Structure formation with suppressed small-scale perturbations

Aurel Schneider[1,2]★

[1]*Center for Theoretical Astrophysics and Cosmology, Institute for Computational Sciences, University of Zurich, CH-8057 Zurich, Switzerland*
[2]*Department of Physics and Astronomy, University of Sussex, Brighton, BN1 9QH, UK*

ABSTRACT

All commonly considered dark matter scenarios are based on hypothetical particles with small but non-zero thermal velocities and tiny interaction cross-sections. A generic consequence of these attributes is the suppression of small-scale matter perturbations either due to free-streaming or due to interactions with the primordial plasma. The suppression scale can vary over many orders of magnitude depending on particle candidate and production mechanism in the early Universe. While non-linear structure formation has been explored in great detail well above the suppression scale, the range around suppressed perturbations is still poorly understood. In this paper, we study structure formation in the regime of suppressed perturbations using both analytical techniques and numerical simulations. We develop simple and theoretically motivated recipes for the halo *mass function*, the expected *number of satellites*, and the halo *concentrations*, which are designed to work for power spectra with suppression at arbitrary scale and of arbitrary shape. As case studies, we explore warm and mixed dark matter scenarios where effects are most distinctive. Additionally, we examine the standard dark matter scenario based on weakly interacting massive particles (WIMP) and compare it to pure cold dark matter with zero primordial temperature. We find that our analytically motivated recipes are in good agreement with simulations for all investigated dark matter scenarios, and we therefore conclude that they can be used for generic cases with arbitrarily suppressed small-scale perturbations.

Key words: cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Within the last decade, numerical simulations have substantially increased in performance and accuracy contributing to what has been denominated the new era of precision cosmology. Statistical measures such as the matter power spectrum can now be calculated to super cent level precision, making it possible to obtain competitive constraints of fundamental cosmological parameters with galaxy surveys. This precision is, however, only obtained in a regime where structure formation behaves in a regular, strictly hierarchical way, owing to a well-behaved linear power spectrum with a close to power-law scaling. As soon as this basic assumption is relaxed, standard numerical methods are not guaranteed to work anymore. In the case of a suppressed power spectrum, as it appears prominently in hot and warm dark matter models, standard numerical methods produce artefacts near the suppression scale by enhancing small non-physical modes present in the initial conditions (ICs; Götz & Sommer-Larson 2003; Wang & White 2007).

Along with the development of numerical schemes, analytical techniques such as the extended Press–Schechter (EPS) approach (Press & Schechter 1974; Bond et al. 1991) and the halo model (Cooray & Sheth 2002) have been put forward to describe non-linear structure formation. They do not achieve the same precision than numerical simulations, but they provide qualitative understanding of the physical processes involved and they have the advantage of not requiring big computer facilities. However, these analytical techniques again only work in the case of a reasonably behaved initial power spectrum with nearly power-law scaling. As soon as the initial power spectrum is significantly suppressed, standard EPS approaches and subsequently the halo model fail to predict the right clustering (Schneider, Smith & Reed 2013).

Understanding structure formation of cosmologies with suppressed perturbations is not merely an academical exercise but a necessity in order to capture the full range of scales of cosmic clustering. As a matter of fact, all reasonable dark matter (DM) scenarios exhibit a suppression of perturbations below a certain scale either due to particle free-streaming or due to interactions in the early Universe. For the prime DM candidate, the weakly interacting massive particle (WIMP), the mass scale of suppression lies somewhere between roughly 10^{-12} and 10^{-4} $h^{-1}\,\mathrm{M}_{\odot}$
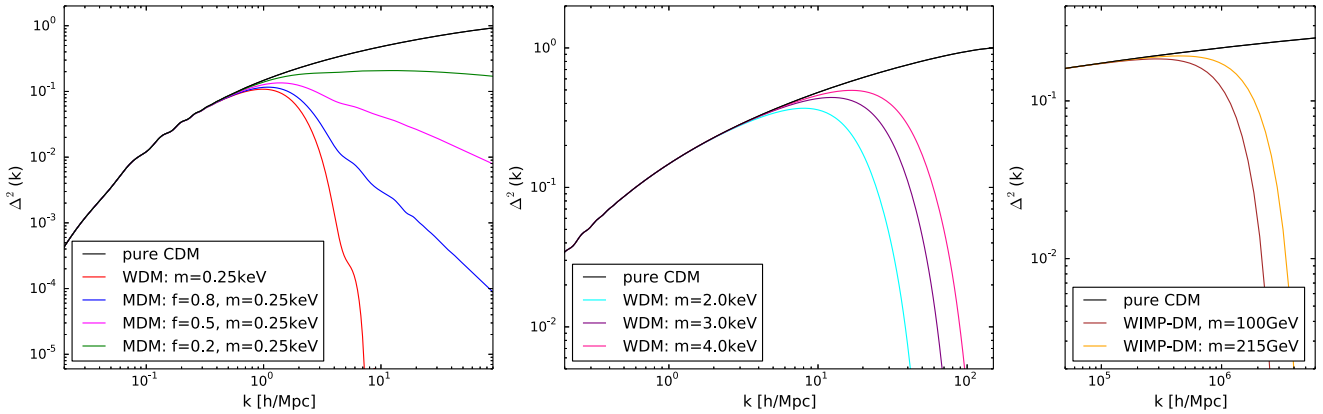
★ E-mail: aurel@physik.uzh.ch

Figure 1. Linear dimensionless power spectrum $\Delta(k) = k^3 P_\chi(k)/2\pi^2$ of the DM scenarios (χ) investigated in this paper. Left: CDM (black), WDM (red) and various MDM models (blue, magenta, green) at redshift 50. Middle: CDM (black) and different WDM models (cyan, purple, pink) at redshift 100. Right: pure CDM (black) and two WIMP-DM scenarios (brown, orange) at redshift 300.

depending on the specific model parameters (Green, Hofmann & Schwarz 2005; Profumo, Sigurdson & Kamionkowski 2005), which is orders of magnitude smaller than the smallest galaxies we observe.[1] Other promising DM candidates such as the sterile neutrino or the gravitino have higher primordial velocities, pushing the regime of power suppression to larger scales where it becomes relevant for galaxy surveys such as in the case of warm or mixed DM. More exotic scenarios like interacting DM (Boehm & Schaeffer 2005), self-interacting DM (Spergel & Steinhardt 2000), asymmetric DM (Petraki & Volkas 2013), or ultralight axion DM (Marsh & Silk 2014) can also lead to suppressed power at rather large scales either because they interact with the cosmic plasma, with dark radiation or because they undergo scalar field oscillations on astrophysical scales. Since the power suppression of different DM scenarios happens at different scales and is of varying shape, it is crucial to properly quantify the non-linear clustering in this regime in order to distinguish between different DM species and to contribute towards a solution of one of the outstanding puzzles in modern physics.

In this paper, we study non-linear structure formation starting from initial power spectra with arbitrary small-scale suppressions. As working examples, we investigate the cases of warm DM, MDM, WIMP DM, and pure cold DM. We run a suite of *N*-body simulations for these models and carefully remove artificial structures which tend to populate simulations of suppressed initial power. At the same time, we develop an EPS method able to cope with arbitrary linear power spectra. We then calculate important quantities, such as the halo mass function, the number of satellites, and the concentration–mass relation of halo profiles.

The paper is structured as follows. In Section 2, we give a short overview of different DM scenarios and discuss how they suppress perturbations. Section 3 contains a summary of the simulations and analysis techniques including a discussion about subtracting artefacts. In Sections 4 and 5, we present our modified Press–Schechter approach and use it to predict mass function, number of Milky Way (MW) satellites and concentrations. Finally, we conclude in Section 6 and give further details about the removal of artefacts in the appendix.

---

[1] The only chance of observing such small structures is probably by directly measuring the flux from DM annihilation.

2 DM MODELS AND THE SUPPRESSION OF PERTURBATIONS

The physical mechanisms leading to the power suppression below a certain scale depend on the DM candidate but is usually either particle free-streaming, tiny interactions with the primordial plasma, or more exotic phenomena in the early Universe. Both the scale and the detailed shape of the suppression strongly vary depending on the characteristics of the DM particle, i.e. its mass, momentum, and interaction cross-sections. In the following, we discuss the most common DM scenarios and outline their effects on non-linear structure formation.

2.1 Warm dark matter (WDM)

DM models with a steep cutoff-suppression at dwarf galaxy scales, similar in shape to the one obtained from particles with a Fermi–Dirac momentum distribution, are usually referred to as WDM. Typical power spectra of WDM models are illustrated in Fig. 1 (red, cyan, purple, and pink lines). Implicitly, the WDM regime is defined to lie in the rather narrow band where the suppression of perturbation has a significant effect on dwarf galaxy formation but is still in agreement with observations. Furthermore, WDM is generally assumed to reduce potential small-scale inconsistencies, such as the *missing satellite* or the *too-big-to-fail* problem, occurring in a Λ cold dark matter (CDM) universe (see Weinberg et al. 2013, for a review on these topics).

Recent studies point out that both requirements – passing observational constraints and alleviating small-scale inconsistencies – seem impossible to combine because constraints from Lyman-α are prohibitively strong in the case of WDM (Viel et al. 2013; Schneider et al. 2014). While some question the accuracy of the Lyman-α measurements, more and more people believe that the small-scale inconsistencies are a consequence of poorly understood baryonic processes of galaxy formation (Brooks & Zolotov 2014).

However, it is possible and worthwhile to use structure formation in order to find upper limits on the scale of DM suppression. The strongest constraints currently come from flux observations of the Lyman-α forest ruling out particle masses below $m_{\mathrm{TH}} = 3.3$ keV at the 2σ confidence level (Viel et al. 2013). Other limits are obtained from dwarf galaxy counts disfavouring particle masses below

$m_{TH} = 2.3$ keV (Polisensky & Ricotti [2011]; Kennedy et al. [2014]). Here, m_{TH} refers to the equivalent thermal mass, assuming a Fermi–Dirac momentum distribution.

The prime candidate for WDM is the hypothetical sterile (or right-handed) neutrino, which can be readily added to the standard model. Sterile neutrinos are well motivated because they provide an explanation for the measured mixing angles of active neutrinos, and because all other fermions exist with both left and right chirality (Drewes [2013]). The particle mass of sterile neutrino DM is expected to lie in the keV-range, what makes it an ideal candidate for WDM or lukewarm DM depending on the suppression scale which is governed by particle mass and momentum distribution.

The recent tentative discovery of an X-ray line at 3.5 keV from galaxy clusters and M31 by two independent research groups (Boyarsky et al. [2014]; Bulbul et al. [2014]) might provide additional motivation for sterile neutrino DM. Since sterile neutrinos decay radiatively, the confirmation of such a signal would be a smoking gun for an $m_{sn} = 7.1$ keV particle, which would translate into a thermal mass of $m \simeq 2.5$–5 keV depending on the production mechanism in the early Universe (Abazajian [2014]; Merle & Schneider [2014]). The measured X-ray line is, however, disputed, since there are some conflicting bounds from galaxies where the signal is not measured (see Iakubovskyi [2014], for a review). Better astronomical data is needed to check if the signal is real and if the flux coincides with the expectations from decaying DM.

