QUASAR LUMINOSITY FUNCTIONS FROM JOINT EVOLUTION
OF BLACK HOLES AND HOST GALAXIES

A. LAPI\textsuperscript{1}, F. SHANKAR\textsuperscript{2}, J. MAO\textsuperscript{3}, G.L. GRANATO\textsuperscript{2,1}, L. SILVA\textsuperscript{4}, G. DE ZOTTI\textsuperscript{1,1}, L. DANENE\textsuperscript{1}

Draft version April 3, 2008

ABSTRACT

We show that our anti-hierarchical baryon collapse scenario (Granato et al. 2004) for the joint evolution of black holes and host galaxies predicts quasar luminosity functions at redshifts $1.5 \lesssim z \lesssim 6$ and local demographic properties in nice agreement with observations. In our model the quasar activity marks and originates the transition between an earlier phase of violent and heavily dust-enshrouded starburst activity promoting rapid black hole growth, and a later phase of almost passive evolution; the former is traced by the submillimeter-selected sources, while the latter accounts for the high number density of massive galaxies at substantial redshifts $z \gtrsim 1.5$, the population of Extremely Red Objects, and the properties of local ellipticals.

1. INTRODUCTION

Since at least a couple of decades observations in the optical and in the X-ray bands have shown that very bright quasars (QSOs) occur very soon in the history of the Universe, and that their average luminosity significantly declines from redshift $z \approx 3$ (e.g., Mathez 1978; Giacconi 1985; Schmidt & Green 1986).

Modern observations confirmed and detailed such a picture (e.g., Fan et al. 2004, 2006; Richards et al. 2005, 2006; Cristiani et al. 2004; Barger et al. 2005; Tozzi et al. 2001; Brandt & Hasinger 2005), measuring the luminosity functions (LFs) up to redshift $z \approx 6$.

As a matter of fact, supermassive black holes (BHs) were found to be ubiquitous in the centers of spheroidal galaxies (Kormendy & Richstone 1995; Magorrian et al. 1998). Moreover, a narrow relationship between the central supermassive BH mass and the stellar mass/luminosity of the spheroidal/bulge component has been firmly established by observations (e.g., Gebhardt et al. 2000; Marconi & Hunt 2003). The BH mass also strictly correlates with the velocity dispersion of the spheroidal component (e.g., Ferrarese & Merritt 2000; Tremaine et al. 2002). Contrariwise, it has been shown that the BH mass is on average at least 10 times smaller in late-type irregular and/or spiral galaxies than in ellipticals with the same luminosity (see Salucci et al. 2000). All together these observations witness the strict connection between the mass of the BH and that of the old stars formed at $z \gtrsim 1$ within the host galaxy. Recently, the studies of QSO host galaxies have been extended to high redshift (e.g., Dunlop et al. 2003; Floyd et al. 2004), with the general results that they are typically early-types.

In fact, the stellar populations of ellipticals are old and essentially coeval (Sandage & Visvanathan 1978; Bernardi et al. 1998; Trager et al. 2000; Terlevich & Forbes 2002; Thomas et al. 2005). A color-magnitude relation is also well established, indicating that brighter spheroids are redder (Bower et al. 1992). The widely accepted interpretation is that brighter objects are richer in metals and that the spread of their star formation epochs is small enough to avoid smearing of their colors; in fact, the slope of the relation does not change with redshift (Ellis et al. 1997; Kodama et al. 1998), supporting this interpretation. The star formation history of spheroidal galaxies is mirrored in the Fundamental Plane (e.g., Jedrzejewski et al. 1987; Dressler et al. 1987) and in its evolution with redshift; ellipticals adhere to this plane with a surprisingly low orthogonal scatter (around 15%), as expected for a homogeneous family.

Furthermore, recent studies (e.g., Treu et al. 2002; van der Wel et al. 2004; Holden et al. 2005) suggest that ellipticals, both in the field and in the clusters, lay on the fundamental plane up to $z \approx 1$, consistent with the hypothesis that massive spheroids are old and quiescent. The progenitors of local massive early-type galaxies have been identified through the $K$-band and Spitzer surveys at substantial redshift $z \gtrsim 1$. Direct evidence that massive galaxies with stellar content $M_* \gtrsim 10^{11} M_\odot$ were in place at $z \gtrsim 2$ is provided by recent $K$-band surveys (e.g., Cimatti et al. 2002; Kashikawa et al. 2003; Fontana et al. 2004). The space density of Extremely Red Objects (EROs) at $z \gtrsim 3$ is only a factor about 5–10 less than that at $z \approx 1$ (e.g., Tecza et al. 2004). The implied phase of extremely high star formation rate (SFR), from several hundreds to thousands solar masses per year, is also witnessed by the submm galaxy counts (e.g., Chapman et al. 2003; 2005).

On the theoretical side, the hierarchical clustering paradigm led to the development of various semi-analytic models for galaxy formation; these share the basic assumption that the main driver in shaping the structure and morphology of galaxies is gravity. In this scenario, the gas cools and form stars following the collapse of dark matter (DM) halos. In the standard cosmology, small DM objects form first and merge together to make larger ones. This scenario then implies that large ellipticals form late, by the merger of disk/bulge systems made primarily of stars, recently indicated as dry mergers (e.g., Naab et al. 2006).

But the early appearance of QSOs with strong luminosity and presumably huge BH mass was rather at variance with respect to this framework (e.g., Bromley et al. 2004). This apparent contradiction led to discuss a higher efficiency in forming massive BHs.
in smaller galaxy halos at higher redshift (e.g., Haehnelt & Rees 1993). A number of theoretical studies of QSO LFs followed this suggestion (e.g., Haiman & Loeb 1998; Wiythe & Loeb 2003; Mahmoud et al. 2005).

Granato et al. (2001, 2004), instead, explored the possibility of reconciling the observed “downsizing” of QSOs and spheroidal galaxies with the bottom-up DM hierarchy. This has been possible by developing a simple model that incorporates the main physical aspects of the DM and baryons residing within galactic halos. This new approach emphasized the role of energy feedbacks both from supernovae (SNae), capable of unbinding the gas in low-mass systems, and from the QSO phase of the supermassive BH, capable of ejecting gas from the largest objects (see also Lapi et al. 2005). These feedbacks can actually reverse the formation sequence of visible galaxies with respect to that of DM halos; hence large galaxies end their star formation, and their BHs shine as QSOs, for first. On the contrary, the star formation and QSO phase is more prolonged in the smaller halos (hence the name anti-hierarchical baryon collapse, or ABC, scenario).

While in previous works we focused our efforts in order to reproduce the properties of spheroidal galaxies, in this paper we explore the constraints to the physical parameters of the model imposed by the LFs of high redshift QSOs. Our plan is as follows: in § 2 we briefly recall the main features of the model by Granato et al. (2004); in § 3 we describe the procedure adopted to compute the QSO LFs, the supermassive BH mass function, and other galactic observables; in § 4 we make a critical comparison of our findings with the other models in the existing literature; finally, in § 6 we summarize our conclusions.

Throughout this work we adopt the cosmology indicated by the WMAP data (Bennett et al. 2003; Spergel et al. 2006), i.e., a flat Universe with matter density \( \Omega_M \approx 0.27 \), baryon density \( \Omega_b \approx 0.044 \) and Hubble constant \( H_0 \approx 71 \) km s\(^{-1}\) Mpc\(^{-1}\).

2. THE MODEL FROM THE GROUND UP

This paper is based on the semi-analytic model developed by Granato et al. (2004), which follows the evolution of baryons within proto-spheroids through simple and physically grounded recipes.

We defer the interested reader to that paper for a full account of the physical justifications and the detailed description of the model; here we provide a short summary of its main features, focusing on the aspects relevant to our discussion on QSO LFs and BH demographics.

