

The Dearth of Halo Dwarf Galaxies: Is There Power on Short Scales?

Marc Kamionkowski

California Institute of Technology, Mail Code 130-33, Pasadena, California 91125

Andrew R. Liddle

*Astrophysics Group, Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BZ, United Kingdom
Isaac Newton Institute, University of Cambridge, Cambridge CB3 0EH, United Kingdom*

(Received 11 November 1999)

N-body simulations of structure formation with scale-invariant primordial perturbations show significantly more virialized objects of dwarf-galaxy mass in a typical galactic halo than are observed around the Milky Way. We show that the dearth of observed dwarf galaxies could be explained by a dramatic downturn in the power spectrum at small distance scales. This suppression of small-scale power might also help mitigate the disagreement between cuspy simulated halos and smooth observed halos, while remaining consistent with Lyman-alpha-forest constraints on small-scale power. Such a spectrum could arise in inflationary models with broken-scale invariance.

PACS numbers: 98.65.-r, 98.62.Gq, 98.80.Cq, 98.80.Es

The broad-brush picture painted by the inflation-inspired hierarchical-clustering paradigm accounts for the smoothness of the cosmic microwave background, its tiny temperature fluctuations, the flatness of the Universe, and the observed distribution of galaxies. However, finer inspection yields some possible—and possibly troubling—discrepancies between the models and observations that still need to be ironed out. One of these is the dearth of substructure in galaxy halos.

Recent high-resolution *N*-body simulations [1,2] have confirmed earlier analytic arguments [3] that suggested that hierarchical-clustering models should produce far more dwarf galaxies around the Milky Way than are observed. There are only 11 dwarf galaxies with internal velocity dispersions greater than 10 km sec^{-1} within the virial radius of the Milky Way halo. However, numerical simulations of structure formation in a hierarchical model indicate that a halo of the Milky Way's mass and circular speed should contain roughly an order of magnitude more dwarf galaxies. These theoretical results are robust to changes in the values of cosmological parameters or in the tilt of the primordial spectrum of perturbations.

One might at first be tempted to dismiss this discrepancy between theory and observations as a consequence of some nasty astrophysics. After all, the simulations consider only gravitational interactions and identify only virialized dark halos, while the real Universe is filled with gas, and dwarf galaxies are identified by their visible matter. It has been speculated that some of the missing halos might be identified with high-velocity clouds [1]. One might also guess that the halos remain invisible because the gas has been expelled by an early generation of supernovae, but even generous estimates of the efficiency of this process fall short of explaining the absence of luminous dwarf galaxies [4]. Even if the baryonic matter could somehow be driven out of the minihalos or kept dark, a fundamental problem would remain: we would still have trouble explaining how

a spiral disk could have formed in the strongly fluctuating potential of such a clumpy halo. This is perhaps the most troubling aspect of the predicted dwarf-halo abundance. See Refs. [1,2,5] for more detailed reviews of such arguments.

In the absence of an obvious astrophysical mechanism, it is natural to think of more exotic explanations. One strategy is to modify the nature of the dark matter, in order to prevent low-mass halos from forming. One option is that the dark matter is warm, for example, a neutrino of mass around a keV, which would suppress the formation of small-scale structure by free streaming out of potential wells [6]. However, such a neutrino would diminish power on scales $2h^{-1}$ to $12h^{-1}$ Mpc to a degree that may conflict with the power inferred from the Lyman-alpha forest [5,7]. Another possibility [5] is that the dark-matter particles interact strongly with each other, but not with ordinary matter [8]. However, the properties required of this particle (elastic-scattering cross sections of order 10^{-25} cm^2) are almost inconceivable in the predominant paradigms of weakly interacting massive particles [9] and axions [10].