2.2 Mixed dark matter (MDM)

In principle, DM can be made of two or more particle species with different properties. Particularly interesting in terms of structure formation is a mixture of cold with warm or hot DM. This leads to qualitatively different initial power spectra with gradual suppression over many orders of magnitude in scale in contrary to the steep cutoff known from WDM. The effect is illustrated in the left-hand panel of Fig. [1], where we plot the dimensionless power spectra of different MDM scenarios (consisting of a CDM part and a thermal WDM part with $m = 0.25$ keV) with increasing mixing fraction $f = \Omega_{WDM}/(\Omega_{CDM} + \Omega_{WDM})$.

The concept of MDM is neither new nor particularly exotic. In fact, it is clear that there must be more than one DM component since neutrinos are known to have non-zero mass and cannot by themselves account for all of the DM. Current observations constrain the sum of neutrino masses to the range of 0.05–0, 23 eV (Fogli et al. [2012]; Planck Collaboration XVI [2014]), yielding a mixing fraction $f \sim 0.004$–0.02. However, the very small neutrino masses result in an overall suppression of the power spectrum at all relevant scales for halo formation, very similar to the case of pure CDM with low σ_8-normalization (Viel, Haehnelt & Springel [2010]).

Instead of two (or more) distinct particles acting as DM components, MDM-like compositions can also arise due to multichannel DM production in the early Universe, yielding momentum distributions that mimic the case of several DM components. Prime example is again the sterile neutrino which can be produced via resonant oscillations with active neutrinos (Shi & Fuller [1999]), where some subdominant part is always produced out of resonance leading to particle momenta from two overlapping distributions (Boyarsky et al. [2009]). The effect can be even stronger if the sterile neutrinos with non-zero mixing angle are produced via the decay of heavy scalars yielding a momentum distribution with two distinct peaks (Merle & Schneider [2014]; Merle & Totzauer [2014]).

2.3 WIMP DM

The most popular group of DM candidates are WIMP with the neutralino as prime candidate. The popularity of WIMP DM comes from the fact that such particles naturally appear in supersymmetric extensions of the standard model, and that they are produced via thermal freeze-out at roughly the right amount to account for the observed DM abundance (this is usually referred to as the *WIMP miracle*; see Bertone, Hooper & Silk [2005], for a summary).

As WIMPs are heavy (with particle masses in the GeV or TeV scales) and weakly interacting, they become non-relativistic very early leading to extremely small suppression scales. Depending on the parameters of the model, the mass scale of WIMP DM suppression is expected to lie between roughly 10^{-12} and 10^{-4} h^{-1} $M_\odot$ (Profumo et al. [2005]).

In the right-hand panel of Fig. [1], we plot the power spectra of neutralino DM with a mass of $m = 100$ (brown line) and $m = 215$ GeV (orange line) and corresponding decoupling temperatures of $T_{dk} = 28$ MeV and $T_{dk} = 33$ MeV. These spectra are compared to the hypothetical case of pure CDM (black line). The suppression of power in WIMP scenarios happens at very high wave-numbers roughly following an exponential cutoff (Green et al. [2005]). We will show later on that such WIMP models suppress structure formation at a halo mass of about 10^{-6} h^{-1} $M_\odot$.

2.4 Other models with suppressed power

There are many more DM candidates with variable suppression scales depending on their interaction and free streaming properties. For example, DM can be self-interacting, as in the case of atomic or mirror DM (Cyr-Racine & Sigurdson [2013]), or it can interact with photons (Boehm et al. [2014]; Wilkinson, Lesgourgues & Boehm [2014a]) and neutrinos (Wilkinson, Boehm & Lesgourgues [2014b]) all yielding strong suppression of small-scale modes. Other possibilities with similar effects on small scales are decaying DM (Kaplinghat [2005]; Wang et al. [2014]), later forming DM (Agarwal et al. [2014]), or ultralight axion DM (Marsh & Silk [2014]).

It is also possible to obtain suppressed small-scale perturbations from effects not related to DM. Inflation could lead to a running of the spectral index, gradually reducing power on small scales (Kosowsky & Turner [1995]), or it could induce a strong cutoff similarly to the case of WDM (Kamionkowski & Liddle [2000]).

3 NUMERICAL SIMULATIONS

We run and analyse numerical simulations of different resolution with linear power spectra representing cold, warm, mixed, and WIMP DM scenarios. The ICs are generated from the linear power spectra illustrated in Fig. [1], which are selected to cover different scales of power suppression as well as a variety of shapes from steep cutoffs to shallow decreases towards large wave-numbers.

The linear power spectra have been calculated with the Boltzmann solver CLASS (Blas, Lesgourgues & Tram [2011]), which comes with an option for multiple DM species of arbitrary mass and momentum distribution (Lesgourgues & Tram [2011]). We use cosmological parameters obtained by *Planck*, i.e. $\Omega_m = 0.304$, $\Omega_b = 0.048$, $\Omega_\Lambda = 0.696$, $H_0 = 68.14$, $n_s = 0.963$, and $\sigma_8 = 0.827$ (Planck Collaboration XVI [2014]). The simulations are set up with the ICs generator MUSIC (Hahn & Abel [2011]) using second-order Lagrangian perturbation theory and initial redshifts of [50, 100, 200, 200, 300] for runs with box-size [256, 64, 8, 4, 0.0001] h^{-1} Mpc. The simulations are performed with PKDGRAV2 (Stadel [2001]),

Table 1. List of simulations performed in this paper. The parameters m, f, T_{kd} refer to (thermal) particle mass, mixing ratio, and decoupling temperature. L [h^{-1} Mpc] designates the box-size and z_f the final redshift. All simulations have been run at both LR and HR (with 512^3 and 1024^3 particles, respectively).

Scenario:		L (h^{-1} Mpc)	z_f
CDM:	Pure	24, 64, 256	0
WDM:	$m = 0.25$ keV	64, 256	0
MDM:	$m = 0.25$ keV, $f = 0.8$	64	0
MDM:	$m = 0.25$ keV, $f = 0.5$	64	0
MDM:	$m = 0.25$ keV, $f = 0.2$	64	0
WDM:	$m = 2.0$ keV	24	0
WDM:	$m = 3.0$ keV	24	0
CDM:	Pure	8	5
WDM:	$m = 2.0$ keV	8	5
WDM:	$m = 3.0$ keV	8	5
WDM:	$m = 4.0$ keV	4	5
CDM:	Pure	10^{-4}	23
WIMP:	$m = 100$ GeV, $T_{kd} = 28$ MeV	10^{-4}	23
WIMP:	$m = 215$ GeV, $T_{kd} = 33$ MeV	10^{-4}	23

applying a standard force softening of 1/50 times the mean particle separation. For the halo finding, we use AHF with particle unbinding and an overdensity criterion corresponding to spherical top-hat collapse (Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009). A list summarizing the main characteristics of the simulations is given in Table 1.

Most of the runs suffer from artificial clumping at the scale of suppressed perturbations. This is not only the case for WDM, where the effect has been observed many times before, but also for WIMP-DM and even for the MDM scenarios with initial power spectra that never drop to zero. In the next sections, we discuss the issue of artificial halo formation in more detail, and we discuss how artefact can be filtered out of halo catalogues from numerical simulations.

3.1 The problem of artificial clumping

Artificial clumping is a serious problem in simulations where the initial power is suppressed below a characteristic scale, usually referred to as the free-streaming or damping scale. Standard *N*-body techniques cannot cope with this setup and produce many small-scale haloes that are resolution-dependent and therefore non-physical (Götz & Sommer-Larson 2003; Wang & White 2007). In this paper, we show that the artificial clumping does not only happen in setups with a steep cutoff and zero physical power below a certain scale, but also in MDM scenarios, where the suppression of perturbations is much smoother and the power spectrum never drops to zero.

Recently, there has been attempts to overcome the problem of artefacts with a novel *N*-body technique tracking the DM sheet in phase-space (Abel, Hahn & Kaehler 2012). With this method, the amount of artificial haloes in WDM simulations is drastically reduced (Angulo, Hahn & Abel 2013). The method is, however, not fully operational yet, since it has difficulties to cope with high-density regions, where the six-dimensional phase-space sheet undergoes multiple foldings. Further studies seem needed in order evaluate whether this promising method can be improved to simultaneously deal with artificial clumping and accurately calculate regions of high density.

In this paper, we use a standard *N*-body technique combined with a post-processing method to identify and remove artificial haloes

from the halo catalogue. The method is explained in the following section.

3.2 Removing artefacts

Several methods of different complexity to identify artificial haloes have been proposed in the past. Diemand, Moore & Stadel (2005) removed haloes on the basis of visual inspection. Schneider et al. (2013) suggested to subtract a power-law component in the halo mass function. Agarwal & Corasaniti (2015) identified artefacts by measuring the spin of haloes found with the friends-of-friends technique. The most complex and arguably most complete method to remove artefacts has been proposed by Lovell et al. (2014), using both the shape of protohaloes in the ICs (i.e. the volume spanned by all particles ending up in a halo) as well as their overlap between realizations with different resolution. In a first step, Lovell et al. (2014) measured the sphericity of protohaloes, removing all objects that are unusually elongated. This selection criterion is based on the observation that artefacts originate from the collapse of larger modes to walls and filaments, pushing already aligned particles together until they cluster into non-physical clumps. In the second step, they analysed the resolution dependence of haloes, removing all objects that do not appear in two simulations of the same initial field but different resolution.

In this paper, we use a similar method than the one proposed by Lovell et al. (2014) and apply it to cosmological simulations. This is more challenging than the original application which was limited to substructures within an MW-sized host haloes. In our case, we use a less stringent limit of 1000 particles for the minimal halo size (as opposed to 10 000 particles in Lovell et al. 2014), which means that we have up to a quarter of a million haloes per simulation with masses spanning more than four orders of magnitude. These haloes are in varying large-scale surroundings and at different stages of their evolution, making it hard to reliably separate real haloes from artefacts. In the following, we show that although a perfect separation is not possible, most artefacts can be reliably identified while only a small fraction of real haloes is erroneously removed. We use the simulations with the fewest and the most artefacts – the CDM and the WDM ($m = 0.25$ keV) run – to illustrate the method and estimate the efficiency of filtering out artefacts without erroneously removing physical haloes. Further information on how halo catalogues of the remaining DM models are affected by the method is provided in the appendix.

