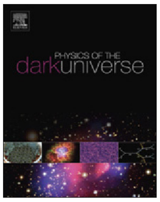
Physics of the Dark Universe 20 (2018) 65–71

Contents lists available at ScienceDirect

Physics of the Dark Universe

journal homepage: www.elsevier.com/locate/dark



A cosmological solution to the Impossibly Early Galaxy Problem

Manoj K. Yennapureddy [a], Fulvio Melia [b,*,1]

[a] Department of Physics, The University of Arizona, AZ 85721, USA
[b] Department of Physics, The Applied Math Program, and Department of Astronomy, The University of Arizona, AZ 85721, USA

ARTICLE INFO

Article history:
Received 26 October 2017
Received in revised form 18 March 2018
Accepted 19 March 2018

Keywords:
Cosmological parameters
Cosmological observations
Cosmological theory
Dark energy
Galaxies
Large-scale structure

ABSTRACT

To understand the formation and evolution of galaxies at redshifts $0 \lesssim z \lesssim 10$, one must invariably introduce specific models (e.g., for the star formation) in order to fully interpret the data. Unfortunately, this tends to render the analysis compliant to the theory and its assumptions, so consensus is still somewhat elusive. Nonetheless, the surprisingly early appearance of massive galaxies challenges the standard model, and the halo mass function estimated from galaxy surveys at $z \gtrsim 4$ appears to be inconsistent with the predictions of ΛCDM, giving rise to what has been termed "The Impossibly Early Galaxy Problem" by some workers in the field. A simple resolution to this question may not be forthcoming. The situation with the halos themselves, however, is more straightforward and, in this paper, we use linear perturbation theory to derive the halo mass function over the redshift range $0 \lesssim z \lesssim 10$ for the $R_{\rm h} = ct$ universe. We use this predicted halo distribution to demonstrate that both its dependence on mass and its very weak dependence on redshift are compatible with the data. The difficulties with ΛCDM may eventually be overcome with refinements to the underlying theory of star formation and galaxy evolution within the halos. For now, however, we demonstrate that the unexpected early formation of structure may also simply be due to an incorrect choice of the cosmology, rather than to yet unknown astrophysical issues associated with the condensation of mass fluctuations and subsequent galaxy formation.

1. Introduction

The structures we see today are believed to have grown gravitationally from tiny fluctuations in the primordial density field. Current theory holds that perturbations started to collapse once their density exceeded a certain critical value, forming bound objects that then assembled together with the surrounding gas and dust to form stars, galaxies and clusters. With dark matter particles decoupling first from the radiation, the early stages of structure formation proceeded principally through the condensation of dark matter halos. Baryonic particles subsequently accreted into the potential valleys created in this fashion once they themselves decoupled from the relativistic background.

The physics responsible for the formation of galaxies in this scenario is still not completely understood, but there is general consensus concerning the rate at which halos formed, specifically their number density distribution as a function of mass and redshift [1–3]. This halo mass function (as it is more commonly known) was first derived analytically by Press & Schechter [4] using several simplifying assumptions, including a spherically symmetric collapse model and a Gaussian initial density field. But though this analysis predicts a reasonable distribution, it nonetheless also underpredicts the number of high-mass halos and overpredicts the low-mass ones compared to detailed numerical simulations. More recently, Sheth & Tormen [1] have shown that this discrepancy may be mitigated by adopting an ellipsoidal collapse model rather than spherical. Even so, these analytical and semi-analytical approaches have for the most part been tested only against numerical simulations. Unfortunately, while Press–Schechter underpredicts the number of high-mass halos, Sheth–Tormen apparently overpredicts them, though a correction factor based on the linear growth rate may have been found. We shall describe this effect following Eq. (21) below. It is more difficult to test these semi-analytic approaches using actual observations because halos cannot be seen directly. The predicted halo distribution must be compared to the data indirectly, through the observation of the galaxy mass function, with an added assumption concerning the evolutionary relationship between them.

The observed halo formation was recently assessed [5] using several previous analyses to compare different techniques for relating the halo and galaxy distributions. For this purpose, these authors employed high redshift surveys in the redshift range $z \sim 4$–8, principally The Cosmic Assembly Near-infrared Deep Extragalactic Survey (CANDELS [6,7]) and the Spitzer Large Area Survey with

---

[*] Corresponding author.
E-mail addresses: manojy@email.arizona.edu (M.K. Yennapureddy), fmelia@email.arizona.edu (F. Melia).
[1] John Woodruff Simpson Fellow.

https://doi.org/10.1016/j.dark.2018.03.003

Hyper-Suprime-Cam (SPLASH [8]), to probe the galaxy luminosity and mass functions, from which the halo distribution may be derived. CANDELS is well suited to find the lower-mass galaxies because it represents a survey over a small area, whereas SPLASH has broad sky coverage and can therefore probe the more massive galaxies.

Obtaining the halo distribution and masses from the galaxy distribution presents quite a challenge. The best way to obtain halo masses from the spatial distribution of galaxies is via galaxy clustering methods [9,10], which do not assume any physical properties of the galaxies themselves, though they must assume a model for the dark matter concentration. Other techniques use the relationship between the luminosity and stellar masses, obtained from template fitting [11]. For example the "abundance matching technique" [12] relates one of the key features in the luminosity or mass function, such as the knee, to a feature in the halo mass function, and then matches the galaxy density and dark-matter halo density to derive halo masses over the whole mass range. Alternatively, one may also assume that the relations derived at low redshift using luminosity to dark-matter mass ratios still apply at high redshifts.

