QUASARS: WHAT TURNS THEM OFF?

ROBERT J. THACKER,1,2 EVAN SCANNAPIECO,3 AND H. M. P. COUCHMAN4

Received 2006 June 2; accepted 2006 August 23

ABSTRACT

While the high-redshift quasar luminosity function closely parallels the hierarchical growth of dark matter halos, at lower redshifts quasars exhibit an antihierarchical turnoff, which moves from the most luminous objects to the faintest. We explore the idea that this may arise from self-regulating feedback, caused by quasar outflows. Using a hybrid approach that combines a detailed hydrodynamic simulation with observationally derived relationships, we calculate the luminosity function of quasars down to a redshift of \( z = 1 \) in a large, cosmologically representative volume. Outflows are included explicitly by tracking halo mergers and driving shocks into the surrounding intergalactic medium, with an energy output equal to a fixed 5% fraction of the bolometric luminosity. Our results are in excellent agreement with measurements of the spatial distribution of quasars on both small and large scales, and we detect an intriguing excess of galaxy-quasar pairs at very short separations. Our results also reproduce an antihierarchical turnoff in the quasar luminosity function; however, this falls short of that observed, as well as that predicted by analogous semianalytic models. The difference can be traced to the treatment of gas heating within galaxies and the presence of in-shock cooling. The simulated galaxy cluster \( L_X-T \) relationship is close to that observed for \( z \approx 1 \) clusters, but the simulated galaxy groups at \( z = 1 \) are significantly perturbed by quasar outflows. Measurements of anomalously high X-ray emission in high-redshift groups, along with detections of 1000 km s\(^{-1}\) winds in poststarburst ellipticals, would provide definitive evidence for the AGN-heating hypothesis.

Subject headings: cosmology; theory — galaxies: evolution — intergalactic medium — large-scale structure of universe — quasars: general

Online material: color figures

1. INTRODUCTION

In the low-redshift universe, active galactic nuclei (AGNs) are not very active. While at high redshifts the quasar luminosity function increases with time, since \( z \approx 2 \) the number density of optically selected AGNs has been dropping dramatically (Schmidt & Green 1983; Boyle et al. 1988, 2000; Koo & Kron 1988; Pei 1995; Fan et al. 2001). Deep X-ray surveys have shown that this downturn occurs antihierarchically, such that the spatial density of AGNs with higher X-ray luminosities peaks earlier than that of lower luminosity AGNs (Steffen et al. 2003; Ueda et al. 2003). Complementary emission-line studies suggest that this trend is driven by a decrease in the characteristic mass of actively growing black holes (Heckman et al. 2004) and is likely to closely parallell the formation history of early-type galaxies (e.g., Granato et al. 2001, 2004). Furthermore, optical and near-infrared observations indicate that the largest galaxies were already in place by \( z \approx 2 \), while smaller ones continued to form stars at much lower redshifts (Pozzetti et al. 2003; Fontana et al. 2004; Glazebrook et al. 2004; van Dokkum et al. 2004; Treu et al. 2005), and a similar trend is observed in the morphological evolution of galaxies (Bundy et al. 2005).

Yet, despite these observations, such widespread galaxy “downsizing” (Cowie et al. 1996) was unexpected. The cold dark matter (ΛCDM) model, while in spectacular agreement with observations (e.g., Spergel et al. 2003), is a hierarchical theory, in which gravitationally bound structures grow by accretion and merging. Superposed on this distribution is the baryonic component, which falls into the dark matter potential wells, shock heats, and must radiate this energy away before forming stars (Rees & Ostriker 1977; Silk 1977). The larger the structure, the longer it takes to cool, and thus, galaxy evolution should be even more hierarchical than structure formation.

Recently, several theoretical studies have shown that the missing element in this picture could be kinetic feedback from AGNs. As bulge and black hole masses are closely related (Gebhardt et al. 2000; Merritt & Ferrarese 2001; see also King 2003, 2005), such outflows would have the largest impact on the largest forming elliptical galaxies, suppressing their formation first (Scannapieco & Oh 2004, hereafter SO04; Binney 2004; Di Matteo et al. 2005; Croton et al. 2006). This would also help to explain the X-ray luminosity-temperature relationship observed in the intracluster medium (ICM) in galaxy clusters. If nongravitational heating were unimportant, the gas density distribution would be self-similar, resulting in \( L_X \propto T^2 \) (Kaiser 1986), but instead the observed slope steepens considerably for low-temperature clusters (e.g., David et al. 1993; Arnaud & Evrard 1999; Helsdon & Ponman 2000). Furthermore, the 100 keV cm\(^2\) level of preheating necessary to explain this discrepancy (Cavaliere et al. 1998; Kravtsov & Yepes 2000; Wu et al. 2000; Babul & Rees 2000) is not arbitrary. Rather, it corresponds to the threshold value for cooling within the age of the universe (Voit & Bryan 2001; Oh & Benson 2003).

Taken together, these observations are strongly suggestive of a model in which the properties of the ICM, the formation history of elliptical galaxies, and the evolution of the quasar luminosity function are all set by self-regulating AGN feedback. In fact, SO04 have shown that the addition of AGN outflows into the semianalytic model developed by Wyithe & Loeb (2002, 2003) can reproduce the drop in the AGN luminosity at low redshifts, as heating this gas slows the accretion of matter onto the supermassive...
black hole. The central AGN engine is then starved of fuel, and a strong suppression at the bright end of the luminosity function occurs. The main drawback of the model, however, is that precise matching of observational results requires fine tuning of the model parameters, such as black hole mass and outflow efficiency.

Several other numerical and analytic investigations have sharpened our understanding of various aspects of this process. Ciotti & Ostriker (2001) conducted a detailed spherically symmetric simulation of infall, modulated by feedback, in elliptical galaxies. Using empirically motivated relationships to derive a model for black hole growth, Haiman et al. (2004) were able to directly retrodict that under the assumption that star formation in spheroids follows black hole mass growth, the luminosity function of quasars must fall dramatically between $z = 2$ and 0. Binney (2004) conducted an analytic study of the impact of AGNs on inhibiting gas cooling in large galaxies, while Sazonov et al. (2005) explored radiative AGN feedback by investigating the impact of photoionization and Compton heating on gas accretion. Robertson et al. (2006) carried out smoothed particle hydrodynamic (SPH) simulations of AGN outflows in individual galaxy mergers and studied the role of feedback in determining the colors of elliptical galaxies and establishing the relationship between stellar velocity dispersion and black hole mass. This suite of simulations was related analytically to the global evolution of quasars and elliptical galaxies in a series of papers by Hopkins et al. (2005c, 2005a, 2005b, 2006). Levine & Gnedin (2005) combined cosmological simulations with an analytic model to constrain the filling factor of AGN outflows as a function of redshift. Scannapieco et al. (2005) emphasized the role that quasars may play in the downsizing of the star-forming galaxy population. Levine & Gnedin (2006) studied the impact of AGN outflows on the matter power spectrum. Menci et al. (2006) used a semianalytical model to study the role of AGN feedback on the color distribution of galaxies from $z = 0$ to 4. The importance of the nature of gas accretion in determining the effectiveness of feedback processes has recently been discussed in Dekel & Birnboim (2006) and Cattaneo et al. (2006). Finally, Croton et al. (2006) combined semianalytic models with the dark matter evolution taken from the “Millennium Simulation” (Springel et al. 2005) to study the impact of a more temporally extended model of AGN feedback on the bright end of the galaxy luminosity function.

In this paper we undertake the first detailed hydrodynamic simulations of quasar outflows in a general cosmological context. Adopting a burst model that associates AGNs with merger events and a global outflow model that is based on our simulations of high-redshift starbursts (Scannapieco et al. 2001, hereafter STD01), we are able to track in detail the impact that quasars have on both their own formation and the properties of galaxy clusters and the intergalactic medium (IGM).

The structure of this work is as follows. In § 2 we describe our overall numerical approach, method for quasar identification, and implementation of outflows. In § 3 we compare our simulation results with measures of quasar clustering and the observed luminosity function in the optical and the X-ray bands. In § 4 we study AGN feedback in a more global context, examining its impact on the properties of the intergalactic and intracluster media. In § 5 we present a discussion and conclusion.