Here we consider an alternative strategy: broken-scale invariance in the primordial power spectrum, arising from some feature in the inflaton potential. In slow-roll inflation, the amplitude of density perturbations on some comoving scale is proportional to $V^{3/2}/V'$, where V is the inflaton potential, and V' its derivative, when the comoving scale under consideration exits the horizon [11]. Suppose that V' is initially very small and that there is then a discontinuity in the second derivative, V'' . The slope V' will then jump, and the density-perturbation amplitude will drop steeply. In this way, the density-perturbation amplitude on some suitably small scale can be suppressed relative to the power on some larger scale. Such an idea has been invoked as a possible explanation of a claimed break in the power spectrum at ~ 100 Mpc scales [12,13] (and in an attempt produce non-Gaussian features [14]). Here,

we use this idea to account for the lack of halo substructure. Below, we show how a suppression of small-scale power can affect the dwarf-galaxy abundance and illustrate a particular inflationary model. We also argue briefly that the absence of small-scale power may help explain the discrepancy between simulated cuspy halos [15] and observed smooth halos.

For simplicity and clarity of presentation, we restrict our discussion here to an Einstein–de Sitter (EdS) model, but everything can readily be generalized to a low-density universe. Although we focus on an EdS cosmogony, we use a cold-dark-matter (CDM) power spectrum [16] with $\Gamma = 0.25$ (e.g., as in a τ CDM model [17]) to ensure that observational constraints from the shape and amplitude of the galaxy power spectrum, the cluster number density, and the COBE normalization are satisfied [18,19]. The solid curve in Fig. 1 shows this power spectrum, along with two other power spectra motivated later. The corresponding rms mass fluctuations are also shown.

We need to understand how changes to the power spectrum of density perturbations will affect the abundance of dwarf galaxies. We are interested in the abundance of minihalos of some given mass that exist within some larger halo of mass $2 \times 10^{12} M_{\odot}$, comparable to that of the Milky Way and the halo mass used by Moore *et al.* [2]. The simulations of Moore *et al.* show that the number of minihalos in the Galaxy today is very close to the number that existed in the protogalaxy; i.e., no more than a small fraction are fully disrupted during their subsequent orbital motion in the Milky Way halo. Thus, our task is simplified: we can calculate the abundance of minihalos in the protogalaxy that later became the Milky Way halo. To do so, we use the conditional mass function [20], given by

$$F(>M_{\text{small}}) = \text{erfc} \left[\frac{\delta_c z_f}{\sqrt{2[\sigma^2(M_{\text{small}}) - \sigma^2(M_{\text{big}})]}} \right]. \quad (1)$$

This equation gives the fraction of the mass in a galactic halo of total mass M_{big} that was in minihalos of mass greater than M_{small} at the redshift z_f at which the protogalaxy broke off from the expansion. Here $\sigma(M)$ is the linear-theory rms fractional mass fluctuation in spheres of radii that on average enclose a mass M , and $\delta_c = 1.7$ is the critical threshold for collapse.

It is the circular velocity of the dwarf galaxies, not the mass, which is observed. We relate the circular speed v_c to the mass by assuming the minihalos underwent collapse at the protogalaxy-formation epoch. Then the circular speed is obtained from $v_c^3 = 10MGH(z_f)$, where G is Newton's constant and $H(z)$ is the expansion rate at the collapse redshift z . To compare with the numerical results of Moore *et al.* [2], we consider a galactic halo of mass $M = 2 \times 10^{12} M_{\odot}$ with a circular speed of 220 km sec^{-1} . Using the conditional mass function, we obtain for the CDM power spectrum the cumulative number of halos

