FORMATION OF THE FIRST NUCLEAR CLUSTERS AND MASSIVE BLACK HOLES AT HIGH REDSHIFT

B. Devecchi\(^1\) and M. Volonteri\(^2\)

\(^1\) Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano Bicocca, Piazza della Scienza 3, 20126 Milano, Italy; bernadetta.devecchi@mib.infn.it
\(^2\) Astronomy Department, University of Michigan, 500 Church Street, Ann Arbor, MI, 48109, USA

Received 2008 September 8; accepted 2008 December 3; published 2009 March 16

ABSTRACT

We present a model for the formation of massive black holes (\(~1000 M_\odot\)) due to stellar-dynamical processes in the first stellar clusters formed at early cosmic times (\(z \sim 10–20\)). These black holes are likely candidates as seeds for the supermassive black holes detected in quasars and nearby quiescent galaxies. The high redshift black hole seeds form as a result of multiple successive instabilities that occur in low metallicity (\(Z \sim 10^{-5} Z_\odot\)) protogalaxies. We focus on relatively massive halos at high redshift (\(T_{\rm vir} > 10^4\) K, \(z \gtrsim 10\)) after the very first stars in the universe have completed their evolution. This set of assumptions ensures that (1) atomic hydrogen cooling can contribute to the gas cooling process, (2) a UV field has been created by the first stars, and (3) the gas inside the halo has been mildly polluted by the first metals. The second condition implies that at low density \(H_2\) is dissociated and does not contribute to cooling. The third condition sets a minimum threshold density for fragmentation, so that stars form efficiently only in the very inner core of the protogalaxy. Within this core, very compact stellar clusters form. The typical star cluster masses are of order \(10^7 M_\odot\) and the typical half mass radii \(\sim 1\) pc. A large fraction of these very dense clusters undergoes core collapse before stars are able to complete stellar evolution. Runaway star–star collisions eventually lead to the formation of a very massive star, leaving behind a massive black hole remnant. Clusters unstable to runaway collisions are always the first, less massive ones that form. As the metallicity of the universe increases, the critical density for fragmentation decreases and stars start to form in the entire protogalactic disk so that (1) accretion of gas in the center is no longer efficient and (2) the core collapse timescale increases. Typically, a fraction \(\sim 0.05\) of protogalaxies at \(z \sim 10–20\) form black hole seeds, with masses \(\sim 1000–2000 M_\odot\), leading to a mass density in seeds of a few \(\gtrsim 10^2 M_\odot/\text{Mpc}^3\). This density allows enough room for black hole growth by accretion during the quasar epoch.

Key words: black hole physics – instabilities – stellar dynamics – galaxies: nuclei – galaxies: formation

Online-only material: color figures

1. INTRODUCTION

Supermassive black holes (BHs) are routinely detected in the center of galaxies, both in nearby quiescent galaxies, and as the engines that power quasars and active galactic nuclei (AGN). Observationally, we can trace quasars until very high redshifts. Luminous quasars are detected in the Sloan survey (e.g., Fan et al. 2001) at \(z \sim 6\), corresponding to a time when the universe was not even one billion years old. This implies that the first BH seeds must have formed earlier on. The currently favored scenario for BH seed formation relies on the remnants of the very first generation of metal-free stars (Population III, PopIII). The first stars are believed to form at \(z \sim 20\) in halos which represent rare high peaks of the primordial density field. Simulations of the fragmentation of zero metallicity protogalaxies suggest a very top-heavy initial stellar mass function, and in particular the production of very massive stars with mass \(>100 M_\odot\). PopIII stars in the mass range \(140 \lesssim m_* \lesssim 260 M_\odot\) are predicted to make pair-instability supernovae. If zero metallicity very massive stars form above \(260 M_\odot\), they will rapidly collapse to BHs with little mass loss (Fryer et al. 2001), leaving behind BHs with masses \(\lesssim 10^2 M_\odot\) (Madau & Rees 2001).

Although this path to BH seed formation seems very natural, large uncertainties exist on the final mass of PopIII stars. Even recent simulations (Gao et al. 2007) have not clarified if PopIII stars are indeed very massive, and in particular if they are above the threshold (\(\gtrsim 260 M_\odot\)) for BH formation (but see Freese et al. 2008; Spolyar et al. 2008; Natarajan et al. 2009). Furthermore, metal abundances in extremely metal poor Galactic halo stars, which are commonly thought to trace the enrichment products of the first generation of stars, are incompatible with the yield patterns of zero metallicity very massive stars (\(\gtrsim 100 M_\odot\), Tumlinson et al. 2004).

Alternative routes to BH seed formation have been explored (Haehnelt & Rees 1993; Loeb & Rasio 1994; Eisenstein & Loeb 1995; Bromm & Loeb 2003; Koushiappas et al. 2004; Begelman et al. 2006, Lodato & Natarajan 2006), typically exploiting gas-dynamical processes in metal-free galaxies. Gravitational instabilities can indeed lead to a vigorous gas inflow into the very central region, supplying the necessary matter for the formation of BH seeds. The typical conditions that lead to efficient gas infall and BH seed formation can be summarized as: (1) the host is massive enough that the virial temperature exceeds \(T_{\rm vir} > 10^4\) K, so that gas is able to cool and collapse via atomic hydrogen cooling; (2) molecular hydrogen does not form as the gas cools and condenses; and (3) the gas has primordial composition, so that metal line cooling is nonexistent.

However, massive halos (\(T_{\rm vir} > 10^4\) K, masses \(\sim 10^7 M_\odot\)) are likely built-up from mini halos (\(T_{\rm vir} < 10^4\) K) that had collapsed earlier on. Some of these halos might have experienced PopIII star formation, and should be enriched with at least some trace amount of metals. Fragmentation and formation of low-mass stars starts as soon as gas is polluted by metals created in the first PopIII stars. Efficient gas collapse, leading to BH seed formation, is mutually exclusive with star formation, as competition for the gas supply limits the mass available.
However, this first episode of efficient star formation can foster the formation of very compact nuclear star clusters (Clark et al. 2008; Schneider et al. 2006) where star collisions can lead to the formation of a supermassive star, possibly leaving a BH remnant with mass in the range $\sim 10^2$–$10^4 \, M_\odot$ (Omukai et al. 2008).

The possibility that an “intermediate-mass” BH could, in principle, form as a result of dynamical interactions in dense stellar systems is a long standing idea (Begelman & Rees 1978; Freitag et al. 2006b, 2006a; Ebisuzaki et al. 2001; Portegies Zwart & McMillan 2002; Miller & Hamilton 2002; Gürkan et al. 2004). During their lifetime, collisional stellar systems evolve as a result of dynamical interactions. In an equal mass system, the central cluster core initially contracts as the system attempts to reach a state of thermal equilibrium; energy conservation leads to a decrease in the core radius as evaporation of the less bound stars proceeds. As a result, the central density increases and the central relaxation time decreases. The core then decouples thermally from the outer region of the cluster. Core collapse then speeds up as it is driven by energy transfer from the central denser region (Spitzer 1987).

This phenomenon is greatly enhanced in multimass systems like realistic star clusters. In this case, the gravothermal collapse happens on a shorter timescale as dynamical friction causes the more massive stars of mass $m$ to segregate in the center on a timescale $t_{\text{df}} = (\langle m \rangle / m) t_{\text{rh}}$ (where $t_{\text{rh}}$ is the half mass relaxation timescale, and $\langle m \rangle$ is the mean stellar mass in the cluster). If mass segregation sets in before the more massive stars evolve out of the main sequence ($\sim 3$ Myr), then a subsystem decoupled from the rest of the cluster can form, where star–star collisions can take place in a runaway fashion that ultimately lead to the growth of a very massive star (VMS; Portegies Zwart et al. 1999).

Yungelson et al. (2008) study the fate of solar composition VMSs in the mass range 60–1000 $M_\odot$. They find that all VMSs are likely to shed most of their mass via winds well before experiencing a supernova explosion. Solar composition VMSs are, therefore, expected to end their lives as objects less massive than $\sim 150 \, M_\odot$, collapsing into BHs with mass $\lesssim 150 \, M_\odot$ or exploding as pair-instability supernovae. The growth of a VMS should be much more efficient at low metallicity. Low metallicity can modify the picture in different ways. First, at subsolar (but still not primordial) metallicity, all stars with low, but nonzero, metallicity.