2. SIMULATIONS OF QUASAR FORMATION

2.1. Overall Numerical Model

Motivated by measurements of the cosmic microwave background, the number abundance of galaxy clusters, and high-redshift supernova distance estimates (e.g., Spergel et al. 2003; Vianna & Liddle 1996; Riess et al. 1998; Perlmutter et al. 1999), we focus our attention on a cold dark matter cosmological model with parameters $h = 0.7$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.046$, $\sigma_8 = 0.9$, and $n = 1$, where $h$ is the Hubble constant in units of $100$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m$, $\Omega_\Lambda$, and $\Omega_b$ are, respectively, the total matter, vacuum, and baryonic densities in units of the critical density, $\sigma_8$ is the variance of linear fluctuations on the $8 \times 10^{26}$ Mpc scale, and $n$ is the “tilt” of the primordial power spectrum. The Eisenstein & Hu (1999) transfer function is used throughout.

As in our earlier work (STD01), simulations were conducted with a parallel OpenMP-based implementation of the Hydra code (Thacker & Couchman 2006) that uses the adaptive particle-particle, particle-mesh algorithm (Couchman 1991) to calculate gravitational forces and the SPH method (Lucy 1977; Gingold & Monaghan 1977) to calculate gas forces. Gas densities and energies are calculated using the standard SPH smoothing kernel method (for exact details see STD01), with the kernel tuned to smooth over 52 particles; and radiative cooling is calculated using standard tables (Sutherland & Dopita 1993). We have kept the metallicity constant at $Z = 0.05$ to mimic a moderate level of enrichment in the galaxy formation process. However, this is an underestimate of the intracluster metallicity, $Z = 0.3$, which seems to be an approximately universal value to intermediate redshifts (Tozzi et al. 2003). Finally, because the epoch of reionization is poorly known, and because we are primarily focusing our attention on mass scales greater than $10^{10} M_\odot$, we do not include a fiducial photoionization background in the simulation.

We also do not include the so-called $\nabla h$ terms (Nelson & Papaloizou 1994; Serna et al. 1996; Springel & Hernquist 2002) in our implementation of SPH. This is potentially a significant concern in this investigation, as we examine gas entropy. However, tests on an expanding spherical shell problem (see STD01) using 1000 particles and no artificial viscosity or cooling show entropy conservation to be accurate at the 6% level, while the combined gravitational and hydrodynamic energy error is around 1.5%. These findings are in broad agreement with the discussions presented in Hernquist (1993) and Springel & Hernquist (2002), in which a small entropy conservation error was always accompanied by a larger energy error when integrating the evolution of the entropic function in the absence of $\nabla h$ terms. While a lower error in the entropy is desirable, for the present phenomenological investigation and given the gains we get from using a code without $\nabla h$ terms, we consider this error acceptable.

We simulated a number of different box sizes and particle numbers to quickly assess the accuracy and numerical resolution dependencies in our model. A single large simulation was then run for statistical purposes, allowing us to probe the bright end of the luminosity function. The specifics of each simulation, including box size and resolution, are given in Table 1. We note that attempting to simulate the formation of the very brightest end of the luminosity function with sufficient resolution to track smaller mergers is a difficult task due to the scarcity of these objects. This is the fundamental motivation behind our progressing to a simulation with $2 \times 640^3$ particles.

2.2. Identification of Quasars

A secondary motivation of this paper is to compare the simulation results directly with semianalytic predictions. We have therefore taken the outflow model of SO04 and adapted it to our simulation as closely as possible, although in some cases, which we highlight, either it was not possible for us to match this model exactly, or we have chosen to make well-motivated changes. While a number of the physical parameters we use in the model could in principle be measured within the simulation, we do not
have high enough resolution within the individual galaxies or the implementation of the subresolution physics (i.e., stellar feedback) necessary to make this approach reasonable. For completeness we reiterate the salient features of the SO04 model in our discussion below.

While in our previous work it was sufficient to track group mass evolution to identify star-forming regions, to evaluate the quasar luminosity function it is necessary to track mergers of groups. We have used the same method as STD01 for identifying groups, which relies on the local baryonic density field to pinpoint centers of mass and a spherical overdensity procedure applied to identify the baryon group. From the baryon group (defined as gas below $10^5$ K) an estimation of the total halo mass is derived by multiplying by $\Omega_b/\Omega_c$. The resulting mass distribution function is in close agreement with that derived from friends-of-friends (FOF), provided suitable limits for the baryon spherical overdensity are chosen (see STD01 for an in-depth discussion of this issue). A secondary benefit of inferring halo mass from gas content is that it can be used as a proxy for the subresolution physics of quasar quenching within gas-starved environments. As gas is removed from systems by AGN-type feedback there is an inferred reduction in circular velocity, which produces an associated reduction in black hole mass and thus quasar luminosity. To track merger events we rely on group labeling, and we label a merger as having occurred when at least 30% of the accreted mass does not come from a single massive progenitor. This procedure means that the first groups to form are also treated as merger events. For each merger event that will be tagged as a quasar we store the details including position and redshift in an output file.

Once a group has been identified as satisfying this criterion, the dynamical time associated with the cold gas disk, which feeds the AGN, and the mass associated with the black hole must be calculated. Note, however, our resolution is insufficient to resolve the inner structure of galaxies, and we thus assume that the baryon mass found is resident in a cold gas disk. For a given redshift $z$ and virial density $\rho_c(z)$, the implied virial radius for a group of $N$ gas particles (below $10^5$ K) with mass $m_g$ is

$$r_v = \left[ \frac{N m_g \Omega_0}{4/3 \pi \rho_c(z)} \right]^{1/3}.$$  

(1)

The circular velocity is then

$$v_c = \left[ \frac{4}{3} \pi G \rho_c(z) r_v \right]^{1/2},$$  

(2)

and the dynamical time $t_d$ associated with a cold gas disk of size $r_o$ is (Mo et al. 1998), where $\lambda = 0.05$ is the spin parameter, is

$$t_d = 0.035 r_v / v_c.$$  

(3)

Note that in order to maintain the same relationship between outflow velocity and black hole mass as Wyithe & Loeb (2002) and SO04, we choose a slightly larger time than was used in these studies. This is because our numerical model also includes thermal energy input to establish the correct initial postshock temperature according to equation (9) below.

While the observed $M_{bh} - \sigma_v$ relation (Merrit & Ferrarese 2001; Tremaine et al. 2002) infers that the black hole mass scales as $\sigma_v^4$, where $\sigma_v \sim 4-4.5$, the $M_{bh}/v_c$ slope is slightly steeper because the $v_c - \sigma_v$ slope is shallower than linear (Ferrarese 2002). Thus, we assume an $M_{bh}/v_c$ relationship given by

$$M_{bh} = 2.8 \times 10^{8} \left( \frac{v_c}{300 \text{ km s}^{-1}} \right)^5 M_{\odot}. \quad (4)$$

Finally, the black hole is assumed to shine at its Eddington luminosity $\left(1.2 \times 10^{38} \text{ ergs s}^{-1} M_{\odot}^{-1}\right)$ for the dynamical time, $t_d$. Note that these values are slightly different from in SO04, but the overall relationship between $v_c$ and luminosity is the same.

2.3. Outflow Implementation

Each AGN in our simulation is assumed to channel a fixed fraction $\epsilon_k$ of its bolometric energy into a kinetic outflow. The amount of energy deposited into the outflow is then

$$E_k = 1.2 \times 10^{38} \epsilon_k \left( \frac{M_{bh}}{M_{\odot}} \right) \left( \frac{t_d}{1 \text{ s}} \right) \text{ ergs}. \quad (5)$$

