HALO SUBSTRUCTURE AND THE POWER SPECTRUM

ANDREW R. ZENTNER

Department of Physics, Ohio State University, 174 West 18th Avenue, Columbus, OH 43210-1173; zentner@pacific.mps.ohio-state.edu

AND

JAMES S. BULLOCK

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138; jbullock@cfa.harvard.edu

Received 2003 April 19; accepted 2003 July 24

ABSTRACT

We present a semianalytic model to investigate the merger history, destruction rate, and survival probability of substructure in hierarchically formed dark matter halos and use it to study the substructure content of halos as a function of input primordial power spectrum. For a standard cold dark matter “concordance” cosmology (ΛCDM; n = 1, σ8 = 0.95) we successfully reproduce the subhalo velocity function and radial distribution profile seen in N-body simulations and determine that the rate of merging and disruption peaks ~10–12 Gyr in the past for Milky Way–like halos, while surviving substructures are typically accreted within the last ~0–8 Gyr. We explore power spectra with normalizations and spectral “tilts” spanning the ranges σ8 ~ 1–0.65 and n ~ 1–0.8, and include a “running-index” model with dα/dlnk = −0.03 similar to the best-fit model discussed in the first-year Wilkinson Microwave Anisotropy Probe (WMAP) report. We investigate spectra with truncated small-scale power, including a broken-scale inflation model and three warm dark matter cases with mw = 0.75–3.0 keV. We find that the mass fraction in substructure is relatively insensitive to the tilt and overall normalization of the primordial power spectrum. All of the CDM-type models yield projected substructure mass fractions that are consistent with, but on the low side, of published estimates from strong lens systems: f0 = 0.4%–1.5% (64th percentile) for subhalos smaller than 10^8 M⊙ within projected cylinders of radius r < 10 kpc. Truncated models produce significantly smaller fractions, f0 = 0.02%–0.2% for mw ~ 1 keV, and are disfavored by lensing estimates. This suggests that lensing and similar probes can provide a robust test of the CDM paradigm and a powerful constraint on broken-scale inflation/warm particle masses, including masses larger than the ~1 keV upper limits of previous studies. We compare our predicted subhalo velocity functions with the dwarf satellite population of the Milky Way. Assuming that dwarfs have isotropic velocity dispersions, we find that the standard n = 1 model overpredicts the number of Milky Way satellites at Vmax ≲ 35 km s⁻¹, as expected. Models with less small-scale power do better because subhalos are less concentrated and the mapping between observed velocity dispersion and halo Vmax is significantly altered. The running-index model, or a fixed tilt with σ8 ~ 0.75, can account for the local dwarfs without the need for differential feedback (for Vmax ≲ 20 km s⁻¹); however, these comparisons depend sensitively on the assumption of isotropic velocities in satellite galaxies.

Subject headings: cosmology: theory — dark matter — galaxies: formation — galaxies: halos — galaxies: structure

1. INTRODUCTION

In the standard cosmological model of structure formation (ΛCDM), the universe is dominated by cold, collisionless dark matter (CDM), made flat by a cosmological constant (Λ), and endowed with initial density perturbations via quantum fluctuations during inflation. The ΛCDM model with ΩM = 1 − ΩΛ = 0.3, h ≈ 0.7 and a scale-invariant spectrum of primordial perturbations [P(k) ∝ k^n, n = 1, σ8 ~ 0.9] is remarkably successful at reproducing a plethora of large-scale observations (e.g., Spergel et al. 2003; Percival et al. 2002). In contrast, several small-scale observations have proven more difficult to explain. Galaxy densities and concentrations appear to be much lower than what is predicted for the standard (n = 1) ΛCDM model (e.g., Debbattista & Sellwood 2000; Côte, Carignan, & Freeman 2000; Borriello & Salucci 2001; Binney & Evans 2001; Keeton 2001; van den Bosch & Swaters 2001; Marchesini et al. 2002; Swaters et al. 2003; McGaugh, Barker, & de Blok 2003; van den Bosch, Mo, & Yang 2003), and the Local Group dwarf galaxy count is significantly below what might naively be expected from the substructure content of ΛCDM halos (Klypin et al. 1999a, hereafter K99; Moore et al. 1999a). In Zentner & Bullock (2002, hereafter ZB02), we showed that the central densities of ΛCDM dark matter halos can be brought into reasonable agreement with the rotation curves of dark matter-dominated galaxies by reducing galactic-scale fluctuations in the initial power spectrum (σ8 ~ 0.75 and n ~ 0.9 is a good match; see Alam, Bullock, & Weinberg 2002, hereafter ABW02; McGaugh et al. 2003; van den Bosch et al. 2003). The present paper is an extension of this work. We explore how changes in the initial power spectrum affect the substructure content of ΛCDM halos, test our findings against attempts to measure the substructure mass fraction via gravitational lensing, and relate our results to the question of the abundance of dwarf satellites in the Local Group.

1 Hubble Fellow.
It is straightforward to see why CDM halos are expected to play host to a large number of distinct, bound substructures, or “subhalos.” In the modern picture of hierarchical structure formation (White & Rees 1978; Blumenthal et al. 1984; Kauffmann, White, & Guiderdoni 1993), low-mass systems collapse early and merge to form larger systems over time. Small halos collapse at high redshift, when the universe is very dense, so their central densities are correspondingly high. When these halos merge into larger hosts, their high densities allow them to resist the strong tidal forces that act to destroy them. While gravitational interactions do serve to unbind most of mass associated with merged progenitors, a significant fraction of these small halos survive as distinct substructure.

Our understanding of this process has increased dramatically in the last 5 years thanks to remarkable advances in N-body techniques that allow the high force and mass resolution necessary to study substructure in detail (Ghigna et al. 1998, 2000; Kravtsov 1999; K99; Klypin et al. 1999a; Kolatt et al. 1999; Moore et al. 1999a, 1999b; Font et al. 2001; Stoehr et al. 2002). For $n = 1$, CDM and CDM simulations, the total mass fraction bound up in substructure is measured at $f \sim 5\%-15\%$ (Ghigna et al. 1998; Klypin et al. 1999a), with a significant portion contributed by the most massive subsystems, $dn/dM \propto M^{-\alpha}$, $\alpha \approx 1.7$. The substructure content of halos seems to be roughly self-similar when subhalo mass is scaled by the host halo mass (Moore et al. 1999a) and the subhalo count is observed to decline at the host halo center, where tidal forces are strongest (Ghigna et al. 1998; Colin, Klypin, & Krastov 2000b; Chen, Kravtsov, & Keeton 2003).

Unfortunately, studies of substructure using N-body simulations suffer from issues of numerical resolution. Simulations with the capability to resolve substructure are computationally expensive. They cannot be used to study the implications of many unknown input parameters and cannot attain both the resolution and the statistics needed to confront observational data on substructure that appear to be on the horizon. Even state of the art simulations face difficulties in the centers of halos, where “overmerging” may be a problem (e.g., Chen et al. 2003; Klypin et al. 1999a) and measurements of the substructure fraction via lensing are highly sensitive to these uncertain, central regions. Our goal is to present and apply a semianalytic model that suffers from no inherent resolution effects and is based on the processes that were observed to govern substructure populations in past N-body simulations. This kind of model can generate statistically significant predictions for a variety of inputs quickly and can be used to guide expectations for the next generation of N-body simulations. Conversely, this model represents in many ways an extrapolation of N-body results into unexplored domains, and it is imperative that our results be tested by future numerical studies. In the present paper, we aim to explore the effect of the power spectrum on the population of surviving subhalos, but in principle these methods are suitable for testing substructure ramifications for a variety of cosmological inputs.

One of the main motivations for this work comes from simulation results that indicate that galaxy-sized CDM halos play host to hundreds of subhalos with maximum circular velocities in the range $10 \text{ km s}^{-1} \lesssim V_{\text{max}} \lesssim 30 \text{ km s}^{-1}$. The Milky Way, as a comparative example, hosts only 11 dwarf satellites of similar size. This “dwarf satellite problem” specifically refers to the gross mismatch between the predicted number of $\Lambda$CDM subhalos and the count of satellite galaxies in the Local Group (K99; Moore et al. 1999a; Font et al. 2001; see also Kauffmann et al. 1993, who indicated that there may be a problem using analytic arguments). The dwarf satellite problem and other small-scale issues led many authors to consider modifications to the standard framework. If the dark matter were “warm” (Pagels & Primack 1982; Colombi, Dodelson, & Widrow 1996; Hogan & Dalcanton 2000; Colin et al. 2000a; Bode, Ostriker, & Turok 2001; Lin et al. 2001; Knebe et al. 2002) or if the primordial power spectrum were sharply truncated on small scales (Starobinsky 1992; Kamionkowski & Liddle 2000), then subgalactic-scale problems may be alleviated without vitiating the overall success of CDM on large scales. Another possibility is that CDM substructure is abundant in all galaxy halos but that most low-mass systems are simply devoid of stars. An intermediate solution may involve a simple modification of the assumed primordial spectrum of density perturbations that gradually lowers power on galactic scales relative to the horizon, e.g., via tilting the power spectrum.

Probing models with low galactic-scale power is motivated not only by the small-scale crises facing standard CDM but also by more direct probes of the power spectrum. While many analyses continue to measure “high” values for $\sigma_8 \sim 1$ (Van Waerbeke et al. 2002; Komatsu & Seljak 2002; Bahcall & Bode 2003) where $\sigma_8$ is the linear, rms fluctuation amplitude on a length scale of $8 h^{-1} \text{ Mpc}$, numerous recent studies relying on similar techniques advocate rather “low” values of $\sigma_8 \sim 0.7$–0.8 (Jarvis et al. 2003; Bahcall et al. 2003; Schuecker et al. 2003; Pierpaoli et al. 2003; Viana, Nichol, & Liddle 2002; Brown et al. 2003; Allen et al. 2003; Hamana et al. 2003; Melchiorri & Silk 2002; Borgani et al. 2001). Similarly, the Ly$\alpha$ forest measurements of the power spectrum are consistent with reduced galactic-scale power (Croft et al. 1998, 2002; McDonald et al. 2000). Set against the normalization of fluctuations on large scales implied by the Cosmic Background Explorer (COBE) measurements of cosmic microwave background (CMB) anisotropy (Bennett et al. 1994), these data suggest that the initial power spectrum may be tilted to favor large scales with $n < 1$.

The recent analysis of the measurements of CMB anisotropy presented by Spergel et al. (2003; see also Verde et al. 2003; Peiris et al. 2003) returns a best-fit spectral index to a pure power law primordial spectrum of $n = 0.99 \pm 0.04$ when only the WMAP data are considered. However, when data from smaller scale CMB experiments, the Two Degree Field (2Df) Galaxy Redshift Survey, and the Ly$\alpha$ forest are included, the analysis favors a mild tilt, $n = 0.96 \pm 0.02$. Interestingly, all of the data sets together yield a better fit if the index is allowed to run: the WMAP team find $dn/dlnk = -0.031 \pm 0.016$. This result is consistent with no running at $\sim 2 \sigma$, and the statistical significance is further weakened when additional uncertainties in the mean flux decrement in the Ly$\alpha$ forest are considered (Seljak, McDonald, & Makarov 2003; Croft et al. 2002), yet such a model certainly seems worth investigating, especially in light of the small-scale difficulties that it may help to alleviate.

We explicitly show how models with reduced small-scale power are expected to help the halo density problem in Figure 1, which is an updated version of Figure 5 in ZB02.
DK01 quoted an approximate upper mass limit of $10^6$–$10^9 M_{\odot}$. They have since concluded that an upper limit of $\sim 10^8$–$10^{10} M_{\odot}$ may be more appropriate (N. Dalal 2003, private communication).

3 See http://online.kitp.ucsb.edu/online/galaxy_c00/white.
models with later structure formation tend to produce halos with more slowly rising rotation curves. If a dwarf galaxy sits in a halo with a slowly rising rotation curve that peaks at $V_{\text{max}} > 50$ km s$^{-1}$, the conversion factor, and thus $V_{\text{max}}$, can be significantly larger. Shifts of this kind in the observed velocity function change the implied velocity (or mass) scale of discrepancy and influence our ideas about the type of feedback that gives rise to the mismatch.

Hayashi et al. (2003, H03) and Stoehr et al. (2002, hereafter S02) suggested that substructure halos experience significant mass redistribution in their centers as a result of tidal interactions and that they are therefore less concentrated than comparable halos in the field. They argue that when this is taken into account, the dwarf satellite mismatch sets in at $V_{\text{max}} \sim 20$ km s$^{-1}$, and that the transition is sudden—below this scale all halos are devoid of observable galaxies. While these conclusions have yet to be confirmed and are dependent on subhalo merger histories and the isotropy of dwarf velocity dispersions, they highlight the need to refine our predictions about halo substructure. They also motivate us to explore how minor changes in cosmological parameters can influence our interpretation of the dwarf satellite problem.