Yungelson et al. (2008) study the fate of solar composition VMSs in the mass range 60–1000 $M_\odot$. They find that all VMSs are likely to shed most of their mass via winds well before experiencing a supernova explosion. Solar composition VMSs are, therefore, expected to end their lives as objects less massive than $\sim 150 \, M_\odot$, collapsing into BHs with mass $\lesssim 150 \, M_\odot$ or exploding as pair-instability supernovae. The growth of a VMS should be much more efficient at low metallicity. Low metallicity can modify the picture in different ways. First, at subsolar (but still not primordial) metallicity, all stars with masses $\geq 40 \, M_\odot$ are thought to collapse directly into a black hole (Heger et al. 2003) without exploding as supernovae. Second, the mass loss due to winds is much more reduced in metal-poor stars, which greatly helps in increasing the mass of the final remnant. In this paper, we investigate the formation of BHs, remnants of VMS formed via stellar collisions in the very first stellar clusters at early cosmic times. We derive the properties of the ensuing BH population that can represent the “seeds” of the supermassive BHs, which we observe in today’s galaxies. In the following, we summarize the main features of our model.

1. We focus on halos with virial temperatures $T_{\text{vir}} \gtrsim 10^4$ K at $z > 5$ after the first episode of star formation, hence with a low, but nonzero, metallicity.

2. Low metallicity gas can fragment and form low-mass stars only if the gas density is above a certain threshold, $n_{\text{crit, Z}}$.

3. Halos possess angular momentum, acquired through interaction with their neighbors; hence, gas collapse ultimately leads to the formation of a disk.

4. If the forming disk is Toomre unstable, instabilities lead to mass infall instead of fragmentation into bound clumps and global star formation in the entire disk.

5. The gas inflow increases the central density, and within a certain, compact, region $n > n_{\text{crit, Z}}$.

6. Star formation ensues and a dense star cluster is formed.

7. If the star cluster goes into core collapse in $\lesssim 3$ Myr runaway collisions of star form a supermassive star, leading to a massive BH remnant.

We describe the physical mechanism for cluster formation in Section 2, where we discuss the conditions leading to very compact stellar clusters in low metallicity ($Z \sim 10^{-5} \, Z_\odot$) protogalaxies. We discuss dynamical instabilities (Section 2.1) driving gas toward the center, and fragmentation occurring at high densities (Sections 2.2 and 2.3). The properties of the star clusters, VMSs, and BHs are described in Sections 2.4 and 2.5.

We present our results in Section 3 and discuss the implications of our model in Section 4.

### 2. BH Seed Formation: Model Set-Up

We start by considering a halo of mass $M_h$, with virial circular velocity $V_h$, virial radius $R_h$, angular momentum $J$. The rotational support can be quantified in terms of the spin parameter

$$\lambda \equiv |J/E|^1/2 / GM_h^{1/2},$$

where $E$ is the total energy of the halo. We further assume that a fraction $m_d \sim 0.05$ of the halo mass is in a gaseous baryonic component than can cool and condense. The collapsing mass will then be $m_d M_h$. The angular momentum of the collapsing baryons will be a fraction $j_d$ of the halo angular momentum, with $j_d/m_d = 1$ if the specific angular momentum of baryons is conserved during collapse. In this work, we focus on relatively massive halos ($T_{\text{vir}} > 10^4$ K, ensuring that gas can cool by atomic hydrogen) after the birth of the first stars in the universe. $10^4$ K halos are likely build up by smaller halos that have already experienced a first episode of star formation. We assume that the first PopIII stars have been able to affect the gas both radiatively (thus precluding subsequent $H_2$ cooling in the halo) and chemically. The first condition implies that, at least at the low density at which the collapse starts, $H_2$ is dissociated and does not contribute to the cooling of the gas (see also Figures 9 and 10 in O’Shea & Norman 2008). The second condition ensures that metals (and dust) in small quantities can cool down the gas efficiently only as the gas density reaches a critical threshold (see Section 2.2; Figures 2, 4, and 5 in Smith et al. 2009). Jointly, the two conditions ensure that at first cooling is driven by atomic hydrogen only.

#### 2.1. Disk Formation and Mass Inflow

We describe the early evolution of the cooling and collapsing baryons in a simple way, assuming adiabatic response of the halo to gas cooling and disk formation (e.g., Mo et al. 1998; Oh & Haiman 2002; Lodato & Natarajan 2006). As we are assuming negligible $H_2$, the tenuous gas cools down by atomic hydrogen only until it reaches $T_{\text{gas}} \sim 4000$ K. At this point, the cooling function of the atomic hydrogen drops by a few orders of magnitude, and contraction proceeds nearly adiabatically.

Given the presence of angular momentum, the contraction in the equatorial plane stops as the gas becomes rotationally supported, and the collapse ultimately leads to the formation of a disk. As pointed out by Mo et al. (1998), the properties of the disk can be determined by assuming a profile for the surface density. $\Sigma$. As the mass assembles in the protodisk, the surface density

$$\Sigma \propto \rho \propto m_{\text{_primitive}} \rho_{\text{inside}} \rho_{\text{outside}}$$

where $m_{\text{primitive}}$ is the mass of the primitive gas, $\rho_{\text{inside}}$ and $\rho_{\text{outside}}$ are the densities at the inside and outside of the disk, respectively. In the case of a spherical collapse, $\rho_{\text{inside}} = \rho_{\text{outside}}$ and the profile can be written as

$$\Sigma \propto m_{\text{primitive}}/s^2, \quad s \propto r,$$

where $s$ is the distance from the center of the collapse. The mass of the primitive gas is $m_{\text{primitive}} = \int_0^R \pi s^2 \rho_{\text{outside}} \, ds$, which gives

$$\Sigma \propto m_{\text{primitive}}/s^2 \propto m_{\text{primitive}}/r^2,$$

where $r$ is the radius of the collapse. The mass of the primitive gas is $m_{\text{primitive}} = \int_0^R \pi s^2 \rho_{\text{outside}} \, ds$, which gives

$$\Sigma \propto m_{\text{primitive}}/s^2 \propto m_{\text{primitive}}/r^2,$$

where $r$ is the radius of the collapse. The mass of the primitive gas is $m_{\text{primitive}} = \int_0^R \pi s^2 \rho_{\text{outside}} \, ds$, which gives

$$\Sigma \propto m_{\text{primitive}}/s^2 \propto m_{\text{primitive}}/r^2.$$
increases. If for the final configuration, the Toomre parameter

\[ Q = \frac{c_s\kappa}{\pi G\Sigma} < Q_c, \quad (1) \]

where \( c_s \) is the sound speed, \( \kappa \) is the epicyclic frequency, and \( Q_c \) is a critical value, then the disk becomes unstable, and it develops barlike structures. We argue that if the destabilization of the system is not too violent, instabilities lead to mass infall instead of fragmentation into bound clumps and global star formation in the entire disk (Shlosman et al. 1990; Lodato & Natarajan 2006). This is the case if the inflow rate is below a critical threshold \( M_{\text{max}} = 2\alpha_c \dot{\rho} \), the disk is able to sustain (where \( \alpha_c \sim 0.12 \) describes the viscosity) and molecular and metal cooling are not important. We substantiate further our argument in Section 2.2. In the following, we discuss what happens to cooling are not important. We substantiate further our argument in Section 2.2. In the following, we discuss what happens to