As in SO04 we adopt $\epsilon_k = 0.05$ throughout this investigation, which is consistent with other literature estimates (e.g., Furlanetto & Loeb 2001; Nath & Roychowdhury 2002). The majority of mass in the outflow at the resolution we can simulate comes from material surrounding the cold gas group. Therefore, as in our previous work, we model the expanding outflow as a spherical shell outside of the virial radius of the system. While the assumption of a spherical shell is a significant oversimplification, given the bipolar nature of outflows, it is worth recalling that within the intracluster medium in galaxy clusters a bipolar outflow will still launch an ellipsoidal cocoon of shocked intracluster gas (Begelman & Cioffi 1989). Hence, we place the expanding outflow at a radius $r_o = 2r_v$, and rearrange all gas below a density threshold of $2.5 \rho_v$ within this radius but outside $r_v$ into an expanding shell of $N_{shell}$ particles. The density threshold prevents us from redistributing cold gas, which is known to be very stable against incoming shocks in SPH simulations. The outflow particles are arranged on two concentric spherical shells of radius $r_o$ and $0.9 r_v$. This multishell structure assures that the outflows will be sufficiently well-resolved radially to be reasonably treated by the SPH solver. Sufficient resolution along each of the shells is assured by arranging the particles in each of the shells to be anticorrelated such that no two particles are within a distance of less than one-half of the average spacing between neighbors. This
minimizes the subrandom particle distribution that arises naturally within SPH. The particles in both shells are then given velocities

\[ v_{\text{outflow}} = v_{\text{cm}} + \frac{L}{r_{\text{shell}}} + v_{\text{ang}}, \]

where \( v_{\text{cm}} \) is the center-of-mass velocity of the galaxy, \( v_s \) is a boost in the radial direction due to the outflow (discussed below), and \( v_{\text{ang}} \) is an axial velocity necessary to conserve angular momentum,

\[ v_{\text{ang}} = \frac{L \times r_{\text{shell}}}{(2r_e)^2 N_{\text{shell}}}, \]

and \( L = \sum_{i=N+1}^{N_{\text{shell}}} r_i \times v_i \). Once we have established the amount of mass available to create the shell, \( M_s = N_{\text{shell}} m_g \), the velocity of the shell \( v_s \) is calculated from

\[ v_s = 1.13 \left( \frac{E_k}{M_s} - G m_g \frac{N_{\text{shell}}}{N_{\text{shell}} + 1} \sum_{i=N+1}^{N_{\text{shell}}} \frac{1}{r_i} \right)^{1/2}, \]

where the second term on the right-hand side denotes the potential energy subtracted as particles are moved from their initial position to the shell.

Note that the prefactor in equation (8) is less than \( \sqrt{2} \), as a fraction of \( E_k \) is channeled into establishing the correct postshock temperature in the outflowing gas. This heating is particularly important, since it will help determine the fraction of impacted gas that is able to cool within a Hubble time. Under the assumption that the shell behaves like a strong shock, the postshock temperature \( T_s \) is

\[ T_s = \frac{3\mu m_p k_B^2}{16 \Lambda_B} = \frac{(13.6 \text{ K}) \nu_s^2}{(1 \text{ km s}^{-1})^2}. \]

While the semianalytic model in SO04 assumes that this heating applies to the galaxy as well, here we only heat the material in the outflowing shell. Our motivation for this choice is the short cooling time of gas in galactic halos and, on a secondary level, the collimated bipolar nature of the outflows. The radial expansion of the shell agrees with analytic predictions (STD01).

Since our resolution is insufficient to provide detailed knowledge of the inner structure of galaxies, we implement star formation on the basis of a merger model. Following a major merger we convert 10% of the gas in the galaxy into star particles. While this method is known to be a good model of high-redshift star formation in low-mass halos (STD01), it does not track quiescent-mode star formation, which is the primary mode of star formation in the higher mass galaxy population. The primary benefit to the simulation is improved calculation speed as SPH particles in cold, dense regions are converted into star particles. We emphasize that the purpose of this study is to focus on the hydrodynamic evolution of the IGM, and we are not attempting to calculate a luminosity function for galaxies; instead, we use them largely as a tracer population.

2.4. The SO04 Outflow Model

Since comparisons to the SO04 model are made a number of times in this paper, we review some of the key features of the outflow implementation in this model. The energy injection in the outflow event is assumed to be a point-source explosion of energy \( E \sim \epsilon_L L_{\text{bol}} t_{\text{dyn}} \), where \( \epsilon_L = 0.05 \) is the wind efficiency. Both \( L_{\text{bol}} \) and \( t_{\text{dyn}} \) are calculated in the same way as discussed earlier. Unlike the simulation outflow, radiative cooling is ignored, as is work against the potential well in which the outflow is embedded. As the semianalytic model does not provide details of the surrounding density field, the outflow is assumed to expand into a medium of uniform overdensity \( \delta_s \).

Under these assumptions the outflow is thus well described by a Sedov-Taylor blast (Sedov 1959). At a given redshift \( z \) and outflow expansion time \( t \), the physical radius of the shell \( R_s \) is described by

\[ R_s = \zeta_0 \left( \frac{E_k}{\rho} \right)^{1/5} \left( \frac{E_{60}}{\delta_s^{1/5}} \right)^{1/5} (1 + z)^{-3/5} t_{\text{Gyr}}, \]

where \( E_{60} \) is the energy in the hot medium in units of \( 10^{60} \text{ ergs} \), \( t_{\text{Gyr}} \) is the expansion time of the outflow in gigayears, the overdensity of the surrounding medium \( \delta_s \) is defined such that \( \delta_s = \rho_s/\rho_m \), rather than the usual definition \( \rho_s/\rho_m - 1 \), and \( \zeta_0 = 1.17 \) for a \( \gamma = 5/3 \) gas (e.g., Shu 1992). The shell velocity is thus

\[ v_s = \frac{2R}{St} = 1500 R_s/M_{\text{Mpc}} (\Delta_0^{1/2} T_s^{1/2} (1 + z)^{-1/2}) \text{ km s}^{-1}, \]

where \( R_s/M_{\text{Mpc}} \) is the shell radius in Mpc. Using equation (9) for the postshock temperature and the postshock density of \( \rho_s = (\gamma + 1)(\gamma - 1) \rho_b \delta_s = 4 \rho_b \delta_s, \)

the postshock entropy, \( S = T/n^{7/3} \), is given by

\[ S_s = (1.8 \times 10^4 \text{ keV cm}^{-2}) E_{60} \delta_s^{5/3} (1 + z)^{-1} R_s/M_{\text{Mpc}}. \]

As emphasized in Oh & Benson (2003), the ability of gas to cool is insightfully described from the perspective of entropy. Under the usually used definition of entropy, \( S = T/n^{7/3} \), the isobaric cooling time can be written

\[ t_{\text{cool}} = \frac{(3/2) n_k T}{n_e^2 \Lambda(T)} = S^{1/2} \left[ \frac{3\mu m_p}{2\mu^2 T^{3/2} \Lambda(T)} \right] = S^{1/2} F(T), \]

where \( \mu = 0.62, \mu_e = 1.18 \), and \( F(T) \) serves as temperature-dependent normalization. By equating the cooling time to the Hubble time, \( t_{\text{H}} \), we can derive a critical entropy value: gas above this entropy limit is unable to cool within the Hubble time and is thus effectively removed from the galaxy formation process. The critical value is

\[ S_{\text{crit}} = (280 \text{ keV cm}^{-2})(1 + z)^{-1} \left[ \frac{E(z)}{(1 + z)^{1/2}} \right]^{2/3} \times \left[ \frac{\Lambda(T_{\text{min}})}{6.3 \times 10^{-22} \text{ ergs s}^{-1} \text{ cm}^{-3}} \right], \]

where \( \Delta(T_{\text{min}}) \) corresponds to the minimum cooling time below \( 10^8 \text{ K} (T_{\text{min}} \sim 2.3 \times 10^4 \text{ K}) \) and \( E(z) = [\Omega_m (1 + z)^3 + \Omega_{\Lambda}]^{1/2} \).

Having defined the critical entropy, we can now evaluate the radius, \( R_{\text{heat}} \), of the region, which the outflow heats above this value by inverting equation (13), to give

\[ R_{\text{heat}} = (5.6 \text{ Mpc}) \zeta_{100, \text{crit}} (z^{-1} P_{\text{en}}^{1/3} \delta_s^{5/9} (1 + z)^{-5/3}). \]
The mass of this region, \( M_{\text{heat}} \), is
\[
M_{\text{heat}}(\delta, z, M) = (4.6 \times 10^{12} M_\odot) S_{200\text{,crit}}(z)^{-1} \times E_{100} \delta_z^{-2/3} (1 + z)^{-2}.
\] (17)

Finally, we can equate this total shocked mass to the radial expansion (eq. [10]) to give the timescale on which this mass is heated, namely,
\[
t_{\text{heat}} = (20 \text{ Gyr}) S_{200\text{,crit}}(z)^{-5/6} E_{100}^{1/3} \delta_z^{-8/9} (1 + z)^{-8/3}. \]
(18)

For low-redshift sources the time can be quite long, which means that after a quasar is visibly extinguished the outflow may still be contributing significantly to local heating.