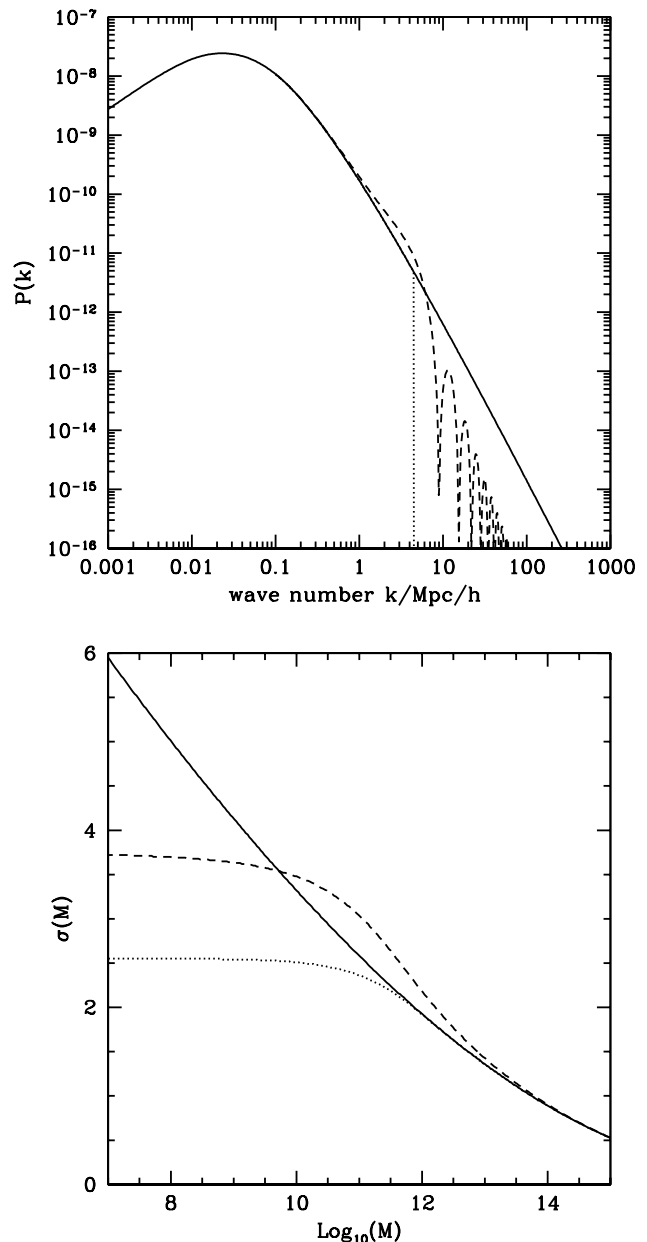


FIG. 1. The upper panel shows the power spectrum for $\Gamma = 0.25$ CDM (solid curve) for a model in which the power spectrum is arbitrarily cut off at $k = 4.5h \text{ Mpc}$ (dotted curve), and the broken-scale-invariance inflation model (dashed curve). The lower panel shows the rms mass fluctuation as a function of the enclosed mean mass M for these three models.

shown by the solid curve in Fig. 2 (cf. Fig. 1 in Ref. [2]). Although our calculation of the number of halos has several shortcomings, the good agreement with the numerical results shown in Fig. 1 of Ref. [2] clearly indicates that we are including the essential physics. In particular, we reproduce the order-of-magnitude excess of halos in the theoretical prediction, as compared to observations, for v_c/v_{global} in the range 0.05 to 0.1.

We now consider how the dwarf-galaxy abundance changes if the power spectrum is modified on short

