A model for halo formation with axion mixed dark matter

David J. E. Marsh, Joseph Silk

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **437,** 2652–2663 (2014) doi:10.1093/mnras/stt2079
Advance Access publication 2013 November 26

A model for halo formation with axion mixed dark matter

David J. E. Marsh[1]⋆ and Joseph Silk[2,3,4]

[1]*Perimeter Institute, 31 Caroline St N, Waterloo, ON N2L 6B9, Canada*
[2]*Institut dAstrophysique, UMR 7095 CNRS, Université Pierre et Marie Curie, 98bis Blvd Arago, F-75014 Paris, France*
[3]*Department of Physics and Astronomy, The Johns Hopkins University, Homewood Campus, Baltimore, MD 21218, USA*
[4]*Beecroft Institute of Particle Astrophysics and Cosmology, Department of Physics, University of Oxford, Oxford OX1 3RH, UK*

Accepted 2013 October 23. Received 2013 October 23; in original form 2013 July 22

ABSTRACT

There are several issues to do with dwarf galaxy predictions in the standard Λ cold dark matter (ΛCDM) cosmology that have suscitated much recent debate about the possible modification of the nature of dark matter as providing a solution. We explore a novel solution involving ultralight axions that can potentially resolve the missing satellites problem, the cusp-core problem and the 'too big to fail' problem. We discuss approximations to non-linear structure formation in dark matter models containing a component of ultralight axions across four orders of magnitude in mass, $10^{-24} \lesssim m_a \lesssim 10^{-20}$ eV, a range too heavy to be well constrained by linear cosmological probes such as the cosmic microwave background and matter power spectrum, and too light/non-interacting for other astrophysical or terrestrial axion searches. We find that an axion of mass $m_a \approx 10^{-21}$ eV contributing approximately 85 per cent of the total dark matter can introduce a significant kpc scale core in a typical Milky Way satellite galaxy in sharp contrast to a thermal relic with a transfer function cut off at the same scale, while still allowing such galaxies to form in significant number. Therefore, ultralight axions do not suffer from the *Catch 22* that applies to using a warm dark matter as a solution to the small-scale problems of CDM. Our model simultaneously allows formation of enough high-redshift galaxies to allow reconciliation with observational constraints, and also reduces the maximum circular velocities of massive dwarfs so that baryonic feedback may more plausibly resolve the predicted overproduction of massive Milky Way Galaxy dwarf satellites.

Key words: elementary particles – galaxies: dwarf – galaxies: halo – cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

There are three outstanding problems in the dwarf galaxy astrophysics of the standard Λ cold dark matter (ΛCDM) cosmology. The controversies on small scales may be summarized as (a) the missing satellites problem (MSP), (b) the cusp-core problem (CCP) and (c) the 'too big to fail' problem (which we refer to here as the massive failures problem, MFP), all reviewed in Weinberg et al. (2013).

The MSP and CCP with CDM structure formation can both be solved by introducing a length-scale into the DM. This can be thermal, coming from free streaming of warm dark matter (WDM), or, as we will discuss below, non-thermal, coming from coherent oscillations of a light scalar field. The thermal solution may suffer from a *Catch 22* issue, whereby galaxy formation occurs too late (Macciò et al. 2012c). We show here that the non-thermal solution both avoids this dilemma and also augurs well for a particle-orientated solution of MFP, a problem for which a feedback solution seems questionable (Boylan-Kolchin, Bullock & Kaplinghat 2012; Garrison-Kimmel et al. 2013b), although not all agree (Brooks et al. 2013).

The paper is organized as follows. We introduce ultralight scalar dark matter and compare the linear theory to WDM in Section 2. In Section 3, we compute the halo mass function (HMF) and model a cut-off in it. In Section 4, we discuss the halo-density profile and core formation. In Section 5, we discuss the relevance of our model for MFP, and in Section 6 we discuss implications for high-redshift galaxy formation. The casual reader can skip to Section 7 where we summarize and discuss our main results, and provide a guide to the relevant figures. Appendix A provides details of our two-component density profile model.

2 ULTRALIGHT SCALAR DARK MATTER

A coherently oscillating scalar field, ϕ, in a quadratic potential $V = m_a^2\phi^2/2$, has an energy density that scales as a^{-3} and thus

⋆E-mail: djemarsh@gmail.com

can behave in cosmology as DM (Turner 1983, 1986).[1] The relic density contains a non-thermal component produced by vacuum realignment, which depends on the initial field displacement ϕ_i. The Klein–Gordon equation is

$$\ddot{\phi} + 3H\dot{\phi} + m_a^2\phi = 0\,, \tag{1}$$

where the Hubble rate $H = \dot{a}/a$. When $H \gg m_a$, the field is frozen by Hubble friction and behaves as a contribution to the cosmological constant (which is negligible for sub-Planckian field values). Therefore, in order to contribute to DM we must have $m_a \gtrsim H_0 \sim 10^{-33}$ eV. Once the mass overcomes the Hubble friction at $m_a \approx 3H(a_{\rm osc})$, the field begins to coherently oscillate. The relic density is then an environmental variable set by the initial field displacement: $\Omega_{\rm a} = \Omega_{\rm a}(m_a, \phi_i) \approx (8\pi G/3H_0^2)a_{\rm osc}(m_a)^3 m_a^2\phi_i^2/2$.

As we will discuss below there is a length-scale, which depends on the inverse mass, below which perturbations in the scalar field energy density will not cluster. Therefore, the clustering of light scalar DM is observationally analogous to that of thermal relic dark matter. In the range 10^{-33} eV $\leq m_a \lesssim 10^{-28}$ eV, the clustering scale is analogous to hot (H)DM, for example composed of thermal relic neutrinos of mass $m_\nu \lesssim 1$eV (Amendola & Barbieri 2006; Marsh et al. 2012). In this section, we will discuss how scalar masses in the range 10^{-24}eV $\lesssim m_a \lesssim 10^{-20}$ eV lead to structure formation that is analogous to WDM in an observationally relevant mass range. Related aspects of structure formation for axion/scalar dark matter in this mass range have been studied in, e.g. Hu, Barkana & Gruzinov 2000; Matos & Urena-Lopez 2000; Arbey, Lesgourgues & Salati 2001, 2003; Bernal & Guzman 2006; Lee & Lim 2009; Park, Hwang & Noh 2012.

While the signatures of thermal relics and ultralight scalars are similar in large-scale structure, there are distinct signatures in the adiabatic and isocurvature cosmic microwave background (CMB) spectra (Marsh et al. 2012, 2013), and future measurements of weak lensing tomography can further break degeneracies (Amendola et al. 2012). For the range of axion masses, we consider $a_{\rm osc} \propto (m_a/{\rm eV})^{-1/2}$ and the redshift $z_{\rm osc}$ is in the range $10^5 \lesssim z_{\rm osc} \lesssim 10^7$. If the axion field is coupled to photons, the rolling field from $\phi = \phi_i$ to $\phi = 0$ at $z_{\rm osc}$ can further affect the CMB. For axions with mass $m_a \lesssim 10^{-28}$ eV that roll after recombination this leads to rotation of CMB polarization (see e.g. Komatsu et al. 2009). For heavier axions rolling at $z \sim 10^6$, photon production in primordial magnetic fields may lead to CMB spectral distortions (Mirizzi, Redondo & Sigl 2009). It is necessary to understand the observational signatures of the parameters m_a and $\Omega_{\rm a}$ if we are to make inferences about the nature of the DM from cosmological constraints, and in particular if hints from the CMB and large-scale structure are pointing to a hot, warm or ultralight scalar component to the DM. In the rest of this paper, we will explore in detail structure formation with ultralight scalars, and similarities and differences with thermal DM.

Such ultralight scalars might arise in a string theory context. It is well known that string theory compactified on sufficiently complicated six-dimensional manifolds contains many axion like particles (ALPs) (Witten 1984; Svrcek & Witten 2006). In Arvanitaki et al. (2010), it was pointed out that since the masses of such axions depend exponentially on the areas of the cycles in the compact manifold, one should expect a uniform distribution of axion masses on a logarithmic scale spanning many orders of magnitude. This phenomenon was dubbed the '*String Axiverse*'. Explicit constructions of the axiverse have been made in M-theory (Acharya, Bobkov & Kumar 2010) and Type IIB theory (Cicoli, Goodsell & Ringwald 2012).

The string axiverse has the potential to provide an elegant solution to the MSP, CCP and possibly MFP, by leading us to expect as natural an axion mixed DM (aMDM) model with many axionic components populating hierarchically different mass regimes. Since the relic density produced via vacuum misalignment is environmental, it can be taken as a free parameter to be constrained observationally, although theoretical priors can be considered (e.g. Aguirre & Tegmark 2005; Tegmark et al. 2006). A component of CDM in the aMDM model from the axiverse arises naturally in the form of the Quantum chromodynamics (QCD) axion (Peccei & Quinn 1977; Weinberg 1978; Wise, Georgi & Glashow 1981; Preskill, Wise & Wilczek 1983; Berezhiani, Sakharov & Khlopov 1992). The mass of the QCD axion is fixed by the pion mass and decay constant, and the axion decay constant f_a. For stringy values of $f_a \sim 10^{16}$ GeV, the QCD axion has a mass around 10^{-10}eV. This is light, but not so light that the sound speed (see below) plays a cosmological role. The requirement that the QCD axion remains light enough, barring accidents, coincidences or fine-tuning, to solve the strong *CP* problem is what guarantees the lightness of the other axions, and as such one should always expect some CDM component alongside the ultralight ALPs (ULAs). Axion mixed dark matter with a QCD axion and supersymmetric neutralino is also expected in many models of beyond the standard model particle physics (see, e.g. Bae, Baer & Lessa 2013).

2.1 Transfer functions

The ULA perturbations, $\delta\phi$, do not behave as CDM: they have a non-zero sound speed[2] $c_a^2 = \delta P/\delta\rho$, which is scale-dependent and given by (Hu et al. 2000; Hwang & Noh 2009; Marsh & Ferreira 2010):

$$c_a^2 = \begin{cases} \frac{k^2}{4m_a^2a^2} & \text{if } k \ll 2m_a a, \\ 1 & \text{if } k \gg 2m_a a. \end{cases} \tag{2}$$

One finds that in a cosmology where axions make up a fraction of the DM, $f_{ax} = \Omega_{\rm a}/\Omega_{\rm d}$, that this causes suppression of the matter power spectrum relative to the case where the DM is pure CDM. Suppression occurs for those modes k that entered the horizon when the sound speed was large. The suppression is centred around a scale k_m, which depends on the mass, and takes the power spectrum down to some value S, which depends on $\Omega_{\rm a}$, times its value in the CDM case.