2.1. The DM sector

As for the treatment of the DM in galaxies, the model basically follows the standard framework of hierarchical clustering, taking also into account the results by Wechsler et al. (2002), and Zhao et al. (2003a; 2003b). Their simulations have shown that the growth of a halo occur in two different phases: a first regime of fast accretion in which the potential well is built up by the sudden mergers of many clumps with comparable masses; and a second regime of slow accretion in which mass is added in the outskirts of the halo, without affecting the central region where the galactic structure resides.

This means that the halos harboring a massive elliptical galaxy once created, even at high redshift, are rarely destroyed; meanwhile, at low redshift they are incorporated within groups and clusters of galaxies. Support to this view comes from studies of the mass structure of elliptical galaxies, which are found not to show strong signs of evolution since redshift \( z \approx 1 \) (Koopmans et al. 2006). Note that, as pointed out in § 1, the BH mass is strictly correlated with properties (mass and velocity dispersion) of the old stars in massive early-type galaxies, formed at least 8 Gyr ago (see Thomas et al. 2005) and, as a consequence, in massive galaxy halos virialized at \( z \gtrsim 1.5 \).

At redshifts \( z \gtrsim 1.5 \), relevant to the aim of this paper, a good approximation of the halo formation rates is provided by the positive term in the cosmic time derivative of the cosmological mass function (e.g., Haehnelt & Rees 1993; Sasaki 1994). For DM halos with mass \( M_{\text{vir}} \) at time \( t_{\text{vir}} \), these formation rates are given by

\[
\frac{d^2 N_{\text{ST}}}{dt_{\text{vir}} dM_{\text{vir}}} = \frac{a d\delta_c(t_{\text{vir}})}{\sigma^2(M_{\text{vir}})} + \frac{2p}{\sigma^2(M_{\text{vir}}) + a^p\sigma^2_p(M_{\text{vir}})} \left[ \frac{d\delta_c(t_{\text{vir}})}{dt_{\text{vir}}} \right] \left[ \frac{d\delta_c(t_{\text{vir}})}{dt_{\text{vir}}} \right] N_{\text{ST}}(M_{\text{vir}}, t_{\text{vir}})
\]

where \( N_{\text{ST}}(M_{\text{vir}}, t) \) is the Sheth & Tormen (1999, 2002) version of the PS mass function (Press & Schechter 1974). In the above equation, \( a = 0.707 \) and \( p = 0.3 \) are constants obtained from comparison of the mass function with the outcome of \( N \)-body simulations; in addition, \( \sigma(M_{\text{vir}}) \) is the mass variance of the primordial perturbation field, computed from the Bardeen et al. (1986) power spectrum with correction for baryons (Sugiyama 1995), and normalized to \( \sigma_8 \approx 0.8 \) on a scale of \( 8 h^{-1} \) Mpc; finally, \( \delta_c(t_{\text{vir}}) \) is the critical threshold for collapse, extrapolated from flat perturbation theory. Note that our adoption of the Sheth & Tormen mass function to construct the rates is mandatory, since it is well-known that the canonical Press & Schechter theory strongly underpredicts the number of massive halos, particularly at the high redshifts relevant to the computation of QSO LFs.

As for the galaxy halo mass range, we set the lower limit to \( M_{\text{min}}^{\text{max}} = 2 \times 10^{11} M_\odot \), since we are interested to follow the history of bright QSOs and of their host galaxies. At the other end, weak lensing observations (e.g., Kochanek & White 2001; Kleinheinrich et al. 2004) and kinematical measurements (e.g., Kronawitter et al. 2000; Gerhard et al. 2001) suggest an upper limit in galaxy halo mass \( M_{\text{max}}^{\text{max}} = 2 \times 10^{13} M_\odot \). At the same mass, moreover, the probability of multiple occupation by galaxies significantly increases (Magliocchetti & Porciani 2003). In fact, the present day galaxy halo mass function, derived after subtracting from the total mass function the one of groups and clusters, exhibits an exponential decline beyond \( M \approx 1.1 \times 10^{13} M_\odot \) (Shankar et al. 2006). The velocity dispersion function of early-type galaxies also requires \( M_{\text{max}} \approx 2 \times 10^{13} M_\odot \) (Cirasuolo et al. 2005).

In view of these results and on the basis of theoretical arguments, one should associate to massive halos a probability, redshift dependent, of hosting a single large galaxy, as opposed to hosting at least a couple of galaxies of comparable mass. However, as a first approximation, we choose to represent this effect with a fixed cut-off in the mass function. These limits in mass and redshift ensure that the positive cosmic time derivative of the mass function is a good approximation for the creation rates of DM halos, as the negative term is negligible.
Fig. 1.—Left panel: The time behavior of the stellar mass, infalling mass, cold gas mass, and star formation rate within halos of various masses, virialized at redshift 4. Right panel: The time behavior of the BH mass, reservoir mass, and BH accretion rates within halos of various masses, virialized at redshift 4.

2.2. The baryonic sector

The evolution of the baryons is much more articulated, due to their strongly collisional nature. The picture underlying our model is the following (see Granato et al. 2004, Cirasuolo et al. 2005 for additional details).

During or soon after the formation of the host DM halo, the baryons falling into the newly created potential well are shock-heated to the virial temperature; this hot gas is (moderately) clumpy and cools fast especially in the central denser regions, so triggering a strong burst of star formation. The radiation drag due to starlight acts on the gas clouds, further reducing their angular momentum; as a consequence, a fraction of the cool gas can fall into a reservoir around the central supermassive BH, and eventually accretes onto it by viscous dissipation, powering the nuclear activity. The energy fed back to the gas by SN explosions and BH activity regulates the ongoing star formation and BH growth; eventually, most of the gas is unbound from the DM potential well, so that the star formation and BH activity come to an (early) end.

In Appendix A and Table A1 we present the basic equations and parameters controlling the evolution of the baryonic component in our model, once the halo mass and the virialization redshift are given. These equations can be numerically integrated to yield, among others, the SFR and the accretion rate onto the central BH as function of cosmic time. In Fig. 1 we plot the basic outputs of the model.

Initially, the cooling is rapid and the star formation is very high; thus the radiation drag is efficient in accumulating mass into the reservoir. The BH starts growing from an initial seed with mass $10^2 M_\odot$ already in place at the galactic center; since there is plenty of material in this phase, the accretion is Eddington (or moderately super-Eddington) limited (e.g., Small & Blandford 1992; Blandford 2004). This regime goes on until the energy feedback from the BH is strong enough to unbind the gas from the potential well, a condition occurring around the peak of the accretion curve. After the peak the SFR drops substantially, the radiation drag becomes inefficient, the consequence storing in the reservoir and the accretion onto the BH decrease a lot. The drop is very pronounced for massive halos $M_{\text{vir}} \gtrsim 10^{12} M_\odot$, while for smaller masses a smoother declining phase can continue for several Gyrs, and the BH and stellar mass can still increase by an appreciable factor.
Before the peak, radiation from stars and from accretion on the supermassive BH is very obscured by the surrounding intra-galactic dust. In fact, these proto-galaxies are extremely faint in the UV-optical rest frame and appear as submm-selected sources. Also the nuclear accretion is heavily obscured; however, since the absorption significantly decreases with increasing X-ray energy of photons, it is easier to detect it in hard X-ray bands. On the other hand, in the proximity of the peak, i.e., when the central supermassive BH is massive and powerful enough to remove most of the gas and dust from the surroundings, the galaxy system will shine as an optical QSO.

The list of the observations well fitted by the model is presented in Table A2, along with a brief description of the main underlying assumptions. In order to translate the SFR and mass in stars into observable quantities such as broad-band luminosity including dust effects, we exploit the GRASIL code (Silva et al. 1998, see http://web.pd.astro.it/granato). As for the QSOs, the bolometric corrections are crucial to determine the expected luminosity in a fixed e.m. band. Finally, we convolve the results obtained for halos of fixed mass and virialization redshift with the rates of halo formation described in § 2.1.

