degeneracy between energy deposition and $n_s$, and to a lesser degree with $A_s$, $\omega_b$ and $\omega_c$, in WMAP data; since our analysis decomposes the biases according to the principal components that generate them, it is now trivial to compute the biases to the cosmological parameters for any arbitrary energy deposition history, in WMAP or in mock Planck data.

For a wide range of energy deposition histories, spanning models of dark matter annihilation and decaying species where annihilation or decay can begin or end abruptly on characteristic timescales shorter than the age of the universe, the coefficients of up to three principal components are potentially measurable by Planck, for energy deposition histories satisfying 95% confidence limits from WMAP, opening up the exciting possibility of distinguishing different models of energy deposition. For a CVL experiment, up to five coefficients are measurable.

For the “standard” WIMP annihilation case, principal component analysis on a large set of WIMP models yields a single principal component $e_{WIMP}(z)$ that describes the effect on the $C_\ell$’s of all the models very well; any model is then parameterized simply by the coefficient of $e_{WIMP}(z)$ (or equivalently, effective $f$). Our analysis confirms previous statements in the literature, and we have provided this “universal $f(z)$” curve for future WIMP annihilation studies.
We performed an accurate MCMC analysis of current WMAP 7 data to impose constraints on the measurable principal component amplitudes, and to forecast constraints for future experiments such as Planck or a CVL experiment. We find good agreement with the Fisher matrix analysis, although the MCMC analysis is required to accurately predict the biases on the cosmological parameters. We have illustrated how it is possible to recover the constraints on an arbitrary energy deposition history from the constraints on the amplitudes of the principal components. The reconstructed constraints are in very good agreement with the constraints obtained by directly sampling a specific energy deposition history, confirming the validity and usefulness of the principal component decomposition.
Chapter 6

Asymmetric and Symmetric Light Dark Matter

6.1 Introduction

Studies of dark matter (DM) have historically focused on particles with weak scale mass \( \sim 100 \text{ GeV} \) [54, 180, 119]. The reason is not only the focus of the high energy physics community on weak scale phenomena, but also because the annihilation cross section for a Weakly Interacting Massive Particle (WIMP) naturally gives rise to the observed cold DM relic abundance. This is the so-called “WIMP miracle.”

More recently there has been a broader interest in light DM, with mass \( m_X \lesssim 10 \text{ GeV} \). Part of the reason for this interest is phenomenological. Direct detection results from DAMA [53], CoGeNT [2, 1], and CRESST [27] claim event excesses that can be interpreted as nuclear scattering of DM with mass \( \sim 10 \text{ GeV} \) (although the mutual consistency of these results is disputed). Meanwhile dark matter with masses of MeV has been studied as a
possible explanation of the INTEGRAL 511 keV line [58, 59, 164, 168, 231, 166, 107].

There is also a theoretical motivation for light DM, as DM with mass $m_X \lesssim 10$ GeV appears in certain classes of models naturally. In supersymmetric hidden sector models, for example, gauge interactions generate light DM masses and give rise to the correct annihilation cross section [166, 123, 125]. The asymmetric DM (ADM) scenario, where the DM particle $X$ carries a chemical potential, analogous to the baryons, provides another approach to light DM (see e.g. [223, 181, 40, 183] and references therein). In these scenarios, both DM ($X$) and anti-DM ($\bar{X}$) particles may populate the thermal bath in the early Universe; however, the present number density is determined not only by the annihilation cross section, but also by the DM number asymmetry $\eta_X$. Depending on the value for $\eta_X$, the DM mass can be as low as $\sim$ keV in ADM models [114], though the natural scale for ADM is set by $(\Omega_{\text{CDM}}/\Omega_b)m_p \approx 5$ GeV.

The purpose of this work is to explore model-independent constraints and predictions for the asymmetric and symmetric limits of light DM with mass $\sim 1$ MeV–10 GeV.\textsuperscript{1} Although both phenomenological and theoretical considerations have motivated the study of light DM candidates, there are still a number of important constraints that should be taken into account in realistic model building. In general, light thermal DM faces two challenges: one is to evade bounds on energy injection around redshifts $z \sim 100 – 1000$ coming from observations of the CMB; the other is to achieve the required annihilation cross section without conflicting with collider physics constraints.

CMB data from WMAP7 strongly limits DM annihilation during the epoch of recom-

\textsuperscript{1}For DM much lighter than $\sim 1$ MeV, DM can only annihilate to neutrinos, new light states that remain relativistic through matter-radiation equality, or hidden sector forces that decay invisibly. In this case, the CMB and collider bounds discussed here do not apply.
bination, and excludes symmetric thermal light DM with mass below $\sim 1 - 10$ GeV if the annihilation is through $s$-wave processes [140, 170, 127]. The CMB bounds may be evaded in the symmetric case if DM dominantly annihilates to neutrinos or if its annihilation is $p$-wave suppressed. When the DM relic density is asymmetric, DM annihilation during recombination can be highly suppressed if the symmetric component is sufficiently depleted, providing a natural way to resolve the tension from CMB constraints for light DM scenarios. Unlike the case of symmetric DM, the CMB places a lower bound on the annihilation cross section for ADM from the requirement of sufficient depletion of the symmetric component. We calculate the minimum annihilation cross section required in order to evade the CMB bound and achieve the correct relic density simultaneously.

However, it is difficult to achieve the needed annihilation rate to Standard Model (SM) particles through a weak-scale mediator. Null results from mono-jet plus missing energy searches at the Tevatron [150, 149, 36] and the LHC [234, 135] strongly constrain such a mediator if DM couples to quarks and gluons. Meanwhile, the mono-photon plus missing energy search at LEP sets limits on the coupling between DM and charged leptons [134] via such a heavy state. These collider constraints are so strong that the annihilation through an off-shell heavy mediator is generally insufficient for ADM to achieve the correct relic density and evade the CMB constraint, if the DM mass is below a few GeV. One way to evade the collider constraints is to invoke a light mediator with mass much less than $\sim 100$ GeV. In this case, DM can annihilate to SM states efficiently via the light state without conflicting with collider bounds. Furthermore, if the mediator is lighter than the DM, a new annihilation channel opens and DM can annihilate dominantly to the mediator directly. In this limit, the mediator particle may couple to the SM sector rather weakly.
The presence of the light mediator has various implications for DM dynamics in galaxies and for cosmology. The light mediator may give rise to significant DM self-interactions (i.e., DM-DM scattering); this is true in both the symmetric and asymmetric limits, since the light state mediates DM-DM interactions as well as anti-DM and DM interactions. These interactions leave footprints in the DM halo dynamics. There are limits on the DM self-interaction cross section coming from observations of elliptical DM halos and elliptical galaxy clusters. We combine these with the relic density constraint to place a lower bound on the mediator mass \( \sim 4 \times 10^{-2} \text{ MeV} - 40 \text{ MeV} \) for DM masses in the range \( \sim 1 \text{ MeV} - 10 \text{ GeV} \). We assume this massive mediator decays to SM relativistic degrees of freedom in the early universe to avoid the overclosure problem, and derive conditions for thermalization of the DM and SM sectors.

These astrophysical and cosmological constraints can be applied to the parameter space of scattering rates in direct detection experiments. We consider DM-nucleon scattering for DM masses of \( 1 - 10 \text{ GeV} \) and DM-electron scattering for DM masses \( 1 \text{ MeV} - 1 \text{ GeV} \). In the case of electron scattering, we combine the astrophysical and cosmological constraints with bounds from beam dump experiments and supernova cooling, which exclude a large region of the allowed parameter space. In addition, the predictions are very different dependent on whether the mediator is heavier or lighter than the DM.

The rest of the chapter is organized as follows. In Section 6.2, we present the relic density calculation for DM in the presence of a chemical potential. In Section 6.3, we study the CMB constraint on ADM models and derive the annihilation cross section required to evade the CMB bound. In Section 6.4, we examine current collider physics constraints on the DM annihilation cross section. In Section 6.5, we study the elliptical halo shape constraint on
the mediator mass. In Section 6.6, we map out the parameter space for DM direct detection. We conclude in Section 6.7.

6.2 Relic Density for Symmetric and Asymmetric Dark Matter

Our starting point is to establish that the correct relic density of \( \Omega_{\text{CDM}} h^2 = 0.1109 \pm 0.0056 \) [195] can be obtained, where we assume that the annihilation cross section \( \langle \sigma v \rangle \) and the asymmetry \( \eta_X \) are floating parameters.

In the usual thermal WIMP scenario, the correct relic density is determined by DM annihilation until freeze-out. For Dirac DM in the symmetric limit, the cold DM relic density is \( \Omega_{\text{CDM}} h^2 \sim 0.11 \left( 6 \times 10^{-26} \text{cm}^3/\text{s} \right) / \langle \sigma v \rangle \). DM may also carry a chemical potential which leads to an asymmetry between the number density of DM and anti-DM. In this case, when the DM sector is thermalized, the present relic density is determined both by the annihilation cross section and the primordial DM asymmetry \( \eta_X \equiv (n_X - n_{\bar{X}})/s \), where \( n_X, n_{\bar{X}} \) are the DM and anti-DM number densities and \( s \) is the entropy density. In the asymmetric limit, neglecting any washout or dilution effects, the correct relic density is obtained for a primordial asymmetry given by

\[
\eta_X \approx \frac{\Omega_{\text{CDM}} \rho_c}{m_X s_0},
\]  

(6.1)

where \( s_0 \approx 2969.5 \text{ \ cm}^{-3} \) and \( \rho_c \approx 1.0540 h^2 \times 10^4 \text{ \ eV/cm}^3 \) are the entropy density and critical density today. In the asymmetric limit, the annihilation cross section is sufficiently large that the thermally-populated symmetric component is a sub-dominant component of the energy density today.
Depending on the strength of indirect constraints on DM annihilation, light DM scenarios must interpolate between the symmetric and asymmetric limits. We thus require precise calculations of the present anti-DM to DM ratio 
\[ r_\infty = \frac{\Omega_{\bar{X}}}{\Omega_X}, \]
which controls the size of indirect signals from DM annihilation. Note that \( r_\infty \) is related to the absolute relic densities by
\[ \Omega_X = \frac{1}{1 - r_\infty} \frac{\eta_X m_X s_0}{\rho_c}, \quad \Omega_{\bar{X}} = \frac{r_\infty}{1 - r_\infty} \frac{\eta_X m_X s_0}{\rho_c}, \]
and the total CDM relic density is \( \Omega_{\text{CDM}} = \Omega_X + \Omega_{\bar{X}} \).