3. DISTRIBUTION AND EVOLUTION OF QUASARS

3.1. Quasar Clustering

In the following three subsections we study the optical properties of quasars, as quantified in the rest-frame \( B \) band. In keeping with our previous investigations, as well as the observations in Elvis et al. (1994), we relate the luminosity in this band to the overall bolometric luminosity by assuming a fixed ratio of \( L_{\text{bol}} = 10.4 L_B \) at all luminosities and redshifts. Fixing this value also allows for direct comparison with Wyithe & Loeb (2003). Finally, the \( L_{\text{bol}} \) of the quasar associated with each outflow is simply computed as \( L_{\text{bol}} = E_k \xi t_j^{-1} \).

We begin by addressing the spatial distribution of quasars, which serves as a check on our merger-based approach. To quantify this distribution, we construct the spatial correlation function of quasars using the center-of-mass information from the outflow data produced in the simulation. In principle, this should be computed while accounting for the finite lifetime of each quasar, according to equation (3). In practice, these times are long enough that such effects can be ignored for distances \( \leq c t_d \approx 20(1 + z)^{-1/2} \text{comoving Mpc} \). Thus, we calculate the three-dimensional real-space correlation function using the simplest estimator,
\[
\xi_{qq}(z, m) + 1 = \frac{DD(z, m)_{\text{k}}}{RR(z, m)_{\text{k}}},
\]
(19)

where \( DD(z, m)_{\text{k}} \) is the number of pairs with a magnitude greater than some limit, separated by a comoving difference corresponding to a bin \( k \), and \( RR(z, m)_{\text{k}} \) is the average number of pairs that would be found at a given separation in a random distribution of points with an overall density equal to the mean density of observable quasars.

We next adopt a fixed magnitude limit in the \( B \) band of 20.84, to allow for comparisons with observations from the 2dF quasar redshift survey (Croom et al. 2001, 2002), which have an overall photometric \( b \)-band limit of 20.9, where \( B \approx b + 0.06 \) (Goldschmidt & Miller 1998). This can be computed as
\[
B = 5.5 - 2.5 \log \left( \frac{L_B}{L_\odot} \right) + 5 \log \left( \frac{d}{10 \text{ pc}} \right) + 2.5(1 - \alpha_v) \log (1 + z),
\]
(20)

where \( d \) is the comoving distance to the quasar and \( \alpha_v = -0.5 \) is the typical slope of the quasar power-law continuum (Wyithe & Loeb 2005).

A detailed examination of short-range quasar correlations is given in the Sloan binary quasar study of Hennawi et al. (2006).
examining the short-scale clustering excess observed by Hennawi et al. (2006). The precise position and magnitude of the clustering excess are reproduced extremely well within our simulation, which we take as both support for our approach and validation of the Hennawi et al. (2006) results.

This turnup, which has also been observed in the high-redshift Lyman-break galaxy population (Ouchi et al. 2005), as well as in a local sample of galaxies (Zehavi et al. 2004), is most likely due to gravitationally bound pairs of quasars that are orbiting each other. This so-called one-halo contribution (e.g., Bullock et al. 2002; van den Bosch et al. 2003; Magliocchetti & Porciani 2003) should become important at distances less than

\[ d_{1\text{halo}} \approx \left[ \frac{m_Q}{(4\pi/3)\Omega_0 \rho_c} \right]^{1/3}, \tag{22} \]

where \( m_Q \) is the mass of the halos associated with the quasars above our magnitude limit. From the large-scale clustering this is \( \approx 2 \times 10^{12} \, M_\odot \), corresponding to the \( \approx 1.0 \) Mpc position of the turnup, lending further weight to this interpretation.

### 3.2. Quasar-Galaxy Cross-Correlation Function

As a complementary investigation, we also examine the cross-correlation function between quasars and galaxies, \( \xi_{qg} \). By cross-correlating these two populations we are directly able to evaluate whether galaxies containing quasars are clustered differently than similar mass quiescent galaxies. Early observational attempts to measure \( \xi_{qg} \) were limited by sample size, and thus a bias toward two-dimensional angular measurements prevailed (e.g., Ellingson et al. 1991; Smith et al. 1995; Croom & Shanks 1999). However, the SDSS and 2dF quasar surveys have enabled cross-correlations in three dimensions below a redshift limit of \( z < 0.3 \) (Croom et al. 2004; Wake et al. 2004), and a study at intermediate redshifts using the DEEP2 data has been undertaken (Coil et al. 2006). To date, these investigations have not uncovered any bias in \( \xi_{qg} \) on scales down to the minimum scale to which they are sensitive, which is around 1 Mpc.

To evaluate \( \xi_{qg} \) we use the quasar catalog from § 3.1, combined with a galaxy catalog evaluated with a FOF group finder on the baryonic material in our simulation. We use a linking length \( b = 0.065 \) to find groups with an outer density limit of \( \delta \simeq 2000 \). With a baryonic mass cut of \( 10^{10.5} \, M_\odot \), we find 37,995 groups, and for \( 10^{11} \, M_\odot \), we find 11,857 groups. The estimator for the cross-correlation function is

\[ \xi_{qg}(z, m)_k + 1 = \frac{D_q D_g(z, m)_k}{R_q R_g(z, m)_k}, \tag{23} \]

where \( D_q D_g(z, m)_k \) is the number of quasar-galaxy pairs above the magnitude limit in bin \( k \), and \( R_q R_g(z, m)_k \) is the number of quasar-galaxy pairs that would be expected if these objects were randomly distributed with the same densities as in our simulation.

Our results are plotted in Figure 2, in which we now employ an absolute magnitude limit of \( M_B = -22 \), to better compare with observations. On scales larger than 1 Mpc, \( \xi_{qg} \) is indistinguishable from \( \xi_{gg} \), in agreement with observations. This is true regardless of whether we choose the \( M_B > 10^{10.5} \, M_\odot \) or \( M_B > 10^{11} \, M_\odot \) galaxy populations. However, on scales below 600 h\(^{-1}\) kpc, \( \xi_{qg} \) exhibits a clustering enhancement. At first glance, this would appear to be consistent with our results from § 3.1, which show that one-halo effects can create an excess of clustering at small scales. However, the explanation cannot be this straightforward. In Figure 2 (bottom) we plot the ratio \( \xi_{qg}/\xi_{gg} \), which shows explicitly the turnover from large-scale agreement to short-scale excess. At large separations, active galaxies are clustered very similarly to the general population, consistent with the DEEP2 results. Yet at small separations, a dramatic change occurs. Despite the fact that the one-halo contribution is implicitly included in \( \xi_{qg} \), the amplitude of the break in the cross-correlation function exceeds that of the galaxy autocorrelation function by a factor of \( \approx 2.5 \).

Thus, it appears that our identification of quasars with mergers enhances their clustering on the smallest scales. At first it seems that this is strongly at odds with previous theoretical studies of mergers (Percival et al. 2003; Scannapieco & Thacker 2003). However, these studies were targeted to separations larger than 1 Mpc, where the two-halo term is the dominant contribution to the correlation function. The excess we find here occurs purely in the one-halo regime, meaning that mergers have an excess of very close neighbors.

In support of our results, local studies of quasar-galaxy clustering have found an excess of galaxies at radii \(< 0.5 \) Mpc, completely analogous to the ones in our simulation. Studying a \( z < 0.3 \) sample of quasars drawn from the Sloan Digital Sky Survey
(SDSS), Serber et al. (2006), have found that \( M_\text{ex} < -23.3 \) quasars are more than 3 times more clustered than \( L^* \) galaxies on \( \leq 0.1 \) \( h^{-1} \) Mpc scales, although they cluster similarly to \( L^* \) galaxies on larger scales. While this study was carried out at a much lower redshift, these quasars have roughly the same intrinsic magnitudes as those in Figure 2, as can be estimated assuming a typical redshift of \( z \approx 1.4 \), which for our simulated sample gives a distance modulus of \( \approx 45 \) or \( M_B \lesssim -24 \). Thus, there seems to be a mounting observational and theoretical evidence that the correlation function of the products of mergers, while only very weakly enhanced at large separations, may nevertheless be more strongly enhanced in the one-halo regime.

In fact, an intriguing possibility is that this is caused by three-body interactions in which a third galaxy removes angular momentum from a nearby close pair. This suggests that dynamical friction may not always be the dominant process driving galaxy mergers. Rather, a significant number may be caused by a process more akin to the formation of tight binaries in dense star clusters (e.g., Rasio et al. 2000). Clearly this issue merits future investigation.