The first step of the algorithm consist of measuring the ellipticity of protohaloes in the ICs using the inertia tensor

$$I_{ij} = m \sum_{\text{particles}} \left(\delta_{ij} \mathbf{x}^2 - x_i x_j \right), \tag{1}$$

where the sum goes over all the protohalo particles and δ_{ij} is the Kroenecker delta. The vector $\mathbf{x}$ corresponds to the particle position with respect to the centre of mass of the protohalo. The eigenvalues of I_{ij} are directly related to the axis $a \geq b \geq c$ of an ellipsoid with the same inertia tensor. This means it is possible to define the axis ratios $s = c/a$ and $q = c/b$ providing a measure of the protohalo sphericity.[2]

In Fig. 2, we show histograms of haloes with ellipticity parameters s (blue) and q (red). The top row corresponds to the cases of

---

[2] In Lovell et al. (2014), only s is used to distinguish between artefacts and real hales. We find it advantageous to use both s and q for the filtering procedure.
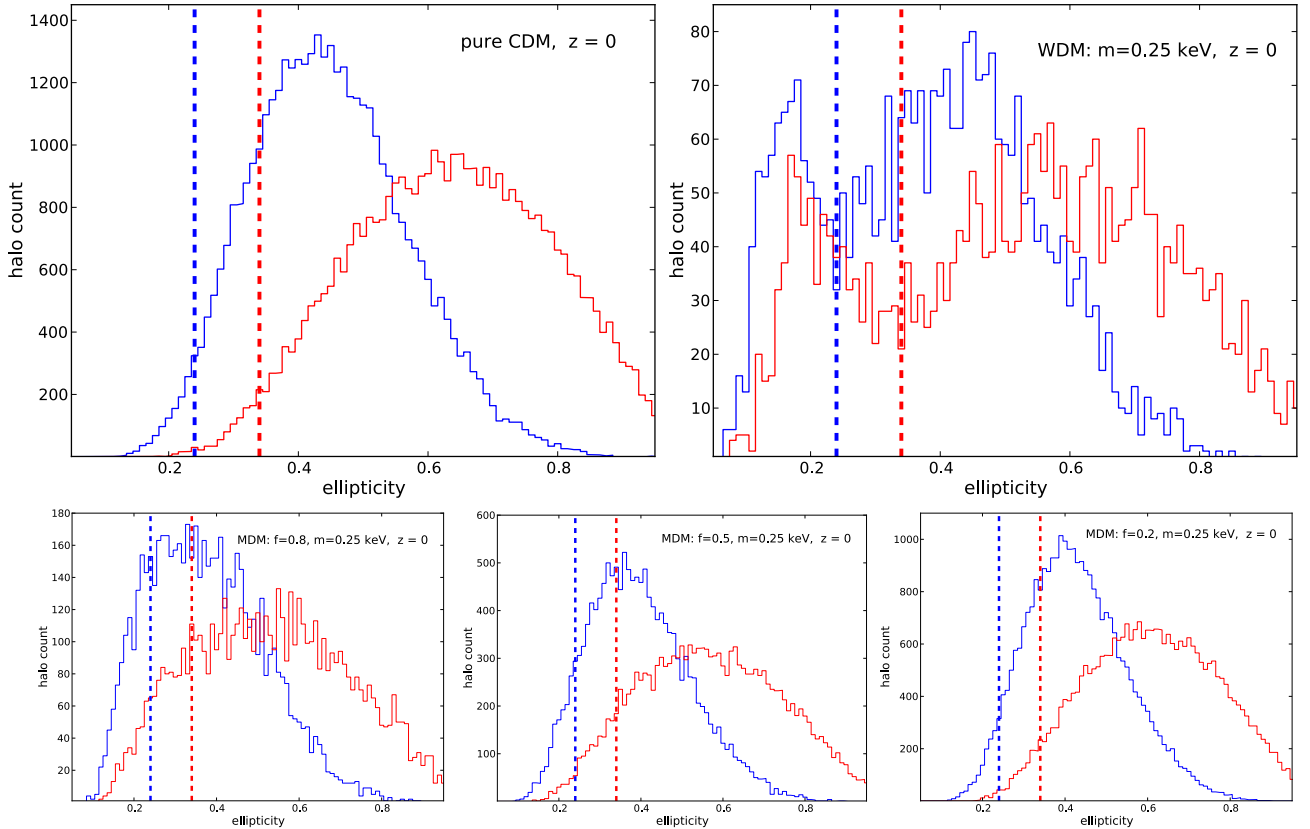
Figure 2. Histograms of haloes with respect to the ellipticity parameters $s = c/a$ (blue) and $q = c/b$ (red), where $a \geq b \geq c$ are the semimajor axes of an ellipsoid with the same inertia tensor than the protohalo. The top row illustrates the cases of CDM (left) and WDM with $m = 0.25$ keV (right), the bottom row the cases of MDM with $f = 0.8$ (left) $f = 0.5$ (middle) and $f = 0.2$ (right). Vertical dashed lines at 0.24 and 0.34 designate the sphericity cut applied to separate real haloes from artefacts.

CDM (left) and WDM with $m = 0.25$ keV (right). While for CDM both parameters follow approximately Gaussian distributions, the WDM case exhibits secondary peaks at low values of s and q. This is a clear indication for the presence of an artificial population of haloes with very elongated protohaloes. The bottom row of Fig. 2 illustrates the same statistics for the MDM models with mixing factor $f = 0.8$ (left), $f = 0.5$ (middle), and $f = 0.2$ (right). The distributions do not show secondary peaks, but they are considerably broader and shifted towards lower values of s and q than in the case of CDM, indicating the presence of a subpopulation of artefacts in these models as well.

Based on the information of Fig. 2, we set cuts of $s_c = 0.24$ and $q_c = 0.34$ on the halo samples, removing all objects with smaller ellipticity parameters in all DM models. We will call this the *sphericity cut* in the following. In the WDM model, the values s_c and q_c lie at the minimum between the two peaks of real and artificial haloes, filtering out 60 per cent of all haloes. In the case of CDM, on the other hand, the same values remove less than 5 per cent of all haloes. This is a small loss considering the fact that changing the halo-finder can translate into 10 per cent differences in the abundance of haloes (Knebe et al. 2011).

The second step of the algorithm consists of additionally filtering out haloes with a poor match between simulations of the same initial field but different resolution. We again follow Lovell et al. (2014) and define the merit function

$$M = \frac{V_{\text{shared}}}{\sqrt{V_{\text{LR}} V_{\text{HR}}}} \qquad (2)$$

for every protohalo, where V_{LR} and V_{HR} are the protohalo volumes in the low-resolution (LR) and high-resolution (HR) simulations, and where V_{shared} is the overlap of the two (i.e. the shared volume). The merit function can be computed with any standard merger-tree algorithm based on particle IDs (see Srisawat et al. 2013), as they can be used to link haloes between different simulations. Since standard merger-tree codes require the same particle numbers between analysed snapshots, we formally increase the particle number of the LR output by splitting every LR particle into n particles corresponding to the HR particles of the same phase-space volume. This simple trick allows us to use the publicly available code MERGERTREE (out of the AHF-package, Knollmann & Knebe 2009) to calculate the number of shared particles between all haloes in the LR and HR simulations. This then trivially leads to the merit function of equation (2) since every particle can be related to a small volume element in the ICs.

In Fig. 3, we illustrate the histograms of haloes with respect to the merit function M for CDM (black), WDM (red), and MDM (blue, green, and magenta). The top panel shows all haloes with more than 1000 particles, i.e. before applying the *sphericity cut*. While the CDM distribution only exhibits one distinct peak at $M \sim 0.8$, the WDM distribution has a peak at $M \sim 0.9$ plus a secondary, much broader feature at $M < 0.7$ which can be attributed to artefacts. The bottom panel of Fig. 3 shows the same statistics for the haloes which are not filtered out by the *sphericity cut*. While the CDM distribution does hardly change, the broad feature at low M in WDM is substantially smaller, indicating that most of the artefacts
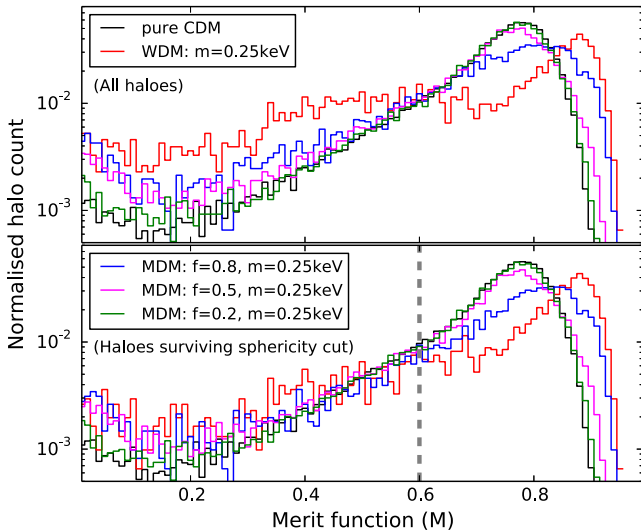
Figure 3. Normalized halo histograms of the merit function (which measures the overlap of protohaloes between simulations of different resolution). The top and bottom panels refer to the halo sample before and after the *sphericity cut*. The vertical dashed line illustrates the *resolution cut*.

have already been filtered out. However, some artefacts survive the *sphericity cut* and still contribute to the histograms of the bottom panel. This is emphasized by the relative excess of WDM and MDM haloes at low values of *M*.

In order to filter out the remaining artefacts visible in the lower panel of Fig. 3, we use the statistics of the merit function to apply a second selection cut of $M_c = 0.6$ (illustrated by the grey dashed line) which we will call the *resolution cut*. This further removal should capture most of remaining artefacts, while keeping the majority of the real haloes in the sample. For the cases of CDM and WDM, the *resolution cut* filters out an additional 15 and 40 per cent of the remaining haloes.

In summary, we note that applying solely the *sphericity cut* leads to rather conservative results in the sense that CDM haloes are hardly affected, but it does not filter out all of the WDM (and MDM) artefacts. Applying both the *sphericity* and the *resolution cut* filters out more artefacts, but it also removes some of the real objects.

The halo mass function of Fig. 4 illustrates how the method of removing artefacts affects the total halo abundance at different masses. Empty and filled symbols illustrate the mass function before and after the removal of haloes. Filled symbols appear twice illustrating the effect of only applying the *sphericity cut* (upper symbols) and both the *sphericity* and *resolution cut* (lower symbols). The coloured area in between therefore illustrates the inherent uncertainty of the method.

4 EPS MODEL

Many aspects of non-linear structure formation can be analytically described by the EPS approach, which assumes linear growth of perturbations followed by immediate halo formation above a certain threshold (see Zentner 2007, for a review). The threshold is derived from an idealized spherical or ellipsoidal collapse calculation. Despite of these oversimplified assumption, the EPS model is able to reproduce fundamental statistics of structure formation, such as halo mass function, accretion history or bias (Press & Schechter 1974; Bond et al. 1991; Sheth & Tormen 1999). This is, however, only true for a pure CDM scenario where the initial power spectrum behaves as a quasi-power-law which is never steeper than k^{-3}. As soon as the initial power spectrum is suppressed further (i.e. becomes steeper than k^{-3}), the standard EPS model completely fails to predict the right clustering. In this section, we show how the formalism needs to be adapted in order to give adequate predictions for cosmologies with arbitrary initial power spectra including steep cutoffs or gradual suppressions. We thereby follow and extend the method developed in Schneider et al. (2013).