To compute \( r_\infty \) we solve the Boltzmann equations for \( n_X, n_{\bar{X}} \) freezeout in the presence of a nonzero chemical potential [243]. In this work, we focus on the case where DM is in thermal equilibrium with the photon thermal bath through freezeout. In general, this assumption may not hold if there is a weakly coupled light mediator coupling DM to the SM. We leave the more general case for future work [200], noting that the effects on the relic density are up to \( \mathcal{O}(10) \), depending on \( m_X \).

The coupled Boltzmann equations for the species \( n_+ = n_X \) and \( n_- = n_{\bar{X}} \) are
\[ \frac{dn_\pm}{dt} = -3H n_\pm - \langle \sigma v \rangle [n_+ n_- - n^{eq}_+ n^{eq}_-] \]
where \( \langle \sigma v \rangle \) is the thermally-averaged annihilation cross section over the \( X \) and \( \bar{X} \) phase space distributions [146]. The Hubble expansion rate is \( H \approx 1.66\sqrt{g_{\text{eff}}}T^2/M_{\text{pl}} \) where \( M_{\text{pl}} \approx 1.22 \times 10^{19} \) GeV is the Planck mass and \( g_{\text{eff}} \) is the effective number of degrees of freedom for the energy density. If there is a primordial asymmetry in \( X \) number, then there is a nonzero chemical potential \( \mu \) which appears in the equilibrium distributions as \( n^{eq}_\pm = e^{\pm \mu/T}n^{eq} \). Here \( n^{eq} \) is the usual equilibrium distribution with \( \mu = 0 \), and thus \( n^{eq}_+ n^{eq}_- = (n^{eq})^2 \).

As usual, we define \( x = m_X/T \) and \( Y_\pm = n_\pm/s \), where \( s = (2\pi^2/45)h_{\text{eff}}(T)T^3 \) is the entropy density and \( h_{\text{eff}}(T) \) is the effective number of degrees of freedom for the entropy.
density. We write the annihilation cross section as \( \langle \sigma v \rangle = \sigma_0 x^{-n} \), with \( n = 0 \) and \( n = 1 \) for \( s\)-wave and \( p\)-wave annihilation processes respectively. Then simplifying Eq. (6.3) gives

\[
\frac{dY_{\pm}}{dx} = -\frac{\lambda}{x^{n+2}} \sqrt{g_*} \left( Y_{\pm} - (Y_{eq})^2 \right),
\]

(6.4)

where \( \lambda \equiv 0.264 M_{pl} m_X \sigma_0 \) and \( Y_{eq} \approx 0.145 (g/h_{eff})^{3/2} e^{-x} \equiv ax^{3/2} e^{-x} \). The effective number of degrees of freedom is \( \sqrt{g_*} = h_{eff} \sqrt{g_{eff}} \left( 1 + \frac{T}{3h_{eff}} \frac{dh_{eff}(T)}{dT} \right) \) [146].

After being generated at some high temperature, the DM asymmetry is a conserved quantity, so we have the constraint

\[
\eta_X = Y_+ - Y_-
\]

(6.5)

which is constant at any given epoch.\(^2\) In order to impose this condition on our numerical solutions, we define the departure from equilibrium \( \Delta \) by \( Y_{\pm} = Y_{eq}^{\pm} + \Delta \), and instead solve the (single) equation for \( \Delta \).

It is helpful to present approximate analytic solutions in the limit of constant \( \sqrt{g_*} \) [243, 152, 175]. Eq. (6.4) can be solved analytically at late times when \( (Y_{eq})^2 \) becomes negligible. In this limit, using Eq. (6.5), we can integrate Eq. (6.4) separately for \( \bar{X} \) and \( X \) to obtain

\[
Y_{\pm}(\infty) \simeq \frac{\pm \eta_X}{1 - [1 \mp \eta_X/(Y_{\pm}(x_f))] e^{\pm \eta_X \lambda \sqrt{g_*} x_f^{-n-1}/(n+1)}}.
\]

(6.6)

These solutions also apply for the symmetric case in the limit of \( \eta_X \rightarrow 0 \). We take the freezeout temperature \( x_f = m_X/T_f \) as derived in [152]:

\[
x_f \simeq \ln [(n+1) \sqrt{g_*} a \lambda] + \frac{1}{2} \ln \frac{\ln^2 [(n+1) \sqrt{g_*} a \lambda]}{[(n+1) \sqrt{g_*} a \lambda] - (\sqrt{g_*})^2 [(n+1) \lambda \eta_X/2]^2}. \]

(6.7)

\(^2\)We assume there is no Majorana mass term for DM, and thus \( X - \bar{X} \) oscillation [114, 95, 92, 64] does not occur. We also assume there is no entropy production in this case and there are no DM-number violating interactions at these temperatures.
Using $Y_{\pm}(\infty)$ given in Eq. (6.6), we can obtain the present ratio of the $\bar{X}$ to $X$ number densities:

$$r_{\infty} \equiv \frac{Y_-}{Y_+}(\infty) \simeq \frac{Y_-(x_f)}{Y_+(x_f)} \exp \left( \frac{-\eta_X \lambda \sqrt{g_*}}{x_f^{n+1}(n+1)} \right).$$

(6.8)

While we can obtain a precise analytic result for $r(x_f) = Y_-(x_f)/Y_+(x_f)$, it turns out that the consequence of neglecting the $(Y^{eq})^2$ in the late-time solution can almost exactly be accounted for by simply setting $r(x_f) = 1$. This gives numerically accurate answers over a wide range of $\eta_X$ and $\langle \sigma v \rangle$ as discussed in [152]. Note that the solution here only converges when $\eta_X \lambda$ is small enough $\sqrt{g_* \eta_X \lambda} < 2x_f^{n+2}$.

### 6.3 CMB Constraints

For both symmetric and asymmetric thermal DM, the DM particles must have a sufficiently large annihilation cross section in order to achieve the correct relic density. This annihilation may have many indirect astrophysical signatures; among these, the most robust prediction (or constraint) is the effect of DM annihilation on the cosmic microwave background (CMB) [225], since the effect only depends on the average DM energy density. We first summarize recent studies of CMB constraints on DM annihilation, and then discuss scenarios which naturally evade these constraints for light DM, focusing on the asymmetric DM scenario.

Energy deposition from DM annihilation distorts the surface of last scattering, which affects the CMB anisotropies and is thus constrained by WMAP7 data. CMB constraints become increasingly severe for smaller DM masses: the energy released in DM annihilations scales as $\sim m_X(n_X)^2 \sim \rho_{\text{CDM}}^2/m_X$, where $\rho_{\text{CDM}}$ is the average energy density in DM. This implies the effect of DM annihilation on the CMB scales as $\sim \langle \sigma v \rangle/m_X$. Though the
precise bound depends on the mass and annihilation channels, WMAP7 limits the amount of annihilation during recombination to below the thermal relic annihilation cross section if $m_X \lesssim 1 - 10$ GeV [254, 139, 140, 170]. Furthermore, Planck data can improve these constraints by up to a factor of 10.

For self-annihilating DM particles such as Majorana fermions or real scalars, the energy deposition rate per volume at redshift $z$ is

$$\frac{dE}{dt dV}(z) = \rho_c^2 \Omega_{CDM}^2 (1 + z)^6 f(z) \frac{\langle \sigma v \rangle_{\text{CMB}}}{m_X},$$

(6.9)

where $\rho_c$ is the critical density at the present time, $\langle \sigma v \rangle_{\text{CMB}}$ is the thermally-averaged annihilation cross section at the epoch of recombination, and $f(z)$ parametrizes the amount of energy absorbed by the photon-baryon fluid at redshift $z$, relative to the total energy released by DM annihilation at that redshift.

The quantity $f(z)$ gives the efficiency of energy deposition at redshift $z$ and thus depends on the spectrum of photons, neutrinos and $e^\pm$ resulting from DM annihilation. In general, the dependence of $f(z)$ on $z$ is mild [254], and an excellent approximation is to take $f(z) \equiv f e_{\text{WIMP}}(z)$ where $f$ is a constant and $e_{\text{WIMP}}(z)$ is a universal function for WIMP DM [127]. In addition, to leading order $f \simeq (1 - f_\nu)$ [170], where $f_\nu$ is the fraction of energy going to neutrinos per annihilation. For DM annihilation channels to charged lepton or pion final states, $f \approx 0.2 - 1$; here annihilation only to $e^\pm$ can give $f \sim 1$.

There is also some mild $m_X$ dependence in $f(z)$ (or $f$), since the spectrum of DM annihilation products depends on $m_X$. Ref. [254] computed detailed efficiency curves $f(z)$ for $m_X > 1 - 10$ GeV, depending on the channel. However, the observed trend is that efficiency does not depend strongly on mass in the range 1-1000 GeV, and furthermore increases for
lower mass.\textsuperscript{3} We will extrapolate results to $m_X < 1$ GeV; we expect this is a conservative approach.