### 3. THE MODEL AT WORK

As a first step we want to properly translate the DM formation rates given by Eq. (1) into BH formation rates; this calls for a relation between the halo mass $M_{\text{vir}}$ and the BH mass $M_\bullet$. In Fig. 2 (left) we plot the one obtained by solving the equations presented in Appendix A, that can be well approximated by the law

$$M_\bullet \approx 8 \times 10^6 \left( \frac{M_{\text{vir}}}{2.2 \times 10^{11} M_\odot} \right)^{3.97} \left( \frac{1 + z_{\text{vir}}}{7} \right) M_\odot.$$ (2)

This relationship is very close to that found by Shankar et al. (2006) by comparing the local BH mass function with the galaxy halo mass function. However, this is only an average relationship and we do expect scatter around it on the basis that the values of the parameters in the equations of Appendix A may naturally vary from halo to halo. In the following we assume that Eq. (2) holds on average with a gaussian dispersion $\Delta M_\bullet$ around the mean. Therefore we convert the halo formation rates into BH formation rates through the convolution

$$\frac{d^2 N_{\text{BH}}}{dt_{\text{vir}} dBH} = \int dM'_{\bullet} \left| \frac{dM_{\text{vir}}}{dM_{\bullet}} \right| M_{\bullet} \frac{d^2 N_\text{ST}}{dt_{\text{vir}} dBH |_{M_{\bullet}}} e^{-\left(M'_{\bullet} - M_{\bullet} \right)^2 / 2 \left(\Delta M_{\bullet} \right)^2} \sqrt{2\pi \left(\Delta M_{\bullet} \right)^2}.$$ (3)

It will be shown below that the dispersion is an important ingredient when comparing model predictions to observations.

#### 3.1. QSO luminosity functions and BH mass function

The QSO LFs can now be computed. Up to the peak, the BH light curve can be well approximated by the simple exponential form

$$L(t) = \frac{\lambda M_\bullet c^2}{t_{\text{Edd}}} e^{\left(t-t_{\text{vir}}-\Delta t_{\text{peak}}\right)/\tau_{\text{ef}}} \theta_H \left(t_{\text{vir}} + \Delta t_{\text{peak}} - \Delta t_{\text{vis}} \lesssim t \lesssim t_{\text{vir}} + \Delta t_{\text{peak}}\right).$$ (4)

Here $t_{\text{Edd}} \approx 4 \times 10^8$ yr is the Eddington timescale, and $\tau_{\text{ef}} \approx \eta t_{\text{Edd}}/(1-\eta) \lambda$ is the $e$-folding time in terms of the BH mass-energy conversion efficiency $\eta$ and of the Eddington ratio $\lambda$; in addition, the Heaviside function $\theta_H$ specifies that the QSO shines

5 Recall that the Heaviside function $\theta_H$ is defined by

$$\theta_H(x) = \begin{cases} 1, & \text{if } x \text{ is true;} \\ 0, & \text{otherwise.} \end{cases}$$
unobscured only during the time interval $\Delta t_{\text{vis}}$ before the peak of its light curve. In this work we have used an Eddington ratio $\lambda$ slightly rising toward high $z$, and specifically from $\lambda \lesssim 1$ at $z \lesssim 2$ to $\lambda \gtrsim 3$ at $z \gtrsim 6$ (see Appendix A and Table A1).

We stress that, of the three relevant timescales entering Eq. (4): the virial time $t_{\text{vir}}$ depends on cosmology; the peak time $\Delta t_{\text{peak}}$ is obtained by solving the system of equations reported in Appendix A; the visibility time $\Delta t_{\text{vis}}$, dependent on dust and gas absorption, is taken as a parameter, since its computation is challenging in semi-analytical models.

The QSO LF at a time $t$ and luminosity $L$ is computed by summing up the contribution of all the sources which virialize at epochs $t_{\text{vir}} \lesssim t$ and have the right delay to shine at the time $t$ with luminosity $L$. Analytically, one has

\[
\Phi(L,t) = \int_{t - \Delta t_{\text{peak}}}^{t} \int_{t_{\text{vir}}}^{t_{\text{peak}}} d\nu_{\text{vir}} \int dM_\bullet \frac{d^2 N_{\text{BH}}}{d\nu_{\text{vir}} dM_\bullet} \delta_D \left( L - \frac{\lambda M_\bullet c^2}{t_{\text{edd}}} e^{(\nu_{\text{vir}} - \nu_{\text{peak}})/\tau_{\text{vis}}} \right),
\]

where $\delta_D$ indicates the Dirac delta function. The time delay $\Delta t_{\text{peak}}$ provided by our code ranges from 0.2 Gyr at redshift $z \gtrsim 5$ where $\lambda \approx 4$ to 1 Gyr at redshifts $z \lesssim 2$, where $\lambda \lesssim 1$, see Fig. 2 (right).

The declining phase of the QSO light curve can be neglected, because after the peak of activity the accretion rate drops to values much lower than the Eddington rate and it becomes much less efficient in radiating energy (see for a review Blandford 2004). Such a result is in line with what was found by Yu & Lu (2004) on the basis of the comparison between the observed QSO LFs and the local supermassive BH mass function.

The last step of our computation is to convert the QSO LFs computed above into the optical and hard X-ray bands through the appropriate bolometric corrections. For the optical band we use $\beta_X = 2$ to $3$ at at $z=8$; as expected, while at $z \approx 6$ the number density of bright QSOs is $\rho(z = 6; M_{1450} \lesssim -27) \approx 10^{59} \text{ Mpc}^{-3}$, the same number density at $z = 8$ is reached only for low luminosity $M_{1450} \lesssim -24$. The
dramatic drop off of the bright QSOs at above \(\zeta = 6\) is due to \(\Delta t_{\text{peak}}\) that, although relatively short (0.2 Gyr), associates the QSOs to halos virializing at significantly higher \(\zeta\), where the abundances of massive halos are much lower.

The flattening of the LFs at the low end is due both to the flatter slope of the host halo formation rate (see Eq. 1) and to the peak time \(\Delta t_{\text{peak}}\), weakly dependent on halo mass (cf. Fig. 2). Since the formation rate at fixed mass varies much more with redshift for massive halo, a peak time independent of mass flattens the LF at low luminosity.

The fit at high luminosity is ensured by the dispersion of 0.3 dex associated to the \(M_\bullet - M_{\text{vir}}\) relation. As can be seen in Fig. 4 (right), the cut off in galaxy halo mass (cf. § 2.1) would yield a drastic drop around \(L_B \approx 10^{13.2} L_\odot\). Introducing a gaussian scatter \(\Delta M_\bullet \approx 0.3\) dex in the relation Eq. 4 is required to obtain a global fit. We check that, for a given mass and virialization time of the host halo, a conceivable variation of the physical parameters from structure to structure can account for such a scatter \(\Delta M_\bullet\) in the relic BH mass; for example, at \(\zeta \approx 6\) and \(M_{\text{vir}} \approx 10^{13.2} M_\odot\) this can be achieved by doubling the clumping factor to \(C \approx 15\), or increasing the seed BH mass to \(10^3 M_\odot\), or decreasing the strength of QSO feedback \(\epsilon_{\text{QSO}}\) by a factor a few. It is worth noticing that also Mahmood et al. (2005) noted this problem, and empirically solved it by inserting a lorentzian tail in the BH formation rates for halo masses above \(10^{13.2} M_\odot\).

