

Quasi-stars and the cosmic evolution of massive black holes

Marta Volonteri¹★ and Mitchell C. Begelman^{2,3}★

¹Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, MI 48109-1042, USA

²JILA, 440 UCB, University of Colorado at Boulder, Boulder, CO 80309-0440, USA

³Department of Astrophysical and Planetary Sciences, 391 UCB, University of Colorado, Boulder, CO 80309-0391, USA

Accepted 2010 July 13. Received 2010 July 13; in original form 2010 March 26

ABSTRACT

We explore the cosmic evolution of massive black hole (MBH) seeds forming within ‘quasi-stars’ (Qs), accreting black holes embedded within massive hydrostatic gaseous envelopes. These structures could form if the infall of gas into the centre of a halo exceeds about $1 M_{\odot} \text{ yr}^{-1}$. The collapsing gas traps its own radiation and forms a radiation pressure-supported supermassive star. When the core of the supermassive star collapses, the resulting system becomes a quasi-star. We use a merger-tree approach to estimate the rate at which supermassive stars might form as a function of redshift, and the statistical properties of the resulting QS and seed black hole populations. We relate the triggering of runaway infall to major mergers of gas-rich galaxies and to a threshold for global gravitational instability, which we link to the angular momentum of the host. This is the main parameter of our models. Once infall is triggered, its rate is determined by the halo potential; the properties of the resulting supermassive star, QS and seed black hole depend on this rate. After the epoch of Qs, we model the growth of MBHs within their hosts in a merger-driven accretion scenario. We compare MBH seeds grown inside quasi-stars to a seed model that derives from the remnants of the first metal-free stars, and also study the case in which both channels of MBH formation operate simultaneously. We find that a limited range of supermassive star/QS/MBH formation efficiencies exists that allows one to reproduce observational constraints. Our models match the density of $z = 6$ quasars, the cumulative mass density accreted on to MBHs (according to Sołtan’s argument) and the current mass density of MBHs. The mass function of Qs peaks at $M_{\text{QS}} \simeq 10^6 M_{\odot}$, and we calculate the number counts for the *James Webb Space Telescope* (JWST) 2–10 μm band. We find that JWST could detect up to several Qs per field at $z \simeq 5$ –10.

Key words: black hole physics – galaxies: formation – galaxies: nuclei – quasars: general – cosmology: theory – infrared: stars.

1 INTRODUCTION

While there is ample evidence that supermassive black holes (BHs) populate the nuclei of most large galaxies and that some BHs with masses exceeding $10^9 M_{\odot}$ formed as early as $z \gtrsim 6$ (e.g. Fan et al. 2001; Barth et al. 2003; Djorgovski et al. 2008; Jiang et al. 2009; Willott et al. 2009), there is little consensus as to the progenitors of these holes. Two schools of thought have persisted since Rees (1978) first devised a flow chart outlining possible routes of massive black hole (MBH¹) formation. According to one line of argument, supermassive BHs grew from the remnants of an early population

of massive stars, the so-called Population III (Pop III), which is believed to have formed in pregalactic minihaloes at $z \gtrsim 20$. According to the other, the precursors of supermassive BHs could have formed by the ‘direct collapse’ of large amounts of gas in much larger haloes at later times.

Each scenario has both positive and negative attributes. Although Pop III remnants were unlikely to have been more massive than a few hundred M_{\odot} each, they would have formed relatively early and thus their growth process would have had a considerable head start. They could have congregated and merged in the cores of merging minihaloes, while simultaneously growing by accretion. Models for the growth of supermassive BHs from stellar-mass seeds (Madau & Rees 2001; Volonteri, Haardt & Madau 2003; Rhoads & Haehnelt 2006; Monaco, Fontanot & Taffoni 2007; Somerville et al. 2008; Volonteri, Lodato & Natarajan 2008) are moderately successful in reproducing the current-day population of supermassive BHs, but have rather more difficulty in producing enough 10^9 – M_{\odot} holes at

*E-mail: martav@umich.edu (MV); mitch@jila.colorado.edu (MCB)

¹We refer here generically to MBHs when the hole mass is above the limit for BH formation in today’s stars, $\simeq 50 M_{\odot}$. This definition comprises both seed BHs and supermassive BHs in galaxies and quasars.

$z \gtrsim 6$ to explain the earliest known quasars (Shapiro 2005; Volonteri & Rees 2005, 2006). Additional worries about the Pop III scenario include the possibility that too many of the remnants would have been ejected from the cores of merging haloes and that their accretion rates would be depressed by the shallow potential wells of the host minihaloes and heating of the ambient gas by stellar radiation and winds (Milosavljević, Couch & Bromm 2009, and references therein).

Seed BHs that formed by direct collapse (e.g. at $z \lesssim 15$) would have had less time to grow, but this would have been partly compensated by their larger initial masses. Various direct collapse models for seed formation have been proposed (Loeb & Rasio 1994; Eisenstein & Loeb; Haehnelt, Natarajan & Rees 1998; Bromm & Loeb 2003; Koushiappas, Bullock & Dekel 2004; Begelman, Volonteri & Rees 2006, hereafter BVR2006; Lodato & Natarajan 2006; Begelman 2010), but so far only limited attempts have been made to place these models into the context of structure formation theories (e.g. Lagos, Cora & Padilla 2008; Volonteri et al. 2008; Van Wassenhove et al. 2010). Because direct collapse models draw only indirectly on star formation lore, there is much less consensus about the initial conditions for direct collapse and the details of how it might have occurred.

In a series of recent papers, we have described a sequence of events that we believe represents a plausible route to MBH formation via direct collapse (BVR2006; Begelman et al. 2008; Begelman 2010). In this picture, the main triggering event is the infall of gas into the centre of a halo at a rate exceeding about $1 M_{\odot} \text{ yr}^{-1}$. Such large rates of infall are possible in haloes with virial temperatures in excess of about 10^4 K, which only become common at $z \lesssim 10$ –15. Questions remain about the ability of gas to accumulate at such a high rate without too much of it fragmenting into stars, but as we argue below, recent simulations as well as analytic calculations suggest that the importance of fragmentation may have been severely overestimated in past work. The collapsing gas traps its own radiation and forms a quasi-static, radiation pressure-supported supermassive star, which burns hydrogen for about a million years while growing to a mass of $\gtrsim 10^6 M_{\odot}$ (Begelman 2010; note that the earlier claim in BVR2006 that hydrogen burning is unable to postpone collapse is erroneous and is corrected in the later paper).

Because of rotation, the BH that forms initially probably comprises only a small fraction of the core, with a mass of $\lesssim 10^3 M_{\odot}$, but it grows rapidly at a rate set by the Eddington limit for the massive gaseous envelope. A novel feature of our model is the prediction that the envelope swells by a factor of $\gtrsim 100$ in radius as it absorbs the energy liberated by BH growth. The resulting object, which we have dubbed a ‘quasi-star’ (QS), resembles a red giant with a luminosity comparable to a Seyfert nucleus. As the BH grows inside it, its photosphere expands and cools until it hits a minimum temperature associated with the Hayashi track, at which point it disperses, leaving behind the naked seed BH.

In this paper, we use a merger-tree approach to estimate the rate at which supermassive stars might form as a function of redshift, and the statistical properties of the resulting QSs and seed BHs. In Section 2 we summarize the properties of supermassive stars, QSs and seed BHs, as a function of time and gas infall rate. The existence of runaway infall depends on a threshold for global gravitational instability, which we model as a function of the gas mass and angular momentum following each halo merger – we discuss these criteria in Section 3. Once infall is triggered, its rate is determined by the halo potential; the properties of the resulting supermassive star, QS and seed BH depend on this rate.

To relate the properties of the seed BHs to the observable distributions of active and quiescent BHs at different redshifts, we apply a simple set of rules for their subsequent growth by accretion and mergers. These are described in Section 4, and in Section 5 we present our results. We compare our direct collapse seed model to a pure Pop III seed model and also study the case in which both channels of MBH formation operate simultaneously. We summarize our conclusions and discuss the prospects of testing this model observationally in Section 6.

2 FORMATION OF SUPERMASSIVE STARS, QUASI-STARS AND SEED BLACK HOLES

Theoretical aspects of the formation and evolution of supermassive stars, QSs and the resulting seed BHs are discussed in BVR2006, Begelman et al. (2008) and Begelman (2010). Here we summarize the formulae that we use as inputs for merger-tree calculations presented below, with emphasis on the results from Begelman et al. (2008) and Begelman (2010).

2.1 Supermassive stars

A supermassive star in the centre of a forming galaxy is unlikely to evolve at constant mass, but rather accumulates mass due to infall. We adopt a constant accretion rate $\dot{M} = \dot{m} M_{\odot} \text{ yr}^{-1}$. For infall rates larger than about $1 M_{\odot} \text{ yr}^{-1}$, the star never relaxes thermally to a fully convective state during its ‘main-sequence’ (hydrogen-burning) lifetime; instead it develops a convective core (with polytropic index $n = 3$) surrounded by a convectively stable envelope (Begelman 2010). The main-sequence lifetime is then

$$t_{\text{MS}} \sim 1.4 \dot{m}^{-1/2} T_{\text{c},8}^{-1} \text{ Myr}, \quad (1)$$

where $10^8 T_{\text{c},8}$ K is the central core temperature, which depends weakly on the mass and CNO abundance as shown in fig. 5 of Begelman (2010). For the calculations below, we adopt a fiducial value of $T_{\text{c},8} = 1.4$, which corresponds to a CNO abundance by mass $Z_{\text{CNO}} \sim 10^{-6}$ ($Z/Z_{\odot} \sim 10^{-4}$). Then, $t_{\text{MS}} \sim \dot{m}^{-1/2}$ Myr and the star’s mass at the end of hydrogen burning is

$$M_* \sim 10^6 \dot{m}^{1/2} M_{\odot}. \quad (2)$$

A partially convective supermassive star has a radius $R_* \sim 3.4 \times 10^{13} \dot{m}$ cm (for electron scattering opacity $\kappa = 0.34 \text{ cm}^2 \text{ g}^{-1}$), independent of mass, and an effective temperature

$$T_* = 1.2 \times 10^5 \dot{m}^{-3/8} \text{ K} \quad (3)$$

at the end of hydrogen burning. Although the values of T_* and R_* suggest that supermassive stars may be strong sources of far-ultraviolet radiation, we cannot confidently calculate the ionizing flux from a rapidly growing supermassive star because the infalling matter is very optically thick. The opacity is strongly scattering-dominated so there will be a colour correction to the effective temperature that hardens the spectrum, but if the infalling matter obscures the supermassive star then the true photosphere could be well outside the star’s nominal radius, leading to softer radiation. By using a blackbody with temperature T_* and radius R_* to estimate ionizing fluxes, we are effectively assuming that the infalling matter is distributed highly anisotropically in solid angle (e.g. due to rotation).