We compute the transfer functions and matter power spectrum in cosmologies containing CDM plus a ULA component using a modified version of the publicly available Boltzmann code CAMB (Lewis 2000; Lewis, Challinor & Lasenby 2000). The modification, which makes use of the fluid treatment of axion perturbations, is described in Marsh et al. (in preparation). The transfer function is defined as

$$T_{\rm ax}(k) = \left(\frac{P_{\rm aMDM}(k)}{P_{\Lambda{\rm CDM}}(k)}\right)^{0.5}. \tag{3}$$

[1] Exponential potentials can also be relevant in the tracking solution Ferreira & Joyce (1997, 1998). Light fields as DM with various potentials have had their background evolution studied in e.g. Matos, Vázquez-González & Magaña (2009).

[2] This is in the cosmological frame, e.g. synchronous or Newtonian gauge, where $\delta\phi \neq 0$.

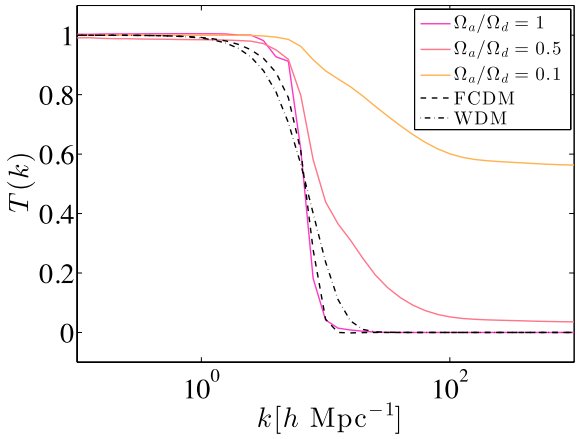

Figure 1. The transfer function, equation (3) for aMDM with $m_a = 10^{-22}$ eV and varying axion fractions to total DM. For comparison, we also plot the FCDM transfer function of Hu et al. (2000) and the WDM transfer function (equation 4) with $m_W \approx 0.84$ keV chosen to match the transfer function half-mode, k_m (equation 5). With this choice and $\Omega_a/\Omega_d = 1$ the axion transfer function at k_m is much steeper than its WDM counterpart.

We compare to the WDM transfer function (in the case where all the DM is warm)

$$T_{\rm WDM}(k) = (1 + (\alpha k)^{2\mu})^{-5/\mu}\,, \tag{4}$$

where $\mu = 1.12$ (Bode, Ostriker & Turok 2001). The mixed C+WDM case is discussed in more detail in e.g. Anderhalden et al. (2013). A well-defined characteristic scale to assign to any such step-like transfer function is the 'half-mode'

$$T(k_m) = 0.5(1 - T(k \to \infty))\,, \tag{5}$$

where $T(k \to \infty) \geq 0$ is the constant plateau value of the transfer function on small scales. This is not the Jeans scale where all structure is suppressed. The Jeans scale is found analytically to be (Hu et al. 2000)

$$k_J = (16\pi G\rho_m)^{1/4} m_a^{1/2}\,. \tag{6}$$

The $\rho^{1/4}$ scaling follows from balancing the growing and oscillating modes in $e^{\Gamma t}$ where $\Gamma^2 = 4\pi G\rho - (k^2/2m)^2$, with $k^2/2m$ coming from the oscillation frequency of the free field.

In Fig. 1, we plot the linear theory aMDM transfer function for a variety of aMDM models, all with $m_a = 10^{-22}$ eV which gives $k_m(10^{-22}\,\mathrm{eV}) = 6.7\,h\mathrm{Mpc}^{-1}$. We compare to equation (4) with $\alpha \approx 0.065\,h^{-1}$ Mpc chosen to give the same k_m. Taking $\Omega_d h^2 = 0.112$, $\Omega_b h^2 = 0.0226$, $h = 0.7$ as our benchmark cosmology, Angulo, Hahn & Abel (2013) gives α in terms of the WDM mass as

$$\alpha = 0.052 \left(\frac{m_W}{\mathrm{keV}}\right)^{-1.15} h^{-1}\,\mathrm{Mpc}\,. \tag{7}$$

Therefore, the matching of half-mode scales gives $m_W \approx 0.83$ keV as equivalent to $m_a = 10^{-22}$ eV.

The logarithmic slope, $\mathrm{d}\ln T(k)/\mathrm{d}\ln k$, evaluated at $k = k_m$ is much steeper for the pure axion model than for the pure WDM model, in agreement with the transfer function of Hu et al. (2000) for 'Fuzzy' (F)CDM, also shown in Fig. 1.

With decreasing fraction of DM in ULAs the slope becomes shallower, and k_m moves out to larger values. The steeper slope for ULAs compared to WDM means that models with the same half-mode will not have $k_J = k_{\rm FS}$ (where $k_{\rm FS}$ is the WDM free-streaming scale, which some authors define differently), and vice versa. Matching Jeans and free-streaming scales, axions will have more power on larger scales relative to WDM; matching the half-mode, axions will have less power on small scales relative to WDM. We choose always to match the transfer function half-mode, since it is well defined for both models.

2.2 Mass scales

We associate characteristic masses to scales k through the mass enclosed within a sphere of radius the half wavelength $\lambda/2 = \pi/k$:

$$M = \frac{4}{3}\pi\left(\frac{\lambda}{2}\right)^3 \rho_0\,, \tag{8}$$

where ρ_0 is the matter density.

In particular, using $k = k_m$ we can expect suppression of the formation of haloes below M_m caused by the decrease in linear power on these scales. We have shown the effects in the transfer function with low axion fraction for illustration, but as we will see in Section 4 the only axion fractions relevant for producing cored density profiles are large, $\Omega_a/\Omega_m \gtrsim 0.85$, and so the characteristic scales will be very close to their values for the pure ULA DM case. Mass scales relevant for halo formation cover axions in the range $10^{-24}\,\mathrm{eV} \lesssim m_a \lesssim 10^{-20}\,\mathrm{eV}$. Axions lighter than this are well probed by the CMB and the linear matter power spectrum (Amendola & Barbieri 2006, Marsh et al., in preparation), while those heavier are probed by supermassive black holes (Arvanitaki & Dubovsky 2011; Pani et al. 2012) and terrestrial experiments (Jaeckel & Ringwald 2010; Ringwald 2012a,b). In Fig. 2, we plot $M_m(m_a)$ for the pure ULA cosmology and find it to be fitted well by a power law $M_m \propto m_a^{-\gamma}$ with $\gamma \approx 1.35$ by least squares over the range of interest. This is very close to the value $\gamma = 4/3$ using the fit of Hu et al. (2000).

In Fig. 2, we also show the Jeans mass, M_J, which is lower by more than two orders of magnitude than M_m. The axion Jeans scale is analogous to the WDM free-streaming scale, where $M_{\rm fs}$ is also some orders of magnitude lower than M_m (Angulo et al. 2013).

Solving equation (4) for the half-mode with WDM and using the fit with $\gamma = 1.35$ to match $M_m(m_W)$ to $M_m(m_a)$, we plot $m_W(m_a)$ in Fig. 3. The power law relating them is $m_W \propto m_a^{0.39}$. Our matching to WDM mass applies to thermal relics like gravitinos, and in Fig. 3 we show the constraint on thermal relics of $m_W > 0.55$ keV from Lyman α forest data reported in Viel et al. (2005). This translates to

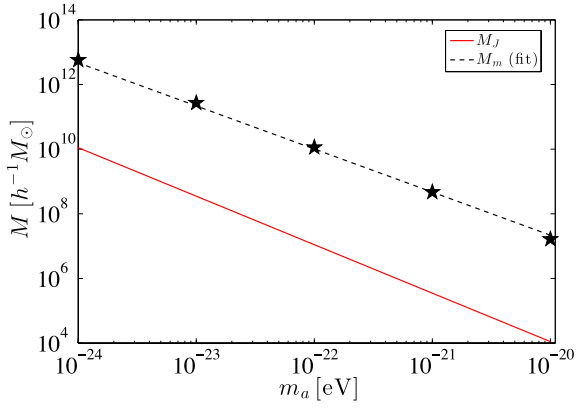

Figure 2. The characteristic mass associated with the half-mode, k_m, of the transfer function as a function of axion mass, $M_m(m_a)$, found from equations (5) and (8). It is well fitted by a power law $M_m \propto m_a^{-\gamma}$ with $\gamma \approx 1.35$. We also show the mass associated with the Jeans scale of equation (6), which is lower by two to three orders of magnitude.

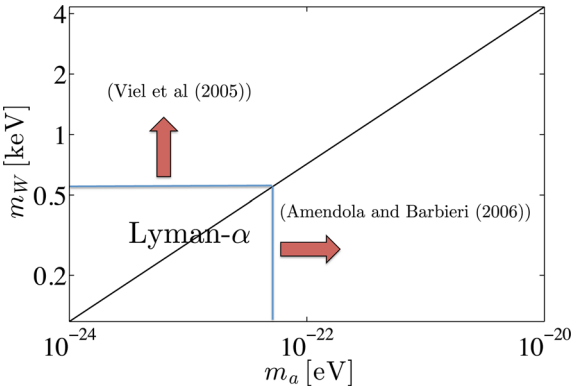

Figure 3. Thermal relic warm dark matter mass in keV chosen to give the same transfer function half-mode, k_m (equation 5), as a ULA, as a function of ULA mass in eV. We also show the Lyman α forest constraints $m_W > 0.55$ keV of Viel et al. (2005) corresponding to $m_a > 5 \times 10^{-23}$ eV, consistent with Amendola & Barbieri (2006).

a constraint on axion mass of $m_a > 5 \times 10^{-23}$ eV, which is consistent with the Lyman α constraints on ULAs reported in Amendola & Barbieri (2006). The more recent Lyman α constraints to WDM, such as Viel et al. (2013) ($m_W \gtrsim 3.3$keV) are much stronger, but there is no corresponding constraint to axions using this data to compare to.

Lyman α constraints are sensitive to the exact shape of the transfer function: since mass goes as radius cubed, small differences between the transfer functions of ULA and WDM models will be amplified to larger differences in the associated mass scales. Lyman α constraints also require careful calibration with hydrodynamical simulations (as done in e.g. Viel et al. 2013). Such simulations are available for CDM and WDM models, but not for ULAs, making the simple comparison of constraints by mass scale perhaps too naive.

The variance of the power spectrum, $\sigma(M)$, is computed by smoothing the power spectrum using a spherical top-hat window function of size R, and is done within CAMB:

$$\sigma(R)^2 = \int_0^\infty \frac{\mathrm{d}k}{k} P(k) W(k|R)^2 \,, \tag{9}$$

$$W(k|R) = \frac{3}{(kR)^3}(\sin kR - kR\cos kR) \,. \tag{10}$$

In Fig. 4, we show the variance associated with the same models as in Fig. 1. The variance for the aMDM models varies little when the fraction is changed between $\Omega_\mathrm{a}/\Omega_\mathrm{d} = 1$ and $\Omega_\mathrm{a}/\Omega_\mathrm{d} = 0.5$, and is comparable to the associated WDM variance.

