Opus 150: Dark Forces in the Universe

by John G. Cramer

Alternate View Column AV-150
Keywords: *column 150 cosmology dark matter WIMP detection annihilation 5th-force positron excess DAMA GLAST/Fermi propulsion energy*
Published in the December-2009 issue of ***Analog Science Fiction & Fact Magazine****;*

This column is a milestone. In 1983, while I was on a one-year sabbatical at the Hahn-Meitner Institute for Nuclear Physics in what was then West Berlin, I received a letter from Stan Schmidt informing me that Jerry Pournelle had decided that he no longer wished to be an Alternate View columnist for Analog and asking if I was interested in taking over as the AV columnist and “alternating” with G. Harry Stine.

This was a problem. At the time I had written about 80 papers for physics journals and a few science-fact pieces for Analog, but I was well aware that writing science-fact for a popular audience is harder and more time-consuming than it looks, and the idea of having to produce a sensible column for every-other issue of Analog on a regular basis was scary. I was not at all sure that I would have anything to write about when the deadlines came around. But I decided that the Analog soapbox was too tempting to pass up.

Fast forward to today. This is column number 150. Somehow, for over 25 years I have managed to meet each deadline with something (I hope) interesting to say about science in general and physics in particular. I think that popularizing science and making it accessible to interested readers is an important activity, and I hope you agree. With that said, let’s consider the subject of this column: possible indications of a new “dark” force in the universe.

It is now clear that our universe is a much stranger place than we had imagined only a decade ago. Its total mass-energy, according to our best cosmological models, divides up as 70% dark energy, 25% cold dark matter, 4% free hydrogen and helium, 0.5% stars (mostly hydrogen), 0.3% neutrinos, and only 0.03% atoms of elements heavier than helium, the stuff that we are mostly made of.

Dark energy, easily the most mysterious of these components, is an intrinsic energy of space spread uniformly through the universe and possessed by each otherwise completely empty volume of space. It creates a repulsive "pressure" that is accelerating the expansion of the universe. Dark matter, the next most mysterious, is some unknown form of mass that does not make or absorb light and interacts gravitationally with itself and with the normal mass of stars. Dark matter clusters around galaxies in a more-or-less spherical "halo", accounts for most of the galactic mass, and causes stars in the outer reaches of a galaxy to orbit the galactic center much faster than they would in the absence of the dark matter halo.

But what is dark matter? It is definitely not ordinary matter (atoms, molecules, electrons) or any of the known fundamental particles including neutrinos. So what's left? Nothing ordinary, so we are pushed into speculating that dark matter is made of a previously unknown family of particles. Some theories that attempt to extrapolate beyond the standard model of particle physics predict new particles: e.g., supersymmetric particles, WIMPs, axions, etc. For the purposes of this column, we'll refer to them all as DMP (i.e., dark matter particles).

How do you look for DMPs? There are basically two techniques: (1) We assume that a speeding DMP can collide with a normal nucleus, giving it enough recoil energy and momentum to trigger a sensitive detector, and (2) We assume that DMPs come in matter and antimatter flavors that can annihilate with each other and produce radiation detectible with sensitive instruments in space. Both types of DMP searches have been going on for some time, and are getting results that are both interesting and confusing.

Gran Sasso, located about 130 km from Rome, is the highest mountain in the Apennines of Central Italy. In 1995, twin highway tunnels connecting Rome to Teramo were cut through the mountain, and at the same time an underground particle physics laboratory was created there, consisting of three large underground low-background chambers shielded from cosmic rays by 1,400 meters of rock. In one of these chambers is the DAMA/LIBRA experiment, operating a cluster of sodium iodide detectors with a total mass of 250 kg (¼ ton) designed to detect small signals arising from the collision of a DMP with a nucleus. When a nucleus recoils from a DMP hit, a small flash of light is made by the resulting ionization, and this light flash can be detected with photomultiplier tubes. The DAMA detectors can observe such signals over background with energies as small as 2 thousand electron volts or 2 keV (here keV means kilo electron volts, the quantity of energy needed to move one electron across a potential of 1000 volts). DAMA has a lower threshold than that of most competing DMP searches. DAMA/LIBRA and its previous incarnation as DAMA/NaI have been operating in the very low radiation Gran Sasso environment for a total of 11 years. During this long period of operation, some interesting and controversial data has been collected.

In its yearly orbit around the Sun, the Earth has a speed of about 30 km/s. Our Sun orbits the galactic center with a speed of about 220 km/s. If our galaxy is embedded in a cloud of

DMPs more or less at rest with respect to the galactic center, the Earth passes through the DMP stream with a varying speed. At some times during its orbit, the Earth's speed adds to the Sun's speed, and at other times it subtracts. Thus, the DMPs streaming through the DAMA/LIBRA detector are expected to hit the detectors with more energy in June than in December.

Therefore, one of the signals examined by the DAMA/LIBRA collaboration is any annual variation in the counting rates in the detector, and indeed they have found such a variation at about the 2% level in counts with energies between 2 and 4 thousand electron volts (2 to 4 keV). As expected, the signal reaches a maximum around June 2, just when the Earth and Sun speeds add. The variation has period of exactly one year with a 0.2% uncertainty. They attribute their observations to the presence of DMPs in the galactic halo and ascribe a confidence level of 8.2 standard deviations to their result.

The problem is that other similar DMP detectors (Zeplin III, CDMS 2008) with a different detection medium and higher energy thresholds see no such effect. Also, the 2-4 keV signals seen by DAMA/LIBRA are lower in energy that is theoretically predicted for DMP-nucleus collisions. Therefore, the DAMA/LIBRA result has been the subject of debate and controversy at several recent international conferences.

As mentioned above, the other way of searching for DMPs is to look for space radiation produced when somewhere a DMP and anti-DMP annihilate. One particle expected from such annihilations is the positron, the antimatter twin of the electron. The PAMELA detector, launched on a Russian satellite by the WIZARD collaboration, has recently reported an observation of the ratio of positrons to (electrons + positrons) in the energy range 1.5 to 90 GeV. (Here, 1 GeV is 10^9 eV; for reference a proton has a mass-energy of 0.938 GeV.)

The theoretical expectation is that such energetic positrons should be produced mainly in gas collisions during the propagation of cosmic ray protons in the galactic medium, and this leads to a positron ratio that should fall steeply with energy. The actual data, however, shows a strong ***increase*** in the ratio with energy, starting at a value of about 0.05 at 10 GeV and rising to above 0.15 at 90 GeV. There are similar reports of a positron excess in the 600-800 GeV range from the balloon-borne ATIC cosmic ray detector, (but these seem to be in conflict with recent GLAST/Fermi results). This excess of energetic positrons is not easily explained, and could be the result of DMP-anti-DMP annihilation.

The WMAP experiment, which has mapped the cosmic microwave background with great precision, has also reported a hard microwave "haze" coming from the center of our galaxy and not readily explained by known galactic emission mechanisms. This haze could be the synchrotron radiation from energetic electrons and positrons made in DMP-anti-DMP annihilation near the galactic center. Studies there with the EGRET gamma ray detector showed an excess of gamma rays at energies above those expected from pi^0 meson decays. It is suggested that the observed gamma rays might arise from the inverse Compton scattering of energetic positrons and electrons colliding with starlight and cosmic background microwaves.

Thus, from several independent sources there is evidence of energetic electrons and positrons, possibly coming from DMP-anti-DMP annihilation. On the other hand, there is no corresponding evidence of any excess of strongly-interacting particles like pions and antiprotons, which would be expected from conventional matter-antimatter annihilations. In particular, the

PAMELA measurements place tight limits the antiproton content of cosmic rays, and EGRET measurements place similar limits on gamma rays from pi^0 meson decays. Neither provides any indication of strongly-interacting annihilation products.

Thus, these results present a paradox. If DMP-anti-DMP annihilations are producing electrons and positrons with energies of 90 GeV or more, why are these decays not making any excess of antiprotons or pi^0 mesons? This is in direct conflict with expectations based on the standard model, so the observations may point to new physics beyond the standard model.

One version of such new physics has recently been suggested by Arkani-Hamed, Finkbeiner, Slatyer, and Weiner (AFSW). They propose a new “dark” force that acts only between the dark-matter particles. In this AFSW model, when there is a DMP-anti-DMP annihilation between dark matter particles that wander into each other, the particle momentarily produced in annihilation is the carrier of the new dark-matter force. And it is assumed that the force-carrier particle has a mass of around 0.1 GeV, so that its low mass constrains its subsequent decay into electron-positron pairs and/or pairs of gamma rays.

Also, because of this new force, as the particles approach each other before annihilation an attraction due to the force occurs that increases the probability that the annihilation will occur. This is called a Sommerfeld enhancement, and it can increase the chances of annihilation by several orders of magnitude. And AFSW model predicts that DMPs should interact primarily with heavy nuclei giving lower recoil energies, explaining why only detectors containing iodine (like DAMA) or lead nuclei show small but detectable recoil signals.

The AFSW theory can, at the expense of postulating a new and previously unknown force, explain all of the observations described above and make predictions that can be tested by new experiments and observations. In particular, the AFSW theory predicts that, as more data is collected, the positron enhancement observed by PAMELA should continue to increase up to the highest energies that the detector can resolve (~270 GeV). As the sources of positrons become better localized in position, the theory predicts that sources of the energetic positrons should be broadly distributed rather than localized at the galactic center (e.g, at the black hole there).

When the LHC begins operation in a few months (a year late because of a major cryogenics rupture) the AFSW theory predicts observation of a striking signature of highly energetic electrons and positrons among the produced particles in the proton-proton collisions. There are also predictions for new observations from the GLAST/FERMI and HESS detectors that have recently been launched.

So there may be a new force in the universe. What are the science-fiction implications of that? If the picture painted by the AFSW theory is correct, the matter content of the universe is mostly dark matter with its own particles, forces, and interactions. It clusters around galaxies in a somewhat different way than does the stars and planets of normal matter, and it may have complexities and structures that we have not imagined. Are there the equivalent of atoms and molecules made of dark matter that are invisible to us except through their gravitational interactions? There are almost no antiprotons in our galactic neighborhood, but apparently there is plenty of dark matter and antimatter. Can we find chunks of dark matter and dark antimatter and annihilate them for energy and propulsion?

So there may be another force in the universe, a fifth “dark” force that acts only between “dark” particles that ignore us as we ignore them, passing invisibly through our stars and planets as if they were not there.

The surface has just been scratched in this area. For science fiction this has the makings of new kinds of energy sources and propulsion, and perhaps even a new “dark’ territory to be explored by space travelers.

AV Columns Online: Electronic reprints of about 150 "The Alternate View" columns by John G. Cramer, previously published in ***Analog*** , are available online at: http://www.npl.washington.edu/av.

References

The AFSW Theory:

“A Theory of Dark Matter”, N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner,
Physical Review ***D79***, 015014 (2009), arXiv preprint 0810.0713v3 [hep-ph].

A Theory of Dark Matter

Nima Arkani-Hamed,[1] Douglas P. Finkbeiner,[2] Tracy R. Slatyer,[3] and Neal Weiner[4]

[1]*School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA*

[2]*Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA*

[3]*Physics Department, Harvard University, Cambridge, MA 02138, USA*

[4]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*

(Dated: January 20, 2009)

We propose a comprehensive theory of dark matter that explains the recent proliferation of unexpected observations in high-energy astrophysics. Cosmic ray spectra from ATIC and PAMELA require a WIMP with mass $M_\chi \sim 500-800\,\mathrm{GeV}$ that annihilates into leptons at a level well above that expected from a thermal relic. Signals from WMAP and EGRET reinforce this interpretation. Limits on $\bar{p}$ and π^0-γ's constrain the hadronic channels allowed for dark matter. Taken together, we argue these facts imply the presence of a new force in the dark sector, with a Compton wavelength $m_\phi^{-1} \gtrsim 1\,\mathrm{GeV}^{-1}$. The long range allows a Sommerfeld enhancement to boost the annihilation cross section as required, without altering the weak-scale annihilation cross section during dark matter freeze-out in the early universe. If the dark matter annihilates into the new force carrier ϕ, its low mass can make hadronic modes kinematically inaccessible, forcing decays dominantly into leptons. If the force carrier is a non-Abelian gauge boson, the dark matter is part of a multiplet of states, and splittings between these states are naturally generated with size $\alpha m_\phi \sim$ MeV, leading to the eXciting dark matter (XDM) scenario previously proposed to explain the positron annihilation in the galactic center observed by the INTEGRAL satellite; the light boson invoked by XDM to mediate a large inelastic scattering cross section is identified with the ϕ here. Somewhat smaller splittings would also be expected, providing a natural source for the parameters of the inelastic dark matter (iDM) explanation for the DAMA annual modulation signal. Since the Sommerfeld enhancement is most significant at low velocities, early dark matter halos at redshift ~ 10 potentially produce observable effects on the ionization history of the universe. Because of the enhanced cross section, detection of substructure is more probable than with a conventional WIMP. Moreover, the low velocity dispersion of dwarf galaxies and Milky Way subhalos can increase the substructure annihilation signal by an additional order of magnitude or more.