The WMAP7 limit on DM energy injection at the 95% C.L. can be written as [140]

$$f \frac{\langle \sigma v \rangle_{\text{CMB}}}{m_X} < \frac{2.42 \times 10^{-27}}{\text{cm}^3/\text{s}} \frac{\text{GeV}}{m_X}. \quad (6.10)$$

This bound\textsuperscript{4} as given assumes DM particles are self-annihilating, \textit{i.e.} Majorana fermions or real scalars. For DM candidates that are Dirac fermions or complex scalars, as in ADM scenarios, the energy injection rate is

$$\frac{dE}{dtdV}(z) = 2 \rho_c^2 \Omega_{\text{CDM}}^2 \frac{r_\infty}{(1 + r_\infty)^2} (1 + z)^6 f(z) \frac{\langle \sigma v \rangle_{\text{CMB}}}{m_X}, \quad (6.11)$$

where we have used $\rho_X + \rho_{\bar{X}} = \rho_{\text{CDM}}$ and $r_\infty = \rho_{\bar{X}}/\rho_X$. Note there is factor of 2 in the energy injection rate relative to the self-annihilating case, accounting for the number of possible annihilations. Comparing Eq. (6.9) and Eq. (6.11), we can translate the bound given in Eq. (6.10) to the Dirac fermion or complex scalar case:

$$\frac{2 r_\infty}{(1 + r_\infty)^2} f \frac{\langle \sigma v \rangle_{\text{CMB}}}{m_X} < \frac{2.42 \times 10^{-27}}{\text{cm}^3/\text{s}} \frac{\text{GeV}}{m_X}. \quad (6.12)$$

We show this constraint for various $r_\infty$ values in Fig. (6.1); the dotted black line gives the thermal relic annihilation cross section in the symmetric case, where we have solved for the relic density numerically and taken $f = 1$.

\textsuperscript{3}Above $m_X, m_\phi > 1$ MeV, most of the annihilation products rapidly cascade down to lower energies and the efficiency $f$ is only mildly sensitive to the initial energy spectrum of annihilation products (normalizing for the total energy). However, photons in the range $\sim 0.1-1$ GeV deposit their energy relatively inefficiently. For annihilation of sub-GeV scale DM, typically a smaller fraction of the total energy goes into photons of these energies, which increases the total efficiency slightly. We thank Tracy Slatyer for this point.

\textsuperscript{4}Note: the results of [170] are slightly weaker by a factor of 1.2-2.
Figure 6.1: WMAP7 95% C.L. constraints on the DM annihilation cross section and mass for asymmetric dark matter and $s$-wave annihilation. We show constraints for various values of $r = r_\infty = \Omega_{\bar{X}}/\Omega_X$, the anti-DM to DM ratio at the present time. The shaded region (blue) is excluded by the WMAP7 data, with different shades corresponding to different $r_\infty$. Along the horizontal contours of constant $r$ are the values of $\langle \sigma v \rangle$ where the correct relic density can be obtained for an efficiency factor $f = 1$. The turnover around $m_X \sim 10$ GeV comes from the drop in SM degrees of freedom when the universe has temperature $\sim 1$ GeV. The solid red line is the intersection of the WMAP7 and relic density contours: it indicates the minimum $\langle \sigma v \rangle$ needed to obtain the observed relic density and satisfy CMB constraints for $s$-wave annihilation.
ADM can evade CMB bounds while still allowing s-wave annihilation.\footnote{In the symmetric limit, one can evade the CMB bounds if DM annihilates via p-wave suppressed interactions. Then $\langle \sigma v \rangle_{\text{CMB}} \simeq (v_{\text{CMB}}/v_f)^2 \langle \sigma v \rangle_f$ and since $v_{\text{CMB}} \sim 10^{-8}$ while $v_f \sim 0.3$, the annihilation cross section at recombination is highly suppressed and WMAP constraints are substantially weakened. An increased branching ratio to neutrinos (smaller $f$) can also alleviate the tension with CMB data for light DM.} The CMB bounds do not completely disappear in the ADM scenario, however, because there is a small symmetric component of DM remaining, $r_\infty$, the size of which depends on $\langle \sigma v \rangle$. Because of the exponential dependence of $r_\infty$ on $\langle \sigma v \rangle$, as shown in Eq. (6.8), the CMB constraints lead to a lower bound on $\langle \sigma v \rangle$. This is shown in Fig. (6.1), where we map out the constraints in the

Figure 6.2: \textit{(Top)} Minimum $\langle \sigma v \rangle$ for efficient annihilation of the symmetric component in an ADM scenario, such that CMB bounds can be evaded, for two different values of the efficiency $f$. The black dotted line gives the thermal relic $\langle \sigma v \rangle$ for the symmetric case. \textit{(Bottom)} The corresponding maximum allowed $r_\infty$, the anti-DM to DM ratio at the present time.
Asymmetric and Symmetric Light Dark Matter

\(\langle \sigma v \rangle_{\text{CMB}} \) and \(m_X\) parameter space, computing the relic density numerically and applying the constraint in Eq. (6.10). The solid line (red) gives the resulting lower bound on \(f \langle \sigma v \rangle_{\text{CMB}}\). This lower bound on \(f \langle \sigma v \rangle_{\text{CMB}}\) translates to an upper bound on the residual symmetric component, \(r_{\infty}\), as shown in Fig. (6.2). We give analytic approximations to these numerical solutions next.

When \(r_{\infty} \ll 1\), we can ignore the \(\bar{X}\) contribution to the total relic density, and the DM asymmetry parameter \(\eta_X\) is set by \(\eta_X \approx \Omega_{\text{CDM}}\rho_c/(m_X s_0)\). For a given \(\eta_X\), the required annihilation cross section at freezeout to achieve a particular residual symmetric component, \(r_{\infty}\), can be obtained by rewriting Eq. (6.8) as

\[
\langle \sigma v \rangle_f \approx \frac{s_0 x_f}{0.264 \Omega_{\text{CDM}} \rho_c \sqrt{g_{*f}} M_{\text{pl}}} \ln \left( \frac{1}{r_{\infty}} \right)
\]

\[
\simeq c_f \times 5 \times 10^{-26} \text{ cm}^3/\text{s} \times \ln \left( \frac{1}{r_{\infty}} \right),
\]

where \(c_f \equiv \left( \frac{x_f}{20} \right) \left( \frac{4}{\sqrt{g_{*f}}} \right)\) is an \(O(1)\) factor. We show the numerical result as the horizontal contours of constant \(r_{\infty}\) in Fig. (6.1); for \(m_X < 1\) GeV we obtain a good approximation to the numerical solution by taking \(c_f = 1\). On the other hand, the CMB bound on the annihilation cross section when \(r_{\infty} \ll 1\) is

\[
\langle \sigma v \rangle_{\text{CMB}} < \frac{2.42 \times 10^{-27}}{2f} \text{ cm}^3/\text{s} \times \left( \frac{m_X}{1 \text{ GeV}} \right) \left( \frac{1}{r_{\infty}} \right).
\]

(6.14)

For \(s\)-wave annihilation, we take \(\langle \sigma v \rangle_f \simeq \langle \sigma v \rangle_{\text{CMB}}\). Since \(\langle \sigma v \rangle_f\) increases with \(\log(1/r_{\infty})\), but the CMB bound on \(\langle \sigma v \rangle_{\text{CMB}}\) increases with \(1/r_{\infty}\), we can evade the CMB constraints by decreasing \(r_{\infty}\). For a given DM mass, thermal ADM is consistent with the CMB constraints if \(r_{\infty}\) satisfies the following condition,

\[
r_{\infty} \ln \left( \frac{1}{r_{\infty}} \right) < \frac{2.42 \times 10^{-2}}{f \times c_f} \left( \frac{m_X}{1 \text{ GeV}} \right).
\]

(6.15)
The numerical result for this bound is shown in Fig. (6.2); a good analytic approximation is given by \( r_\infty < r_0 / \ln (1/r_0) \), with \( r_0 \simeq 2 \times 10^{-2} (m_X / \text{GeV}) / f \). Taking \( f \sim 1 \), we can see that \( r_\infty \) has to be smaller than \( 5 \times 10^{-3} \) and \( 2 \times 10^{-6} \) for \( m_X \sim 1 \) GeV and 1 MeV, respectively.

Likewise, we can combine Eq. (6.13) and Eq. (6.14) to place a lower bound on \( \langle \sigma v \rangle_f \):

\[
\frac{\langle \sigma v \rangle_f}{c_f \times 5 \times 10^{-26} \text{ cm}^3/\text{s}} \gtrsim 5 \times 10^{-26}
\]

\[
\begin{cases}
\ln \left( 40 c_f f \times \frac{1}{m_X} \text{GeV} \right) + \ln \ln \left( 40 c_f f \times \frac{1}{m_X} \text{GeV} \right), & m_X \lesssim f \times 10 \text{ GeV}, \\
2, & m_X \gtrsim f \times 10 \text{ GeV}.
\end{cases}
\]

Note if \( m_X \) is larger than \( f \times 10 \) GeV, the CMB constraints do not apply and the annihilation cross section is set by the relic density requirement. The analytic approximation in Eq. (6.16) agrees well with the numerical results, which are shown in Fig. (6.2).

With these constraints on the minimum annihilation cross section, we now turn to discussing what classes of models can generate the needed annihilation cross section consistent with collider constraints.

