Gravitational probes of dark matter physics

Matthew R. Buckley\textsuperscript{a}, Annika H. G. Peter\textsuperscript{b,c,}\textsuperscript{*}

\textsuperscript{a}Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA
\textsuperscript{b}CCAPP and Department of Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, OH 43210, USA
\textsuperscript{c}Department of Astronomy, The Ohio State University, 140 W. 18th Ave., Columbus, OH 43210, USA

Abstract

The nature of dark matter is one of the most pressing questions in physics. Yet all our present knowledge of the dark sector to date comes from its gravitational interactions with astrophysical systems. Moreover, astronomical results still have immense potential to constrain the particle properties of dark matter in the near future. We introduce a simple 2D parameter space which classifies models in terms of a particle physics interaction strength and a characteristic astrophysical scale on which new physics appears, in order to facilitate communication between the fields of particle physics and astronomy. We survey the known astrophysical anomalies that are suggestive of non-trivial dark matter particle physics, and present a theoretical and observational program for future astrophysical measurements that will shed light on the nature of dark matter.

Keywords: dark matter, galaxies, cosmology, particles
PACS: 95.35.+d, 98.62.-g, 98.80.-k, 14.80.-j

1. Introduction

For particle physicists, dark matter is clear evidence of the existence of new physics beyond the Standard Model. Significant theoretical effort has been applied to construct viable models of particle dark matter, and the resulting model-space is enormous. While many of these models contain particles which are “minimal” dark matter candidates in the cosmological sense (i.e., cold, functionally collisionless, non-baryonic matter), many others contain dark matter candidates that deviate at some level from the cold dark matter (CDM) paradigm \cite{1, 2, 3}. Well known examples include warm dark matter (WDM) \cite{4, 5} and self-interacting dark matter (SIDM) \cite{6}. Given the variety of models available, experimental input is necessary to narrow down the possibilities. While many experiments have searched for signals of particle dark matter using a variety of traditional particle physics techniques \cite{7, 8, 9} (e.g., collider production \cite{10, 11, 12}, direct detection \cite{13, 14, 15, 16, 17}, and indirect detection \cite{18, 19, 20}), so far all results have been negative.

Consequentially, the primary source of positive statements about the microphysical properties of dark matter continues to be the astrophysical and cosmological study of dark matter’s gravitational interactions. This is perhaps not surprising. After all, it was through gravity that dark matter was discovered \cite{1, 21, 22}, and its status as a non-relativistic (“cold,” or at least no more than “warm”) component of the Universe at the time of structure-formation verified \cite{23, 24, 25}. As astronomical surveys and galaxy simulations continue to improve, so too will our ability to explore the distribution of dark matter in the Universe and potentially discover deviations from the CDM paradigm, as predicted by many of the theoretical dark matter models. In many cases these astrophysics techniques are sensitive to dark matter properties orthogonal to those measured by particle physics experiments. Specifically, traditional particle physics techniques probe the microphysical interactions between dark matter and the Standard Model. However, the distribution of dark

\*Principal corresponding author

\textit{Email addresses:} mbuckley@physics.rutgers.edu (Matthew R. Buckley), apeter@physics.osu.edu (Annika H. G. Peter)

Preprint submitted to Elsevier September 27, 2018
matter in the Universe may be governed by additional interactions within a “dark sector” that are secluded from the Standard Model, in addition to interactions between dark matter and the Standard Model. Such interactions would not appear in experiments probing the Standard Model-dark matter scattering, but would appear in astrophysical measurements by altering the spatial distribution of dark matter in the Universe with respect to the CDM prediction.

The microphysical properties of dark matter — particle mass, number of species, interaction strength of each particle with Standard Model and dark particles — can imprint themselves on the macroscopic distribution of dark matter in the Universe. To illustrate the power of astrophysics to reveal the microphysical properties of dark matter, it is useful to consider the decidedly non-minimal Standard Model. For example, because the Standard Model includes charged particles and photons, gas can cool. This enables the formation of stars and galaxies. One might imagine that hypothetical researchers in the dark sector would be able to identify at least some of the major components and non-gravitational forces of the Standard Model based on gravitational evidence of the clustering of Standard Model particles. We explore this thought experiment in more detail in Section 2 of this work. As this conceptual exercise will demonstrate, astronomical observations constitute measurements of dark matter microphysics.

There is immense potential for discovery in the intersection of particle and astrophysical measurements of dark matter. Positive signals of deviations from the gravitational predictions of CDM would give theoretical physicists much-needed experimental guidance about parameters that are not easily measured in particle physics experiments. If, on the other hand, all astrophysical studies of dark matter were to find correspondence with the CDM predictions, the improved knowledge of dark matter distributions would be extremely useful for reducing major sources of theoretical uncertainties in the particle physics experiments [26, 27, 28, 29, 30]. Likewise, results from particle experiments can either restrict the possible model space relevant to novel astrophysical signals, or suggest specific deviations from the CDM paradigm which can be confirmed by later surveys.

Though the connection between astrophysical measurements and particle physics models of dark matter is recognized for particular models (notably, sterile neutrinos [31, 32, 33, 34, 35]), the full potential is perhaps not appreciated in the broader context. This is in part due to the separate languages used by particle physicists and astrophysicists to describe dark matter, and in part due to the fact that astrophysical and particle-physics measurements of dark matter often depend on very different subsets of the parameter space of dark matter microphysics. Finally, the “natural” set of metrics used to constrain or define dark matter depend strongly on the type of experiment or observation used to find dark matter.

We have two primary goals for this paper. First, we want to present a compact parameter space to capture the phenomenology of dark matter models relevant to both particle physicists and astrophysicists. To that end, we propose a simple two-dimensional parameter space, with one axis closely tied to particle dark matter searches and the other closely tied to astronomical searches, to orient dark-matter hunters in both fields. We show where models of current interest lie in this space, and where potential discrepancies between observation and CDM predictions are located. Second, we advocate a long-term plan for sharpening astrophysical constraints on dark-matter microphysics. This project will require close collaboration between physicists and astronomers (many of whom may not have dark matter as a primary area of research), which motivates our simple parameter space which translates between the two fields.

The two-dimensional description is a simple mapping which can be applied to a wide range of models and makes clear the sensitivity of both particle- and astrophysical-probes to the dark sector. By considering where specific models fall in this parameter space, particle physicists may identify new constraints or observables from astrophysics, while astrophysicists may find applications for their results of which they were previously unaware.

The two parameters we adopt are:

1. The characteristic interaction strength between dark matter and the Standard Model, $\Lambda^{-1}$; and
2. The characteristic largest dark matter halo mass where significant deviations from CDM predictions should arise, $M_{\text{halo}}$. The halo is the fundamental unit of cosmological dark matter on non-linear scales, a self-gravitating collection of dark-matter particles in equilibrium formed when a perturbation in the homogeneous universe detaches from the Hubble flow, and the structure within which galaxies live.
The former parameter (defined to have dimensions of inverse energy) gives an estimate of the ability of particle physics experiments to discover dark matter through interactions with baryonic (used in the astronomical sense: all Standard Model particles but neutrinos and photons) matter. The latter gives an estimate of the type of astrophysical system whose dynamics must be understood in order to discover the presence of unique dark matter physics.

We emphasize that these two parameters are not intended to encompass the full behavior from a particular model of dark matter. Given the breadth of theoretical work on this topic, many well-motivated models of dark matter can have unique phenomenology that would be difficult to capture in any low-dimensional set of parameters. However, flattening the model space down to these two parameters allows for direct comparison of a variety of models, which is particularly useful to scientists in one community (particle physics or astrophysics) looking to make connections with the results in the other.

As part of this exercise, we highlight current challenges and new opportunities in dark matter astrophysics. We show what the current theory systematics are in mapping between dark matter microphysics and $M_{\text{halo}}$, and where the current observational constraints on deviations from CDM lie. Importantly, we outline a systematic plan for tightening theory systematics and improving observational constraints. Some ideas have been discussed in the literature, but have not been unified into a comprehensive approach. We also identify a range of $M_{\text{halo}} \lesssim 10^5 M_\odot$ which is a critical test of the CDM paradigm, but is not expected to be probed by any present or near-future set of observations. This calls for new ideas, and the community must rise to meet this challenge.

We begin our paper in Section 2 with our thought experiment, considering what a dark-matter scientist could discover of the particle physics in the Standard Model, using only evidence from our gravitational imprint. Having demonstrated the interplay between particle physics and astronomy, in Section 3, we define our two figures of merit, and demonstrate their applicability to various dark matter models. In Section 4, we discuss a series of observations that suggest that some non-trivial physics may be at play in the dark sector. These observations are at this point just hints, and in Section 5, we discuss near-term astrophysical probes and theoretical improvements that can increase the existing sensitivity to the astrophysical figure of merit, as well as more accurately defining $M_{\text{halo}}$ given known astrophysical systematics. This program will conclusively answer the question whether these hints are of dark matter origin, as well as bring more dark matter models into experimental reach. We conclude with a discussion of the prospects for the joint astrophysical-particle physics investigation of dark matter model-space, and the important lessons about the physics of the dark sector that can be learned even in the absence of positive results.

2. Dark Matter Discovery of the Standard Model

Before developing a set of parameters to characterize the physics of dark matter models, it is interesting and informative to consider a thought experiment which makes clear the power of astrophysical studies to uncover novel particle physics in the dark sector. While we do not know if dark matter has non-trivial internal particle physics associated with it, our own Standard Model is decidedly non-trivial, with multiple particles and interactions. If the Standard Model sector were as invisible to a hypothetical dark matter observer as the dark sector is to us, one might assume that none of this complexity could be inferred. However, this is not the case, as we will show. The goal of the following thought experiment is two-fold. First, and most importantly, we illustrate the power of astrophysics to unveil some major aspects of dark-matter physics. Second, we show which aspects of a secluded model will be most challenging to discover with astrophysics.

We imagine a hypothetical “dark-matter scientist,” capable of perceiving the dark matter in the Universe but not the baryonic matter. For the purposes of the thought experiment, this implies some massless or very low mass photon equivalent, i.e., a “dark photon” which has nontrivial constraints from

\footnote{Following the usual nomenclature we will use “dark” to denote the sector of physics invisible to us but visible to our hypothetical dark-matter scientist, while “visible” corresponds to the baryonic sector which is detectable to us but invisible to the dark-matter scientist.}
data 39 40 11. To avoid the issues of extra light degrees of freedom in the Cosmic Microwave Background (CMB), we assume the dark sector decoupled from the visible sector early on, and was colder than the Standard Model bath at the time of recombination 12 37. We assume that this dark photon provides the dark-matter scientist with a “dark CMB” which allows them precision measurements of the Universe’s energy budget. Imagine that, just as we visible-sector scientists discovered dark matter through a combination of astronomical and cosmological measurements (through, e.g., rotation curves 22 or the CMB), the dark-matter scientist discovered hints of the visible sector in the sky.

Having divined its existence, what else can our dark-matter scientist learn of this mysterious matter? Particle physics searches would prove difficult. That is true even if the dark-matter scientist lives in a dark matter halo that also contains a significant number of baryons, rather than in a baryon-poor filament or void. This is because the visible galaxy, composed of baryons, inhabits the very center of the halo — the radius enclosing half the mass of the baryonic component is at best a few percent of the virial radius for dark matter 43 44. The dark-matter scientist would have to be deep inside the potential well of their halo in order to have a chance at detecting baryons using the dark matter equivalent of direct detection experiments 13 14 17, assuming a weak-scale coupling between the two sectors. Indirect detection of dark sector byproducts from baryonic self-annihilation would also be ineffective, as protons are not their own antiparticle, and the Universe is overwhelmingly matter-dominated 45. Thus, there would be little annihilation at the center of any halo. A dark-matter scientist could probe interactions between the sectors at or below the strength of the weak nuclear force only in the dark-matter equivalent of a particle accelerator (where they would face the equivalent of all the experimental difficulties inherent in searching for dark matter at the LHC).

Yet, from astrophysics and cosmology, our dark-matter scientist may infer much about the structure of the Standard Model. With the dark CMB and measurements of the clustering of dark-matter halos, they would discover that the dark matter could not account for all of the non-relativistic energy, and that some missing matter exists, with Ω_b ∼ 0.05. Moreover, this missing matter must have many light degrees of freedom, in order to explain the redshift of matter-radiation equality. In fact, the light degrees of freedom from the visible sector would overwhelmingly dominate cosmology over those from the dark sector.

Depending on the redshift of dark decoupling relative to matter-radiation equality, the dark-matter scientist could obtain precision measurements of the number of light degrees of freedom in the Standard Model (including evidence for the neutrinos, though this would be difficult to disentangle from a non-abelian unbroken gauge sector), as well as the temperature of the visible sector.

From the two-point correlation function of dark matter halos, the dark-matter scientist would see the imprint of baryon acoustic oscillations (BAO) 46, which would indicate that, whatever this mysterious matter was, it must be strongly self-coupled. If the scientist then turned their attention to the inner structure of dark matter halos, they would see gravitational evidence for a surprising deviation from the spherical density distribution they are accustomed to finding at large radii, replaced by some kind of disky structure.

The combination of evidence of a baryonic disk, matter-radiation equality, and BAO would tell our dark-matter scientist that the visible matter must not just be strongly self-coupled, but that it must have some light force carrier capable of radiating away the excess kinetic energy after the virialization of the halo. A reasonable conclusion would be that the light degrees of freedom, inferred from the scale factor of matter-radiation equality, are related to these light force carriers. This would be suggestive of some unbroken (or nearly unbroken) gauge group (assuming that dark-matter scientists are at least as familiar with quantum field theory as we are). A non-abelian gauge group for this light force carrier would be a possibility in addition to the actual solution of the electromagnetic U(1), but in either case the dark-matter scientist would recognize the need for the visible matter to be overall neutral under this force. This would suggest at least two particles in the baryonic sector with opposite charges: H (heavy) and L (light) 2.

At this point, the dark-matter scientist would not be able to tell whether H and L are two different particles with different masses, or just antiparticles of each other (in which case, of course, m_H = m_L).

2 Alternatively, N particles charged under the fundamental of some non-abelian SU(N), as with the three quarks of SU(3)_C.
However, as we will describe below, the observation that the baryonic matter has undergone gravitational collapse requires an efficient mechanism to radiate away energy, which is suggestive of at least one light charged particle. As we will see, if the dark-matter scientists consider the possibility that the kinetic energy of the baryons during gravitational collapse is much larger than the binding energy (so that free-free emission is possible), this would lead to the conclusion that $m_H \gg m_L$. Thus, while other possibilities might lead to self-consistent astrophysics, it will be reasonable for them to consider this limit.

Now the dark-matter scientist would want to determine the coupling between the $H$ and $L$ — what we know as the electromagnetic fine structure constant $\alpha$. There are two options: either the $H$ and $L$ form bound states in the baryonic halo, or they are ionized. The former possibility requires binding energies much higher than the typical kinetic energy of the particles in the disk ($B \gg kT_{\text{vir}}$, where $T_{\text{vir}}$ is the temperature of particles in a halo of mass $M_{\text{vir}}$ — see Section 3.2.1 for precise definitions). Cooling would proceed through the collisional excitation [17]. At minimum, cooling could be facilitated by hyperfine de-excitation, which would only be possible if the virial temperature of the halo were sufficiently larger than the hyperfine transition energy ($kT_{\text{vir}} \sim GM_H M_{\text{vir}}/R_{\text{vir}} \gg E_{\text{hf}}$), or

$$\alpha_{\text{bound states}} < 10^{-1/2} \left( \frac{M_{\text{vir}}}{10^{12} M_\odot} \right)^{1/6} \left( \frac{m_H/m_L}{m_p/m_e} \right)^{1/2},$$

where $M_{\text{vir}}$ is the virial mass of a halo [11]. This is the requirement on $\alpha$ for the existence of bound states and the possibility that cooling may proceed through hyperfine emission. In general, though, the resulting radiation transport and cooling functions are complicated [18, 19], especially if dark nuclear physics exists. It also is more important for low-mass halos rather than high-mass halos, and generally the range of $\alpha$ and $m_H/m_L$ that allow bound states to cool efficiently is small. For example, if hyperfine transitions are the dominant cooling mode, we require

$$\alpha_{\text{HF cooling}} \gtrsim 0.1 \left( \frac{m_H}{m_p} \right)^{3/2} \left( \frac{M_{\text{vir}}}{10^{12} M_\odot} \right)^{1/6}$$

if $m_H/m_L \gg 1$. This is the requirement on $\alpha$, that it lie between the go-criterion limit of Eq. (1) and the requirement that the baryonic matter cool in a Hubble time (Eq. (2)), if the matter exists as bound states. This scenario is an attractive possibility for low-mass dark-matter halos if one prefers $\alpha$ to remain in the perturbative limit, and is one that is realized in the visible sector on dwarf scales before reionization [17].

Turning to the possibility that the binding energy of the $H - L$ system is much lower than the average kinetic energy of particles in the halo, which is possible if

$$\alpha_{\text{ionized states}} \lesssim 10^{-2} \left( \frac{M_{\text{vir}}}{10^{12} M_\odot} \right)^{1/3} \left( \frac{m_H/m_L}{m_p/m_e} \right)^{1/2}$$

then cooling would proceed through thermal bremsstrahlung. This limit on $\alpha$ is what is required for the gas to be ionized for a halo of mass $M_{\text{vir}}$. An argument of this sort was used in Ref. [50] to argue for the typical mass scale of galaxies. In the case that $m_H \gg m_L$, then the dark-matter scientist would determine the volumetric energy loss rate is [51]

$$\frac{d^2 E}{dVdt} \approx \left( \frac{2\pi k_B T_{\text{vir}}}{3m_L} \right)^{1/2} \frac{2 \hbar^2 \alpha^3}{3m_L} n_L n_H,$$

where $n_L$ and $n_H$ are the number densities of $L$ and $H$. Because Coulomb interactions keep the $L$ and $H$ species in kinetic equilibrium, $n_L = n_H$ under the assumption that the galaxy is neutral under whatever force is allowing the cooling of the $H - L$ mixture. The total energy radiated away over a Hubble time is approximately equal to the kinetic energy in the baryonic sector at the initial virialization of the halo:

$$E_{\text{tot}} \sim \frac{fM_{\text{vir}} k_B T_{\text{vir}}}{m_H},$$

where
where $f = \Omega_b/\Omega_{DM}$. Setting the volumetric loss rate times the lifespan of the Galaxy over the entire virial radius of the baryonic disk equal to the total energy lost, and using the virial density $\tilde{\rho} = \Delta_r \rho_c$ (where $\rho_c$ is the critical density of the Universe and $\Delta_r \approx 200$ [23]; Section 3.2.1) to determine $n_H = n_L = f\tilde{\rho}/m_H$, the dark physicist would finally come to the conclusion that

$$\alpha_{\text{brem}} \gtrsim 10^{-7/3} \left( \frac{M_{\text{vir}}}{10^{12} M_\odot} \right)^{1/9} \left( \frac{m_H/m_L}{m_p/m_e} \right)^{1/2} \left( \frac{m_L}{m_e} \right).$$

(6)

For visible matter to exist as unbound states and cool in halos in a Hubble time, $\alpha$ must lie between the limits of Eqs. (3) and (5).

Interestingly, applied to Milky Way-mass halos and using the actual masses of the Standard Model particles, this minimum bound on $\alpha$ from the thermal bremsstrahlung estimate is close to the actual value of $\alpha \sim 1/137$. Note that increasing the possible mass of the “unknown” baryonic particles would eventually lead the dark physicist to a non-perturbative theory.

From this rough calculation, the dark-matter scientist would also be able to conclude that visible matter is not composed of a thermal bath of particle-antiparticle, as the implied $\alpha$ leads to a scattering rate for a bath of matter-antimatter partners which is much too large to allow a significant relic abundance to survive into this late era of the Universe’s life [32]. Given that $H$ and $L$ are not antiparticles, the possibility that $m_H$ and $m_L$ are parametrically different would be an attractive avenue of research — resulting in self-consistent results for the dark-matter scientist.

Having now concluded that baryons are not a thermal relic, the dark-matter scientist could infer that there must be some form of baryogenesis at work in the early Universe, and therefore that the baryonic sector contains CP violation in some form. They would be however hard-pressed to come up with any theory about the origin of that CP violation — a situation that we in the visible sector, with far more information about our own particle content, can certainly sympathize with.

Let us summarize what the dark-matter scientist would find in our thought experiment, and what open questions would still exist. Via traditional cosmological and astronomical measurements, the dark-matter scientist would find that a new sector of matter must exist, and could quantify what fraction of the Universe’s energy budget it must contain. They could determine that this sector must have a number of light degrees of freedom, and that the sector must have been tightly coupled early in the Universe’s history but not today. Assuming that at least a few of these light degrees of freedom might be the force carriers that allow this sector to be tightly coupled, they might assume that there is something like a $U(1)$ symmetry (electromagnetism, or akin to it) in the visible sector. Their suspicions would be strengthened by observing that halos often have dense pancake-like features at the center, indicative of a light force carrier being radiated away through collisional cooling. They would be able to constrain $\alpha$ and two particle masses from cooling rate considerations. Between this and calculations of relic densities in the early Universe relic densities as well as baryogenesis, the dark-matter scientist would conclude that the two charged particles would likely not be anti-particles, but instead two species with opposite charges.

This set of conclusions is non-trivial. However, it is also useful to consider the parts of the Standard Model which might be harder for the dark-matter scientist to infer. Would the dark-matter scientist be able to discover the three generations of lepton we know exist in the visible sector, and not just the electron? Would the dark-matter scientist discover nuclear physics, revealing that our heavy particle, the proton, is a composite particle? Would they discover three generations of quarks? Would the extra light degrees of freedom be identified with neutrinos, or as additional $U(1)$ forces? For some of these questions, the answer is a qualified yes. For example, stars in the visible sector might act as microlenses for dark matter astronomers, as they do for visible-sector applications [53]. The abundance and stability of such objects may hint at nuclear fusion-type processes. But for some of these questions — e.g., three generations of quarks and leptons — the answer might be no.

This thought experiment is not an attempt to argue that a dark-matter scientist could immediately determine the unique properties of the baryons. As we demonstrated, there are several possible branches that our hypothetical researcher may wander down; it is not clear that all other options would not also lead to self-consistent results. Indeed, one might expect the same level of vigorous debate as to the nature of this
mysterious extra component of the Universe as found among our baryonic theoretical physicists. However, the dark-matter scientist would be able to map out some of the most important features of the Standard Model, like electromagnetism, which were also the first Standard Model features that were described by modern theory by visible-sector scientists.

3. Metrics for Dark Matter Models

As our thought experiment demonstrates, much may be learned about the complicated Standard Model particle physics through measurements of the gravitational imprint of baryons if we were dark-matter scientists surveying the Universe. We can uncover non-trivial dark matter physics in the same manner. A comprehensive characterization of dark matter microphysics requires a combination of approaches: laboratory-based particle physics searches for interactions with the Standard Model, and the astronomical searches for interactions within a dark sector and also (as we will see) with the Standard Model. To organize these searches, we need a compact space in which to classify models in terms of their observability in the laboratory and in the sky. Our goal with this section is to motivate a specific choice for this space, and to show how particle dark matter models inhabit it. The space is designed to be well-matched to the ways particle physicists and astronomers think about dark matter, making the mapping between the particle and astronomical spaces
Figure 2: Representative example of the types of plots used to classify dark matter candidates in a parameter space useful to particle physicists. Particle masses are given in energy units (GeV), instead of mass units such as grams or GeV/c², because c = 1 by convention in particle physics (i.e., mass = energy). The cross section σ\text{int} is defined as the interaction cross section between dark matter and a Standard Model particle, where the relevant particle is different for different dark matter models, depending on which interaction is tested for a specific model in a specific experiment. A pb is a picobarn, or 10⁻³⁶ cm².

transparent and straightforward, and compact but informative enough so that one might define “figures of merit” to quantify how well future experiments and observations will constrain dark matter models.

We classify dark matter models by their interaction strength with the Standard Model, Λ⁻¹, and the cosmological scale at which we expect to see a deviation from the Cold Dark Matter (CDM) paradigm, M\text{halo}. The former defines the sensitivity of particle physics detectors and the latter defines the largest size of the systems that must be understood in order to discover model-specific dark matter structures. We consider each axis of this parameter space in turn, and classify some well-known dark matter models by where they fall in the resulting two-dimensional parameter space. We summarize our estimates for these models in Figure[1] with details given in the text. Because one of our goals is to enable better communication between particle physicists and astronomers, we take a more pedagogical approach in defining parameters than is standard for either field.

For a more thorough review of the particle physics of many (though not all) dark matter models, see Refs. [54, 7, 55].
3.1. Standard Model Interaction Strength

Defining a single parameter describing the interaction of dark matter with the Standard Model which can be applied to the wide variety of dark matter models is not straightforward, and any approach is bound to obscure some information about the models. A common approach in particle physics circles is to classify dark matter models in a two-dimensional parameter space, but even defining this simplified parameter space capturing the phenomenology of particle physics is challenging. We plot one commonly-used set of parameters in Figure 2. Typically, one dimension of the particle physics space is the particle mass, which is a well-motivated and straightforward choice. From an experimental perspective, the particle mass conveys information about the energy scale (kinetic energy in the case of WIMP direct detection, total available energy in the case of creation in the lab or indirect detection in the sky) required to find dark matter particle. The second parameter is typically chosen to be an interaction cross section (as in Figure 2) between dark matter and whatever Standard Model particle is most relevant for a particular dark matter model. This cross section is not uniquely defined between models, because it depends on the type of Standard Model particle involved in the interaction, the energy scale, and quantum numbers. Despite these limitations and ambiguities, this axis provides a quantity that captures some sense of how easy or difficult it would be to find a particular type of dark matter. As demonstrated in Figure 2, the volume of the space is enormous, and it is not obvious how to compare the distinguishing power of different types of experiment fairly in this space. Moreover, this particle-physics parametrization fails to capture any of the experimental constraints on dark matter coming from astrophysics and cosmology.