In the sections that follow, we investigate the suppression of halo formation at and below M_m in axion models in more detail.

3 THE HALO MASS FUNCTION

The MSP arises with CDM due to a larger expected number of low mass haloes than the number of low mass satellites observed in the Local Group (see e.g. Primack 2009, for a review).

To quantify this problem in various models, we adopt the Press–Schechter (PS) approach (Press & Schechter 1974) to compute the abundance of haloes of a given mass: the HMF. In the usual formalism this gives

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M} = -\frac{1}{2}\frac{\rho_0}{M} f(\nu) \frac{\mathrm{d}\ln\sigma^2}{\mathrm{d}\ln M} \,, \tag{11}$$

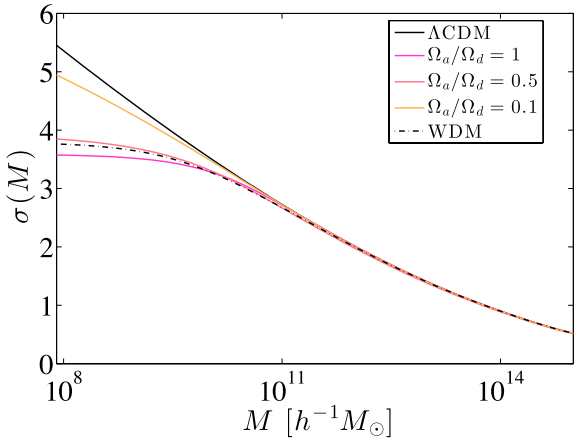

Figure 4. Variance $\sigma(M)$ for ΛCDM, and aMDM with various $\Omega_\mathrm{a}h^2$ at fixed total $\Omega_\mathrm{d}h^2 = 0.112$ and axion mass $m_a = 10^{-22}$ eV.

$$\nu \equiv \frac{\delta_c}{\sigma} \,, \tag{12}$$

where $\mathrm{d}n = n(M)\mathrm{d}M$ is the abundance of haloes within a mass interval $\mathrm{d}M$. For the function $f(\nu)$, we use the model of Sheth & Tormen (1999) (ST):

$$f(\nu) = A\sqrt{\frac{2}{\pi}}\sqrt{q}\nu(1 + (\sqrt{q}\nu)^{-2p})\exp\left[-\frac{q\nu^2}{2}\right] \,, \tag{13}$$

with parameters $\{A = 0.3222, p = 0.3, q = 0.707\}$. The remaining ingredient in this approach is the critical overdensity, δ_c, and what to do on mass scales $M < M_m$, both of which we now discuss.

3.1 Mass-dependent critical density from scale-dependent growth

In the case where all of the DM is made up of ULAs, as we saw in Fig. 1, there is no structure formed below k_J, and so we should expect no peaks in the density field, and thus no haloes, below the mass scale M_J. However, applying the PS formalism described above with a constant barrier δ_c leads to a non-zero mass function for $M < M_\mathrm{J}$. In the case of WDM, this discrepancy is modelled in Smith & Markovic (2011) by the addition of a smooth step in $\mathrm{d}n/\mathrm{dlog}\,M$ at $M = M_\mathrm{fs}$. In the analytical results of Benson et al. (2012), a much sharper cut-off was seen, and was attributed in part to a strong mass dependence in δ_c, which was seen to increase rapidly below M_fs. A shallower cut-off was seen in the numerical results of Angulo et al. (2013). In the recent work of Schneider, Smith & Reed (2013) the cut-off due to free-streaming in WDM was investigated, and also found to be shallower than Benson et al. (2012). Schneider et al. (2013) advocate a sharp k-space window function to match simulations and remove spurious structure thus providing the source of the cut-off: investigating different cut-offs and mass functions in aMDM will be the subject of a future work.

In the absence of numerical simulations for ULA DM, or an existing treatment of the excursion set and spherical collapse in these models, one does not know what form the cut-off in the HMF near M_J should take. In addition, for mixed dark matter models where the small-scale power is not entirely erased but only suppressed, one does not know how much (additional, *ad hoc*) suppression to introduce. In this sub-section, we make a physically motivated argument for a mass-dependent increase in δ_c at low M that should

account in some way for additional suppression in the HMF for $M < M_m$.

Since we use results from CAMB, we take the overdensity δ to evolve with redshift, and in an Einstein-de Sitter (EdS) universe, take the critical overdensity to be fixed, $\delta_c = \delta_{\rm EdS} \approx 1.686$. Alternatively, one can view the overdensity as being fixed, and take δ_c to evolve with redshift as $\delta_c(z) = D_0 \delta_{\rm EdS}/D(z)$ (Percival, Miller & Peacock 2000; Percival 2005), which accounts for the growth between z and $z = 0$. The growth factor $D(z)$ is given by

$$D(z) = \frac{5\Omega_{\rm m}}{2H(z)} \int_0^a \frac{{\rm d}a'}{(a' H(a')/H_0)^3} \,. \tag{14}$$

In the aMDM model, there is scale-dependent growth (see e.g. Acquaviva & Gawiser 2010; Marsh et al. 2012), and we use this to model the change in δ_c with scale. For the relatively heavy axions, we consider here the growth at the pivot scale $k_0 = 0.002\, h{\rm Mpc}^{-1}$ is the same as in ΛCDM, while it is much smaller at $k > k_m$. We take $\delta_c(k)$ at $z = 0$ to be altered by an amount $D(k_0)/D(k)$, and normalize by the same ratio in ΛCDM (to take account of the small amount of scale-dependent growth there). In the interests of simplicity, we will only be concerned with examples of the HMF at $z = 0$ and take $\delta_c(z = 0, k = k_0) = \delta_{\rm EdS}$, which is good to within a few per cent for ΛCDM (Percival 2005).[3] At redshift $z = 0$ our model takes

$$\delta_c(k) = \mathcal{G}(k)\delta_{\rm EdS} \,, \tag{15}$$

$$\mathcal{G}(k) := \frac{D(k)_{\Lambda{\rm CDM}}}{D(k)_{\rm aMDM}} \,. \tag{16}$$

Two, not entirely unrelated, issues arise with this model when trying to extract the growth from a Boltzmann code. The first is that to use this model we must disentangle growth from transfer function, which is by definition somewhat problematic in the case of scale-dependent growth. Defining the transfer function as the piece which depends solely on k this can only be done with the logarithmic derivative ${\rm d}\log\delta/{\rm d}\log a = {\rm d}\log D/\log a$, which does not give us the absolute value at $z = 0$ that we seek. We take a more practical definition suited to numerical computation. In ΛCDM, the transfer function freezes in somewhere around the decoupling epoch (Eisenstein & Hu 1997), when matter domination is total. This provides a definition of the scale-dependent growth at $z = 0$ which is easily accessible from a numerical solution for $\delta(k, z)$, normalized such that $D(k = k_0) = 1$:

$$\frac{D(k)}{D_0} := \frac{\delta(k,0)}{\delta(k,z_h)} \frac{\delta(k_0,z_h)}{\delta(k_0,0)} \,, \tag{17}$$

where z_h is chosen so that in ΛCDM the transfer function has frozen in, and $k_0 \ll k_m$ is the pivot scale. Using CAMB, we find $z_h \approx 300$ works well. We then use exactly the same definition to set the scale of $D(k)$ in the aMDM case.

Scale-dependent growth causes the mass-dependent critical density to increase below $M \approx M_m$. Fig. 5 shows $\mathcal{G}(M)$ for these models. There is the obvious trend that $\mathcal{G}(M)$ decreases with increasing CDM fraction.

The second issue is that if the axions completely dominate the matter density then the overdensity will become vanishingly small for $k \gg k_m$ even at high redshift, and so we are faced with the problem of dividing zero by zero to set the scale of $D(k)$. This is a numerical precision problem and, when combined with Baryon Acoustic Oscillations (BAO) distortions, leads to the spikey/oscillatory behaviour of $\mathcal{G}(M)$ for $\Omega_{\rm c}/\Omega_{\rm d} \lesssim 0.01$ in Fig. 5.

[3] A more advanced treatment of spherical collapse in coupled quintessence cosmologies in Tarrant et al. (2012) found that even in ΛCDM $\delta_c(z = 0)$ can differ from $\delta_{\rm EdS}$ by more than this amount.

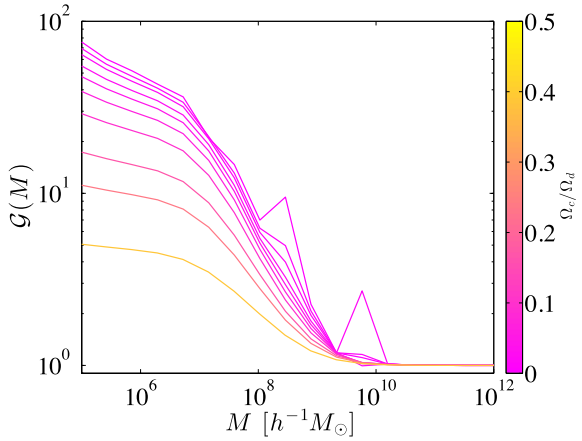

Figure 5. The mass-dependent critical density from scale-dependent growth, equation (16), is given by $\delta_c(M) = \mathcal{G}(M)\delta_{\rm EdS}$. We show $\mathcal{G}(M)$ for aMDM with various $\Omega_{\rm c}/\Omega_{\rm d}$ at fixed total $\Omega_{\rm d}h^2 = 0.112$ and axion mass $m_a = 10^{-22}$ eV. The spikes at low fraction are due to BAO distortions and numerical instability defining scale-dependent growth via a ratio.

3.2 HMF results

In Fig. 6, we plot the HMF with a fixed axion mass of $m_a = 10^{-22}$ eV for a variety of values of the axion density, with fixed total $\Omega_{\rm d}\, h^2 = 0.112$. We see suppression of the mass function beginning at M_m, with the amount of suppression increasing and the asymptotic slope of the mass function decreasing as we raise the axion density. We show results taking δ_c fixed, and those with mass-dependent $\delta_c(M)$, modelled for as above.

