ETHOS – An effective parametrization and classification for structure formation: the non-linear regime at $z \gtrsim 5$

Sebastian Bohr$^1$*, Jesús Zavala$^1$, Francis-Yan Cyr-Racine$^2$, Mark Vogelsberger$^3$, Torsten Bringmann$^4$ and Christoph Pfrommer$^5$

$^1$Centre for Astrophysics and Cosmology, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland
$^2$Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
$^3$Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
$^4$Department of Physics, University of Oslo, Box 1048, N-0315 Oslo, Norway
$^5$Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We propose two effective parameters that fully characterise galactic-scale structure formation at high redshifts ($z \gtrsim 5$) for a variety of dark matter (DM) models that have a primordial cutoff in the matter power spectrum. Our description is within the recently proposed ETHOS framework and includes standard thermal Warm DM (WDM) and models with dark acoustic oscillations (DAOs). To define and explore this parameter space, we use high-redshift zoom-in simulations that cover a wide range of non-linear scales from those where DM should behave as CDM ($k \sim 10\, h\, \text{Mpc}^{-1}$), down to those characterised by the onset of galaxy formation ($k \sim 500\, h\, \text{Mpc}^{-1}$). We show that the two physically motivated parameters $h_{\text{peak}}$ and $k_{\text{peak}}$, the amplitude and scale of the first DAO peak, respectively, are sufficient to parametrize the linear matter power spectrum and classify the DM models as belonging to effective non-linear structure formation regions. These are defined by their relative departure from Cold DM ($k_{\text{peak}} \to \infty$) and WDM ($h_{\text{peak}} = 0$) according to the non-linear matter power spectrum and halo mass function. We identify a region where the DAOs still leave a distinct signature from WDM down to $z = 5$, while a large part of the DAO parameter space is shown to be degenerate with WDM. Our framework can then be used to seamlessly connect a broad class of particle DM models to their structure formation properties at high redshift without the need of additional $N$-body simulations.

Key words: cosmology: dark matter – galaxies: haloes – methods: numerical

1 INTRODUCTION

Dark matter (DM) is a crucial ingredient in the formation of structures in the Universe as it makes up the majority of its matter content. Although the most likely explanation for DM is the particle hypothesis, its specific nature remains a mystery. The CDM model of structure formation has now emerged as the standard paradigm, and it has been shown to be consistent with the observed large scale structure of the Universe (e.g. Springel et al. 2005). At smaller (galactic) scales however, the CDM model has faced a number of significant challenges over the last decades: (i) the under-abundance of low-mass galaxies (either satellites or in the field) (Klypin et al. 1999; Moore et al. 1999; Zavala et al. 2009; Papastergis et al. 2011; Klypin et al. 2015), (ii) the core-cusp problem in low-surface brightness galaxies and possibly in dwarf spheroidals (de Blok & McGaugh 1997; Walker & Peñarrubia 2011), (iii) the "too-big-to-fail problem" (Boylan-Kolchin et al. 2011; Papastergis et al. 2015), (iv) the plane of satellites problem (Pawlowski et al. 2013), and (v) the diversity problem of rotation curves in dwarf galaxies (Oman et al. 2015). We note that with recent observations of ultra-faint galaxies, the too-big-to-fail problem becomes a diversity problem as well for the broad distribution of stellar kinematics in dwarf spheroidals in the Milky Way (Zavala et al. 2019).

There is a long history of attempts to provide a satisfactory solution to these issues based on either: (i) incompleteness, biases and systematic uncertainties in observations (e.g. Koposov et al. 2008; Kim et al. 2018, for the “missing satellites problem”), (ii) invoking a strong influence of uncertain baryonic physics in dwarf galaxies (e.g. impulsive super-
Figure 1. Diagram illustrating the different sets of parameters that characterise a given DM particle model and the connections between them within the ETHOS framework, both in Cyr-Racine et al. (2016) and in this work. The particle physics space parameters such as the DM particle mass $m_\chi$, coupling constants $g_i$ (e.g. DM-DR), internal parameters $h_i$ such as the mediator mass and degrees of freedom and the present day DR to CDM temperature ratio $\xi$, were mapped in Cyr-Racine et al. (2016) into effective parameters fully describing the linear DM power spectrum (see section 3.1). In this work, we make a re-parametrisation, defining new ETHOS parameters that have both a more straightforward interpretation in terms of the linear power spectrum and a clearer physical interpretation (amplitude $h_{\text{peak}}$ and scale $k_{\text{peak}}$ of the first DAO peak, amplitude of the second peak $h_2$, and damping of higher order peaks $\tau$; see section 3). The redefined ETHOS parameter space can be connected naturally to that defined in section 3.1, and thus to the particle physics space. Crucially, it is also sufficient to characterise non-linear structure formation for a variety of relevant DM models (such as WDM and models with DAOs) in the high-redshift Universe.

Figure 1. Diagram illustrating the different sets of parameters that characterise a given DM particle model and the connections between them within the ETHOS framework, both in Cyr-Racine et al. (2016) and in this work. The particle physics space parameters such as the DM particle mass $m_\chi$, coupling constants $g_i$ (e.g. DM-DR), internal parameters $h_i$ such as the mediator mass and degrees of freedom and the present day DR to CDM temperature ratio $\xi$, were mapped in Cyr-Racine et al. (2016) into effective parameters fully describing the linear DM power spectrum (see section 3.1). In this work, we make a re-parametrisation, defining new ETHOS parameters that have both a more straightforward interpretation in terms of the linear power spectrum and a clearer physical interpretation (amplitude $h_{\text{peak}}$ and scale $k_{\text{peak}}$ of the first DAO peak, amplitude of the second peak $h_2$, and damping of higher order peaks $\tau$; see section 3). The redefined ETHOS parameter space can be connected naturally to that defined in section 3.1, and thus to the particle physics space. Crucially, it is also sufficient to characterise non-linear structure formation for a variety of relevant DM models (such as WDM and models with DAOs) in the high-redshift Universe.

Density perturbations due to either collisionless damping effects at galactic scales if DM is made of extremely light bosons with $\lambda >> k$ de Broglie wavelength (fuzzy DM; for a review see Bose et al. 2016) or quantum effects at galactic scales if DM is made of extremely light bosons with $\mathcal{O}(1 \text{ kpc})$ de Broglie wavelength (fuzzy DM; for a review see Hui et al. 2017), also giving rise to extended DM cores (Robles & Matus 2012; Mocz et al. 2017).

Whether the CDM challenges are due to missing DM cores (Pontzen & Governato 2012), tidal effects from the Milky-Way disk to alleviate the too-big-to-fail problem (Garrison-Kimmel et al. 2019), and suppression of galaxy formation at the dwarf mass scale due to cosmic reionisation (Gnedin 2000; Sawala et al. 2016) to explain the underabundance of low-mass galaxies; and (iii) additional DM physics, i.e. departures from the CDM hypothesis such as: Warm Dark Matter (WDM; for a review see Adhikari et al. 2017) where the relativistic motion of the DM particles in the early Universe reduces the abundance and inner DM densities of galactic-scale haloes relative to CDM (e.g. Colín et al. 2000; Lovell et al. 2012); self-interacting DM (SIDM; for a review see Tulin & Yu 2018) where DM particles have strong self-interactions redistributing energy in the centre of haloes, thus resulting in DM cores (e.g. Spergel & Steinhardt 2000; Vogelsberger et al. 2012); and quantum effects at galactic scales if DM is made of extremely light bosons with $\mathcal{O}(1 \text{ kpc})$ de Broglie wavelength (fuzzy DM; for a review see Hui et al. 2017), also giving rise to extended DM cores (Robles & Matus 2012; Mocz et al. 2017).

We note that previous works have proposed analytical formulae to describe the linear power spectra of different DM models, usually written as a transfer function relative to CDM (e.g. for WDM Bode et al. 2001; Viel et al. 2005). More recently, Murgia et al. (2017) proposed a formula for the transfer function that can seemingly accommodate WDM, fuzzy DM, and also certain ETHOS models. Crucially however, this formula does not describe DAOs since they were deemed not relevant for the properties of interest in Murgia et al. (2017), namely, for the 1D Lyman–$\alpha$ flux power spectrum, and for the number of observable Milky-Way subhaloes (i.e. those that can host a luminous satellite). As we demonstrate and quantify in this work, DAOs are quite relevant for a range of ETHOS models (see also Bose et al. 2016).
Moreover, our approach differs from previous ones since the parametrization we propose goes beyond providing a fit to the power spectrum, with the parameters having a clear physical interpretation.