PACS numbers: 95.35.+d

I. PAMELA/ATIC AND NEW DARK FORCES

Thermal WIMPs remain one of the most attractive candidates for dark matter. In addition to appearing generically in theories of weak-scale physics beyond the standard model, they naturally give the appropriate relic abundance. Such particles also are very promising in terms of direct and indirect detection, because they must have some connection to standard model particles.

Indirect detection is particularly attractive in this respect. If dark matter annihilates to some set of standard model states, cosmic ray detectors such as PAMELA, ATIC and Fermi/GLAST have the prospect of detecting it. This is appealing, because it directly ties the observable to the processes that determine the relic abundance.

For a weak-scale thermal particle, the relic abundance in the case of s-wave annihilation is approximately set by

$$\Omega h^2 \simeq 0.1 \times \left(\frac{\langle \sigma v \rangle_{\mathrm{freeze}}}{3 \times 10^{-26}\,\mathrm{cm}^3\mathrm{s}^{-1}} \right)^{-1}. \tag{1}$$

For perturbative annihilations, s-wave dominates in the late universe, so this provides an approximate upper limit on the signal that can be observed in the present day. Such a low cross section makes indirect detection, whereby the annihilation products of dark matter are detected in cosmic ray detectors, a daunting task.

However, recent experiments have confirmed the longstanding suspicion that there are more positrons and electrons at 10s-100s of GeV than can be explained by supernova shocks and interactions of cosmic ray protons with the ISM. The experiments are

- **PAMELA**
  The Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics has reported results [1] indicating a sharp upturn in the positron fraction $(e^+/(e^+ + e^-))$ from $10 - 100\,\text{GeV}$, counter to what is expected from high-energy cosmic rays interacting with the interstellar medium (ISM). This result confirms excesses seen in previous experiments, such as HEAT [2, 3] and AMS-01 [4]. One possible explanation for this is dark matter annihilation into e^+e^- [5, 6, 7], but this requires a large cross section [8].

- **ATIC**
  The Advanced Thin Ionization Calorimeter is a balloon-borne cosmic ray detector which studies electrons and positrons (as well as other cosmic rays) up to $\sim$ TeV energies, but cannot distinguish positrons and electrons. The primary astrophysical sources of high-energy electrons are expected to be supernovae: electrons are accelerated to relativistic speeds in supernova remnants and then diffuse outward. The ATIC-2 experiment reported a $4 - 6\sigma$ excess (over a simple power law) in its $e^+ + e^-$ data [9] at energies of $\sim 300 - 800\,\text{GeV}$, with a sharp cutoff in the $600 - 800\,\text{GeV}$ range. Dark matter would seem a natural candidate for this as well, with its mass scale determining the cutoff.

- **WMAP**
  Studies of the WMAP microwave emission from the galactic center show a hard component not spatially correlated with any known galactic emission mechanism. This "WMAP haze" [10, 11] can be explained as synchrotron radiation from electrons and positrons produced from dark matter annihilation in the galactic center [12].

- **EGRET**
  Gamma-ray measurements in the galactic center (inner 5°) provide hints of an excess at $10-50\,\text{GeV}$ [13]. Strong et al. reanalyzed EGRET data and found a harder spectrum at these energies than previously derived, using the improved EGRET sensitivity estimates of [14]. Despite poor spatial resolution, Strong et al. found an excess in this energy range above the expected π^0 gamma-ray emission from cosmic ray protons interacting with the interstellar medium (see Fig. 8 of [13]). Such γ's could naturally arise from inverse-Compton scattering of high-energy electrons and positrons off of starlight and the cosmic microwave background (CMB).

Taken together, these make a compelling case for excessive electronic production in the galaxy. While individual astrophysical explanations may exist for each signal (pulsar wind nebulae for PAMELA and ATIC [15, 16, 17], for instance, supernovae for the WMAP haze [63]), the data cry out for a unified explanation. Dark matter annihilations provide an appealing candidate.

In addition to the above, there are two other anomalies that are worth mentioning. The INTEGRAL 511 keV signal indicates $\sim 3 \times 10^{42} e^+/\text{s}$ annihilating in the galactic center, far more than expected from supernovae. The spectrum suggests that these positrons are injected with relatively low energies ($E \lesssim$ few MeV), and so form a distinct population from those above. eXciting dark matter (XDM) [18] can naturally explain such low energy positrons with $\sim 1\,\text{MeV}$ excited states of the dark matter, while still producing the high-energy positrons via annihilation [19, 20]. Lastly, there is the DAMA/LIBRA indication of an annual modulation consistent with that expected from dark matter induced nuclear scattering [21]. Such a signal is difficult to reconcile with null results of other experiments, but can be reconciled with $\sim 100\,\text{keV}$ excited states in the "inelastic dark matter" (iDM) scenario [22, 23, 24]. Although we are motivated by the specific signals above (PAMELA/ATIC as well as the haze and EGRET), the picture we are led to for explaining them naturally incorporates the necessary ingredients to explain the INTEGRAL and DAMA signals as well.

Focusing on only the high-energy positrons and electrons, there are a number of challenges to any model of dark matter. These are:

- **A large cross section**
  Studies of PAMELA and ATIC signals seem to require a cross section *much* larger than what is allowed by

thermal relic abundance. Boost factors of $\mathcal{O}(100)$ or more above what would be expected for a thermal WIMP are required to explain these excesses [25, 26].

- **A large cross section *into leptons***
  Typical annihilations via Z bosons produce very few hard leptons. Annihilations into W bosons produce hard leptons, but many more soft leptons through the hadronic shower. Higgs bosons and heavy quarks produce even softer spectra of leptons, all of which seem to give poor fits to the data. At the same time, absent a leptophilic gauge boson, it is a challenge to construct means by which dark matter would annihilate directly to leptons.

- **A *low* cross section into hadrons**
  Even if a suitably high annihilation rate into leptons can be achieved, the annihilation rate into hadronic modes must be low. Limits from diffuse galactic gamma rays [27], as well as gamma rays from the center of the galaxy constrain the production of π^0's arising from the dark matter annihilation. PAMELA measurements of antiprotons tightly constrain hadronic annihilations as well [26]. Consequently, although quark and gauge boson annihilation channels may occur at some level, the dominant source of leptons must arise through some other channel.

The combination of these issues makes the observed high-energy anomalies – especially ATIC and PAMELA – difficult to explain with thermal dark matter annihilation. However, we shall see that the inclusion of a new force in the dark sector simultaneously addresses all of these concerns.

New Forces in the Dark Sector

A new interaction for the dark sector can arise naturally in a variety of theories of physics beyond the standard model, and is thus well motivated from a theoretical point of view. Although there are strong limits on the self-interaction scattering cross section from structure formation [28, 29], the presence of *some* new force-carrying boson should be expected, with only the mass scale in question. A light boson could arise naturally if its mass scale is generated radiatively [30], or if it were a pseudo-goldstone boson.

One of the important modifications that can arise with a new light boson is an enhancement of the annihilation cross section via a mechanism first described by Sommerfeld [31]. The presence of a new force carrier can distort the wave function of the incoming particles away from the plane-wave approximation, yielding significant enhancements (or suppressions) to annihilation cross sections [64]. Equivalently, ladder diagrams involving multiple exchanges of the force carrier must be resummed (Fig. 1). As we shall describe, the Sommerfeld enhancement can only arise if the gauge boson has a mass $m_\phi \lesssim \alpha M_{DM} \sim$ few GeV. Thus, with the mass scale of $\sim 800\,\text{GeV}$ selected by ATIC, and the large cross sections needed by both ATIC and PAMELA, the mass scale for a new force carrier is automatically selected. Interactions involving W and Z bosons are insufficient at this mass scale.

Once this new force carrier ϕ is included, the possibility of a new annihilation channel $\chi\chi \to \phi\phi$ opens up, which can easily be the dominant annihilation channel. Absent couplings to the standard model, some of these particles could naturally be stable for kinematical reasons; even small interactions with the standard model can then lead them to decay only into standard model states. If they decay dominantly into leptons, then a hard spectrum of positrons arises very naturally. Motivated by the setup of XDM, Cholis, Goodenough and Weiner [19] first invoked this mechanism of annihilations into light bosons to provide a simple explanation for the excesses of cosmic ray positrons seen by HEAT without excessive antiprotons or photons. Simple kinematics can forbid a decay into heavier hadronic states, and as we shall see, scalars lighter than $\sim 250\,\text{MeV}$ and vectors lighter than $\sim$ GeV both provide a mode by which the dark matter can dominantly annihilate into very hard leptons, with few or no π^0's or antiprotons.

In the following sections, we shall make these points more concretely and delineate which ranges of parameters most easily explain the data; but the essential point is very simple: if the dark matter is $\mathcal{O}(800\,\text{GeV})$ and interacts with itself via a force carrier with mass $m_\phi \sim$ GeV, annihilation cross sections can be considerably enhanced at present times via a Sommerfeld enhancement, far exceeding the thermal freeze-out cross section. If that boson has a small mixing with the standard model, its mass scale can make it kinematically incapable of decaying via a hadronic shower, preferring muons, electrons and, in some cases, charged pions, and avoiding constraints from π^0's and antiprotons.

If the force-carriers are non-Abelian gauge bosons, we shall see that other anomalies may be incorporated naturally in this framework, explaining the INTEGRAL 511 keV line via the mechanism of “eXciting dark matter” (XDM), and the DAMA annual modulation signal via the mechanism of “inelastic dark matter” (iDM). We shall see that the excited states needed for both of these mechanisms [$\mathcal{O}(1\,\mathrm{MeV})$ for XDM and $\mathcal{O}(100\,\mathrm{keV})$ for iDM] arise naturally with the relevant mass splitting generated radiatively at the correct scale.

II. SOMMERFELD ENHANCEMENTS FROM NEW FORCES

A new force in the dark sector can give rise to the large annihilation cross sections required to explain recent data, through the “Sommerfeld enhancement” that increases the cross section at low velocities [65]. A simple classical analogy can be used to illustrate the effect. Consider a point particle impinging on a star of radius R. Neglecting gravity, the cross section for the particle to hit and be absorbed by the star is $\sigma_0 = \pi R^2$. However, because of gravity, a point coming from a larger impact parameter than R will be sucked into the star. The cross section is actually $\sigma = \pi b_{max}^2$, where b_{max} is the largest the impact parameter can be so that the distance of closest approach of the orbit is R. If the velocity of the particle at infinity is v, we can determine b_{max} trivially using conservation of energy and angular momentum, and we find that

$$\sigma = \sigma_0 \left(1 + \frac{v_{esc}^2}{v^2}\right) \tag{2}$$

where $v_{esc}^2 = 2G_N M/R$ is the escape velocity from the surface of the star. For $v \ll v_{esc}$, there is a large enhancement of the cross section due to gravity; even though the correction vanishes as gravity shuts off ($G_N \to 0$), the expansion parameter is $2G_N M/(Rv^2)$ which can become large at small velocity.

The Sommerfeld enhancement is a quantum counterpart to this classical phenomenon. It arises whenever a particle has an attractive force carrier with a Compton wavelength longer than $(\alpha M_{DM})^{-1}$, i.e. dark matter bound states are present in the spectrum of the theory. (We generically refer to $\alpha \sim \text{coupling}^2/4\pi$, assuming such couplings are comparable to those in the standard model, with $10^{-3} \lesssim \alpha \lesssim 10^{-1}$.)