Because they are strongly dominated by radiation pressure, non-rotating supermassive stars have very small binding energies and can be destabilized by tiny corrections due to general relativity (Hoyle & Fowler 1963; Iben 1963a,b; Fowler 1966). However, a

relatively small amount of rotation (compared to breakup) can increase the binding energy dramatically, stabilizing stars as massive as $\sim 10^8 M_\odot$ (Fowler 1966). The gas falling on to supermassive stars is expected to have substantial, though sub-Keplerian, rotation, and we will conservatively assume that supermassive stars can exist up to masses of $10^7 M_\odot$, above which they undergo core collapse even if hydrogen is not exhausted.

2.2 Quasi-stars and seed black holes

It is probable that only a tiny fraction of the mass in the core of a supermassive star has low enough angular momentum to collapse directly to a BH. Any additional growth of the hole must be accompanied by the outward transport of angular momentum which, barring an efficient relief valve such as a polar jet piercing the star, implies the outward transport of a large amount of energy. The energy liberated by the formation and initial growth of the BH must be absorbed by the stellar envelope. While the BH is still small – perhaps only a few hundred solar masses – the liberated energy becomes comparable to the binding energy of the envelope, which therefore inflates by a large factor. It probably does not disperse, since that would cut off the BH growth and therefore the energy supply.

This leads to a structure dubbed a ‘QS’ (BVR2006; Begelman et al. 2008), which resembles a red giant or Thorne–Żytkow object (Thorne & Żytkow 1977) except that the power source for inflating the envelope comes from black hole accretion instead of nuclear shell burning around a compact core. As the envelope inflates, the core density declines, which reduces the BH growth rate and the energy feedback rate until a stable equilibrium is reached when the energy output from BH growth equals the Eddington limit for the total mass of the hole + envelope. The effective temperature of this bloated object is

$$T_{\text{QS}} \sim 5000 \epsilon_{-1}^{-1/5} m_{\text{BH},4}^{-2/5} m_{*,6}^{7/20} \text{ K}, \quad (4)$$

where the BH and QS masses are expressed in units of 10^4 and $10^6 M_\odot$, respectively. The parameter $\epsilon = 0.1 \epsilon_{-1}$ represents an efficiency factor for the conversion of BH growth into energy. In all of our models below, we adopt $\epsilon_{-1} = 1$; if some of the energy were able to escape without percolating through the envelope, it would imply a smaller ϵ and allow faster BH growth during the QS phase.

Since the BH grows more rapidly than the QS envelope, T_{QS} decreases with time. But it cannot drop below the temperature corresponding to the Hayashi (1961) track. For Pop III abundances, the minimum temperature is about 4000 K (which we adopt as the fiducial value); for near-solar enrichment, this number drops to ~ 3000 K or slightly lower. Unlike gas pressure-supported red giants, which remain stable as they hover close to the minimum temperature, QSs lose dynamical equilibrium and disperse once they reach this limit. This is because QSs are supported against collapse by radiation pressure, which also governs internal energy transport – implying that there are not enough degrees of freedom available to allow the envelope to adjust quasi-statically. From equation (4), we see that the BH mass cannot exceed about 1 per cent of the QS mass. For $\dot{m} > 3 \epsilon_{-1}^{8/9} T_{\text{c},8}^{14/9}$, this leads to a final QS mass that substantially exceeds the mass of its supermassive star precursor,

$$m_{*,6} \sim 1.4 \epsilon_{-1}^{4/9} \dot{m}^{8/9}, \quad (5)$$

and a BH mass (at the time of dispersal),

$$m_{\text{BH},4} \sim 3 \epsilon_{-1}^{-1/9} \dot{m}^{7/9}. \quad (6)$$

The lifetime of the QS phase is then

$$t_{\text{QS}} \sim 1.4 \epsilon_{-1}^{4/9} \dot{m}^{-1/9} \text{ Myr}. \quad (7)$$

For smaller infall rates, the final QS mass will be similar to the mass of the supermassive star (equation 2) and the BH mass will be $\lesssim 10^4 M_\odot$.

3 WHERE QUASI-STARS FORM: MODELLING THE HOST

We investigate the formation and evolution of QSs and MBHs via cosmological realizations of the merger hierarchy of dark matter haloes from early times to the present in a Λ cold dark matter (Λ CDM) cosmology [Wilkinson Microwave Anisotropy Probe 5 (WMAP5); Komatsu et al. 2009]. We track the merger history of 175 parent haloes with present-day masses in the range $10^{11} < M_h < 10^{15} M_\odot$ with a Monte Carlo algorithm (Volonteri et al. 2003). The mass resolution of our algorithm reaches $10^5 M_\odot$ at $z = 20$, and the most massive haloes are split into up to 600 000 progenitors. We also generate 12 Monte Carlo realizations of the merger hierarchy of a $M_h = 2 \times 10^{13} M_\odot$ halo at $z = 6$, starting from $z = 25$.

The formation of QSs requires large gas inflows, $\dot{M} \gtrsim 1 M_\odot \text{ yr}^{-1}$, within the inner region (sub-parsec scales) of a galaxy. Following BVR2006, we associate these inflows with global dynamical instabilities driven by self-gravity, the so-called bars-within-bars mechanism (Shlosman, Frank & Begelman 1989; Shlosman, Begelman & Frank 1990). Self-gravitating gas clouds become bar-unstable when the level of rotational support surpasses a certain threshold. A bar can transport angular momentum outwards on a dynamical time-scale via gravitational and hydrodynamical torques, allowing the radius to shrink. Provided that the gas is able to cool, this shrinkage leads to even greater instability, on shorter time-scales, and the process cascades. This mechanism works on a dynamical time ($\approx \text{Myr}$ for high-redshift galaxies) and can operate over many decades of radius (from hundreds of pc to sub-parsec scales).

Global bar-driven instabilities have now been observed in high-resolution numerical simulations of gas-rich galaxies (Levine et al. 2008; Wise, Turk & Abel 2008; Mayer et al. 2009; Regan & Haehnelt 2009). While BVR2006 suggested that fragmentation and global star formation are suppressed only if the gas temperature remains close to the virial temperature (e.g. due to lack of metals and molecular hydrogen), these simulations find strong inflows that occur before most of the gas fragments and forms stars also at solar metallicities, possibly due to the continuous generation of supersonic turbulence (Begelman & Shlosman 2009). In contrast to BVR2006, we therefore relax the assumption that QSs can be formed only out of zero-metallicity gas. Inspired by Mayer et al. (2009), we instead assume that inflows are triggered by gas-rich major mergers.

The gas content of galaxies is regulated by the competition between stellar wind/supernova feedback that depletes galaxies of gas and replenishment of the gas reservoir via mergers with gas-rich haloes. In particular, we assume that Pop III stars form in metal-free haloes with $T_{\text{vir}} > 2000$ K (Yoshida et al. 2006), where we model metal enrichment by the ‘high feedback, best guess’ model of Scannapieco, Schneider & Ferrara (2003). Scannapieco et al. (2003) model metal enrichment via pair-instability supernovae winds, by following the expansion of spherical outflows into the Hubble flow. They compute the comoving radius, at a given redshift, of an outflow from a population of supernovae that exploded at an earlier time. Using a modification of the Press–Schechter technique

(Scannapieco & Barkana 2002), they then compute the bivariate mass function of two haloes of arbitrary mass and collapse redshift, initially separated by a given comoving distance. From this function, they calculate the number density of supernovae host haloes at a given comoving distance from a ‘recipient’ halo of a given mass M_h that form at a given redshift z . By integrating over this function, one can calculate the probability that a halo of mass M_h forms from metal-free gas at a redshift z . When a halo forms in our merger tree, we calculate the probability that it is metal-free (hence, it can form Pop III stars) and determine if this condition is satisfied using Monte Carlo techniques. If a Pop III star forms in a halo, we conservatively assume that all gas is expelled, leading to a nil gas fraction, $f_{\text{gas}} = 0$. Gas is replenished only via mergers with gas-rich haloes, that is, haloes that have never experienced mass loss through Pop III outflows. When two haloes merge, we sum their gas masses. We allow QS formation only in mergers where the final gas fraction exceeds a specified threshold, $f_{\text{gas}} > f_{\text{thr}}$. We have varied f_{thr} between 0.025 and 0.25 and found that the model results are largely insensitive to the choice of f_{thr} .

When selecting the sites of QS formation, we additionally require that the merger remnant has very low angular momentum, which ensures the optimal conditions for gas infall (BVR2006). We parametrize the angular momentum of a dark matter halo with mass M_h via its halo spin parameter, $\lambda_{\text{spin}} \equiv J_h E_h^{1/2} / (GM_h^{3/2})$, where J_h is the total angular momentum and E_h is the binding energy. The exact angular momentum threshold below which inflows are triggered, λ_{thr} , is a free parameter that we constrain against observations (see Section 5).