The introduction of scale-dependent growth via $\mathcal{G}(M)$ in Fig. 6 causes the HMF to be sharply cut off at around $M \approx 10^8\, h^{-1}\,{\rm M}_\odot \approx 0.01 M_m$ with large axion fraction. This is in agreement with the

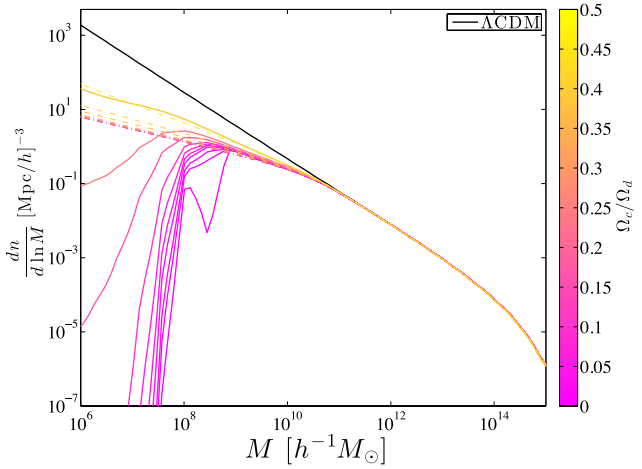

Figure 6. Halo mass function computed directly from CAMB for ΛCDM, and aMDM with various $f_c = \Omega_{\rm c}/\Omega_{\rm d}$ at fixed total $\Omega_{\rm d}\, h^2 = 0.112$ and axion mass $m_a = 10^{-22}$ eV. Solid lines have $\delta_c(M)$ while dashed lines have $\delta_c = \delta_{\rm EdS}$. At $f_c = 0.5$ the cut-off is no longer present in the mass range shown, the difference between $\delta_{\rm EdS}$ and $\delta_c(M)$ having largely vanished. The spikes at low fraction are due to BAO distortions and numerical instability defining scale-dependent growth via a ratio.

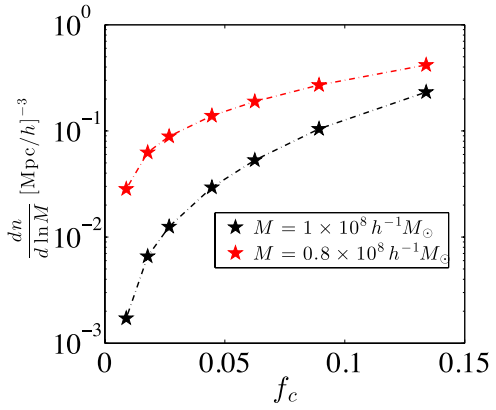

Figure 7. The HMF evaluated at $M = 1 \times 10^8\, h^{-1}\,\mathrm{M}_\odot \approx 0.01 M_m$ and $M = 0.8 \times 10^8\, h^{-1}\,\mathrm{M}_\odot$ for aMDM as a function of $f_c = \Omega_\mathrm{c}/\Omega_\mathrm{d}$ at fixed total $\Omega_\mathrm{d}\, h^2 = 0.112$ and axion mass $m_a = 10^{-22}$ eV. The HMF decreases rapidly below the cut-off, at around 1 per cent of the half-mode mass. Varying the CDM fraction between 1 and 15 per cent can change the value of the HMF at the cut-off by two orders of magnitude.

cut-off of Smith & Markovic (2011) and the numerical results of Angulo et al. (2013) for WDM: the HMF falls below its ΛCDM value at the half-mode mass, but only cuts off completely at a lower mass, intermediate between the Jeans (free-streaming for WDM) mass and the half-mode mass. By considering fragmentation of proto-halo objects formed in WDM cosmologies Angulo et al. (2013) found a smoother cut-off in the HMF than the sharp cut-off of Benson et al. (2012) coming from analytic results. By the time we reach the Jeans scale of $M_\mathrm{J} \approx 1.1 \times 10^7\, h^{-1}\,\mathrm{M}_\odot$ the mass function for $\Omega_\mathrm{a}/\Omega_\mathrm{d} = 1$ is vanishingly small, more than eight orders of magnitude below its ΛCDM value.

In Fig. 7, we plot the HMF evaluated at various masses near $M = 0.01 M_m$ as a function of $f_c = \Omega_\mathrm{c}/\Omega_\mathrm{a}$ at low f_c to investigate the effect of a small admixture of CDM on the value of the mass function at the cut-off. Varying f_c between 1 and 15 per cent can change the value of the HMF near the cut-off by two orders of magnitude. The small admixture of CDM can help an aMDM model form dwarf haloes near the HMF cut-off.

The low values of $f_c \lesssim 0.13$, as we will see in Section 4.3, are those relevant for core formation with aMDM. At larger values of f_c approaching the equally mixed DM $f_c = 0.5$, the sharp cut-off in the HMF has vanished, although it is still significantly reduced compared to ΛCDM. The large admixture of CDM, if the need for cores is foregone, still remains relevant to the MSP and introduces no potentially problematic cut-off in the HMF.

Scale-dependent growth in aMDM induces a cut-off in the HMF similar to the cut-off observed in numerical simulations of WDM. In Angulo et al. (2013), the cut-off in WDM simulations could be fit by introducing non-spherical filtering to compute $\sigma(R)$. By assigning masses to radii differently for WDM compared to CDM after accounting for formation of haloes by fragmentation this cut-off was made less severe. In order to discuss the assignment of masses to haloes in aMDM, we now move on to model the halo density profile and its normalization.

4 HALO DENSITY PROFILE

The *Catch 22* (Maccio' et al. 2012a) of solving the CCP with WDM is that the WDM particle mass required to introduce a core of sufficient size in a dwarf galaxy serves to cut off the HMF at exactly the mass of the dwarf, so that it is never formed. At the same time, WDM allowed by constraints from Large Scale Structure (LSS) does not form cores of relevant (kiloparsec) size. In order to ascertain whether the *Catch 22* applies to aMDM, or indeed to the case of pure axion DM, we must model the expected core size.

We follow Hu et al. (2000) and associate a core size to the Jeans scale within the halo, $r_{J,h}$, below which the density will be assumed constant.[4] The Jeans scale within the halo is related to the linear Jeans scale, r_J, by scaling the energy density in equation (6)

$$r_{J,h} = \left(\frac{\rho_0}{\rho(r_{J,h})}\right)^{1/4} r_J\,. \tag{18}$$

Thus, we can determine the linear Jeans scale (and so the ULA mass) necessary to provide a given core size inside a dwarf halo, if we know the external profile $\rho(r)$. The assumption inherent in equation (18) is that the coherent effects in the scalar field giving rise to the Jeans scale survive in the non-linear regime when mode mixing becomes important and the linear derivation of the sound speed in equation (2) may break down. *N*-body/lattice simulations of the axion field are needed to test this assumption.

Assuming that collapse occurs as in ΛCDM, Hu et al. (2000) computed $\rho_0/\rho(r_{J,h})$ and found that a core of size $r_{J,h} \sim 3.4$kpc is obtained in a dwarf halo of mass $10^{10}\,\mathrm{M}_\odot$ for an axion of mass $m_a = 10^{-22}$ eV. As we have seen, the HMF for such a ULA is only cut-off for $M \lesssim 10^8\, h^{-1}\,\mathrm{M}_\odot$ suggesting that axions do not suffer the *Catch 22* of WDM.

In the following section, we address this is in a more detailed model of aMDM. First, we compute halo parameters with the pure ULA variance, normalize our cored halo profile, and find the relationship between ULA mass and core size in a representative Milky Way satellite. The picture that emerges is qualitatively the same as Hu et al. (2000), but quantitatively different. Secondly, we extend this picture to a two-component profile and ask whether cores can be maintained as a small admixture of CDM is added.

4.1 The NFW Profile

For the external profile, $\rho(r)$, outside of the Jeans scale where the ULA behaves as CDM, we use the universal radial density profile of Navarro, Frenk & White (1997) (hereafter, NFW):

$$\frac{\rho(r)}{\rho_0} = \frac{\delta_\mathrm{char}}{(r/r_s)(1 + r/r_s)^2}\,, \tag{19}$$

where the scale radius $r_\mathrm{s} = r_{200}/c$, with r_{200} the virial radius, c the concentration parameter and δ_char the characteristic density.

The characteristic density is assumed to be proportional to the density of matter in the Universe at the collapse redshift of the halo, z_coll. The definition of $z_\mathrm{coll}(M)$ is fixed for NFW and follows from PS (see also Lacey & Cole 1993) :

$$\mathrm{erfc}\left(\frac{\delta_{c,\mathrm{EdS}}(D(z_\mathrm{coll})^{-1} - 1)}{\sqrt{2(\sigma^2(fM) - \sigma^2(M))}}\right) = \frac{1}{2}\,, \tag{20}$$

The NFW profile is fitted with $f = 0.01$. As above we work in the convention where δ_c is constant but the overdensities themselves

[4] See also Arbey et al. (2001) and Arbey et al. (2003) who studied the effect of scalar DM of mass $m_a \approx 10^{-23}$ eV on galaxy rotation curves in the presence of baryons, and core formation in the Bose condensate. Yet another model of cores is considered by Bernal, Matos & Nunez (2003). None of these models consider the altered cosmology and structure formation.

evolve with linear growth factor $D(z)$. The characteristic density is then

$$\delta_{\rm char} = C\Omega_{\rm m}(1+z_{\rm coll})^3\,, \tag{21}$$

where $C = 3.4\times10^3$ fits the cosmologies of NFW. This constant varies with the power spectrum, and we find for our *Wilkinson Microwave Anisotropy Probe 7* cosmology that

$$C = 1.24\times10^4\,, \tag{22}$$

gives the correct normalization such that M in the definition of $z_{\rm coll}$ matches M_{200} from the virial radius.

The virial radius is defined as the radius at which the average enclosed density is 200 times the mean density, in terms of the halo mass M at redshift $z = 0$ it is given by

$$r_{200}(M,z) = \left(200\frac{4}{3}\pi\right)^{-1/3}\left(\frac{\rho_0}{h^2{\rm kpc}^{-3}\,{\rm M}_\odot}\right)^{-1/3} \times\left(\frac{M_{200}}{h^{-1}\,{\rm M}_\odot}\right)^{1/3} h^{-1}\,{\rm kpc}\,. \tag{23}$$

For the NFW profile, the concentration and scale radius, with the correct choice of C, are defined such that $M = M_{200}$. For the cored profile that we discuss below, the scale radius of the external NFW profile does not have the same relationship with the true virial radius, and we normalize separately for M_{200}.

Finally, the concentration is defined from the characteristic density, $\delta_{\rm char}$, by

$$\delta_{\rm char} = \frac{200}{3}\frac{c^3}{\ln(1+c)-c/(1+c)}\,. \tag{24}$$

The definition of $z_{\rm coll}$ in equation (20), and hence the concentration defined from it will go to zero for a variance that flattens out at low masses, as is the case for aMDM with small CDM fraction. The lower concentration of low mass haloes in comparison to ΛCDM will be relevant for MFP, which we discuss in Section 5. Since $z_{\rm coll}$ is also lower, in Section 6 we discuss the collapsed mass fraction and potential conflicts with observations of high-redshift galaxies.