Let us study this enhancement more quantitatively. We begin with the simplest case of interest, namely, a particle interacting via a Yukawa potential. We assume a dark matter particle χ coupling to a mediator ϕ with coupling strength λ. For s-wave annihilation in the nonrelativistic limit, the reduced two-body wavefunction obeys the radial Schrödinger equation,

$$\frac{1}{m_\chi}\psi''(r) - V(r)\psi(r) = -m_\chi v^2 \psi(r), \tag{3}$$

where the s-wave wavefunction $\Psi(r)$ is related to $\psi(r)$ as $\Psi(r) = \psi(r)/r$, v is the velocity of each particle in the center-of-mass frame (here we use units where $\hbar = c = 1$), and for scalar ϕ the potential takes the usual Yukawa form,

$$V(r) = -\frac{\lambda^2}{4\pi r} e^{-m_\phi r}. \tag{4}$$

The interaction in the absence of the potential is pointlike. As reviewed in the appendix, the Sommerfeld enhancement in the scattering cross section due to the potential is given by

$$S = \left| \frac{\frac{d\psi_k}{dr}(0)}{k} \right|^2 \tag{5}$$

where we solve the Schrödinger equation with boundary conditions $\psi(0) = 0$, $\psi(r) \to \sin(kr + \delta)$ as $r \to \infty$. In the recent dark matter literature, a different but completely equivalent expression is used, with

$$S = |\psi(\infty)/\psi(0)|^2 \tag{6}$$

where we solve the Schrödinger equation with the outgoing boundary condition $\psi'(\infty) = i m_\chi v \psi(\infty)$ [32].

Defining the dimensionless parameters

$$\alpha = \lambda^2/4\pi,\ \epsilon_v = \frac{v}{\alpha},\ \epsilon_\phi = \frac{m_\phi}{\alpha m_\chi}, \tag{7}$$

and rescaling the radial coordinate with $r' = \alpha m_\chi r$, we can rewrite Eq. 3 as,

$$\psi''(r') + \left(\epsilon_v^2 + \frac{1}{r'} e^{-\epsilon_\phi r'}\right)\psi(r') = 0. \tag{8}$$

In the limit where the ϕ mass goes to zero ($\epsilon_\phi \to 0$), the effective potential is just the Coulomb potential and Eq. 8 can be solved analytically, yielding an enhancement factor of,

$$S \equiv |\psi(\infty)/\psi(0)|^2 = \frac{\pi/\epsilon_v}{1 - e^{-\pi/\epsilon_v}}. \tag{9}$$

For nonzero m_ϕ and hence nonzero ϵ_ϕ, there are two important qualitative differences. The first is that the Sommerfeld enhancement saturates at low velocity–the attractive force has a finite range, and this limits how big the enhancement can get. At low velocities, once the deBroglie wavelength of the particle $(Mv)^{-1}$ gets larger than the range of the interaction m_ϕ^{-1}, or equivalently once ϵ_v drops beneath ϵ_ϕ, the Sommerfeld enhancement saturates at $S \sim \frac{1}{\epsilon_\phi}$. The second effect is that for specific values of ϵ_ϕ, the Yukawa potential develops threshold bound states, and these give rise to resonant enhancements of the Sommerfeld enhancement. We describe some of the parametrics for these effects in the appendix, but for reliable numbers Eq. 8 must be solved numerically, and plots for the enhancement as a function of ϵ_ϕ and ϵ_v are given there. As we will see in the following, we will be interested in a range of $m_\phi \sim 100$ MeV - GeV; with reasonable values of α, this corresponds to ϵ_ϕ in the range $\sim 10^{-2} - 10^{-1}$, yielding Sommerfeld enhancements ranging up to a factor $\sim 10^3 - 10^4$. At low velocities, the finite range of the Yukawa interaction causes the Sommerfeld enhancement to saturate, so the enhancement factor cannot greatly exceed this value even at arbitrarily low velocities. The nonzero mass of the ϕ thus prevents the catastrophic overproduction of gammas in the early universe pointed out by [33].

Having obtained the enhancement S as a function of ϵ_v and ϵ_ϕ, we must integrate over the velocity distribution of the dark matter in Earth's neighborhood, to obtain the total enhancement to the annihilation cross section for a particular choice of ϕ mass and coupling λ. We assume a Maxwell-Boltzmann distribution for the one-particle velocity, truncated at the escape velocity:

$$f(v) = \begin{cases} N v^2 e^{-v^2/2\sigma^2} & v \leq v_{\rm esc} \\ 0 & v > v_{\rm esc}. \end{cases} \tag{10}$$

The truncation does not significantly affect the results, as the enhancement factor drops rapidly with increasing velocity. The one-particle rms velocity is taken to be 150 km/s in the baseline case, following simulations by Governato et al. [34]. Fig. 1 shows the total enhancement as a function of m_ϕ/m_χ and the coupling λ for this case.

One can inquire as to whether dark matter annihilations in the early universe experience the same Sommerfeld enhancement as dark matter annihilations in the galactic halo at the present time. This is important, because we are relying on this effect precisely to provide us with an annihilation cross section in the present day much larger than that in the early universe. However, it turns out that particles leave thermal equilibrium long before the Sommerfeld enhancement turns on. This is because the Sommerfeld enhancement occurs when the expansion parameter $\alpha/v = 1/\epsilon_v$ is large. In the early universe, the dark matter typically decouples at $T_{\rm CMB} \sim m_\chi/20$, or $v \sim 0.3c$. Since we are taking $\alpha \lesssim 0.1$ generally, the Sommerfeld enhancement has not turned on yet. More precisely, in the early-universe regime $\epsilon_v \gg \epsilon_\phi$, so we can use Eq. 9 for the massless limit. Where $\epsilon_v \gg 1$, Eq. 9 yields $S \approx 1 + \pi/2\epsilon_v$: thus the enhancement should be small and independent of m_ϕ. Fig. 2 shows this explicitly. We are left with the perturbative annihilation cross section $\sigma \sim \alpha^2/m_\chi^2$ which gives us the usual successful thermal relic abundance.

At some later time, as the dark matter velocities redshift to lower values, the Sommerfeld enhancement turns on and the annihilations begin to scale as $a^{-5/2}$ (before kinetic decoupling) or a^{-2} (after decoupling). From decoupling until matter-radiation equality, where the Hubble scale begins to evolve differently, or until the Sommerfeld effect is saturated, dark matter annihilation will produce a uniform amount of energy per comoving volume per Hubble time.

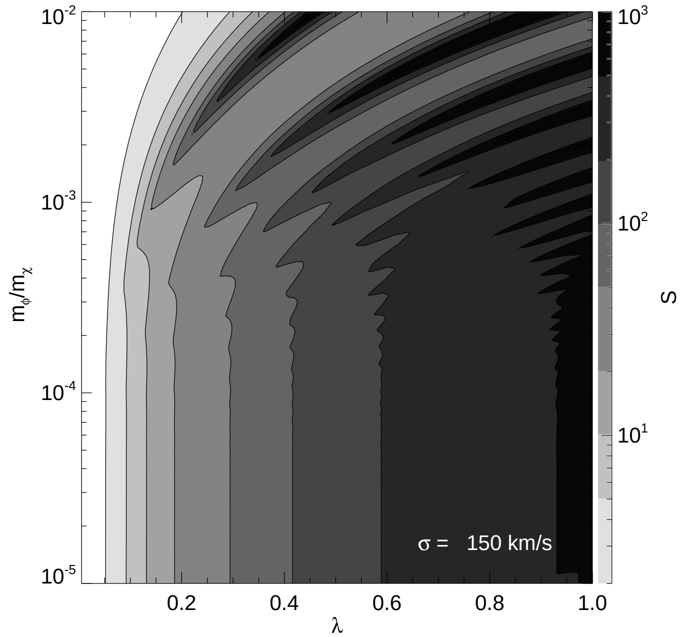

FIG. 1: Contours for the Sommerfeld enhancement factor S as a function of the mass ratio m_ϕ/m_χ and the coupling constant λ, $\sigma = 150$ km/s.

This uniform spread of energy injected could have potentially interesting signals for observations of the early universe. An obvious example would be a possible effect on the polarization of the CMB, as described in [35, 36, 37]. Because at the time of matter-radiation equality, the dark matter may have slowed to velocities of $v \sim 10^{-6}c$ or slower, the large cross sections could yield a promising signal for upcoming CMB polarization observations, including *Planck*. However, we emphasize that saturation of the cross section at low v avoids the runaway annihilations discussed by [33].

III. MODELS OF THE SOMMERFELD FORCE AND NEW ANNIHILATION CHANNELS

What sorts of forces could give rise to a large Sommerfeld enhancement of the dark matter annihilation? As we have already discussed, we must have a *light* force carrier. On the other hand, a massless particle is disfavored by the agreement between big bang nucleosynthesis calculations and primordial light element measurements [38], as well as constraints from WMAP on new relativistic degrees of freedom [39]. Thus, we must have massive degrees of freedom, which are naturally light, while still coupling significantly to the dark matter. There are three basic candidates.

- The simplest possibility is coupling to a light scalar field, which does give rise to an attractive interaction. However, given that we need an $\mathcal{O}(1)$ coupling to the DM fields, this will typically make it unnatural for the scalar to stay as light as is needed to maximize the Sommerfeld enhancement, unless the dark matter sector is very supersymmetric. This can be a challenge given that we are expecting the dark matter to have a mass $\mathcal{O}(\text{TeV})$. Consequently, the natural scale for a scalar which couples to it would also be $\mathcal{O}(\text{TeV})$, although this conclusion can be evaded with some simple model-building.

- The scalar could be naturally light if it is a pseudoscalar π with a goldstone-like derivative coupling to matter $1/F J_\mu \partial^\mu \pi$. This does lead to a long-range spin-dependent potential of the form $V(\vec{r}) = \frac{1}{r^3}(\vec{S}_1 \cdot \vec{S}_2 - 3\vec{S}_1 \cdot \hat{r}\vec{S}_2 \cdot \hat{r})$, but the numerator vanishes when averaged over angles, so there is no long-range interaction in the s-wave and hence no Sommerfeld enhancement.

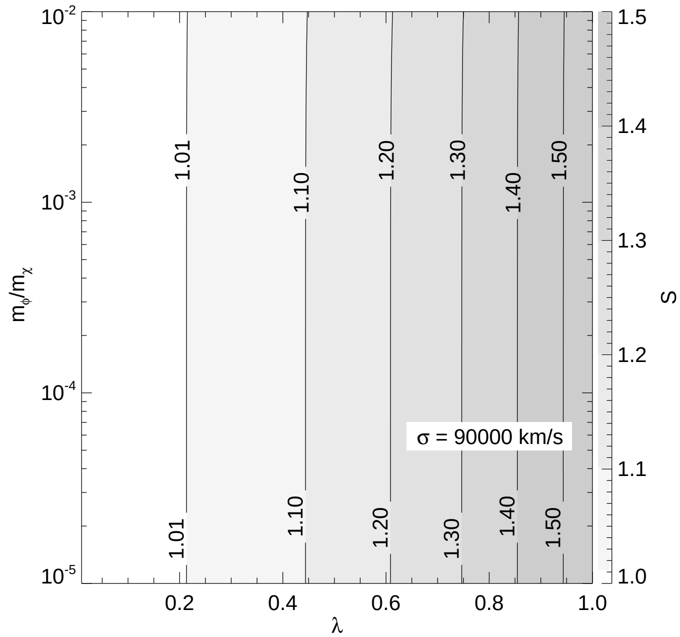

FIG. 2: Contours for the Sommerfeld enhancement factor S as a function of the mass ratio m_ϕ/m_χ and the coupling constant λ, at a temperature $T_{\rm CMB} = m_\chi/20$ (the enhancement is integrated over a Boltzmann distribution with $\sigma = 0.3c$).

- Finally, we can have a coupling to spin-1 gauge fields arising from some dark gauge symmetry $G_{\rm dark}$. Since the gauge fields must have a mass $\mathcal{O}$(GeV) or less, one might worry that this simply begs the question, as the usual explanation of such a light gauge boson requires the existence of a scalar with a mass of $\mathcal{O}$(GeV) or less. However, because that scalar need not couple directly to the dark matter, it is sufficiently sequestered that its small mass is technically natural. Indeed, the most straightforward embedding of this scenario within SUSY [30] naturally predicts the breaking scale for $G_{\rm dark}$ near $\sim$ GeV. Alternatively, no fundamental scalar is needed, with the vector boson possibly generated by the condensate of a strongly coupled theory.