The final configuration of the structure would then be characterized by an outer surface density profile of a Mestel disk and a central denser region:

\[
\Sigma(R) \sim \begin{cases} 
\Sigma_0 \left( \frac{R}{R_0} \right)^{-1} & R > R_{\text{tr}}, \\
\Sigma_0 \left( \frac{R}{R_0} \right)^{-\gamma} & R < R_{\text{tr}},
\end{cases} 
\]

where \( \Sigma_0 \) can be written as \( \Sigma_0(R_0/R_h)^{1-\gamma} \) by imposing continuity at \( R_h \). At this point, the parameters of the final disk are \( \Sigma_0, R_0, R_{\text{tr}} \), and the fractional mass of the halo \( m_h \) that participate to the infall. They can be determined by imposing the conditions of mass and angular momentum conservation, \( Q = Q_c \) (for \( R > R_{\text{tr}} \)) and by imposing that mass \( m_h M_h \) is added to that already in place inside \( R_{\text{tr}} \). For simplicity, we assume that the dark matter halo hosting the protogalaxy follows an isothermal profile. The resulting disk properties are as follows:

\[
\Sigma_0 = \frac{10 m_d (m_d/j_d)^2 H \nu_0 (1 - m_d/m_d)^3}{16\pi G\lambda^3} \quad (4)
\]

\[
R_0 = \frac{2\sqrt{2}}{(1 - m_d/m_d)} \frac{j_d}{m_d} R_h \quad (5)
\]

\[
m_d = m_d \left( 1 - \frac{8\lambda}{m_d T_v (T_{\text{vir}}/T_{\text{vir}})^{1/2}} \right) \quad (6)
\]

\[
R_{\text{tr}} = \frac{m_d M_h}{4\pi \Sigma_0 R_0} \quad (7)
\]

where \( H_0 \) is the Hubble constant at redshift \( z \). The vertical structure of the disk is determined by solving the equation for hydrostatic equilibrium. For a disk that is isothermal and self-gravitating, this implies a \( z \) dependency for the density \( \propto \cosh^{-2}(h/H) \), where \( H \) is the vertical scale height, and it is equal to \( H(R) = \sqrt{c_s^2/\pi G\Sigma(R)} \). We can, therefore, express the inner disk density as

\[
n(R, h) = n_0 \left( \frac{R}{R_0} \right)^{-5/3} \cosh^{-2}(h/H(R)) \quad (8)
\]

where \( n_0 \) is a function of the disk parameters, \( \Sigma_0, R_0 \), and \( \gamma \) (see Appendix A).

2.2. Fragmentation of the Disk

A necessary condition for the inflow process described in the previous section is that star formation does not take place in the entire disk. If this happens, the gas that would otherwise flow into the central region, is consumed as it is converted into stars. Rice et al. (2005) proposed that fragmentation in a thin disk sets in when the gravitationally induced stress exceeds a critical value. Describing angular momentum transport in terms of the \( \alpha \) prescription for viscous dissipation (i.e., that the torque strength can be expressed in terms of \( \alpha \); Shakura & Sunyaev 1973), then the critical threshold for fragmentation \( \alpha_c \) determines how much angular momentum can be transported in a steady state. In this sense, the fragmentation boundary is ultimately due to the inability of the disk to redistribute angular momentum on a sufficiently short timescale. In the classical analysis (Gammie 1996), fragmentation develops as the condition \( t_{\text{cool}} = t_{\text{dyn}} \) is reached. In the framework of Rice et al. (2005), we can expect that the same effect develops if the mass inflow from the halo
induces too strong a stress. Lodato & Natarajan (2006) apply this framework to disks of primordial composition. By requiring the mass-accretion rate from the halo \( M_h = m_d \frac{V_d}{\pi} \) to be less than the maximum value \( M_{\text{max}} = 2 \alpha_c c_s^3 \), the disk is able to sustain, they argue that to avoid fragmentation it must be

\[
\frac{T_{\text{vir}}}{T_{\text{gas}}} < \left( \frac{4 \alpha_c}{m_d} \right)^{1/3},
\]

where \( \alpha_c \sim 0.12 \) (Clarke et al. 2007) is the critical value for fragmentation.

We assume \( T_{\text{gas}} = 4000 \text{ K} \), which corresponds to \( M_{\text{max}} \sim 10^{-2} M_\odot \text{ yr}^{-1} \). The joint conditions, \( m_d > 0 \) (Equation (6)) and Equation (9) then impose an upper limit to the virial temperature \( T_{\text{vir}} \ll 1.8 \times 10^4 \text{ K} \). We, therefore, consider halos with \( 10^4 \text{ K} \leq T_{\text{vir}} \leq 1.8 \times 10^4 \text{ K} \).

By considering these gas and virial temperatures, we can ensure that the fragmentation threshold is not attained at least at \( R > R_\odot \). However, \( T_{\text{vir}} = 4000 \text{ K} \) requires the absence of coolants other than atomic hydrogen. If \( H_2 \) contributes efficiently as a coolant, the gas temperature drops, and Equation (9) implies that massive halos are subject to strong fragmentation. For instance, if gas cools to \( T_{\text{gas}} \sim 200 \text{ K} \), only halos with \( T_{\text{vir}} \leq 1000 \text{ K} \) satisfy Equation (9). At redshift \( z \sim 20 \), this corresponds to halo masses around a few \( 10^5 M_\odot \).

The issue of \( H_2 \) cooling suppression in the presence of a strong UV field has been addressed in a number of recent studies on the first galaxies. Even if a strong UV background is not already in place, a “local” UV field can be established by the presence of PopIIIIs in surrounding halos. Eventually, \( H_2 \) can be suppressed by previous PopIII star formation in the smaller systems assembled to form the final \( T_{\text{vir}} \sim 10^4 \text{ K} \) halo (see Ciardi 2008 or Ciardi & Ferrara 2005 for a discussion on this issue). O’Shea & Norman (2008) have shown that halos with \( T_{\text{vir}} \sim 10^4 \text{ K} \) embedded in a UV background with strength higher than \( \sim 3 \times 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1} \) are not subject to efficient cooling and fragmentation in the disk (see their Figure 9(b)), and it is only in the central region (\( r < 10 \text{ pc} \)) that \( T_{\text{gas}} \) drops down to \( \sim 10^3 \text{ K} \), leading to the formation of a central single massive star born in isolation. Halos subject to a weaker UV field show the same radial behavior but shifted at lower temperature. If \( H_2 \) cooling in fact does not act on large scales, it does not affect the inflow process described in Section 2.1 so that material can still be transported to small distances by dynamical instabilities.

In metal-free conditions, the gaseous density in the central region increases until \( H_2 \) is activated (O’Shea & Norman 2008; Omukai et al. 2008, and references therein) and the formation of a Pop III star can proceed. On the other hand, if the gas has been enriched to a certain level, fragmentation can take place and an entire stellar cluster is formed (Clark et al. 2008). It is this last situation that we want to examine in more details, i.e., the possibility that a cluster of stars is formed instead of a single massive star.

2.3. Critical Metallicity for Fragmentation

Various authors have suggested that the presence of a certain amount of metals is the key ingredient in order to produce efficient fragmentation (Schneider et al. 2006; Bromm et al. 1999, 2002; Omukai et al. 2008; Clark et al. 2008 and references therein). Let us define a critical metallicity \( Z_{\text{crit}} \) where transition from PopIII to “normal” star formation occurs. \( Z_{\text{crit}} \) depends on the coolants at work. Bromm et al. (2001) simulated the collapse of a halo of \( 2 \times 10^6 M_\odot \) at different values of \( Z \). They estimate \( Z_{\text{crit}} \sim 10^{-3.5} \) and show that when \( Z > Z_{\text{crit}} \), a rotationally supported disk that fragments vigorously can form. More recently, Clark et al. (2008) simulate the collapse of a rotating cloud polluted by dust. They show that a tightly packed cluster of protostars is formed in the center.

To estimate the value of \( Z_{\text{crit}} \) at which gas start to fragment, one can compare the cooling rate \( \Lambda_{\text{cool}} \) to the adiabatic heating rate \( \Gamma_{\text{ad}} \). Fragmentation requires the \( \Lambda_{\text{cool}} \geq \Gamma_{\text{ad}} \). Once this happens, the temperature starts to decrease as the density increases until cooling stops to be efficient. The condition \( \Gamma_{\text{ad}} < \Lambda_{\text{cool}} \), in fact, constrains the metallicity at a given temperature and density to be greater than a critical value \( Z_{\text{crit}} \). For an initially isothermal gas, this condition provides a relationship between metallicity and density. Equivalently, if the gas is characterized by a fixed metallicity \( z \), then only those regions with density greater than a given threshold \( n_{\text{crit},z} \) are able to cool down efficiently and fragment into stars. This is the condition that we apply to collapsing disks to determine if they can develop into stellar clusters.