In the remainder of this paper we present our study of CDM substructure. In § 2 we describe our semianalytic model, provide some illustrative examples, and compare our results for standard ΛCDM with previous $N$-body results. In § 3 we briefly describe the input power spectra that serve as the basis for this study. In § 4 we present our results on subhalo mass functions and velocity functions. We make predictions aimed at measuring substructure mass fractions via gravitational lensing and address the dwarf satellite problem in light of some of our findings in this section. In § 5 we discuss some shortcomings of our model and how they might be improved on in future work. In § 6 we summarize our work and draw conclusions from our results. In this study we vary only the power spectrum and work within the context of the so-called concordance cosmological model with $\Omega_M = 0.3$, $\Omega_{\Lambda} = 0.7$, $h = 0.72$, and $\Omega_B h^2 = 0.02$ (e.g., Turner 2002; Spergel et al. 2003).

## 2. Modeling Halo Substructure

In order to determine the substructure properties of a dark matter halo we must model its mass accretion history as well as the orbital evolution of the subsystems once they are accreted. For the first step, we rely on the extended Press-Schechter (EPS) formalism to create merger histories for each host system. We give a brief description of our EPS merger trees in § 2.1. In § 2.2 we discuss our model for the density structure of accreted halos and the host system, and in § 2.3 we describe our method for following the orbital evolution of each merged system. We show tests and examples of this model in § 2.4.

### 2.1. Merger Histories

We track diffuse mass accretion and satellite halo acquisition of host systems by constructing merger histories using the EPS method (Bond et al. 1991; Lacey & Cole 1993, hereafter LC93). In particular, we employ the merger tree algorithm of Somerville & Kolatt (1999, hereafter SK99). This allows us to generate a list of the masses and accretion redshifts of all subhalos greater than a given threshold mass that merged to form the host halo. We describe the method briefly here and encourage the interested reader to consult LC93 and SK99 for further details.

A merger tree that reproduces many of the results of $N$-body simulations can be constructed using only the linear power spectrum. For convenience, we express this in terms of $\sigma(M)$, the rms fluctuation amplitude on mass scale $M$ at $z = 0$. As in LC93, let $S(M) \equiv \sigma^2(M)$, $\Delta S \equiv S(M) - S(M + \Delta M)$, $w(t) \equiv \delta_c(t)$, and $\delta w \equiv w(t) - w(t + \Delta t)$. Here $\delta_c(t)$ is the linear overdensity for collapse at time $t$ associated with our choice of cosmology (see LC93 or White 1996). The probability that a halo of mass $M$, at time $t$, accreted an amount of mass associated with a step of $\Delta S$ in a given time step implied by $\delta w$ is

$$P(\Delta S, \delta w)d(\Delta S) = \frac{\delta w}{\sqrt{2\pi\Delta S^3/2}} \exp \left[ \frac{-\delta w^2}{2\Delta S^2} \right] d(\Delta S). \quad (1)$$

Merger histories are constructed by starting at a chosen redshift and halo mass and stepping back in time with an appropriate time step. SK99 tell us that if the minimum mass of a progenitor that we wish to track is $M_{\text{min}}$, then each time step must be small in order to reproduce the conditional mass functions of EPS theory: $\delta w \lesssim \{M_{\text{min}}/[dS(M)/dM]\}^{1/2}$.

We build merger trees by selecting progenitors at each time step according to equation (1) and treating events with $\Delta M < M_{\text{min}}$ as diffuse mass accretion. At each step we identify the most massive progenitor with the host halo and all less massive progenitors with accreted subhalos, and we continue this process until the host mass falls below $M_{\text{min}}$. In practice, we use a slightly modified version of the SK99 scheme. At each stage we demand that the number of progenitor halos in the mass range that we consider be close to the mean value. As discussed in BKW00, this modification considerably improves the agreement between the analytically predicted progenitor distribution and the numerically generated progenitor distribution. In what follows we set $M_{\text{min}} = 10^7 M_\odot$. Our fiducial $z = 0$ host mass is $1.4 \times 10^{12} M_\odot$, but we vary these choices in order test sensitivity to the host mass and redshift.

### 2.2. Halo Density Structure

Whether a merged system survives or is destroyed depends on the density structure of the subhalo and on the gravitational potential of the host system. Therefore, it is worthwhile to describe our assumptions about CDM density profiles in some detail. The size of a virialized dark matter halo can be quantified in terms of its virial mass $M_{\text{vir}}$, or equivalently its virial radius $R_{\text{vir}}$, or virial velocity $V_{\text{vir}}^2 \equiv GM_{\text{vir}}/R_{\text{vir}}$. The virial radius of a halo is defined as the radius within which the mean density is equal to the virial overdensity $\Delta_{\text{vir}}$ multiplied by the mean matter density of the universe $\rho_M$, so that $M_{\text{vir}} = 4\pi \rho_M \Delta_{\text{vir}}(z) R_{\text{vir}}^3/3$. The virial overdensity $\Delta_{\text{vir}}$ can be estimated using the spherical top-hat collapse approximation and is generally a function of $\Omega_M$, $\Omega_{\Lambda}$, and redshift (e.g., Eke, Navarro, & Frenk 1998). We compute $\Delta_{\text{vir}}$ using the fitting function of Bryan & Norman (1998). In the cosmology considered here, $\Delta_{\text{vir}}(z = 0) \simeq 337$, and at high redshift $\Delta_{\text{vir}} \rightarrow 178$, approaching the standard cold dark matter (i.e., $\Omega_M = 1$) value.
The gross structure of dark matter halos has been described by several analytic density profiles that have been proposed as good approximations to the results of high-resolution \( N \)-body simulations (Moore et al. 1999b; Power et al. 2003). In the interest of simplicity, we choose to model all halos with the density profile proposed by Navarro, Frenk, & White (1997, hereafter NFW):

\[
\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}.
\]

For the NFW profile, the amount of mass contained within a radius \( r \) is

\[
M(< r) = M_{\text{vir}} \frac{g(x)}{g(c_{\text{vir}})}
\]

where \( x \equiv r/r_s \), \( g(y) \equiv \ln(1 + y) - y/(1 + y) \) and the concentration parameter is defined as \( c_{\text{vir}} \equiv R_{\text{vir}}/r_s \). Restating equation (3) in terms of a circular velocity profile yields \( V^2_\text{circ}(r) = V^2_{\text{circ}}(c_{\text{vir}})g(x)/g(c_{\text{vir}}) \). The maximum circular velocity occurs at a radius \( r_{\text{max}} \approx 2.16r_s \), with a value \( V^2_{\text{max}} \approx 0.216V^2_{\text{circ}}/g(c_{\text{vir}}) \).

As a result of the study by Wechsler et al. (2002, hereafter W02) and several precursors (e.g., Zaroubi & Hoffman 1993; NFW; Avila-Reese, Firmani, & Hernández 1998; Bullock et al. 2001, hereafter B01), we now understand that dark matter halo concentrations are determined almost exclusively by their mass assembly histories. The gross structure of dark matter halos has been observed trends with halo mass and redshift using a simulation by ENS01 for WDM halos with masses smaller than the “free-streaming” mass (see § 3). The four WDM halos simulated by ENS01 with masses small enough to be appreciably affected by free-streaming all had \( c_{\text{vir}} \) values that were \( \approx 2 \sigma \) lower than the B01 model. On the basis of these data, ENS01 proposed a model in which halo collapse time depends not only on the amplitude of the power spectrum \( \sigma(M) \), but on an effective overdensity amplitude, \( \sigma_{\text{eff}} \equiv -\sigma(M) \ln \sigma(M)/d \ln M \). This results in a \( c_{\text{vir}}(M) \) relation that increases with mass for masses smaller than the truncation scale and decreases at larger masses as in \( \Lambda \)CDM. By defining an effective overdensity in this way, ENS01 were able to account for the low \( c_{\text{vir}} \) values observed in their \( \Lambda \)CDM simulations. The slope of the \( c_{\text{vir}}-M_{\text{vir}} \) relation for ENS01 is shallower than the slope predicted by the B01 relation, and therefore the ENS01 model also leads to less concentrated halos at small mass (\( M \lesssim 10^{10} M_\odot \)) even for identical input power spectra. This disparity grows larger when tilted and/or running spectra are considered, as in this paper.

Unfortunately, the ENS01 model cannot be applied in the WDM cases we explore because in these models \( \sigma(M) \) is very flat on scales smaller than the free-streaming mass and the ENS01 model breaks down when \( d\sigma(M)/dM \) becomes very small. In the ENS01 model, WDM halos smaller than \( \approx 1\% \) of the free-streaming mass never collapse because \( \sigma_{\text{eff}} \ll 1 \). In addition to this practical problem, the ENS01 predictions are not supported by the results of Avila-Reese et al. (2001) and Colin et al. (2000a). Using \( \approx 25 \) halos, Avila-Reese et al. found WDM halo concentrations to be roughly constant with mass down to several orders of magnitude below the free-streaming scale, in accordance with the B01 model predictions. In light of these difficulties and the discordant results of different \( N \)-body studies, we have not explored the implications of the ENS01 model in this work. This is not an indictment of the ENS01 model. Rather, the results of ENS01 highlight the uncertainty in assigning halo concentrations to low-mass systems, especially with power spectra that vary rapidly with scale.

---

4 Preliminary results from new simulation data show promising agreement with the B01 model all the way down to \( M \approx 10^7 M_\odot \). (P. Colin 2003, private communication.)
Our choice of the B01 relation is a matter of pragmatism and represents a conservative choice in that halos are assigned the higher of the two predictions of $c_{\text{vir}}$ at small mass. Lower $c_{\text{vir}}$ values (in line with ENS01 expectations) would result in less substructure and larger deviations from the standard ΛCDM model than the predictions in §4.

2.3. Orbital Evolution

With the accretion history of the host halo in place and with a recipe in hand that fixes the density structure of host and satellite halos, the next step is to track the orbital evolution of accreted systems. This is necessary in order to account for the effects of dynamical friction and mass loss due to tidal forces. These processes cause most of the accreted subhalos either to sink to the center of the host halo and become “centrally merged” or to lose most of their mass and be “tidally disrupted” and no longer identifiable as distinct substructure. We model these effects using an improved version of the BKW00 technique, borrowing heavily from the dynamical evolution model proposed by Taylor & Babul (2001, hereafter TB01; see also Taylor & Babul 2003a) and the dynamical friction studies of Hashimoto, Funato, & Makino (2003, hereafter HFM03) and Valenzuela & Klypin (2003; see also O. Valenzuela & A. Klypin, in preparation).

We denote the mass of an accreted subhalo as $M_{\text{sat}}$, its outer radius as $R_{\text{sat}}$, and the accretion redshift as $z_{\text{acc}}$. We set the subhalo concentration to the median value given by the B01 model for this mass and redshift. Although initially set by the virial mass and radius of the in-falling subhalo, $M_{\text{sat}}$ and $R_{\text{sat}}$ are allowed to evolve with time, as described in more detail below. We track the orbit of each subhalo in the potential of its host from the time of accretion $t_{\text{acc}}$ until today ($t_0 \approx 13.5$ Gyr in this cosmology) or until it is destroyed. The mass accretion history also yields the host halo mass at each time step. We fix the density profile of the host at each accretion time using the median B01 expectation for a halo of that mass. As we mentioned earlier, the scale radius and central density of the host remain approximately constant.

For the purpose of tracking each subhalo orbit, we assume the host potential to be both spherically symmetric and static. We update the host halo profile using the B01 model for each accretion event, but hold it fixed while each orbit is integrated. While the approximation of a static host potential for each orbit is not ideal, it allows for an extremely simple prescription that significantly reduces the computational expense of our study. Moreover, this approximation is not bereft of physical motivation. As we discussed above, halos observed in numerical simulations appear to form dense central regions early in their evolution after which their scale radii and central densities remain roughly fixed with time. Additionally, we have run test examples that include an evolving halo potential (set by the results of W02) and find that this addition has a negligible effect on the statistical properties of substructure that we are concerned with here.