4.1 Halo mass function: the standard approach

The most used application and the starting point of the EPS model is the halo mass function. In the standard scenario it can be written as

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M} = \frac{1}{2}\frac{\bar{\rho}}{M} f(\nu)\frac{\mathrm{d}\ln \nu}{\mathrm{d}\ln M}, \qquad (3)$$

where n is the number density of haloes, M the halo mass, ν the peak-height of perturbations, and $\bar{\rho}$ the average density of the Universe (Press & Schechter 1974). The first crossing distribution $f(\nu)$ is obtained with the excursion-set approach, which follows random walk trajectories counting events of first up-crossing of the collapse
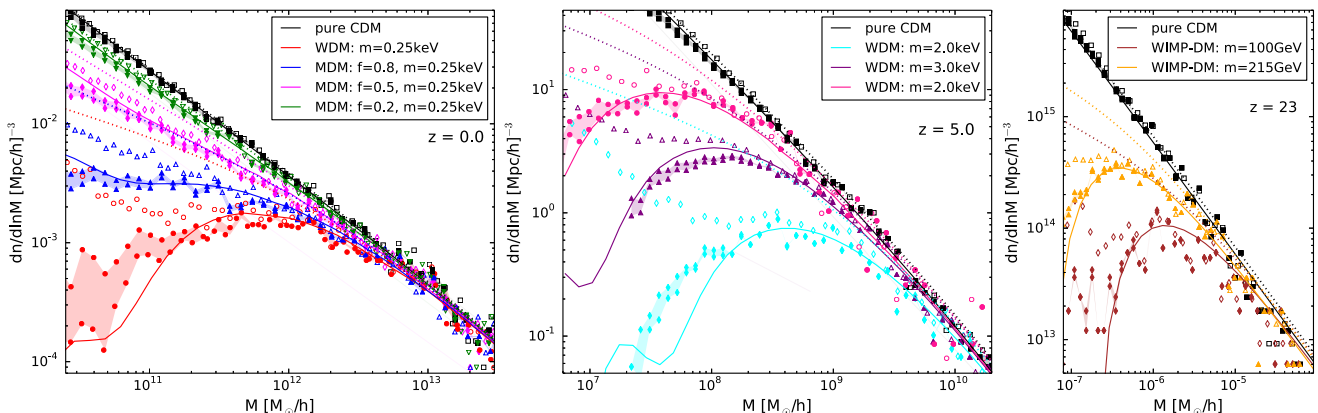

Figure 4. Halo mass functions for various DM models. Left: CDM (black), WDM (red), and MDM (blue, magenta, green) at redshift zero. Middle: CDM (black) and WDM (cyan, purple, pink) at redshift 5. Right: pure CDM (black) and WIMP DM (brown, orange) at redshift 23. The colour-shaded regions account for the uncertainty due to the removal procedure of artefacts (see the text). Solid lines represent the sharp-*k* model, dotted lines the standard Sheth–Tormen mass function.

threshold (Bond et al. 1991). The spherical collapse model predicts a constant threshold, leading to

$$f(\nu) = \sqrt{\frac{2\nu}{\pi}} e^{-\nu/2}. \tag{4}$$

The more realistic case of ellipsoidal collapse, on the other hand, leads to a mass-dependent threshold and a first-crossing distribution of the form

$$f(\nu) = A\sqrt{\frac{2q\nu}{\pi}} \left[1 + (q\nu)^{-p}\right] e^{-q\nu/2} \tag{5}$$

with $A = 0.3222$, $p = 0.3$, and $q = 1$ (Sheth & Tormen 1999). The peak height ν is defined as

$$\nu = \frac{\delta_{c,0}^2}{S_\chi(R)D^2(a)}, \tag{6}$$

where $\delta_{c,0} = 1.686$ and where

$$D(a) = \frac{5\Omega_m}{2}H(a)\int \frac{da}{a^3 H(a)^3} \tag{7}$$

is the universal growth factor of perturbations. Here, we have introduced the Hubble parameter $H = \dot{a}/a$. Finally, the variance S_χ is defined as

$$S_\chi(R) = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} P_\chi(k) W^2(k|R), \tag{8}$$

where χ stands for the DM scenario imposed by the linear power spectrum $P_\chi(k)$. There is a filter function in Fourier-space $W(k|R)$ appearing in equation (8), which is a priori unconstrained. One obvious choice used in the standard EPS approach is a top-hat function in real space with enclosed mass $M = 4\pi\bar{\rho}R^3/3$, transforming into

$$W(k|R) = \frac{3}{(kR)^3}\left[\sin(kR) - 3\cos(kR)\right] \tag{9}$$

in Fourier space. Other common choices for $W(k|R)$ are a sharp-k filter (a top-hat function in Fourier space) or a Gaussian filter (Bond et al. 1991).

The usual recipe to construct the EPS halo mass function consists of combining equations (3–9) leading to very good agreement with CDM simulations for halo masses below $M = 10^{13} \, h^{-1} \, \mathrm{M}_\odot$. Significantly larger haloes, on the other hand, are underpredicted by the model. Motivated by this discrepancy at the largest scales, Sheth & Tormen (1999) heuristically modified the first-crossing distribution by setting $q = 0.707$, what leads to very good agreement with simulations over all mass scales, but it technically consists of allowing for one free parameter. This modified model is commonly referred to as the Sheth–Tormen mass function.

In Fig. 4, we show the expected good agreement between the Sheth–Tormen mass function (dotted lines) and simulations of a CDM cosmology (black squares), which holds over a wide range of halo masses and redshifts. However, Fig. 4 also shows the complete failure of the Sheth–Tormen mass function to reproduce the halo abundance of WDM, MDM or WIMP DM simulations (coloured symbols) around the scale of suppression (the coloured area between the two sets of filled symbols designate the uncertainty in the method to filter out artificial haloes, the empty symbols illustrate the raw data, including real haloes and artefacts; see Section 3.2 for more details).

4.2 Halo mass function: the sharp-k model

The inability of the standard EPS approach to model cosmologies with suppressed power spectra has been reported by several papers

in the past (Barkana, Haiman & Ostriker 2001; Benson et al. 2013; Schneider et al. 2013; Hahn & Paranjape 2014), and different methods to solve this problem have been proposed. A simple and elegant way of modifying the model is to change the filter from a top-hat function in real space to a top-hat function in Fourier space, called sharp-k filter, i.e.

$$W(k|R) = \Theta(1 - kR), \tag{10}$$

where Θ is the Heaviside step function. This possibility was initially discussed in a paper by Bertschinger (2006) and has been adapted to WDM by Benson et al. (2013) and Schneider et al. (2013). The reason why the sharp-k filter does a better job than the standard top-hat filter lies in the asymptotic behaviour of equation (8) towards small radii. With a sharp-k filter, this integral naturally depends on the shape of the power spectrum for any radius. With a top-hat filter, on the other hand, the integral becomes completely insensitive to the shape of the power spectrum as soon as the latter decreases faster than k^{-3}. Instead, the mass function becomes solely driven by the shape of the filter function and strongly deviates from the halo abundance in simulations (see Schneider et al. 2013, for a more detailed discussion).

For the case of a sharp-k filter, the functional form of the halo mass function (i.e. equation 3) can be further simplified. A straightforward calculation leads to the expression

$$\frac{dn_{SK}}{d\ln M} = \frac{1}{12\pi^2}\frac{\bar{\rho}}{M}\nu f(\nu)\frac{P(1/R)}{\delta_c^2 R^3}. \tag{11}$$

For the sake of simplicity, we have dropped the χ-index standing for the DM model. Together with equations (5), (8), and (10) as well as with an appropriate relation between filter scale and mass, we get a closed set of equations yielding the halo number density per mass scale.

The sharp-k filter has an important drawback we have not discussed so far, namely that it has no well-defined mass M associated with its filter scale R. This means that, apart from the proportionality $M \propto R^3$ due to the spherical symmetry of the filter, the halo mass is unconstrained. Introducing a free parameter c, the relation between filter scale and mass can be written as

$$M = \frac{4\pi}{3}\bar{\rho}(cR)^3, \tag{12}$$

where $c = 2.5$ gives the best match to simulations.[3] The calibration of the parameter c is done once and does not change for different DM scenarios.

The sharp-k model has two main advantages compared to the standard EPS approach: first, it consistently uses the same filter (the sharp-k function) for the entire excursion-set calculation, while the EPS approach usually relies on the sharp-k filter to calculate the first crossing distribution and on the top-hat filter to connect it to the number density of haloes. Secondly, the sharp-k method is based on the original first-crossing distribution from ellipsoidal collapse, while the Sheth–Tormen model introduces a heuristic parameter adjustment. The second advantage is counterbalanced by the fact that the sharp-k filter does not have a well-defined mass. At the end, both the sharp-k and the Sheth–Tormen model have one free parameter, which needs to be adjusted to simulations.

The halo mass function from the simulations (symbols), the Sheth–Tormen model (dashed lines), and the sharp-k model (solid

---

[3] The best value for c depends somewhat on the halo-definition and the halo-finding routine. In Schneider et al. (2013), we used a slightly higher value $c = 2.7$.

lines) are all plotted in Fig. 4. For the case of CDM (black), the sharp-*k* mass function closely follows both the simulation measurements and the Sheth–Tormen model. For the case of WDM (red, cyan, purple, pink), MDM (green, magenta, blue), and WIMP DM (brown, orange), the sharp-*k* mass function gives a reasonably good match to simulations, while the Sheth–Tormen approach fails to match the flattening or the turnaround visible in simulations.

In Schneider et al. (2013), the sharp-*k* model has been reported to underestimate the halo abundance when the suppression scale lies in the exponential tail of the halo mass function (i.e. for $\nu \gg 1$), which generally happens at very high redshift. It turns out, however, that this discrepancy between the sharp-*k* model and the data is greatly reduced for haloes defined by a spherical overdensity instead of a friends-of-friends linking criterion (see Watson et al. 2013, for a comparison of the two). We therefore do not use the correction model proposed by Schneider et al. (2013).

4.3 Conditional mass function

Another important application of the EPS model is the conditional mass function, which gives the abundance of haloes per mass and look-back redshift z_1, eventually ending up in a single host halo at redshift z_0. As the conditional mass function provides a connection between haloes at different redshifts, it acts as the starting point of more evolved quantities such as the halo collapse redshift, the number of satellites, and halo merger trees. The conditional mass function is given by

$$\frac{dN(M|M_0)}{d\ln M} = -\frac{M_0}{M} S f(\delta_c, S|\delta_{c,0}, S_0) \frac{d\ln S}{d\ln M} \qquad (13)$$

(Lacey & Cole 1993). For the sharp-*k* model this can be simplified to

$$\frac{dN_{SK}(M|M_0)}{d\ln M} = \frac{1}{6\pi^2} \frac{M_0}{M} f(\delta_c, S|\delta_{c,0}, S_0) \frac{P(1/R)}{R^3}, \qquad (14)$$

where the filter scale R and the mass M are related by equation (12).