Santoro & Shull (2006) studied the condition for fragmentation in a low metallicity universe assuming that PopIII stars are the main sources of pollution. The \( Z_{\text{crit}}-n_{\text{crit},z} \) relation depends on the ratio between the different species of coolants (see Figure 9 in Santoro et al. 2006). In all cases, the critical metallicity increases with decreasing density until \( n \) reaches the critical value for collisional de-excitation of the dominant coolants. We adopt as a reference the curve corresponding to PopIII stars in the intermediate mass range (mass range \( 185–205 M_\odot \); however, the solar abundance ratio produces a similar pattern as well, e.g., solid and dotted curves in Figure 10 in Santoro & Shull 2006). We discuss how our results depend on the specific \( Z_{\text{crit}}-n_{\text{crit},z} \) relation in Section 3.2.2.

The metallicity of gas in a given halo most probably depends on its mass and on the redshift. The cosmic metallicity history (MEH) has been investigated both from an observational and a theoretical perspectives (Scannapieco et al. 2003; Tornatore et al. 2007; Savaglio 2006; Savaglio et al. 2005; Prochaska et al. 2003; Prochaska et al. 2007; Kulkarni et al. 2005; Li 2008). The observed mean metallicity decreases with redshift, but the rate of decrease depends on the type of sources studied. From studies of QSO–DLAs, \( Z/Z_\odot \) scales as \( 10^{-0.6} \beta \) with \( \beta = 0.36 \) (Li 2008), while measurement based on GRBs point toward a shallower decline, so that at \( z \sim 3–4 \) already \( Z = 10^{-1} Z_\odot \) (Savaglio 2006). The differences in the slope and in the normalization are often ascribed to different histories of metal enrichment in QSOs and GRBs hosts (Li 2008 and references therein).

We model the MEH based on fits presented in Li 2008 that we refer to for additional details. In our reference model, we adopt \( \beta = 0.36 \), that is \( Z/Z_\odot = 0.35 \times 10^{-0.36z} \). The case \( \beta = 0.18 \), leading to \( Z/Z_\odot = 0.63 \times 10^{-0.18z} \), is discussed in Section 3.2.3. We further allow a logarithmically uniform scatter in \( Z \) of \( \Delta \log(Z) = 1.5 \), based on the observed scatter in \( Z/Z_\odot \) from measurements of the QSO–DLAs. We wish to stress that our treatment of MEH is highly simplified. In a forthcoming paper, we will determine self-consistently the evolution of the MEH from stellar winds pollution.

2.4. Stellar Cluster Formation

The central star cluster forms within the region where gas at a given metallicity has reached the critical density for fragmentation. Given a specific density profile (Equation (2)), the requirement \( n > n_{\text{crit},z} \) translates into a condition \( R < R_{\text{SF}} \).
where we define $R_{\text{SF}}$ as the radius where the density in the plane $h = 0$ reaches the value $n_{\text{crit},Z}$ for the gas metallicity $Z$, that is:

$$n(R = R_{\text{SF}}, h = 0) = n_{\text{crit},Z}. \quad (10)$$

As long as $n(R_{\text{SF}}, 0) < n_{\text{crit},Z}, n(R, h)$ is defined by Equation (8) and the radius within which gas fragments and stars form can be written as

$$R_{\text{SF}} = R_{\text{tr}} \left[ \frac{\Sigma_0 R_0}{c_s R_{\text{tr}}} \frac{\pi G}{2 \mu m_{\text{H}} n_{\text{crit},Z}} \right]^{1/y}, \quad (11)$$

where the metallicity dependence is included implicitly in $n_{\text{crit},Z}$ (see Appendix A), $\mu$ is the molecular weight, and $m_{\text{H}}$ is the proton mass. The half mass radius can be expressed as a function of the cluster radius in a very simple way, as

$$R_h = 2^{1/(y-1)} R_{\text{SF}}. \quad (12)$$

For a given $\gamma$, the mass in stars of the cluster can be calculated as

$$M_{\text{cl}} = \epsilon_{\text{SF}} 2\pi \int_0^{R_{\text{SF}}} \Sigma(R) R dR$$

$$= \epsilon_{\text{SF}} \left[ \frac{2\pi}{2 - \gamma} \frac{\Sigma_0^2 R_0^2}{c_s} \frac{\pi G}{2 \mu m_{\text{H}} n_{\text{crit}}^2} \left( \frac{m_{\text{H}}}{y-1} \right) \right]^{2\gamma - 2}, \quad (13)$$

where $\epsilon_{\text{SF}}$ is the fraction of gas converted into stars. We assume $\epsilon_{\text{SF}} = 0.25$, consistently with the star formation efficiency in the low redshift universe (Lada & Lada 2003). Given the metallicity, $Z$, it is now possible to determine the extent of the region inside which star formation is allowed. The cluster properties for a given halo are, therefore, uniquely described for any given metallicity.

At high metallicities, the case $n(R_{\text{SF}}, 0) > n_{\text{crit},Z}$ becomes common. This implies $R_{\text{SF}} > R_{\text{tr}}$, that is, star formation takes place in the region of the disk where no inflow is taking place. In other words, once the gas reaches $R_{\text{SF}}$, it is no longer able to be routed efficiently in the inner region as star formation starts to consume gas. In the remainder of the paper, we will assume conservatively that if $R_{\text{SF}} > R_{\text{tr}}$ no cluster formation occurs, as stars are formed in the disk rather than in the central compact region.

### 2.5. Runaway Instability of the Central Cluster

The conditions under which “mass segregation instability” can occur have been investigated in a series of papers, focusing on clusters in the present day universe. A successful core collapse requires that the core collapse time be less than the main sequence lifetime of the most massive stars (mass losses from supernovae expand the core and increase interaction times). The main sequence lifetime of massive stars asymptotes to about 2.5 Myr because all stars go off the main sequence when they have consumed about 15% of their hydrogen, and for high-mass stars with luminosities approaching Eddington, $L \propto M$ (not $M^{3.5}$ as is the case for lower masses) and then the lifetime $\propto 0.15 M/L \sim \text{const.}$

With a Monte Carlo code, Gürkan et al. (2004) found that typically the mass of the collapsing core is $10^{-3}$ times that of the entire cluster. Similarly, Portegies Zwart & McMillan (2002) related the mass of the VMS with the parameters of the cluster taking into account both numerical simulations and analytical arguments. Portegies Zwart & McMillan (2002) have shown that core collapse occurs on a timescale

$$t_{cc} \simeq 3 \text{ Myr} \left( \frac{R_h}{1 \text{pc}} \right)^{3/2} \left( \frac{M_{\odot}}{5 \times 10^8 M_{\odot}} \right)^{1/2} \times \left( \frac{10 M_{\odot}}{m} \right) \left( \frac{8.5}{\ln \lambda_C} \right), \quad (14)$$

where $\ln \lambda_C$ is the Coulomb logarithm of dynamical friction (Binney & Tremaine 1987). The same simulations also find that the mass of the VMS can reach values as high as $10^3 M_{\odot}$ (see also Freitag et al. 2006b). The final mass of the VMS, however, depends on complex phenomena related to both the dynamics and hydrodynamics of the collisions. At solar metallicity, the growth of a star with mass greater than $\sim 100 M_{\odot}$ can be highly problematic as mass loss occurs both during the main sequence phase and at the end of the evolution when the star collapses into a BH. The growth of a VMS should, in principle, be much more efficient for metal-poor stars as in this case as mass loss should be strongly reduced. Recent models of stellar evolution at $Z \sim 10^{-5} Z_{\odot}$ have shown that mass loss due to stellar winds during the main sequence phase is almost unimportant and that the main contribution to the reduction of the stellar mass is due to the effect of rotation. This can reduce the mass of the star by a factor of order 2–4 (Meynet et al. 2008). If the final mass achieved by the VMS is greater than $\sim 260 M_{\odot}$, then after the main sequence it collapses into a BH retaining most of its mass (Heger et al. 2003).

To estimate the final mass of the VMS, we follow the treatment outlined in Portegies Zwart & McMillan (2002). We assume that the mass of the BH seed, $M_{\text{BH}}$, corresponds to the final mass of the VMS:

$$M_{\text{BH}} = m_* + 4 \times 10^{-3} M_{\odot} f_c \ln \lambda_C \ln \left( \frac{t_{\text{MS}}}{t_{cc}} \right). \quad (15)$$

Here $t_{\text{MS}} = 3 \text{ Myr}$ is the main sequence lifetimes of massive ($> 40 M_{\odot}$, Hirschi 2007) stars, $m_*$ is the initial mass of the seed star that experiences runaway growth, and $f_c$ is the factor used to calibrate the analytical expectation with direct numerical simulations. Portegies Zwart & McMillan (2002) find $f_c = 0.2$ for $\ln (\lambda_C) = \ln (0.1 M_{\odot}/m)$. We adopt the same values as Portegies Zwart & McMillan (2002) for both parameters, further assuming $m_* = m = 10 M_{\odot}$ consistent with the characteristic stellar mass in $10^{10}$-$10^8 M_{\odot}$ halos at redshift $\sim 10$ (see Figures 2 and 3 in Clarke & Bromm 2003).