Upon accretion onto the host, each subhalo is assigned an initial orbital energy based on the range of binding energies observed in numerical simulations (K99; A. V. Kravtsov 2002, private communication). We place each satellite halo on an initial orbit of energy equal to the energy of a circular orbit of radius $R_{\text{circ}} = \eta R_{\text{vir}}$, where $R_{\text{vir}}$ is the virial radius of the host at the time of accretion and $\eta$ is drawn randomly from a uniform distribution on the interval [0.4, 0.75]. We assign each satellite an initial specific angular momentum $J = \epsilon J_{\text{circ}}$, where $J_{\text{circ}}$ is the specific angular momentum of the aforementioned circular orbit and $\epsilon$ is known as the “orbital circularity.” Past studies drew $\epsilon$ from a uniform distribution on the interval [0.1, 1] (BKW00) to match the circularity distribution of surviving subhalos in simulations reported by Ghigna et al. (1998). However, the orbits of surviving halos are biased relative to the orbits of all accreted systems because subhalos on radial orbits are preferentially destroyed. We find that we better match the Ghigna et al. (1998) result for surviving satellites if we draw the initial $\epsilon$ from the simple, piecewise-linear distribution depicted in Figure 2. The initial radial position of each subhalo is set to $R_{\text{ini}} = R_{\text{circ}}$ and for all noncircular orbits, we set the subhalo to be initially in-falling so that $dR/dt < 0$.

To calculate the trajectories of subhalos, we treat them as point masses under the influence of the NFW gravitational potential of the host halo. We model orbital decay by dynamical friction using the Chandrasekhar formula (Chandrasekhar 1943). The Chandrasekhar formula was derived in the context of a highly idealized situation; however, numerical studies indicate that this approximate relation can be applied more generally (e.g., Valenzuela & Klypin 2003 have performed a new test that supports the use of this approximation). Using the Chandrasekhar approximation, there is a frictional force exerted on the

---

5 The exception to this is the case of a late-time merger of halos of comparable mass in which case the central densities and scale radii of the participating halos may change considerably (W02).
subhalo that points opposite to the subhalo velocity:

\[ F_{\text{DF}} \approx \frac{4\pi}{V_{\text{orb}}} \frac{\ln(\Lambda)G^2 M^2_{\text{sat}} \rho(r)}{r} \left[ \text{erf}(X) - \frac{2X}{\sqrt{\pi}} \exp(-X^2) \right]. \tag{4} \]

In equation (4), \( \ln(\Lambda) \) is the Coulomb logarithm, \( r \) is the radial position of the orbiting satellite, and \( \rho(r) \) is the density of the host halo at the satellite radius. The quantity \( V_{\text{orb}} \) is the orbital speed of the satellite halo and \( X \equiv V_{\text{orb}} / (2\sigma) \), where \( \sigma \) is the one-dimensional velocity dispersion tensor, which is the dispersion of particles in the host halo. For an NFW profile, the one-dimensional velocity dispersion can be determined using the Jeans equation. Assuming an isotropic velocity dispersion

\[ \sigma^2 \left( \frac{x = r}{r_s} \right) = \frac{V_{\text{vir}}^2}{g(c_{\text{vir}})} x(1 + x)^2 \int_x^\infty \frac{g(x')}{x'^3(1 + x')^2} \, dx'. \tag{5} \]

We find the following approximation useful and accurate to 1% for \( x = 0.01 - 100\):

\[ \sigma(x) \approx V_{\text{vir}} \frac{1.4393x^{0.354}}{1 + 1.1756x^{0.725}}. \tag{6} \]

There has been much debate on the appropriate way to assign the Coulomb logarithm in equation (4). Dynamical friction is caused by the scattering of background particles into an overdense “wake” that trails the orbiting body and tugs back on the scatterer. The Coulomb logarithm is interpreted as \( \ln(h_{\text{max}} / h_{\text{min}}) \), where \( h_{\text{max}} \) is the maximum relevant impact parameter at which background particles are scattered into the wake and \( h_{\text{min}} \) is the minimum relevant impact parameter. A common approximation is to choose a constant value of the Coulomb logarithm (perhaps by calibrating to the results of numerical experiments as in TB01), but some studies indicate that this approach significantly underestimates the dynamical friction timescale. As such, we may expect this to be true for the typical timescale of order orbital period, \( T \), for which we expect to be relevant timescale for tidal mass stripping.

As the satellite orbits within the host potential, it is stripped of mass by the tidal forces that it experiences. First, we estimate the instantaneous tidal radius of the subhalo, \( r_t \), at each point along its orbit. In the limit that the satellite is much smaller than the host, the tidal radius is given by the solution to the equation (von Hoerner 1957; King 1962)

\[ r_t^3 \approx \frac{M_{\text{sat}}(<r_t)}{M_{\text{host}}(<r)} \frac{2 + \omega^2 R^2/GM_{\text{host}}(<r)}{\ln M_{\text{host}}(<r)/\ln r^3}, \tag{10} \]

where \( r \) is the radial position of the satellite, \( M_{\text{host}}(<r) \) is the host’s mass contained within this radius (see eq. [3]), \( M_{\text{sat}}(<r_t) \) is the satellite’s mass contained within \( r_t \), and \( \omega \) is the instantaneous angular speed of the satellite. Equation (10) is merely an estimate of the satellite’s tidal limit. For a satellite on a circular orbit, it represents the distance from the satellite center to the point along the line connecting the satellite and the host halo center where the tidal force on a test particle just balances the attractive force of the satellite. In reality, the tidal limit of a satellite cannot be represented by a spherical surface; some particles within \( r_t \) will be unbound while others without may be bound. Nevertheless, TB01 showed that this can serve as a very useful approximation.

As the tidal radius shrinks, unbound mass in the periphery is stripped. Tidal forces are strongest, and \( r_t \) smallest, when the orbit reaches pericenter; however, all of the mass outside of \( r_t \) is not stripped instantaneously, but after a timescale set by the change in the host halo potential on the length scale of the orbiting satellite (which does not vary monotonically with time), we set its value by determining the radius within which the mass profile retains the appropriate bound mass using equation (3).

We fix the scale radius of the subhalo, \( r_{\text{sat}} \), at the value defined at the epoch of accretion.

Our approximations for dynamical friction and tidal stripping are least accurate when the mass of the satellite is not very small compared with the mass of the host. However, as \( F_{\text{DF}} \propto M^2_{\text{sat}} \), it is in precisely these cases that we expect the satellite to merge quickly with the host and no longer be identifiable as distinct substructure. As such, the precise dynamics should not have a significant effect on our results in these cases, particularly because our main
predictions involve low-mass substructure. However, more detailed modeling will be important for investigations that focus on more massive substructures, for example, explorations that use disk thickening as a test of the ΛCDM cosmological model (e.g., Font et al. 2001).

The final ingredients for our semianalytic model of halo substructure are the criteria for declaring subhalos to be tidally disrupted and centrally merged. Let \( r_{\text{sat}} \) be the radius at which the subhalo’s initial velocity profile attains its maximum, and \( M_{\text{sat}}(< r_{\text{sat}}) \) be the mass of the satellite originally contained within the radius \( r_{\text{sat}} \). We declare a subhalo to be centrally merged with the host if its radial position relative to the center of the host becomes smaller than \( r_{\text{sat}} \). We declare a satellite tidally disrupted if the mass of the satellite becomes less than \( M_{\text{sat}}(< r_{\text{sat}}) \). This criterion is partially motivated by the numerical study of H03, who find that NFW subhalos are completely tidally destroyed short after \( r_{t} \) becomes less than \( r_{\text{sat}} \). Of course, the distinction between centrally merged and tidally destroyed satellites is somewhat arbitrary as subhalos are typically severely tidally disrupted as they approach the center of the host potential. Fortunately, for the issues we explore here, the precise nature of a satellite halo’s destruction is not important. We discuss this issue further in a forthcoming extension of our work (A. Zentner & J. Bullock, in preparation).

In reporting results concerning the velocity function of substructure, we invoke a further modification. H03 noted that subhalos that experienced significant tidal stripping suffered not only mass loss at radii \( \gtrsim r_{t} \), but mass redistribution in their central regions, at radii smaller than \( r_{t} \). To account for this, we determine whether or not the tidal radius of each surviving subhalo was ever less than \( r_{\text{sat}} \). If so, we follow the prescription of H03 to account for mass redistribution and scale the maximum circular velocity of the satellite via

\[
V_{\text{final}}^{\text{max}} = \left( \frac{M_{\text{final}}^{\text{sat}}}{M_{\text{initial}}^{\text{sat}}} \right)^{1/3} V_{\text{initial}}^{\text{max}},
\]

where \( V_{\text{initial}}^{\text{max}} \) is the maximum circular velocity of the satellite according to its initial density profile, \( M_{\text{final}}^{\text{sat}} \) is its final mass, and \( M_{\text{initial}}^{\text{sat}} \) is its initial mass before being tidally stripped. In practice, this rescaling has a fairly small effect on our velocity functions. Roughly \( \sim 30\% \) of surviving halos meet this condition for \( r_{t} \). For those halos that do experience this kind of mass loss, the typical reduction in \( V_{\text{max}} \) is \( \lesssim 25\% \).

We are currently in the process of checking this model against idealized N-body experiments designed to mimic the type of orbital histories that we encounter here (J. Bullock, K. Johnston, & A. Zentner, in preparation). Preliminary results show promising agreement.

### 2.4. Tests and Examples

Figure 3 shows three example calculations of subhalo trajectories aimed at demonstrating how various factors affect the orbital evolution of a satellite system. Each satellite was started on the same initial orbit, \( \epsilon = \eta = 0.5 \), but the satellite properties were varied: \( M_{\text{sat}}^{0} = 10^{8} M_{\odot} \), \( c_{\text{vir}} = 15 \) (solid line); \( M_{\text{sat}}^{0} = 10^{8} M_{\odot} \), \( c_{\text{vir}} = 7.5 \) (dashed line); and \( M_{\text{sat}}^{0} = 5 \times 10^{9} M_{\odot} \), \( c_{\text{vir}} = 15 \) (short-dashed line). The upper and lower panels depict the evolution of orbital radius and mass of the subhalo, respectively.

The accretion time was set at 8 Gyr in the past, \( a = (1+z)^{-1} \approx 0.45 \) for this cosmology. The host halo parameters were chosen to match reasonable expectations for a Milky Way-sized progenitor at that time: \( M_{\text{host}}^{0} = 5 \times 10^{11} M_{\odot} \) (\( R_{\text{vir}} \approx 110 \) kpc) and \( c_{\text{vir}} = 6 \). While the subhalo represented by the solid line experiences gradual tidal mass loss and slight orbital decay as a result of dynamical friction, its core survives for the full time period. The less concentrated subhalo (dashed line) is more strongly affected by tides and is completely disrupted \( \sim 3.5 \) Gyr after being incorporated into the host. (Although not shown, a similar effect is seen if the host halo concentration is increased and the subhalo concentration is held fixed.) In the case of the massive subhalo, dynamical friction causes the orbit to decay more quickly and the subhalo experiences more frequent pericenter passages. Consequently, disruption occurs \( \sim 6 \) Gyr after accretion. Notice that because the stripping process is gradual (unless orbits are very radial) and the timescales involved are of order \( \sim 1 \) Gyr, the accretion time is also important in determining survival probability. If any of these subhalos were accreted more recently, their chance of survival to the present day would increase accordingly. The combination of factors illustrated here—accretion time, satellite mass, and the relative concentrations of host and satellite—will be important in later sections for understanding the factors that set the subhalo population from one cosmology to the next.

Figure 4 shows the ensemble-averaged, cumulative velocity function for the substructure population of Milky Way–like host halos computed in our standard ΛCDM cosmology. The host properties at \( z = 0 \) are \( M_{\text{vir}} = 1.4 \times 10^{12} \)
$M_{\odot}$, $c_{\text{vir}} \approx 13.9$, and $V_{\text{max}} \approx 187 \text{ km s}^{-1}$. The lines represent the means of 200 merger tree realizations, and the error bars represent the sample variances over these realizations.

In particular, the thick solid line shows the surviving subhalo population at $z = 0$. For comparison, the thin dashed line is the best-fit velocity function reported by K99 on the basis of simulations of substructure in ΛCDM halos. This line is plotted over the range that their resolution and sample size allowed them to probe. The apparent agreement between our semianalytic model and the N-body result is excellent, and lends confidence in our ability to apply this model to different power spectra.