Bearing these issues with the traditional two-dimensional space in mind, our goal is to quantify the interaction of dark matter with the Standard Model with a single parameter with a compact volume that retains distinguishing power among dark matter models. As a case study, let us consider two of the best-known dark matter models: weakly interacting dark matter (WIMPs) [56], and axion dark matter [57]:

WIMP dark matter — plotted in Figure 2 as “supersymmetry/extra dimensions” — is relatively heavy (on the scale of the electroweak breaking scale, 246 GeV), and generally has interaction cross sections with Standard Model of approximately the same scale as neutrinos. The high mass of WIMP dark matter allows for direct and indirect detection searches with signals well-separated in energy from backgrounds, while the weak-scale interactions allow for the possibility of dark matter production at the Large Hadron Collider (LHC; see Ref. [58] for a review). However, for any given implementation of WIMP dark matter, including even the constrained version found in the Minimal Supersymmetric Standard Model (MSSM), the relevant cross section for these three detection techniques can vary significantly, depending on the details of the theory.

Axion models, in contrast, have phenomenology largely driven by a single parameter: the mass scale $f_a$. This single scale sets both the mass of the axion and the interaction with the Standard Model, both being proportional to inverse powers of $f_a$ (though specific implementations of axion models have different $O(1)$ numbers which modify these relations). Viable models of axion dark matter require very high values for $f_a \gtrsim 10^{10}$ GeV, which in turn implies very small dark matter mass $\sim \mu eV$ (see Ref. [59] for a review).

To describe models of dark matter as different as WIMPs and axions in a single variable, we define a particle physics parameter combining the dimensionless coupling constant ($\lambda$) that controls the strength of the interaction between the visible and dark particles with the relevant mass (energy) scale $M$ for the interaction. This scale $M$ is typically the mass scale of the particle(s) mediating the interaction (like the $Z$ boson mediating electron-neutrino elastic scattering in the Standard Model), rather than the mass of the dark matter itself. Larger interaction rates are the result of larger $\lambda$ or smaller $M$, all else being equal. Therefore, we define our figure of merit as

$$\Lambda^{-1} \equiv \frac{\lambda^2}{4\pi M},$$

where the extra factor of $4\pi$ is inserted by convention to account for phase space: this factor typically appears alongside $\lambda^2$ in cross section calculations. Larger values of this parameter correspond to stronger interactions with the Standard Model, while smaller values correspond to dark matter models which are less coupled to visible matter.

However, it is important to recognize that, due to the difference in techniques that can be applied to search for different classes of dark matter, even though smaller $\Lambda^{-1}$ corresponds to weaker coupling to the
Standard Model, the experimental reach in $\Lambda^{-1}$ is not the same for all types of dark matter. If axions and WIMP dark matter interacted identically with the Standard Model fields, then one would expect that any experiment that could probe one class of dark matter model would (at least in principle) be able to probe the other, albeit with an overall rate that differed by some eighteen orders of magnitude (the squared ratio between the scale $f_a$ and the electroweak breaking scale). This is not at all the case. Axion dark matter contains an interaction between axions and two photons, allowing for axion-photon conversion. Such interactions are completely absent in WIMP models. In contrast, WIMP dark matter is heavy enough to impart significant energy in a WIMP-nucleon scattering event, while axions are far too light. Thus, low-background resonance cavities allow searches for axions with low interaction rates (i.e., high $f_a$) in a $\Lambda^{-1}$ range inaccessible to WIMP dark matter searches, and low-background nuclear recoil detectors allow WIMP searches that are insensitive to axions.

Thus, while it is in general true that lower values of $\Lambda^{-1}$ are more difficult to experimentally probe, one cannot place a universal experimental lower limit on $\Lambda^{-1}$ for all models of dark matter. With this caveat, comparing different models via their $\Lambda^{-1}$ does allow useful heuristic for characterizing different types of dark matter. In general, experiments using a particular technique sensitive to dark matter-Standard Model interactions of a particular form and dark matter masses in a particular range will find smaller $\Lambda^{-1}$ harder to detect.

We now estimate $\Lambda^{-1}$ for various dark matter models.

**Cold dark matter**

The minimum possible interaction a dark matter can have with the Standard Model is a gravitational-strength interaction, and is the closest we can come to a pure non-relativistic, non-interacting CDM paradigm (as described further in the next subsection). Note that there is no mechanism for production of dark matter in this pure-gravity model — only an unspecified appeal to Planck-scale physics. As this is the most minimal model of dark matter (at least as it pertains to dark matter after it is produced), and the behavior of all other dark matter candidates in the evolving Universe can be thought of as departures from this model, we discuss it first. Here, the mediation scale $M$ is the Planck scale, $(10^{19}$ GeV), and so

$$\Lambda_{\text{gravity}}^{-1} \sim 1/m_{\text{Pl}} \sim 10^{-19} \text{ GeV}^{-1}. \quad (8)$$

Note that several of the models which we will discuss later have $\Lambda^{-1} < \Lambda_{\text{gravity}}^{-1}$ (the result of dark matter-Standard Model interactions being mediating by a series of interactions, all of which are very weak). In such cases, the non-gravitational interaction would be subdominant, and the most important dark matter-Standard Model interaction would be the one mediated by the Planck scale. Given the incredibly small coupling that this implies, the difference is somewhat academic, as both would imply interaction strengths far below any potential particle physics detection strategy.

**WIMPs**

Turning to dark matter models that have a plausible production mechanism, WIMP dark matter should have interactions mediated by electroweak-scale particles, with $M$ on the order of the $W$ or the Higgs mass. Couplings are expected to be on the order of the gauge couplings $g \sim 0.65$ or $g' \sim 0.35$. Thus, while specific implementations of WIMP dark matter might have additional mixing angles and tunings reducing the interaction rate, we expect

$$\Lambda_{\text{WIMP}}^{-1} \sim \frac{g'^2}{4\pi m_H} - \frac{g^2}{4\pi m_W} \sim (6-40) \times 10^{-5} \text{ GeV}^{-1}. \quad (9)$$

This estimate applies to many models of WIMP dark matter, including two of the most commonly discussed: supersymmetry (where the dark matter is a linear combination of the superpartners of the photon, the $Z$ boson, and the Higgs) and extra-dimensions (where the dark matter is a Kaluza-Klein excitation of the photon or $Z$). These specific iterations of WIMP dark matter are shown in Figure 1.
**Thermal relic dark matter**

WIMP dark matter is a specific example of dark matter produced through thermal freeze-out of non-relativistic particles in the early Universe. A necessary component of this is a thermally averaged annihilation cross section in the early Universe of \(\langle \sigma v \rangle \approx 3 \times 10^{-26} \text{ cm}^3/\text{s} \) \[60\]. In the case of WIMP dark matter, this annihilation is mediated by the weak nuclear force, however production of thermal dark matter requires only a sufficiently large cross section, not that it proceeds through the weak force. Assuming this annihilation is between dark matter and the Standard Model, then in general the cross section should be proportional to \(\Lambda^{-2} \), giving

\[
\Lambda_{\text{thermal}}^{-1} \approx 5 \times 10^{-5} \text{ GeV}^{-1}.
\]

Note this is somewhat below the low-end of the WIMP \(\Lambda^{-1}\) range; this is not a contradiction, as usually additional \(O(1)\) factors appear in any detailed calculation of \(\langle \sigma v \rangle\) that one would perform in a given WIMP model.

**Dark Photons**

For freeze-out to occur while the dark matter is non-relativistic, thermal dark matter is typically assumed to have a mass between a few GeV and a few tens of TeV. The particle that mediates the interaction between this relatively massive dark matter and the Standard Model is often of a similar mass. For example, in true WIMP dark matter, the mediating particle would be a \(W\), \(Z\), or Higgs boson, all with masses of \(O(100 \text{ GeV})\). However, the mediator could potentially be much lighter, and have escaped detection in Standard Model experiments through highly-suppressed couplings. Such scenarios are interesting because light mediators result in long-range forces which can significantly alter the distribution of dark matter on astrophysical scales, as we will explore in more detail in the next subsection.

As an example, we consider here “dark photons,” (see Ref. \[61\] for a review of the relevant particle physics) where the mediating particle connecting dark matter to the Standard Model is a photon analog with a mass below \(\sim 1 \text{ GeV}\). This would include the possibility that the dark photon is massless. A “dark photon” is a quantum superposition of the mediator and a (slight admixture) of the Standard Model photon. The amount of “photon” contained in the dark photon is the mixing parameter \(\epsilon \ll 1\). The particle physics constraints on dark photons themselves are a subject of much active research in a wide variety of experiments (see Ref. \[62\] for a status update).

In these dark photon models, the dark matter can achieve a thermal cross section through pair annihilation into pairs of dark photons, followed by later decay of the dark photons into pairs of Standard Model particles \[63, 64\]. This divorces the thermal cross section from the interaction rate in a present day experiment, which must connect dark matter to the Standard Model through the dark photon. Thus, we might estimate

\[
\Lambda_{\text{dark photon}}^{-1} \sim \frac{\epsilon g_D e}{\max(m_{\chi}^2/m_{\gamma}, m_{\chi})},
\]

where the requirement of thermal freeze-out sets the ratio of dark photon coupling to dark matter mass to \(g_D^2/m_{\chi} \sim \Lambda_{\text{thermal}}^{-1}\). Taking \(\epsilon \sim 10^{-7} - 10^{-3}\) as required by null search results \[62\] and \(g_D \sim e\) as suggested — although not required — by many Beyond the Standard Model scenarios results in

\[
\Lambda_{\text{dark photon}}^{-1} \sim 10^{-12} - 10^{-8} \text{ GeV}^{-1}.
\]

We note that the concept of a “dark photon” encompasses a wide variety of models (some of which do not assume dark matter is a thermal relic), and the relevant couplings can likewise vary even more than in this rough estimate.

**Asymmetric dark matter**

We can apply the thermal result to a non-thermal model of dark matter. Asymmetric dark matter (ADM) \[65\] produces dark matter through some mechanism similar to baryogenesis (or possibly through a shared mechanism, see Ref. \[86\] for a comprehensive review). However, just as the asymmetry in the number of protons and antiprotons is
only relevant in the Universe today because the number density of thermally produced proton-antiproton pairs is very small (a factor of $10^{10}$ less than the asymmetric proton density), the asymmetric density of dark matter can only be relevant if the symmetric component has been annihilated away. This requires annihilation cross sections in the early Universe at least as large as those of a thermal relic \cite{52}, and so
\begin{equation}
\Lambda_{\text{ADM}}^{-1} \gtrsim 5 \times 10^{-5} \text{ GeV}^{-1}. \tag{13}
\end{equation}

**Axions, QCD and otherwise**

Switching to dark matter that was never in thermal equilibrium with the Standard Model plasma, axions (as previously mentioned) can be produced if the axion field in the early Universe is not located at the minimum of the potential generated by interactions with colored particles. If such a “misalignment angle” occurs after the Universe cools through the QCD phase transition, the axion field will oscillate around the potential minimum. As it settles to the minimum, non-relativistic axions will be produced, providing a dark matter candidate. Assuming the misalignment angle is $\mathcal{O}(1)$ and not unnaturally small, the observed cold dark matter relic abundance realized when $f_a \sim 10^{10} - 10^{12}$ GeV \cite{59}. Larger $f_a$ are possible if the misalignment angle is suppressed for some reason. The coupling between axions and the Standard Model of relevance for experiments such as ADMX \cite{87} is the axion-photon coupling, which is $\lambda^{2}/4\pi = \alpha$ (times known Standard Model parameters). As a result,
\begin{equation}
\Lambda_{\text{axion}}^{-1} \sim e^{2}/(4\pi f_a) \sim 10^{-11} - 10^{-15} \text{ GeV}^{-1}. \tag{14}
\end{equation}

A true QCD axion has a very tight relationship between the coupling of the axion to Standard Model particles and the mass of the axion itself, both being dependent on the single new physics parameter $f_a$ to various powers (times known Standard Model parameters). However, the idea of dark matter as the result of a very low mass field oscillating coherently in the early Universe can be abstracted away from the QCD axion (though in doing so one loses the nice connection with a solution to the Strong CP problem). For reasons that will be made clear when we discuss astrophysical scales, such dark matter is often called “fuzzy” dark matter, and can have very low masses, down to $\sim 10^{-21} - 10^{-22}$ GeV \cite{88, 89, 90, 91, 92, 93, 94, 95, 96}. The interactions with Standard Model particles are very model-dependent, but one might expect
\begin{equation}
\Lambda_{\text{FDM}}^{-1} \sim 10^{-17} - 10^{-19} \text{ GeV}^{-1}. \tag{15}
\end{equation}

However, it must be emphasized that this estimate is very rough, and is highly sensitive to (usually unspecified) details of a Fuzzy Dark Matter model.

**Sterile neutrinos**

Sterile neutrinos are a non-thermally produced dark matter candidate, created in the early Universe through a small mixing angle $\sin \theta$ with the active neutrino species. Due to this mixing, this type of dark matter is unstable, with a lifetime of
\begin{equation}
\tau_{\text{sterile-}\nu} \sim (10^{22} \text{ s}) \sin^{-2} \theta (m_\chi \text{ keV})^{-5}. \tag{16}
\end{equation}

From null searches for the resulting monochromatic gamma-ray, the mixing angle is constrained to be $\theta^2 \lesssim 10^{-5}(m_\chi/\text{keV})^{-5}$ \cite{35, 33, 97}. The mass of sterile neutrino dark matter is constrained to be above $\sim 0.4$ keV by phase space arguments \cite{98} — tighter constraints arise for specific production mechanisms in the context of astrophysical constraints, as discussed in the next section. The interaction figure of merit in the Universe today then is
\begin{equation}
\Lambda_{\text{sterile-}\nu}^{-1} \sim e^{2} \sin \theta/(4\pi m_W) \lesssim 10^{-7} \text{ GeV}^{-1}. \tag{17}
\end{equation}

**Gravitinos**

Supersymmetric dark matter is often considered to be a thermally produced neutralino, with $\Lambda^{-1}$ therefore given by the WIMP values. However, another possibility is gravitino dark matter from supergravity
produced through decay of heavier superpartners. The present-day interaction of gravitinos with the Standard Model would then occur through interactions mediated by particles at the supersymmetric breaking scale $m_F$, which could be anywhere from $10^6$ to $10^{11}$ GeV up to the scale of Grand Unification ($10^{16}$ GeV). For production of gravitinos at TeV-scale colliders (say, at the LHC), the effective coupling of the gravitinos with colliding Standard Model particles will be generated through these mediators with mass $m_F$. The coupling will increase with collision energy, but will be suppressed by $m_F^2$, resulting in a figure of merit of

$$\Lambda_{\text{gravitino}}^{-1} \sim \text{TeV}/m_F^2 \lesssim 10^{-29} - 10^{-9} \text{ GeV}^{-1}. \quad (18)$$

As previously mentioned, for values of $\Lambda^{-1}$ below the scale relevant for purely gravitationally mediated dark matter, the gravitational mediation dominates.

**Primordial Black Holes**

Finally, dark matter might be a relic population of primordial black holes, created early in the history of the Universe [102, 103, 104, 105]. Black holes lighter than $10^{-18} M_\odot$ are ruled out, as their Hawking radiation lifetime is shorter than the age of the Universe. Black holes between $10^{-7} M_\odot - 10 M_\odot$ can be constrained by microlensing [106] searches for MACHO dark matter, those between $10^6 M_\odot - 10^9 M_\odot$ by millilensing of compact radio sources [107], and those above $10^9 M_\odot$ by dynamical constraints on gravitationally bound systems [108, 109]. Dynamical constraints from wide binary stars in the stellar halo of the Milky Way and the stellar distribution within the Milky Way’s satellite population further exclude PBH as a significant contributor to the dark matter budget for PBH masses greater than a few times $10^3 M_\odot$ [110, 111, 112, 113]. Combining all of these constraints (and others) nominally excludes primordial black holes as 100% of dark matter if the mass spectrum is monochromatic [114, 105]. However, astrophysical uncertainties leave three plausible mass ranges where black holes could comprise all of the dark matter: $10^{-17} M_\odot - 10^{-16} M_\odot$, $10^{-13} M_\odot - 10^{-9} M_\odot$, and $1 M_\odot - 10^3 M_\odot$.

As objects with macroscopic masses (though often subatomic radii), it is somewhat difficult to map primordial black holes into our $\Lambda^{-1}$ parameter. A naive identification of $\Lambda^{-1}$ with the Schwarzschild radius would result in an enormous value for the “particle physics” parameter, but this would be misleading, both because no particle physics experiment on Earth can directly access the energy equivalent of even the lightest possible black hole dark matter candidate, and the interaction rate of primordial black holes in the Universe is very low (as their geometric cross section and low number density results in an interaction time which is very small compared to galactic distance scales). As any reasonable definition for $\Lambda^{-1}$ in this case would be much smaller than their gravitational effects, we can safely take

$$\Lambda_{\text{PBH}}^{-1} \sim \Lambda_{\text{gravity}}^{-1} \sim 10^{-19} \text{ GeV}^{-1}. \quad (19)$$

**Baryons**

To put these dark models in context, we consider the subdominant component of the matter distribution most familiar to us: baryons. To astronomers, “baryon” means “every Standard Model particle but neutrinos and photons,” which may be heretical to those with a traditional particle physics upbringing. Because our Universe is charge-neutral and primarily hydrogen by mass, the dominant species of “baryon” by mass is the proton, with a mass of 1 GeV. Determining $\Lambda^{-1}$ is subtle because the baryons can interact through a number of different forces. At astrophysical scales, the most important interactions are electromagnetic. Direct proton-proton scattering in a halo with $v/c \sim 10^{-3}$ via the electromagnetic force resulting in a significant exchange of momentum would have an approximate

$$\Lambda^{-1} \sim \alpha/m_p v^2 \sim 10^4 \text{ GeV}^{-1}. \quad (20)$$

Strong-force nucleon-nucleon scattering has

$$\Lambda^{-1} \sim \text{GeV}^{-1}, \quad (21)$$
and weak-scattering would be
\[ \Lambda^{-1} \sim \alpha/m_W \sim 10^{-4} \text{ GeV}^{-1}. \] (22)

This representative set of dark matter models gives an idea of the large variation in the figure of merit controlling the “visibility” of dark matter to the particle physics experiments. The range of \( \Lambda^{-1} \) values for representative dark matter models we have considered are shown in Figure 1 (horizontal axis), along with the astrophysical parameter discussed next (vertical axis). Note that this parameter spans \( \sim 20 \) orders of magnitude — a wide range, to be sure, but much smaller than the range of either the mass or cross section parameters shown in Figure 2. Moreover, the experimentally accessible range is significantly smaller. The dark matter searches targeting dark matter that was in thermal equilibrium with the Standard Model in the early Universe can cover much of the parameter space above \( \sim 10^{-5} \) GeV, though some dark matter models with very small values of \( \Lambda^{-1} \) can be detected using specialized detection detection techniques (for example, axions and sterile neutrinos).

3.2. A Measure of Astrophysical Scales

The immense range of scales controlling the interaction of dark matter with the visible particles is directly related to the explosion of viable dark matter models that theoretical physicists have devised over the years. This is not to undersell the theoretical motivations underpinning most of these dark matter scenarios. Rather it is an acknowledgment that, without direct experimental evidence, theory alone cannot provide a unique solution to the problem of dark matter. Some of these solutions lend themselves to direct searches in particle physics experiments, but others do not have obvious search strategies given current technology.

However, even for models of dark matter which are invisible to our particle detectors [115, 116], their gravitational imprint on matter and light is unavoidable. This allows for astrophysical probes of dark matter by investigating the structure of the gravitationally bound dark matter halos. Indeed, all of our present knowledge of dark matter comes from such investigations: we know that dark matter was non-relativistic at the time of structure formation [117, 118], at most has self-interactions below the scale of the strong nuclear force [119], and forms gravitationally bound structures down to a mass of at least \( 10^8 M_\odot \) [120, 121, 122, 123].

If dark matter were a purely non-collisional, non-interacting, non-relativistic particle — a pure CDM model — then the power spectrum of the gravitationally-bound dark matter structures (“halos”) existing in our Universe today would be set by the primordial density perturbations of the inflaton field (experimentally known to be close to the Harrison-Zel’dovich spectrum [124]), followed by gravitational evolution as the Universe expanded. Overdensities in the matter field would grow as modes of the density fluctuations entered the cosmological horizon. Once the perturbations became unstable, matter would decouple from the Hubble flow and collapse in on itself and virializing (converting from bulk to thermal kinetic energy) through a process called “violent relaxation” to form bound dark matter halos. Violent relaxation is the process by which the orbits of particles flowing into perturbations get scrambled because of the rapid time evolution of the potential as the perturbations collapse (Sec. 4.10 of Ref. [125]), eventually virializing (e.g., coming to equilibrium) when the potential no longer changes on timescales shorter than the dynamical time. This would result in hierarchical structures with gravitationally-bound dark matter masses extending from super-clusters down to the mass of individual particles of dark matter. The evolution of this power spectrum to the present day and into the non-linear regime has been well-studied [126, 127, 117]. This hierarchy of scales is the strongest, the most striking prediction of CDM cosmology.

3.2.1. Dark Matter Halo Primer

Given the importance of dark matter halos have to the astrophysical probes of dark matter (and therefore to this paper), we provide a primer on the basic concepts for the particle physicist, who may be more used to thinking of individual particles of dark matter. The fundamental macroscopic unit of dark matter is the halo — an overdensity which decoupled from the Hubble flow to collapse and virialize into a gravitationally bound clump. Halos are triaxial [130, 131, 132], but we characterize them in terms of spherically-averaged quantities. Dark matter halos are defined by their virial mass (which we will typically refer to as \( M_{\text{vir}} \) in
Figure 3: Dark matter candidates in an astronomically relevant parameter space. The horizontal axis is dual-labeled, and represents primordial effects. The bottom axis shows primordial effects quantified in terms of the characteristic free-streaming wavenumber of the model; the top axis shows how this translates to non-linear scales, quantified in terms of the half-mode halo mass (see text for details). The vertical axis quantifies evolutionary effects on dark-matter halos and cosmology according to a characteristic interaction or decay rate.

this paper), which is the mass enclosed in a virial radius $R_{\text{vir}}$, in which the average density is some number $\Delta$ times the critical density $\rho_c$ of the Universe:

$$M_{\text{vir}} = \frac{4\pi}{3}(\Delta \times \rho_c)R_{\text{vir}}^3. \quad (23)$$

There is no universal choice for $\Delta$, with different choices having different physical motivation [133] though the variation is only a factor of a few. The most common choices are the virial density $\Delta_v(z)$, which is related to spherical collapse [134]: $\Delta_{200c} = 200$ [135]; and $\Delta_{200b} = 200\Omega_m$. In galaxy cluster literature, $\Delta_{500c} = 500$ is common. Because structural parameters and substructure counts often depend on the definition of the virial radius, it is important to keep track of which virial radius is meant in a particular paper, and translate to another choice of “virial” if necessary. For the order-of-magnitude estimates of interest in this paper, we can treat all these choices of $\Delta$ as approximately the same.

Often we describe halos in terms of velocity instead of mass, because it is typically more robust to uncertainties in the definition of “virial,” also because it represents the characteristic speed or temperature
Figure 4: Dark matter halo scales relative to galaxy scales. Cluster-sized halos typically contain thousands of satellite galaxies, each in their own subhalo. Galaxy-sized halos typically have a few hundred (at most) satellites down to ultrafaint scales. Dwarf galaxies are expected to only host a handful of satellites. For each class of object, we show the typical halo mass if the object is a central or host galaxy (halo), the stellar mass of the central galaxy, and the total stellar mass contained within the halo (including host and satellite contributions). The scaling of other halo and galaxy properties are given as a function of host halo mass, including the half-mode wavenumber $k_{hm}$.

| $z = 0$ | Dwarf | Galaxy | Cluster |
|---------|-------|--------|---------|
| $M_{\text{vir}}$ (central) : | $\sim 10^8 - 10^{11} \, M_\odot$ | $\sim 10^{11} - 10^{14} \, M_\odot$ | $\sim 10^{14} - 10^{15} \, M_\odot$ |
| $M_*$ (central) : | $\sim 10^2 - 10^9 \, M_\odot$ | $\sim 10^9 - 10^{11} \, M_\odot$ | $\sim 10^{12} \, M_\odot$ |
| $M_*$ (total) : | $\sim 10^{-4} \, M_{\text{vir}}$ | $\sim 0.03 \, M_{\text{vir}}$ | $\sim 0.01 \, M_{\text{vir}}$ |

The virial velocity

$$v_{\text{vir}} = \sqrt{GM_{\text{vir}}/R_{\text{vir}}}$$

is the speed a particle would have on a circular orbit at the virial radius. It is also common to see the maximum velocity

$$v_{\text{max}} = \max \left( v_c(r) \right),$$

with the peak of the circular velocity curve,

$$v_c(r) = \sqrt{GM(<r)/r},$$

used as an alternate. Because $v_c(r)$ peaks deep inside the halo, $v_{\text{max}}$ is a good mass proxy when the definition of mass is ambiguous.

Halos grow by ingesting other halos and matter that is not yet subsumed into halos, a process known as hierarchical structure formation. Halos that have not been wholly dissolved by the host halo (a.k.a. the
central halo) are called subhalos. These subhalos are typically tidally stripped of much of their dark matter\cite{136}, with 70-90% mass stripping being typical. Halos reach their maximum mass either at $z = 0$ (if they are central halos), or just before they fall into bigger systems (if they are to become subhalos)\cite{137, 138}. The maximum mass of a halo throughout cosmic time is called peak mass $M_{\text{peak}} = \max(M(z))$. For subhalos in particular, where the definition of halo mass is not uniquely defined, it is usually more useful to characterize them in terms of $v_{\text{max}}$ or in terms of the maximum of $v_{\text{peak}}$ throughout cosmic time, $v_{\text{peak}}$. If $v_{\text{max}}/v_{\text{peak}} < 1$, you can safely assume that the subhalo has been quite significantly stripped.