At this juncture it is worth discussing an important point. As we have emphasized, to produce a Sommerfeld *enhancement*, we need an attractive interaction. Scalars (like gravitons) universally mediate attractive forces, but gauge fields can give attraction or repulsion, so do we generically get a Sommerfeld enhancement? For the case of the Majorana fermion (or real scalar) in particular, the dark matter does not carry any charge, and there does not appear to be any long-range force to speak of, so the question of the enhancement is more interesting. As we discuss in a little more detail in the appendix, the point is that the gauge symmetry is broken. The breaking dominates the properties of the asymptotic states. For instance, the dark matter must be part of a multiplet with at least two states, since a spin-1 particle cannot have a coupling to a single neutral state. The gauge symmetry breaking leads to a mass splitting between the states, which dominates the long-distance behavior of the theory, determining which state is the lightest, and therefore able to survive to the present time and serve as initial states for collisions leading to annihilation. However, if the mass splitting between the states is small enough compared to the kinetic energy of the collision, the gauge-partner DM states will necessarily be active in the collision, and eventually at distances smaller than the gauge boson masses, the gauge breaking is a negligible effect. Since the asymptotic states are in general roughly equal linear combinations of positive and negative charge gauge eigenstates, the remnant of asymptotic gauge breaking at short distances is that the incoming scattering states are linear combinations of gauge eigenstates, so the two-body wavefunction will be a linear combination of attractive and repulsive channels. While two-body wavefunction in the repulsive channel is indeed suppressed at the origin, the attractive part is enhanced. Therefore there is still a Sommerfeld enhancement, suppressed at most by an $\mathcal{O}(1)$ factor reflecting the attractive component of the two-body state. (This is of course completely consistent with the Sommerfeld enhancements seen

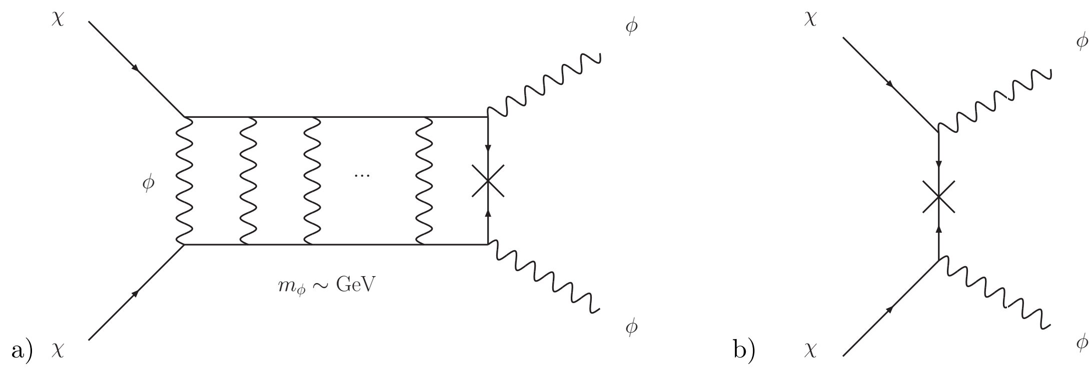

FIG. 3: The annihilation diagrams $\chi\chi \to \phi\phi$ both with (a) and without (b) the Sommerfeld enhancements.

for ordinary WIMP annihilations, mediated by W/Z/γ exchange).

Because of the presence of a new light state, the annihilation $\chi\chi \to \phi\phi$ can, and naturally will, be significant. In order not to spoil the success of nucleosynthesis, we cannot have very light new states in this sector, with a mass $\lesssim 10$ MeV, in thermal equilibrium with the standard model; the simplest picture is therefore that all the light states in the dark sector have a mass $\sim$ GeV. Without any special symmetries, there is no reason for any of these particles to be exactly stable, and the lightest ones can therefore only decay back to standard model states, indeed many SM states are also likely kinematically inaccessible, thus favoring ones that produce high energy positrons and electrons. This mechanism was first utilized in [19] to generate a large positron signal with smaller π^0 and $\bar{p}$ signals. Consequently, an important question is the tendency of ϕ to decay to leptons. This is a simple matter of how ϕ couples to the standard model. (A more detailed discussion of this can be found in [30].)

A scalar ϕ can couple with a dilaton-like coupling $\phi F^{\mu\nu} F_{\mu\nu}$, which will produce photons and hadrons (via gluons). Such a possibility will generally fail to produce a hard e^+e^- spectrum. A more promising approach would be to mix ϕ with the standard model Higgs with a term $\kappa\phi^2 h^\dagger h$. Should ϕ acquire a vev $\langle\phi\rangle \sim m_\phi$, then we yield a small mixing with the standard model Higgs, and the ϕ will decay into the heaviest fermion pair available. For $m_\phi \lesssim 200\,\mathrm{MeV}$ it will decay directly to e^+e^-, while for $200\,\mathrm{MeV} \lesssim \mathrm{m}_\phi \lesssim 250\,\mathrm{MeV}$, ϕ will decay dominantly to muons. Above that hadronic states appear, and pion modes will dominate. Both e^+e^- and $\mu^+\mu^-$ give good fits to the PAMELA data, while e^+e^- gives a better fit to PAMELA+ATIC.

A pseudoscalar, while not yielding a Sommerfeld enhancement, could naturally be present in this new sector. Such a particle would typically couple to the heaviest particle available, or through the axion analog of the dilaton coupling above. Consequently, the decays of a pseudoscalar would be similar to those of the scalar.

A vector boson will naturally mix with electromagnetism via the operator $F'_{\mu\nu} F^{\mu\nu}$. This possibility was considered some time ago in [40]. Such an operator will cause a vector ϕ_μ to couple directly to charge. Thus, for $m_\phi \lesssim 2m_\mu$ it will decay to e^+e^-, while for $2m_\mu \lesssim m_\phi \lesssim 2m_\pi$ it will decay equally to e^+e^- and $\mu^+\mu^-$. Above $2m_\pi$, it will decay $40\%\, e^+e^-$, $40\%\, \mu^+\mu^-$ and $20\%\, \pi^+\pi^-$. At these masses, no direct decays into π^0's will occur because they are neutral and the hadrons are the appropriate degrees of freedom. At higher masses, where quarks and QCD are the appropriate degrees of freedom, the ϕ will decay to quarks, producing a wider range of hadronic states, including π^0's, and, at suitably high masses $m_\phi \gtrsim 2\,\mathrm{GeV}$, antiprotons as well [66]. In addition to XDM [18], some other important examples of theories under which dark matter interacts with new forces include WIMPless models [41], mirror dark matter [42] and secluded dark matter [43].

Note that, while these interactions between the sectors can be small, they are all large enough to keep the dark and standard model sectors in thermal equilibrium down to temperatures far beneath the dark matter mass, and (as mentioned in the previous section), we can naturally get the correct thermal relic abundance with a weak-scale dark matter mass and perturbative annihilation cross sections. Kinetic equilibrium in these models is naturally maintained down to the temperature $T_{\mathrm{CMB}} \sim m_\phi$ [44].

IV. A NON-ABELIAN $G_{\rm dark}$: INTEGRAL, DIRECT DETECTION, AND DAMA

Up to this point we have focused on a situation where there is a single force-carrying boson ϕ, whether vector or scalar. Already, this can have significant phenomenological consequences. In mixing with the standard model Higgs boson, there is a nuclear recoil cross section mediated by ϕ. With technically natural parameters as described in [18], the rate is unobservable, although in a two-Higgs doublet model the cross section is within reach of future experiments [45].

In contrast, an 800 GeV WIMP which interacts via a particle that couples to charge is strongly constrained. Because the ϕ boson is light and couples to the electromagnetic vector current, there are strong limits. The cross section per nucleon for such a particle is [43]

$$\begin{aligned} \sigma_0 &= \frac{16\pi Z^2 \alpha_{\rm SM}\alpha_{\rm Dark}\epsilon^2\mu_{\rm ne}^2}{A^2 m_\phi^4} \\ &= \left(\frac{Z}{32}\right)^2\left(\frac{73}{A}\right)^2\left(\frac{\epsilon}{10^{-3}}\right)^2\left(\frac{\alpha_{\rm Dark}}{137^{-1}}\right)\left(\frac{\mu_{\rm ne}}{938\,{\rm MeV}}\right)\left(\frac{1\ {\rm GeV}}{m_\phi}\right)^4 \times 1.8\times 10^{-37}\,{\rm cm}^2, \end{aligned} \tag{11}$$

where $\alpha_{\rm Dark}$ is the coupling of the ϕ to the dark matter, ϵ describes the kinetic mixing, $\mu_{\rm ne}$ is the reduced mass of the DM-nucleon system and $\alpha_{\rm SM}$ is the standard model electromagnetic coupling constant. With the parameters above, such a scattering cross section is excluded by the present CDMS [46] and XENON [47] bounds by 6 orders of magnitude. However, this limit can be evaded by splitting the two Majorana components of the Dirac fermion [22] or by splitting the scalar and pseudoscalar components of a complex scalar [48, 49]. Since the vector coupling is off-diagonal between these states, the nuclear recoil can only occur if there is sufficient kinetic energy to do so. If the splitting $\delta > v^2\mu/2$ (where μ is the reduced mass of the WIMP-nucleus system) no scattering will occur. Such a splitting can easily arise for a $U(1)$ symmetry by a $U(1)$ breaking operator such as $\frac{1}{M}\psi^c\psi h^* h^*$ which generates a small Majorana mass and splits the two components (see [23] for a discussion).

Remarkably, for $\delta \sim 100\,{\rm keV}$ one can reconcile the DAMA annual modulation signature with the null results of other experiments [22, 23, 24], in the "inelastic dark matter" scenario. We find that the ingredients for such a scenario occur here quite naturally. However, the splitting here must be $\mathcal{O}(100\,{\rm keV})$ and the origin of this scale is unknown, a point we shall address shortly.

Exciting Dark Matter from a Non-Abelian Symmetry

One of the strongest motivations for a $\sim$ GeV mass ϕ particle prior to the present ATIC and PAMELA data was in the context of eXciting dark matter [18]. In this scenario, dark matter excitations could occur in the center of the galaxy via inelastic scattering $\chi\chi \to \chi\chi^*$. If $\delta = m_{\chi^*} - m_\chi \gtrsim 2m_e$, the decay $\chi^* \to \chi e^+ e^-$ can generate the excess of 511 keV x-rays seen from the galactic center by the INTEGRAL [50, 51] satellite. However, a large (nearly geometric) cross section is needed to produce the large numbers of positrons observed in the galactic center, necessitating a boson with mass of the order of the momentum transfer, i.e., $m_\phi \lesssim M_\chi v \sim$ GeV, precisely the same scale as we require for the Sommerfeld enhancement. But where does the scale $\delta \sim$ MeV come from? Remarkably, it arises radiatively at precisely the appropriate scale[67].

We need the dark matter to have an excited state, and we will further assume the dark matter transforms under a non-Abelian gauge symmetry. Although an excited state can be present with simply a $U(1)$, this only mediates the process $\chi\chi \to \chi^*\chi^*$. If this requires energy greater than $4m_e$ it is very hard to generate enough positrons to explain the INTEGRAL signal. If we assume the dark matter is a Majorana fermion, then it must transform as a real representation of the gauge symmetry. For a non-Abelian symmetry, the smallest such representation will be three-dimensional [such as a triplet of SU(2)]. This will allow a scattering $\chi_1\chi_1 \to \chi_2\chi_3$. If m_3 is split from $m_2 \sim m_1$ by an amount $\delta \sim$ MeV, we have arrived at the setup for the XDM explanation of the INTEGRAL signal.

Because the gauge symmetry is Higgsed, we should expect a splitting between different states in the dark matter multiplet. This could arise already at tree-level, if the dark matter has direct couplings to the Higgs fields breaking the gauge symmetry; these could naturally be as large as the dark gauge breaking scale $\sim$ GeV itself, which would be highly undesirable, since we need these splittings to be not much larger than the DM kinetic energies in order to get a Sommerfeld enhancement to begin with. However, such direct couplings to the Higgs could be absent or very small (indeed most of the Yukawa couplings in the standard model are very small). We will assume that such a direct coupling is absent or negligible. The gauge breaking in the gauge boson masses then leads, at one loop, to splittings between different dark matter states, analogous to the familiar splitting between charged and neutral components of a Higgsino or Dirac neutrino. These splittings arise from infrared effects, and so are completely calculable, with sizes generically $\mathcal{O}(\alpha m_Z)$ in the standard model or $\mathcal{O}(\alpha m_\phi) \sim$ MeV in the case at hand. Thus we find that the MeV splittings needed for XDM arise *automatically* once the mass scale of the ϕ has been set to $\mathcal{O}(GeV)$.