But though each of these techniques yields somewhat different outcomes in a quantitative sense, they all agree qualitatively [5]. These earlier findings show quite emphatically that the halo distribution estimated from galaxies at $z \gtrsim 4$ in the CANDELS and SPLASH surveys is inconsistent with the evolution of the halo mass function and the galaxy luminosity and mass functions predicted by standard ΛCDM [9,10,13,12]—a situation termed "The Impossibly Early Galaxy Problem" [5]. Various possible remedies were considered by these authors to reconcile the observed and predicted halo mass distributions, including possible errors introduced in calibrating the data using relations derived at lower redshifts, which may not be applicable for $z \gtrsim 4$. None of the remedies worked, however. If anything, this extended study showed that the high-redshift galaxies appear normal, suggesting that the relations derived at lower redshifts are probably also applicable at these higher redshifts.

The tension between the predictions of ΛCDM and the 'measured' halo mass function may be resolved with a better understanding of the underlying physics, e.g., regarding start formation and galaxy evolution. On the other hand, the current uncertain situation may simply be an indication that there are insurmountable problems with the use of ΛCDM as the background cosmology. In this paper, we will proceed under this assumption—i.e., that the problems elucidated by Steinhardt et al. [5] are real, and seek to find a solution to the surprisingly early formation of massive halos. To balance the discussion, however, we acknowledge the fact that this point of view is not universally accepted—a situation largely due to uncertainties in the simulations used to fully interpret the data and halo mass function.

Understanding the evolution of galaxies and their observational signatures, such as their UV luminosity or their redshift-dependent clustering, necessarily relies on modelling dark-matter evolution [14], and cosmological hydrodynamical simulations [15,16], complemented by analytical and semi-analytical calculations [17,18]. Complications arise in part because the observed UV luminosity function depends strongly on redshift (at least from $z \sim 4$ to 10), and various combinations of inputs and assumptions produce degenerate results [19,20]. It is fair to say that the degree of tension between the observations and predictions of the standard model depends on one's point of view.

But as noted, there are good reasons to suspect that real problems with the formation of structure do exist in the standard model. Some of these have to do with the unusually early appearance of supermassive black holes at $z \sim 6$–7 [21,22] and galaxies at $z \sim 10$–12 (see refs. cited in [23]). In addition, a rather compelling case may be made that a problem exists [5] based on the following points: (1) the halo mass function at $0 \lesssim z \lesssim 8$ is inferred using 3 or 4 different techniques, not just one, and all of the results agree at least qualitatively; (2) the fact that these techniques all require a blending of observational and simulational (i.e., model-dependent) factors to arrive at a mutually consistent picture measure that it is difficult to understand exactly what the results mean, because such an approach is very compliant to the assumptions one makes. For example, abundance matching forces agreement between observation and theory even in the absence of a strong physical motivation for the underlying model. The uncertainties (e.g., in how to match star-forming galaxy UV luminosities with halo formation in both mass and time) leave unresolved questions concerning how galactic evolution impacts our understanding of halo evolution. Nonetheless, forcing consistency between the observations and predictions of the standard model comes at a considerable cost.

One may understand this situation as follows. Much of the analysis in this paper is based on the 'standard' ratio of halo to stellar-mass, which arises from two considerations: First is the expectation that 10% of the baryonic matter condensed into stars [24]. Second, is the ratio of dark matter to baryonic matter, which is observed to be about 6:1 [25]. As we shall detail below, Steinhardt et al. [5] attempted to reconcile the disparity between theory and observation by introducing several modifications to the underlying physical processes. In order to fit the derived halo mass function in ΛCDM, however, they found that only a change by 0.8 dex in the ratio of dark matter to baryonic matter would suffice. But such a drastic change could come about only with a complete absence of dark matter at redshift 8, or if essentially 100% of the baryons condensed into stars at higher redshifts. Both of these scenarios constitute implausible physics, such as the need to convert all of the baryons into stars instantly upon halo virialization [5]. Other attempted remedies have equally unlikely requirements. So perhaps a better way to characterize the problem with the halo mass function is to say that it can only be made consistent with expectations of the standard model with the adoption of unlikely, new physics.

Given the unsettled debate concerning the formation and evolution of galaxies, we stress that our focus in this paper is not to model the galaxies themselves. We merely use some key observations of galactic profiles to infer the mass and time evolution of halos which, in principle, constitutes a much simpler, cleaner objective. For a complete assessment of problems with the formation of structure, it will eventually be necessary to study both the formation of halos and the galaxies within them, but this is a much more challenging analysis than we are attempting here. Such elaborate simulations for the formation and clustering of galaxies are outside the scope of the present paper. The outcome of this subsequent work will be reported elsewhere.

In the present context, the difficulty that ΛCDM has in accounting for the observed halo mass function has much in common with the growing tension between the measured cosmological growth rate, $b\sigma_8(z)$, and its value predicted by the standard model, particularly in the redshift range $0 < z < 1$, where a significant curvature expected in the functional form of $b\sigma_8(z)$ is absent in the data [26]. Admittedly, the errors in the measured values of $b\sigma_8(z)$ are still too large to rule out any model, but this is precisely why a comparison of the measured halo mass function with theory is very probative. If it turns out that *both* the halo distribution at high redshift and $b\sigma_8(z)$ at lower redshift are in tension with the growth of structure expected in standard cosmology, a compelling argument can be made that an alternative expansion scenario must be seriously considered. In this paper, we therefore compare the measured halo mass function, not only with the prediction of ΛCDM, but