D’Onghia & Navarro (2007) investigate the spins of haloes that are remnants of major mergers and compare them to the ‘global’ population using cosmological N -body simulations. They find that haloes that are still unrelaxed after a major merger tend to have higher-than-average spins. They quantify the spin parameter distributions for post-merger haloes and relaxed haloes. In both cases the distributions are lognormal, but with different parameters: $\bar{\lambda}_{\text{spin}} = 0.028$ and $\sigma_\lambda = 0.58$ for relaxed haloes while $\bar{\lambda}_{\text{spin}} = 0.04$ and $\sigma_\lambda = 0.65$ for unrelaxed systems. When a halo forms, we pick its spin parameter from the distribution for relaxed haloes. After a major merger, we pick a new λ_{spin} from the unrelaxed lognormal distribution and compare it to λ_{thr} . If the halo has a spin parameter below the threshold, we consider the halo a candidate for QS formation.²

For haloes that meet both criteria, $f_{\text{gas}} > f_{\text{thr}}$ and $\lambda_{\text{spin}} < \lambda_{\text{thr}}$, we calculate the gas inflow rate as $\dot{M} = v_c^3 / G$, where v_c is the circular velocity.³ If $\dot{M} > 1 M_\odot \text{ yr}^{-1}$, then we model supermassive star, QS and seed BH formation as described in Section 2. We limit the mass of the forming QS to less than the mass of available gas, $f_{\text{gas}} M_h$.

² The spin parameter for the gaseous component, λ_{gas} , can be different from λ for the halo. Gottlöber & Yepes (2007) find, however, that statistically $\lambda_{\text{gas}} / \lambda_{\text{halo}} \simeq 1.4$ and that the tendency is for this ratio to decrease towards unity as redshift increases. We expect therefore that statistically our approach is valid.

³ A halo of mass M_h collapsing at redshift z has a circular velocity

$$v_c = 142 \left[\frac{M_h}{10^{12} M_\odot} \right]^{1/3} \left[\frac{\Omega_m}{\Omega_m^z} \frac{\Delta_c}{18\pi^2} \right]^{1/6} (1+z)^{1/2} \text{ km s}^{-1}, \quad (8)$$

where Δ_c is the overdensity at virialization relative to the critical density. For a *WMAP5* cosmology, we adopt here the fitting formula (Bryan & Norman 1998) $\Delta_c = 18\pi^2 + 82d - 39d^2$, where $d \equiv \Omega_m^z - 1$ is evaluated at the collapse redshift, so that $\Omega_m^z = \Omega_m(1+z)^3 / [\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2]$.

From stability considerations, we also assume that the supermassive star that precedes the QS undergoes core collapse if its mass exceeds $10^7 M_\odot$ even if it has not exhausted its core hydrogen.

We further assume that if a seed MBH is already present, QS formation is suppressed. This criterion comes from the fact that any QS with a BH mass exceeding about 1 per cent of the envelope mass will violate the minimum temperature condition and be dispersed by radiation pressure (Begelman et al. 2008). Thus, it is very difficult to lay down a massive envelope of gas around an existing, naked MBH.

Summarizing, our model assumptions are as follows:

- (1) inflow triggered by gas-rich major mergers (mass ratio between 1:10 and 1:1);
- (2) inflow if gas fraction $f_{\text{gas}} > f_{\text{thr}}$ and spin parameter $\lambda_{\text{spin}} < \lambda_{\text{thr}}$;
- (3) central mass accretion rate $\dot{M} = v_c^3 / G$;
- (4) if $\dot{M} > 1 M_\odot / \text{yr}$, then QS and seed MBH form, with masses given by
 - (i) $M_{\text{QS}} = 1.4 \times 10^6 \epsilon_{-1}^{4/9} \dot{m}^{8/9} M_\odot$
 - (ii) $M_{\text{BH}} = 3 \times 10^4 \epsilon_{-1}^{-1/9} \dot{m}^{7/9} M_\odot$;
- (5) QS formation suppressed if MBH already present in halo.

We compare two models, one in which QSS are the only route to MBH seed formation and the other in which Pop III remnants offer a complementary channel. In both cases, we assume that Pop III stars form in metal-free haloes with $T_{\text{vir}} > 2000 \text{ K}$.⁴ In the QS-only scenario, we assume that Pop III stars form but leave no massive remnant (that is, the star mass is below $260 M_\odot$; Fryer, Woosley & Heger 2001). Therefore, Pop III stars in this model only provide stellar wind/supernova feedback that depletes galaxies of gas. In the Pop III remnants+QS scenario, we assume a logarithmically flat initial mass function, $dN/d \log M = \text{constant}$, between 10 and $600 M_\odot$, where the upper limit comes from Omukai & Palla (2003), and suppose that seed MBHs form when the progenitor star is in the mass range of 40–140 or 260–600 M_\odot (Fryer et al. 2001). The remnant mass is taken to be one-half the mass of the star. Once a seed MBH forms, we subsequently treat it in exactly the same way, regardless of its formation mechanism.

4 HOW BLACK HOLE SEEDS GROW: MODELLING THE COSMIC EVOLUTION

We base our model for MBH mass growth on a set of simple assumptions, supported by both simulations of quasar triggering and feedback (Springel, Di Matteo & Hernquist 2005), and analysis of the relationship between MBH masses (M_{BH}) and the properties of their hosts in today’s galaxies (Ferrarese & Merritt 2000; Gebhardt et al. 2000; Ferrarese 2002). The main features of the models have been discussed elsewhere (Volonteri et al., 2003; Volonteri & Natarajan 2009, and references therein). We summarize in the following the relevant assumptions. MBHs in galaxies undergoing a major merger (i.e. having a mass ratio of $> 1:10$) accrete mass and

⁴ We neglect the effect of a redshift-dependent Lyman–Werner background that progressively increases the minimum mass necessary for Pop III star formation (Trenti & Stiavelli 2009). As shown below, the differences in the numbers of forming QSS and MBHs, for the cases with and without Pop III remnants, are negligible. If the Lyman–Werner background reduces the number of forming Pop III stars, the result will fall in-between the two cases considered here.

become active. Each MBH accretes an amount of mass, $\Delta M = 9 \times 10^7 (\sigma/200 \text{ km s}^{-1})^{4.24} M_\odot$, where σ is the velocity dispersion after the merger. This relationship scales with the $M_{\text{BH}}-\sigma$ relation, as it is seen today (Gültekin et al. 2009):

$$M_{\text{BH}} = 1.3 \times 10^8 \left(\frac{\sigma}{200 \text{ km s}^{-1}} \right)^{4.24} M_\odot; \quad (9)$$

the normalization in ΔM was chosen to take into account the contribution of mergers, without exceeding the mass given by the $M_{\text{BH}}-\sigma$ relation.

We link the correlation between the BH mass and the central stellar velocity dispersion of the host with the empirical correlation between the central stellar velocity dispersion and the asymptotic circular velocity (v_c) of galaxies (Ferrarese 2002; see also Baes et al. 2003; Pizzella et al. 2005):

$$\sigma = 200 \left(\frac{v_c}{304 \text{ km s}^{-1}} \right)^{1.19} \text{ km s}^{-1}. \quad (10)$$

The latter is a measure of the total mass of the dark matter halo of the host galaxy. We calculate the circular velocity from the mass of the host halo and its redshift.

The rate at which mass is accreted scales with the Eddington rate for the MBH, and we set a fixed Eddington ratio of $f_{\text{Edd}} = 0.3$, based on observations of quasars (Kollmeier et al. 2006). This is also the strong upper limit for radiatively efficient spherical accretion on to Pop III remnants (Milosavljević et al. 2009). Accretion starts after a halo dynamical time-scale and lasts until the MBH, of initial mass M_{in} , has accreted ΔM . The lifetime of an active galactic nucleus (AGN) therefore depends on how much mass it accretes during each episode:

$$t_{\text{AGN}} = \frac{t_{\text{Edd}}}{f_{\text{Edd}}} \frac{\epsilon}{1 - \epsilon} \ln(M_{\text{fin}}/M_{\text{in}}), \quad (11)$$

where ϵ is the radiative efficiency ($\epsilon \simeq 0.1$), $t_{\text{Edd}} = 0.45 \text{ Gyr}$ and $M_{\text{fin}} = \min [(M_{\text{in}} + \Delta M), 1.3 \times 10^8 (\sigma/200 \text{ km s}^{-1})^{4.24} M_\odot]$.

We further assume that when two galaxies hosting MBHs merge, the MBHs themselves merge within the merger time-scale of the host haloes, which is a plausible assumption for MBH binaries formed after gas-rich galaxy mergers (Dotti et al. 2007, and references therein). We adopt the relations suggested by Taffoni et al. (2003) for the merger time-scale.

5 MODEL RESULTS: QUASI-STARS AND BLACK HOLES

5.1 Constraining the quasi-star and black hole formation efficiency

As described in Section 3, our model depends mostly on one parameter, λ_{thr} . We constrain λ_{thr} from below by using the number density of $z \simeq 6$ quasars. We obtain a (less certain) upper limit on λ_{thr} by estimating the maximum permitted contribution of supermassive stars to reionization, and later show (in Section 5.3) that this constraint is consistent with the Sołtan (1982) constraint on integrated quasar light.