We would like χ_2 and χ_1 to stay similar in mass, which can occur if the breaking pattern approximately preserves a custodial symmetry. However, if they are too degenerate, we are forced to take $\epsilon < 10^{-5}$ in order to escape direct-detection constraints. On the other hand, we do not want *too* large of a splitting between χ_2 and χ_1, as this would suppress the rate of positron production for INTEGRAL. Thus, we are compelled to consider $\delta_{21} \sim 100 - 200\,\mathrm{keV}$, which puts us precisely in the range relevant for the inelastic dark matter explanation of DAMA. (See Fig. 4.)

All of these issues require detailed model-building, which we defer to future work; however, existence proofs are easy to construct. The Lagrangian is of the form

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \mathcal{L}_{\rm Dark} + \mathcal{L}_{\rm mix}. \tag{12}$$

As a familiar example, imagine that $G_{\rm Dark} = SU(2) \times U(1)$, with gauge bosons $w_{\mu I}$ and b_μ which we collectively refer to as $a_{\mu i}$, and the dark matter multiplet χ transforming as a triplet under $SU(2)$ and neutral under the $U(1)$; it could be either a scalar or fermion. We also assume that some set of Higgses completely break the symmetry. Working in unitary gauge, the tree-level dark sector Lagrangian is

$$\mathcal{L}_{\rm Dark} = \mathcal{L}_{\rm Gauge\ Kin.} + \frac{1}{2} m_{ij}^2 a_i^\mu a_{\mu j} + \cdots \tag{13}$$

where the m_{ij}^2 makes all the dark spin-1 fields massive, and $\cdots$ refers to other fields such as the physical Higgses that could be present. At one loop, this broken gauge symmetry will induce splittings between the 3 real DM states all of $\mathcal{O}(\alpha_{Dark} m_{Dark})$ as just discussed above. The leading interaction between the two sectors is via kinetic mixing between the new $U(1)$ and the photon (which is inherited from such a mixing with hypercharge):

$$\mathcal{L}_{\rm mix} = \frac{1}{2} \epsilon b_{\mu\nu} F^{\mu\nu} \tag{14}$$

We put an ϵ in front of this coupling because it is natural for this coupling to be small; it can be induced at one loop by integrating out some heavy states (of any mass between the GeV and Planck scales) charged under both this new $U(1)$ and hypercharge. This can easily make $\epsilon \sim 10^{-4} - 10^{-3}$. Even without a $U(1)$, a similar mixing could be achieved with an "S-parameter" type operator ${\rm Tr}[(\Phi/M)^p G_{\mu\nu}] F_{\mu\nu}$, where Φ is a dark Higgs field with quantum numbers such that $(\Phi/M)^p$ transforms as an adjoint under $G_{\rm Dark}$; it is reasonable to imagine that the scale M suppressing this operator is near the weak scale.

Going back to PAMELA/ATIC, the non-Abelian self-couplings of the vector bosons can have an interesting effect on the annihilation process. For large enough $\alpha_{\rm Dark}$, the gauge bosons radiate other soft and collinear gauge bosons leading to a "shower"; this can happen when $\alpha_{\rm Dark} \log^2(M_\chi/m_\phi) \gtrsim 1$. While for quite perturbative values of $\alpha_{\rm Dark}$ this is not an important effect, it could be interesting for larger values, and would lead to a greater multiplicity of softer e^+e^- pairs in the final state.

Before closing this section, it is important to point out that it is not merely a numerical accident that the excited states relevant for INTEGRAL and DAMA can actually be excited in the DM-DM collisions and DM collisions with direct-detection nuclei, but is rather a parametric consequence of maximizing the Sommerfeld enhancement for the annihilation cross section needed to explain ATIC/PAMELA. As we already emphasized in our discussion of Sommerfeld enhancement with vector states, vector boson couplings necessarily connect different dark matter mass eigenstates, and therefore there is no enhancement for the annihilation cross section needed to explain ATIC/PAMELA if the mass splittings are much larger than the kinetic energy available in the collision. But this parametrically implies that dark matter collisions should also have the kinetic energy needed to create the excited states, as necessary for the INTEGRAL signal. It is also interesting to note that the condition needed for the large geometric capture cross

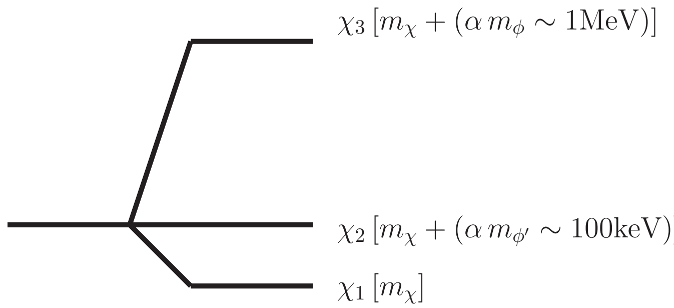

FIG. 4: Spectrum of exciting dark matter.

section, $m_\phi \lesssim Mv$, also tells us that the Sommerfeld enhancement is as large as can be at these velocities, and has not yet been saturated by the finite range of the force carrier. Furthermore, since the mass of the heavy nuclei in direct detection experiments is comparable to the dark matter masses, the WIMP-nucleus kinetic energy is also naturally comparable to the excited state splittings. In this sense, even absent the direct experimental hints, signals like those of INTEGRAL and inelastic scattering for direct detection of dark matter are parametric predictions of our picture.

V. SUBSTRUCTURE AND THE SOMMERFELD ENHANCEMENT

As we have seen, the Sommerfeld enhancement leads to a cross section that scales at low energies as $\sigma v \sim 1/v$. This results in a relatively higher contribution to the dark matter annihilation from low velocity particles. While the largest part of our halo is composed of dark matter particles with an approximately thermal distribution, there are subhalos with comparable or higher densities. Because these structures generally have lower velocity dispersions than the approximately thermal bulk of the halo, the Sommerfeld-enhanced cross sections can make these components especially important.

Subhalos of the Milky Way halo are of particular interest, and N-body simulations predict that many should be present. There is still some debate as to what effect substructures can have on indirect detection prospects. However, some of these subhalos will have velocity dispersions of order 10 km/s [52], and a simple examination of Fig. 5 shows that dramatic increases of up to two orders of magnitude in the Sommerfeld enhancement can occur for these lower velocities!

Although in our model there are no direct π^0 gammas, there are still significant hopes for detection of dwarf galaxies through dark matter annihilation. If the DM is also charged under $SU(2) \times U(1)$ there may be a subdominant component of annihilations into W^+W^- for instance, much larger than the s-wave limited thermal cross section, which could yield significant photon signals from the hadronic shower[68] . The copious high energy positrons and electrons, produced even more abundantly than expected for a nonthermal WIMP generating the ATIC or PAMELA signal, can produce inverse Compton scattering signals off of the CMB. Even a loop suppressed annihilation into $\gamma\gamma$ may be detectable with this enhancement.

This presumes that the local annihilation does not have significant enhancement from a low-velocity component, however. If the local annihilation rate is also enhanced by substructure, then our expectations of the enhancement for dwarf galaxies would not be as large. As a concrete example, we can consider the possibility of a “dark disk.” Recently, it has been argued that the old stars in the thick Milky Way disk should have an associated dark matter component, with dynamics which mirror those stars, and a density comparable to the local density [53]. Because of the low velocity dispersion (of order 10 km/s), our estimates of what is a reasonable Sommerfeld boost may be off by an order of magnitude.

As a consequence of this and other substructure, the Sommerfeld boost of Fig. 1 should be taken as a lower bound, with contributions from substructure likely increasing the cross section significantly further. We will not attempt to quantify these effects here beyond noting the possible increases of an order of magnitude or more suggested by Fig.

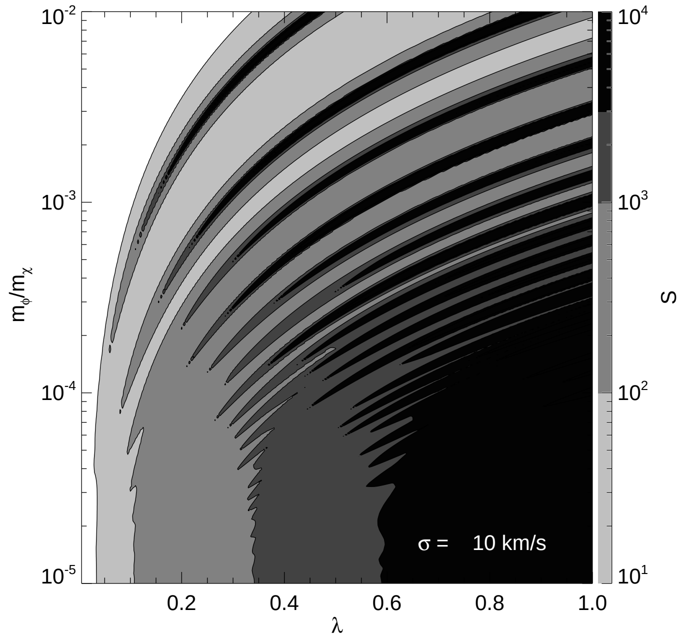

FIG. 5: Contours for the Sommerfeld enhancement factor S as a function of the mass ratio m_ϕ/m_χ and the coupling constant λ, $\sigma = 10$ km/s.

5. However, understanding them may be essential for understanding the size and spectrum if this enhancement of the cross section is responsible for the signals we observe.

VI. OUTLOOK: IMPLICATIONS FOR PAMELA, PLANCK, FERMI/GLAST, AND THE LHC

If the excesses in positrons and electrons seen by PAMELA and ATIC are arising from dark matter, there are important implications for a wide variety of experiments. It appears at this point that a simple modification of a standard candidate such as a neutralino in the minimal supersymmetric standard model (MSSM) is insufficient. The need for dominantly leptonic annihilation modes with large cross sections significantly changes our intuition for what we might see, where, and at what level. There are a few clear consequences looking forward.

- The positron fraction seen by PAMELA should continue to rise up to the highest energies available to them (~ 270 GeV), and the electron + positron signal should deviate from a simple power law, as seen by ATIC.
- If the PAMELA/ATIC signal came from a local source, we would not expect additional anomalous electronic activity elsewhere in the galaxy. Dark matter, on the other hand, should produce a significant signal in the center of the galaxy as well, yielding significant signals in the microwave range through synchrotron radiation and in gamma rays through inverse-Compton scattering. The former may already have been seen at WMAP [10, 12] and the latter at EGRET [13, 14]. Additional data from Planck and Fermi/GLAST will make these signals robust [54].
- We have argued that the most natural way to generate such a large signal is through the presence of a new, light state, which decays dominantly to leptons. It is likely these states could naturally be produced at the LHC in some cascade, leading to highly boosted pairs of leptons as a generic signature of this scenario [30].
- Although the cross section is not Sommerfeld enhanced during freeze-out, it can keep pace with the expansion rate over large periods of the cosmic history, between kinetic decoupling and matter-radiation equality. This

can have significant implications for a variety of early-universe phenomena as well as the cosmic gamma-ray background [54].

- The Sommerfeld enhancement is increasingly important at low velocity. Because substructure in the halo typically has velocity dispersions an order of magnitude smaller than the bulk of the halo, annihilations can be an order of magnitude higher, or more. With the already high local cross sections, this makes the prospects for detecting substructure even higher. With the mass range in question, continuum photons would be possibly visible at GLAST, while monochromatic photons [which could be generated in some models [30] would be accessible to air Cerenkov telescopes, such as HESS (see [55] for a discussion, and [56] for HESS limits on DM annihilation in the Canis Major overdensity)].

We have argued that dark matter physics is far richer than usually thought, involving a multiplet of states and a new sector of dark forces. We have been led to propose this picture not by a flight of fancy but rather directly from experimental data. Even so, one can justifiably ask whether such extravagances are warranted. After all, experimental anomalies come and go, and it is entirely possible that the suite of hints that motivate our proposal are incorrect, or that they have more conventional explanations. However, we are very encouraged by the fact that the theory we have presented fits into a very reasonable picture of particle physics, is supported by overlapping pieces of experimental evidence, and that features of the theory motivated by one set of experimental anomalies automatically provide the ingredients to explain the others. Our focus in this paper has been on outlining this unified picture for dark matter; new experimental results coming soon should be able to tell us whether these ideas are even qualitatively on the right track. Needless to say it will then be important to find a specific and simple version of this theory, with a small number of parameters, to more quantitatively confront future data.