The observations by Richards et al. (2006, see their Fig. 20) show that the number density of very luminous QSOs with \(M_{1500} \lesssim -27\) peaks between \(\zeta = 2\) and \(\zeta = 3\). Our model reproduces this trend; specifically, the rise from high redshift to \(\zeta \approx 2.5\) is due to the strong increase of the formation rate for very massive halos, which overwhelms the decrease of the BH mass (see Eq. 2). But at redshifts \(\zeta \lesssim 2\) the latter starts to dominate the evolution causing the fall in the number density of bright sources; this is because after Eq. 4 large BH masses at low redshift correspond to halos so massive as to exceed the cut-off of the imposed at \(M_{\text{vir}}^{\max} \approx 10^{13.2} M_\odot\). Recall that this roughly describes the decreasing probability of getting one galaxy largely dominating in mass over possible companions in very massive halos; in other words, galaxy groups are more easily formed at lower redshift (see Wechsler et al. 2002; Zhao et al. 2003a).

To reproduce the optical data, we adopt a visibility time \(\Delta t_{\text{vis}} \approx 5 \times 10^7 - 10^8\) yr, with the longer value applying at redshifts \(\zeta \lesssim 2\). Similarly, the X-ray LFs requires visibility times around \(\Delta t_{\text{vis}} \approx 3 \times 10^8\) yr at \(1.5 \lesssim \zeta \lesssim 3\). The visibility time turns out to be a factor of 5–10 shorter than the time \(\Delta t_{\text{peak}}\) spent by the BH to grow to its final mass. This result implies absorption to have a leading role in determining the duration of the QSO phase; in fact, dust is largely present in the pre-QSO phase as witnessed by submm observations.

Since hard X-ray photons produced by the QSO are much less absorbed than UV photons, the hard X-ray emission from the growing supermassive BH is more easily detected. Actually this X-ray emission has been already revealed from the center of submm-selected galaxies (Alexander et al. 2005; Borys et al. 2005). A specific discussion on this issue is presented elsewhere (Granato et al. 2006).

As mentioned in Appendix A, the accretion rate onto the central BH has been taken at most slightly super-Eddington \(\lambda = M_\bullet / M_{\text{Edd}} \lesssim 4\), as expected when it is limited by radiation pressure (see Small & Blandford 1992; King & Pounds 2003). The best fit to the LF is obtained by allowing the Eddington ratio to decrease from \(\lambda \approx 4\) at \(\zeta \gtrsim 6\) to \(\lambda \approx 1\) at \(\zeta \gtrsim 1.5\).

If radiation pressure keeps the BH growth at around the Eddington limit, outflows with mass rates \(M_{\text{out}} \sim M_{\text{Edd}}\) are expected, as shown in Appendix A. In massive galaxies the outflows remove most of the cold and infalling gas. The model predicts the existence of clouds of quite chemically enriched gas flowing out from the host galaxies (the expelled cold gas). Spectroscopical studies of narrow associated absorption lines in QSOs have detected enriched gas outflows (Srianand & Petitjean 2000; D’Odorico et al. 2004). The average metallicity of the gas expelled from large galaxies \(M_{\text{vir}} \gtrsim 10^{12} M_\odot\) is \(Z \sim 1 - 2 Z_\odot\), a value that after proper dilution contributes to the metal abundance \(Z \sim Z_\odot/3\) of the intergalactic medium in the central regions of clusters (see review by Voit 2004).
Exploiting Eq. (2) it is straightforward to derive the bias and the clustering properties of QSOs at a given BH mass. The result is presented in Fig. 4 (right), which shows the agreement with the data points from the analysis of the 2dF QSO Redshift Survey (Croom et al. 2005; also Porciani et al. 2004). Note that the model predicts a large increase of the bias with redshift, since bright QSOs with \( M_\bullet \gtrsim 10^8 M_\odot \) form in extremely biased peaks (cf. Eq. (2)), which are presently mostly confined in rich galaxy clusters.

The accreted mass function predicted by our model is plotted in Fig. 7 (left), and is found to be in good agreement with the semiempirical determination by Shankar et al. (2004) and Marconi et al. (2004). This shows that for \( z \gtrsim 1.5 \) about 60% of the relic supermassive BH mass density is in place; the rest of the mass will be accreted at later times onto BHs with final mass \( M_\bullet \lesssim 10^8 M_\odot \). This result well agrees with observations, showing that at \( 0.5 \lesssim z \lesssim 1.5 \) QSO of relatively low luminosity \( L_X \lesssim 3 \times 10^{44} \text{ erg s}^{-1} \) provide about 60% of the X-ray background (Ueda et al. 2003). The 40% in mass still to be accreted can account for this fraction, if the correction from X-ray to bolometric luminosity decreases with redshift and/or with luminosity, i.e., the radiative efficiency increases in the X-ray band; by the way, observations seems to confirm this trend. The accreted mass function is consistent with the Soltan argument; we check this by redoing the calculations of Shankar et al. (2004) with the Eddington ratios adopted in this paper, and exploiting the Ueda et al. (2003) LFs corrected for very obscured sources.

4.2. Host galaxy properties

The model interfaced to the GRASIL code provides also good fits to a variety of galactic properties. Most of the results have been presented in a series of previous papers (Granato et al. 2004, Cirasuolo et al. 2005, Silva et al. 2005) and are summarized in Table A2. Note that the current set of parameters differs from the one adopted in Granato et al. (2004), but the general scenario and in particular the results concerning the galaxy properties are essentially unaffected.

Fig. 4 shows that small and large halos exhibit quite different behavior. In halos with masses \( M_{\text{vir}} \lesssim 10^{12} M_\odot \) the interplay between SFR, BH accretion, and respective feedbacks limits the growth of the central BH, which never reaches the power needed to remove all the gas. Therefore the SFR is low and prolonged (\( \Delta t \gtrsim \Delta t_{\text{peak}} \)), and a substantial amount of stars is formed even after the peak of the QSO activity. Contrariwise, in large galaxies with \( M_{\text{vir}} \gtrsim 10^{12} M_\odot \) the growth of the BH produces winds, which are able to stop star formation in a short time lapse \( \Delta t \approx \Delta t_{\text{peak}} \approx 0.2 [(1 + z)/7]^{-1.5} \text{ Gyr} \).

On the other hand, in the pre-QSO phase the SFR varies from several hundreds to several thousands, within halos of mass ranging from \( 10^{10} \) to \( 10^{12} M_\odot \). Almost all the associated power is emitted in the far IR. The model fits the 850 \( \mu \text{m} \) counts; in particular, it predicts a large surface density \( N(\gtrsim 1 \text{ Jy}, z \gtrsim 5) \approx 600 \text{ (sq. deg.)}^{-1} \) of pre-QSO host galaxies at substantially high redshift among bright objects selected at 850 \( \mu \text{m} \) (Silva et al. 2005). The large number (respect to that of the QSO at the same redshift) is mainly ascribed to the fact that the pre-QSO phase lasts a factor of 10 longer than the QSO phase, i.e., \( \Delta t_{\text{peak}} / \Delta t_{\text{vir}} \approx 10 \) at high redshift.

A direct consequence of the extremely high SFR in massive halos is the early appearance of galaxies with large mass in stars \( M_\star \propto \text{SFR} \times \Delta t_{\text{peak}} \) at high redshift. These galaxies, at redshift corresponding to cosmological time \( t_{\text{H}} \gtrsim t_{\text{vir}} + \Delta t_{\text{peak}} \), have already hosted their own QSO and are evolving passively, i.e., without major addition of newly formed stars or large mass accretion onto the central BH. Interestingly, the 850 \( \mu \text{m} \) counts can be adjusted also with SFR lower than predicted by our model, provided that a flatter initial mass function is assumed (Baugh et al. 2005). The latter approach yields a much lower number density of massive, high-redshift galaxies; stronger constraints on this number density would help in discriminating between the two scenarios (e.g., Tecza et al. 2004; Fontana et al. 2004; Greve et al. 2005; Caputi et al. 2006). The predicted number density of galaxies with \( M_\star \gtrsim 10^{11} M_\odot \) after the QSO phase at redshift \( z \approx 5 \) is a factor of 200 below that at \( z \approx 1.5 \). Once more the identification of such objects would be an important test of the model (see Bouwens & Illingworth 2006).
The short $\Delta t_\mathrm{vis}$ of the order of half a Gyr is responsible of the observed enhancement of $\alpha$ elements in massive galaxies (see also Romano et al. 2002; Granato et al. 2004), since the $Fe$ enrichment is strictly related to the explosion of Type-Ia SNe and these are delayed by a time around 1 Gyr. The behavior of the average stellar metallicity $\langle Z_* \rangle$ as a function of the host halo mass is well described by the approximate expression $\langle Z_* \rangle \approx -0.247 + 0.023 \log (M_{\mathrm{vir}}/M_\odot)$; it increases for increasing halo/galaxy mass, a trend in agreement with observations of elliptical galaxies (see Thomas et al. 2005).