In this work we study structure formation within ETHOS in the high-redshift Universe down to $z=5$. This choice was partly done to avoid entering the regime where DM self-interactions (another relevant ingredient in ETHOS) start to have a relevant impact in the centre of DM haloes. We are also only considering the impact of DM physics in structure formation without taking into account the role of baryonic physics, which clearly plays a role in DM clustering, albeit considerably smaller at high-redshift relative to the low-redshift Universe. In this way we can isolate the potential difference between CDM and other DM models, purely due to DM physics, without the influence of baryons; this is in fact needed to disentangle the impact of both effects. Our plan in future work is to extend the spirit of this work, by defining the space of structure formation parameters that are relevant for the physics of galaxies, to include both SIDM and baryonic physics. We also choose $z=5$ as the lowest redshift we examine since it is roughly the maximum redshift where data from the Lyman-α flux-power spectra have been used to constrain the DM power spectrum at small scales (e.g. Viel et al. 2013). We use this both to exemplify how our parametrization can be used to potentially constrain DM models and to define the maximum scale where new DM physics can play a role in galactic-scale structure formation: DM models with a non-linear power spectrum significantly deviating from CDM at $k \lesssim 10 h$ Mpc$^{-1}$ are not compatible with the data (Iršič et al. 2017; although see Garzilli et al. 2019). On the other hand, we set the relevant minimum scale to be given by the atomic cooling limit (specifically, the primordial gas in haloes with a virial temperature $\lesssim 10^4$ K cannot cool via atomic transitions; see White & Rees 1978). Galaxy formation is thus suppressed for DM haloes with masses below $\sim 10^8 M_\odot h^{-1}$ (corresponding to non-linear scales of $\sim 500$ h Mpc$^{-1}$). In summary, we study non-linear structure formation down to $z=5$ within the ETHOS framework using DM-only N-body simulations focusing on the non-linear scale range $10 h$ Mpc$^{-1} \lesssim k \lesssim 500 h$ Mpc$^{-1}$ (halo virial masses in the range $10^8 M_\odot \lesssim M_{vir} \lesssim 10^{10} M_\odot$).

This paper is organized as follows. In Section 2, we describe our simulation setup and the zoom-in method we use to cover the dynamic range of interest. The convergence properties of our simulations is discussed in Appendix A. In Section 3, the new ETHOS parametrization is constructed and connected to that in Cyr-Racine et al. (2016) (see also Appendix B). In Section 4, we present our main results on how different structure formation models are classified within the new parametrization based on both the non-linear power spectrum and the halo mass function. Finally, our conclusions are given in Section 5.

## 2 NUMERICAL METHODOLOGY

The cosmological dark-matter-only $N$-body simulations used in this work were performed with the code AREPO (Springel 2010). Initial conditions for the simulations were generated using the code MUSIC (Hahn & Abel 2011) with the cosmological parameters set to $\Omega_m = 0.31069$, $\Omega_{\Lambda} = 0.68931$, $H_0 = 67.5$ km/s/Mpc, $n_s = 0.9653$ and $\sigma_8 = 0.815$, where $\Omega_m$ and $\Omega_{\Lambda}$ are the contributions from matter and cosmological constant to the matter-energy density of the Universe today, respectively, $H_0$ is the Hubble constant today, $n_s$ is the spectral index, and $\sigma_8$ is the mass variance of linear fluctuations in 8 Mpc $h^{-1}$ spheres at $z = 0$. This choice of cosmological parameters is consistent with a Planck cosmology (Planck Collaboration et al. 2018). The linear power spectrum used as an input for MUSIC for the different DM models we explore is computed with a modified version of CLASS$^1$ (Archidiacono et al. 2017, 2019)$^2$.

For this work, we are interested in the matter power spectrum at the non-linear scales relevant for dwarf galaxies at high redshift from $k \sim 10$ to $500 h$ Mpc$^{-1}$. This range roughly corresponds to halo masses $\sim 10^{10}$ to $10^7 M_\odot$, which are close to the limits where significant deviations from CDM are possible, and galaxy formation becomes highly inefficient, respectively. Achieving a fair representation of the power spectrum and the halo mass function at such small scales was not feasible with a uniform simulation box due to the following stringent limitation. At a fixed spatial resolution, reducing the size of the cosmological box, reduces the minimal scales probed, but at the cost of missing the power transferred from larger to smaller scales in the non-linear evolution. Thus, the power spectrum at the scales and redshifts of interest would be biased towards lower values. This problem can be alleviated by having a sufficiently large simulation box, but in order to resolve $500 h$ Mpc$^{-1}$, the amount of particles and thus the computational cost increases dramatically. To achieve our goals we therefore rely instead on cosmological zoom-in simulations by using the method described in the following.

### 2.1 Small-scale power spectrum with zoom simulations

In a zoom-in cosmological simulation, the computational resources are focused on a smaller subregion within a large cosmological box. This subregion is simulated at the desired highest resolution, while the volume around contains low resolution elements that still preserve an accurate representation of large scale properties of the density field. The region of interest is usually a halo and its immediate environment, and the procedure to construct the initial conditions for zoom simulations consists of: i) run a low resolution parent cosmological simulation within a cosmological box large enough to provide a fair representation of the clustering properties at the scales of the box in the range of redshifts of interest (we found that a box size of 40 Mpc $h^{-1}$ per side satisfies this for $z \geq 5$); ii) select a volume within the parent simulation encompassing the region of interest at the redshift of interest; iii) trace back the particles within this region to the starting redshift of the resimulation; this represents the Lagrangian volume for the zoom simulation; (iv) finally, an initial conditions code like MUSIC is used to

---

1. Blas et al. (2011) (class-code.net)
2. Note that in the first ETHOS paper (Cyr-Racine et al. 2016), this implementation is done with CAMB (Lewis & Bridle 2002).
generate a multi-layered resolution coverage of the resimulation specifying the volume that covers this Lagrangian region with the highest resolution required. For more details of the general procedure see Östroe et al. (2014).

In our case we followed the previous standard zoom-in procedure but not focusing on a particular halo in the parent simulation but rather on a smaller subregion with the main requirement for it to have a similar power spectrum to the larger parent box at the resolved scales. In the following we describe how we find the optimal sub-region according to this requirement.

We find that at the scales of interest the power spectrum has roughly a similar shape, but with an amplitude that correlates strongly with the overdensity $\delta_{\text{sub}}$ of the subregion, as can be seen in Fig. 2. This type of cosmic variance is a well known effect that has been studied in the past, particularly in the linear regime where $\delta_{\text{sub}} \ll 1$. In this regime, it is possible to correct for this bias by e.g. using the separate universe approach, where each subregion is treated as a separate universe with a different cosmology, in this case a universe with different background density (see e.g. Chiang et al. 2014; Li et al. 2014). As is clear from the bottom panel of Fig. 2, the distribution of $\delta_{\text{sub}}$ is broad, covering values that are clearly non-linear anymore. This is because we are looking at smaller scales where the impact of non-linear effects is stronger and the variance of $\Delta^2$ for regions with the same overdensity can be quite large, weakening the correlation between the amplitude of the power spectrum and the overdensity. Instead of trying to generalize the separate universe approach into the non-linear case, we decided to carefully select our high resolution sub-region so that it has a power spectrum that is as similar as possible to the one of the larger lower resolution region, at the scales that both can resolve. In this way a correction becomes unnecessary. This is sufficient for our purposes since we are only interested in an average measure of the power spectrum down to small (galactic) scales, rather than in its variance. Nevertheless, Fig. 2 gives an impression of the (cosmic) variance to be expected in the power spectrum for survey volumes that are small $\lesssim 10$ Mpc.

### 2.2 Performance of the zoom-in simulation technique

As a benchmark test for the reconstruction ability and resource advantage of the method described above, we performed four CDM simulations in a $(40\ Mpc h^{-1})^3$ volume down to $z = 5$. The baseline is a uniform simulation with $1024^3$ particles, while the other three are zoom simulations where the low-resolution region corresponds to $512^3$ particles. The first of these has a $(12.5\ Mpc h^{-1})^3$ zoom region with an effective resolution of $1024^3$ particles, the second one has a $(6.25\ Mpc h^{-1})^3$ zoom region with an effective resolution of $2048^3$ particles, and the third one has a $(6.25\ Mpc h^{-1})^3$ zoom region with an effective resolution of $4096^3$ particles.