Dark matter, by its nature, cannot be directly observed, and so the masses and velocities of the halos are not direct astronomical observables. However, halos are the structures in which galaxies are embedded. Galaxies are gravitationally bound collections of stars and/or gas with significant spread in metallicity\cite{139}. The galaxy at the center of the “host” or “central” halo is called the host or central galaxy. Galaxies inhabiting subhalos are called satellite galaxies. Galaxies and their halos, including massive galaxies like the Milky Way, may sometimes be embedded in even larger halos of dark matter, forming galaxy clusters.

One of the most important developments in galaxy evolution theory and observation in recent decades is the realization that galaxy and halo mass (the stellar-mass–halo-mass relation, \textit{SMHM}) are strongly correlated with each other for halos more massive than $10^{11} M_\odot$. This likely also holds true for smaller halos as well (see Figure\\cite{140, 141, 142, 143}), although there are beginning to be hints in simulations\cite{140, 141, 142} and observations\cite{143} that the scatter becomes large. For small halos, it is possible that there is a floor in galaxy mass on scales. As we show in the next section, there is significant uncertainty in how smaller galaxies populate halos. A heuristic mapping between galaxies and halos is shown in Figures\\cite{140, 141, 142, 143}.

Galaxies are small compared to halos — both in volume and in mass. The typical half-light radius of a galaxy is of order a couple of percent of the virial radius\cite{143, 144, 145, 146}. The mass in cold gas and stars is never more than a few percent of the total mass of a halo\cite{138, 151}, although large ($M_{\text{vir}} > 10^{12} M_\odot$, Milky Way mass and above) halos retain significant reservoirs of shock-heated gas throughout their volume. The stellar and gas contents of galaxies are used to trace the dark matter potential wherever the stellar and gas density is high enough to admit an observation. Spiral galaxies are “rotationally supported” — stars and gas travel in nearly circular, coplanar orbits. Observations of this motion are used to create rotation curves, the circular velocity as a function of distance $r$, Eq.\ (26). Spheroidal and elliptical galaxies are “dispersion supported” — stars and gas exhibit random motions, characterized by a velocity dispersion $\sigma(r)$, and the potential can be recovered under assumptions of virial or hydrostatic equilibrium.

The standard \textit{Hubble Sequence} galaxies, ellipticals and spirals, typically inhabit halos of $M_{\text{vir}} \sim 10^{11} M_\odot$. Giant elliptical galaxies — dispersion-supported systems with stellar mass $M_\star = 10^{11} - 10^{12} M_\odot$ — are hosted by large halos ($M_{\text{vir}} \gtrsim 10^{13} M_\odot$), whereas spiral galaxies are more common at scales of $M_{\text{vir}} \sim 10^{12} M_\odot$ (this is the approximate mass of the Milky Way). Dwarf galaxies span an enormous range of stellar mass and origin. Dwarfs with $M_\star \gtrsim 10^5 M_\odot$ are generally forming stars today if they are central galaxies, and have a large gas-to-stellar mass ratio\cite{152}. If they are satellites, they are typically devoid of gas and recent star formation\cite{153}. In the Milky Way, such galaxies are called classical dwarfs. Dwarfs with $M_\star \lesssim 10^5 M_\odot$ are classified as ultrafaints, and are thought to be ancient relics of star formation prior to the reionization of the Universe\cite{154}. Reionization heats the intergalactic medium to the point where it cannot condense into small halos, preventing small halos from forming stars at all but the earliest times\cite{155}. We return to this point in Section\\cite{156}.

While dark matter halos grow by ingesting smaller halos, much of the galaxy formation occurs in situ. Satellite galaxies destroyed by the host dissolve into the stellar halo (called \textit{intracluster light on cluster scales}). Because the efficiency of star formation drops rapidly with decreasing halo mass (Figure\\cite{140}, the stellar halo of the Milky Way contributes only $\sim 1\%$ of the total stellar mass of the Milky Way\cite{125, 156}. Most of the stellar halo derives from a single or a few massive progenitors\cite{157}. The kinematics of halo
stars may be used to map out the dark matter distribution of the host.

3.2.2. Particle Physics and Halo Mass

In the standard Cold Dark Matter paradigm, the hierarchical structure of large dark matter halos containing smaller halos, which themselves contain smaller halos (i.e., the halo of a galaxy cluster containing many galactic halos, which contain subhalos of varying sizes, and with their own substructure) should continue down to extremely small scales. However, if there exists interactions either between dark matter particles or between dark matter and the Standard Model, then number and structure of these halos can be modified. Depending on the nature of this new interaction, we should expect the modification to only become apparent in halos of some characteristic physical size.

Therefore, deviations from the CDM model can be expressed in terms of the largest physical scale on which deviations from CDM appear noticeable in simulations or analytic theory. This scale is often expressed in terms of a halo mass, \( M_{\text{halo}} \), to make contact with astronomical tests on non-linear scales. This mass scale can also be recast in terms of a distance scale, typically in terms of a comoving wavenumber \( k \) (an inverse distance), to make contact with linear theory. This wavenumber might be set by free-streaming \( k_{\text{fs}} \) or kinetic decoupling \( k_{\text{kd}} \). We will comment further on exactly how this link is made in the discussion of warm dark matter candidates below. Figure 4 gives a sense of scale of the kinds of dark-matter halos which are traced by different types of visible structure, and a corresponding wave number; we expand on this figure in Section 4.

Measurements which probe dark matter gravitational structures on a variety of scales can therefore be used to constrain the theoretical model space. Deviations from the CDM model can take two forms, illustrated in Figure 3. The deviations can be primordial, when non-trivial physics terminates or modifies the gravitational collapse of dark matter at some comoving wavenumber \( k \). It translates to a characteristic mass scale today below which the number of dark matter halos is significantly reduced. Note that if the inflationary power spectrum has non-trivial structure at small scales (rather than the nearly scale-free spectrum measured on large scales) or if the reheating temperature is low, there may be non-trivial structure in the halo mass function even for CDM-like particle dark matter models [158, 159].

Alternatively, the deviations can be evolutionary, driven by the interactions of dark matter over the history of the Universe since the original structure formation, with a characteristic rate \( \Gamma \). These effects can erase existing structure or change the cosmic density and velocity distributions of dark matter from their non-evolutionary baseline. While evolutionary deviations are typically characterized by a time scale (e.g., a lifetime for unstable dark matter, or a scattering rate for interacting dark matter), as shown with the vertical axis in Figure 3, the precise choice for this axis has the same type of ambiguity as the \( \sigma_{\text{int}} \) interaction cross section in Figure 2. For example, interaction rates are certainly density-dependent (and likely velocity-dependent as well), and so change from location to location in the Universe. Instead of classifying the evolutionary deviations of particle models by their late-time interaction or decay rates, we choose to characterize evolutionary processes in terms of a halo mass scale \( M_{\text{halo}} \) as well. Aside from being pragmatic, this choice has a firm basis in phenomenology. For example, the kinematics of late-time processes, such as late-time decays or self-interactions, have drastically different impact on halos of different masses [160, 161]. While relativistic decays leave their mark on halos of all sizes, non-relativistic decays perturb small halos significantly while leaving large halos untouched [160]. Hidden sector models typically yield strongly velocity-dependent self-interaction cross sections, leading to a “resonant” effect when the halo virial velocity is well-matched to peak of the self-interaction cross section [162, 163, 119]. Thus, models with evolutionary effects can be mapped to \( M_{\text{halo}} \).

The effects of “primordial” and “evolutionary” deviations cannot be cleanly divided in any realistic model of dark matter particle physics — for example, extra interactions which suppress the inner density of dark matter halos also may lead to a suppression in the primordial power spectrum [164]. Moreover, any suppression in the matter power spectrum additionally leads to a delay in the formation time of halos relative to a pure \( \Lambda \)CDM model with a Harrison-Zel’dovich inflationary power spectrum. Thus, there may be extra \textit{time-dependent} signatures of new dark matter physics even if the origin is purely primordial [165, 166, 167, 168]. Despite this intrinsic mixing it is useful to divide the possible sources of power spectrum deviations in this way for the purposes of this paper, always keeping in mind that applications to any
particular model of particle physics necessarily involves a more detailed calculation which includes both the primordial power spectrum and its evolution. It is convenient to define a single parameter to characterize a scale on which dark-matter microphysics leads to significant departures from CDM predictions, shown for the models described below in Figure 1.

3.2.3. Primordial Deviations

In the standard ΛCDM model of the Universe, inflation lays down a nearly scale-invariant power spectrum of potential perturbations, described by their wavelength $\lambda$ or wavenumber $k = 2\pi/\lambda$. If dark matter is pure CDM, there is no time-evolution in the linear power spectrum other than the standard suppression of growing modes for large $k$ before matter-radiation equality. The non-linear halo mass function $dn/dM_{\text{vir}}$ (number density of halos per unit mass) is thus also nearly scale invariant below cluster scales (Figure 3), $dn/dM_{\text{vir}} \propto M^{-1.9}_{\text{vir}}$ [169]. Non-minimal models of dark matter which influence the power spectrum of dark matter halos during the period of initial Jeans collapse typically lead to cut-off at the comoving wavenumber $k_{\text{cut}}$, although this cut-off is sometimes heavy in features and is rarely a step function.

However, the effects of a truncation in the matter power spectrum typically become apparent on much larger scales in the non-linear halo mass function. The effects of non-minimal dark-matter physics on the linear matter power spectrum are usually expressed in terms of a transfer function

$$T_{\text{DM}}(k) = \left( \frac{P_{\text{DM}}(k)}{P_{\text{CDM}}(k)} \right)^{1/2},$$

(27)

where $P_{\text{CDM}}(k)$ is the linear CDM power spectrum and $P_{\text{DM}}(k)$ is the power spectrum for an arbitrary dark matter model. Ref. [170] define a “half-mode mass” $M_{\text{hm}}$ corresponding to the wavenumber $k_{\text{hm}}$ at which $T(k_{\text{hm}}) = 1/2$. This is also approximately the point at which the halo mass function drops by a factor of order two with respect to CDM when modes become non-linear.

The half-mode wavenumber can be converted to an equivalent halo mass, below which the number density of halos would deviate significantly from the predictions of hierarchical structure formation in CDM:

$$M_{\text{hm}} = \frac{4\pi}{3} \bar{\rho} \left( \frac{\pi}{k_{\text{hm}}} \right)^3,$$

(28)

where $\bar{\rho}$ is the comoving matter density of the Universe [170, 171]. In other words, this is mass enclosed in an average patch of the Universe with radius $\pi/k_{\text{hm}} = \lambda_{\text{hm}}/2$. For primordial effects, we choose

$$M_{\text{halo}} = M_{\text{hm}}.$$  

(29)

For the most frequently encountered (thermal relic warm dark matter) transfer functions, $k_{\text{cut}} \sim 10 k_{\text{hm}}$, or $M_{\text{hm}} \sim (a \text{ few}) \times 10^3 M_{\text{cut}}$, where $M_{\text{cut}}$ is the halo mass corresponding to $k_{\text{cut}}$ [170]. We caution that when encountering a new model, it is always best to calculate the half-mode mass directly from the power spectrum.

There are two primary physical pathways leading to an early-time truncation or suppression in the matter power spectrum. First, for particles may free-stream out of small density perturbations if they “born hot”, i.e., have a (semi-) relativistic momentum distribution in the early universe. This typically leads to a smooth truncation of the matter power spectrum. Second, interactions with Standard Model or other hidden-sector particles can lead to truncations that have acoustic-type oscillations structure. Often, both types of primordial features are present in the power spectrum.

Thermal relics

Assuming dark matter is a thermal relic, the power spectrum is exponentially suppressed once the dark matter kinetically decouples from the Standard Model particle bath. This occurs at a temperature $T_{\text{kd}}$, with the approximate relation [172, 173, 174, 175]

$$M_{\text{halo}} \sim (10^{-1} M_\odot) \left( \frac{T_{\text{kd}} g_{\text{eff}}}{50 \text{ MeV}} \right)^{1/4} \text{ }^{-3},$$

(30)
where \( g_{\text{eff}} \) is the number of effective degrees of freedom in the thermal bath at \( T_{\text{kd}} \). For an arbitrary thermal relic model with mass \( m_\chi \) and scattering via the \( \ell \)th partial wave is given by

\[
m_\chi T_{\text{kd}} \sim \left[ \frac{\xi(3)}{\pi^2} \left( \frac{90 g_{\text{eff}}}{8 \pi^2} \right)^{1/2} m_{\text{Pl}} m_\chi \sigma_0^{\text{el}}(m) \right]^{1/(3+\ell)}.
\]  

(31)

Here \( \sigma_0^{\text{el}} \) is the elastic scattering cross section. For 100 GeV dark matter, if \( \sigma_0^{\text{el}} \) is the thermally averaged cross section \((3 \times 10^{-26} \text{ cm}^3/\text{s})\) then

\[
T_{\text{kd}} \sim 17(154) \text{ MeV}
\]  

(32)

for \( \ell = 0(1) \). Note that the assumption that \( \sigma_0^{\text{el}} \) is exactly the thermal cross section is not typically realized in a specific model of thermally produced dark matter. That said, in this simplified model, our assumptions result in a minimum halo mass of

\[
(M_{\text{halo}})^{\text{thermal}} \sim 10^{-5} M_\odot - 10^{-2} M_\odot.
\]  

(33)

Free streaming can further damp the matter power spectrum, with

\[
k_{fs} \approx (10^6 \text{ Mpc}^{-1}) \sqrt{\frac{(m_\chi/100 \text{ GeV})(T_{\text{kd}}/30 \text{ MeV})}{1 + 0.05 \ln(T_{\text{kd}}/30 \text{ MeV})}}
\]  

(34)

The relative importance of free-streaming and kinetic decoupling depends on the specifics of the model. For the thermally-coupled models considered here, free-streaming is less important than kinetic decoupling, and so is ignored.

Note that our numbers for \( M_{\text{halo}} \) may appear large to experienced practitioners. As described above, this is because we are considering \( M_{\text{halo}} \) to be the scale on which deviations from CDM begin to become important, rather than (as is usually done) calculating the minimum possible halo mass. The difference between the two halo masses can be orders of magnitude in scale. This prediction can be refined in specific models of thermal relic dark matter, where the elastic scattering cross section can be calculated from fundamental parameters and related to the annihilation cross section.

**WIMPs**

For supersymmetric WIMP dark matter, scanning over the parameter space of the minimal supersymmetric Standard Model, \( T_{\text{kd}} \) can vary from 15 – 1500 MeV [177], resulting in

\[
(M_{\text{halo}})^{\text{SUSY}} \sim 10^{-8} M_\odot - 10^{-2} M_\odot.
\]  

(35)

A similar scan over the parameters of Kaluza-Klein dark matter from models with extra spatial dimensions found

\[
(M_{\text{halo}})^{\text{extra-dim}} \sim 10^{-2} M_\odot - 1 M_\odot.
\]  

(36)

**Dark Photons**

Dark matter charged under a new force with a massless or nearly massless gauge boson can induce dark acoustic oscillations in the early Universe. This will suppress the number of small halos [40], as well as decrease the density profile of smaller dark matter halos [164]. While it is in principle possible to construct dark photon models that suppress structure on almost any scale (by appropriate selection of masses and couplings), we take the parameters of Ref. [164] as an upper limit, and estimate

\[
(M_{\text{halo}})^{\text{dark photon}} \lesssim 10^{12} M_\odot.
\]  

(37)
Asymmetric dark matter

Asymmetric dark matter models could stay in kinetic equilibrium for much longer than thermal dark matter, due to the larger interaction cross section required to suppress the thermal component of dark matter. Thus, considering only the interactions required to annihilate away the thermal component, we can estimate
\begin{equation}
(M_{\text{halo}})_{\text{ADM}} \lesssim 10^{-1} M_\odot,
\end{equation}
with the upper limit in a particular model set either by the decoupling of the dark matter from the baryons, or by the collapse of the baryon sound speed at the QCD phase transition, depending on the details of the model. If additional dark matter physics is added to an asymmetric model (for example, dark photons), then $M_{\text{halo}}$ could be much larger.

Axions, QCD and otherwise

It has been suggested that axion dark matter formed from field misalignment could form small-scale Bose-Einstein condensates (BEC), taking the form of high-density "droplets" of axions with a mass \cite{178, 179, 180, 181, 182, 183, 184, 185}
\begin{equation}
(M_{\text{halo}})_{\text{axion-droplet}} \sim 10^{-10} M_\odot - 10^{-13} M_\odot,
\end{equation}
with the droplet mass proportional to $f_a^2$. These droplets then cluster into larger cosmological structures, as in other forms of dark matter. This primordial deviation from the cold dark matter power spectrum is distinct from the possibility of large-scale axion BECs, which would effect the structure of galaxies, and will be discussed in the next section.

The axion-like Fuzzy Dark Matter models also can alter the structure of dark matter halos, in this case by erasing structure below the de Broglie wavelength of the dark matter. By focusing on ultra-light $10^{-21} - 10^{-22}$ eV particles, this wavelength can be made larger than dwarf galaxies, suppressing the number of structures below \cite{186}:
\begin{equation}
(M_{\text{halo}})_{\text{FDM-pr}} \sim 10^{10} M_\odot \left( \frac{m_\chi}{10^{-22} \text{ eV}} \right)^{-4/3}.
\end{equation}
These models have evolutionary effects in addition to this cut-off in the primordial power spectrum (see also Ref. \cite{186}). Note that the choice of dark matter mass is driven purely by astrophysical phenomenology.

Sterile neutrinos

Sterile neutrinos are a warm dark matter candidate \cite{32, 31, 34, 187}, with a mass range of $\sim 0.4 - 10^5$ keV. The lower end of this mass range is set by astrophysics: for lighter sterile neutrinos, the structure of dark matter halos would deviate too much from observations of the Lyman-$\alpha$ forest or from Milky Way satellite counts \cite{115, 122, 123}. The mixing angle to the active neutrinos $\sin^2 \theta$ is proportional to $m_\chi^{5/2}$ \cite{97}. The free-streaming length depends strongly on the sterile neutrino production mechanism \cite{188}. Contrary to how limits are almost always presented in the literature, there is no general one-to-one mapping between sterile neutrino mass and a free-streaming scale. However, for the specific choice of the neutrino minimal standard model ($\nu$MSM) \cite{97}, and for some non-resonant production models (requiring a primordial lepton asymmetry) \cite{32, 34, 187}, the free-streaming length is approximately
\begin{equation}
k_f \sim \left( 0.5 \text{ Mpc}^{-1} \right) \left( \frac{m_\chi}{\text{keV}} \right)^{5/6}.
\end{equation}
This results in a range of $M_{\text{halo}}$ for sterile neutrinos of
\begin{equation}
(M_{\text{halo}})_{\text{sterile-$\nu$}} = 10^{-6} - 10^{11} M_\odot.
\end{equation}
The presence of small-scale structure has long been used to constrain neutrino dark matter \cite{189}.

This astrophysical figure of merit is therefore correlated to the particle physics parameter $\Lambda^{-1}$, which is proportional to $\sin^2 \theta \propto m_\chi^{-5/2}$. As a result, $\log \Lambda^{-1} \propto \frac{5}{6} \log M_{\text{halo}}$. As mentioned previously, the exact production mechanism for the sterile neutrinos can modify this relationship, introducing additional dependence of $M_{\text{halo}}$ on $\sin^2 \theta$, but for our purposes, this rough estimate suffices.
Gravitinos

Similarly, gravitino dark matter has a dark matter mass set by the supersymmetry breaking scale $m_F$

$$m_\chi \sim \frac{m_F^2}{m_{Pl}}.$$  \hfill (43)

With $m_F$ in the range of $10^6 - 10^{11}$ GeV, the gravitino masses can range from $\sim 100$ eV up to 100 TeV. The resulting free-streaming length sets a minimum halo mass range of

$$(M_{\text{halo}})^{\text{gravitino}} \approx 10^{-17} M_\odot - 10^{13} M_\odot.$$ \hfill (44)

As with the sterile neutrinos, we expect the particle physics and astrophysics figures of merit to be correlated, with $\log \Lambda \propto -\frac{1}{3} \log (M_{\text{halo}})$.

Self-interacting dark matter (SIDM)

In recent years, there has been a great deal of interest in models of SIDM \cite{6, 190}, driven in large part by the possibility of evolutionarily altering the structure of dark matter halos relative to CDM (see our discussion of evolutionary deviations below). Motivations for these models \cite{191} will be discussed in more detail in Section 4. Here, we consider a subset of SIDM models which also cause a primordial deviation from the cold dark matter power spectrum.

Such deviations can occur if the SIDM model contains a relatively light force carrier, such as the dark photon discussed in the previous section \cite{37, 172, 192, 193, 39} or dark pion from a confining sector \cite{194}. Though massless states in the dark sector face severe constraints from CMB measurements \cite{42, 40}, a relatively low-mass force mediator can keep the dark matter in kinetic equilibrium with itself long after decoupling from the Standard Model. This can suppress and alter primordial structure both through collisional damping and “dark” acoustic oscillations (analogous to baryon acoustic oscillations) \cite{40}.

The resulting deviations are more complicated than a straightforward exponential suppression of the power spectrum, and can include a reduction of the central densities of collapsed dark matter halos, caused by a delay in the formation time. A full calculation of the effects of a particular model of SIDM on the primordial power spectrum requires specialized $N$-body simulation \cite{164}, but it has been demonstrated that SIDM models with light mediators can affect the primordial power spectrum on scales corresponding to

$$(M_{\text{halo}})^{\text{SIDM-pr}} \lesssim 10^{11} M_\odot.$$ \hfill (45)

There are numerous mechanisms for the generation of SIDM dark matter, so we cannot assign a single particle physics parameter $\Lambda$ to all of these models.

Primordial Black Holes

As discussed above, it is possible for primordial black holes to be 100% of the dark matter in the Universe if their mass spectrum falls in one of three windows:

$$(M_{\text{halo}})^{\text{PBH}} \approx 10^{-(17-16)} M_\odot, 10^{-(14-9)} M_\odot, 10^{(0-3)} M_\odot,$$ \hfill (46)

due to a combination of constraints from MACHO microlensing searches, disruption of gravitationally bound binary objects, and searches for Hawking radiation from the evaporation of light black holes \cite{105, 105}. We treat these ranges of PBH masses as $M_{\text{halo}}$ for this model.

Gravity-Only Interactions

If dark matter interacts only via gravity, the mechanism for dark matter production is completely unspecified. As a result, no general statement can be made about the deviation of gravity-mediated dark matter structure from the predictions of pure cold dark matter.
Baryons

As in Section 3.1, we set the primordial effects of baryon physics in contrast to these dark-matter models. Because baryons are tightly coupled to photons in the early Universe, sound waves propagate up until the point of recombination. Not only does this leave an imprint in the cosmic microwave background, but leads to the set of “wiggles” in the matter power spectrum, the BAO that we discussed in Section 2. In contrast to the dark photon model discussed above, the BAO do not lead to a cut-off in the matter power spectrum because baryons are a subdominant form of matter. In a baryon + CDM model, the scale on which these wiggles occur is \( \sim 100 \) Mpc, well in the linear regime today [195]. These result in features in the two-point correlation function of galaxies rather than in galaxy or halo number counts.

A second primordial effect imprints itself onto the matter power spectrum in different patches of the Universe — in other words, there is a large scale-dependence of this small-scale effect. It arises from the relative velocity between dark and baryonic perturbations at the epoch of recombination [196]. The relative velocity between baryons and dark matter suddenly becomes supersonic for baryons as photons break free from the previously tightly coupled baryon-photon fluid. This bulk flow of baryons with respect to dark matter suppresses the growth of small halos at early times [197, 198], on scales of \( M_{\text{vir}} \sim 10^8 M_{\odot} \), but the suppression is modulated on \( \sim 100 \) Mpc scales.

3.2.4. Evolutionary Deviations

The structure of dark matter halos may also be altered by late-time effects. These effects are distinct from the suppression of power in the primordial spectrum, though many models of dark matter may cause both primordial and evolutionary deviations. Primordial deviations may lead to time-dependent changes in the growth of structure relative to CDM, notably that structure forms later when there exists a cut-off in the matter power spectrum [165, 199, 200]. But we focus this section on evolutionary deviations that have their origin in ongoing interactions or decays, rather than time-dependent but primordial-in-origin effects.

WIMPs and thermal relics

Dark matter-dark matter interactions become important in reshaping halos when the per-particle interaction rate \( \Gamma \sim \langle \sigma v \rangle \rho / m_\chi \) is of order the inverse of the local dynamical time \( t_d \sim r/v \sim 1 \sim 100 \) Myr, or higher. The central regions of halos have densities of order \( \sim 0.1 M_{\odot}/pc^3 \sim 10^{-23} \) g/cm\(^3\) or less [201]. If we assume that any dark matter self-interaction cross section is of order the thermal relic cross section \( \langle \sigma v \rangle \sim 10^{-26} \) cm\(^3\)/s, and that the WIMP is \( \sim 100 \) GeV in mass, then \( \Gamma \sim 10^{-27} \) s\(^{-1} \sim 10^{-11} \) Gyr\(^{-1} \). Thus, the relative evolution of WIMP-like thermal relic dark matter halos vs. pure CDM halos is completely negligible.

Axions: QCD and otherwise

Axion models may form high-density “droplets” from BEC effects in the early Universe, of order \( M_{\text{halo}} \sim 10^{-10} M_{\odot} \) in size (see our previous discussion) [184]. It has also been argued that Milky Way-sized dark matter halos composed of axions would form caustics from long-range BEC effects [202, 203, 204, 205]. These deviations do not come in the form of the complete suppression of structure on any mass scale, but rather alterations of the density and velocity distributions on the physical scale of \( \sim 10 \) kpc, which is to say a mass scale of

\[
(M_{\text{halo}})^{\text{axion-BEC}} \sim 10^{12} M_{\odot},
\]

(47)

Recent studies of axion physics call the existence of the long-range BEC into question [184], finding that the attractive interactions that are typical between axions cannot lead to long-range correlations [206]. Ref. [184] argues that repulsive interactions may allow long-range forces and large-scale BEC to exist, but Ref. [206] shows that it is difficult to construct a light scalar theory with repulsive, not attractive, interactions. We include the possibility of large-scale BEC here for completeness even though the current conventional wisdom for QCD axions comes down against it (see also Ref. [207]).