The conditional first-crossing distribution again depends on the assumed model for non-linear collapse. The case of spherical collapse is given by

$$f(\delta_c, S|\delta_{c,0}, S_0) = \frac{(\delta_c - \delta_{c,0})}{\sqrt{2\pi(S - S_0)}} \exp\left[-\frac{(\delta_c - \delta_{c,0})^2}{2(S - S_0)}\right], \qquad (15)$$

which is a simple recalibration of equation (4), obtained by shifting the starting point of trajectories from (0, 0) to $(\delta_{c,0}, S_0)$ in the $(\delta_c - S)$ plane. In the case of ellipsoidal collapse, the threshold depends on the variance S and no simple recalibration is possible. As a consequence, every point in the $(\delta_c - S)$ plane requires an independent excursion-set calculation, and no general analytical expression exists for the conditional first-crossing distribution (Sheth & Tormen 1999). To keep it simple, we therefore use the spherical collapse model (i.e. equation 15) whenever we compute the conditional mass function in this paper.

In Fig. 5, the conditional mass function is plotted for the case of a $10^{13} \, h^{-1} \, M_\odot$ host and a look-back redshift of $z = 1.1$. The symbols correspond to the average number of progenitor haloes out of 84 host systems with masses close to $10^{13} \, h^{-1} \, M_\odot$ (upper symbols stand for the conservative and lower symbols for the more radical method to filter out artefacts, see Section 3.2 for further information). The solid and dotted lines are predictions from the sharp-*k* and the standard EPS approach, respectively. While the sharp-*k* model gives a good match to simulations of all DM scenarios, the standard EPS model works for CDM but fails for all scenarios with suppressed small-scale perturbations.
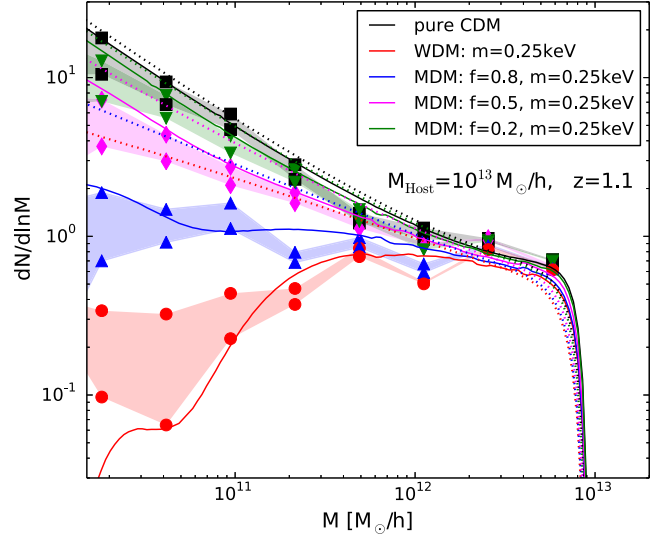
Figure 5. Conditional mass functions for a $M_0 = 10^{13} \, h^{-1} \, M_\odot$ host and a look-back redshift of $z = 1.1$. Coloured symbols refer to simulation outputs (with circumjacent shaded regions representing the uncertainty due to artefact subtraction), while the solid and dotted lines represent the sharp-*k* model and the standard Press–Schechter model, respectively. The colour-coding is the same as in the previous plots.

4.4 Estimating the number of dwarf satellites

Each DM scenario has to produce a sufficient amount of substructures to account for the observed MW satellites. While some (or most) of the substructures could be dark due to inefficient star formation, fewer substructures than observed means the failure of the DM scenario. In the case of WDM, comparing numbers of simulated substructures with observed satellites has led to tight constraints on the thermal particle mass ruling out masses below 2 keV (Polisensky & Ricotti 2011; Kennedy et al. 2014).

The EPS approach can be used to estimate the average number of dwarf galaxies orbiting a galaxy like the MW. This means it is possible to check whether a certain DM scenario is likely to be in agreement with observations without running expensive numerical zoom-simulations of an MW halo. In principle, finding the number of progenitors from an EPS approach consists of constructing the full merger-tree of the host-halo and counting all branches directly connected to the trunk. This is a rather cumbersome calculation which lies beyond the scope of this paper. Instead, we follow a simplified procedure presented in Giocoli, Pieri & Tormen (2008) and adapt it to the sharp-*k* model.

The total number of progenitors over all redshifts $N_{Sat}(\delta_c > \delta_{c,0})$ can be computed by integrating the conditional mass function over δ_c, i.e.

$$\frac{dN_{Sat}}{d\ln M} = \frac{1}{N_{norm}} \int_{\delta_{c,0}}^{\infty} \frac{dN_{SK}}{d\ln M} d\delta_c \qquad (16)$$

with $N_{norm} = 1$ for now. This integral overcounts the actual number of progenitors (and therefore the expected number of substructures) because it includes multiple counts of structures simultaneously existing at different redshifts. This issue can be accounted for by normalizing the integral so that equation (16) matches the outcome of ΛCDM simulations. We use data based on the Aquarius project of a simulated MW halo with mass $M_0 = 1.77 \times 10^{12} \, h^{-1} \, M_\odot$, which has 157 satellites of $M > 10^8 \, h^{-1} \, M_\odot$ (Lovell et al. 2014). Integrating equation (16) over the mass range between $10^8 \, h^{-1} \, M_\odot$

Table 2. Number of predicted MW satellites with mass above $10^8 \, h^{-1} \, M_\odot$ from the model (equation 17) and from simulations by Lovell et al. (2014). For the case of CDM the perfect agreement is by construction, since the model is calibrated to this number.

Scenario		N_{Sat} (EPS model)	N_{Sat} (Simulation)
WDM:	$m = 1.5$ keV	9	12
WDM:	$m = 2.0$ keV	25	27
WDM:	$m = 2.3$ keV	32	30
CDM:	Pure	158	158

and M_0 and equating it to the number of simulated satellites, leads to the normalization constant $N_{norm} = 44.5$. For the case of a sharp-k filter, equation (16) then becomes

$$\frac{dN_{Sat}}{d \ln M} = \frac{1}{44.5} \frac{1}{6\pi^2} \left(\frac{M_0}{M} \right) \frac{P(1/R)}{R^3 \sqrt{2\pi(S - S_0)}}. \tag{17}$$

This relation provides an estimate for the number of satellites per mass M of a host halo with mass M_0, where S and S_0 are the corresponding variances and $P(1/R)$ is the linear power spectrum at the inverse of the filter scale R. A further integration over M finally leads to the total amount of satellites for a DM scenario with arbitrary power spectrum.

The model is clearly an oversimplification and cannot capture the full complexity of hierarchical structure formation. First, it does not give the final subhalo mass prior to the merger with the host, but rather some average mass of the subhalo formation history. Secondly, it completely ignores tidal striping which reduces the mass of substructures and becomes significant as soon as a satellite passes close to the centre of the host.

In Table 2, we give the number of satellites with $M > 10^8 \, h^{-1} \, M_\odot$ predicted by the EPS model for different WDM scenarios. The numbers agree surprisingly well with satellite counts from WDM simulations of Lovell et al. (2014). The predicted number of satellites lies close to the outcome of simulations well within the expected statistical fluctuation between different host haloes. These results strongly suggests that despite its oversimplifications, the model (i.e equation 17) can be used to estimate the amount of dwarf satellites for scenarios with suppressed power spectra. It consists of a very simple and useful test to check if a DM scenario is likely to be ruled out by dwarf galaxy counts, or if it has the potential to account for potential *small-scale problems* of ΛCDM.

As an example, we can use the model developed above to constrain the MDM parameter space. There are 11 classical satellites plus additional 15 satellites recently observed by the Sloan Digital Sky Survey (SDSS) within the viral radius of the MW (Wolf et al. 2010). As SDSS covers only 28 per cent of the sky, the latter number needs to be multiplied by 3.5 to account for full sky coverage (assuming spherically distributed dwarfs). Altogether this leads to an estimate of $11 + 50 = 61$ dwarf satellites from observations (see Polisensky & Ricotti 2011, for more details about how to estimate the number of MW satellites).

The total mass of these objects are very difficult to infer, since the only information about the DM halo comes from stellar kinematics at around the half-light radius. Fitting different possible profiles and extrapolating to the viral radius, Brooks & Zolotov (2014) find that the masses of all classical and SDSS dwarfs should lie above $10^8 \, h^{-1} \, M_\odot$. We therefore use $M_{min} = 10^8 \, h^{-1} \, M_\odot$ as a lower limit for our analysis.
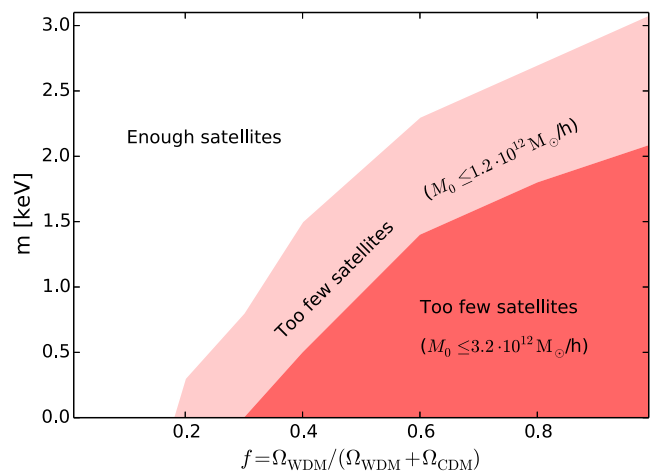

Figure 6. Exclusion plot of MDM based on satellite counts, assuming an MW mass of 1.2×10^{12} and $3.2 \times 10^{12} \, h^{-1} \, M_\odot$ (average and maximum estimates for the MW mass from Guo et al. 2010).

The predicted number of satellites is very sensitive to the host mass, which is uncertain by a factor of a few in the case of the MW. In our analysis, we take the bounds $5.5 \times 10^{11} < M_0 < 3.2 \times 10^{12}$ $h^{-1} \, M_\odot$ on the MW mass obtained by Guo et al. (2010).

Fig. 6 provides constraints on the mixed DM scenario obtained by combining the estimated satellite count with predictions from equation (17). The constraints depend on both the thermal mass (m) of the warm/hot DM part and the mixing fraction (f). As smaller host-haloes have fewer satellites, lower estimates of the MW mass translate into stronger constraints. We account for this by investigating exclusion limits for two MW masses, one corresponding to the mean value ($M_0 \le 1.2 \times 10^{12} \, h^{-1} \, M_\odot$, pink shaded region) and one to the maximum value ($M_0 \le 3.2 \times 10^{12} \, h^{-1} \, M_\odot$, red shaded region) given by Guo et al. (2010).

For the extreme case of $f = 1$ (corresponding to WDM), the constraints lie at 2–3 keV, which is roughly in agreement with estimates from Polisensky & Ricotti (2011) and Kennedy et al. (2014). On the other hand, MDM with fractions smaller than $f \sim 0.2$–0.3 cannot be constrained by satellite counts. A similar lower limit on f was found by Anderhalden et al. (2013) based on satellite counts in MDM simulations and by Boyarsky et al. (2009) based on data from the Lyman-α forest.

5 CONCENTRATION–MASS RELATION

The concentration of a halo is defined as the ratio between virial radius and scale radius, assuming a broken power-law profile, where the virial radius delimitates the system and the scale radius defines the break between the two power-law regimes. Usually an NFW profile (Navarro et al. 1997) is assumed, which has an inner slope of r^{-1} and an outer slope of r^{-3}. In a pure CDM scenario, halo concentrations are slightly mass-dependent, decreasing on average towards larger halo masses with an important scatter between individual haloes. In WDM scenarios, halo concentrations have been reported to increase, turnover, and decrease on average towards large masses (Avila-Reese et al. 2001; Bode, Ostriker & Turok 2001; Schneider et al. 2012).