In Fig. 3, we can see that all three zoom simulations give a good reconstruction of the baseline power spectrum at large scales. As expected, the one with the same effective resolution as the baseline ($1024^3$ particles; red line) shows almost the same power spectrum as the uniform one at all scales, with nearly the same level of Poisson noise. The other two zoom simulations can resolve the power spectrum at smaller scales by factors of 4 (green) and 8 (blue) relative to the uniform simulation. This test shows that we can measure the power spectrum across a large dynamical range using the zoom simulation technique described in Section 2.1. More importantly, it is possible to achieve this with a reduced computational cost as we show in Table 1. For instance, the zoom simulation with same effective resolution as the uniform one (red and black lines in Fig. 3) uses only a small fraction ($\lesssim 1/7$) of the core hours and less than half the memory of the uniform simulation. Even our highest resolution zoom simulation uses about the same core-hours and memory compared to the uniform run, while improving the scales that can be probed by a factor of 8, making it a very
our goal in this Section is to present a new parametrization of the linear power spectrum for DM models that have a primordial power spectrum cutoff with or without DAOs within the ETHOS framework. This parametrization is purely phenomenological but it is constructed with two objectives in mind: (i) although it parametrizes the linear power spectrum, its parameters should be sufficient to describe with good precision the non-linear power spectrum and (ii) the parameters should have a clear physical interpretation. To accomplish this, our starting point is the work of Murgia et al. (2017) who suggested the following parametrization for the cutoff of non-CDM (nCDM) models in terms of the linear transfer function \( T_L^2(k) = P_{\text{nCDM}}(k)/P_{\text{CDM}}(k) \):

\[
T_L(k) = [1 + (ak)^\nu]^{\gamma},
\]

where \( a \) is a measure of the cutoff scale length, and \( \beta \) and \( \gamma \) the shape of the cutoff. This is a generalization of the fitting formula for WDM, where \( \beta = 2\nu \) and \( \gamma = -5/\nu \) with \( \nu = 1.12 \), and allows for much higher variety in the shape of the cutoff. However, it only describes a single cutoff in the power spectrum, while we want to include models with DAOs as well. We remark however that the transfer of power from large to smaller scales in the non-linear evolution tends to erase the DAOs (e.g. Buckley et al. 2014). Since one of the goals of our parametrization is to reproduce with good precision the non-linear evolution of the power spectrum down to \( z = 5 \), we thus start by looking at the accuracy to which Eq. (1) can be expected to account for the non-linear regime.

This can be seen in Fig. 4, where the red line corresponds to Eq. (1). Comparing the results from our simulations using this parametrization and the power spectrum with several DAOs (black lines) as initial conditions, we find that it is not sufficient to capture with precision the amplitude and features of the non-linear power spectrum at small scales for models which have strong DAO features (i.e. where the first oscillations are near the CDM amplitude). As can be seen in Fig. 4 (red line), for this particular strong DAO model, this parametrization underestimates the power at \( k \gtrsim 100 \, h \, \text{Mpc}^{-1} \). For instance, by up to 48% and 24% at \( k = 500 \, h \, \text{Mpc}^{-1} \) for \( z = 8 \) and 5 respectively. In Section 4.1 we quantify in detail the impact of DAOs in the non-linear power spectrum for a broad range of scales and amplitudes of the DAOs.

In a first attempt to improve Eq. (1) to account for DAO models, we add a term that includes the first DAO peak by modellling it with a Gaussian:

\[
T_L(k) = [1 + (ak)^\nu]^{\gamma} - \sqrt{h_{\text{peak}}} \exp \left( -\frac{1}{2} \left( \frac{k - k_{\text{peak}}}{\sigma k_{\text{peak}}} \right)^2 \right),
\]

where \( h_{\text{peak}} \) and \( k_{\text{peak}} \) give the amplitude (relative to CDM) and position of the first peak (these two will be the most important parameters throughout this work), and \( \sigma \) controls how narrow the Gaussian is. Eq. (2) improves the agreement with the full non-linear power spectrum as can be seen in Fig. 4, but is still not good enough to reconstruct the full power at the smallest scales, where it underestimates the power by 30% and 14% for \( z = 8 \) and 5 respectively. Therefore, the power provided by the secondary peaks in sDAO models remains relevant down to \( z = 5 \).

To gain precision in our parametrization for models that have DAOs, we extend Eq. (2) by adding terms that model the secondary peaks of the DAOs. These peaks can be described by two features, their envelope and oscillations. The oscillations of the higher order peaks are very regular and

| Dimensionless power spectra | \( \Delta^2 \) at \( z = 5 \) for a uniform simulation (black) and three zoom simulations with different zoom volumes and resolution levels. Note that the power spectrum from the zoom simulations can reach the same larger scales as the uniform simulation by including the low resolution particles. Thus, all lines completely overlap for \( k \lesssim 30 \, h \, \text{Mpc}^{-1} \). The dashed lines show the expected Poisson (shot) noise (\( \Delta^2_{\text{shot}} = k^2 V/(2\pi^2 N) \), where \( N \) is the number of particles and \( V \) the volume they occupy) for the corresponding simulation. |
can be fitted with a cosine function, whose frequency is determined by \( k_{\text{peak}} \). The envelope can be parametrized with the amplitude of the second peak \( h_2 \) and two error functions, one giving the steep rise on the left side (similar to the Gaussian describing the first peak) and the other controlling the damping on the right (the oscillations are not fully symmetrical and thus a Gaussian is not enough to describe their shape). The full fitting function is then given by:

\[
T_L(k) = \left[ 1 + (ak)^2 \right] - \sqrt{h_{\text{peak}}} \exp \left( -\frac{1}{2} \left( \frac{k - k_{\text{peak}}}{\sigma k_{\text{peak}}} \right)^2 \right)
\]

\[
+ \frac{\sqrt{h_2}}{4} \text{erfc} \left( \frac{k - 1.805k_{\text{peak}}}{\tau k_{\text{peak}}} - 2 \right) \cos \left( \frac{1.1083\pi k}{k_{\text{peak}}} \right),
\]

where \( \text{erfc}(x) = 1 - \text{erf}(x) \) is the complementary error function. This full parametrization would have 8 parameters, but they are not all independent and can be simplified for ETHOS models, where they are fixed by \( h_{\text{peak}} \) and \( k_{\text{peak}} \) (see Section 3.1 below). The parameter \( \alpha \) can be determined by the scale at which the transfer function dropped to \( 1/2 \) \((k_{1/2})\), which is connected to \( k_{\text{peak}} \):

\[
\sigma = \frac{d}{k_{\text{peak}}} \left[ \left( \frac{1}{\sqrt{2}} \right)^{1/\beta} - 1 \right]^{1/\beta},
\]

where \( d \) controls the ratio between \( k_{\text{peak}} \) and \( k_{1/2} \), which is in the range \( 2.4 - 3 \).

Fitting the power spectrum cutoff of ETHOS models with Eq. (1) leads in all cases to large negative values for \( \gamma \), whose precise value make almost no difference in the reconstruction of the cutoff; thus, we have fixed \( \gamma = -20 \). From the remaining parameters, \( h_{\text{peak}} \) and \( k_{\text{peak}} \) are the most relevant free parameters since the former determines the position of the first DAO peak as well as describing the position of the cutoff (see Eq. (4)), and the latter the amplitude of the first DAO. These parameters are responsible for the leading order effects on the non-linear power spectrum. Regarding the remaining parameters: \( \beta \) is responsible for the cutoff shape, \( \tau \) controls the damping of the DAOs, and \( \sigma \) gives the width of the first peak. Physically, \( k_{\text{peak}} \) is connected to the DM sound horizon and \( \tau \) to the Silk damping scale (the physical interpretation of the key parameters is described in Section 3.1 and Appendix B). The effects of the parameters and the quality of our final parametrization in the linear transfer function can be seen in Fig. 5.

3.1 Connection with the ETHOS framework and physical interpretation of the parameters: \( h_{\text{peak}}, k_{\text{peak}} \) and \( h_2 \)

Having accomplished the goal of providing a parametrization for DAO models that is simpler than a fully general parametrization of the linear power spectrum (as provided in Cyr-Racine et al. 2016) but still precise enough to describe their non-linear evolution, we proceed now to establish the connection between these phenomenological parameters and the physical parameters of the ETHOS framework.
(Cyr-Racine et al. 2016) in regards to the effects of the DM-DR interactions in generating the power spectrum cutoff and the DAOs. We recall that with such a connection, it is then possible to have a complete mapping between the particle physics parameters of the models explored in Cyr-Racine et al. (2016) and the parameters relevant for nonlinear structure formation.

The physics of the DAOs in the linear power spectrum is captured within the modelling presented in Cyr-Racine et al. (2016) by the parameters \( n \) and coefficient \( a_n \) that control the redshift scaling of the DM drag opacity \( \chi \) with \( (1+z)^{n+1} \), plus a set of coefficients \( a_l \) that parameterize the angular dependence of the DM-DR scattering cross section. For this work, we will refer only to models that have single values for \( n \) and \( a_n \), and a set of constant \( a_{l>2} \) values. A specific particle physics scenario contained within these constraints is that of a massive fermionic DM particle interacting with a massless fermion via a massive vector mediator as in van den Aarssen et al. (2012), which corresponds to the case \( n = 4 \), \( a_{l>2} = 3/2 \) with different values of \( a_2 \) providing cutoff scales for the power spectrum. This specific model has been studied with simulations in the past (Vogelsberger et al. 2016), particularly the benchmark model referred to as ETHOS-4 in table 1 of Vogelsberger et al. (2016).