4.2 Halo jeans scale for pure axion DM

In this sub-section, we consider the core size and normalization of haloes in a pure ULA dark matter model. We assume collapse occurs as in ΛCDM, with $D(z)$, but use the axion variance, $\sigma(M)$.

For definiteness, we consider haloes with the simplest possible cored profile

$$\rho_{\rm core}(M,r) = \theta(r-r_{J,h}(M))\rho_{\rm NFW}(M,r) + \theta(r_{J,h}(M)-r)\rho_{\rm NFW}(M,r_{J,h}(M))\,, \tag{25}$$

where $\theta(x)$ is the Heaviside function, although much of what we say below will apply to any cored profile with core radius $r_c = r_{J,h}$ fixed by equation (18).[5] In particular, the choice of a Heaviside function introduces sharp transitions into the density profile, and as such is only for illustration. The external NFW profile is consistent with what is observed in the WDM simulations of Maccio' et al. (2012a).

In order to find the Jeans scale within a halo we must solve equation (18) for an NFW profile with external profile normalization fixed at M_s (the 'scale mass'), and shape fixed by scale radius $r_s(M_s) = r_{200}(M_s)/c(M_s)$ to find $r_{J,h}(M_s)$. This is *not* the Jeans scale within a halo of mass $M = M_s$: the mass M_s is the mass that an equivalent NFW profile would have. We will discuss normalization of M shortly.

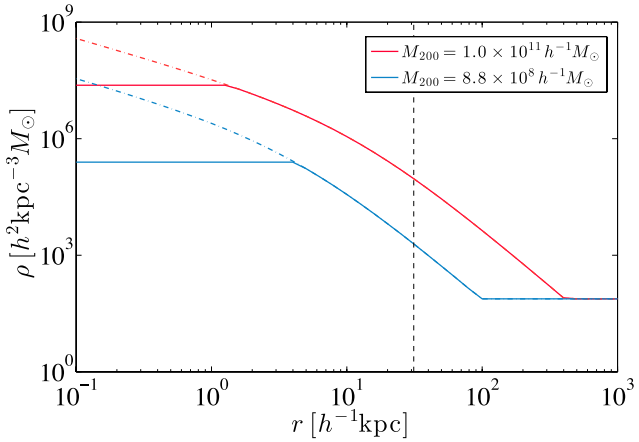

Figure 8. Halo profiles of equation (25) (solid lines) with various M_{200} compared to their parent NFW profiles of mass $M_s \approx M_{200}$ (dot–dashed lines). The axion mass is $m_a = 10^{-22}$ eV and we show the linear Jeans scale, $r_J = 31.2\,h^{-1}$ kpc (vertical dashed line). As long as halo becomes overdense outside of r_J it can continue to be overdense inside until it reaches the halo Jeans scale, $r_{J,h}$ satisfying equation (18). More massive haloes are more dense at r_J and the halo Jeans scale is smaller. No profiles form with $M_s < M_{\rm low}$ when the NFW parent has not become overdense outside of r_J (although they may form by fragmentation).

We use the variance for the axion model to compute our NFW halo parameters: using the ΛCDM variance, the halo Jeans scales with fixed M_s will be different. With low M_s relative to M_m the concentration is lower when the correct variance is used, which causes the Jeans scale to be smaller by the increase in scale radius relative to r_{200}. On the other hand, the Jeans scale within intermediate and high-mass objects is found to be larger with the correct variance. For example, with $m_a = 10^{-21}$ eV the shift in halo Jeans scale inside dwarf galaxies can be of the order of $0.1\,h^{-1}$ kpc.

The Jeans scale decreases in higher density environments and therefore the positive real solution of equation (18) is a monotonically decreasing function of M_s. At low enough M_s, then, one finds $r_{J,h} > r_J$. This cannot be a physical solution. Solutions to equation (18) giving $r_{J,h} > r_J$ can only occur when $\rho/\rho_0 < 1$, which represents a void and not a halo.[6] This break at $r_J = r_{J,h}$ will occur at a certain mass, $M_{\rm low}(M, r_J)$, which is a function of the linear Jeans scale. It is found by solving

$$\rho_{\rm NFW}(M_{\rm low}, r_J) = \rho_0\,. \tag{26}$$

This tells us that no cored haloes with normalization $M_s < M_{\rm low}$ exist for fixed r_J. Cored halo profiles of various masses are shown in Fig. 8 and compared to their parent NFW profiles, with the linear Jeans scale shown for scale. It is clear that no haloes should form that do not already become overdense outside r_J. In the example, with $r_J = 31.2\,h^{-1}$ kpc, this implies that haloes with $M_s = 4\times10^8\,h^{-1}\,{\rm M}_\odot$ only just form.

[5] Other cored profiles are studied in e.g. Zavala, Vogelsberger & Walker (2013) for self-interacting DM.

[6] Solutions to equation (18) with $r_{J,h} > r_J$ also reminds us of another overlooked feature of pure ULA dark matter: not only are haloes cored, having a maximum density, but voids will also be 'cored', having a maximum underdensity. This suggests other possible probes/virtues of ULAs (Peebles & Nusser 2010), but we will not consider this possibility further here.

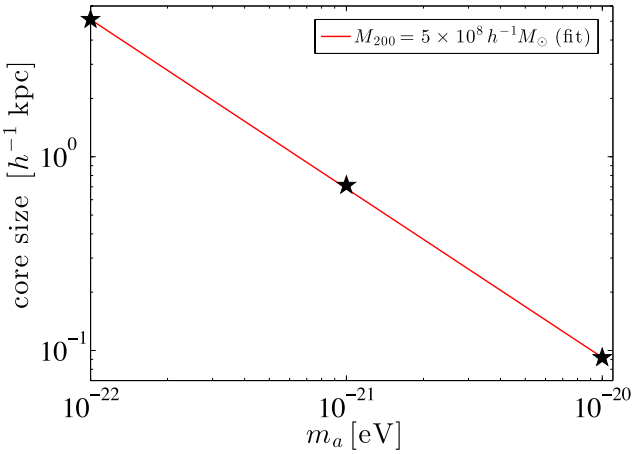

Figure 9. The expected core size in a typical Milky Way satellite of mass $M_{200} = 5 \times 10^8\, h^{-1}\,\mathrm{M}_\odot$ as a function of axion mass, m_a. The relationship is well fitted by a simple power law, shown as $m_a^{-0.87}$, but due to the dependence of the concentration on m_a this is not the scaling of r_J with m_a. The core size is fairly significant, $\gtrsim 0.1\, h^{-1}$ kpc, across the entire mass range, which demonstrates that axions satisfying all large-scale structure constraints can provide a potentially viable resolution of the cusp-core problem in CDM haloes.

Given that no haloes form with external scale $M_s < M_{\rm low}$, what is the minimum halo mass in this model? A halo is normalized to mass M_{200} by the integral out to the virial radius, r_{200}

$$4\pi\rho_0 \int_0^{r_{200}} r^2 \rho_{\rm core}(M, r)\, {\rm d}r = M_{200} \equiv 200\rho_0 \frac{4}{3}\pi r_{200}^3 . \quad (27)$$

This defines $M_{200}(M_s)$. Since the Jeans scale within haloes is a monotonically decreasing function of M_s, at some scale mass $M_{\rm non\text{-}vir}$ the solution for r_{200} will occur when $\rho_{\rm NFW}(M_{\rm non\text{-}vir}, r_{J,h}(M_{\rm non\text{-}vir})) = 200\rho_0$. Haloes with scale mass $M_s < M_{\rm non\text{-}vir}$ will not be virialized in the sense that their average density never exceeds 200 times the background density. In order to assign mass to these haloes one cannot use M_{200}. The total enclosed mass is found by integrating out to $\rho(r_1) = \rho_0$. With this alternative definition of mass it is clear that the lowest mass object formed at $M_s = M_{\rm low}$ is at exactly the Jeans mass, $M_{\rm J}$.

Haloes with $M < M_{200}(M_{\rm non\text{-}vir})$ will also need to have masses assigned to radii coming from the linear filtering in $\sigma(M|R)$ in the HMF in a different manner than is applicable to CDM. This is accounted for by fits to simulations in Angulo et al. (2013) and is partly responsible for the less severe cut-off in the HMF found in that work. By comparison, if we made the same assignment in the HMF with pure axion DM, the cut-off found by introducing scale-dependent growth might also become less severe.

We typically find that, for $M_s > M_{\rm non\text{-}vir}$, r_{200} for the cored profile is approximately the same as for the parent NFW, and that $M_{200} \approx M_s$. Approaching $M_{\rm non\text{-}vir}$, M_{200} drops below M_s.

Finally, having normalized our haloes, we show in Fig. 9 the core size expected in the typical Milky Way dwarf galaxy of mass $M_{200} = 5 \times 10^8\, h^{-1}\,\mathrm{M}_\odot$ as a function of axion mass for the heavier axions $m_a \geq 10^{-22}$ eV allowed by the relevant Lyman α constraints. The core size is well fitted by a single power law in axion mass. The best-fitting power law has $r_{J,h} \sim m_a^{-0.87}$, while the power law given in Hu et al. (2000) is $r_{J,h} \sim m_a^{-2/3}$. Hu et al. (2000) used approximate formulae to solve equation (18) and do not give the dependence of the concentration on axion mass. The dependence of the concentration on axions mass comes from computing NFW parameters with the correct variance and leads to the different power law.

We find that there is a considerable core size for all masses considered. In particular, even our heaviest axion with $m_a = 10^{-20}$ eV has a core size of $r_{J,h} = 0.1\, h^{-1}$ kpc. This heaviest axion has a characteristic mass scale of $M_m = 10^8\, h^{-1}\,\mathrm{M}_\odot$ and would not affect the formation rate of these dwarf galaxies in any dramatic way. This demonstrates that pure ULA dark matter does not suffer from the *Catch 22* of WDM: ULAs allowed by even the most stringent large-scale structure constraints, which would barely affect the HMF at Milky Way satellite masses, can still give significant cores to dwarf galaxies.

4.3 A mixed dark matter halo profile

Having understood the simple cored profile in pure axion DM, we now move on to consider the two-component dark matter subclass of aMDM (see Medvedev 2013 and references therein for recent work on two-component DM haloes). In this sub-section, we continue to assume ΛCDM growth. We study axion mass $m_a = 10^{-22}$ eV and a benchmark halo with $M_{200} = 5 \times 10^8\, h^{-1}\,\mathrm{M}_\odot$.