### 6.4 Light Mediators

Thus far, we have treated the annihilation cross section \( \langle \sigma v \rangle \) as a free parameter. To proceed we must specify the physics that generates this cross section. First, DM may annihilate directly to SM particles through heavy mediators with mass greater than the weak scale. This coupling to the SM implies light DM can be produced in abundance in colliders. We review constraints from missing (transverse) energy searches at collider experiments and from direct detection experiments, which conflict with the \( \langle \sigma v \rangle \) required to obtain the observed relic density. In this case, thermal light DM is ruled out in both the symmetric
and asymmetric scenarios. Second, DM can annihilate via new light states which have a mass below the typical momentum transfer scale in the colliders. In this case, the collider constraint can be evaded. If the new state is lighter than DM, it can be very weakly coupled to the SM.

### 6.4.1 Collider and Direct Detection Constraints on Light DM with Heavy Mediators

In the heavy mediator case, a convenient way to parametrize the DM-SM coupling is via higher dimensional operators, which is valid if the mediator mass is heavier than the relevant energy scale. Here we give two typical examples,

\[
\mathcal{O}_1 : \frac{\bar{X} \gamma_{\mu} X f \gamma_{\mu} f}{\Lambda_1^2} \quad \text{and} \quad \mathcal{O}_2 : \frac{\bar{X} X f f}{\Lambda_2^2},
\]

where \( X \) is DM, \( f \) is a SM fermion, and \( \Lambda_{1,2} \) are cut-off scales for \( \mathcal{O}_{1,2} \). The cut-off scale, in terms of the parameters in the UV-complete models, is \( \Lambda = m_\phi / \sqrt{g_X g_f} \), where \( m_\phi \) is the mediator mass, and \( g_X \) and \( g_f \) are coupling constants of DM-mediator and SM-mediator interactions respectively.

In the limit of \( m_X \gg m_f \), the DM annihilation cross sections at freezeout are given by

\[
\langle \sigma v \rangle_1 \approx \frac{N_f}{8\pi} \frac{m_X^2}{\Lambda_1^4} \quad \text{and} \quad \langle \sigma v \rangle_2 \approx \frac{N_f}{8\pi} \frac{m_X^2}{\Lambda_2^4} \frac{1}{x_f},
\]

for \( \mathcal{O}_1 \) and \( \mathcal{O}_2 \) respectively. \( N_f \) is the color multiplicity factor of fermion \( f \), and \( x_f = m_X/T \approx 20 \), with \( T \) the temperature. Note that the annihilation cross section through \( \mathcal{O}_2 \) is \( p \)-wave suppressed. Now we can estimate the limit on the cut-off scales \( \Lambda_1 \) and \( \Lambda_2 \) by
requiring the correct relic density

\[
\begin{align*}
\Lambda_1 &\lesssim 370 \text{ GeV} \left( \frac{N_f}{3} \right)^{1/4} \left( \frac{m_X}{10 \text{ GeV}} \right)^{1/4} \left( \frac{6 \times 10^{-26} \text{ cm}^3/\text{s}}{\langle \sigma v \rangle} \right)^{1/4}, \\
\Lambda_2 &\lesssim 100 \text{ GeV} \left( \frac{N_f}{3} \right)^{1/4} \left( \frac{m_X}{10 \text{ GeV}} \right)^{1/4} \left( \frac{6 \times 10^{-26} \text{ cm}^3/\text{s}}{\langle \sigma v \rangle} \right)^{1/4} \left( \frac{20}{x_f} \right)^{1/4},
\end{align*}
\]

(6.19)

(6.20)

where the limit is relevant for both the asymmetric and symmetric cases. Since the annihilation cross section is \( p \)-wave suppressed for \( \mathcal{O}_2 \), we need a smaller cut-off scale to obtain the correct relic abundance. Now we review various constraints on the cut-off scales \( \Lambda_{1,2} \).

- **Direct Detection Constraints**

If DM couples to quarks, the operators \( \mathcal{O}_{1,2} \) can lead to direct detection signals with the DM-nucleon scattering cross section: \( \sigma_{n_1,2} \sim \mu_n^2 / \Lambda_{1,2}^4 \), and \( \mu_n \) is the DM-nucleon reduced mass. For a DM mass \( \sim 10 \text{ GeV} \), taking the value of \( \Lambda_{1,2} \) given in Eqs. (6.19) and (6.20), we expect the DM-nucleon scattering cross section to be \( \sigma_{n_1} \sim 10^{-38} \text{ cm}^2 \) and \( \sigma_{n_2} \sim 10^{-36} \text{ cm}^2 \). However, the current upper bound on \( \sigma_n \) from direct detection experiments for DM with mass \( m_X \gtrsim 10 \text{ GeV} \) is \( \sigma_n \lesssim 10^{-42} \text{ cm}^2 \) [29], which is much smaller than the predicted values from requiring the correct thermal relic density. For DM with mass below a few GeV, the recoil energies are too small and direct detection bounds are currently very weak or nonexistent.

- **Tevatron and LHC Constraints**

The DM-quark interactions given in \( \mathcal{O}_{1,2} \) can lead to signals of mono-jet plus missing transverse energy at hadron colliders, while the Tevatron data for this signal matches the SM prediction well. We require that \( \mathcal{O}_{1,2} \) do not give rise to sizable contributions to this signal. The lower bounds on \( \Lambda_{1,2} \) are \( \sim 400 \text{ GeV} \) and \( \sim 400 \text{ GeV} \) [150, 149, 36]
respectively, for DM masses $m_X \lesssim 10$ GeV that we are interested in. Recent LHC results give a stronger limit on $\Lambda_1 \gtrsim 700$ GeV [135]. Therefore the Tevatron and LHC searches have excluded both thermal symmetric DM and ADM in the whole range of light DM if the DM particles annihilate to light quarks through $\mathcal{O}_1$ and $\mathcal{O}_2$.

- **LEP Constraints**

If DM particles couples to the electron through $\mathcal{O}_{1,2}$, the mono-photon search at LEP sets a limit on the cut-off scale: $\Lambda_1 \gtrsim 480$ GeV and $\Lambda_2 \gtrsim 440$ GeV for DM mass $m_X \lesssim 10$ GeV [134]. Note the limit also applies to the case where DM couples to three generations of charged leptons universally. One may avoid the limit by coupling DM only to $\mu$ or $\tau$. However this approach usually involves model building complications and severe flavor constraints.

Thus we conclude that for $\mathcal{O}_{1,2}$, DM does not have the correct relic abundance for symmetric DM and ADM due to the combination of direct detection and collider constraints. The direct detection constraints can be relaxed by suppressing the direct detection scattering cross section; this can happen for example if the scattering off nuclei is velocity suppressed, notably through an axial interaction. However, the collider bounds are still severe for higher dimensional operators involving interactions with light quarks or electrons [150, 149, 36, 134, 234, 135].

### 6.4.2 Light Dark Matter with Light Mediators

One simple way to evade the collider constraints for light DM is to invoke light mediators with masses much smaller than the typical transverse momentum of the colliders $p_T \sim$
\( \mathcal{O}(100 \text{ GeV}) \) (or the center of mass energy \( \sim 200 \text{ GeV} \) for LEP). In this limit, the effective theory approach breaks down and the collider bounds become much weaker [36, 134, 135, 151]. In general, if the mediator mass is much less than the \( p_T \) probed at colliders, there exists a large parameter space for light DM scenarios to achieve the correct relic density. We consider a hidden sector with Dirac DM coupled to a light mediator which could be a spin-1 or spin-0 particle; for ease of notation we always refer to it as \( \phi \). We write the Lagrangians as

\[
\mathcal{L}_V = g_X \bar{X} \gamma^\mu X \phi_\mu + g_f \bar{f} \gamma^\mu f \phi_\mu + m_X \bar{X} X + m_\phi \phi^\mu \phi_\mu, \tag{6.21}
\]

\[
\mathcal{L}_S = g_X \bar{X} X \phi + g_f \bar{f} f \phi + m_X \bar{X} X + m_\phi \phi^2, \tag{6.22}
\]

where \( m_\phi \) is the mediator mass. We consider two cases for the mediator mass: a mediator with \( m_\phi > 2m_X \) and lighter mediator with \( m_\phi < m_X \).

In the case of \( p_T \gg m_\phi > 2m_X \), the DM particles can annihilate to SM particles through the \( s \)-channel process. There is a collider bound on \( g_f \) because an on-shell mediator which decays to \( X \bar{X} \) can be produced, potentially contributing to the mono-jet plus missing transverse energy signal. Tevatron data has been employed to place an upper bound on \( g_f < 0.015/\sqrt{Br(\phi \rightarrow X \bar{X})} \) for \( m_\phi < 20 \text{ GeV} \) [151], where \( Br(\phi \rightarrow X \bar{X}) \) is the branching ratio of \( \phi \) decay to the DM pair. In this case the annihilation cross section is given by

\[
\langle \sigma v \rangle_V \approx 4\alpha_X g_f^2 m_X^2 N_f^c / m_\phi^4 \quad \text{and} \quad \langle \sigma v \rangle_S \approx \alpha_X g_f^2 m_X^2 N_f^c / 2m_\phi^4 x_f,
\]

where \( \alpha_X \equiv g_X^2 / 4\pi \). To see how the collider constraint affects the annihilation cross section in this case, we take the conservative limit \( g_f \lesssim 0.015 \), setting \( Br(\phi \rightarrow X \bar{X}) \approx 1 \). From the relic density constraint,

\footnote{In this paper, we do not consider the intermediate case \( m_\phi \sim 2m_X \), where there is a resonance in the \( s \)-channel annihilation of \( X \bar{X} \).}
we then obtain an upper bound on the mediator mass,

$$m_\phi \lesssim 13 \text{ GeV} \left( \frac{\alpha_X}{10^{-1}} \right)^{1/4} \left( \frac{10^{-25} \text{ cm}^3/\text{s}}{\langle \sigma v \rangle} \right)^{1/4} \left( \frac{m_X}{1 \text{ GeV}} \right)^{1/2}. \quad (6.23)$$

This bound\textsuperscript{7} is consistent with our assumption that $m_\phi \gg m_X$.