3.3. Optical Quasar Luminosity Function

To construct the luminosity function for each luminosity and redshift bin, we calculate the number of quasars in this bin times the total time these objects are shining and divide by the time interval, the width of the bin, and the volume of the simulation. That is, for a given redshift bin \( i \) and a given luminosity bin \( j \) the luminosity function is simply

\[
\Psi_{i,j} = \frac{1}{V \Delta t \Delta L_{B,j}} \sum_{k \in \text{bin}_{i,j}} N_{d,k},
\]

where the sum is over all quasars with redshifts and luminosities associated with the \( i, j \) bin, which spans a time interval \( \Delta t \) and a range of luminosities \( \Delta L_{B,j} \). In Figure 3 we plot the resulting luminosity function for our fiducial 1020 simulation.

The dotted line in this plot gives the Wyithe & Loeb (2003) estimate of the luminosity function, which is simply based on a merger prescription and does not account for feedback. Comparing this estimate with our simulation results uncovers a clear turndown in the number of \( L_B \geq 10^{13} L_\odot \) quasars at \( z \leq 2 \). However, comparing the simulation results with the measured points makes it immediately clear that this turndown is not as strong as seen in the observations. This means that the suppression is much weaker in the simulation than in the semianalytic SO04 model, which was found to be a good fit to the observations. This is a surprising result that needs to be understood in more detail, and, to explore this further, we recalculate the luminosity function but impose by hand the precise gas-heating methodology used in the semianalytic approach.

To impose the SO04 outflow model described in § 2.4 we first use equation (17) to define the amount of gas heated above the critical entropy. This mass, which we denote \( M_{\text{ex}} \) but is equivalent to \( M_{b,\text{heat}} \), defines a reservoir of mass that we need to remove from our simulation catalog. We search to a radius \( R = \min(R_s, R_{\text{heat}}) \), where \( R_s \) is the radius of the thin shell as a function of time (eq. [10]) and \( R_{\text{heat}} \) (eq. [16]) is the maximum radius of the region heated above \( S_{\text{crit}} \). If a quasar is found within \( R \), we subtract \( M_{\text{ex}} \) from the gas mass associated with it, and if \( M_{\text{ex}} \) exceeds this mass, we remove it from our catalog altogether. In all cases, we assume \( S_{\text{crit}} \) is 60 keV cm\(^2\), corresponding to the metallicity value of \( Z = 0.05 \) used in our simulation, and we take an average postshock overdensity, \( \delta_s \), of 20.

The resulting luminosity function is plotted in Figure 3 (bottom panels). In this case, there is significantly better agreement between the data (and the semianalytic model) for this revised luminosity function. While this model does not precisely reproduce the strong knee observed in the SO04 model and the magnitude of the turndown at higher luminosities, the improved agreement is compelling. It is thus clear that the primary cause of the difference between these models is not their varying approaches to modeling quasars themselves, but rather the simplified exclusion conditions we derive through using \( M_{\text{ex}}, R_s, \) and \( R_{\text{heat}} \).

While the SO04 outflow model captures most of the salient features of shock heating, there are two major effects that it fails to address. First, it does not differentiate between material within galaxies and material in the intergalactic medium, while in our simulation we do not apply the outflow heating to the host galaxy itself. Second, the outflow model assigns a single density to all the gas associated with an object that is overaken by an outflow, whereas heating processes in the simulation are directly affected by the detailed substructure in this gas.

To explore the relative impact of these two effects, we examine the distribution of progenitor halo labels in the three largest outflows in our simulation at an epoch of \( z = 1.3 \). These systems have baryonic masses ranging from \( 4.7 \times 10^{12} \) to \( 5.1 \times 10^{12} M_\odot \), and the majority of particles within them have a single label associated with the previous outflow event. To determine the number of particles with dissimilar labels, we place a sphere at the center of mass of the group and then fit a radius (by hand) to enclose the outer boundary of the particles with the main group label. The typical radius of this outer boundary was 250 kpc. For all systems we found a significant amount of substructure, as evidenced by all groups having at least 18, and typically more than 40, distinct labels each with at least 20 particles. The total number of particles with labels distinct from the main group was always close to the merger mass limit, indicative of a system just about to undergo the outflow event associated with the merger.
This results seems to suggest that both effects may be contributing to the reduced suppression, as each outflow is associated with a large group with a single index (that might have been disrupted if in-galaxy heating was included), which merges with a collection of remnants with many indices (that might not have been accreted if in-shock heating were more efficient). Ultimately, the stronger agreement for the revised catalog versus that of the original simulation indicates that both of these issues need to be explored further. On the simulation side this would involve contrasting the results with a model that includes more efficient ejection of gas from galaxies themselves. In this regard, more targeted convergence studies may be necessary, as the level of mixing between subclumps and heated gas is notoriously difficult to capture in SPH, as emphasized in the early-type galaxy context by Naab et al. (2005). Likewise, in the semianalytic model, the efficiency of heating could be parameterized by adding an additional parameter to account for in-shock cooling, although it would be necessary to conduct detailed simulations of this process to precisely calibrate this number. These are significant issues, which we return to in §5.

Finally, as a test of convergence, in Figure 4 we compare the luminosity function in our fiducial simulation with that derived from our 35 $h^{-1}$ Mpc$^3$ simulations. In this case we do not attempt to impose the SO04 outflow model. At all redshifts we obtain good agreement between runs over the luminosity range spanned by these smaller simulation volumes, although the 0400 results are very noisy. Note that while in the $z$ = 2.25 column, the luminosity function in the 3200 run has fewer large quasars than the other runs, this is due to small-number statistics and the fact that it was stopped earlier than the other simulations due to the excessively large number of time steps required. Similarly, $L_B \gtrsim 10^{13} L_\odot$ quasars are so rare that they cannot be compared between the 1020 simulation and the test simulations, motivating our choice of an extremely large volume for this run.

3.4. Hard X-Ray AGN Luminosity Function

Our simulation can also be directly compared with observations of the hard X-ray luminosity function (HXLF), which we construct from the bolometric luminosities using a similar approach to that for the optical quasar luminosity function. To convert from the bolometric luminosity to $L_X$ we use results from Marconi et al. (2004), who calculated the expected correction for $\approx 10^{12} L_\odot$ AGNs with a spectral template motivated by recent observations. The ratio of the bolometric luminosity to the hard X-ray band (2–10 keV) is given by a third-degree polynomial,

$$\log \left[ \frac{L}{L(L_{2-10\text{ keV}})} \right] = 1.54 + 0.24 \mathcal{L} + 0.012 \mathcal{L}^2 + 0.0015 \mathcal{L}^3,$$

where $\mathcal{L} = \log (L_{\text{bol}}) - 12$ and $L_{\text{bol}}$ is given in $L_\odot$.

Early observational work used Advanced Satellite for Cosmology and Astrophysics (ASCA) data (Boyle et al. 1998) and BeppoSAX data (La Franca et al. 2002) to show that the HXLF was evolving strongly between $z = 0$ and 1.5, consistent with pure luminosity evolution. More recently, Cowie et al. (2003) used Chandra data in two redshift bins ($z = 0.1–1$ and 2–4) to argue that the AGN number density for luminosities lower than $10^{44}$ ergs s$^{-1}$ seems to peak at a lower redshift than those of higher luminosity. This antihierarchical evolution was demonstrated definitively by Ueda et al. (2003), who carried out a comprehensive compilation of HEAO-1 (Piccinotti et al. 1982; Grossan 1992), ASCA (Ueda et al. 2001; Akiyama et al. 2003), and Chandra (Brandt et al. 2001) data to derive the comoving spatial densities of AGNs in three luminosity ranges between $\log (L_X) = 41.5$ and 48.

In Figure 5 we compare our derived spatial densities to the observational data in the $\log (L_X) = 43–44.5$ and 44.5–48 luminosity bands in this sample. Both the qualitative and quantitative predictions of the simulation agree with the measurements: the downsizing trend is apparent in both luminosity bands, and the overall normalizations agree well. However, below $z \simeq 2$, we are faced with two minor issues. First, in the lower luminosity bin it appears that the luminosity function is turning down slightly more quickly, as the observations suggest that the turnardown in this luminosity range occurs after $z = 1$. Second, the brightest bin, while turning down at the observed epoch of $z = 2$, does not perfectly follow the observational trend. Imposing the SO04 exclusion conditions improves this fit somewhat, although these differences are small compared to the measurement errors from the observations. Imposing these conditions has no impact on the lower luminosity bin. In general, these results are consistent with those in §3.3.