Again, we take the halo density profile to be NFW outside of the halo Jeans scale, while inside the halo Jeans scale we take the axion component to be smooth. Inside this smooth background we then assume the CDM component of the DM collapses as usual and forms its own NFW halo. Because the axion component is smooth inside this radius, we superpose the profiles and therefore below the Jeans scale the ratio of axions to CDM is not constant but decreases at small radius. This is in agreement with the simulations of mixed cold plus warm DM in Anderhalden et al. (2013). The assumed two-component profile is given in Appendix A.

The size of the cored region depends on the fraction of DM that is cold: $f_c = \Omega_{\rm c}/\Omega_{\rm d}$. For $r < r_{\rm core}$ (see equation A3) there is a cusp while for $r_{\rm core}(f_c, M_s) < r < r_{J,h}(M_s)$ there is a core. We plot the halo profile for $f_c = 0.13$ in our benchmark halo in Fig. 10, while we plot $r_{\rm core}(f_c)$ for our benchmark halo in Fig. 11. We judge the

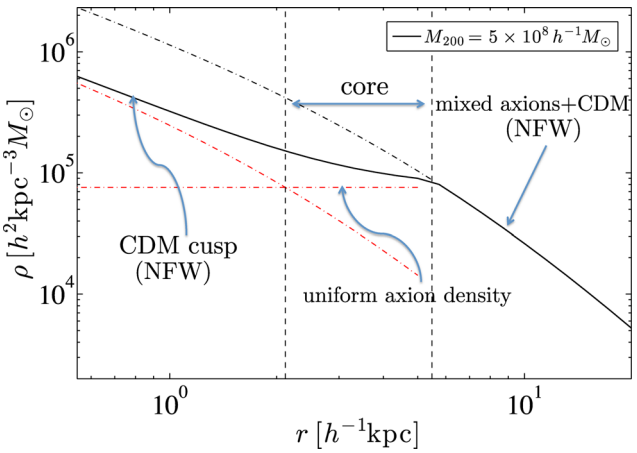

Figure 10. The mixed dark matter halo profile equation (A1) with 13 per cent CDM, $f_c = 0.13$. The outer region is an NFW profile of mixed axions and CDM. The halo mass is $M_{200} = 5 \times 10^8 h^{-1}\,\mathrm{M}_\odot$ and the halo Jeans scale $r_{J,h} = 5.4\, h^{-1}$ kpc (outer vertical dashed line) corresponds to axion mass $m_a = 10^{-22}$ eV with this particular DM composition. At the halo Jeans scale the axions stop clustering and form a uniform component, while the CDM forms an NFW cusp. There is a core down to $r_{\rm core} = 2.1\, h^{-1}$ kpc where the cusp takes over from the uniform piece (inner vertical dashed line).

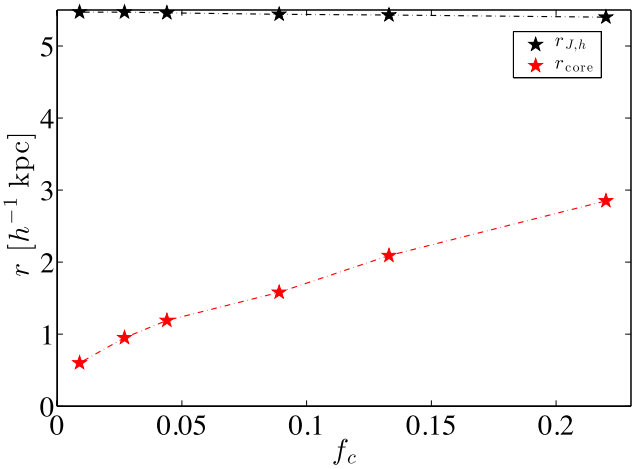

Figure 11. Core radius and halo Jeans scale as a function of CDM fraction, $f_c = \Omega_{\rm c}/\Omega_{\rm d}$, in the two-component halo (equation A1). There is a core for $r_{\rm core} < r < r_{J,h}$, while there is a cusp for $r < r_{\rm core}$, so that increasing f_c makes halo profiles more cuspy, as expected. The halo mass $M_{200} = 5 \times 10^8 h^{-1}\,{\rm M}_\odot$ and the axion mass $m_a = 10^{-22}$ eV.

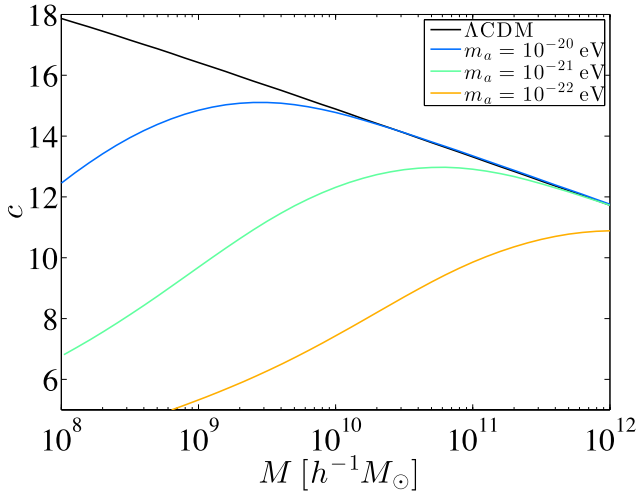

Figure 12. Halo concentration parameter, $c(M)$ for ΛCDM and various axion masses. The flattening of the variance near M_m causes a later formation time for low mass haloes, and due to the presence of $f = 0.01$ in equation (20) leads to a lower concentration for haloes below fM_m. The lowered concentration can help alleviate the too big to fail' problem.

core to be of significant size if it persists down to <50 per cent of the halo Jeans scale. With $f_c = 0.13$ the benchmark axion mass and halo corresponds to a core in the range $2.1\,h^{-1}$ kpc $\lesssim r \lesssim 5.4\,h^{-1}$ kpc.

Although this benchmark core is significant, it is actually not present on sub-kiloparsec scales, which suggests that while introducing a fraction of CDM with an axion of mass 10^{-22} eV may raise the HMF for low-mass dwarf galaxies to acceptable levels, it may not provide a totally adequate solution to the CCP. As we will see below, a more preferable solution to all the problems outlined may be given instead by a higher axion mass.

5 TOO BIG TO FAIL?

The so-called 'too big to fail' problem (here 'MFP') was introduced by Boylan-Kolchin, Bullock & Kaplinghat (2011). In ΛCDM, there are predicted to be massive sub-haloes of the Milky Way of high concentration and circular velocity that cannot host bright satellites, and are not observed. One astrophysical solution to this problem is feedback (Garrison-Kimmel et al. 2013a). WDM (Lovell et al. 2012) and C+WDM (Maccio' et al. 2012b; Medvedev 2013) are also known to help this problem, since the flattened variance leads to later formation times and lower concentration for these most massive sub-haloes.

Since the variance, $\sigma(M)$ in aMDM also flattens at low masses, just like WDM it will lead to a lower concentration for Milky Way sub-haloes compared to ΛCDM if the sub-halo mass is lower than fM_m (see equation 20). In Fig. 12, we plot $c(M_{200})$ for three representative axion masses, $m_a = 10^{-22}, 10^{-21}\,10^{-20}$ eV and compare to ΛCDM, confirming that this is the case.

MFP can also be expressed as the non-observation of large numbers of satellites with maximum circular velocity $v_{\rm max} \gtrsim 40\,{\rm km\,s^{-1}}$. The circular velocity as a function of r is given by

$$v(r) = (GM(<r)/r)^{1/2}\,. \tag{28}$$

In Fig. 13, we plot $v(r)$ for a halo of mass $M_{200} = 2 \times 10^{10}\,h^{-1}\,{\rm M}_\odot$. In ΛCDM this halo has $v_{\rm max} > 40\,{\rm km\,s^{-1}}$. We compare this to $v(r)$ where the DM is made up of an axion of mass $m_a = 10^{-22}$ eV, and to WDM. The axion and WDM haloes are chosen to have the same $r_{\rm max}$ as ΛCDM, having $M_{200} = 10^{10}\,h^{-1}\,{\rm M}_\odot$. Both the axion and

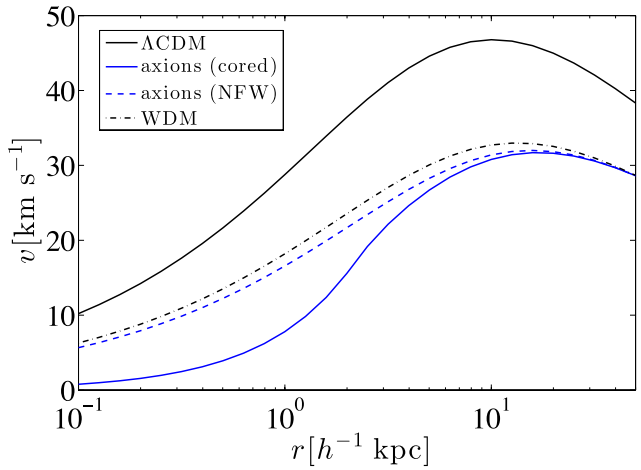

Figure 13. The circular velocity profile for a one component DM halo in various models, derived from the NFW profile. For ΛCDM we have chosen a halo with $M_{200} = 2 \times 10^{10}\,h^{-1}\,{\rm M}_\odot$ which has a maximum circular velocity $v_{\rm max} > 40\,{\rm km\,s^{-1}}$, demonstrating that CDM suffers from the 'too big to fail' problem. We compare to axion and WDM models with $M_{200} = 10^{10}\,h^{-1}\,{\rm M}_\odot$ chosen such that they have the same $r_{\rm max}$ as ΛCDM. Both axions and WDM suppress $v_{\rm max}$ by a factor of about 1.5 relative to ΛCDM, demonstrating their relevance for the solution to MFP. The core in the axion density profile does not affect the suppression of $v_{\rm max}$. We choose axion mass $m_a = 10^{-22}$ eV and equivalent WDM mass of $m_W \approx 0.84$ keV.

WDM models reduce the maximum circular velocity considerably: $v_{\rm max} \approx 30\,{\rm km\,s^{-1}}$.

For WDM we do not model the effect of a possible core, and so for fair comparison we show the axion model with both a cored and NFW profile. The effect of the core in the axion model is small, so that the assumption of core formation does not affect an axion solution to MFP.