Although the parameters \( n, a_n \) and the set \( \{a_l\} \) are sufficient to characterize the linear power spectrum within the ETHOS framework, they obscure somewhat the physical mechanism behind the DAOs, and they also lack the simple phenomenological interpretation of the parameters described above \( \{k_{\text{peak}}, h_{\text{peak}}, b_2\} \). Because of this, we first attempt to approximate the results of the full calculation of the linear power spectrum based on a Boltzmann code (modified version of CLASS; Archidiacono et al. 2017, 2019) with a simple physical model based on the tight coupling limit approximation (between DM and DR) in analogy with the photon-baryon plasma (see e.g. Hu & Sugiyama 1996). This attempt is described in Appendix B. Although we find that this approximation is not accurate enough, particularly in describing the damping envelope of the DAOs, it does provide relevant insights into the relevance of the sound horizon scale and the DM decoupling epoch as the physical quantities behind the DAO features. Therefore, we decided to try a phenomenological approach based on these quantities. To test this approach we explore a set of 84 ETHOS models as described above with the set of values: \( \{n = (3-15, 20), \log_{10}(a_n) = (0, 1, 2, 3, 5, 7)\} \), and fixing \( a_{l>2} = 3/2 \).

The sound horizon scale \( r_{\text{DAO}} \approx c_{s} \eta_{k_{\text{peak}}} \), where \( c_{s} \) is the DM sound speed and \( \eta_{k} \) is the conformal time of DM decoupling defined by

\[
\int_{\eta_{k_{\text{peak}}}}^{\eta_0} -\chi d\eta = 1
\]

with \( \chi \) being the DM drag opacity due to the DM-DR interactions, should give the largest scale affected by acoustic oscillations and thus should be connected to \( k_{\text{peak}} \). We found

\[
3 \text{ We refer specifically to the case where DR-DR interactions are irrelevant.}
\]

\[
4 \text{ More specifically, } a_{l} \text{ is the ratio between the opacity of the } l^{th}-\text{moment to that of the dipole moment of the DR multipole hierarchy given by the angular dependence of DM-DR scattering.}
\]
by the DM mean free path (due to the DM-DR interactions) at DM decoupling, which is given by the inverse of \( \dot{\kappa}_x(\eta_1) \) (see Fig. 7):

\[
h_{\text{peak}} = 0.21 \left( \frac{\dot{\kappa}_x(\eta_1)}{\mathcal{H}(\eta_1)} \right)^{0.66} - 0.31,
\]

where the relevant quantity is actually the ratio \( \dot{\kappa}_x(\eta_1)/\mathcal{H}(\eta_1) \), with \( \mathcal{H} \) being the Hubble rate (relative to the conformal time). This ratio is actually equal to the ETHOS parameter \( n \) (see Appendix B). Thus, for \( n \gg 1 \), the DM drag visibility function \( \kappa e^{-\kappa t} \) is narrower, which implies a faster decoupling time scale; indeed, the DM-DR plasma is clearly in the tightly coupled regime at \( \eta_1 \), and thus the damping by DR diffusion is only significant for the smallest scales. On the contrary when \( n \lesssim 1 \), the DM drag visibility function is broader so that the timescales for decoupling (which occurs mostly in the weakly coupled regime) are larger. Thus, the DM mean free path at decoupling is relatively large and DR can diffuse substantially, lowering the value of \( h_{\text{peak}} \). While for \( n \lesssim 9 \), \( n \) is the only factor in determining \( h_{\text{peak}} \), that is not true anymore for \( n \gtrsim 10 \). In the latter case, \( h_{\text{peak}} \) spreads around the fit in Fig. 7 depending on the value of \( a_n \), the larger \( a_n \) the larger \( h_{\text{peak}} \).

Finally, the ratio of the first two DAO peaks \( h_{\text{peak}}/h_2 \), is connected to the ratio of the DR to DM decoupling times \( \eta_{\text{DR}}/\eta_{\text{D}} \) (see Fig. 8), where \( \eta_{\text{DR}} \) is the conformal time of DR decoupling defined by

\[
\int_{\eta_{\text{DR}}}^{\eta_{\text{D}}} -\kappa_{\text{DR}} d\eta = 1
\]

with \( \kappa_{\text{DR}} \) being the DR opacity to DM scattering. This relation can be approximated by a parabola, but with different parameters for different values of \( n \).

There are two additional features in Fig. 7 that we highlight: (i) for models with very large \( n \gtrsim 15 \) it is possible to have \( h_{\text{peak}} > 1 \); (ii) for models with \( n \gtrsim 10 \), the value of \( h_{\text{peak}} \) depends not only on \( n = \dot{\kappa}_x(\eta_1)/\mathcal{H}(\eta_1) \), but also on the specific value of \( a_n \) (the larger \( n \), the stronger the dependence). For the latter models it is also true that the second DAO peak is significantly larger than the first (see Fig. 8). We decide to exclude these models, i.e. those with \( n \gtrsim 10 \), from our analysis for two reasons: (i) they would have a power spectrum that exceeds that of CDM at certain scales, and (ii) the parameter \( h_2 \) would no longer be a secondary parameter in determining the non-linear power spectrum. We notice however, that there is potentially interesting phenomenology in these models, which we leave for a future work. With this exclusion and using the strong correlations seen in Figs. 6-8, we have accomplished our goals at least for the regime of weak to moderately strong \((h_{\text{peak}} \sim 0.6) \) DAO models, i.e., we have found a way to connect the parameters \( h_{\text{peak}} \) and \( k_{\text{peak}} \) in our parametrization to \( n \) and \( a_n \), respectively, in the original ETHOS framework, as well as to connect \( h_{\text{peak}} \), \( k_{\text{peak}} \) and \( h_2 \) to physical quantities that are responsible for the DM-DR decoupling.

Within our parametrization, it is possible to include models with stronger DAO features \((h_{\text{peak}} \sim 1)\), but that do not exceed greatly the CDM power spectrum, such as the benchmark sDAO model analysed in Bose et al. (2019) to show the distinct features this type of models leave in the \( Ly\alpha \) forest 1D flux spectrum. To do so, we need to change the value of \( a_{12} \) from 3/2 to a value of \( O(10) \) for \( n \sim 9 \). In this way, we can create power spectra that have strong DAO features but without having \( h_{\text{peak}}/h_2 < 1 \) (see Fig. 9).

This modification breaks the relations for the peak heights in Fig. 7 and 8, while the one for \( k_{\text{peak}} \) stays unchanged. We used these models with increased \( a_{12} \) for our strong DAO cases.

### 3.2 Final parameter space

Given all previous considerations, we work with a 2D parameter space with a set \( \{k_{\text{peak}}, h_{\text{peak}}\} \). We explore simulations within a range of \( k_{\text{peak}} \) between 35-300 \( h \) Mpc\(^{-1}\) and...
4 RESULTS

To characterise the differences between DM models and find out which features survive the non-linear evolution, we look at the matter power spectrum and the halo mass function at high redshift $z > 5$.

4.1 Matter power spectrum

We evaluate the power spectrum of the simulations at different scales $k_{\text{probe}}$ between 10 h Mpc$^{-1}$ and 500 h Mpc$^{-1}$. This range roughly covers the relevant range where new DM physics can play a role in the physics of galaxies: the larger scales are bounded by current constraints over deviations from CDM (for instance from Ly-$\alpha$ forest measurements e.g. Irišić et al. 2017), while the smaller scales are bounded by the minimum scales at which galaxies can form, where we use the atomic cooling limit as a reference. At $z = 5$, the virial mass of a halo corresponding to a non-linear scale of $k = 500$ h Mpc$^{-1}$ is $\sim 5 \times 10^8 M_\odot h^{-1}$, which is just below the atomic cooling limit at this redshift, $\sim 10^9 M_\odot h^{-1}$.

To quantify the difference of a given model with respect to CDM, we define two different diagnostics: (i) we compute the ratio $R(k_{\text{probe}})$ of the power spectra with respect to CDM at $k_{\text{probe}}$:

$$R(k_{\text{probe}}) = \frac{\Delta^2_{\text{CDM}}(k_{\text{probe}})}{\Delta^2(k_{\text{probe}})},$$

and consider $R(k_{\text{probe}}) \leq 1.1$ as essentially indistinguishable from CDM (since we set our convergence goal to 5%; see Appendix A); (ii) to distinguish different non-cold DM models, $R(k_{\text{probe}})$ is not sufficient enough, as two models with the same ratio at a given $k$ can have different behaviour on larger scales. To capture this with a single number, we define the following dimensionless integrated quantity$^5$

$$I(k_{\text{probe}}) = \int_{k_{\text{min}}}^{k_{\text{probe}}} \frac{\Delta^2_{\text{CDM}}(k)}{\Delta^2(k)} d\ln k,$$  

where we choose $k_{\text{min}} = 10$ h Mpc$^{-1}$ since, as we mentioned above, the models we are interested in have the same power at this scale. By construction, a larger value of $R(k_{\text{probe}})$ also results in a larger value of $I(k_{\text{probe}})$, thus we need to normalize it to a reference case in order to define a comparative quantity across the different structure formation models.