3.5. Wind Properties

Finally, we consider the question of the observable properties of the outflows in our simulation. The postshock temperatures of these outflows are several keV, and thus one might expect them to only be observable through X-ray absorption line studies. However, experience with starburst-driven winds leads us to expect...
that some fraction of this material will be in lower ionization states associated with much colder clumps of gas. Indeed, one of the most useful lines in studying $\approx 200$ km s$^{-1}$ starburst winds has proven to be Na$\lambda$, with an ionization potential of only 5.1 eV. (e.g., Heckman et al. 2000; Schwartz & Martin 2004). Thus, the large outflows in our simulation are likely to be similarly observable through absorption lines seen in optical spectroscopy.

With this in mind, we constructed histograms of the initial velocities of these outflows. As the properties of the host galaxies are not directly calculable from our simulation itself we approximate these using the observed black hole to total bulge mass ratio of $\approx 500$ (Marconi & Hunt 2003). Although the winds are put in instantaneously in our simulation, observationally they should be thought of as surrounding galaxies with an age equal to $r_{\text{sh}}/v_{\text{th}}$, the initial shell radius over the initial velocity. For the majority of galaxies in our simulation this gives an age $\approx 30$ Myr, and thus these are winds associated with poststarburst early-type (E+A) galaxies, rather than ongoing mergers. With these properties in mind, we show the distribution of velocities of outflows associated with poststarburst bulges with total masses $> 10^{11.2} M_\odot$ in the top panel of Figure 6 and outflows associated with $> 10^{12} M_\odot$ poststarburst bulges in the bottom panel.

In both cases, the outflow velocities are $\approx 1000$ km s$^{-1}$ with a significant number of cases reaching or even exceeding $2000$ km s$^{-1}$. Furthermore, a simple calculation shows that global winds with such high velocities are extremely difficult to attain in starburst-driven winds. To see this, we can consider a population of stars in which supernovae drive a mass in gas equal to the total mass in stars. Assuming a Salpeter initial mass function then gives a maximum velocity of $[2 \times 10^{51} \text{ergs/(300 } M_\odot)]^{1/2} = 580$ km s$^{-1}$ even if the full kinetic energy of each supernovae is applied to the ejecting material. Thus, winds with velocities above $1000$ km s$^{-1}$ would require not only negligible radiative losses, but extremely efficient star formation, such that the ratio of gas to stars was $\leq 25\%$. This means that the detection of $\approx 1000$ km s$^{-1}$ outflows from poststarburst elliptical galaxies would provide extremely strong evidence for the picture proposed here.

4. OTHER IMPLICATIONS

4.1. Impact on the Intergalactic Medium

While our study has been focused on the properties of the quasar population itself, our simulations naturally have predictions for the more tenuous gas surrounding large galaxies. In fact, as discussed above, the most clear observational evidence for widespread nongravitational heating lies not in the galaxy population, but rather in the properties of the diffuse gas in galaxy clusters.

To examine the impact of outflows on this material, we first focus on the total amount of gas that has been shocked to $S > S_{\text{crit}}$, such that it no longer participates in fueling further generations of quasars. The redshift evolution of the mass fraction of this gas is plotted in Figure 7 (top), which shows that quasar feedback is primarily a low-redshift phenomenon. Thus, above $z \approx 3$ less than $3\%$ of the gas in the simulation has been affected, consistent with the lack of suppression of the luminosity function at these redshifts. At lower redshifts, however, the $S > S_{\text{crit}}$ mass fraction grows prodigiously, preventing roughly $20\%$ of the gas in the simulation from cooling. This is consistent with the turn-down in the luminosity function seen in Figure 3, which, while not as efficiently quenched as the semianalytic results, nevertheless differs substantially from the pure merger predictions.

As a test of convergence, we also plot in this panel the $S > S_{\text{crit}}$ mass fraction from each of our smaller simulations. These range from the 0400 run, in which particles are 64 times more massive than in the 1020 simulation, to the 3200 run, in which particles are 0.125 times the mass of those in the 1020 run. As increasing resolution adds a large number of low-mass, high-redshift outflows, the mass fractions at high redshift increase monotonically with resolution. At lower redshift, however, the mass fractions approach each other asymptotically, and in the important $z \leq 3$ range, this quantity is largely constant across
runs. However, this mass convergence does not give a complete picture of the effect of resolution.

In Figure 7 (bottom) we plot the evolution of the volume filling factor of $S \geq S_{\text{crit}}$ gas in each of these runs. To calculate these quantities in the SPH method it is necessary to first smooth the particle data onto a grid. In the 1020 case we do so on a 1340$^3$ mesh, so that the smoothing scale for the filling factor at expansion factor $a$ is $0.155a$ Mpc, which is considerably above our minimum smoothing length. In the other runs we use a mesh that is twice the size of the particle resolution, for example, the $2 \times 160^3$ run was smoothed on to a $320^3$ mesh.

The data point in this panel gives the upper limit of the volume filling factor provided by the Ly$\alpha$ forest (Penton et al. 2004; Levine & Gnedin 2005), which is well above our results, as expected from the semianalytic estimates in SO04. For comparison, we also plot the results from a range of models taken from the N-body + semianalytic study of Levine & Gnedin (2005). While this is substantially higher than our results, this is to some degree due to the fact that they computed the full volume impacted by quasar outflows, rather than only the volume heated above $S_{\text{crit}}$.

Furthermore, it is clear from this figure that the volume filling factors are significantly different across simulations, even at the lowest redshifts. This result seems at odds with our mass fraction measurements. To explore this issue further, in Figure 8 we choose a fixed redshift of $z = 3$ and plot the mass fraction and volume filling fraction above a threshold entropy $S_{\text{thr}}$, which we allow to vary. For all three models, the mass fraction is only a weak function of $S_{\text{thr}}$ for all entropy values near $S_{\text{crit}}$. This means that small differences in entropy have only a small impact on the number of particles prevented from cooling, and therefore both the $S > S_{\text{crit}}$ mass fraction (shown in Fig. 7) and the luminosity function (shown in Fig. 4) are similar across runs. Essentially, at $z = 3$, the particles are divided into two types, those whose entropies are well above $S_{\text{crit}}$ and those that are far below this critical value.

In the middle panel of Figure 8, we plot the volume filling factor as a function of $S_{\text{thr}}$, again at $z = 3$. In this case, near $S_{\text{crit}}$ the volume filling factor is a strong function of our threshold entropy. This suggests that the high-entropy gas is largely found in low-density environments, so that changes in $S$ around $S_{\text{crit}}$ pass through a region where the volume can change rapidly, but there is actually little mass to modify the overall mass fraction. In the bottom panel of Figure 8 we plot the differential over-density $\Delta M/\Delta V$, that is, the change in the mass fraction over the change in the volume filling factor, as a function of $S$. This confirms that the majority of $S \approx S_{\text{crit}}$ gas is in environments only a few times denser than the mean and that the density of this material is increasing strongly as a function of entropy.

Figure 9 illustrates why this is the case. Here we show the entropy distribution in slices taken from our 1020 simulation at the final output redshift. It is clear from this plot that the boundary between $S > S_{\text{crit}}$ and subcritical gas (indicated by the white lines) lies at the very edges of cosmological halos, where the density is declining precipitously. These boundaries are defined by comparatively few SPH particles, and thus their positions can change rapidly following small changes in the particle distribution. Finally, the thin slice shown in the smallest scale panel in Figure 9 uncovers the presence of $S \leq S_{\text{crit}}$ subclusters within a larger (high-entropy) region. Again, the presence of this cooling substructure was not included in the SO04 models and is one of
the key causes of the differences between their semianalytical results and those presented here.