6 HIGH-REDSHIFT OBJECTS

We now move on to discuss the collapse redshift of objects to see whether ULAs can accommodate observations of high-redshift

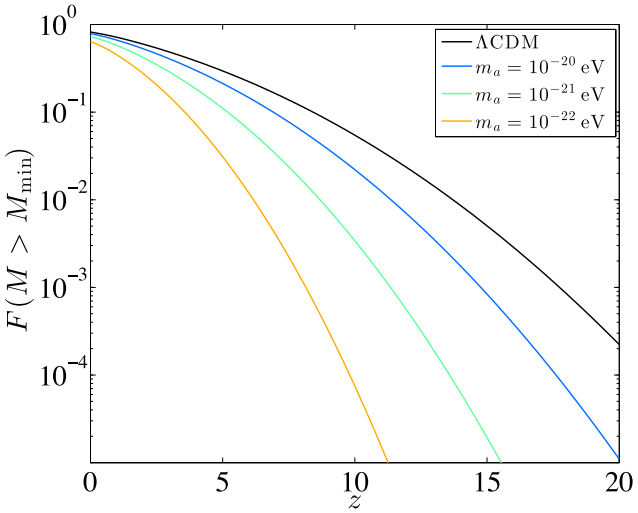

Figure 14. Total collapsed mass fraction, equation (29), with $M_{\rm min} = 10^6\,h^{-1}\,{\rm M}_\odot$ with varying axion mass. We ignore scale-dependent growth and take the axions to make up all of the DM.

galaxies. We consider collapse with ΛCDM growth given by $D(z)$, ignoring for the moment the scale-dependent growth, which will only serve to amplify effects below the characteristic mass. All the effects of axions therefore come in the variance, $\sigma(M)$. For the large axion fractions relevant for core formation in the two-component halo of Section 4.3 the effects of CDM on the variance are virtually negligible, so in our examples we take $\Omega_{\rm a}/\Omega_{\rm d} = 1$ and investigate only the effects of varying axion mass.

The total fraction of objects collapsed with $M > M_{\rm min}$ at redshift z is

$$F(M > M_{\rm min}, z) = {\rm erfc}\left(\frac{\delta_{c,{\rm EdS}}\,D(z_{\rm coll})^{-1}}{\sqrt{2}\sigma(M_{\rm min})}\right). \quad (29)$$

We plot this assuming ΛCDM growth for $M_{\rm min} = 10^6\,h^{-1}\,{\rm M}_\odot$ in Fig. 14. For our benchmark mass $m_a = 10^{-22}$ eV the collapsed mass fraction $F(z \gtrsim 6) \lesssim 0.01$ putting such a light axion in considerable tension with observations of high-redshift galaxies.

Using the scale-dependent growth to assign a mass-dependent critical density for collapse as in equation (16) provided a good working model for the cut-off in the HMF; however, we cannot naively apply such a prescription for δ_c into equation (29): such a cut-off would make the collapsed mass fraction non-monotonic. The ansatz of equation (16) applies only in the HMF and takes the HMF as fundamental, so the analogue of equation (29) is properly defined in this framework as the integral of the HMF. The redshift dependence of such an integral requires us to know the full function $D(k, z)$. Alternatively one could make a fit for the HMF with the ansatz that the cut-off remains always a fixed geometric distance between M_m and $M_{\rm J}$ as they evolve with redshift. We do not explore this effect of the cut-off on the collapsed mass fraction as a function of redshift, but the qualitative effects are obvious: scale-dependent growth should amplify the effects we have seen already in Fig. 14.

Knowing the HMF at $z = 10$ allowed the authors of Pacucci, Mesinger & Haiman (2013) to place constraints on a WDM thermal relic of $m_W > 0.9$ keV using high redshift observations. Using our mass scale conversions of Section 2.2 we might expect such observations to constrain $m_a \gtrsim$ few $\times\, 10^{-22}$ eV, as the simple argument based on collapsed mass fraction with ΛCDM growth given above anticipates.

Suppression of galaxy formation at high redshift has been invoked as a possible solution to another problem of structure formation in ΛCDM: the discrepancy in the evolution of the stellar mass function between observations and models, highlighted in Weinmann et al. (2012). ULAs were invoked, along with WDM, by these authors as a solution. Again, as with Lyman α constraints, access to hydrodynamical simulations with WDM allowed for a detailed comparison of models to observations, and showed a WDM solution to most likely be unviable. Do ULAs remain a viable solution? We have shown in previous sections that there are enough differences between ULAs and WDM that this is possible, but without simulations one cannot quantify this. However, in this section we have confirmed the suppression of halo formation at high redshift necessary for ULAs to be an interesting candidate for further study in this regard.

7 SUMMARY AND DISCUSSION

By studying large-scale structure, we have probed ULAs with masses in the range $10^{-24} \leq m_a \leq 10^{-20}$ eV. Across a fair portion of this range, such ultralight fields can evade large-scale structure constraints while still being different enough from standard CDM on scales relevant to three main problems of structure formation: the MSP, the CCP, and the MFP. We have primarily studied a benchmark ULA of mass $m_a = 10^{-22}$ eV and shown that it is able to solve the MSP, CCP and MFP, avoiding the so-called *Catch 22* of a WDM solution. If this axion constitutes all of the DM, however, then it may come into tension with observations of the Lyman α flux power spectrum (which constrains WDM at masses $m_W \gtrsim 3.3$ keV), high-redshift galaxies (at $z \gtrsim 6$) and the existence of very low mass dwarf galaxies ($M \lesssim 5 \times 10^7\,h^{-1}\,{\rm M}_\odot$). These tensions can be relieved in two ways: by introducing a fraction of CDM or increasing the axion mass. Introducing a fraction of CDM retains adequate solutions to all problems, but cores may yield to cusps at unacceptably large radii. We advocate a higher mass axion of $m_a \gtrsim 10^{-21}$ eV as potentially the best solution.

If all the DM is constituted of an axion or other light scalar of mass $m_a = 10^{-22}$ eV then in the linear power spectrum structure formation is suppressed below some characteristic scale (Fig. 1). The half-mode for this suppression is the same as the half-mode for a WDM particle of mass $m_W \approx 0.84$ keV, just on the edge of the bounds coming from Lyman α forest flux power spectrum constraints (Fig. 3). By considering scale-dependent growth it has been shown that such an axion will cut off the HMF for $M \lesssim 10^8\,h^{-1}\,{\rm M}_\odot$. Introducing a fraction of CDM, $f_c = \Omega_{\rm c}/\Omega_{\rm d}$, the cut-off is made less severe, disappearing completely and leaving only a small suppression to dwarf galaxy formation when $f_c \approx 0.5$ (Fig. 6). This mixed dark matter model, aMDM, may therefore be relevant to the MSP. Such a mix of DM may be natural given certain theoretical priors (Aguirre & Tegmark 2005; Tegmark et al. 2006; Bousso & Hall 2013), or in non-thermal cosmologies expected after moduli stabilization (e.g. Acharya et al. 2010; Acharya, Kane & Kuflik 2012).

If the axion linear Jeans scale can be considered to scale into non-linear environments as the fourth root of the relative density contrast (equations 6, 18), requiring coherence of field oscillations to be maintained, then ultralight axions can give cores to dwarf galaxy density profiles (Hu et al. 2000). By considering a simple cored profile given by an external NFW profile outside the Jeans scale, it was shown that in such a scenario no haloes are formed below the linear Jeans mass, $M_{\rm J}$. Axions in the mass range 10^{-22} eV $\leq m_a \leq 10^{-20}$ eV can give kiloparsec scale cores to dwarf

galaxies of mass $M = 5 \times 10^8\, h^{-1}\, \mathrm{M}_\odot$, and are thus relevant to the CCP of CDM halo density profiles (Figs 8, 9). Since the HMF is only cut off below the dwarf mass, and axions in this mass range are allowed by large-scale structure constraints, we can conclude that in this simple model ultralight axions, or other FCDM candidates, do not suffer from the *Catch 22* that might affect WDM (Maccio' et al. 2012a).

For an axion mass at the low end of the range allowed by large-scale structure constraints, $m_a \approx 10^{-22}$ eV, to form lighter dwarf galaxies of $M \lesssim 5 \times 10^7\, h^{-1}\, \mathrm{M}_\odot$, however, the mixed aMDM is necessary. By considering a two-component density profile below the axion Jeans scale it was shown that an admixture of $f_c \approx 13$ per cent will significantly increase the mass function for light dwarves (Fig. 7), while still allowing for a core on scales greater than a kiloparsec (Figs 10 and 11), although such a core may in fact be too large.

The flattened variance in aMDM (Fig. 4) in the NFW formalism leads to later formation times and consequently lower concentrations for low mass haloes compared to CDM (Fig. 12). This, combined with maximum circular velocity remaining low, $v_{\mathrm{max}} < 40\,\mathrm{km s}^{-1}$, in typical dwarves (Fig. 13), also suggests that ULAs may, just like WDM, play a role in the resolution of the MFP (Boylan-Kolchin et al. 2011).

While avoiding the *Catch 22* in an axion cusp-core and missing satellites resolution an axion mass as low as $m_a = 10^{-22}$ eV comes into tension with high-redshift observations since the collapsed mass fraction becomes very small at $z \gtrsim 6$ (Fig. 14). A heavier axion of mass $m_a \gtrsim 10^{-21}$ eV would be in less tension with observations of high-redshift galaxies (and more recent Lyman α forest constraints to WDM) and could still introduce a kiloparsec scale core to dwarf galaxies and significantly lower the concentration of these galaxies. The formation of dwarves would still be significant, yet also reduced relative to ΛCDM, so that such heavier axions remain relevant to the cusp-core, missing satellites, and 'too big to fail' problems of CDM. Suppression of galaxy formation at high redshift relative to ΛCDM may also be a factor in resolving conflicts between models and observations of the stellar mass function (Weinmann et al. 2012).

With large axion fractions, Marsh et al. (2013) showed that isocurvature constraints imply such a model would be falsified by any detection of tensor modes at the percent level in the CMB by Planck. Axions lighter than those we study are well constrained as components of the DM by observations of the CMB and the linear matter power spectrum (Amendola & Barbieri 2006, Marsh et al., in preparation), while heavier axions are probed, and in some cases ruled, out by the spins of supermassive black holes (Arvanitaki & Dubovsky 2011; Pani et al. 2012), and terrestrial axion searches (Jaeckel & Ringwald 2010; Ringwald 2012a).

We have not discussed the role of baryons in this model, the knowledge of this role being incomplete in even the most state of the art simulations. We have preferred to focus on simple and idealized DM only models where the relevant physics is well understood. The baryonic disc has only a modest effect on the rotation curve, with DM haloes still necessary. Adiabatic contraction of baryons may lead to the enhancement of DM cusps, and thus more need for a core-forming component (Zemp et al. 2012). On the other hand, baryonic feedback is a process driven by supernova explosions driving outflows of gas, which can remove DM cusps in dwarf galaxies while leaving a thick stellar disc (Governato et al. 2012). Baryons may also transfer angular momentum to the halo and modestly effect the spin-up of a massive halo. In massive galaxies, it is unlikely that baryons have a significant effect on the DM halo profile. None of these effects, most notably feedback, solve the excess baryon problems of the MSP and MFP.