In large galaxies the SN heating can remove only a small fraction of the cold gas involved in star formation, which is rapidly metal-enriched due to the high SFR; thus, when the QSO reaches the peak, the energy released removes from the halo most of the initial, metal-poor gas. On the contrary, in small galaxies SN explosions are more efficient in removing the enriched cold gas, which is continuously replaced by the initial diffuse and metal-poor gas, keeping the metallicity of the cold gas low. As a result large galaxies, in spite of the shorter $\Delta t_\mathrm{vis}$, exhibit stellar populations with larger metal content than smaller galaxies do.

The power of the QSO outflows ensures that the mass turned into stars for large halos $M_{\mathrm{vir}} > 10^{12} M_\odot$ is only about 30% of the initial cosmic baryons; for smaller halos the long lasting action of SNe is able to remove even a larger fraction of gas, letting only less than 10% of the initial gas to turn into stars (Granato et al. 2004).

Cirasuolo et al. (2005) found that the observed velocity dispersion function of spheroidal galaxies (Sheth et al. 2003) and the Faber-Jackson relation can be reproduced under the hypotheses that the old stellar populations are located in galaxy halo virialized at redshift $z \gtrsim 1.5$ and that the relation $\sigma \approx 0.55 V_{\mathrm{vir}}$ holds; here the velocity dispersion $\sigma$ refers to old stellar populations in central galaxy regions. They also showed (cf. their Fig. 4) that the same $\sigma - V_{\mathrm{vir}}$ relation plugged into the $M_* - M_{\mathrm{vir}}$ for redshift $z \gtrsim 1.5$, well fit the data in the plane $M_* - \sigma$ with their dispersion (Ferrarese & Merritt 2000; Tremaine et al. 2002). Interestingly, our results predict a bending of the relationship in this plane, since for small halos SNe are able to keep the SFR at low level enough to hamper the growth of the reservoir around the seed BH and eventually of the BH itself (cf. Eq. [A3]; also Granato et al. 2004).

5. COMPARISONS WITH PREVIOUS MODELS

Reproducing the QSO and host galaxy LFs along with their statistical properties and relationships is a severe challenge for analytical, semi-analytical and numerical galaxy formation and evolution models.

In analytical models the standard methodology to estimate the QSO LFs is to convert the DM halo formation rates into BH formation rates through a link between the mass of the supermassive BH and that of the host halo; then a description of the QSO light curve (usually Eddington-limited accretion) is assumed to link the BH mass to the QSO luminosity. The QSO LFs are then built up according to the approximate expression

$$
\Phi(L, z) \approx \Delta t_{\mathrm{vis}} \frac{d^2 N_{\mathrm{ST}}}{d \log L d M_{\mathrm{vir}}} \left| \frac{d M_{\mathrm{vir}}}{d M_*} \right| \left| \frac{d M_*}{d L} \right|
$$

the above equation corresponds to Eq. (5), under the assumptions that the duty-cycle of QSO activity $\Delta t_{\mathrm{vis}}$ is much shorter than the cosmological time and that the QSO appears as soon as the host halo virializes, i.e. $\Delta t_{\mathrm{peak}} \approx 0$.

In two papers, Whythe & Loeb (2003) and Mahmood et al. (2004) estimated the QSO LFs as function of cosmic time by assuming: (i) no delay between virialization and QSO peak luminosity; (ii) a visibility time $t_{\mathrm{vis}} \approx t_{\mathrm{disc}}$ where the dynamical time of the disc is $t_{\mathrm{disc}} = 0.035 R_{\mathrm{vir}}/V_{\mathrm{vir}} \approx 7 \times 10^7 (1+z)^{-3/2}$ yr; (iii) rates derived from the extended Press & Schechter theory (Lacey & Cole 1993); (iv) $M_* - M_{\mathrm{vir}}$ relation deduced from the relations $M_* - \sigma$ and $\sigma - M_{\mathrm{vir}}$ observed by Ferrarese (2002).

The first and second assumptions are quite critical at high redshift. If the time scale of QSO activity is associated to the dynamical timescale of a galaxy disc, one has first to wait to built up the disc, which implies a delay time $\Delta t \gtrsim \max[t_{\mathrm{cool}}, t_{\mathrm{dyn}}]$, where both times are much larger than $t_{\mathrm{disc}}$.

The specification of the DM halo formation rates is a crucial step for analytical models; most of the works in the literature adopt the extended Press & Schechter theory (Bond et al. 1991). In this paper we have used instead the rates provided by the
positive time derivative of the Sheth & Tormen mass function. Notice that in both cases the rates are given by the product of
the mass function (Press & Shechter or Sheth & Tormen, respectively) by a factor, weakly dependent on time (Kitayama & Suto
1996). As a matter of fact, mass functions derived from numerical simulations are much better approximated by the Sheth &
Tormen than by the Press & Schechter theory; the latter at high redshift and at the large halo masses, relevant for QSOs, can be
smaller by a factor of 10 (e.g., Springel et al. 2005).

As for the last assumptions recalled above, these authors assume a $M_\bullet-M_{\rm vir}$ relationship quite at variance with ours; on average,
they predict at higher redshift larger BH mass at fixed halo mass (see also discussion in Shankar et al. 2006). Recall that we
obtain the relation presented in Eq. (2) as a self-consistent output of the equations in Appendix A, that follow the details of the
building up of the central BH mass.

In summary, the comparison of our results with those of Whythe & Loeb (2003) and Mahmood et al. (2004) is complex,
because of quite different assumptions. Despite of these, similar QSO LFs are produced due to compensations; at high redshift
their lower $t_{\rm vis} \lesssim 10^7$ yr and formation rates of DM halos are compensated by: (i) neglecting the time needed to built up the
galactic structure and (ii) assuming a relationship $M_\bullet-M_{\rm vir}$ that privileges larger BH masses. The difference emerges quite
clearly at redshifts around $7-8$, where we predict a dramatic drop off of the LF at high luminosity, because of the delay (see
§ 4.1), whereas these authors found number densities larger by a factor of 10.

On the other hand, differences are quite clear on the host galaxy side. In our model the time $\Delta t_{\rm peak}$ is a very important and
active phase, within which the massive host galaxies are forming stars at rates around $10^7-10^3M_\odot$ yr$^{-1}$ (cf. Fig. 1) and the
central BH is growing. This phase is directly explored by submm surveys and by the observations of the X-ray emission from the
-growing QSO detected by Alexander et al. (2005).

More recently, several authors tried to follow the evolution of both QSOs and host galaxies using results from numerical
simulations, in which the effects of SNe explosions and winds from the central active nucleus are taken into account (see
Hopkins et al. 2006 and references therein). These authors took an approach complementary to that presented in this paper.
Instead of starting from first cosmological principles, they simulated many realizations of collisions between two stable, isolated
disk galaxies endowed with a central supermassive BH. A two phases interstellar medium was used to describe star formation and
SN feedback (Springel & Hernquist 2003). The accretion onto the BHs was estimated from local gas density and sound speed
limited to the Eddington rate. These authors assumed that 5% of the bolometric luminosity due to accretion is transferred to the
surrounding gas, as thermal energy.