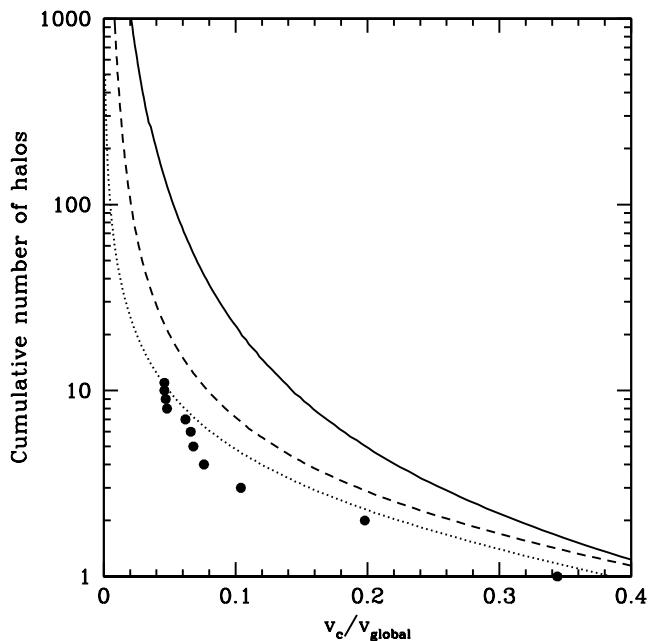


FIG. 2. The cumulative number of minihalos for the power spectra shown in Fig. 1 as a function of the circular speed v_c of the halo divided by the circular speed v_{global} of the galactic halo. The points show the Milky Way satellites. Compare with Fig. 1 in Ref. [2]. Line styles are as in Fig. 1.

scales. Clearly, if power is reduced on dwarf-galaxy scales ($\approx 10^{10} M_\odot$) relative to that at the galactic scale ($\sim 10^{12} M_\odot$), the number of dwarf galaxies will be suppressed. To investigate what is required, we first sharply cut off the power spectrum at $k = 4.5h \text{ Mpc}^{-1}$, shown by the dotted curve in Fig. 1; the lower panel shows how the rms mass fluctuations are modified. The corresponding prediction for halo substructure is shown as the dotted curve in Fig. 2 and is in striking agreement with the observations.

However, such a cutoff is purely phenomenological, and much stronger motivation is needed. In particular, the question arises as to whether inflation can produce such a power spectrum. The heuristic arguments given above suggest that the power spectrum can undergo a sharp drop on small scales if there is a discontinuity (or near discontinuity) in the slope of the potential, and this is known as the broken-scale-invariance (BSI) model. Remarkably, there is an exact solution for the power spectrum in this situation, due to Starobinsky [21], even though the slow-roll approximation does not apply. Further, this spectrum has a universal form, depending only on the change in the derivative across the discontinuity, in the sense that the form of the spectrum is preserved if the discontinuity is smoothed out, provided the smoothing is over a sufficiently short range of scalar-field values. The underlying reason is that the field briefly becomes kinetic-energy dominated as it goes over the discontinuity, so the precise form of the potential becomes irrelevant.

The dashed curve in Fig. 1 shows this power spectrum for a suitably chosen location and amplitude of the discontinuity (the amplitude is $p = 10$, in the notation of Ref. [13]). The universal form features a modest rise, followed by a series of oscillations asymptoting to the original spectral shape at a much lower amplitude; in this case the power is reduced by a factor $p^2 = 100$ on short scales. Such a model represents the fastest possible cutoff in power that can be obtained from a single-field inflation model. (We suspect that this is true in multifield models too, as additional fields will supply extra friction, slowing down the fields' evolution and hence stretching features in k space.)

This power spectrum gives rise to the dashed curve shown in Fig. 2. This model is almost as successful as the cutoff power spectrum; it produces roughly 20 minihalos with $v_c/v_{\text{global}} \gtrsim 0.04$, in much better agreement than the original power spectrum with the $\mathcal{O}(10)$ that are observed. We conclude that the BSI model provides a promising possibility for reconciling the predictions with observations.

Our modification to the power spectrum sets in only on very short scales, and so does not affect successes of the standard paradigm on much larger scales, such as the cluster number density or galaxy correlation function. However, we need to check that the loss of short-scale power is not inconsistent with object abundances at high redshift. The most powerful and direct constraint comes from the abundance of damped Lyman-alpha systems [22]. Observations exist at redshifts 3 and 4, giving comparable constraints. Redoing the calculation of Ref. [18] to include updated observations [23] gives a 95% confidence lower limit of

$$\sigma(10^{10} h^{-1} M_\odot, z = 0) > 2.75 + h, \quad (2)$$

in a critical-density universe, which is (just) satisfied by the BSI model (though not by the cutoff power spectrum).