The radial distribution of substructure at $z = 0$ for the same ensemble of halos is shown in Figure 5. Open circles show the differential number density profile of subhalos with $M_{\text{sat}} > 10^{6} M_{\odot}$ normalized relative to the total, volume-averaged number density of subhalos within $R_{\text{vir}}$ that meet the same mass requirement. The solid pentagons show the same quantity for more massive subhalos, $M_{\text{sat}} > 6 \times 10^{8} M_{\odot}$. The line shows the NFW dark matter profile for the host system normalized relative to the average (virial) density within the halo. Observe that the subhalo profile traces the dark matter profile at large radius ($\rho \propto r^{-3}$) but flattens toward the center as a consequence of tidal disruption. This result agrees well with that presented in Figure 3 of Colin et al. (1999). Using an N-body analysis of a cluster-sized host, Colin et al. (1999) showed that the number density of systems with $M_{\text{sat}}$ greater than 0.04% of the host mass traces the background halo profile at large radius, begins to flatten at $r \sim 0.2 R_{\text{vir}}$, and is roughly a factor of 5 below the background at $r \sim 0.07 R_{\text{vir}}$ (their innermost point). The solid pentagons in Figure 5 correspond to the same mass fraction relative to the host. Notice that at $r = 0.07 R_{\text{vir}} \approx 20 \text{ kpc}$, the factor of ~5 mismatch is reproduced. Ghigna et al. (1998) observed the same qualitative behavior for subhalos in a standard CDM simulation of a cluster-sized halo. Chen et al. (2003) have measured the substructure profile using a high-resolution galaxy-size halo with $M_{\text{sat}} \lesssim 0.0015 M_{\text{host}}$, corresponding to subhalos intermediate in mass between those represented by the open circles and solid pentagons in Figure 5. Chen et al. (2003) similarly find core behavior setting in at a radius of ~30 kpc, but they also find a stronger overall suppression in substructure counts within $r \lesssim 70 \text{ kpc}$. Our results suggest that some of the observed suppression may be caused by merging in the central regions of their simulated halos. Ongoing studies by other workers lead to similar conclusions (J. Taylor 2003, private communication). Only the next generation of numerical simulations can reliably test this. That we produce a reasonable approximation to the number density profile of substructure is an indication of the soundness of our model.

We quote results relative to $R_{\text{vir}}$ and $M_{\text{host}}$ because the host halo in Colin et al. (1999) is significantly more massive than the halos that we consider.

---

Fig. 4.—Velocity functions of progenitor and surviving subhalo populations derived using our fiducial ($\mu = 1, \sigma = 0.95$) ΛCDM cosmology and a 200 halo ensemble of $1.4 \times 10^{12} M_{\odot}$ systems at $z = 0$. Shown are all accreted halos (dashed line) and the fraction of those that are tidally destroyed (short-dashed line) and centrally merged (dotted line). The solid line shows the surviving population of subhalos at $z = 0$, and for comparison, the thin dashed line shows the surviving population derived by K99 using N-body simulations. The error bars represent the sample variance.

Fig. 5.—Radial number density profile of substructure derived from 200 model realizations of a $M_{\text{sat}} = 1.4 \times 10^{12} M_{\odot}$ host halo at $z = 0$ in our fiducial ($\mu = 1, \sigma = 0.95$) ΛCDM cosmology. The open circles show the number density of subhalos with $M_{\text{sat}} > 10^{6} M_{\odot}$ divided by the average number density of systems meeting this mass threshold within the virial radius of the host system. The points reflect the radial profile averaged over all realizations, and the error bars reflect the sample variance. Solid pentagons show the same result for $M_{\text{sat}} > 6 \times 10^{8} M_{\odot}$ subhalos. The variance (not shown) is significantly larger for the higher mass threshold because there are significantly fewer such systems in each host. For reference, the solid line shows NFW density profile of the host at $z = 0$. The virial radius for a host halo of this size is $R_{\text{vir}} \approx 285 \text{ kpc}$, and the typical NFW scale radius is $r_{s} \approx 20 \text{ kpc}$. We do not plot predictions beyond $r = 0.75 R_{\text{vir}} \approx 215 \text{ kpc}$ because this is the maximum circular radius we assign to in-falling, bound systems.
3. MODEL POWER SPECTRA

The initial power spectrum of density fluctuations is conventionally written as an approximate power law in wavenumber \( k \), \( P(k) \propto k^n \), corresponding to a variance per logarithmic interval in wavenumber of \( \Delta^2(k) = k^3 P(k)/2\pi^2 \). If the fluctuations were seeded during an early inflationary stage, as is commonly supposed, then the initial spectrum is likely to be nearly scale-invariant, with \( n \approx 1 \). Any deviation from power law behavior, or “running” of the power law index with scale is likewise expected to be small, \( |dn/d\ln k| < 0.01 \) (Kosowsky & Turner 1995). In addition to these theoretical prejudices, large-scale observations of galaxy clustering and CMB anisotropy seem to favor nearly scale-invariant models that can be parameterized in this way. In this paper we explore the effects on halo substructure of taking \( n \neq 1 \) and allowing for scale-dependence in the power law index and more dramatic features in the power spectrum. In this section we give a brief description of the power spectra that we explore.

Table 1 summarizes the relevant features of our example power spectra. The second and third columns list the primordial spectral index evaluated at the pivot scale of the COBE measurements \( k_{COBE} \approx 0.0023 \ h \ Mpc^{-1} \), and the running of the spectral index.\(^7\) We neglect any variation in the running with scale and explicitly set \( d^2n(k)/d\ln k^2 = 0 \). Except for the RI case, we normalize all models to the COBE measurements of the CMB anisotropy using the fitting formulae of Bunn, Liddle, & White (1996; also Bunn & White 1997). The fourth column of Table 1 gives the implied value of \( \sigma_8 \). We calculate spectra using the transfer functions of Eisenstein & Hu (1999). In Figure 6 we illustrate the implied \( \sigma(M) \) for these models.

Many of the spectra listed in Table 1 are motivated by particular models of inflation. We invoke an inverse power law potential that gives rise to a mild tilt \( n \approx 0.94 \), as well as a model in which the inflation has a logarithmically running mass and which can give rise to significant tilt and running for natural parameter choices (Stewart 1997a, 1997b; Covi & Lyth 1999; Covi, Lyth, & Roszkowski 1999; Covi, Lyth, & Melchiorri 2003). We employ specific inflationary potentials mainly as a conceptual follow-up to ZB02, which highlighted the fact that various levels of tilt may occur naturally within the paradigm of inflation and that \( n \approx 1 \) is not demanded by this paradigm. For the purposes of this paper, one may regard our choices simply as spanning a range of observationally viable input power spectra. The values of tilt and \( \sigma_8 \) that we consider range from \( n \approx 0.84 \) with \( \sigma_8 = 0.65 \) to \( n = 1 \) and \( \sigma_8 = 0.95 \). The model with \( \sigma_8 = 0.75 \) was specifically chosen to match galaxy central densities, as described in ZB02. We also explore the best-fit, RI model of the WMAP team (Spergel et al. 2003), with \( d\ln k = -0.03 \). We refer to this as the “RI model.” Note that Spergel et al. (2003) quote a value of \( n = 0.93 \) evaluated at \( k = 0.05 \ Mpc^{-1} \). The value listed in Table 1 is larger because we quote it at a smaller wavenumber, \( k = k_{COBE} \).

In addition to tilted ΛCDM models, we consider spectra with abrupt reductions in power on small scales. In the “broken scale-invariance” (BSI) example, we adopt an idealized inflation model introduced by Starobinsky (1992) that exhibits the most rapid drop in power possible for a single field model. Kamionkowski & Liddle (2000) studied this type of model as a way to mitigate the dwarf satellite problem, but our choice of parameters is slightly different from theirs (see ZB02).

We also consider WDM scenarios in which the primordial power spectra are scale-invariant but small-scale fluctuations are subsequently filtered by free-streaming. The free-streaming scale is set by the primordial velocity dispersion of the warm particles. In the canonical case of a “neutrino-like,” thermal relic with two internal degrees of freedom, the free-streaming scale can be expressed in terms of the warm particle mass \( m_w \) and relic abundance, \( \Omega_w h^2 \):

\[
R_f \approx 0.11 \left( \frac{\Omega_w h^2}{0.15} \right)^{1/3} \left( \frac{m_w}{\text{keV}} \right)^{-4/3} \text{Mpc} . \quad (13)
\]

We calculate WDM spectra assuming the same flat

---

\(^7\) We use the definition of running employed by Spergel et al. (2003) rather than that given in, for instance, Kosowsky & Turner (1995). These definitions differ by a factor of 2.

---

### TABLE 1

| Model Description          | \( n(k_{COBE}) \) | \( dn(k)/d\ln k \) | \( \sigma_8 \) | Comments                              |
|----------------------------|------------------|--------------------|---------------|---------------------------------------|
| Scale-invariant.............| 1.00             | 0.000              | 0.95          |                                       |
| Inverted power law inflation| 0.94             | -0.002             | 0.83          |                                       |
| Running-mass inflation I....| 0.84             | -0.008             | 0.65          | See Stewart 1997a, 1997b              |
| Running-mass inflation II...| 0.90             | -0.002             | 0.75          |                                       |
| WMAP best-fit RI model......| 1.03             | -0.03              | 0.84          | WMAP best fit; see Spergel et al. 2003 |
| BSI inflation...............| 1.00             | 0.000              | 0.97          | Exhibits sharp decline in power at \( k \geq 1 \ h \ Mpc^{-1} \), power suppressed for \( M \leq 10^{10} M_\odot \) |
| WDM, \( m_w = 3.0 \text{ keV} \) | 1.00             | 0.000              | 0.95          | \( M_f \approx 8.3 \times 10^9 M_\odot \) |
| WDM, \( m_w = 1.5 \text{ keV} \) | 1.00             | 0.000              | 0.95          | \( M_f \approx 1.3 \times 10^{10} M_\odot \) |
| WDM, \( m_w = 0.75 \text{ keV} \) | 1.00             | 0.000              | 0.94          | \( M_f \approx 2.1 \times 10^{11} M_\odot \) |

Notes.—Col. (1) gives a brief description of the inflation or WDM model used to predict the power spectrum. In the text we distinguish the first five models by their tilts and/or their values of \( \sigma_8 \). We label the WDM models by the warm particle mass. Col. (2) and (3) give the tilt \( n(k_{COBE}) \) on the pivot scale of the COBE data \( k_{COBE} \approx 0.0023 \ h \ Mpc^{-1} \) and the running of the spectral index \( dn(k)/d\ln k \), respectively. We have explicitly assumed the “running-of-running” to be small and have taken \( d^2n(k)/d\ln k^2 = 0 \). Col. (4) contains the values of \( \sigma_8 \) implied by the tilt or warm particle mass, the COBE normalization, and our fiducial cosmological parameters, except in the case of the WMAP best-fit RI model, in which case the value of \( \sigma_8 \) reflects their best-fit normalization.
cosmology with $\Omega_M = \Omega_W + \Omega_B = 0.3$, and use the approximate WDM transfer function given by Bardeen et al. (1986), $P(k) = \exp[-kR_0 - (kR_0)^2]P_{\text{CDM}}(k)$. Several studies have placed approximate constraints on WDM masses based on either the argument that there must be enough power on small scales to reionize the universe at sufficiently high redshift ($z_{\text{re}} \gtrsim 6$) or by probing the power spectrum on small scales directly with the Ly$\alpha$ forest (Barkana, Haiman, & Ostriker 2001; Narayan et al. 2000). These authors essentially find that $m_W \approx 0.75$ keV assuming a neutrino-like thermal relic; however, this constraint may be significantly more restrictive if measurements of $z_{\text{re}} \approx 17$ by the WMAP collaboration (Kogut et al. 2003; Spergel et al. 2003) are confirmed (Somerville, Bullock, & Livio 2003). As such, we consider three illustrative examples in what follows, $m_W = 0.75$ keV, 1.5 keV, and 3.0 keV. The corresponding “free-streaming” masses, below which the fluctuation amplitudes are suppressed, are listed in Table 1.

4. RESULTS

4.1. Accretion Histories

Our first results concern the merger histories of halos that are approximately Milky Way–sized, with $M_{\text{vir}} = 1.4 \times 10^{12}$ $M_\odot$ at $z = 0$. For the $n = 1$, $\Lambda$CDM model, we present results based on 200 realizations. For all other models, our findings are based on 50 model realizations.

Figure 7 focuses on the mass distribution of accreted halos, integrated over the entire merger history of the host. We plot $df / d \log(M_{\text{sat}})$, the fraction of mass in the final halo that was accreted in subhalos of a given mass per logarithmic interval in subhalo mass, $M_{\text{sat}}$. Observe that the mass fraction accreted in subhalos of a given mass is relatively insensitive to the shape of the power spectrum. Although similarity from model to model may be somewhat surprising at first, it follows directly from repeated application of equation (1). In particular, the shape of the progenitor
distribution for $M_{\text{sat}} \ll M_{\text{host}}$ must follow $df/d\log(M_{\text{sat}}) \propto M_{\text{sat}}^{-1/2}$, and the turnover occurs because mass conservation suppresses the number of major mergers. The shape shown in Figure 7 and its insensitivity to the power spectrum is discussed in detail by LC93.

While the total mass function of accreted substructure is relatively independent of the power spectrum, the merger histories themselves are not. In models with less power on galaxy scales, halos assemble their mass later and experience more recent mergers and disruption events.