The concentration–mass relation is a direct consequence of the connection between concentrations and the halo accretion history, as shown explicitly in numerical studies of pure CDM

cosmologies (Wechsler et al. 2002; Ludlow et al. 2014). In a universe governed by hierarchical clustering, small haloes collapse first while the average density is high what translates into an increased value for the concentration. This connection has been utilized in various analytically motivated models for the concentration–mass relation (Navarro et al. 1997; Bullock et al. 2001; Eke, Navarro & Steinmetz 2001).

Arguably, the most widely used model for the concentration–mass relation has been developed by Bullock et al. (2001) and assumes a concentration of

$$c(z) = K \frac{(1+z)}{(1+z_c)}, \qquad \langle D(z_c) \rangle_\chi = \frac{\delta_{c,0}^2}{S_\chi(FM)}, \qquad (18)$$

with the free parameters K and F. The collapse redshift z_c is derived from the second relation, where $\langle D(z_c) \rangle_\chi$ stands for the growth factor of collapse of an average perturbation, assuming a DM scenario χ. The Bullock-model approximately works for pure CDM, but it breaks down for the case of WDM where it predicts the concentration–mass relation to become constant instead of the observed downturn towards small masses (Eke et al. 2001; Schneider et al. 2012).

The reason for this failure comes from an inaccuracy in the estimation of the redshift of collapse. While the Bullock model uses the average collapse redshift of perturbations with mass M (i.e. the second relation of equation 18), what is actually required is the average collapse redshift of a halo that has collapsed beforehand and still exists today. This subtle difference of perspective becomes important for large masses ($FM > M_*$, with M_* being defined by the relation $S(M_*) = \delta_{c,0}^2$) where the Bullock model assigns negative collapse redshifts which is inherently contradictory. The inaccuracy of the Bullock model becomes more important for WDM, where haloes around the suppression scale collapse out of very shallow perturbations (that have been nearly entirely destroyed by free-streaming) and, therefore, only exist for a short time before they get accreted by a larger halo. As a result, small haloes which survive until today tend to have very low collapse redshifts. In the following, we derive a new relation for the average halo collapse redshift which takes this effect into account.

5.1 Halo collapse redshift

Halo collapse is not an instantaneous event but a lengthy processes of accretion and merging, and defining a formation time is therefore intrinsically ambiguous. The usual definition for the collapse redshift z_c is the moment when a halo has accreted a fraction F of its final mass M. Every halo has an individual accretion history and a different collapse redshift.

The distribution of collapse redshifts (g_χ) for all progenitors of one halo with mass M is simply given by the conditional first-crossing distribution

$$g_\chi(\delta_c) = f\left[\delta_c, S_\chi(FM) | \delta_{c,0}, S_\chi(M)\right], \qquad (19)$$

where $\delta_c = \delta_{c,0}/D(z_c)$ is the redshift-dependent collapse threshold of a given progenitor (with $\delta_{c,0} = 1.686$ in the spherical collapse model). Equation (19) is a well-behaved probability function with an average value

$$\langle \delta_c \rangle = \frac{\int d\delta_c \delta_c g_\chi(\delta_c)}{\int d\delta_c g_\chi(\delta_c)}. \qquad (20)$$
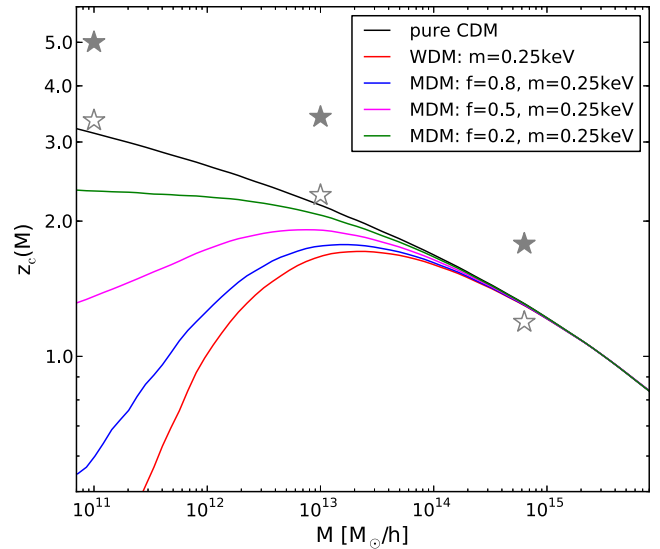

Figure 7. Redshift of collapse of haloes with mass M at redshift zero. The curves are drawn from equation (21) with $F = 0.05$. The grey stars are extracted from measured halo accretion rates in CDM simulations (Zhao et al. 2009). Filled stars correspond to the absolute values, empty stars are normalized to the black curve.

Solving this equation explicitly with the help of equation (20) yields a relation for the average growth-factor

$$\langle D(z_c) \rangle_\chi = \left[1 + \sqrt{\frac{\pi}{2}} \frac{1}{\delta_{c,0}} \sqrt{S_\chi(FM) - S_\chi(M)}\right]^{-1} \qquad (21)$$

(for a given DM scenario χ) from which it is straight forward to derive the corresponding redshift of collapse. Equation (21) is a clear improvement compared to equation (18) of the Bullock model, since it provides a reasonable value for the collapse redshift even in the case $FM > M_*$. Additionally, it allows the z_c–M relation (and therefore the concentration–mass relation as we will see in the next section) to turnover towards small masses, provided the function $S_\chi(M)$ becomes flat enough which is the case for WDM and some MDM scenarios.

The problem with the derivation above is that does not yield the collapse redshift of the main halo but rather an average collapse redshift of all its progenitors. In general, the true halo collapse redshift can only be determined by constructing the full EPS halo merger-tree, which lies beyond the scope of this work.[4]

In Fig. 7, the average collapse redshift from equation (21) is plotted against halo mass at redshift zero. Here, we used the fraction $F = 0.05$ and the sharp-k filter to calculate S_χ, but similar results are obtained with a top-hat filter if the fraction is changed to $F = 0.01$. For CDM (black line), z_c increases monotonically towards smaller masses, while for WDM (red line) and MDM (blue, magenta, and green line) it increases, turns over, and decreases again. Very small and very large WDM and MDM haloes are thus rather young, while medium-sized haloes are considerably older. It is important to note that this peculiar behaviour is no sign for

---

[4] Assuming the special case of $F \geq 0.5$, it is possible to compute the collapse distribution (and therefore the average collapse redshift) by integrating the conditional mass function (13) over the mass range [FM, M] followed by a derivative with respect to δ. This is, however, not a valuable option for the purpose of fitting concentrations, since F needs to act as a free parameter in order to reproduce the measurements from simulations.

the failure, but rather a natural prediction of hierarchical structure formation: suppressed perturbations can only collapse if they sit on top of larger perturbations, but this means that they only live for a short time before they get cannibalized by the collapsing larger perturbation.

As mentioned earlier, equation (21) provides the average collapse redshift of all progenitors and is therefore expected to underpredict the actual collapse redshift of the halo. In Fig. 7, we show that this is indeed the case: the average collapse redshift measured in CDM simulations (solid grey stars, obtained from Jiang & van den Bosch 2014, based on simulations from Zhao et al. 2009) is about a factor of 1.5 larger than the one predicted by our simplified model. It is, however, interesting to realize that the slope of the CDM z_c–mass relation described by equation (21) agrees well with the one from simulations. This is illustrated by the empty grey stars which are the same data normalized to the black line. In the next section, we will see that this relative agreement is sufficient to derive a physically motivated recipe for the concentration–mass relation.

5.2 Concentrations

Theory-based arguments can be used to understand the qualitative behaviour of the concentration–mass relation, but there is no consistent theoretical model that provides a prediction. It is therefore common practice to measure concentrations in simulations and to use fits for further investigations. Here, we present a simple method which is based on prior knowledge of pure CDM concentrations and predicts the concentration–mass relation for any scenario with suppressed small-scale perturbations.

Let us suppose, we know a function describing the concentration–mass relation for the CDM case. In this paper, we use two different fitting functions: the first corresponds to a simple power-law proposed by Macciò, Dutton & van den Bosch (2008),

$$c_{CDM}(M) = \alpha \left(\frac{10^{12} h^{-1} \, M_\odot}{M} \right)^{\gamma}, \tag{22}$$

the second includes an additional parameter $\mathcal{M}$ accounting for an increase at large masses

$$c_{CDM}(M) = \alpha \left(\frac{10^{12} h^{-1} \, M_\odot}{M} \right)^{\gamma} \left[1 + \left(\frac{M}{\mathcal{M}} \right)^{0.4} \right], \tag{23}$$

as described by Klypin et al. (2014). Based on one of these fits, it is possible to obtain a concentration–mass relation for any DM scenario χ by simply assuming that haloes with the same collapse redshifts end up having the same concentrations. In other words, we can use equation (21) to equate

$$\langle D \rangle_{CDM}(M_{CDM}) \equiv \langle D \rangle_{\chi}(M_{\chi}), \tag{24}$$

and determine a function $M_{CDM}(M_{\chi})$ relating the halo mass of any DM scenario χ to the corresponding halo mass of CDM. This function can then be used in either equation (22) or equation (23) to obtain the concentration–mass relation of scenario χ, i.e.

$$c_{\chi}(M_{\chi}) = c_{CDM}[M_{CDM}(M_{\chi})]. \tag{25}$$

The procedure explained here is physically motivated, as it simply assigns the same value for the concentration to haloes with the same collapse redshift, independently of the DM scenario.

In Fig. 8, we compare the concentrations from simulations (filled and empty symbols) to the model described by equation (25) utilizing the fitting functions from equation (22, dotted lines) and

equation (23, solid lines). The left-hand panel shows the concentrations of CDM (black) and WDM (cyan, purple) at redshift 0. While the concentration–mass relation of the CDM model exhibits the expected monotonic increase towards small masses, there is a clear turnover visible in the WDM scenarios. The model given by equation (25) is able to reproduce this behaviour for either of the underlying fitting functions (using the parameters $\alpha = 10.84$, $\gamma = 0.085$, and $\mathcal{M} = 5.5 \times 10^{17} \, h^{-1} \, M_\odot$).