The population of $z \simeq 6$ quasars probes the high-luminosity end of the luminosity function, roughly $L \geq 10^{47} \text{ erg s}^{-1}$. Recent surveys (and references therein Jiang et al. 2009) yield a number density of these luminous sources, powered by billion-solar-mass MBHs (assuming sub-Eddington luminosities and negligible beaming), of $\sim 1 \text{ Gpc}^{-3}$. This comoving density roughly corresponds to the number density of haloes with a mass of $\lesssim 10^{13} M_\odot$ in a *WMAP5* cosmology. With the scalings described in Section 4, we can express

the mass of a halo hosting a 10^9-M_\odot MBH as follows:

$$M_{\text{h}} = 5 \times 10^{13} \left[\frac{M_{\text{BH}}}{10^9 M_\odot} \right]^{0.84} \left[\frac{\Omega_{\text{m}}}{\Omega_{\text{m}}^z} \frac{\Delta_c}{18\pi^2} \right]^{-1/2} (1+z)^{-3/2} M_\odot, \quad (12)$$

which is in basic agreement with the number-density-based estimate. If Pop III remnants, with a mass of $\sim 100 M_\odot$, provided the only route to MBH formation, equation (11) would imply that for an Eddington ratio of 0.3 and a radiative efficiency of 0.1, $t_{\text{AGN}} = 2.68 \text{ Gyr}$ is needed for an MBH to reach $10^9 M_\odot$. This is more than twice the age of the Universe at $z = 6$. Even decreasing ϵ to 0.05-0.06 (appropriate for either a Schwarzschild BH or a BH fed by a counter-rotating accretion disc) does not bring the time-scale within an acceptable range ($< 0.95 \text{ Gyr}$ in a *WMAP5* cosmology). For Pop III remnants to reach billion solar masses by $z = 6$, very specific conditions must be met: accretion occurs *continuously* and at rates *close to* $f_{\text{Edd}} = 1$, if not higher. If either of these conditions fails, Pop III remnants alone cannot reach the required MBH masses. Clearly, our approach here is very conservative. The rate at which MBHs in the early Universe accrete is unknown, and it is possible that the most massive high-redshift MBHs may actually be fed at the Eddington limit (or above; Volonteri & Rees 2005). One open question is the role of feedback, if accretion is radiatively efficient. Pelupessy, Di Matteo & Ciardi (2007), Alvarez, Wise & Abel (2009) and Milosavljević et al. (2009) suggest that the growth of the MBH is severely limited by thermal feedback to rates not higher than 30 per cent of the Eddington rate.

On the other hand, at very high accretion rates, the excess radiation might be effectively trapped, and a BH accreting at super-critical values does not necessarily radiate at super-Eddington luminosity (Begelman 1979; Begelman & Meier 1982). Also, when the accretion rate is high, the formation of collimated outflows is common (e.g. blazars). Such collimated jets may not cause feedback directly on the host (which is pierced through), but deposit their kinetic energy at large distances, leaving the host unscathed (in a different context, see Vernaleo & Reynolds 2006). This is likely if at large accretion rates photon trapping decreases the disc luminosity, while concurrently the presence of a jet helps to remove angular momentum, thus promoting efficient accretion (Volonteri, Ghisellini & Haardt, in preparation). This is supported by the strong cosmological evolution of blazars detected by the Burst Alert Telescope onboard the *Swift* satellite (Ajello et al. 2009). The derived luminosity function indicates that beamed, jetted sources become dominant at the highest MBH masses ($> 10^9 M_\odot$) and redshifts ($z > 5$), when compared to the cosmic evolution of radio-quiet sources (Ghisellini et al. 2010).

Fig. 1 shows the mass growth of the central BH in the main halo of merger trees corresponding to $M_{\text{h}} = 2 \times 10^{13} M_\odot$ haloes at $z = 6$. We see that the requirement set by $z = 6$ quasars, powered by 10^9-M_\odot MBHs, requires λ_{thr} to be larger than 0.01.

The case $\lambda_{\text{thr}} = 0.01$ is only marginally consistent with the constraints. This is because the probability of MBH formation in a given halo is less than 10^{-2} , given the functional form of the spin parameter probability distribution. A halo needs $\simeq 100$ major mergers to have a probability of MBH formation close to unity (provided, of course, that v_c is large enough to give the required \dot{M}). In the 12 realizations of $M_{\text{h}} = 2 \times 10^{13} M_\odot$ haloes at $z = 6$, we have not found a single case in which the main halo of the tree forms an MBH if $\lambda_{\text{thr}} = 0.01$. The MBH hosted in the central halo is instead acquired during a merger with another halo hosting a central MBH. This probability is lower if MBH formation is less efficient (lower

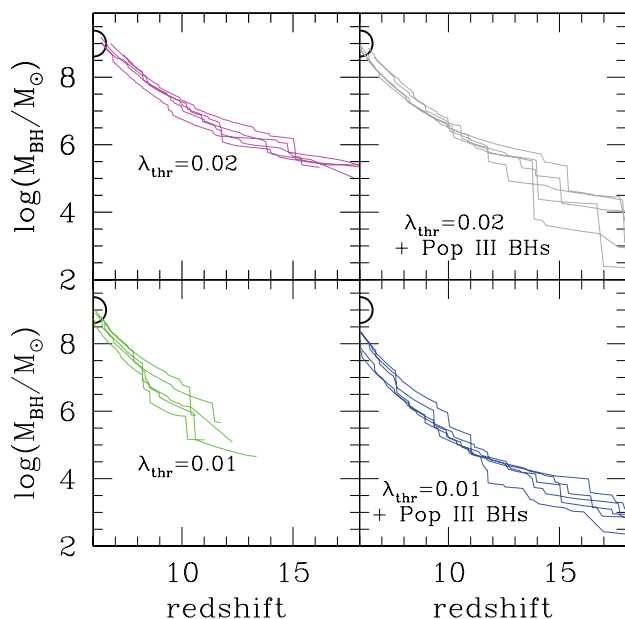


Figure 1. Growth of the central BH in the main halo of merger trees corresponding to $M_h = 2 \times 10^{13} M_\odot$ haloes at $z = 6$ (we show here only five realizations out of 12 for clarity). Lower panels: relatively low-efficiency QS/supermassive star formation case ($\lambda_{\text{thr}} = 0.01$). Upper panels: efficient QS/supermassive star formation case ($\lambda_{\text{thr}} = 0.02$). Left-hand panels: Pop III stars leave no remnant that can impede QS and MBH seed formation. Right-hand panels: Pop III stars leave remnant MBHs. The big circle is centred at $z = 6$ and $M_{\text{BH}} = 10^9 M_\odot$.

λ_{thr}), leading to the main halo acquiring an MBH only at relatively late cosmic times ($z \simeq 12$ – 14), thus postponing its growth. When Pop III stars leave behind massive remnants, the situation worsens. Since we assume that QS formation is suppressed in the presence of an MBH, even fewer haloes can form massive seeds that can accelerate the growth of an MBH by boosting the merger rate.⁵ $z = 6$ quasars therefore provide a lower limit to QS and seed formation efficiency: λ_{thr} must be larger than 0.01.

We estimate the possible contribution of supermassive stars to hydrogen reionization as follows. Since the hydrogen recombination time-scale at $z > 6$ is shorter than the then-Hubble time, to derive an upper limit to the contribution to reionization we can integrate the number of ionizing photons over time, that is, the sum of all the ionizing photons emitted by all the forming supermassive stars. To estimate the number of ionizing photons emitted by a single supermassive star, we assume the photon output of a blackbody with temperature and radius given by T_* and R_* , respectively, from Section 2, and multiply the output rate by the main-sequence lifetime t_{MS} (equation 1). Since the mass function of QSs – and the preceding supermassive stars – peaks at $\simeq 10^6 M_\odot$ (see Section 5.2), we calculate the photon output for $m = 1$ and use $T_{c,8} = 1.4$ as before.

⁵ Here, we have ignored the effect of ‘gravitational recoils’ on the MBH population. For MBHs merging in a gas-rich environment, as expected for these high-redshift systems, Volonteri, Gultekin & Dotti (2010) have shown that the probability of ejection is negligible. This is because the accreting gas exerts gravitomagnetic torques that suffice to align the spins of both MBHs with the angular momentum of the large-scale gas flow, leading to aligned spin-orbit configurations that yield recoil velocities of $< 100 \text{ km s}^{-1}$. Even for the ‘worst-case scenario’ of isotropically distributed spins, the probability of ejection for MBH binaries in the massive haloes which host very high-redshift quasars effectively drops to zero at $z \simeq 13$ – 19 .

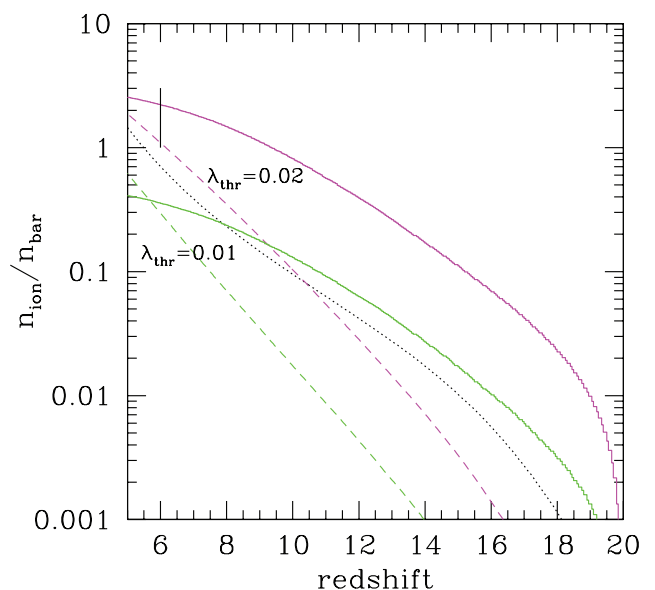


Figure 2. Number of reionizing photons per baryon for two QS formation efficiencies, $\lambda_{\text{thr}} = 0.02$ and $\lambda_{\text{thr}} = 0.01$. Cases where Pop III remnants offer an additional channel for MBH formation yield almost identical results. Solid curves: supermassive star contribution to the reionizing budget. Dashed curves: contribution from quasars associated with QS remnants where we assume a classic multicolour disc spectrum up to $kT_{\text{max}} \sim 1 \text{ keV}$ (M_{BH}/M_\odot)^{-1/4} (Shakura & Sunyaev 1973), and a non-thermal power-law component $L_\nu \propto \nu^{-\alpha}$, with $\alpha \approx 1$ at higher energies. With this choice, an MBH with a mass of $\simeq 10^4 M_\odot$ radiates 20 per cent of its bolometric power as hydrogen-ionizing photons, with $\langle E_\gamma \rangle = 140 \text{ eV}$ (for details on the implementation, see Volonteri & Gnedin 2009). Dotted curve: contribution from normal stars (from Gnedin 2008).

This gives an ionizing photon output of 1.6×10^{67} ionizing photons per supermassive star.