It will be important to investigate this model further in the future with numerical *N*-body and other non-linear studies in order to verify whether our simple predictions stand up to detailed scrutiny both theoretically and observationally. It is possible that non-linear effects such as oscillons (see e.g. Gleiser & Sicilia 2009) may play a role. Also, at non-linear order additional terms in the effective fluid description of the axion will be generated, such as anisotropic stresses Hertzberg (2012), which could alter the simple picture of structure formation with a sound speed and Jeans scale dominating effects at short distances.

Furthermore, Lyman α constraints play a key role in determining the validity of WDM models to resolve small-scale crises in CDM, the constraints of Viel et al. (2013) appearing to all but rule out WDM in this regard. Applying such constraints reliably to ULAs will require developing hydrodynamical simulations with them. Finally, a thorough development of the halo model with ULAs building on the groundwork laid here (as was done for WDM by Smith & Markovic 2011) will be invaluable in understanding weak lensing constraints to ULAs obtainable with future surveys (Amendola et al. 2012; Marsh et al. 2012).

ACKNOWLEDGEMENTS

DJEM acknowledges the hospitality of the BIPAC at Oxford University, where this work was initiated, and of Institute d'Astrophysique de Paris, where it was completed. DJEM also acknowledges useful discussions with Andrew Pontzen, Daniel Grin, Scott Tremaine, Enrico Pajer, Jesus Zavala Franco, Simeon Bird and Tom Abel, and an email exchange with Robert Smith. The authors thank Renée Hlozek for comments on the manuscript, and the anonymous referee for helpful comments and reference suggestions. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. The research of JS has been supported at IAP by the ERC project 267117 (DARK) hosted by Université Pierre et Marie Curie - Paris 6 and at JHU by NSF grant OIA-1124403.

REFERENCES

Acharya B. S., Bobkov K., Kumar P., 2010, J. High Energy Phys., 2010, #105
Acharya B. S., Kane G., Kuflik E., 2012, Int. J. Mod. Phys. A, 27, 1230012
Acquaviva V., Gawiser E., 2010, Phys. Rev. D, 82, 082001
Aguirre A., Tegmark M., 2005, J. Cosmol. Astropart. Phys., preprint (arXiv:hep-th/0409072)
Amendola L., Barbieri R., 2006, Phys. Lett. B, 642, 192
Amendola L. et al., 2012, Living Rev. Relativ., 16, 6
Anderhalden D., Diemand J., Bertone G., Maccio A. V., Schneider A., 2013, J. Cosmol. Astropart. Phys., preprint (arXiv:1206.3788)
Angulo R. E., Hahn O., Abel T., 2013, MNRAS, 434, 3337
Arbey A., Lesgourgues J., Salati P., 2001, Phys. Rev. D, 64, 123528
Arbey A., Lesgourgues J., Salati P., 2003, Phys. Rev. D, 68, 023511
Arvanitaki A., Dubovsky S., 2011, Phys. Rev. D, 83, 044026
Arvanitaki A., Dimopoulos S., Dubovsky S., Kaloper N., March-Russell J., 2010, Phys. Rev. D, 81, 123530
Bae K. J., Baer H., Lessa A., 2013, preprint (arXiv:1306.2986)
Benson A. J. et al., 2012, MNRAS, 428, 1774
Berezhiani Z. G., Sakharov A. S., Khlopov M. Y., 1992, Sov. J. Nucl. Phys., 55, 1063
Bernal A., Guzman F. S., 2006, Phys. Rev. D, 74, 063504
Bernal A., Matos T., Nunez D., 2003, preprint (arXiv:0303455)

Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Bousso R., Hall L., 2013, Phys. Rev. D, 88, 063503
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2012, MNRAS, 422, 1203
Brooks A. M., Kuhlen M., Zolotov A., Hooper D., 2013, ApJ, 765, 22
Cicoli M., Goodsell M., Ringwald A., 2012, J. High Energy Phys., 2012, 146
Eisenstein D. J., Hu W., 1997, ApJ, 496, 605
Ferreira P. G., Joyce M., 1997, Phys. Rev. Lett., 79, 4740
Ferreira P. G., Joyce M., 1998, Phys. Rev. D, 58, 023503
Garrison-Kimmel S., Rocha M., Boylan-Kolchin M., Bullock J., Lally J., 2013a, MNRAS, 433, 3539
Garrison-Kimmel S., Rocha M., Boylan-Kolchin M., Bullock J., Lally J., 2013b, MNRAS, 433, 3539
Gleiser M., Sicilia D., 2009, Phys. Rev. D, 80, 125037
Governato F. et al., 2012, MNRAS, 422, 1231
Hertzberg M. P., 2012, preprint (arXiv:1208.0839)
Hu W., Barkana R., Gruzinov A., 2000, Phys. Rev. Lett., 85, 1158
Hwang J.-c., Noh H., 2009, Phys. Lett. B, 680, 1
Jaeckel J., Ringwald A., 2010, Annu. Rev. Nucl. Part. Sci., 60, 405
Komatsu E. et al., 2009, ApJS, 180, 330
Lacey C. G., Cole S., 1993, MNRAS, 262, 627
Lee J.-W., Lim S., 2009, J. Cosmol. Astropart. Phys., 2010, 007
Lewis A., 2000, available at: http://camb.info/
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 473
Lovell M. R. et al., 2012, MNRAS, 420, 2318
Maccio' A. V., Paduroiu S., Anderhalden D., Schneider A., Moore B., 2012a, MNRAS, 424, 1105
Maccio' A. V., Ruchayskiy O., Boyarsky A., Munoz-Cuartas J. C., 2012b, MNRAS, 428, 882
Macciò A. V., Paduroiu S., Anderhalden D., Schneider A., Moore B., 2012c, MNRAS, 424, 1105
Marsh D. J. E., Ferreira P. G., 2010, Phys. Rev. D, 82, 103528
Marsh D. J. E., Macaulay E., Trebitsch M., Ferreira P. G., 2012, Phys. Rev. D, 85, 103514
Marsh D. J. E., Grin D., Hlozek R., Ferreira P. G., 2013, Phys. Rev. D, 87, 121701
Matos T., Urena-Lopez L. A., 2000, Phys. Rev. D, 63, 063506
Matos T., Vázquez-González A., Magaña J., 2009, MNRAS, 393, 1359
Medvedev M. V., 2013, preprint (arXiv:1305.1307)
Mirizzi A., Redondo J., Sigl G., 2009, J. Cosmol. Astropart. Phys., 2009, 001
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493 (NFW)
Pacucci F., Mesinger A., Haiman Z., 2013, MNRAS, 435, L53
Pani P., Cardoso V., Gualtieri L., Berti E., Ishibashi A., 2012, Phys. Rev. Lett., 109, 131102
Park C.-G., Hwang J.-c., Noh H., 2012, Phys. Rev. D, 86, 083535
Peccei R. D., Quinn H. R., 1977, Phys. Rev. Lett., 38, 1440
Peebles P. J. E., Nusser A., 2010, Nature, 465, 565
Percival W. J., 2005, A&A, 443, 819
Percival W. J., Miller L., Peacock J. A., 2000, MNRAS, 318, 273
Preskill J., Wise M., Wilczek F., 1983, Phys. Lett. B, 120, 127
Press W. H., Schechter P., 1974, ApJ, 187, 425
Primack J. R., 2009, New J. Phys., 11, 105029
Ringwald A., 2012a, Phys. Dark Universe, 1, 116
Ringwald A., 2012b, preprint (arXiv:1209.2299)
Schneider A., Smith R. E., Reed D., 2013, MNRAS, 433, 1573
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Smith R. E., Markovic K., 2011, Phys. Rev. D, 84, 063507
Svrcek P., Witten E., 2006, J. High Energy Phys., 2006, 051
Tarrant E. R. M., van de Bruck C., Copeland E. J., Green A. M., 2012, Phys.Rev. D, 85, 023503
Tegmark M., Aguirre A., Rees M. J., Wilczek F., 2006, Phys. Rev. D, 73, 023505
Turner M. S., 1983, Phys. Rev. D, 28, 1243
Turner M. S., 1986, Phys. Rev. D, 33, 889
Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502
Weinberg S., 1978, Phys. Rev. Lett., 40, 223
Weinberg D. H., Bullock J. S., Governato F., Kuzio de Naray R., Peter A. H. G., 2013, preprint (arXiv:1396.0913)
Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Maccio A. V., 2012, MNRAS, 426, 2797
Wise M. B., Georgi H., Glashow S. L., 1981, Phys. Rev. Lett., 47, 402
Witten E., 1984, Phys. Lett. B, 149, 351
Zavala J., Vogelsberger M., Walker M. G., 2013, MNRAS, 431, L20
Zemp M., Gnedin O. Y., Gnedin N. Y., Kravtsov A. V., 2012, ApJ, 748, 54

APPENDIX A: DETAILS OF THE MIXED DARK MATTER PROFILE

In Section 4.2, we discussed the simple cored profile. The two-component profile of Section 4.3 is given by

$$\rho_{\rm aMDM}(M, r, f_c) = \theta(r - r_{J,h}(M_s))\rho_{\rm NFW}(M_s, r) + \theta(r_{J,h}(M_s) - r)[(1 - f_c)\rho_{\rm NFW}(M_s, r_{J,h}(M_s)) + \rho_{\rm NFW}(M_\star(M_s, f_c), r)], \quad \text{(A1)}$$

where $\theta(x)$ is the Heaviside function and $f_c = \Omega_{\rm c}/\Omega_{\rm d}$. $M_s \neq M_{200}$ is the mass of the external NFW profile, the 'scale mass'. The scale mass of the internal NFW profile for the CDM, $M_\star(M_s, f_c)$, is fixed by continuity and the requirement that at the halo Jeans scale the ULAs and CDM are in their relative cosmic abundance

$$\rho_{\rm NFW}(M_\star(M_s, f_c), r_{J,h}(M_s)) = f_c\rho_{\rm NFW}(M_s, r_{J,h}(M_s)). \quad \text{(A2)}$$

The CDM NFW profile inside the axion Jeans scale must be assigned a concentration, which must be computed from the variance in a given cosmology. We take the appropriate variance to be the one for the cosmology as a whole. This assumes that, with hierarchical structure formation, this CDM inner region will be made from a lighter halo itself formed earlier in cosmic history.

To find the size of the cored region for a given CDM fraction f_c we simply find the radius $r_{\rm core}$ that solves

$$(1 - f_c)\rho_{\rm NFW}(M_s, r_{J,h}(M_s)) = \rho_{\rm NFW}(M_\star(M_s, f_c), r_{\rm core}). \quad \text{(A3)}$$

As $f_c \to 1$, $r_{\rm core} \to r_{J,h}(M_s)$, the cored region disappears and equation (A1) goes to the standard NFW case.

This paper has been typeset from a TEX/LATEX file prepared by the author.