We choose the WDM case as the reference and normalize Eq. (10) for a given model by the value of $I(k_{\text{probe}})$ of a WDM model with the same value of $R(k_{\text{probe}})$:

$$\hat{I}(k_{\text{probe}}) = \left( \frac{I(k_{\text{probe}})}{I_{\text{WDM}}(k_{\text{probe}})} \right) R(k_{\text{probe}}).$$

Defined in this way, for a fixed $R(k_{\text{probe}})$, all models with $\hat{I}(k_{\text{probe}}) = 1$ have a non-linear power spectrum at the given redshift which is essentially indistinguishable from a WDM of the same $R(k_{\text{probe}})$, regardless of how different the linear power spectrum of these models is relative to WDM.

Figure 10 shows the results for our simulations for $z = 5$ and $k_{\text{probe}} = 500$ h Mpc$^{-1}$ in the leading order space of parameters $k_{\text{peak}}$ and $h_{\text{peak}}$ (Eq. 3). We recall that we ran 50 simulations for models within this parameter space, which are then used to interpolate the values of $R(k_{\text{probe}})$ and $\hat{I}(k_{\text{probe}})$ to fill in Fig. 10. The line contours show $R(k_{\text{probe}})$, which increases from right to left with the models to the right of the blue dashed line ($R = 1.1$) being virtually indistinguishable from CDM, while those to the left become ever more divergent from CDM. The colour scale shows the value of $\hat{I}(k_{\text{probe}})$ and therefore quantifies how different a model is compared with a WDM model that has the same value of $R$. Note that the WDM models in this plot lie at the bottom $k_{\text{peak}}$ axis ($h_{\text{peak}} = 0$). Using the connection between $\alpha$ and $m_{\text{WDM}}$ from Viel et al. (2005)

$$\alpha = 0.049 \left( \frac{m_{\text{WDM}}}{1 \text{ keV}} \right)^{-1.11} \frac{1}{\Omega_{\Lambda}}^{0.11} \frac{\Omega_m}{0.25}^{0.11} \left( \frac{h}{0.7} \right)^{1.22} h^{-1} \text{ Mpc},$$

Table 2. Parameters used to construct the linear power spectra used in our simulations (Eq. 3) as a function of the amplitude $h_{\text{peak}}$ of the first DAO peak relative to CDM. For $h_{\text{peak}}$ in the range [0.2, 0.6], the amplitude of the second DAO peak, $h_2$ is given by $h_2 = A \exp(Bk_{\text{peak}}) + C$, where the (A, B, C) values are given in the column.

| $h_{\text{peak}}$ | $h_2$ | $\tau$ | $\sigma$ | $\beta$ | $d$ |
|------------------|-------|--------|----------|--------|-----|
| 0.0              | 0.0   | 0.0    | 0.0      | 2.24   | 3.0 |
| 0.2              | (0.067, -0.086, 0.011) | 0.34 | 0.23 | 3.1 | 2.93 |
| 0.4              | (0.221, -0.025, 0.0129) | 0.34 | 0.22 | 3.61 | 2.61 |
| 0.6              | (0.572, -0.008, 0.1490) | 0.38 | 0.2 | 3.91 | 2.44 |
| 0.8              | 0.88  | 0.55   | 0.2      | 4.0    | 2.46 |
| 1.0              | 1.08  | 0.67   | 0.2      | 4.05   | 2.5 |

$^5$ Note that we use a quadratic dependence on the ratio $1/f_{\text{NL}}^2 = \Delta^2_{\text{CDM}}/\Delta^2$ in the integrand in Eq. (10) instead of a linear one in order to enhance the difference between models.
we can compute \( m_{\text{WDM}} \) from \( h_{\text{peak}} \):

\[
\frac{m_{\text{WDM}}}{1 \text{ keV}} = \left[ 0.050 \left( \frac{k_{\text{peak}}}{h \text{ Mpc}^{-1}} \right) \left( \frac{\Omega_{\Lambda}}{0.25} \right)^{0.11} \left( \frac{h}{0.7} \right)^{1.22} \right]^{1/2}.
\]  

(13)

For a value of \( I \) close to one (green colour), the model’s power spectrum up to \( k_{\text{ probe}} = 500 h \text{ Mpc}^{-1} \) at \( z = 5 \) will be indistinguishable from WDM, while larger values mean that the power spectrum shape for \( k < k_{\text{ probe}} \) is truly distinct from WDM regardless of the value of \( R(k_{\text{ probe}}) \). The region in the top left where \( I(k_{\text{ probe}}) \) has the largest values corresponds to models with strong DAO features \( (h_{\text{ peak}} \lesssim 0.7 \) and \( k_{\text{ peak}} \lesssim 65 h \text{ Mpc}^{-1} \); labeled as DAOs) where the impact of the DAO features has not been erased by the non-linear evolution down to \( z = 5 \), and thus still leaves a signature in the power spectrum up to \( k_{\text{ probe}} \). As is apparent most of the parameter space outside of the latter DAO region has values of \( I \) close to 1, and thus any models here are essentially degenerate (up to \( k_{\text{ probe}} = 500 h \text{ Mpc}^{-1} \) and at \( z = 5 \)) with a WDM model with the same value of \( R \). This degeneracy is either caused by the non-linear evolution erasing the DAO features, especially for the weak DAO models in the lower part of the plot \( (h_{\text{ peak}} \lesssim 0.6) \), or because the DAO features appear at smaller scales \( k > k_{\text{ probe}} \) than we are interested in (for models with \( k_{\text{ peak}} \gtrsim 100 h \text{ Mpc}^{-1} \)).

The hashed region on the lower left was constructed taking as a reference the constraints on the thermal WDM particle mass from current Lyman-\( \alpha \) forest data from Iršič et al. (2017): \( m_{\text{WDM}} < 3.5 \text{ keV} \) (2\( \sigma \) C.L.). To do this, we follow the contour line corresponding to this WDM model, up to the value of \( h_{\text{ peak}} \) that remains degenerate with this
WDM models (\( h_{\text{peak}} \approx 0.47 \)) using a value of \( \tilde{I} = 1.01 \) as the dividing threshold. We then continue the Lyman-\( \alpha \) constraint line towards larger values of \( R \) along this \( \tilde{I} \) threshold. We remain within this threshold because we expect that beyond, the Lyman-\( \alpha \) analysis based on WDM model would no longer be valid due to the impact of the DAO features. In this way, the hashed region on the lower left is our expectation for the exclusion region from Lyman-\( \alpha \) data. Drawing this region more precisely would require a full analysis of the predictions of the 1D flux power spectrum within our framework. Finally, we have indicated with a red symbol the location of the benchmark ETHOS-4 model defined in Vogelsberger et al. (2016), which is seemingly barely allowed within the Lyman-\( \alpha \) exclusion region. This is a point that was noticed in Vogelsberger et al. (2016) where the linear power spectrum of this model was considered to have a WDM equivalent (in terms of the cutoff) with a thermal particle mass of \( m_{\text{WDM}} = 3.66 \text{ keV} \). Our results confirm this correspondence and indicate that the ETHOS-4 model has a power spectrum that is nearly indistinguishable at \( z = 5 \) from a WDM model with \( m_{\text{WDM}} \approx 3.5 \text{ keV} \).

From the slopes of the contour lines in Fig. 10, we can see that the degeneracies are not between weak DAO and WDM models with the same power spectrum cutoff (i.e. \( k_{\text{peak}} \)), but that the additional power coming from the DAOs at scales smaller than \( k_{\text{peak}} \) still matters and can only be accounted for by WDM models with a cutoff at smaller scales. This is shown more clearly on the left panels of Fig. 11, which show the power spectrum for a weak DAO model (ETHOS-2 in Vogelsberger et al. (2016); \( h_{\text{peak}} = 0.2, k_{\text{peak}} = 46.5 \text{ h Mpc}^{-1} \)) and a WDM model with \( m = 2.3 \text{ keV} \) at \( z = 127 \) (top) and \( z = 5 \) (bottom). Despite having distinct linear power spectra, these models are nearly degenerate at \( z = 5 \); they have both the same value of \( R(k_{\text{peak}} = 500 \text{ h Mpc}^{-1}) \) and \( \tilde{I}(k_{\text{probe}} = 500 \text{ h Mpc}^{-1}) \). Moreover, in order to match the weak DAO model, the WDM model needs to sustain more power at larger scales in the linear power spectra to compensate for its steeper cutoff. The right panels of Fig. 11 show the distinct behaviour of a strong DAO model. In this case the additional power from the secondary DAO peaks have an impact in the non-linear power spectrum down to \( z = 5 \) that cannot be replicated by a WDM model; in order for the sDAO and the WDM models to have the same power at \( k_{\text{peak}} = 500 \text{ h Mpc}^{-1} \), the WDM linear cutoff needs to occur at significantly smaller scales than the one for the sDAO model. Thus, the WDM model has more power at intermediate scales by \( z = 5 \).