The middle panel of Fig. 8 shows concentrations of CDM (black), WDM (red), and MDM (blue, magenta, green) at redshift 0. As expected, the CDM concentrations increase monotonically towards small masses, while the WDM concentrations exhibit a turnover. For the case of MDM, the concentration–mass relation is either flattened compared to CDM (green dots) or it turns over as in the WDM scenario (magenta and blue dots) depending on the mixing fraction. Additionally to this general trend, the red and blue dots (representing the most extreme cases in terms of suppression) exhibit a second upturn at small masses below $10^{12} \, h^{-1} \, M_\odot$. This feature can be understood by comparing the collapse redshifts: haloes with masses around this secondary upturn have very low collapse redshifts comparable to the ones from the largest haloes with masses above $10^{15} \, h^{-1} \, M_\odot$. As a result, their concentrations should coincide, and an increase of the concentration–mass relation towards very large masses directly translates into an upturn towards small masses around the suppression-scale. This is the reason why the model based on the fit with large-scale upturn (equation 23) is in good agreement with the simulations, while the model based on the simple power-law fit (equation 22) fails at the smallest mass scales.[5] For the fitting parameters we used $\alpha = 10.96$, $\gamma = 0.12$, and $\mathcal{M} = 5.5 \times 10^{17} \, h^{-1} \, M_\odot$ in the former and $\alpha = 10.96$ and $\gamma = 0.11$ in the latter case. These values are very close to results from Klypin et al. (2014) and Dutton & Macciò (2014), respectively, except for a small change of normalization due to different halo mass definitions.

In the right-hand panel of Fig. 8, the same models are plotted at redshift 1.1. As a general trend, the concentration–mass relations are flatter than at redshift 0 which is in agreement with former studies (Prada et al. 2012; Dutton & Macciò 2014). The secondary upturn towards small masses is, however, more prominent than at redshift 0. This is a direct consequence of the fact that the large-scale upturn of pure CDM concentrations shifts towards smaller masses if higher redshifts are considered (Klypin et al. 2014). In this panel, we used the parameters $\alpha = 5.9$, $\gamma = 0.1$, and $\mathcal{M} = 4.4 \times 10^{16} \, h^{-1} \, M_\odot$ for the fit with large-scale upturn and $\alpha = 6.1$, $\gamma = 0.085$ for the simple power-law fit.

In summary, the model based on equation (25) is able to reproduce the concentration–mass relation of scenarios with various small-scale power suppressions provided the concentrations of pure CDM are well known. In agreement with previous papers, we observe that concentrations seem to be tightly connected to the halo collapse redshifts. In the presence of a strong suppression of perturbations (such as in WDM or some MDM scenarios), this may lead to a wave-like feature in the concentration–mass relation, reflecting the low collapse redshifts of haloes around the suppression scale.

---

[5] The origin of the upturn of concentrations towards very high halo masses is disputed: while Ludlow et al. (2014) argue that the largest haloes are not fully virialized, Klypin et al. (2014) claim that many large haloes grow out of unusually spherical perturbations, leading to a boost of the concentrations.
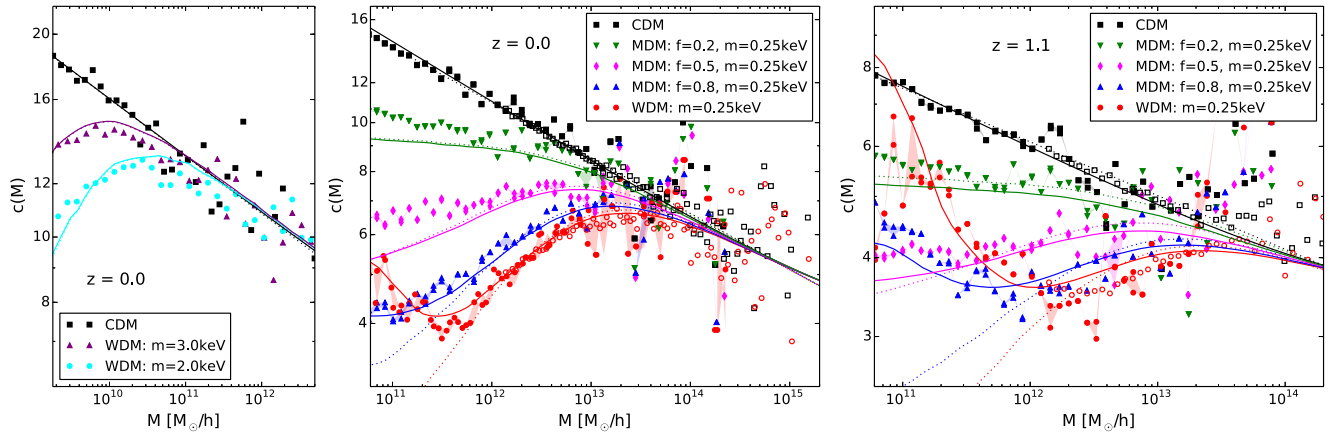
Figure 8. Concentration–mass relation for CDM (black), different WDM (red, cyan, purple) and different MDM models (blue, magenta, green) with circumjacent shaded regions representing the uncertainty due to artefact subtraction. Symbols illustrate the measurements from simulations, solid, and dotted lines correspond to the models presented in the text (based on CDM fitting functions with and without upturn on cluster scales).

6 CONCLUSIONS

The standard model of cosmology predicts a suppression of perturbations below a characteristic scale depending on the DM candidate. For the case of DM consisting of WIMP, the suppression scale is orders of magnitudes below the range of current astronomical observations, and only a potential detection of the WIMP annihilation signal could eventually change this. For alternative DM candidates, such as the sterile neutrino, the suppression scale is expected to be large enough to have an effect on current galaxy observations. In order to distinguish between different DM scenarios, it is therefore crucial to understand non-linear structure formation around the scale of suppressed perturbations.

In this paper, we study structure formation of various suppressed initial power spectra using both numerical simulations and analytical techniques. The simulations are deliberately chosen to cover a variety of different scenarios from steep cutoffs to shallow suppressions occurring on a wide range of scales. We study warm and WIMP DM both exhibiting steep cutoffs as well as mixed DM with much shallower suppressions depending on the mixing fraction.

Many of the simulations performed here suffer from artificial clumping, a well-known problem in WDM simulations (Götz & Sommer-Larson 2003; Wang & White 2007; Elahi et al. 2014; Reed et al. 2014). We therefore apply a post-processing scheme similar to the one presented by Lovell et al. (2014), which filters out artefacts based on both protohalo sphericity and overlap between two simulations with the same density field but different resolution.

Along with the simulations, we develop analytically motivated models for the halo and subhalo mass functions, the halo collapse redshift, and the concentration–mass relation which are designed to work for arbitrary small-scale suppressions. The main findings of the paper are summarized in the following list.

(i) We further investigate the EPS model based on the sharp-*k* filter function introduced by Schneider et al. (2013). We show that it provides a simple prescription of the halo mass function, matching the simulation outcomes for all models studied here. This includes various warm DM, mixed DM, WIMP-DM, and pure cold DM scenarios, covering suppression scales at very different scales and with a variety of different shapes. It is therefore fair to assume that the sharp-*k* model can be used to predict the halo abundance for cosmologies with arbitrary small-scale suppressions.

(ii) We present a method to estimate the number of substructures in a host halo, and show that it agrees very well with warm DM simulations from Lovell et al. (2014). Based on this, we provide approximative constraints on mixed DM models by comparing the expected number of substructures to the observed number of MW satellites.

(iii) Using the conditional mass function, we determine the average collapse redshift of progenitors. Normalizing this relation to measurements from CDM simulations yields a simple estimation for the collapse redshifts of a halo in any DM scenario. Well above the suppression scale, smaller haloes tend to be older than larger ones. This is not the case around the suppression scale where haloes are very young on average. We argue that, despite this turnaround in the z_c–mass relation for models with suppressed perturbations, structure formation remains a hierarchical process in the sense that there is now sign of halo-formation via fragmentation.

(iv) Based on the collapse redshifts of haloes, we develop a recipe for the concentration–mass relation in the presence of arbitrary suppressed small-scale perturbations and show that it agrees surprisingly well with simulations. For the case of warm DM and some mixed DM scenarios, the concentration–mass relation decreases towards smaller masses when approaching the suppression scale. At even smaller scales, however, the relation grows again, an effect tightly connected to the recently observed increase of concentrations at the largest scales in pure CDM simulations (Prada et al. 2012; Klypin et al. 2014).

The findings presented in this paper are potentially useful for analytical and semi-analytical modelling of the non-linear Universe. In the presence of suppressed small-scale perturbations, analytical methods are particularly justified because numerical simulations notoriously suffer from artefacts and need to be analysed with great care.

REFERENCES

Abazajian K. N., 2014, Phys. Rev. Lett. 112, 161303

Abel T., Hahn O., Kaehler R., 2012, MNRAS, 427, 61

Agarwal S., Corasaniti P.-S., 2015, preprint (arXiv:1503.03503)

Agarwal S., Corasaniti P.-S., Das S., Rasera Y., 2014, preprint (arXiv:1412.1103)

Anderhalden D., Schneider A., Macciò A. V., Diemand J., Bertone G., 2013, J. Cosmol. Astropart. Phys., 03, 014

Angulo R. E., Hahn O., Abel T., 2013, MNRAS, 434, 3337

Avila-Reese V., Colín P., Valenzuela O., D'Onghia E., Firmani C., 2001, ApJ, 559, 516

Barkana R., Haiman Z., Ostriker J. P., 2001, ApJ, 558, 482

Benson A. J. et al., 2013, MNRAS, 428, 1774

Bertone G., Hooper D., Silk J., 2005, Phys. Rep., 405, 279

Bertschinger E., 2006, Phys. Rev. D, 74, 063509

Blas D., Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 07, 034

Bode P., Ostriker J. P., Turok N., 2001, ApJ., 556, 93

Boehm C., Schaeffer R., 2005, A&A, 438, 419

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Boyarsky A., Lesgourgues J., Ruchayskiy O., Viel M., 2009, J. Cosmol. Astropart. Phys., 5, 12

Boyarsky A., Ruchayskiy O., Iakubovskyi D., Franse J., 2014, Phys. Rev. Lett., 113, 251301

Brooks A. M., Zolotov A., 2014, ApJ, 786, 87

Bulbul E., Markevitch M., Foster A., Smith K. R. M., Loewenstein M., Randall S. W., 2014, ApJ. 789, 13

Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559

Bïhm C., Schewtschenko J. A., Wilkinson R. J., Baugh C. M., Pascoli S., 2014, MNRAS, 445, 31L

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Cyr-Racine F.-Y., Sigurdson K., 2013, Phys. Rev. D, 87, 103515

Diemand J., Moore B., Stadel J., 2005, Nature, 433, 389

Drewes M., 2013, Int. J. Mod. Phys. E, 22, 1330019

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Eke V. R., Navarro J. F., Steinmetz M., 2001, ApJ, 554, 114

Elahi P. J., Mahdi H. S., Power C., Lewis G. F., 2014, MNRAS, 444, 2333

Fogli G. L., Lisi E., Marrone A., Montanino D., Palazzo A., Rotunno A. M., 2012, Phys. Rev. D, 86, 013012

Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399

Giocoli C., Pieri L., Tormen G., 2008, MNRAS, 387, 689

Götz M., Sommer-Larsen J., 2003, Ap&SS, 284, 341

Green A. M., Hofmann S., Schwarz D. J., 2005, J. Cosmol. Astropart. Phys., 08, 003

Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hahn O., Paranjape A., 2014, MNRAS, 438, 878