It is important to stress that Equation (15) provides an upper limit to $M_{\text{BH}}$. However, once the seed is born, the BH is still embedded into a dense cluster of stars. Even if its growth has been limited before the collapse of the VMS, the remnant can still gain mass by accretion of stars (for the growth rate of a massive BH hosted in a cluster of stars see for example the models of Marchant & Shapiro 1979). If the combination of star formation and stellar feedback of the cluster stars does not deplete of gas in the inner region, this surviving gas can supply an additional reservoir of material for BH growth. Both these processes can contribute to bring the BH mass to values as high as Equation (16) would suggest.

Clusters forming in unstable disks might have some degree of rotation. In this case, the gravo-gyro instability (Inagaki & Hachisu 1978) could contribute to accelerate the core collapse. The gravo-gyro instability is believed to occur in systems that
Figure 1. Cluster properties as a function of the spin parameter \(\lambda \equiv J/|E|^{1/2}/GM_r^{1/2}\) (i.e., the fraction of halo support given by rotation) for two critical densities for fragmentation, \(n_{\text{crit}} = 10^3 \, \text{cm}^{-3}\) (solid line), and \(n_{\text{crit}} = 10^4 \, \text{cm}^{-3}\) (dashed line). Curves are truncated when \(R_{\text{SF}}\) equals \(R_0\).

Clusters exhibit a radial gradient of the angular speed. In analogy with viscous transport, angular momentum is transferred outward. The core of the star cluster then contracts because of a deficit in the centrifugal force. Ernst et al. (2007) showed that the gravitogyro instability occurs in clusters with equal-mass stars, but in systems with two-mass components the effect of rotation seems negligible, as mass segregation and rotation compete in leading the evolution of the stellar cluster. For this reason, we neglect its contribution in this work.

2.6. Cluster Properties

Various cluster properties: size (Equation (11)), half-mass radius (Equation (12)), mass (Equation (13)) are shown in Figure 1 as a function of the spin parameter for two representative values of the critical density for fragmentation. We also present BH masses (Equation 15), where we have further imposed the condition \(t_{\text{cc}} < 3 \, \text{Myr}\).

As \(\lambda\) increases, less mass inflows within \(R_{\text{tr}}\), and \(R_{\text{SF}}\) are reached at smaller radii at a given \(n_{\text{crit}}\). Therefore, \(R_{\text{SF}}\) and \(M_{\text{cl}}\) decrease with increasing \(\lambda\), in contrast to the disk size \(R_0\). Note how at low critical densities (corresponding to high metallicities) only a very small fraction of systems can undergo core collapse, as on the one hand cluster formation is suppressed for large \(\lambda\) (as \(R_{\text{SF}} > R_0\)), on the other hand clusters are too massive and large at small \(\lambda\) for fulfilling the condition \(t_{\text{cc}} < 3 \, \text{Myr}\). Clusters undergoing core collapse have preferentially large spin parameters, within the region where clusters form. Figure 2 shows cluster sizes, masses, core collapse timescales (Equation (13)) and BH masses as a function of metallicity. As the metallicity increases, the critical density for fragmentation decreases, and cluster formation is eventually suppressed when \(n(R_{\text{tr}}, 0) > n_{\text{crit}, Z}\) (\(R_{\text{SF}} > R_0\)). At metallicities below \(10^{-5} \, Z_\odot\), fragmentation is impossible for the specific choice of the \(Z_{\text{crit}}-n_{\text{crit}, Z}\) relationship (intermediate PopIII mass case in Santoro & Shull (2006)).

Figure 2. Cluster properties as a function of metallicity for 3-spin parameters: \(\lambda = 0.018\) (solid line), \(\lambda = 0.01\) (dashed line), \(\lambda = 0.005\) (dotted line). Each curve begins at low metallicity when halos starts star formation and ends when \(R_{\text{SF}}\) equals \(R_0\); this happens at different critical metallicities for different spin parameters. Here, the \(Z_{\text{crit}}-n_{\text{crit}, Z}\) relationship follows the intermediate PopIII mass case in Santoro & Shull (2006).

Figure 3. Parameter space (virial temperature, spin parameter) for cluster and BH formation. Here, we select halos with \(T_{\text{vir}} > 10^5 \, \text{K}\) at \(z = 12\), and derive the disk and cluster properties, assuming a single critical density for fragmentation (top left: \(10^3 \, \text{cm}^{-3}\), top right: \(2 \times 10^3 \, \text{cm}^{-3}\), bottom left: \(2.5 \times 10^3 \, \text{cm}^{-3}\), bottom right: \(3 \times 10^3 \, \text{cm}^{-3}\); for the chosen \(Z_{\text{crit}}-n_{\text{crit}, Z}\) relationship, these densities correspond respectively to \(\log(Z/ Z_\odot) = -4.2; -4.3; -4.4; -4.5\). The black shaded area shows the range of temperatures and spin parameters where disks are Toomre unstable and the joint conditions, \(m_A > 0\) (Equations (6) and (9)), are fulfilled. The lighter shaded area selects the systems where \(R_{\text{SF}} < R_0\). The hatched area picks the subsample of clusters where \(t_{\text{cc}} < 3 \, \text{Myr}\), where VMSs and BH seeds can form.

(A color version of this figure is available in the online journal.)
The parameter space (viral temperature, spin parameter), where the multiple instabilities are efficient, is shown in Figure 3. Here, we select halos with $T_{\text{vir}} > 10^4$ K at $z = 12$ and derive the disk and cluster properties, assuming a single critical density for fragmentation (from $10^3$ cm$^{-3}$ to $10^4$ cm$^{-3}$). The higher the critical density for fragmentation (i.e., the lower the metallicity, see Section 2.3) the more compact are the clusters, and the shorter is $t_{\text{cc}}$. When $n_{\text{crit},Z} < 10^3$ cm$^{-3}$, no clusters can undergo core collapse in less than 3 Myr. When $n_{\text{crit},Z} > 10^3$ cm$^{-3}$, all forming clusters undergo core collapse in less than 3 Myr.

3. RESULTS

3.1. Summary of the General Procedure

We first summarize the procedure taken in order to determine the properties of the BH seed population. We calculate the mass of halos that at redshift $z$ correspond to virial temperatures $10^4$ K $\lesssim T_{\text{vir}} \lesssim 1.8 \times 10^4$ (Barkana & Loeb 2001), and we determine their frequency using a modified version of the Press & Schechter formalism (Sheth & Tormen 1999) in a WMAP5 cosmology (Dunkley et al. 2009). To each halo, we assign a value of the spin parameter, $\lambda$, extracted from the probability distribution found in the Millennium simulations (Bett et al. 2007):

$$P(\log \lambda) = A \left( \frac{\lambda}{\lambda_0} \right)^3 \exp \left[ -\zeta \left( \frac{\lambda}{\lambda_0} \right)^{3/\xi} \right],$$

where $\lambda_0 = 0.0043$ is the peak location, $\xi = 2.509$ and the normalization reads $A = 3 \ln 10 \xi^{\xi-1}/\Gamma(\xi)$, with $\Gamma$ being the gamma function. This set of assumptions allows us to calculate the initial disk properties, $\Sigma_0$, $R_0$, $Q_\text{c}$, $R_\text{b}$ and $m_\text{a}$.

We assign to each halo a metallicity, $Z$ by extrapolating at higher redshift the fit to the observational constraints of the MEH ($Z \propto 10^{-\beta Z}$), taking also into account the observed metallicity scatter. We then calculate $n_{\text{crit},Z}$ from a given $Z_{\text{crit}} - n_{\text{crit},Z}$ relation. If a protogalaxy has $Q < Q_\text{c}$, we determine the properties of the stellar cluster ($R_{\text{SF}}$, $M_\text{cl}$ and $R_b$). We then check if the cluster can develop runaway instability via Equation (14), and we select the systems where $t_{\text{cc}} < 3$ Myr. For these unstable clusters, we calculate the expected mass of the seed BH from Equation (15). In Table 1, we summarize all the different cases we describe in the following sections.