Fuzzy dark matter, in addition to erasing structure on scales below the de Broglie wavelength, also acts as a quantum fluid, rather than a collection of discrete particles. This can result in a “solitonic core” [208, 209, 96], which affects the inner kpc of the galaxies that form above the primordial \( M_{\text{halo}} \) cut-off. Dark-matter-only
N-body simulation indicates that this softens the dark matter cusp predicted by CDM into a core \cite{209}, resulting in deviations from the predictions of CDM on the scale of

$$(M_{\text{halo}})^{\text{FDM-ev.}} \sim 10^{8} M_{\odot} - 10^{10} M_{\odot}.$$  \hspace{1cm} (48)

In addition, quantum interference outside the core can lead to clumpy structure throughout the rest of the halo \cite{210, 205, 208, 211, 200, 212}.

**Sterile neutrinos**

In Section 3.1, we found that the sterile neutrino lifetime $\tau_{\text{sterile-}\nu} \sim (10^{22} \text{ s}) \sin^{-2} \theta \left(\frac{m_{\chi}}{\text{keV}}\right)^{-5}$. For typical values of $m_{\chi}$ and $\sin^{2} \theta$, this decay time is many orders of magnitude larger than the age of the Universe. Only a tiny fraction of sterile neutrinos decay, so sterile neutrinos have negligible evolutionary deviations from CDM. Their deviations are essentially purely primordial.

**SIDM**

Interactions lead to the redistribution of energy and angular momentum of particles in individual dark matter halos, typically on scales where the per-particle interaction rate $\Gamma \sim \rho \sigma v / m_{\chi}$ is of order $1/t_{d}$, where $t_{d}$ is the dynamical time and $\sigma$ is the (velocity-dependent) self-interaction cross section \cite{6, 161, 213}. Note that the cross section to which we refer here is the momentum-transfer or viscosity cross section, depending on the specific SIDM model \cite{163, 41}. The redistribution of energy and angular momentum leads to the formation of flat, spherically symmetric density cores. For velocity-independent cross sections, these cores become significant at approximately the same fraction of the virial radius across a wide range of halo mass (up to cluster scales) \cite{214}. For velocity-dependent cross sections, cores can show up for only a narrow range of halo mass, leaving halos outside this mass range relatively unperturbed with respect to CDM \cite{215, 21, 162, 216, 119, 190}.

Under some circumstances, interactions trigger central densities that may be higher than predicted in CDM theory, at least in a subset of halos. First, if the scattering cross section (even purely elastic) is sufficiently high, it is possible to trigger a “gravothermal catastrophe,” in which case the halo center becomes increasingly dense with time \cite{217, 218, 219, 220}. Cross sections typically need to be $\gtrsim 10 \text{ cm}^{2}/\text{g}$ for this process to begin \cite{221}. Second, if some fraction of dark matter has significant self-interactions, including massless mediators allowing for radiative cooling, then more exotic deviations are possible \cite{41}, for example the formation of dark disks in spiral galaxies \cite{222}.

Interactions may become important when halos accrete smaller halos, in the process of hierarchical structure growth. Originally, a major motivation for SIDM models was to eliminate subhalos by the ejection of particles from, and eventual evaporation of, subhalos through particle interactions with the host halo \cite{6}. Although this mechanism is largely inefficient for currently allowed cross sections \cite{223, 224, 225}, which are of order the scale of the strong nuclear force, the idea that mergers can probe interactions is alive on cluster scales ($M_{\text{halo}} \sim 10^{14} M_{\odot} - 10^{15} M_{\odot}$) \cite{226, 227, 228, 229, 230, 231, 232}. We will discuss this more in Section 5.

Thus, SIDM models can affect dark matter structure on the scales

$$(M_{\text{halo}})^{\text{SIDM-ev.}} \lesssim 10^{15} M_{\odot}. \hspace{1cm} (49)$$

**Baryons**

Baryons are special for two reasons: their interaction rates are high, and they can dissipate energy in these interactions. Thus, after halos virialize, baryons can continue to cool, as we showed in Section 2. Baryons condense in the centers of halos. As we discuss in Section 4, baryons can alter the total matter (baryon + dark matter) density profile on scales

$$M_{\text{vir}} \gtrsim 10^{8} M_{\odot}. \hspace{1cm} (50)$$

24
Below this scale, baryons cool inefficiently, which we see by the very high mass-to-light ratios in dwarf galaxies. The dynamical effects of baryons on the matter profile for small halos are limited in scope. Above this scale, baryons do affect halo structure, by an amount still under debate. Thus,

\[ 10^8 M_\odot < M_{\text{halo}} < 10^{15} M_\odot. \]  

(51)

For completeness we note that if baryons scattered significantly but elastically, it would also alter the matter density profile, but not dramatically unless the scattering were high enough to trigger a gravothermal catastrophe \[^{125}.\]

To conclude this section, we note that the scale on which departures are expected can span a range of halo mass below \( M_{\text{halo}} \). To discover dark matter, the key is to see consistency in departures from CDM across a range of scales.

4. Open Problems: Hints and Known Systematic Uncertainties

While for particle theorists, it is natural to consider a wide range of particle models for dark matter, for astronomers, one paradigm reigns: CDM. All observations on small scales are couched in terms of how well they are or are not perceived to be in line with the two main predictions of the CDM paradigm. In CDM,

1. There exists a nearly scale-free hierarchy of bound dark matter structures, called halos, down to astronomically irrelevant scales.
2. Dark matter halos are stratified in density, sharply cusped in their centers.

Any deviation from this paradigm may suggest non-minimal extensions to the CDM paradigm, of the type we described in Section 3. In recent years, there have been hints from data implying that there are deviations from the predictions of CDM, affecting the structure and evolution of astrophysical systems. These hints are the topic of a vigorous debate in the astronomical community, largely over the interplay between dark matter and baryonic physics. For particle theorists, the hints have influenced the development of specific types of dark matter models. This is reflected in the number of dark matter models in Figure 1 which have \( M_{\text{halo}} \) on the range \( 10^8 M_\odot - 10^{10} M_\odot \). In both fields, we must fully understand both baryonic effects on halos if we are to continue exploring dark matter interactions via astrophysics, even if the current set of hints are the result of the interplay between CDM and baryonic physics.

In this section, we discuss the present state of observations, which have been interpreted as hints of physics beyond CDM; and the major theoretical uncertainty complication in transforming hints to measurements, the dynamical interplay between galaxy formation and dark matter physics. We describe the current state of these possible hints of new physics from astrophysical observations, namely (see also Ref. \[^{233}:\]

1. The cusp/core problem (see Figure 5 for a summary),
2. The Missing Satellites problem (Figure 5),
3. “Too Big To Fail” (Figure 7),
4. The baryonic Tully-Fisher relation (Figure 8) and “too big to fail in the field” (Figure 6).

We will end the section with a summary on the state of simulations that include galaxy evolution physics, and show that it is not currently possible to unambiguously determine whether these hints are evidence of new physics or the result of the gravitational interplay between dark matter and baryons. We summarize our main conclusions about these hints of possible deviations from CDM in Figure 9. A key insight in this section is that a major observational uncertainty behind the lack of resolution to these problems is the not-well-quantified relationship between galaxy mass and halo mass on dwarf-galaxy scales. In Section 5, we describe a path forward to break through present-day barriers to progress, summarized in Figure 10. We recommend that readers less familiar with galaxies and halos consult Section 3.2.1 before continuing.
4.1. Hints

There are a number of identified “problems” on small scales. We discuss each in turn, and show where they land in astrophysical parameter $M_{\text{halo}}$. A summary of these problems can be found in Figure 9. For an expanded discussion of these problems, see Ref. [233].

Cusp/Core Problem

From the time of the earliest CDM simulations and calculations, it was found that dark matter halos should achieve extremely high densities at their centers (e.g., Ref. [236]). The primordial fine-grain phase-space density, which sets the upper bound on the central phase-space density of the halo, is high compared to any material with non-zero velocity dispersions or interactions [237]. Thus, both the central phase-space density and the configuration-space mass density (the integral of the phase-space density over velocity space) are high compared to the hot baryonic halo gas permeating large dark-matter halos. By the mid-1990’s, it was clear from dark-matter-only simulations that CDM halos did not follow the scale-free $\rho \propto r^{-2}$ isothermal profile, but instead had another nearly universal profile, a broken Navarro-Frenk-White (NFW) power-law profile [234, 135].

\[ \rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}. \]  

(52)

Since then, there have been refinements to the inner slope of the density profile [235, 239], but in general the $\rho \propto r^{-1}$ (at small $r$) cusped profile fits CDM-only simulations well, on average.

Evidence that this simulation-derived profile did not match observed mass profiles appeared around the same time as the original NFW papers (Refs. [234, 135]). The first hint came from low surface-brightness galaxies (LSBs): spiral galaxies with unusually diffuse disks [240]. Because these disks have low baryon density, the rotation curves of these galaxies should be dark matter dominated. It was found that these...
galaxies have rotation curves which rise (almost) linearly as a function of radius \( r \), which are characteristic of halos having \( \rho \propto r^{-\alpha} \) with \( \alpha \approx 0 \), that is, a core (Figure 5). There are actually two discrepancies: in the shape of the density profile, where this can be measured; and the mean density within some radius, which is more robustly measured for most galaxies. This discrepancy in the density profile between CDM simulations and observations at small radii has become known as the “cusp/core problem.”

Since then, deviations from NFW predictions have been found in all systems for which dynamical estimates of the density profile are possible. We step through the evidence from small-halo to large-halo scales. Notably, there is as yet no observation of the density profile of the smallest known galaxies: ultrafaint dwarfs (stellar mass \( < 10^5 M_\odot \), \( M_{\text{vir}} \) likely of order \( \sim 10^6 M_\odot - 10^7 M_\odot \)). This is on account of their paucity of stars (but see Refs. 247, 248 for ideas on photometric estimates of density profiles). For these systems, we are in the near-term limited to estimates of the mean density within the half-light radius of each system. The density is remarkably constant as a function of stellar mass, a fact which is actually consistent with CDM interpretations 201, 249.

For somewhat larger objects, like the classical dwarf satellites of the Milky Way (with stellar mass \( M_* \gtrsim 10^5 M_\odot \), many of which were recently star-forming) or the late-type (spiral) dwarf galaxies in the Little Things surveys 250, 251, \( (M_{\text{halo}} \sim 10^{10} M_\odot - 10^{11} M_\odot ) \), it is clear that their central densities are lower than expected in CDM-only simulations 252, 253. The existence of cores appear likely outside the Milky Way 254, 255.

For the Milky Way satellites, there is still a vigorous debate as to whether they have cusped or cored density profiles (e.g., Refs. 256, 257, 258, 259, 260, 261, 262). This ambiguity arises because these systems are dispersion-supported, and there are well-known degeneracies between the orbit anisotropies, 3D stellar density, and dark matter density profile that are difficult to disentangle with the observational tools available 125, 262. As we describe in Section 5 proper motion measurements in the 2020’s with the WFIRST satellite may break this degeneracy.

One major surprise in this area is the recently discovered Milky Way companion Crater 2 263. It is bright like the classical satellites, but its surface brightness is so low that it was only recently discovered. It has an unusually low central density (although see Ref. 143), even for a dwarf galaxy, and it lies far from the center of the Milky Way, so tidal effects should be small. When better understood, this object may help clarify the cusp/core problem.

The story becomes more complicated at larger scales. For LSBs and larger spiral galaxies (\( M_{\text{halo}} \sim 10^{12} M_\odot \)), halos also appear less dense than expected from CDM-only simulations (in addition to the references above, see Refs. 264, 265). Interestingly, while the total mass profile of cluster-scale halos follows the NFW prediction well 266, 267, 268, the most massive galaxy clusters also appear to have underdense halos (\( M_{\text{halo}} \sim 10^{15} M_\odot \)). Typically, the center of a cluster is occupied by a massive galaxy: the brightest cluster galaxy (BCG). Because BCGs are compact objects relative to dark-matter halos, and dominate the gravitational potential where they sit, this implies that the density profile of the dark matter outside the BCG is shallower than NFW.

The one outlier to this trend of low-density halos is found in of elliptical galaxies in group-scale halos (\( M_{\text{halo}} \sim 10^{13} M_\odot - 10^{14} M_\odot \)). Strong gravitational lensing by these systems consistently shows that the total mass density profile is consistent with \( \rho \propto r^{-2} \) for a large part of the halo 269, 270. The dark matter profiles are typically found to be either consistent with or steeper than NFW 271, 265, 272.

To summarize, in the parameter \( M_{\text{halo}} \), the dark-matter solution to the cusp/core problem would have to show visible deviations from pure CDM on scales of \( M_{\text{halo}} \sim 10^9 M_\odot - 10^{15} M_\odot \).

Originally, the cusp/core model motivated investigations into WDM models, because of their much-lower-than CDM primordial phase-space densities 273, 5. This would imply that a dark matter solution is primordial in origin, and would thus occupy the moderate \( k_{\text{th}} \), low \( \Gamma_{\text{int}} \) region of parameter space in Figure 3. However, it has since been recognized that the cores produced in WDM cosmologies in the absence of baryons are far too small to be consistent with observations 274, 275. To obtain the relatively large core sizes needed to resolve the cusp/core problem, the number densities of those same halos would also be massively suppressed. This is the “Catch 22 problem” 276, 277, 278. Thus, most dark-matter interpretations of the core/cusp problem now focus on ultralight dark matter 96 (another primordial solution) or self-interacting dark matter (large evolutionary effect) 119, 279. Notice that astrophysics
is already informing the parameter space for viable particle physics models of dark matter.

“Missing Satellites” problem, and other substructure problems
While the cusp/core problem revolves around the smooth distribution of matter within individual halos, the missing satellites problem and other substructure problems arise from the non-smooth components. In a hierarchical Universe, large objects grow by the accretion of smaller objects. However, the accretion of these smaller objects is incomplete in many cases, as is demonstrated by the existence of satellite galaxies and unmixed stellar streams. Until the late 1990’s, simulations did not have enough resolution to reliably identify halo substructure in Milky Way-mass systems. When the simulations did achieve sufficient resolution[4] it was discovered that Milky Way-like systems have as much substructure as galaxy clusters do, if normalized to host properties [280, 281]. In the context of substructure, because of the ambiguity in the definition of subhalo mass (Section 3.2.1) abundances are typically presented in terms of cumulative number of subhalos of a given $v_{\text{max}}$ per unit host halo, $N(> v_{\text{max}})$ (Figure 6), often normalized by the host $v_{\text{host}}$, $N(> v_{\text{max}})/v_{\text{host}}$.

In Refs. [280, 281], this substructure velocity function was compared against the velocity function of known Milky Way satellites, which numbered a dozen at the time (see Ref. [283] for a compilation of known

[4] However, see Ref. [285] for a serious warning about numerical overmerging even today.
satellites and their properties). Converting the observed line-of-sight stellar velocity dispersion $\sigma_{\text{LOS}}$ to $v_{\text{max}}$ is non-trivial, but for many years a conversion of $v_{\text{max}}$ (or $v_{\text{vir}}$, the virial speed) equal to $\sqrt{3} \sigma_{\text{LOS}}$ was made on the assumption that the stellar velocity dispersion tensor was isotropic. As demonstrated in Figure 6, there was a significant mismatch in the velocity functions. If one interprets the mismatch in the velocity functions "vertically" (i.e., assuming our match between observational and simulated velocities is perfect, thus indicating that the discrepancy is in the vertical axis of the plot), it suggests that there are far fewer large satellites than expected in CDM. This was dubbed the "missing satellites problem." It inspired the theoretical development of WDM and SIDM models [6], and is a significant driver of the prominence these two classes of theory have in the particle and astronomical worlds.

Since this time, there have been a number of important developments and restatements of this problem. Importantly, dozens of new satellites of the Milky Way and Andromeda galaxies have been discovered in the past fifteen years, notably in the SDSS (e.g., Ref. [280]), PANDAS [287], DES (e.g., Ref. [288]), PAN-STARRS (e.g., Ref. [289]), MagLites [290], HSC (Hyper Suprime-Cam: [291]), and VST ATLAS surveys [292]. Most of these new satellites have properties distinct from those of the classical (pre-SDSS) satellites: they are tiny ($M_\ast \ll 10^5 M_\odot$), ancient (consistent with being reionization fossils [293, 294]), and extremely low in surface brightness. All of these properties allowed them to remain hidden before the advent of large, deep surveys. Taking into account survey selection functions, it is now thought that there could be hundreds (if not thousands) of satellites of the Milky Way (Figure 6) [295, 296, 297, 298, 299, 122, 123]. Simultaneously, the connection between galaxies and the halos they live in has been much better quantified (e.g., [146, 148, 151, 300]), as has the effect of reionization to prevent star formation in small halos [301, 155, 154, 302]. While the velocity function of these satellites remains unknown, it is the case that if the mismatch of velocity functions in Refs. [280, 281] is interpreted as missing satellites, then the new discoveries and estimated completeness corrections ameliorate this counting problem essentially entirely (Ref. [123]: Section 4.2).

This puts significant pressure on any dark-matter model with primordial effects. If the missing satellite problem’s initial incarnation in 2000 had a dark matter solution, the particle model would have lead to either primordial or evolutionary deviations on the scale of $10^7 M_\odot \lesssim M_{\text{halo}} \lesssim 10^{10} M_\odot$ (though with at least an order of magnitude uncertainty on both ends). As seen in Figure 1, a number of dark-matter models result in deviations in this regime. This is perhaps not surprising, because the missing satellite problem motivated many of these models (e.g., Ref. [6]). However, the reevaluation of this problem in light of the new dwarf galaxy surveys suggests that primordial deviations from CDM can appear on scales likely no greater than $M_{\text{halo}} \sim 10^8 M_\odot$ (evolutionary deviations up to cluster scales are still allowed for some models) [122, 123]. Refs. [122] and [123] show that WDM models equivalent to a thermal relic mass $m \approx 4 – 8$ keV are under severe pressure.

The remaining uncertainty in the solution to the missing satellites problem revolves around how and why, precisely, galaxies inhabit halos; and at what scale halos remain totally dark (i.e., devoid of luminous gas and stars). This problem can only be probed by gravity, that is to say, only by astrophysics, rather than particle physics. We discuss some possibilities in Section 4.2.

One way around the baryonic ambiguities is to find a way to count subhalos regardless of their baryonic content. This is the approach of strong gravitational lensing [303, 304]. (We will discuss other probes of dark subhalos, including stream gaps, in Section 5.) Strong gravitational lensing occurs when the light from a source is significantly distorted by a lens along the line of sight to the observer. Most dramatically, this leads to “Einstein ring” or “Einstein cross” systems, in which the source appears as a ring around the lens or where it is split into multiple images. It is sensitive to substructure in three different ways, probing three different derivatives of the gravitational potential and thus different properties of the substructure [305]:

1. Substructures can change the positions of images, either in the ring or multiple-image cases [305];
2. Substructure can perturb the relative brightnesses of images [306];
3. Substructure can induce time-delay anomalies [307], on account of the extra bending of the light travel path relative to the smooth lens case.

\[\text{http://www.astro.uvic.ca/~alan/Nearby_Dwarf_Database.html}\]
Studies of Einstein ring systems are sometimes referred to as “gravitational imaging”; tests of substructure using the fluxes of Einstein crosses use “flux-ratio anomalies” among the images in the crosses. Currently, the best constraints on the substructure mass function come from flux ratio anomalies and gravitational imaging studies [308, 309, 310, 311, 312]. Interestingly, these studies prefer an amount of structure consistent with or in excess of CDM predictions, albeit with large theoretical and observational uncertainties [308]. In Section 5, we discuss near-term prospects to reduce the uncertainties and home in on dark matter physics with strong lensing and new ideas for probes of tiny subhalos in the Milky Way halo using both the phase-mixed stellar halo and individual stellar streams.

The missing satellites and substructure lensing problems are both nearing twenty years old, but have evolved rapidly in just the past few years. We discuss prospects for continued progress in Section 5. The classic and excellent Refs. [313, 314] provide more in-depth discussion of the missing satellites problem, Ref. [123] discusses the likely resolution; Refs. [303, 305, 304] provide illuminating descriptions of strong and substructure lensing; and Ref. [315] for an introduction to stellar streams.

“Too Big to Fail” (TBTF)

An alternative way to view the missing satellites problem is to compare the central densities of satellites derived from observations to those from simulations. This can be more unambiguously measured and related in simulations and in observations than a relation between \( \sigma_{LOS} \) and \( v_{\text{max}} \) [318, 201, 319, 320, 249, 321]. A mismatch similar to the formulation of the missing satellites problem in terms of \( v_{\text{max}} \) persists, with CDM
predicting many more dense subhalos than the observed number of dense satellites. This formulation of the missing satellites problem was the direct progenitor of the “Too Big to Fail” (TBTF) problem. Specifically, TBTF was the observation that the largest subhalos in CDM dark-matter-only simulations had circular velocities larger than those measured in the large Milky Way classical dwarfs (see Figure 7). That is, the most massive subhalos in simulations, which should be “too big to fail” to produce stars and therefore should host visible satellites, did not seem to be present in observations.

The TBTF problem is a statement about the central densities of the classical dwarf satellites of the Milky Way ($M_*>10^5\,M_\odot$, i.e., the classical dwarfs known prior to SDSS) as measured and in simulations. Although there is a degeneracy between the velocity anisotropy of stellar orbits, the mass profile of dark matter, and the stellar density profile, the degeneracies can be approximately broken at one point in a dwarf galaxy: the half-light radius $r_{1/2}$

This occurs because (so far) we can only measure line-of-sight velocities of stars, and because there are so few stars, especially going toward and into the ultrafaint dwarf regime. Let us consider the first point. Line-of-sight velocities are mostly dominated by radial orbits at the center, and tangential orbits near the outskirts of the galaxy. At the half-light radius, we sample comparable proportions of stars on radial or tangential orbits. Because the strongest degeneracy is between mass and the relative distributions of tangential and radial orbits, which are totally degenerate for a spherically symmetric equilibrium system, we can only break the degeneracy if we have the other two components of the velocity vector (discussed in Section 5) or if we measure at the half-light radius. However, even with proper motion measurement of individual stars in dwarfs to get the other components of the velocity vector, there is still considerable uncertainty on the density profile because of the low number of stars to trace the gravitational potential

Fortunately, the degeneracy breaks at one point in the dwarf, and so one may estimate the mass enclosed within approximately the half-light radius $r_{1/2}$ for multiple distinct stellar populations, we can find the mass within the half-light radius for each population, and infer a slope for the mass profile as well.

The mass within the half-light radius, $M_{1/2} = M(<r_{1/2})$, can also be reinterpreted in terms of the circular velocity: the velocity of a particle on a circular orbit at the half-light radius ($v_{1/2}$: Eq. 20). Ref. 322 plotted $v_{1/2}$ for the classical satellites of the Milky Way, as well as the $v_c$ (a proxy for the enclosed mass) curves of the most massive subhalos of Milky Way analogs in several CDM-only simulations. In principle, one may match an observed satellite to a simulated subhalo if $v_{1/2}$ lies on top of one of the $v_c$ lines from a simulations.

Based on our expectation from galaxy evolution theory, the biggest satellites (in terms of number of stars) should live in the biggest subhalos. Successful models of galaxy formation — at least for large galaxies where they can be empirically tested — suggest that the stellar mass of a galaxy is highly correlated with halo mass, the SMHM relation (see Section 3.2.1). There is some evidence that the scatter in the SMHM relation increases with decreasing halo mass, through a combination of feedback and reionization regulation of gas flow and star formation $^{140,141}$, but the overall correlation is observed in simulations down to ultrafaint galaxy scales $^{326}$. We show the range of (sub)halo masses expected to host the classical Milky Way satellites according to the SMHM relation of Moster et al. $^{151}$ in Figure 7. (Although this is empirically tested and constrained for galaxies more massive than the Magellanic Clouds, anything below a stellar mass of $M_*>5\times10^7\,M_\odot$ requires extrapolation and assumptions about the surface brightness distribution function of galaxies around that mass scale $^{300,151,327}$.)

If we have a strong prior that CDM is the correct phenomenological model for dark matter, that dark-matter density profiles are unaffected by baryonic physics, and that dwarf galaxies follow the extrapolated SMHM relation, we expect the $v_{1/2}$ points of the classical galaxies to lie on the NFW $v_c(r)$ curves of the most massive simulated Milky Way subhalos. However, on the small scales of Milky Way satellites, Ref. 322 found that the largest subhalos in CDM dark-matter-only simulations had circular velocity curves far above the $v_{1/2}$ points for known satellites. In fact, it appeared that only lower-mass simulated subhalos had circular velocity curves that were more consistent with the observed $v_{1/2}$ points, but even then it was a challenge to match some surprisingly low-density satellites (e.g., Fornax, which is the brightest Milky Way satellite other than the Magellanic Clouds and the disrupting Sagittarius dwarf). We illustrate this point in Figure 7 showing $v_{1/2}$ for the classical satellites and the $v_c$ profiles of SMHM-matched subhalos with typical density.
profiles in CDM.

This observation can be interpreted in several ways. The TBTF interpretation is that the big subhalos \((\gtrsim 10^{10} M_\odot)\) are devoid of baryons. Thus, these massive objects are not seen in astronomical surveys, leaving only the smaller subhalos \((\lesssim 10^{10} M_\odot)\) with stars in the sample of galaxies to be matched with simulation. However, this idea is at odds with our intuition regarding galaxy evolution. While we do expect star formation to become stochastic (that is, difficult to predict from halo mass alone) on small scales \((M_{\text{vir}} \lesssim 10^8 M_\odot - 10^9 M_\odot)\) \cite{228}, we expect that scale to be much smaller than the scale on which TBTF would require that halos be dark \cite{122,123,152,228}. Again, this is the source of the moniker “Too Big to Fail,” as such massive subhalos are expected to be “too big” to fail to produce stars.