If $m_\phi < m_X$, DM can annihilate to the mediator directly and the annihilation cross section is determined primarily by the hidden sector coupling $g_X$:

$$\langle \sigma v \rangle_V = \frac{\pi \alpha_X^2}{m_X^2} \frac{m_X^2}{2 m_\phi^2 m_X} \sqrt{1 - \left( \frac{m_\phi}{m_X} \right)^2}, \quad \langle \sigma v \rangle_S = \frac{9 \pi \alpha_X^2}{2 m_X^2} \frac{T}{m_X} \sqrt{1 - \left( \frac{m_\phi}{m_X} \right)^2}, \quad (6.24)$$

for the vector and scalar mediators respectively. Meanwhile $g_f$ determines how the DM sector couples to the SM sector. As for the collider physics, the production of $X\bar{X}$ occurs through an off-shell mediator; since this is a three-body process, the bound is rather weak. Tevatron data requires $g_f \lesssim 0.2$ if the mediator couples to quarks universally [151].

Although $g_f$ does not appear to play an important role in the relic density, this coupling controls the width (lifetime) of $\phi$ and is relevant for cosmology. The width $\Gamma_\phi$ of the mediator is

$$(\Gamma_\phi)_V = \frac{4N_c}{3} \frac{m_\phi}{16\pi} g_f^2 \sqrt{1 - \left( \frac{2m_f}{m_\phi} \right)^2}, \quad (\Gamma_\phi)_S = 2N_f \frac{m_\phi}{16\pi} g_f^2 \sqrt{1 - \left( \frac{2m_f}{m_\phi} \right)^2}, \quad (6.25)$$

where the lifetime $\tau_\phi = \Gamma_\phi^{-1}$. In Section 6.2, we assumed the DM particles to be in thermal equilibrium with the SM thermal bath in the early universe, and in this case the standard freezeout picture and cosmology apply. Now, we check the condition for thermalization of the two sectors. If the mediator decay rate is larger than the Hubble expansion rate at temperatures $T > m_\phi$, then the inverse decay processes can keep $\phi$ in chemical equilibrium with the SM thermal bath [122]. At these temperatures, the decay rate is given by $\Gamma_\phi \sim$ \textsuperscript{7}Note that in this case there are also strong bounds on $m_\phi$ from neutrino experiments [107]; however, we have checked that it is still possible to obtain the correct relic density and that the direct detection predictions are unaffected.
\( g_f^2 m_\phi^2 / (16\pi T) \), where the factor of \( m_\phi / T \) accounts for the effect of time dilation. In order for the mediator to stay in thermal equilibrium with the SM thermal bath through DM freezeout, we require \( \Gamma_\phi \gtrsim H \) at temperatures \( T \sim m_X \). This gives a constraint on \( g_f \):

\[
g_f \sim \sqrt{\frac{16\pi \Gamma_\phi}{m_\phi}} \gg 8 \times 10^{-8} \left( \frac{\sqrt{g_{\text{eff}}}}{9} \right)^{1/2} \left( \frac{m_X}{\text{GeV}} \right)^{3/2} \left( \frac{100 \text{ MeV}}{m_\phi} \right) .
\]

(6.26)

If \( g_f \) is less than the bound given in Eq. (6.26), the DM sector can have a different temperature from the SM sector and the standard freezeout calculation can be modified in a number of ways. We have checked that these effects lead to change in the minimum annihilation cross section by less than a factor \( \mathcal{O}(10) \), compared to the results we derived, in Sections 6.2-6.3. Furthermore, the massive mediator is a late-decaying particle and in the case where the mediator decays to the SM states, can modify standard nucleosynthesis (BBN). There are stringent constraints on the hadronic decay of long-lived particles from the \(^4\text{He}\) fraction, which requires that the lifetime of the mediator be less than \( 10^{-2} \) s [187, 188, 178]. This leads to a lower bound of \( g_q \gtrsim 1.6 \times 10^{-11} \sqrt{1 \text{ GeV}/m_\phi} \) for a vector mediator, where we take \( N_f^c = 3 \). For leptonic decay modes, we take the lifetime of the mediator \( \tau_\phi \lesssim 1 \) s, and obtain a slightly weaker bound, \( g_e \gtrsim 5 \times 10^{-11} \sqrt{10 \text{ MeV}/m_\phi} \), for a vector mediator with \( N_f^c = 1 \).

Finally, we comment on the calculation of the relic density and application of the CMB constraints in the light mediator case. When \( m_\phi < m_X \), \( \bar{X}X \) can annihilate to \( \phi\phi \), but \( \phi \) decays to standard model particles rapidly compared to the relevant time scales at recombination so that the CMB constraints are unchanged. The only difference between a heavy mediator and light mediator with large width is whether there is a contribution to the effective degrees of freedom, \( g_\ast \), from the light mediator. A slightly higher \( g_\ast \) in the light mediator case gives rise to smaller \( r_\infty \), which in turn weakens the lower bound on \( \langle \sigma v \rangle \) from
CMB constraints.

In addition, we have neglected the Sommerfeld enhancement effect. As we will discuss in the following section, the mediator mass is bounded from below by DM halo shapes; this limits the size of any Sommerfeld enhancement. In addition, since $\langle \sigma v \rangle \approx \pi \alpha_X^2 / m_X^2$, for light DM the coupling $\alpha_X$ can be much smaller and still satisfy the relic density constraint. For the DM masses considered here, we have checked that the Sommerfeld enhancement effect is negligible for $s$-wave and $p$-wave annihilation processes at both freezeout and during recombination, if we take $\alpha_X$ and $m_\phi$ close to their minimum allowed values.

### 6.5 Halo Shape Constraints on the Mediator Mass

The presence of the light mediator allows for significant DM self-interactions, which can have non-trivial effects on DM halo dynamics. A number of astrophysical observations constrain DM self-interactions, for example observations of the Bullet Cluster [211], elliptical galaxy clusters [214] and elliptical DM halos [120, 121]. Among these, the upper bound on DM self-interaction from the ellipticity of DM halos is the strongest [120]. DM self-interactions can erase the velocity anisotropy and lead to spherical DM halos, so the observed ellipticity of DM halos constrains the DM self-scattering rate. Because the strength of self-interaction increases as the mediator mass decreases, we can use the elliptical halo shape constraint to place a lower limit on the mediator mass. Note that in the case of $m_\phi = 0$, the ellipticity of the DM halos then places a strong upper limit on the hidden sector coupling $g_X$ [213]; it is only possible to obtain the correct relic density if $m_X \gtrsim 10^3$ GeV [120, 4]\(^8\).

\(^8\)This limit can be relaxed if the hidden sector is much colder than the visible sector when DM freezes out. In this case, DM can achieve the correct relic density with a smaller annihilation cross section [124].
The effect of DM self-interactions on DM halo shapes can be parametrized by the average rate for DM particles to change velocities by $O(1)$ [121]:

$$\Gamma_k = \int \! d^3v_1 d^3v_2 f(v_1) f(v_2) (n_X v_{\text{rel}} \sigma_T) \frac{v_{\text{rel}}^2}{v_0^2}, \quad (6.27)$$

where $n_X$ is the DM density in the DM halo, $v_{\text{rel}} = |\vec{v}_1 - \vec{v}_2|$, and $f(v)$ is the DM velocity distribution in the DM halo, for which we take $f(v) = e^{-v^2/v_0^2}/(v_0 \sqrt{\pi})^3$. $\sigma_T$ is the scattering cross section weighted by the momentum transfer: $\sigma_T = \int \! d\Omega (d\sigma/d\Omega)(1 - \cos \theta_*)$.

The form of $\sigma_T$ depends on the particle physics nature of DM self-interactions and the relevant momentum scales. If the mediator is lighter than the typical momentum transfer in collisions, DM particles interact through long-range forces and $\sigma_T$ depends on velocity. In the opposite limit where the mediator is heavy compared to momentum transfer, DM self-interactions are contact interactions and $\sigma_T$ is independent of $v_{\text{rel}}$. In this case, we can take the $\sigma_T$ out of the velocity integrals in Eq. (6.27) and the calculation is straightforward. We first will derive the upper bound on the DM self-interaction cross section assuming a contact interaction, and then show that this limit applies in deriving the minimum mediator mass.

We consider the well-studied elliptical galaxy NGC720 [65, 169], taking our bound from the observed ellipticity at a radius of 5 kpc. The DM density profile is fit with local density 4 GeV/cm$^3$ and radial velocity dispersion $\bar{v}_r^2 = v_0^2/2 \simeq (240 \text{ km/s})^2$. We require the average time for DM self-interactions to create $O(1)$ changes on DM velocities to be larger than the galaxy lifetime $t_g \sim 10^{10}$ years i.e. $\Gamma_k^{-1} > t_g$. This gives the upper bound

$$\sigma_T \lesssim 4.4 \times 10^{-27} \text{ cm}^2 \left( \frac{m_X}{1 \text{ GeV}} \right) \left( \frac{10^{10} \text{ years}}{t_g} \right). \quad (6.28)$$

The reader should bear in mind that this is an analytic estimate and detailed N-body simulations studying a range of elliptical galaxies are required for a robust bound.
Figure 6.3: Lower limit on the mediator mass from combining relic density and DM self-interaction constraints. We show the case of a vector mediator; the result for a scalar mediator is similar and is given in Eq. (6.31). We consider DM self-interaction constraints from elliptical halo shapes and elliptical cluster shapes. Bullet cluster constraints do not give a lower bound on $m_\phi$. The dashed red line indicates the bound on the mass from elliptical halo shapes if CMB bounds are also applied, assuming efficiency $f \approx 1$.