We now discuss the properties that nuclear clusters possess at birth and the resulting BH seed population. We start describing our reference model A in Section 3.1, while in Section 3.2, we discuss how our results depend on PopIII stars metallicity patterns and on the rate of metal enrichment of the universe.

3.2. Model A

The first halos reach the critical metallicity for fragmentation at redshift $\sim 14$. This is also when the first stellar clusters form. The first clusters have masses of the order of $10^5 M_\odot$ and $R_b \sim 0.5$–1 pc; in such compact clusters, core collapse starts early ($t_{\text{cc}} \sim 0.1$ Myr at $z = 14$). At later cosmic times, the average gas metallicity increases, so that the critical density for fragmentation decreases. A lower critical density implies that $R_{\text{SF}}$ increases (Equation 11), lengthening the core collapse timescale. Therefore, clusters form less concentrated and more massive, and their core collapse timescale continues to increase with decreasing redshift. This behavior is evident in Figure 4 for the mean quantities ($\langle M_\text{cl} \rangle$, $\langle R_b \rangle$, $\langle t_{\text{cc}} \rangle$, and $\langle M_{\text{BH}} \rangle$). The mean seed mass as a function of redshift is shown in Figure 4 (lower right panel). Unlike $\langle M_\text{cl} \rangle$ and $\langle R_b \rangle$, $\langle M_{\text{BH}} \rangle$ shows no redshift dependence as the increase in $M_\text{cl}$ and $t_{\text{cc}}$ compensates (see Equation 15) leading to a roughly constant $M_{\text{BH}}$.

In Figure 5, the entire cluster mass function (solid histogram) is compared to the mass function of systems able to form a BH seed (dashed histogram). The clusters that do not form BH seeds are the very last to form, when the metallicity of the universe is already significant. The BH mass function is also shown in Figure 5. The mass function is peaked at $\sim 1000 M_\odot$ with a long tail at low masses, and a steep drop at high masses. The distribution of core-collapse timescales is shown in Figure 6 (left panel), and in Figure 7 (bottom panel) we show the sharp redshift boundaries to MBH formation (and the influence of the metal enrichment history; see Section 3.3). Figure 8 shows the fraction, $f_{\text{BH}}$, of halos hosting a BH seed. Seeds start to form at $z \sim 14$ in coincidence with the first, very compact, stellar clusters. The typical core collapse timescale increases with cosmic time, $t_{\text{cc}} \sim 3$ Myr at $z \sim 9$, and by $z \sim 8$ $f_{\text{BH}}$ drops rapidly. From this point on even the most
No. 1, 2009 FORMATION OF MASSIVE BLACK HOLE SEEDS 309

Figure 5. Top: cluster mass function for model A integrated over all redshift. The shaded area corresponds to the subpopulation of clusters able to form BH seeds. Bottom: BH seed mass function integrated over all redshifts.

Figure 6. Top: mass function of seed BHs integrated over all redshifts. Bottom: mean BH mass as a function of redshift. Lines color and style as in Figure 6. (A color version of this figure is available in the online journal.)

Concentrated and least enriched disks are unable to create central stellar concentrations with $t_{cc} < 3$ Myr, and seed formation is completely suppressed. Figure 8 also shows the integrated comoving mass density of seeds $\rho_{seed}$. At $z \sim 10$, the mass density saturates at a value of $\sim 300 M_\odot$ Mpc$^{-3}$. This mass density should be considered a lower limit for the total black holes mass density $\rho_{BH}$ as we are completely neglecting BH growth after seeds formation. We will discuss the implications of this formation mechanism on the evolution of the supermassive black hole population in a forthcoming paper. We note, however, that this seed density is similar to that expected from PopII stellar remnants (roughly a factor of 3 larger, Volonteri et al. 2003), and we therefore expect that most observational constraints can be fulfilled at the same level.

3.3. Impact of the Uncertainties on Metal Enrichment onto the Seed Population

We now discuss how our results depend on our choice of parameters. We compute the seed population for BHs formed over cosmic time varying (1) the history of metal enrichment (models B and C) and (2) the ratio of the coolants that determine the $Z_{crit} - n_{crit}$ relationship (models D and E).

Mass functions of protogalactic nuclear clusters for all models defined in Table 1 are shown in Figure 6. Cluster mass functions, $M_{cl}$, span two orders of magnitude, between $10^4 M_\odot$ and $10^6 M_\odot$, with a peak around $\sim 10^5 M_\odot$. Figure 6 also shows the distribution of core collapse timescale of all clusters. The dashed vertical line at 3 Myr marks the limit for VMS formation. As discussed in the previous section, clusters satisfying the condition for the onset of runaway instability are clustered at small masses and radii: typical masses of runaway unstable clusters are around a few $10^5 M_\odot$, and typical radii are $\sim 1$ pc (see Equation (14)). This naturally points toward the very first clusters: the first systems that form are indeed those that more easily can give birth to BH seeds. Consequently, these are also the most metal-poor clusters, so that our picture is consistent with requiring that VMS can more easily grow in low metallicity environments.

As times goes on, both $\langle M_{cl} \rangle$ and $\langle R_{cl} \rangle$ grow: as a consequence $\langle t_{cc} \rangle$ increases. Even if clusters continue to form, BH seeds cannot be created any longer (cf. Figure 7). The mass functions of BH seeds are shown in Figure 7. The seed mass distributions show a characteristic shape with a peak at a few $10^3 M_\odot$ and a long tail at lower masses with very little redshift dependence. This general picture is valid for all models; we now discuss in turn the dependences on specific model parameters.
Figure 8. Top: fraction of halos hosting a BH seed as a function of redshift for models in Table 1. Bottom: comoving mass density of the seed BH for the same models. Colors and lines styles as in Figure 6.

3.3.1. Changing MEH

The MEH is one of the most uncertain parameters. We explore two extreme cases in models B (β = 0.18, thick (blue in the online version) dot-dashed curves in Figure 6 and onward) and C (β = 0.36 and Δ log(Z) = 0, thin (green in the online version) dot-dashed curves in Figure 6 and onward). The MEH determines, together with the chosen Zcrit−nt−Z relation, when seeds form. The duration of the BH seeds formation epoch is indeed given by a combination of the assumed metallicity spread, Δ log(Z), and of the slope of the Z(z). Once Z increases over Zcrit, fragmentation is activated in a more extended region of the disk, the inflow is reduced, and runaway instability cannot proceed efficiently. The seed formation epoch is, therefore, longer either if Z has larger Δ log(Z) or if metal enrichment is rather inefficient.

Model B has the most efficient metal enrichment, and BHs appear already at redshift 30. On the other hand, BH formation is also suppressed very early, at z ∼ 18. At this early cosmic epoch, very few halos were massive enough for efficient atomic line cooling, thus leading to a comoving seed mass density of only a few M⊙. As previously noted for model A ρseed does not necessary coincide with ρBH, as we are neglecting seeds growth. In model B seeds form at higher redshift, and mass can be built up for a longer period, likely increasing BH masses by accretion.

Model C has the same redshift dependence of the MEH as our reference model A, but we assume no scatter. A null Δ log(Z) (model I) produces a short burst of seeds very concentrated in time, as seed formation is allowed only for a very sharp range of Z. Clusters and BHs form in a burst at z = 11–12 when Z ∼ 10−5 (the minimum in the Zcrit−nt−Z relation). This burst is very efficient, with a high fraction of halos undergoing cluster and BH formation, and the resulting ρseed ∼ 300 M⊙ Mpc−3 is similar to our reference case.

3.3.2. Changing Zcrit−nt−Z

We explore the effect of different yield patterns for PopIII stars in our results in models D (blue long-dashed curves in Figure 6 and onward) and E (green long-dashed curves). The adopted Zcrit−nt−Z relation for model D, clusters begin to form earlier. With increasing cosmic time, Rst increases to Rv, thus precluding the formation of a seed. In model D, this happens already at z ∼ 12, while BH formation proceeds all the way to z = 5 in model E. Consequently, the seed mass density ρseed ranges from ∼100 to 300 M⊙ Mpc−3 at z = 5, with model D, which forms seed early on, having the lower seed density.