Acknowledgements

We would like to thank Ilias Cholis, Lisa Goodenough, Peter Graham, Juan Maldacena, Patrick Meade, Michele Papucci, Nati Seiberg, Leonardo Senatore, David Shih, Tomer Volansky, and Yosi Gelfand for many stimulating discussions. We thank Matias Zaldarriaga for an enjoyable discussion about classical and quantum Sommerfeld effects, and David Shih for pointing out an error in our discussion of the parametrics of the Sommerfeld effect for the Yukawa potential in v.1 of this paper, though our result is unchanged. We also thank Matt Strassler, Patrick Meade, and Tomer Volansky for a discussion of showering in annihilations to the non-Abelian G_{Dark} gauge bosons. The work of N.A.-H. is supported by the DOE under Grant No. DE-FG02-91ER40654. NW is supported by NSF CAREER Grant No. PHY-0449818 and DOE OJI Grant No. DE-FG02-06ER41417.

APPENDIX A: A QUICK REVIEW OF SOMMERFELD ENHANCEMENT

The Sommerfeld enhancement is an elementary effect in nonrelativistic quantum mechanics; in this appendix we will review it in a simple way and discuss some of the parametrics for how the enhancement works for various kinds of interactions.

Consider a nonrelativistic particle moving around some origin. There is an interaction Hamiltonian $H_{ann} = U_{ann}\delta^3(\vec{r})$ localized to the origin, which e.g. annihilates our particle or converts to another state in some way. Imagine that the particle is moving in the z direction so that its wavefunction is

$$\psi_k^{(0)}(\vec{x}) = e^{ikz} \tag{A1}$$

then the rate for this process is proportional to $|\psi^{(0)}(0)|^2$. But now suppose we also have a central potential $V(r)$ attracting or repelling the particle to the origin. We could of course treat V perturbatively, but again at small velocities the potential may not be a small perturbation and can significantly distort the wave-function, which can be determined by solving the Schrödinger equation

$$-\frac{1}{2M}\nabla^2\psi_k + V(r)\psi_k = \frac{k^2}{2M}\psi_k \tag{A2}$$

with the boundary condition enforcing that the perturbation can only produce outgoing spherical waves as $r \to \infty$:

$$\psi \to e^{ikz} + f(\theta)\frac{e^{ikr}}{r} \,\text{as}\, r \to \infty \tag{A3}$$

Now, since the annihilation is taking place locally near $r = 0$, the only effect of the perturbation V is to change the value of the modulus of the wave-function at the origin relative to its unperturbed value. Then, we can write

$$\sigma = \sigma_0 S_k \tag{A4}$$

where the Sommerfeld enhancement factor S is simply

$$S_k = \frac{|\psi_k(0)|^2}{|\psi_k^{(0)}(0)|^2} = |\psi_k(0)|^2 \tag{A5}$$

where we are using the normalization of the wavefunction ψ_k as given by the asymptotic form of Eq. A3.

Finding a solution of the Schrödinger equation with these asymptotics is a completely elementary and standard part of scattering theory in nonrelativistic QM, which we quickly review for the sake of completeness. Any solution of the Schrödinger equation with rotational invariance around the z axis can be expanded as

$$\psi_k = \sum_l A_l P_l(\cos\theta) R_{kl}(r) \tag{A6}$$

where $R_{kl}(r)$ are the continuum radial functions associated with angular momentum l satisfying

$$-\frac{1}{2M}\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}R_{kl}\right) + \left(\frac{l(l+1)}{r^2} + V(r)\right)R_{kl} = \frac{k^2}{2M}R_{kl} \tag{A7}$$

The $R_{kl}(r)$ are real, and at infinity look like a spherical plane wave which we can choose to normalize as

$$R_{kl}(r) \to \frac{1}{r}\sin(kr - \frac{1}{2}l\pi + \delta_l(r)) \tag{A8}$$

where $\delta_l(r) \ll kr$ as $r \to \infty$. The phase shift $\delta_l(r)$ is determined by the requirement that $R_{kl}(r)$ is regular as $r \to 0$. Indeed, if the potential $V(r)$ does not blow up faster than $1/r$ near $r \to 0$, then we can ignore it relative to the kinetic terms, and we have that $R_{kl}(r) \sim r^l$ as $r \to 0$; all but the $l = 0$ terms vanish at the origin. We now have to choose the coefficients A_l in order to ensure that the asymptotics of Eq. A3 are satisfied. Using the asymptotic expansion of e^{ikz}

$$e^{ikz} \to \frac{1}{2ikr}\sum_l (2l+1)P_l(\cos\theta)\left[e^{ikr} - (-1)^l e^{-ikr}\right] \tag{A9}$$

determines the expansion to be

$$\psi_k = \frac{1}{k}\sum_l i^l(2l+1)e^{i\delta_l}P_l(\cos\theta)R_{kl}(r) \tag{A10}$$

It is now very simple to determine $\psi_k(0)$, since as we just commented, $R_{kl}(r = 0)$ vanishes for every term other than $l = 0$. Thus, we have

$$S_k = \left|\frac{R_{k,l=0}(0)}{k}\right|^2 \tag{A11}$$

We can furthermore make the standard substitution $R_{k,l=0} = \chi_k/r$, then the Schrödinger equation for χ turns into a one-dimensional problem

$$-\frac{1}{2M}\frac{d^2}{dr^2}\chi_k + V(r)\chi_k = \frac{k^2}{2M}\chi_k \tag{A12}$$

which we normalize at infinity with the condition

$$\chi_k(r) \to \sin(kr + \delta) \tag{A13}$$

and since $R_{k,l=0}$ goes to a constant as $r \to 0$, we have to have that $\chi \to 0$ as $r \to 0$, or

$$\chi_k(r) \to r\frac{d\chi_k}{dr}(0) \text{ as } r \to 0 \tag{A14}$$

Effectively, we can imagine launching χ from $\chi = 0$ at $r = 0$ with different velocities $\frac{d\chi_k}{dr}(0)$, and these will evolve to some waveform as $r \to \infty$, but the correct $\chi'(0)$ is determined by the requirement that the waveform at infinity have unit amplitude.

Summarizing, then, the Sommerfeld enhancement is

$$S_k = \left|\frac{\frac{d\chi_k}{dr}(0)}{k}\right|^2 \tag{A15}$$

where χ_k satisfies the 1D Schrödinger equation A12 with boundary conditions Eqs.A13, A14. As a sanity check, let us see how this works with vanishing potential. The solution that vanishes as $r \to 0$ is $\chi_k(r) = A\sin(kr)$, and matching the asymptotics forces $A = 1$. Then $\chi_k'(0) = k$ and $S_k = 1$.

In the recent literature on the subject, a different expression for the Sommerfeld enhancement is used, arising from the use of the optical theorem. We are instructed to solve the same 1D Schrödinger equation (Eq. A12), this time with no special boundary conditions at $\chi = 0$, but with boundary conditions so that $\chi \propto e^{+ikr}$ as $r \to \infty$. Then, the Sommerfeld enhancement is said to be

$$S_k = \frac{|\chi_k(\infty)|^2}{|\chi_k(0)|^2} \tag{A16}$$

It is very easy to show that that these two forms for S_k agree exactly. To see this, let us begin by denoting $\chi_1(r)$ to be the solution to the Schrödinger equation (Eq. A12) with the boundary condition $\chi_1(r) \to \sin(kr + \delta)$ as $r \to \infty$ (Eq. A13). As shown above, $\chi_1(0) = 0$. Let $\chi_2(r)$ be the linearly independent solution with the boundary condition $\chi_2(r) \to \cos(kr + \delta)$ as $r \to \infty$, and define $A \equiv \chi_2(0)$. Now Eq. A12 has a conserved Wronskian

$$W = \chi_1(r)\chi_2'(r) - \chi_2(r)\chi_1'(r). \tag{A17}$$

It is easy to verify directly from the differential equation that $W'(r) = 0$; this is true (Abel's theorem) because there are no χ' terms in the differential equation. But comparing the values of the conserved Wronskian at zero and ∞,

$$W(\infty) = -k(\sin^2(kr + \delta) + \cos^2(kr + \delta)) = -k = W(0) = -A\chi_1'(0). \tag{A18}$$

So then $|\chi_1'(0)| = k/|A|$, and our new expression for the Sommerfeld enhancement, Eq. A15, is just $S_k = 1/|A|^2$. Now, our second form $S_k = |\chi(\infty)|^2$, where $\chi(r)$ satisfies the boundary conditions $\chi'(r) \to ik\chi(r)$ as $r \to \infty$, and $\chi(0) = 1$. By the asymptotic behavior at large r, we can identify $\chi(r)$ as the linear combination $\chi(r) = C(\chi_2(r) + i\chi_1(r))$, where C is some complex constant. But then $\chi(0) = CA = 1$, and $S_k = |C|^2$, so as previously we obtain $S_k = 1/|A|^2$. Thus the two formulae for the Sommerfeld enhancement are equivalent.

1. Attractive Coulomb Potential

Let us see how this works in some simple examples. We are solving the Schrödinger equation for a particle of mass M and asymptotic velocity v, with potential

$$V(r) = -\frac{\alpha}{2r} \tag{A19}$$

which we solve with the boundary conditions of Eqs. A13,A14.

We simplify the analysis by working with the natural dimensionless variables, with the unit of length normalized to the Bohr radius, i.e. we take

$$r = \alpha^{-1}M^{-1}x \tag{A20}$$

Then the Schrödinger equation becomes

$$-\chi'' - \frac{1}{x}\chi = \epsilon^2\chi \tag{A21}$$

where we have defined the parameter

$$\epsilon_v \equiv \frac{v}{\alpha} \tag{A22}$$

Of course we can solve this exactly in terms of hypergeometric functions to find the result obtained by Sommerfeld

$$S_k = \left|\frac{\frac{\pi}{\epsilon_v}}{1 - e^{-\frac{\pi}{\epsilon_v}}}\right| \tag{A23}$$

Note that as $\epsilon_v \to \infty$, $S_k \to 1$ as expected; there is no enhancement at large velocity. For the attractive Yukawa at small velocities we have the enhancement

$$S_k \to \frac{\pi\alpha}{v} \tag{A24}$$

while for the repulsive case, there is instead the expected exponential suppression from the need to tunnel through the Coulomb barrier

$$S_k \sim e^{-\frac{\pi|\alpha|}{v}} \tag{A25}$$

To get some simple insight into what is going on, let us re-derive these results approximately. For x much smaller than $1/\epsilon_v^2$, we can ignore the kinetic term. In the WKB approximation, the waves are of the form $x^{1/4}e^{i\sqrt{x}}$, so the amplitudes grow like $x^{1/4}$. In order to match to a unit norm wave near $x \sim 1/\epsilon_v^2$, we have to scale the wavefunction at small x by a factor $\sim \epsilon_v^{1/2}$, so that near the origin

$$\chi \sim x\epsilon_v^{1/2} \sim \alpha M r \epsilon_v^{1/2} \tag{A26}$$

from which we can read off the derivative at the origin $\frac{d\chi}{dr}(0) = \epsilon_v^{1/2}\alpha M$, and with $k = Mv$ we can determine

$$S_k \sim |\frac{\epsilon_v^{1/2}\alpha M}{Mv}|^2 = \frac{\alpha}{v} \tag{A27}$$

which is correct parametrically. We can also arrive at this result from the second form for S_k. The computation is even more direct here. We have waveforms growing as $x^{1/4}$ towards $x \sim \frac{1}{\epsilon_v^2}$. In the region near $x \sim 1/\epsilon_v$, we must transition to a purely outgoing wave. This is a transmission/reflection problem, and ingoing and outgoing waves from the left will have comparable amplitude. When we continue these back to the origin, we will then have an amplitude reduced by a factor $\sim \epsilon_v^{1/2}$. Then,

$$S_k \sim \left(\frac{1}{\epsilon_v^{1/2}}\right)^2 \sim \frac{\alpha}{v} \tag{A28}$$

2. Attractive Well and Resonance Scattering

Let us do another example, where

$$V = -V_0\theta(L - r), V_0 \equiv \frac{\kappa^2}{2M} \tag{A29}$$

The solution inside is $\chi(r) = A \sin k_{in} r$, where $k_{in}^2 = \kappa^2 + k^2$, while outside we write it as $\sin(k(r-L)+\delta$. Then, matching across the boundary at $r = L$ gives

$$A \sin k_{in} L = \sin\delta,\; k_{in} A \cos k_{in} L = k \cos\delta \tag{A30}$$

Squaring these equations and adding them we can determine

$$A^2 = \frac{1}{\sin^2 k_{in} L + \frac{k_{in}^2}{k^2}\cos^2 k_{in} L} \tag{A31}$$

and so

$$S_k = \frac{A^2 k_{in}^2}{k^2} = \frac{1}{\frac{k^2}{k_{in}^2}\sin^2 k_{in} L + \cos^2 k_{in} L} \tag{A32}$$

Now for $k^2/2M \ll V_0$, we have $k_{in} = \kappa + k^2/(2\kappa) + \cdots$. Clearly, if $\cos\kappa L$ is not close to zero, there is no enhancement. However, if $\cos\kappa L = 0$, then we have a large enhancement

$$S_k \to \frac{\kappa^2}{k^2} \tag{A33}$$

This has a very simple physical interpretation in terms of resonance with a zero-energy bound state. Our well has a number of bound states, and typically the binding energies are of order V_0. We see that if $\cos\kappa L$ is not close to zero, we have to have $A \sim k/k_{in}$ be small. However, if accidentally $\cos\kappa L = 0$, then there is a zero-energy bound state: the wavefunction can match on to $\psi = 1$ for $r > L$ smoothly, giving a zero-energy bound state. The enhancement is of the form

$$S \sim \frac{V_0}{E - E_{\rm bound}} \sim \frac{\kappa^2}{k^2} \tag{A34}$$

Note this formally diverges as $v \to 0$, but is actually cut off by the finite width of the state as familiar from Breit-Wigner.