Figure 12 is equivalent to Fig. 10 but at different redshifts (\( z = 5, 8, 10 \)) and scales \( k_{\text{probe}} \) (50, 150, 300 \( \text{h Mpc}^{-1} \)). Note that we use the same colour scale for \( I \) for all cases in order to ease the comparison between the different panels. Focusing on the values of \( R(k_{\text{peak}}) \) represented by the contour lines first, we can see that they shift towards larger \( k_{\text{peak}} \) as the redshift increases (from top to bottom) or as \( k_{\text{probe}} \) increases (from left to right). The former trend is expected since at higher redshift the damping envelope in the non-linear power spectrum is more sensitive to the differences across DM models in the linear

4.2 Halo mass function

The halo mass function provides another relevant measure to characterise structure formation models. It is also more sensitive to the differences across DM models in the linear

Figure 11. Comparison of the power spectra from a couple of DAO models and the comparable WDM models with \( m = 2.3 \text{ keV} \) (left) and \( m = 3.4 \text{ keV} \) (right) at the initial conditions for our simulations (\( z = 127 \) top panels) and at the end of the simulations (\( z = 5 \); bottom panels). The left panels are for a \( wDAO \) model (ETHOS-2 in Vogelsberger et al. 2016), while the right panels are for an \( sDAO \) model. On the left, the different models have clearly different power spectra at the initial conditions, but become completely degenerate at \( z = 5 \). On the right panels on the other hand, the models remain different even at \( z = 5 \), despite having the same power at \( k = 500 \text{ h Mpc}^{-1} \). This shows strong DAO models have truly distinct power spectra features relative to the WDM model.
Figure 12. Division of structure formation models in the parameter space $h_{\text{peak}}$ and $k_{\text{peak}}$ of Eq. (3) based on the power spectra at different scales $k_{\text{probe}}$ and redshifts. See Fig. 10 for a description of the different elements in this plot.

Figure 13 shows an example of a comparison of the halo mass function at $z = 5$ between the CDM model, and a WDM, weak DAO and strong DAO model with the same cutoff scale in the initial power spectrum. It can be seen that even though the three models are designed to start deviating from CDM at roughly the same mass, the slope of the halo mass function at smaller scales is very different. The halo mass function for the WDM model stays roughly flat towards the left of the cutoff mass until the slope rises again artificially due to the presence of spurious haloes caused by well-known discreteness effects in models with a primordial power spectrum cutoff. The limiting mass below which one can no longer trust the halo mass function is well described by a formula that depends on the cutoff scale of the model and the spatial resolution of the simulation (see Wang & White 2007). As can be seen in Fig. 13, this formula describes reasonably well the scale at which spurious haloes start to dominate not only for the WDM model, but also for the weak and strong DAO models. For the CDM model, the mass function can be trusted to even lower masses until the simulation particle number is too low to resolve haloes (typically $\lesssim 100$). On the other hand, the additional small scale power of the DAO models keeps the slope of the halo mass function steeper at small masses, relative to the WDM case, and for strong DAOs, the halo mass function is even parallel to the CDM case, albeit with a reduced normalization.

Figure 14 shows the structure formation models in the parameter space ($h_{\text{peak}}, k_{\text{peak}}$) as characterised by the halo mass function of our simulations at $z = 5$ and at a halo mass of $M_{\text{probe}} = 10^8M_\odot/h$. This figure is analogous to Fig. 10, with the contours showing in this case the ratio between the CDM halo mass function and that of a given model at $M_{\text{probe}} = 10^8M_\odot/h$:

$$R_M(M_{\text{probe}}) = \left(\frac{M_{\text{CDFM}}}{MF}\right)_{M_{\text{probe}}}$$

(14)

where $MF = dn/d\log M$ is the differential halo mass function. On the other hand, the colour contours represent the number of haloes with $M > M_{\text{probe}}$ normalized to the number of
We can measure the halo mass function. We can see that the mass function is so close to the same contour line, since in this case the cutoff in the halo mass functions of a WDM model with the same ratio at $k_{\text{peak}}$ (irrespective of the value of $h_{\text{peak}}$) do not have very different slopes for their halo mass functions yet and thus, they all look alike. On the contrary, for $k_{\text{peak}} < 100 \, h \, \text{Mpc}^{-1}$, the DAO models have halo mass functions with slopes that are clearly steeper (and thus distinguishable) than that of the WDM model below the cutoff mass. Therefore, in order for the corresponding WDM model to lie on the same contour line (i.e. to have the same halo mass function at $M_{\text{probe}}$), it needs to have a cutoff scale at a relatively smaller mass (larger $k_{\text{peak}}$), and thus will necessarily have more haloes with $M > M_{\text{probe}}$ than the DAO model (see Eq. 15). We observe that for $k_{\text{peak}} \lesssim 100 \, h \, \text{Mpc}^{-1}$, the slope in the halo mass function towards smaller masses is related to $h_{\text{peak}}$, which can be seen by looking at how the contour lines bend ever more sharply towards lower values of $k_{\text{peak}}$ as $h_{\text{peak}}$ increases, eventually becoming nearly flat for $k_{\text{peak}} \sim 0.6$ at $k_{\text{peak}} \sim 30–60 \, h \, \text{Mpc}^{-1}$. This implies that for these models, the actual mass cutoff (given by $k_{\text{peak}}$) does not matter any longer since they all have the same mass function at $10^{9} M_{\odot} h^{-1}$ haloes. Naturally, these models are still distinguishable since they have different halo abundances at larger masses.

In contrast to the division of structure formation models based on the non-linear power spectrum (Fig. 10), in this case represented by the halo mass function, the distinctive DAO region (black dashed line in Fig. 14) occupies a larger region of the parameter space, reaching into the regime of the weak DAO models. For instance, the red star in Fig. 14 corresponds to the ETHOS-4 model used in Vogelsberger et al. (2016) and it appears at the border of our definition of the DAO structure formation region. It is thus clear that even though weak DAO models are degenerate with WDM models in their non-linear power spectrum, this degeneracy is broken for the halo mass function.

At different redshifts (Fig. 15), the DAO region (black dashed line) remains almost unchanged (shrinking slightly at high redshift); the same is true for the CDM-like region (blue dashed line). It is only the contour lines of constant $R_{M}$ that change across redshift, with the ratio of the halo mass function at $10^{9} M_{\odot} h^{-1}$ becoming higher at larger redshift for all the region below $k_{\text{peak}} \lesssim 100 \, h \, \text{Mpc}^{-1}$. Therefore, the halo mass function provides a diagnostic to classify structure formation models that is less susceptible to being erased by the non-linear evolution than the power spectrum. A more detailed analysis of the halo mass function for the DAO models studied here will be presented in the future.

5 CONCLUSIONS

There are multiple ways in which non-standard DM physics can introduce a cutoff in the linear matter power spectrum whose shape can range from an exponential featureless free-streaming collisionless damping (as in thermal WDM models) to a shallower collisional damping driven by DM-dark radiation interactions with strong DAOs. The effective theory of structure formation (ETHOS), introduced in Cyr-Racine et al. (2016) aims at connecting the particle physics parameters of a variety of DM models into effective parameters that characterise the linear power spectrum. In this way, DM particle models can be classified in terms of a set of parameters that fully describe the linear power spectrum, particularly the characteristics of the small-scale cutoff and DAOs. It is however, not trivial to characterise the signature that these different departures from the linear CDM power spectrum leave in the non-linear regime of structure formation. It may indeed be possible that the gravitational coupling between different scales erases features like the DAOs making all models essentially indistinguishable from the standard WDM cutoff at the scales that are relevant for galaxy formation and evolution.$^{6}$ $10 \, h \, \text{Mpc}^{-1} \lesssim k \lesssim 500 \, h \, \text{Mpc}^{-1}$. In this

$^{6}$ Relevant departures from CDM are bounded at large scales by current constraints based on e.g. observations of the Lyman-$\alpha$ 1D flux spectrum, and at small scales by the suppression of galaxy formation below the atomic cooling limit.
Figure 14. Division of structure formation models in the effective parameter space $h_{\text{peak}}$ and $k_{\text{peak}}$ (see Eq. 3) according to their halo mass functions at $z = 5$ for $M_{\text{probe}} = 10^8 M_\odot h^{-1}$. The contour lines correspond to the ratio $R_M(M_{\text{probe}})$ of the CDM halo mass function relative to a given model at $M_{\text{probe}}$ ($\hat{I}_M(M_{\text{probe}})$; see Eq. 14). The colour scale shows the number of haloes with $M > M_{\text{probe}}$ normalized to a WDM model with the same number of haloes at $M = M_{\text{probe}}$ (see Eq. 15), where a value of 1 corresponds to areas that are degenerate with WDM. The black dashed line ($\hat{I}_M(M_{\text{probe}}) < 0.85$) on the left encompasses the area where DAO features survive until $z = 5$. The region to the right of the blue dashed line, $R_M(M_{\text{probe}}) = 1.1$, can be considered as nearly indistinguishable from CDM down to $M_{\text{probe}}$. The ETHOS-4 model used as a benchmark in Vogelsberger et al. (2016) is indicated by the purple star.