This narrow escape makes it look as if the model is quite marginal. However, we have analyzed only the critical-density case, and the damped-Lyman-alpha constraint is much weaker in a low-density model. The observational constraint on $\sigma(M)$ at redshift 4 is almost the same (no more than 10% lower for Ω_0 values of interest [24]). However at redshift 4 suppression of the growth of perturbations has yet to set in, and the normalization to COBE (and/or cluster abundance) means σ is higher by a factor $1/\Omega_0$. With this additional factor all our models would easily pass the damped-Lyman-alpha system test where $\Omega_0 \sim 0.3$. Recent detailed calculations verify this conclusion [25].

There are two other key short-scale predictions worth considering. One is the power spectrum of perturbations obtained from the Lyman-alpha forest [7,26], which has recently been extended to shorter scales [27]. However the simulations of Ref. [25] show that such power can be generated by nonlinear effects, so these observations are unlikely to be in conflict with our theoretical spectra.

The second test is whether the model has early enough structure formation to reionize the Universe by a redshift 5, as required by the Gunn-Peterson test. Estimates using the technique of Ref. [28] would suggest this is marginal in the critical-density case. However, once again in the low-density case we benefit from the much higher early-time normalization of the power spectrum, which increases the predicted redshift by a factor $\approx 1/\Omega_0$. We thus conclude that this BSI model is likely to be consistent in the observationally favored low-density flat model, though a more careful investigation is merited.

The suppression of small-scale power may also help to explain the discrepancy between the cuspy inner halos observed in simulations and the smooth halos inferred from observations. Although some halo-formation models [29] would suggest that halos should be less cuspy with suppressed small-scale power, numerical simulations seem to contradict such arguments [30]. Even so, suppression of small-scale power might delay halo formation and thus lead to smaller halo central densities [31,32], and this could help mitigate the conflict between predicted and observed halo core densities. It would also help alleviate possible problems with disk kinematics and the zero point of the Tully-Fisher relation in CDM models [33]. More numerical work will be needed to assess the effects of this BSI model on halo central densities.

In conclusion, the arguments [1–3,15] suggesting that the standard structure-formation paradigm predicts too much substructure in galactic halos have now been refined to an extent that requires that the problem be taken very seriously and that radical answers to the discrepancy be considered. We have proposed here that the discrepancy indicates a dramatic lack of short-scale power in the primordial power spectrum. While such a sharp feature in the spectrum would not be expected *a priori*, it can be obtained from inflation, with the BSI model being the best available example. This model has already been invoked to introduce features on much larger scales of around $100h^{-1}$ Mpc [12,13], but the observational motivation for a short-scale break appears stronger. If our suggestion is correct, the lack of short-scale power will soon become apparent in new observations of high-redshift phenomena.

We thank B. Guiderdoni, B. Moore, J. Peacock, P. Steinhardt, and M. Steinmetz for discussions, and we thank the Isaac Newton Institute for hospitality. M.K. acknowledges the support of the DoE, NSF, NASA, and the Alfred P. Sloan Foundation.

-
- [1] A. A. Klypin, A. V. Kravtsov, and O. Valenzuela, *Astrophys. J.* **522**, 82 (1999).
 [2] B. Moore *et al.*, *Astrophys. J. Lett.* **524**, L19 (1999).
 [3] G. Kauffman, S. D. M. White, and B. Guiderdoni, *Mon. Not. R. Astron. Soc.* **264**, 201 (1993).