Other astrophysical constraints have been derived for $\sigma/m_X$, assuming a hard sphere scattering cross section $\sigma$. A similar bound derived from shapes of elliptical galaxy clusters is $(\sigma/m_X \lesssim 10^{-25.5} \text{ cm}^2/(m_X/ \text{ GeV}))$ [214]. Specifically, this estimate is obtained from the inner regions of the galaxy cluster MS2137-23, at a radius of 70 kpc with dark matter density $\sim 1$ GeV/cm$^3$. Cosmological simulations of cluster-sized objects support this estimate within an order of magnitude [273]; however, the bound is still based on a single cluster. There is also a bound derived from the Bullet Cluster $(\sigma/m_X \lesssim 2 \times 10^{-24} \text{ cm}^2/(m_X/ \text{ GeV}))$ [211], reproduced in simulations of the collision by [235]. Note that this result is not derived from the shapes of the merging clusters but from requiring that the subcluster does not lose a
significant fraction of its mass in passing through the larger cluster; however, we have found
that the bound is too weak in this case to give a minimum mediator mass.

For the vector and scalar interactions considered here, the force is described by a Yukawa
potential \( V(r) = \pm \alpha_X e^{-m_\phi r} / r \). Depending on the mediator, and whether we are in the
asymmetric limit, the sign may be positive or negative. For the vector case, we have both
\( XX \) interactions (+) and \( X \bar{X} \) interactions (-) unless we are in the asymmetric limit. For the
scalar case, the sign is always negative. However, in the limit of a contact interaction, the
sign of the potential does not matter. The momentum transfer cross section for scattering
through \( t \) and \( u \)-channel processes in the Born approximation is

\[
\sigma_T \approx \frac{4\pi \alpha_X^2 m_X^2}{m_\phi^4},
\]

which is subject to the bound in Eq. (6.28). We have assumed a contact interaction,
\( m_X v_{\text{rel}} / m_\phi \ll 1 \); we will justify later that this is a valid assumption in deriving the bounds
below.

On the other hand, the relic density constraint places a lower bound on the annihilation
cross section \( \langle \sigma v \rangle \gtrsim 10^{-25} \text{cm}^3/\text{s} \) for light DM and thus on \( \alpha_X \):

\[
\alpha_X |_V \gtrsim 5 \times 10^{-5} \left( \frac{\langle \sigma v \rangle}{10^{-25} \text{cm}^3/\text{s}} \right)^{1/2} \left( \frac{m_X}{\text{GeV}} \right),
\]

\[
\alpha_X |_S \gtrsim 11 \times 10^{-5} \left( \frac{\langle \sigma v \rangle}{10^{-25} \text{cm}^3/\text{s}} \right)^{1/2} \left( \frac{m_X}{\text{GeV}} \right) \left( \frac{x_f}{20} \right)^{1/2},
\]

for vector and scalar coupling respectively. Note that we assume \( m_\phi < m_X \) and take the
annihilation cross sections in Eq. (6.24).

Since \( \alpha_X \) cannot be arbitrarily small, \( m_\phi \) cannot be made arbitrarily small. Combining

Chapter 6: Asymmetric and Symmetric Light Dark Matter

the bound on $\alpha_X$ with Eq. (6.28), we obtain a lower bound on the mediator mass:

$$m_\phi|_V \gtrsim 7 \text{ MeV} \left( \frac{\langle \sigma v \rangle}{10^{-25}\text{cm}^3/\text{s}} \right)^{1/4} \left( \frac{m_X}{\text{GeV}} \right)^{3/4},$$

$$m_\phi|_S \gtrsim 11 \text{ MeV} \left( \frac{\langle \sigma v \rangle}{10^{-25}\text{cm}^3/\text{s}} \right)^{1/4} \left( \frac{x_f}{20} \right)^{1/4} \left( \frac{m_X}{\text{GeV}} \right)^{3/4}$$  \hspace{1cm} (6.31)

for the vector and scalar mediator cases, where we take the elliptical galaxy with $t_g = 10^{10}$ years. Note that because the bound on $m_\phi$ scales as $\sigma_T^{-1/4}$ in the contact interaction limit, the result is not very sensitive to the precise bound on $\sigma_T$.

In deriving the above bound on $m_\phi$, we have assumed that $m_\phi \gg m_X v_{\text{rel}}$ and that the Born approximation is valid. Now we check that the bound given in Eq. (6.31) is consistent with these assumptions. The condition $m_\phi \gg m_X v_{\text{rel}}$ is satisfied for $1 \text{ MeV} < m_X < 10 \text{ GeV}$, since from Eq. (6.31) we have $m_\phi/m_X \sim 10^{-2}(m_X/\text{GeV})^{-1/4}$ but $v_{\text{rel}} \sim 10^{-3}$. In this limit the Born approximation is valid if the following condition is satisfied

$$m_X \left| \int_0^\infty r V(r)dr \right| = \frac{m_X \alpha_X}{m_\phi} \ll 1.$$  \hspace{1cm} (6.32)

From Eq. (6.30), we can see $v_{\text{rel}} \gg \alpha_X$ in the DM mass range we are interested in, and thus this condition is also satisfied if $m_\phi \gg m_X v_{\text{rel}}$. We emphasize that we cannot extrapolate the lower mass bound given in Eq. (6.31) to $m_X \gtrsim 50 \text{ GeV}$ because the Born approximation breaks down. For these higher masses, in general one has to solve the scattering problem numerically [63]. In the classical limit where $m_X v_{\text{rel}} \gg m_\phi$, there is a fitting formula available in [189] for the transfer cross section, which has been used to study self-interactions via a light mediator for DM masses greater than $\sim 100 \text{ GeV}$ [121, 173, 122, 203].

In Fig. (6.3) we show the lower limit on $m_\phi$ for the vector case, including the result derived from the more conservative bounds from elliptical cluster shapes. We also show the slightly
stronger result if we take the CMB constraint on the cross section,\(^9\) given in Eq. (6.16). There is a turnover for the elliptical cluster bounds because the contact interaction limit breaks down; here we use the full cross section, again in the Born approximation, given in [121]. The bounds from the Bullet Cluster, which we derive following [120], do not give rise to a lower bound on \(m_\phi\).

### 6.6 Direct Detection

Given the experimental effort needed to detect DM directly, it is important to map out the parameter space of direct detection cross sections, subject to the astrophysical and cosmological constraints we have discussed. Current experiments are not sensitive to DM-nucleon scattering if the DM mass is below \(\sim 1\) GeV because of the energy thresholds. It has been suggested that DM-electron scattering may provide an alternative way for the detection of light DM [112]. We consider DM-nucleon scattering for \(m_X \gtrsim 1\) GeV and DM-electron scattering for \(1\) MeV \(\lesssim m_X \lesssim 1\) GeV.

We compute the range of allowed elastic scattering cross sections within the framework of light DM annihilating via hidden sector mediators, assuming mediator couplings to electrons or light quarks. We consider both lighter mediators, \(m_\phi < m_X\), and heavier mediators, where we focus on the case \(m_\phi \gg m_X\). When \(m_\phi < m_X\) the mediator can be very weakly coupled to the SM, and so the scattering cross sections can be much smaller than when \(m_\phi \gg m_X\).

However, there is still a lower limit on the cross section coming from the lower bounds on the couplings of the mediator to the DM and SM fermions, \(\alpha_X\) and \(g_f\) respectively. The lower bound on \(\alpha_X\) is derived from requiring that relic density and CMB constraints are satisfied.

\(^9\)In the scalar case, annihilation is \(p\)-wave suppressed and thus CMB constraints don’t apply.
We consider two possible lower bounds on $g_f$: from requiring the thermalization between the DM and SM sectors, or from requiring decay of the mediator before BBN. When $m_\phi \gg m_X$, the lower limit on the cross section arises purely from the relic density and CMB constraints.

Meanwhile, we obtain upper bounds on the electron scattering cross section from the combination of halo shape bounds and requiring that the mediator does not significantly affect the electron anomalous magnetic moment. Including supernova and beam dump constraints on the dark force coupling [55] then carves out a nontrivial part of the parameter space for electron scattering.

Fig. (6.4) summarizes our results for the case where the mediator is a vector. We show the possible DM-nucleon (left panel) and DM-electron (right panel) scattering cross sections as a function of DM mass. The green shaded region is the parameter space for $m_\phi < m_X$ which is allowed by the constraints from the relic density, BBN, and DM halo shape constraints; in the electron case we include beam dump and supernova cooling constraints. The lighter green area is set by the additional assumption that the mediator has large decay width and thus that the two sectors are in thermal equilibrium. In the nucleon scattering case, $m_\phi \gg m_X$ is ruled out by CRESST-I and XENON10. In the electron scattering case, the red shaded region gives the allowed cross sections for $m_\phi \gg m_X$. In the following sections we derive these results and present more details.