4.2. Impact on Clusters and Groups

Recent Chandra observations of Hydra A (Nulsen et al. 2005) have uncovered evidence for AGN activity within clusters at intermediate redshifts. While cooling-flow clusters at lower redshift seem inconsistent with the idea of powerful shocks (e.g., Voit & Donahue 2005; Croton et al. 2006), it remains an intriguing possibility that these systems underwent an earlier period of strong feedback and have now settled into a quiescent state. The observations of radio-quiet clusters by Donahue et al. (2005) that show extremely long cooling times despite an absence of inferred black hole activity are broadly consistent with this hypothesis.
These results prompted the numerical investigations of Sijacki & Springel (2006), who showed that their model of AGN activity has comparatively little effect on the cluster $L_X$-$T$ relationship when all material within the virial radius is included. Therefore in this section we examine the effect of our outflow model on the $L_X$-$T$ relationship and the cluster entropy profile. It is worth noting that as our raw luminosity function shows we have an excess of bright quasars at $z = 1.25$, the results we derive in this section are an upper bound on the effect of outflows on clusters.

To find cluster groups we first begin from a FOF $b = 0.2$ catalog of the 1020 simulation. The centers of mass are evaluated for these groups and then used as the beginning stage of an iterative spherical overdensity group finder that searches radially outward until the group is below the density threshold. The center of mass is then evaluated and used as the beginning point for the radial search, with this process being repeated five times. This technique has the advantage of biasing against mergers, since the center of mass of mergers is usually sufficiently offset from the merging groups to stop the spherical-overdensity convergence process early, and the group is then discarded due to the low amount of mass found. We find 1272 groups by this process.

To estimate the bolometric luminosity of the simulated clusters we use

$$L_{\text{bol}} = \sum_{i=1}^{N_{\text{clus}}} \frac{m_i \rho_i}{(\mu m_p)^2} \Lambda(T_i),$$

and the emission-weighted temperature is given by

$$T_{\text{ew}} = \frac{\sum m_i \rho_i \Lambda(T_i) T_i}{\sum m_i \rho_i \Lambda(T_i)}.$$  

In these formulae, $\Lambda(T_i)$ is the pure bremsstrahlung estimator of Pearce et al. (2000; see also Navarro et al. 1995; Muanwong et al. 2001, for similar approaches), and $m_i$, $\rho_i$, and $T_i$ are the mass, density, and temperature of particle $i$. While we could use the emission curve associated with the $Z = 0.05$ metallicity gas used in the simulation, the strong peak caused by collisional excitation of He$^+$ at $10^5$ K will give a very high weighting to the cores of clusters, which in fact would probably have cooled significantly if we were using a $Z = 0.3$ metallicity gas. Overall, as emphasized in Muanwong et al. (2001), our choice of a pure bremsstrahlung estimator will bias our luminosities low.

Under the assumption of self-similar evolution for spherically symmetric clusters, and also that bremsstrahlung emission dominates in the X-ray band, the luminosity-temperature relationship can be scaled by the redshift. Thus, the redshift evolution in the $L_X$-$T$ relationship can be scaled by dividing the luminosity by $E(z) \Delta(z)/\Delta(0)^{1/2}$. While we could thus scale $L_X$ at $z = 0$ clusters back to our final redshift, a recent analysis by Maughan et al. (2006) of the Wide Angle ROSAT Pointed Survey (WARPS) clusters in the region $0.6 < z < 1.0$, provides an unscaled $L_X$-$T$ relationship at $z \approx 1$. Thus, in what follows we use their results as a comparison. We also note that there are no observations of galaxy groups at these epochs.

We plot our results for the $L_X$-$T$ relationship in Figure 10 and also show the fit of Maughan et al. (2006). It is immediately noticeable that our raw data show a very large scatter in luminosities. Since the luminosity of clusters is linearly weighted by the density, we decided to plot radial profiles of our clusters to determine whether any “overcooling” effects (e.g., Thacker et al. 2000) might be present (see Fig. 12). The radial profile shows that the most massive clusters do indeed show a sudden upturn in density in their cores along the lines observed in test problems. We therefore applied a density cut (filter) at $\delta < 5 \times 10^{-4}$ to the gas in our clusters, which cuts out this problem region. The resulting
been used to smooth the data. A least-squares fit for our 1 keV and brighter unfiltered profile reported elsewhere (e.g., Voit et al. 2005). The results are strongly supportive of the conclusions of Sijacki & Springel (2006) best fit of $5^{+1}_{-0.82}/C18/C19$.

We have presented results from a suite of cosmological simulations that self-consistently follow the evolution of quasars and the outflows associated with them. By tracking the merger history of halos and applying the quasar model of Wyithe & Loeb (2003), we have been able to make direct predictions for the spatial distribution and luminosity function of these objects. Our results are in excellent agreement with the observed correlation function of quasars on both small and large scales, reproducing both the power-law behavior measured by the 2dF redshift survey on $\geq 1$ Mpc scales and the strong break measured from the SDSS on $\leq 1$ Mpc scales.

Furthermore, we predict that the quasar-galaxy cross-correlation function should show a small-scale upturn relative to the galaxy-galaxy correlation function. This occurs on sub-Mpc scales and therefore can only be measured from a very large sample of quasars, making its detection difficult even in the large DEEP2 survey. However, it is worth noting that the Dark Energy Survey will produce a quasar catalog spanning 5000 deg$^2$, which, combined with photometric redshifts for 300 million galaxies and spectroscopic follow-up as needs be, should be able to suppress statistical uncertainties to a sufficiently low level to measure this upturn. We also note that this clustering excess is in qualitative agreement with the increase in the integrated galaxy overdensity at small scales for the SDSS quasar sample (Serber et al. 2006). Although the origin of the excess is uncertain, there is an interesting possibility that the presence of an ancillary galaxy could accelerate the merger process via a three-body interaction. Given the extended nature of the mass distribution associated with galaxies, it is not clear whether this mechanism will work in a similar way to stellar interactions. We plan to investigate this intriguing idea in the near future.

As the correlation function is dominated by more populous low-luminosity quasars, our spatial results should be largely interpreted as lending support to our dark matter modeling choice of merger model. Similarly, as it tracks the number of quasars formed as a function of the gas mass accreted onto galaxies, the luminosity function is more sensitive to the details of our feedback modeling. In this case, our simulation qualitatively reproduces the observed antihierarchical behavior, but the turnover is much weaker than observed. Matching the suppression at the brighter end requires that we increase the efficiency of heating from the AGN outflows to mimic that assumed in semianalytic models. This is equivalent to suppressing one key physical process, namely in-shock cooling in the presence of substructure, and including the ejection of gas from quasar host galaxies. These results emphasize how sensitive the luminosity function is to issues in baryon physics and how the treatment of these issues in semianalytic models is still quite approximate. Plotting the velocities in our simulation uncovers a large sample of $\approx 1000$ km s$^{-1}$ outflows around poststarburst bulges. As these velocities are almost impossible to attain in starbursts, detection of such high-velocity global winds around large E+A galaxies would provide extremely strong evidence for the picture proposed here.

Investigation of the mass fractions and filling factors of gas above $S_{\text{crit}}$ showed good convergence in the mass fractions at low redshift, but less so in the filling factors. The differences between runs are most noticeable at high redshift, where successively higher resolution leads to the first generation of AGN outflows occurring at earlier epochs. While mass fractions show fairly strong convergence below $z \approx 5$, the tendency of gas above $S_{\text{crit}}$ to occupy low-overdensity regions makes an accurate calculation of the volume filling factor difficult due to sampling issues.

![Figure 12. Density, temperature, and entropy profiles for the four largest relaxed clusters at $z = 1.2$ in physical Mpc. A 100 particle moving average has been used to smooth the data.](image-url)

Data are plotted in Figure 10 (right), and show a much smaller scatter. A least-squares fit for our 1 keV and brighter unfiltered cluster catalog gives

$$L_{\text{bol}} = 2.82 \times 10^{42} \left( \frac{T_{\text{ew}}}{\text{keV}} \right)^{3.58},$$

while the filtered catalog gives,

$$L_{\text{bol}} = 2.09 \times 10^{42} \left( \frac{T_{\text{ew}}}{\text{keV}} \right)^{3.18},$$

which is closer to, albeit slightly less luminous than, the Maughan et al. (2006) best fit of $5.4 \times 10^{42}(T_{\text{ew}}/\text{keV})^{2.92}$. Overall, these results are strongly supportive of the conclusions of Sijacki & Springel (2006).