They found that during a major merger the central BHs are fed with enough gas to yield a luminosity depending exponentially
on the time. Their approach includes also estimates of absorption by dust. The energy injected by the QSO in the interstellar
medium is enough to unbind the gas itself on a time scale $\Delta t_{\rm vis} \sim 10^8$ yr, within which the luminosity varies from about $10^{33} L_{\rm peak}$
to a maximum $L_{\rm peak}$, with a weak dependence on $L_{\rm peak}$ (cf. Eq. [7] of Hopkins et al. 2006). In the final step, by exploiting the
QSO light curve obtained from the simulations, they compute the rate of appearance of QSOs $n(L_{\rm peak})$ with luminosity $L_{\rm peak}$,
starting from the observed LFs. A good representation of their finding is a log-normal distribution of peak luminosity. At high
redshift $z \gtrsim 2$ pure density evolution or pure luminosity evolution is assumed.

The time scale $\Delta t_{\rm vis}$ for the growth of the BH derived from numerical simulations of low redshift galaxy mergers at large
halo mass has the same meaning of our time $\Delta t_{\rm peak}$ when the maximum luminosity is attained. However, at high mass and high
redshift the two timescales have quite different values $\Delta t_{\rm vis} \sim 10^8 \gtrsim \Delta t_{\rm peak} \approx 0.2 [(1+z)/7]^{-1.5}$ Gyr. The long time before the
peak is extremely relevant at high redshift $z \gtrsim 4$, when $\Delta t_{\rm vis} \sim t_{\rm ff}$; in fact, there we expect that the model of two colliding discs
requires too long time (recall that the discs themselves requires at least $R_{\rm vir}/v_{\rm vir}$ to set up). Therefore the extension of the model
at $z \gtrsim 2$ is quite complex, as discussed by Hopkins et al. (2005) in their § 3.2. However, it is interesting to note that at lower
redshift $z \lesssim 1.5$ the two times are comparable $\Delta t_{\rm vis} \sim \Delta t_{\rm peak}$. Our model is not aimed at following the evolution of QSOs at lower

![Figure 6](image-url)
redshift $z \lesssim 1$; this later phase is better described by models that include “interactions” (minor mergers, fly-bys, disk instabilities, etc. etc.) among galaxies as triggers of nuclear activity (e.g., Kauffmann & Haehnelt 2000; Cavaliere & Vittorini 2000; Menci et al. 2003).

6. CONCLUSIONS

In their thoughtful review paper, Brandt & Hasinger (2005) list some “key outstanding problems” of AGN astrophysics that need investigation. These include: the detailed cosmic history of supermassive BH accretion; the nature of AGN activity in young, forming galaxies, and the connection between supermassive BH growth and star formation in submm galaxies; the clustering properties of AGNs.

In this paper we address these issues in the framework of the anti-hierarchical baryon collapse (ABC) scenario developed by Granato et al. (2004). We have shown that the condensation of baryons within DM halos both in stars and in BHs can be described with a simple physical model, whose main equations are listed in Appendix A. The model yields the time dependence of the SFR, of the accretion rate $\dot{M}$ onto the central BH, of the total mass in stars $M_\star$ and of the final BH mass $M_\bullet$ for any given halo mass and virialization epoch. The GRASIL code then provides the SED from X-ray to radio bands of evolving stellar populations (successfully reproducing the epoch-dependent galaxy LFs in different spectral bands, as well as a variety of relationships among photometric, dynamical and chemical properties, as shown in previous papers; see Table A2), while observationally determined bolometric corrections allow us to convert the accretion rates onto supermassive BHs into luminosities in optical and X-ray bands.

In the ABC scenario, the growth rate of BHs is proportional to the SFR, and the latter is more effectively slowed down by SN feedback in smaller halos, resulting in more efficient growth of more massive BHs, especially at higher redshifts. Both star formation and BH growth in massive halos are stopped by the feedback from the active nucleus as soon as it becomes powerful enough to sweep out the residual interstellar medium.

Most of the BH accretion is heavily dust obscured and occurs in host galaxies with very intense star formation. For massive objects, the growth phase lasts 15–20 $e$-folding times of Eddington-limited accretion, but the AGNs are detectable by current hard X-ray surveys only in the last several $e$-folding times, i.e. have hard X-ray visibility times of $\sim 3 \times 10^8$ yr. Interestingly, Borys et al. (2005) estimate that the BH masses associated to the X-ray emitting AGNs detected in submm galaxies are, on average, $\sim 50$ times lower that those of associated to local spheroidal galaxies with similar stellar masses, consistent with being $\sim 4$ $e$-folding times before reaching their final mass. At earlier times, submm galaxies are expected to host intrinsically weak and highly obscured nuclei, undetectable with current X-ray telescopes.

The model accounts for the hard X-ray AGN LF at various redshifts. A crucial ingredient, to this end, is the rapid quenching of the accretion rate onto the biggest BHs, predicted by the model as the consequence of the AGN feedback which sweeps out the residual interstellar medium.

The optical (B-band) visibility time is shorter than that for the more penetrating hard X-rays. The observed epoch-dependent B-band LFs are accurately reproduced for a $\Delta t_{\text{vis}}$ of order of the $e$-folding time, indicating that only when the AGN is approaching its maximum luminosity it can clear up the surrounding region. The redshift dependence of the space density of optically bright quasars is controlled, in the ABC scenario, by two competing factors. On one side, according to the hierarchical clustering paradigm, the formation rate of very massive halos hosting them is increasing rapidly with decreasing redshift. On the other side, the BH to host halo mass ratio decreases with decreasing redshift, so that at low redshifts relatively more massive host are required for a given BH mass; but the galactic halo mass function sinks down exponentially for large halo masses, and is actually cut off at $M_{\text{vir}} \approx 10^{13.2} M_\odot$. In our model, the first factor dominates for $z \gtrsim 2.5$ and the second dominates at lower $z$, thus accounting for the increase in the bright QSO space density with decreasing $z$ down to $z \approx 2.5$, as well as for the subsequent decrease.

The redshift dependence of the bolometric QSO LF is illustrated by Fig. 8. It is broadly reminiscent of luminosity evolution,
but with significant deviations from that simple description particularly at high luminosities. We note in particular the flattening at high luminosity around $z \approx 5$, borne out by SDSS data (Fan et al. 2001) and the very sharp drop (by about 2 orders of magnitude) of the QSO LF between $z \approx 6$ and $z \approx 8$, due to the dearth of massive halos at high redshift.

Since luminous high-$z$ quasars are associated to very massive halos, the model implies that their clustering properties are similar to those of massive spheroidal galaxies and of the bright submm galaxies detected by the SCUBA surveys. We have shown that the redshift-dependent bias factor predicted by our model for a typical BH mass of $10^{8} M_{\odot}$ matches the observational determination by Croom et al. (2005).

During the relatively short phase when massive galaxies are in the process of expelling their gas and dust, their nuclei should appear as powerful X-ray sources $L_{X} \gtrsim 10^{45} \text{erg s}^{-1}$, but still somewhat obscured in the optical band; they thus may resemble Type 2 QSOs. Spectroscopic studies, aimed to determine velocity, amount and chemical abundances of the gas ejected during the QSO phase, would be extremely informative; in fact, the model predicts that at early times in the Universe large amounts of gas were evacuated from the galaxy halos to the intergalactic medium.

We warmly thank S. Bressan and M. Cirasuolo for helpful discussions. Work supported by grants from ASI, INAF and MIUR.