In this work we address this question by performing a large number of cosmological simulations within the ETHOS framework. Our goal is to define a reduced set of simple yet physically motivated parameters that allow to distinguish DM models based on differences in how structure formation proceeds (at the scales relevant for the physics of galaxies). In this first work, we have concentrated on DM-only simulations at high redshift ($z \geq 5$). Our analysis and main results can be summarised as follows:

- We have implemented a zoom-in simulation technique to efficiently cover a wide range of scales (0.2–500 $h$ Mpc$^{-1}$) and accurately reconstruct the (average) matter power spectrum in this range (see Fig. 3). The computational cost of this method is significantly less than a uniform simulation with equivalent range (see Table 1).

- We introduced a new analytic formula (Eq. (3), Fig. 5) to describe the linear transfer function (relative to CDM) of models with a primordial cutoff, which accommodates both WDM and models with DAOs. This formula is accurate enough to reproduce the non-linear power spectrum for the scales of interest compared to the full calculation with a Boltzmann code (see Fig. 4). Crucially, only two parameters in this formula, $\{h_{\text{peak}}, k_{\text{peak}}\}$, the amplitude and scale of the first DAO peak, are sufficient to characterise non-linear structure formation at high-redshift for WDM and a large class of ETHOS models with DAOs. Moreover, we found a simple physical interpretation for these two main parameters where $k_{\text{peak}}$ is connected to $\eta_x$ the time of DM decoupling from the DR (see Fig. 6) and $h_{\text{peak}}$ is determined by the ratio of the

$\eta_x$, the time of DM decoupling from the DR (see Fig. 6), and $h_{\text{peak}}$ is determined by the ratio of the

$\nu_{x}$. In the case of WDM, $k_{\text{peak}}$ is connected to $k_{1/2}$, the scale at which the transfer function squared is equal to 1/2; see Eq. (4).
DM drag opacity to the Hubble rate at the time of DM decoupling $k(\theta_k)/\mathcal{H}$ (see Fig. 7).

- Using 50 simulations down to $z = 5$ within the $(h_{\text{peak}}, k_{\text{peak}})$ parameter space (a new effective space in ETHOS), we have been able to classify DM models into regions with distinct non-linear structure formation at galactic scales (CDM, WDM, DAOs), quantified by the non-linear power spectrum and the halo mass function at high redshift.

- As far as the non-linear matter power spectrum is concerned, we find that only a small region within this effective parameter space, corresponding to relatively small values of $k_{\text{peak}}$ and large values of $h_{\text{peak}}$ (strong DAO models), still preserves a signature of the DAOs at $z = 5$ at the galactic scales corresponding to the smallest galaxy-forming haloes $500\, h^{-1} \text{Mpc}$ (Fig. 10). The rest of the relevant parameter space is either degenerate with WDM or indistinguishable from CDM. This distinct DAO region expands at higher redshifts and contracts at smaller scales (see Fig. 12).

- We find that it is possible to break (to a certain extent) the degeneracies between weak DAO models (small values of $h_{\text{peak}}$) and WDM models seen in the non-linear power spectrum by characterising structure formation models using the halo mass function instead. This is because the halo mass function retains a memory of the linear power spectrum, having a slope that is very sensitive to the value of $h_{\text{peak}}$.

In this way, the distinct DAO region covers a much larger region of the parameter space and changes only slightly with redshift (see Figs. 14–15).

Using our results, it is possible to use the new analytic formula we propose (Eq. 3) to fit the linear power spectrum of a broad class of ETHOS models with DAOs, and use the values of $(h_{\text{peak}}, k_{\text{peak}})$ to determine to which structure formation region they belong to in the non-linear high-redshift regime (CDM-like, WDM-like or DAO), without performing additional N-body simulations. Notice that this is valid for any DM particle model with a primordial power spectrum with DAOs that can be fitted accurately with our formula (up to the second DAO peak). In other words, given the values of $(h_{\text{peak}}, k_{\text{peak}})$, our method allows to infer the value of the non-linear power spectrum and halo mass function at any relevant scale/mass at high-redshift $z \sim 10 – 5$.

Furthermore, the effective parameters for structure formation we propose here represent a potentially powerful way to constrain the parameter space of a variety of particle physics models by using observations in the high-redshift Universe, such as the Lyman-$\alpha$ 1D flux power spectrum. We should remark however that in order to accurately exploit this avenue, we need to incorporate the baryonic physics that is relevant for the intergalactic medium into the ETHOS parameter space, which is something we plan to do in the future. Another avenue we will explore is to extend our results towards lower redshifts, where DM self-interactions have a significant impact in the centre of DM haloes, and thus need to be incorporated as an additional parameter to classify structure formation regimes.

**ACKNOWLEDGEMENTS**

SB and JZ acknowledge support by a Grant of Excellence from the Icelandic Research Fund (grant number 173929). CP acknowledges support by the European Research Council under ERC-CoG grant CRAGSMAN-646955. The simulations were performed on resources provided by the Icelandic High Performance Computing Centre at the University of Iceland, and the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

**REFERENCES**

Adhikari R., et al., 2017, J. Cosmology Astropart. Phys., 2017, 025

Archidiacono M., Bohr S., Hannestad S., Helboe Jørgensen J., Lesgourgues J., 2017, J. Cosmology Astropart. Phys., 2017, 010

Archidiacono M., Hooper D. C., Murgia R., Bohr S., Lesgourgues J., Viel M., 2019, J. Cosmology Astropart. Phys., 2019, 055

Blas D., Lesgourgues J., Tran T., 2011, J. Cosmology Astropart. Phys., 2011, 034

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

Bezrukov C., Riazuelo A., Hansen S. H., Schaeffer R., 2002, Phys. Rev. D, 66, 083505
To determine the minimum scale at which we can trust our measurements of the power spectrum and the halo mass function in our simulations, we performed convergence tests for a few DM models using three resolution levels for each. These models cover representative regions of the parameter space we explore and, based on our analysis, they bracket the possible range of convergence variations. The convergence reported here is thus a fair representation of the convergence for all the parameter space explored in this paper. The three resolution levels were done within a $(6.25\, \text{Mpc} h^{-1})^3$ Lagrangian zoom region at $z = 5$ with $1024^3$ (LR), $2048^3$ (MR), and $4096^3$ (HR) effective particle resolution. We set the goal to determine for the two lower resolution levels, the wavenumber $k_{\text{conv}}$ at which the power spectrum differs by $5\%$ with respect to the highest resolution. The upper panel of Fig. A1 shows this convergence test for the models highlighted with large circles within the parameter space shown in the inset. Notice that since we subtract the shot-noise from the power spectra, the power falls off at the smallest scales. This behaviour is responsible for a smaller convergence scale (i.e. larger $k_{\text{conv}}$) for the WDM models (green and red lines) compared to the sDAO models; this also applies in general to all wDAO models. In models with a steep linear power spectrum cutoff, the non-linear true power is expected to be highly suppressed at sufficiently small scales and thus, there is not much power left at the unresolved scales. However, the sDAO case is similar to CDM, there is still significant power left at the unresolved scales and therefore, the $5\%$ convergence level is reached at larger scales (smaller $k_{\text{conv}}$). Despite this difference across different DM models, Fig. A1 shows that there is at least a factor of 3 improvement in $k_{\text{conv}}$ between the LR and MR simulations. The power spectra of the latter are converged to $k_{\text{conv}} > 250\, \text{hMpc}^{-1}$ and thus, assuming at least another factor of 2 improvement for the HR simulations, all models are converged to better than $5\%$ at $k \sim 500\, \text{hMpc}^{-1}$.

The bottom panel of Fig. A1 is equivalent to the top panel but for the halo mass function. It shows the ratio of the halo mass functions of the two lower resolution levels to that of the high resolution. It can be seen that the LR simulations drop below $5\%$ convergence at $\sim 2 \times 10^9 \, M_\odot / h$, while for the MR simulation this threshold occurs at $\sim 3 \times 10^9 \, M_\odot / h$. We highlight however, that for the $2.5\, \text{keV}$ model discreteness effects cause the well-known effect of spurious haloes (Wang & White 2007), which appear in this case at $M < 10^8 \, M_\odot / h$, dominating the signal. Although all models with a primordial power spectrum cutoff suffer from spurious haloes we find that in all cases, for our highest resolution, we are free from this effect at a halo mass of $10^8 \, M_\odot / h$. We therefore set this mass as our lower mass limit for all cases and report a convergence of the halo mass function to better
The goal in Section 3.1 is to connect two different parametrizations for the linear power spectrum. On the one hand, the parameters we have defined in this work \(\{n_{\text{peak}}, \kappa_{\text{peak}}, h_2\}\), and on the other the parameters used in Cyr-Racine et al. (2016), essentially \(\{a_n, n, \alpha_{\text{DR}}\}\). As we show in Section 4, the former can be used directly to quantify the differences between different structure formation models in the non-linear regime, while the latter can be connected directly to the particle physics parameters of a given model. In this Appendix we explore the connection between these two sets of parameters, which allows to obtain a physical interpretation of the final parametrization we have used in this work.