4. DISCUSSION

We described a model for BH seed formation as a result of multiple successive instabilities. On a large scale, gravitational torques in Toomre unstable primordial disks can pile up a significant amount of gas in the central region of high redshift halos (e.g., Lodato & Natarajan 2006). We focus on relatively massive halos at high redshift (T < 108 K, z ≥ 10) after the very first stars in the universe have completed their evolution. This set of assumptions ensures that (1) atomic hydrogen cooling can contribute to the gas cooling process, (2) a UV field from the first stars has heated the gas, precluding H2 cooling except in the highest density regions, and (3) the gas inside the halo has been mildly polluted by the first metals. The second condition implies that, at least at the low density, at which the collapse starts, H2 is dissociated and does not contribute to the cooling of the gas (see also Figures 9 and 10 in O’Shea & Norman2008). The third condition ensures that metals (and dust) in small quantities can cool down the gas efficiently only as the gas density reaches a critical threshold. Jointly, the two conditions ensure that cooling is driven at first by atomic hydrogen only.

At low metallicity (Z ∼ 10−5 Z⊙), fragmentation and low-mass star formation can occur only where the density is above a metallicity-dependent critical threshold (Zcrit−nt−Z⊙, Santoro & Shull 2006), corresponding to the central densest region of the protogalaxies. Within this limited region, where star formation takes place efficiently, very compact stellar clusters form. The typical stellar masses are of order 103 M⊙ and the typical half mass radii ∼1 pc. Eventually, a large fraction of these very dense clusters undergo core collapse before stars are able to complete stellar evolution and a VMS can form (e.g., Güürkan et al. 2004; Portegies Zwart et al. 2004).

Clusters unstable to runaway collisions are always the first, less massive (∼105 M⊙) ones. As the metallicity of the universe increases, the critical density for fragmentation decreases and stars start to form in the entire protogalactic disk so that (1) accretion of gas in the center is no longer efficient and (2) the core collapse timescale increases. As a result, less and less compact clusters form, and less of them are subject to rapid core collapse.

We computed the properties of the BH population for a set of models in dependence of various parameters. The epoch of seed formation is determined by the time at which gas in the center of the halo can start to fragment. This redshift depends on the metal enrichment history and on the exact shape of the Zcrit−nt−Z relation. If metal pollution is very efficient and the universe was enriched early, the BH formation epoch ends very early, when only a few halos were massive enough for efficient
atomic line cooling. The mass and number densities of BH seeds are consequently very low. The fraction of halos hosting a BH seed depends also on the fraction of the unstable disk (hence, the critical Toomre parameter, $Q_c$) and on the frequency of halos with $Z \sim Z_{\text{crit}, \min}$. Decreasing $Q_c$ has a twofold effect on the efficiency of BH seeding. First, as already noted by Lodato & Natarajan (2007), it decreases the number of bar-unstable disks as the Toomre criterion is satisfied for higher surface densities (requiring very small spin parameters). Additionally, the higher surface density for bar instabilities implies smaller unstable regions ($R_0$), but a larger $R_{\text{SF}}$ at a fixed particle density. Cluster formation is, therefore, truncated at $R_{\text{SF}}$, but a larger $R_{\text{SF}}$ still the population of seeds is comparable to the case of $R_0$.

Most of our assumptions have been quite conservative, but still the population of seeds is comparable to the case of Population III star remnants discussed, for instance in Volonteri et al. (2003). The fraction of high-redshift galaxies seeded with Population III star remnants discussed, for instance in Volonteri et al. (2003). The fraction of high-redshift galaxies seeded with Population III star remnants discussed, for instance in Volonteri et al. (2003). The fraction of high-redshift galaxies seeded with Population III star remnants discussed, for instance in Volonteri et al. (2003).

We are extremely grateful to Cole Miller for providing insightful comments on the manuscript. M.V. wishes to thank Marc Freitag for encouraging discussions during the very early stages of this work.

Appendix A

Disk and Cluster Structure

A.1. Disk Parameters

Disk parameters $\Sigma_0$, $R_0$, $R_{\text{tr}}$, and $m_a$ have been summarized in Section 2.1. We only insert here the complete expression for $R_{\text{tr}}$ with the $\gamma$ dependency

$$R_{\text{tr}} = \frac{2 - \gamma}{\gamma - 1} \frac{m_a M_h}{2 \pi \Sigma_0 R_0}. \quad (A1)$$

As stated in Section 2.1, the particle density of the inner disk can be written as

$$n(R, h) = n_0 R_0 \left( \frac{R_0}{R} \right)^\gamma \cos^{-2}(h/H(R)), \quad (A2)$$

where $H$ can be computed imposing hydrostatic equilibrium, and it is found out to be $H = \left( \frac{c_s^2}{2} \right) \left( \frac{\pi G}{2 m_a n_{\text{crit}, Z}} \right)$ (see e.g., Oh & Haiman 2002). To simplify the notation, we have defined $f(R) = R_0/R_\text{tr}(R_\text{tr}/R)^\gamma$. We relate $n_0$ and $\Sigma_0$ by imposing that the surface density calculated starting from Equation (A2), follows the profile described by Equation (3):

$$\Sigma_0 f(R) = \mu m_a n_0 f(R) \int_{-\infty}^{+\infty} \cos^{-2}(z/h) dz$$

$$= 2 \mu m_a n_0 f(R) H = \sqrt{2} c_s \left( \frac{n_0 \mu m_a f(R)}{\pi G} \right)^{1/2}. \quad (A3)$$

Resolving for $n_0$, we find

$$n_0 = \frac{\Sigma_0^2}{c_s^2} \frac{\pi G f(R)}{2 \mu m_a}. \quad (A4)$$

A.2. Star Formation Radius

Once the density profile of the disk is defined, the star formation radius can be determined by imposing $n(R = R_{\text{SF}}, h = 0) \equiv n_{\text{crit}, Z}$. We allow clusters to form only if star formation is induced inside $R_{\text{tr}}$ and not in the external part of the disk. The condition for this to happen is that $n_{\text{tr}} \equiv n(R = R_{\text{tr}}, h = 0) < n_{\text{crit}, Z}$. $R_{\text{SF}}$ is then computed by imposing

$$n_0 R_0 \left( \frac{R_0}{R_{\text{SF}}} \right)^\gamma = n_{\text{crit}, Z}. \quad (A5)$$

Inserting the expression for $n_0$

$$\frac{\pi G \Sigma_0^2}{2 \mu m_a c_s^2} \left( \frac{R_0}{R_{\text{tr}}} \right)^2 \left( \frac{R_{\text{tr}}}{R_{\text{SF}}} \right)^{2\gamma} = n_{\text{crit}, Z}. \quad (A6)$$

Resolving for $R_{\text{SF}}$

$$R_{\text{SF}} = R_{\text{tr}} \left[ \frac{\Sigma_0 R_0}{c_s R_{\text{tr}}} \sqrt{ \frac{\pi G}{2 \mu m_a n_{\text{crit}, Z}} } \right]^{1/\gamma}$$

$$= R_{\text{tr}}^{2/5} \left[ \frac{\Sigma_0 R_0}{c_s} \sqrt{ \frac{\pi G}{2 \mu m_a n_{\text{crit}, Z}} } \right]^{3/5}, \quad (A7)$$

where in the last expression, we have inserted $\gamma = 5/3$ explicitly.

As stated in Section 2.3, $n_{\text{crit}, Z}$ depends on the metallicity of the gas. The curves in Figure 10 of Santoro & Shull (2006) can be fitted by the expression

$$\log(Z/Z_\odot) = a \log^2(n_{\text{crit}, Z}) + b \log(n_{\text{crit}, Z}) + c, \quad (A8)$$

where the values of $a$, $b$, and $c$ have been calculated for the three curves and are reported in Table 2.