Iakubovskyi D., 2014, Adv. Astron. Space Phys., 4, 9

Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193

Kamionkowski M., Liddle A. R., 2000, Phys. Rev. Lett., 84, 4525

Kaplinghat M., 2005, Phys. Rev. D, 72, 063510

Kennedy R., Frenk C., Cole S., Benson A., 2014, MNRAS 442, 2487

Klypin A., Yepes G., Gottlober S., Prada F., Hess S., 2014, preprint (arXiv:1411.4001)

Knebve A. et al., 2011, MNRAS, 415, 2293

Knollmann S. R., Knebe A., 2009, ApJ, 182, 608

Kosowsky A. Turner M. S., 1995, Phys. Rev. D, 52, 1739

Lacey C., Cole S., 1993, MNRAS, 262, 627L

Lesgourgues J., Tram T., 2011, J. Cosmol. Astropart. Phys., 09, 032

Lovell M., Frenk C., Eke V., Gao L., Jenkins A., Theuns T., 2014, MNRAS, 439, 300

Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, MNRAS, 441, 378L

Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940

Marsh D., Silk J., 2014, MNRAS, 437, 2652

Merle A., Schneider A., 2014, preprint (arXiv:1409.6311)

Merle A., Totzauer M., 2015, preprint (arXiv:1502.01011)

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Petraki K., Volkas R. R., 2013, Int. J. Mod. Phys. A, 28, 1330028

Planck Collaboration XVI, 2014, A&A, 571, A16

Polisensky E., Ricotti M., 2011, Phys. Rev. D, 83, 043506

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Press W. H., Schechter P., 1974, ApJ, 187, 425

Profumo S., Sigurdson K., Kamionkowski M., 2005, Phys. Rev. Lett., 97, 1301

Reed D., Schneider A., Smith R. E., Potter D., Stadel J., Moore B., 2014, preprint (arXiv:1410.1541)

Schneider A., Smith R. E., Macciò A. V, Moore B., 2012, MNRAS, 424, 684

Schneider A., Smith R. E., Reed D., 2013, MNRAS, 433, 1573

Schneider A., Anderhalden D., Macciò A. V., Diemand J., 2014, MNRAS, 441, 6L

Sheht R., Tormen G., 1999, MNRAS, 308, 119

Shi X., Fuller G. M., 1999, Phys. Rev. L, 82, 2832

Spergel D. N., Steinhardt P. J., 2000, Phys. Rev. Lett., 84, 3760

Srisawat C. et al., 2013, MNRAS, 436, 150

Stadel J. G., 2001, PhD thesis, Univ. Washington

Viel M., Haehnelt M. G., Springel V., 2010, J. Cosmol. Astropart. Phys., 06, 015

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G. A., 2013, Phys. Rev. D, 88, 3502

Wang J., White S. D. M., 2007, MNRAS, 380, 93

Wang M.-Y., Peter A. H. G., Strigari L. E., Zentner A. R., Arant B., Garrison-Kimmel S., Rocha M., 2014, MNRAS, 445, 614

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, ApJ, 568, 52

Weinberg D., Bullock J., Governato F., Kuzio de Naray R., Peter A. H. G., 2013, preprint (arXiv:1306.0913)

Wilkinson R. J., Lesgourgues J., Boehm C., 2014a, J. Cosmol. Astropart. Phys., 04, 026

Wilkinson R. J., Boehm C., Lesgourgues J., 2014b, J. Cosmol. Astropart. Phys., 05, 011

Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Munoz R. R., Simon J. D., Avedo F. F., 2010, MNRAS, 183, 1220

Zentner A. R., 2007, Int. J. Mod. Phys. D, 16, 763

Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, ApJ, 707, 354

APPENDIX A: SOME MORE ABOUT REMOVING ARTEFACTS

In the main text, we have discussed how the removal of artefacts affects the halo samples of the CDM, WDM ($m = 0.25$ keV) and MDM simulations at redshift zero. Here, we focus on the remaining DM models studied in this paper, providing the details of the CDM and WDM ($m = 2, 3, 4$ keV) simulations at redshift 5 as well as the pure CDM and WIMP DM models at redshift 23.

The distinction between real haloes and artefacts is somewhat easier towards high redshifts, since most of the haloes are in a less evolved state. This allows us to choose higher values for the sphericity parameters plus a less restrictive minimal number of 500 particles per halo.

For the CDM and WDM models (with $m = 2, 3, 4$ keV) at redshift 5, we use $s_c = 0.3$ and $q_c = 0.4$ for the *sphericity cut*, which are values that filter out a maximum of WDM artefacts while removing less than 5 per cent of all CDM haloes. The distributions of sphericity parameters s (blue) and q (red) are shown in the top panels of Fig. A1. The distributions of the WDM model with $m = 2$ keV is
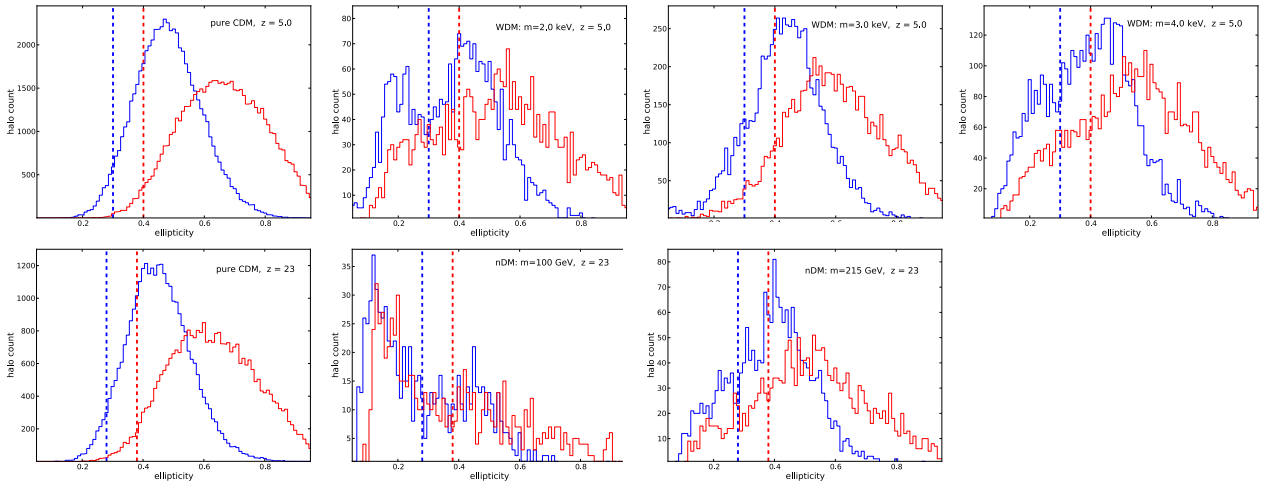
Figure A1. Histograms of the halo ellipticity parameters s (blue) and q (red) not shown in the main text. Top: CDM and WDM with $m = 2, 3, 4$ keV (from left to right) at redshift 5. Bottom: pure CDM and WIMP DM with neutrino masses of $m = 100, 215$ GeV (from left to right) at redshift 23. Vertical dashed lines illustrate the sphericity cut.
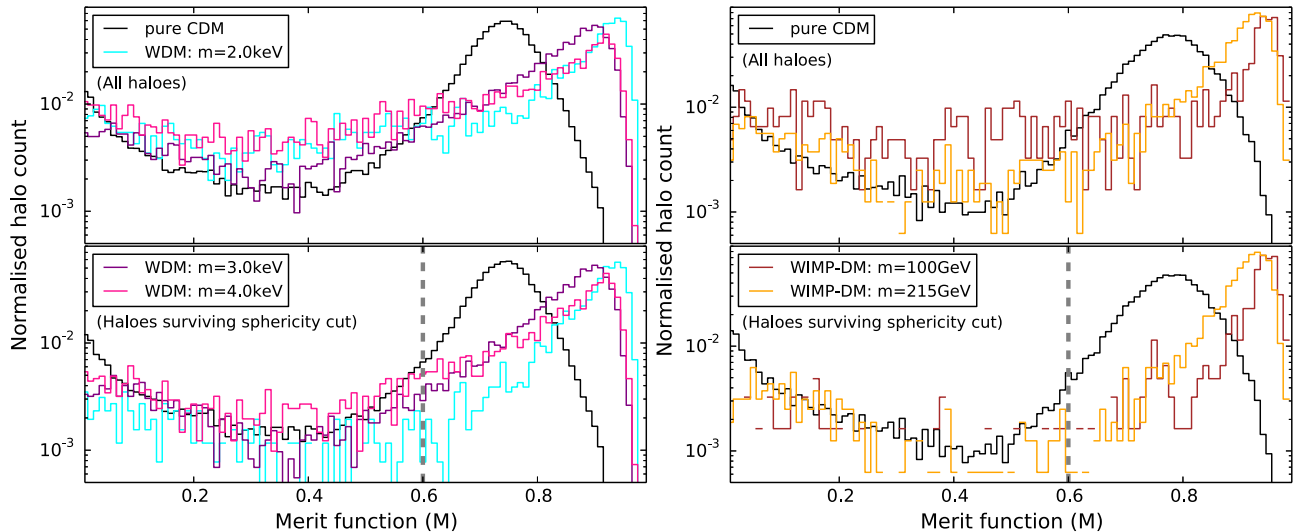

Figure A2. Histograms of the merit function M (i.e. equation 2) of all haloes before (top) and after (bottom) the sphericity cut. Left: CDM (black) and WDM with $m = 2, 3, 4$ keV (cyan, magenta, pink) at redshift 5. Right: pure CDM (black) and WIMP-DM with neutrino masses of $m = 100, 215$ GeV (brown, orange) at redshift 23. Vertical dashed lines correspond to the resolution cut.

characterized by two peaks, while the other WDM models show broadened distributions compared to CDM, indicating the presence of artificial haloes.[6] The distribution of the merit function (quantifying the overlap of haloes from different resolution) is illustrated in the left-hand side of Fig. A2. The top panel includes all haloes (before the *sphericity cut*), clearly showing the excess of artefacts in the WDM simulations at $M < 0.6$. This excess is greatly reduced after the *sphericity cut*, as illustrated in the bottom panel of Fig. A2. Further artefacts are removed with the *resolution cut* $M_c = 0.6$, illustrated as vertical dashed line in the lower panel of Fig. A2.

For the models of pure CDM and WIMP DM (with neutralino mass of $m = 100$ and $m = 215$ GeV) at redshift 23, we apply a

sphericity cut of $s_c = 0.28$ and $q_c = 0.38$, which again removes a maximum of artefacts while affecting pure CDM by less than 5 per cent. The sphericity distributions of these DM models are presented in the bottom panels of Fig. A1, showing the usual picture of a Gaussian distribution for pure CDM and much broader, occasionally double-peaked distributions for WIMP DM. The distributions of the merit function are illustrated on the right-hand side of Fig. A2, the top and bottom panel showing the cases before and after the *sphericity cut*. We again observe an excess of haloes in the WIMP DM models at low M with respect to pure CDM which is greatly reduced once the *sphericity cut* is applied. The *resolution cut* $M_c = 0.6$ is illustrated as vertical dashed line in the bottom panel of Fig. A2.

---

[6] The WDM model with $m = 4$ keV is run with a smaller box size than the other WDM models of the series, a fact that leads to a further broadening of the histogram.

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