There are three commonly invoked alternative explanations to explain the mismatch between observed \(v_{1/2}\) and the CDM-only circular velocity curves. First, the Milky Way may have a smaller dark matter halo than assumed by Ref. \cite{322}. The mass of the most-major subhalos depends on the parent halo’s mass, and so reducing the mass of the Milky Way would reduce the expected mass of the largest subhalos orbiting it. Several have suggested that a Milky Way halo mass of \(\sim 8 \times 10^{11} M_\odot\) has no TBTF problem if CDM-only simulations are to be believed \cite{228,330,331,332}. While such a low halo mass is consistent with some Milky Way mass estimates \cite{333}, it is inconsistent with others \cite{334}. A low mass also makes the presence of the Magellanic Clouds much more problematic \cite{335,336}, and is inconsistent with expectations from the SMHM relation \cite{300,151}.

Second, the Milky Way could simply be an outlier among Milky-Way-mass systems \cite{330,332}. There is significant halo-to-halo scatter in satellite and subhalo populations \cite{337,338}, so perhaps the Milky Way lies some ways off the mean satellite population.

The third explanation, one closely tied to the cusp/core problem, is that the mass profile of CDM-only simulated halos may just be a poor match to reality. Connecting the enclosed mass measured at the half-light radius to a total halo mass requires an assumption about the density profile of the dark matter, since the virial radius is one to two orders of magnitude larger than the half-light radius (in Figure 7 we plot \(v_c\) curves out to the virial radius) \cite{43,149,44,150}. If the density profiles are modified by physics which is not correctly modeled by the CDM-only simulations, the matching between \(v_{1/2}\) and \(M_{\text{vir}}\) changes significantly \cite{221,224,261,285}. Such a modification can be achieved either with new dark-matter physics or via baryonic supernova feedback (see below), but due to the computational cost of high-resolution simulations of galaxies with either baryonic physics or non-trivial dark matter physics, the exact goodness of fit of any solution remains to be determined. As shown in Figure 7, if either dark-matter physics or baryons can produce cores in dark-matter halos, the \(v_c\) curves of the halos expected to host the Milky Way classical dwarfs (according to the SMHM relation) pass through the classical dwarf \(v_{1/2}\) points.

Interpreted in the context of dark matter in Figure 1, the TBTF problems suggests we should look for new dark-matter physics in halos of \(\sim 10^{10} M_\odot\). New physics could show up on smaller scales than this, but a solution to TBTF specifically must address this scale. The solution to this problem alone can either be primordial (suppression of moderate-sized halos) or evolutionary (modification of density profiles).

**The Baryonic Tully-Fisher (BTF) relation and field dwarf counts**

One of the longest standing distance measures for galaxies is the Tully-Fisher (TF) relation \cite{339}. The distance measure exploits the tight power-law correlation between the intrinsic luminosity of spiral galaxies and their spectroscopically determined rotation speed — that is, a tight correlation between the luminosity of a galaxy and the depth of its potential well. The TF relation, as derived from stellar luminosity, has much higher scatter for small galaxies (roughly at the scale of the Large Magellanic Cloud, \(M_{\text{halo}} \approx 10^{11} M_\odot\)), because such galaxies are so gas-dominated. Big galaxies have comparatively little cold gas, but low-mass isolated field dwarf galaxies are, to first order, clouds of gas lightly sprinkled with stars \cite{340,341,342}. When cool gas and stars are summed, the tight power-law correlation between the rotation velocity of galaxies and baryonic mass reappears \cite{343,344} called the baryonic Tully-Fisher (BTF) relation \cite{345}.

However, the TF relation (shown in Figure 5) mapping from rotation curves to \(M_{\text{halo}}\) is in tension with other measurements of the relation between the halo mass and baryonic mass of galaxies. If the furthest measured point of the rotation curve is a good tracer of the halo potential (allowing a mapping between velocity and halo mass), it appears that dwarf galaxies live in systematically smaller halos than expected.
from the extrapolated SMHM relation [341, 347, 348, 349]. This again suggests that if CDM is the correct description of dark matter, there must be a “too big to fail” problem for small galaxies in the field (that is, far from a large Milky Way-type galaxy) not just for satellite galaxies of the Milky Way. As there are many more small halos than big halos in CDM, if the TF relation is to hold, there must be many halos which are dark but relatively massive.

This counting problem shows up in field in other ways. Using the Alfafox radio data (measuring 21 cm radiation from atomic hydrogen, and used to quantify the BTF relation [348]), Ref. [282] turned their galaxy counts into a velocity function for these isolated galaxies, just as done for the missing satellites work. They found a flattening of the velocity function below 80 km/s relative to the CDM $v_{\text{max}}$ function (Figure 6, bottom panel). A similar discrepancy was found on a similar scale by Ref. [350], which focused on optically selected galaxies within 10 Mpc from the Milky Way rather than the radio-selected sample of Ref. [282]. These measurements again suggest a “Too Big to Fail” problem in the field.

One of the major problems with this interpretation is the difficulty in relating the observed velocity data to halo masses. There is a close relation between the characteristic speed of the rotation curve of large galaxies and the virial velocity (Eq. (24)) of their hosts [346]. However, we know this because we have a number of different tools available to measure halo masses for large galaxies, notably galaxy-galaxy weak gravitational lensing [351]. For small halos, fewer tools are available, and the mapping between observed rotation speed and halo mass is more ambiguous. Often, the rotation curves are still rising when the gaseous or stellar tracers disappear, well within the radius at which $v_{\text{max}}$ occurs [349]. Typically, even if the rotation curves are not well-matched to NFW density profiles, the last point on the rotation curve is used to assign galaxies to a halo mass under the assumption that the NFW profile is a reasonable description outside that radius [347, 282], which may not be the case. If the halos are cored rather than cusped, we break the neat relation between the measured rotation curve and the virial mass, and in fact a low-amplitude rotation curve can be mapped to a halo mass that is substantially higher than inferred if one assumes that the halos
Figure 9: Summary of the halo mass $M_{\text{halo}}$ where hints of deviations from $\Lambda$CDM have been claimed, as well as the range of $M_{\text{halo}}$ where baryons are expected to influence the structure of halos. “TBTF” is “Too big to fail” and “BTF” is the baryonic Tully-Fisher relation.

have NFW density profiles (Figure 7). Thus, the mismatch between the predicted and observed velocity functions, or the TF-derived SMHM relation and the abundance-matching prediction, may be hints of a cusp/core problem instead. As we will see below, the interpretation of the problem, additional systematic uncertainties, and the necessity of either new baryonic or dark-matter-physics solutions are hotly debated [352, 353, 354, 355, 356, 357].

Interpreted in the context of our astrophysical parameter $M_{\text{halo}}$, solutions to the BTF problem in the form of new dark matter physics are needed for the $M_{\text{halo}} \sim 10^9 M_\odot - 10^{11} M_\odot$ mass range.

Are all these problems aspects of one central problem?

We described the four most commonly cited problems with CDM on small scales, summarized in Figure 9. However, we argue here that these problems are different views of one central problem, which is that the relationship between the mass of galaxies and that of their dark matter halos is uncertain on small scales.

Why is this the central problem, and how will a reduced uncertainty help illuminate dark matter microphysics? Let us consider first the missing satellites problem, the TBTF problem, and the BTF relation. The existence of these problems as they are stated depends on an accurate mapping from the observed kinematics of baryons to a halo mass. If we assume that the currently proposed mappings are correct, then the horizontal axes in Figure 6 and the vertical axis in Figure 8 are correct and the discrepancy between the velocity functions in the Milky Way or the field can be interpreted by their difference in the vertical (Figure 6) or horizontal (Figure 8). That is, there is a mismatch in simulated CDM halo and observed galaxy abundances; the problem is a counting problem.

However, if the mapping between the internal kinematics of galaxies and their host halo is incorrect, either because of systematics or because the dark matter mass profile of halos is fundamentally different from CDM-only simulations, then the problem is instead in the horizontal axis of Figure 6 and vertical axis of Figure 8. It turns the problem into a cusp/core problem. If interpreted as such, the important implication is that by assuming CDM-only density profiles, we underestimated the halo mass for fixed baryonic mass or for fixed baryonic kinematics (Figure 7). Cored halos generically have higher ratio of virial velocity (or virial mass) to the velocity of the baryonic tracer in the inner part of the halo than do cusped halos. Thus, if we understood exactly what types of halos are inhabited by different types of galaxies, there would be no ambiguity in interpreting galaxy abundances in the context of dark matter theories—we would know if we had a serious counting problem on our hands, or mostly a cusp/core problem.

Likewise, if we knew how to match galaxies with halos, we could assess the origin of the cusp/core problem. Currently, baryonic solutions to this problem in the context of baryonic feedback (discussed below) depends on how much energy is available in the baryonic sector to push the dark matter around in halos [358, 352]. The amount of energy required to push dark matter out of halo centers depends on the depth of the potential well, which depends on the halo mass. If we knew how to match galaxies to halos, we
could determine if the pattern of central dark matter densities as a function of galaxy mass were consistent with a CDM + baryon physics interpretation, or not.

Thus, pushing direct measurements of halo mass down to $M_{\text{vir}} \lesssim 10^{11} M_\odot$ with, for example, galaxy-galaxy lensing [359] (see Section 5) should be a high priority for the field. Because several of the small-scale “problems” are Milky Way-specific, we also need better constraints on the Milky Way halo mass, which may be obtained in the future with distant satellite orbits and with stellar streams [334, 360, 361, 315, 362].

4.2. The Known Unknowns of Theoretical Predictions for Astrophysical Dark Matter: It’s All About the Baryons

These four discrepancies between the dark matter predictions and observations have been immense interest to both astrophysicists and particles physicists, driving the creation of a large number of dark matter models which modify dark matter halos on the scale of $M_{\text{vir}} \sim 10^8 M_\odot - 10^{15} M_\odot$. However, the biggest known problem for dark matter structure predictions is that most predictions are made without baryons. Although baryons are relatively minor contributors in the energy budget of the Universe, they are dynamically important precisely in the parts of the Universe where dark-matter-induced deviations from CDM might appear. Thus, a true mapping between dark matter models and $M_{\text{halo}}$ requires an accurate treatment of baryons. Here, we consider the impact of baryons on the “hints” described above, and highlight current uncertainties. We focus largely on hydrodynamic simulations with CDM, as hydrodynamic simulations with other types of dark matter are still in their infancy [363, 364, 365]. We turn to non-CDM hydrodynamic simulations at the end of the section.

An alternative way to cast this discussion is in the context of Figure 1, where baryons are placed at $M_{\text{halo}} \sim 10^8 M_\odot - 10^{15} M_\odot$ — that is, we expect the existence of baryons, and in particular the non-trivial interactions that baryonic physics allows (scattering, cooling, stellar formation, supernovae, etc.) to modify the structure of dark matter halos in this mass range. Baryons are a form of matter, with a unique microphysical description. If we want to understand the microphysics of the particles making up astrophysical systems (which include baryons) from the observations of the structure of these macroscopic systems, we need a dictionary to translate between the particle and galactic scales. Simulations serve as this dictionary, and so we we discuss the current state of the art.

Cusp/Core Problem

When hydrodynamic simulations were first performed, it appeared that baryons exacerbated the cusp/core problem. Baryons cooled efficiently in halos, and dragged dark matter in with them as they deepened the gravitational potential wells of their systems. Thus, halos appeared denser in hydrodynamic simulations than in dark-matter-only simulations [366, 367]. There were additional problems with the simulations at the time, namely that they were too efficient at forming stars (this is still somewhat of a problem on Milky Way scales), and that the disks of galaxies were too dense (the “angular momentum catastrophe”) [368]. Ref. [369] showed that a key reason for these problems was insufficient spatial resolution. Star formation in simulations is implemented using semi-analytic prescriptions rather than ab initio with all relevant physics self-consistently implemented (star formation remains a very active area of research [370]). Star formation is inherently clumpy, taking place in individual molecular clouds ($\sim$ pc scale). Due to its spatial concentration, the radiation and kinetic energy flux from the death throes of stellar evolution have a (relatively) easy time “punching” through the interstellar medium and out of the galaxy, removing significant quantities of gas in the process. However, the simulations of the early 2000’s were averaging star formation over very large scales relative to the scales on which stars actually form. As a result, in these simulations, energy and gas had a hard time getting out of the interstellar medium, and stars formed very efficiently. When the simulations of Ref. [369] and others began resolving star formation at more realistic scales, star formation became clumpy, energy and gas escaped from the interstellar medium, and dark matter halos responded by becoming less dense than the CDM-only simulations predicted [371].

Despite differences in code and the implementation of semi-analytic star formation prescriptions, there are now largely consistent results about the sign of the effect of baryons on halos on dwarf scales [255, 284, 372, 373, 374, 375, 376, 377, 378, 379, 352, 261]. However, there are disagreements in the details. In
some simulations, halos show true cores, while in others they appear to have low-density cusps (Figure 5). Moreover, below some mass scale, there are never enough baryons at the centers of halos to be dynamically important. On cluster scales, simulations show anything from steeper-than-NFW cusps to mild cores [380, 377, 379]. Determining why there are these differences among simulations, and which (if any) is closest to reality, is the subject of vigorous ongoing activity.

Thus, there are indications that including baryons in simulations goes at least part of the way to resolve the cusp/core problem. It is as yet unclear if baryons are solely responsible. The halo mass scale on which baryons can affect the central density of dark matter is around $M_{\text{halo}} \sim 10^9 \, M_\odot - 10^{15} \, M_\odot$, where the bottom of the range is uncertain by about an order of magnitude because of the uncertainty on the scale on which baryons significantly affect halo structure [358, 381, 352].

“Missing Satellites” problem, and other substructure problems

Baryonic solutions to the missing satellites problem involve three separate physical effects:

1. subhalo mass functions and survival,
2. the mapping between the line-of-sight velocities of individual stars in galaxies and the $v_{\text{max}}$ of dark-matter halos, and
3. the probability that some subhalos have no visible baryons in them at all.

Together, these solutions give strong evidence that the missing satellites problem is resolved in the context of CDM, but also that new methods are required to test the halo mass function below $M_{\text{vir}} \sim 10^8 \, M_\odot$. Galaxy evolution theorists consider the missing satellites problem solved.

Why? The missing satellites problem was initially formulated in the context of dark-matter-only simulations. However, baryons can significantly reduce the abundance of dark matter halos at fixed $v_{\text{max}}$ in several ways. First, baryonic outflows from halos can prevent halos from growing by accretion as fast as they would in the absence of baryons. This is a $\sim 20\%$ effect on the mass function (although it varies between simulation codes) [382, 383]. Second, once halos fall into larger halos, the small halos are more easily destroyed in hydrodynamic simulations as compared to their dark-matter-only analogues. This is largely driven by the extra tidal field of the central galaxy in the host [384, 284, 372], but also because the lower central density of subhalos in hydrodynamic simulations makes satellite galaxies more vulnerable to tidal heating or disruption [385, 372, 225, 143, 386]. This leads to a $O(10\%)$ reduction in the subhalo mass function. However, neither of these effects is the dominant driver of the solution to the missing satellites problem.

Next, we consider the mapping of the observed properties of galaxies to those found in simulations. In Section 4.1, we saw that the missing satellites problem was conceptualized as a velocity problem, and that we needed to relate line-of-sight stellar velocity dispersion $\sigma_{\text{LOS}}$ to the halo $v_{\text{max}}$ to compare an observed property of a galaxy to a halo. It is now possible to bypass the $v_{\text{max}} \rightarrow \sigma_{\text{LOS}}$ conversion in hydrodynamic simulations, and instead compare the simulated stellar $\sigma_{\text{LOS}}$ distribution functions of simulated and observed satellites [357, 358, 201]. If the missing satellites problem is cast in terms of a luminosity or stellar mass function, there is a similar level of agreement on classical dwarf scales [373] [387]. However, these simulations can only resolve classical satellites, not yet the ultrafaint galaxies that are expected to dominate the Milky Way satellite luminosity function below a stellar mass $M_* \lesssim 10^5 \, M_\odot$ [290, 295, 299, 122, 123]. Thus, the velocity functions are now a good match to classical satellites, but there are still many more subhalos in simulations than ultrafaint dwarfs orbiting the Milky Way.

The dominant solution to the missing satellites problem comes when we consider which halos are actually occupied by visible baryons, which is especially important for ultrafaint dwarfs. It is an interesting and instructive historical note that the existence of ultrafaint dwarfs was predicted before their discovery in SDSS [155] [301] [153]. The origin of this prediction resulted from the following arguments. Galaxies form from gas enriched by the first generation of stars, and cannot form in arbitrarily small halos [389, 390]. Even if molecular hydrogen and metals are only minute constituents of the baryon budget before reionization ($z \geq 6$), it can allow gas to collapse and cool in halos of $M_{\text{vir}} (z) \sim 10^6 \, M_\odot - 10^7 \, M_\odot$ ($v_{\text{max}} \sim \text{a few km/s}$) [47, 391, 328]. There are thousands of subhalos larger than those in the Milky Way today.
However, and very importantly, the reionization of the Universe sets an important limit on how many of such small galaxies may form. When the Universe reionizes, the intergalactic medium is heated and pressurized; it can no longer collapse into small halos, and what gas exists in small halos at reionization may be boiled out [392]. This has two consequences. First, few halos reach the halo mass threshold for cooling before reionization, so only a minority of halos host ancient stellar populations. Second, galaxies forming after reionization can only form in much larger halos \( v_{\text{max}} \gtrsim 20 \text{ km/s} - 50 \text{ km/s} \) than before [391]. Depending on the exact prescription for star formation and the redshift of reionization, the Milky Way is estimated to have a few hundred to a few thousand reionization fossils in its halo in a CDM cosmology [393, 394, 395, 129, 123]. In fact, the number of completeness-corrected dwarf satellites projected to inhabit the Milky Way halo based on present-day observations matches almost exactly the number predicted by the SMHM relation and simulation-tested reionization models [123], shown in Figure 6.

Where does this leave us? The current state of theory and simulations suggest that baryon physics solves the missing satellites problem. It should be noted, though, that there do not exist fully self-consistent simulations of the Milky Way satellite population down to ultrafaint scales (see Ref. [396] for a novel variant of standard “zoom-in” simulations).

The new frontier for tests of primordial deviations from CDM is on the \( M_{\text{vir}} \sim 10^8 \text{ M}_\odot \) scale and below, where halos are expected to be sans baryons in galaxy evolution theory. Methods that do not rely on luminous tracers, like substructure lensing or stellar stream perturbations, are the future. However, the highest resolution hydrodynamic CDM simulations focus almost exclusively on Milky Way analogs or isolated dwarfs. The host halos of strong lenses tend to be significantly more massive than the Milky Way, and require resolution down to \( M_{\text{vir}} \lesssim 10^6 \text{ M}_\odot \) scales. We discuss this point in greater detail in Section 5. Note that evolutionary deviations are possible on scales above \( 10^8 \text{ M}_\odot \).

Too Big to Fail

As with the missing satellites problem and the core/cusp problem, hydrodynamic CDM simulations indicate that baryons play an important role in shaping the central densities of Milky Way satellites to solve TBTF [388, 261]. This is not surprising, given the core/cusp interpretation of TBTF; if baryons can push dark matter out of the cores of halos, then the central densities of halos can be low even if the total halo mass is high. However, the surprisingly low densities of Fornax (a big satellite), Canis Venatici I, and Sextans remain difficult to reconcile with simulations of CDM + baryons [397, 398, 387, 388]. The conclusion is that baryons can address TBTF on halo mass scales \( M_{\text{halo}} \sim 10^{10} \text{ M}_\odot \), but are still not bringing about perfect agreement. Fornax, in particular, remains an intriguing problem [399].

Baryonic Tully-Fisher Relation

This arguably is the hardest problem to solve in the context of baryons. Let us consider problems with BTF separately from the problem found in the number of dwarf galaxies in the field. Despite the relatively good agreement between observations and the APOSTLES simulations for the TBTF problem [261], the field galaxies in this simulation suite consistently show high rotation curves for fixed stellar mass relative to observations, below \( M_* \sim 10^8 \text{ M}_\odot \) \( (v_{\text{max}} \lesssim 40 \text{ km/s}) \) [349]. This discrepancy persists even when the authors “measure” (i.e., use the rotation speed derived from the simulated mass distribution) the simulated rotation curves in the same regions of the galaxies within which rotation curves are measured for real galaxies. The problem can be stated one of two ways. If one believes the simulated stellar masses are accurate, the problem is that the simulated dark-matter halos are too dense. On the other hand, if we believe the velocity numbers instead, it suggests that star formation remains too efficient for fixed halo mass. We return to this point after we discuss the counting problem.

The counting problem begins for even higher halo masses, \( v_{\text{max}} \lesssim 80 \text{ km/s} \), which in the context of the Milky Way corresponds to an analog of the Magellanic Clouds (Figure 3). It is curious that the missing satellites problem can be solved on those scales, while the problem in the field begins on higher mass scales. While other attempts to reconcile the counting problem in the field below this scale in the context of CDM have not worked well [400, 353], a new study by Ref. [357] indicates that the problem may be the way in which simulators compare their results to observations. Many groups create rotation curves using the distribution of simulation gas, star, and dark matter particles or fields. Instead, Ref. [357] “observed”
the atomic hydrogen in their simulated galaxies, analyzing their simulated observations the same way as real observations. The agreement between theory and observation was greatly improved, especially for the lowest-atomic-gas-mass objects. The origin of the agreement is that much of the highest-velocity material in small galaxies is in gas whose column density is too low to produce flux above the background level of the experiment, and thus the rotation curves for these dwarfs is systematically observationally underestimated. For the smallest gas-containing galaxies, all of the material is too low in density to be picked up with the current 21-cm surveys.

The idea that there are systematics remaining in comparing the velocity profiles of real and simulated galaxies is appealing, in no small part because other simulation methods of matching galaxies to halos work remarkably well. In particular, the relationship between stellar and dark matter mass (the SMHM relation; Section 3.2) found in hydrodynamic + CDM simulations [326, 142, 401] shows good agreement with extrapolations from empirical measurements for larger masses [151] and inferences based on Local Group satellite populations [144, 123]. It is attractive therefore to assume that the problem lies not with the efficiency of star formation in simulation, but rather in the interpretation of rotation curves. These lines of evidence point to the problem with BTF lies in the assignment of galaxies to halos based on rotation curves, both on the simulation and observational sides.

**Baryons and non-CDM models**

Our figure of merit $M_{\text{halo}}$ encodes the characteristic size of dark matter haloes where non-gravitational physics enters and causes a departure of the power spectrum or internal properties of the haloes from the predictions of CDM. The interpretation of the hints in Section 4.1 depends at minimum on accurate predictions for CDM with baryons, to set the benchmark for CDM cosmology. Tests of specific dark-matter models other than CDM must also incorporate the physics of baryons in order to define $M_{\text{halo}}$, the halo mass scale(s) at which deviations from CDM arise.

At present, there are many more non-CDM, dark-matter-only simulations than simulations with baryons (e.g., Refs. 224, 161, 213, 194, 183, 210, 402, 403, 404, 405, 406, 212, 407). Most of the deviations from CDM, quantified by $M_{\text{halo}}$, shown in Figure 1 and Section 3 were made in the context of dark-matter-only simulations.

As with CDM, the field is transitioning from making accurate predictions for the wrong problem (increasingly high-resolution studies of structure formation without baryons) to increasingly precise and more robust predictions for the right problem (simulations of dark matter and baryons). In the context of dark matter physics, the two classes of astronomical systems simulated most often are isolated dwarf galaxies [365, 374, 408, 409, 410, 412, 413, 414, 365, 415] and Milky Way analogs [373, 358, 359, 414, 365, 415] (although some are simulated with simplified models of the baryonic component [416, 365]). The former is a popular choice for simulations because one may achieve much higher spatial resolution for fixed computation time than for more massive systems. Hydrodynamic simulations are much more computationally costly to run than dark-matter-only simulations (∼ an order of magnitude more, even without potentially important physics like radiative transfer), so this is a practical strategy. The latter is popular because of the Milky Way-centric missing satellites and TBTF problems.

How are non-CDM hydrodynamic simulation results different than those without galaxy evolution physics? The most important finding so far is that baryon physics brings CDM and non-CDM predictions closer together, at least when it comes to halo density profiles and bright galaxy counts. As pointed out in Ref. 417, dark matter with self interactions behaves as a fluid, and is in hydrostatic equilibrium unless the system is undergoing a major merger. Thus, in baryon-dominated central potentials, the self-interacting dark matter nearly forms a cusp unlike the low-density cores found in dark-matter-only simulations [363, 416, 359, 415, 407]. For non-baryon-dominated systems, there remain differences in the density profiles between CDM and SIDM halos, but they are not as extreme as the dark-matter-only simulations suggest [412]. Galaxies in cosmologies with a truncation in the matter power spectrum also show convergent evolution if residing in halo masses above the cutoff [411, 365, 412], even if there are some differences in the star-formation histories at early times [199, 411, 413]. However, any cutoff in the power spectrum does lead to distinct differences in galaxy and halo counts regardless of baryon physics [123].
In summary, we described a number of small-scale hints that CDM is an inadequate description of dark matter on scales of $10^9 M_\odot \lesssim M_{\text{vir}} \lesssim 10^{15} M_\odot$. These hints are shown in Figure 9. They highlights a region in Figure 1 for which dark matter candidates with novel physics may warrant significantly more study. However, we also know of another form of matter, baryons, that also affect halos on these scales (again, see Figure 9). Baryons are important for $10^8 M_\odot \lesssim M_{\text{vir}} \lesssim 10^{15} M_\odot$ and above. It is not clear if baryons can solve all small-scale “problems” with CDM, but the physics goes in the direction of better, rather than worse, agreement. Baryons also affect non-CDM predictions significantly, especially with respect to halo density profiles. Critically, we argued that a central observational problem preventing further progress in determining the dark or baryonic origin of these hints is the fact that the mapping between galaxies and the halos they inhabit is still highly uncertain on the dwarf galaxy scale.