3. Attractive Yukawa Potential

Now let us examine

$$V(r) = -\frac{\alpha}{2r} e^{-m_\phi r}. \tag{A35}$$

Working again in Bohr units, we have

$$V(x) = -\frac{1}{x} e^{-\epsilon_\phi x} \tag{A36}$$

where

$$\epsilon_\phi \equiv \frac{m_\phi}{\alpha M}. \tag{A37}$$

Now, if $\epsilon_\phi \ll \epsilon_v^2$, the Yukawa term is always irrelevant and we revert to our previous Coulomb analysis.

However, if $\epsilon_\phi \gg \epsilon_v^2$, our analysis changes; we will use the second expression for the Sommerfeld enhancement for simplicity. The potential turns off exponentially around $x \sim 1/\epsilon_\phi$. Now, the effective momentum is

$$k_{\rm eff}^2 = \frac{1}{x} e^{-\epsilon_\phi x} + \epsilon_v^2 \tag{A38}$$

and the quantity

$$\left| \frac{k'_{\rm eff}}{k_{\rm eff}^2} \right| \tag{A39}$$

determines the length scale the potential is varying over relative to the wavelength; so long as it is small, the WKB approximation is good, and we have a waveform growing as $k_{\rm eff}^{-1/2} e^{i \int^x dx' k_{\rm eff}(x')}$. Note that for $1 \ll x \ll 1/\epsilon_\phi$, the WKB approximation is manifestly good. Let us now take the arbitrarily low velocity limit, where $\epsilon_v \to 0$. Then in the neighborhood of $x \sim 1/\epsilon_\phi$ we have $k_{\rm eff}^2 \sim \epsilon_\phi e^{-\epsilon_\phi x}$, and

$$\left| \frac{k'_{\rm eff}}{k_{\rm eff}^2} \right| \sim \sqrt{\epsilon_\phi} e^{\frac{1}{2}\epsilon_\phi x} \sim \frac{\epsilon_\phi}{k_{\rm eff}} \tag{A40}$$

so the WKB approximation breaks down when $k_{\rm eff} \sim \epsilon_\phi$, where the WKB amplitude is $\sim \epsilon_\phi^{-1/2}$. The potential then varies more sharply than the wavelength, and we have a reflection/transmission problem, with an O(1) fraction of the amplitude escaping to infinity. The enhancement is then

$$S \sim \frac{1}{\epsilon_\phi} \sim \frac{\alpha M}{m_\phi} \tag{A41}$$

We did this analysis for $\epsilon_v \to 0$, but clearly it will hold for larger ϵ_v, till $\epsilon_v \sim \epsilon_\phi$, at which point it matches smoothly to the $\frac{1}{\epsilon_v}$ enhancement we get for the Coulomb problem. The crossover with $\epsilon_v \sim \epsilon_\phi$ is equivalent to $Mv \sim m_\phi$, when the deBroglie wavelength of the particle is comparable to the range of the interaction. This is intuitive–as the particle velocity drops and the deBroglie wavelength becomes larger than the range of the attractive force, the enhancement saturates. Of course if ϵ_ϕ is close to the values that make the Yukawa potential have zero-energy bound states, then the enhancement is much larger; we can get an additional enhancement $\sim \epsilon_\phi/\epsilon_v^2$ up to the point where it gets cut off by finite width effects.

In this simple theory it is of course also straightforward to solve for the Sommerfeld enhancement numerically. We show the enhancement as a function of ϵ_ϕ and ϵ_v in Figs. 6 and 7.

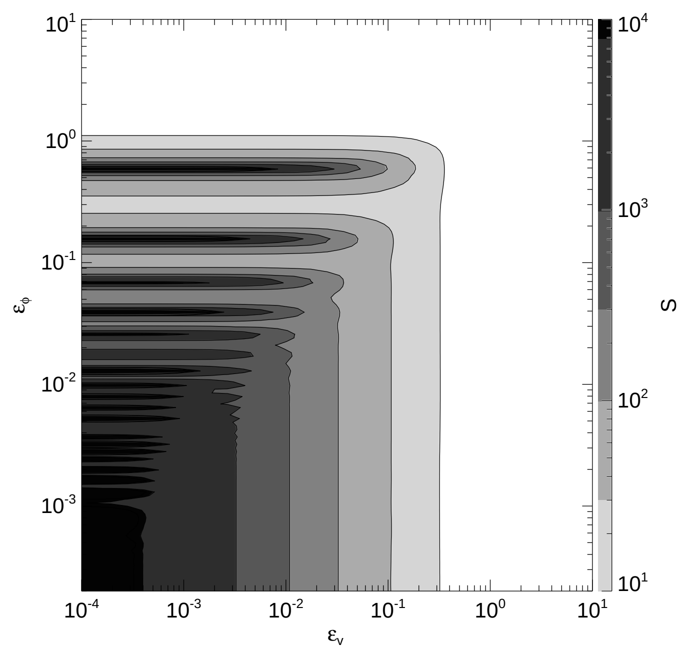

FIG. 6: Contour plot of S as a function of ϵ_ϕ and ϵ_v. The lower right triangle corresponds to the zero-mass limit, whereas the upper left triangle is the resonance region.

4. Two-particle annihilation

Let us finally consider our real case of interest, involving two-particle annihilation. To keep things simple, let us imagine that the two particles are not identical, for instance they could be Majorana fermions with opposite spins; we

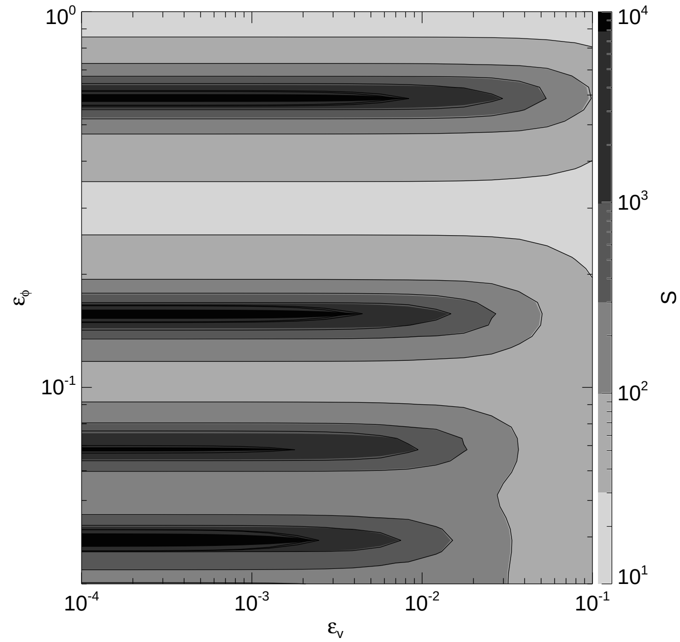

FIG. 7: As in Fig. 6, showing the resonance region in more detail.

can restrict to this case because none of the interactions we consider depend on spin, and the annihilation channels we imagine can all proceed from this spin configuration. Let us imagine that there are a number of states χ_A of (nearly) equal mass, with the label A running from $A = 1, \cdots, N$. The two-particle Hilbert space is labeled by states $|\vec{x}, A; \vec{y}, B\rangle$ and a general wavefunction is $\langle\psi|\vec{x}, A; \vec{y}, B\rangle \equiv \psi_{AB}(\vec{x}, \vec{y})$. There is some short-range annihilation Hamiltonian $U\delta^3(\vec{x}_1 - \vec{x}_2)$ into decay products $|final\rangle$; where U is an operator

$$\langle final|U|AB\rangle \equiv M^{ann}_{AB} \tag{A42}$$

Now, suppose there is also a long-range interaction between the two particles with a potential. The general Schrödinger equation is of the form

$$-\frac{1}{2M}\left(\nabla_x^2 + \nabla_y^2\right)\psi_{AB}(x, y) + V_{ABCD}(x - y)\psi_{CD}(x, y) = E\psi_{AB}(x, y) \tag{A43}$$

As usual we factor out the center-of-mass motion by writing $\psi_{AB}(\vec{x}, \vec{y}) = e^{i\vec{P}\cdot(\vec{x}+\vec{y})}\phi_{AB}(\vec{x} - \vec{y})$, and we have

$$-\frac{1}{M}\nabla_r^2\phi_{AB}(\vec{r}) + V_{ABCD}(\vec{r})\phi_{CD}(\vec{r}) = E_{CM}\phi_{AB}(\vec{r}) \tag{A44}$$

where E_{CM} is the energy in the center-of-mass frame.

We have an initial state with some definite A_*, B_*, and we take as an unperturbed solution

$$\phi^{(A^*B^*)(0)}AB = \delta_A^{A_*}\delta_B^{B^*}e^{ikz} \tag{A45}$$

The annihilation cross section without the interaction V is proportional to

$$\sigma^{(0)}_{ann} \propto |M^{ann}_{A^*B^*}|^2 \tag{A46}$$

However with the new interaction, the annihilation cross section is

$$\sigma_{ann} \propto |\phi^{(A^*B^*)}_{CD}(0)M_{CD}|^2 \tag{A47}$$

So we can write

$$\sigma_{ann} = \sigma_{ann}^{(0)} \times S_k^{A^*B^*} \tag{A48}$$

where

$$S_k^{A^*B^*} = \frac{|\phi_{CD}^{A^*B^*}(0) M_{CD}^{ann}|^2}{|\phi_{CD}^{A^*B^*}(\infty) M_{CD}^{ann}|^2} \tag{A49}$$

Just as above, in reducing the problem to an S-wave computation, we can replace $\phi_{CD}^{A^*B^*}(0)$ with $(\chi'(r=0))_{CD}^{A^*B^*}$ in the obvious way.