REFERENCES

Alexander D.M., Bauer, F.E., Chapman, S.C., Smail, I., Blain, A.W., Brandt, W.N., & Ivison, R.J., 2005, ApJ, 632, 736
Bond, J.R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440
Bennett, C.L., et al., 2003, ApJS, 148, 1
Bardeen J., Bond J., Kaiser N. & Szalay A., 1986, ApJ, 304, 15
Barger, A.J., Cowie, L.L., Mushotzky, R.F., Yang, Y., Wang, W.-H., Steffen, A.T., & Capak, P., 2005, AJ, 129, 578
Baugh, C.M., Lacey, C.G., Frenk, C.S., Granato, G.L., Silva, L., Bressan, A., Benson, A. J., & Cole, S., 2005, MNRAS, 356, 1191
Bernardi, M., et al., 1998, ApJ, 508, L143
Blandford, R.D., 2004, in Coevolution of Black Holes and Galaxies, ed. L.C. Ho (Cambridge: Cambridge Univ. Press), 153
Bond, J.R., Cole, S., Efstathiou, G., & Kaiser, N., 1991, ApJ, 379, 440
Borys C., Smail I., Chapman S.C., Blain A.W., Alexander D.M. & Ivison R.J., 2005, ApJ, 635, L143
Bouwens, R., & Illingworth, G., 2006, NewAR, 50, 152
Bower, R.G., Lucey, J.R., Ellis, R.S., 1992, MNRAS, 254, 589
Brandt, W.N., & Hasinger, G., 2005, ARA&A, 43, 827
Brandt, W.N., & Gallagher, S.C., 2000, NewAR, 44, 461
Bromley, J.M., Somerville, R.S., & Fabian, A.C., 2004, MNRAS, 350, 456
Caputi, K.I., McLure, R.J., Dunlop, J.S., Cirasuolo, M., & Schael, A.M., 2006, MNRAS, 366, 609
Cavaliere, A., & Vittorini, V., 2000, ApJ, 543, 599
Chapman, S.C., Blain, A.W., Smail, I., & Ivison, R.J., 2005, ApJ, 622, 772
Chapman, S.C., Blain, A.W., Ivison, R.J., & Smail, I. 2003, Nature, 422, 695
Chartas, G., Brandt, W.N., & Gallagher, S.C., 2003, ApJ, 595, 85
Cimatti, A., et al., 2002, A&A, 392, 595
Cirasuolo M., Shankar F., Granato G.L., De Zotti G. & Danese L., 2005, ApJ, 629, 816
Cristiani S., et al., 2004, ApJ, 600, L119
Croom, S.M., et al., 2005, MNRAS, 356, 415
Croom S.M. et al., 2004, MNRAS, 349, 1397
D’Odorico V., Cristiani S., Romano D., Granato G.L., Danese L., 2004, MNRAS, 351, 976
Dressler, A., et al., 1987, ApJ, 313, 42
Fan, X., et al., 2004, ApJ, 128, 515
Fan, X. et al., 2001, AJ, 121, 1203
Ferrarese, L., & Merritt, D., 2000, ApJ, 539, L9
Floyd, D.J.E., et al. 2004, MNRAS, 355, 196
In this Appendix we provide an overview of the physical model by Granato et al. (2004), recalling the basic equations and parameters that control the evolution of the baryonic component.

The baryonic content of a given DM halo with mass $M_{\text{vir}}$ is partitioned in three gaseous phases: a hot diffuse medium with mass $M_{\text{inf}}$ infalling and/or cooling toward the center; cold gas with mass $M_{\text{cold}}$ condensing into stars; low-angular momentum gas stored in a reservoir around the central supermassive BH, and eventually viscously accreting onto it. In addition, two condensed phases are present, namely, the stellar mass $M_\star$ (Eq. A2) and the gas mass $M_{\text{res}}$.

The evolution of the baryonic content is described by the system of differential equations:

$$M_{\text{inf}} = -M_{\text{cond}} - M_{\text{inf}}^{\text{QSO}},$$

$$M_{\text{cold}} = M_{\text{cond}} - M_\star - M_{\text{res}} - M_{\text{cold}}^{\text{SN}} - M_{\text{cold}}^{\text{QSO}},$$

$$M_{\text{res}} = M_{\text{inflow}} - M_\star,$$

where $M_{\text{cond}}$ is defined by Eq. (A2). At the virialization redshift $z_{\text{vir}}$ we set $M_{\text{inf}} \approx M_{\text{vir}}/6$ following the universal baryon to DM ratio and featuring a primordial chemical composition, $M_\star \approx M_{\text{seed}}$ in terms of a seed mass originated by some process in the early universe, and the other baryonic components to zero.

We now explicit the various terms on the r.h.s. in Eqs. (A1). The hot gas condenses/cools down at the rate

$$M_{\text{cond}} = \frac{M_{\text{inf}}}{\max(\Delta t_{\text{cool}}, \Delta t_{\text{dyn}})},$$

in terms of the dynamical and cooling timescales $\Delta t_{\text{dyn}}$ and $\Delta t_{\text{cool}}$ at the virial radius, respectively. As for the latter, it includes the appropriate cooling function (Sutherland & Dopita 1993) and it allows for a clumping factor $C$ in the baryonic component. Since at high redshift major mergers between massive halos are very frequent, we neglect here the effect of the angular momentum, since it is lost by dynamical friction through mergers of mass clouds $M_\star$ on time scale $\tau_{\text{fr}} \approx 0.2 (\xi / \ln \xi)^{-1} \Delta t_{\text{dyn}}$, where $\xi = M_{\text{vir}}/M_\star$.
(see e.g. Mo & Mao 2004); major mergers imply \( \xi \sim \) a few. In this context it is also worth noticing that the gas that collapses and cools enough to form stars is only a fraction less than 30\% of that associated to the virialized halo (cf. Fig. 1).

The cold gas turns into stars at the rate

\[
M_\star = \int \frac{dM_{\text{cold}}}{\max[t_{\text{cool}}-t_{\text{dyn}}]},
\]

where now \( t_{\text{cool}} \) and \( t_{\text{dyn}} \) refer to the mass shell \( dM_{\text{cold}} \).

The energy feedback from the ensuing Type-II SNae remove gas from the cold phase at the rate

\[
M_{\text{SN cold}}^\text{SN} \approx \epsilon_{\text{SN}} \left( \frac{V_{\text{vir}}}{500 \text{ km s}^{-1}} \right)^{-2} M_\star,
\]

in terms of an efficiency \( \epsilon_{\text{SN}} \). The amount of cold mass removed is proportional to the number of SN explosions (assuming a average energy around \( 10^{51} \text{ ergs per SN} \) hence to the SFR), and inversely proportional to the depth of the halo potential well.

The radiation drag (see Kawakatu & Umemura 2002) from star formation trigger the inflow of cold gas into a reservoir of low-angular momentum around the central supermassive BH, at the rate

\[
\dot{M}_{\text{inflow}} \approx \alpha_{\text{RD}} \times 10^{-3} \dot{M}_\star \left(1-e^{-t_{\text{BH}}/\tau} \right) M_\odot \text{ yr}^{-1},
\]

in terms of the strength \( \alpha_{\text{RD}} \), and of the effective optical depth

\[
\tau = \tau_0 \left( \frac{Z}{Z_\odot} \right) \left( \frac{M_{\text{cold}}}{10^{12} M_\odot} \right) \left( \frac{M_{\text{vir}}}{10^{13} M_\odot} \right)^{-2/3},
\]

normalized through the parameter \( \tau_0 \); the metallicity \( Z \) is computed from the code self-consistently, see Granato et al. (2004) for further details.