We show an example of this in Figure 8. Here we plot the average accretion rate of subhalos with $M_{\text{sat}} > 10^8 M_\odot$ for host halos in the standard $n = 1$, ACDM model, the RI model, and our lowest normalization case ($n = 0.84$, $\sigma_8 = 0.65$). The total accretion rate is divided in two pieces: dashed lines show those subhalos that are eventually destroyed, and solid lines show the accretion times of subhalos that survive until $z = 0$. For the standard ($n = 1$, $\sigma_8 = 0.95$) case, the event rate peaks sharply about $\sim 12$ Gyr in the past, while the $\sigma_8 = 0.65$ case has a broader distribution, peaking later at $\sim 9$ Gyr ago and with a long tail of accretion events extending toward the present day.

The shift in accretion times in models with less small-scale power plays an important role in regulating the number of surviving subhalos. As we discussed in relation to Figure 3, a finite amount of time is required for an orbit to decay or for a system to become unbound and in many cases the longer a subhalo orbits in the background potential, the more probable its disruption becomes. The later accretion times in models with less power partially compensate for the fact that subhalos in these models are less centrally concentrated and more susceptible to disruption at each pericenter passage. Particular results for substructure populations in each model are given in the following subsections.

That we expect a characteristic merger/disruption phase in each halo’s past is intriguing, as this phase is approximately coincident with the estimated ages of galactic thick disks, $t_{\text{td}} \sim 8-10$ Gyr (e.g., Quillen & Garnett 2000 for the Milky Way), which seem to be ubiquitous and roughly coeval (Dalcanton & Bernstein 2002). In this context, the age distributions of thick disks might serve as a test of this characteristic accretion time, which varies as a function of normalization and cosmology. We reiterate that the lookback times shown for the dashed lines in Figure 8 are the times that the subhalos were accreted. The distributions of central merger rates and tidal destruction rates peak at slightly more recent times and their widths are broader, with longer tails toward the present epoch.

It is interesting to note that the surviving halos in Figure 8 represent a distinctly different population of objects than the destroyed systems—they tend to have been accreted more recently. We are inclined to speculate that the star formation histories of galaxies that were destroyed after being accreted could be distinctly different from those of the surviving (dwarf satellite) galaxies as well. This may have implications for understanding whether the stellar halo of our Galaxy formed from disrupted dwarfs or some other process. While the global structure of the stellar halo seems consistent with the disruption theory (Bullock, Kravtsov, & Weinberg 2001), the element ratios of stellar halo stars and stars in dwarf galaxies are not consistent with a common history of chemical evolution (Shetrone, Côte, & Sargent 2001). The results shown in Figure 8 provide general motivation to model dwarf galaxy evolution and Milky Way formation in a cosmological context.

### 4.2. Mass and Velocity Functions

We present our results on CDM halo substructure beginning with the abundance of satellites in Milky Way–like galaxies. We plot the mass function of subhalos $N(M_{\text{sat}})$, or the number of subhalos with mass greater than $M_{\text{sat}}$ as a function of $M_{\text{sat}}$, for each of our models in Figure 9. The host halo mass is again fixed at $1.4 \times 10^{12} M_\odot$ at $z = 0$. From this figure, we see that even in the significantly tilted, low-normalization model ($\sigma_8 \sim 0.65$), the number of satellite halos with mass greater than $10^8 M_\odot$ is roughly equal to that in the standard $n = 1$ model. The systematic differences between models are small compared with the scatter. The suppression is weak because several competing effects tend to compensate for the reduced concentrations of the subhalos. In tilted models with reduced small-scale power, subhalos are typically accreted at later times. In addition, host halos are less concentrated and correspondingly less capable of disrupting their satellites.

In contrast, the BSI model shows a substantial decrease (a factor of $\sim 3$) in the number of surviving satellite halos at fixed mass. The reason for the dramatic reduction in this case is easy to understand. First, power is reduced only on scales smaller than a critical scale around $\sim 10^{10} M_\odot$ (see Fig. 6), so the concentration and accretion history of the $\sim 10^{12} M_\odot$ host halo are minimally altered while the concentrations of the small subhalos are drastically reduced (see ZB02). In other words, the host halo has a density structure similar the $n = 1$ model host and is just as capable of tidally disrupting satellites, but the satellites are significantly more
susceptible to disruption. A second difference is that galaxy-size halos in the BSI model, unlike the tilted models, accrete 40% fewer low-mass \((\leq 10^7 M_\odot)\) halos over their lifetimes, and this further widens the disparity between the BSI and tilted-ΛCDM models.

It is conventional to discuss the substructure population of Milky Way–like halos in terms of the velocity function. In Figure 10 we show our results for the cumulative velocity functions of subhalos for a fixed host mass of \(M_{\text{host}} = 1.4 \times 10^{12} M_\odot\). Notice that the velocity functions show a stronger trend with power spectrum than the mass functions (Fig. 9), but the effect is still modest compared with the statistical scatter. For the most extreme tilted model, the total number of subhalos with \(V_{\text{max}} \gtrsim 10 \text{ km s}^{-1}\) is only a factor of \(\sim 2\) lower than in the standard, scale-invariant case. In the case of the tilted models, the reduction in the velocity function is largely due to the fact that the subhalos are less concentrated, so the \(V_{\text{max}}\) values are correspondingly smaller for fixed halo masses (see eq. [3] and the discussion that follows).

This effect is illustrated explicitly in Figure 11, where rather than fixing the host mass at \(z = 0\), we have fixed its maximum circular velocity at \(V_{\text{max}} = 187 \text{ km s}^{-1}\), the value of a typical \(n = 1\), \(M_{\text{host}} = 1.4 \times 10^{12} M_\odot\) halo at \(z = 0\). Normalizing our host halos by \(V_{\text{max}}\) rather than mass is perhaps a more reasonable choice because \(V_{\text{max}}\) is more closely related to observations.\(^8\) Models with less galactic-scale power require a more massive host in order to obtain the same value of \(V_{\text{max}}\) and their velocity functions shift correspondingly. For example, a host with \(V_{\text{max}} = 187 \text{ km s}^{-1}\) in the \(\sigma_8 = 0.65\) model requires \(M_{\text{host}} \approx 2.2 \times 10^{12} M_\odot\). With

\(^8\) \(V_{\text{max}} = 187 \text{ km s}^{-1}\) is somewhat smaller than a typical rotation velocity for a galaxy like the Milky Way (\(V_{\text{MWmax}} \sim 220 \text{ km s}^{-1}\)), but this value is in line with expectations for the dark matter halo once the effects of baryonic in-fall have been included (e.g., Klypin, Zhao, & Somerville 2002).
curves correspond to BSI halos of \( M(\text{curve}) = 10^{12} M_\odot \) and \( M(\text{dotted curve}) = 10^{11} M_\odot \) at \( z = 0 \). The lower set of thin curves correspond to BSI halos of \( M_{\text{host}} = 1.4 \times 10^{12} M_\odot \) at \( z = 0 \) (solid curve) and \( M_{\text{host}} = 3 \times 10^{12} M_\odot \) (long-dashed curve) all at \( z = 0.6 \). The crosses reflect an analytic fit to the \( n = 1 \) results, as discussed in the text (see eq. [14]).

**Fig. 12.**—Average differential mass fraction, \( df/dM_{\text{sat}} \), normalized relative to host mass and satellite mass. The upper set of (bold) curves were computed for the \( n = 1 \) cosmology with \( M_{\text{host}} = 1.4 \times 10^{12} M_\odot \) at \( z = 0 \) (solid curve) and \( M_{\text{host}} = 10^{11} M_\odot \) (dotted curve), and \( 3 \times 10^{12} M_\odot \) (long-dashed curve), and \( 10^{13} M_\odot \) (dot-dashed curve) all at \( z = 0.6 \). The bottom set of thin curves have very similar scatter. The bottom set of errors reflect the same range determined from 50 realizations for the BSI model (other CDM-type models show similar scatter) and the bottom set of errors reflect the same range determined from 50 realizations of the BSI spectrum.

Rather than the total mass fraction, lensing measurements are sensitive to the mass fraction in substructure projected onto the plane of the lens at a halocentric distance of

\[
df = \left( \frac{x}{x_0} \right)^{-\alpha} \exp \left( -\frac{x}{x_0} \right) ,
\]

with \( x \equiv M_{\text{sat}}/M_{\text{host}} \), \( \alpha = 0.6 \) and \( x_0 = 0.07 \pm 0.05 \). The quoted range in \( x_0 \) characterizes well the rms scatter from realization to realization (not shown). This function (with \( x_0 = 0.07 \)) is shown as the set of crosses in Figure 12. As expected, the mass fractions are somewhat lower for the BSI model halos. The other CDM-type models all yield differential mass functions similar to those of the \( n = 1 \) case. While in the next section we present results for a particular choice of host mass, the self-similarity demonstrated here implies that results at a fixed satellite-to-host mass ratio, \( x \), can be scaled in order to apply these results to any value of \( M_{\text{host}} \).

**Fig. 13.**—Fraction of the parent halo mass that is bound up in substructure in the mass range between \( 10^9 M_\odot \) and \( M_{\text{halo}} \) as a function of \( M_{\text{halo}} \). The host halo in each case has \( M = 3 \times 10^{12} M_\odot \) at \( z = 0.6 \). Lines reflect the mean over all realizations, and results are shown for the \( n = 1 \) model (solid line), RI model (dot-short-dashed line), \( \sigma_0 = 0.75 \) (dashed line), \( \sigma_0 = 0.65 \) (dot-long dashed line), and BSI (dotted line). The error bars on the top set of lines reflect the 90 percentile range determined using 200 merger tree realizations for the \( n = 1 \) case (the other models in the top set of lines have very similar scatter). The bottom set of errors reflect the same range determined using 50 realizations of the BSI model.
order the Einstein radius of the lens, \( R_E \sim 5-15 \) kpc. In Figure 14 we plot \( f_{\text{sat}}>10^6 M_\odot \) projected through a cylinder of radius 10 kpc centered on the host halo for the same set of halos shown in Figure 13. The large and small error bars reflect the 90 and 64 percentile ranges, respectively, in measured projected mass fractions derived using 200 \( n = 1 \) realizations (top set) and 50 BSI realizations (bottom set). A downward arrow is plotted instead of a lower, large error tick if at least 5% of the realizations had \( f = 0 \) in that bin. A downward arrow with no accompanying lower error bar means that at least 18% of the realizations were without projected substructure in that bin.

The projected mass fractions are not as severely suppressed relative to the volume-averaged mass fractions as one might expect given that tidal forces act systematically to destroy substructure near host halo centers (see Fig. 5). The reason is that we are examining substructure in a cylindrical volume and picking up subhalos with large halocentric radii. We illustrate this effect in Figure 15, where we compare the mass fraction in cylindrical projection radius \( \rho \) with the mass fraction in spherical shells with the same value of spherical radius \( r \). Notice that the mass fraction in spherical regions is significantly reduced in the center, while the projected mass fraction is less severely affected. Of course, the mass fraction approaches the global value at large radii. Figures 16 and 17 demonstrate how the mass fractions change as a function of projection radius for various subhalo mass cuts for the \( n = 1 \) and BSI models, respectively. Notice that the relative drop in mass fraction as a function of projection radius is more pronounced in the BSI model.
than in the $n = 1$ case. This reflects the fact that tidal disruption is more important in the BSI case and corelike behavior of the subhalo radial distribution sets in at a larger radius in this model.

### 4.4. Warm Dark Matter and Gravitational Lensing

In the previous section we demonstrated that the substructure mass fraction is sensitive to abrupt changes in the power spectrum and used the BSI model as a specific example. In this section we investigate these differences in the context of warm dark matter (WDM). We label the different WDM models by the warm particle mass $m_W$ and assume the canonical case of a "neutrino-like" thermal relic with two internal degrees of freedom, $g_W = 2$.

Figure 18 shows the total mass fraction of $3 \times 10^{12} M_\odot$ host halos at $z = 0.6$ as a function of $M_{\text{sat}}$ implied by our three WDM model power spectra compared with our standard LCDM case. For substructure smaller than $\sim 10^7 M_\odot$, the differences between the models are as large as an order of magnitude or more, and even the largest WDM particle mass (3 keV) provides a potentially measurable suppression of substructure. Figure 19 shows the mass fraction in projected cylinders of radius 10 kpc.