The minimal critical density for fragmentation can be computed from Equation (A8). Inserting the expression for the critical density into Equation (A7), one finds

$$\log(R_{\text{SF}}) = \frac{\gamma - 1}{\gamma} \log \left( \frac{2 - \gamma}{\gamma - 1} \frac{m_a M_h}{2 \pi \Sigma_0 R_0} \right) + 1 = \frac{1}{2} \log \left( \frac{\pi G}{2 \mu m_a} \right). \quad (A9)$$

Table 2

List of the Values of $a$, $b$, and $c$ for Fitting Different $Z_{\text{crit}} - n_{\text{crit}}$ Curves.

| $Z_{\text{crit}} - n_{\text{crit}}$ | 1      | 2      | 3      |
|-----------------------------|--------|--------|--------|
| a                           | 0.03517305 | 0.0317305 | 0.0317305 |
| b                           | -0.582132 | -0.572132 | -0.62132 |
| c                           | -4.1    | -2.75   | -1.1   |
where in the last expression, we have first expressed $n_{\text{crit}, Z}$ as a function of $Z$ and then we have inserted the redshift dependence of the metallicity $Z$. All constants except $\gamma$ have been included in $A$, $B$, $C$, $C'$, $D$, and $D'$:

\begin{align*}
A &= \log\left(\frac{m_a M_h}{2\pi \Sigma_0 R_0}\right) \\
B &= \frac{b}{4a} + \log\left(\frac{\Sigma_0 R_0}{c_z}\right) + \frac{1}{2} \log\left(\frac{\pi G}{2\mu m_a}\right) \\
C &= \frac{b}{4a} \sqrt{1 - \frac{4ac}{b^2}} \\
C' &= C' \sqrt{1 + D' \delta} \\
D &= \frac{b}{4a} - \frac{b^2}{b^2 - 4ac} \\
D' &= \frac{D'}{1 + D' \delta}.
\end{align*}

Finally, for $\gamma = 5/3$

$$
\log(R_{\text{SF}}) = k_1 \gamma + k_2 \sqrt{1 + k_3 \beta z},
$$

where all constants are collected in $k_1 \equiv \frac{2}{3}(A - 0.3) + 3/5B$, $k_2 \equiv 3/5C$ and $k_3 \equiv D$.

A.3. Black Hole Masses

Black hole masses depend on the stellar mass of the clusters and on the half-mass radius, via the core collapse timescale.

Cluster masses and timescales to core collapse can be estimated by coupling Equations (A7) and (A11) to Equation (12):

\begin{align*}
M_{cl} &= 2\pi \epsilon \int_0^{R_{\text{SF}}} \Sigma_0 R_0 \frac{\gamma - 1}{\gamma} R^{1-\gamma} dR \\
&= 2\pi \epsilon_{\text{SF}} \Sigma_0 R_0 \int_0^{R_{\text{SF}}/\gamma} \frac{R^{1-\gamma}}{2-\gamma} dR \\
&= 2\pi \epsilon_{\text{SF}} \Sigma_0 R_0 \left[\frac{2-\gamma}{2-\gamma - 1} m_a M_h \right] \int_0^{R_{\text{SF}}/\gamma} R^{1-\gamma} dR \\
&= \epsilon_{\text{SF}} \left[\frac{2\pi}{2-\gamma} \Sigma_0 R_0 \left(\frac{\pi G}{2\mu m_a n_{\text{crit}}}\right)^{1/5} \left(\frac{3}{2} m_a M_h\right)^{4/5}\right],
\end{align*}

where in the last line we assume $\gamma = 5/3$.

The core collapse timescale is defined in Equation (14). The half-mass radius can be simply expressed as $R_{\text{SF}} = 2^{1/(1-\gamma)} R_{\text{SF}}$. We use Equation (A11) to express the dependence of $t_{cc}$ on $M_{cl}$ as

\begin{align*}
t_{cc} &= \frac{\tau_0 R_{\text{SF}}^{1/2} M_{cl}^{1/2}}{\epsilon_{\text{SF}} \Sigma_0 R_0^{1/(1-\gamma)} m_a M_h} \\
&= \frac{\tau_0}{6\pi \epsilon_{\text{SF}} \Sigma_0 R_0^{1/(1-\gamma)}} \left(\frac{3}{2} m_a M_h\right)^{3/2} M_{cl}^{1/2} \times \left(\frac{2-\gamma}{2\pi \epsilon_{\text{SF}} \Sigma_0 R_0^{1/(1-\gamma)}}\right)^{1/2} \\
&= \tau_0 2^{1/4} \left[\frac{1}{6\pi \Sigma_0 R_0^{1/(1-\gamma)}}\right]^{3/2} \left(\frac{3}{2} m_a M_h\right)^{-3/2} M_{cl}^{5/2},
\end{align*}

where $\tau_0$ is the normalization of Equation (14), and its value in Myr, $M_\odot$ and $pc$ is 0.019. In the last Equation, we assume $\gamma = 5/3$.

$M_{\text{BH}}$ can now be found by inserting Equation (A14) into Equation (15).

\begin{align*}
M_{\text{BH}} &= m_a + 4 \cdot 10^3 f_c \ln \lambda M_{cl} \ln \left(\frac{g}{M_{cl}^{2/(1-\gamma)}}\right) \\
&\propto M_{cl} \ln \left(\frac{\tau}{M_{cl}^{2/(1-\gamma)}}\right),
\end{align*}

where $g$ is defined as

$$
g = \frac{\tau_{\text{MS}}}{\tau_0} 2^{1/(1-\gamma)} \left[\frac{2-\gamma}{2\pi \epsilon_{\text{SF}} \Sigma_0 R_0^{1/(1-\gamma)}}\right]^{1-\gamma} \left(\frac{3}{2} m_a M_h\right)^{-3/2} M_{cl}^{5/2}.
$$

From Equation (A15), we can argue that $M_{\text{BH}}$ first grows with increasing $M_{cl}$, it attains a maximum value and then it decreases, as more massive clusters have longer core collapse timescales with $t_{cc} \propto M_{cl}^2$. This last proportionality results from the dependence of $R_{\text{SF}}$ (and consequently $R_\text{h}$) on $M_{cl}$.

\section*{REFERENCES}

Abel, T., Bryan, G. L., & Norman, M. L. 2000, \textit{ApJ}, 540, 39

Barkana, R., & Loeb, A. 2001, \textit{Phys. Rep.}, 349, 125

Begelman, M. C., & Rees, M. J. 1978, \textit{MNRAS}, 185, 847

Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, \textit{MNRAS}, 370, 289

Bett, P., Eke, V., Frenk, C. S., Jenkins, A., Helly, J., & Navarro, J. 2007, \textit{MNRAS}, 376, 215

Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton, NJ: Princeton University Press), 747

Bromm, V., Coppi, P. S., & Larson, R. B. 1999, \textit{ApJ}, 527, L5

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, \textit{ApJ}, 564, 23

Bromm, V., Ferrara, A., Coppi, P. S., & Larson, R. B. 2001, \textit{MNRAS}, 328, 969

Bromm, V., & Loeb, A. 2003, \textit{Nature}, 425, 812

Ciardi, B., 2008, First Stars III, 990, 535

Ciardi, B., & Ferrara, A. 2005, \textit{Space Sci. Rev.}, 116, 625

Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2008, \textit{ApJ}, 672, 757

Clarke, C. J., & Bromm, V. 2003, \textit{MNRAS}, 343, 1224

Clarke, C. J., Harper-Clark, E., & Lodato, G. 2007, \textit{MNRAS}, 381, 1543

Di Matteo, T., Colberg, J., Springel, V., Hernquist, L., & Sijacki, D. 2008, \textit{ApJ}, 676, 33

Dunkley, J., et al. 2009, \textit{ApJS}, 180, 306

Ebisuzaki, T., et al. 2001, \textit{ApJ}, 562, L19

Eisenstein, D. J., & Loeb, A. 1995, \textit{ApJ}, 443, 11

Fan, X., et al. 2001, \textit{AJ}, 122, 2833

Freese, K., Bodenheimer, P., Spolyar, D., & Gondolo, P. 2008, \textit{ApJ}, 685, L101

Freitag, M., Gürkan, M. A., & Rasio, F. A. 2006a, \textit{MNRAS}, 368, 141

Freitag, M., Rasio, F. A., & Baumgardt, H. 2006b, \textit{MNRAS}, 368, 121

Fryer, C. L., Woosley, S. E., & Heger, A. 2001, \textit{ApJ}, 550, 372

Gammie, C. F. 1996, \textit{ApJ}, 462, 725

Gao, L., Yoshida, N., Abel, T., Frenk, C. S., Jenkins, A., & Springel, V. 2007, \textit{MNRAS}, 378, 449