The total number density of ionizing photons emitted by the supermassive star population can then be approximated as

$$n_{\text{ion}}(z) \approx 1.6 \times 10^{67} \sum_{i=0, j(z)} n_{\text{SMS}, i} \text{ ph Mpc}^{-3}, \quad (13)$$

where the sum is over the number density of supermassive stars (n_{SMS}) that form during each time-step between our starting redshift ($z_{\text{max}} = 20$, time-step 0) and the time-step $j(z)$ corresponding to the redshift of interest, z .

We compare this number density of photons to the number density of baryons:

$$n_{\text{bar}} = f_{\text{bar}} \Omega_{\text{m}} \frac{\rho_{\text{crit}}}{\langle m \rangle} \simeq 7.4 \times 10^{66} \text{ Mpc}^{-3}, \quad (14)$$

where $f_{\text{bar}} \simeq 0.17$, $\Omega_{\text{m}} \simeq 0.3$, $\rho_{\text{crit}} = 9.2 \times 10^{-30} \text{ g cm}^{-3}$ and $\langle m \rangle \simeq 1.7 \times 10^{-24} \text{ g}$. Therefore,

$$\frac{n_{\text{ion}}(z)}{n_{\text{bar}}} = \frac{1.6}{0.74} \times \sum_{i=0, j(z)} n_{\text{SMS}, i}. \quad (15)$$

This calculation is shown in Fig. 2. If we consider a reionization criterion $1 < n_{\text{ion}}/n_{\text{bar}} < 3$ at $z = 6$ (Gnedin 2008), we see that $\lambda_{\text{thr}} > 0.02$ tends to overproduce ionizations, provided that the ionizing radiation escapes from the halo core. As noted in Section 2, this is far from certain because the gas falling on to the supermassive star is optically thick. Our estimate of the ionization rate should therefore be regarded as an upper limit, corresponding to the case where the photospheric radiation of the supermassive star is unobscured (e.g. as the result of disc-like inflow) or at least not degraded. The

ionization rate from supermassive stars therefore provides a weak constraint. However, we note that an additional limit on QS and BH formation, discussed in Section 5.3 (Soltan’s argument), reinforces the limits discussed in this section. In the following, we assume $0.01 < \lambda_{\text{thr}} < 0.02$ and show results for $\lambda_{\text{thr}} = 0.01$ (low efficiency) and $\lambda_{\text{thr}} = 0.02$ (high efficiency).

The quasar population that results from MBHs formed in QSs does not contribute significantly to hydrogen reionization (see Fig. 2) for the low-efficiency case, while the high-efficiency case produces many more MBHs that produce significant ionizing radiation. High-efficiency MBH formation ($\lambda_{\text{thr}} = 0.02$) is also constrained to be an upper limit by the contribution of quasars to helium reionization. Taking into consideration ionization energy thresholds for He II⁶ and H (54.4 versus 13.6 eV) and number densities of the two species (12 times more H than He), there are roughly three times more He ionizing photons per He atom than photons above 13.6 eV per H atom for our quasar spectrum. As He recombines $\simeq 6$ times faster than H, the number of ionizing photons per He atom emitted in one recombination time is only roughly half of the corresponding value for H, and there will be just a small delay between H and He reionizations. In the low-efficiency model He reionization occurs at $z \simeq 3$, while in the high-efficiency case He is reionized at $z \simeq 3.5\text{--}5$ (depending on the clumpiness of the intergalactic medium and the escape fraction of ionizing photons). Although observational constraints on He reionization are still weak, they suggest that He II reionization occurs at $z \sim 3\text{--}4$ (Faucher-Giguère et al. 2008, and references therein). Hence, while the low-efficiency model is consistent with He reionization, the high-efficiency case clearly represents a strong upper limit to MBH seed production, as, if MBHs accrete as efficiently as assumed here, He is reionized at much higher redshift. We note that hard photons from supermassive stars do not contribute much to He reionization (the number of ionizing photons for the He atom is roughly one-third of the number of ionizing photons for the H atom shown in Fig. 2).

5.2 Mass functions and formation times of quasi-stars

Fig. 3 shows the mass function of QSs, integrated over all redshifts of formation. The mass function peaks at $\simeq 10^6 M_{\odot}$, corresponding to the minimum gas infall rate $\dot{M} = v_c^3/G = 1 M_{\odot} \text{ yr}^{-1}$. Since the infall rate is related to the circular velocity of the host halo, it is also linked to the mass of the halo. The halo mass function in a Λ CDM cosmology is a steep function of the halo mass; the QS mass function therefore peaks at the lowest allowed mass ($\simeq 10^6 M_{\odot}$).

A QS ends its life with a photospheric temperature of about 4000 K for Pop III abundances (which we adopt as the fiducial value). We model the emission as a blackbody at $T_{\text{QS}} = 4000$ K for the entire lifetime of the QS (equation 7) and calculate number counts in the 2–10 μm band of the *James Webb Space Telescope* (*JWST*), assuming a sensitivity of 10 nJy at 2 μm . These are shown in Fig. 4 for our reference models. In the high-efficiency case ($\lambda_{\text{thr}} = 0.02$), *JWST* could detect up to a few QSs per field at $z \simeq 5\text{--}10$. In each panel of Fig. 4, the lower curves show the expected number counts for fields at the depth of the *Hubble Ultra Deep Field* (HUDF), observed with the Wide Field Camera 3 (WFC3) onboard the *Hubble Space Telescope* in the F160W band, to an AB magnitude limit of 28.8 (Bouwens et al. 2010). We conclude that it is possible that QSs can already be seen in deep WFC3 images, although it is not very likely.

⁶ The ionization threshold for He I, 24.6 eV, is low enough that He I and H reionization are expected to occur almost concurrently.

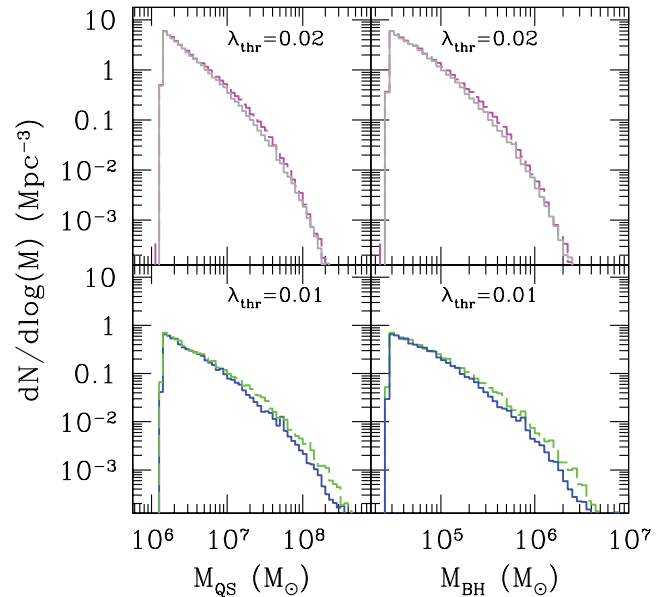


Figure 3. Left: mass function of QSs, integrated over all redshifts of formation. Right: mass function of seed BHs. Lower set of histograms: relatively low-efficiency QS formation case ($\lambda_{\text{thr}} = 0.01$). Upper set of histograms: efficient QS formation case ($\lambda_{\text{thr}} = 0.02$). Dashed curves: Pop III stars leave no remnant that can impede QS and MBH seed formation. Solid curves: Pop III stars leave remnant MBHs.

Although the peak of QS formation occurs at $z > 5$, some QSs form at redshift as low as $z \simeq 2$. These late bloomers form in haloes with relatively low cosmic bias. We recall that the bias of a halo is the number of standard deviations that the critical collapse overdensity represents on the halo mass scale, M_h : $\nu_c = \delta_c(z)/\sigma(M_h, z)$. At any redshift, we can identify the characteristic collapsing mass (i.e. $\nu_c = 1$) and its multiples. The higher the ν_c , the more massive and rarer the halo, and the higher its bias and clustering strength. Fig. 5 exemplifies our results for two cases: a Milky Way-sized halo ($M_h = 2 \times 10^{12} M_{\odot}$ at $z = 0$, lower bias) and a cluster-size halo ($M_h = 10^{15} M_{\odot}$ at $z = 0$, higher bias). We therefore expect that the latest forming QSs should be found in the field rather than in a high-density environment. As one goes to lower and lower galaxy masses, the number of haloes where QSs can form decreases, as the requirement $\dot{M} \propto v_c^3 > 1 M_{\odot} \text{ yr}^{-1}$ selects only haloes with mass larger than $\sim 10^9 M_{\odot}$ as possible QS/MBH hosts. Given that the probability of QS/MBH formation in a given halo is 1–10 per cent (due to the lognormal distribution of spin parameters), and that major mergers between $\sim 10^9\text{--}M_{\odot}$ haloes are very rare events in the merger history of galaxies with masses below $\sim 10^{11} M_{\odot}$, we expect dwarf galaxies to be rarely seeded with MBHs (results for the frequency of MBHs in dwarf galaxies are very similar to those found in Van Wassenhove et al. 2010). In order to check whether the infrequency of QS/MBH formation at low halo masses is compensated by the larger abundance of small galaxies, we performed 150 merger trees each for haloes with a mass of 5×10^9 , 10^{10} and $5 \times 10^{10} M_{\odot}$. We find that formation of QS/MBHs becomes statistically negligible for haloes with mass below $10^{10} M_{\odot}$ and that QSs in low-mass galaxies do not dominate the overall population.