The differences in mass fractions seen for the different models in Figures 18 and 19 come about because subhalos become less concentrated relative to their host halos as the WDM particle mass is decreased and power is suppressed on larger scales, much like the BSI case. In true WDM models there are additional processes that, in principle, can alter the formation and density structure of dark matter halos. In Figures 18 and 19 we have accounted only for the effect of the power spectrum on substructure mass fractions and assumed that the density structure of WDM halos is identical to that for CDM halos. For high-mass systems, this is a sensible approximation (Colín et al. 2000a; Avila-Reese et al. 2001), but this approximation should break down at small masses and lead to further suppression of substructure.
One consequence of a WDM particle with nonnegligible velocity dispersion is that gravitational clustering is resisted by structures below the effective Jeans mass of the warm particles (e.g., Hogan & Dalcanton 2000; Bode et al. 2001):

$$M_J \approx 6 \times 10^3 \left(\frac{\text{keV}}{m_w}\right)^4 \left(\frac{\Omega_m h^2}{0.15}\right)^{1/2} \left(\frac{2}{g_w}\right) (1+z)^{3/2} M_\odot.$$

(15)

For both the $m_w = 1.5$ and $m_w = 3.0$ keV models, $M_J \ll 10^7 M_\odot$ when $z \lesssim 10$, so all halos of interest in this context are minimally affected. The situation is somewhat more complicated in the $m_w = 0.75$ keV model, where $M_J \lesssim 10^7 M_\odot$ for redshifts $z \gtrsim 2$. We therefore expect that the formation of these halos should be suppressed compared with the predictions of the EPS formalism. This suppression should only have a minor effect on our predictions because we restrict ourselves to satellite masses $\gtrsim 10^5 M_\odot$ and most surviving subhalos are accreted at $z \lesssim 2$. In the interest of simplicity, we chose to ignore this effect here. As a result, we may significantly overpredict substructure mass fractions at low $M_{\text{inj}}$ in these cases. In the context of this study, this is a conservative approach because the true mass fraction would be reduced by these effects, bringing it further away from the measured substructure mass fractions and standard ΛCDM predictions.

In addition to the effective Jeans suppression, WDM halos, unlike their CDM counterparts, cannot achieve extremely high densities in their centers because of phase space constraints (Tremaine & Gunn 1979). In the early universe the primordial phase space distribution of the WDM particles is a Fermi-Dirac distribution with a maximum of $f_{\text{max}} = g_w/h_{\text{pl}}^3$ at low energies ($h_{\text{pl}}$ is Planck’s constant). For a collisionless species, the phase space density is conserved and this maximum phase space density may not be exceeded within WDM halos. If we define the phase density as $Q \equiv \rho / (2\pi\sigma^2)^{3/2}$, then the maximum allowed phase density is (Hogan & Dalcanton 2000)

$$Q_{\text{max}} \approx 5.2 \times 10^{-4} \left(\frac{m_w \text{ keV}}{\text{keV}}\right)^4 \left(\frac{g_w}{2}\right) \frac{M_\odot}{\text{pc}^{-3}} \left(\frac{\text{km s}^{-1}}{\text{km s}^{-1}}\right)^3.$$  

(16)

This limit implies that WDM halos cannot achieve the central density cusps of the kind observed in simulated CDM halos. Instead, we expect a core in the density profile. For viable WDM models, the phase space core is expected to be dynamically unimportant for any halo massive enough to host a visible galaxy (ABW02). However, for the lowest mass subhalos ($M \lesssim 10^7 M_\odot$) the presence of phase space-limited cores may be important because halos with large cores are less resistant to tidal forces than cuspy halos.

We have attempted to estimate (crudely) how the phase space limit affects the substructure population of WDM halos by adopting our standard model of halo accretion and orbital evolution, but allowing the density structure of the appropriately small subhalos to be set by the phase space limit. For these calculations we used the phenomenological density profile of Burkert (1995),

$$\rho_B(r) = \frac{\rho_0}{(1 + r/r_B)[1 + (r/r_B)^2]},$$

(17)

The Burkert profile resembles the NFW form at large radius, but it features a constant density core at its center and thus a velocity dispersion that approaches a constant at small $r$: $\sigma^2_B \approx 0.55 V^2_{\text{max}}$.

For Burkert profiles, $V^2_{\text{max}} \approx 0.86 V^3_{\text{vir}} c_B / g_B(c_B)$, where $c_B \equiv r_{\text{vir}} / r_B$ is the Burkert concentration and

$$g_B(y) \equiv \ln(1+y^2) + 2 \ln(1+y) - 2 \tan^{-1}(y).$$

Solving for the phase density in the core ($r \ll r_B$) and equating it with the maximum phase density of equation (16) yields the following relation for the maximum attainable value of $c_B$:

$$c_B^{3/2} g_B^{1/2}(c_B) \approx \frac{111}{(1+z)^3} \left(\frac{0.15}{\Omega_m h^2}\right) \left(\frac{178}{\Delta_{\text{vir}}}\right) \left(\frac{V_{\text{vir}}}{\text{km s}^{-1}}\right)^3 \times \left(\frac{g_w}{2}\right) \left(\frac{m_w}{\text{keV}}\right).$$

(18)

We assigned Burkert concentrations $c_{\text{vir}}$ for each halo according to the B01 model. We converted from NFW concentration to Burkert concentration $c_B$ by interpreting the B01 value of $r_B$ as the radius at which $d\ln \rho(r)/d\ln r_{\text{vir}} = -2$. This implies that $r_B \approx 0.65 r_{\text{vir}}$, or $c_B \approx 1.5 c_{\text{vir}}$. With this correspondence, the adopted Burkert profile achieves the maximum of its rotation curve at $r_{\text{max}}^B \approx 0.99 r_{\text{max}}$, where $r_{\text{max}}$ is the radius at which the corresponding NFW halo achieves $V_{\text{max}}$. Similarly, $V_{\text{max}}$ of the adopted Burkert profile is within 10% of the corresponding NFW $V_{\text{max}}$ for all relevant concentrations ($1 \leq c_{\text{vir}} \leq 25$). Second, we computed the maximum value of $c_B$ allowed by the phase space constraints using equation (18). We then assigned each halo the smaller of these two values of $c_B$ at the time of accretion. In this way we guaranteed that the phase space constraint was met by all halos. We have checked that this prescription for Burkert halos does not yield any systematic bias in our results by applying it all of our CDM models. We found that it gave nearly identical results to that of our standard NFW model, which is not surprising in the context of our model and disruption criteria.

We present our estimates of cumulative mass fractions in WDM models, including the effect of the phase space constraint, in Figure 20. It is clear, at least from this rough estimate, that the Tremaine-Gunn limit plays an important role only for the most extreme WDM models, $m_w \lesssim 1$ keV, and only the smallest halos, $\lesssim 10^6 M_\odot$. However, we emphasize that our new assumptions about WDM halos have not been tested with $N$-body simulations. Simulations have yet to examine the density structure of halos that saturate the phase space bound, and most studies have ignored the initial velocity dispersion of the WDM particles (Colin et al. 2000a; Avila-Reese et al. 2001; Knebe et al. 2002), but the Burkert profile assumption seems plausible. With these precautions in mind, Figure 18 may be regarded as an approximate upper limit on the substructure mass fraction for WDM halos. Any phase space bound or the effects of primordial velocity dispersions on halo formation and density...
structure should lead to enhanced disruption, resulting in lower mass fractions.

One physical process that might affect WDM (and BSI) models that we have not considered is top-down fragmentation (e.g., Knebe et al. 2002). It is possible that power can be transported from large scales to small in truncated models, resulting in a population of low-mass halos that would not be accounted for in Press-Schechter theory. While such a process could result in a higher substructure abundance than that estimated using our model, there are reasons to believe that the effect should be fairly small. Systems that form in this manner collapse quite late, and their density structure likely would be very diffuse compared with their hierarchically formed brethren. Therefore, it is less likely that systems formed via fragmentation could survive tidal disruption once incorporated into a galactic halo.

4.5. The Dwarf Satellite Problem

Comparisons between the predicted subhalo population and the observed dwarf galaxy abundance are usually made by comparing counts as a function of maximum circular velocity, \( V_{\text{max}} \). This is a sensible mode of comparison because it sidesteps the complicated issues of star formation and feedback in these poorly understood galaxies. Yet there are considerable uncertainties, even for this method of comparison, and it is likely that efforts to compare predictions as a function of dwarf luminosity (Somerville 2002; Benson et al. 2002) in tandem with velocity comparisons will be needed in order to fully understand the nature of the dwarf satellite problem.

For most satellites, the quantity that is observed and used to infer the halo \( V_{\text{max}} \) is the line-of-sight stellar velocity dispersion, \( \sigma_* \). As discussed in S02 and H03, the mapping between \( \sigma_* \) and \( V_{\text{max}} \) depends on the theoretical expectation for the density profile of the subhalo, as well as on the stellar mass distribution of the galaxy. An additional complication concerns the unknown velocity anisotropy of the stars in the system.

A phenomenologically motivated approximation for the stellar distribution in a dwarf galaxy is the spherically symmetric King profile (King 1962),

\[
\rho_k(r) = \frac{k}{r} \left[ \frac{\cos^{-1}(z)}{z} - \sqrt{1 - z^2} \right],
\]

where \( z \equiv [1 + (r/r_c)^2]/[1 + (r/r_t)^2] \), \( r_c \) and \( r_t \) are the core and tidal radii of the King profile, and \( \rho_k(r > r_t) = 0 \). The normalization is not important in what follows.

If we assume that a stellar system described by equation (19) is in equilibrium and embedded in a spherically symmetric dark matter potential characterized by the circular velocity profile \( V_c(r) \), then the radial stellar velocity dispersion profile \( \sigma_r(r) \), can be computed via the Jeans equation

\[
r \frac{d}{dr} \left( \rho \sigma_r^2 \right) = -\rho V_c^2 \iffiffiffiffiffiffiffiffiffiffiffiffi \frac{d}{dr} \left( \rho \sigma_r^2 \right) ,
\]

assuming a constant mass-to-light ratio. If a galaxy has a measured stellar profile (the King parameters in this case) and measured value of \( \sigma_* \), then the Jeans equation places only one constraint on the rotation curve of the system, \( V_c(r) \). We expect the halo velocity profile to be at least a two-parameter function (e.g., the NFW profile), so determining \( V_{\text{max}} \) requires some theoretical input for the expected form of \( V_c(r) \) in order to provide a second constraint.

Motivated by dark matter models, we assume that the global rotation curve is set by an NFW profile associated with the dwarf galaxy halo. The rotation curve for an NFW halo is fully described by specifying two parameters, and a natural pair is \( V_{\text{max}} \) and \( r_{\text{max}} \). For any given cosmology, the relation between \( V_{\text{max}} \) and \( r_{\text{max}} \) is expected to be rather tight, and this provides a second, theoretically motivated constraint that sets the \( V_{\text{max}} - \sigma_* \) mapping implied by equation (20).

The \( V_{\text{max}} - r_{\text{max}} \) relationships for surviving subhalos in two of our models are shown in Figure 21. The lower set of points corresponds to our standard \( n = 1 \) model, and the higher set of points is derived from our \( \sigma_0 = 0.65 \) model. In each case, we plot one point for each surviving halo in 10 model realizations. The strong correlation, \( r_{\text{max}} \propto V_{\text{max}}, \gamma \approx 1.3, \) follows directly from the input correlations between \( M_{\text{halo}}(z) \) and \( c_{\text{vir}} \) (see § 2.3 and B01). The normalizations and slopes are influenced by the cosmology,
accretion times, and (mildly) orbital history of the subhalos.\textsuperscript{10}

The thick solid and dashed lines in Figure 21 show the locus of points in the $V_{\text{max}}$-$r_{\text{max}}$ plane that correspond to the central values of the observed line-of-sight velocity dispersions for Carina and Draco respectively, given their measured King profile parameters. Our adopted $\sigma_s$ values and King parameters are listed in Table 2 along with appropriate references. The light solid lines illustrate how these contours expand when we include the $\pm 1\sigma$ measurement error in $\sigma_s$ for Carina. A similar (although narrower) band exists for Draco, but we have omitted it for the sake of clarity. Consistency with the observed King parameters and velocity dispersions requires each dwarf to reside in a halo with structural parameters that lie within the region of overlap between the contours and the model points. For example, in the $n = 1$ model Carina is expected to reside in a halo with $V_{\text{max}} \approx 11$ km s$^{-1}$ and $r_{\text{max}} \approx 1$ kpc. For the $\sigma_s = 0.65$ model, Carina is expected to sit in a larger halo, with $V_{\text{max}} \approx 29$ km s$^{-1}$ and $r_{\text{max}} \approx 10$ kpc. Similar comparisons hold for Draco and all of the Local Group dwarf satellites and these comparisons can be made in a similar way for any cosmology. The point is that the maximum velocities that are assigned to satellite galaxies are cosmology dependent. Therefore, "observed" velocity functions are also cosmology dependent because theoretical inputs are used to convert from $\sigma_*$ to $V_{\text{max}}$.