Ultimately, Figure 1 must be made in the context of dark matter + baryon predictions. Regardless of what the true nature of dark matter is, its cosmic distribution is governed by the gravitational effects of baryons in addition to dark physics, and so the baryons must be well understood before non-gravitational interactions in the dark sector can be identified or ruled out with any confidence. We now turn to discussing what the future prospects are for this effort.

5. Future Directions

In this section, we lay out a plan for dark matter constraints from astronomy. As the previous sections have demonstrated, the space of dark matter theory is enormous, and encompasses particles far removed from the canonical CDM or WIMP dark matter picture. Critically, an entire axis characterizing the effects of dark sector physics on astrophysical structures ($M_{\text{halo}}$ in Figure 1), is orthogonal to the particle interactions which have been the primary interest of many dark matter physicists (characterized by $\Lambda^{-1}$ in Figure 1). The study of dark matter structures has to-date been the source of all positive statements about the properties of dark matter, and has informed many of the constraints on the possible nature of this mysterious material. Additionally, we showed in Section 4 that there are tantalizing hints for non-minimal dark matter on a wide variety of halo mass scales, but we also highlighted some of the present-day systematic uncertainties in turning these hints into measurements of dark matter microphysics.

Here, we advocate a specific plan to turn the promise of measurements of dark matter from astronomy into reality. We focus on two specific paths. First, we outline a theory program to make the mapping between the specific types of dark matter models described in Section 3 and $M_{\text{halo}}$ (and, more generally, what observables are relevant on those scales) much more precise. Importantly, we will show why this program is relevant for particle physicists even if dark matter turns out to be a vanilla WIMP. Second, we describe some ideas of how to leverage the next decade’s impressive increase in the quantity and quality of large astronomical survey data for dark matter science, to tightly constrain deviations from CDM as a function of $M_{\text{vir}}$. We highlight the parts of the $M_{\text{halo}}$ space which are especially in need of new ideas for observations. Our main message is that the capabilities to make dramatic improvements in astrophysical dark matter constraints are here now, both in theory and observation.

5.1. Theory

The ability of astronomical observations to measure dark-matter physics is predicated on an accurate mapping between particle theory space and astronomical observable, which is tied to a specific $M_{\text{vir}}$ scale or scales. Building on Section 4.2, we advocate for a specific theory program, highlighting how this approach leads to improvements in traditional astroparticle dark-matter searches as well as for measurements of dark-matter microphysics with astronomy. The key is accurate theoretical predictions for structure formation for different classes of dark matter models that include the effects of galaxy evolution physics on the dark matter distribution, and that these predictions are tuned to observationally relevant scales.

What are the necessary requirements to this program? Let us start with the big picture first: theoretical predictions for $M_{\text{halo}}$ must be made for the types of astronomical systems most amenable to study, and the theory program must be flexible enough to respond to new (and unexpected) signatures of dark matter physics and influence observational programs.
The former is easier to plan for than the latter, but still requires a potentially new approach. Here we give an example of the issues the community faces to make these predictions, and show what kinds of new approaches may be required. As we described in Section 4.3 most high-resolution hydrodynamic simulations (especially those focused on cusp/core problem or missing satellites problems) focus on isolated dwarfs or Milky Way analogs, out of a combination of technical feasibility issues and taste in problems. However, as we describe below in Section 5.2 substructure lensing of group-scale dark matter halos is a promising observational method to measure the halo mass function down to low masses ($\sim 10^6 M_\odot - 10^7 M_\odot$, depending on a precise definition of halo mass). Accurate predictions for these systems as a function of dark matter physics, after including galaxy evolution systematics, are essential to realizing the potential of these systems for dark matter physics.

The challenges are:

- The large dynamic range of the problem. The host halo masses are $\sim 10^{13} M_\odot - 10^{14} M_\odot$, or 10 to 100 times more massive than the Milky Way, yet we need to accurately resolve substructure that is $10^7 - 10^8$ times less massive in size. This does not even include line-of-sight effects on up to Gpc scales, which may be as or more important than structure within individual halos. \cite{118,119,120}.

- Each system has unique properties and uncertainties in those properties — halo mass and density profile, central galaxy matter distribution, large-scale environment that may add to the lensing signal, and specific realization of substructure — and not all lensing configurations are equally useful for dark matter constraints. \cite{303}.

- We have ever-increasing ensemble sizes of these lens systems, and that the necessary physics (dark and baryonic) to understand these systems to the level required for dark matter constraints is either missing or not well quantified.

Even in the absence of baryons, predictions for these systems are challenging in detail, in part because of line-of-sight effects and the difficulty in following small subhalos as they evolve in their hosts. \cite{118,121,285,120}. Uncertainties in galaxy evolution physics, detailed in Section 4.2 add significant computing time and uncertainties even for a fixed dark matter model. \cite{383,122}.

How do we make predictions for observations of ensembles of objects as a function of dark matter physics, given the enormous computational cost to model even a single system for fixed dark matter physics? The solution is almost certainly a hybrid program of simulations of various types, and semi-analytic and analytic modeling \cite{123,124,125,126}. Simulations are powerful, but are so costly that they are best suited to case studies. They are also incredibly important to isolate specific physical effects, which may then be modeled analytically. Different types of simulations — large volume but low-resolution cosmological simulations, cosmological zoomed simulations, and high-resolution idealized simulations — are necessary for different parts of the problem. But it will be a challenge to model substructure lensing observations with simulations alone. Semi-analytic models (SAMs) layer parameterized models for galaxy formation physics on halo merger trees. They are fast and flexible tools, and particularly well suited to creating realizations of large ensembles of observational systems. It is straightforward to quantify and parametrize uncertain physics, and to connect the model to likelihood functions to constrain dark-matter physics. However, they must be calibrated with simulations, which may be idealized in order to isolate specific physical processes (e.g., dynamical friction). Efforts to incorporate new dark-matter physics into semi-analytic models are underway \cite{167,127,168,128,129}.

In addition to tailoring an astrophysics theory program to model known classes of interesting systems, it is also important to pay attention to novel signatures of dark matter physics. Due to the small-scale structure problems (described in Section 4.1), a great deal of theoretical thought has been devoted to dark matter physics scenarios that modify CDM predictions on the scale of $10^9 M_\odot - 10^{10} M_\odot$. However, this is not the unique scale that can be affected by new physics in the dark sector (merely the scale at which hints currently appear), and other ranges of $M_{\text{vir}}$ have not received as much theoretical consideration. While theory should always take cues from experiment and observation, it can also proactively suggest new targets and benchmarks against which observation can be tested. As we show below, new ideas are desperately
needed for large swaths of $M_{\text{vir}}$, possibly leading to better use of astronomical objects already identified as interesting. For example: the implications for two-body gravitational relaxation in ultrafaint dwarf galaxies.

We now turn to more technical issues, namely the mapping between dark-matter microphysics and the macroscopic implementation of that physics in simulations. Importantly, the algorithms to model dark matter properties and their accuracy used to simulate dark matter and galaxies together must be well-matched to the observed system under consideration. Leaving aside improvements in the hydrodynamic modeling of gas and stellar evolution in galaxies, for which a vigorous program is underway in the galaxy simulation community (e.g., Refs. [430, 431, 432, 433, 434, 435, 436]), there are other technical challenges to achieve an accurate calibration for the $M_{\text{halo}}$ figure of merit. Fundamentally, the problem is again a matter of scales; in simulations, we represent enormous patches of phase space with single simulation particles. In the case of self-interacting dark matter, significant progress has been made to map microphysical parameters onto cosmological phenomenological parameters (e.g., Refs. [437, 438]). However, the interactions between particles that drive the exchange of energy and momentum at late times pose problems, as the dark matter cannot be modeled in either of the “simple” limits of collisionless or maximally collisional (i.e., fluid), for which there is a vast body of literature. To model the macroscopic effects of these moderately frequent microscopic collisions, various authors recommend using different forms of the cross section (viscosity, transfer) in different limits (e.g., a Debye-like plasma limit) to model interactions of simulation particles, but it is not clear which is on the firmest theoretical ground. Velocity-dependent cross sections are especially problematic if the cross section is highly peaked in velocity space in a region of phase space poorly sampled by simulation particles. Even in the case of non-interacting sterile neutrinos and light thermal WDM models, the non-cold momentum distribution is difficult to map to non-linear scales. There have been recent technical advances in calculating transfer functions and removing numerical artifacts from simulations [438, 402, 188]. More work is needed on this subject.

In summary, the key points of the astrophysics theory program are:

1. The highest priority for predictions are for the types of astronomical objects for which observations currently or will soon exist (e.g., the exponentially growing ensemble of substructure lenses), where predictions should include simulations of observations (a la Ref. [357]). However, the theory program must be flexible enough to go on “fishing expeditions” for novel signatures of dark matter microphysics.
2. Predictions for the effects of dark matter on structure formation must be made with the baryonic physics of galaxy evolution included. The uncertainties in galaxy evolution physics should be marginalized over in predictions.
3. It is important that the algorithms used to model non-minimal dark matter physics in simulations be on firm theoretical footing. This is an example of an issue where close collaboration between particle physicists and simulators is essential.
4. Predictions for observations as a function of dark matter microphysics will require a hybrid approach of simulations and semi-analytic modeling. Simulations have high spatial resolution but are extremely costly. Semi-analytic models must be informed by simulations, but are fast and well-suited to statistical studies of astronomical objects and for parameter constraints from observations.
5. As much as possible, phenomenological models of dark matter for cosmic structure evolution that can encompass many microphysical models (e.g., Refs. [439, 402, 188]) should be used in order to minimize the computational overhead for theory predictions.

Applications to Direct and Indirect Detection

We are proposing a program to improve our understanding of astrophysical structures in order to aid the discovery of new physics in the dark sector. However, imagine that — ten years from now — a new WIMP-like particle is found at the LHC, and that astronomical observations are completely consistent with a minimal inflationary ΛCDM Universe. How will the theory program described above, meant to carefully calibrate $M_{\text{halo}}$, and identify specific signatures of new dark-matter physics in observations, be helpful in this scenario? Would all the work done to search for departures from the predictions of CDM across multiple scales of $M_{\text{halo}}$ be in vain, if a particle candidate for cold dark matter was found?
Importantly, the theory program outlined in this paper quantifies the baseline CDM+baryons model from which $M_{\text{halo}}$ (the scale of deviations from CDM) is defined. If a CDM-like particle is discovered, we can use the baseline CDM+baryon astronomical model to greatly improve the “traditional” astroparticle experiments searching for WIMPs: direct and indirect detection. Any particle discovered at the LHC which is “dark matter-like” requires consistent positive signals from one or both of these classes of experiments to unambiguously identify the new particle discovered at colliders as the same particle that makes up $\sim 25\%$ of the Universe’s energy budget.

The program we describe will lead to much better constraints on the density and velocity distribution of dark matter in halos, and thus will reduce astrophysical uncertainties and improve the experimental reach of direct and indirect detection experiments. In other words, with better astrophysical modeling of dark matter, we sharpen measurements of $\Lambda^{-1}$ with direct and indirect detection methods.

Indirect detection searches for dark matter annihilation or decay in the Universe today can improve signal over background by targeting high-density regions — be it the Galactic Center [441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456], satellite galaxies [457, 458, 459, 460, 26, 27], or galaxy clusters [461, 462, 463, 464]. Improving our understanding of these systems will both provide more targets for indirect detection [465, 466], and remove sources of astrophysical uncertainty (e.g., the density profile of halos) which limit our knowledge of the underlying particle physics parameters [467].

Similarly, direct detection of dark matter relies on high-velocity dark matter scattering on target nuclei (in the case of WIMP-like dark matter) [13], or on resonant conversion of dark matter into visible particles (as in the case of the ADMX search for axions [87]). Currently, constraints on the particle properties of dark matter from direct and indirect detection are completely degenerate with the poorly quantified dark matter density and velocity structure of halos. In all cases, knowledge of the local density and velocity distribution of dark matter can improve the sensitivity of the searches, reduce the astrophysical uncertainties, and sharpen particle physics particle constraints [468, 469, 28]. Therefore, even in cases where there is no physics in the dark sector beyond that consistent with CDM (i.e., $M_{\text{halo}}$ immeasurable small), the results of astrophysical studies that probe this parameter will be of practical use in the search for dark matter.

5.2. Observations

Fundamentally, the theory and observational programs are intertwined. Theory predictions are required to interpret observations, as well as suggest new types of observations that have the potential to be good probes of dark matter physics. On the other hand, as illustrated in Section 4.1, observations can reveal unexpected phenomenology that demands a theoretical explanation. In this section, we will highlight near-term prospects for observations relevant to different $M_{\text{halo}}$ scales, and show where new ideas and new types of observation are needed.

A major theme in this section is that current and next-generation wide-field astronomical surveys designed for dark-energy and time-domain science should yield major results for dark matter, at a small marginal cost. There are many opportunities for motivated dark-matter enthusiasts and cosmologists interested using their dark-energy-finding skills on the other dark side. Moreover, with the next astronomy decadal survey just around the corner, now is the time to start identifying new types of observational projects that will extend the reach of dark matter astrophysics in the future beyond the projects currently under construction.

Throughout this section, we will refer to a number of astronomical and cosmological surveys and facilities. We provide a summary of these surveys in Table 1.

We organize this discussion by $M_{\text{vir}}$ scale, working from large to small scales. This section is summarized in Figure 10.

Cosmological scales ($M_{\text{vir}} \gg 10^{15} M_\odot$; homogeneous and linear-regime Universe)

Measurements from the cosmic microwave background and galaxy clustering at low redshift tell a consistent story about the contents of the Universe, about the abundance of dark matter, dark energy, and baryons. And yet in this age of precision cosmology, a few points of tension remain between the Universe at high- and low-redshift. We briefly discuss these, their implications for dark matter, and upcoming projects that will be able to address these tensions head-on.
Table 1: Summary of present and near-term astronomical surveys or instruments whose capability to probe aspects of dark matter physics is discussed in the text. The surveys are all multi-purpose — for example, although listed under astrometry, JWST will also dramatically enable substructure lensing measurements and distance measurements to Local Volume ultrafaint and ultradiffuse galaxies.

Measurement of the Hubble constant $H_0$ has been an active area of research for nearly a century. At the present, measurements from the nearby cosmic distance ladder [493, 494, 495] and from the CMB and large-scale structure [496, 497, 124, 498, 499] achieve a precision of a few percent, but the two types of measurement are in tension with each other at the $2\sigma - 4\sigma$ level (central values ranging from $H_0 = 67 \text{ km/s/Mpc}$ to $73 \text{ km/s/Mpc}$). Measurements in the local Universe consistently find larger values of $H_0$ than those from cosmological scales (but see Ref. [499]). It is at present difficult to reconcile these measurements in the standard cosmological model [500, 501, 502], although it should be noted that the level of discrepancy is insufficient to definitively point to new physics.

In addition, there exists tension between high- and low-redshift estimates of the amplitude of matter fluctuations, quantified in terms of $\sigma_8$, although the tensions are not as severe as for $H_0$ [503]. In this case, early-Universe estimates of $\sigma_8$ are larger than those from low-redshift cosmological probes.

In both cases, non-minimal dark matter solutions have been invoked to reconcile the measurements of these fundamental cosmological parameters. Decaying dark matter can reduce the amplitude of matter fluctuations with time, as well as change the expansion history of the Universe compared to the standard $\Lambda$CDM case [504, 505, 506, 507, 508]. Hidden sector dark matter with a dark photon can also ameliorate these tensions, by introducing light degrees of freedom (either weakly or strongly coupled) before recombination. This changes the expansion history of the Universe at early times (notably changing the redshift of matter-radiation equality), and thus affects acoustic horizon [493, 509, 510, 511, 512, 513]. It also affects the growth of small-scale structure, through a combination of free streaming at late times and the shift in matter-radiation equality (before which time the growth of structure is suppressed).

The current and next generation of telescopes will be critical for determining if these anomalies are real, or statistical fluctuations. On the most local scale, the Hubble Space Telescope (HST) and the Gaia mission will enable 1% measurements of Cepheid stars, the bottom of the cosmic distance ladder [514, 515]. Hence, low-$z$ measurements of the Hubble constant should approach percent-level precision in just a few years. Stage-IV CMB [477, 474] and galaxy [484, 482, 516, 472, 485, 180] surveys (see Table 1) should shrink error bars on the number of light degrees of freedom, $N_{\text{eff}}$, and $\sigma_8$, by factors of several to factors of nearly ten [512, 473, 509, 510, 511, 512, 513]. We also note that LIGO observations of “standard sirens” can provide an orthogonal measurement of $H_0$ with competitive error bars in the near future [518, 519]. It is important to continue to consider novel effects from dark matter physics that may affect observables in these experiments, especially as it pertains to $H_0$, $N_{\text{eff}}$, and $\sigma_8$ [512, 520, 513].
| $M_{\text{halo}}$ | Probes |
|-----------------|--------|
| $10^2 M_\odot$ | Gravitational nanolensing (time domain) | Gravitational waves from compact-object DM (multi-messenger) |
| $10^4 M_\odot$ | Milky Way stellar halo perturbations (astrometry) | Milky Way stellar halo perturbations (astrometry) |
| $10^8 M_\odot$ | Substructure lensing | Substructure lensing subhalo mass functions of group & cluster halos (galaxy surveys, ground- and space-based spectroscopy) |
| $10^{11} M_\odot$ | Dwarf & ultradiffuse galaxy counts as a function of z (wide-field galaxy surveys, targeted surveys) | Lya forest (spectroscopy) |
| $10^{14} M_\odot$ | Initial mass function | Stellar-mass—halo mass relation w/ cosmological tools on wide-field surveys |
| $10^{15} M_\odot$ | Cluster component offsets (lensing, wide-field surveys) | Cluster mass from wide-field surveys |
| $10^{17} M_\odot$ | Local measurements of $H_0$ (astrometry) | Galaxy survey & CMB measurements of $H_0$, $\sigma_8$, $N_{\text{eff}}$ |

Figure 10: $M_{\text{halo}}$ scales accessible with future probes, color-coded by what aspect of dark matter astrophysics they will address. Observations marked in blue will weigh in on the (sub)halo mass function, those marked in purple will enable halo mass estimates, while those in orange will facilitate measurements of the mass distribution within halos. Red marks cosmic-scale measurements, while those probes marked in green measure aspects of dark matter astrophysics not captured by the other probes. This figure summarizes Section 5.2, and the halo mass ranges map onto our section titles. Please see each relevant subsection for more details.

Clusters: $10^{14} M_\odot \lesssim M_{\text{vir}} \lesssim 10^{15} M_\odot$

Galaxy clusters, the largest gravitationally bound objects in the Universe, are excellent laboratories for dark matter physics on account of their deep potential wells and abundance of substructure. In Section 4.1, we already saw some evidence pointing to galaxy clusters ($M_{\text{vir}} \sim 10^{14} M_\odot$) having density profiles shallower than NFW, an intriguing hint for dark matter science. Currently, the density profiles of cluster-scale halos are well fit with self-interaction cross sections of order $\sigma/m \sim 0.01 - 0.1 \text{ cm}^2/\text{g}$ (notably lower than suggested for dwarf systems) [119, 407]. Two other classes of observation shed light on other aspects of dark matter physics on this scale.

First, enabled by the Hubble Frontier Field initiative [521], strong lensing maps of a handful of clusters reveal an exquisite amount of substructure down to $M_{\text{vir}} \sim 10^{9.5} M_\odot$ (with the caveat that this is the tidally stripped — not infall — subhalo mass) [522, 523, 524]. The amount of substructure is consistent with CDM predictions, although large subhalos are perhaps too abundant in the merging cluster Abell 2744 [522].

Future surveys will be hard-pressed to improve on the data from the Frontier Fields for individual systems: the excellent spatial resolution combined with the extreme depth of the field (measured in 100’s of kiloseconds in exposure) which are needed to find many faint strongly lensed background galaxies will be hard to beat without a dedicated campaign. However, constraints may be improved by stacking clusters, and
with dedicated campaigns. Expanding the number of well-studied clusters — especially relaxed morphologies which make for more straightforward comparisons with theory — would be possible with WFIRST. The large footprint of the WFIRST camera makes it possible to achieve similar depth to the Frontier Fields with far fewer pointings than with HST.

Second, strong lensing maps enable a study of the distribution of dark matter, gas, and galaxies in clusters, making them sensitive to offsets between these components. During the process of cluster growth, either by the minor merger of individual galaxies or small groups, or through major mergers (with a mass ratio between primary components of 10:1 or smaller), it is expected that galaxies trace the dark matter component. This is because galaxies (as compact objects on the scale of clusters and comprising only \( \sim 2\% \) of the cluster mass budget), effectively behave as collisionless particles during cluster mergers, just like collisionless dark matter. Since gas strongly interacts, there may be separation between the gas and the galaxies/dark matter \([525]\). Gas is easily ram-pressure stripped during the merger process. Galaxies can be traced by their light, gas by its X-ray emission, and dark matter by gravitational lensing.

In observations of cluster-cluster mergers, or of galaxies falling onto clusters, there are frequently observations of separations between the galaxies and the dark matter, in addition to the expected separation between galaxies/dark matter and gas (e.g., \([526, 527, 229]\)). It has been speculated that this may be a sign of SIDM, where the self-interactions (characterized by a cross section per mass \( \sigma/m \)) induce a drag-like force on halos \([225, 226, 229, 232]\). Because the column density of dark matter in clusters is much higher than in smaller halos, the interaction probability of dark matter particles is likewise higher in clusters \([232]\). However, recent works shows that the separation between dark matter and galaxies is expected to be small in SIDM during the merger, in some cases much smaller than the observed separations, even for large \((\sigma/m \sim 10 \text{ cm}^2/\text{g})\) cross sections \([227, 228, 232, 230, 231]\). The smallness arises because the galaxies are gravitationally bound to the halo, and thus respond gravitationally to the halos' motion and evolution. The recent work casts doubt on the Bullet Cluster limit of Ref. \([528]\). The observed offsets between galaxies and halos remains a mystery, although some of it is almost certainly a result of sparse galaxy sampling effects on centroiding the galaxy distribution \([529]\), and lensing systematics \([530]\).

However, although offsets predicted by SIDM are smaller than originally anticipated, they are not negligible. Moreover, with an ever-increasing sample of merging clusters, powered by wide-field radio surveys \([531, 532, 533]\), continuing into the SKA era, we will be able to follow those systems that are most likely to be clean tests of SIDM (Table 1). The wide-field strong lensing enabled by WFIRST will be particularly valuable here, assuming that the systematics underlying the anomalously large offsets described above can be brought under control.

Perhaps more excitingly, Ref. \([232]\) showed that, for SIDM models, separations between galaxies and the center of the dark matter halo are expected in the relaxed merger remnant, and that these offsets may be much more significant, both in magnitude and duration, than the transient offsets found during the merger process. This is because SIDM halos are cored, and dynamical friction is ineffective in cored (i.e., harmonic oscillator) potentials \([531, 534]\). As such, galaxies at the center of the halo “slosh” \([232]\) or “wobble” \([536]\) about the center of the halo. Departures of \( \sim 100 \text{ kpc} \) are expected for self-interaction cross sections of order \( \sim 1 \text{ cm}^2/\text{g} \), rather than the \( \sim 10 \text{ kpc} \) for offsets during the merger for a similar cross section. However, new hydrodynamics simulations \([407]\) hint that the amplitude of the wobble may be somewhat smaller than found in the dissipationless simulations of Ref. \([232]\). Clearly more simulation work is needed on this front. Intriguingly, there is already an abundance of relaxed cluster data that could be used for a sloshing study \([537, 538, 539, 540]\). Sloshing is a smoking gun for a shallow density profile toward the center of halos, and investigation thereof should be a high priority for the field. Finding the “true” centers of halos is important for dark energy science, which has motivated nearly all of the work performed so far on galaxy cluster centers; the synergy with dark matter science makes this measurement even more compelling.

These measurements can also be used to measure the shapes of cluster-scale halos. Dark matter self-interactions can alter the shapes of halos, making them rounder than expected from CDM cosmologies \([213]\). While cross section constraints are presently at the \( 1 \text{ cm}^2/\text{g} \) level — not competitive with these other measurements — we expect improvements in constraints as more hydrodynamic simulations of clusters are made \([407]\). The extraction of SIDM constraints from cluster scales is currently theory-limited rather than limited by observation, a problem which will be exacerbated without serious theory effort as the observational
data sets are expected to mushroom in size in the LSST and WFIRST eras.

Finally, we close this section with a discussion of halo masses. Because of the exponential drop-off of the cluster mass function and its exquisite sensitivity to the cosmological growth function, clusters are an important dark energy probe. Characterizing the mass function correctly depends on a faithful mapping between cluster masses measured in simulation and observation. There is enormous effort in this direction in the dark energy community (e.g., Refs. [511, 572, 573]); the dark matter community will benefit from this work, as the cluster mass function is also sensitive to late-decaying dark matter models [503, 506]. Better mappings of the baryon and dark content of clusters will also be beneficial for more accurate estimates of self-interaction cross sections from observations of offsets and density profiles.

**Galaxies and galaxy groups** ($10^{11} M_\odot \lesssim M_{\text{vir}} \lesssim 10^{14} M_\odot$)

On this scale, the connection between dark-matter halos and the galaxies that inhabit them is well-measured, on average, with the plethora of available tools (including lensing [571], two-point statistics [514], abundance matching [137]). The stellar mass of the central galaxy (Section 3.2.1) is estimated to between $M_\ast \sim 10^9 M_\odot$ to $\sim 10^{11} M_\odot$ for this range of halo masses. The consistency with vanilla ΛCDM is generally good for objects in this mass range, but with a new class of object (ultradiffuse galaxies) beginning to perhaps muddy the waters again. The outstanding “hint” (Section 4.1) is the cusp/core problem, although the puzzle over ultradiffuse galaxy origins is starting to hint at new additions to field galaxy (and halo) counts. In this section, we explore this new ultradiffuse direction and the big question for this halo mass range: What is the central dark matter density profile, and does that profile make sense in the context of baryonic physics for a ΛCDM cosmology? Addressing these questions is essential to assess the (baryonic or dark) origin of the observed cusp/core problem, and to properly counting dark matter halos (Section 4.1).