Fig. 3 also shows the mass function of seed BHs, integrated over all redshifts of formation. The mass function peaks at $\simeq 10^4 M_{\odot}$, but rare supermassive seeds, with masses of $\simeq 10^6 M_{\odot}$, are possible. The more efficient QS and BH formation is, the steeper

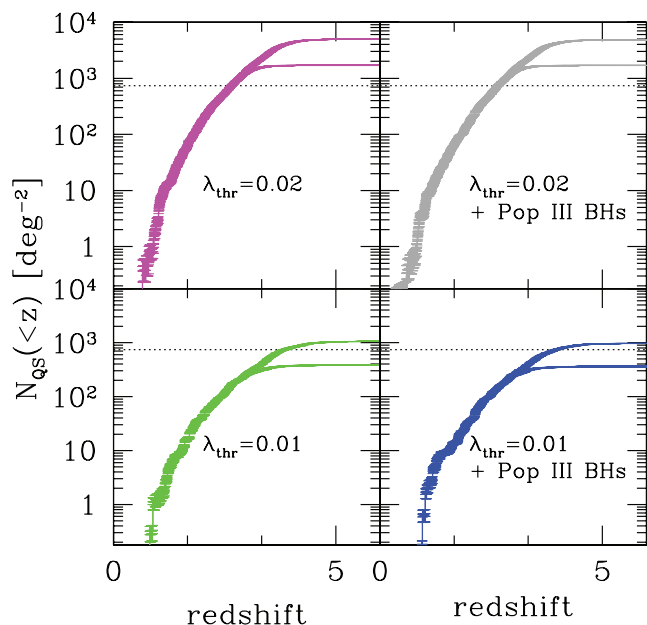


Figure 4. Number counts of Qs for a *JWST* sensitivity of 10 nJy at 2 μm (upper curves) and for the HUDF WFC3/infrared in the F160W band, to an AB magnitude limit of 28.8 (Bouwens et al. 2010; lower curves). We assume $T_{\text{QS}} = 4000$ K. Lower panels: relatively low-efficiency QS formation case ($\lambda_{\text{thr}} = 0.01$). Upper panels: efficient QS formation case ($\lambda_{\text{thr}} = 0.02$). Left-hand panels: Pop III stars leave no remnant that can impede QS and MBH seed formation. Right-hand panels: Pop III stars leave remnant MBHs. The dotted horizontal line shows the limit of one source per *JWST* field (2.2×2.2 arcmin 2).

the mass function. This is because more seeds form at the highest redshifts when the hosts have relatively low mass and low circular velocity.

5.3 Integrated properties of the black hole population

Fig. 6 shows the number density of MBHs: depending on the efficiency of QS and MBH formation (λ_{thr}), the number density can be dominated by either Pop III or QS MBHs. In the low-efficiency case ($\lambda_{\text{thr}} = 0.01$) the number density of MBHs is dominated by the Pop III channel, if available; the opposite is true for the high-efficiency case. Note how the light Pop III remnant seeds are much more often actively accreting (compare left-hand and right-hand panels). This is an effect of imposing self-regulation of the MBH mass, i.e. in a given accretion episode an MBH cannot grow to exceed the $M_{\text{BH}}-\sigma$ relation (Volonteri & Natarajan 2009). Since QS MBHs are born *above* the expected $M_{\text{BH}}-\sigma$ relation for their host, accretion is effectively damped. Light Pop III remnants, instead, are *undermassive* and accrete steadily.

The number density of MBHs at $z = 0$ can be compared to the number density of galaxies, using as a constraint, for instance, the galaxy luminosity function as measured in the Sloan Digital Sky Survey (e.g. Benson et al. 2007). At stellar masses of $\sim 10^8 M_{\odot}$ the galaxy number density is $\sim 0.3\text{--}1 \text{ Mpc}^{-3}$, which is well above our MBH density at $z = 0$ (see also fig. 11 in Carlberg, Sullivan & Le Borgne 2009).

In contrast to the number density of MBHs, the mass density is dominated by QS MBHs (Fig. 7). We show both the mass density locked into MBHs and the cumulative mass density accreted on to MBHs (related to the Sołtan constraint: Sołtan 1982). Sołtan’s argument offers an additional constraint on our models. We com-

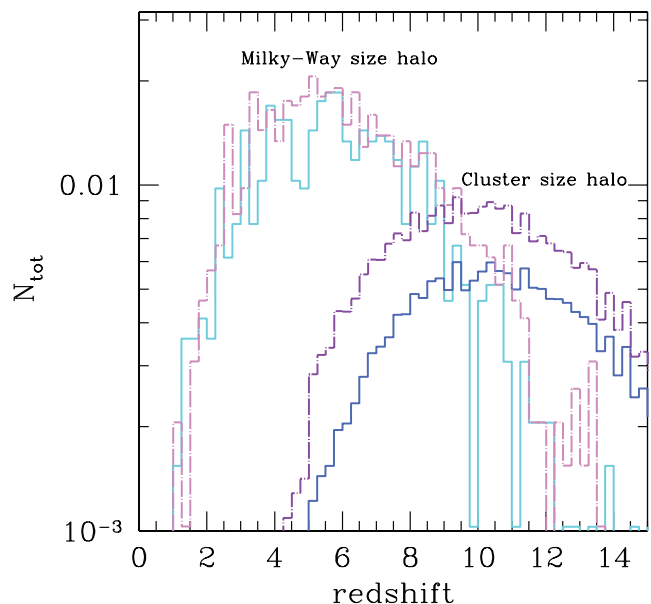


Figure 5. Formation times of Qs for two differently biased systems: an Milky Way-size halo and a cluster-size halo. Dot-dashed curves: Pop III stars leave no remnant that can impede QS and MBH seed formation. Solid curves: Pop III stars leave remnant MBHs. Here N_{tot} is normalized to a comoving cubic Mpc.

pare the cumulative mass density accreted on to MBHs with that derived from the Hopkins, Richards & Hernquist (2007) bolometric luminosity function of quasars:

$$\rho_{\text{QSO}}(<z) = \frac{(1-\epsilon)}{\epsilon c^2} \int_0^z \int \frac{L' \Phi(L', z)}{H_0(1+z) \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} dL' dz, \quad (16)$$

where the mass accretion rate, $\dot{M}_{\text{acc}} = L\epsilon^{-1}c^{-2}$, is converted into MBH mass growth, $\dot{M}_{\text{BH}} = (1-\epsilon)L\epsilon^{-1}c^{-2}$, with ϵ being the energy conversion coefficient. The boundaries assume a minimum radiative efficiency $\epsilon = 0.06$ and a maximum radiative efficiency $\epsilon = 0.3$ (Schwarzschild to rapidly rotating Kerr BH). All our models fit comfortably within the allowed range. We see from Fig. 7 that Sołtan’s argument provides an additional constraint to the upper limit on MBH formation via QS seeds. If the QS formation efficiency were much larger than $\lambda_{\text{thr}} = 0.02$, the cumulative accreted mass density at $z = 4$ would be overestimated.

Note that the value of the mass density locked into MBHs at $z = 0$, $\rho_{\text{BH},0} = (3.2\text{--}5.4) \times 10^5 M_{\odot} \text{ Mpc}^{-3}$ (e.g. Merloni & Heinz 2008, and references therein) is another independent constraint that can be evaluated from Fig. 7. The low-efficiency case is only marginally consistent with $\rho_{\text{BH},0}$. For comparison, we show here a case of MBH evolution that considers the Pop III remnant channel only. This is also hardly consistent with $\rho_{\text{BH},0}$ and Sołtan’s argument. If BH seeds are rare or very light, a constant $f_{\text{Edd}} = 0.3$ is insufficient to explain the growth of the BH population and its current demography.

The mass function of MBHs is shown in Fig. 8. We also show the mass function predicted by Merloni & Heinz (2008), derived using a completely different technique. Merloni & Heinz (2008) solve the continuity equation for the MBH mass function using the $z = 0$ one as a boundary condition and the luminosity function as a tracer of the MBH growth rate. There is no a priori reason to expect that enforcing the $M_{\text{BH}}-\sigma$ relation and assuming a fixed Eddington rate would match the $z > 0$ mass function of MBHs

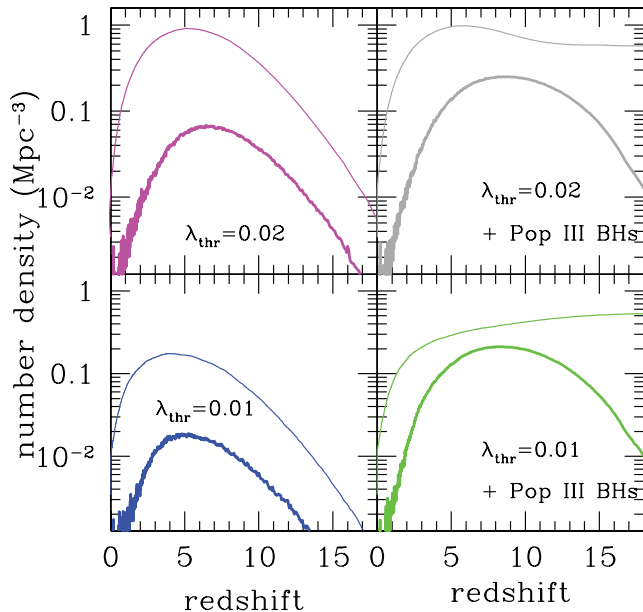


Figure 6. Number density of MBHs versus redshift, integrated over all MBH masses. Lower panel: relatively low-efficiency QS formation case ($\lambda_{\text{thr}} = 0.01$). Upper panel: efficient QS formation case ($\lambda_{\text{thr}} = 0.02$). Left-hand panels: Pop III stars leave no remnant that can impede QS and MBH seed formation. Right-hand panels: Pop III stars leave remnant MBHs. Thick curves (the lowest curve in each set): active MBHs only. Note that we do not include information regarding lifetime or a luminosity threshold. Depending on the efficiency of QS and MBH formation (λ_{thr}), the number densities can be dominated by Pop III or QS MBHs (e.g. QS MBHs always dominate in the $\lambda_{\text{thr}} = 0.02$ case, while in the $\lambda_{\text{thr}} = 0.01$ case Pop III remnants can dominate the MBH population *in number*).

predicted by Merloni & Heinz (2008). The very good agreement at masses of $>10^7 M_{\odot}$ suggests that, albeit very simple, our accretion prescription is well motivated. The disagreement at low masses could be due to incompleteness in the Merloni–Heinz inventory of low-mass MBHs, since their model works backwards from $z = 0$, and it does not assume an initial seed mass below which an MBH mass cannot decrease. If the only channel of MBH formation is via QS seeds, then the mass function cuts off below $\log(M_{\text{BH}}/M_{\odot}) = 4.5$, where the seed mass function drops as well. If Pop III remnants offer an alternative route to MBH formation, then the mass function is double-peaked, with each peak corresponding to a different mechanism for the formation of seeds. The peak at $\log(M_{\text{BH}}/M_{\odot}) \simeq 4.5$ becomes more pronounced with a higher efficiency of QS (and resulting seed BH) formation. Additional factors, such as mass and redshift-dependent accretion rates (we used a simple fixed-Eddington rate model) and obscuration, will modulate the exact shape of the mass function of observable MBHs.