Lower redshift observations of groups (e.g., Xue & Wu 2000) show that the power-law exponent for groups is close to 5. For systems below 1 keV (regardless of whether or not they are core filtered) we do not observe any steepening of the systems below 1 keV (regardless of whether or not they are core filtered) indicating that the gas in these systems is quite strongly perturbed. Thus, rather than analyzing relaxed systems, we are in fact analyzing groups that are radiating significantly due to the presence of an outgoing shock. In the event that this shock is sufficient to heat gas above $S_{\text{crit}}$, we might well expect, in the absence of significant further accretion and AGN activity, that these groups expand and cool below $z = 0$. To quantify this hypothesis, we plot the cluster and core entropy versus temperature in Figure 11. Our results are in broad agreement with the analysis of nearby clusters (Ponnan et al. 2003; Finoguenov et al. 2005) and do indeed show that groups tend to show an excess of entropy in their cores. Therefore, we tentatively suggest that the X-ray emission of galaxy groups at $z > 1$ may well be a “smoking gun” for the AGN-heating hypothesis.

In addition to the radial density profiles, we also show the temperature and entropy profiles in Figure 12. The only unusual feature in the temperature profiles is that the second cluster from the left has a very slightly inverted temperature profile following an extremely strong outflow event. The resulting entropy profile is $S \propto r^{1.7}$, while the remaining profiles all match the $S \propto r^{1.1}$ profile reported elsewhere (e.g., Voit et al. 2005).
These results are also further complicated by the known dependence of shock resolution on particle number (e.g., Thacker et al. 2000) by which accurate modeling of shock jumps in spherical collapse is reached once $N_{\text{coll}} > 30,000$. Nonetheless, the results do elucidate that the impact of quasar outflows is largely felt at low redshifts, in agreement with the downsizing trend. Perhaps the most interesting issue we have not yet explored with regards to filling factors is the relative impact of including a more bipolar outflow. However, there is no reason to expect a difference beyond a factor of 2 as the gas will flow toward low-overdensity regions, as observed in supernova outflow calculations (STD01).

In somewhat denser environments, our results for the cluster $L_{\times}-T$ relation are in broad agreement with observations. This is especially encouraging, since our cluster modeling is not as sophisticated as other more targeted studies. In this case, the predominant view is that AGN heating is best modeled in terms of “hot bubble” ejection from the brightest cluster galaxy (BCG), which then mixes with the surrounding material. Interestingly, the exact nature of this mixing is not well understood, and substantial differences are found between Eulerian and Lagrangian simulations, which by definition differ significantly in their treatment of advection. While SPH simulations inhibit the development of instabilities that promote mixing due to the necessary use of flow interpenetration suppression, it has the advantage of exhibiting significantly less numerical diffusion than many Eulerian methods. Ultimately, input from laboratory experiments may well be necessary to help determine the correct mixing behavior. We also note that due to our inefficient star formation model, our simulations include a significant amount of cold gas in BCGs that will tend to promote an “overcooling” instability, which others have avoided by applying phase decoupling (Pearce et al. 2000). We believe that this is a significant contributor to the lack of an entropy core in our radial profiles, and removing this central core produces an $L_{\times}-T$ relationship that is significantly less noisy, while not exhibiting a large change in normalization at intermediate cluster mass scales. In addition, the lack of a strong turnaround in the quasar luminosity function at $z = 1.2$ also promotes extremely strong outflows in these clusters that may well be transporting high-entropy gas from the core out to the edges of the cluster without significantly raising the entropy of the gas immediately surrounding the BCG.

Perhaps the most intriguing result to come out of our $L_{\times}-T$ study is the prediction that the gas in $z \approx 1$ galaxy groups should be strongly perturbed by AGN activity, which is in the process of turning off at that epoch. This effect is confined to small, high-redshift clusters for two main reasons. At lower redshifts, the comparative paucity of AGN activity will allow group gas to evolve largely adiabatically, decreasing $L_{\times}$ dramatically to establish the steep $L_{\times}-T$ relation ($L_{\times} \propto T^{3}$) observed locally. In more massive $z \approx 1$ clusters, on the other hand, the effect of outflows is largely masked when averaging over the material within $r_{200}$. Thus, galaxy groups at $z \approx 1$ represent the key mass and redshift range at which the AGN-heating hypothesis is most likely to be testable through observations.

Finally, our simulation results indicate that simplifications in current semianalytic models may well be downplaying critical physics. While much attention has been paid to drawing-broad conclusions from comparisons between observations and these models, the differences we have uncovered in this investigation are troubling. While, as expected, postprocessing of our simulation results was able to reproduce the semianalytic behavior, it is clear that in-shock cooling due to substructure in low-overdensity halos is an issue that must be considered carefully in semianalytic calculations. Fortunately, constraining the effect by simulation would not be difficult, and ultimately a single parameter could be used to quantify this behavior. On the simulation side, it is also clear that a more detailed understanding of gas ejection from quasar hosts will be necessary for definitive conclusions to be reached. Nonetheless, the insight gained from this initial simulation study clearly serves to highlight the promise of a self-regulating picture of quasar formation. The simple merger model of Wyithe & Loeb (2003), when supplemented with an outflow model, appears to represent a significant step forward in understanding the evolution of quasar clustering and the cause of their antihierarchical low-redshift turnover.

We thank Andrea Ferrara, Piero Madau, and Mattias Steinmetz for hosting the IGM-Galaxy Interactions workshop at the KITP where part of this work was undertaken. We are grateful to Alison Coil, Darren Croton, Marc Davis, Joseph Hennawi, Lars Bildsten, J. P. Ostriker, & Christy Tremonti for their many useful comments, particularly in regards to our comparisons with observations. We also wish to thank the anonymous referee for his/her careful reading of our paper, which resulted in several significant improvements. Computing was performed at WestGrid (under a RAC grant), SHARCNET, and HPCVL. We thank Rob Simmonds of WestGrid for helping arrange a period of dedicated use of their IBM p595 system during initial testing. R. J. T. acknowledges funding from the CITA National Fellow program and NSERC of Canada via the operating grants of Professors Larry Widrow and Richard Henriksen. E. S. was supported by the National Science Foundation under grants PHY 99-07949 and AST 02-05956.

REFERENCES

Akiyama, M., Ueda, Y., Ohhta, K., Takahashi, T., & Yamada, T. 2003, ApJS, 148, 275
Allen, S. W., Schmidt, R. W., & Fabian, A. C. 2002, MNRAS, 334, L11
Arnaud, M., & Evrard, E. 1999, MNRAS, 305, 631
Babul, A., Balogh, M. L., Lewis, G. F., & Poole, G. B. 2002, MNRAS, 330, 329
Belgianum, M. C., & Cioffi, D. F. 1989, ApJ, 345, L21
Binney, J. 2004, MNRAS, 347, 1093
Boyle, B. J., Shanks, T., Croom, S. M., Smith, R. J., Miller, L., Loaring, N., & Heymans, C. 2000, MNRAS, 317, 1014
Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, MNRAS, 235, 935
Boyle, B. J., et al. 1998, MNRAS, 296, 1
Brandt, W. N., et al. 2001, AJ, 122, 2810
Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80
Bullock, J. S., Wechsler, R. H., & Somerville, R. S. 2002, MNRAS, 329, 246
Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, ApJ, 625, 621
Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaziot, J. 2006, MNRAS, 370, 1651
Cavaliere, A., Menci, N., & Tozzi, P. 1998, ApJ, 501, 493
Ciotti, L., & Ostriker, J. P. 2001, ApJ, 551, 131
Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., & Davis, M. 2006, ApJ, 644, 671
Couchman, H. M. P. 1991, ApJ, 368, L23
Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
Cowie, L. L., et al. 2003, ApJ, 584, L57
Croom, S. M., & Shanks, T. 1999, MNRAS, 303, 411
Croom, S. M., Shanks, T., Boyle, B. J., Smith, R. J., Miller, L., Hoyle, R. L., & Heyle, F. 2001, MNRAS, 325, 483
Croom, S. M., et al. 2004, in ASP Conf. Ser. 311, AGN Physics with the Sloan Digital Sky Survey, ed. G. T. Richards & P. B. Hall (San Francisco: ASP), 457
Croxton, D., et al. 2006, MNRAS, 365, 11
Davé, R., Hernquist, L., Katz, N., & Weinberg, D. H. 1999, ApJ, 511, 521
David, L. P., Slyz, A., Jones, C., Forman, W., Vrtilek, S. D., & Arnaud, K. A. 1993, ApJ, 412, 479

No. 1, 2006 QUASARS: WHAT TURNS THEM OFF? 99