In Table 2 we show estimates for halo $V_{\text{max}}$ values for the observed Milky Way satellites under the assumption that $\beta = 0$ along with a similar analysis for $\beta = 0.15$ (values in parentheses). We estimate halo $V_{\text{max}}$ values for six different power spectra, relying on the model-dependent $r_{\text{max}}$-$V_{\text{max}}$ relationship for substructure in each case and taking the central values of the measured velocity dispersions for each halo. Taking the quoted $\pm 1\sigma$ range for the measured velocity dispersions typically leads to a shift in $V_{\text{max}}$ of $\approx 30\%$, which is considerable compared with the inherent scatter in the $M_{\text{vir}}$-$V_{\text{max}}$ relation. For reference, we have also included the adopted $V_{\text{max}}$ values from the original K99 work on the dwarf satellite problem. As expected, the implied $V_{\text{max}}$ values become larger as we explore models with less galactic-scale power. Our estimates for the $n = 1$, $\beta = 0$ case are close to those of K99.

The left panel of Figure 22 shows the Milky Way satellite counts for each model, assuming $\beta = 0$, along with the predicted velocity functions for each model. In addition to the satellites listed in upper portion of Table 2, we have also included the Small Magellanic Cloud (SMC; with $V_{\text{max}} = 60$ km s$^{-1}$, estimated by Stanimirović 2000\textsuperscript{11} to include a substantial contribution from the baryonic component) and the Large Magellanic Cloud (LMC; with $V_{\text{max}} = 50$ km s$^{-1}$; van der Marel et al. 2002) in our cumulative velocity functions. In the standard case ($n = 1$) the discrepancy sets in at $V_{\text{max}} \approx 30$ km s$^{-1}$, requiring roughly one in 10 halos with $V_{\text{max}} \approx 10-20$ km s$^{-1}$ to be the host of an observed galaxy. The extremely tilted model with $\sigma_s \approx 0.65$ actually underpredicts the dwarf count for large systems. Interestingly, dwarfs in the $dn/d\ln k = -0.03$ RI model as well as the $\sigma_s = 0.75$ case are consistent with inhabiting the 10 most massive subhalos, with only Sextans standing as an outlier. The BSI model also looks to be in good agreement with the data for $V_{\text{max}} \approx 12$ km s$^{-1}$, lending support to the conclusions of Kamionkowski & Liddle (2000). However, the problem of Local Group satellites is not completely "solved" in any of these models because the velocity function continues to rise below the velocity scale of Sextans in all cases. What changes in the low-power models is the nature of the discrepancy. In one extreme, the mismatch sets in at $V_{\text{max}} \approx 30$ km s$^{-1}$ and gradually becomes worse for smaller systems. In the other extreme, the mismatch seems to imply a sharp threshold for dwarf galaxy formation at a scale near $V_{\text{max}} \approx 10-20$ km s$^{-1}$.

Unfortunately, a detailed accounting of the mismatch is difficult, even for a given cosmology. The dwarf $V_{\text{max}}$ estimate is very sensitive to the velocity anisotropy parameter, $\beta$. For example, assuming $\beta = 0.15$ leads to $V_{\text{max}}$ values that are significantly lower than in the isotropic case because rotational support has been traded for pressure support. The velocity function comparisons with $\beta = 0.15$ are shown in the right panel of Figure 22 and the $V_{\text{max}}$ values are listed in parentheses in Table 2. In this case, only the $\sigma_s \approx 0.65$ case and the BSI model can account for the dwarf

\textsuperscript{10} The scatter in the $V_{\text{max}}$-$r_{\text{max}}$ plane should be larger than that shown here because we have not included the expected scatter in the input $c_p-M_{\text{gal}}$ relation. For $\sigma(\log c_p) \approx 0.14$ (B01; W02), the implied scatter is $\sigma(\log r_{\text{max}}) \approx 0.18$ at fixed $V_{\text{max}}$. For $\sigma(\log c_p) \approx 0.08$ (Jing 2000), the implication is $\sigma(\log r_{\text{max}}) \approx 0.11$.

\textsuperscript{11} See http://astron.berkeley.edu/~sstanimi.
### Table 2: Characteristics of Milky Way Satellites

| Satellite        | \(\sigma_s\) (km s\(^{-1}\)) | \(r_c\) (kpc) | \(r_t\) (kpc) | \(V_{\text{max}}\) (K99) (km s\(^{-1}\)) | \(V_{\text{max}}\) \(\sigma_s = 0.83\) (km s\(^{-1}\)) | \(V_{\text{max}}\) (RI Model) (km s\(^{-1}\)) | \(V_{\text{max}}\) (BSI) (km s\(^{-1}\)) | \(V_{\text{max}}\) \(\sigma_s = 0.65\) (km s\(^{-1}\)) |
|------------------|-------------------------------|---------------|---------------|------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|
| Sagittarius.....| 11.7 ± 0.7                    | 0.44          | 3.0           | 20                                       | 17 (13)                                    | 15 (13)                                    | 21 (17)                                    | 26 (19)                                    | 35 (25)                                    |
| Fornax...........| 10.5 ± 1.5                    | 0.46          | 2.3           | 18                                       | 15 (11)                                    | 14 (13)                                    | 19 (17)                                    | 21 (19)                                    | 30 (21)                                    |
| Draco.............| 9.5 ± 1.6                     | 0.18          | 0.93          | 17                                       | 21 (15)                                    | 31 (20)                                    | 38 (25)                                    | 41 (26)                                    | 64 (44)                                    |
| Ursa Minor...... | 9.3 ± 1.8                     | 0.20          | 0.64          | 16                                       | 21 (15)                                    | 31 (20)                                    | 37 (25)                                    | 41 (26)                                    | 69 (43)                                    |
| Leo I.............| 8.8 ± 0.9                     | 0.22          | 0.82          | 15                                       | 17 (12)                                    | 24 (16)                                    | 28 (19)                                    | 30 (20)                                    | 51 (32)                                    |
| Carina............| 6.8 ± 1.6                     | 0.21          | 0.69          | 12                                       | 11 (8)                                     | 14 (10)                                    | 18 (13)                                    | 18 (16)                                    | 29 (19)                                    |
| Leo II............| 6.7 ± 1.1                     | 0.16          | 0.48          | 12                                       | 13 (9)                                     | 19 (13)                                    | 23 (16)                                    | 24 (16)                                    | 40 (26)                                    |
| Sculptor.........| 6.6 ± 0.7                     | 0.11          | 1.5           | 11                                       | 13 (9)                                     | 18 (12)                                    | 21 (15)                                    | 23 (15)                                    | 38 (24)                                    |
| Sextans...........| 6.6 ± 0.7                     | 0.34          | 3.2           | 11                                       | 8 (7)                                      | 10 (7)                                     | 11 (8)                                     | 12 (9)                                     | 14 (10)                                    |
| SMC ..............| 60 < \sigma_s \leq 60*        | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* | 60 < \sigma_s \leq 60* |
| LMC..............| 50 < \sigma_s \leq 50*        | 50 < \sigma_s \leq 50* | 50 < \sigma_s \leq 50* | 50 < \sigma_s \leq 50* | 50 < \sigma_s \leq 50* | 50 < \sigma_s \leq 50* | 50 < \sigma_s \leq 50* |

### Notes
- The names of the satellite galaxies of the Milky Way are given in column (1). We consider only those galaxies with galactocentric distances smaller than 300 kpc. Cols. (2)–(4) give the measured line-of-sight velocity dispersion and the King core and tidal radii for each satellite. The exceptions are the LMC and the SMC. These galaxies have measured rotation speeds, listed in Cols. (6)–(11). All velocities are expressed in units of km s\(^{-1}\) and distances are in kiloparsecs. Cols. (5) gives the value of \(V_{\text{max}}\) assigned to the halo of each satellite by K99. Cols. (6)–(11) and in the first nine table rows give the value of \(V_{\text{max}}\) that we estimated for each satellite on the basis of its measured velocity dispersion and King profile parameters and assuming the primordial power spectra specified at the top of each column. Values of \(V_{\text{max}}\) listed without parentheses were calculated assuming an anisotropy parameter of \(\beta = 0\), and those inside parentheses assume \(\beta = 0.15\). Except for the case of Draco, we use the velocity dispersions and King profile core and tidal radii for the Milky Way satellites given in the review article by Mateo 1998. For Draco we use the parameters quoted by Odenkirchen et al. 2001. The quoted maximum rotation speeds for the LMC and SMC were taken from van der Marel et al. 2002 and Stanimirović 2000. The LMC rotation curve is observed to be flat from 4 kpc out to greater than 8.9 kpc.
- For many of the low-power models (indicated by asterisks), the flat portion of the curve (\(\tau_{\text{max}}\)) is expected to be at larger radius. In order to explain this in the context of these models, we must suppose that baryonic in-fall plays an important role in setting the properties of dark matter rotation curves (Blumenthal et al. 1986). In this case, the measured value of \(V_{\text{max}}\) (\(\tau_{\text{max}}\)) is larger (smaller) than it would be for the pristine halo prior to baryonic contraction. The SMC rotation curve is even more likely to be influenced by baryons (Stanimirović 2000), and baryonic in-fall is likely to be of some importance for all cases. While not demanded by the data, the effects of baryonic in-fall could be important for all satellites, thus the listed \(V_{\text{max}}\) values should be considered lower limits. Lastly, the large value of \(\tau_{\text{max}}\) associated with Draco in the \(\sigma_s = 0.65\) case may be difficult to reconcile with the kinematic data of Kleyna et al. 2002 and may disfavor a model with such low power.

![Fig. 22](image_url)
population without a differential feedback mechanism. The rest of the models overpredict the counts, with the greatest apparent discord in the $n = 1$ case. The specific choice of $\beta = 0.15$ serves mainly to illustrate the effect of a minor anisotropy. We chose this value because it is typical of what is seen in the central regions of simulated dark matter halos (Colin et al. 2000b), and therefore it seems a reasonable possibility for the anisotropy parameter of particles in dwarf galaxies.

5. CAVEATS

In this study we employed a simple, semianalytic model based on many previous studies (LC93; SK99; BKW00; TB01; B01; W02; HFM03) and designed to produce large numbers of halo realizations with minimal computational effort. In developing this model, we have made many simplifying assumptions. In this section we draw attention to many of these shortcomings and discuss how they might affect our results and be improved on in future work.

Among the most obvious omissions in this work is the neglect of any disk or bulge component in each halo. We have specifically chosen to ignore the effects of central galaxies because the physics of dark halo formation is relatively well understood compared with that of galaxy formation. This allows us to ground our work against dissipationless $N$-body simulations. In order to include a galactic component, one is forced to adopt many poorly constrained models and assumptions regarding gas accretion, cooling, angular momentum distributions, feedback, and the effects of substructure on the host galaxy itself. Once a reliable framework for the dark matter has been developed, we can use this as a foundation for more speculative (yet interesting) explorations involving the baryonic components.

A central (disk) galaxy would add to the dynamical friction force experienced by subhalos orbiting near the plane of the disk and cause halos on highly inclined orbits to be tidally heated during rapid encounters with the disk potential (e.g., Gnedin & Ostriker 1999; Gnedin, Hernquist, & Ostriker 1999; TB01). These effects lead to enhanced satellite disruption. Conversely, subsystems that are massive enough to host galaxies might be rendered more resistant to tidal disruption because their central densities would be enhanced by the presence of cool baryons. For low-mass halos, the net effect of a central galaxy would likely be to reduce the substructure count, mainly at small radii. Even without including these effects, we find that the substructure fraction drops significantly at small radii because of the dark matter potential, and that a large part of the projected central mass fraction comes from subhalos at large radii that are picked up in projection. Nevertheless, projected mass fractions are rather sensitive to the size of the core in the subhalo radial distribution (Chen et al. 2003), so if the core region were larger as a result of a central galaxy, the implied lensing signal would be reduced relative to our estimates. We find that eliminating all substructure within 20 kpc of the halo center reduces the projected mass fractions in subhalos less massive than $10^{10} M_\odot$ by $\sim 30\%$.

In addition to the considerable uncertainties associated with galaxy formation, there are potential shortcomings in our efforts to model substructure properties in the context of collisionless dark matter physics alone. For example, we have allowed for only a mild redistribution of mass within the tidal radii of the orbiting subhalos up until the time that the subhalo is totally disrupted. The work of S02 and H03 suggests that this effect may be larger, but these results may have been compromised by limited numerical resolution or inappropriate assumptions regarding initial subhalo orbits and/or accretion times. In this sense, our approach represents a conservative extreme because we assume that the surviving subhalo density structure is typically very similar to that of halos in the field. We study the effects of tidal mass redistribution in a forthcoming paper (J. Bullock, K. Johnston, & A. Zentner, in preparation). In addition, we have adopted a halo concentration relation (B01) that has not been confirmed for $M \lesssim 10^9 M_\odot$. Similarly, the EPS merger tree calculations have yet to be tested in the very low-mass regime. In light of these extrapolations, it is imperative that our results be tested and updated using the next generation of numerical studies.