6 DISCUSSION AND CONCLUSIONS

We have used a merger-tree approach to estimate the rate at which supermassive ($\gtrsim 10^6 M_{\odot}$) stars might have formed as a function of redshift, and the statistical properties of the resulting Qs and seed BHs. The key to the formation of supermassive stars is a large gas infall rate ($\dot{M} \gtrsim 1 M_{\odot} \text{ yr}^{-1}$), driven by global gravitational instability in the potential of a recently merged halo. We use the analyses presented by Begelman et al. (2008) and Begelman (2010) to estimate the properties of the resulting supermassive star, QS and seed BH as a function of this rate.

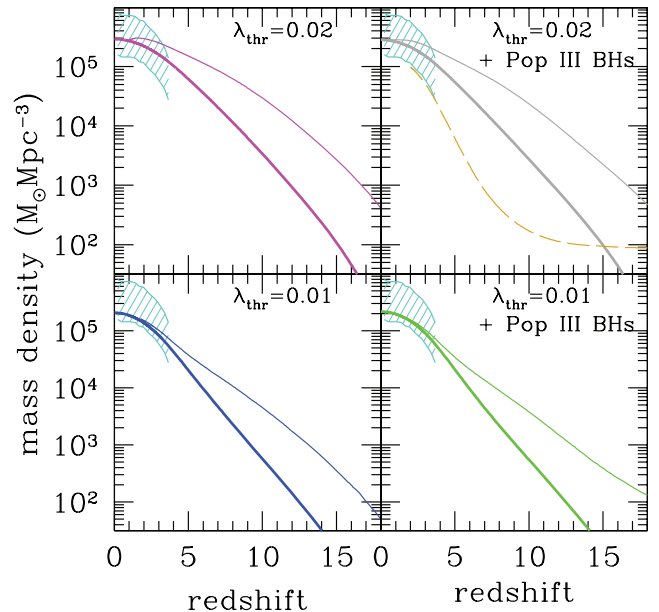


Figure 7. Mass density of MBHs versus redshift. Panels and lines are as in Fig. 6. The thick line in each panel shows the integral over time of the accreted mass (cumulative accreted mass density). The dashed curve in the upper-right panel shows the total mass density in MBHs when we consider Pop III remnants only. Hatched area: Sołtan’s constraint from the luminosity function of quasars using the Hopkins et al. (2007) bolometric luminosity function. The boundaries assume a minimum radiative efficiency $\epsilon = 0.06$ and a maximum $\epsilon = 0.3$ (Schwarzschild to rapidly rotating Kerr MBH).

We relate the properties of the seed BHs to the observable distributions of active and quiescent BHs at different redshifts, by applying a simple set of rules for their subsequent growth by accretion and mergers. We compare our direct collapse seed model to a pure Pop III seed model and also study the case in which both channels for MBH formation operate simultaneously.

Our results can be summarized as follows.

- (i) A limited range of supermassive star/QS/MBH formation efficiencies exists that allows one to reproduce observational constraints (the density of $z = 6$ quasars, the cumulative mass density accreted on to MBHs, the current mass density of MBHs and reionization). These constraints translate into $0.01 \leq \lambda_{\text{thr}} \leq 0.02$.
- (ii) The mass function of QSs peaks at $M_{\text{QS}} \simeq 10^6 M_{\odot}$ and decreases almost as a power law with slope $\simeq -(1.5-2)$. The more efficient QS and MBH formation is, the steeper the mass function. This is because more seeds form at the highest redshifts when the hosts have relatively low mass and circular velocity.
- (iii) Modelling QS emission as a blackbody at $T_{\text{QS}} = 4000 \text{ K}$, we calculate the number counts for the *JWST* 2–10 μm band. Assuming a sensitivity of 10 nJy at 2 μm , we find that *JWST* could detect up to several Qs per field at $z \simeq 5-10$.
- (iv) The redshift of formation of Qs increases with the cosmic bias of their hosts. We therefore expect that the latest forming Qs should be found in the field rather than in a high-density environment.
- (v) The number density of MBHs can be dominated by either the descendants of Pop III remnants or QS MBHs. In the low-efficiency case ($\lambda_{\text{thr}} = 0.01$) the number density of MBHs is dominated by the Pop III channel, if available; the opposite is true for the high-efficiency case ($\lambda_{\text{thr}} = 0.02$). In contrast to the number density of MBHs, the mass density is always dominated by QS MBHs.

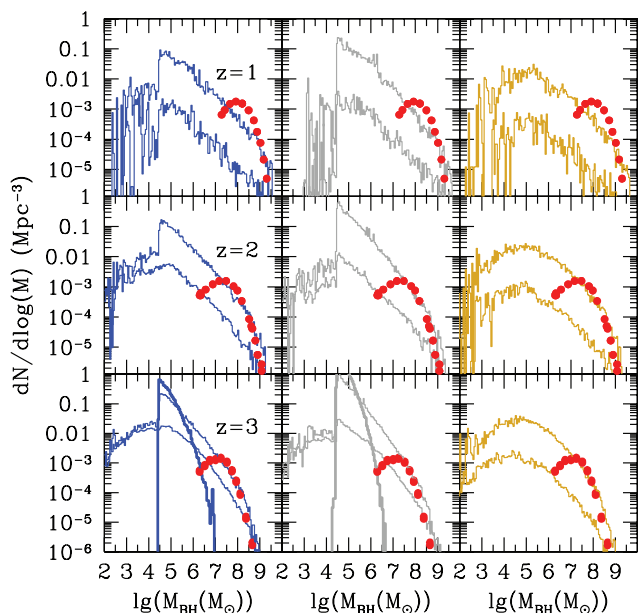


Figure 8. Mass function of MBHs in three redshift bins: $0.95 < z < 1.025$, $1.9 < z < 2.1$ and $2.8 < z < 3.2$. The bin width was chosen to span similar cosmic times (~ 0.32 Gyr). In each panel, we show two curves; the upper curve is the mass function of all BHs. The lower curve is the mass function of active BHs. We also show (big dots) the mass function predicted by Merloni & Heinz (2008). The very good agreement at masses of $> 10^7 M_{\odot}$ suggests that our MBH growth prescription is well motivated. Thick lines in the bottom panels: IMF of seed MBHs left behind by QSS. Left: $\lambda_{\text{thr}} = 0.01 + \text{Pop III MBHs}$. Middle: $\lambda_{\text{thr}} = 0.02 + \text{Pop III MBHs}$. If we switch off the Pop III channel, the mass function drops to zero at $\log(M_{\text{BH}}/M_{\odot}) = 4.5$, where the seed mass function drops as well. Right: Pop III MBHs only.

(vi) If the only channel of MBH formation is via QS seeds, then the mass function of MBHs cuts off below $\log(M_{\text{BH}}/M_{\odot}) = 4.5$, where the seed mass function drops as well. If Pop III remnants offer an alternative route to MBH formation, then the mass function is double-peaked, with each peak tracing a different seed formation mechanism. The more efficient QS formation is, the more pronounced the peak at $\log(M_{\text{BH}}/M_{\odot}) \simeq 4.5$ becomes.

This model differs from the previous proposed mechanisms (e.g. Lodato & Natarajan 2006) in various respects. First, gas accumulation in the central regions of protogalaxies is described by global, rather than local, instabilities (e.g. via the Toomre stability criterion). This implies that the spin parameter threshold for collapse cannot be derived by demanding that the Toomre parameter, $Q = \frac{c_s \kappa}{\pi G \Sigma}$ (where Σ is the surface mass density, c_s is the sound speed, $\kappa = \sqrt{2} V_h / R$ is the epicyclic frequency and V is the circular velocity of the disc), approaches a critical value, Q_c , of order unity. We instead assume a fixed spin parameter threshold (cf. BVR2006). Secondly, we have here relaxed the assumption that large gas infall rates can occur only at zero metallicity to avoid fragmentation. The predicted mass functions have distinctly different shapes (compare Fig. 3 with fig. 2 in Lodato & Natarajan 2007), and MBH seeds tend to form later in the model presented here ($z \simeq 5\text{--}10$ instead of $z \simeq 16\text{--}18$, when zero metallicity is required). Differences between models can possibly be looked for at the highest redshifts, as at later times the growth of MBHs by accretion and mergers likely washes out dissimilarities.

The most direct prediction of this work is shown in Fig. 4, where we estimate the detectability of QSS in a *JWST* field of view. At low

metallicities, QSS at redshifts of a few may resemble featureless blackbodies, with colours reminiscent of brown dwarfs. Direct redshift measurements may not be feasible, in which case QSS would have to be identified via their massive hosts. Their low numbers – a consequence of their short lifetimes – will make it even more challenging to find them with a telescope having a relatively small field of view. It is possible that some QSS could have formed in metal-enriched regions and even at relatively low redshifts ($z \gtrsim 2$). These might be relatively easy to detect but extremely rare. Other, secondary characteristics of QSS might conceivably aid in their detection, e.g. if accretion on to the BH deep in the core leads to the formation of a jet that pierces the stellar surface.