### 6.6.1 Nucleon Scattering

We first consider nucleon scattering in the mass range $1$ GeV $\lesssim m_X \lesssim 10$ GeV, taking universal couplings to the light quarks given by $g_q$. The DM-nucleon scattering cross section
Figure 6.4: (Left) Nucleon scattering through a vector mediator. The green shaded region indicates the allowed parameter space of direct detection cross sections. The lighter green region imposes the bound of thermal coupling between the two sectors (“large width”) while the larger shaded region only requires mediator decay before BBN. Also shown is the lower bound for the heavy mediator ($m_\phi \gg m_X$) case. (Right) Electron scattering through a vector mediator, for $m_\phi < m_X$ (green) and $m_\phi \gg m_X$ (red); the intersection of the two regions is shaded brown. We show the projected sensitivity of a Ge experiment, taken from [112]. Beam dump, supernova, and halo shape constraints apply here and carve out the region of large $\sigma_e$ at low $m_X$. For more details, see the text. In the lighter green region, the condition of thermal equilibrium between the visible and hidden sectors is imposed.

is given by

$$\sigma_n = 4\alpha_X g_n^2 \mu_n^2 \mu_\phi^2,$$

(6.33)

where $\mu_n$ is the WIMP-nucleon reduced mass, and $g_n = 3g_q$ is the $\phi_\mu$-nucleon coupling constant. The upper bounds here are set by results from direct detection experiments, in particular CRESST-I [26] and XENON10 [25]. We have taken a contact interaction; this is a good approximation over much of the parameter space because the momentum transfer is generally less than the minimum mediator mass allowed by the ellipticity of DM halos, as
discussed in Section 6.5. We note that momentum-dependence can be relevant for scattering off heavier nuclei such as xenon if we take $m_\phi$ to be close to this minimum value, and thus can change the upper limit from XENON10 [133, 117, 115]. However, the lower limit is obtained in the limit that $m_\phi \approx m_X$ and thus momentum dependence will not be important. We therefore consider the bounds on a contact interaction for simplicity.

To determine the lower limit on this cross section, we bound $\alpha_X$ and $g_q$ from below in the case that the mediator is lighter than the DM, $m_\phi < m_X$. For thermal DM and masses $m_X > 1$ GeV, a lower bound on $\alpha_X$ is determined primarily by the relic density. As described in Section 6.3, CMB constraints are only important in this mass range if $\phi_\mu$ decays dominantly to electrons, for which the efficiency factor is $f \sim 1$. For $\phi_\mu$ coupling primarily to quarks, $f \approx 0.2$ and CMB bounds don’t apply above $m_X \sim 2$ GeV. Then the minimum annihilation cross section is $\langle \sigma v \rangle \approx \pi \alpha_X^2 / m_X^2 \approx 10^{-25}$ cm$^3$/s, giving a bound of $\alpha_X \gtrsim 5.2 \times 10^{-5} (m_X / \text{GeV})$. Requiring thermal equilibrium between the hidden and visible sectors, we take the bound on $g_q$ in Eq. (6.26), with $\sqrt{g_{\text{eff}}} \approx 9$. Combining the limits above results in a lower bound on the nucleon scattering cross section:

$$\sigma_n \gtrsim 10^{-48} \text{cm}^2 \times \left( \frac{m_X}{\text{GeV}} \right)^4 \left( \frac{\text{GeV}}{m_\phi} \right)^6 \left( \frac{\mu_n}{0.5 \text{ GeV}} \right)^2.$$  

(6.34)

Since $m_\phi < m_X$, this quantity is saturated for any $m_X$ if we set $m_\phi$ to its maximum value of $m_\phi \sim m_X$. This bound is indicated by the “Large width” line in Fig. (6.4). Coincidentally, the lower limit here is similar to the best achievable sensitivity for WIMP-nucleon scattering if the dominant irreducible background is coherent scattering of atmospheric neutrinos off of nuclei [260, 216, 136]. However, these studies focused on WIMP DM; for light DM, solar neutrinos become much more important and the best achievable sensitivity may be several orders of magnitude weaker.
The lower bound on $\sigma_n$ given in Eq. (6.34) is derived by requiring the two sectors be in thermal equilibrium. We may relax this assumption, and just demand the mediator decay by nucleosynthesis. This gives $g_{\phi} \gtrsim 1.6 \times 10^{-11} \sqrt{1 \text{ GeV}/m_{\phi}}$, as discussed in Section 6.4.2. For such $g_{\phi}$ the two sectors are decoupled through freezeout; then the relic density calculation is slightly more complicated and depends on the thermal history of the sectors. The change in the relic density then modifies the bound on $\alpha_X$. We have checked that the full calculation generally only changes the bound on $\alpha_X$ by an $O(1)$ factor [200], so here we take the bound on $\alpha_X$ from the large $\phi$ width case for simplicity. In this limit, the lower bound on $\sigma_n$ is given by

$$\sigma_n \gtrsim 5 \times 10^{-54} \text{cm}^2 \left( \frac{m_X}{\text{GeV}} \right) \left( \frac{\text{GeV}}{m_{\phi}} \right)^5 \left( \frac{\mu_n}{0.5 \text{ GeV}} \right)^2$$

labeled as “Decay before BBN” in Fig. (6.4).

For reference, we also give the lower bound on the cross section in the case where $m_{\phi} \gg m_X$. Here DM annihilation occurs directly to SM final states through $\phi_{\mu}$, with annihilation cross section $\langle \sigma v \rangle = 4\alpha_X g_n^2 m_X^2 / m_{\phi}^4$. Since the same combination of parameters enters in both the annihilation cross section and the nucleon scattering cross section, we can directly apply the relic density constraint to obtain

$$\sigma_n \gtrsim 5 \times 10^{-37} \text{cm}^2 \left( \frac{1 \text{ GeV}}{m_X} \right)^2 \left( \frac{\mu_n}{0.5 \text{ GeV}} \right)^2.$$  

This is the “$m_{\phi} \gg m_X$” line in Fig. (6.4). However, this scenario is ruled out by the direct detection limits on the cross section.
Figure 6.5: (Left) Constraints on mediator mass $m_\phi$ and coupling to electrons $g_e$ for $m_\phi < m_X$. The shaded region is excluded from electron anomalous magnetic moment, beam dump experiments, and supernova cooling [55]. The red dashed line shows the $g_e$ value used to derive the corresponding red dashed line (“C”) in the right plot. (Right) Constraints on electron scattering from Fig. 6.4. The boundaries A, B, and C are discussed in more detail in the text.

### 6.6.2 Electron Scattering

We consider scattering off electrons for DM in the mass range $1$ MeV $< m_X < 1$ GeV. The DM-electron scattering cross section is

$$\sigma_e = 4\alpha_X g_e^2 \frac{\mu_e^2}{m_\phi^2}. \quad (6.37)$$

The lower bound on the scattering cross section can be derived in the same way as in the nucleon case, taking $m_\phi < m_X$. Here both CMB and relic density constraints apply, since $m_X < 1$ GeV and the energy deposition efficiency $f \approx 1$ for decay to electrons. We take the bound on the annihilation cross section in Eq. (6.16) with $c_f \approx 1$, giving a lower limit
on $\alpha_X$:

$$\alpha_X \gtrsim 4 \times 10^{-7} \left( \frac{m_X}{10 \text{ MeV}} \right) \sqrt{\ln \left( \frac{40 \text{ GeV}}{m_X} \right)}.$$

(6.38)

As in the nucleon case, a lower bound on the DM-electron scattering cross section can be derived by assuming that the hidden and visible sectors are in thermal equilibrium. Analogously to Eq. (6.34), we find

$$\sigma_e \gtrsim 3 \times 10^{-51} \text{cm}^2 \times \left( \frac{m_X}{10 \text{ MeV}} \right)^4 \left( \frac{10 \text{ MeV}}{m_\phi} \right)^6 \left( \frac{\mu_e}{0.5 \text{ MeV}} \right)^2 \sqrt{\ln \left( \frac{40 \text{ GeV}}{m_X} \right)},$$

(6.39)

where we take $\sqrt{g_{\text{eff}}} \approx 3$.

Again, it is possible that the DM sector thermal bath evolves independently from the SM sector and in this case we only require the mediator to decay before BBN. From Section 6.4.2, we take the bound $g_e \gtrsim 5 \times 10^{-11} \sqrt{10 \text{ MeV}/m_\phi}$. The minimum scattering cross section is

$$\sigma_e \gtrsim 3 \times 10^{-53} \text{cm}^2 \left( \frac{m_X}{10 \text{ MeV}} \right)^5 \left( \frac{10 \text{ MeV}}{m_\phi} \right)^5 \left( \frac{\mu_e}{0.5 \text{ MeV}} \right)^2 \sqrt{\ln \left( \frac{40 \text{ GeV}}{m_X} \right)}.$$

(6.40)

If the annihilation goes through a heavier mediator $m_\phi \gg m_X$, we derive the strongest lower bound on the scattering cross section by applying CMB and relic density constraints:

$$\sigma_e \gtrsim 4 \times 10^{-39} \text{cm}^2 \left( \frac{10 \text{ MeV}}{m_X} \right)^2 \left( \frac{\mu_e}{0.5 \text{ MeV}} \right)^2 \ln \left( \frac{40 \text{ GeV}}{m_X} \right).$$

(6.41)

For electron scattering there are no direct experimental bounds on $\sigma_e$. However, for $m_\phi < m_X$, there are bounds on $\sigma_e$ arising from indirect constraints, namely halo shape bounds and from searches for new light gauge bosons [55]. The halo shape constraint requires that the self-scattering cross section satisfy $\sigma_T/m_X < 4.4 \times 10^{-27} \text{cm}^2/\text{GeV}$ with $\sigma_T \simeq 4\pi\alpha_X^2 m_X^2/m_\phi^4$. If $m_\phi < m_X$ then constraints on new light gauge bosons rule out parts of the $(m_\phi, g_e)$ parameter space; we show beam dump, supernova cooling and electron anomalous magnetic
moment constraints\textsuperscript{10} in Fig. (6.5) (left panel). Here we make use of the convention in [55], where \( g_e = \epsilon e \), with the kinetic mixing parameter \( \epsilon \equiv \epsilon_Y \cos \theta_W \) and \( e \) electric charge. The solid line (and shaded region) indicates the constraint.