Of course one contribution to V_{ABCD} comes from the small mass splittings between the states. If we write the (almost equal) common mass term for the DM states as $(M + \Delta M_A)\chi_A\chi_A$, the ΔM's show up in the potential as

$$V_{ABCD}^{split} = (\Delta M_A + \Delta M_B)\delta_{AC}\delta_{BD} \tag{A50}$$

For the Sommerfeld enhancement, we also need some long-range attractive interaction. As we have discussed, vectors are possibly the most promising candidate. The leading coupling to spin-1 particles $a_{\mu i}$ is

$$g\bar{\chi}_A\bar{\sigma}^\mu\chi_B T_{AB}^i a_{i\mu} \tag{A51}$$

and ignoring the mass of the gauge boson this gives us a $1/r$ contribution to the effective potential

$$V_{ABCD}^{gauge}(\vec{r}) = -\alpha\frac{1}{r}T_{AC}^i T_{BD}^i \tag{A52}$$

while taking the vector masses into account gives both Yukawa exponential factors and a more complicated tensor structure. In total,

$$V_{ABCD} = V_{ABCD}^{split} + V_{ABCD}^{gauge} \tag{A53}$$

Now, it is obvious that in our basis, the T_{AB}^i should be antisymmetric

$$T_{AB}^i = -T_{BA}^i \tag{A54}$$

since the gauge symmetry must be a subgroup of the $SO(N)$ global symmetry preserved by the large common mass term $M\chi_A\chi_A$, and the $SO(N)$ generators are antisymmetric. Thus, the coupling to vectors in this basis is necessarily off-diagonal. Let us look at a simple example, where $N = 2$ and we have a single Abelian gauge field. The ϕ_{AB} span a 4-dimensional Hilbert space, with $|11\rangle, |12\rangle, |21\rangle, |22\rangle$ as a basis. Since the gauge boson exchange necessarily changes $1 \leftrightarrow 2$, V_{ABCD} is block diagonal, operating in two separate Hilbert spaces, spanned by $(|11\rangle, |22\rangle)$ and $(|12\rangle, |21\rangle)$. Since we are ultimately interested in scattering with 11 initial states, let us look at the first one, where we have

$$\mathbf{V} = \begin{pmatrix} 2\Delta M & -\frac{\alpha}{r} \\ -\frac{\alpha}{r} & 0 \end{pmatrix}, \quad \text{in the basis} \quad |11\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, |22\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{A55}$$

Clearly, if ΔM is enormous, we will not have any interesting Sommerfeld enhancement in (11) scattering, since in this case there is no long-range interaction between 11 at all, so let us assume that ΔM is smaller than the kinetic energy of the collision. Now it is clear that as $r \to \infty$, the mass splitting dominates the potential, and obviously particle 1 is the lightest state. However, at smaller distances, the gauge exchange term dominates. This is not diagonal in the same basis, and has "attractive" and "repulsive" eigenstates with energies $\pm\alpha/r$. Note that the asymptotic $|11\rangle$ state is an equal linear combination of the attractive and repulsive channels. While the repulsive channels suffer a Sommerfeld suppression, the attractive channel is Sommerfeld enhanced. Note that as long as ΔM is parametrically smaller than the kinetic energy, its only role in this discussion was to split the two asymptotic states, and therefore determine what the natural initial states are. Note also that if ΔM is large but not infinite, the mixing with 2 generates an attractive potential between 11 of the form $V_{eff}(r) = -\alpha^2/(\Delta M r)$. It would be interesting to understand these limits of the multistate Sommerfeld effect in parametric detail; we defer this to future work.

It is very easy to see that our conclusion about the presence of a Sommerfeld effect is general for any gauge interaction. We think of $V^{gauge}_{(AB)(CD)}$ as a matrix in the Hilbert space spanned by (AB). Note that since the T^i are antisymmetric, they are also traceless, and as a consequence, the matrix $V^{gauge}_{(AB)(CD)}$ is also traceless and so has both positive and negative eigenvalues, reflecting the obvious fact that gauge exchange gives us both attractive and repulsive potentials. Let us go to a basis in (AB) space where V^{gauge} is diagonal, and denote eigenvectors with negative (attractive) eigenvalues as $f^{attractive}_{AB}$ and repulsive ones as $f^{repulsive}_{AB}$. We can think of the initial wavefunction in AB space $\delta^{A^*B^*}_{AB}$ as a state in (AB) space and expand it in terms of these eigenvectors as

$$\delta^{(A^*B^*)}_{AB} = C^{A^*B^*}_{att} f^{attractive}_{AB} + C^{A^*B^*}_{rep} f^{repulsive}_{AB} \tag{A56}$$

we can determine the coefficients by dotting left-hand side and right-hand side into the eigenvectors, so that

$$\delta^{(A^*B^*)}_{AB} = f^{attractive}_{A^*B^*} f^{attractive}_{AB} + f^{repulsive}_{A^*B^*} f^{repulsive}_{AB} \tag{A57}$$

Then, since the repulsive components are exponentially suppressed at the origin while the attractive components are enhanced, we get a Sommerfeld enhancement as long as $f_{A^*B^*} \neq 0$. In particular, for scattering the same species, this is true so long as $f_{A^*A^*} \neq 0$. Said more colloquially, these Majorana states are linear combinations of "positive" and "negative" charged states; so long as they have any component which would mutually attract, there is a Sommerfeld enhancement from that component alone.

---

[1] O. Adriani et al. (2008), 0810.4995.
[2] S. W. Barwick et al. (HEAT), Astrophys. J. **482**, L191 (1997), astro-ph/9703192.
[3] J. J. Beatty et al., Phys. Rev. Lett. **93**, 241102 (2004), astro-ph/0412230.
[4] M. Aguilar et al. (AMS-01), Phys. Lett. **B646**, 145 (2007), astro-ph/0703154.
[5] G. L. Kane, L.-T. Wang, and T. T. Wang, Phys. Lett. **B536**, 263 (2002), hep-ph/0202156.
[6] D. Hooper, J. E. Taylor, and J. Silk, Phys. Rev. **D69**, 103509 (2004), hep-ph/0312076.
[7] I. Cholis, D. P. Finkbeiner, L. Goodenough, and N. Weiner (2008), 0810.5344.
[8] E. A. Baltz, J. Edsjo, K. Freese, and P. Gondolo, Phys. Rev. **D65**, 063511 (2002), astro-ph/0109318.
[9] J. Chang et al., Nature **456**, 362 (2008).
[10] D. P. Finkbeiner, Astrophys. J. **614**, 186 (2004), astro-ph/0311547.
[11] G. Dobler and D. P. Finkbeiner, Astrophys. J. **680**, 1222 (2008), 0712.1038.
[12] D. Hooper, D. P. Finkbeiner, and G. Dobler, Phys. Rev. **D76**, 083012 (2007), 0705.3655.
[13] A. W. Strong, R. Diehl, H. Halloin, V. Schönfelder, L. Bouchet, P. Mandrou, F. Lebrun, and R. Terrier, Astron. Astrophys. **444**, 495 (2005), arXiv:astro-ph/0509290.
[14] D. J. Thompson, D. L. Bertsch, and R. H. O'Neal, Jr., Astrophys. J. Supp. **157**, 324 (2005), arXiv:astro-ph/0412376.
[15] F. A. Aharonian, A. M. Atoyan, and H. J. Voelk, Astron. Astrophys. **294**, L41 (1995).
[16] D. Hooper, P. Blasi, and P. D. Serpico (2008), 0810.1527.
[17] H. Yuksel, M. D. Kistler, and T. Stanev (2008), 0810.2784.
[18] D. P. Finkbeiner and N. Weiner, Phys. Rev. **D76**, 083519 (2007), astro-ph/0702587.
[19] I. Cholis, L. Goodenough, and N. Weiner (2008), 0802.2922.
[20] I. Cholis, G. Dobler, D. P. Finkbeiner, L. Goodenough, and N. Weiner (2008), 0811.3641.
[21] R. Bernabei et al. (DAMA), Eur. Phys. J. **C56**, 333 (2008), 0804.2741.
[22] D. R. Smith and N. Weiner, Phys. Rev. **D64**, 043502 (2001), hep-ph/0101138.
[23] D. Tucker-Smith and N. Weiner, Phys. Rev. **D72**, 063509 (2005), hep-ph/0402065.
[24] S. Chang, G. D. Kribs, D. Tucker-Smith, and N. Weiner (2008), 0807.2250.
[25] I. Cholis, L. Goodenough, D. Hooper, M. Simet, and N. Weiner (2008), 0809.1683.
[26] M. Cirelli, M. Kadastik, M. Raidal, and A. Strumia (2008), 0809.2409.
[27] E. A. Baltz et al., JCAP **0807**, 013 (2008), 0806.2911.
[28] D. N. Spergel and P. J. Steinhardt, Physical Review Letters **84**, 3760 (2000), astro-ph/9909386.
[29] R. Davé, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, Astrophys. J. **547**, 574 (2001), arXiv:astro-ph/0006218.
[30] N. Arkani-Hamed and N. Weiner (2008), 0810.0714.

[31] A. Sommerfeld, Annalen der Physik **403**, 257 (1931).
[32] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, Phys. Rev. D **71**, 063528 (2005), arXiv:hep-ph/0412403.
[33] M. Kamionkowski and S. Profumo (2008), 0810.3233.
[34] F. Governato, B. Willman, L. Mayer, A. Brooks, G. Stinson, O. Valenzuela, J. Wadsley, and T. Quinn, Mon. Not. R. Astron. Soc. **374**, 1479 (2007).
[35] P. J. E. Peebles, S. Seager, and W. Hu, Astrophys. J. **539**, L1 (2001), astro-ph/0004389.
[36] N. Padmanabhan and D. P. Finkbeiner, Phys. Rev. **D72**, 023508 (2005), astro-ph/0503486.
[37] M. Mapelli, A. Ferrara, and E. Pierpaoli, Mon. Not. Roy. Astron. Soc. **369**, 1719 (2006), astro-ph/0603237.
[38] G. Steigman, Ann. Rev. Nucl. Part. Sci. **57**, 463 (2007), 0712.1100.
[39] G. Hinshaw et al. (WMAP) (2008), 0803.0732.
[40] B. Holdom, Phys. Lett. **B166**, 196 (1986).
[41] J. L. Feng and J. Kumar, Phys. Rev. Lett. **101**, 231301 (2008), 0803.4196.
[42] R. Foot and R. R. Volkas, Phys. Rev. **D52**, 6595 (1995), hep-ph/9505359.
[43] M. Pospelov, A. Ritz, and M. B. Voloshin, Phys. Lett. **B662**, 53 (2008), 0711.4866.
[44] D. P. Finkbeiner, N. Padmanabhan, and N. Weiner, **805** (2008), 0805.3531.
[45] D. P. Finkbeiner, T. R. Slatyer, and N. Weiner (2008), 0810.0722.
[46] Z. Ahmed et al. (CDMS), Phys. Rev. Lett. **102**, 011301 (2009), 0802.3530.
[47] J. Angle et al. (XENON), Phys. Rev. Lett. **100**, 021303 (2008), 0706.0039.
[48] T. Han and R. Hempfling, Phys. Lett. **B415**, 161 (1997), hep-ph/9708264.
[49] L. J. Hall, T. Moroi, and H. Murayama, Phys. Lett. **B424**, 305 (1998), hep-ph/9712515.
[50] G. Weidenspointner et al., Astron. Astrophys. **450**, 1013 (2006), astro-ph/0601673.
[51] G. Weidenspointner et al. (2007), astro-ph/0702621.
[52] L. E. Strigari, S. M. Koushiappas, J. S. Bullock, and M. Kaplinghat, Phys. Rev. **D75**, 083526 (2007), astro-ph/0611925.
[53] J. I. Read, G. Lake, O. Agertz, and V. P. Debattista (2008), 0803.2714.
[54] Work in Progress, Institute for Advanced Study, New York University, Harvard University collaborations (2008).
[55] G. D. Mack, T. D. Jacques, J. F. Beacom, N. F. Bell, and H. Yuksel, Phys. Rev. **D78**, 063542 (2008), 0803.0157.
[56] HESS Collaboration: F. Aharonian, ArXiv e-prints (2008), 0809.3894.
[57] M. Cirelli, A. Strumia, and M. Tamburini, Nuclear Physics B **787**, 152 (2007), arXiv:0706.4071.
[58] M. Cirelli and A. Strumia (2008), 0808.3867.
[59] M. Pospelov and A. Ritz (2008), 0810.1502.
[60] V. Barger, W. Y. Keung, D. Marfatia, and G. Shaughnessy (2008), 0809.0162.
[61] F. Donato, D. Maurin, P. Brun, T. Delahaye, and P. Salati (2008), 0810.5292.
[62] J. F. Beacom, N. F. Bell, and G. Bertone, Phys. Rev. Lett. **94**, 171301 (2005), astro-ph/0409403.

[63] Although, it should be noted that such an explanation seems to fail because the spectrum would not be sufficiently hard; see Fig. 6 of [11].
[64] The importance of this effect in the context of multi-TeV scale dark matter interacting via W and Z bosons was first discussed by [32], and more recently emphasized with regards to PAMELA in the context of “minimal dark matter,” with similar masses by [57, 58].
[65] Another possibility for sufficiently light bosons would be radiative capture [59].
[66] Recently, [60] explored the possibility that PAMELA could distinguish the spin of the dark matter, which could be relevant here in annihilations to vector bosons.
[67] Note that in the early universe, all these states will be populated at freeze-out, and will participate in maintaining thermal equilibrium. However, unlike coannihilation in the MSSM, which typically occurs between a Bino and a slepton, there is not expected to be a large effect, on the connection between the present annihilation rate and that at freeze-out. This is because the self-annihilation and coannihilation cross sections are similar or even equal here, while in the MSSM the coannihilation cross section is much larger.
[68] Although larger annihilation rates into W^+W^- or other standard model states are possible, the cross section must be lower than the limits of [26, 61] arising from antiprotons. Moreover, strong limits from final state radiation [62] can also arise, although these depend more sensitively on the nature of the halo profile.