- [4] M.-M. Mac Low and A. Ferrara, *Astrophys. J.* **513**, 142 (1999).
 [5] D. N. Spergel and P. J. Steinhardt, *Phys. Rev. Lett.* **84**, 3760 (2000).
 [6] R. Schaeffer and J. Silk, *Astrophys. J.* **332**, 1 (1988); S. D. Burns, astro-ph/9711304; S. Colombi, S. Dodelson, and L. M. Widrow, astro-ph/9505029.
 [7] R. A. C. Croft *et al.*, *Astrophys. J.* **520**, 1 (1999).
 [8] E. D. Carlson, M. E. Machacek, and L. J. Hall, *Astrophys. J.* **398**, 43 (1992); M. E. Machacek, *Astrophys. J.* **431**, 41 (1994); A. A. de Laix, R. J. Scherrer, and R. K. Schaefer, *Astrophys. J.* **452**, 495 (1995).
 [9] G. Jungman, M. Kamionkowski, and K. Griest, *Phys. Rep.* **267**, 195 (1996).
 [10] M. S. Turner, *Phys. Rep.* **197**, 67 (1990); G. Raffelt, *Phys. Rep.* **198**, 1 (1990).
 [11] See, e.g., A. R. Liddle and D. H. Lyth, *Phys. Rep.* **231**, 1 (1993).
 [12] F. Atrio-Barandela *et al.*, *JETP Lett.* **66**, 397 (1997).
 [13] J. Lesgourgues, D. Polarski, and A. A. Starobinsky, *Mon. Not. R. Astron. Soc.* **297**, 769 (1998).
 [14] L. Wang and M. Kamionkowski, *Phys. Rev. D* **61**, 063504 (2000).
 [15] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **462**, 563 (1996); B. Moore *et al.*, *Astrophys. J. Lett.* **499**, L5 (1998).
 [16] J. M. Bardeen *et al.*, *Astrophys. J.* **304**, 15 (1986).
 [17] S. Dodelson, G. Gyuk, and M. S. Turner, *Phys. Rev. Lett.* **72**, 3754 (1994).
 [18] A. R. Liddle *et al.*, *Mon. Not. R. Astron. Soc.* **281**, 531 (1996).
 [19] S. Dodelson, E. Gates, and M. S. Turner, *Science* **274**, 69 (1996).
 [20] R. G. Bower, *Mon. Not. R. Astron. Soc.* **248**, 332 (1991); C. Lacey and S. Cole, *Mon. Not. R. Astron. Soc.* **262**, 627 (1993).
 [21] A. A. Starobinsky, *JETP Lett.* **55**, 489 (1992).
 [22] H. J. Mo and J. Miralda-Escudé, *Astrophys. J.* **430**, L25 (1994); G. Kauffman and S. Charlot, *Astrophys. J.* **430**, L97 (1994); C.-P. Ma and E. Bertschinger, *Astrophys. J.* **434**, L5 (1994).
 [23] L. J. Storrie-Lombardi, R. G. McMahon, and M. J. Irwin, *Mon. Not. R. Astron. Soc.* **283**, L79 (1996).
 [24] A. R. Liddle *et al.*, *Mon. Not. R. Astron. Soc.* **282**, 281 (1996).
 [25] M. White and R. Croft, astro-ph/0001247.
 [26] A. Nusser and M. Haehnelt, astro-ph/9906406.
 [27] P. McDonald *et al.*, astro-ph/9911196.
 [28] A. R. Liddle and D. H. Lyth, *Mon. Not. R. Astron. Soc.* **273**, 1177 (1995).
 [29] D. Syer and S. D. M. White, *Mon. Not. R. Astron. Soc.* **293**, 337 (1998); A. Nusser and R. Sheth, *Mon. Not. R. Astron. Soc.* **303**, 685 (1999); E. L. Lokas, astro-ph/9901185.
 [30] B. Moore *et al.*, astro-ph/9903164.
 [31] J. Navarro and M. Steinmetz, *Astrophys. J.* **528**, 607 (2000).
 [32] M. Steinmetz (private communication).
 [33] See, e.g., J. Navarro and M. Steinmetz, astro-ph/0001003.