Finally, our model does not treat the substructure population self-consistently. We have neglected any subhalo-subhalo interactions, which could serve to increase the internal heating of substructure and modify dynamical friction timescales as orbital energy is exchanged between subhalos and traded for internal energy. We have adopted the approximation that all in-falling halos are “distinct” and have no subhalos of their own (see Taylor & Babul 2003b for a study of merger tree “pruning”). However, our “tree-level” calculations suggest that the substructure mass fraction is uniformly $\sim 10\%$ regardless of host mass, so we expect the that this correction would typically affect our derived mass fractions by $\lesssim 10\%$. Considering the assumptions that have gone into our calculations and the current level of observational precision, this level of error is acceptable; however, it may need to be improved on as observations zero in on the masses of the subclumps responsible for the lensing signals and the mass fractions in these subclumps.

6. CONCLUSIONS AND DISCUSSION

The abundance of substructure in dark matter halos is determined by a continuous competition between accretion and disruption. Accreted subhalos with dense cores are resistant to disruption, but over time their orbits decay, their mass is stripped away, and they are often destroyed. The model that we presented here allows us to follow the complicated interplay between density structure, orbital evolution, and accretion time in order to determine how changes in the power spectrum affect the final substructure population in galaxy-sized halos. For a fixed set of cosmological parameters, changes in the power spectrum manifest themselves by changing collapse times for halos, where less power leads to later accretion times and lower densities. We have specifically focused on tilted models that help to relieve the central density crisis facing CDM and that may be favored by joint CMB and large-scale structure analyses (Spergel et al. 2003). We have also considered a BSI inflation model and WDM models, where the power is sharply reduced on small scales.

For a large class of CDM-type models, including models with significant tilt and running, we find that the fraction of mass bound up in substructure, $f$, for galaxy-mass halos is relatively insensitive to the slope of the primordial power spectrum. This is because both the host halos and their accreted subsystems collapse later in these models, in a
roughly self-similar way as power is reduced. Note that this result would have been roughly expected if we were varying only the overall normalization in the models because the relative redshifts of collapse would be invariant (we assume host halos are small enough that they collapse before \( z \sim 0 \)). Our investigation suggests that this intuitive description holds even for tilted and RI models, at least over the parameter range we have explored. All indications are that this insensitivity to the tilt of the power spectrum is a rather robust result and should hold even if some unknown factor has caused our overall normalization in predicted mass fractions to be in error (e.g., our exclusion of central galaxies). Interestingly, the shape of the mass function, \( f(x = M_{\text{sat}}/M_{\text{host}}) \), is also relatively insensitive to the mass of the host halo (eq. [14]), and a similar shape holds for all of the tilted models we explored.\(^{12}\)

The similarity in mass fractions breaks down for models with sharp features in their power spectra, like our BSI and WDM models. In these models, low-mass halo formation is delayed significantly relative to the formation time of their hosts. Consequently, fewer subhalos are dense enough to withstand the tidal field they experience on accretion. We find that for the relevant WDM and BSI models, the mass fraction in substructure is reduced by a factor of \( \geq 3 \) compared with the standard/tilted ΛCDM models.

Inspired by recent attempts to measure substructure mass fractions using multiply imaged quasars, we applied our model to ensembles of host halos with \( M = 3 \times 10^{12} M_{\odot} \) at \( z = 0.6 \), which represents the expectation for massive lens galaxies (DK01; DK02). For the ΛCDM/tilted cases, we found substructure mass fractions within a 10 kpc projected radius in systems less massive than \( M = 10^8, 10^9 \), and \( 10^{10} M_{\odot} \) of \( f_8 \approx 0.2\% - 0.4\% \), \( f_9 \approx 0.4\% - 1.5\% \), and \( f_{10} \approx 0.6\% - 2.5\% \) at the 64 percentile range. These estimates are consistent with, but on the low side of, first attempts to measure the substructure fraction using multiply imaged quasars by DK01, who obtain \( f \approx 0.6\% - 7\% \) at 90% confidence, with an upper mass limit of \( 10^8 - 10^{10} M_{\odot} \) (N. Dalal 2003, private communication). The lensing results disfavor the BSI model, which leads to mass fractions \( f_8 \approx 0.01\% - 0.06\% \), \( f_9 \approx 0.02\% - 0.2\% \), and \( f_{10} \approx 0.03\% - 0.4\% \) at 64%. This is true unless the break scale in the power spectrum is pushed to such a small value that this model no longer has the attractive feature of alleviating the central density problem. A \( m_{\text{vir}} = 0.75 \text{ keV} \) WDM model is similarly disfavored, and even our highest mass WDM case, \( m_{\text{vir}} = 3 \text{ keV} \), has a typical projected fraction (\( f_9 \approx 0.4\% \)) that is low compared with the DK01 estimate. Again, this indicates that if the warm particle is a thermal relic, the mass must be large enough that WDM no longer mitigates the small-scale problems of standard ΛCDM. Yet these results are interesting because they show how lensing may be used as one of the few probes of the WDM particle mass in the range \( \gtrsim 1 \text{ keV} \) or a break in the primordial power spectrum at large wavenumber.

Clearly these conclusions must be regarded with some caution. In addition to the uncertainties of modeling discussed in § 5, other issues make drawing definite conclusions difficult. For example, we have only accounted for the substructure within the virial radius of the host halo, yet the anomalous flux ratios of lensed images are sensitive to the presence of small halos along the line of sight to the lens. Keeton (2003) showed that field halos can have a significant lensing effect even if they are separated from the lens by several tenths in redshift and in hierarchical, CDM-type models, small field halos are ubiquitous. Chen et al. (2003) showed that the relative effect from halos outside the virial radius of the lens is typically a few percent but may be as large as 20%–30% of that from subhalos, depending on assumptions about the subhalo population. Also, as the mass fraction in substructure of a given mass depends on the mass of the host (eq. [14]), it may be important to constrain the host halo mass in order to fully exploit the ability of lensing measurements of substructure to probe cosmology and structure formation.

We compared our model predictions for the cumulative subhalo velocity function, \( N(> V_{\text{max}}) \), with the satellite galaxy count of the Milky Way. The approach here was to estimate \( V_{\text{max}} \) for each satellite galaxy’s dark matter halo on the basis of its observed line-of-sight velocity dispersion, \( \sigma_v \). We emphasized in § 4.5 that the mapping between \( \sigma_v \) and \( V_{\text{max}} \) is sensitive to theoretical prejudice regarding the density structure of the dwarf galaxy’s halo as well as the unknown velocity anisotropy parameter of the system, \( \beta \). For a fixed value of \( \beta \), less concentrated host halos imply larger values of \( V_{\text{max}} \) because halo rotation curves are more slowly rising and stars probe only the inner \( \sim 1 \text{ kpc} \) of the halo. Interestingly, this implies that tilted models and truncated models do significantly better than \( n = 1 \), ΛCDM in reproducing apparent dwarf counts, even though their mass fractions are similar. While our estimates of \( V_{\text{max}} \) cannot be considered robust because of the simplicity of our model and the fact that the \( \sigma_v - V_{\text{max}} \) mapping is very sensitive to the inner structure of the subhalos, the general trends that we would like to persist in more elaborate studies and N-body simulations. Moreover, our results reveal yet another reason that it is difficult to consider the dwarf satellite problem a serious challenge to ΛCDM theory: the nature of the dwarf satellite problem is very sensitive to cosmology, the power spectrum, and assumptions about the shape of the velocity ellipsoid for stars in dwarf galaxies.

When we fix \( \beta = 0 \), our \( dn/d \ln k = -0.03 \) RI model, \( \sigma_v = 0.75 \) model, and the BSI case all do well in matching the known satellite population of the Milky Way for \( V_{\text{max}} \gtrsim 20 \text{ km s}^{-1} \). Our lowest power model (\( \sigma_v = 0.65, n \approx 0.84, \text{ and mild running} \)) actually underpredicts the dwarf count for \( V_{\text{max}} \gtrsim 30 \text{ km s}^{-1} \). However, this result is achieved only for the optimistic assumption of isotropic velocities. If we adopt a small level of anisotropy, \( \beta = 0.15 \), consistent with the centers of simulated dark matter halos, agreement for most models is worsened. Only the BSI and \( \sigma_v = 0.65 \) models show good agreement in this case. Yet even with \( \beta = 0.15 \), the RI and \( \sigma_v = 0.75 \) models still compare more favorably than the \( n = 1 \) case with \( \beta = 0 \).

What do these results imply for the dwarf satellite problem? In all models, including those with truncated power, the velocity function of subhalos continues to rise below the scale of the smallest observed Milky Way satellite, \( V_{\text{max}} \lesssim 10 \text{ km s}^{-1} \). No matter how one modifies the power spectrum, some kind of feedback is required to explain the local satellite population. Different power spectra (even different values of \( \beta \)) seem to indicate that different types of feedback are needed. For example, in

\(^{12}\) Note that these self-similar trends break down on cluster-mass scales since recent accretion, which is a strong function of the overall power, likely plays an important role in these objects.
models with $\sigma > 0.8$ (the precise number depends on typical $\beta$ values and the degree of running/tilt), the feedback must be \textit{differential}. That is, for $V_{\text{max}} \approx 8-30$ km s$^{-1}$, only one of every $\sim 5-10$ halos in this range should form stars. On the other hand, in models like the $dn/dln k = -0.03$ RI model, the BSI case, and our $\sigma_s \approx 0.75$ model with $\beta = 0$, the discrepancy seems to set in suddenly at $V_{\text{max}} \approx 10-20$ km s$^{-1}$, suggesting that nearly all halos smaller than this are completely devoid of stars. In this case, the feedback mechanism must provide a sharp transition.

The feedback mechanism proposed by BKW00 accommodates the need for only $\sim 10\%$ of subsystems to actually host observable galaxies by suggesting that only those systems that formed before reionization were able to retain their gas and eventually form stars. However, if reionization were to occur very early (e.g., $z \gtrsim 15$), many fewer than $10\%$ of these dwarf-sized systems could have collapsed before reionization, so that almost all systems smaller than $V_{\text{max}} \approx 30$ km s$^{-1}$ would be dark. This would be more in line with what we see for the low-power models. This is an intriguing result. The best-fit power spectrum of the \textit{WMAP} team (Spergel et al. 2003) leads to similar substructure mass fractions as standard CDM, alleviates the central density problem and dwarf satellite discrepancy, and forces us to consider feedback mechanisms that predict a sharp transition between luminous and nonluminous galaxies. Additionally, the possible detection of early reionization by the \textit{WMAP} team (Kogut et al. 2003) provides a feedback mechanism that results in a sharp transition. Of course, explaining early reionization in models with low small-scale power may be problematic (Somerville et al. 2003). Another feedback scenario that leads to a sharp transition is photo-evaporation (Barkana & Loeb 1999; Shaviv & Dekel 2003). Nonetheless, the uncertainty associated with $\beta$ in determining satellite galaxy $V_{\text{max}}$ values suggests that efforts to model dwarf galaxy luminosities as well as dynamical properties will be required to resolve this issue (Somerville 2002; Benson et al. 2002).

Remaining uncertainties in understanding the precise nature of the dwarf satellite problem highlight the need to focus on attempts to measure CDM substructure by other means. Continued efforts to detect substructure via gravitational lensing (e.g., DK01; Keeton 2003; Keeton, Gaudi, & Petters 2003; Moustakas & Metcalfe 2003) or by probes within our own Galaxy (e.g., Johnston et al. 2002; Ibata et al. 2002a, 2002b; Mayer et al. 2002; Font et al. 2001) offer useful avenues for doing so. Modeling of the kind presented here may play an important role in interpreting results of ongoing observational programs and bring us one step closer to confirming or refuting one of the fundamental predictions of the CDM paradigm.

We are pleased to thank Neal Dalal, Kathryn Johnston, Charles Keeton, James Kneller, Chris Kochanek, Leon Koopmans, Andrey Kravtsov, Eugene Lim, Ari Muller, Benton Metcalf, Leonidas Moustakas, Joel Primack, David Rusin, Bob Scherrer, Rachel Somerville, Gary Steigman, Rob Swaters, Terry Walker, Risa Wechsler, and David Weinberg for many helpful discussions. We are also grateful to Neal Dalal, Michael Schirber, Risa Wechsler, David Weinberg, and Joshua Winn for valuable comments on an early draft of this manuscript. J. S. B. acknowledges Rosemary Wyse for her generous hospitality during visits to Johns Hopkins University, where a substantial fraction of this work was done. A. R. Z. is supported by Ohio State University and by US DOE contract DE-FG02-91ER40690. J. S. B. is supported by NASA through Hubble Fellowship grant HF-01146.01-A from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.