As a simple application of the constraints discussed above, we derive the upper bound on the cross section by rewriting \( \sigma_e \):

\[
\sigma_e = \frac{4\mu_e^2}{\sqrt{4\pi m_X}} \sqrt{\frac{m_X}{\alpha_X}} \left( \frac{g_e}{m_\phi} \right)^2 \lesssim 3.5 \times 10^{-35} \text{ cm}^2 \left( \frac{\mu_e}{0.5 \text{ MeV}} \right)^2 \sqrt{\frac{10 \text{ MeV}}{m_X}}. \tag{6.42}
\]

Here we have applied the halo shape constraint and taken \( (g_e/m_\phi)^2 \lesssim 10^{-1} e^2/\text{GeV}^2 \), arising from measurements of the electron anomalous magnetic moment [229].

To explain more complicated constraints on the \((m_X, \sigma_e)\) plane from the supernova cooling and beam dump experiments for \( m_\phi < m_X \), we show again the allowed parameter space for electron scattering cross sections, but highlight boundaries of the constraints by labeling “A”, “B”, and “C” in the right panel of Fig. (6.5). We can map excluded regions on the \((m_\phi, g_e)\) plane to these constraints:

- **Constraint “A”**:

  For \( m_\phi < m_X \lesssim 8 \text{ MeV} \), supernova plus beam dump constraints require \( g_e \lesssim 1.3 \times 10^{-9} \). This places a stringent upper bound on the cross section, which we derive by taking \( m_\phi \) to its minimum value of \( m_\phi = 2m_e \approx 1 \text{ MeV} \), and then setting \( \alpha_X \) to the maximum value allowed by halo shape constraints: \( \alpha_X < 9.5 \times 10^{-6} \sqrt{10 \text{ MeV}/m_X} \).

\textsuperscript{10}In general there are also constraints from low-energy \( e^+e^- \) colliders, fixed target experiments, and neutrino experiments [107]. We find these do not significantly affect our results. In the case of kinetic mixing, bounds from measurements of the muon anomalous magnetic moment also apply. We do not include them in this work.
This upper bound is then
\[
\sigma_e \lesssim 6 \times 10^{-45} \text{ cm}^2 \left( \frac{\mu_e}{0.5 \text{ MeV}} \right)^2 \sqrt{\frac{10 \text{ MeV}}{m_X}}. \tag{6.43}
\]

Note that the constraint changes somewhat if we also consider \( m_\phi < 1 \) MeV. In this case, supernova cooling constraints still require \( g_e \lesssim 1.3 \times 10^{-9} \) but halo shapes allow for a somewhat smaller \( m_\phi \). As a result, the upper bound is slightly weaker if we allow \( m_\phi < 1 \) MeV: \( \sigma_e \lesssim 6 \times 10^{-44} \text{ cm}^2 \left( \mu_e/0.5 \text{ MeV} \right)^2 \left( 10 \text{ MeV}/m_X \right)^{-2} \).

- **Constraint “B”:**

This constraint applies for the large width case. In contrast with constraint A, taking \((m_\phi, g_e) = (1 \text{ MeV}, 1.3 \times 10^{-9})\) is in conflict with the condition of thermal equilibrium between the two sectors if the DM mass \( m_X \gtrsim 5 \) MeV. Furthermore, for \( m_X \gtrsim 20 \) MeV, the region \((m_\phi \sim 20 \text{ MeV}, g_e \sim 3 \times 10^{-8})\) opens up. These competing effects lead to the kink in line B.

- **Constraint “C”:**

For \( m_X \gtrsim 8 \) MeV, then supernova and beam dump constraints allow a region of larger \( g_e \): for example, \((m_\phi \sim 8 \text{ MeV}, g_e \sim 6 \times 10^{-4})\) is now allowed. The red dashed lower bound on \( g_e \) in the left panel of Fig. 6.5 then gives rise to the constraint “C”. The lower bound on the cross section here comes from setting \( m_\phi \sim m_X \), applying the red dashed lower bound on \( g_e \), and setting \( \alpha_X \) to its minimum value from CMB constraints.

We make two final notes. First, in the heavy mediator case, the beam dump constraints do not apply and the CMB constraints are in general much stronger. As a result, the high \( \sigma_e \), low \( m_X \) region which is excluded in the light mediator case is again allowed indicated by the
light red shaded region in Fig. (6.4). Second, if we remove the constraint $m_\phi > 1$ MeV, $\phi$ will decay invisibly, and only the supernova constraints are relevant. Then a small region of parameter space with $g_e \sim 1.3 \times 10^{-9}$ and $m_\phi < 1$ MeV opens up, as discussed above under constraint “A.”

We have verified the bounds discussed above by performing a general scan of the hidden sector parameter space. Fig. (6.6) illustrates our method. We begin by mapping out the parameter space of $(m_\phi, g_e)$ and require either large $\phi$ width or $\phi$ decay before BBN. We combine this with the constraints in [55], given by the solid curve in the top panels of Fig. (6.6). In doing so, we impose the limit $1$ MeV $< m_\phi < m_X$ for the case of $m_\phi < m_X$ and $m_\phi > 2m_X$ in the case where $m_\phi \gg m_X$. The lower limit of $m_\phi > 1$ MeV is imposed in order to allow for $\phi$ decay to electrons. If the halo shape constraint gives a stronger lower bound on $m_\phi$, then we take $(m_\phi)_{\text{min,halo}} < m_\phi < m_X$ for the $m_\phi < m_X$ case, where $(m_\phi)_{\text{min,halo}}$ is minimum mediator mass allowed by the halo shape constraint. This generates the sampled points in $(m_\phi, g_e)$ that we have shown. For a fixed $(m_\phi, g_e)$, a range of values for $\alpha_X$ is allowed, giving rise to a range of allowed scattering cross sections. We sample random $\alpha_X$ values, subject to the halo shape constraint and the relic density constraint as in Eq. (6.38). This then gives a randomly sampled $\sigma_e$ value, which we indicate by the color of the point in Fig. (6.6). For a fixed $m_X$ value, because of the range of allowed $m_\phi$ and $\alpha_X$ values, excluded regions in $g_e$ do not directly map to an excluded region in $\sigma_e$. An excluded region in $\sigma_e$ only arises if a sufficiently large region of $g_e$ is excluded, as shown in the left plot of Fig. (6.6). We thus verify the possible values of $\sigma_e$ in this way, imposing all the constraints self-consistently.
Figure 6.6: For fixed $m_X$ and a mediator with mass $m_\phi < m_X$, we generate random values of $(m_\phi, g_e)$ allowed by beam dump, supernova, $a_e$, and BBN constraints. We show a sample of allowed points in the $(m_\phi, g_e)$ parameter space; the solid curve is extrapolated from the constraints in [55], also shown in left panel of Fig. (6.5). For each $(m_\phi, g_e)$ point, we then sample the allowed $\alpha_X$ satisfying halo shape and relic density constraints, and compute the corresponding elastic scattering cross section $\sigma_e$. The color of the point is determined by $\sigma_e$. (Left) $m_X = 20$ MeV, where the minimum mediator mass is $m_\phi = 1$ MeV. (Right) $m_X = 100$ MeV, where the minimum mediator mass $m_\phi \gtrsim 3$ MeV is set by halo shape constraints.

6.7 Conclusions

Given the unknown nature of DM, it is important to carry out broad-based studies of models of DM. In this chapter, we have examined constraints on thermal DM with mass $1$ MeV $\lesssim m_X \lesssim 10$ GeV, a mass range interesting for numerous phenomenological and theoretical reasons. We considered bounds from cosmology, colliders and astrophysics, and derived implications of these constraints on direct detection.

CMB constraints on DM annihilation present the most serious challenge for light thermal DM, excluding symmetric thermal relic DM with $s$-wave annihilation if $m_X \lesssim 1 - 10$ GeV. Two natural ways to evade this constraint are to have a DM number asymmetry or velocity suppressed annihilation. In the asymmetric case, we found the constraint on the annihilation
cross section such that the symmetric component efficiently annihilates away; the minimum cross section is larger than the usual thermal relic cross section by a factor of a few, depending on the mass.

Achieving this minimum cross section is difficult if annihilation occurs through a weak scale (or heavier) mediator. Collider and direct detection constraints have forced the presence of relatively light mediator states in the hidden sector in order to achieve the correct relic abundance and evade the CMB bounds. On the other hand, we found that the DM halo shape bounds on DM self-interactions require that the mediator is not too light. We examined constraints from elliptical galaxy NGC720 and elliptical clusters, and derived a lower bound on the mass of the mediator particle.

We also calculated the range of scattering cross sections allowed within this scenario. Although the lowest bound which is cosmologically consistent is well below the reach of any current or envisioned direct detection experiments, we showed that several cosmologically interesting benchmarks could be reached. For example, in the case of scattering off nucleons, a hidden sector in thermal contact with the SM at $T \sim m_X$ can be ruled out if an experiment can reach cross sections with $\sigma_n \lesssim 10^{-48}$cm$^2$. In the case of scattering off electrons, the scenario where $m_\phi \gg m_X$ can be probed by direct detection. Beam dump and supernova constraints carve out a significant fraction of the available parameter space if $m_\phi < m_X$. 
Chapter 7

Conclusions

We have explored direct detection of dark matter through its scattering off of standard model particles and indirect detection of dark matter through its annihilation or decay. In considering the direct detection of dark matter, we focus on two classes of models proposed as explanations of the DAMA/LIBRA annual modulation signal: inelastic dark matter and light dark matter with mass 1 MeV to 10 GeV. We show that directional detection experiments provide a powerful test of inelastic dark matter models. We map out the parameter space of elastic scattering cross sections for light dark matter, subject to astrophysical and collider constraints.

Dark matter annihilations or decay to high energy electrons and positrons give rise to gamma ray, cosmic ray, and microwave signals. We present a model-independent fit of recent anomalous signals in PAMELA, Fermi, and WMAP data to dark matter annihilation. Annihilations can also be observed indirectly in the cosmic microwave background, and we derive model independent constraints on the annihilation rate during the epoch of recombination.
Bibliography