The viscous time \( t_{\text{visc}} = R_{\text{crit}} t_{\text{dyn}}^{\text{BH-res}} \) of the accretion from the reservoir to the BH is related to the dynamical time of the system \( BH \) + reservoir \( t_{\text{dyn}}^{\text{BH-res}} \), and to the critical Reynolds number \( Re_{\text{crit}} \sim 10^2 - 10^3 \), see Granato et al. (2004) for added details; thus the gas in the reservoir accretes onto the supermassive BH at the rate

\[
M_{\star}^{\text{visc}} \approx 5 \times 10^3 k_{\text{acc}} \left( \frac{V_{\text{vir}}}{500 \text{ km s}^{-1}} \right)^3 \left( \frac{M_{\text{res}}}{M_\star} \right)^{3/2} \left( \frac{M_{\text{visc}}}{M_{\text{res}}} \right)^{1/2} M_\odot \text{ yr}^{-1},
\]

in terms of an efficiency \( k_{\text{acc}} \). For reasonable values of \( 10^{-4} \lesssim k_{\text{acc}} \lesssim 10^{-2} \), the accretion rate \( M_{\star}^{\text{visc}} \) easily exceeds the Eddington rate for high redshift and massive halos. Several possible alternatives might occur in this situation: (i) very rapid growth of the central BH in a dynamical timescale, followed by a slower accretion when less mass is available in the reservoir; (ii) limited BH growth by radiation pressure, followed by strong mass outflows \( M_{\star} \sim M_{\text{Edd}} \) (e.g., Small & Blandford 1992; King & Pounds 2003; Begelman 2004). We select the latter option and the actual accretion rate is set

\[
\dot{M}_\star = \min[M_{\star}^{\text{visc}}, \lambda M_{\star}^{\text{Edd}}],
\]

where

\[
M_{\star}^{\text{Edd}} \approx 1.3 \left( \frac{M_\star}{10^8 M_\odot} \right) M_\odot \text{ yr}^{-1},
\]

is the Eddington rate and \( \lambda \lesssim 4 \) is the maximum allowed Eddington ratio. Since \( M_{\star}^{\text{visc}} \) is larger at higher redshift, we allow also \( \lambda \) mildly varying with \( z \), see Table A1.

In presence of Eddington limited or super-Eddington accretion, QSO- driven outflows are expected, see above; large mass outflows have been confirmed by X-ray observations of BAL QSOs (Brandt & Gallagher 2000; Chartas et al. 2003). Winds

---

**TABLE A1**

**MODEL PARAMETERS**

| Parameter | Value | Short description |
|-----------|-------|-------------------|
| \( C \) | 7 | clumping factor |
| \( \epsilon_{\text{SN}} \) | 0.05 | strength of SN feedback |
| \( \alpha_{\text{RD}} \) | 2.5 | strength of radiation drag |
| \( \tau_0 \) | 1 | zero-point of optical depth |
| \( M_{\text{seed}} \) | \( 10^2 M_\odot \) | mass of BH seed |
| \( k_{\text{acc}} \) | \( 10^{-2} \) | strength of viscous accretion |
| \( \lambda \) | \( 0.8-4 \) | Eddington ratio |
| \( \eta \) | 0.15 | radiative efficiency |
| \( \epsilon_{\text{QSO}} \) | 1.3 | strength of quasar feedback |

*NOTE:* We let the maximum allowed Eddington ratio \( \lambda \) to depend on the redshift as: \( \lambda = 4 \) for \( z \gtrsim 6 \), \( \lambda = 3 \) for \( 5 \lesssim z \lesssim 6 \), \( \lambda = 1.7 \) for \( 3 \lesssim z \lesssim 5 \), \( \lambda = 1 \) for \( 2 \lesssim z \lesssim 3 \), and \( \lambda = 0.8 \) for \( 1.5 \lesssim z \lesssim 2 \); the empirical fit \( \lambda(z) \approx -1.15 + 0.75(1+z) \) works well in the redshift range \( 1.3 \lesssim z \lesssim 6 \).
Table A2
Overview of Model Results

| Property or Statistics | External Inputs | Reference |
|------------------------|-----------------|-----------|
| **QSOs and BHs**       |                 |           |
| $M_\bullet - M_{vir}$   | no add          | Fig. 2 (left) |
| $\Delta t_{peak} - M_{vir}$ | no add          | Fig. 2 (right) |
| Opt. LFs ($z$)         | $\Delta t_{vis} \approx 5 \times 10^7$ yr, $\Delta M_\bullet \approx 0.3$ dex | Figs. 3-6 |
| X-ray LFs ($z$)        | $\Delta t_{vis} \approx 3 \times 10^8$ yr | Figs. 5-6 (right) |
| BH mass function       | no add          | Fig. 7 (left) |
| QSO clustering         | no add          | Fig. 7 (right) |
| $M_\bullet - \sigma$   | no add          | Fig. 6 of G04 |
| **Spheroidal Galaxies**|                 |           |
| VDF                    | $\sigma/V_{vir}$ | Fig. 1 of C05 |
| Faber-Jackson          | $\sigma/V_{vir}$ | Fig. 2 of C05 |
| LF ($z$)               | IMF, SEDs (GRASIL) | Fig. 10 of G04 |
| Metal abundances       | IMF, chemical yields | Fig. 9 of G04 |
| K-band counts          | IMF, dust modeling (GRASIL) | Fig. 1 of S05 |
| 850 $\mu$m counts      | IMF, dust modeling (GRASIL) | Fig. 12 of G04 |
| EROS                   | IMF, dust modeling (GRASIL) | Fig. 6 of S05 |

Note. — * specific assumptions or parameters, not included in Table A1. References include G04: Granato et al. (2004), C05: Cirasuolo et al. (2005), Silva et al. (2005).

Can form just above the disc by a combination of radiation and gas pressure; following the model by Murray et al. (1995) we estimate the mass outflow rate as $M_w \approx 2.8 f_c N_{22} (M_\bullet/10^8 M_\odot)^{1/2} M_\odot$ yr$^{-1}$ in terms of the covering factor $f_c$ of the hydrogen column density $N_{22}$ normalized to $10^{22}$ cm$^{-2}$. Then the kinetic luminosity of the outflow reads

$$L_K = \frac{1}{2} M_w v_\infty^2 = 4.2 \times 10^{44} \left( \frac{f_c}{0.1} \right) \left( \frac{N_{22}}{10} \right) \left( \frac{M_\bullet}{10^8 M_\odot} \right)^{3/2},$$

(A10)

where $v_\infty \propto M_\bullet^{1/2}$ is the asymptotic speed of the outflow. Eventually, QSO outflows remove gas from both the hot and the cold phases at the rates

$$M_{inf, cold}^{QSO} = 3 \times 10^3 \epsilon_{QSO} \left( \frac{V_{vir}}{500 \text{ km/s}} \right)^{-2} \left( \frac{M_\bullet}{10^8 M_\odot} \right)^{3/2} \frac{M_{inf, cold} + M_{inf}}{M_{inf}} M_\odot \text{ yr}^{-1},$$

(A11)

in terms of the strength $\epsilon_{QSO} = (f_\epsilon/0.5)(f_c/0.1)(N_{22}/10)$, where $f_\epsilon$ is the fraction of the kinetic luminosity transferred to the gas. Since the power of the outflows increases with the BH mass $M_w \sim M_{Edd}$, most of the energy transferred from the QSO to the halo gas is released very close to the peak of the activity. It is worth noticing that the power transferred to the gas components $L_\epsilon = f_\epsilon L_K$ is only a fraction of the Eddington luminosity $(L_\epsilon/L_{Edd}) \approx 1.4 \times 10^{-2} \epsilon_{QSO} (M_\bullet/10^8 M_\odot)$.

The basic outputs of the model are the evolution of the SFR $M_\star$, and the BH accretion rate $M_\bullet(t)$ as a function of the galactic age $t$ for any given value of the halo mass $M_{vir}$ and of the virialization redshift $z_{vir}$. Once the star formation history, $M_\star$, of a galaxy has been computed, its luminosity in any chosen band is obtained as a function of $t$ from the GRASIL code, which yields the chemical and the spectro-photometric evolution from the radio to the X-ray band, allowing for the effect of dust absorption and reradiation.