Answering both of these questions requires understanding two properties of galaxies at these scales: the link between galaxies and their halos (the SMHM relation) as a function of galaxy properties, and the initial mass function (IMF) of stellar populations. The former sets the energy budget for baryons or dark matter physics to alter the dark matter profile shape from the CDM prediction of NFW, and the latter determines how well we can infer the density profile of the central regions of dark-matter halos from dynamical mass measurements.

While the link between galaxy and halo mass in the galaxy-group range is well-established for typical high surface brightness galaxies (e.g., Refs. [571, 512]), there is increasingly strong evidence that the average halo mass varies significantly as a function of galaxy properties for fixed stellar mass [548, 559]. In fact, Refs. [571, 548] find that red (non-star-forming) galaxies live in halos at least twice as massive as those galaxies that are blue (and star-forming).

In recent years, a major surprise (and new mystery) on these mass scales is the discovery of extreme outliers to the SMHM relation. A new class of ultradiffuse (extremely low projected luminosity density) galaxies has been discovered, largely in galaxy clusters (because that is where people looked first [549, 550, 551, 552, 553, 128, 554]), but also in the field [555, 556, 557]. Many, if not most, of these galaxies are “conventional” dwarf galaxies in low-mass halos (stellar masses $M_\ast \ll 10^9 M_\odot$ in halos of $M_{\text{vir}} \lesssim 10^{11} M_\odot$), which on average have lower luminosity densities than large galaxies [532]. However, others appear to be dwarf-luminosity galaxies living in LMC- to Milky-Way-mass dark-matter halos [558, 559, 560, 561]. The most famous example is Dragonfly 44 in the Coma cluster, which is a galaxy with a stellar mass only 0.5% that of the Milky Way’s, but residing in a halo approximately as massive [558]. Understanding these systems is important both for galaxy evolution and dark-matter studies, and suggests that the current census of even high-mass halos is incomplete. Thus, ultradiffuse galaxies hint at a new counting problem on galaxy-halo-mass-scales, but also further complicate the cusp/core problem.

Because hydrodynamic simulations predict that the halo profile is a strong function of star-formation history and halo mass (see the discussion in Section 4.2), an accurate observational mapping between galaxies and their halos is essential to assessing the proposed baryonic solutions to small-scale problems. We strongly endorse efforts to further tease apart the connection between galaxies and their halos as a function of galaxy properties other than mass, down to much lower stellar masses ($M_\ast \lesssim 10^{10} M_\odot$), which will be enabled by clever new methods (e.g., Refs. [559, 560]) applied to the next generation of deep wide-field surveys (Table 1 optical survey list) [484, 482, 471, 472, 485, 483]. The spectroscopic capability of thirty-meter-class giant
optical telescopes (Table 1) is essential to characterizing the halo masses of ultradiffuse galaxies—their low surface brightness currently makes follow-up spectroscopy infeasible for samples greater than $O(10^2)$ objects for ten-meter-class telescopes [558, 562].

While the total matter profile of galaxy- and group-scale systems are well-measured observationally, interpreting measurements in the context of dark-matter halo profiles is complicated by low-mass stars [563, 246, 564, 264, 270, 272]. Low-mass ($< M_{\odot}$) stars dominate the number count and mass of stars in galaxies, yet contribute little to the total luminosity. Thus, determining the stellar mass function at birth (the IMF) is critical for accurately inferring the mass of stars in a galaxy from the galaxy light. Only then can one subtract the baryons from the total mass profile to determine the dark-matter distribution in the inner parts of dark-matter halos. Recently, Refs. [565, 566] used stellar absorption features of elliptical galaxies to demonstrate that elliptical galaxies tend to have relatively fewer low-mass stars than the Milky Way. Since then, evidence has suggested that the initial mass function varies from galaxy to galaxy, and also within galaxies [567, 568, 569]. This is an active area of research, and has significant implications for the dynamical modeling of galaxies. It should be noted that the mass distribution of halos outside the luminous region can be measured with gravitational lensing and are consistent with NFW profiles, although it is not expected that baryonic physics affects halos so far from the centrally-concentrated stellar and cold gas components [570].

Thus, we expect significant progress in assessing solutions to the cusp/core problem on the scale of large-ish galaxies, and to a more accurate accounting of large halos with the discovery of more ultradiffuse galaxies. Next, we consider the visible and invisible substructure of those galaxies.

**Dwarf galaxies and linear scale counterparts ($10^8 M_\odot \lesssim M_{\text{vir}} \lesssim 10^{11} M_\odot$)**

It is in this mass range where the most significant hints of problems with CDM occur, as described in Section 4.1. It is therefore in this mass range where observational progress is especially important and (hopefully) imminent if there is sufficient effort by dark-matter-oriented (astro)physicists. Progress is expected on at least three fronts:

1. Robustly constraining the abundance, mass profiles, and orbits of Local Group dwarf galaxies.
2. Matching galaxies with halos.
3. Counting small galaxies (and their halos) at a variety of epochs.

Significant progress here will go a long way to discovering the dark or baryonic origin of the hints described in Section 4.1.

We consider the Milky Way satellites first. The Milky Way’s satellite dwarf galaxies are central to hints of new dark physics, and understanding them is essential to revealing the origin of the hints. Their importance comes from the fact that we can study intrinsically faint systems best if they are close to us.

In Section 4.1, we argued that the missing satellites problem is considered solved, but there are still important related unresolved problems that will only be answered in the LSST era with a more complete survey of dwarf galaxies. The most pressing issue is the precise number and distribution of satellite galaxies within the Milky Way, and field dwarfs in the Local Group [571, 123]. Based on completeness corrections of the known Milky Way satellites, Ref. [123] predict a relatively low number of Milky Way satellites more luminous than Segue I ($\sim 100 - 150$) if the satellites are centrally concentrated, but up to 2000 if they are distributed closer to the prediction of Refs. [384, 572]. If the latter is found to be the case, we may have a “too many satellites problem” — that there are more satellite galaxies than we can comfortably accommodate with a CDM subhalo mass function without putting galaxies in truly tiny halos. Observations with LSST are essential to matching galaxies with subhalos in the Milky Way.

At present, there are two major sources of uncertainty about the kinematics of the Milky Way dwarfs, which arise from the fact that we have typically only line-of-sight velocities for bright stars in these galaxies (i.e., only one of the three velocity coordinates). First, the mass of the Milky Way halo is uncertain to a factor of two. Statements about how unusual (or not) the Milky Way’s satellite population is (e.g., the presence of the Magellanic Clouds) are highly sensitive to the mass of the Milky Way halo. With full three-dimensional bulk velocities for dwarf galaxies and globular clusters — enabled in the future with a combination of line-of-sight velocities from ten- and thirty-meter-class telescopes and plane-of-sky proper
motions with space-based astrometry (Table 1) — it is possible to put significantly better constraints on the Milky Way halo mass [334, 573]. Stellar streams are also an essential tool in the Milky Way mass estimation toolkit, since stars in a stream have almost identical initial conditions, and map out a large fraction of the Milky Way potential [360, 361, 315]. The discovery of new tidal streams with deep wide-field surveys, and their characterization in phase space with spectroscopy and proper motions, are important for maximizing their use in mapping the potential well of the Milky Way.

Second, solutions to the TBTF and dwarf-scale cusp/core problems are hampered by the lack of three-dimensional velocity data for individual stars in galaxies. There is a major degeneracy between the mass profile of dwarfs and the orbital structure of stars in those galaxies. This degeneracy can be broken if the full three-dimensional stellar velocity vectors are measured [324, 573]. The uncertainties related to the internal velocities of satellites will be significantly ameliorated when it is possible to measure stellar motion in the plane of the sky, in addition to line-of-sight velocities. Although much good work has been done with HST and will be done with JWST, the field will truly open up with the astrometry performed by the Gaia and WFIRST satellites [472, 362]. Interestingly, the Gaia satellite does not go deep enough to probe the internal velocities for most Milky Way satellites (although it does for bulk velocities [574, 575, 576, 577]), but it will play a critical role in anchoring the astrometric reference frame for WFIRST studies. We strongly endorse the work of the WINGS science investigation team and the WFIRST astrometric working group’s effort to explore the implications of WFIRST astrometry [362] for Local Group dark matter science.

While proper motions will enable incredible studies of the internal kinematics and structure of classical dwarfs ($M_r \gtrsim 10^5 M_\odot$; see Section 4.1) in the Local Group, we need new ideas to measure densities and density profiles of galaxies that are either small and/or far away. The smallest galaxies, the ultrafaints, contain insufficient stars for the shape of their dark matter density profile to be measured, even with full three-dimensional stellar kinematics, although the observational depth of WFIRST will enable the use of some of the more abundant faint stars for proper motions. However these are the galaxies which are expected to have cusps in CDM regardless of baryonic physics, which increases the importance of observational tests of their central density profiles. For small galaxies beyond the Milky Way, follow-up considerations are different depending on whether galaxies can be resolved into individual stars or not. For both, even line-of-sight velocities will be a challenge—thirty-meter-class telescopes will be able to follow up only a small fraction of new discoveries. With ten-meter-class telescopes, line-of-sight velocities for individual stars are obtained only for distances less than 1 Mpc. Beyond that, we need light integrated over many stars and (especially) gas to get a signal in a decent amount of time. The presence of nebular emission lines is essential to obtaining good velocity information in a non-prohibitive amount of time. However, we expect the faintest satellites, the ultrafaints, to be devoid of ionized gas (although they might possibly have small reservoirs of neutral gas in the field [599]). Thus, the the only tool available to us is photometry. We strongly endorse the development of novel ways to estimate the density profile of small galaxies with photometry alone, as proposed by Ref. [248].

The second important pathway to progress is determining how to match galaxies with halos, and to find ways to constrain the matter power spectrum on dwarf scales that do not necessarily depend on the exact mapping of galaxies to halos. The former is essential for determining if baryons can solve the cusp/core problem in a CDM framework, and the latter is important to look for “primordial” effects of dark matter on structure formation. In both cases, we strongly encourage the cosmology community to extend their tools to low(er)-mass halo regimes if at all possible. As discussed above, galaxy-galaxy weak lensing and two-point correlation functions can be used to match galaxies to halos statistically [571, 578, 544, 547]. Ref. [579] presented first results from a study of atomic hydrogen-selected galaxies at mass scales similar to the Magellanic Clouds. Intriguingly, a recent study of cosmic shear with CFHTLenS suggests that the matter power spectrum is consistent with CDM down to scales of $M_{\text{halo}} \approx 5 \times 10^9 M_\odot$ [580]. With the flowering of wide-field surveys (Table 1), we expect constraints on the matter power spectrum and on the statistical matching of galaxies with halos, to become robust on dwarf scales, although there are important caveats [581].

---

20R. Sanderson, WFIRST in the 2020’s meeting, http://www.stsci.edu/~dlaw/WFIRST2020s/slides/sanderson.pdf
Finally, we expect significant advances in the counting of individual halos by the galaxies (or gas clouds) they host in the near term, and at a wide range of redshifts.

At high \((z \sim 10)\) redshifts, dwarf galaxies are expected to reionize the Universe; there are simply not enough big galaxies or quasars at early times to yield sufficient ionizing photons \([582, 583, 584, 585, 586, 587, 588]\). Thus, dwarf searches are of high interest at high redshift. The effect of the matter power spectrum on \(k \sim 0.1 \text{ Mpc}^{-1}\) scales (Figure 4) may be observed at \(z \sim 10 - 20\) 21-cm radiation \([589]\) (Table 1), and are sensitive to the first stars rather than to the first galaxies \([590]\). At \(z \sim 10\), dwarf searches are one of the primary drivers for the upcoming James Webb Space Telescope (JWST) \([591, 592]\), and of major interest for WFIRST \([21]\).

The UV luminosity function of galaxies near the epoch of reionization can be an effective probe of power spectrum cutoffs (already excluding thermal relic WDM masses below 2.4 keV), with significant improvement expected in the JWST and WFIRST eras \([593, 429]\). The star-formation histories of nearby galaxies, determined today with HST but in the future with JWST, can also be used as a “time machine” to constrain the reionization-era galaxy stellar mass function, and hence, the matter power spectrum \([199, 411, 594, 595, 596]\).

Again, we encourage efforts to constrain using primordial dark matter physics using the galaxy luminosity or stellar mass function at early times.

Moving down in redshift, from measurements of the Lyman-\(\alpha\) forest at \(z \sim 3 - 6\) \([597, 598, 599, 600, 118]\), the CDM paradigm is known to be consistent with observations down to \(M_{\text{halo}} \sim 10^9 \, M_{\odot}\) \([601, 118]\). The Lyman-\(\alpha\) forest is a series of Lyman-\(\alpha\) absorption features from gas clouds backlit by quasars. Because gas density roughly traces the matter field, it can be used to constrain the matter power spectrum at moderate redshift, albeit with uncertainties related to the thermal history of the intergalactic medium (IGM) \([118]\).

The resulting information about the number of gas clouds at varying redshift can be used to constrain deviations from CDM. For example, free-streaming of “warm” (semi-relativistic) dark matter would suppress the power spectrum on scales probed by the Lyman-\(\alpha\) forest. This constrains thermal relic dark matter to be heavier than 3.5 – 5.3 keV, depending on assumptions about how smoothly the temperature of the IGM changes with time \([118]\). In the future, measurements from X-SHOOTER on the VLT, WEAVE-QSO on the William Herschel Telescope, and the DESI survey will expand the resolution, sightlines, and redshift of the Lyman-\(\alpha\) forest dataset \([480, 602, 603]\), improving these limits. We strongly support continued efforts to measure the Lyman-\(\alpha\) forest and isolate systematic uncertainties in their interpretation.

Statistical samples of dwarf galaxies are being obtained in the local Universe using a variety of methods, the sample size of which will increase vastly with LSST in particular. A galaxy luminosity function (or stellar mass function), robust to surface brightness detection limits, is a critically important input for abundance matching and two-point clustering statistics to associate galaxies with halos, as well as to disentangle baryonic from dark matter effects on the core/cusp problem, and to solve the “TBTF in the field” problem (Section 4.1).

The SHIELD survey focuses on finding field dwarfs in atomic hydrogen \([492]\). Other radio and intensity mapping surveys (Table 1) will or already have uncovered more. Optical surveys — either specialized like MADCASH \([604]\), or general-purpose like the Dark Energy Survey \([288]\) (Table 1) — will be able to reach ultrafaint, or nearly ultrafaint \([338]\) scales in the Milky Way and beyond.

As mentioned earlier in this section, the biggest challenge after finding these systems is to characterize their kinematics, internal and bulk, and measure their distances from us. The low surface brightness of many of these objects and the paucity of ionized and neutral gas in the lowest-stellar-mass systems makes both extremely challenging. Distances for nearby dwarfs are currently obtained either with variable stars or using the “tip of the red giant branch” (TRGB) method \([605, 606]\), both of which require resolved stellar populations. Beyond a few megaparsecs, this requires space-based angular resolution, notably HST, JWST, and WFIRST. We strongly endorse efforts to find and characterize dwarf galaxies using a variety of methods, in a variety of environments, and at extremely low surface brightnesses; and to develop new methods to characterize these galaxies. We also endorse the application of statistical techniques developed for higher-redshift data \([607, 608]\) to the near field.

The broad message of this section is that enormous new and powerful data sets are imminent or already extant, and we must learn how to mine them for dark matter science. We strongly encourage the use of

\[21\] S. Finkelstein, WFIRST in the 2030’s meeting
http://www.stsci.edu/~dlaw/WFIRST2020s/slides/finkelstein.pdf
“standard” tools in statistical cosmology to approach dwarf-scale problems, to complement the current approaches largely driven by the galaxy evolution community. We anticipate many opportunities for important progress in resolving both the core/cusp and counting issues on dwarf scales in the next decade. However, work will be needed needed to propagate the astrophysical results out to theorists, as well as bringing new theoretical ideas to the simulators and observers to test.

Invisibles \((10^5 \, M_\odot \lesssim M_{\text{vir}} \lesssim 10^8 \, M_\odot)\)

Below halo masses of \(\sim 10^8 \, M_\odot\), we expect dark matter halos to be largely devoid of baryons \((122, 144, 142, 123)\). Thus, measurements of dark matter halos on these scales must rely on gravity alone. The two major methods under consideration on these scales are gravitational lensing and Milky Way stellar halo and stream perturbations (Section 3.2.1). We note that although on these scales dark matter searches are typically discussed in terms of halo counts or halo mass functions, the distribution of matter within halos can affect their observability. Hence, it may be possible to unravel both primordial and evolutionary deviations from CDM, even without the presence of baryons. The most exciting aspect of tests in this mass scale is that the promise of lensing and stellar perturbations are about to be realized in new data sets. Here, we describe why the data sets are about to grow exponentially in size, and what the remaining roadblocks are in mapping observations to \(M_{\text{halo}}\).

As discussed in the previous section on dwarfs, weak lensing may possibly used to probe this regime \([580]\). However, in this regime of halo mass, we usually discuss strong gravitational lensing as a probe of the abundance of small halos \([303, 305, 304]\). As we showed in Section 4, substructure along the line of sight connecting us to a galaxy- or galaxy-group-scale lens (including substructure in the lens and structure outside of it but along the line of sight) can perturb the apparent flux, position, and arrival time of light from strongly lensed background galaxy or active galactic nucleus (AGN). Up until \(\sim 2010\), most work focused on flux-ratio anomalies from seven radio-loud, quadruple-image AGN. Excitingly, the flux-ratio anomalies were consistent with CDM, albeit with large uncertainties \([308]\). Since 2010, the method of “gravitational imaging” \([609, 309]\) has gained traction, mapping out perturbations in a few dozen Einstein ring galaxies, which are also consistent with CDM but again with large uncertainties \([610, 611, 612]\).

Why so few targets, and what is the path to stronger tests of dark matter microphysics? Radio-loud AGN were desired for flux-ratio anomalies because they are less prone to contamination by the microlensing induced by stars in the lens galaxy in optical bands, as the image size in the lens plane is comparable to the Einstein radius of stars as well as small halos. The mid-IR and radio emission from AGN comes from a much larger region, and it much less susceptible to microlensing (see Ref. \([311]\) for an illustration). But such AGN are rare. Up until recently, there were only seven known suitable systems. In addition, a further complication is that for both flux-ratio anomalies and gravitational imaging, the angular size of systems is small (on the order of arcsecond(s)), meaning high-resolution imaging is required to study these systems.

What is changing on the data side? First, new wide-field surveys are finding more suitable candidates \([479, 613, 614]\) — even SDSS is still being mined for lenses \([615]\). Hundreds to thousands of quadruply lensed AGN are expected to be discovered, with commensurate numbers of Einstein ring systems \([616]\). Second, new facilities are enabling follow-up of these candidates. New adaptive optics facilities provide the high-resolution imaging and spectroscopy required for these studies \([610, 617]\), the mid-infrared capabilities and grisms of JWST will enable many more quadruply-imaged AGN to be used for substructure studies, and the high-quality astrometry available in the radio will enable low-mass subhalo detections via gravitational imaging. Third, new methods are being pioneered that allow most quadruply-lensed AGN to be used to measure flux-ratio anomalies without microlensing contamination with space telescopes (HST, JWST, and WFIRST \([617, 311]\)), or enhance the science return on gravitational imaging surveys \([618, 619]\). Finally, Ref. \([620]\) opened a window to study new sources with new facilities: clumpy, high-redshift galaxies with the ALMA interferometer \([621]\). There is much to be done even with HST \([515]\). The landscape of what is possible is changing rapidly.

Likewise, the prospects for finding small subhalos in the Milky Way halo are also changing with the advent of new data and new analysis methods. As satellites of the Milky Way, especially the mysterious ultra-dense-in-stars globular clusters, are disrupted by tidal forces from the Galaxy, stars disperse along tidal stream. Dark matter subhalos can “punch” through the streams, perturbing stellar orbits and leaving persistent gaps.
To probe the Galactic subhalo population and not the Galactic bar or molecular clouds, we require streams with large enough orbital pericenters so that they are insensitive to baryonic substructure, but small enough that the progenitor may be stripped. The discovery of suitable streams in new wide-field surveys, including Gaia, is already happening, and will continue into the LSST era. A combination of deep imaging and proper motions will lead to a large number of exquisitely mapped stellar streams. It is expected that constraints on dark-matter models with primordial deviations from CDM can be significantly tighter than achieved with the Lyman-alpha forest or dwarf galaxy counts, with limits comparable to or better than from the Lyman-alpha forest anticipated. A new study suggests that halos as small as $10^6 M_\odot - 10^7 M_\odot$ may be discovered by their effects on the smooth stellar halo as well, or even at lower masses. Both the theoretical modeling of the stellar halo and its streams, and the discovery of new streams, have advanced dramatically in just the past few years — a trajectory whose derivative will only increase in the near future. The upside of the Milky Way is that its halo can be mapped to high precision with current and future surveys; the downside is that it is a sample of one. It will be extraordinarily powerful if a consistent signature of small halos is discovered in the Milky Way using stellar halo methods, and in other galaxies using other methods.

Because of this new wealth of data, for both lensing and stellar halo methods, what is desperately needed is a clean mapping between dark matter theory and observables. There is progress on both fronts, especially in the case of CDM. But this is a specific instance where progress toward dark matter constraints will soon be theory-limited.

Tests of CDM: the final frontier ($10^{-6} M_\odot \lesssim M_{\text{vir}} \lesssim 10^5 M_\odot$)

One of the strongest astrophysical predictions of the WIMP-ΛCDM paradigm is the hierarchy of halos down to Earth-mass scales. A measurement of halos on these scales would be an astounding confirmation of this theory. Although the predictions for the existence of substructure are strong for CDM, fluffy, extended, and small halos are difficult to detect.

A number of ideas have bubbled up in recent years to detect such small halos, but require additional validation. A number of authors have suggested different ways to use time-domain observations, in all of the strong, weak, and microlensing regimes, to detect halos in this mass regime. These ideas need to be validated in an observational context, but they point to a very important fact: the current and next generation of optical wide-field surveys are time-domain surveys by design. We encourage, in the strongest terms, scientists studying dark matter to think about the new axes of possibility, like the time axis, opened up by next-generation facilities.

Other attempts to discover dark matter on these small scales are model-dependent, but high-impact if those models for dark matter are correct. If dark matter consists of primordial black holes, they can be detected with gravitational waves at LIGO, stellar microlensing, or stellar dynamics in ultrafaint dwarf galaxies. If dark matter consists of WIMPs, the substructure abundance can be backed out of indirect detection observations if a signal is seen in different types of cosmic environments.

This is arguably the most challenging regime for astronomical dark matter searches, but is the one with potentially the largest payoff. We strongly encourage new ideas in this regime, not just ones that exploit new observational capabilities, but ones that can be brought to fruition with dedicated new facilities.

Dark matter vs. inflation

Identifying a theory of dark matter from astronomical observations and experiments means seeing a consistent signature across $M_{\text{vir}}$ scales and in the laboratory. Figure 1 is our recommendation for how to begin characterizing consistency. However, so far we discussed predictions in the paradigm of a primordial scale-free power spectrum from inflation. Inflation can lead to a boost in small-scale power relative to the standard slow-roll case either “primordially” or through altering the thermal history of the Universe.
or a suppression relative to the standard case. There are a number of innovative methods to suss out the former, many which tie into either indirect dark matter searches (tiny compact structures can boost the signal) or gravitational lensing perturbations. The latter effect (modification of the thermal history) can easily be confused with a dark-matter-physics-induced truncation of the matter power spectrum. A novel idea is to use spectral distortions of the CMB to distinguish between the two. This may be possible with the PIXIE satellite.

We endorse further attempts to distinguish between inflationary and dark-matter effects on the power spectrum, and in exploring the connection between inflation and the indirect detection of dark matter.

New ideas

In this section, we have laid out the prospects for both theoretical and observational progress to making an accurate map between dark-matter theory space and astronomical observables (i.e., a more accurate and nuanced version of Figure 1), and to constraining the nature of dark matter more generally. Our main message is that we expect potentially transformational progress in the next decade as theorists and observers come together to maximize the return on new observational and computational facilities and techniques. However, there are some problems that require new ideas; and other problems where the current path to progress is straightforward but slow, and where new insights would be extraordinarily valuable.

There are two major points we would like to make before we conclude.

• First, arguably the strongest and cleanest predictions for CDM are on small scales. In terms of the cusp/core problem, the cleanest systems to study are the ones where the baryons are the fewest and have the most boring behavior: in ultrafaint dwarf galaxies and small, dark subhalos. But the smoking gun of CDM theory is the abundance of tiny, solar-system-sized dark matter halos. A discovery of such small halos would be definitive proof of the CDM paradigm. However, it is extremely challenging to measure the abundance of and matter distribution within such small objects. There simply are not enough stars in ultrafaint dwarfs to distinguish between core and cusp scenarios kinematically. Finding dark subhalos relies on gravity, but halos are extended objects, which complicates searches (compare to primordial black holes, which would be relatively easier to discover on account of their compactness). New ideas that will enable the discovery of the hierarchy of CDM-predicted subhalos with forthcoming or existing facilities, or which could be enabled by bespoke new facilities, are highly desirable.

• Second, there are many new astronomical facilities coming on-line in the next decade, which we listed at the beginning of the section. While dark matter science is not a primary driver for these facilities, we showed here that they have the potential to significantly advance this field. Now is the time to think about and advocate for minor changes in survey design or guest-observer programs that have small marginal costs yet yield enormous progress in dark matter astrophysics.