We start by recalling that in Cyr-Racine et al. (2016) the DR and DM-drag opacities for a given ETHOS model are respectively written as:

\[
\begin{align*}
\kappa_{\text{DR}}(z) &= -\eta_{\text{DR}} \left( 1 + \frac{z}{\eta_{\text{DR}}} \right)^n, \\
\kappa_{\text{1}}(z) &= \frac{4}{3} \Omega_{\text{DR}} h^2 \left( 1 + \frac{z}{\eta_{\text{DR}}} \right)^{n+1}.
\end{align*}
\]

if we just consider a single pure power law, where \(\Omega_{\text{DR}}\) and \(\Omega_{\text{DR}} h^2\) are the DM and DR densities in units of the critical density, \(h\) is the dimensionless Hubble constant, \(a_n\) and \(n\) are given by the particle physics model, and \(\eta_{\text{DR}}\) is an arbitrary normalization factor chosen to be \(\eta_{\text{DR}} = 10^3\). Eqs. (B1) is an approximation that assumes that a single power law is sufficient for a given value of \(n\), which is a good approximation for instance in the case of DM interacting with massless radiation via a massive mediator (models of this type were simulated in Vogelsberger et al. 2016). Assuming that DM and DR decouple well within the radiation dominated era (where the Hubble rate, relative to the conformal time goes as \(H = \eta^{-1}\); note that we use the conformal time in units of Mpc), Eqs. (B1) can be rewritten in terms of the conformal time, using \(1 + z \propto \eta^{-1}\),

\[
\begin{align*}
\kappa_{\text{DR}}(\eta) &= -\eta^{-1} \eta_{\text{DR}} \Omega_{\text{DR}} h^2 \left( 1 + \frac{\eta}{\eta_{\text{DR}}} \right)^n, \\
\kappa_{\text{1}}(\eta) &= -\eta^{-1} \eta_{\text{DR}} \Omega_{\text{DR}} h^2 \left( 1 + \frac{\eta}{\eta_{\text{DR}}} \right)^{n+1}.
\end{align*}
\]

where \(\eta_{\text{DR}}\) is the conformal time of DR decoupling defined by \(\int_{\eta_{\text{DR}}}^{\infty} -\kappa_{\text{DR}} d\eta = 1\), and \(\eta_{\text{DR}}\) is the conformal time when \(\kappa_{\text{DR}} = 0\). Note that this definition of the decoupling time takes the weakly coupled regime into account, i.e., a broad visibility function will lead to a later decoupling time. The DR decoupling time \(\eta_{\text{DR}}\) can be related directly with \(a_n\) and \(n\):

\[
\eta_{\text{DR}} = \left[ \frac{\Omega_{\text{DR}} h^2 a_n}{n - 1} \left( \frac{1}{H_0 \Omega_{\text{DR}} h^2 \left( 1 + \eta_{\text{DR}} \right)} \right)^n \right]^{\frac{1}{n+1}},
\]

where \(\Omega_{\text{DR}}\) is the radiation density in units of the critical density, and \(H_0\) is the Hubble constant today. In a similar way, we can define the conformal time for DM decoupling by \(\int_{\eta_{\text{DR}}}^{\infty} -\kappa_{\text{DR}} d\eta = 1\), and thus \(\eta_{\text{DR}}\) can be written as:

\[
\eta_{\text{DR}} = \left( \frac{n}{n - 1} \eta_{\text{DR}} \right)^{1/\eta_{\text{DR}}}. \tag{B4}
\]

We can use the tight coupling approximation (e.g. Hu &...
for the DM-DR plasma to roughly capture the acoustic oscillations in the DM fluid as well as the effect of DR diffusion damping. We then propose that the DM density fluctuations have the following $k-$dependence:

$$\delta_{\chi}(k) \approx \cos(kr_{\text{DAO}})e^{-k^2/k_{\text{D}}^2},$$  \hspace{1cm} (B5)

where the sound horizon scale is given by:

$$r_{\text{DAO}} = \frac{\int_{0}^{\eta_{\text{DR}}} \frac{d\eta}{\sqrt{6(1+\mathcal{R})}}} {\int_{0}^{\eta_{\text{DR}}} \frac{d\eta}{\sqrt{6(1+\mathcal{R}^2)}}},$$

where $c_\chi$ is the dimensionless DM sound speed. The parameter $k_{D}$ in Eq. (B5) gives the damping envelope of the DAOs, but due to a finite width of the visibility function, the damping envelope should be weighted by the DM visibility function $\mathcal{D}_{\text{DR}} = -\frac{\delta_{\text{DR}}}{\delta_{\text{DR}}^2}(e^{-k_{\text{D}}^2/k_{\text{D}}^2})$:

$$D_n(k) = \int_{0}^{\eta_{\text{DR}}} \frac{d\eta}{\sqrt{6(1+\mathcal{R})\mathcal{D}_{\text{DR}}}} \frac{4}{5\eta_0^2} \frac{\eta^2}{1+\mathcal{R}},$$

Therefore, under these approximations the linear transfer function is given by:

$$T^2(k) = \cos^2 \left(\frac{k \eta_0}{\sqrt{3}}\right) D_n^2(k),$$ \hspace{1cm} (B9)

where the input parameters $(\eta_{\chi}, \eta_{\text{DR}}, k_{D})$ are ultimately given by the cosmological parameters assumed $(\Omega_\chi, \Omega_\Lambda, H_0)$ and the set of values $(\Omega_{\text{DR}}, n, a_n, a_2)$ for a given ETHOS model.

Eq. (B9) results in damped DAOs that resemble the behaviour of the linear power spectrum of ETHOS models generated by the Boltzmann solver (see Fig. B1). This approach also predicts correctly that the amplitude of the DAOs scales with the value $n$. Unfortunately, the agreement is only qualitative, neither the position of the peaks nor the damping envelope agree with the full calculation. An example of this can be seen by comparing the red solid and dashed lines in Fig. B1, the latter of which is the ETHOS-2 model in Vogelsberger et al. (2016) (with $n = 4, a_4 = 1784.05 \text{ h } \text{Mpc}^{-1}$, $a_{\Omega_\chi} = \Omega_{\chi} = 3/2$; see their table 1).

The previous approach ignores the effect of the velocity perturbations on the density perturbation $\delta_{\chi}$, the so-called velocity overshoot. We improved the modelling by including this effect which is part of a more rigorous treatment (see Hu & Sugiyama 1996). Modelling the velocity perturbation as $\theta_{\chi} = -\sin(k r_{\text{DAO}}) D_{n}(k)$ moves the first DAO peak to better agree with the numerical results, but the other peaks and damping envelope were still not in agreement. Exchanging the sine and cosine functions for the full analytic solutions, in which case the potentials are given by Bessel functions, gives the evolution of $\delta_{\chi}$ and $\theta_{\chi}$ until DR decoupling with good accuracy. The position of the peaks of the DAOs are captured quite accurately with this modification, but the approach still does not capture the transition from the tightly coupled to the weakly coupled regimes correctly. The main reason for this seems to be that the exponential diffusion damping in Eq. (B8) is not an accurate representation of the numerical results for the first few DAO peaks at large scales; it is only a good approximation at much smaller scales. The damping of the first DAO peaks deviates strongly from the exponential behaviour because the timescale for DM decoupling is large compared to the oscillation frequency causing the DM to spend a longer time in the weakly coupled regime. Trying to model the DM decoupling timescale by weighting $\delta_{\chi}$ and $\theta_{\chi}$ with the visibility function, improves slightly the result of this analytical approach, but it remains inaccurate.

Instead of increasing the complexity of the modelling, which would eventually take us closer and closer to a full approach of the Boltzmann solver, but would defeat the purpose of having a simple physical interpretation, we choose instead to use a phenomenological approach as described in Section 3.1. The starting point is to notice that the ratio $k_{\chi}/H \sim (1 + \mathcal{R})$ is the relevant ratio of timescales (or length scales) in the DM decoupling process, with $k_{\chi}/H \gg 1$ corresponding to the tightly coupled regime and $k_{\chi}/H \ll 1$ to the decoupled regime. We found that the value of this ratio at the DM decoupling time $\eta_{\chi}$ is actually strongly correlated with the amplitude of the first DAO peak $h_{\text{peak}}$ (Fig. 7) since it is actually equal to $n$ as can be seen through Eqs. (B2-B4), which controls how fast the transition is from the tightly coupled to decoupled regimes, and thus how narrow the DM drag visibility function is.

Figure B1. The solid lines are the linear transfer function for different values of $n$ computed with Eq. (B9) using the a fixed value of $\eta_{\text{DR}}$ chosen to match that of the ETHOS-2 model in Vogelsberger et al. (2016) (dashed lines) and ETHOS-2. Note that by fixing $\eta_{\text{DR}}$, the value of $a_n$ for a given $n$ is given by Eq. (B3).

This paper has been typeset from a TeX/LaTeX file prepared by the author.