The shape of the MBH mass function below $\log(M_{\text{BH}}/M_{\odot}) = 4.5$ is an additional diagnostic. Direct dynamical measurement of MBH masses at the low-mass end is extremely difficult, due to the necessity of resolving the dynamics of stars within the sphere of influence of the putative MBH, i.e. the region where the Keplerian potential of the MBH dominates over the overall galaxy potential. This region typically subtends much less than an arcsecond in nearby galaxies, at or below the resolution limit of existing 8–10 m class telescopes. Future 20–30 m telescopes are likely to increase the sample of low-mass MBHs.

Gravitational waves produced during the inspirals of compact objects into MBHs – extreme-mass-ratio inspirals, in particular – are expected to provide accurate constraints on the mass function of BHs at low redshift. The proposed space-based gravitational wave detector, the *Laser Interferometer Space Antenna (LISA)*, can probe the mass function in the $10^4\text{--}10^7 M_{\odot}$ range. Gair, Tang & Volonteri (2010) find that with as few as 10 events, *LISA* should constrain the slope of the mass function below $\sim 10^6 M_{\odot}$ to a precision of ~ 0.3 , which is the current level of observational uncertainty in the low-mass slope of the BH mass function (Greene & Ho 2007). The combination of electromagnetic and gravitational wave observations in the coming years will improve the currently limited constraints on what route, or routes, lead to MBH seed formation.

Numerical simulations (e.g. Mayer et al. 2009) can test how strongly the ‘bars-within-bars’ inflow mechanism is tied to the angular momentum of the merging galaxies, and either validate our hypothesis that a single parameter λ_{thr} can describe the efficiency of the cascade process or indicate that additional parameters influence the formation of QSS and their nested BHs. Even the existence of a second parameter may not have a strong effect on the results: for instance, we found that our results are not very sensitive to varying the threshold gas fraction, as long as it is above $f_{\text{gas}} \simeq 0.025$. It will also be necessary to study the global dynamical behaviour of self-gravitating inflows in the inner regions, where the total gravitational potential is expected to approach a Keplerian shape and the dominant means of angular momentum transport may change from a flow involving large-scale bars ($m = 2$) to a flow resembling an eccentric disc ($m = 1$; e.g. Hopkins & Quataert 2010).

It is even more critical to check how much of the infalling gas actually reaches the small radial scales necessary to build the supermassive star. We have suggested that fragmentation and star formation may be less important than previously thought in quenching such inflows. Testing these hypotheses will require numerical experiments with high spatial resolution and a large dynamic range.

ACKNOWLEDGMENTS

MV acknowledges support from the SAO Award TM9-0006X and a Rackham faculty grant. MCB acknowledges support from NSF

grant AST-0907872. MV is grateful to Francesco Haardt for help with reionization models.

REFERENCES

- Ajello M. et al., 2009, *ApJ*, 699, 603
 Alvarez M. A., Wise J. H., Abel T., 2009, *ApJ*, 701, L133
 Baes M., Buyle P., Hau G. K. T., Dejonghe H., 2003, *MNRAS*, 341, L44
 Barth A. J., Martini P., Nelson C. H., Ho L. C., 2003, *ApJ*, 594, L95
 Begelman M. C., 1979, *MNRAS*, 187, 237
 Begelman M. C., 2010, *MNRAS*, 402, 673
 Begelman M. C., Meier D. L., 1982, *ApJ*, 253, 873
 Begelman M. C., Shlosman I., 2009, *ApJ*, 702, L5
 Begelman M. C., Volonteri M., Rees M. J., 2006, *MNRAS*, 370, 289 (BVR2006)
 Begelman M. C., Rossi E. M., Armitage P. J., 2008, *MNRAS*, 387, 1649
 Benson A. J., Džanović D., Frenk C. S., Sharples R., 2007, *MNRAS*, 379, 841
 Bouwens R. J. et al., 2010, *ApJ*, 709, L133
 Bromm V., Loeb A., 2003, *ApJ*, 596, 34
 Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
 Carlberg R. G., Sullivan M., Le Borgne D., 2009, *ApJ*, 694, 1131
 Djorgovski S. G., Volonteri M., Springel V., Bromm V., Meylan G., 2008, in Kleinert H., Jantzen R. T., Ruffini R., eds, *Proc. XI Marcel Grossmann Meeting on General Relativity*. World Scientific, Singapore (arXiv:0803.2862)
 D’Onghia E., Navarro J. F., 2007, *MNRAS*, 380, L58
 Dotti M., Colpi M., Haardt F., Mayer L., 2007, *MNRAS*, 379, 956
 Eisenstein D. J., Loeb A., 1995, *ApJ*, 443, 11
 Fan X. et al., 2001, *AJ*, 121, 54
 Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, *ApJ*, 681, 831
 Ferrarese L., 2002, *ApJ*, 578, 90
 Ferrarese L., Merritt D., 2000, *ApJ*, 539, L9
 Fowler W. A., 1966, *ApJ*, 144, 180
 Fryer C. L., Woosley S. E., Heger A., 2001, *ApJ*, 550, 372
 Gair J. R., Tang C., Volonteri M., 2010, *Phys. Rev. D*, 81, 104014
 Gebhardt K. et al., 2000, *ApJ*, 539, L13
 Ghisellini G. et al., 2010, *MNRAS*, 442
 Gnedin N. Y., 2008, *ApJ*, 673, L1
 Gottlöber S., Yepes G., 2007, *ApJ*, 664, 117
 Greene J. E., Ho L. C., 2007, *ApJ*, 667, 131
 Gültekin K. et al., 2009, *ApJ*, 698, 198
 Haehnelt M. G., Natarajan P., Rees M. J., 1998, *MNRAS*, 300, 817
 Hayashi C., 1961, *PASJ*, 13, 450
 Hopkins P. F., Quataert E., 2010, *MNRAS*, 407, 1529
 Hopkins P. F., Richards G. T., Hernquist L., 2007, *ApJ*, 654, 731
 Hoyle F., Fowler W. A., 1963, *MNRAS*, 125, 169
 Iben I. Jr, 1963a, *AJ*, 68, 281
 Iben I. Jr, 1963b, *ApJ*, 138, 1090
 Jiang L. et al., 2009, *AJ*, 138, 305
 Kollmeier J. A. et al., 2006, *ApJ*, 648, 128
 Komatsu E. et al., 2009, *ApJS*, 180, 330
 Koushiappas S. M., Bullock J. S., Dekel A., 2004, *MNRAS*, 354, 292
 Lagos C. D. P., Cora S. A., Padilla N. D., 2008, *MNRAS*, 388, 587
 Levine R., Gnedin N. Y., Hamilton A. J. S., Kravtsov A. V., 2008, *ApJ*, 678, 154
 Lodato G., Natarajan P., 2006, *MNRAS*, 371, 1813
 Lodato G., Natarajan P., 2007, *MNRAS*, 377, L64
 Loeb A., Rasio F. A., 1994, *ApJ*, 432, 52
 Madau P., Rees M. J., 2001, *ApJ*, 551, L27
 Mayer L., Kazantzidis S., Escala A., Callegari S., 2009, *Nat*, in press (arXiv:0912.4262)
 Merloni A., Heinz S., 2008, *MNRAS*, 388, 1011
 Milosavljević M., Couch S. M., Bromm V., 2009, *ApJ*, 696, L146
 Monaco P., Fontanot F., Taffoni G., 2007, *MNRAS*, 375, 1189
 Omukai K., Palla F., 2003, *ApJ*, 589, 677
 Pelupessy F. I., Di Matteo T., Ciardi B., 2007, *ApJ*, 665, 107
 Pizzella A., Corsini E. M., Dalla Bontà E., Sarzi M., Coccato L., Bertola F., 2005, *ApJ*, 631, 785
 Rees M. J., 1978, in Berkhuijsen E. M., Wielebinski R., eds, *Proc. IAU Symp. 77, Structure and Properties of Nearby Galaxies*. Reidel, Dordrecht, p. 237
 Regan J. A., Haehnelt M. G., 2009, *MNRAS*, 393, 858
 Rhoads K. J., Haehnelt M. G., 2006, *MNRAS*, 373, 623
 Scannapieco E., Barkana R., 2002, *ApJ*, 571, 585
 Scannapieco E., Schneider R., Ferrara A., 2003, *ApJ*, 589, 35
 Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
 Shapiro S. L., 2005, *ApJ*, 620, 59
 Shlosman I., Frank J., Begelman M. C., 1989, *Nat*, 338, 45
 Shlosman I., Begelman M. C., Frank J., 1990, *Nat*, 345, 679
 Sołtan A., 1982, *MNRAS*, 200, 115
 Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *MNRAS*, 391, 481
 Springel V., Di Matteo T., Hernquist L., 2005, *MNRAS*, 361, 776
 Taffoni G., Mayer L., Colpi M., Governato F., 2003, *MNRAS*, 341, 434
 Thorne K. S., Zytzkow A. N., 1977, *ApJ*, 212, 832
 Trenti M., Stiavelli M., 2009, *ApJ*, 694, 879
 Van Wassenhove S., Volonteri M., Walker M. G., Gair J. R., 2010, *MNRAS*, in press, arXiv:1001.5451 (doi:10.1111/j.1365-2966.2010.17189.x)
 Vernaleo J. C., Reynolds C. S., 2006, *ApJ*, 645, 83
 Volonteri M., Gnedin N. Y., 2009, *ApJ*, 703, 2113
 Volonteri M., Natarajan P., 2009, *MNRAS*, 400, 1911
 Volonteri M., Rees M. J., 2005, *ApJ*, 633, 624
 Volonteri M., Rees M. J., 2006, *ApJ*, 650, 669
 Volonteri M., Haardt F., Madau P., 2003, *ApJ*, 582, 559
 Volonteri M., Lodato G., Natarajan P., 2008, *MNRAS*, 383, 1079
 Volonteri M., Gültekin K., Dotti M., 2010, *MNRAS*, 404, 2143
 Willott C. J. et al., 2009, *AJ*, 137, 3541
 Wise J. H., Turk M. J., Abel T., 2008, *ApJ*, 682, 745
 Yoshida N., Omukai K., Hernquist L., Abel T., 2006, *ApJ*, 652, 6

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.