Astrophysics in the Nightmare Scenario

We end this section with an investigation of the role astrophysics might play in the “worst-case” particle physics scenario: no dark matter signal is seen in any of the particle physics searches in the next decade. In this scenario, there is no evidence found for physics beyond the Standard Model at the Large Hadron Collider; axion experiments (such as ADMX) continue to report null results; no definitive signal of WIMP dark matter in direct-detection experiments; and no sign of dark-matter annihilation from gamma-ray and neutrino telescopes. The lack of an axion signal would imply that QCD axions are excluded as dark matter unless the vacuum misalignment angle $\theta \ll 1$ or most axions cluster in Bose stars. Would such null results imply that dark matter cannot largely consist of WIMPs or axions, or are we just sitting in an unfortunate (for detection) point in the parameter space? Gravitational probes could help resolve the issue in the following ways:

• If probes of the small-scale power spectrum at early times (via the Lyman-$\alpha$ forest probing $z \sim 2$ or 21-cm fluctuations during reionization at $z \sim 10 - 20$) show indications of a small-scale cut-off in the
power spectrum, and this is confirmed at low redshift by measurements of the subhalo mass function (e.g., with strong lensing or observations of Local Group satellite galaxies), then this would point towards dark matter paradigms with primordial deviations in $M_{\text{halo}}$. This could, for example, help motivate the development of X-ray satellites to look for line emission from sterile neutrinos in nearby galaxies, and the guide further theoretical work on non-trivial dark sectors.

- If the matter power spectrum is consistent with CDM at early times (e.g., the CMB) but there are indications of late-time structure evolution (e.g., a truncation in the halo or subhalo mass function where $M_{\text{halo}}$ is larger today than indicated by the Lyman $\alpha$ forest), this would indicate that the dark matter decays or interacts with other dark-sector particles on cosmological time scales. This would focus theoretical work on models that have evolutionary effects on $M_{\text{halo}}$. If there were no Standard Model particles emerging from dark matter particle interactions or decays, it would imply that the dark sector is complex but almost entirely decoupled from the Standard Model.

- Finally, if all gravitational probes of dark matter indicated that both large-scale and small-scale structure evolve as expected in the dark energy plus CDM cosmological framework, this would indicate that dark matter should be cold. In this case, it is possible that dark matter could consist, say, of heavy (multi-TeV) WIMPs that are simply hard to create at colliders, due to the high energy scale; or detect via direct or indirect detection due to the fact that the dark-matter-particle number density scales inversely with mass.

In any of these three scenarios, astronomical observations play a key role in sharpening the allowed dark matter parameter space, and guiding particle physicists and experimentalists to new directions.

6. Conclusions

As direct proof of physics beyond the Standard Model, dark matter is one of the most significant open questions in particle physics today. We know remarkably little about dark matter, and what little we do know has been discovered by astrophysics via its gravitational imprint on baryonic matter.

These astrophysical measurements are indirectly sensitive to particle physics properties of dark matter, in particular those that modify the initial distribution of dark matter away from the predictions of cold dark matter, or provide mechanisms for energy exchange between dark matter particles at late times. These attributes are largely orthogonal to those accessible to particle physics experiments (e.g., direct detection, indirect detection, and collider production). Therefore, any comprehensive approach to the study of dark matter physics must incorporate results from both fields.

Constructing a joint effort requires a common language and a clear sense of the open questions in both fields. To facilitate this, in this paper we have developed a simplified parameter space which allows theoretical models of dark matter to be characterized by the strength of their particle physics interaction $\Lambda$ and the astrophysical scale at which non-trivial deviations from CDM would appear $M_{\text{halo}}$. Our goal with this parameter space is that it will highlight and enable the advantages of interdisciplinary collaboration in dark matter physics.

Having constructed this common dictionary between particle and astroparticle properties, we consider a number of hints in the existing data that are suggestive of new physics beyond pure CDM. These revolve around discrepancies between observation and simulations at mass scales of $10^9 M_\odot \lesssim M_{\text{vir}} \lesssim 10^{15} M_\odot$. Resolving the origin of these hints requires a more accurate consideration of baryonic effects, and is a necessary step in any effort to use astronomical observations to probe the particle physics of dark matter.

Regardless of the ultimate origin of these particular anomalies, the example of the “Crisis in Small Scale Structure” demonstrates the potential for collaboration between particle physicists and astrophysicists in unraveling the mysteries of dark matter. In Section 5, we outline some opportunities to use existing and near-future astronomical surveys to cover more of the $M_{\text{halo}}$ parameter space. We note that dark matter structures below $M_{\text{vir}} \lesssim 10^{7} M_\odot$ are both largely unconstrained and of particular interest to a large class of theoretical models. New ideas are needed here.
Particle physics searches for dark matter are at a mature stage, and their sensitivity will only increase in the coming years. However, they cannot test all aspects of dark matter phenomenology, and astronomical observations have enormous discovery potential. The development of a comprehensive framework to use astronomical measurements to constrain dark matter theory, which can in turn motivate and guide future studies, will be the challenge to the field of dark matter science.

Acknowledgements

We would like to thank Alyson Brooks, Keith Bechtol, Francis-Yan Cyr-Racine, Alex Drlica-Wagner, Manoj Kaplinghat, Stacy Kim, Leonidas Moustakas, Chanda Prescod-Weinstein, Kris Sigurdson, and David Weinberg for enlightening discussions. This research has made use of the matplotlib python package [665]. NASA’s Astrophysics Data System, and INSPIRE. M.R.B. is supported by DOE grant de-sc0017811. A.H.G.P. is supported by the NSF Grant No. AST-1615838.

It was brought to our attention that our choice of terminology used in Section II in an earlier version of this paper could be marginalizing to people of color. We sincerely apologize for this, and have changed the language in question. We thank those members of our community who raised this issue, and challenged us to do better.

References

[1] F. Zwicky, Die Rotverschiebung von extragalaktischen Nebeln, Helv. Phys. Acta 6 (1933) 110–127.

[2] G. R. Blumenthal, H. Pagels, J. R. Primack, Galaxy formation by dissipationless particles heavier than neutrinos, Nature299 (1982) 37. doi:10.1038/299037a0

[3] C. S. Frenk, S. D. M. White, M. Davis, Nonlinear evolution of large-scale structure in the universe, ApJ271 (1982) 417–430. doi:10.1086/161209

[4] C. J. Hogan, J. J. Dalcanton, New dark matter physics: clues from halo structure, Phys. Rev. D62 (2000) 063511. arXiv:astro-ph/0002330

[5] J. J. Dalcanton, C. J. Hogan, Halo cores and phase space densities: Observational constraints on dark matter physics and structure formation, Astrophys. J. 561 (2001) 35–45. arXiv:astro-ph/0004381 doi:10.1086/323207

[6] D. N. Spergel, P. J. Steinhardt, Observational evidence for self interacting cold dark matter, Phys. Rev. Lett. 84 (2000) 3760–3763. arXiv:astro-ph/9909326 doi:10.1103/PhysRevLett.84.3760

[7] G. Bertone, D. Hooper, J. Silk, Particle dark matter: Evidence, candidates and constraints, Phys. Rept. 405 (2005) 279–390. arXiv:hep-ph/0404175 doi:10.1016/j.physrep.2004.08.031

[8] J. L. Feng, Dark Matter Candidates from Particle Physics and Methods of Detection, Ann. Rev. Astron. Astrophys. 48 (2010) 495–545. arXiv:1003.0904 doi:10.1146/annurev-astro-082708-101659

[9] C. Patrignani, et al., Review of Particle Physics, Chin. Phys. C40 (10) (2016) 100001. doi:10.1088/1674-1137/40/10/100001

[10] Y. Bai, P. J. Fox, R. Harnik, The Tevatron at the Frontier of Dark Matter Direct Detection, JHEP 12 (2010) 048. arXiv:1005.3797 doi:10.1007/JHEP12(2010)048

[11] D. Abercrombie, et al., Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum arXiv:1507.00966

[12] B. Penning, The pursuit of dark matter at colliders?an overview, J. Phys. G45 (6) (2018) 063001. arXiv:1712.01391

[13] M. W. Goodman, E. Witten, Detectability of Certain Dark Matter Candidates, Phys. Rev. D31 (1985) 3059. doi:10.1103/PhysRevD.31.3059

[14] I. Wasserman, Possibility of Detecting Heavy Neutral Fermions in the Galaxy, Phys. Rev. D33 (1986) 2071–2078. doi:10.1103/PhysRevD.33.2071

[15] K. van Bibber, P. M. McIntyre, D. E. Morris, G. G. Raffelt, A Practical Laboratory Detector for Solar Axions, Phys. Rev. D39 (1989) 2089. doi:10.1103/PhysRevD.39.2089

[16] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, K. A. van Bibber, Experimental Searches for the Axion and Axion-Like Particles, Ann. Rev. Nucl. Part. Sci. 65 (2015) 485–514. arXiv:1602.00039 doi:10.1146/annurev-nucl-102014-022120

[17] L. Baudis, Dark matter detection, J. Phys. G43 (4) (2016) 044001. doi:10.1088/0954-3899/43/4/044001

[18] F. W. Stecker, The Cosmic Gamma-Ray Background from the Annihilation of Primordial Stable Neutral Heavy Leptons, Astrophys. J. 223 (1978) 1032–1036. doi:10.1086/156336

[19] J. Silk, K. A. Olive, M. Srednicki, The Photino, the Sun and High-Energy Neutrinos, Phys. Rev. Lett. 55 (1985) 257–259, [283(1985)]. doi:10.1103/PhysRevLett.55.257

[20] T. R. Slatyer, Indirect Detection of Dark Matter, in: Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016, 2018, pp. 297–353. arXiv:1710.05137 doi:10.1142/9789813239348_0005
[116] G. Bertone, D. Cumberbatch, R. Ruiz de Austri, R. Trotta, Dark Matter Searches: The Nightmare Scenario, JCAP 1201 (2012) 004. arXiv:1107.5813, doi:10.1088/1475-7516/2012/01/004.

[117] B. A. Reid, et al., Cosmological Constraints from the Clustering of the Sloan Digital Sky Survey DR7 Luminous Red Galaxies, Mon. Not. Roy. Astron. Soc. 404 (2010) 60–85. arXiv:0907.1659, doi:10.1111/j.1365-2966.2010.16276.x.

[118] V. Iršič, M. Vieš, T. A. M. Berg, V. D’Odorico, M. G. Haehnelt, S. Cristiani, G. Cupani, T.-S. Kim, S. López, S. Ellison, G. D. Becker, L. Christensen, K. D. Denney, G. Worseck, J. S. Bolton, The Lyman α forest power spectrum from the XQ-100 Legacy Survey, MNRAS406 (2010) 3432–3445. arXiv:1002.01761, doi:10.1109/mrsalst.2012.63372.

[119] M. Kaplinghat, S. Tulin, H.-B. Yu, Dark Matter Halos as Particle Colliders: Unified Solution to Small-Scale Structure Puzzles from Dwarfs to Clusters, Phys. Rev. Lett. 116 (4) (2016) 041302. arXiv:1506.03839, doi:10.1103/PhysRevLett.116.041302.

[120] M. Vieš, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, A. Riotto, Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest, Phys. Rev. D71 (2005) 063534. arXiv:astro-ph/0501562, doi:10.1103/PhysRevD.71.063534.

[121] U. Seljak, A. Slosar, P. McDonald, Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints, JCAP 0610 (2006) 014. arXiv:astro-ph/0606335, doi:10.1088/1475-7516/2006/10/014.

[122] P. Jethwa, D. Erkal, V. Belokurov, The upper bound on the lowest mass halo, Mon. Not. Roy. Astron. Soc. 473 (2) (2018) 2060–2083. arXiv:1612.07834, doi:10.1109/mrsalst.2012.63372.

[123] S. Y. Kim, A. H. G. Peter, J. R. Hargis, There is No Missing Satellites Problem arXiv:1711.06267.

[124] P. A. R. Ade, et al., Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. 594 (2016) A13. doi:10.1051/0004-6361/201525830.

[125] J. Bailin, M. Steinmetz, Internal and external alignment of the shapes and angular momenta of lambda-CDM halos, Mon. Not. Roy. Astron. Soc. 359 (2005) 1029–1040. arXiv:astro-ph/0408163, doi:10.3847/0004-637X/754/5/538.

[126] J. Diemand, B. Moore, J. Stadel, Earth-mass dark-matter haloes as the first structures in the early Universe, Nature433 (2005) 389–391. arXiv:astro-ph/0501589, doi:10.1038/nature03270.

[127] V. Springel, C. S. Frenk, S. D. M. White, The large-scale structure of the Universe, Nature440 (2006) 1137–1144.

[128] B. Allgood, R. A. Flores, J. R. Primack, A. V. Kravtsov, R. H. Wechsler, A. Faltenbacher, J. S. Bullock, The shape of dark matter halos: dependence on mass, redshift, radius, and formation, Mon. Not. Roy. Astron. Soc. 367 (2006) 1781–1796. arXiv:astro-ph/0508497, doi:10.1111/j.1365-2966.2005.09944.x.

[129] B. Diemer, A. V. Kravtsov, Dependence of the outer density profiles of halos on their mass accretion rate, Astrophys. J. 824 (1) (2016) 10. arXiv:1604.06462, doi:10.3847/0004-637X/824/1/10.

[130] G. A. Gooley, A. H. G. Peter, T. Yang, B. Willman, B. F. Griffen, A. Frebel, An observer’s guide to the (Local Group) dwarf galaxies: predictions for their own dwarf satellite populations, MNRAS471 (2017) 493–508. arXiv:1610.00708, doi:10.1093/mnras/stx1900.

[131] Y. F. Jing, Y. Suto, Triaxial modeling of halo density profiles with high-resolution N-body simulations, Astrophys. J. 574 (2002) 538. arXiv:astro-ph/0202064, doi:10.1086/341065.

[132] J. Bailin, M. Steinmetz, Internal and external alignment of the shapes and angular momenta of lambda-CDM halos, Astrophys. J. 627 (2005) 647–665. arXiv:astro-ph/0408163, doi:10.1086/430397.

[133] B. Allgood, R. A. Flores, J. R. Primack, A. V. Kravtsov, R. H. Wechsler, A. Faltenbacher, J. S. Bullock, The shape of dark matter halos: dependence on mass, redshift, radius, and formation, Mon. Not. Roy. Astron. Soc. 367 (2006) 1781–1796. arXiv:astro-ph/0508497, doi:10.1111/j.1365-2966.2005.09944.x.

[134] R. M. Reddick, R. H. Wechsler, J. L. Tinker, P. S. Behroozi, The Connection between Galaxies and Dark Matter Structures in the Local Universe, Astrophys. J. 771 (2013) 30. arXiv:1207.2160, doi:10.1088/0004-637X/771/1/30.

[135] P. S. Behroozi, R. H. Wechsler, Y. Lu, O. Hahn, M. T. Busha, A. Klypin, J. R. Primack, Mergers and Mass Accretion for Infalling Halos Both End Well Outside Cluster Virial Radii, Astrophys. J. 787 (2014) 16. arXiv:1310.2239, doi:10.1088/0004-637X/787/2/156.

[136] B. Willman, J. Strader, ‘Galaxy,’ Defined, Astron. J. 144 (2012) 76. arXiv:1203.2608, doi:10.1088/0004-6256/144/3/76.

[137] A. Fitts, et al., FIRE in the Field: Simulating the Threshold of Galaxy Formation, Mon. Not. Roy. Astron. Soc. 471 (3) (2017) 3547–3562. arXiv:1611.02281, doi:10.1093/mnras/stx1757.

[138] S. Garrison-Kimmel, J. S. Bullock, M. Boylan-Kolchin, E. Bardwell, Organized Chaos: Scatter in the relation between stellar mass and halo mass in small galaxies, Mon. Not. Roy. Astron. Soc. 464 (3) (2017) 3108–3120. arXiv:1603.04855, doi:10.1093/mnras/stx2564.

[139] F. Munshi, A. M. Brooks, E. Applebaum, D. R. Weisz, F. Governato, T. R. Quinn, Going, going, gone dark: Quantifying the scatter in the faintest dwarf galaxies arXiv:1705.06286.

[140] R. Errani, J. Peharrubia, M. G. Walker, Systematics in virial mass estimators for pressure-supported systems, ArXiv e-print arXiv:1805.00484.

[141] B. A. Gooley, A. H. G. Peter, J. L. Carlín, A. Frebel, K. Bechtol, B. Willman, The predicted luminous satellite populations around SMC- and LMC-mass galaxies - a missing satellite problem around the LMC?, MNRAS472 (2017) 1060–1073. arXiv:1703.06321, doi:10.1093/mnras/stx2001.
E. Papastergis, E. A. K. Adams, J. M. van der Hulst, An accurate measurement of the baryonic Tully-Fisher relation with heavily gas-dominated ALFALFA galaxies, A&A493 (2016) A39. arXiv:1602.09807 doi:10.1051/0004-6361/201628410

L. V. Sales, J. F. Navarro, K. Omura, A. Fattahi, I. Ferrero, M. Abadi, R. Faw, R. A. Crain, C. S. Frenk, T. Sawala, M. Schaller, J. Schaye, T. Theuns, S. D. M. White, The low-mass end of the baryonic Tully-Fisher relation, MNRAS464 (2017) 2419–2426. arXiv:1602.02155 doi:10.1093/mnras/stx2461

A. Klypin, I. Karachentsev, D. Makarov, O. Nasonova, Abundance of field galaxies, MNRAS454 (2015) 1798–1810. arXiv:1405.4823 doi:10.1093/mnras/stv2040

R. Mandelbaum, U. Seljak, G. Kaulfmann, C. M. Hirata, J. Brinkmann, Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the sdss: stellar mass, luminosity, morphology, and environment dependencies, Mon. Not. Roy. Astron. Soc. 368 (2006) 715. arXiv:astro-ph/0511644 doi:10.1111/j.1365-2966.2006.10156.x

J. I. Read, O. Agertz, M. L. M. Collins, Dark matter cores all the way down, MNRAS459 (2016) 2573–2590. arXiv:1508.04143 doi:10.1093/mnras/stw713

S. Trujillo-Gomez, A. Schneider, E. Papastergis, D. S. Reed, G. Lake, Another baryon miracle? Testing solutions to the missing dwarfs’ problem, ArXiV e-print:arXiv:1610.09335

A. Schneider, S. Trujillo-Gomez, E. Papastergis, D. S. Reed, G. Lake, Hints against the cold and collisionless nature of dark matter from the galaxy velocity function, ArXiV e-print:arXiv:1611.09362

A. Kamada, M. Kaplinghat, A. B. Pace, H.-B. Yu, How the Self-Interacting Dark Matter Model Explains the Diverse Galactic Rotation Curves, ArXiV e-print:arXiv:1611.02716

P.Creasey, O. Samei, V. L. Sales, H.-B. Yu, M. Vogelsberger, J. Zavala, Spreading out and staying sharp - creating diverse rotation curves via baryonic and self-interaction effects, MNRAS468 (2017) 2283–2295. arXiv:1612.03903 doi:10.1093/mnras/stw652

A. M. Brooks, E. Papastergis, C. R. Christensen, F. Governato, A. Stålpp, T. R. Quinn, J. Wadsley, How to Reconcile the Observed Velocity Function of Galaxies with Theory, ArXiV e-print:arXiv:1701.07835

J. Peaňarrubia, A. Pontzen, M. G. Walker, S. E. Koposov, The Coupling between the Core/Cusp and Missing Satellite Problems, ApJL759 (2012) L42. arXiv:1207.2777 doi:10.1088/2041-8205/759/2/L42

Y. Zu, R. Mandelbaum, Mapping stellar content to dark matter haloes using galaxy clustering and galaxy-galaxy lensing in the SDSS DR7, MNRAS454 (2015) 1161–1191. arXiv:1505.02781 doi:10.1093/mnras/stv2062

A. Bonaca, M. Geha, A. H. W. Kuepper, J. Diemand, K. V. Johnston, D. W. Hogg, Milky Way Mass and Potential Recovery Using Tidal Streams in a Realistic Halo, ApJ795 (2014) 94. arXiv:1406.6063 doi:10.1088/0004-637X/795/1/94

A. M. Price-Whelan, D. W. Hogg, K. V. Johnston, D. W. Hogg, Inferring the Gravitational Potential of the Milky Way with a Few Precisely Measured Stars, ApJ794 (2014) 4. arXiv:1405.6721 doi:10.1088/0004-637X/794/1/4

The WFIRST Astrometry Working Group, R. E. Anderson, A. Bellini, S. Casertano, J. R. Lu, P. Melchior, D. Bennett, M. Shao, J. Rhodes, S. Malhotra, S. Gaudi, M. Fall, E. Nelan, P. Guhathakurta, J. Anderson, S. Ho, M. Libralato, Astrometry with the WFIRST Wide-Field Imager, ArXiV e-print:arXiv:1712.05420

M. Vogelsberger, J. Zavala, C. Simpson, A. Jenkins, Dwarf galaxies in CDM and SIDM with baryons: observational probes of the nature of dark matter, Mon. Not. Roy. Astron. Soc. 444 (4) (2014) 3684–3698. arXiv:1405.5216 doi:10.1093/mnras/stu1713

A. B. Fry, F. Governato, A. Pontzen, T. Quinn, M. Tremmel, L. Anderson, H. Menon, A. M. Brooks, J. Wadsley, All about baryons: revisiting SIDM predictions at small halo masses, Mon. Not. Roy. Astron. Soc. 452 (2) (2015) 1468–1479. arXiv:1501.00497 doi:10.1093/mnras/sts1330

M. R. Lovell, V. Gonzalez-Perez, S. Bose, A. Boyarsky, S. Cole, C. S. Frenk, O. Ruchayskiy, Addressing the too big to fail problem with baryon physics and sterile neutrino dark matter, MNRAS468 (2017) 2836–2849. arXiv:1611.00005 doi:10.1093/mnras/stx2621

G. R. Blumenthal, S. M. Faber, R. Flores, J. R. Primack, Contraction of Dark Matter Galactic Halos Due to Baryonic Infall, Astrophys. J. 301 (1986) 27. doi:10.1086/163867

O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, D. Nagai, Response of dark matter halos to condensation of baryons: Cosmological simulations and improved adiabatic contraction model, Astrophys. J. 616 (2004) 16–26. doi:10.1086/383516

F. Governato, L. Mayer, J. Wadsley, J. P. Gardner, B. Willman, E. Hayashi, T. Quinn, J. Stadel, G. Lake, The Formation of a Realistic Disk Galaxy in A-dominated Cosmologies, ApJ607 (2004) 688–696. arXiv:astro-ph/0207044 doi:10.1086/383516

F. Governato, et al., At the heart of the matter: the origin of bulgeless dwarf galaxies and Dark Matter cores, Nature 463 (2010) 203–206. arXiv:0911.2237 doi:10.1038/nature08640

M. R. Krumholz, Notes on Star Formation, ArXiV e-print:arXiv:1511.03457

A. Pontzen, F. Governato, How supernova feedback turns dark matter cusps into cores, Mon. Not. Roy. Astron. Soc. 421 (2012) 3464. arXiv:1106.0499 doi:10.1111/j.1365-2966.2012.20571.x

A. M. Brooks, A. Zolotov, Why Baryons Matter: The Kinematics of Dwarf Spherial Satellites, Astrophys. J. 786 (2014) 87. arXiv:1207.2468 doi:10.1088/0004-637X/786/2/87

A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, J. Wadsley, Baryons Matter: Why Luminous Satellite Galaxies Have Reduced Central Masses, Astrophys. J. 761 (2012) 71. arXiv:1207.0007 doi:10.1088/0004-637X/761/1/71

A. Pontzen, A. Zolotov, Y. Dubois, J. Read, Cusp-core transformations in dwarf galaxies: observational predictions, Mon. Not. Roy. Astron. Soc. 429 (2013) 3068. arXiv:1206.4895 doi:10.1093/mnras/sts563

K. S. Arraki, A. Klypin, S. More, S. Trujillo-Gomez, Effects of baryon removal on the structure of dwarf spheroidal...
A. L. Erickcek, The Dark Matter Annihilation Boost from Low-Temperature Reheating, Phys. Rev. D92 (10) (2015) 103505. [arXiv:1504.03335] doi:10.1103/PhysRevD.92.103505

T. Bringmann, P. Scott, Y. Akrami, Improved constraints on the primordial power spectrum at small scales from ultracompact minihalos, Phys. Rev. D85 (2012) 125027. [arXiv:1110.2484] doi:10.1103/PhysRevD.85.125027

F. Li, A. L. Erickcek, N. M. Law, A new probe of the small-scale primordial power spectrum: astrometric microlensing by ultracompact minihalos, Phys. Rev. D86 (2012) 043519. [arXiv:1202.1284] doi:10.1103/PhysRevD.86.043519

T. Bringmann, P. Scott, Y. Akrami, Improved constraints on the primordial power spectrum at small scales from ultracompact minihalos, Phys. Rev. D85 (2012) 125027. [arXiv:1110.2484] doi:10.1103/PhysRevD.85.125027

H. A. Clark, G. F. Lewis, P. Scott, Investigating dark matter substructure with pulsar timing I. Constraints on ultracompact minihaloes, Mon. Not. Roy. Astron. Soc. 456 (2) (2016) 1394–1401. [Erratum: Mon. Not. Roy. Astron. Soc. 464, no. 2, 2468 (2017)]. [arXiv:1509.02938] doi:10.1093/mnras/stw2582, 10.1093/mnras/stv2743

J. Chluba, J. Hamann, S. P. Patil, Features and New Physical Scales in Primordial Observables: Theory and Observation, Int. J. Mod. Phys. D24 (10) (2015) 1530023. [arXiv:1505.01834] doi:10.1142/S0218271815300232

J. Chluba, Which spectral distortions does ΛCDM actually predict?, Mon. Not. Roy. Astron. Soc. 460 (1) (2016) 227–239. [arXiv:1603.02496] doi:10.1093/mnras/stw945

T. Nakama, J. Chluba, M. Kamionkowski, Shedding light on the small-scale crisis with CMB spectral distortions, Phys. Rev. D95 (12) (2017) 121302. [arXiv:1703.10559] doi:10.1103/PhysRevD.95.121302

G. Cabass, A. Melchiorri, E. Pajer, μ distortions or running: A guaranteed discovery from CMB spectrometry, Phys. Rev. D93 (8) (2016) 083515. [arXiv:1602.05578] doi:10.1103/PhysRevD.93.083515

J. D. Hunter, Matplotlib: A 2D Graphics Environment, Computing in Science and Engineering 9 (2007) 90–95. doi:10.1109/MCSE.2007.55